République Algérienne Démocratique et Populaire

Université Saad Dahlab Blida

Faculté des Sciences

Département d'Informatique

Mémoire de fin d'études

Pour l'obtention du diplôme de Master en Informatique

Option: génie logiciel

Gestion de la sécurité d'une application Web à l'aide d'un IDS comportemental optimisée par l'algorithme des K-means

Réalisé par: Mohammed Mahmoud Yousef

Sous la supervision de: Mr Djenouri youcef

l'encadreur: Mr Djahra walid

MA-004-286-1

République Algérienne Démocratique et Populaire

Université Saad Dahlab Blida

Faculté des Sciences

Département d'Informatique

Mémoire de fin d'études

Pour l'obtention du diplôme de Master en Informatique

Option: génie logiciel

Gestion de la sécurité d'une application Web à l'aide d'un IDS comportemental optimisée par l'algorithme des K-means

Réalisé par: Mohammed Mahmoud Yousef

Sous la supervision de: Mr Djenouri youcef

l'encadreur: Mr Djahra walid

REMERCIEMENTS

Avec un grand plaisir je remercie Allah qui m'a aidé et m'a donné la patience, le courage et la force d'achever ce travail.
Je tiens à remercier en cette occasion tout le corps professoral et administratif de département d'informatique de l'université SAAD DAHLEB de Blida pour la richesse et la qualité de leurs enseignements et qui déploient de grands efforts pour assurer à leurs étudiants une formation actualisée.
Je tiens à remercier sincèrement Mr Youcef Djenouri, qui, en tant que promoteur de mémoire, s'est toujours montré à l'écoute et très disponible tout au long de la réalisation de ce mémoire, ainsi pour l'orientation, la confiance, l'aide et le temps qu'il a bien voulu me consacrer et sans lui ce mémoire n'aurait jamais vu le jour.
J'exprime également ma gratitude aux membres du jury, qui m'ont honoré en acceptant de juger ce modeste travail.
Je tiens à remercier sincèrement mes parents et mes amis, qui m'ont donné le courage.
Je souhaite d'adresser mes remerciements les plus sincères aux personnes qui m'ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire.

Dédicace

Je dédie ce mémoire à

Mes parents :

Ma mère, qui a œuvré pour ma réussite, de par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude.

Mon père, qui peut être fier et trouver ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie. Puisse Dieu faire en sorte que ce travail porte son fruit ; Merci pour les valeurs nobles, l'éducation et le soutient permanent venu de toi.

A mon très chères frère Othmane et mes sœurs Hanane, Noor el Houda et Ikrame, qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de générosité.

A mon Encadreur Mr Djahra qui m'a beaucoup aidé durant cette année, Il nous a donnez tous les moyens pour réussir dans ce stage, Ainsi les ingénieurs Oussama, Khaled et Yassine qu'il n'a épargné aucun effort pour nous aider.

A mes amies Abdou, Issam, Nacer, Amine, Anis, Walid, Mokran, et Salim qui m'ont toujours entourés par leur amour et amitié, d'être toujours à mes cotés, de partager avec moi des moments de joie et d'humeur, de me soutenir au moment les plus difficiles, les mots ne suffisent guère pour exprimer l'attachement, l'amour et l'affection que je porte pour vous.

A tous ceux qui, par un mot, m'ont donné la force de continue.

Table des matières

Liste des figures

Liste des abréviations

IDS	Intrusion Detection System
AIDS	Application Intrusion Detection System
CDDL	Common Development and Distribution License
CSI	Computer Security Institue
DDoS	Distributed Denial of Service
DoS	Denial of Service
FTP	File Transport Protocol
HIDS	Host Intrusion Detection System
HTML	Hyper Text Markup Language
HTTP	Hyper Text Transfer Protocol
ICMP	Internet Control Message Protocol
IP	Internet Protocol
IPV6	Internet Protocol Version 6
IPSec	Internet Protocol Security
IPS	Intrusion Prevention System
IEC	International Electrotechnical Commission
LAN	Local Area Network
VLAN	Virtual Local Area Network
WAN	Wide Area Network
NIDS	Network Intrusion Detection System
OSI	Open Systems Interconnection
SBIDS	Stack Based Intrusion Detection System
TCP	Transmission Control Protocol
VPN	Virtual Private Network
XML	eXtensible Markup Language
PHP	Hypertext Preprocessor
CSS	Cascading Style Sheets
BD	Base de Donner

Introduction Générale

> Dans la « société de l'information », la sécurité des systèmes informatiques constitue un enjeu crucial. Le contrôle de l'information traitée et partagée au sein de ces systèmes est un problème d'autant plus délicat que le nombre d'utilisateurs de ces systèmes est important. Relier ces Systèmes entre eux au sein de réseaux informatiques, eux-mêmes interconnectés, complexifie donc la tâche des responsables de la sécurité.

La sécurité d'un système informatique repose en premier lieu sur la mise en place d'une politique de sécurité. Celle-ci peut être définie comme un ensemble de règles permettant d'assurer trois propriétés:

- la confidentialité des données : seuls les utilisateurs autorisés peuvent consulter une information donnée ;
- l'intégrité des données : seuls les utilisateurs autorisés peuvent modifier une information donnée ;
- la disponibilité du système : le système doit être capable de rendre le service prévu en un temps borné.

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 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 des systèmes de détection Intrusion (IDS).

L'objectif de la détection d'intrusions est d'automatiser la tâche d'audit. Il s'agit bien, théoriquement, de détecter de manière automatique les violations de politique de sécurité, qu'on appelle intrusions. Dans la pratique, les outils actuels ne sont cependant pas configurés directement par la politique. Aussi, s'ils détectent certaines intrusions, ils détectent aussi des tentatives d'intrusions infructueuses, ce qui peut être souhaité, ou non. En outre, la relative naïveté des algorithmes de détection conduit à un nombre élevé d'alertes, dont une part significative est en fait constituée de fausses alertes (faux positifs). Enfin, certaines intrusions peuvent ne pas être détectées (faux négatifs).

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 une application web à l'aide d'un système de détection d'intrusion comportementale à base de l'algorithme K-means.

Le premier chapitre est un chapitre descriptif pour la sécurité des réseaux, sur lequel on va définir les menaces, les logiciels malveillants et une politique de sécurité ainsi les principaux mécanismes de sécurité.

Le second chapitre est 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 troisième chapitre est consacré à la conception de notre application, le grand problème des IDS, et les solutions proposée pour optimiser et diminuer les fausses alertes détecter par les IDS comportementales.

Le dernier chapitre est consacré à la réalisation de notre application (une boutique en ligne), et l'implémenté avec un système de détection d'intrusion.

Problématique:

Afin de remplir les objectifs des IDS, diverses méthodes de détections d'intrusions ont été proposées, parmi ces méthodes, plusieurs travaux ont été menés sur les algorithmes de classification. Cependant les algorithmes de classification exploités jusqu'à présent dans le domaine de la sécurité possédant eux même des limites qui peuvent entacher la détection d'intrusions :Un arbre de décision par exemple présente un très gros default dans le cas ou des instances de l'ensemble de tests ne satisfont aucune règle de la base d'apprentissage .Ce qu'il motiver l'apparence de nombreux de fausses alertes dans le system et nuire la fiabilité et l'efficacité du system.

Pour remédier à ces défauts en particulier et de résoudre les limites des IDS en général, la détection d'intrusions doit s'orienter vers des nouvelles techniques de détection pour mieux assurer la sécurité de réseaux. Pour cela, nous proposons une procédure de détection d'intrusions, qui consiste à utiliser des méthodes de classification pour but de démunie les fausses alertes déclenché par les system de détection d'intrusion afin d'avoir un system plus fiable et plus performent.

Objective:

- Créer une boutique Online
- Créer un system de détection d'intrusion par l'approche comportemental
- Créer un system de détection d'intrusion en utilise l'algorithme K-means
- Faire une comparaison entre l'approche comportementale et l'approche utilisé le K-means.

CHAPITRE I

La sécurité informatique

I.1 Introduction:

L'informatique et en particulier l'Internet jouent un rôle majeur dans le domaine des réseaux. Un grand nombre d'applications critiques d'un point de vue de leur sécurité sont déployées dans divers domaines comme le domaine militaire, la santé, le commerce électronique, etc. La sécurité des réseaux devient alors une problématique essentielle tant pour les individus que pour les entreprises ou les états. Il est donc important de définir une politique de sécurité pour ces réseaux et de veiller à son respect. Néanmoins les mécanismes de sécurité préventifs mis en place ne sont pas incontournables. Il est nécessaire de mettre en œuvre des outils permettant de détecter toute violation de la politique de sécurité, c'est-à-dire toute intrusion.

Tout au long de ce chapitre, notre intérêt se porte sur les principales menaces pressant sur la sécurité des réseaux ainsi que les mécanismes de défense.

I.2 Menaces sur la sécurité informatique

La sécurité informatique est la protection de l'information et des systèmes d'information contre l'accès, l'utilisation, la divulgation, la perturbation, la modification ou la destruction afin de garantir la confidentialité, l'intégrité et la disponibilité.

I.2.1 Les principes de la sécurité informatique

La sécurité des systèmes informatique repose sur trois principes clés: la confidentialité, l'intégrité et la disponibilité. Vu le contexte de l'application de ces principes, certains d'entre eux peuvent avoir plus d'importance que d'autres. Par exemple, la confidentialité est la plus importante dans le cadre d'une transmission des messages secrets entre deux agences de sécurité nationale ou internationale, si quelqu'un arrive à décrypter le message transmis, la sécurité sera compromise et l'information sera divulguée. Par contre la disponibilité est la plus importante pour les sites de e-commerce, la non-disponibilité est catastrophique pour des sites comme amazone et eBay [1]

- **La confidentialité :** La confidentialité consiste à préserver la révélation non autorisée d'information sensible. La révélation pourrait être intentionnelle comme les attaques qui visent a casser le chiffrement des données et lire les informations, ou involontaire dû au manque de vigilance ou de l'incompétence des individus qui manient les informations
- **L'intégrité :** l'intégrité consiste à garantir trois buts principaux :

➢ Préserver le changement des informations par les utilisateurs non autorisés
➢ Préserver le changement non autorisé ou involontaire d'information par les utilisateurs autorisés
➢ Préserver la cohérence interne et la cohérence externe
  - La cohérence interne: consiste à assurer la cohérence des données interne. Par exemple dans une organisation on assure que le nombre total des articles maintenus par cette organisation est égal à la somme des mêmes articles dans la base de données.

- La cohérence externe: consiste à assurer que la cohérence entre les données dans la base de données et le monde réel est maintenue. Par exemple dans une entreprise on assure que le nombre des articles vendus est le même nombre dans la base de données.

- **La disponibilité :** La disponibilité assure que les utilisateurs autorisés ont un accès opportun et non interrompu aux informations dans le système et le réseau [2]

I.2.2 Les causes de l'insécurité

Au sein d'un réseau informatique, on distingue généralement cinq types de faille qui peuvent causer l'état d'insécurité :

- **Les failles physiques :** généralement dans une entreprise ou une administration la sécurité d'accès aux matériels informatiques n'a pas une grande importance. Il suffit de trouver des prétextes comme faire des tests, de la maintenance ou le nettoyage pour accéder. L'exploitation de cet accès physique pour voler un mot de passe, effacer des données, usurper l'identité d'un autre ou injecter des programmes malveillants peut causer des dégâts catastrophiques pour une entreprise.
- **Les failles réseaux :** les réseaux informatiques sont fondés sur des normes et des standards bien réfléchis où plusieurs organismes collaborent pour les perfectionner. Malgré tous les efforts faits, il existe certaines failles ou détournements de fonctionnement des standards exploitables. Le problème avec les failles réseau c'est la complexité de leurs corrections qui varie d'après la taille du réseau. À titre d'exemple, corriger les failles réseau d'internet est utopique, c'est la raison pour laquelle on se contente de faire des améliorations comme le passage vers IPV6 ou IPSec.
- **Les failles systèmes :** les systèmes d'exploitation sont de plus en plus sophistiqués, ils intègrent différent mécanisme de sécurité comme les mots de passe, les logs, séparation des privilèges…etc. La complexité, la mauvaise configuration ainsi que les faiblesses de certains mécanismes des systèmes d'exploitation représentent un danger pour les utilisateurs. Par exemple la complexité d'un mécanisme de sécurité pousse les utilisateurs à le désactiver, de plus la mauvaise configuration peut engendrer l'arrêt ou la saturation du système.
- **Les failles applicatives :** les failles applicatives sont des failles très connues et très répondues. Ils peuvent être causés par la mauvaise conception, non-traitement des exceptions, faille dans le langage de programmation. Ces failles peuvent engendrer beaucoup de problèmes qui influencent le fonctionnement du système.
- **Les failles Web :** le monde du web représente la combinaison de différent protocole, réseaux, système et application. Les failles web peuvent être causées par l'une des failles précédemment citées ou par des failles qui résident au niveau des protocoles et des standards du fonctionnement du web.

I.2.3 Les différents types d'attaque informatique

Une attaque informatique est littérairement définie par toute tentative de détruire, exposer, modifier, désactiver, voler ou obtenir un accès non autorisé ou toute utilisation non autorisée d'une information, logiciels, physique comme un serveur, services, des personnes et de leurs qualifications, et les biens incorporels (ISO/IEC 27000, 2009).

Pour l'aspect technique, on peut définir une attaque par l'exploitation de l'une des failles précédemment citées pour des fins illégales. Il existe cinq formes d'attaque que nous détaillons comme suit :

- **L'attaque passive :** les attaques passives représentent tout acte qui nous permet de faire l'analyse et le décryptage du trafic, la surveillance des communications, et la capture des informations d'authentification. Les attaques passives peuvent entraîner la divulgation des informations ou des données à un attaquant sans que la victime soit consciente. L'interception du mot de passe, numéros de carte de crédit, des emails représentent tous des attaques passives.
- **L'attaque active :** les attaques actives comprennent toute tentative a pour but de contourner ou arrêter les fonctions de protection, introduire un code malveillant et de voler ou modifier des informations. Les attaques actives peuvent entraîner la divulgation et la diffusion des données, un déni de service, ou la modification des données.[3]
- **L'attaque de proximité ou externe :** les attaques de proximité (externes) représentent l'utilisation de la proximité physique du réseau ou du système qui a été obtenue grâce à l'entrée clandestine ou un accès ouvert afin de modifier, collecter ou refuser l'accès à l'information.
- **L'attaque interne :** les attaques internes peuvent être intentionnelles ou non intentionnelles. Les attaques intentionnelles représentent les tentatives d'espionner, de voler ou d'endommager des informations, utiliser l'information de manière frauduleuse, ou interdire l'accès à d'autres utilisateurs autorisés. Les attaques non intentionnelles représentent le résultat d'une mauvaise manipulation, la négligence ou le manque de connaissances.
- **L'attaque de distribution :** les attaques de distribution représentent toute modification malveillante du matériel ou du logiciel en usine ou lors de la distribution. Ces attaques consistent à introduire un code malveillant dans un produit comme un port dérobé pour obtenir un accès non autorisé à des informations ou une fonction système.

I.2.4 Exemples des attaques informatiques

Il existe plusieurs types d'attaques très connues dans le monde de l'informatique, nous détaillons ici trois exemples d'attaques informatiques très réputées par leurs dangerosités et les dégâts qui peuvent en causer.

I.2.4.1 L'attaque man-in-the-middle:

L'attaquant s'introduit entre deux systèmes sans que l'un d'entre eux aperçoive l'existence d'un troisième système qui fait passer les échanges réseau.

Pour réussir une telle attaque, il faut que la machine de l'attaquant soit physiquement entre les deux machines victimes ou que l'attaquant arrive à modifier le routage réseau afin que sa machine devienne un des points de passage Le schéma suivant illustre le fonctionnement de l'attaque man-in-the-middle.[4]

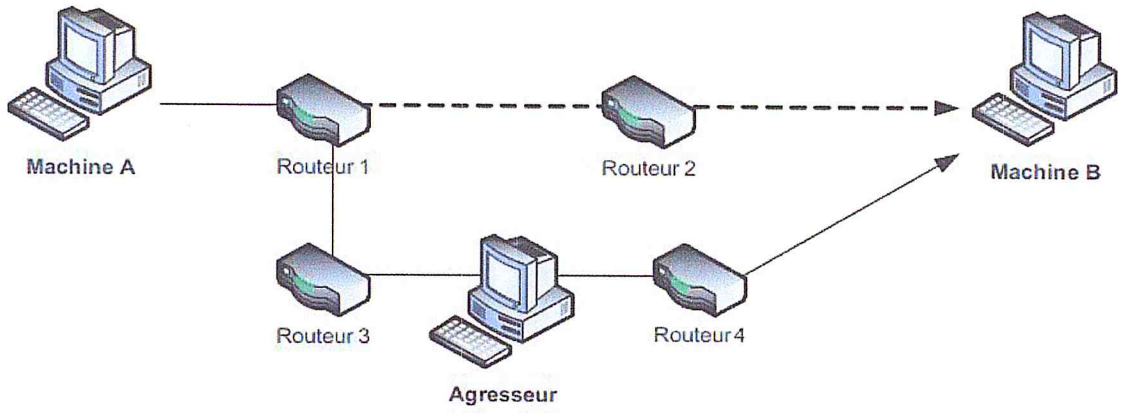


Figure I.1 L'attaque man-in-the-middle

I.2.4.2 L'attaque de déni de service distribué (DDoS):

Elle représente la version distribuée de l'attaque de déni de service. Le but de cette variante de l'attaque DoS est que la victime n'arrive pas à isoler les attaquants vu le nombre important des machines utilisées pour réaliser cette attaque, il faut premièrement pénétrer par diverses méthodes des systèmes dits "handlers" et agents. Où l'attaquant contrôle un ensemble de systèmes "handlers" qui contrôlent eux-mêmes un ensemble de systèmes agents. Le hacker lance l'attaque en ordonnant les systèmes "handlers", qui eux-mêmes ordonnent les agents

Le schéma suivant illustre le fonctionnement de l'attaque DDoS.

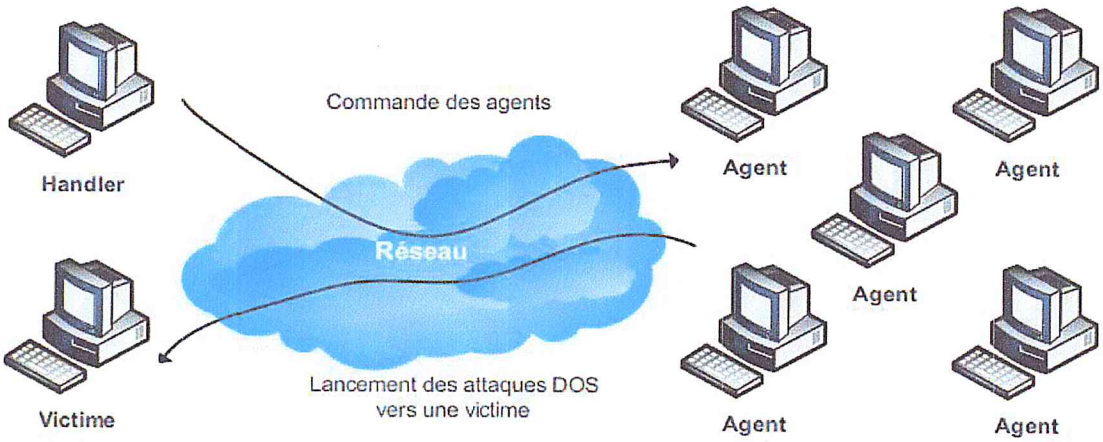


Figure I.2 L'attaque DDoS

I.2.4.3 Attaque par virus:

un virus informatique est tout programme capable de se reproduire par lui-même. Un virus informatique peut prendre la forme d'une routine ou d'un programme une fois activé il utilise tous les moyens pour empoisonner la vie de l'utilisateur. Les virus informatiques représentent le type d'attaque le plus fréquent. Le cycle de vie d'un virus commence par la création, puis la reproduction, ensuite l'activation, ensuite le découvrir et en fin le détruire. Il existe plusieurs types de virus qu'on peut les résumer par :

- virus de secteur d'amorçage.
- virus d'infection des fichiers (parasites).
- virus non-résidents mémoire.
- virus résidents mémoire.
- virus multiformes.
- virus furtifs.
- virus polymorphes (mutants).
- virus réseau et vers (worms).
- virus flibustiers (bounty hunters).
- bombes logiques.
- Chevaux de Troie.

I.2.5 L'impact des attaques informatiques

Au début de l'histoire des réseaux informatiques les attaques informatiques ont été munies par des experts, leur nombre a été limité voir très limité. Maintenant les outils de piratage et des attaques informatiques sont disponibles aux amateurs avec quelques dollars. De plus, les pertes qui peuvent être engendrées par une attaque informatique sont de plus en plus graves.

On parle aujourd'hui des milliards de dollars de perte, des pays paralysés, des projets

stratégiques sabotés, des programmes présidentiels divulgués tout ça à cause des attaques informatiques qui varient dans le but, l'ampleur et la dangerosité.

La figure suivante montre les pertes causées par des attaques informatiques dans une étude munie par le CSI (Computer Security Institue) sur 194 organisations. Le montant total de perte durant l'année 2007 est de 66.930.950 dollars American. Les fraudes financières occupent la première place avec 21.124.750 $USA, la deuxième place est occupée par les

virus avec 8.391.800 $USA [5]

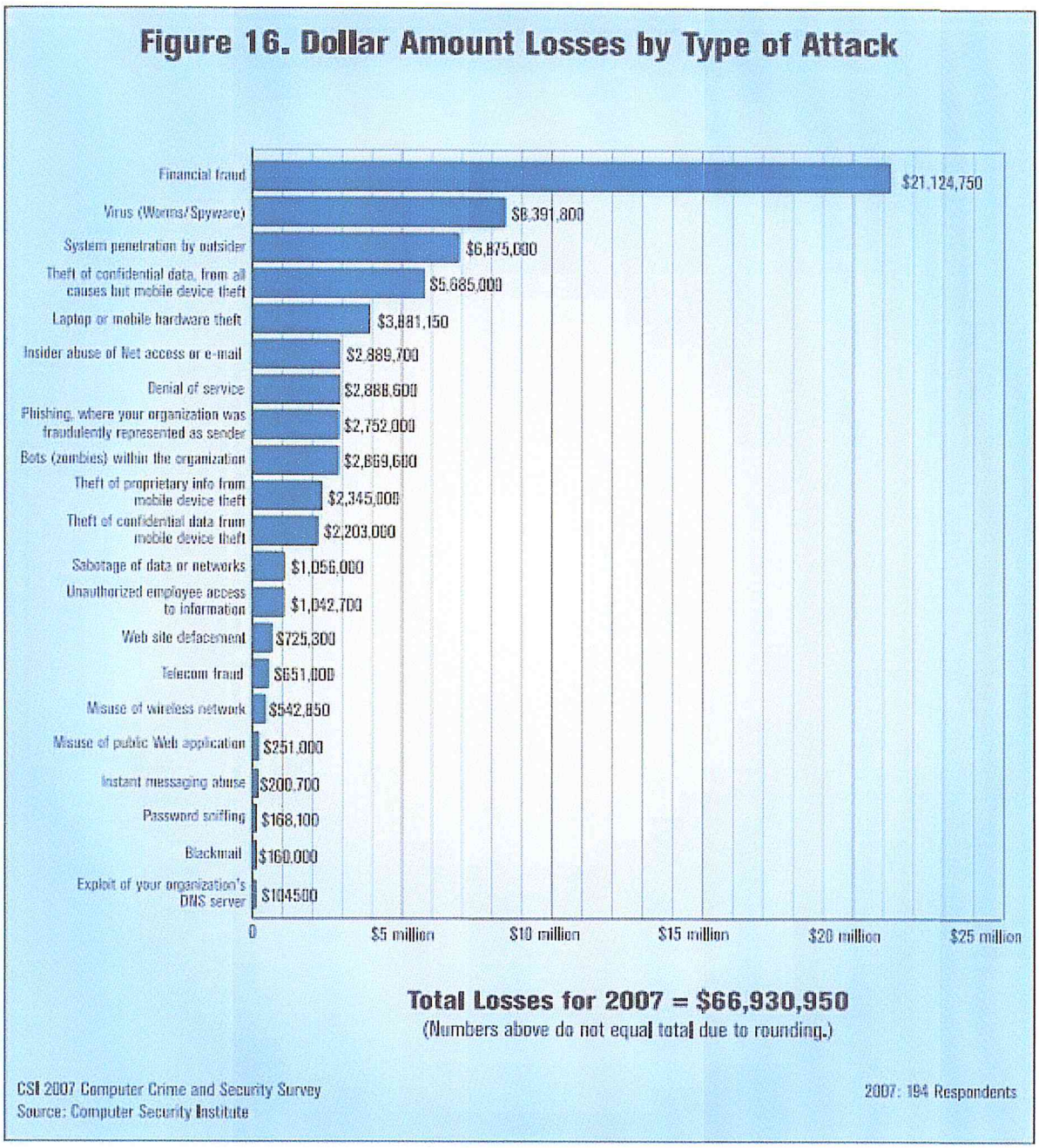


Figure I.3 Le rapport des pertes causées par des attaques informatiques

I.2.6 La gestion des risques lie aux incidents informatiques

La gestion des risques comprend trois processus: l'estimation des risques, la réduction des risques, et l'évaluation et l'estimation

Le processus de l'estimation des risques comprend :

- L'identification et évaluation des risques
- L'identification et évaluation de l'impact des risques
- La recommandation des mesures pour la réduction des risques.

On peut évaluer quantitativement les risques par l'équation suivante :

$$Risque = \frac{menace \times vulnérabilité}{contre-mesure}.$$

La menace représente le type d'action susceptible de nuire dans l'absolu, tandis que la vulnérabilité représente le niveau d'exposition face à la menace dans un contexte particulier, la contre-mesure est l'ensemble des actions mises en œuvre pour prévenir les menaces.

Le processus de la réduction des risques comprend les taches suivantes:

- Prioriser les mesures de réduction des risques recommandés par le processus de l'estimation des risques.
- Implémenter les mesures de réduction des risques recommandés par le processus de l'estimation des risques.
- Maintenir les mesures de réduction des risques recommandés par le processus de l'estimation des risques.

Le processus d'évaluation et d'estimation inclut un processus d'évaluation continu. Par exemple, l'autorité approbatrice désignée des États-Unis d'Amérique est la responsable de déterminer si le risque résiduel dans le système est acceptable ou que des mesures de contrôle et de protection supplémentaires devraient être implémentées pour accomplir l'accréditation d'un système informatique [6].

I.3 Mécanismes et les Outils de la sécurité :

À cause des menaces provenant des logiciels malveillants, Il faut mettre en place des mécanismes pour s'assurer la confidentialité, l'intégrité et la disponibilité des services. Parmi ces mécanismes, on peut citer :

I.3.1 Cryptage :

Le chiffrement est un procédé de cryptographie grâce auquel on souhaite rendre la compréhension d'un document impossible à toute personne qui n'a pas la clé de (dé)chiffrement. Ce principe est généralement lié au principe d'accès conditionnel Bien que le chiffrement puisse rendre secret le sens d'un document, d'autres techniques cryptographiques sont nécessaires pour communiquer de façon sûre.

La Figure I.4 montre le fonctionnement de chiffrement. [8]

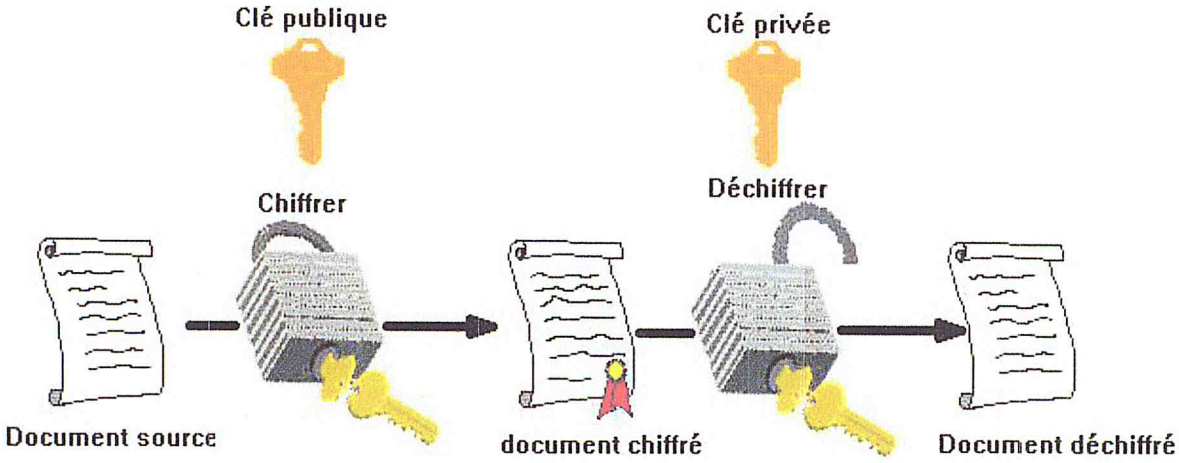


Figure I.4 Cryptage .

I.3.2 Pare-Feu :

C'est un ensemble de différents composants matériels (physique) et logiciels (Logique) qui contrôle le trafic intérieur/extérieur selon une politique de sécurité. Un système pare-feu fonctionne la plupart du temps grâce à des règles de filtrage indiquant les adresses IP autorisées à communiquer avec les machines aux réseaux, il s'agit ainsi d'une passerelle filtrante.

Il permet d'une part de bloquer des attaques ou connexions suspectes d'accéder au réseau interne. D'un autre côté, un firewall sert dans de nombreux cas également à éviter la fuite non contrôlée d'informations vers l'extérieur. Il propose un véritable contrôle sur le trafic réseau de l'entreprise, Il permet donc d'analyser, de sécuriser et de gérer le trafic réseau.

La Figure I.5 schématise le fonctionnement d'un pare-feu.

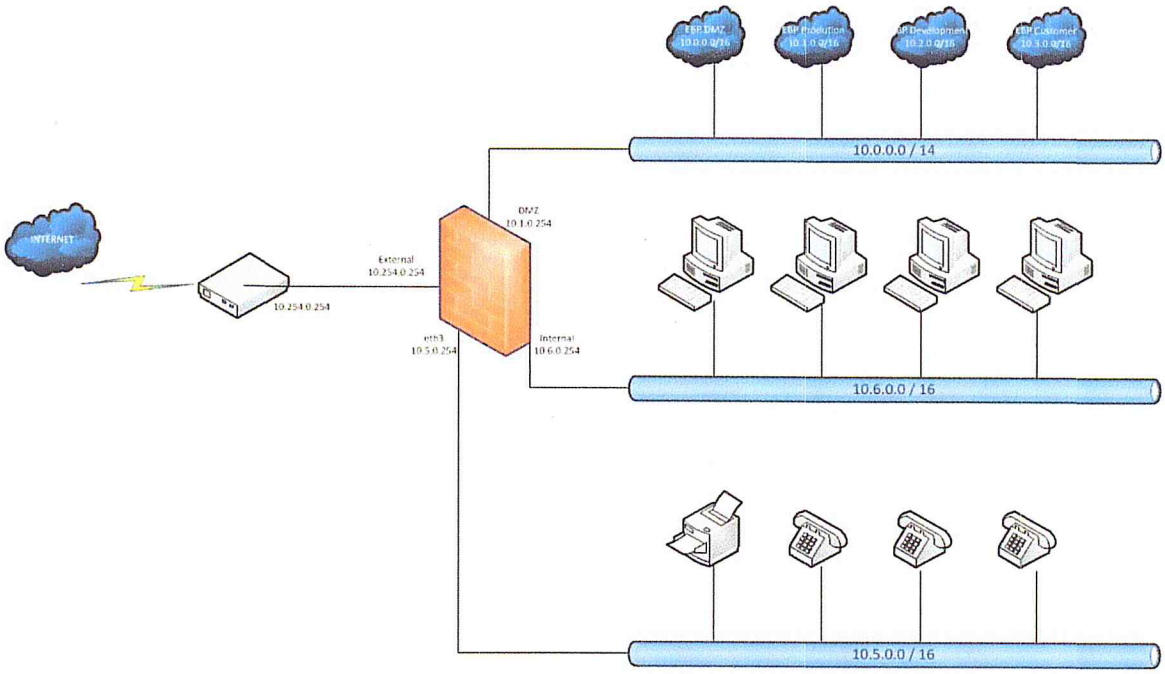


Figure I.5 Pare-feu.

I.3.3 Antivirus:

Les antivirus sont des logiciels conçus pour identifier, neutraliser et éliminer des logiciels malveillants. Ceux-ci peuvent se baser sur l'exploitation de failles de sécurité, mais il peut également s'agir de programmes modifiant ou supprimant des fichiers, que ce soit des documents de l'utilisateur de l'ordinateur infecté, ou des fichiers nécessaires au bon fonctionnement de l'ordinateur. [9]

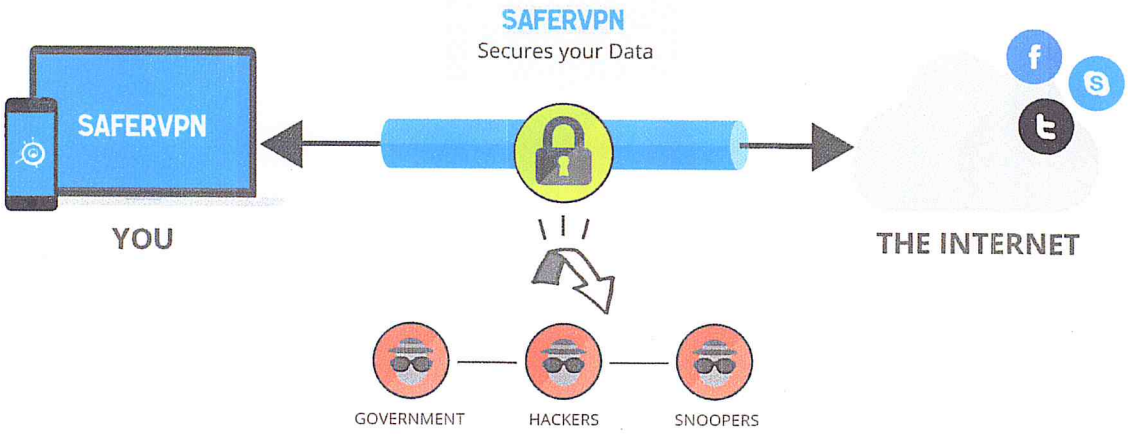
Un antivirus vérifie les fichiers et courriers électroniques, les secteurs de boot (pour détecter les virus de boot), mais aussi la mémoire vive de l'ordinateur, les médias amovibles (clefs USB, CD, DVD, etc.), les données qui transitent sur les éventuels réseaux (dont internet), etc.

I.3.4 VPN :

Dans les réseaux informatiques, le réseau privé virtuel (Virtual Private Network en anglais, abrégé en VPN) est une technique permettant aux postes distants de communiquer de manière sûre, tout en empruntant des infrastructures publiques (internet). [10]

Un VPN repose sur un protocole, appelé protocole de tunnelisation, c'est-à-dire un protocole permettant aux données passant d'une extrémité à l'autre du VPN d'être sécurisées par des algorithmes de cryptographie.

La Figure I.6 montre le principe de protocole de tunnelisation



1.

I.3.5 Système de détection d'intrusions:

Un système de détection d'intrusion (ou IDS : Intrusion Détection System) est un mécanisme destiné à repérer des activités anormales ou suspectes sur la cible analysée (un réseau ou un hote). Il permet ainsi d'avoir une connaissance sur les tentatives réussies comme échouées des intrusions.

Il faut distinguer deux aspects dans le fonctionnement d'un IDS : le mode de détection utilisé et la réponse apportée par l'IDS lors de la détection d'une intrusion.

I.4 Mise en place d'une politique de sécurité :

La mise en œuvre d'une politique de sécurité globale est assez difficile, essentiellement par la diversité des aspects à considérer. Une politique de sécurité peut se définir par un certain nombre de caractéristiques : les niveaux où elle intervient, les objectifs de cette politique et enfin les outils utilisés pour assurer cette sécurité.

Chaque aspect différent doit être pris en compte, de façon à atteindre les objectifs de sécurité désirés, en utilisant de façon coordonnée les différents outils à disposition. [11]

I.4.1 Politique de sécurité :

Lors de la configuration d'un réseau qu'il s'agisse d'un réseau local (LAN), d'un réseau local virtuel (VLAN) ou d'un réseau étendu (WAN), il est important de définir dès le départ des politiques de sécurité. Celles-ci forment des règles électroniques programmées et stockées dans un dispositif de sécurité, qui ont pour objectif de contrôler certains aspects spécifiques, tels que les droits d'accès.

Les politiques mises en œuvre doivent contrôler les accès à des zones définies du réseau et expliquer comment interdire l'accès à certaines zones à des utilisateurs non autorisés. Les utilisateurs ayant accès à certaines parties du réseau doivent également être soumis à des règles précises.

- **Qui doit appliquer et gérer ces politiques ?:** la personne ou le groupe chargé de gérer et d'entretenir le réseau et sa sécurité doit avoir accès à l'ensemble de ses zones. La fonction de gestion des politiques de sécurité doit donc être confiée à des personnes particulièrement dignes de confiance et disposant des compétences techniques nécessaires.
- **Diffuser les politiques** [12]**:** une politique risque d'être inutile si toutes les parties concernées n'ont pas connaissance ou ne la comprennent pas. Il est capital de développer des mécanismes de diffusion des politiques existantes, des changements de politique, des nouvelles politiques et des alertes de sécurité relatives aux virus ou à des attaques imminentes.
- **Contrôler l'application des politiques:** pour qu'elles soient parfaitement efficaces, il est essentiel de contrôler que les politiques sont bien appliquées. Il est donc important d'établir une vérification de la bonne mise en place de celles-ci.
- **Mots de passe** [12] **:** le moyen le plus simple et le plus classique de s'assurer que seules les personnes autorisées peuvent accéder à une certaine partie du réseau est de protéger certaines zones du réseau par un mot de passe. Cependant, les plus puissantes infrastructures de sécurité sont inefficaces si les mots de passe ne sont pas eux-mêmes protégés.

  Les règles d'or ou politiques à suivre, en matière de mots de passe sont les suivants :
  - ✓ Changer régulièrement les mots de passe ;
  - ✓ Choisir des mots de passe aussi dénués de sens que possible ;
  - ✓ Ne jamais divulguer les mots de passe.

  Dans l'avenir, certains mots de passe pourraient être remplacés par des technologies de biométrie identifiant les utilisateurs en fonction de caractéristiques physiques comme leurs empreintes digitales, optiques ou vocales.
- **Certificats numériques** [12]**:** les certificats numériques ou clé publiques sont les équivalents électroniques d'un permis de conduite ou d'un passeport et sont émis par une autorité de certification. Les certificats numériques sont généralement utilisés à des fins d'identification, lors de l'établissement de tunnels sécurisés sur Internet, comme c'est le cas dans les réseaux virtuels privés(VPN).

I.5 L'avenir de la sécurité :

I.5.1 Nouveaux protocoles, nouvelles menaces :

Depuis quelques années, des protocoles qui ne sont peut-être plus nouveaux d'un point de vue chronologique, mais qui méritent encore ce qualificatif par l'innovation qu'ils ont incarné par rapport aux protocoles traditionnels d'Internet, fondent sur le modèle client-serveur, posent aux administrateurs de réseaux de nouvelles questions, notamment dans le domaine de la sécurité.

I.5.2 Les cinquante prochaines années, selon Alan Cox :

Alan Cox est un des principaux développeurs du noyau Linux, qu'il a notamment contribué à doter de la capacité de préemption. Il est intéressant de relever ce qu'il considère comme des de progrès de la sécurité des systèmes informatiques, en partant de son jugement sur la situation actuelle d'insécurité, qu'il estime insoutenable :

- L'essor des systèmes de vérification de code, et surtout de leur utilisation ;
- L'amélioration des méthodes de développement, avec des langages comme Java qui règlent la majeure partie des problèmes d'allocation mémoire ;
- Une gestion plus fine et plus restrictive de l'attribution des privilèges aux utilisateurs ;
- Les techniques de défense en profondeur se répandent : ainsi, le choix d'adresses aléatoires (ou plutôt imprévisibles) pour l'implantation des objets en mémoire, le verrouillage par le matériel ou par le logiciel de certaines régions de mémoire rendues non exécutables, l'usage de systèmes sécurisés.[6]

I.5.3 Détection et prévention d'intrusion :

La vague suivante de produits de sécurité fut celle des systèmes de détection et de prévention d'intrusions, dont le modèle libre est le logiciel Snort. Ces logiciels utilisent une base de données de signatures de vers et d'autres logiciels malfaisants, un peu à la manière d'un antivirus, et se sont révélés relativement efficaces contre la grande épidémie de vers des années 2001 à 2004, mais leur vogue décline au fur et à mesure que leur efficacité diminue.

I.6 Conclusion:

Aucun système d'information n'est sûr à 100% ! Il est impossible de garantir la sécurité totale d'un système pour les raisons suivantes :

- Les bugs dans les programmes courants et les systèmes d'exploitation sont nombreux ;
- La cryptographie a ses faiblesses : les mots de passe peuvent être cassés ;
- Même un système fiable peut être attaqué par des personnes abusant de leurs droits ;
- Plus les mécanismes de sécurité sont stricts, moins ils sont efficaces ;
- On peut s'attaquer aux systèmes de sécurité eux-mêmes.

Pour une machine connectée à un réseau, le problème aujourd'hui n'est plus de savoir si elle va se faire attaquer, mais quand cela va arriver ; une solution possible est alors d'essayer de repousser les risques dans le temps par la mise en œuvre de divers moyens destinés à augmenter le niveau de sécurité.

CHAPITRE II

SYSTEME DE DETECTION D'INTRUSION

II.1 Introduction

Une propriété de valeur doit être protégée contre le vol et la destruction. Certaines maisons sont équipées de systèmes d'alarme qui peuvent décourager des voleurs, prévenir les autorités dans le cas d'une effraction et même avertir les propriétaires que leur maison est en feu. De telles mesures sont nécessaires pour assurer l'intégrité des maisons et la sécurité de leurs propriétaires.

La même assurance d'intégrité et de sécurité devrait également être appliquée aux systèmes et données informatiques. L'internet a facilité le flux d'informations, personnelles, financières et autres. En même temps, il a également promu autant de dangers. Les utilisateurs malveillants et les craqueurs recherchent des proies vulnérables comme les systèmes sans correctifs, les systèmes affectés par des chevaux de Troie et les réseaux exécutant des services peu sûrs. Des alarmes sont nécessaires pour prévenir les administrateurs et les membres de l'équipe de sécurité qu'une effraction s'est produite afin qu'ils puissent répondre en temps réel au danger.

Les systèmes de detection d'intrusions ont été conçus pour jouer le rôle d'un tel système d'alarme. [13]

Deux approches ont été proposées à ce jour dans ce but: l'approche comportementale et l'approche par signatures. La première se base sur l'hypothèse que l'on peut définir un comportement « normal » de l'utilisateur et que toute déviation par rapport à celui-ci est potentiellement suspecte. La seconde s'appuie sur la connaissance des techniques employées par les attaquants: on tire des signatures d'attaque et on recherche dans les traces d'audit leur éventuelle survenue. [14]

Dans Ce chapitre nous présentons tout d'abord la notion de système de détection d'intrusions ainsi que son architecture. Nous présentons également la classification des IDS, dans ce cadre plusieurs critères sont prises en compte nous commençons par la classification selon la méthode d'analyse qui découpe les IDS en deux approaches (Comportementale et par signatures), enfin nous allons mettre le point sur la détection d'intrusions Web.

2.3.1. Définition d'un système de détection d'intrusion

On peut définir un système de détection d'intrusion (IDS) comme tout outil, méthode et ressource qui nous aident à prévoir ou identifier toute activité non autorisée dans un réseau.

Une partie du nom du système de détection d'intrusion est trompeuse, les systèmes de détection d'intrusion actuels ne détectent pas les intrusions, mais ils détectent les activités réseau qui peuvent être une intrusion ou non. La détection d'intrusion est typiquement une partie d'un système de protection total installé autour d'un système ou appareil. Il n'est pas une mesure de protection autonome.

Les IDS traditionnellement suivent deux critères :

- **Fiabilité :** toute intrusion doit effectivement donner lieu à une alerte. Une intrusion non signalée constitue une défaillance de l'IDS, appelée faux négatif.

  (Voir Figure II.1)

- **Pertinence des alertes :** toute alerte doit correspondre à une intrusion effective. Toute « fausse alerte » (appelée également faux positif) diminue la pertinence de l'IDS. (Voir Figure II.1)

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

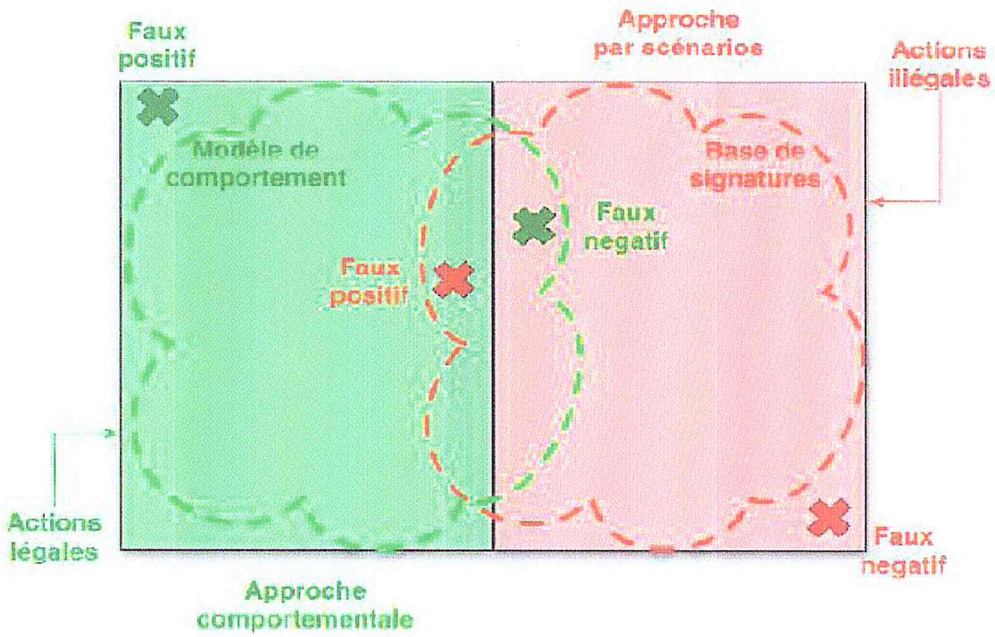


Figure II.1 Problèmes des IDS.

Les IDS proposent les fonctions suivantes:

- Détection d'attaques (actives ou passives) ;
- Génération des rapports ;
- Outils de corrélation avec d'autres éléments de l'architecture de sécurité ;
- Réaction aux attaques par le blocage de route ou la fermeture de connexion ;
- Transfert d'activités.[15]

2.3.2. Les avantages d'un système de détection d'intrusion

Les systèmes de détection d'intrusion offrent beaucoup d'avantages comme :

- Une efficacité plus grande que celle de la détection manuelle des intrusions.
- L'utilisation d'une base de connaissance plus grande pour prédire les intrusions.
- La capacité de traiter un large volume de données.
- Produit une alerte presque en temps réel ce qui réduit le dommage potentiel des
- attaques.
- Des mesures de contre-attaque automatique comme la fermeture des sessions, désactivation des comptes utilisateur, lancement des scripts automatiques.
- L'ajout d'une valeur préventive forte.
- La création automatique des rapports et le jugement de la suite d'événements.[16]

2.4. Taxonomie des systèmes de détection d'intrusion

Il existe de nombreux systèmes de détection d'intrusion, ces IDSs peuvent être classifiés d'après plusieurs critères. Cinq critères de classification des systèmes de détection d'intrusion ont été introduits par [16] :

La figure 2 résume ces cinq critères de classification des IDSs:

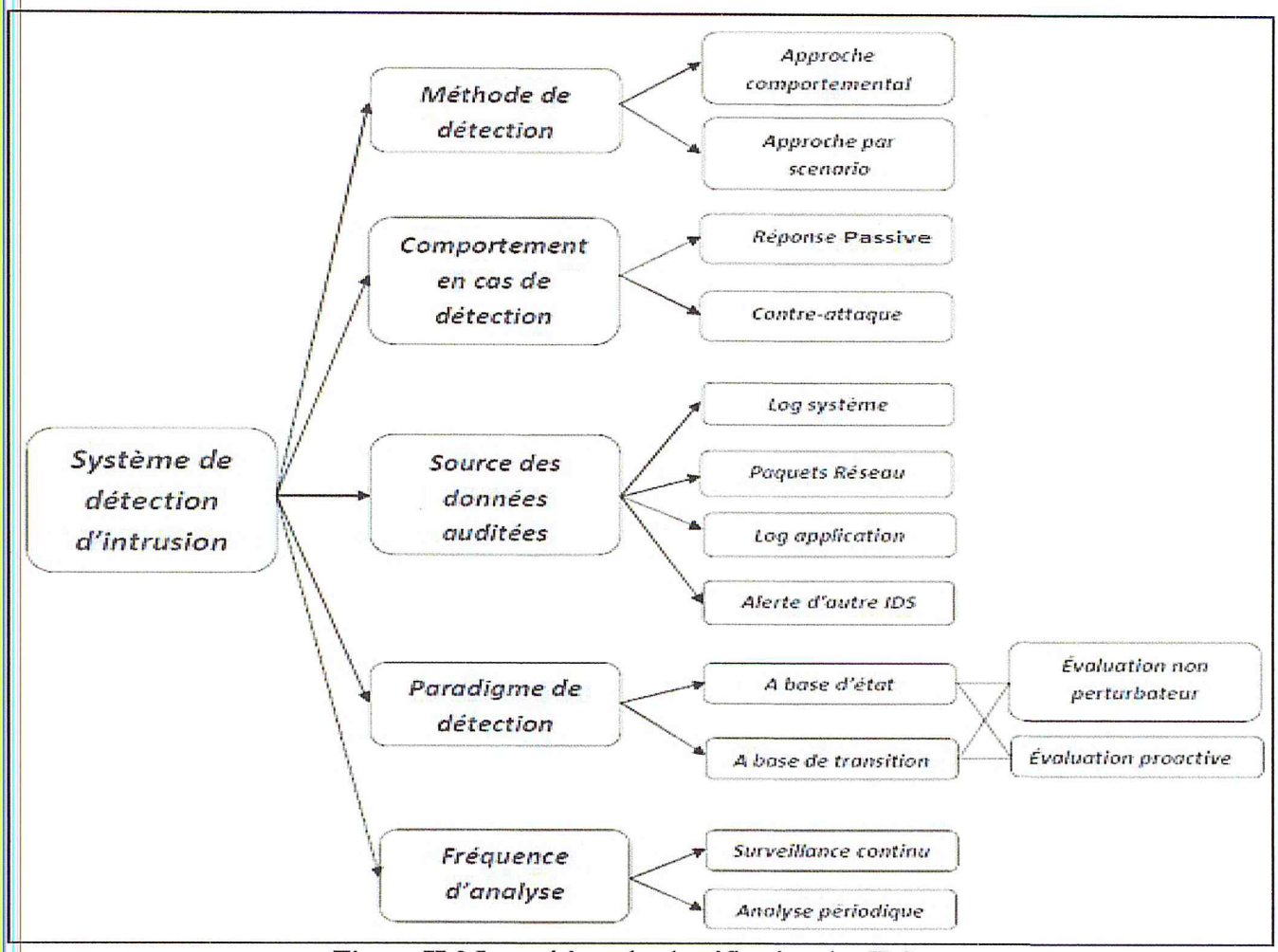


Figure II.2 Les critères de classification des IDSs

II.2 Architecture d'un IDS :

Nous décrivons dans cette section les trois composants qui constituent classiquement un système de détection d'intrusions [15].

La Figure II.3 illustre les interactions entre ces trois composants.

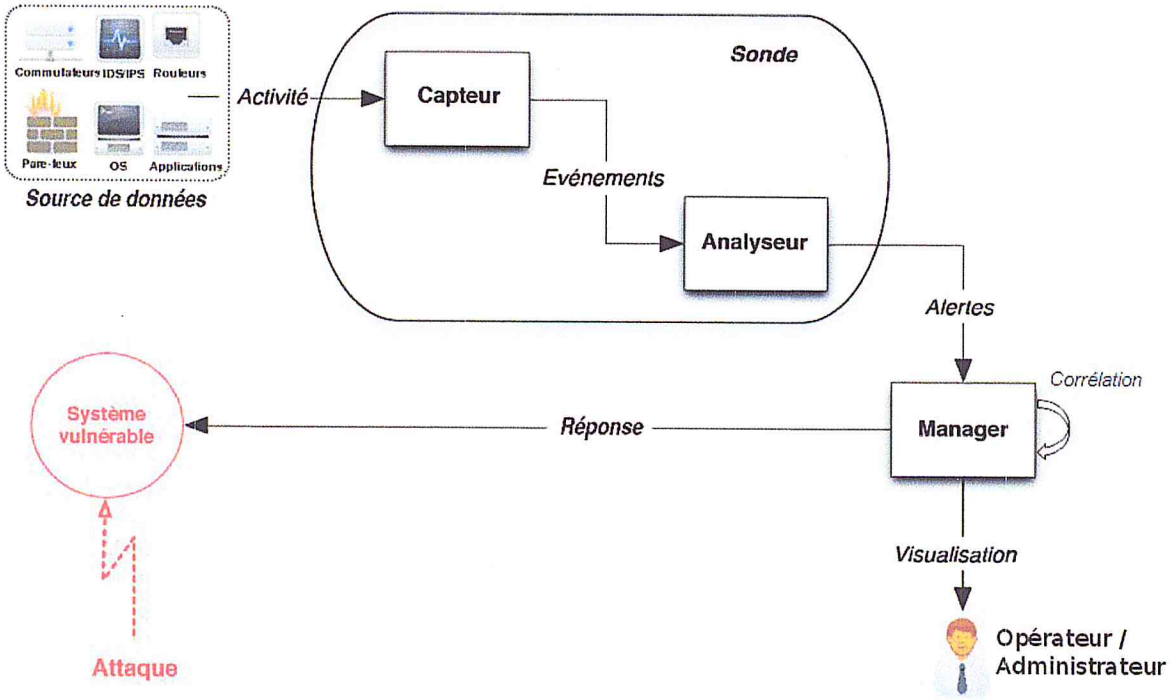


Figure II.3 Architecture classique d'un IDS

II.2.1 Capteur :

Le capteur observe l'activité du système par le biais d'une source de données et fournit à l'analyseur une séquence d'événements qui renseignent de l'évolution de l'état du système. Le capteur peut se contenter de transmettre directement ces données brutes, mais en général un prétraitement est effectué.

On distingue classiquement trois types de capteurs en fonction des sources de données utilisées pour observer l'activité du système : les capteurs système, les capteurs réseau et les capteurs applicatifs.

II.2.2 Analyseur :

L'objectif de l'analyseur est de déterminer si le flux d'événements fourni par le capteur contient des éléments caractéristiques d'une activité malveillante.

II.2.3 Manager:

Le gestionn collecte les alertes produites par le capteur, les met en forme et les présente à l'opérateur. Éventuellement, le manager est chargé de la réaction à adopter qui peut être :

- Confinement de l'attaque, qui a pour but de limiter les effets de l'attaque ;
- Eradication de l'attaque, qui tente d'arrêter l'attaque ;
- Recouvrement, qui est l'étape de restauration du système dans un état sain ;
- Diagnostic, qui est la phase d'identification du problème.

Du fait du manque de fiabilité des systèmes de détection d'intrusions actuels, les réactions sont rarement automatisées, car elles peuvent se traduire par un déni de service en cas de faux positif.

II.3 Méthodes d'analyses :

La technologie des systèmes de détection d'intrusions permet d'analyser les données recueillies de trois façons [17] :

II.3.1 Analyse centralisée :

L'IDS possède plusieurs capteurs, il centralise les alertes pour les analyser sur une seule machine. Ce type d'analyse présente l'avantage d'avoir une vue globale sur toutes les machines protégées. Toutefois, il a l'inconvénient d'occupation très longue du réseau pour acheminer l'information.

II.3.2 Analyse locale :

Chaque machine dispose d'un capteur et analyse l'information à son niveau. Avec ce type d'analyse le trafic réseau est diminué mais les attaques distribuées peuvent échapper à la détection.

II.3.3 Analyse distribuée :

Des petits programmes appelés agents sont déployés sur les noeuds du réseau. Pour les besoins d'analyse un agent est envoyé sur une machine pour traiter l'information.

II.4 Classification des systèmes de détection d'intrusions :

Plusieurs critères permettent de classer les systèmes de détection d'intrusions, la méthode d'analyse étant le principal. On peut citer aussi d'autres critères de classification des IDS : les sources de données à analyser, le type de réaction, le mode d'utilisation, etc.

II.4.1 Classification selon la méthode d'analyse :

Deux grandes approches ont été proposées dans la littérature : la détection d'intrusions par signatures et la détection d'intrusions comportementale :

L'approche par signatures qui cherche à répondre à la question : "le comportement actuel de l'utilisateur et/ou l'application contient-il une attaque connue ?", tandis que l'approche comportementale cherche à répondre à la question : "le comportement actuel de l'utilisateur et/ou l'application est-il cohérent avec son comportement passé ?". Ces deux approches sont opposées mais complémentaires.

II.4.1.1 Approche par signatures :

L'approche par signatures (Misuse Detection) ressemble beaucoup aux techniques utilisées par l'antivirus, le principe commun à toutes les techniques de cette classe consiste à utiliser une base de données, contenant des spécifications de signatures d'attaques. Le détecteur d'intrusions compare le comportement observé du système à cette base et remonte une alerte si ce comportement correspond à une signature prédéfinie. « Ainsi, tout ce qui n'est pas explicitement défini est autorisé » et tout ce qui est explicitement défini est interdit [18].

Un IDS par signatures se compose :

- D'une ou plusieurs sondes, générant un flux d'événements, qui peuvent être de type réseau ou hôte;
- D'une base de signatures : le taux de couverture de l'IDS dépend essentiellement de la qualité de la base de données puisque seules les attaques dont la signature est présente dans la base sont susceptibles d'être détectées. Les signatures sont décrites à l'aide de langages de description d'attaques. Les signatures sont la plupart du temps définies par un opérateur bien que des travaux récents permettent la génération automatique des signatures. [19]

La base de signatures doit également être maintenue :

- Les nouvelles attaques détectées par la communauté doivent être intégrées à la base.
- Suivant les choix de l'administrateur de sécurité, les signatures qui ne correspondent plus à une possible intrusion peuvent être enlevées de la base.

La maintenance de la base de signatures est une tâche importante. Sans maintenance, l'IDS ne peut détecter les nouvelles attaques.

D'un système de reconnaissance de motifs dans le flux d'événements : il est chargé d'identifier les motifs présents dans la base de signatures, dans le flux d'événements. Différents systèmes de reconnaissance de motifs ont été définis dans la littérature. Cela va de systèmes simples à base de règles ou de correspondances de chaînes de caractères (string matching) à des systèmes bien plus complexes à base de systèmes experts ou de modélisation d'états qui peuvent apporter suffisamment d'abstraction pour détecter des attaques inconnues mais qui font partie d'une même classe d'attaques. On pourra consulter la classification d'Axelsson pour plus de détails sur ces systèmes. [20]

Il existe plusieurs mécanismes pour mettre en oeuvre cette approche :

- **Analyse par comparaison :** le principe de cette approche (Pattern Matching) est de faire correspondre à chaque signature d'attaque un motif (Pattern) qui est sous forme d'une chaîne de caractères. Durant l'analyse du flux de données qui est aussi une chaîne de caractères, le système de détection d'intrusions tente dereconnaître les motifs d'attaques déjà connus.

  Le principal inconvénient de cette méthode est que seules les attaques reconnues par les signatures seront détectées, il est donc nécessaire de mettre à jour régulièrement la base de signatures.

  Un autre inconvénient est que les motifs sont en général fixes. Or, une attaque n'est pas toujours identique à 100%, le moindre octet différent par rapport à la signature provoquera la non-détection de l'attaque.

- **Recherche de motifs dynamiques :** le principe de cette méthode est le même que précédemment mais les signatures des attaques évoluent dynamiquement. L'IDS est de ce fait doté de fonctionnalités d'adaptation et d'apprentissage.

- **Détection par inférence :** elle est fondé sur le principe de l'inférence de Bayes Dans ce modèle, les attaques connues constituent des hypothèses, pouvant expliquer les faits observés. On considère qu'une attaque donnée se traduit par des symptômes, pouvant apparaître sous forme d'événements dans l'audit, mais aussi de données statistiques comme dans le cas de la détection d'anomalies.

  Etant donné un ensemble de symptômes, l'inférence bayésienne permet de calculer la probabilité de chaque signature d'attaque connue. Lorsqu'une signature affiche une probabilité élevée, une alerte est levée.

- **Analyse de protocoles :** cette méthode se base sur une vérification de la conformité des flux, ainsi que sur l'observation des champs et paramètres suspects dans les paquets.

  L'analyse protocolaire est souvent implémentée par un ensemble de préprocesseurs, où chaque préprocesseur est chargé d'analyser un protocole particulier (FTP, HTTP, ICMP,..). Du fait de la présence de tous ces

préprocesseurs, les performances dans un tel système s'en voient fortement dégradées.

L'intérêt fort de l'analyse protocolaire est qu'elle permet de détecter des attaques inconnues, contrairement au pattern matching qui doit connaître l'attaque pour pouvoir la détecter.

- **Système expert :** technique qui repose sur une base de connaissances et un moteur d'inférence. La base de connaissances est composée elle-même d'une base de règles décrivant les attaques et d'une base de faits contenant les évènements relatifs aux attaques. Durant la phase de détection, le moteur d'inférence est capable de détecter les attaques en utilisant la base des connaissances. [21]

- **Analyse de transition d'états :** dans cette méthode les attaques sont représentées sous forme d'un ensemble d'états par lesquels passe le système à surveiller. Les états sont définis par des conditions sur les variables du système. Les transactions représentent les actions suivant les évènements qui surviennent.

- **Analyse heuristique :** ce type de détecteur d'intrusions nécessite une maintenance active : puisque par nature il ne peut détecter que les attaques dont les signatures sont dans sa base, cette base doit être régulièrement mise à jour en fonction de la découverte de nouvelles attaques. Aucune nouvelle attaque ne peut par définition être détectée, ce qui implique un taux plus élevé de faux négatifs. Le problème se pose essentiellement pour les attaques très récentes, dont les signatures n'ont pas encore pu être incluses dans la base.

II.4.1.2 Approche comportementale :

La détection d'intrusions comportementale (Anomaly Detection) a été la première approche proposée et développée. Anderson [22] propose de détecter des violations de la politique de sécurité du système en observant le comportement des utilisateurs et en le comparant à un modèle du comportement considéré comme normal, appelé profil.

D'une manière générale, l'approche comportementale comporte deux phases : une phase d'apprentissage où le profil est constitué en observant le comportement de l'entité surveillée et une phase de détection pendant laquelle l'IDS observe le comportement de l'entité, mesure la similarité entre ce dernier et le profil et émet une alerte si la déviation est trop importante.

L'idée principale de cette approche est de considérer toute déviation, toute anomalie dans le comportement comme une intrusion. Cette hypothèse est certainement fausse : des événements ou des comportements rares peuvent tout à fait être légitimes du point de vue de la politique de sécurité du système. Le système est susceptible d'émettre des faux positifs. Tant que le nombre de faux positifs reste suffisamment faible, la méthode peut être valide. Cela conduit à poser deux questions essentielles, dans le domaine de la détection d'intrusions comportementale, sur le caractère correct et complet du modèle de comportement normal. La Figure II.4 représente se caractère.

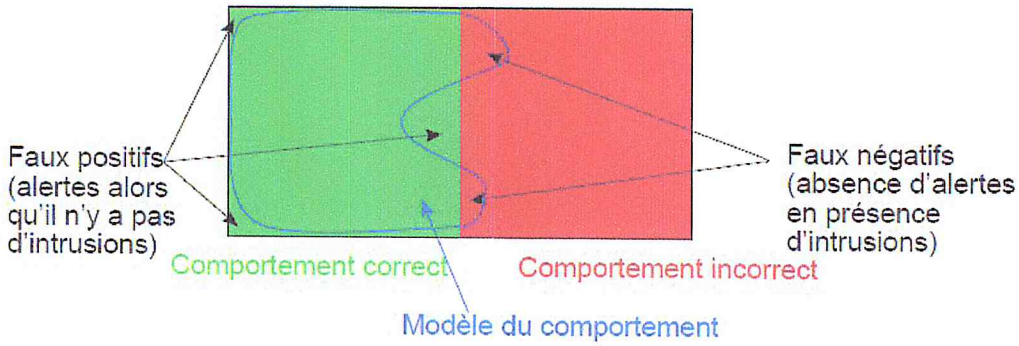


Figure II.4 Caractère complet et correct du modèle de comportement normal.

Le modèle de comportement normal est dit correct s'il ne modélise que le comportement légitime, du point de vue de la politique de sécurité, de l'entité surveillée. Toutes les intrusions sont alors détectées par l'IDS : il n'y a pas de faux négatif. L'idéal est d'obtenir un modèle à la fois complet et correct. Le choix de la méthode de modélisation, des attributs à considérer dans les événements observés vont avoir un impact important sur le caractère correct et complet du modèle.

Plusieurs méthodes de modélisation ont été proposées pour établir le profil de l'entité surveillée:

- **Modèles statistiques :** dans cette méthode, le profil est établi en observant la valeur de certains paramètres du système considéré comme des variables aléatoires. Pour chaque paramètre du système, un modèle statistique est utilisé pour établir la distribution de la variable aléatoire correspondante. Une fois le modèle établi, un vecteur distance est calculé entre le flux d'événements observés et le profil. Si la distance dépasse un certain seuil, une alerte est émise.Classiquement, la détection d'une anomalie repose sur un modèle statistique du comportement des utilisateurs. Denning a ainsi identifié trois familles de modèles statistiques [23] :

- Les modèles simples utilisant des seuils sur des variables. Ces variables peuvent correspondre à la fréquence d'apparition d'un événement ;
- Les modèles utilisant les moments statistiques (moyenne, écart-type…).Ils reposent sur l'hypothèse que le comportement « normal » d'un utilisateur peut être modélisé par une loi statistique, ce qui n'est pas toujours le cas ;
- Les modèles dérivés du modèle de Markov. Les événements ne sont alors plus considérés indépendamment les uns des autres mais en séquence.

Le modèle statistique présente des avantages ainsi que des inconvénients :

- **Avantage :**
  - ✓ Permet de détecter des attaques inconnues.
  - ✓ Habitudes des utilisateurs apprises automatiquement.

- **Inconvénients :**
  - ✓ Difficulté de construire un modèle universel.
  - ✓ Complexité en termes de maintenance.

- **Approche probabiliste:**

  Dans cette approche la construction des profils se base sur la probabilité qu'un évènement ait lieu par rapport à une séquence d'autres évènements. Toute activité ne respectant pas le modèle probabiliste provoquera la génération d'une alerte. Malgré les avantages suivants, il existe des inconvénients :

  - **Avantage :**
    - ✓ Construction du profil simple et dynamique.
    - ✓ Réduction de faux positifs.
  - **Inconvénients :**
    - ✓ Risque de déformation progressive du profil par des attaques répétées.
    - ✓ Mise en place d'un mécanisme d'observation du profil.

- **Systèmes experts :** le profil est établi en observant le flux d'événements pour en déduire un certain nombre de règles qui décrivent le comportement normal du système. Pendant la phase de détection, le système applique les règles au flux d'événements et vérifie si ce flux d'événements respecte ou non les règles apprises.

- **Réseaux de neurones :** le principe repose sur le fait que chaque utilisateur peut être identifié à son comportement. Le profil associé à chaque utilisateur reflète donc ces informations dans le cadre d'une utilisation « normale », c'est-à-dire légitime. Il est possible de représenter efficacement ce profil par un réseau de neurones, conçu pour reconnaître des suites d'opérations caractéristiques de l'utilisateur. Le réseau enregistre les opérations de l'utilisateur durant une fenêtre temporelle donnée, puis tente de prédire la prochaine opération. Un échec de prédiction correspond ainsi à une déviation par rapport au profil et donne potentiellement lieu à une alerte.

  La détection par réseaux de neurones a été largement développée et testée dans les années 80 et 90 [24], mais la plupart des projets se sont soldés par des échecs. Néanmoins, cette méthode fait actuellement l'objet de nouvelles recherches, avec quelques résultats prometteurs.

  Les réseaux de neurones ont des avantages et des inconvénients :

  - **Avantages :** adaptés pour la détection :
    - ✓ Chevaux de Troie ;
    - ✓ Détournement d'identité ;
    - ✓ Contournement d'identification.
  - **Inconvénients :** mise en oeuvre :
    - ✓ Construction du réseau.
    - ✓ Paramétrage du réseau.
    - ✓ Complexité.
    - ✓ Problèmes spécifiques liées aux réseaux de neurones.

- **Approche Immunologique:**
  Cette méthode a été proposée par Forrest et al. [25] et vise à détecter les comportements anormaux d'applications en observant les séquences d'appels système qu'effectue l'application surveillée.
  Pendant la phase d'apprentissage, toutes les séquences d'appels système d'une taille donnée sont stockées dans une base de séquences et constitue le profil.
  Lors de la phase de détection, une alerte est émise lorsqu'une séquence d'appels système effectués par l'application n'est pas présente dans la base de séquences.
  L'approche immunologique a des avantages et des inconvénients:

  - **Avantages :**
    - ✓ Capacités de détecter de nouvelles attaques.
    - ✓ Besoin de peu de maintenance.
  - **Inconvénients :**
    - ✓ Risque d'attaques lors de la construction des profils.
    - ✓ Pas adapté au changement d'entité modélisée.
    - ✓ Evolution des profils au cours du temps peut être vu comme une faille.

  L'avantage majeur de l'approche comportementale par rapport à l'approche par signatures est de ne pas chercher à caractériser les intrusions mais le comportement attendu du système et donc de pouvoir détecter des intrusions inconnues.

  De manière générale, les IDS sont fondés sur cette approche sont fiables car une intrusion génère souvent une anomalie dans le comportement observé. Par contre, ces

  IDS sont, généralement, peu pertinents. Il y a relativement peu d'études de performances des IDS comportementaux en termes de faux positifs.

  De plus, la phase d'apprentissage présente quelques problems pour la plupart des méthodes de modélisation, il faut s'assurer que la base d'apprentissage soit exempte d'intrusions. Dans le cas contraire, l'IDS risquerait d'apprendre des comportements intrusifs et ne serait donc pas capable de les détecter ensuite. Le comportement de l'entité surveillée peut également évoluer au cours du temps, il est possible de modifier le profil pendant la phase de détection pour que ce dernier représente toujours le plus fidèlement possible le comportement de l'entité. Dans ce cas, le système peut apprendre progressivement des comportements intrusifs introduits par un attaquant.

La Figure II.5 montre les différentes étapes de fonctionnement d'un IDS dans les deux approches :

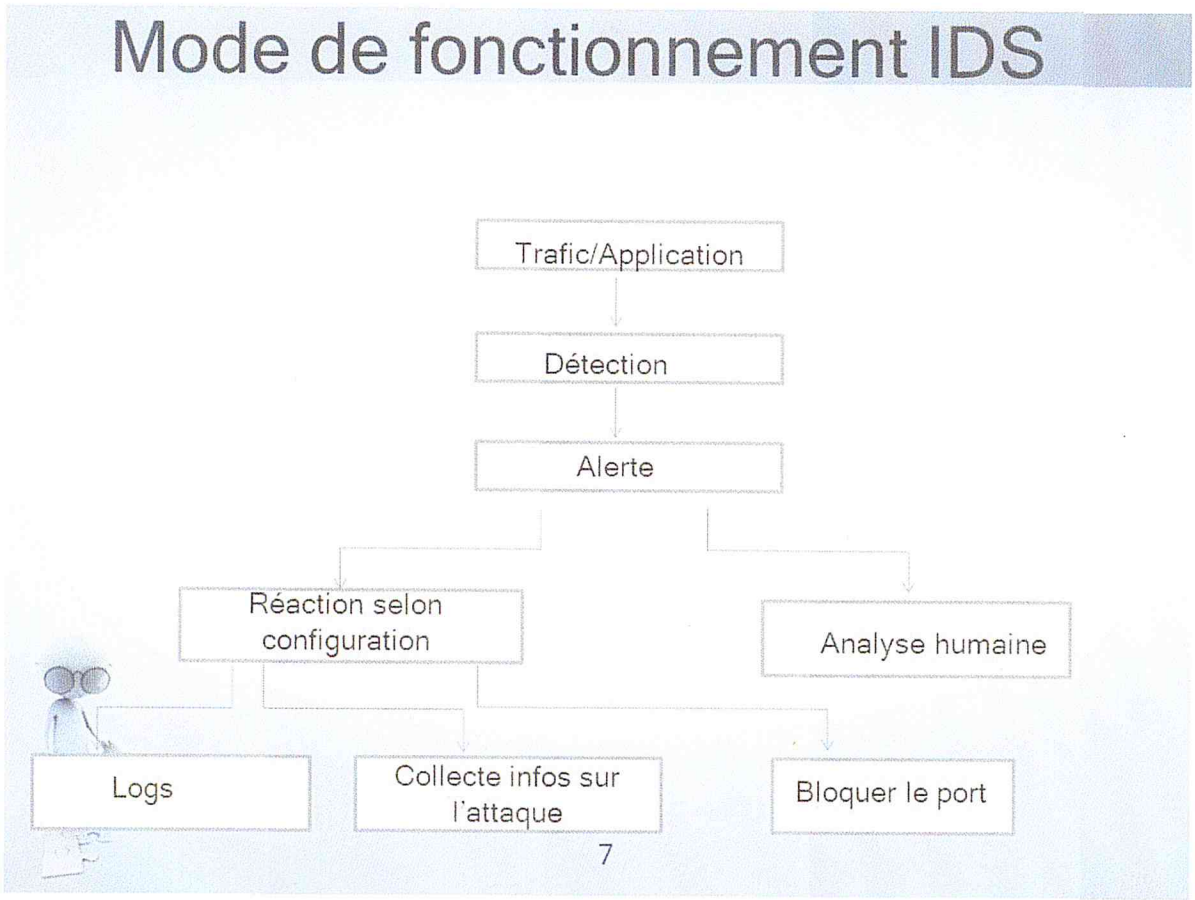


Figure II.5 Fonctionnement d'un IDS.

II.4.1.3 Approche hybride:

L'approche hybride est une approche qui combine les deux approches (l'approche comportementale et l'approche par signatures). Dans un premier lieu l'approche comportementale cherche à trouver de possibles intrusions ensuite ces dernières sont passées à l'approche par signatures pour la mise à jour de sa base.

II.4.2 Classification selon le type de ressources analysées :

Les sources possibles de données à analyser sont une caractéristique essentielle des systèmes de détection d'intrusions. Les données proviennent soit de fichier généré par le système d'exploitation, soit de fichier généré par des applications soit encore d'informations obtenues en écoutant le trafic réseau.

II.4.2.1 Sondes réseau :

Les sondes réseau (NIDS : Network based IDS) récoltent tout événement matérialisé sur le réseau. Ces événements sont soit à l'état brut (non encore interprétés), soit qu'ils ont déjà traversé quelques noeuds (routeurs, commutateurs, pare-feu, points d'accès, etc.) avant d'arriver à la sonde. L'avantage avec ce type de sondes est que l'analyse peut être effectuée sur un système dédié sans affecter les performances du réseau surveillé.

Cependant, avec l'augmentation des débits (on parle aujourd'hui de quelques Gigabits par seconde) l'analyse en temps réel devient de plus en plus difficile à cause du grand volume de données récoltées. L'autre complication pour les sondes réseau est l'utilisation des protocoles cryptographiques, ce qui empêche l'analyse du contenu. La

Figure II.6 représente un exemple d'un NIDS :

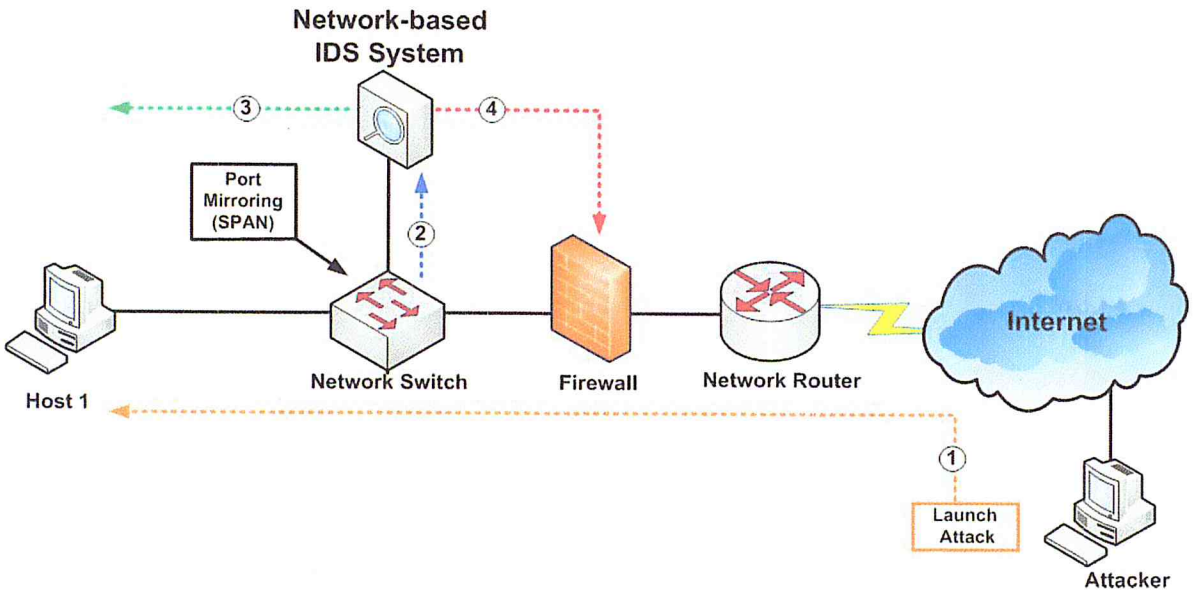


Figure II.6 Exemple de NIDS

II.4.2.2 Sondes systèmes :

Les sondes système (HIDS : Host based IDS) se greffent sur les systèmes surveillés et récoltent seulement les événements matérialisés sur ces systèmes. Ces événements proviennent principalement du noyau et des modules du système d'exploitation. (Voir l'exemple d'un HIDS sur la Figure II.7).

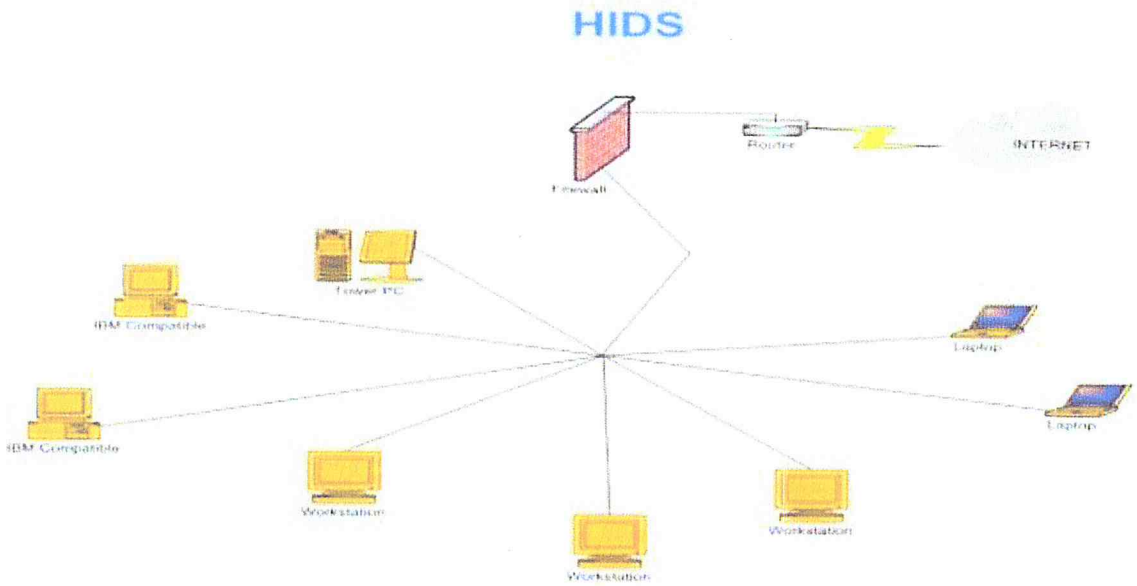


Figure II.7 Exemple de HIDS .

Ces IDS utilisent deux types de sources pour fournir une information sur l'activité : les logs et les traces d'audit du système d'exploitation. Chacun a ses avantages :

Les traces d'audit sont plus précises, détaillées et fournissent une meilleure information ; les logs, qui ne fournissent que l'information essentielle, sont plus petits et peuvent être mieux analysés en raison de leur taille.

L'inconvénient avec ce type de sondes est que les performances des systèmes surveillés sont affectées (ce qui n'est pas le cas avec des sondes réseau). Cette perte en performances est le prix à payer pour une analyse plus précise et plus fine sur les appels systèmes du noyau. En plus, on n'est plus confronté au problème des données chiffrées auquel font face les sondes réseau puisqu'au niveau du système les événements ne sont pas chiffrés.

II.4.2.3 Systèmes de détection d'intrusions hybrides :

Généralement utilisés dans un environnement décentralisé, ils permettent de réunir les informations de diverses sondes placées sur le réseau. Leur appellation hybride provient du fait qu'ils sont capables de réunir aussi bien des informations provenant d'un système HIDS qu'un NIDS.

II.4.2.4 Sondes applicatives :

Les sondes applicatives (AIDS : Application based IDS) surveillent les journaux spécifiques des applications indépendamment du système sur lequel elles se trouvent.

L'avantage est qu'on est à un niveau où plusieurs événements élémentaires sont regroupés ensemble pour former « un plus gros » événement qui est sémantiquement plus riche.

II.4.2.5 IDS basé sur la pile :

Les systèmes de détection d'intrusion basés sur une pile (SBIDS pour Stack-Based

IDS), travaillent étroitement avec la pile TCP/IP, octroient la consultation des paquets lorsqu'ils montent à travers les couches OSI et permettent ainsi à l'IDS de retirer les paquets de la pile avant que le système d'exploitation ou l'application n'ait eu la possibilité d'élaborer la charge virale. L'IDS basé sur une pile peut être efficace contre certaines formes de chiffrement en retraçant les paquets après qu'ils aient été déchiffrés par la pile TCP/IP [26].

II.4.3 Classification par type de réaction :

De toute évidence, un détecteur d'intrusions qui ne réagit pas ne sert pas à grand-chose. Le minimum que doit assurer un tel outil lorsqu'il détecte une attaque est de consigner cette information dans un journal. Après, si on a plus d'ambition, on peut espérer qu'il riposte aux attaques et qu'il identifie et localise de façon précise et complète l'intrus. On distingue alors deux types de détecteurs d'intrusions: passifs et actifs.

II.4.3.1 Détecteurs d'intrusions passifs :

Les systèmes de détection d'intrusions passifs (IDS) sont seulement par rapport aux attaques. Cela dit, ils peuvent effectuer plusieurs opérations de façon active comme récolter des informations sur l'attaquant, alerter l'administrateur en lui envoyant un courrier électronique, produire un rapport sur les attaques détectées, etc.

II.4.3.2 Détecteurs d'intrusions actifs :

Ont pour objectif d'empêcher (dans la mesure du possible) le succès d'une attaque ou du moins de limiter son impact. Le terme IPS (Intrusion Prevention Systems) est de plus en plus utilisé, dans la presse spécialisée et par les services commerciaux des compagnies œuvrant dans ce domaine, pour désigner les détecteurs d'intrusions actifs.

Par opposition aux détecteurs d'intrusions passifs, les détecteurs actifs peuvent changer la configuration des systèmes ou du réseau. Ainsi, les réponses peuvent aller de la simple déconnexion de l'intrus, l'arrêt d'un processus: le verrouillage d'un compte utilisateur, jusqu'au changement des règles de filtrage d'un pare-feu ou même des règles de routage d'un routeur, Pour plus d'efficacité, les IPS sont généralement positionnés en coupure dans l'architecture du réseau (comme un pare-feu).

Un IPS possède de nombreux inconvénients. Le premier est qu'il bloque toute activité qui lui semble suspecte. Or, il est impossible d'assurer une fiabilité à 100% dans l'identification des attaques. Les faux positifs sont donc très dangereux pour les IPS. Le deuxième inconvénient est qu'un pirate peut utiliser sa fonctionnalité de blocage pour mettre hors service un système. Et enfin, le troisième inconvénient et non le moindre: un IPS est peu discret. En effet, à chaque blocage d'attaque, il montre sa présence. Cela peut paraître anodin, mais si un pirate remarque la présence d'un IPS, il tentera de trouver une faille dans celui-ci afin de réintégrer son attaque, mais cette fois en passant inaperçu.

Voilà pourquoi les IDS passifs sont souvent préférés aux IPS. Cependant, il est intéressant de noter que plusieurs IDS ont été dotés d'une fonctionnalité de réaction automatique à certains types d'attaques. [27]

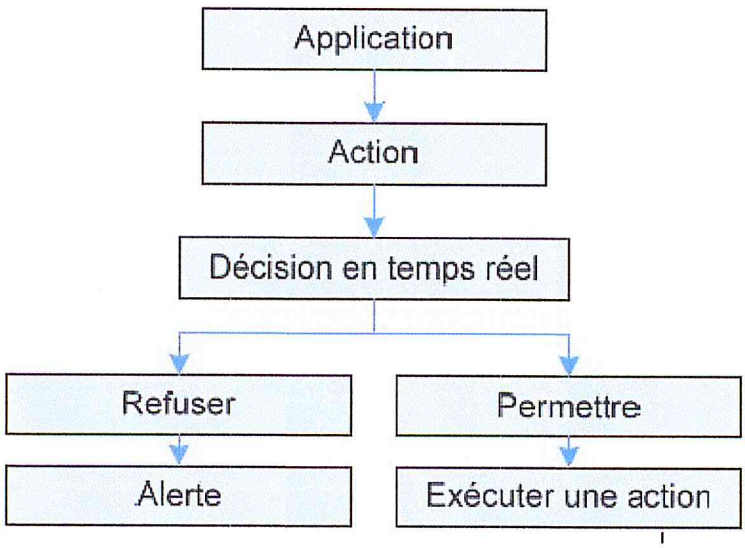


Figure II.8 Fonctionnement d'un IPS.

Enfin, il est possible d'unir ces deux méthodes pour obtenir un système hybride comme le système de la Figure II.9. [27]

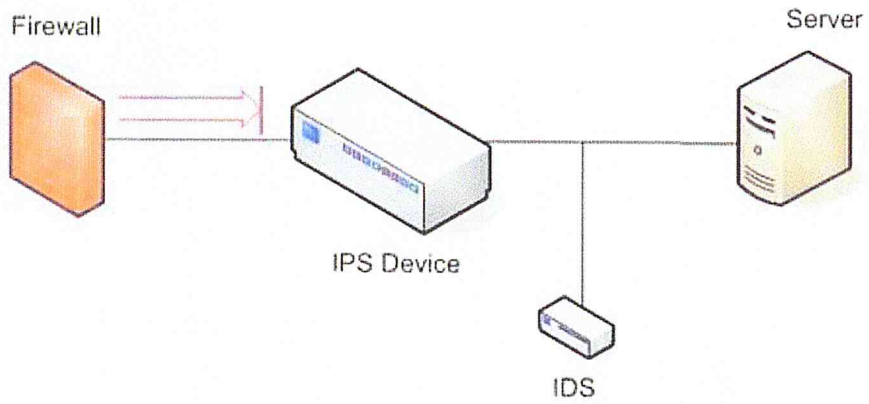


Figure II.9 Exemple d'un système hybride

II.4.4 Classification par mode d'utilisation :

Selon les besoins, on peut choisir le mode d'utilisation d'un détecteur d'intrusions.

II.4.4.1 Analyse en temps réel :

L'analyse en temps réel (online) est une analyse effectuée « au fur et à mesure de la réception des événements » [28]. Le principal avantage est que les alertes sont lancées dès que les attaques sont détectées. Cet état de veille coute cher en termes de ressources et nécessite des algorithmes plus complexes que ceux d'une analyse en différé.

II.4.4.2 Analyse en différé :

L'analyse en différé (offline) est une analyse effectuée sur des données stockées (non fraîches). Vu le caractère non-pressant de ce type d'analyse (à tête reposée), les algorithmes sont plus simples, sauf que la détection des attaques se fait après coup et nécessite le stockage des événements avant leur analyse.

II.5 Détection d'intrusions Web :

Les serveurs Web sont un environnement de test intéressant pour la détection d'intrusions, d'une part, par leur importance et par l'universalité du protocole HTTP et d'autre part, par le nombre de vulnérabilités les frappant.

Les serveurs Web sont la vitrine des entreprises, associations, états, voir des individus par l'intermédiaire des blogs sur Internet. Ils sont, dans certains cas, une source de revenus importants (commerce en ligne par exemple). De plus en plus d'applications Web sont déployées sur Internet.

Les outils de détection d'intrusions utilisés pour détecter les attaques contre les serveurs Web utilisent principalement une approche par signatures bien que des approches comportementales soient apparues récemment. Nous allons donc présenter les différents IDS spécifiques au Web suivant leur approche de détection.

II.6 Discutions générale:

Un système de détection d'intrusion peut être divisé en deux approches qui sont les comportements sur la base (anomalie) et fondée sur la connaissance.

L'approche de comportement en fonction est également connue comme système basé sur une anomalie alors que l'approche par scénario basée sur la connaissance est connue comme système basé sur l'utilisation abusive. L'utilisation par scénario ou la signature sur la base IDS est un système qui contient un certain nombre de la description d'attaque ou de signature qui sont en correspondance avec un flux de données d'audit à la recherche de preuves d'attaque modélisée.

Les données d'audit peuvent être recueillies à partir du trafic réseau ou un journal d'application. Cette méthode peut être utilisée pour détecter précédente attaque connue et le profil de l'attaquant doit être révisé manuellement lorsque de nouveaux types d'attaque sont découverts. Par conséquent, les attaques inconnues dans le profil d'intrusion réseau et caractéristique pourrait ne pas être prise en utilisant cette technique, le système basé sur l'anomalie identifie l'intrusion en identifiant le trafic ou l'application qui est présumée être une activité normale sur le réseau ou l'hôte. Le système basé sur des anomalies construit un modèle qui cherche une activité anormale telle que les activités qui ne confirment pas au modèle établi.

La mise en œuvre du système de détection d'intrusion réactive. Par conséquent, il est important de réduire les fausses alerts générées par les deux techniques Du système. Bien que fausse alerts est une préoccupation majeure dans le développement du système de détection d'intrusion en particulier le système de détection d'intrusion d'anomalie sur la base, mais le système a pleinement atteint. les organisations objectif par rapport à la Le succès d'un IDS dépend de la décision sur un ensemble de caractéristiques que le système va utiliser pour détecter l'attaquant en particulier les attaques rapides. En effet, le mécanisme d'une attaque rapide ne nécessite que quelques secondes et la technique utilisée par l'attaquant pour lancer l'attaque est également différent. Au meilleur de notre connaissance, il n'y a pas de classification complète des caractéristiques de ce système de détection d'intrusion peut utiliser pour détecter les attaques de réseau sur la base des attaques particulièrement rapides.

Paramètres	IDS par scénario	IDS comportementale
Les faux négatifs	par définition ne détecte pas les nouvelles attaques qui conduisent à Faux négatif	Taux de fausse alerte négative est relativement faible
Mise à jour	Il nécessite la mise à jour régulière de la base de données, de sorte qu'il peut détecter une nouvelle attaque	Aucune mise à jour régulière n'est nécessaire
Sensibilité à l'attaque	il trouve des problèmes de détecter les attaques évasives	Taux de détection à des techniques évasives sont nettement plus élevés par rapport à la signature sur la base.
Couverture des réseaux IDS	Il est capable de détecter les agressions extérieures, mais il ne peut pas détecter attaque interne.	Il peut détecter les attaques internes et externes.

Ce tableau présent une comparaison entre l'approche comportementale et celle par scénario. on voit que l'IDS avec l'approche comportementale et plus efficace et fiable que l'IDS avec l'approche par scénario sur plusieurs niveaux : le taux de faux négatifs, la sensibilité et la couverture de réseaux, ce qu'il motive le choix de l'IDS comportementale sur le compte de l'IDS par scénario.

II.7 Conclusion:

La plupart des IDS sont fiables, Ce qui explique qu'ils sont souvent intégrés dans les solutions de sécurité. Les avantages qu'ils présentent face aux autres outils de sécurités les favorisent, mais d'un autre côté cela n'empêche pas que les meilleurs IDS présentent aussi des lacunes et quelques inconvénients. Nous comprenons donc bien qu'ils sont nécessaires mais ne peuvent pas se passer de l'utilisation d'autres outils de sécurité visant à combler leurs défauts, ces defaut qui peuvent nuire l'utilisation efficace et reduire la fiabilité de ce type de system, ca nous pousse de chercher des solution pour certain de ces defauts

CHAPITRE III

CONCEPTION

III.1 Introduction:

Ce chapitre présente la conception de notre choix de K-means, pour réaliser et claustré les informations des utilisateurs selon des clusters indépendant qui va permet par la suit de détecter les intrusions. Cette partie consiste à décrire les spécifications de notre algorithme ainsi que les principes sur lequel l'application va être réalisée

III.2 IDS de détection d'anomalies :

Afin d'éviter les utilisations malveillantes de l'application Web réalisée, il faut développer un système capable de détecter et d'identifier les intrusions. Pour éviter les tâches fastidieuses de mise à jour de la base des modèles d'intrusions, notre système doit pouvoir s'adapter de manière autonome pour intégrer dynamiquement la détection de nouvelles intrusions en se basant sur l'approche comportementale qui comporte deux phases:

III.2.1 Phase d'apprentissage :

Dans cette phase, chaque client sera orienté dans une classe à base de son profile. La classification des clients est faite par rapport aux critères suivants : temps de connexion, prix moyen d'achat, fréquence d'achat ;

- **Temps de connexion T :**
  Représente la moyenne des temps de connexions d'un Client, trois choix sont possibles, T1 s'il ne dépasse pas le temps minimale, T3 s'il est supérieur au temps maximale et T2 s'il est entre T1 et T3.
- **Prix moyen d'achat P :**
  Deux choix sont possibles, P1 si le prix est inférieur au prix moyen et P2 si le prix est supérieur au prix moyen.
- **Fréquence d'achat F:**
  On trouve 2 choix, F1 si le client ne dépasse pas le nombre Moyenne d'article achetés et F2 s'il dépasse le seuil definie d'achat.

Ces trois critères donnent naissance à douze classes différentes, chaque client va être classé après la phase d'apprentissage dans la classe correspondante parmi les classes créées,

La Figure III.1 montre les différentes classes possibles pour un client.

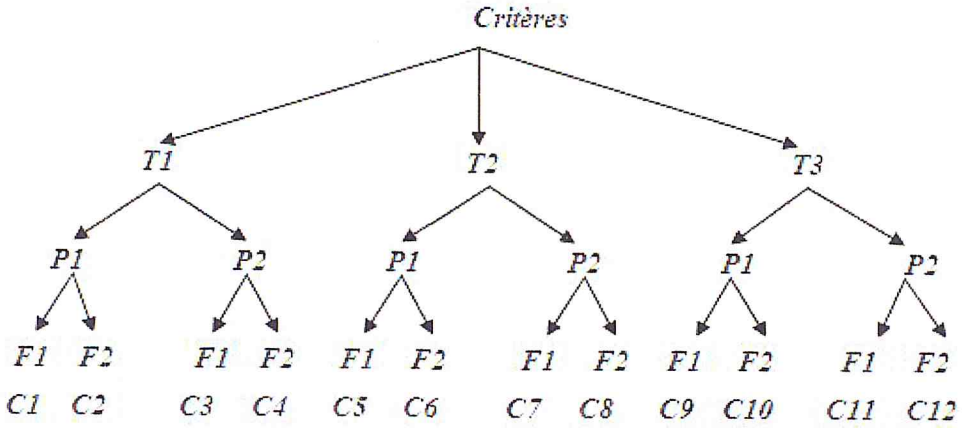


Figure III.1 Classification d'un client en fonctions des 3 critères.

III.2.2 Phase de détection :

Cette étape est valable uniquement pour les anciens clients (qui ont été déjà classés). Si l'un de ces clients vient de se connecter une nouvelle fois, l'IDS va le suivre pour récupérer son profil grâce au temps de connexion, prix d'achat et fréquence d'achat afin d'obtenir son nouveau comportement (sa nouvelle classe), pour enfin mesurer la similarité entre son nouveau comportement et son profil déterminé dans la phase d'apprentissage.

Si sa nouvelle classe est différente de la classe ancienne, l'IDS considère le changement de classe comme attaque.

III.3 la méthod utilisé :

Les IDS comportemental utilise une simple analyse de données dont le principe est de minimiser les faux positifs à base d'un seuil qui égal à 25%. Ce seuil représente la valeur max tolérée d'un changement de profil pour un client donné par rapport à son propre profil pour parler d'un faux positif. Si le seuil est dépassé, il s'agit d'une attaque.

Alors Plus le seuil est petit, plus le nombre de faux positif est réduit.

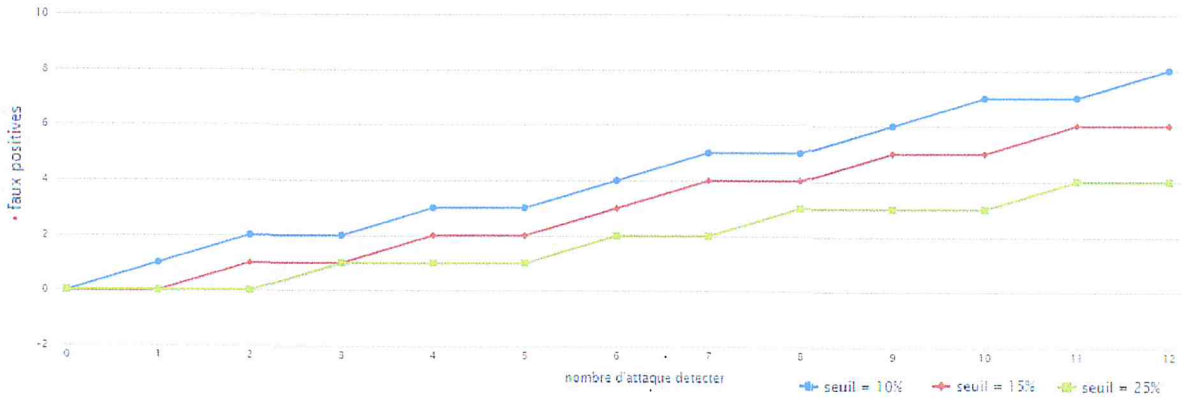


Figure III.2 Variation de faux positifs en fonction de nombre d'attaques.

Cette variation ne présente pas tout la realité, par exemple l'augmentation du taux de faux positive réduit le taux de faux négative et vice versa.

Donc le choix du seuil est relié avec les exigences professionnels de chaque system, ce choix et le choix de la fiabilité au surcharge de la pertinence ou la pertinence au surcharge de la fiabilité,

En effet, on veut améliorer ce travail en utilisant un algorithme qui permet de réduire le nombre de faux positifs et négative.

III.4 L'algorithme de clustering K-means:

L'algorithme de clustering K-means est l'un des plus simples algorithmes qui permettent de résoudre le problème de classification bien connu. Cet algorithme de clustering a été développé par MacQueen en 1967. K-means vise à partitionner un ensemble de données fournies en clusters (k grappes), où k est une constante prédéfinie ou définie par l'utilisateur. L'idée principale est de définir k centres de gravité, pour chaque cluster.

k-means est un algorithme itératif qui minimise la somme des distances entre chaque individu et le centroïde. Le choix initial des centroïdes conditionne le résultat final. Admettant un nuage d'un ensemble de points. Afin de construire des catégories de ce nuage de points, k-Means change les points de chaque cluster jusqu'à ce que la somme ne puisse plus diminuer. Le résultat est un ensemble de clusters compacts et clairement séparés, sous réserve de choisir la bonne valeur k du nombre de clusters [29].

Explication de l'algorithme :

L'application de l'algorithme k-Means se fait en suivant les étapes suivante :

1) Choix de k, le nombre de cluster à créer.

2) Choix des centres des clusters da manière aléatoire à partir des objets en entrée. a

La procédure adoptée pour le choix des centres des Clusters initiaux est extrêmement importante car elle a un impact direct sur le résultat final du Clustering. Il est donc très important de choisir des clusters bien séparés.

Algorithme :[29]

Entrée

Ensemble de N données, noté par x

Nombre de groupes souhaité, noté par k

Sortie

Une partition de K groupes $\{C_1, C_2, \ldots, C_k\}$

Début

1) Initialisation aléatoire des centres c_k;

Répéter

2) Affectation : générer une nouvelle partition en assignant chaque objet au groupe dont le centre est le plus proche ;

$$x_i \in C_k \; si \; \forall j \; |x_i - \mu_k| = \min_j |x_i - \mu_j|$$

Avec le centre de la classe K ;

3) Représentation : Calculer les centres associe à la nouvelle partition ;

$$\mu_k = \frac{1}{N} \sum_{x_i \in C_k} x_i$$

Jusqu'à convergence de l'algorithme vers une partition stable ;

Fin.

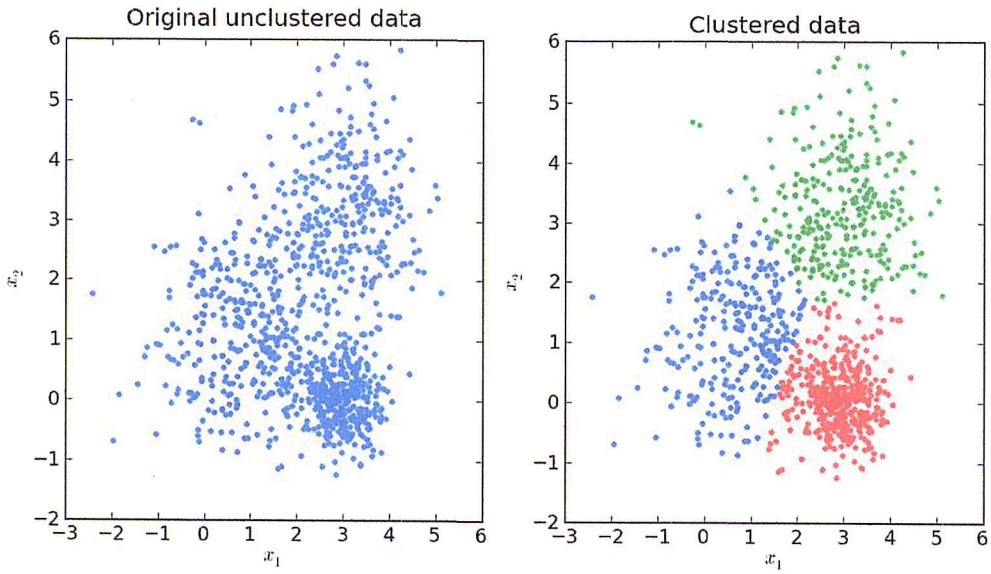


Figure III.3 Sélection des centres.

1- Parcourir tous les objets afin de les affecter ou les réaffecter au cluster approprié en se basant sur la minimisation de la distance entre l'objet ET le centre du cluster.

2- Calculer les centres de chaque cluster puisqu'ils peuvent changé après affectation des objets.

3- Refaire les étapes (1) et (2) jusqu'aucun changement du calcul des centres des clusters ou une stabilité des objets.

III.5 Organigramme :

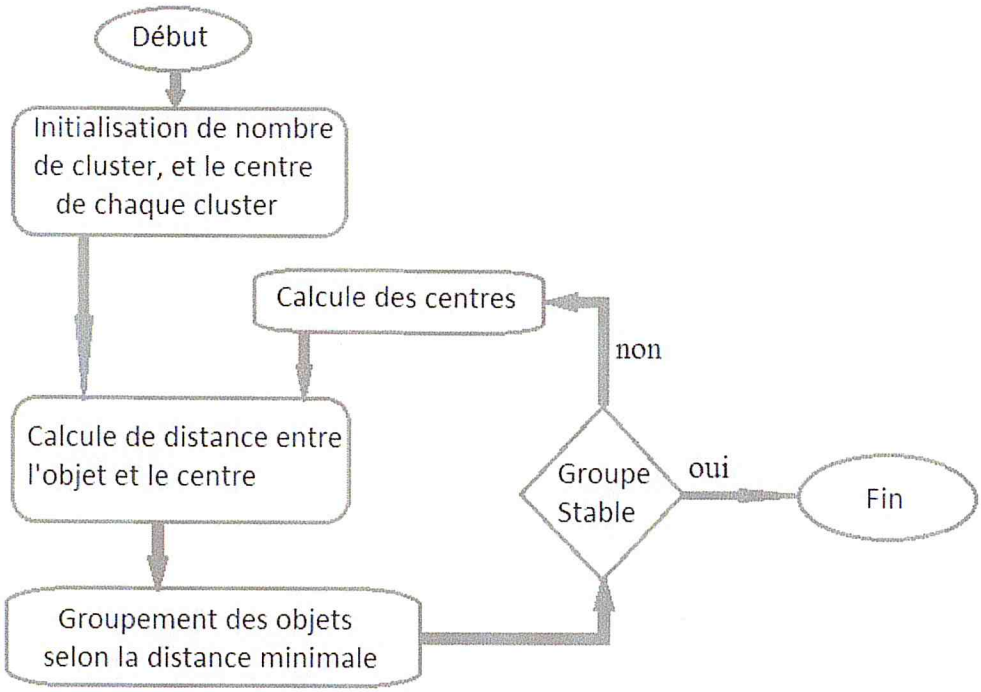


Figure III.4 Organigramme de l'algorithme k-means.

L'algorithme consiste à grouper les points selon un critère bien déterminé.

L'entrée de l'algorithme est le nombre k de groupes (cluster). Une fois le nombre de groupes saisi, l'algorithme choisi arbitrairement k points comme centres «initiaux » des k groupes.

L'étape suivante consiste à calculer la distance entre chaque individu (point) et les k centres, la plus petite distance est retenue pour inclure cet individu dans le groupe ayant le centre le plus proche.

Une fois tous les individus groupés, on aura k sous-nuages (cluster) disjoints du nuage total. Pour chaque groupe, l'algorithme calcule le nouveau centre de gravité. L'algorithme s'arrête lorsque les groupes construits deviennent stables.

III.6 Implémentation de l'algorithme dans l'application web:

L'algorithme k-means est très populaire du fait qu'il est très facile à comprendre et à mettre en œuvre. Il permet de regrouper les points en cluster. Dans notre projet, on a des classes et non pas des points. Donc le problème qui se pose dans notre cas c'est convertir les classes en points qui ont des dimensions (X, Y). Pour cela, on a proposé de calculer la moyenne des trois critères de chaque classe.

Par exemple pour le **Temps T**, on a supposé que :

- Si T<X1 : la moyenne du T=X1/2 mn ;
- Si X1<T<X2 : la moyenne du T=(X1+X2)/2 mn ;
- Si T>X2 : la moyenne du T=X2+X1 mn ;
-

Pour le **Prix d'achat P**, on a supposé que :

- Si P<Y1 : la moyenne du P=Y1/2 € ;
- Si P>Y1 : la moyenne du P=Y1+Y1/2 € ;

Pour le **Nombre de produits achetés F**, on a supposé que :

- Si F<Z1 : la moyenne du F=Z1/2 ;
- Si F>Z1 : la moyenne du F=Z1+Z1/2 ;

D'après le résultat de l'algorithme de clustering k-means, les clusters sont les suivant:

Cluster1= {C1, C2}

Cluster2= {C5, C6, C9, C10}

Cluster3= {C3, C4, C7, C8, C11, C12}

Dans la phase de détections, la vérification de comportement de nos clients reste toujours entre les class des clients (les class ancien et les nouvelle class), et plus précisément entre les clusters de chaque client (ancien cluster avec le nouveaux cluster)

Le changement de class ne déclenche aucun alerte, par contre le changement de cluster déclenche une alerte et bloque la procédure d'achat.

III.7 Diagramme récapitulatif :

En résumé, les étapes pour sécuriser l'application Web, sont décrites une à une dans l'organigramme suivant :

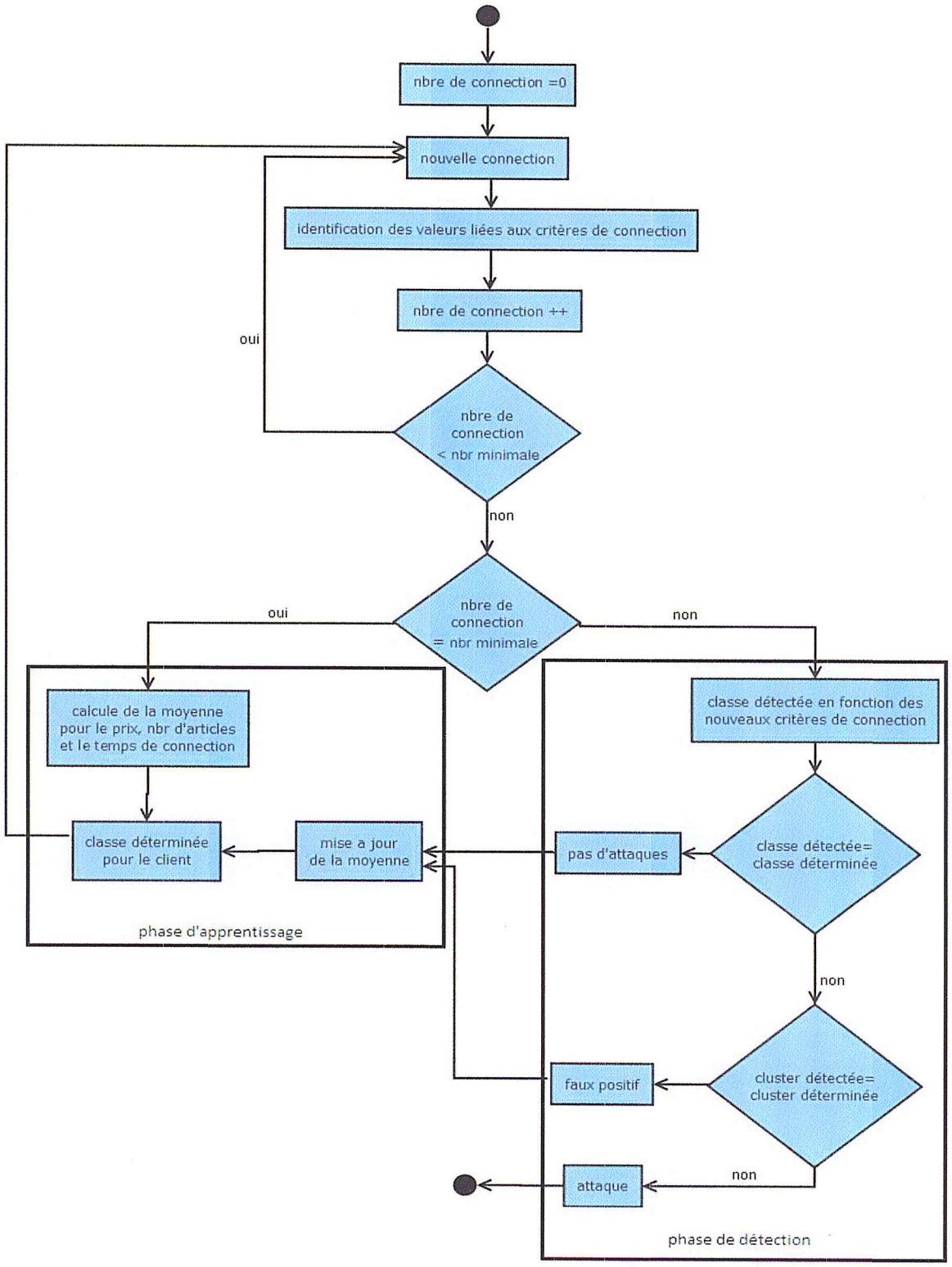


Figure III.5 Organigramme de détection d'une attaque

Ce diagramme présent la procédure globale de notre system, l'utilisation de ce system réduit le nombre de fauss alertes et spécialement les faux positifs, avec une mise à jour régulière de notre base de données, qui permis de suivre nos clients et avoir des résultats plus précises

III.8 Résultat obtenu avec le K-means :

D'après l'implémentation de l'algorithme de clustering K-means et le résultat obtenu, la figure suivante montre la variation de faux positifs dans l'approche comportementale et l'approche comportementale avec l'algorithme k-means.

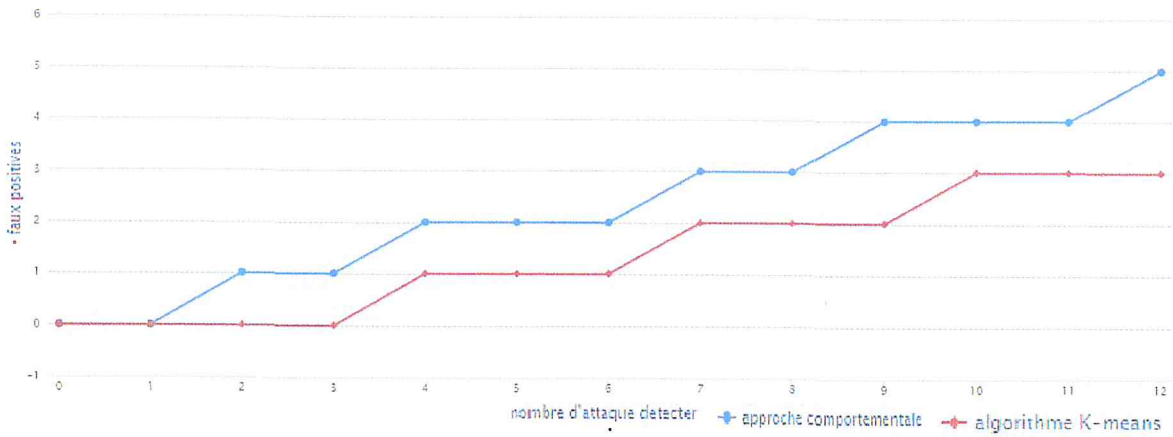


Figure III.6 Nombre d'attaques détectées

La courbe montre que l'algorithme k-means permet de détecter plus d'attaques et mois de faux positifs par rapport à l'approche comportementale car l'algorithme k-means permet de classer les objets dans les clusters, donc si le client change leur comportement dans le même cluster aucune attaque ni faux positifs sera déclenchée.

Conclusion:

Au cours de ce chapitre on a étudié l'approche comportementale qui se base sur l'hypothèse qu'on peut définir un comportement « normal » de l'utilisateur et que toute déviation par rapport à celui-ci est potentiellement suspecte. L'inconvénient de cette approche c'est les risques des fausses alertes (faux positifs). En fin on a implémenté un algorithme de clustering k-means qui permet de regrouper les classes en k-clussters afin de diminuer le nombre de faux positifs.

CHAPITRE IV

REALISATION

IV.1 Introduction :

Ce chapitre présente la description d'une solution proposée dont le but est de sécuriser une application web. Cette partie comprend trois étapes. La première étape consiste à décrire la réalisation en détail de cette application Web (boutique en ligne) La deuxième étape consiste à sécuriser l'application Web réalisée en se basant sur l'approche comportementale des IDS et enfin la dernière étape consiste à appliquer l'algorithme K-means afin de diminuer le nombre des faux-positifs.

IV.2 Outils de réalisation :

Cette partie présente les principaux outils utilisés pour la mise en place de l'application. La réalisation de cette dernière a été faite sous la plateforme Java et PHP en utilisent netbeans et WAMP server.

IV.2.1 PHP :

HyperText Preprocessor, plus connu sous son sigle **PHP** est un langage de programmation orienté objet compilé libre.

Il est Principalement utilisé pour produire des pages web dynamiques via un serveur généralement Apache, mais pouvant également fonctionner comme n'importe quel langage interprété de façon locale. PHP est un langage impératif disposant depuis la version 5.

IV.2.2 CSS

Le terme CSS est l'acronyme anglais de Cascading Style Sheets qui peut se traduire par "feuilles de style en cascade". Le CSS est un langage informatique utilisé sur internet pour mettre en forme les fichiers HTML ou XML. Ainsi, les feuilles de style, aussi appelées, les fichiers CSS, comprennent du code qui permet de gérer le design d'une page en HTML.

IV.2.3 HTML

HTML est un langage informatique utilisé pour créer des pages web. L'acronyme signifie

HyperText Markup Language, ce qui signifie en français "langage de balisage d'hypertexte".

Cette signification porte bien son nom puisqu' effectivement ce langage permet de réaliser de l'hypertexte à base d'une structure de balisage.

Ce n'est pas à proprement parler d'un langage de programmation, mais plutôt d'un langage qui permet de mettre en forme du contenu. Les balises permettent de mettre en forme le texte et de placer des éléments interactifs, tel des liens, des images ou bien encore des animations.

IV.2.4 Java :

Les modules conçus ont été réalisés sous Java dont les principales vertus, sont résumées dans les points suivants :

- Java est un langage orienté objet : un programme Java est centré complètement sur les objets et fournit un ensemble prédéfini de classes facilitant la manipulation des entrées-sorties, la programmation réseau, système, graphique…
- Le langage Java est distribué : il est conçu pour développer des applications en réseau, les manipulations des objets distants ou locaux se font de la même manière.
- Le langage Java est robuste et sûr : il est fortement typé ; il élimine bien des erreurs d'incohérence de type à la compilation et ne supprime pas tous les problèmes de sécurité mais les réduit fortement.
- Le langage Java est interprété : un programme Java n'est pas compilé en code machine ; il est transformé en code intermédiaire interprété.
- Le langage Java est portable et indépendant des plates-formes : un IDS ne doit ni dépendre de l'architecture matérielle, ni du système d'exploitation.

On a utilisé NetBeans version 7.4 qui est placé en open source par Sun sous licence CDDL (Common Development and Distribution License). En plus de Java, NetBeans permet également de supporter différents langages, comme C, C++, XML et HTML. Il comprend toutes les caractéristiques d'un IDE moderne (éditeur en couleur, projets multi-langage, refactoring, éditeur graphique d'interfaces et de pages Web).

IV.2.5 Choix de MySQL :

MySQL est un serveur de BD relationnelles open-source qui stocque les données dans des tables séparées plutôt que de tout rassembler dans une seule table. Cela améliore la rapidité et la souplesse de l'ensemble. Les tables sont reliées par des relations définies, qui rendent possible la combinaison de données entre plusieurs tables durant une requête. Le SQL (Structured Query Language) : le langage standard pour les traitements de bases de données .

IV.3 Réalisation de l'application Web :

IV.3.1 Description de boutique en ligne :

La boutique en ligne réalisée est sous le nom de Moonboutique, elle propose une sélection de matériaux et de logiciels informatiques, elle contient les éléments suivants :

- Un catalogue électronique en ligne, présentant l'ensemble des produits disponible à la vente ;
- Un moteur de recherche permettant de trouver facilement un produit à l'aide de critères de recherche (Product Name) ;

- Un système de caddie virtuel (appelé parfois panier virtuel). Ce dernier permet de conserver la trace des achats du client tout au long de son parcours et de Modifier les quantités pour chaque référence ;

La Figure IV.1 présente la page d'accueil de Moonboutique :

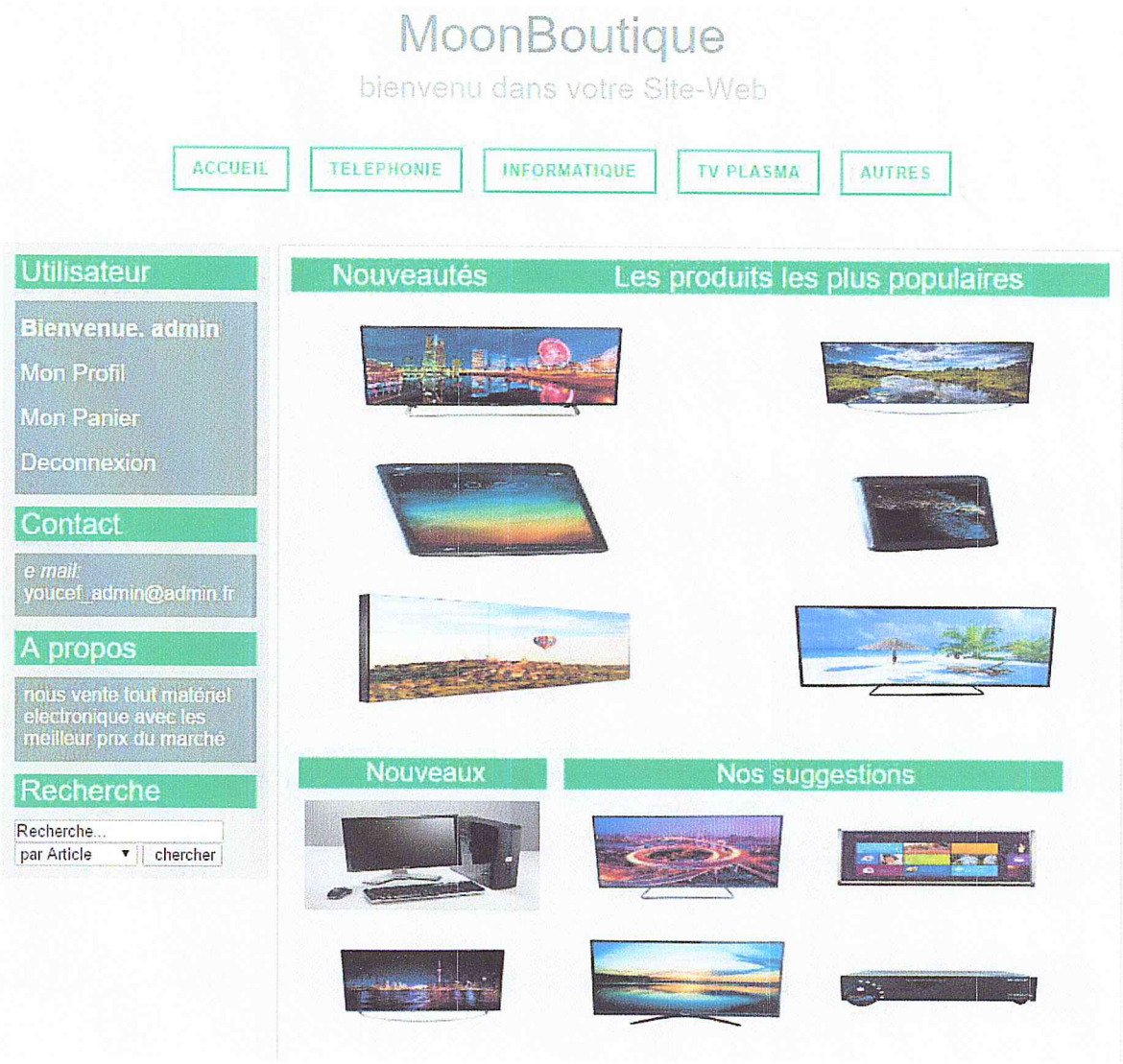


Figure IV.1 Page d'accueil d'InformatiqueBoutique.

A travers une boutique en ligne et comme dans un magasin réel, on peut choisir et payer des articles. L'acheteur (qui est obligatoirement un client) doit s'inscrire à cette boutique si c'est un nouveau client ;

La Figure IV.2 présente l'inscription d'un client :

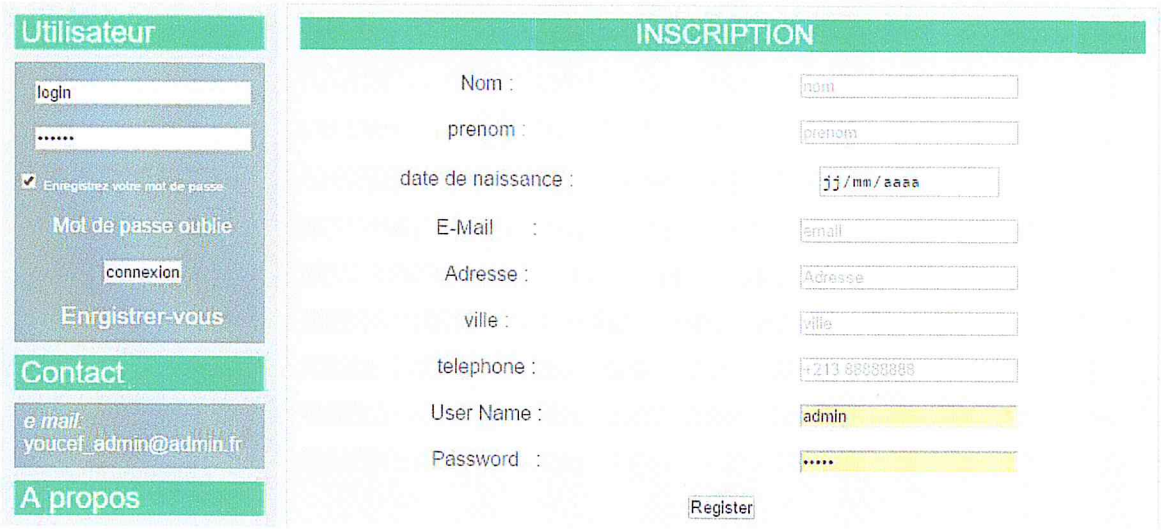


Figure IV.2 Inscription d'un client

Si le client est déjà inscrit (client ancien) il suffit donc de se connecter avec son Pseudo et son mot de passe comme il est montré dans la figure suivante :

Figure IV.3 Connexion d'un client

Quand le client veut savoir les détails d'un produit il suffit uniquement de glisser le curseur vers ce dernier. S'il veut l'acheté, il doit cliquer sur ce produit (ajouter a mon panier)

Figure IV.4 détails produit.

et automatiquement va s'ajouter au panier virtuel (basket) ;

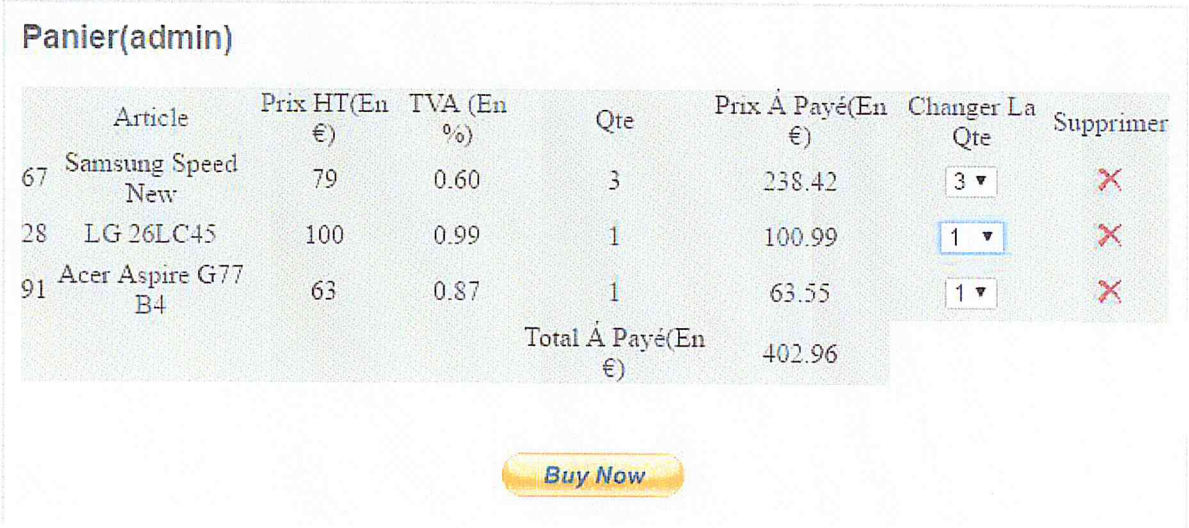


Figure IV.5 panier virtuel.

On peut ajouter ou modifier ou même supprimer les produits facilement depuis notre page d'administration

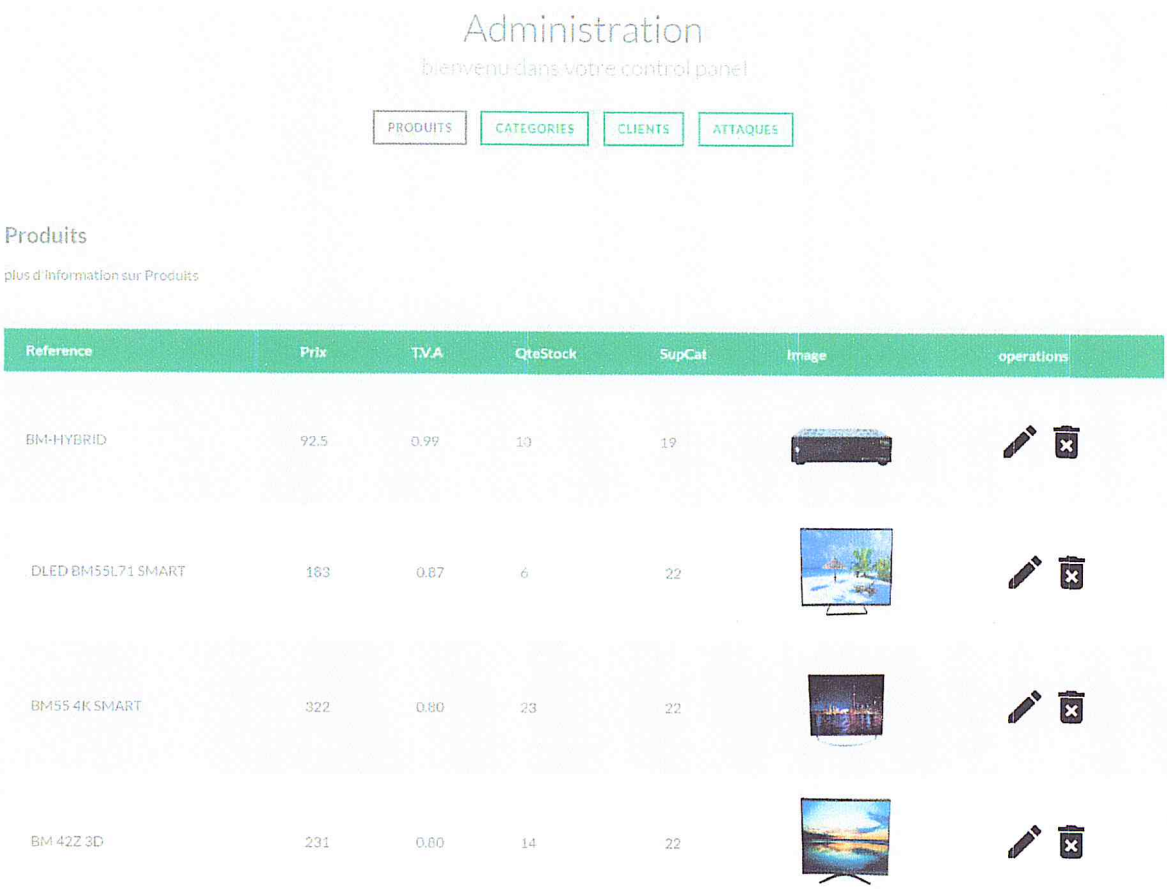


Figure IV.6 Administration des produits

Les produits de notre boutique online sont classés par catégories, on peut par la suit manager ces catégories

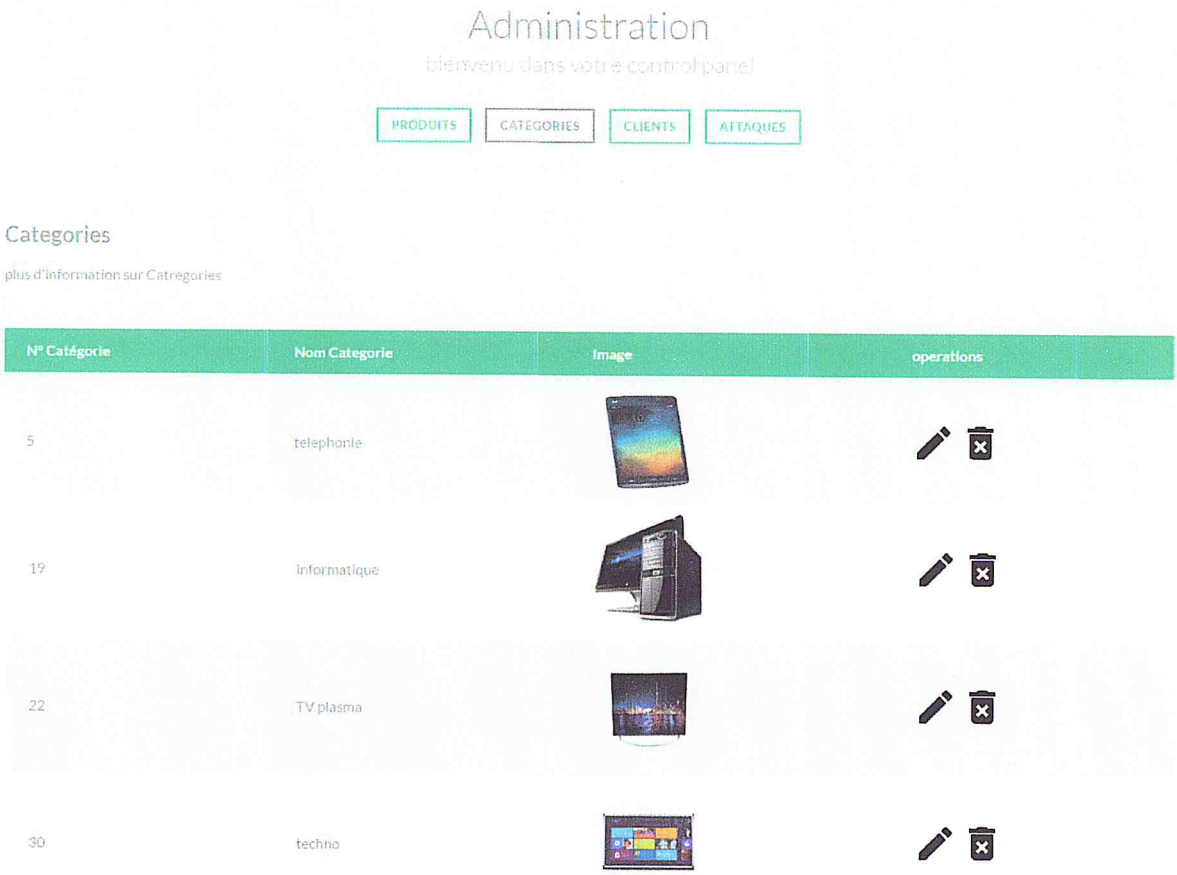


Figure IV.7 Administration des catégories

La gestion des Clients permet de vérifier les informations personnelles de nos clients, et les contacter en cas d'un bloque temporaire de leur comptes

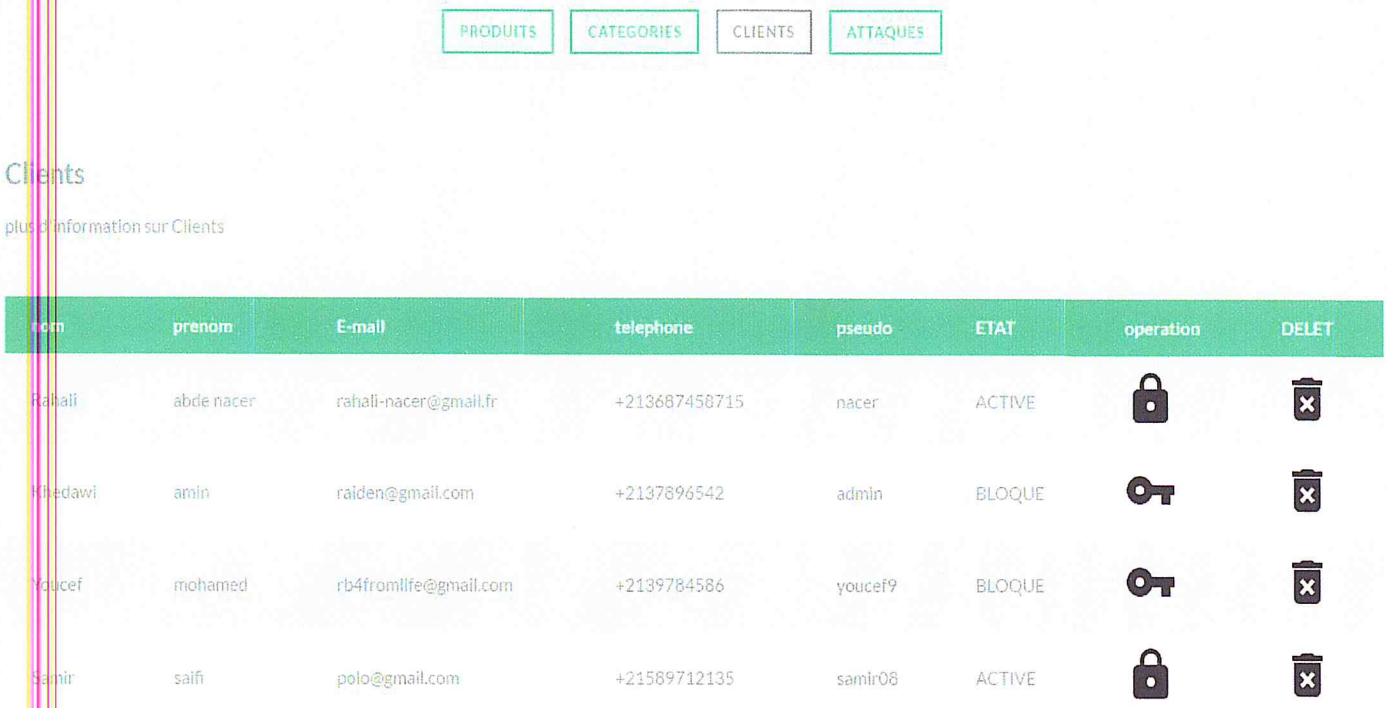


Figure IV.8 Administration des clients

IV.4 Réalisation de l'IDS avec Kmeans:

Pour la réalisation de de notre IDS utilisent l'algorithme Kmeans, on doit définir chaque classe (intervalle d'utilisation) par un point, alors on va convertir les classes en points qui ont des dimensions (X, Y). Pour cela, on a proposé de calculer la moyenne des trois critères de chaque classe selon la méthode suivante:

Par exemple pour le **Temps T**, on a supposé que:

- Si T<60 : la moyenne du T=30 mn ;
- Si 60<T<120 : la moyenne du T=90 mn ;
- Si T>120 : la moyenne du T=150 mn ;

Pour le **Prix d'achat P**, on a supposé que:

- Si P<500 : la moyenne du P=250 € ;
- Si P>500 : la moyenne du P=750 € ;

Pour le **Nombre de produits achetés F**, on a supposé que:

- Si F<5 : la moyenne du F=3 ;
- Si F>5 : la moyenne du F=8 ;

Donc d'après cette étape, les classes seront transférées en points suivants:

Classes	Points
C1 (T<60, P<500, F<5)	(30, 250, 3)
C2(T<60, P<500, F>5)	(30, 250, 8)
C3 (T<60, P>500, F<5)	(30, 750, 3)
C4 (T<60, P>500, F>5)	(30, 750, 8)
C5 (60<T<120, P<500, F<5)	(90, 250, 3)
C6 (60<T<120, P<500, F>5)	(90, 250, 8)
C7 (60<T<120, P>500, F<5)	(90, 750, 3)
C8 (60<T<120, P>500, F>5)	(90, 750, 8)
C9 (T>120, P<500, F<5)	(150, 250, 3)
C10 (T>120, P<500, F>5)	(150, 250, 8)
C11 (T>120, P>500, F<5)	(150, 750, 3)
C12 (T>120, P>500, F>5)	(150, 750, 8)

Figure III.9 Tableau de conversation des classes en points.

Après qu'on a implémenté cet algorithme à l'application web, avec une initialisation de 12 points (classes) et de 3 clusters.

IV.4.1 Période d'apprentissage:

Les nouveaux utilisateurs dans notre application de vente en ligne (qui en mois de 30 connexion) doivent être suivre par l'enregistrement de leur comportement dans la base de donnée pour détecter leur class et leur utilisation normal de l'application

Pour notre projet on a déjà supposé qu'on a des clients avec leur profile (sa classe ancienne).

La table suivante montre le comportement des clients :

pseudo	Classe_ancienne	Classe_nouvelle	prix	nbr	temps
jonson	6	6	385	7	87
youcef,	5	5	430	2	69
admin	1	1	263	3	35
nacer	7	7	948	4	87

Figure IV.10 Comportement des clients

Le champ **classe_ancienne** a une valeur de classe du profile pendant la phase d'apprentissage. Le champ **classe_nouvelle** a une valeur de numéro de la nouvelle classe correspondante au comportement d'un client après une nouvelle connexion. Mais dans cette table le champ classe_nouvelle = classe_ancienne car le client n'a pas encore changé son profil. Le champ **temps** présente le temps de connexion d'un client(le temps de connexion T).

Le champ **nbr**présente le nombre de produits achetés (la fréquence F) ;

Le champ **prix** présente le prix total des produits achetés (le prix moyen d'achat) ;

IV.4.2 Période de détection :

Pour les anciennes utilisateurs de notre application de vente en ligne (31 connexion au minimum) Le system va faire la vérification du comportement des clients a chaque nouvelle connexion et détecter leur nouvelle classe et la comparais avec leur profil obtenu dans la phase d'apprentissage (ou dans les mises à jour d'utilisation) afin de détecter une attaque en cas d'un grand changement

Par exemple la table suivante présente le profil du client Youcef pendant la phase d'apprentissage:

pseudo	Classe_ancienne	Classe_nouvelle	prix	nbr	temps
jonson	6	6	385	7	87
youcef	5	5	430	2	69
admin	1	1	263	3	35
nacer	7	7	948	4	87

Figure IV.11 Table de comportement du client Youcef

Ici, le client Youcef à un temps moyen de connexions qui dépasse 30 mn, son prix moyen d'achat est inférieur à 500€, il a une fréquence d'achat moins de 5 produits. Donc après le classement en fonction de ces critères sa nouvelle classe est la classe C5.

Lorsque le client Youcef a fait une nouvelle connexion, la table comportement sera modifiée, comme il est indiqué dans les figures suivantes :

pseudo	Classe_ancienne	Classe_nouvelle	prix	nbr	temps
jonson	6	6	385	7	87
youcef	5	1	430	2	48
admin	1	1	263	3	35
nacer	7	7	948	4	87

Figure IV.12 Table de comportement du client Youcef (2)

Puisque le client Youcef a changé son profile (classe_nouvelle!= classe_ancienne) alors

La détection nécessite une autre vérification (les clusters) avant de déclencher l'alerte

Dans chaque tentative d'attaque ou un grand changement du comportement du client le system enregistre les informations sur cette attaque (ainsi que le comportement générale avant cette attaque)

L'administrateur va consulter ces informations dans la page d'attaque

Figure IV.13 liste des attaques

Dans cette partie, toujours l'IDS observe le comportement du client c'est-à-dire mesure la similarité entre sa classe_ancienne et sa nouvelle_classe mais en plus et plus particulièrement mesure la similarité entre son ancien cluster et son nouveau cluster.

Si le client change sa classe dans le même cluster, aucune alerte n'est déclenchée, mais s'il change le cluster (c'est-à-dire Cluster_ancien != Cluster_nouveau), une alerte sera déclenchée et le client sera bloqué et il finit son rôle d'achats.

La figure III.18 suivante montre le comportement des clients :

pseudo	Classe_ancienne	Classe_nouvelle	Cluster_ancien	Cluster_nouveau	prix	nbr	temps
jonson	6	5	2	2	423	4	72
nacer	7	6	3	2	480	6	87
admin	7	7	3	3	684	4	94

Figure IV.14 Table de comportement des clusters.

Toujours la classification des clients se base sur les critères (Temps, Prix, Fréquence).

Par exemple la classe_ancienne du client nacer est la classe 7 (60<T<120, P>500, F<5); et son cluster est le cluster 3.

Après une nouvelle connexion, le client nacer a changé sa classe du C7 à C6 et son cluster a été changé (Car C7 et C6 ne sont pas dans le même cluster) ; donc une alerte se déclenche et le client ne pourra pas poursuivre son achats.

Pour le client Jonson, il a changé sa classe de C6 à C5 mais ces derniers sont dans le même cluster (cluster2) donc il pourra poursuivre ses achats (aucune alerte ne sera déclenchée).

-Après la détection d'attaque, le système ses bloquer le client c'est-à-dire que ce dernier ne peut pas poursuivre ses achats et finir son rôle par l'affichage d'un message d'alerte.

Figure IV.15 Message d'alerte corresponds à la détection d'attaque.

Conclusion :

Au cours de ce dernier chapitre, on a réalisé une petite application web (boutique en ligne) et on a implémenté un système de détection d'intrusion IDS dont le but est de sécuriser cette application web à l'aide d'un ensemble de programmes. On a implémenté un algorithme de clustering k-means qui permet de regrouper les classes en k-clusters afin de diminuer le nombre de faux positifs.

Conclusion générale

La défense en profondeur des réseaux passe par une bonne stratégie préventive pour protéger ses réseaux et leurs interconnexions de façon sécurisée. Cette approche doit être complétée une fois le réseau est opérationnel pour permettre de détecter des anomalies qui peuvent être révélatrices.

Ce travail nous a permis d'avoir une idée plus claire sur les applications du domaine de la sécurité informatique. On a également découvert les IDS et leurs approches, plus précisément l'approche comportementale. Cette application qu'on a élaborée présente des avantages comme la détection rapide des anomalies ainsi qu'un taux de fausses alertes limité.

On a amélioré les performances de notre IDS comportementale à travers un algorithme de clustering K-means qui permet de détecter des anomalies avec un taux minimum de fausses alertes par rapport à l'approche comportementale.

En outre, il est important de noter que le risque nul d'être piraté n'existe pas et il faut s'avoir s'appuyer au mieux sur les outils (nouvellement) disponibles afin de tendre vers cet idéal.

Bibliographie :

[1] Cole et al « Les menaces informatique », IEEE Computer Society Press, juin 2005.

[2] Guy Pujolle, « Les réseaux », Rapport de stage, Laboratoire Spécification et Vérification , janvier 2008.

[3] La sécurité des réseaux, support de cours, Mercredi, 8. novembre 2006.

[4] Brigitte Ulmann, « Cisco et la sécurité », Novembre 2004.

[5] Richardson, «les attaques informatiques», novembre 2008.

[6] Stiven carl, « La sécurité informatique », avril 2005.

[7] Nicolas Baudoin et Marion Karle, « NT Réseaux –IDS et IPS », 2000, support de cours,Enseignant Etienne Duris en 2003-2004.

[8] Isabelle Facon, « Les enjeux de sécurité en Asie centrale : la politique de la réussie », 2000.

[9] Eric Berthomier, « Sécurité des Réseaux », Mars 2005.

[10] Ritcherd simon et robert smond « Les enjeux de la sécurité », janvier 2010.

[11] *Nicolas* Baudoin et Marion Karle, « NT Réseaux –IDS et IPS » support de cours, Enseignant Etienne Duris en 2004-2005.

[12] Brigitte Ulmann, « Cisco et la sécurité », Novembre 2004.

[13] http://web.mit.edu/rhel-doc/4/RH-DOCS/rhel-sg-fr-4/ch-detection.html, consulté le : Juin 2013.

[14] Jacobcarter, , « An Intrusion to Intrusion Detection », Crossroads Student Magazine,november 2014.

[15] Hervé Debar, Marc Dacier et Andreas Wespi, « A Revised Taxonomy for Intrusion-Detection Systems – Annales des Télécommunications », 55, n° 7-8,2000.

[16] Endorf et al « Les system de détection d'intrusion », janvier 2004

[17] Madjid Ouharoun, « Modélisation des IDS par des jeux probabilistes », Mémoire de maitrise, Université du *Québec* Canada, 2011.

[18] Robert Longeon «Guide de la sécurité des systèmes d'information », livre vol.94, 2009

[19] T. Steven et al. « An attack language for state-based intrusion detection », Journal of Computer Security, pages 71–103, 2002.

[20] Stefan Axelsson, « Intrusion detection systems : A taxonomy and survey »,Technical Report 99-15, Dept. of Computer Engineering, Chalmers, Université deTechnology, Mars 2000.

[21] Giovanni Vigna et al. « A stateful intrusion detection system for world-wide Webservers », In Proceedings of the Annual Computer Security ApplicationsConference (ACSAC 2003), pages 34–43, Las Vegas, Novembre/December 2003.

[22] P. Anderson, « Computer security threat monitoring and surveillance », Rapporttechnique, Company de Washington, Avril 1980.

[23] D. Denning « An Intrusion-Detection Model – IEEE transaction on SoftwareEngineering », 1987.

[24] H. Debar, M. Becker et D. Siboni, « A Neural Network Component for anIntrusion Detection System, Proceeding of the IEEE Symposium of Research in Computer Security and Privacy », 1992.

[25] Stephanie Forrest et al. « A sense of self security for unix processes », IEEE Computer Society Press, Mai 1996.

[26] Nathalie Dagorn, « Détection et prévention d'intrusion : présentation et limites ».Rapport de recherche, Université de Nancy1, 2006.

[27] Alex michel « Vers un détection d'intrusions à fiabilité etpertinence prouvables », Thèse de doctorat, Université de Technology, Australie,2006.

[28] J. Olivain, « Plate-forme de détection d'intrusions », Rapport de stage, Laboratoire Spécification et Vérification, Décembre 2003.

[29] Kardi Teknomo, «K-Means Clustering» mars 2008.

Résumé :

Un système de détection d'intrusion (IDS) est un mécanisme écoutant le trafic réseau de manière furtive afin de repérer des activités 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 présentes 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 web (boutique en ligne) en utilisant l'approche comportementale optimisé par l'algorithme K-means.

Mots-Clefs : IDS, approche comportementale, approche par signatures, faux positifs, faux négatifs et algorithme K-means.

Abstract:

An intrusion detection system (IDS) is a mechanism for listening to the network traffic stealth to identify anomalous or suspicious activities 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 positives and false negatives. Our goal is to manage the security of a web application (online store) using the optimized K-means algorithm behavioral approach.

Keywords: IDS, behavioral approach, approach signatures, false positives, false negatives and K-means algorithm.

ملخص:

نظام كشف التسلل (IDS) هو آلية للإستماع إلى الشبكة خلسة لأجل تحديد الأنشطة الغير طبيعية أو المشبوهة، و يسمح أيضا باتخاذ إجراءات وقائية على خطر حركة التسلل. أساليب كشف التسلل يعتمد على نهجين رئيسيين: نهج السلوكية و نهج السيناريو. كل منهما لديه نقاط قوة و لكن أيضا نقاط ضعف و التي تتمثل في الأخطاء الإيجابية و الأخطاء السلبية. هدفنا هو إدارة أمن تطبيق ويب (متجر على شبكة انترنت) باستخدام نهج السلوكية عن طريق خوارزمية K-means.

الكلمات المفتاحية : نظام كشف التسلل، نهج السلوكية، نهج السيناريو، الأخطاء الإيجابية، الأخطاء السلبية و خوارزمية K-means