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

N° d'ordre : .....

UNIVERSITE DE M'SILA

FACULTE DES MATHEMATIQUES ET DE L'INFORMATIQUE

Département d'Informatique

MEMOIRE de fin d'étude

Présenté pour l'obtention du diplôme de MASTER

Domaine : Mathématiques et Informatique

Filière : Informatique

Spécialité :Réseaux

Par: BENAICHA Salah eddine

SUJET

Système de détection d'intrusion réseau basé sur les Algorithmes Génétiques

Soutenu publiquement le : 26/06/2013 devant le jury composé de :

Malek Khadija	**Université de M'sila Président**
SAOUDI Lalia	**Université de M'sila Rapporteur**
BOUHOUITA Salah Eddine	**Université de M'sila Examinateur**

Promotion : 2012 /2013

Table des matières

Chapitre 3:Approche basée Algorithme Génétique pour les NIDS

Introduction générale:

Contexte scientifique :

Les réseaux locaux et Internet se développent à un rythme incroyable au cours des dernières années. Alors que nous bénéficions de l'avantage que la nouvelle technologie nous a apporté, les systèmes informatiques sont exposés à des menaces croissantes de sécurité qui proviennent de l'externe ou de l'interne. Différentes mais complémentaires, de nouvelles technologies ont été développées et déployées pour protéger les systèmes informatiques des organisations contre les attaques réseau, telles que, les anti-virus, Les firewall, le chiffrement des messages, les protocoles sécurisés, protection par authentification…etc. Malgré les différents mécanismes de protection, il est presque impossible d'avoir un système totalement sécurisé. Par conséquent, la détection d'intrusion devient une technologie de plus en plus importante qui surveille le trafic réseau et identifie les intrusions sur le réseau tel que les comportements anormaux de réseau, l'accès non autorisé au réseau et les attaques malveillantes sur les systèmes informatiques.

Il existe deux grandes catégories de systèmes de détection d'intrusion (IDS): la détection par scénarios et la détection des anomalies. Les systèmes de détection par scénarios détectent les intrus avec des motifs connus, et les systèmes de détection d'anomalies identifiées les déviations de comportements normaux du réseau et d'alerte pour les attaques inconnues potentiels. Quelques IDS intègrent à la fois la détection par scénario et l'anomalie, ces systèmes sont les systèmes de détection de forme hybride. Les IDS peuvent également être classés en deux catégories en fonction de l'endroit où ils recherchent des intrusions. Un IDS hôte surveille les activités associées à un hôte particulier, et d'un IDS à base réseau écoute du trafic réseau.

Problématique :

Les IDS deviennent un dispositif essentiel pour l'obtention de la sécurité. Néanmoins, ils restent plus au moins complexes. Cette complexité est montrée dans les deux approches de détection (anomalies et par scénarios).

Plusieurs outils du marché utilisent l'approche par scénario, on peut citer à titre d'exemple : Snort, RealSecure, NFR, Dragon, cet approche ne détecte que les attaques connues par le motif (Signatures d'attaques) et le temps de recherche reste très lent (multiplicités des

signatures). Donc afin de remédier à ces problèmes, les techniques de l'intelligence artificielle sont apparues.

Méthode :

L'approche par scénario utilise plusieurs techniques que l'on peut regrouper dans trois classes:

- les approches à base de règle, c'est-à-dire les systèmes experts,
- les approches basées sur la signature,
- les algorithmes génétiques GA.

Dans ce mémoire, nous présentons une approche GA fondée sur la détection d'intrusions réseau. L'approche GA est choisi en raison de certaines de ses propriétés intéressantes, par exemple, aucune information est nécessaire de trouver une solution globale optimale ou sous-optimale, les capacités d'auto-apprentissage,définition de nouvelle règles de détection...etc. En utilisant le GA pour la détection d'intrusion réseau s'est avérée une approche efficace.

Les algorithmes génétiques sont utilisés afin d'optimiser la recherche de scénarios d'attaques dans les fichiers d'audit grâce à son bon équilibre exploration/exploitation, cela permet d'obtenir en un temps de traitement raisonnable, le sous-ensemble des attaques potentiellement présentes dans les traces d'audit.

Dans ce travail, nous déposons en œuvre un IDS basé sur l'approche présentée. Ce système est expérimenté en utilisant les données DARPA (NSL-KDD), qui est devenu le standard de test pour les systèmes de détection d'intrusion.

Objectifs :

L'objectif de ce mémoire est l'implémentation d'une approche GA dans les systèmes de détection d'intrusions basée réseau, contient deux modules où chacun fonctionne à une étape différente. Dans l'étape d'entrainement, un ensemble de règles sont produites à partir des données d'audit en utilisant GA dans un environnement différé. Dans l'étape de détection d'intrusions en ligne, les règles produites sont employées pour classifier les connexions réseau entrantes en temps réel. Mais le but principal est l'optimisation de nombre de signatures dans les données d'audit afin de minimiser le temps de recherche, et l'augmentation de taux de détection d'attaques.

Travaux antérieurs:

Cette section présente brièvement quelques-unes des applications des techniques de l'intelligence artificielle pour la détection d'intrusions, afin de comparer leur travail avec notre travail.

GA et GP ont été utilisés pour la détection d'intrusion réseau de différentes manières. Certaines approches utilisent directement GA pour dériver les règles de classification, tandis que d'autres utilisent des méthodes différentes IA pour l'acquisition de règles, où les GA sont utilisés pour sélectionner les fonctions adéquate ou de déterminer les paramètres optimaux de certaines fonctions [2][3].

L'effort au début de l'utilisation du GA pour la détection d'intrusion peut être daté à 1995.

Bridges et al. [2] ont développé une méthode qui intègre des techniques d'exploration de données floues et des algorithmes génétiques pour détecter les mauvais usages du réseau et les anomalies. Dans cette approche, un GA est utilisé pour trouver les paramètres optimaux des fonctions floues ainsi que de sélectionner des fonctions réseau les plus pertinents.

Lu et al. [4] présente une approche qui utilise GP permet de dériver directement un ensemble de règles de classification à partir des données du réseau historique. L'approche utilise le cadre support confidence dans la fonction d'adaptation et elle est capable de détecter de façon générale, ou classer précisément les intrusions réseau. Cependant, l'utilisation de GP rend plus difficile la mise en œuvre et plus de données ou de temps sont nécessaires pour rendre le système plus adapté.

WeiLi[5] a écrit une proposition pour l'utilisation de GA dans un NIDS et Ren Hui Gong[6] a suivi avec sa mise en œuvre. Li a établi les bases pour la création d'un système utilisant des algorithmes génétiques qui analyse des ensembles de données DARPA, et Gong a proposé une autre implémentation en utilisant ECJ (A Java-based Evolutionary Computation Research System). Gonga fourni un pseudo code et des diagrammes de classes. Li a proposé d'utiliser l'ensemble des données DARPA [WEB-1] du MIT Lincoln Laboratory de l'entrainement et de test. Dans les deux casla proposition de Li et l'approche de Gong, se base sur une fonction d'adaptation et un type de chromosome pour les algorithmes génétiques.

Structure du mémoire :

Ce mémoire est articulé autour de quatre chapitres :

Le premier chapitre propose un état de l'art sur la sécurité informatique, les attaques ainsi que les mécanismes de défense.

Le deuxième chapitre de ce mémoire présente les systèmes de détection d'intrusions, différentes approches, différentes classifications, domaines impliqués dans la détection d'intrusions, et quelques solutions IDS du marché.

Le troisième chapitre présente les algorithmes génétiques, les données de DARPA, et enfin l'application de l'approche GA dans les IDS.

Le dernier chapitre est la phase d'implémentation et d'expérimentation, il présente l'environnement d'implémentation, architecture de l'algorithme de détection, les données de test et d'entrainement choisit et enfin l'interprétation et discussion des résultats.

Nous conclurons ce mémoire par une conclusion générale et des perspectives.

Conclusion générale:

Le nombre d'attaques contre les entreprises ne cessent d'augmenter ce qui peut entrainer des pertes conséquentes, ainsi, le besoin des entreprises en sécurité informatique devient de plus en plus important.

Plusieurs politiques et outils ont été développés pour fournir des mécanismes de défenseefficaces.

Aucun système n'est parfait. Les pirates informatiques sont toujours en avance pour exploiter les failles de sécurité. Pour cela des mécanismes de défense dynamiques utilisés tels que les systèmes de détection d'intrusions (IDS) ont été introduit.Le but des IDS est de surveiller l'activité d'un réseau ou d'un hôte donné, afin de détecter toute tentative d'intrusions.

La performance d'un IDS, est mesurée en termesde taux de détection et de faux positifs.

Dans notre travail de recherche nous nous sommes intéressés aux algorithmes génétiques, qui sont des algorithmes d'optimisation, afin de garantir les objectifs précédents et améliorer le temps de recherche dans les données d'audit utilisées.

Des résultats satisfaisants sont produits en termes de taux de détectiontrès élevé (99%), renforcée par un faible taux de faux positifs(3%). Les résultats sont obtenus après plusieurs améliorationsde l'approche utilisée, tels que le choix de la population initiale par chaque type d'attaques.

Les données de détection d'intrusions DARPA restent toujours le meilleur corpus pour faire la phase d'entrainement et de test, maisbeaucoup de nouveaux protocoles ont été développées, et beaucoup type d'attaques ont été produits, d'où la nécessité d'enrichir cet ensemble de données.

Perspective :

En perspective, il serait intéressant d'améliorer les performances de notre approche à travers l'approche comportementalepour construire un IDS hybride, capable de détecter des nouveaux attaques précéder par l'approche par signature, et le choix des attributs spécifiques pour chaque type d'attaque.

Bibliographie

[1] M. Moradi and M. Zulkernine, "A Neural Network Based System for Intrusion Detection and Classification of Attacks," IEEE International Conference on Advances in Intelligent Systems, Luxembourg, Nov-2004.

[2] S. Bridges and R. Vaughn, "Fuzzy Data Mining And Genetic Algorithms Applied To Intrusion Detection," Proceedings of 12th Annual Canadian Information Technology Security Symposium, 2000.

[3] J. Gomez and D. Dasgupta, "Evolving Fuzzy Classifiers for Intrusion Detection," Proceedings of the IEEE, 2002.

[4] W. Lu and I. Traore, "Detecting New Forms of Network Intrusion Using Genetic Programming," University of victoria, Canada, 2004.

[5] W. Li, "A Genetic Algorithm Approach to Network Intrusion Detection," SANS Institute, USA, 2004.

[6] R. H. Gong, M. Zulkernine, and P. Abolmaesumi, "A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection," IEEE, University Kingston, Ontario, Canada, 2005.

[7] B. Nicolas and K. Marion, "NT Réseaux : IDS et IPS," 2003,2004.

[8] E. Thierry, *LES IDS : Les systèmes de détection des intrusions informatiques*, Nouvelle (12 février 2004). Paris: Dunod / 01 Informatique, 2004.

[9] C. Duret and N. Gaillard, *Les attaques Internet et les moyens de s'en protéger*. 2002.

[10] K. Phung, "La sécurité dans les réseaux hauts débit," Institut de la francophonie pour l'informatique (IFI), Mai 2005.

[11] U. Brigitte, *Cisco et la sécurité*. 2004.

[12] L. Wenke, J. Salvatore, and W. Kui, "A data mining framework for building intrusion detection models," IEEE Symposium on Security and Privacy, 1999.

[13] L. Wenke, A. Rahul, K. Kam, B. Sunil, H. Pragneshkumar, T. Thuan, and J. Salvatore, "A data mining and CIDF based approach for detecting novel and distributed intrusions," Recent Advances in Intrusion Detection, 2000.

[14] J. Cannady, *Artificial neural networks for misuse detection*. 1998.

[15] M. Ludovic and A. Véronique, "Détection d'intrusion dans un système informatique : méthodes et outils," TSI, 1996.

[16] C. Jabou, M. Schillings, and A. Hantach, "TER Detection d'anomalies sur le réseau," Université Paris Descartes, 2008.

[17] Wenke Lee, Salvatore J. Stolfo, and Kui W. Mok. Algorithms for mining system audit data.

[18] T. Mahbod, B. Ebrahim, L. Wei, and G. Ali, "A Detailed Analysis of the KDD CUP 99 Data Set," *IEEE*, 2009.

[19] S. Gunadiz, "Algorithmes d'intelligence artificielle pour la classification d'attaques réseaux à partir de données TCP," Université UMBB, 2011.

[20] H. Mohammad Sazzadul, M. Abdul, and B. Abu Naser, "AN IMPLEMENTATION OF INTRUSION DETECTION SYSTEM USING GENETIC ALGORITHM," *International Journal of Network Security & Its Applications (IJNSA)*, Mar-2012.

[21] N. Ben Amor, S. Benferhat, and Z. Elouedi, "Réseaux bayésiens naïfs et arbres de décision dans les systèmes de détection d'intrusions," *RSTI-TSI*, 2006.

[22] A. Souquet and F.-G. Radet, *ALGORITHMES GENETIQUES*. 2004.

[23] E. G. David, "Algorithmes Génétiques," France, 1994.

[24] D. Whitley, "Genetic Algorithm Tutorial," Departement of Computer Science,Colorado state University, Mar-1993.

[25] B. E. Lavender, "IMPLEMENTATION OF GENETIC ALGORITHMS INTO A NETWORK INTRUSION DETECTION SYSTEM (netGA), AND INTEGRATION INTO nProbe," CALIFORNIA STATE UNIVERSITY, SACRAMENTO, 2010.

[26] K. Vivek, M. Sonali, and V. Swati, "Intrusion Detection System using Genetic Algorithm and Data Mining: An Overview," International Journal of Computer Science and Informatics ISSN, 2012.

[27] M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," *Submitted to Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA)*, 2009.

Les sites web:

[WEB-1] MIT Lincoln Laboratory, DARPA datasets, MIT,USA, http://www.ll.mit.edu/IST/ideval/data/data_index.html (consulté le 25/03/2013).

[WEB-2]: URL http://www.commentcamarche.net , consulté le : 28/02/2013

[WEB-3]http://www.enterasys.com/products/ids/, consulté le: 12/05/2013

[WEB-4] : DARPA

URL http://www.ll.mit.edu/mission/communications/cyber/CSTcorpora/ideval/data/ (consulté le : 19/01/2013)

[WEB-5]: UCI KDD archive, KDD 1999 data set

URL http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (consulté le: 19/01/2013)

[WEB-6]: KDD-NSL

URL http://nsl.cs.unb.ca/NSL-KDD/ (consulté le: 10/04/2013)

[WEB-7]: URL http://WWW.pmsi.fr/Skanda2.htm.(Consulté le: 16/03/2013)

ملخص:

نظام كشف التسلل هو نظام للكشف عن الأنشطة غير الطبيعية أو المشبوهة الموجهة ضد نظام المعلومات. يوجد نهجين لكشف التسلل: نهج السيناريو ونهج السلوكية. كل لديه نقاط قوة ونقاط ضعف والتي تتمل في الأخطاء الايجابية والأخطاء السلبية. هدفنا هو انجاز نموذج فعّال باستخدام الخوارزميات الجينية في نظام كشف التسلل الشبكي, الخوارزميات الجينية استعملت من أجل تحسين البحث لسيناريوهات الهجمات في معطيات التدقيق. هذه المقاربة, من خلال موازنته الاستكشاف/الاستغلال, يسمح بالحصول على مدة معالجة معقولة, للمجموعة الثانوية للهجمات الموجودة في معطيات التدقيق مع نسبة الكشف العالي معززة بمستوى منخفض من النواحي الايجابية الخاطئة.

الكلمات المفتاحية : الخوارزمية الجينية, نظام كشف التسلل,نهج السيناريو,نظام كشف التسلل الشبكي,NSL-KDD, الهجوم, الأخطاء الاجابية.

Abstract:

An intrusion detection system (IDS) is a system making it possible to detect abnormal or suspect activities directed against an information system. Two approaches coexist in the intrusions detection: approach by signatures (misuse detection) and behavioral approach (anomaly detection). Each of the two present strong points, but also, weaknesses which are the false-positives and false-negative.

Our objective is to build an effective NIDS analysis engine, based on the genetic algorithms, the genetic algorithms are used in order to optimize the search for attacks models in the audit files. This approach, dueto its good balance exploration/exploitation, makes it possible to obtain in a reasonable processing time, the subset of the attacks potentially present in the traces of audit with a very high detection rate reinforced by a low level of false-positives.

Keywords:genetic algorithm, IDS, approach by signatures, NIDS, NSL-KDD, Attack, false-positives

Résumé :

Un système de détection d'intrusions (IDS) est un système permettant de détecter des activités anormales ou suspectes dirigées contre un système d'information. Deux approches coexistent dans la détection d'intrusions: l'approche par signatures (*misusedetection*) et l'approche comportementale (*anomalydetection*). Chacune des deux présentes des points forts, mais aussi des faiblesses qui sont les faux positifs et les faux négatifs.

Notre objectif est de construire une approche basée sur les algorithmes génétiquespour les NIDS, les algorithmes génétiques sont utilisés afin d'optimiser la recherche de modèles d'attaques dans les fichiers d'audit. Cette approche, grâce à son bon équilibre exploration/exploitation, permet d'obtenir en un temps de traitement raisonnable, le sous ensemble des attaques potentiellement présentes dans les traces d'audit avec un taux de détection très élevée renforcée par un faible taux de faux positifs.

Mots-clefs :algorithme génétique, IDS, approche par signatures, NIDS, NSL-KDD, attaque, faux positifs.