République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique

UNIVERSITE Dr. TAHAR MOULAY SAIDA
FACULTE : TECHNOLOGIE

DEPARTEMENT : INFORMATIQUE

MEMOIRE DE MASTER

OPTION :
SIC

Thème

Conception Et Implémentation D'un Systeme De Détection D'intrusion Par Les Abeilles Sociales

Présenté par :

-Khader Mohamed EL Amine
-Bourbig Mokhtar

Encadré par :

-Dr Lokbani Ahmed Chaouki

-Dr Boudia Mohamed Amine

Promotion : Juin & 2018

Remerciement

En préambule à ce mémoire nous remerciant ALLAH qui nous aide et nous donne la patience et le courage durant ces langues années d'étude.
Nous souhaitant adresser nos remerciements les plus sincères aux personnes qui nous ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire ainsi qu'à la réussite de cette formidable année universitaire.
Ces remerciements vont tout d'abord au corps professoral et administratif de la Faculté (Dr.MOULAY TAHAR) des technologie, pour la richesse et la qualité de leur enseignement et qui déploient de grands efforts pour assurer à leurs étudiants une formation actualisée.

Nous tenant à remercier sincèrement Monsieurs, (Dr **Ahmed Chaouki LOKBANI**) et (Dr **Amine BOUDIA**.), qui, en tant que Directeurs de mémoire, se sont toujours montrés à l'écoute et très disponible tout au long de la réalisation de ce mémoire, ainsi pour l'inspiration, l'aide et le temps qu'ils ont bien voulu nous consacrer et sans qui ce mémoire n'aurait jamais vu le jour.
On n'oublie pas nos parents pour leur contribution, leur soutien et leur patience.

Enfin, nous adressons nos plus sincères remerciements à tous nos proches et amis, qui nous ont toujours encouragée au cours de la réalisation de ce mémoire.
Merci à tous et à toutes.

DEDICACE

A mon Père Larbi

A ma Chère Mère

A ma Femme

A Mon Fils Taha Yacine

Dont le mérite, les sacrifices et les qualités humaines

m'ont permis de vivre ce jour.

A mon Frères

Tayeb et Mustpaha

A tous les gens m'aiment

(Mohamed EL Amine)

DEDICACE

A mon Père Aissa

A ma très chère maman

Qu'ils trouvent en moi la source de leur fierté

A qui je dois tout

A mon frère Sid Ahmed

A qui je souhaite un avenir radieux plein de réussite

A mes Amis

A tous ceux qui me sont chers

(Mokhtar)

Résumé :

Un système de détection d'intrusion (IDS) est un mécanisme écoutant le trafic réseau de manière furtIIIe afin de repérer des actvités anormales ou suspectes et permettant aussi d'avoir une action de prévention sur les risques d'intrusions. Les méthodes de détection d'intrusions reposent essentiellement sur deux approches : l'approche comportementale et l'approche par signatures. Chacune des deux présente des points forts, mais aussi des faiblesses qui sont les faux positifs et les faux négatifs. Notre objectif et de gérer la sécurité d'une application detection intrusion par bio-inspiré les abeilles et l'algorithme naïve bayse.

Mots clefs : Sécurité Informatique, système de détection d'intrusion, Data Mining, la protection des abeilles sociales, Kdd Cup'99.

Abstract:

An intrusion detection system (IDS) is a mechanism for listening to the network traffic stealth to identify anomalous or suspicious actvities and to also have a prevention of the risk of intrusions. The methods of intrusion detection based on two main approaches: the behavioral approach and the approach signatures. Each of the two has strengths, but also weaknesses that are false positves and false negatives. Our goal is to manage the security of a web application detection using the bio-inspired bees and optimized bees and naïve bayse.

Keywords : Computer Security, Intrusion detection system, Data mining, order Recognition, social bees protection, Kddcup'99.

Sommaire

Liste des figures

Liste des Tableux

Chapitre 3

Introduction générale

Les réseaux informatiques sont devenus beaucoup plus important qu'ils en aient il y a quelques années. De nos jours les entreprises dès leur création n'hésitent pas à mettre en place un réseau informatique pour faciliter la gestion de leur infrastructure, c'est pour cela que la sécurité de ces réseaux constitue un enjeu crucial. La sécurité d'un système informatique repose en premier lieu sur la mise en place d'une politique de sécurité.

Une fois la politique de sécurité définie, il convient de la mettre en œuvre au sein du système informatique. Deux approches non exclusives sont envisageables: la prévention des attaques et leur détection. La première approche, en appliquant un contrôle a priori sur les actions effectuées au sein du système, s'assure que les utilisateurs ne pourront pas violer la politique. Cette approche évite que le système ne se trouve dans un état corrompu, nécessitant une analyse et une correction.

De ce fait, des mécanismes de prévention sont présents sur les systèmes informatiques, il s'agit souvent de contrôle d'accès. Cependant, de tels mécanismes possèdent leurs propres limitations, qui peuvent porter sur des aspects théoriques des modèles sous-jacents ou sur leur implémentation. Ces limitations justifient le recours à des mécanismes de détection d'intrusions (IDS).

Afin de qualifier un IDS, on s'intéresse à sa fiabilité, qui est sa capacité à émettre une alerte pour toute violation de la politique de sécurité, et à sa pertinence, qui est sa capacité à n'émettre une alerte qu'en cas de violation de la politique de sécurité.

Un IDS est parfaitement fiable en absence de faux négatif ; il est parfaitement pertinent en l'absence de faux positif.

Notre travail s'articule autour de ce domaine dont il consiste à sécuriser un réseau informatique à l'aide d'un système de détection d'intrusions.Le premier chapitre est un chapitre descriptif pour la sécurité des réseaux, sur lequel on va definer

les menaces, les logiciels malveillants et une politique de sécurité ainsi les principaux mécanismes de sécurité.

Le second chapitre sera consacré à présenter une architecture globale d'un IDS

la définition et le mode de fonctionnement de ce dernier.

Ainsi la classification des IDS et enfin la méthode de détection d'une intrusion.Le dernier chapitre on parle de l'implémentation de notre modèle ,d'abord on définit notre corpus de donnée intitulé KDD 1999, ensuite on définit notre modèle et on applique notre modèle

sur le corpus et on discute sur les résultats et les tests et a la fin on conclure avec la définition du fonctionnement bioinspiré des abeilles sociales et le modèle artificiel.

Chapitre I :

Introduction à la sécurité informatique

Dans ce chapitre :

1. La sécurité informatique.
2. Services et mécanismes de sécurité.
3. Terminologie de la sécurité informatique.
4. Quelques possibilités en matière de sécurité réseaux.
5. Conclusion.

1. La sécurité informatique

De nos jours, l'information dans l'entreprise est d'une importance capitale, ce qui rend sa protection une fonction primordiale. L'institut britannique de standardisation **[08]** définit la sécurité informatique par la préservation des trois propriétés suivantes:

- ✓ **La confidentialité** : assurer que l'information est accessible uniquement aux utilisateurs autorisés (empêcher la divulgation non autorisée de données).
- ✓ **L'intégrité** : assurer l'exactitude et la complétude de l'information (empêcher la modification non autorisée de données).
- ✓ **La disponibilité**: assurer l'accès à l'information aux utilisateurs autorisés (empêcher l'utilisation non autorisée de ressources informatiques d'une façon générale).

UBEZEN **[01]** a rajouté une autre propriété concernant l'utilisation de l'information :

- ✓ **La responsabilité** : assurer qu'une action peut être liée sans doute à son initiateur.

De façon générale, la sécurité informatique peut être définie par l'ensemble des moyens matériels, logiciels et humains mis en œuvre pour minimiser les vulnérabilités d'un système d'information, et le protéger contre les menaces accidentelles ou intentionnelles, provenant de l'intérieur ou de l'extérieur de l'entreprise.

La vulnérabilité est souvent définie comme une faille dans le système permettant la violation de la politique de sécurité. Une bonne politique de sécurité représente l'ensemble de règles qui fixent les actions autorisées et interdites dans le domaine de sécurité. Cette politique doit preserver les propriétés de sécurité citées ci-dessus.

Du point de vue organisationnel, nous pouvons découper le domaine de la sécurité informatique de la façon suivante **[03]:**

- ➢ La sécurité logicielle gère la sécurité au niveau logiciel du système d'information (par exemple : l'intégration des protections logicielles comme l'antivirus).
- ➢ La sécurité du personnel comprend la formation et la sensibilisation des personnes utilisant ou travaillant avec le système d'information.
- ➢ La sécurité physique regroupe : la politique d'accès aux bâtiments, la politique d'accès aux matériels informatiques, et les règles de sécurité pour la protection des équipements réseaux.
- ➢ La sécurité procédurale définit les procédures et les règles d'utilisation du système d'information.
- ➢ La sécurité réseau s'occupe de : l'architecture physique et logique du réseau, la politique d'accès aux différents services, la gestion des flux d'informations sur les réseaux, et surtout les points de contrôle et de surveillance du réseau.
- ➢ La veille technologique souvent oubliée permet d'évaluer la sécurité au cours du temps afin de maintenir un niveau suffisant de protection du système d'information.

2. Services et mécanismes de sécurité

Le célèbre modèle en couches OSI (Open Systems Interconnection) est décrit dans la norme multi parties ISO 7498, intitulée “Interconnexion des Systèmes Ouverts - Modèle de reference de base”. Dans la partie intitulée “Architecture de sécurité” **[01]**, se trouvent des definitions et des concepts de base de la sécurité.

De façon générale, les mécanismes de sécurité permettent de mettre en œuvre des services de sécurité. Ces services peuvent être la confidentialité (des données ou du flux de données), l’authentification (d’une entité ou de l’origine des données), le contrôle d’accès, l’intégrité ou encore la non répudiation (avec preuve de l’origine ou preuve de la remise).
Les mécanismes peuvent être le chiffrement, l’authentification, l’intégrité, la signature numérique et d’autres encore.

Les services de sécurité

Les principaux besoins de sécurité que peut avoir l’émetteur d’un message sont les suivants **:**

- E1 : le message ne doit être connu que de son destinataire,
- E2 : le message doit parvenir au bon destinataire,
- E3 : le message reçu doit être identique au message émis,
- E4 : le destinataire ne doit pas pouvoir nier avoir reçu le message**.** Et les besoins du destinataire peuvent être :
- D1 : le message ne doit être connu que de lui (et de l’émetteur),
- D2 : l’émetteur du message doit être connu avec certitude,
- D3 : le message reçu doit être identique au message émis,
- D4 : l’émetteur ne doit pas pouvoir nier avoir émis le message.

Les besoins E1 et D1 sont identiques. Ils sont satisfaits par la mise en œuvre d’un service de confidentialité, définie dans la norme 7498-2 comme la “propriété d’une information qui n’est ni disponible, ni divulguée aux personnes, entités ou processus non autorisés”.

Les besoins E2 et D2 sont symétriques. Chaque entité doit s’assurer de l’identité de l’autre, ce qui implique de mettre en œuvre un service d’authentification, défini comme la “confirmation qu’une entité homologue d’une association est bien l’entité déclarée”, et même, dans le cas du destinataire, un service d’authentification de l’origine des données, ou “confirmation que la source des données est telle que déclarée”.
Les besoins E3 et D3 sont identiques. L’égalité entre le message émis et le message transmis est assurée par un service d’intégrité (des données).

qui est la "propriété assurant que des données n'ont pas été modifiées ou détruites de façon non autorisée".

Enfin, les besoins E4 et D4 sont symétriques. Le service correspondant est la non repudiation qui empêche la répudiation, c'est-à-dire "le fait, pour une des entités impliquées dans la communication de nier avoir participé aux échanges, totalement ou en partie". Dans un cas il s'agira de non répudiation avec preuve de l'origine, dans l'autre de non répudiation avec preuve de la remise.

A tout cela s'ajoute le service de contrôle d'accès, ou "précaution prise contre l'utilisation non autorisée d'une ressource", et qui peut s'appliquer à divers types d'accès (utilisation de ressources de communication, lecture, écriture ou suppression d'une ressource d'information, exécution d'une ressource de traitement).

Parfois, la simple observation du flux de données fournit de l'information à un ennemi.

C'est ce qu'on appelle l'analyse de trafic, qui permet de détecter la présence, l'absence, la quantité, la direction, ou la fréquence de telles ou telles données, qu'elles soient compréhensibles ou non.

On peut alors renforcer la confidentialité des données en assurant également la confidentialité du flux de données, c'est-à-dire un "service de confidentialité fournissant une protection contre l'analyse de trafic". La confidentialité, tout comme l'intégrité, peut être sélective par champ, c'est-à-dire ne s'appliquer qu'à une partie des champs contenus dans le message transmis.

Les mécanismes de sécurité

Les différents services de sécurité décrits précédemment sont mis en œuvre grace des mécanismes, dont la plupart sont de nature cryptographique. La cryptographie est la "discipline incluant les principes, moyens et méthodes de transformation des données, dans le but de cacher leur contenu, d'empêcher que leur modification passe inaperçue et/ou d'empêcher leur utilisation non autorisée" **[08].**

Afin d'assurer la confidentialité des données et/ou du flux de données, on fait appel à un mécanisme de chiffrement, qui est la "transformation cryptographique de données produisant un cryptogramme", unité de données dont "le contenu sémantique n'est pas compréhensible".

L'opération inverse du chiffrement est le déchiffrement.

Lorsqu'il est effectué de bout en bout, le chiffrement a lieu "à l'intérieur ou au niveau du système extrémité source, le déchiffrement correspondant ne se produisant qu'à l'intérieur, ou au niveau du système extrémité de destination".

S'il n'est effectué qu'à chaque liaison du système (dans quel cas les données sont en clair à l'intérieur des entités relais), il s'agit de chiffrement de liaison.

La confidentialité du flux de données exige en outre un mécanisme de bourrage de trafic, consistant à produire des "instances de communications parasites, des unités de données parasites

des données parasites dans des unités de données". Cet échange continu de données, transportant ou non de l'information, permet d'éviter qu'un tiers ne sache quand deux entités sont entrées en communication.

Le service d'authentification (d'entité homologue) est fourni par un mécanisme d'échange d'authentification, "destiné à garantir l'identité d'une entité par échange d'informations". (Typiquement, cet échange est constitué d'un nombre choisi au hasard envoyé par l'entité qui souhaite authentifier l'autre, et d'une réponse de cette dernière obtenue en appliquant un mécanisme cryptographique à ce nombre et à un secret connu d'elle seule).

L'authentification de l'origine des données peut être obtenue grâce à un mécanisme de signature numérique. Il s'agit de "données ajoutées à une unité de données permettant à un destinataire de prouver la source et l'intégrité de l'unité de données et protégeant contre la contrefaçon (par le destinataire, par exemple)". Le même terme désigne aussi la transformation cryptographique qui produit ces données. Pour produire une signature, il faut une information privée, c'est-à-dire connue du seul signataire. Pour la vérifier, il suffit d'une information publique. Il doit cependant être matériellement impossible de déduire l'information privée de l'information publique correspondante.

L'intégrité des données est assurée par un mécanisme du même nom. Un tel mécanisme peut consister à produire une valeur de contrôle cryptographique, à partir des données à protéger et d'un secret partagé par les entités en communication. Dans ce cas, la vérification par le destinataire consiste à recalculer cette valeur et à la comparer avec celle reçue. Si elles sont égales, il y a présomption d'intégrité. Mais on peut également utiliser un mécanisme de signature numérique qui, en plus de l'origine des données, garantit également leur intégrité. Par ailleurs, il peut être nécessaire de recourir en outre à des mécanismes visant à éviter le rejoue (répétition frauduleuse de tout ou partie des données), tels que la numérotation, l'horodatage ou le chaînage cryptographique des données.

Pour obtenir le non répudiation avec preuve de l'origine, on peut utiliser un mécanisme de signature numérique. En effet, la caractéristique essentielle de ce mécanisme est que la signature ne peut être produite qu'en utilisant l'information privée du signataire. On peut donc, en vérifiant la signature, prouver à tout moment à une tierce partie (par exemple un juge ou un arbitre) que seul le détenteur unique de l'information privée peut avoir produit la signature. Il est cependant possible d'utiliser aussi des mécanismes de chiffrement ou d'intégrité.

La non répudiation avec preuve de la remise peut aussi reposer sur un mécanisme de signature, produite cette fois par le destinataire du message. Les deux services de non répudiation, et plus particulièrement le second, peuvent aussi faire appel à un mécanisme de notariat.

Ce mécanisme met en jeu une tierce partie, appelée notaire, qui garantit certaines propriétés relatives à des données communiquées entre deux ou plusieurs entités, telles que leur intégrité, leur origine,

l'heure d'émission, etc. Le notaire s'interpose alors entre les entités communicantes.

Les mécanismes de contrôle d'accès peuvent utiliser des éléments variés tels que l'identité authentifiée de l'entité, une information sur cette entité, une liste de droits d'accès, des "étiquettes" de sécurité spécifiant des niveaux de sensibilité, etc.

La politique de contrôle d'accès choisie peut être de type discrétionnaire (l'utilisateur définit les droits d'accès aux informations dont il a la responsabilité) ou de type par mandat (l'autorisation d'accès dépend des droits du demandeur, du niveau de sensibilité des informations et d'attributs spécifiques).

Le contrôle de routage permet d'acheminer l'information à travers des sous réseaux, liaisons ou relais considérés comme sûrs. Il peut, soit spécifier explicitement les chemins autorisés, soit tenir compte du niveau de sensibilité des informations dans le choix des chemins utilisés.

3. Terminologie de la sécurité informatique

La sécurité informatique utilise un vocabulaire bien défini que nous utilisons dans cette thèse. Il est nécessaire de définir certains termes **[21]:**

- **Les vulnérabilités** : ce sont les failles de sécurité dans un ou plusieurs systèmes. Tout système vu dans sa globalité présente des vulnérabilités, qui peuvent être exploitables ou non.
- **Les attaques (exploits):** elles représentent les moyens d'exploiter une vulnérabilité. Il peut y avoir plusieurs attaques pour une même vulnérabilité mais toutes les vulnérabilités ne sont pas exploitables.
- **les contre-mesures** : ce sont les procédures ou techniques permettant de résoudre une vulnérabilité ou de contrer une attaque spécifique (auquel cas il peut exister d'autres attaques sur la même vulnérabilité).
- **Les menaces** : ce sont des adversaires déterminés capables de monter une attaque exploitant une vulnérabilité.

Les attaques peuvent à première vue être classées en deux grandes catégories :

- **Les attaques passives**: consistent à écouter sans modifier les données ou le fonctionnement du réseau. Elles sont généralement indétectables mais une prévention est possible.
- **Les attaques active** : consistent à modifier des données ou des messages, à s'introduire dans des équipements réseaux ou à perturber le bon fonctionnement de ce réseau. Notons qu'une attaque active peut être exécutée sans la capacité d'écoute. De plus, il n'y a généralement pas de prévention possible pour ces attaques, bien qu'elles soient détectables (permettant ainsi une réponse adéquate).

Profils et capacités des attaquants

Les attaquants peuvent être classés non seulement par leurs connaissances (débutant, expert, etc.) mais également suivant leur capacité d'attaques dans une situation bien définie.

Ainsi, on dénombre les capacités suivantes **[21]** :

- Transmission de messages sans capacité d'écoute (IP spoofing);
- Écoute et transmission de messages ;
- Écoute et perturbation des communications (blocage de paquets, DoS (Denial of Service) et DDoS (Distribute DoS));
- Écoute, perturbation et transmissions de messages;
- Écoute et relais de messages (attaques type man-in-the-middle).

Une autre caractéristique des attaquants est leur contrôle unidirectionnel ou bidirectionnel sur les communications, du fait de la nature asymétrique de celles-ci. En effet, la plupart des canaux de transmissions sur Internet ou sur tout autre réseau hétérogène sont unidirectionnels et empruntent des chemins différents suivant les règles de routage. Ainsi, de nombreux protocoles de sécurité sont également unidirectionnels et il faut établir plusieurs canaux pour permettre un échange en "duplex". Ces canaux qui sont au nombre de 2 minimums, sont la plupart du temps gérés de façon totalement indépendante par les protocoles de sécurité. C'est le cas pour SSL/TLS (Secure Socket Layer/Transport Layer Security) mais également pour IPSec.
(IP sécurisé) dont les associations de sécurité sont unidirectionnelles et indépendantes, chacune définissant son propre jeu de clés, algorithmes, etc.

4. Quelques possibilités en matière de sécurité réseaux

Actuellement, toute une série d'outils et de techniques permettent à un administrateur de sécuriser facilement son réseau et les machines qui le composent. Chacune de ces techniques se base sur des principes fondamentalement différents, mais celles-ci ont un but commun : permettre une connexion entre Internet (réseau non sécurisé) et le réseau de l'entreprise concernée, en assurant la sécurité des équipements et des informations disponibles sur ce réseau, tout en tenant compte des contraintes de plus en plus présentes, telles que les interconnexions de réseaux, les besoins de « contacts électroniques » pour le personnel (mails, transferts de fichiers, accès Web, etc.), les systèmes d'informations complexes, et autres **[03].**

Nous allons citer et expliquer brièvement quelques outils de sécurité courants, pour nous permettre par la suite de distinguer entre les systèmes de détection d'intrusions (IDSs), l'objectif de cette thèse, et les Firewalls, à cause de la confusion qui peut exister entre eux.

Les Firewalls

Le mot Firewall (Pare-feu) signifie qu'on instaure une série de protections en un point particulier entre deux entités connectées, en l'occurrence entre Internet et le réseau interne d'une entreprise. En pratique, le Firewall consiste en une architecture, plutôt qu'un matériel ou un logiciel

précis. Cette architecture intègre alors une série de composants matériels et logiciels, qui tentent précisément d'assurer le niveau de sécurité requis**. [06]**

L'architecture la plus utilisée actuellement est basée sur une « Zone démilitarisée », communément appelée DMZ (Demilitarized Zone). Elle consiste à placer un réseau intermédiaire entre l'accès Internet et le réseau interne (éventuellement plusieurs). Cette DMZ sera isolée, aussi bien vis-à-vis de l'Internet que du réseau local, par des systèmes de filtrage (filtres de paquets entrant et sortant). Ensuite, les éventuels serveurs nécessaires à l'entreprise devant continuer à être accessibles de l'extérieur seront connectés directement sur cette DMZ, de manière à les séparer du réseau interne. Par exemple, on pourra y trouver un serveur Web, un serveur DNS (Domain Name Service), un serveur de mails, un serveur FTP (File Transfer Protocol), etc. Dans le cas où l'un de ces serveurs serait compromis, le filtrage entre la DMZ et le réseau interne doit être capable d'assurer une protection suffisante au réseau interne.

Bien évidement, cette architecture doit être adaptée plus précisément à la structure d'une entreprise précise, et éventuellement intégrer des composants supplémentaires, tels que des Proxys (machine intermédiaire entre les ordinateurs d'un réseau local et le Web) et autres dispositifs.

Les filtres de paquets

Un filtre de paquet, tout comme son nom l'indique, permet de filtrer les paquets circulant sur un réseau. Plus précisément, on peut même dire que le filtrage s'effectue sur les paquets traversant une interface réseau. Celui-ci fonctionne en analysant le contenu de ces paquets, principalement en observant les valeurs de certains champs des en-têtes des protocoles IP (Internet Protocol), ICMP (Internet Control Message Protocol), UDP (User Datagramme Protocol) et TCP (Transmission Control Protocol). Cela permet par exemple d'interdire des paquets provenant d'une source précise, étant destinés à une destination précise, des paquets réceptionnés sur une interface précise, des paquets avec des ports sources ou cibles précis, d'intégrer des contraintes d'heures éventuelles d'après l'horaire d'une entreprise, etc. **[06]**

Au niveau de la configuration, on fait établir une série de règles de filtrage qui reflète la politique de sécurité de l'entreprise. Les paquets ne satisfaisant pas aux règles de filtrage seront alors bloqués (supprimés), et peuvent entraîner éventuellement la génération d'un message d'erreur (via un protocole comme ICMP).

Les scanners et les outils relatifs à la sécurité

Etant donné que les hackers (pirates informatique) trouvent de plus en plus les outils nécessaires à la réalisation de leurs attaques, les entreprises travaillant dans le domaine de la sécurité ont petit à petit commencé à proposer leurs propres outils de vérification de vulnérabilités. C'est ainsi qu'on commences à avoir apparaître toute une série de scanners, qui offrent de nombreuses possibilités.

Il est primordial à l'heure actuelle d'effectuer de nombreux tests de sécurité réguliers,

car ces tests permettent de mettre en avance des modifications dans l'architecture et dans la configuration du réseau et des machines qui le composent. Ces outils sont décomposés en toute une série de catégories, dont notamment

- Les scanners de vulnérabilités.
- Les scanners orientés réseaux.
- Les scanners orientés hosts (machines).
- Les sniffers.

Les systèmes de détection d'intrusions

Un IDS a pour fonction d'analyser en temps réel ou différé les évènements en provenance des différents systèmes à travers le réseau, de détecter et de prévenir les attaques. Les IDS ont donc un rôle d'alarme (la comparaison avec une alarme anti-vol placée dans le hall d'une maison, qui détecte des mouvements ou des ouvertures de portes, correspond d'ailleurs assez bien). Les buts sont nombreux :

- Collecter des informations sur les intrusions.
- Gestion centralisée des alertes.
- Effectuer un premier diagnostic sur la nature de l'attaque permettant une réponse rapide et efficace.
- Réagir activement à l'attaque pour la ralentir ou la stopper.

5. Conclusion

Bien que le domaine de la sécurité informatique soit très vaste, et qu'il est difficile de le cerner par une définition, la sécurité peut être considérée par le niveau de confiance donné à la confidentialité, l'intégrité et la disponibilité de l'information. La préservation de ces propriétés nécessite la mise

en place des services de sécurité qui seront implémentés par des mécanismes de sécurité.

Ces services peuvent être la confidentialité (des données ou du flux de données),l'authentification (d'une entité ou de l'origine des données), le contrôle d'accès, l'intégrité ou encore la non répudiation (avec preuve de l'origine ou preuve de la remise). Les mécanismes peuvent être le chiffrement, l'authentification, l'intégrité, la signature numérique...etc.

Dans cette thèse on s'intéresse au mécanisme de contrôle d'accès, et plus précisément à la surveillance du flux de données. Cette fonction est assurée par les systèmes de détection d'intrusion. Ils analysent en temps réel ou différé les évènements en provenance des différents systèmes à travers le réseau, détectent, préviennent les attaques et éventuellement prennent des contre-mesures.

Enfin, il faut noter que la sécurité ne peut être assurée à cent pour cent, et que les outils ne sont pas parfaits. Ils possèdent toujours des failles. Cependant, avec une bonne politique de sécurité, et un bon déploiement des outils, la sécurité peut être très proche des niveaux acceptés.

Chapitre II :

Détection des intrusions

Dans ce chapitre :

1. Introduction.
2. Intrusions informatiques.
3. Systèmes de détection d'intrusions.
4. Classification des systèmes de détection d'intrusion
5. L'analyse basée connaissance versus l'analyse comportementale
6. Les IDS Bases Hôtes Versus Les IDS Bases Réseau
7. les architectures d'implémentation des IDS
8. Une vue générale de quelques systems de détection d'intrusions existants
9. Emplacement de l'IDS
10. Evaluation de l'IDS
11. Conclusion

1. Introduction

De nos jours l'outil informatique est omni présent dans notre vie quotidienne. Que ce soit pour l'achat d'articles, faire des transactions bancaires, l'envoi de courrier ou encore la réservation de places de cinéma, nous dépendons énormément de cette technologie et nous ne pouvons plus nous en passer.L'arrivé d'Internet est définie comme l'âge d'or de l'informatique. Avec l'interconnexion des réseaux, la rapidité de diffusion et d'acquisition de l'information, Internet a révolutionné notre vie. Comme toute innovation ou nouvelle technologie, Internet peut avoir des conséquences regrettables en cas de mauvaise utilisation. De nouveaux genres d'espionnage industriel, guerre internations ou abus ont vu le jour. Par conséquent une communauté de gens malveillants s'est fondée et de nouveaux objectifs se sont créés tel que le détournement d'informations, la percée des secrets personnels et cela peut aller jusqu'à la destruction d'informations vitales.

2. Intrusions informatiques

Avec l'arrivée d'internet, de nouveaux marchés ont vu le jour et de nouvelles perspectives sont apparues. La plupart des opportunités sont dans le domaine commercial, où les entreprises peuvent exposer leurs produits et services au monde entier grâce à des sites web. Des transactions avec des sommes colossales sont effectuées chaque jour, ce qui expose les entreprises à différentes menaces.

Autrefois, on s'était beaucoup plus intéressé aux avantages de cette nouvelle technologie et rares étaient ceux qui pensaient ou mettaient quelques moyens et ressources pour assurer un minimum de sécurité. **[01]**

Le plus grave est que plusieurs entreprises courent des risques sans le savoir, et que les administrateurs réseau ou les personnes chargées d'assurer la sécurité informatique ignorent de quoi ils devraient se protéger. On estime aujourd'hui à moins de 4 mn le temps moyen pour qu'un PC non protégé connecté à Internet subisse une tentative d'intrusion ou soit contaminé par un programme malicieux. **[04].**

Différentes techniques ont été mises en place par des communautés des pirates informatiques. Chacune d'entre elles touche un certain aspect de l'outil informatique. Nous pouvons catégoriser les différentes techniques de piratage informatique en deux classes **[05]** :

- ✓ **Attaques réseaux** : Cette classe d'attaques, regroupe l'ensemble des techniques mises au point, permettant d'exploiter les faiblesses des réseaux ou bien les attaques qui ciblent des composants réseau.
- ✓ **Attaques applicatives**: Cette deuxième catégorie contient les techniques basées sur les faiblesses et les bugs des applications, permettant ainsi d'exploiter ces dernières pour des fins malveillantes.

Types de pirates informatiques

Les pirates informatiques se sont organisés en communautés et selon leur appartenance, leurs objectifs diffèrent ainsi que leurs manières de procéder. Deux grandes catégories se sont distinguées

- ✓ **Hackers** : ce sont d'excellents développeurs, spécialistes des réseaux. Ils existent depuis longtemps. Dés l'apparition de l'informatique et avec les premiers mini-ordinateurs, ils ont contribué à faire de l'informatique ce qu'elle est aujourd'hui. Certains d'entre eux ont participé au développement d'Unix, d'autres ont créé Internet. **[11].** Les hackers partagent librement l'information et ne causent jamais la perte ou la destruction d'informations. **[09]**
- ✓ **Crackers** : Les crackers quant à eux sont dangereux et peuvent causer des dégâts. Si les hackers s'intéressent à découvrir les failles et les bugs, les crackers les utilisent pour des fins destructives. Ils contournent les protections par mots de passe, utilisent des techniques de brute force, effacent des données, déstabilisent et mettent hors d'usage des systèmes …etc. **[10].**

La différence fondamentale entre hackers et crackers est que les hackers construisent les choses, les crackers (pirates) les démolissent. **[07]**

Différentes phases d'une attaque

Quelque soit les attaques menées, elles ont toutes la même démarche **[12]**. Une attaque est constituée de ce qui est connu sous le nom des « 5 P », ces cinq verbes anglais définissent ce que c'est qu'une attaque :

- **Prob** : C'est la partie d'audit qui sert à collecter des informations relatives à la cible. Cette collecte est facilitée par des outils tel que Whois, DNS Lookup, Un scan des ports ou encore un scanner de vulnérabilités.
- **Penetrate** : C'est la phase de pénétration, en utilisant les informations collectées.
- **Persist** : Une fois l'attaque réussie, et le système pénétré, l'attaquant tente de garder un contrôle sur le système. Il crée ainsi un compte avec des droits d'administrateur, ou installe une porte dérobée (Cheval de trois).
- **Propagate** : L'attaquant vérifie s'il y'a d'éventuelles cibles à partir du réseau local, si l'attaque peu être propagée sur d'autres machines.
- **Paralyze** : A la fin de son attaque, l'attaquant peut utiliser la victime pour mener d'autres attaques ou détruire des informations vitales, ou carrément mettre hors d'usage le système.

Les conséquences du piratage informatique peuvent être très graves et les chiffres le montrent clairement : des centaines de systèmes ont été piratés, des quantités énormes d'informations confidentielles et de secrets industriels ont été volés.

Ce problème est devenu très sérieux. Les infrastructures informatiques gouvernementales et nationales sont les cibles préférées des pirates. Ensuite, les organisations commerciales, les banques et en derniers lieu les simples utilisateurs **[14].**

Voici un petit historique des plus grandes attaques connues avant le bug de l'an 2000

- Le 11 mai 1999 : Le site web de la maison blanche a été mis hors service.
- Le 21 mai 1999 : le GAO (General Accounting Office) dit « Nous avons réussi à pénétrer quelques systèmes à missions critiques, dont un responsable du calcul de position terre-orbite pour un vaisseau spatial ».
- Le 1er juin 1999 : Des pirates chinois ont attaqué une poignée de sites gouvernementaux américains le site de la maison blanche a été mis hors service.
- Le 7 octobre 1999 : Des pirates russes réussissent à s'introduire dans le réseau du département de défense américain et ont réussi à dérober une très grande quantité d'informations du département d'armement nucléaire et de recherche.
- Le 7 octobre 1999 : Une attaque réussie contre la NASA, une attaque très massive et très discrète, les attaquants ont pris une très grande quantité d'informations (des listes de fichiers des répertoires personnels… etc.). Les attaquants ont installé des backdoors, pour pouvoir se reconnecter plus tard. les backdoors ont été découvert bien après l'attaque…etc

Nous remarquons que dans un laps de temps très réduit, un grand nombre d'attaques ont eu lieu touchants les plus grandes institutions gouvernementales qui sont supposées être sécurisées. Nous pouvons donc imaginer les dégâts causes sur des systèmes moins importants. Le passage à l'an 2000 a coute très chère aux sociétés de l'information d'après le rapport publie par CSI (Computer Security Institute). montre les pertes causées par le piratage informatique. Nous pouvons voir clairement sur la figure que durant les deux années qui ont suivi l'an 2000, une grande activité des pirates a été constatée, et par conséquent, le total des pertes s'est vu augmenter d'une manière très remarquable.

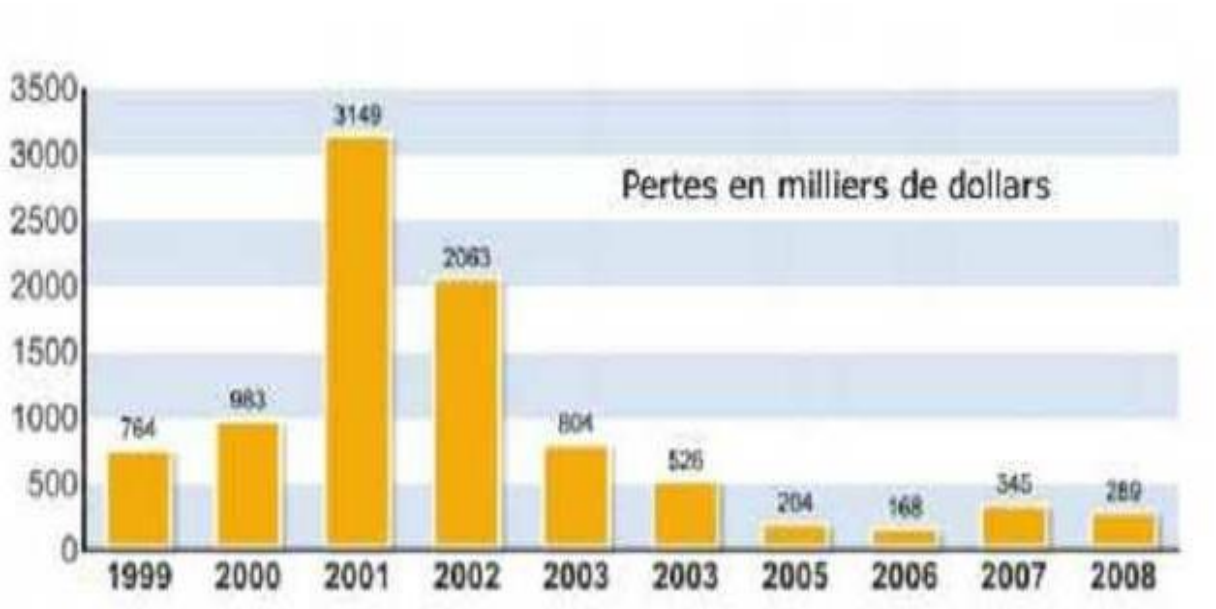

FIG II.1 Rapport de CSI des pertes causées par le piratage informatique en millier de dollars. **[08]**

Différentes attaques sont derrières toutes ces pertes, que ce soit des attaques de virus, des attaques par déni de service ou des détournements de sessions, nous retrouvons sur la figure **FIG II.2** les différentes techniques de piratage citées plus haut présentent avec pour chacune le taux de perte cause.

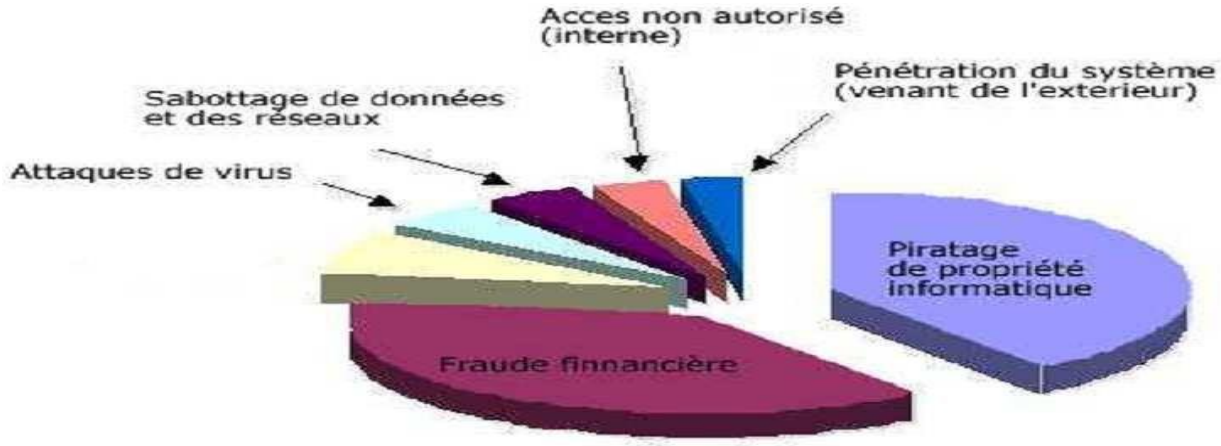

FIG II.2 Pourcentage des différents types de piratage informatique selon les dégâts causés**[07].**

2.3Contre-mesures

Bien évidemment, face à de telles menaces, les informaticiens et surtout les développeurs ont réagi, en créant des mécanismes de protection, garantissant ainsi un certain niveau de sécurité. **[15]**

Voici les contres mesures les plus connues :

- Pares feu
- Antivirus
- Scanners de vulnérabilités
- Patchs
- Systèmes de leurre

Malgré l'existence de différents mécanismes de protection (pare feu, antivirus, scanner de vulnérabilités), chaque système de protection peut présenter une menace particulière pour les systèmes d'informations, car chacun d'entre eux à ses points faibles.

D'autres méthodes et mécanismes existent tels que les systèmes de détections d'intrusions (Intrusion Detection System). Objet de ce mémoire de fin d'études, ce sont des systèmes permettant de détecter des tentatives d'attaques ou des attaques réussies. Ils permettent de retrouver les traces des attaques déjà connues ou de détecter une nouvelle forme d'attaques. Les systèmes de détection d'intrusions servent à contrôler le trafic autorisé par le pare- feu et prennent des décisions si le trafic observé est suspect. Ces derniers peuvent découvrir les attaquants qui parviennent à pénétrer le pare-feu ou le trafic qu'ils jugent suspects et l'annoncent aux administrateurs du système, qui peuvent prendre des mesures pour éviter d'éventuels dégâts.

3. Systèmes de détection d'intrusions

C'est quoi la détection d'intrusion ?

La détection d'intrusion est le processus qui consiste à surveiller les évènements se produisant dans un ordinateur ou dans un réseau informatique, et de les analyser pour découvrir des signes d'intrusions, définies comme des tentatives de compromission de la confidentialité, l'intégrité, la disponibilité et la responsabilité, ou pour dévier des mécanismes de sécurité **[16].**

Les intrusions sont provoquées par : l'accès d'attaquants externes aux systèmes via des réseaux ouverts comme Internet, des utilisateurs autorisés qui essayent de gagner des privilèges additionnels pour lesquels ils ne sont pas autorisés, ou des utilisateurs autorisés qui abusent de leurs privilèges **[03]**.Les systèmes de détection d'intrusions (IDS : Intrusion Detection System) sont les systèmes logiciels ou matériels qui automatisent cette tâche de surveillance et d'analyse **[03]**.Debar **[17]** simplifie le système de détection d'intrusion dans un détecteur qui analyse les informations en provenance du système surveillé (voir la figure **Fig.II.3**).

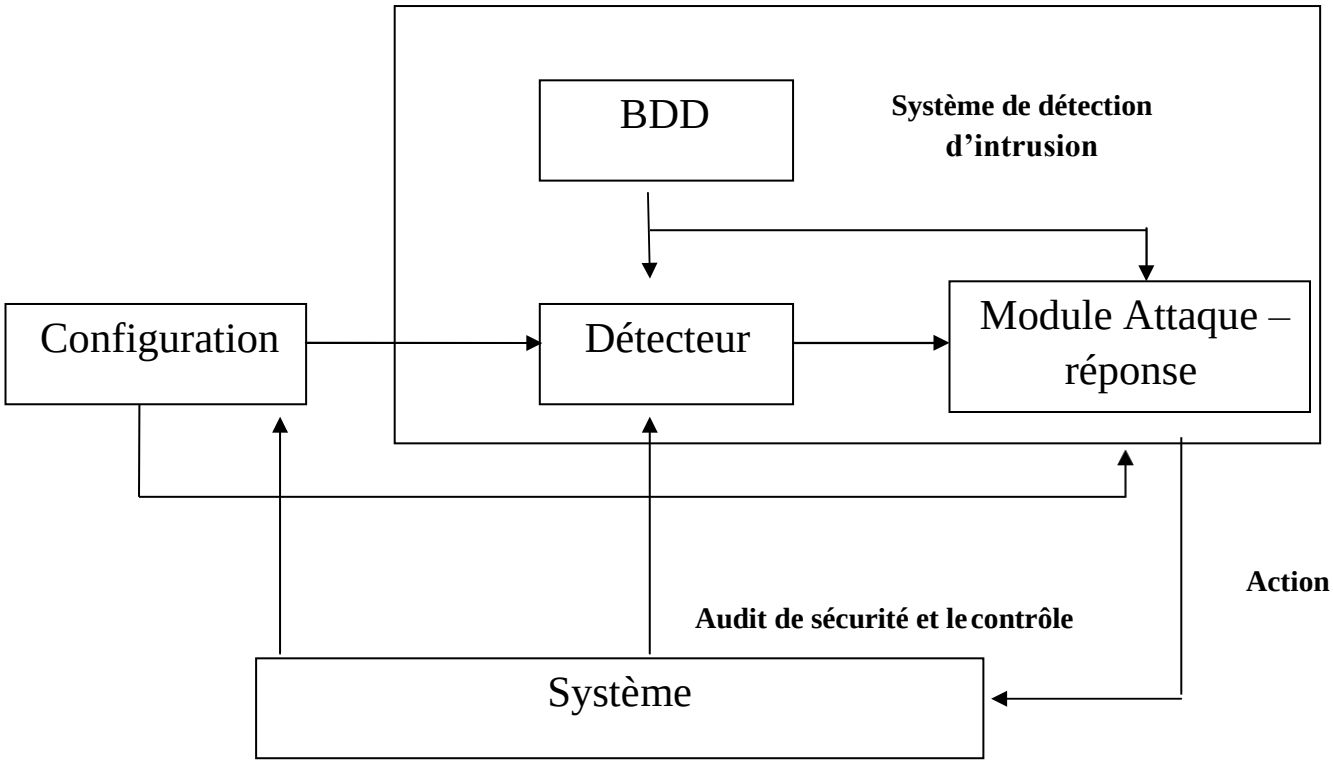

Fig. II.3 Modèle simplifié d'un système de détection d'intrusions **[18]**

Un système de détection d'intrusions à un niveau très macroscopique **[19]**peut être décrit comme un détecteur. Ce détecteur est un moteur d'analyse qui reçoit des données de trois sortes de ressources (**Fig. II.3**). L'analyse de ces données génère une décision d'évaluation de la probabilité que ces actions peuvent être considérées comme des symptômes d'intrusions. Ces données sont :

- ✓ Des informations de configuration relatives à l'état actuel du système.
- ✓ Des informations à long terme relatives à la technique utilisée pour détecter les intrusions par exemple une base de connaissances d'attaques.
- ✓ Des informations venant du système à protéger qui sont les informations d'audit décrivant les événements qui apparaissent dans le système.

Efficacité des systèmes de détection d'intrusions

Philip dans**[18]** définit trois critères pour évaluer l'efficacité des systèmes de détection d'intrusion

- ➢ **L'exactitude (accuracy) :** on parle de l'exactitude quand le système de détection d'intrusion déclare comme malicieux une activité légale. Ce critère correspond au faux positif.
- ➢ **La performance (performance) :** la performance de système de détection d'intrusion est le taux de traitement des évènements. Si ce taux est faible, la détection en temps réel est donc impossible.

- **La complétude (completeness) :** on parle de la complétude quand le système de détection d'intrusion rate la détection d'une attaque. Ce critère est le plus difficile, parce que il est impossible d'avoir une connaissance globale sur les attaques. Ce critère correspond au vrai négatif.

Debar dans **[10]** a rajouté également les deux critères suivants :

- **La tolérance aux fautes (Fault tolerance)** : le système de détection d'intrusion doit lui- même résisté aux attaques, particulièrement au déni de service. Ceci est important, parce que plusieurs systèmes de détection d'intrusion s'exécutent sur des matériels ou logiciels connus comme vulnérables aux attaques.
- **La réaction à temps (Timeliness)** : le système de détection d'intrusion doit s'exécuter et propager les résultats de l'analyse le plus tôt possible, pour permettre à l'officier de sécurité de réagir avant que des graves dommages n'aient lieu. Ceci implique plus qu'un calcul de performance, parce qu'il ne s'agit pas seulement de temps de traitement des évènements, mais aussi le temps nécessaire pour la propagation et la réaction à cet évènement.

Que doit assurer la détection d'intrusion?

La détection d'intrusion permet aux organisations de protéger leurs systems contre les menaces qui ne cessent de croître à cause de l'augmentation de la connectivité du réseau public (Internet), et la confiance accordée aux systèmes informatiques qui comportent des bugs.

La question pour les professionnels de sécurité ne devraient pas être s'il faut utiliser la détection d'intrusion, mais quels dispositifs utiliser et quelles sont leur capacité de détection d'intrusion.

Les systèmes de détection d'intrusion ont gagné l'acceptation d'être un élément nécessaire dans l'infrastructure de la sécurité informatique de chaque organisation. En effet, il y a plusieurs raisons pour acquérir et utiliser les systèmes de détection d'intrusion :

- ✓ Pour détecter les attaques et autres violations de sécurité qui ne sont pas empêchées par d'autres outils de sécurité.
- ✓ Pour documenter les menaces existantes dans une organisation, c'est-à-dire découvrir les vulnérabilités avant qu'elles ne soient exploitées par un attaquant.
- ✓ Pour agir en tant que contrôle de qualité pour la conception de sécurité, particulièrement dans les grandes et complexes entreprises.
- ✓ Pour fournir des informations utiles au sujet des intrusions qui ont eu lieu, et faire des diagnostics, recouvrement, et corrections des facteurs causatifs.

- ✓ Pour arrêter les intrusions afin de limiter les dégâts. Malheureusement cela n'est pas toujours possible à cause de la complexité et la diversité des intrusions, et la naissance de nouveaux types d'intrusions liées au développement des nouvelles technologies d'information. Les contre-mesures actives sont souvent optionnelles dans la quasi-totalité des systèmes de détection d'intrusion.

Modèle de processus de la détection d'intrusion

La majorité des systèmes de détection d'intrusion peuvent être décrits en terme de trois composants fonctionnels fondamentaux **[16]** :

- ❖ **La source d'informations (sonde)** : Les différentes sources des évènements utilisées pour déterminer les intrusions qui ont eu lieu. Ces sources peuvent être fournies par les différents niveaux du système d'information : les réseaux, les hôtes, et les applications.
- ❖ **L'analyse** : La partie du système de détection d'intrusions qui réellement organise et donne un sens aux évènements dérivés des sources d'informations, décidant quand ces évènements indiquent que des intrusions se produisent ou ont déjà eu lieu. Les principales approches communes d'analyse sont : détection d'abus (The misuse detection) ou encore dite approche par scénarios et détection d'anomalie (Anomaly detection) ou encore dite approche comportementale qui seront expliquées par la suite.
- ❖ **La réponse** : L'ensemble de contre-mesures que le système prend une fois qu'il détecte des intrusions. Celles-ci sont typiquement groupées dans des mesures actives et passives, les mesures actives comportent une certaine interposition automatisée de la part du système, alors que les mesures passives rapportent des résultats issus de l'analyse aux responsables, qui sont alors prévenus pour agir et prendre une action basée sur ces rapports.

4. Classification des systèmes de détection d'intrusion

Le domaine de la détection d'intrusions est encore jeune mais en plein développement. Nous dénombrons à l'heure actuelle environ une centaine de systèmes de détection d'intrusions, que ce soit des produits commerciaux ou du domaine public **[03].** Il est donc devenu très utile d'utiliser des critères pour classifier ces systèmes de détection d'intrusion, c'est ce que nous allons présenter dans cette section.Le domaine de la détection d'intrusions est encore jeune mais en plein développement. Nous dénombrons à l'heure actuelle environ une centaine de systèmes de détection d'intrusions, que ce soit des produits commerciaux ou du domaine public **[03].** Il est donc devenu très utile d'utiliser des critères pour classifier ces systèmes de détection d'intrusion, c'est ce que nous allons présenter dans cette section.

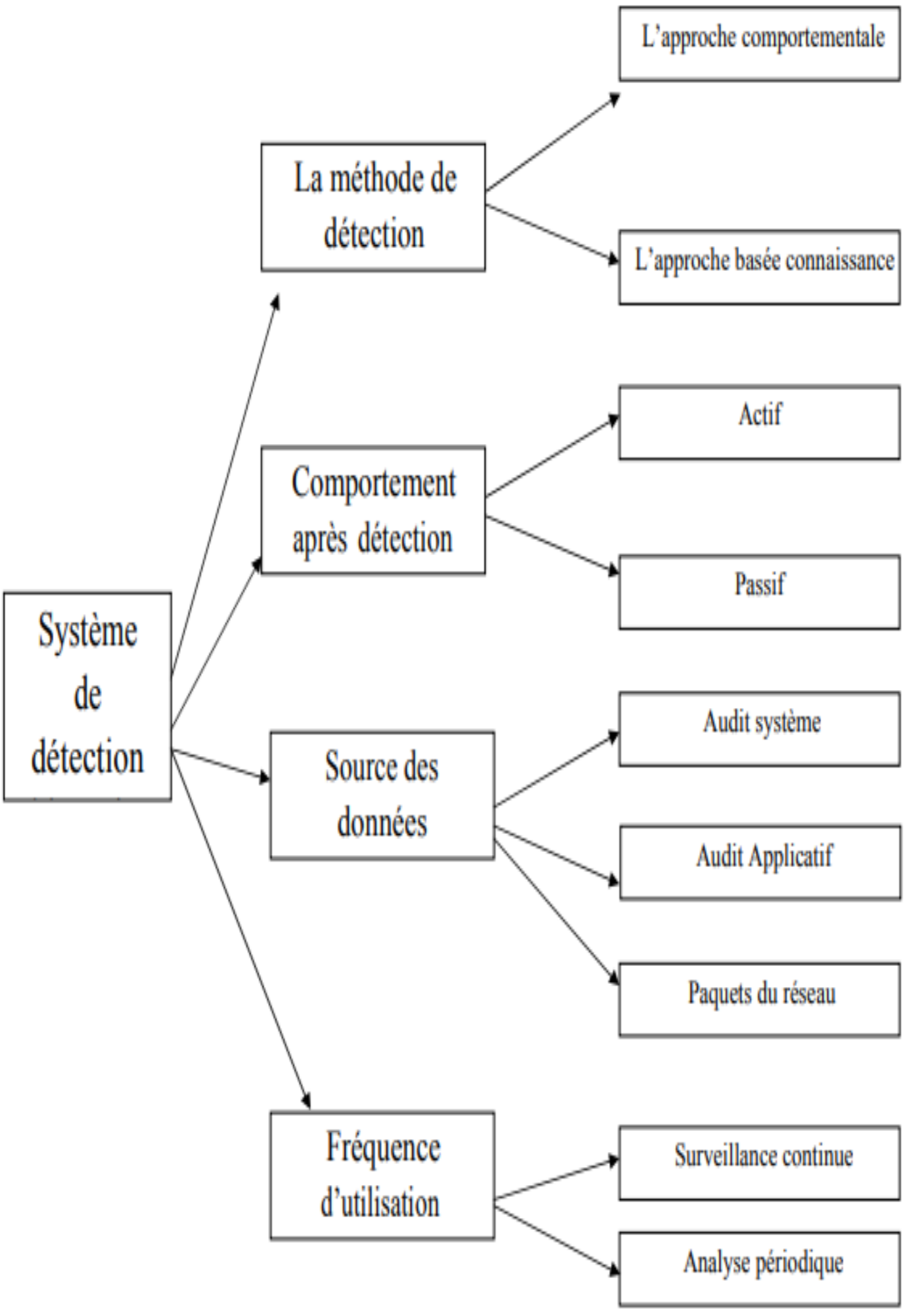

Fig. II.4 Taxonomie des systèmes de détection d'intrusion **[03]**

Il existe plusieurs critères qu'on peut utiliser pour classifier les différents systèmes de détection d'intrusion, dont les principaux sont résumés dans la figure **Fig. II.3 [04].**

A- La méthode de détection

La détection d'intrusions repose sur deux approches de base :

- ✓ L'approche comportementale.
- ✓ L'approche basée connaissance.

A-1) L'approche comportementale

Cette approche est connue aussi par l'approche de détection d'anomalies. Elle consiste à définir un profil de l'activité normale d'un utilisateur et à considérer les déviations significatIIIes de l'actIIIité d'utilisateur courante par rapport aux profils de comportement normaux comme anomalie..

A-2) L'approche basée connaissance

Cette approche définit des signatures soupçonneuses basées sur les vulnérabilités connues de système et la politique de sécurité. Une intrusion est signalée lorsque la trace d'une attaque connue est présente dans les traces d'audit.

Ces deux méthodes d'analyse constituent la partie importante des systèmes de détection d'intrusions. Pour cette raison, elles seront détaillées dans les sections suIIIantes.

B- Le comportement après la détection d'intrusions

Le comportement d'un IDS après la détection d'une intrusion est l'ensemble des actions prises par le système lorsqu'il détecte une attaque. Ces réponses peuvent être actIIIes ou bien passIIIes.

B-1) Réponse active

La réponse active implique des actions automatisées prises par un IDS quand le système détecte une intrusion. Par exemple interrompre le progrès d'une attaque pour bloquer ensuite l'accès suivant de l'attaquant.

B-2) Réponse passive

Dans ce cas, quand une attaque est détectée, le système de détection d'intrusions ne prend aucune action. Il génère seulement une alarme pour notifier l'administrateur de système qui va prendre des mesures en se basant sur les rapports générés par le système de détection d'intrusions.

C) La nature des données analysées

Les systèmes de détection d'intrusions sont classés en fonction de l'origine des données qui seront exploitées pour détecter des actions intrusives. La source de données utilisée est une caractéristique essentielle pour classer les systèmes de détection d'intrusions. On distingue trois catégories de sources d'informations :

- ✓ Les audits systèmes.
- ✓ Les audits applicatifs.
- ✓ Le trafic réseau.

C-1) Les audits systèmes

Les audits systèmes sont produits par le système d'exploitation d'un hôte. Ces données permettent à un IDS de contrôler les activités d'un utilisateur sur un hôte. Elles peuvent être également de plusieurs types, par exemple :

- ➢ **Historique des commandes systèmes** : tous les systèmes d'exploitation possèdent des commandes pour obtenir des informations instantanées sur les processus actifs courants dans un ordinateur. Grâce à ces commandes, l'IDS peut avoir des informations précises sur les événements systèmes.
- ➢ **Accounting** : l'accounting fournit des informations sur l'usage des ressources partagées par les utilisateurs. Ces ressources sont par exemple : le temps processeur, la mémoire, L'espace disque, les applications lancées, etc.
- ➢ **Systèmes d'audit de sécurité** : les systèmes d'exploitation sont dotés par ce service pour définir des événements, les associer à des utilisateurs et assurer leurs collectes dans un fichier d'audit. L'IDS possède potentiellement des informations sur toutes les actions effectuées par un utilisateur.

L'avantage de ces données systèmes réside dans leur fiabilité et leur granularité fine, qui permettent un diagnostic précis des actions effectuées sur un hôte par un attaquant. Cependant, le volume d'événements généré par les audits systèmes est très volumineux ce qui implique un impact très important sur les performances de la machine surveillée. Les IDS qui se basent sur cette catégorie des sources de données sont appelés : Les IDS basés hôte « Host Based Intrusion Detection System ».

C-2) Les audits applicatifs

La troisième catégorie de source de données est constituée des audits applicatifs. Les données à analyser sont produites directement par une application, par exemple des fichiers logs générés par les serveurs ftp et les serveurs Web. L'avantage de cette catégorie est que les données produites sont très synthétiques, elles sont sémantiquement riches et leur volume est modéré. On note que ces types d'informations sont généralement intégrés dans les IDS basés hôte.

Vu de l'importance des IDS basés hôte et basés réseau, une étude détaillée de ces deux types d'IDS sera exhibée dans les prochaines sections.

D)La fréquence d'utilisation

La fréquence d'utilisation d'un système de détection d'intrusions peut exister selon deux formes :

D-1) Surveillance périodique

Ce type de système de détection d'intrusions analyse périodiquement les différentes sources de données à la recherche d'une éventuelle intrusion ou une anomalie passée.

D-2) Surveillance en temps réel

Les systèmes de détection d'intrusions en temps réel fonctionnent sur le traitement et l'analyse continue des informations produites par les différentes sources de données. La détection d'intrusions en temps réel permet de limiter les dégâts produits par une attaque car elle permet de prendre des mesures qui réduisent le progrès de l'attaque détectée.

5. L'analyse basée connaissance versus l'analyse comportementale

Comme nous l'avons vu dans la section précédente, il existe deux approches pour détecter les intrusions dans les systèmes informatiques **[12][17][19].** L'approche basée connaissance qui se base sur la définition d'un modèle constitué des actions interdites dans les systèmes informatiques et l'approche comportementale qui est basée sur la définition d'un modèle constitué des actions autorisées.

L'analyse basée connaissance

Cette approche de détection est désignée en anglais par le terme « Misuse Detection », qui signifie dans la littérature la détection d 'une mauvaise utilisation, et il existe plusieurs traductions françaises adoptées pour cette approche, par exemple l'approche par signatures ou par scénarios.

Elle est caractérisée par l'existence d'une base de connaissances qui comporte des modèles d'attaque connus a priori qui sont appelés les signatures. Elle examine les activités du système et du réseau en cherchant des événements ou l'ensemble des événements qui décrivent une attaque connue. Ainsi, dans cette approche, tous ce qui n'est pas explicitement interdit est autorisé. Cette approche possède un certain nombre d'avantages et d'inconvénients. **[05] [01]**

Les avantages de l'analyse basée connaissance

- L'analyse basée connaissance est très efficace pour la détection d'attaque avec un taux très bas des alarmes de type positif faux.
- Les alarmes générées sont significatives.

Les inconvénients de l'analyse basée connaissance

- Cette analyse basée connaissance permet seulement la détection des attaques qui sont connues au préalable. Donc, la base de connaissances doit être constamment mise à jour avec les signatures de nouvelles attaques.
- Le risque que l'attaquant peut influencer sur la détection après la reconnaissance des signatures.

L'analyse comportementale

Dans l'analyse comportementale, un modèle de comportement normal du système surveillé est préalablement construit. Ce modèle est appelé profil de comportement normal qui sera utilisé comme une référence dans la détection. Au cours de la surveillance du système, toute déviation significative du comportement courant de système contrôlé par rapport au comportement normal de référence donne lieu à une attaque. Cette approche possède aussi un certain nombre d'avantages et d'inconvénients **[20] [18].**

Les avantages de l'analyse comportementale

- L'analyse comportementale n'exige pas des connaissances préalables sur les attaques.
- Elle permet la détection de la mauvaise utilisation des privilèges.
- Elle permet de produire des informations qui peuvent être employées pour définir des signatures pour l'analyse basée connaissance.

Les inconvénients de l'analyse comportementale

- Les approches comportementales produisent un taux élevé des alarmes de type positif faux en raison des comportements imprévisibles d'utilisateurs et des réseaux.
- Ces approaches nécessitent des phases d'apprentissage pour caractériserles profils de comportement normaux.
- Les alarmes générées par cette approche ne sont pas significatives.

Cette étude comparative entre les deux approches d 'analyses utilisées par les systèmes de détection d'intrusions montre l'existence d'une complétude entre ces deux méthodes. Cette complétude qui permettra de surmonter les inconvénients relatifs à chaque méthode d'analyse. Pour cette raison, il est préférable d'adopter les deux techniques d 'une manière parallèle pour obtenir un système de détection d 'intrusions efficace **[14**] Cependant, les systèmes de détection d'intrusions commerciaux disponibles emploient seulement la technique basée signature, ce qui motive les efforts de recherche croissants pour construire des détecteurs d'anomalies efficaces pour des buts de détection d' intrusions. L'effort principal de cette recherche est concentré sur les systèmes de détection d'intrusions qui sont basés sur la technique comportementale. Pour cette raison, nous présenterons dans la section suivante les différentes approches utilisées dans la méthode de détection comportementale.

Les techniques utilisées dans l'approche comportementale

Un système de détection d'intrusions basé sur la détection d'anomalies contrôle les activités du système afin de les classer comme normales ou anomalies. Il procède à construire des profils d'un comportement normal pour les activités des utilisateurs et à observer les déviations significatives de l'activité de l'utilisateur courante par rapport à la forme normale établie. D'une façon générale, la détection d'anomalies est composée de deux phases :

- **Une phase d'apprentissage** : le système apprend le comportement normal d'un utilisateur ou un système. Il crée ainsi « le profil normal » d'un utilisateur ou d'un système à partir des données collectées.
- **Une phase de détection** : le système compare les traces d'audit courantes ou le trafic réseau aux profils pour vérifier s'il n'y a pas une activité intrusive. Si la différence entre le profil et les traces d'audit est significative, une alarme est déclenchée.

Pour pouvoir formaliser le comportement normal d'un système, des approches diverses ont été utilisées **[14]** . Cette section sera consacrée à une présentation générale de ces différentes approches.

L'approche statistique

L'approche statistique est utilisée pour la génération d'un modèle de comportement normal d'un système. Elle consiste à générer le profil de comportement normal à partir d 'un ensemble de variables aléatoires, échantillonnées à des intervalles réguliers dans le temps, ces variables peuvent être par exemple :

- Le temps CPU utilisé.
- Le nombre de connexions établi durant une période de temps.
- Les fichiers les plus fréquemment utilisés.
- Les entrées/sorties effectuées.
- Etc.

Dans cette approche, Denning a proposé un ensemble de modèles statistiques, leur but est de définir à partir de n observations X1, X2,….Xn sur une variable donnée x, si la valeur Xn+1 de l'observation n+1 est anormale. Parmi ces modèles, on peut citer les modèles suivants :

- **Le modèle opérationnel** : ce modèle est très simple, une anomalie est détectée par la comparaison de la valeur d'une nouvelle observation avec un seuil fixe qui est défini d'une manière intuitive en se basant sur les données historiques.
- **Le modèle de déviation standard et moyen** : Ce modèle définit un seuil d'anomalie par l'estimation d'un intervalle de confiance. L'intervalle de confiance est la moyenne et l'écart type des n observations qui peuvent être considérées normales. Si la valeur d'une nouvelle observation est en dehors de cet intervalle alors elle est considérée anormale.
- **Le modèle de covariances** : Il est similaire au modèle précédent mais il se base sur la corrélation de plusieurs variables pour tirer des conclusions.

Ces approches ont été adoptées dans le développement de plusieurs systèmes de détection d'intrusions, on peut citer par exemple :

- MIDAS « Multics Intrusion Detection and Alerting System » **[18].**
- NIDES « Next Generation Real time Intrusion Detection Expert System » **[11]**

L'approche de la machine learning

Le but principal de l'utilisation de la machine learning est l'extraction automatique des caractéristiques des activités normales qui sont critiques pour la détection d'anomalies. A partir des données d'audit, le modèle de la machine learning essaye d'identifier des règles pour définir les comportements normaux. Ces règles seront employées pour déterminer si des événements nouvellement observés sont anormaux ou non.

Parmi les travaux qui sont basés sur cette approche, le système de détection d'intrusions basé règle TIM « Time based Inductive Machine » **[09]** proposé par Teng et son groupe. TIM génère des règles qui essayent de prédire les événements futurs en se basant sur des événements qui se sont déjà produits dans le passé.

Durant la phase de détection, les règles possédant des parties gauches qui correspondent à la séquence d'événements observée seront sélectionnées et l 'évènement prédit de cette règle sera comparé avec le dernier évènement qui apparaît dans la séquence d 'évènements observée. Si cet évènement dévie d'une manière significative de ceux prédits dans la règle, alors TIM alerte l'officier de sécurité.

L'approche de réseaux de neurones

Les réseaux de neurones sont utilisés dans la détection d'anomalies afin d'exploiter leurs capacités d'apprentissage. L'idée de base est d'utiliser les mécanismes d'apprentissage des réseaux de neurones pour apprendre les profils de comportements normaux des utilisateurs ou d'un système.

Plusieurs travaux ont été élaborés qui ont essayé d'abord d'apprendre à un réseau de neurones le comportement normal d'un système pour qu'il puisse par la suite de décider si un ensemble d'action est normal ou suspect. Parmi ces travaux, nous citons le travail de Debar qui a proposé l'utilisation des réseaux de neurones pour construire un modèle du comportement des utilisateurs du système informatique. Le travail proposé s'intéresse à l'aspect dynamique du comportement et à sa présentation sous des séries d'actions temporelles.

L'approche de datamining

Le but de cette approche est l'exploitation des techniques de datamining pour extraire des anomalies à partir des grandes quantités de données du trafic réseau. Parmi les travaux existants, on peut citer ADAM « Audit Data Analysis and Mining » **[17]** qui est un système de détection d'intrusions qui exploite des techniques de datamining pour construire des profils du trafic réseau normaux.

ADAM utilise les règles d'association pour construire des profils du trafic de réseau normaux qui seront employées par la suite pour détecter les comportements incorrects de trafic de réseau. Pour détecter des anomalies, ADAM extrait les règles d 'association à partir des données du trafic réseau et qui seront comparées aux profils du réseau. Si n'importe quelle règle d'association produite à partir des données de trafic de réseau rassemblées n'est pas incluse dans les profils, alors cette règle est considérée comme une indication d'un comportement incorrect.

L'approche immunologique

Vu que la détection d'anomalies est une application directe de la métaphore immunitaire, plusieurs travaux tentent de calquer la manière dont le système immunitaire naturel procède pour la distinction entre le comportement normal et le comportement suspect afin de construire des systèmes de détection d'intrusions efficaces. Parmi ces travaux nous citons le système LYSIS **[17],** qui a intégré des différentes propriétés et mécanismes inspirés par le système immunitaire humain. Il se base principalement sur l'algorithme de la sélection négative proposé par Forrest **.**Dans ce travail, une population de détecteurs est générée aléatoirement, puis en se basant sur les modèles de comportement normaux des utilisateurs, les détecteurs qui identifient ces modèles seront éliminés, en d'autre terme élimination des détecteurs qui détectent le soi. La population des détecteurs restants procède à contrôler les opérations effectuées dans le système de telle sorte que s'il y a une correspondance entre un détecteur et l'opération courante dans le système alors cette opération est considérée litigieux ou anormal.

6. Les IDS Bases Hôtes Versus Les IDS Bases Réseau

En raison des multiples possibilités d'attaques des systèmes informatiques et les réseaux. Il existe différents types de systèmes de détection d'intrusions **[19**], [**01**], qui varient selon l'endroit qu'ils surveillent et ce qu'ils contrôlent (les sources d'information).

L'IDS basé hôte (host- based IDS)

L'IDS basé hôte contrôle un seul hôte. Il analyse des informations rassemblées d'un système d'ordinateur individuel, ce qui permet à l'IDS basé hôte d'analyser des activités avec une grande fiabilité et précision en déterminant exactement les processus et les utilisateurs impliqués dans une attaque particulière. Les avantages et les inconvénients **[19**] de l'IDS basé hôte sont :

Les avantages d'un IDS basé hôte

- La capacité de contrôler les activités locales des utilisateurs avec précision.

- Capable de déterminer si une tentative d'attaque est couronnée de succès.
- La capacité de fonctionnement dans des environnements cryptés.
- L'IDS basé hôte fonctionne sur les traces d'audit des systèmes d'exploitation ce qui lui permet de détecter certains types d'attaques (ex : Cheval de Troie).

Les inconvénients d'un IDS basé hôte

- La vulnérabilité aux attaques du type déni de service puisque l'IDS peut résider dans l'hôte cible par les attaques.
- La difficulté de déploiement et de gestion, surtout lorsque le nombre d'hôtes qui ont besoin de protection est large.
- Ces systèmes sont incapables de détecter des attaques contre de multiples cibles dans le réseau.

L'IDS basé réseau (Network- based IDS)

Bien que le système de détection d'intrusions basé hôte a montré des résultats encourageants mais son problème majeur est la détection des intrusions essayées à travers le réseau. Pour détecter cette sorte d'intrusion, l'IDS a besoin de contrôler des événements multiples produits sur plusieurs hôtes. En effet, une proportion large d'intrusions est réalisée via les réseaux et en conséquence l'utilisation des informations sur le trafic réseau rend l'IDS plus efficace. Ce problème motive l'évolution des IDS basés hôte vers l'IDS basé réseau système de détection d'intrusions basé réseau détecte des attaques en capturant et analysant des paquets du réseau. Les avantages et les inconvénients **[19]** **[05]** de ce type d'IDS sont :

Les avantages d'un IDS basé réseau

- L'IDS basé réseau est capable de contrôler un grand nombre d 'hôte avec un petit coût de déploiement.
- Il n'influence pas sur les performances des entités surveillées.
- L'IDS basé réseau est capable d'identifier les attaques de /à multiples hôtes.
- L'IDS basé réseau assure une grande sécurité contre les attaques parce qu'il est invisible aux attaquants.

Les inconvénients d'un IDS basé réseau

- L'IDS basé réseau ne peut pas fonctionner dans des environnements cryptés.
- Ce type d'IDS ne permet pas d'assurer si une tentative d'attaque est couronnée de succès.

7. les architectures d'implémentation des IDS

L'architecture d'implémentation d'un système de détection d'intrusions qui est considérée comme une stratégie de contrôle décrit la manière de contrôle effectuée par les éléments d'un système de détection d'intrusions. Nous distinguons trois approches d'implémentation **[15**], **[14], [19**]: Monolithique, hiérarchique et coopérative.

L'approche monolithique (centralisée)

Les premières mises en œuvre des systèmes de détection d'intrusions ont employé une architecture monolithique sous laquelle les données rassemblées seront analysées à un point central. Puisque le contrôle de l'activité des utilisateurs d'un seul hôte ne révèle pas les attaques impliquant des hôtes multiples. L'IDS basé réseau a été développé, qui analyse le trafic de réseau pour déduire les anomalies venant du réseau.

Bien qu'un IDS basé réseau avec un serveur central a montré des résultats prometteurs pour des réseaux à petite échelle. Cependant, cette approche ne peut pas supporter un grand réseau à cause de la quantité énorme des données des différents hôtes qui doivent être analysée par le serveur central, ce qui engendre une dégradation sévère des performances de réseau. Un exemple d'un système de détection d'intrusions qui se base sur l'approche monolithique est le système NADIR **[18],**

L'approche hiérarchique

Cette approche a été proposée pour surmonter les problèmes de l'approche monolithique. Elle est caractérisée par l'existence des secteurs de contrôle hiérarchiques. Chaque IDS contrôle un secteur avec l'élimination du transfert des données d'audit rassemblées par les hôtes locaux à un point central. Chaque IDS à n'importe quel niveau de contrôle exécute une analyse locale et envoie ses résultats d'analyse au niveau suivant dans la hiérarchie.

L'approche hiérarchique montre la meilleure incrémentabilité « scalability » en permettant des analyses locales aux secteurs de contrôle distribués. Cependant, les problèmes vus précédemment demeurent toujours. En plus, le changement de la topologie du réseau cause un changement aussi bien dans la hiérarchie de réseau et dans les mécanismes de rassemblement des rapports d'analyse locaux. Ainsi, la difficulté de détecter les attaques qui visent le niveau le plus haut de la hiérarchie. Un exemple de système de détection d'intrusions hiérarchique : GrIDS**[15], [12],** EMERALD **.**

L'approche cooperative (distribuée)

Cette approche a été suggérée pour résoudre les problèmes de l'approche précédente. Elle essaye de distribuer les responsabilités d'un serveur central à un nombre de systèmes de détection d'intrusions coopératifs. La différence de cette approche avec l'approche hiérarchique est qu'il n'y a aucune hiérarchie entre les IDS distribués ce qui signifie que l'échec de n'importe quel IDS n'empêche pas la détection d'attaques coordonnées. Parmi les systèmes de détection d'intrusions coopératifs, nous pouvons citer par exemple le système CSM **[16] [18]**et le système AAFID **[04].**

8. Une Vue Générale De Quelques Systèmes De Détection D'intrusions Existants

Il existe plusieurs systèmes de détection d'intrusions qui ont été développés. Dans cette section, nous présenterons quelques systèmes de détection d'intrusions existants.

IDES

IDES (Intrusion-Detection Expert System) a été développé par SRI International. Il représente le modèle de référence pour un grand nombre de systèmes de détection d'intrusions. Il a été conçu pour surveiller un seul hôte et il traite uniquement les données d'audit. Ce système de détection d'intrusions est indépendant du système surveillé, il fonctionne sur une machine dédiée, reliée au système par un réseau. Afin de détecter les violations de sécurité en temps réel, IDES s'appuie aussi bien sur une approche statistique que sur un système expert **[20], [18]**. Ainsi, il est constitué de deux éléments importants :

- **Le détecteur d'anomalie** : qui est responsable de la détection des comportements atypiques, en utilisant des méthodes statistiques du modèle de Denning **.**
- **Le système expert** : qui est chargé de détecter les attaques suspectes en s'appuyant sur une base de connaissances de scénarios d'attaques connus.

NIDES

NIDES (Next- Generation IDES **[18]** est une version améliorée du système de détection d'intrusions IDES. Il assure la détection d'intrusions sur plusieurs hôtes (distribuées) en se basant toujours sur les données d'audit. Il n'y a aucune analyse du trafic réseau. Il utilise les mêmes algorithmes qu'IDES.

NADIR

NADIR (Network Anomaly Detection and Intrusion Reporter **[18]**, **[04]** est un système expert qui a été conçu pour le réseau ICN (Integrated Computing Network) du Laboratoire National Los Alamos. Son but est d'analyser les activités réseaux des utilisateurs et d 'ICN en se basant sur les règles du système expert qui définissent la politique de sécurité et les comportements suspects. L'inconvénient majeur de ce système est qu'il ne peut être porté sur d'autres réseaux, étant donné que les protocoles réseaux d'ICN ne sont pas standards.

DIDS

DIDS (Distributed Intrusion Detection System)**, [18]**, **[04]** est un système de détection d'intrusions basé réseau qui se base sur l'approche hiérarchique. Afin d'éviter la dégradation des performances de système, DIDS délègue certaines analyses locales aux hôtes locaux. Son architecture se compose de trois entités :

- **Le « Host Monitor »** : Il en existe un par hôte. Il collecte les données de l 'hôte surveillé, fait une première analyse simple sur ces données puis transmet les événements pertinents au « DIDS Director ».
- **Le « LAN Monitor »** : Il en existe un pour chaque segment LAN. Il surveille le trafic sur le LAN, collecte les informations réseaux et reporte au « DIDS Director » les activités suspectes et non autorisées qui se sont produites sur le réseau.
- **Le « DIDS Director »** : Il analyse les rapports reçus du « LAN Monitor » et des « Host Monitor » afin de détecter les attaques potentielles.

GrIDS

GrIDS (Graph-Based Intrusion Detection System) **[15] [16]** a été conçu pour détecter des attaques à grande échelle. GrIDS considère les réseaux larges comme une agrégation de sous réseaux. Les données concernant l'activité des hôtes et le trafic réseau entre ces hôtes sont rassemblées dans des graphes d'activité qui révèlent la structure causale de l'activité réseau.

Les nœuds d'un graphe d'activité correspondent aux hôtes constituant le réseau alors que les arêtes représentent l'activité réseau entre les différents hôtes.

Durant la phase de détection, GrIDS analyse les caractéristiques des graphes d'activité et compare ces graphes à des formes intrusives connues. S'il y a des similitudes entre ces graphes et des attaques connues, il en informe l'officier de sécurité.

CSM

CSM (Cooperating Security Manager) est un système de détection d'intrusions qui peut être utilisé dans un environnement de réseau distribué. Son principal objectif est de détecter les activités intrusives de façon non centralisée car utiliser un directeur central qui coordonnerait toutes les activités limiterait la taille du réseau « le problème d'incrémentabilité ». Pour cela, CSM doit s'exécuter sur chaque hôte connecté au réseau. Ainsi, au lieu de reporter les activités anormales à un directeur central, les CSM communiquent entre eux pour détecter d 'une manière coopérative les intrusions réseaux. Les composants principaux de ce système de détection d'intrusions sont :

- Un système de détection d'intrusions local (IDS) : qui assure la détection d'intrusions pour un hôte local.
- Un gestionnaire de sécurité : qui coordonne la détection d'intrusions distribuée entreles CSM.
- Un gestionnaire d'intrus (IH : intruder handling component) : dont le rôle est d'entreprendre les actions nécessaires lorsqu'une intrusion est détectée.

AAFID

Le système AAFID (Autonomous Agent for Intrusion Detection) **[04]** est la première tentative d'utilisation des agents autonomes pour les systèmes de détection d'intrusions basés réseau où plusieurs agents indépendants opèrent de manière coopérative pour assurer la surveillance du système cible. La décision finale du système est le résultat de coopération entre ces différents processus.

Les recherches actuelles visent à améliorer les systèmes de détection d'intrusions en raison de la complexité croissante des environnements à protéger, qui sont de plus en plus larges et dynamiques. Ainsi, la nature des intrusions actuelles et futures nous incite à développer des outils adaptatifs et automatiques. Une solution prometteuse consiste à s'inspirer à partir des métaphores biologiques pour résoudre ces problèmes. Cela est réalisé via l'exploitation des concepts et des méthodes d'identification et de détection du système immunitaire humain, qui est capable d'assurer la protection du corps contre les différents intrus d'une manière robuste, autonome, distribuée et adaptative. Alors, pourquoi de ne pas concevoir des systèmes immunitaires afin de protéger les systèmes et les réseaux informatiques.

Le système immunitaire constitue un intérêt croissant des recherches vu de sa capacité de traitement des informations. En particulier, il assure des calculs d'une manière distribuée et parallèle. Il peut apprendre des nouvelles informations et identifier les différents modèles d'une manière

décentralisée. Il détecte et répond aux envahisseurs étrangers d'une façon distribuée. L'approche immunologique est une solution prometteuse pour la détection d'anomalies vue l'analogie puissante qui existe entre l'objectif du système immunitaire humain et celui du système de détection d'intrusions ainsi que la capacité du système immunitaire humain à protéger le corps contre les intrus. Ce système présente un intérêt croissant des différents travaux existants pour exploiter ces méthodes d 'identification et de détection dans des systèmes de détection d'intrusions.Pour cette raison, dans ce travail nous nous intéresserons par ce domaine de recherche.

9. Emplacement de l'IDS

Le choix de l'IDS est très influencé par son éventuel emplacement au sein du réseau. En effet, la topologie réseau impose quelques règles à respecter si on veut l'IDS soit efficace.

L'emplacement de l'IDS doit également tenir compte du type d'intrusions à détecter (internes, externes, les deux). Si dans un réseau, il existe un seul point de connexion à internet, le meilleur emplacement de l'IDS est qu'il soit juste après le routeur. Si dans un autre réseau, différents points de connexions à internet existent, un IDS est placé pour chaque point de connexion comme nous pouvons le voir dans la **Fig. II**.5. Par contre pour détecter les intrusions internes un IDS doit être placé à chaque segment du réseau. **[14]**

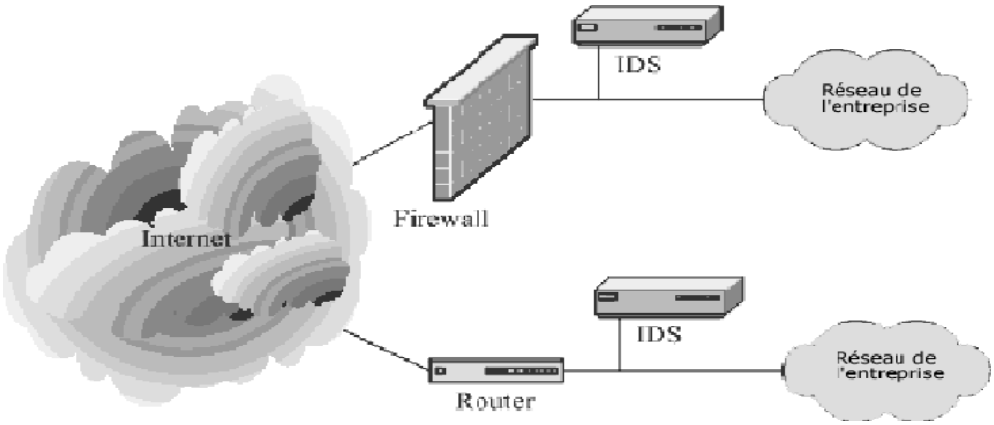

Fig. II.5. Emplacement de l'IDS au sein d'un réseau.

10. Evaluation d'un IDS

Des mesures permettent de comparer et de mesurer l'efficacité des IDS. Les IDS sont des éléments très importants dans une stratégie de sécurité, pour cela le choix de l'IDS est très décisif et doit être basé sur les caractéristiques de ce dernier, les tâches qu'il devra accomplir, son emplacement …etc. mais également selon des mesures qui permettent d'évaluer son efficacité. La technique la plus utilisée pour évaluer un IDS est le scanner de vulnérabilités. Cependant, ce dernier reste peu fiable.

Les mesures permettant de mieux choisir son IDS et mesurer son efficacité sont :**[09]**

- Qualité des informations fournies par l'IDS : le taux de faut positif.
- Réponse de l'IDS dans un environnement surchargé.
- La possibilité de mettre à jour la base des signatures ou de modifier certaines signatures.
- La séparabilité des fonctions d'administration (architecture distribuée).…etc.

Les IDS à leurs tours peuvent faire l'objet d'attaques et certaines de leurs faiblesses sont liées au système d'exploitation (saturation de la mémoire ou de la carte réseau) **[07].**

11. Conclusion :

Dans ce chapitre, nous avons présenté le système de détection d 'intrusions et nous avons également étudié d'une manière détaillée les différents types d'IDS selon différents critères de classification avec la présentation générale des différentes techniques utilisées pour la détection d'intrusions.

Afin d'obtenir un système de détection d'intrusions compétent et efficace, il est souhaitable d'utiliser les deux techniques de détection comportementale et basée connaissances en parallèle pour surmonter les problèmes liés à chacune de ces deux techniques de détection. Cependant, les systèmes de détection d'intrusions commercialisés emploient seulement la technique de détection basée connaissance, ce qui motive les différents efforts de recherche dans le domaine de la détection d'anomalies.

Pour cette raison, différentes approches sont utilisées pour implémenter la technique de la détection d'anomalies. Parmi ces approches deerses, nous intéresserons.

Chapitre III :

Implémentaion et Résultat

1. Introduction
2. Les Abeilles socieles
3. le modele informatique :
4. Corpus utilisé
5. Outils d'evaluation
6. Résultat et discussion.
7. Conclusion

Introduction

Le biomimétisme est une approche scientifique révolutionnaire qui consiste à imiter les plus belles inventions de la nature (la capacité énergétique de la photosynthèse, solidité du corail, résistance du fil de soie de l'araignée. . .) pour les adapter au service de l'homme dans la nature, plusieurs espèces sont caractérisées par le comportement social.Les bancs de poissons, les nuées d'oiseaux, et les troupeaux d'animaux terrestres, sont le résultat du besoin biologique qui leur pousse à vivre en groupe.Ce comportement est également un des principaux caractéristiques des insectes sociales (abeilles, termites, fourmis. . .). De ces principes là, les chercheurs se sont inspirés pour développer des méthodes basées sur les comportements de ces ani-maux, et ont donné naissance à ce que l'on appelle par métaheuristique.

Ce mot concerne toutes les méthodes qui modélisent l'interaction des agents (ani-maux) qui sont en mesure de s'auto-organiser. Elles représentent des méthodes de résolution de problème combinatoires qui consistent à réitérer certains processus jusqu'à obtenir la solution optimale.
L'une des insectes les plus organisées et les plus rigoureuses dans leur travails est l'abeille.
Les abeilles possèdent une très grande capacité de communication. Et grâce à sont intelligence, une méthode appelée méthode des abeilles a été déve-lopper. Dans cette méthode, les abeilles artificielles représentent des agents qui en collaborant les une avec les autres, résolvent des problèmes complexes d'optimisa-tion combinatoire.

Les abeilles sociales

Comme les fourmis, les abeilles sont des insectes sociaux. Elles sont obligées de vivre en colonie très organisée, formée d'ouvrières, de faux-bourdon et d'une seule reine, et où chacune a un travail bien précis à faire.
Les abeilles se nourrissent essentiellement de pollen et de miel. Elles vont butiner les fleurs pour prendre le nectar.
Au cours de sa courte vie (environ 45 jours), l'ouvrière fait plusieurs métiers : elle nettoie les cellules, nourrit les larves, elle range le pollen et le nectar dans les alvéoles, elle ventile la ruche en agitant rapidement ses ailes, elle construit les rayons avec la cire qu'elle produit, elle garde le trou de vol pour chasser les intrus, elle devient butineuse, porteuse d'eau et récolte du pollen et du nectar jusqu'à la fin de sa vie.
L'abeille est capable, par la danse ou par la production de substances chimiques appelées « phéromone », de communiquer aux autre abeilles l'endroit où elle a découvert de la nourriture. Elle danse en rond quand elle a trouvé du pollen à faible distance (moins de 25 mètres).

Elle utilise une danse très compli-quée dite la danse frétillante (figure 8), ou danse en huit, si la nourriture se trouve moins de 10 kilomètres. La direction de la nourriture est exprimée par rapport la position du soleil. La distance est exprimée par le nombre et la vitesse des tours effectués par l'abeille sur elle-même. Afin de survivre à l'hiver, les abeilles doivent recueillir et stocker environ 15 à 50 Kg de nectar.

Les faux bourdons ne servent que pour la reproduction. Ils sont incapables de se nourrir eux-mêmes (les ouvrières les nourrissent) et ils n'ont pas de dard pour protéger la ruche. protection de ruche des intrus par l'abeille sociale afin de développer un Système de Détection d'Intrusion que nous avons appelé « IDSBees **»[21].**

— Durant les trois premiers jours, elle joue le rôle de nettoyeuse et veille à la propreté des cellules. Sa deuxième mission est celle de nourricière, elle distribue la gelée royale à toutes les larves, qui don-neront naissance aux jeunes abeilles, et aux reines et ce jusqu'au environ du dixième jour suivant sa naissance **[20].**

— Du 11ième au 20ième jours, les abeilles exécutentdes travaux de nettoyage, débarrassent la ruche des détritus, des cadavres de leurs sœurs. Elles vont aussi à la rencontre des butineuses rentrantes pour les décharger du nectar récolté en le disposant dans les alvéoles, et de s'occuper également du pollen ramené par leurs compagnes **[20].**

— Pendant la troisième phase les ouvrières magasinières procèdent encore à la construction des cel-lules de miel de réserve et celles des nymphes .

— Du 18ième au 21ième jour, elles deviennent les gardiennes en prenant part à la défense de la ruche et montent la garde au trou de vol à l'affût des pillards comme des bourdons, guêpes ou abeilles de ruches voisines. Elles communiquent grâce à ses antennes avec les abeilles qui entrent dans la ruche. Celles qui ne font pas partie de la colonie sont repoussées. Les voleuses de miel sont chassées à coup de dard .

La collecte est la dernière et plus longue tâche d'une ouvrière s'étalant du 21ème jour jusqu'à sa mort ou Elle part récolter le pollen et le nectar des fleurs pour la production de miel

Durant la quatrième phase de leurs vies, les abeilles ouvrières servent de gardes à l'entrée de la ruche en interdisant l'accès aux intrus. A ce stade, un changement physique se produit au niveau du corps de l'abeille ou leurs glandes à venin se développent et commencent à produire du venin **[20].**

Signalons que les gardiennes ne quittent jamais leurs postes, même dans le cas d'une défense offen-sive qui se réalise sur une autre entrée, les gardiennes des entrées ne rejoindront pas l'essaim car elles ne peuvent pas laisser une entrée sans surveillance ce qui sera une vulnérabilité et pourra être détectée et utilisée par l'intrus **[22].**

Toutes les abeilles se ressemblent énormément, pourtant les abeilles étrangères qui entrent dans la ruche sont immédiatement identifiées. Les scientifiques qui ont étudié la question à savoir comment les abeilles accomplissent cela en sont arrivés à de surprenantes conclusions : L'odeur de la ruche est le plus important facteur qui permet aux abeilles de se reconnaître et de se distinguer les unes des autres. Celles qui n'ont pas l'odeur distinctive de la ruche représentent donc un danger et sont immédiatement expulsées ou tuées par les abeilles sentinelles les sentinelles font preuve d'une réaction énergique, en utilisant leurs aiguillons contre toute créa-ture perçue comme n'appartenant pas à la ruche.

Le Modèle Informatique

Avant de détailler l'approche de notre contribution, nous devons d'abord décrire le modèle naturel du fonctionnement du système anti-intrusion .

Dans les approches conventionnelles, la construction du modèle d'intrusion se faisait à partir d'un Training set généralisé englobant toutes les connexions (intrusion et non-intrusion) sur tous les ports du réseau. Générant ainsi un seul modèle d'apprentissage d'intrusion qui peut contenir un déséquilibre entre les ports, réduisant ainsi la sécurité et la fiabilité du système de détection d'intrusion puisqu'il ne prend pas en considération toutes les intrusions possibles sur tous les ports du réseau. nous avez partitionner KDD en 4 base d'apprentissage, chaque base d'apprentissage correspond à une attaque bien précice, car dans KDD il y a 4 type d'attaque)

État initial

Appliquer sur chaque sous base d'apprentissage spécialisée relative à un port du réseau algorithme de classification : Naives Bayes

Algorithme Naïve bayse

La **classification naïve bayésienne** est un type de classification bayésienne probabiliste simple basée sur le théorème de Bayes avec une forte indépendance (dite naïve) des hypothèses. Elle met en œuvre un classifieur bayésien naïf, ou classifieur naïf de Bayes, appartenant à la famille des classifieurs linéaires.**[19]**

Un terme plus approprié pour le modèle probabiliste sous-jacent pourrait être « modèle à caractéristiques statistiquement indépendantes ».**[18]**

En termes simples, un classifieur bayésien naïf suppose que l'existence d'une caractéristique

pour une classe, est indépendante de l'existence d'autres caractéristiques. Un fruit peut être considéré comme une pomme s'il est rouge, arrondi **[19]** et fait une dizaine de centimètres.

Même si ces caractéristiques sont liées dans la réalité, un classifieur bayésien naïf déterminera que le fruit est une pomme en considérant indépendamment ces caractéristiques de couleur, de forme et de taille.

1. **Former la base de test KDD.**

Faire l'évaluation du modèle d'apprentissage cité au point 3 avec la base de test cite au point 4.

Corpus utilisé

Depuis 1999, KDD'99 est le corpus le plus utilisé pour l'évaluation des anomalies et la détection d'in-trusion (Benchmark). Ce corpus a été construit par Stolfo **[22]**, à partir de la base des données collectées par le programme d'évaluation des systèmes de détection d'intrusion DARPA'98. La taille de DARPA'98 est d'environ 4 gigaoctets de (binaires) relevés sur 7 semaines

de trafic réseau, constitué d'en-viron 5 millions d'enregistrements de connexion d'environ 100 octets chacune.

Le KDD'99 est constitué d'un ensemble de données d'environ 4.900.000 vecteurs de connexion unique dont chacune contient 41 colonnes dont une est étiqueté comme normal ou une attaque, avec exactement un type d'attaque spécifique.

Les attaques peuvent être classées dans l'une des quatre catégories suivantes :

1. Attaques par déni de service (DoS).
2. Attaque U2R.
3. Attaque R2L.
4. Attaque probing.

Il est important de noter que les données de la base de test incluant des types d'attaques spécifiques qui ne figure pas dans le training set. Certains experts d'intrusion croient que la plupart des nouvelles attaques sont des variantes d'attaques connues. Le training set contient 24 types d'attaque, avec 14 types supplémentaires dans la base de test. Le nom et description détaillée des types d'attaque sont répertoriés dans le travail de chercheur Lippmann**[17]**.

Les fonctionnalités de KDD99 peuvent être classées en trois groupes :

— Caractéristiques de base : cette catégorie englobe tous les attributs qui peuvent être extraites à partir d'une connexion TCP / IP.

— Caractéristique de trafique de réseau : cette catégorie inclut des fonctionnalités qui sont

calculées par rapport à un intervalle de fenêtre et est divisé en deux groupes :

— caractéristiques "même hôte"

— caractéristiques "même service"

— Caractéristique de contenu.

Outils d'évaluation

La matrice de confusion, dans la terminologie de l'apprentissage supervisé, est un outil servant à mesurer la qualité d'un système de classification (Voir Tableau). Chaque colonne de la matrice re-présente le nombre d'occurrences d'une classe estimée, tandis que chaque ligne représente le nombre d'occurrences d'une classe réelle (ou de référence). Un des intérêts de la matrice de confusion est qu'elle montre rapidement si le système parvient à classifier correctement, c'est-à-dire lorsque les VP et VN doivent être maximales alors automatiquement les FN et FP sont minimales.

	Classé comme intrusion	Classé comme non-intrusion
Réellement Intrusion	VP	FN
Réellement Non-Intrusion	FP	VN

Tableau III 1 – Matrice de confusion (IDS)

De cette matrice nous calculerons toutes les métriques d'évaluation suivante : Rappel, Précision, F-mesure, Kappa statistique, Entropie et Précision dont les formules sont citées .

L'editeur Netbeans de langage java

NetBeans est un projet open source ayant un succès et une base d'utilisa-teur très large, une communauté en croisance constante, et près 100 partenaires mondiaux et des centaines de milliers d'utilisateur à travers le monde. Sun Micro-systems a fondé le projet open source NetBeans en Juin 2000 et continue d'être le sponsor principal du projet.**[22]**

Fonctionnement

FIG III 1 Interface de Menu

Nous disposons de deux ensembles de données KDD99 : l'ensemble d'apprentissage et l'ensemble de test dont nous trouvons deux types de données : données normales et les données représentant des attaques. Dans cette section nous allons présenter les différents modules et la participation des différents ensembles de données à la construction des différents modèles de détection d'intrusion.

6-Résultat et discussion :

- Dans notre travaille on a prendre la base d'apprentissage KDD 1999 avec un pourcentage de 20% (25192 instances) , et on a prendre la base de test KDD 1999 complet (22544 instances). le table si dessous 7 contient les résultats de notre modèle avec différents nombre d'attributs de la KDD .**[19]**

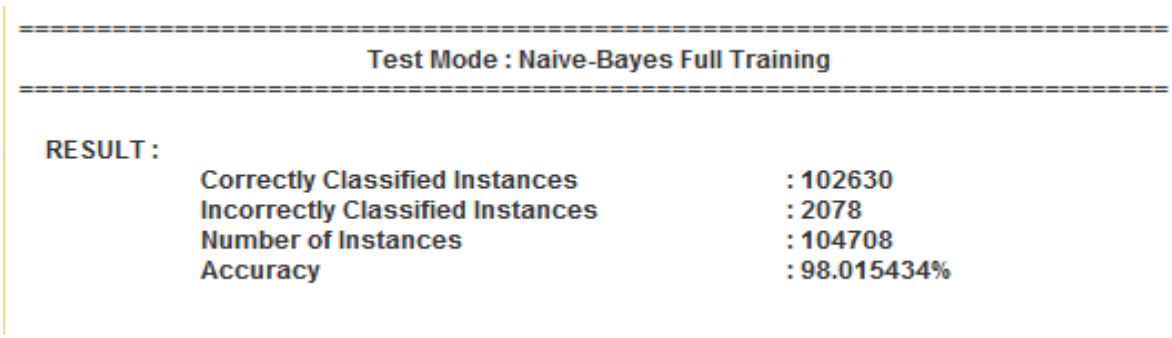

FIG III 2 les résultats de la détection d'intrusion avec notre modèle I

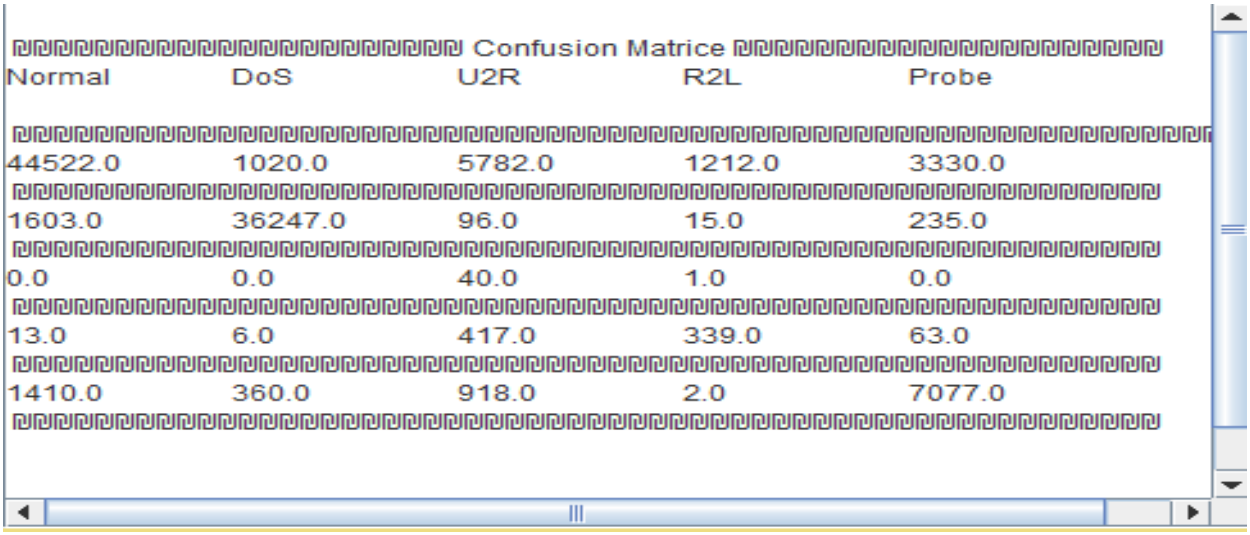

FIG III 3 Confusion de Matrice

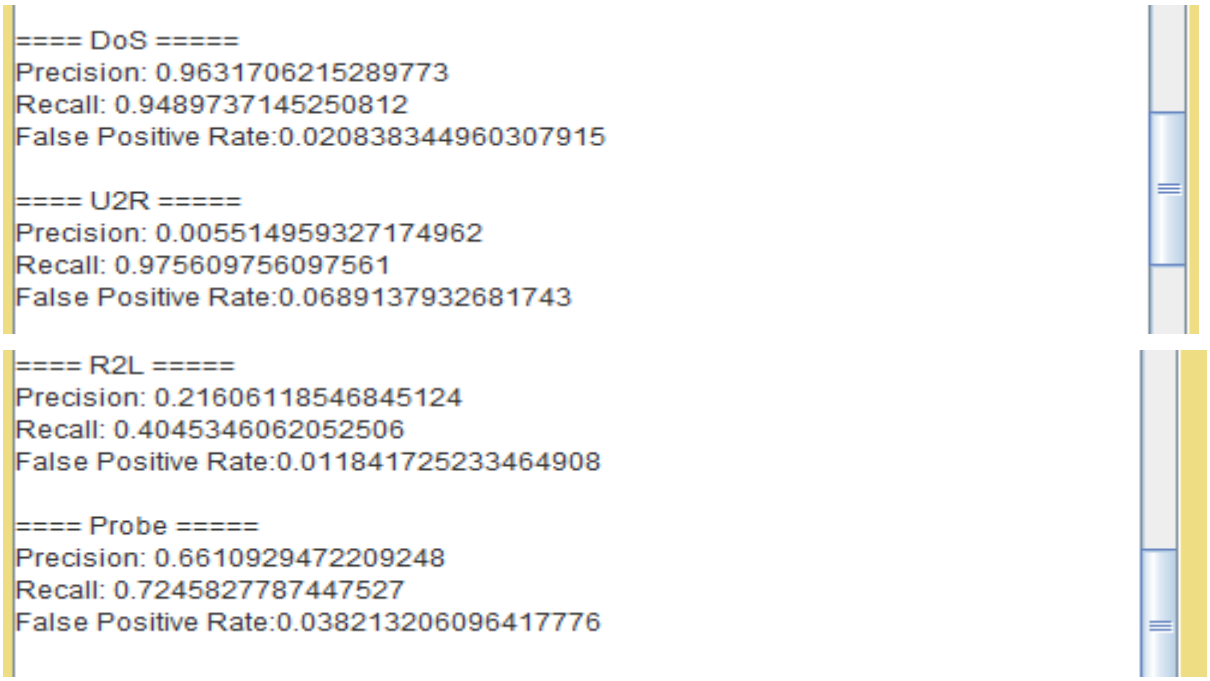

FIG III 4 les résultats de la détection d'intrusion avec notre modèle II

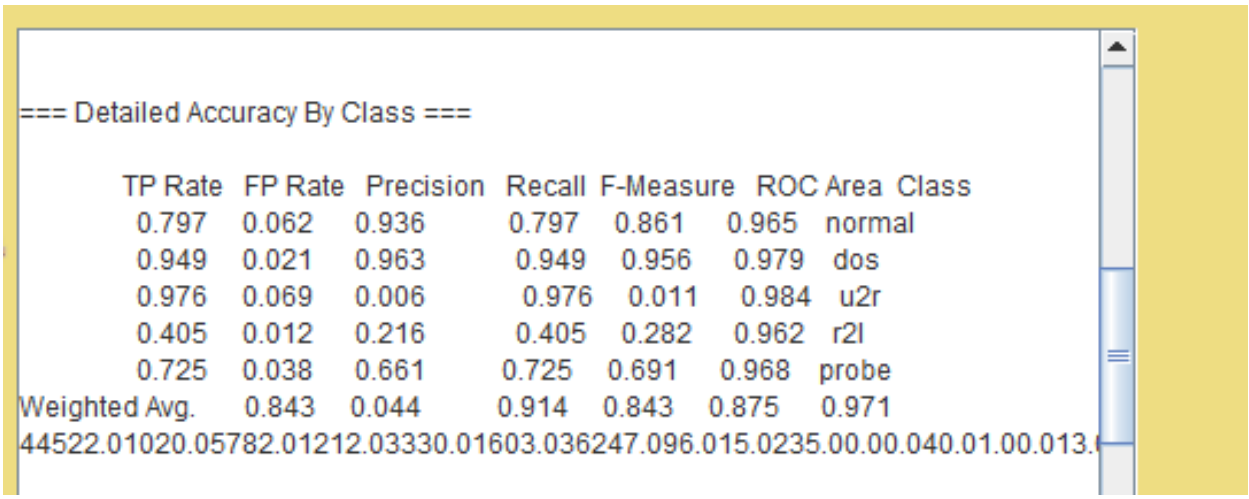

FIG III 5 les résultats de la détection d'intrusion avec notre modèle III

Avec l'addition a chaque fois le nombre de colonnes de la kdd on voie que la precision s'augmente progressivement de 5 colonnes à 20 colonnes, et à partir de 20 colonnes elle commence a se stabiliser avec un pourcentage de 90.5% un peut prêt .

Pour le rappel , avec l'addition a chaque fois le nombre de colonnes de la kdd on voie qu'il commence a diminué progressivement de 5 à 20 colonnes , et à partir de 20 colonnes elle commence a se stabiliser avec un pourcentage de 61% un peut

L'ensemble de données KDD99 (apprentissage et test) est fournit sous forme de deux fichiers texte.

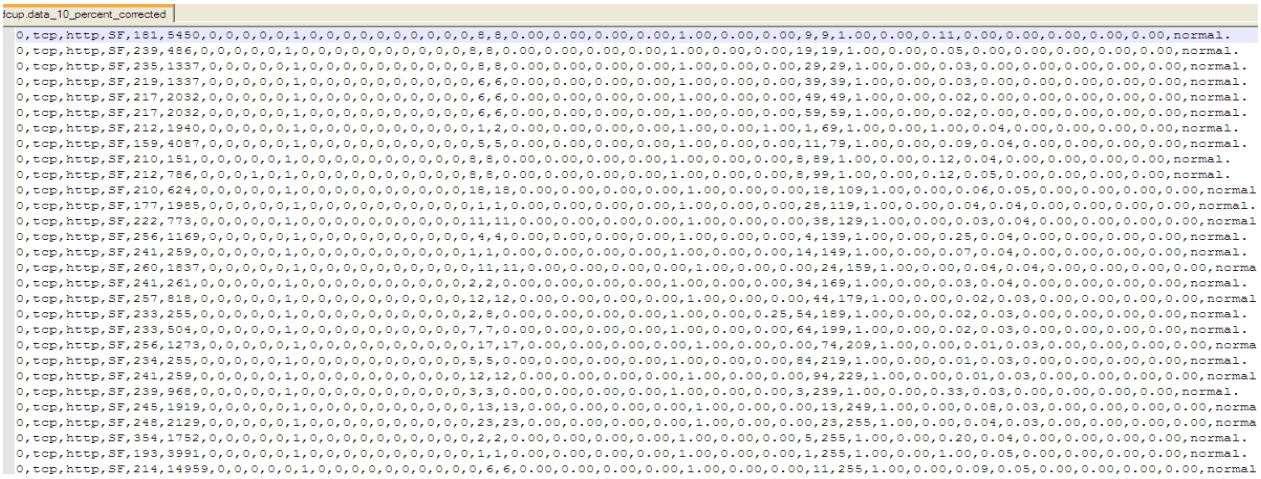

FIG III.6 : base de données KDD99 fichier texte

Numérisation de données

Cette étape consiste à convertir les attributs symboliques en numériques. La procédure est comme suit :

En premier, Les noms d'attaques (comme buffer_overflow, guess_passwd, etc.) sont rangées dans une des cinq classes : 0 pour Normal, 1 pour Probe, 2 pour DOS, 3 pour U2R et 4 pour R2L **[12]**. Les attributs symboliques comme protocol_type (3 symboles différents), service (70 symboles différents) et flag (11 symboles différents) sont transformés à des valeurs entières de 0 à N-1 où N est le nombre de symboles

- **Les propriétés concernant le pc de traitement**

- notre traitement est fait dans un PC portable de marque HP , avec **un pro-cesseur** : Intel (R) Core (TM) i3-3230M CPU 2.60GHz 2.60GHz , **une Mémoire installé (RAM)** : 4 Go (3.88 Go utilisable) , **un système d'exploitation** 64 bits , **processeur** x64 , **Memoire graphique total disponible** : 5571 Mo .

Comparaison :

Pour démontrer l'efficacité de notre approche, nous l'avons comparé avec des approches similaires en termes de détection des attaques connues, nouvelles attaques et de faux positifs. Le tableau suivant résume les résultats.

D'après le tableau, il ressort clairement que notre approche est très compétitive avec les solutions proposées dans la littérature.

En effet, à part quelques approches **[23][12]** qui sont efficaces puisqu'ils ont atteint un taux de détection de 100%, toutes les autres méthodes ont été dépassées sur le taux de détections des nouvelles attaques et de faux positifs, le taux global de l'approche **[23]** basé aussi sur les Abeilles avec l'algorithme Naive bayse.

L'efficacité de notre approche basée sur les abeilles et les règles d'associations a été démontrée et surpasse presque toutes les autres méthodes

Méthode	Taux de détection						Fausses alarmes
	Tout	Nouvelles attaques					
	Globale	Globale	U2R	PROB	DOS	R2L	
Notre approche	94.05%	89.39%	96.67%	99.58%	99.61%	64.73%	1.92%
[23]	92.30%	-	60.96%	10.06%	0.36%	2.20%	0.57%

Tableau III 2: Approche[23] vs Notre Approche

Conclusion

Vu les évolutions actuelles, les systèmes informatiques vont vraisemblablement continuer à s'immiscer d'avantage dans notre quotidien. Le développement de ces systèmes s'accompagne automatiquement de nombreux défis technologiques tels que la mobilité, l'autonomie,et la réactIIIité, etc. parmi ces défis, la question cruciale de la sécurité demeure un enjeu majeur en raison de la dématérialisation croissante de l'information et des risques de plus en plus importants liés à la complexité de ces systèmes. Cette thèse traite la sécurité des systèmes informatiques. Les contributions que nous avons réalisées se focalisent principalement sur la protection de ces systèmes par la détection d'attaques informatiques, qui suscitent actuellement un intérêt croissant autant dans le monde académique que professionnel et surtout industriel en raison de la propagation de ces menaces informatiques représentées dans les intrusions, qui sont notre problématique, et qui se font aujourd'hui de plus en plus oppressantes. Donc la mise en place d'une bonne stratégie de défense contre les intrusions passe par l'amélioration des techniques de détections des IDS. Pour cela, il faut améliorer les algorithmes de détection, afin qu'ils puissent traiter
les données de manière efficace.

Conclusion générale

Conclusion générale

Les contributions que nous avons réalisées se focalisent principalement sur la protection de ces systèmes par la détection d'attaques informatiques, qui suscitent actuellement un intérêt croissant autant dans le monde académique que professionnel et surtout industriel en raison de la propagation de ces menaces informatiques représentées dans les intrusions, qui sont notre problématique, et qui se font aujourd'hui de plus en plus oppressantes. Donc la mise en place d'une bonne stratégie de défense contre les intrusions passe par l'amélioration des techniques de détections des IDS. Pour cela, il faut améliorer les algorithmes de détection, afin qu'ils puissent traiter les données de manière efficace. Nous avons entamé nos contributions en proposant un nouveau modèle artificiel calqué du monde naturel qu'est le monde des abeilles (Le bio mimétisme), par l'utilisation des techniques du data mining, sur les quelles nous nous sommes basés pour identifier les intrus. Pour cerner cette problématique complexe, nous avons proposé une idée novatrice qui a dominé nos contributions en se basant sur le fait que les modèles des approaches traditionnelles des IDS sont construits à partir d'un ensemble d'apprentissage générale de tout le réseau et qui génère un seul modèle d'apprentissage d'intrusion. Ce qui nous a amené proposer une nouvelle approche plus spécifique et plus décentralisée, en prenant aléatoirement le même nombre de connexions, pour les bases d'apprentissage et les bases de Test, correspondant au nombre de ports logique du réseau.
Par conséquent, au lieu d'avoir un seul modèle d'apprentissage (approche traditionnelle) nous aurons ainsi comme nombre de modèles d'apprentissage différents le nombre des ports logiques du réseau, ce qui maximise la sécurité et la robustesse du réseau.
Nous prévoyons dans un avenir proche de finaliser les deux sous autres approches, en l'occurrence :
— Reconnaissance de comportement dans la ruche : détection d'intrusion par comportement (qui fait actuellement l'objet d'un article)
— Stratégie de réponse a une intrusion : la réponse actIIIe (défense offensIIIe). Ceci pour donner plus d'ampleur à notre système de détection d'intrusion et pour qu'il soit complet et pourquoi pas opérationnel.
Comme, nous prévoyons d'apporter certaines améliorations en fonction des résultats en s'appuyant sur les points forts et essayer de corriger les faiblesses maximales de notre système basé sur

Bibliographie

Bibliographie

[01] U. Aickelin & J. Greensmith & J. Twycross « Immune System Approaches toIntrusion Detection - A review », School of Computer Science, UnIIIersity of Nottingham, 2004

[02] Axelsson. S.: « Intrusion Detection Systems: A Taxonomy and Survey ». Technical Report No 99-15, Dept. of Computer Engineering, Chalmers UnIIIersity of Technology, Sweden, 2000,

[03] Rebecca Bace and Peter Mell. Intrusion Detection Systems. NIST Special Publication on Intrusion Detection Systems ,2000.

[04] J. S. Balasubramaniyan & J. O. Garcia-Fernadez & D. Isacoff & E. H. Spafford & D. Zamboni. « An Architecture for Intrusion Detection using Autonomous Agents
». Technical Report Coast-TR-98-05, Computer Sciences Department, Purdue UnIIIersity, 1998.

[05] J. Balthrop & S. Forrest & M. Glickman. « Revisiting lysis: Parameters and normal behaviour ». Proceedings of the Congress on Evolutionary Computation, pages 1045- 1050, 2002.

[06] Hubert GUERRIAT et Michel ITTELET. “Aperçu sur le statut du Milan noir (Milvus migrans) en Belgique”. In : *Aves* 19.3 (1982), p. 183–191.

[07] Statistiques communiqués par la BSA (Business Software Aliance), 2002.

[08] International Standards Organization. Information Processing Systems - OSI – Basic Reference Model - Part 2: Security Architecture. ISO 7498-2, February 2000.

[09] David BURGERMEISTER, Jonathan KRIER, Système de détection d’intrusion, 2006. (http://dbprog.developpez.com);

[10] Pr Manuel CASTELLS, Le développement d'Internet, qui était exponentiel, trouve actuellement sa limite, Journal le monde, 2000.

[11] Jackson. K & DuBois. D & Stallings. C « The NIDES Statistical Component Description and Justification » Technical Report, Computer Science Laboratory, SRIInternational, Menlo Park, CA, March, 1994.

[12] W. Jansen & P. Mell, T.Karygiannis & D.Marks «Mobile Agents in Intrusion Detection And Response », 2000

[13] KDD'99 datasets The UCI KDD ArchIIIe, Irvine CA, USA, 1999 http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.htm

[14] J. Kim « Intrusion Detection », PhDThesis, UnIIIersity College London, 2002.

[15] C. Kruegel & T. Toth, « Applying Mobile Agent Technology to Intrusion Detection », Technical UnIIIersity of Vienna, Distributed Systems group, 2002.

[16] LABED Ines « la détection d'intrusions » thèse de magister informatique unIIIersité mentouri de Constantine 2006.

[17] Li. Y& Wu. N & Jajosia. S & Sean Wang. X « Enhancing Profiles for Anomaly Detection Using Time Granularities » Proc. 1st ACM workshop on Intrusion Detection Systems, Athens, Greece, Nov. 2000.

[18] Mykerjee. B & Heberlein. L.T & Levitt .K.N « Network Intrusion Detection », IEEE Network, Vol 8, No 3, pp .26-41, 1994.

[19] Thèse LOKBANI Le problème de sécurité par le Data Mining\Ahmed Chaouki LOKBANI 2017

[20] Thèse BOUDIA : « Optimisation, intégration des données et découverte de connaissances à partir des données du web »

[21] Lokbani, A. C., Lehireche, A., & Hamou, R. M. (2013, June). Experimentation of Data Mining Technique for System's Security: A Comparative Study. In International Conference in Swarm Intelligence (pp. 248-257). Springer, Berlin, Heidelberg.

[22] Lokbani, A. C., Lehireche, A., Hamou, R. M., & Amine, A. (2014). Synthesis of Supervised Approaches for Intrusion Detection Systems. In Network Security Technologies: Design and Applications (pp. 44-57). IGI Global.

[23] Lee, J. H., Lee, J. H., Sohn, S. G., Ryu, J. H., & Chung, T. M. (2008, February). Effective value of decision tree with KDD 99 intrusion detection datasets for intrusion detection system. In Advanced Communication Technology, 2008. ICACT 2008. 10th International Conference on (Vol. 2, pp. 1170-1175). IEEE. Cité 22 fois