Université Abou Bekr Belkaid
Tlemcen Algérie
جامعة أبي بكر بلقايد
تلمسان الجزائر

République Algérienne Démocratique et Populaire
Université Abou Bekr Belkaid– Tlemcen
Faculté des Sciences
Département d'Informatique

Mémoire de fin d'études

pour l'obtention du diplôme de Master en Informatique

Option : *Réseaux et Systèmes Distribués (R.S.D)*

Thème

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

Réalisé par :

- **BENDELLA Zineb**

Présenté le 26 Novembre 2013 devant le jury composé de :

- *M^r BENAMMAR Abdelkrim* *(Président)*
- *M^r BENMAMMAR Badr* *(Encadreur)*
- *M^r BENMOUNA Youcef* *(Examinateur)*
- *M^r BELABED Amine* *(Examinateur)*

Année universitaire: 2012-2013

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 ABOU BAKR BELKAID de Tlemcen 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 Badr Benmammar****, qui, en tant que encadreur 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 mon Mari, 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

A l'aide de DIEU tout puissant, qui trace le chemin de ma vie, j'ai pu arriver à réaliser ce modeste travail que je dédie:

A la mémoire de ma grande mère paternel ;

A mon très cher ***père*** *et ma très chère* ***mère*** *qui n'ont pas cessé de m'encourager et de se sacrifier pour que je puisse franchir tout obstacle durant toutes mes années d'étude que Dieu me les garde en très bonne santé ; Aucune dédicace ne pourra compenser les sacrifices de mes parents;*

A ma plus belle étoile qui puisse exister dans l'univers, mon très cher mari ***Ali****, celui à qui je souhaite une longue vie pleine de joie, de bonheur et de santé;*

A ma ***grand-mère*** *pour toutes ses prières ;*

A ma chère sœur du monde ***Meriem****, que je leur souhaite une longue vie pleine de joie et de réussite.*

A mon frère ***Abd Elghani****, et mon petit chère frère* ***Mohamed*** *;*

Aux petites sœurs adorables ***Amira*** *&* ***Hassnaa*** *;*

A ma belle famille, aux chers ***parents de mon mari****,* ***Karima****,* ***Rabie****,* ***Issam*** *et* ***Ilyes****, je leur souhaite une vie pleine de bonheur ;*

A mes ***oncles****, mes* ***tantes****, mes* ***cousines****, mes* ***cousins*** *spécialement* ***Sofiane****, et à toute ma famille ;*

A ***Yahouni Zakaria****, qui m'a aider tout au long de ce mémoire, je lui souhaite une vie pleine de joie, de bonheur et de réussite ;*

A mes enseignants et surtout ***Mr Badr Benmammar*** *, mon encadreur;*

Et à tous ceux qui m'aiment et qui me cannaient de proche ou de loin.

Bendella Zineb

Table des matières

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 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 : [16]

- *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 de détection *Intrusion Detection System*s (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 dernier chapitre est consacré à la réalisation de notre application (une boutique en ligne), nous avons donc implémenté un système de détection d’intrusion avec deux approches : une approche comportementale et une approche utilisant l’algorithme K-means.

CHAPITRE I :

Sécurité des Réseaux

I. Introduction :

L'informatique et en particulier l'Internet jouent un rôle grandissant 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 pesant sur la sécurité des réseaux ainsi que les mécanismes de défense.

II. Sécurité des réseaux:

II.1 Définition:

La sécurité d'un réseau est un ensemble de moyens techniques, organisationnels, juridiques et humains nécessaires et mis en place pour conserver, rétablir, et garantir sa sécurité. En général, la sécurité d'un réseau englobe celle du système informatique sur lequel il s'appuie. [1]

II.2 Évaluation de la sécurité d'un réseau:

La sécurité d'un réseau peut s'évaluer sur la base d'un certain nombre de critères de sécurité. On distingue généralement trois principaux critères de sécurité [1]:

- ***Disponibilité*** : Elle consiste à garantir l'accès à un service ou à une ressource.
- ***Intégrité*** : Elle consiste à s'assurer que les données n'ont pas été altérées durant la communication (de manière fortuite ou intentionnelle).
- ***Confidentialité*** : Elle consiste à rendre l'information inintelligible à d'autres personnes que les seuls acteurs concernés.

En plus des ces trois critères, on peut ajouter les critères suivants:

- ***Authentification*** : Elle consiste à assurer l'identité d'un utilisateur, c'est-à dire de garantir à chacun des correspondants que son partenaire est bien celui qu'il croit être.
- ***Non répudiation*** : Elle consiste à garantie qu'aucun des correspondants ne pourra nier la transaction.

L'évaluation de la sécurité d'un système informatique est un processus très complexe basé en général sur une méthodologie. Cette évaluation passe par une analyse de risques. Cette dernière pesant sur un système informatique elle même s'appuie sur un ensemble de métriques définies au préalable [1].

III. Les causes pour sécuriser les réseaux :

III.1 Les enjeux :

a) **Enjeux économiques** : Les organismes ou entreprises à but lucratif on presque toujours la même finalité : c'est de réaliser des bénéfices sur l'ensemble de leurs activités. Cette réalisation est rendue possible grâce à son système d'information considéré comme moteur de développement de l'entreprise. D'où la nécessité de garantir la sécurité de ce dernier. La concurrence fait que des entreprises s'investissent de plus en plus dans la sécurisation de leurs systèmes d'information et dans la qualité de service fournit aux clients [1].

b) **Enjeux politiques :** La plupart des entreprises ou organisations se réfèrent aux documents officiels de sécurité élaborés et recommandés par l'État. Ces documents contiennent généralement des directives qui doivent être appliquées par toute structure engagée dans un processus de sécurisation du réseau. Dans le cadre du chiffrement des données par exemple, chaque État définit des cadres et mesures d'utilisation des algorithmes de chiffrement et les recommande aux entreprises exerçant sur son territoire. Le non respect de ces mesures et recommandations peut avoir des conséquences graves sur l'entreprise. A ce niveau, l'enjeu est plus politique parce que chaque État souhaite être capable de décrypter toutes les informations circulant dans son espace [1].

c) **Enjeux juridiques** : Dans un réseau, on retrouve de l'information multiforme (numérique, papier, etc.). Le traitement de celle ci doit se faire dans un cadre bien définit et dans le strict respect des lois en vigueur. En matière de juridiction, le non respect des lois et exigences relatives à la manipulation des informations dans un système d'information peut avoir des conséquences graves sur l'entreprise. [1]

III.2 Les vulnérabilités :

Tous les systèmes informatiques sont vulnérables. Peu importe le niveau de vulnérabilité de ceux ci. Une vulnérabilité est une faille ou une faiblesse pouvant être exploitée par une personne mal intentionnée pour nuire.

Les vulnérabilités des systèmes peuvent être classées en catégorie (humaine, technologique, organisationnelle, mise en œuvre). [1]

a) **Vulnérabilités humaines** : L'être humain de par sa nature est vulnérable. La plupart des vulnérabilités humaines proviennent des erreurs (négligence, manque de compétences, surexploitation, etc.), car ne dit t-on pas souvent que l'erreur est humaine? Un SI étant composé des humains, il convient d'assurer leur sécurité si l'on veut garantir un maximum de sécurité dans le SI. [1]

b) **Vulnérabilités technologiques :** Avec la progression exponentielle des outils informatiques, les vulnérabilités technologiques sont découvertes tous les jours. Ces vulnérabilités sont à la base dues à une négligence humaine lors de la conception et la réalisation. Pour être informé régulièrement des vulnérabilités technologiques découvertes, il suffit de s'inscrire sur une liste ou des listes de diffusion mises en place par les CERT(Computer Emergency Readiness ou Response Team). [1]

c) **Vulnérabilités organisationnelles :** Les vulnérabilités d'ordre organisationnel sont dues à l'absence des documents cadres et formels, des procédures (de travail, de validation) suffisamment détaillées pour faire face aux problèmes de

sécurité du système. Quand bien même ces documents et procédures existent, leur vérification et mises à jour ne sont pas toujours bien assurées. [1]

d) **Vulnérabilités mise en œuvre :** Les vulnérabilités au niveau mise en œuvre peuvent être dues à la non prise en compte des certains aspects lors de la réalisation d'un projet. [1]

III.3 Menaces :

On peut également classer les menaces en deux catégories selon qu'elles ne changent rien (menaces *passives*) ou qu'elles perturbent effectivement le réseau (menaces *actives*). [2]

a) ***Les menaces passives*** : consistent essentiellement à copier ou à écouter l'information sur le réseau, elles nuisent à la confidentialité des données. Dans ce cas, celui qui prélève une copie n'altère pas l'information elle-même. [2]

b) ***Les menaces actives*** : sont de nature à modifier l'état du réseau. [2]

III.4 Risques :

Les risques se mesurent en fonction de deux critères principaux : la *vulnérabilité* et la *sensibilité*.

La vulnérabilité désigne le degré d'exposition à des dangers. Un des points de vulnérabilité d'un réseau est un point facile à approcher. Un élément de ce réseau peut être très vulnérable tout en présentant un niveau de sensibilité très faible : le poste de travail de l'administrateur du réseau, par exemple, dans la mesure où celui-ci peut se connecter au système d'administration en tout point du réseau. [2]

La sensibilité désigne le caractère stratégique d'un composant du réseau. Celui-ci peut être très sensible, vu son caractère stratégique mais quasi invulnérable, grâce à toutes les mesures de protection qui ont été prises pour le prémunir contre la plupart des risques. [3]

IV. Les logiciels malveillants :

Un logiciel malveillant est un programme développé dans le but de nuire à un système informatique, sans le consentement de l'utilisateur infecté. Plusieurs types de logiciels malveillants ont été proposées nous citons les plus répandus :

IV.1 Virus :

Un virus est un morceau de programme informatique malicieux, conçu et écrit pour qu'il se reproduise. Cette capacité à se répliquer, peut toucher votre ordinateur, sans votre permission et sans que vous le sachiez. En termes plus techniques, le virus classique s'attachera à un de vos programmes exécutables et se copiera systématiquement sur tout autre exécutable que vous lancez.[4]

Les virus peuvent s'avérer particulièrement dangereux et endommager plus ou moins gravement les machines infectées. Le virus peut se répandre à travers tout moyen d'échange de données numériques comme l'Internet, et notamment par l'intermédiaire des messages électroniques ou de leurs pièces attachées.

IV.2 Vers :

Un ver (ou *worm*) est un type de virus particulier qui se propage par le réseau. Le vers contrairement aux virus, une fois implantés et activés dans un ordinateur, sont des programmes capables de se propager d'un ordinateur à un autre via le réseau, sans intervention de l'utilisateur et sans exploiter le partage de fichiers.

IV.3 Cheval de Troie :

Un cheval de Troie (*Trojan horse*) est un programme qui exécute des instructions sans l'autorisation de l'utilisateur. Ces instructions sont généralement nuisibles à l'utilisateur, et qui une fois installé sur un ordinateur y effectue des actions cachées et pernicieuses .[5]

Le cheval de Troie contrairement au ver ne se réplique pas.

IV.4 Logiciel Espion :

Un logiciel espion (ou spyware) est un programme, conçu dans le but de collecter des données personnelles sur ses utilisateurs et de les envoyer à son concepteur, ou à un tiers via Internet ou tout autre réseau informatique, sans avoir obtenu au préalable une autorisation explicite et éclairée desdits utilisateurs.[1]

Une variété particulièrement toxique de logiciel espion est le keylogger (espion dactylographique), qui enregistre fidèlement tout ce que l'utilisateur tape sur son clavier et le transmet à son honorable correspondant ; il capte ainsi notamment identifiants, mots de passe et codes secrets. [5]

IV.5 Spam :

Le spam est une vraie problématique. Il encombre les résultats de recherche ce qui gène l'utilisateur. Un spam peut être définit comme étant un email anonyme, non sollicité, indésirable et envoyé en grand nombre de façon automatique sans l'accord de son destinataire.

IV.6 Cookies :

Un cookie est un petit fichier très simple, en fait un texte, enregistré sur le disque dur de l'ordinateur d'un internaute à la demande du serveur gérant le site Web visité. Il contient des informations sur la navigation effectuée sur les pages de ce site. L'idée originelle est de faciliter l'utilisation ultérieure du site par la même personne.

Un cookie n'étant pas exécutable, il ne peut contenir de virus.

IV.7 Bombe logique :

Une Bombe logique est une partie d'un programme malveillant (virus, cheval de Troie, etc.) qui reste dormante dans le système hôte jusqu'à ce qu'un instant ou un événement survienne, ou encore que certaines conditions soient réunies, pour déclencher des effets dévastateurs en son sein. [6]

Les bombes logiques sont généralement utilisées dans le but de créer un déni de service en saturant les connexions réseau d'un site, d'un service en ligne ou d'une entreprise.

IV.8 Porte dérobée :

C'est un moyen de contourner les mécanismes de contrôle d'accès. Elle s'agit d'une faille du système de sécurité due à une faute de conception accidentelle ou intentionnelle.

V. Mécanismes 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 :

V.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. [9]

La Figure I.1 montre le fonctionnement de chiffrement.

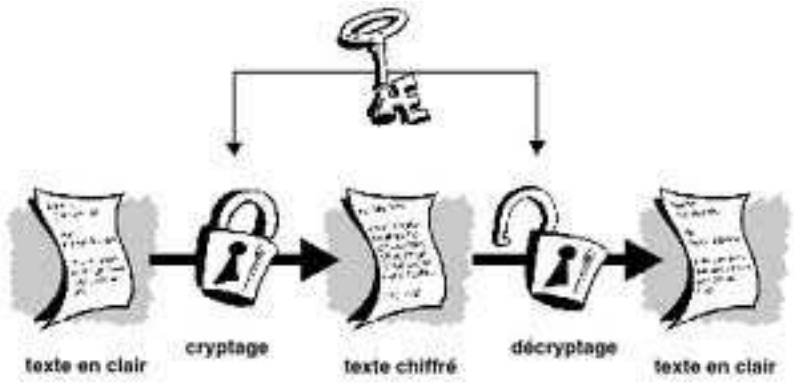

Figure I.1. Cryptage[9] .

V.2 Pare-Feu :

C'est un ensemble de différents composants matériels (physique) et logiciels (logique) qui contrôlent 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 a 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.2 schématise le fonctionnement d'un pare-feu.

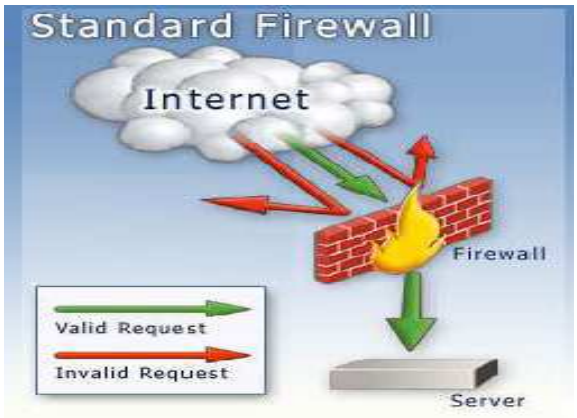

Figure I.2. Pare-feu [2].

V.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. [7]

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.

V.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.[10]

La Figure I.3 montre le principe de protocole de tunnelisation.

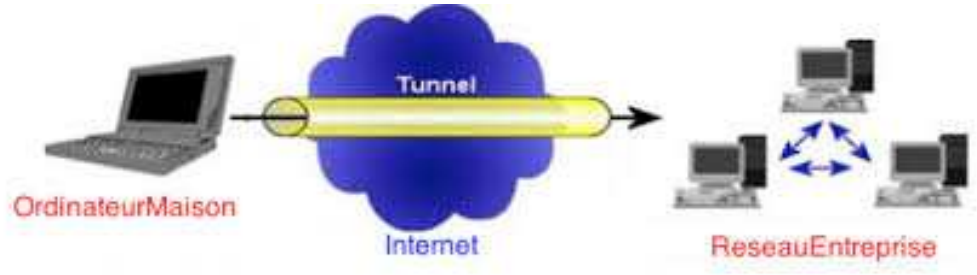

Figure I.3. Principe de VPN .

V.5 IDS :

Un système de détection d'intrusion (ou IDS : Intrusion Detection System) est un mécanisme destiné à repérer des activités anormales ou suspectes sur la cible analysée (un réseau ou un hôte). Il permet ainsi d'avoir une connaissance sur les tentatives réussies comme échouées des intrusions.[11]

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.

V.6 IPS :

Un système de prévention d'intrusion (ou IPS, Intrusion Prevention System) est un outil similaire aux IDS, sauf que ce système peut prendre des mesures afin de diminuer les risques d'impact d'une attaque. C'est un IDS actif, il détecte un balayage automatisé, l'IPS peut bloquer les ports automatiquement.[11]

VI. 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.[12]

Nous allons tout d'abord parler des différents aspects d'une politique de sécurité, avant de définir les objectifs visés, puis de voir les outils disponibles pour appliquer cette politique.[12]

Une politique de sécurité s'élabore à plusieurs niveaux. [12]

- sécuriser l'accès aux données de façon logicielle (authentification, contrôle d'intégrité).
- sécuriser l'accès physique aux données : serveurs placés dans des salles blindées avec badge d'accès…

- Un aspect très important pour assurer la sécurité des données d'une entreprise est de sensibiliser les utilisateurs aux notions de sécurité, de façon à limiter les comportements à risque : si tout le monde peut accéder aux salles de serveurs, peut imposte qu'elles soient sécurisées !

- De même, si les utilisateurs laissent leur mot de passe écrit à côté de leur PC, son utilité est limitée…

- Enfin, il est essentiel pour un responsable de sécurité de s'informer continuellement, des nouvelles attaques existantes, des outils disponibles…de façon à pouvoir maintenir à jour son système de sécurité et à combler les brèches de sécurité qui pourraient exister.

VII. Conclusion :

Dans ce chapitre, on a présenté les principales notions et concepts de la sécurité des systèmes informatiques et des réseaux, dont on a décrit plus particulièrement les menaces provenant des logiciels malveillants et introduit une politique de sécurité. Ainsi différentes méthodes et mécanismes connus pour sécuriser les réseaux.

A travers les différentes sections qu'on a montré, on conclu qu'aucun réseaux n'est sùr à 100% et il est impossible de garantir la sécurité totale d'un réseau.

CHAPITRE II :

Système de Détection d'Intrusion

I. 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.[30]

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 détection 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. Je présente également la classification des IDS, dans ce cadre plusieurs critères sont pris en compte nous commençons par la classification selon la méthode d'analyse qui découpe les IDS en deux approches (comportementale et par signatures), enfin nous allons mettre le point sur la détection d'intrusions Web.

II. Définition :

La détection des intrusions est le processus de surveillance des événements se trouvant dans un système des ordinateurs ou du réseau et les analysant pour détecter les signes des intrusions, défini comme des tentatives pour compromettre la confidentialité, l'intégrité, la disponibilité ou éviter des mécanismes de sécurité de l'ordinateur ou du réseau. L'intrusion est causée par les attaques accédant au système via l'Internet, autorisée l'utilisateur du système qui essaye de gagner les privilèges supplémentaires pour lesquels ils n'ont pas autorisés, et autorisé les utilisateurs qui abusent les privilèges donnés. Le système de détection des intrusions est un logiciel ou un matériel qui automatise des surveillances et les processus analysés.[15]

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.

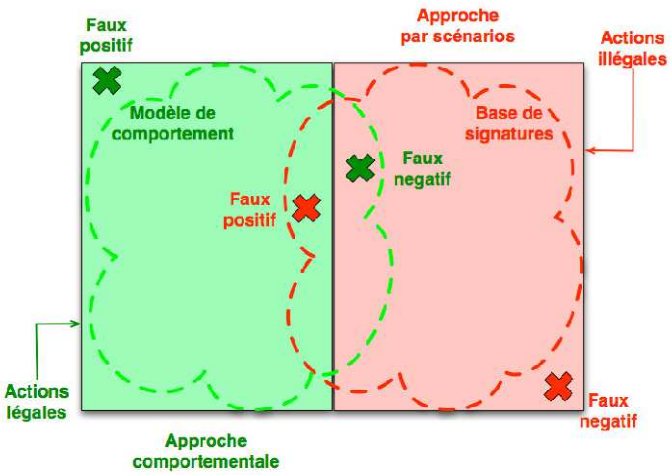

Figure II.1. Problèmes des IDS [16].

Un IDS a quatre fonctions principales : l'analyse, la journalisation, la gestion et l'action.

- ***Analyse:*** Analyse des journaux du système pour identifier des intentions dans la masse de données recueillie par l'IDS. Il y a deux méthodes d'analyse : L'une basée sur les signatures d'attaques, et l'autre sur la détection d'anomalies. [17]
- ***Journalisation:*** Enregistrement des évènements dans un fichier de log. Exemples d'évènements : arrivée d'un paquet, tentative de connexion. [17]
- ***Gestion:*** Les IDS doivent être administrés de manière permanente. On peut assimiler un IDS à une caméra de sécurité. [17]
- ***Action:*** Alerter l'administrateur quand une attaque dangereuse est détectée. [17]

III. Types des IDS :

III.1 IDS réseaux :

Le rôle essentiel d'un IDS réseau (NIDS) est l'analyse et l'interprétation des paquets circulant sur ce réseau.

L'implantation d'un NIDS sur un réseau se fait de la façon suivante : des capteurs sont placés aux endroits stratégiques du réseau et génèrent des alertes s'ils détectent une attaque. Ces alertes sont envoyées à une console sécurisée, qui les analyse et les traitent éventuellement. Cette console est généralement située sur un réseau isolé, qui relie uniquement les capteurs et la console.[12]

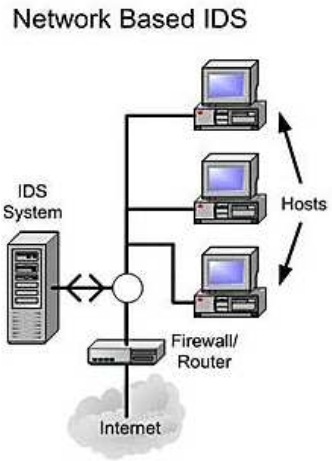

Figure II.2: Architecture d'un NIDS.[18]

III.2 IDS Host :

Les HIDS (Host IDS) analysent le fonctionnement et l'état des machines sur lesquelles ils sont installés. L'intégrité des systèmes est alors vérifiée périodiquement et des alertes peuvent êtres levées. [18]

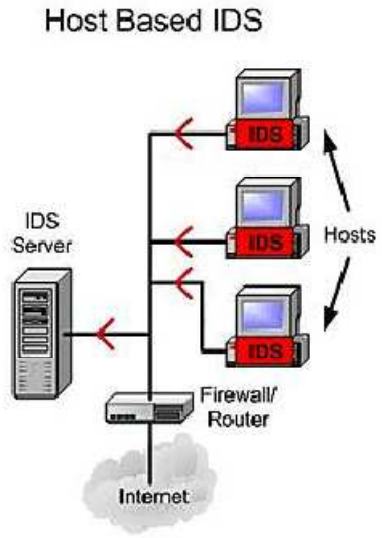

Figure II.3: Architecture d'un HIDS [18]

III.3 IDS Hybride :

Les IDS hybrides rassemblent les caractéristiques des NIDS et HIDS. Ils permettent, en un seul outil, de surveiller le réseau et les terminaux. Les sondes sont placées en des points stratégiques, et agissent comme NIDS et/ou HIDS suivant leurs emplacements. Toutes ces sondes remontent alors les alertes à une machine qui va centraliser le tout, et agréger/lier les informations d'origines multiples. Ainsi, on comprend que les IDS hybrides sont basés sur une architecture distribuée, où chaque composant unifie son format d'envoi. Cela permet de communiquer et d'extraire des alertes plus pertinentes. Les avantages des IDS hybrides sont multiples :[18]

- Moins de faux positif.
- Meilleure corrélation (la corrélation permet de générer de nouvelles alertes à partir de celles existantes).
- Possibilité de réaction sur les analyseurs.

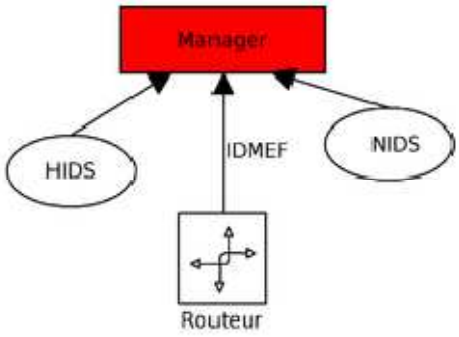

Figure II.4: Architecture d'un IDS Hybride [18]

III.4 Système de prévention d'intrusion (IPS) :

Un système de prévention d'intrusion est un dispositif capable de détecter des attaques, connues et inconnues, et de les empêcher d'être réussies. L'IPS n'est pas un observateur : il fait partie intégrante du réseau. Il est placé en ligne et examine tous les paquets entrants ou sortants.[19]

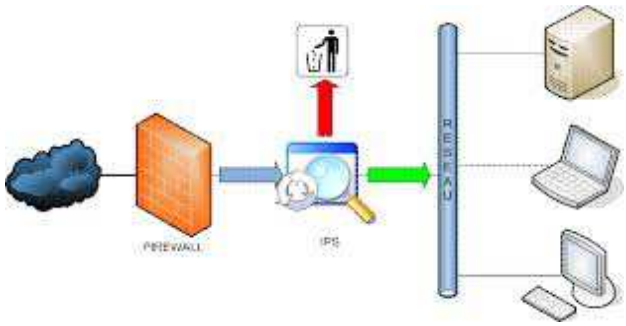

Figure II.5: Architecture d'un IPS

III.5 KIDS/KIPS:

Ce type d'IDS analyse les appels systèmes et bloque tout accès suspect au système . Ainsi les KIPS sont des solutions rarement utilisés sur des serveurs souvent sollicités.

IV. Architecture d'un IDS :

Cette section décrit les trois composants qui constituent classiquement un système de détection d'intrusions. La Figure II.6 illustre les interactions entre ces trois composants.

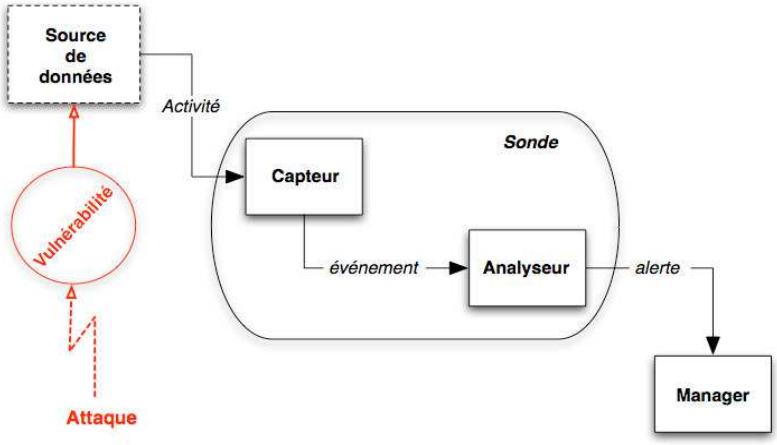

Figure II.6. Architecture d'un IDS. [16]

VII.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. [16]

VII.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. [16]

VII.3 Manager :

Le manager 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 : [16]

- 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.

V. Mode de fonctionnement d'un IDS:

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. Il existe deux modes de détection, la détection d'anomalies et la reconnaissance de signatures. De même, deux types de réponses existent, la réponse passive et la réponse active. [12]

V.1 Modes de détection :

a) La détection d'anomalies :

Elle consiste à détecter des anomalies par rapport à un profil "de trafic habituel". La mise en œuvre comprend toujours une phase d'apprentissage au cours de laquelle les IDS vont "découvrir" le fonctionnement "normal" des éléments surveillés. Ils sont ainsi en mesure de signaler les divergences par rapport au fonctionnement de référence. [12]

b) La reconnaissance de signature :

Cette approche consiste à rechercher dans l'activité de l'élément surveillé les empreintes (ou signatures) d'attaques connues. Ce type d'IDS est purement réactif ; il ne peut détecter que les attaques dont il possède la signature. De ce fait, il nécessite des mises à jour fréquentes. [12]

De plus, l'efficacité de ce système de détection dépend fortement de la précision de sa base de signature.

Une signature permet de définir les caractéristiques d'une attaque, au niveau des paquets ou au niveau protocole.

V.2 Répense passive et active :

Il existe deux types de réponses, la réponse passive et la réponse active.

a) La réponse passive : La réponse passive d'un IDS consiste à enregistrer les intrusions détectées dans un fichier de log qui sera analysé par le responsable sécurité. Certains IDS permettent de logger l'ensemble d'une connexion

identifiée comme malveillante. Ceci permet de remédier aux failles de sécurité pour empêcher les attaques enregistrées de se reproduire, mais elle n'empêche pas directement une attaque de se produire. [12]

b) **La réponse active :** La réponse active au contraire a pour but de stopper une attaque au moment de sa détection. [12]

VI. Classification des IDS :

Plusieurs critères permettent de classer les systèmes de détection d'intrusions, la méthode d'analyse étant le principal. Deux méthodes dérivant de cette dernière existent aujourd'hui : l'approche comportementale et l'approche par scénarios.

On peut citer aussi d'autres critères de classification des IDSs : la fréquence d'utilisation, les sources de données à analyser, le comportement de l'IDS après intrusion.[20]

VI.1 Approche comportementale :

L'approche comportementale est fondée sur une description statistique des sujets. L'objectif est de détecter les actions anormales effectuées par ces sujets (par exemple, des heures de connexion anormales, un nombre anormal de fichiers supprimés ou un nombre anormal de mots de passe incorrects fournis au cours d'une connexion).

Le comportement normal des sujets est appris en observant le système pendant une période donnée appelée phase d'apprentissage (par exemple, un mois). Le comportement normal, appelé comportement sur le long terme, est enregistré dans la base de données et comparé avec le comportement présent des sujets, appelé comportement à court terme. Une alerte est générée si une déviation entre ces comportements est observée. Dans cette approche, le comportement sur le long terme est, en général, mis à jour périodiquement pour prendre en compte les évolutions possibles des comportements des sujets. Je considère traditionnellement que l'avantage principal de l'approche comportementale est de pouvoir être utilisée pour détecter de nouvelles attaques. Autrement dit, en signalant toute déviation par rapport au profil, il est possible de détecter a priori toute attaque qui viole ce profil, même dans le cas où cette attaque n'était pas connue au moment de la construction du profil. [20]

Cependant, cette approche présente également plusieurs inconvénients. Tout d'abord, le diagnostic fournit par une alerte est souvent flou et nécessite une analyse complémentaire.

Ensuite, cette approche génère souvent de nombreux faux positifs car une déviation du comportement normal ne correspond pas toujours à l'occurrence d'une attaque. Citons à titre d'exemples, en cas de modifications subites de l'environnement de l'entité modélisée, cette entité changera sans doute brutalement de comportement. Des alarmes seront donc déclenchées.
Pour autant, ce n'est peut-être qu'une réaction normale à la modification de l'environnement. [20]

En outre, les données utilisées en apprentissage doivent être exemptes d'attaques, ce qui n'est pas toujours le cas. Enfin, un utilisateur malicieux peut habituer le système (soit pendant la phase d'apprentissage, soit en exploitation si l'apprentissage est continu) à des actions malveillantes, qui ne donneront donc plus lieu à des alertes. Le problème de la détection d'intrusions est couramment approché d'une façon radicalement différente qui est l'approche par scénario. [20]

VI.2 Approche par scénario :

La détection d'intrusions peut également s'effectuer selon une approche par scénario. Il s'agit de recueillir des scénarios d'attaques pour alimenter une base d'attaques. Le principe commun à toutes les techniques de cette classe consiste à utiliser une base de données, contenant des spécifications de scénario d'attaques (on parle de signatures d'attaque et de base de signatures). 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. Le principal avantage d'une approche par scénario est la précision des diagnostics qu'elle fournit par rapport à ceux avancés par l'approche comportementale. Il est bien entendu que l'inconvénient majeur de cette approche est qu'elle ne peut détecter que des attaques dont elle dispose de leur signature. Or, définir de façon exhaustive la base de signatures est une des principales difficultés à laquelle se heurte cette approche. La génération de faux négatifs est à craindre en présence des nouvelles attaques. En effet, contrairement à un système de

détection d'anomalies, 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 de données, cette base doit être régulièrement (sans doute quotidiennement) mise à jour en fonction de la découverte de nouvelles attaques. Aucune nouvelle attaque ne peut par définition être détectée. [20]

D'autre part, il existe de nombreuses attaques difficiles à détecter car elles nécessitent de corréler plusieurs événements. Dans la plupart des produits commerciaux, ces attaques élaborées sont décomposées en plusieurs signatures élémentaires. Cette décomposition peut générer de nombreux faux positifs si un mécanisme plus global n'est pas développé pour corréler les alertes correspondant à ces différentes signatures élémentaires. Chacune de ces deux approches peut conduire à des faux positifs (détection d'attaque en absence d'attaque) ou à des faux négatifs (absence de détection en présence d'attaque). [20]

VI.3 Autres critères :

Parmi les autres critères de classification existants, nous pouvons citer entre autres :

VI.3.1 Les sources de données à analyser :

Les sources possibles de données à analyser sont une caractéristique essentielle des systèmes de détection d'intrusions. Les données proviennent, soit de fichiers générés par le système d'exploitation, soit de fichiers générés par des applications, soit encore d'informations obtenues en écoutant le trafic sur le réseau. [20]

VI.3.2 Le comportement de l'IDS après intrusion :

Une autre façon de classer les systèmes de détection d'intrusions, consiste à voir quelle est leur réaction lorsqu'une attaque est détectée. Certains se contentent de déclencher une alarme (réponse passive). [20]

VI.3.3 La fréquence d'utilisation :

Une autre caractéristique des systèmes de détection d'intrusions est leur fréquence d'utilisation : périodique ou continue. Certains systèmes de détection d'intrusions analysent périodiquement les traces d'audit à la recherche d'une éventuelle intrusion ou anomalie passée. Cela peut être suffisant dans des contextes peu sensibles.

La plupart des systèmes de détection d'intrusions récents effectuent leur analyse des traces d'audit ou des paquets réseau de manière continue afin de proposer une détection en quasi temps réel. Cela est nécessaire dans des contextes sensibles (confidentialité) ou commerciaux (confidentialité, disponibilité). C'est toutefois un processus coûteux en temps de calcul car il faut analyser à la volée tout ce qui se passe sur le système. [20]

VII. Détection d'intrusions Web :

Les méthodes de détection d'intrusions utilisées à l'heure actuelle reposent essentiellement sur l'observation d'événements et leur analyse. La collecte d'informations constitue donc la première étape dans tout les IDS.

Le rôle des outils de détection d'intrusions consiste alors à exploiter cette masse d'informations, appelée audit, de manière à y détecter des événements signalant potentiellement une intrusion.[21]

VII.1 Approche comportemental :

La détection d'anomalies consiste à définir, dans une première phase, un certain comportement du système, des utilisateurs, des applications, etc. Dans une seconde phase, on observe l'entité ainsi modélisée et tout écart par rapport au comportement de référence est signalé comme étant suspect.

Cette approche recouvre en fait deux problèmes distincts : la définition du comportement «normal» (souvent appelé profil) d'une part, la spécification des critères permettant d'évaluer le comportement observé par rapport à ce profil d'autre part. [21]

Le principal investissement lors de la mise en œuvre d'un détecteur d'anomalies est la construction du profil. Cette étape est délicate, car le profil doit refléter à la fois une certaine politique de sécurité.

Le profil peut donc contenir des règles impératives imposées par l'administrateur.

La mise en service d'un détecteur d'anomalies est donc précédée d'une phase d'apprentissage au cours de laquelle le profil, initialement construit uniquement à partir d'une politique de sécurité, évolue, afin que toute utilisation jugée normale soit reconnue comme telle. Dans certains cas, cet apprentissage continue également après la

mise en service : le profil évolue constamment afin de suivre au mieux l'utilisation réelle du système.

Les différentes approches de détection d'anomalies se distinguent essentiellement par le choix des entités modélisées dans le profil et l'interprétation qui est faite des divergences par rapport à ce profil. [21]

VII.2 Approche par scénario :

Le problème de la détection d'intrusions est également couramment approché d'une façon radicalement différente, en visant à détecter des signes de scénarii d'attaques connues. Le principe commun à toutes les techniques de cette classe consiste à utiliser une base de données, contenant des spécifications de scénarii d'attaques (on parle de signatures d'attaque et de base de signatures).

Le détecteur d'intrusions confronte le comportement observé du système à cette base et lève une alerte si ce comportement correspond à l'une des signatures. [21]

VII.3 Approche hybride :

Pour tenter de compenser quelques inconvénients de chacune des techniques, certains systèmes utilisent une combinaison de l'approche comportementale et l'approche par scénario. L'IDS comportementale permet de filtrer les requêtes normales et ainsi seules les requêtes détectées comme anormales sont passées à l'IDS par signatures. [21]

VIII. 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.

CHAPITRE III:

Conception et Réalisation

I. 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) en utilisant le paradigme MVC2 à travers la version 2 du Framework Struts. La deuxième étape consiste à sécuriser l'application Web réalisée on 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.*

II. 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 on se basant sur le Framework Struts2 avec l'utilisation de MYSQL comme serveur de base de données.

II.1 Langage de programmation 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).

II.2 Choix du Framework Struts2 :

Le Framework Web Apache Struts est un logiciel gratuit open-source pour créer des applications Web Java basées sur JSP (*Java Server Pages*).
Struts2 implémente le modèle d'architecture MVC2 (Modèle - Vue - Contrôleur), il permet notamment de séparer la partie modèle (programmation, traitement des informations) de la partie présentation (affichage). Le modèle représente le code métier ou la base de données, la vue représente le code de conception de pages et le contrôleur représente le code de navigation (servlet unique).

La Figure III.1 schématise l'architecture de Struts2 :

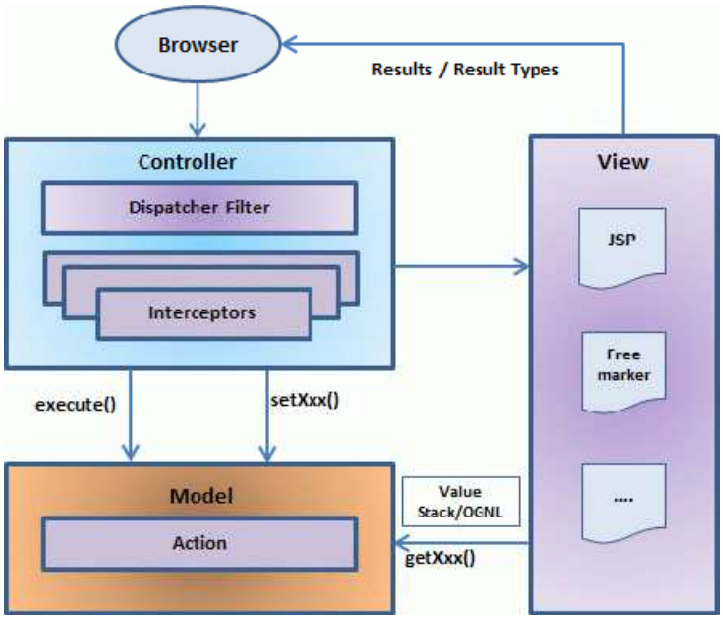

Figure III.1. Architecture de Struts2

Le cycle de vie d'un client dans Struts 2 se déroule comme suit:

- L'utilisateur envoie une demande au serveur pour demander à une ressource.
- Le FilterDispatcher ressemble à la demande et détermine ensuite les mesures appropriées. Celui-ci voit passer toutes les demandes des clients.
- Les fonctionnalités intercepteurs configurés s'appliquent comme une validation.
- L'action sélectionnée est exécutée pour effectuer l'opération demandée.
- Encore une fois, le contrôleur C demande à la page JSP correspondant à la clef de navigation de s'afficher.
- Enfin, le résultat est préparé par la vue et renvoie le résultat à l'utilisateur.

II.3 Choix de MySQL :

MySQL est un serveur de BDD relationnelles open-source qui stocke 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 [22].

On a choisi EasyPHP comme outil pour créer la BDD qui constitue le langage intermédiaire entre cette base et l'utilisateur de la base.

Notre BDD contient les tables suivantes :

- baskets (basketID, basketSession, productID, productPrice);
- clients (nom, login, pwd, id, basketSession, crédit, tmp) ;
- comportement (login, classeancienne, classenouvelle, Cluster_ancien, Cluster_nouveau, time, nb, prix) ;
- history (login, trace, date) ;
- news (actualité) ;
- products (productID, productName, productImage, productPrice, Description, cat, nb) ;

III. Réalisation de l'application Web :

III.1 Description de boutique en ligne :

La boutique en ligne réalisée est sous le nom de InformatiqueBoutique, elle propose une sélection de matériaux et de logiciels informatique, 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 (productName) ;
- 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 III.2 présente la page d'accueil d'InformatiqueBoutique :

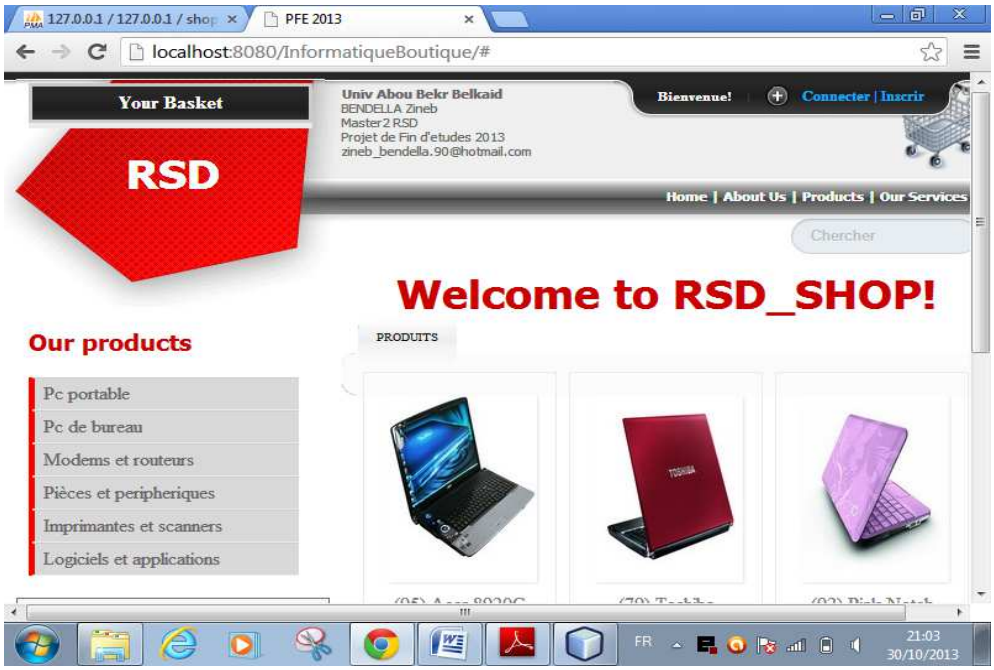

Figure III.2. Page d'accueil d'InformatiqueBoutique.

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 III.3 présente l'inscription d'un client :

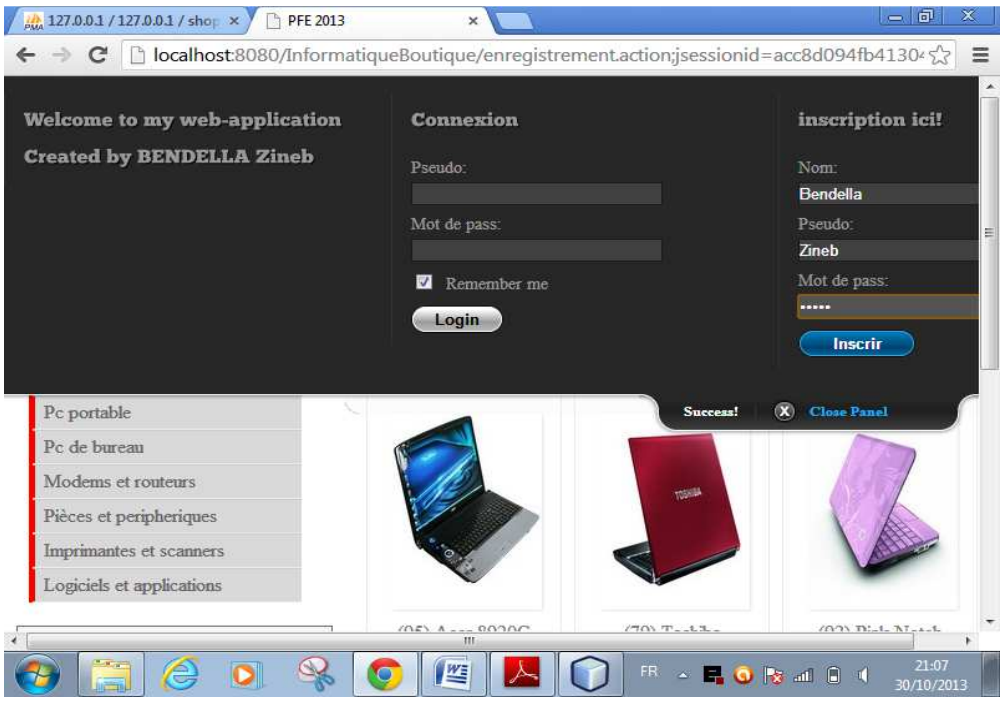

Figure III.3 Inscription d'un client

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 :

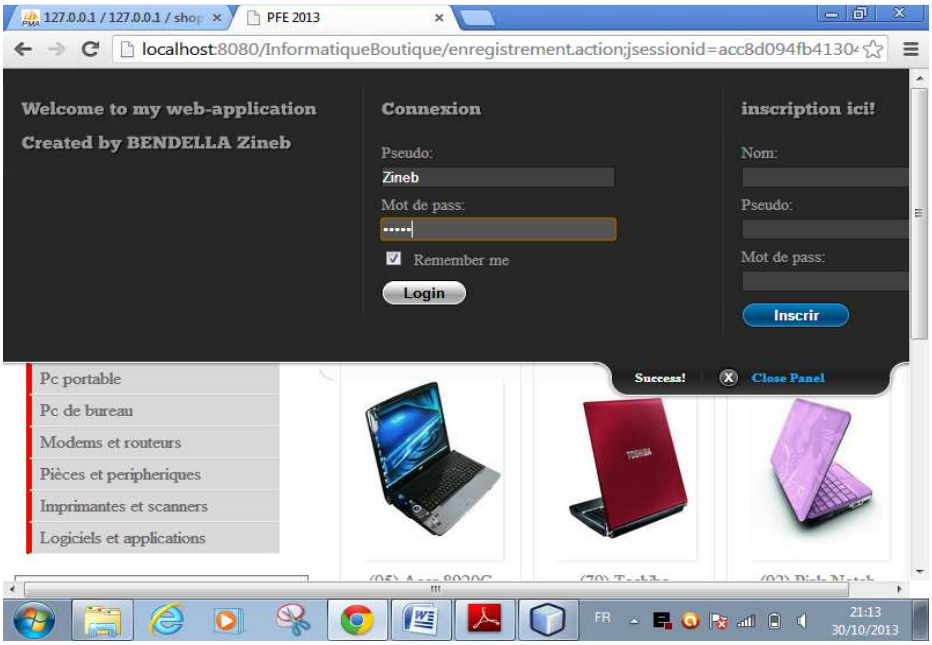

Figure III.4. Connexion d'un client

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 (ADD TO BASKETS) et automatiquement va s'ajouté au caddie virtuel (basket) ;

Figure III.5. Caddie virtuel.

IV. Sécuriser l'application Web:

IV.1 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 :

IV.1.1 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 une demi-heure (30 mn), T2 s'il est entre une demi-heure et une heure (entre 30 et 60 mn) et T3 s'il est supérieur à une heure (dépasse 60 mn).
- ***Prix moyen d'achat P*** : deux choix sont possibles, P1 si le prix est inferieur à 3000€ et P2 si le prix est supérieur à 3000€.
- ***Fréquence d'achat F***: on trouve 2 choix, F1 si le client ne dépasse pas en moyenne 10 produits achetés et F2 s'il achète en moyenne plus de 10 produits.

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.6 montre les différentes classes possibles pour un client.

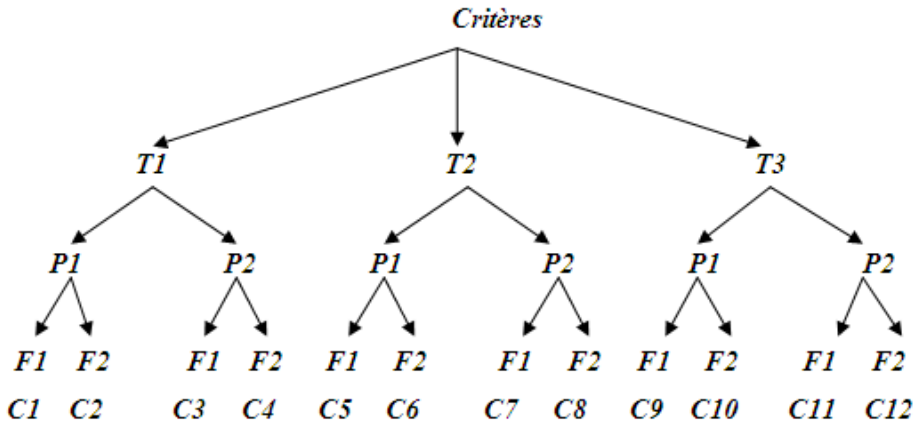

Figure III.6. Classification d'un client en fonctions des 3 critères.

Cette phase d'apprentissage, plus précisément la classification des clients a été faite l'année passée par les auteurs du [31]. Et pour notre projet on déjà supposé qu'on a des clients avec leur profile (sa classe_ancienne).

La table suivante montre le comportement des clients :

login	classe_ancienne	classe_nouvelle	time	nb	prix
client	6	6	45	11	1480
meriem	7	7	36	7	3330
Zineb	5	5	30	3	600
Zinouba	1	1	16	2	880

Figure III.7. 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 ne change pas encore leur profil.

Le champ ***time*** présente le temps de connexion d'un client(le temps de connexion T).

Le champ ***nb*** 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.1.2 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.

Par exemple la table suivante présente le profile de la cliente Zineb pendant la phase d'apprentissage:

login	classe_ancienne	classe_nouvelle	time	nb	prix
client	6	6	45	11	1480
meriem	7	7	36	7	3330
Zineb	5	5	30	3	600
Zinouba	1	1	16	2	880

Figure III.8. Table de comportement du client Zineb (1)

Ici, la cliente Zineb à un temps moyen de connexions qui dépasse 30 mn, son prix moyen d'achat est inférieur à 3000€, il a une fréquence d'achat moins de 10 produits. Donc après le classement en fonction de ces critères sa nouvelle classe est la classe C5.

Lorsque la cliente Zineb a fait une nouvelle connexion, la table comportement sera modifiée, comme indiqué dans les figures suivantes :

login	classe_ancienne	classe_nouvelle	time	nb	prix
client	6	6	45	11	1480
meriem	7	7	36	7	3330
Zineb	5	1	16	2	1100
Zinouba	1	1	16	2	880

Figure III.9. Table de comportement du client Zineb (2)

Puisque la cliente Zineb a changée leur profile (classe_nouvelle != classe_ancienne) alors automatiquement une alerte sera déclanchée.

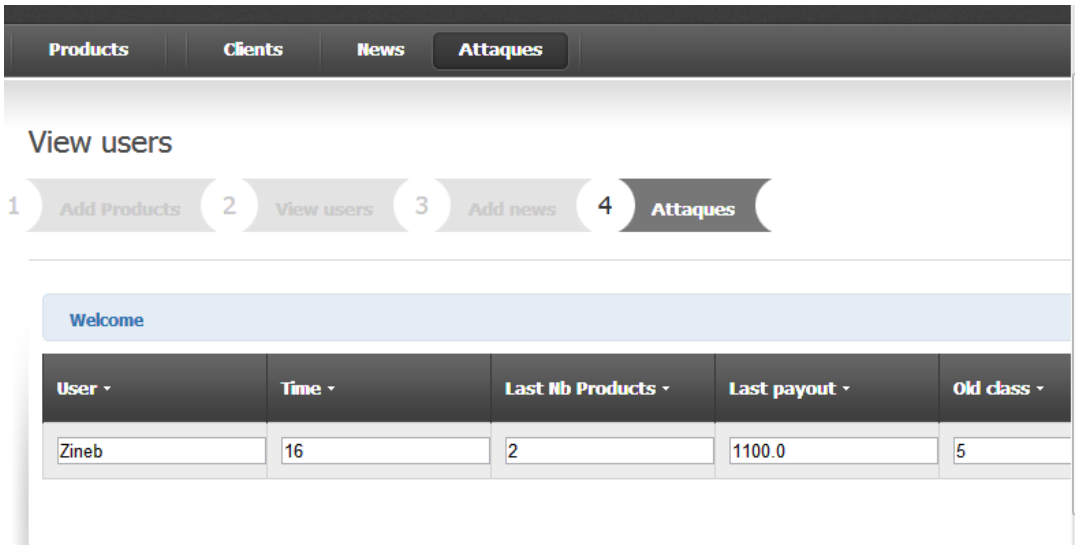

Figure III.10 liste des attaques(1)

Après la détection d'attaque, le système va 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.

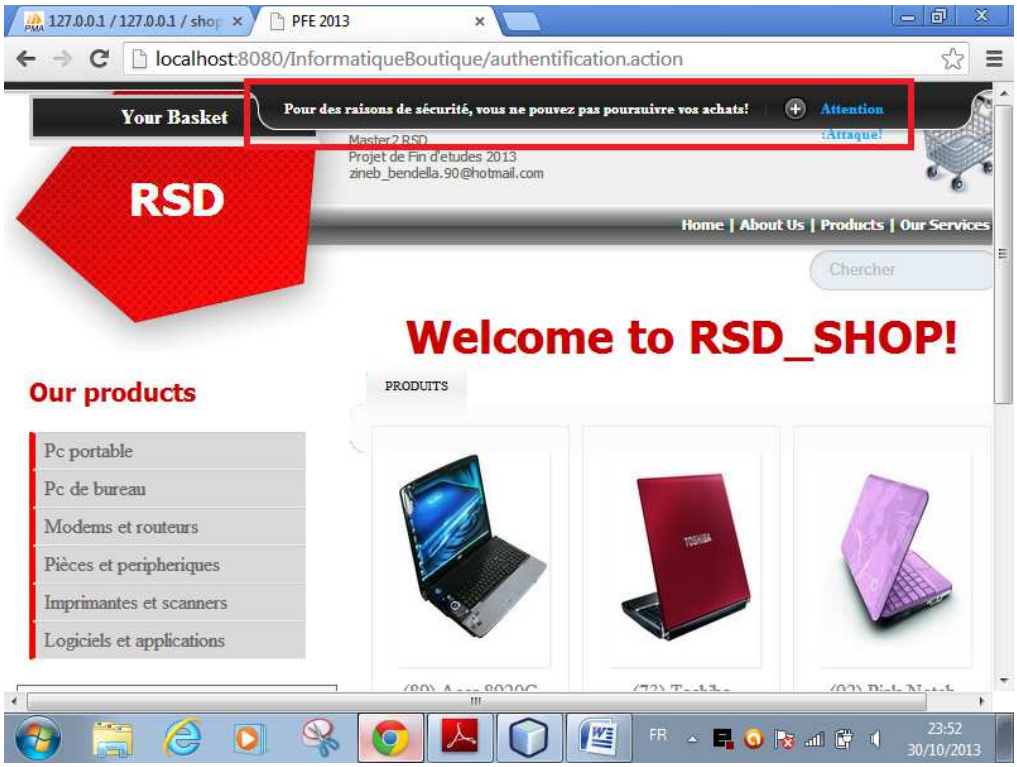

Figure III.11. Message d'alerte corresponds à la détection d'attaque(1).

IV.1.3 Les faux positifs :

Dans [31], les auteurs ont utilisés une simple analyse de données dont le principe est de minimiser les faux positifs à base d'un seuil qui est égal à 25% (0,25). 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.

Si par exemple le client Ali a fait un changement de profil de la classe 1 tels que le temps moyen de connexions à l'application Web est égale à 28 mn, le prix moyen d'achat est égale à 700, la fréquence moyenne d'achat est égale à 7 vers la classe 5 où les critères sont : le temps de connexion à l'application Web vaut 30 mn, le prix d'achat est de 810, la fréquence d'achat est égale à 5 alors la valeur de changement (Val) est calculée comme suit :

$$Val1 = \frac{|\mathbf{30 - 28}|}{\mathbf{30}} = 0.06$$

$$Val2 = \frac{|\mathbf{5 - 7}|}{\mathbf{7}} = 0.28$$

$$Val3 = \frac{|\mathbf{810 - 700}|}{\mathbf{810}} = 0.13$$

Val = (Val1 + Val2 + Val3) / 3
= (0.06 + 0.28 + 0.13) / 3
= 0.15

Puisque Val est < 0.25 donc ici on parle de faux positif malgré le changement de classes de C1 à C5.

D'après le PFE de l'année précédente [31], plus le seuil est petit, plus le nombre de faux positif est réduit. En effet, on veut améliorer ce travail en utilisant un algorithme qui permet de réduire le nombre de faux positifs.

IV.2 Implémentation de l'algorithme de clustering K-means:

IV.2.1 Définition de 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 dévelloppé 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, un 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 [24].

❖ ***Algorithme :[24]***

- **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 μ_k le centre de la classe K ;

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

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

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

Fin.

❖ ***Explication de l'algorithme :***

L'application de l'algorithme k-Means se fait en suivant les étapes suivante[27] :

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.

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 séparées.[25]

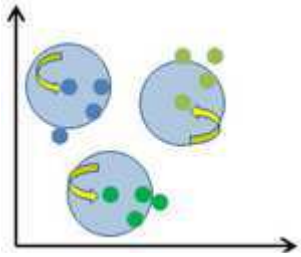

Figure III.12. Sélection des centres.[26]

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.

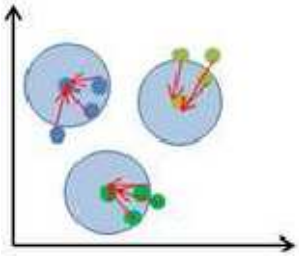

Figure III.13. Affectation des objets. [26]

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

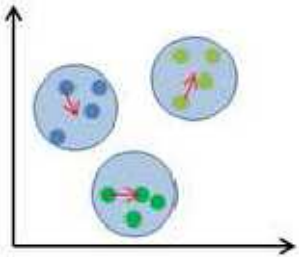

Figure III.14. Recalcule les centres des clusters. [26]

3- Refaire les étapes (3) et (4) jusqu'aucun changement du calcul des centres des clusters ou une stabilité des objets.

IV.2.2 Organigramme :

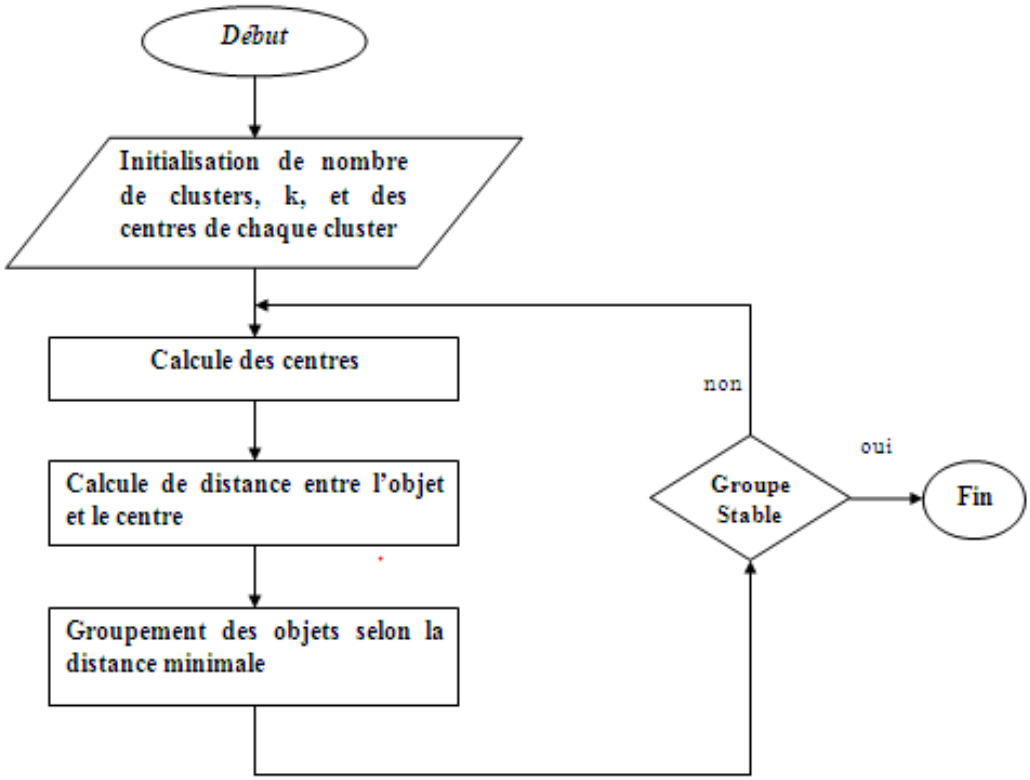

Figure III.15. Organigramme de l'algorithme k-means.

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 choisit 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.

IV.2.3 Implémentation de l'algorithme dans l'application web:

L'algorithme de 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<30 : la moyenne du T=20 mn ;
- Si 30<T<60 : la moyenne du T=45 mn ;
- Si T>60 : la moyenne du T=90 mn ;

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

- Si P<3000 : la moyenne du P=2500 € ;
- Si P>3000 : la moyenne du P=3500 € ;

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

- Si F<10 : la moyenne du F=5 ;
- Si F>10 : la moyenne du F=12 ;

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

Classes	Points
C1 (T<30, P<3000, F<10)	(20, 2500, 5)
C2 (T<30, P<3000, F>10)	(20, 2500, 12)
C3 (T<30, P>3000, F<10)	(20, 3500, 5)
C4 (T<30, P>3000, F>10)	(20, 3500, 12)
C5 (30<T<60, P<3000, F<10)	(45, 2500, 5)
C6 (30<T<60, P<3000, F>10)	(45, 2500, 12)
C7 (30<T<60, P>3000, F<10)	(45, 3500, 5)
C8 (30<T<60, P>3000, F>10)	(45, 3500, 12)
C9 (T>60, P<3000, F<10)	(90, 2500, 5)
C10 (T>60, P<3000, F>10)	(90, 2500, 12)
C11 (T>60, P>3000, F<10)	(90, 3500, 5)
C12 (T>60, P>3000, F>10)	(90, 3500, 12)

Figure III.16. 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. La figure suivante représente le résultat de l'algorithme de clustering k-means :

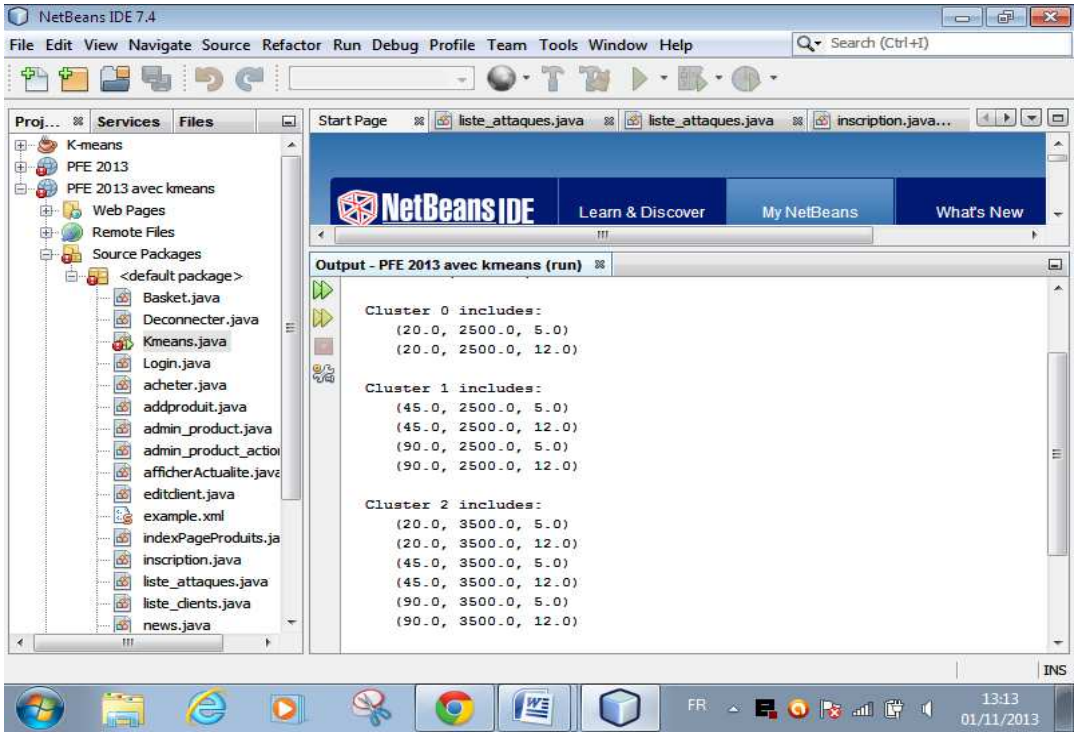

Figure III.17. Résultat de l'algorithme k-means.

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 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 déclanchée, mais s'il change le cluster (c'est-à-dire Cluster_ancien != Cluster_nouveau), une alerte sera déclanchée et le client sera bloqué et il finit son rôle d'achats.

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

login	classe_ancienne	classe_nouvelle	Cluster_ancien	Cluster_nouveau	time	nb	prix
Ali	1	5	1	2	41	4	700
Zineb	7	8	3	3	53	12	3530
Zinouba	7	7	3	3	52	7	3110

Figure III.18. 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 Ali est la classe C1 (T<30, P<3000, F<10); et son cluster est le cluster2 ;

Après une nouvelle connexion, le client Ali a changé sa classe du C1 à C5 et son cluster (car C1 et C5 ne sont pas dans le même cluster) ; donc une alerte déclenche et le client pourra pas poursuivre son achats.

Pour la cliente Zineb, elle a changé sa classe de C7 à C8 mais ces derniers sont dans le même cluster (cluster3) donc elle pourra poursuivre leur achats (aucune alerte sera déclenchée).

Figure III.19. Message d'alerte corresponds à la détection d'attaque(2).

IV.2.4 Résultat obtenu avec le K-means :

D'aprés l'imlémentation 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.

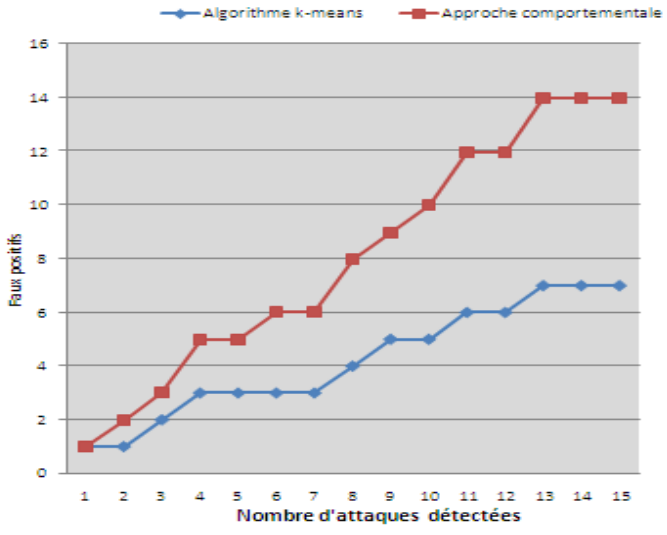

Figure III.20. Variation de faux positifs

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.

V. 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 et de BDD.

En première étape, on a étudié l'approche comportemental qui 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. 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.

Conclusion **g**énérale

La défense en profondeur des réseaux passe par une bonne stratégie préventive pour penser ses réseaux et leurs interconnexions de façon sécurisée. Cette approche doit être complétée une fois le réseau en opération pour permettre de détecter des anomalies qui peuvent être révélatrices. [32]

Ce travail m’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 précisement 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.

Références **b**ibliographiques

[1] Elie MABO, La sécurité des systèmes informatiques (Théorie), support de cours, 2010.

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

[3] Dominique SERET, Ahmed MEHAOUA et Neilze DORTA, « RESEUX ET TELECOMMUNICATIONS », support de cours, Université René Descartes – Paris, 2006.

[4] Laurence Monaco, « Quelques définitions », 2010.

[5] Laurent Bloch et Christophe Wolfhugel, « Sécurité Informatique-principes et méthodes », livre vol.276, P.57, 2007.

[6] Laurent Poinsot «Introduction à la sécurité informatique», support de cours, Université Paris 13.

[7] Vincent Erceau & Romain Colombier, « GMSI Informatique », Projet SAS, 2011.

[8] Emonet Jeam-Burno, « Algorithme de chiffrement – mesures de performances réseaux », 2005.

[9] Rabehi Sidi Mohamed El Amine, « Mise en place d'un serveur radius sous linux pour la sécurisation d'un réseau 802.11 », Projet de fin d'étude, Université Abou Bakr Belkaid, Tlemcen-Algérie, 2011.

[10] Rachid NAIT BEKOU et Younès MOUSSAHHIL, « Etude de fiabilité et conception d'une solution VPN », Mémoire de Projet de fin d'étude, Université Mohammed V SOUSSI, Maroc, 2004.

[11] Elies Jebri, « Introduction à la sécurité», support de cours, 2008 disponible sur url : https://www.google.fr/search?newwindow=1&q=ddata.over-blog.com%2F...%2F0%2F...%2FIntroduction_a_la_securite_2008_elies.pdf, consulté le : Mai 2013.

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

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

[14] Aurobindo Sundaram, Met à jour le: Février 2005, « An Intrusion to Intrusion Detection », ACM Crossroads Student Magazine, disponible sur : http://www.acm.org/crossroads/xrds2-4/intrus.html, Consulté le : Juin 2013.

[15] Tran Van Tay, « LE SYSTEME DE DETECTION DES INTRUSIONS ET LE SYSTEME D'EMPECHEMENT DES INTRUSIONS », Rapport de stage de fin d'études, 2005.

[16] Jacob Zimmermann et al. , « Vers une détection d'intrusion à fiabilité et pertinence prouvable», Thèse de doctorat, Université de Technology, Australie, 2006.

[17] Fissale TCHAKALA, « Optimisation de la sécurité dans un environnement de travail bancaire», mémoire de fin d'étude pour l'obtention de Licence professionnelle, Université de Lomé, Togo, 2011.

[18] Michaël AMAND et Mohamed NSIRI , « Etude d'un système de détection d'intrusion comportemental pour l'analyse du trafic aéroportuaire », Rapport de projet LENAC, 2011

[19] Lalaina KUHN, « VolP & Security :IPS» support de cours, Ecole d'Ingénieurs du Canton de Vaud.

[20] Jabou Chaouki, Schillings Michaël et Hantach Anis, « TER Détection d'anomalies sur le réseau », Rapport de projet, Université Paris Descartes, 2009.

[21] Jacob Zimmermann et Ludovic Mé, « Les systèmes de détection d'intrusions : principes Algorithmiques », disponible sur url : http://www.researchgate.net/publication/239917861_Les_systemes_de_d'etection_d'intrusions_principes_algorithmiques.

[22] MySQL un serveur de bases de données relationnelles disponible sur le site : http://www.futura-sciences.com.

[23] Belhabib abdelkader et Lagha Omar, « Développement d'une application à base de l'algorithme de classification k-means », mémoire de fin d'étude pour l'obtention du diplôme Licence, Université Abou Bakr Belkaid, Tlemcen-Algérie, 2012.

[24] Z.Guellil et L.Zaoui, « Proposition d'une solution au problème d'initialisation cas du K-means », livre : CIIA, volume 547 of CEUR Workshop Proceedings, CEUR-WS.org, Université des Sciences et de la Technologie, Oran – Algérie, 2009.

[25] M.Emre Celebi, Hassan A. Kingravi, Patrico A.Vela, « A Comparative Study of efficient initialization methods for the k-means clustering algorithm», Expert Systems with Applications,2013, Vol.40.

[26] http://www.codeproject.com/Articles/439890/Text-Documents-Clustering-using-K-Means-Algorithm, consulté le Septembre 2013.

[27] ELFOUZI Ilhame, « Clustering des News », projet de fin d'étude, UNIVERSITE DE NICE SOPHIA ANTIPOLIS, 2013.

[28] NetBeans est un environnement de développement intégré disponible sur le site :

http://netbeans.org/.

[29] Ibrahim Mohamed Amine et Tebourbi Hamdi, « Installation et Configuration d'un système de détection d'intrusion (IDS) », Mémoire de licence, Université de 7 novembre, Carthage, 2009.

[30] http://stuff.mit.edu/afs/athena/project/rhel-doc/4/RH-DOCS/rhel-sg-fr-4/ch-detection.html.

[31] Asma CHIKH et Amina DJENNANE, «Sécurité d'une application Web à l'aide d'un système de détection d'intrusions comportementale », mémoire de fin d'étude Master, Université Abou Bakr Belkaid, Tlemcen-Algérie, 2012.

[32] http://www.securite-informatique.gouv.fr/autoformations/securite_reseaux_3/co/secu_reseau_3_2.html, consulté Octobre 2013.

Liste **d**es **f**igures

Liste **d**es **a**bréviations

BDD	**B**ase **D**e **D**onnées
CDDL	**C**ommon **D**evelopment and **D**istribution **L**icense
CERT	**C**omputer **E**mergency **R**esponse **T**eam
HIDS	**H**ost **I**ntrusion **D**etection **S**ystem
HTML	**H**yper **T**ext **M**arkup **L**anguage
HTTP	**H**yper **T**ext **T**ransfer **P**rotocol
IDE	**I**ntegrate **D**evelopment **E**nvironment
IDS	**I**ntrusion **D**etection **S**ystem
IP	**I**nternet **P**rotocol
IPS	**I**ntrusion **P**revention **S**ystem
JSP	**J**ava **S**erver **P**age
KIDS	**K**ernal **I**ntrusion **D**etection **S**ystem
MVC	**M**odel **V**iew **C**ontroller
NIDS	**N**etwork **I**ntrusion **D**etection **S**ystem
SQL	***S**tructured **Q**uery **L**anguage*
SI	***S**ystème d'**I**nformation*
URL	**U**niform **R**esource **L**ocator
VPN	**V**irtual **P**rivate **N**etwork
XML	e**X**tensible **M**arkup **L**anguage

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 présente des points forts, mais aussi des faiblesses qui sont les faux positifs et les faux négatifs. Notre objectif et de gérer la sécurité d'une application 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