République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Saad Dahlab de Blida1

Faculté des sciences

Département Informatique

Mémoire de Master
Pour l'obtention du diplôme de master en informatique

Option
Génie des Systèmes Informatiques

Thème

Conception et implémentation d'une solution à base des honeypots pour l'investigation des crimes informatiques.

Encadré par :
Dr. BOUSTIA Narimene

Présenté par :
Mr. MANSOUR Ilyes
Mr. HAMMADA Adel

Soutenu le : 26/09/2016
Devant le jury :

Mme. Bensetteti Souad	**Présidente**
Mr. Benhabiles Halim	**Examinateur**
Mme. Boustia Narhimene	**Promotrice**

Année universitaire 2015-2016

Résumé

Ce document porte sur l'étude et la mise en place d'une plate forme à base des honeypots pour l'investigation des crimes cybernétiques qui constituent aujourd'hui l'une des nouvelles formes de criminalité et de délinquance, dont les conséquences peuvent être graves à l'échelle nationale et internationale. L'objectif de stage est la conception et l'implémentation d'une solution à base des honeypots en spécifient les composants matériels et logiciels nécessaire à la mise en place de telle plate forme en intégrant les fonctionnalités de journalisation,de supervision et d'analyse de données qui permettent de mesurer, détecter et d'analyser les nouvelles techniques d'attaques.

Le travail réalisé se décompose principalement en deux parties, la première partie traite l'expérimentation des honeypots a forte interaction comme honeynet et la réalisation des attaques sur la plate forme mise en place. Tandis que la seconde partie s'attache à appliquer les concepts de data warehousing et la modélisation dimensionnelle dans l'analyse des données de sécurité collectés via les honeypots déployés.

Mots-clé :
Cybercriminalité, piratage, attaque, honeypot, investigation, modélisation dimensionnelle, entrepôt de données.

Abstract

This paper focuses on the study and implementation of a platform based honeypots for the investigation of cyber crimes that are now one of the new forms of crime and delinquency, the consequences can be serious nationally and internationally. The objective of the course is the design and implementation of a solution based honeypots to specify the hardware and software necessary for the establishment of such a platform by integrating the functionality of logging and monitoring to measure, detect and analyze new attack techniques.

The work is divided into two main parts, the first part deals with the experimentation of a high interaction honeypots as honeynet and carrying out attacks on the platform implementation. While the second part attempts to apply the concepts of data warehousing and dimensional modeling in the analysis of security data collected via the deployed honeypots.

Keywords :

Cybercrime, hacking, attack, honeynet, investigation, modeling dimensional, data warehouse.

Remerciements

Nous remercions avant tout dieu qui nous a aidé à accomplir ce travail.

Nous tenons tout particulièrement à remercier très sincèrement Docteur Boustia Narimene, pour nous avoir dirigés tout au long de ce travail, pour l'aide précieuse, les conseils éclairés et les encouragements qu'ils n'ont cessés de nous prodiguer, qu'ils trouvent ici l'expression de notre profonde reconnaissance.

Nous tenons à exprimer Nos vifs remerciements à : Madame Benstiti chef de département informatique ,Docteur Zahra Fatima Zohra et à l'ensemble du personnel du département pour leurs conseils, encouragements et soutients.

Nos vifs remerciements s'adressent également à l'ensemble du personnel de l'institut national de criminalistque et de criminologie de la Gendarmerie Nationale.

Nous remercions également les membres du jury d'avoir accepter de juger avec volonté ce travail.

Enfin, une pensée concerne bien évidemment tous nos proches, nos amis et en particulier, nos parents, et toutes les personnes qui ont contribué, de près ou de loin, à la réussite de ce projet.

Dédicaces

ILYES

Je dédie ce travail à mes très chers parents.

A mes frères , sœurs, leurs époux et leurs enfants.

A ma femme et mes enfants : Aridj et Abdellah

A tout mes amis.

A tous ceux qui m'ont aidé de prés ou de loin.

Et enfin à la mémoire de mes grands-parents et mes grands-mères.

ADEL

A tous ceux qui me sont chers et tout particulièrement mes parents

A ma femme Nour el Houda et ma petite fille Ritel Bayane

A mes frères , sœurs

A tout mes amis.

Et enfin à tous ceux qui m'ont aidé de prés ou de loin

Sommaire

Introduction 1

1 Sécurité informatique 5

1.1 Définition 5

1.2 Les critères de sécurité 6

1.3 Terminologies et concepts de base 7

1.4 Le Hacker 8

1.4.1 Qu'est-ce qu'un hacker ? 8

1.4.2 Classes des hackers 8

1.4.3 Motivations des Hackers 9

1.4.4 Les phases de hacking 10

1.5 La cybercriminalité 12

1.5.1 Définition 12

1.5.2 Croissance 13

1.6 Mécanismes de sécurité 14

1.6.1 Authentification et contrôles d'accès aux ressources . . 14

1.6.2 Scanners de vulnérabilités 14

1.6.3 La cryptographie 14

1.6.4 Protection anti-virus 15

1.6.5 Les pare-feux (Firewall) 15

1.6.6 Les systemes de détection d'intrusions (IDS) 15

1.6.7 Les honeypots 16

1.7 Résumé 16

2 Le concept des honeypots **17**

2.1 Définition 17

2.2 Objectifs 18

2.3 Caractéristiques souhaitées 19

2.4 Classes de Honeypots 19

2.4.1 Selon le but recherché 20

2.4.2 Selon le degré d'interactivité 21

2.5 Principe de fonctionnement 23

2.5.1 Capture des données 23

2.5.2 Contrôle des données 24

2.5.3 Collecte des données 25

2.5.4 Analyse des données 25

2.6 Honeynet 25

2.7 Architectures de honeynet 26

2.7.1 Première génération :GENI 26

2.7.2 Deuxième génération :GENII 27

2.7.3 Troisième génération :GENIII 28

2.8 Résumé 29

3 Conception et Implémentation **31**

3.1 Conception 31

3.1.1 Les honeypots (réseau interne) 33

3.1.2 Le système honeywall 33

3.1.3 Le système de management 33

3.1.4 Le réseau externe 33

3.1.5 L'entrepôt de données de sécurité 34

3.2 Implémentation 34

3.2.1 Description de la plateforme 34

3.2.2 Installation de la plateforme 35

3.2.3 Installation et Configuration du Honeywall 36

3.2.4 Architecture de Honeywall 38

3.3 L'entrepôt de données de sécurité (Module d'analyse de données) 40
3.3.1 Contexte 40
3.3.2 Modélisation et Conception du Data Warehouse 40
3.3.3 Alimentation du Data Warehouse 47
3.3.4 Résumé 48

4 Expérimentation et Analyse de données **49**
4.1 Expérimentation 49
4.1.1 Étude de cas 50
4.2 Analyse de données et présentation 53
4.2.1 Analyse dimensionnelle des données 53
4.2.2 Présentation des données (Reporting) 54
4.3 Résumé 56

Conclusion **58**

A Le Data Warehouse **1**

A.1 Qu'est ce qu'un Data Warehouse 1
A.2 Historique des Data Warehouse 2
A.3 Structure des données d'un Data Warehouse 3
A.4 Les éléments d'un Data Warehouse 4
A.5 Architecture d'un Data Warehouse 5
A.6 Modélisation des données de l'entrepôt 5
A.7 Le concept OLAP : 7

Table des figures

1.1 Les critères de sécurité 6
1.2 Les attaques passives et actives 7
1.3 Les phases de hacking 10
1.4 Le nombre de crime enregistré sur internet par an. 13
1.5 Le nombre de menaces engendrées par Internet 13

2.1 Mécanisme de contrôle des données 24
2.2 Architecture de la première génération des honeynets [21]. . . 26
2.3 Architecture de la deuxième génération des honeynets 27
2.4 Architecture de la Troisième génération des honeynets 29

3.1 Conception de la solution à base de honeynet virtuel 32
3.2 Architecture de honeynet virtuel 35
3.3 Interface d'installation de honeywall Roo 1.4 36
3.4 Menu de configuration 37
3.5 Interface de configuration Walleye 37
3.6 Architecture de système honeywall [24] 38
3.7 Interface principal de walleye 39
3.8 Les attaques enregistrées 39
3.9 Le Modèle en étoile de l'entrepôt de données de sécurité . . . 46
3.10 Processus ETL 47
3.11 View de données sources 48

4.1 Exécution des commandes d'attaque 51
4.2 La page de site altérée. 51

4.3 Récupération de compte et mot de passe du victime 52
4.4 Analyse de trafic capturé par le Honeywall 52
4.5 Déploiement de cube créé 54
4.6 Nombre d'attaque par protocole par jour 55
4.7 Nombre d'attaques par @ ip source vers la même @ ip destination 55
4.8 Nombre d'attaque par jour selon des critères de sélection 56

A.1 Évolution des bases de données décisionnelles. 3
A.2 Structure des données d'un Data Warehouse. 4
A.3 Architecture globale d'un Data Warehouse 5

Liste des tableaux

2.1 Classes des honeypots selon le degré d'interactivité[20]. 22

3.1 Table Dimension Temps. 42
3.2 Table Dimension Machine Source. 43
3.3 Table Dimension Machine Destination. 43
3.4 Table Dimension Protocole. 44
3.5 Table Dimension Action. 44
3.6 Table Fait Attaque. 45

Introduction

Aujourd'hui la cybercriminalité constitue l'une des nouvelles formes de criminalité et de délinquance, dont les conséquences peuvent être graves à l'échelle nationale et internationale. Le caractère virtuel des échanges qui débutent sur Internet permet l'apparition d'un bon nombre de crimes informatiques par nature immatériel commises par des hackers qui exploitent les failles matérielles ou logicielles afin de pouvoir accéder à des données confidentielles, prendre le contrôle des machines ou répandre des logiciels malveillants (vers, chevaux de troie. . .) en dépit des lois.

Pour contrer ces crimes, plusieurs méthodes d'investigation et des contre mesures ont été développées sur la base des attaques connues, mais pour prévenir les prochaines attaques, il devient nécessaire aujourd'hui de connaitre les techniques des attaquants, d'en savoir plus sur leur motivation et sur ce qu'ils recherchent. Pour cela il convient d'analyser et étudier leurs nouvelles techniques afin d'anticiper au mieux les prochaines attaques et d'élaborer des nouveaux mécanismes de protection adaptés et d'améliorer les techniques d'investigations des crimes informatiques.

Les "honeypots" sont un des moyens de contre mesure permettant de mettre en place des processus proactifs pour mieux lutter contre les crimes informatique. Dans le but de mesurer, détecter et prévenir les activités illicites, les honeypots représentent une formidable technologie dont l'utilisation permet de leurrer et tromper les attaquants pour mieux les appréhender.

Le travail présenté dans ce mémoire s'inscrit dans ce contexte. Notre objectif est de concevoir et de mettre en place une solution à base des honeypots pour l'investigation des crimes informatiques qui réunie les différentes fonctionnalités

nécessaire à l'expérimentation et l'étude de différents modes opératoires des attaques. Elle comporte aussi un module d'analyse de données de sécurité collectées via les honeypots qui traite, analyse et mesure les attaques commises contre les honeypots déployés afin de faciliter la tache des administrateurs de sécurité en matière de gestion et d'analyse de grand nombre d'alertes et d'événements et de permettre aussi d'effectuer des statistiques et des analyses efficaces en utilisant la technologie de datawarehousing et la modélisation dimensionnelle.

Ce mémoire présente, dans le premier chapitre les notions de bases de la sécurité informatique et les différentes phases de hacking. Il met l'accent sur l'évolution de la criminalité informatique et les mécanismes de sécurité et de contre-mesure pour faire face aux délinquants informatiques.

Le deuxième chapitre de ce document tente de définir les termes "Honeypots" et "Honeynets", d'introduire les concepts de base relatifs aux systèmes leurres. Il décrive en détail le principe de fonctionnement des honeypots, leurs classes et leurs architectures.

Le chapitre suivant s'attardera plus particulièrement sur la conception et l'implémentation de notre solution a base des honeypots qui a été mise en œuvre et exploité avec nombreuses possibilités d'utilisation. Il montre aussi la conception du modèle dimensionnel de données et l'implémentation de l'entrepôt de données de sécurité.

Le dernier chapitre s'attarde à présenter les résultats obtenus au cours de ce stage avec l'expérimentation de plate-forme mise en place avec des cas d'attaques réels. Le module d'analyse de données de sécurité collectées via les honeypots a également été testé et sera détaillé.

Pour finir, la conclusion générale synthétise les principaux résultats obtenus dans le cadre de nos travaux et présente quelques perspectives.

Partie Théorique

Chapitre 1

Sécurité informatique

La sécurité informatique devient un véritable préoccupation des institutions et entreprises avec le développement des réseaux, la dématérialisation des échanges d'information et surtout avec l'apparition de la criminalité informatique qui évoluent exponentiellement.

Dans ce chapitre,nous présentons certains aspects de la sécurité informatique et les différents phases de haking, ensuite nous mettons l'accent sur l'évolution de la criminalité informatique et les mécanismes de sécurité et de contre-mesure pour faire face aux délinquants informatiques.

1.1 Définition

La sécurité informatique est l'ensemble des moyens techniques, organisationnels, juridiques et humains nécessaires et mis en place pour conserver, rétablir, et garantir la sécurité des systèmes informatiques. Elle est principalement liée à la sécurité de l'information et des systèmes d'information . D'une manière générale, la sécurité informatique consiste à assurer que les ressources matérielles ou logicielles d'une organisation sont uniquement utilisées dans le cadre prévu [1].

1.2 Les critères de sécurité

Pour remédier aux failles et pour contrer les attaques, la sécurité informatique se base sur un certain nombre de critères qui permettent de mettre en place une réponse appropriée à chaque menace. Les principaux services de sécurité [1] indiqués dans la figure 1.1 sont :

* **La confidentialité** : les informations ne doivent être accessible qu'aux seules personnes autorisées ou habilitées [2].

* **L'intégrité** : les ressources et les données qu'elles contiennent doivent être protégées contre les changements imprévus.

* **La disponibilité** : permettant de maintenir le bon fonctionnement du système d'information, les ressources et les données qu'elles contiennent doivent être disponibles à tout moment aux personnes autorisées.

* **L'authentification** : consiste à assurer que seules les personnes autorisées aient accès aux ressources.

* **La non-répudiation** : pour éviter la contestation par l'émetteur de l'envoi de données, la non répudiation est une propriété qui assure que l'auteur d'un acte ne peut ensuite nier l'avoir effectué (signature de l'acte) et que le récepteur ne peut ultérieurement dénier avoir reçu un message.

FIGURE 1.1 – Les critères de sécurité

1.3 Terminologies et concepts de base

Afin de comprendre mieux l'aspect de la sécurité informatique, il est primordiale de définir quelques notions de base à savoir :

* **attaque** : est une agression contre une machine par une personne n'ayant pas les droits sur elle. Elle est une tentative de contournement des contrôles de sécurité sur un matériel ou service informatique (serveur, routeur, application, etc.).Le succès de l'attaque dépend de la vulnérabilité du matériel ou service attaqué, mais si elle réussit, l'attaquant peut avoir un accès illimité au système d'information et engendrer alors des dégâts importants [3].

* **Types d'attaques** : Les attaques sont classées en deux grandes catégories à savoir, les attaques passives qui consistent à écouter sans modifier les données ou le fonctionnement du réseau et les attaques actives qui consistent à modifier les données, à s'introduire dans des équipements réseau ou à perturber le fonctionnement de réseau.

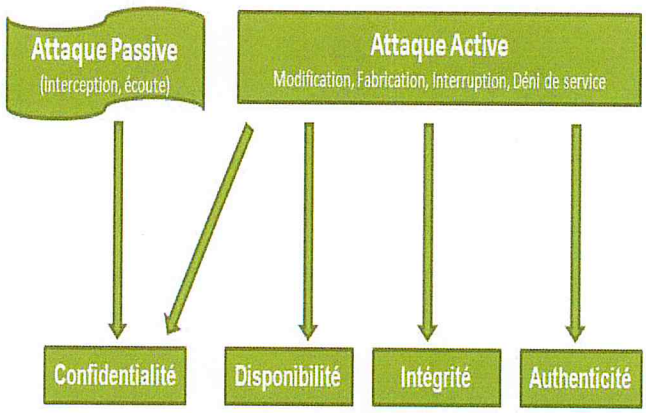

FIGURE 1.2 – Les attaques passives et actives

* **Intrusion** : événement ou combinaison d'événements permettant d'avoir indûment accès (sans autorisation) à un système et ses ressources.

* **Vulnérabilité (Vulnerability)** : défaut ou faiblesse dans la conception d'un système, son implémentation, fonctionnement ou administration et qui pourrait être exploité pour violer la politique de sécurité.

* **Menace (Threats)** : danger potentiel pouvant exploiter une vulnérabilité pour violer la politique de sécurité causant éventuellement des dégâts.

* **Exploits** : elles représentent les moyens d'exploiter une vulnérabilité. Il peut y avoir plusieurs attaques pour une même vulnérabilité, mais toutes les vulnérabilités ne sont pas exploitables [4].

* **Les risques (Risk)** :se définissent comme une combinaison de menaces exploitant une vulnérabilité et pouvant avoir un impact. De manière générale, les risques sont soit des causes (attaques, pannes...) soit des conséquences (fraude, intrusion, divulgation...).

* **Les contre-mesures** : ce sont les procédures ou techniques permettant de résoudre une vulnérabilité ou de contrer une attaque spécifique [5].

1.4 Le Hacker

1.4.1 Qu'est-ce qu'un hacker ?

Le terme hacker a eu plus d'une signification depuis son apparition à la fin des années 50. A l'origine ce nom désignait d'une façon méliorative les programmeurs expérimentés, puis il servit au cours des années 70 à décrire les révolutionnaires de l'informatique, qui pour la plupart sont devenus les fondateurs des plus grandes entreprises informatiques. C'est au cours des années 80 que ce mot a été utilisé pour catégoriser les personnes impliquées dans le piratage de jeux vidéos, en désamorçant les protections de ces derniers, puis en revendant des copies.

Aujourd'hui ce mot désigné à la fois celui qui écrit le code et celui qui l'exploit , même si ces deux groupes de hackers ont des objectifs différents, ils emploient des techniques similaires afin de pirater les systèmes informatiques [6].

1.4.2 Classes des hackers

En réalité il existe de nombreuses classes de hackers (attaquants) catégorisés selon leurs expériences et selon leurs motivations [7] :

- **White hat hacker** : un hacker qui pénètre par effraction dans des systèmes

ou des réseaux dans l'objectif d'aider les propriétaires du système à mieux le sécuriser. Ce sont des personnes effectuant des tests d'intrusions en accord avec leurs clients et la législation en vigueur afin de qualifier le niveau de sécurité de leurs systèmes. Les objectifs des white hat hackers sont en règle générale un des suivants : L'apprentissage, l'optimisation des systèmes informatiques et la mise à l'épreuve des technologies jusqu'à leurs limites afin de tendre vers un idéal plus performant et plus sûr.

- Black hat hacker : sont des personnes s'introduisant dans les systèmes informatiques dans un but nuisible ; ils sont des créateurs de virus, cyber-espions, cyber-terroristes et cyber-escrocs, agissant dans le but soit de nuire soit de tirer profit de leurs actes illégaux. Parmi ses motivations on peut citer : l'envie de nuire (détruire des données, empêcher un système de fonctionner) ; l'intérêt financier ; terrorisme ; espionnage "classique" ou industriel et Chantage.

- Grey hat hacker : un hacker hybride entre les chapeaux blancs et chapeaux noirs. Ils n'hésitent pas à pénétrer dans les systèmes sans y être autorisés, ils n'ont pas pour but premier de nuire. C'est souvent l'exploit informatique qui les motive, une façon de faire la preuve de leur savoir-faire.

- Les scripts kiddies : (gamins du script, crashers, lamers) sont des jeunes utilisateurs du réseau utilisant des programmes trouvés sur Internet, généralement de façon maladroite, pour vandaliser des systèmes informatiques afin de s'amuser.

- Les hacktivistes : sont des hackers dont la motivation est principalement idéologique et politique, ils emploient leurs connaissances en informatique pour diffuser et promulguer leurs opinions. Ses actions les plus spectaculaires sont notamment le piratage de sites informatiques en altérant les données, en détournant des serveurs, en remplaçant des pages d'accueil afin de détourner la signification et l'engagement de ces sites.

1.4.3 Motivations des Hackers

Les motivations des attaquants peuvent être de différentes sortes [8] :

- Obtenir un accès au système ;
- Voler des informations, telles que des secrets industriels ou des propriétés

Intellectuelles. Et s'informer sur l'organisation (entreprise de l'utilisateur, etc.) ;

- Capter des informations personnelles sur un utilisateur ;
- Récupérer des données bancaires ;
- Entraver le bon fonctionnement d'un service ;
- Utiliser les ressources du système de l'utilisateur, notamment lorsque le réseau sur lequel il est situé possède une bande passante élevée.
- Vérification de la sécurisation d'un système.
- Terrorisme, Manifestation politique, Espionnage "classique" ou industriel.
- Par simple "jeu", par défi.

1.4.4 Les phases de hacking

Les hackers réalisent généralement une attaque en utilisant cinq phases communes. Il est important de comprendre ces phases d'attaques de piratage afin de mieux se défendre contre eux. La figure 1.3 illustre les cinq phases [9] que les hackers suivent généralement pour hacker un système.

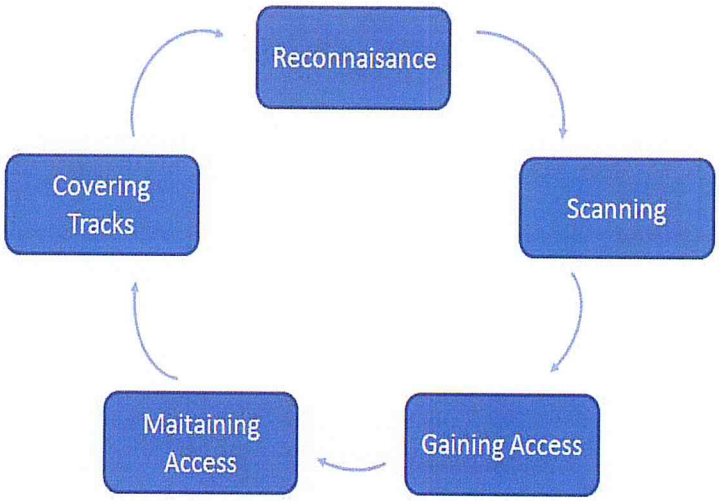

FIGURE 1.3 – Les phases de hacking

- **Phase 1 : Reconnaissance** La première phase de réalisation d'une attaque est relative à la collecte d'information sur la cible et la recherche des vulnérabilités d'un système, cela consiste essentiellement à prendre connais-

sance des mécanismes et des niveaux de sécurité en vigueur concernant l'identification, l'authentification, le contrôle d'accès, cryptographie, la surveillance et à identifier les failles techniques organisationnelles et humaines de l'environnement.

- **Phase 2 : Balayage (Scanning)** le balayage consiste à prendre l'information découverte pendant la phase reconnaissance et de l'exploiter à examiner le réseau. Les intrus cherchent de toute information qui peut les aider à commettre l'attaque telle que des noms d'ordinateur, adresses IP, et comptes d'utilisateur.
- **Phase 3 : Intrusion (Gaining Access)** C'est la phase réelle de hacking. Les Vulnérabilités découvertes pendant les phases de reconnaissance et de balayage sont maintenant exploitées pour accéder. La méthode d'intrusion utilise les failles et les exploits existent. par exemples incluent des débordements de tampon, déni du service (DOS), et hacking de session, etc.
- **Phase 4 : Garder l'accès (maintaining Access)** Une fois qu'un intrus a accédé, il veut garder l'accès pour des futures exploitations et attaques. Parfois, les hackers gardent l'accès avec des backdoors, des rootkits, et Trojans. Une fois que l'intrus attaque le système, ils peuvent l'employer comme base pour lancer des attaques additionnelles. Dans ce cas, le système attaqué est désigné parfois sous le nom de système zombi.
- **Phase 5 : Elimination de traces (Covering Tracks)** La phase d'élimination de traces a pour objectifs principaux de faire en sorte que l'attaque ne puisse être détectée et que l'attaquant ne laisse pas de trace pouvant servir à son identification. Pour contribuer à cela, il tente de rester anonyme, il peut alors utiliser des alias (pseudonymes), usurper l'identité numérique d'utilisateurs ou encore brouiller les pistes en passant par plusieurs systèmes intermédiaires (relais). Les intrus essayent d'enlever toutes les traces de l'attaque, telles que des dossiers de notation ou alarmes du système de détection d'intrusion. Pour continuer à utiliser le système attaqué, enlever l'évidence de hacking, et éviter l'action judiciaire[10].

1.5 La cybercriminalité

1.5.1 Définition

La cybercriminalité constitue l'une des nouvelles formes de criminalité et de délinquance transnationales dont laquelle les systèmes et les réseaux informatiques sont un outil, une cible ou un lieu d'attaques criminelles.

Elle englobe trois catégories d'activités criminelles :

- La première concerne les formes traditionnelles de criminalité, comme les fraudes ou les falsifications ;
- La seconde concerne les infractions liées aux contenus illicites par voie électronique (violence sexuelle contre les enfants, incitation à la haine raciale...) ;
- La dernière connait des infractions propres aux réseaux électroniques (attaques visant les systèmes d'information, déni de service, piratage...)[11].

La criminalité informatique répond à cinq postulats de base, par ailleurs valables pour d'autres formes de criminalité et qui correspondent en fait à la théorie criminologie des rapports coût/avantages dite théorie économique du crime de Gary S. Becker.

- Tout système informatique est vulnérable et comporte des failles.
- Toute personne ayant accès à un système informatique est susceptible de découvrir ces failles.
- Quiconque découvre ces failles peut être tenté de les utiliser à son profit.
- Si les risques sont faibles, ces failles seront utilisées : la délinquance est souvent à la recherche de la solution.
- Ces failles seront d'autant plus facilement utilisées que l'on opère de l'intérieur de l'entreprise et que les enjeux sont importants [12].

1.5.2 Croissance

- L'évolution de l'internet à connue aussi une évolution dans les crimes informatique, comme montre le rapport (Internet Crime Report) du (IC3 [1]) une croissance du nombre de crime sur internet (figure 1.4).

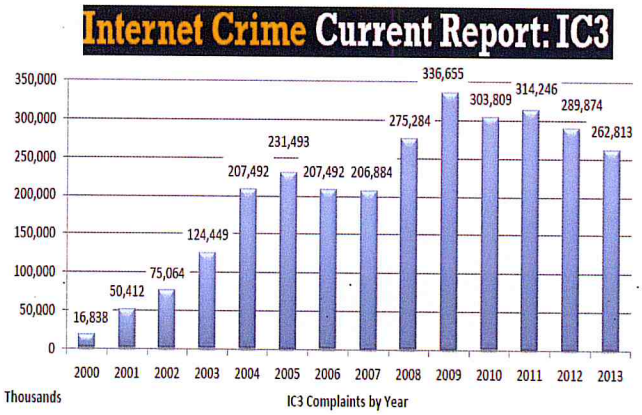

FIGURE 1.4 – Le nombre de crime enregistré sur internet par an.

- McAfee dans sont rapport annuel,Publié en 2009,sur les menaces de sécurité engendrées par Internet, confirme une nette augmentation du nombre de menaces conçues pour commettre des crimes informatiques et qui vise davantage à voler les données à l'insu des utilisateurs.

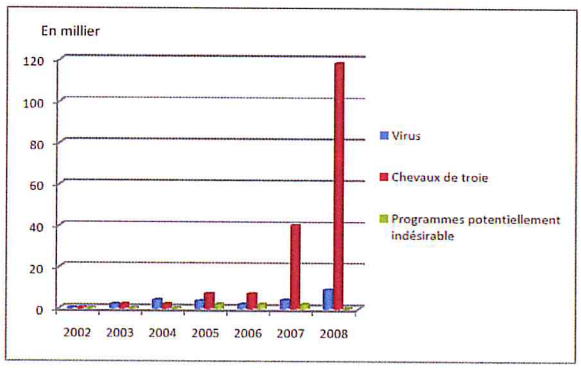

FIGURE 1.5 – Le nombre de menaces engendrées par Internet

1. The Internet Crime Complaint Center www.ic3.gov

1.6 Mécanismes de sécurité

Pour faire face aux dégats qu'engendre le piratage, des mécanismes et outils de protection ont été d'eveloppés en tant que contre-mesure, on peut citer les plus repandus :

1.6.1 Authentification et contrôles d'accès aux ressources

Il s'agit en premier lieu de la sécurité physique des équipements et installations. L'authentification est un mécanisme permettant de prouver l'identité d'un utilisateur (mots de passe, cartes à puce, méthodes biométriques, etc.) et de lui accorder uniquement les privilèges nécessaires pour l'accomplissement de ses tâches.

1.6.2 Scanners de vulnérabilités

Les scanners de vulnérabilités automatisent la découverte des failles de sécurité. Ils sont utilisés par les attaquants pour localiser les faiblesses du réseau cible. De plus les administrateurs peuvent en tirer profit pour corriger les vulnérabilités de leurs systèmes informatique. Cependant, malgré le grand nombre de vulnérabilités détectées, les scanners d'aujourd'hui sont inaptes à déterminer toutes les faiblesses possibles. De plus, la mise à jour de ces produits ne suit pas le rythme de la découverte des nouvelles vulnérabilités.

1.6.3 La cryptographie

Les informations sensibles et confidentielles sont souvent cryptées pour empêcher leur lecture même si elles étaient dérobées ou accédées frauduleusement. La cryptographie garantit la confidentialité, l'intégrité, la non répudiation et l'authenticité des données mais elle ne constitue pas une solution unique et suffisante de sécurité.

1.6.4 Protection anti-virus

Les logiciels anti-virus sont largement utilisés pour les stations de travail et ordinateurs personnels. Ils détectent et protègent contre les virus pouvant se propager à travers des fichiers, courrier électronique, etc.

1.6.5 Les pare-feux (Firewall)

Un pare-feu est un outil permettant de contrôler le trafic circulant entre l'intérieur et l'extérieur d'un périmètre de sécurité. Le périmètre de sécurité constitue la limite entre le réseau que l'on considère comme sûr ou que l'on désire protéger et le reste de l'Internet [13].

Un pare-feu assure d'abord une fonction de filtrage des flux entrants et sortants puisqu'il est un passage obligé pour tout échange. Le risque d'attaques distantes est alors réduit puisque le pare-feu ne laisse passer que le trafic d'une certaine plage d'adresses IP et relatif à certains ports et services. Malgré leurs grands intérêts, les pare feux présentent quelques lacunes. En effet, un attaquant peut exploiter les ports laissés ouverts pour pénétrer au réseau local. Les scripts constituent aussi des sources d'intrusion que les pare feux échouent à détecter. Ainsi l'opération supplémentaire d'encapsulation/décapsulation des données permet à l'attaquant de contourner le pare feu.

1.6.6 Les systemes de détection d'intrusions (IDS)

la détection d'intrusion concerne l'ensemble des pratiques et mécanismes utilisés pour la détection d'erreurs pouvant conduire à une défaillance de sécurité, et/ou pour la détection d'attaques [14]. Un IDS est un périphérique ou processus actif qui analyse l'activité du système et du réseau pour détecter toute entrée non autorisée et ou toute activité malveillante. La manière dont un IDS détecte des anomalies peut beaucoup varier ; cependant, l'objectif principal de tout IDS est de prendre sur le fait des auteurs avant qu'ils ne puissent vraiment endommager les ressources.

Les IDS protègent un système contre les attaques, les mauvaises utilisations et les compromis. Ils peuvent également surveiller l'activité du réseau, analyser les configurations du système et du réseau contre toute vulnérabilité, analyser l'intégrité de données et bien plus.

1.6.7 Les honeypots

Un honeypot se définit comme un système informatique connecté à un réseau, volontairement vulnérable à une ou plusieurs failles et visant à attirer les attaquants afin d'étudier leur comportement[15].

1.7 Résumé

Au cours de ce chapitre, nous avons présenté certains aspects et notions de la sécurité informatique, les différentes phases de haking,les mécanismes de sécurité et de contre-mesure nécessaires, ainsi que l'évolution de la criminalité informatique.

Dans le chapitre suivant nous nous intéresserons à l'un des mécanismes de sécurité en l'occurrence les honeypots. A la base de ces derniers, nous proposerons une solution pour contrer des attaques informatiques.

Chapitre 2

Le concept des honeypots

Aujourd'hui la masse d'informations circulant entre les divers systèmes reliés au réseau mondial (Internet) est de plus en plus importante. Et dans le même temps le nombre d'attaques menées depuis et vers un système informatique croît de manière incessante.

Pour contrer ces attaques plusieurs méthodes de contre-mesures se sont développées sur la base des attaques connues, mais pour prévenir les prochaines attaques, il devient nécessaire aujourd'hui de connaître les méthodes des attaquants, d'en savoir plus sur leur motivations et sur ce qu'ils recherchent, pour cela en utilisant la technologie des honeypots. Dans ce chapitre nous décrivons en détail le concept des honeypots ,le principe de fonctionnement, les classes et leurs architectures.

2.1 Définition

L'idée d'observer des attaquants sur des systèmes informatiques grâce à des leurres est née dans les années quatre-vingt. Stoll a observé une attaque dans le laboratoire Berkeley. Il a pris la décision de distribuer des faux documents et il a surveillé les accés à celles-ci. En suite, il a pu identifier l'attaquant. Un peu plus tard dans les années quatre vingt-dix, Cheswick a eu l'idée de déployer des faux services accessibles depuis Internet afin d'observer des attaques. Dans les années 2000, Spitzner a introduit la terminologie des honeypots qui est de nos jours acceptée dans la majorité des communautés de sécurité informatique[16].

Un honeypot est utilisé pour attirer le pirate informatique sur une ressource vulnérable et l'empêcher ensuite d'utiliser cette ressource de manière non contrôlée pour rebondir vers l'extérieur.Une définition simplifiée du terme "honeypot" pourrait être : *Un Honeypot est une ressource dédiée à être découverte, exploitée et attaquée*[17]. Généralement, un honeypot est un système dédié, simulant au mieux un système opérationnel classique.

Il est également possible d'émuler un réseau complet de système de production (avec ses services Web, les données vives d'une entreprise), on parle alors de *honeynet* comme un ensemble de honeypots destiné à émuler un service complet d'entreprise.

2.2 Objectifs

Un honeypot est un système leurre qui peut être relié directement au réseau Internet ou bien disposé au sein du réseau de l'entreprise. Son objectif est de permettre[18] :

- D'étudier la nature du trafic (d'attaque) à destination de ce système.

- En cas de compromission, d'analyser les données générées par l'activité du pirate et d'apprendre sur les outils, tactiques et motivations des pirates.

- De réduire considérablement le taux de fausses alarmes et le nombre d'attaques non détectées liés généralement à l'utilisation de système de détection d'intrusions (IDS).

- De mesurer l'activité illicite ou anormale afin d'ajuster le niveau de protection nécessaire à la sécurité du réseau de l'entreprise.

2.3 Caractéristiques souhaitées

Pour répondre à la mission qui lui est confié, un honeypot doit intégrer les caractéristiques suivantes :

- Attirer l'attaquant. En effet, un honeypot qui ne subit pas d'attaques ne peut pas être compromis et dans ce cas ne sert à rien. Un honeypot doit pouvoir offrir des services réseaux attrayants, tels que SMTP [1], FTP [2], ... pour que sa présence se révèle utile.

- Contrôler l'activité sur le système leurre. Dans le cas où le système est compromis, l'attaquant ne doit pas pouvoir mener des actions illicites contre les systèmes du réseau interne de l'organisation ou bien contre des systèmes reliés à l'Internet. Sinon, le système leurre perd tout son intérêt et peut se montrer dangereux pour l'organisation.

- Collecter un maximum d'informations. Le fait que le honeypot puisse collecter les informations, permet de les analyser ultérieurement après l'attaque et la compromission du système leurre si celle-ci réussi.

- Alerter en cas d'attaque. Afin de pouvoir observer l'activité de l'attaquant encore faut-il être prévenu que celui-ci tente de compromettre le système.

2.4 Classes de Honeypots

Il existe deux grandes classes de honeypots catégorisées selon le but recherché et selon le degré d'interactivité.

1. Simple mail Transfert Protocol
2. File transfert Protocol

2.4.1 Selon le but recherché

Marty Roesh, le créateur de Snort [3] définit deux catégories de honeypots en fonction du but recherché : les honeypots orientés production et les honeypots orientés recherche.

Honeypots orientés production

Cette première classe de honeypot vise à réduire les vulnérabilités possibles dans une organisation en aidant à prendre les bonnes mesures de sécurité. L'objectif d'un tel honeypot est de détecter les attaques et vérifier que les mesures de sécurité sont appliquées. Cette catégorie de honeypot est généralement plus facile à réaliser et à déployer car elle nécessite généralement moins de fonctionnalités, ce qui a pour conséquence de diminuer le niveau de risque lié à leur déploiement dans l'entreprise. Généralement, les honeypots orientés production sont utilisés par les entreprises désireuses d'améliorer la protection de leur réseau, par exemple pour renforcer la capacité à détecter les attaques.

Honeypots orientés recherche

Cette seconde classe de honeypots est conçue pour récolter le maximum d'informations sur la communauté pirate. L'objectif d'un tel honeypot n'a pas d'impact direct sur la sécurité d'une organisation spécifique mais est plutôt destiné à rechercher et étudier les futures menaces et ainsi à mieux les anticiper. Les honeypots orientés recherche sont principalement destinés aux laboratoires de recherche des universités, des sociétés ou des services de défense ainsi qu'aux entités du CERT [4] pour qui la veille technologique en matière de sécurité constitue l'activité principale.

3. IDS orienté réseau, http ://www.snort.org/
4. Computer Emergency Response Team, www.cert.org

2.4.2 Selon le degré d'interactivité

Cette classification de Lance Spitzner[19] est basée sur le niveau d'interaction entre l'attaquant et le honeypot, trois niveaux d'interactions peuvent être définis de la manière suivante (table 2.1) :

Faible interactivité avec l'attaquant.

Dans ce cas un honeypot est facile à installer et à configurer et se contente d'émuler certains services ou systèmes d'exploitations. Un attaquant peut potentiellement mener une série de scans, se connecter aux services seulement, Les informations collectées sont d'un intérêt limité : qui s'est connecté, sur quel port et quand. L'avantage est que le niveau de risque est très faible puisque l'attaquant n'interagit pas avec un protocole complexe ni un système d'exploitation. Mais c'est aussi un désavantage car l'observation du pirate avec le système n'est pas possible. En effet, le processus qui émule le service ne fait qu'écouter et ne renvoie aucune réponse. (moins risqué, peu de maintenance)

Forte interactivité avec l'attaquant.

Un honeypot avec un tel niveau d'interactivité est un système non émulé dans lequel le pirate peut interagir directement avec le système d'exploitation une fois le système compromis. Aucune hypothèse n'est faite sur le comportement du pirate, ainsi toute son activité est capturée. Les informations collectées sur l'activité du pirate sont plus nombreuses et ont une valeur très importante puisque le pirate se comporte comme sur un vrai système de production. Et donc l'analyse de ces données est riche d'enseignements. Mais, un tel système demande une installation plus complexe et une surveillance de l'activité beaucoup plus intense car en offrant plus d'interactivité le pirate cherchera rapidement à obtenir tous les privilèges sur le système leurre pour ensuite l'utiliser comme instrument pour mener des attaques vers d'autres systèmes(internes ou extérieurs à l'organisation). Le contrôle du trafic sortant est donc un point essentiel. (plus risqué, plus complexe à maintenir, capture de données plus intéressantes).

Moyenne interactivité avec l'attaquant.

On peut définir également, entre les deux niveaux d'interactivité précédents, un niveau intermédiaire où certains services émulés par le pot de miel ne se contentent pas seulement d'enregistrer les connexions et les informations reçues mais émulent des protocoles complexes en envoyant des réponses adaptées aux informations demandées par l'attaquant. Celui-ci a l'illusion d'un véritable service mais n'interagit pas directement avec le système d'exploitation. Les risques de compromissions sont donc limités et les informations collectées sont plus intéressantes que dans le cas de systèmes ayant une faible interactivité. Cependant, l'installation d'un tel système leurre demande un développement complexe et une très bonne connaissance des protocoles émulés. De plus, les services développés ne doivent pas souffrir des même faiblesses que les implémentations des protocoles originaux.

Niveau d'interactivité	**Informations collectées**	**Exemples**
Honeypots à faible interaction	Heure et date de l'attaque, protocole, adresse IP source,adresse IP destination et les ports source et destination	Spectre, Honeyd
Honeypots à moyen interaction	Information sur les attaque complexes comme les botnets	Nepenthes, Mwcollect, Honeytrap, SGNET
Honeypots à forte interaction	Information sur la compétence de l'attaquant et informations possible sur les attaques Zero-day	**Honeynet**, Argos, Minos

TABLE 2.1 – Classes des honeypots selon le degré d'interactivité[20].

2.5 Principe de fonctionnement

Un honeypot est une ressource en attente d'être attaquée afin d'en étudier les techniques de compromissions. Cette ressource n'est pas destinée à offrir un quelconque service de production et pour cette raison tout trafic à destination ou en provenance du honeypot est suspect par défaut. Ceci signifie que chaque nouvelle connexion initiée vers un honeypot est potentiellement le début d'une attaque. De la même manière, chaque connexion initiée depuis un honeypot révèle que celui-ci est compromis. Les paragraphes suivants introduisent les points importants à prendre en compte dans le déploiement des honeypots[18].

2.5.1 Capture des données

A l'instar des IDS, la collecte d'informations *(Data Capture)* s'effectue à partir de deux sources de données distinctes. La première source de données est le système leurre lui-même (activité système) et la seconde provient des fichiers journaux enregistrés par les pare-feu et autres IDS (activité réseau).

Informations issues du système leurre

Les informations capturées sur le système leurre peuvent être classées selon deux catégories :

- Les informations qui génèrent des flux de données comme par exemple les frappes de touches clavier.
- Les informations pouvant être recueillies par les fonctions d'administrations comme l'utilisation de la mémoire, du CPU et la liste des processus courant.

Informations issues du trafic réseau

Elles représentent un bon moyen permettant l'analyse de l'activité du pirate. Ces informations peuvent être enregistrées soit par le pare-feu qui contrôle les flux sortant du honeypot, soit par une sonde d'écoute réseau (sniffing).L'intérêt de la capture d'informations à partir du réseau est que tous les paquets sont

enregistrés et il est possible d'analyser après coup la totalité de l'activité intrusive et de disposer de l'historique complet de l'attaque et de la compromission.

2.5.2 Contrôle des données

Le contrôle de données *(Data Control)* permet de limiter les risques lors de la compromission du honeypot en mettant en oeuvre des mécanismes de contrôle destinés à bloquer ou inhiber tout trafic pouvant s'avérer dangereux pour les réseaux externes au honeypot (Internet ou réseaux de production de l'entreprise).

Le contrôle des données doit offrir des mécanismes de protection redondants et efficaces sans toutefois dévoiler la présence du honeypot à l'attaquant. En effet, lorsque le honeypot est compromis l'attaquant cherche un accès Internet de manière à consolider son attaque. Une certaine liberté d'action doit être accordée au pirate pour qu'il puisse réaliser son attaque mais ses actions devront être surveillées et rester sous contrôle[18].

On cherchera, par exemple, à limiter le nombre maximum de connexions possibles par protocole. Au-delà de cette limite, ses connexions seront bloquées. On pourra également chercher à désactiver la dangerosité de ses requêtes en modifiant dynamiquement celles-ci pour les rendre inopérantes.

Plus la liberté d'action du pirate est grande et plus on en apprendra sur ses motivations et sur ses outils et ses méthodes, mais en contrepartie le risque encouru sera aussi plus important. [5]

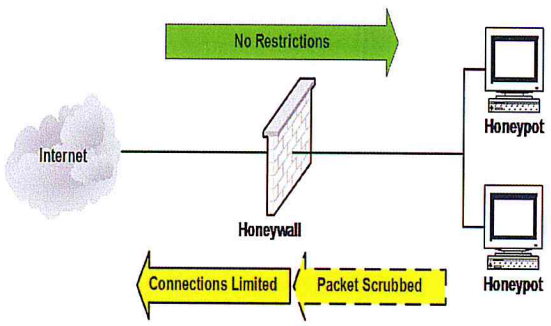

FIGURE 2.1 – Mécanisme de contrôle des données

5. www.tracking-hackers.com

2.5.3 Collecte des données

La collecte des données *(Data Collection* est réservée aux organisations qui déploient de multiple honeynets afin d'accroître leur potentiel de recherche. The Honeynet Project est un exemple de ce type d'organisation. L'objectif de la collecte de données est de centraliser l'administration et de corréler les informations en provenance des différents sites. On utilisera une base de noms cohérente (IP/DNS) entre les différents honeynets, une référence temporelle commune (exemple : le protocole NTP [6] pour la synchronisation des horloges) et un mécanisme de transmission assurant la confidentialité, l'intégrité et l'authenticité des données recueillies.

2.5.4 Analyse des données

L'analyse des données *(Data Analysis)* regroupe deux fonctions : la première est la possibilité du Honeynet à émettre des alertes aux administrateurs du Honeynet lorsqu'il est compromis et la seconde concerne l'analyse des données enregistrées par le module Data Capture afin de recueillir des informations sur les techniques d'attaques, les motivations et le comportement des hackers.

2.6 Honeynet

Honeynet du Honeynet Project [7] est une catégorie de honeypot à forte interaction disposé en réseau imitant une architecture complète de systèmes de production (ex : stations windows,stations Lunix, serveurs Microsoft, ...), à laquelle on ajoute des dispositifs dédiés à la capture et au contrôle des données (routeur, pare-feu, IDS, serveur de logs, ...). Il existe deux types de honeynet en fonction de leur mise en œuvre : les **honeypots physique** et **les honeypots virtuels** [18].

6. Network Time Protocol

7. The Honeynet Project is a leading international security research organization, dedicated to investigating the latest attacks and developing open source security tools to improve Internet security

2.7 Architectures de honeynet

Du point de vue architecture, il existe différentes manières de déployer un Honeynet, mais celles présentées ici suivent la proposition du Honeynet Project, qui distingue trois générations de honeynets : GenI, GenII et GenIII.

2.7.1 Première génération :GENI

Cette génération de Honeynet remonte à la fin des années 90. Son objectif principal est d'étudier la communauté blackhat et constitue la première solution honeypot à forte interactivité, capable de capturer suffisamment d'information pour permettre d'étudier les nouvelles attaques et techniques de la communauté blackhat.

L'architecture typique de cette première génération de honeynet est décrite sur la Figure 2.2.

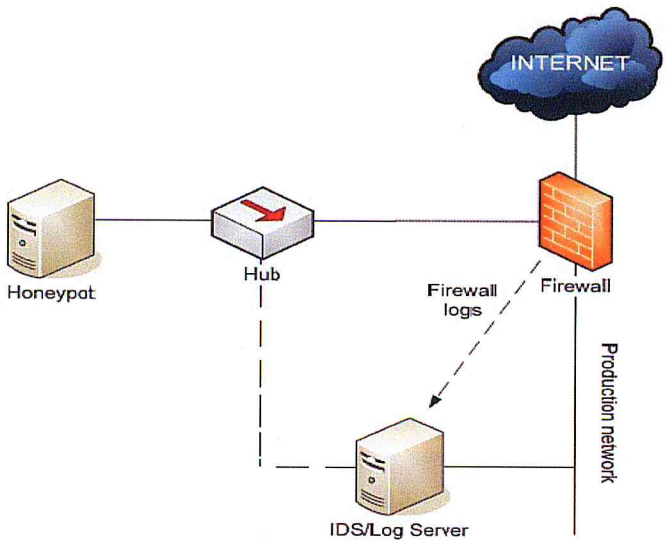

FIGURE 2.2 – Architecture de la première génération des honeynets [21].

Elle est essentiellement basée sur l'utilisation d'un réseau dédié, protégé par un pare-feu pour limiter le risque de rebond vers l'extérieur. L'utilisation d'équipements supplémentaires renforce encore le niveau de sécurité :

– un routeur pour masquer le pare-feu du point de vue Honeynet et bloquer des attaques basées sur ICMP, l'usurpation d'adresses IP ou encore les attaques de type Déni de Service (DoS [8]).

– un IDS pour réaliser l'enregistrement de l'activité réseau et générer des alertes lorsqu'il reconnaît un paquet ou plusieurs paquets comme une attaque connue.[22]

2.7.2 Deuxième génération :GENII

Pour cibler des attaquants plus confirmés, la deuxième génération de honeynets est préférable car ils sont plus difficiles à détecter et offrent des mécanismes de contrôles et de captures de données plus sophistiqués.

La Figure 2.3 représente une architecture type de honeynet de deuxième génération[23].

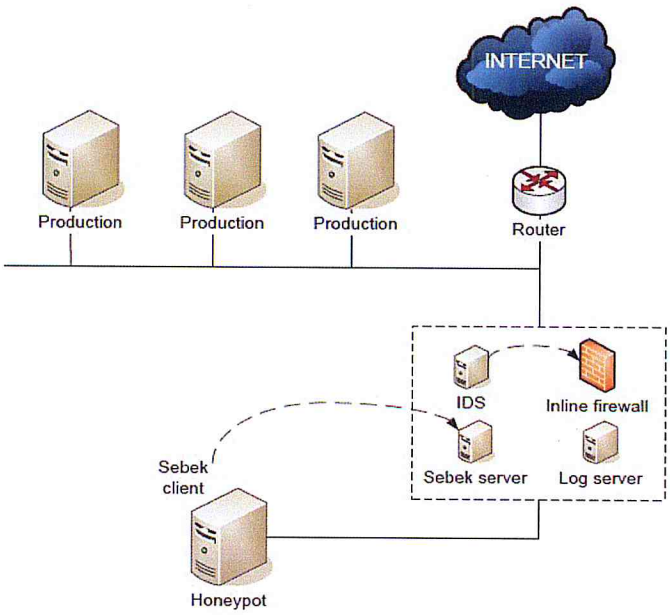

FIGURE 2.3 – Architecture de la deuxième génération des honeynets

8. Denial of Service

Elle dispose qu'un seul équipement ("Honeynet Sensor") pour effectuer les rôles dévolus au pare-feu et à l'IDS dans l'architecture de première génération. Le pare-feu fonctionne au niveau de la couche 2 du modèle OSI, c'est à dire qu'il ne possède pas d'adresse IP.

Cette architecture a été introduit en 2001 et présente plusieurs avantages. Tout d'abord elle a le mérite de simplifier considérablement le déploiement du honeynet car elle nécessite moins d'équipements et surtout elle est totalement transparente vis-à-vis de la couche réseau. En fonctionnant au niveau 2, cette solution est également plus furtive (pas de routage au niveau du pare-feu donc la valeur TTL [9] n'est pas décrémentée, ce qui permet de rendre invisible le pare-feu à des outils tels que traceroute) et plus robuste (le pare-feu n'a pas d'adresse IP à offrir à l'attaquant). Enfin, cette deuxième génération offre des outils de contrôle de données plus puissants, permettant non seulement de limiter le nombre de connexions mais également de modifier à la volée les flux de données dangereux.

2.7.3 Troisième génération :GENIII

La troisième génération a été libéré à la fin de 2004, elle est caractérisé par la mise en place d'un dispositif unique qui gère le contrôle de données et les mécanismes de capture et d'analyse de données de l'honeynet appelé le Honeywall qui est réalisé en tant que pont transparent.

Cette architecture (Figure 2.4) a la même architecture de la deuxième génération,la seule différence étant des améliorations en matière de déploiement, de gestion des Honeynets, d' administration à distance de l' interface graphique, et l' intégration d' analyse des données.

9. Time To Live

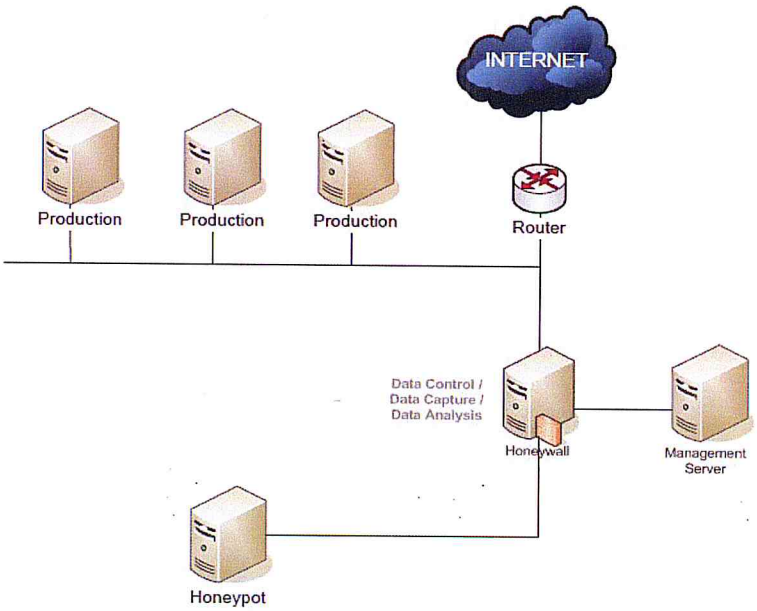

FIGURE 2.4 – Architecture de la Troisième génération des honeynets

Elle intègre 3 interfaces sur le Honeywall. Deux interfaces ont agi comme un pont entre la réseau externe et le réseau interne des honeypots, tandis que la troisième interface a été utilisé pour la gestion et les tâches de configuration.

2.8 Résumé

Dans ce chapitre, nous avons présenté le concept des honeypots et leurs principes de fonctionnement, leurs architectures, leurs importances dans le domaine de la sécurité informatique et essentiellement leurs utilités dans la détection et analyse des nouvelles techniques d'attaques.

Nous aborderons dans le chapitre suivant l'implémentation et la conception de notre solution, tout en basant sur la technologie des honeypots.

Partie Pratique

Chapitre 3

Conception et Implémentation

Dans le but d'aider les enquêteurs dans leur travail d'investigation des crimes informatique et mettre dans leurs disposition un outil d'analyse des données de sécurité et étudier les nouvelles techniques d'attaques. Nous présentons dans ce chapitre,la conception et l'implémentation de la solution proposée en spécifiant les composants matériels et logiciels nécessaire à la mise en place de telle plate forme en intégrant les fonctionnalités de journalisation et de supervision. Ensuite nous détaillons la conception du modèle dimensionnel de données basé sur le modèle en étoile et l'implémentation de l'entrepôt de données de sécurité collectés via les honeypots déployées.

3.1 Conception

La première étape de cette thèse consiste à la réalisation d'une plate-forme d'investigation des crimes informatiques à base des honeypots. L'objectif est de détecter, étudier et mesurer les différents modes opératoires des attaques informatiques afin de pouvoir les analyser et les utiliser dans le cadre de recherche.

Dans notre projet, nous avons choisi le honeynet du Honeynet project, qui est un honeypot a forte interaction et outil flexible qui peut être modifié et construit selon notre proposition et nos besoins.

Un Honeynet est un réseau qui contient un ou plusieurs honeypots dont la fonction est d'enregistrer les flux de données et les intrusions dans le réseau à des

fins de recherche. Il est utilisé pour collecter des informations sur le trafic réseau , et capturer les nouveaux outils et tactiques d'attaques.Il aide aussi a trouver les attaque de type zero day où les pirates tentent d'exploiter les vulnérabilités des systèmes avant d'être connu par les développeurs des logiciel de sécurité et les administrateurs des réseaux.

La conception de notre solution est basée sur l'architecture honeynet de troisième génération [1] qui est caractérisé par la mise en place d'un système Honeywall qui assure les trois fonctions de honeynet a savoir le contrôle, la capture et l'analyse des données. nous avons opté pour la solution des honeynet virtuel qui nous permet d'exécuter un Honeynet complet avec plusieurs systèmes d'exploitation sur le même dispositif physique. Cette solution a l'avantage d'être facile à déployer et simple à gérer mais elle nécessite un matériel plus puissant.

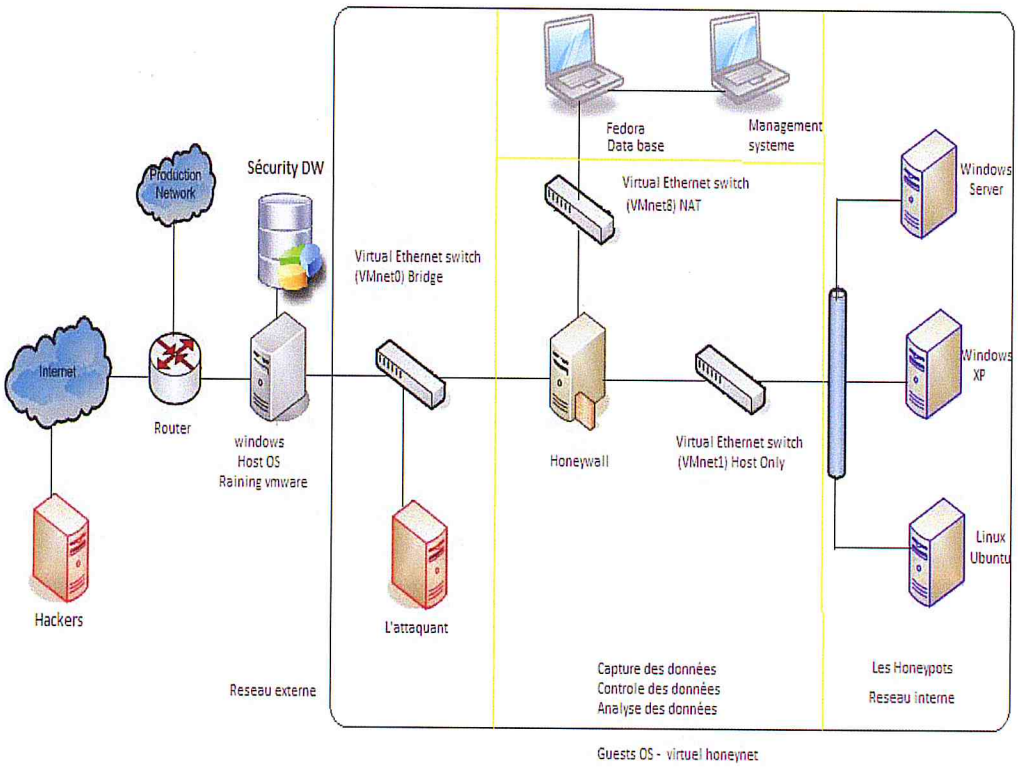

FIGURE 3.1 – Conception de la solution à base de honeynet virtuel

1. voir chapitre 2

Cette conception Elle comporte cinq modules :

- Les honeypots (réseau interne).
- Le système Honeywall.
- Le système de management.
- Le réseau externe.
- L'entrepôt de données de sécurité.

3.1.1 Les honeypots (réseau interne)

Est un réseau interne qui émule les systèmes et les services d'un réseau de production. Nous avons choisi un serveur windows 2003 qui offert les services réseaux suivant : SMTP ,FTP, Mail ,HTTP, Base de données en raison de leur forte utilisation en production afin d'attirer les attaquants. Deux autres postes clients différents : un honeypot avec système d'exploitation windows XP et l'autre avec un système Linux (Ubuntu).

3.1.2 Le système honeywall

Dans cette conception le système Honeywall joue le rôle d'un passerelle entre le réseau interne (les honeypots) et le réseau externe et qui enregistre tout le trafic entre les deux réseaux. Il est indétectable par les attaquants parce qu'il intègre deux interfaces (eth0, eth1) fonctionnant au niveau 2 de modèle OSI (pas d'adresse IP).

3.1.3 Le système de management

Il assure la gestion des Honeypots, l'administration à distance de honeywall et la gestion des données capturés.

3.1.4 Le réseau externe

IL représente les attaquants externes (blackhat hackers) qui tentent d'attaquer notre plateforme et aussi il comporte une machine virtuelle tourne sous

le système Kali Lunix pour tester notre configuration Honeynet et réaliser des attaques internes.

3.1.5 L'entrepôt de données de sécurité

Le datawarhouse de sécurité est un outil d'analyse de données collectées via les honeypots dans le but de faciliter l'analyse et la prise de décision des administrateurs de sécurité et de réseau. [2]

3.2 Implémentation

3.2.1 Description de la plateforme

Nous avons défini précédemment la conception et l'architecture globale de la plate-forme. L'implémentation faite est représentée sur le schéma ci-dessous. La plate-forme est exécutée sur une machine windows8 faisant tourner VMWare Workstation. Sur ces machines VMWare, il est possible de faire fonctionner des vrais systèmes tel que Windows XP, Linux (Ubuntu) et le système Honeywall.

La mise en place de la plate-forme nécessite plusieurs outils distincts à savoir :

1. VMware Workstation 11.0.0 : VMware Workstation est un logiciel de virtualisation qui permet le fonctionnement de plusieurs systèmes d'exploitation en même temps sur les architectures Intel x86. Il a été développé par "VMware Incorporation".
2. Honeywall Roo version 1.4 : c'est un système open source basé sur linux CentOS, contient tous les outils et les fonctionnalités nécessaires pour créer, configurer et maintenir un honeynet de troisième génération.
3. Honeypots : Windows server 2003, Windows xp sp2, Linux ubuntu
4. Système de management : fedora 11.04, windows7.
5. kali lunix : Kali Linux est une distribution Linux Open source basée sur Debian visant à les tests de pénétration avancée et audit de sécurité. Kali c'est

2. plus de détail dans la section 3.3

le successeur de BackTrack qui contient des centaines d'outils destinés à diverses tâches de sécurité de l'information, tels que les tests de pénétration, Forensics et Reverse Engineering.

La fonction des honeypots a forte interactivité est réalisée via le système Honeywall qui fait l'enregistrement de tous le trafic entre les honeypots installées et le réseau externe. Dans cette architecture le honeywall a trois interfaces réseau : la première eth0 est en vmnet0 et se connecte au périphérique Ethernet virtuelle du vmware. Le deuxième adaptateur réseau eth1 est utilisé pour connecter le système honeywall aux honeypots. La troisième interface réseau eth2 est utilisé pour connecter au système de management comme montré dans la figure 3.2.

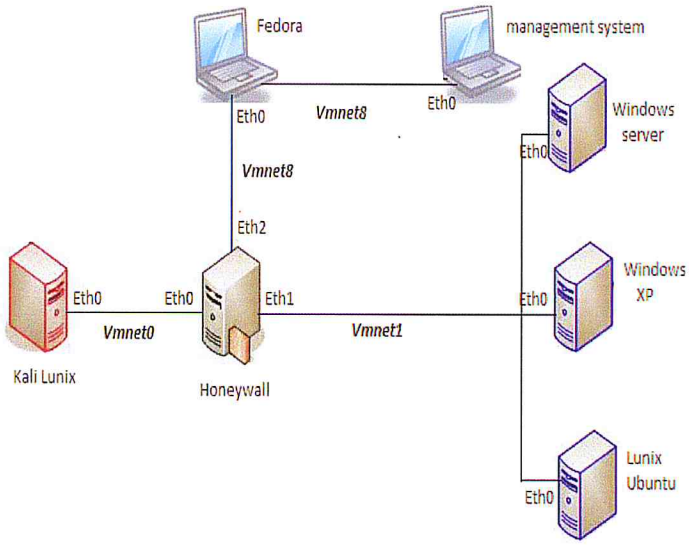

FIGURE 3.2 – Architecture de honeynet virtuel

3.2.2 Installation de la plateforme

La section suivante décrive Les étapes de l'installation et la configuration des différents systèmes pour créer un honeynet virtuel vmware. Ces étapes sont classée comme suit :

- l'installation du vmware 11.0.0 sur la machine physique.
- l'installation du système honeywall roo 1.4.

– l'installation des honeypots :windows server, Windows XP et ubuntu.

– l'installation du système d'exploitation fedora qui contient la base de donnée du honeywall.

– l'installation du système de management.

– l'installation du système Kali Lunix pour attaquer les Honeypots.

– l'attribution des adresses ip aux différents systèmes installés.

– l'ajout de trois interfaces réseau dans le système honeywall. la configuration du réseau vmware selon l'architecture, comme indiqué dans la figure 3.2, et affecter une adresse ip à l'interface eth2, alors que eth0 et eth1 n'ont pas des adresses ip .

3.2.3 Installation et Configuration du Honeywall

L'installation et la configuration du honeywall est l'étape la plus importante dans la mise en place d'un honeynet car le Honeywall est un dispositif de passerelle qui sépare les honeypots et les réseaux externes. Tout trafic à destination ou à partir des honeypots doit passer par le honeywall. Cette passerelle est traditionnellement un dispositif de pontage de couche 2, ce qui signifie le dispositif devrait être invisible de toute personne interagissant avec les honeypots.

– L'installation est faite sur la version du honeywall cdrom roo 1.4 (figure 3.3).

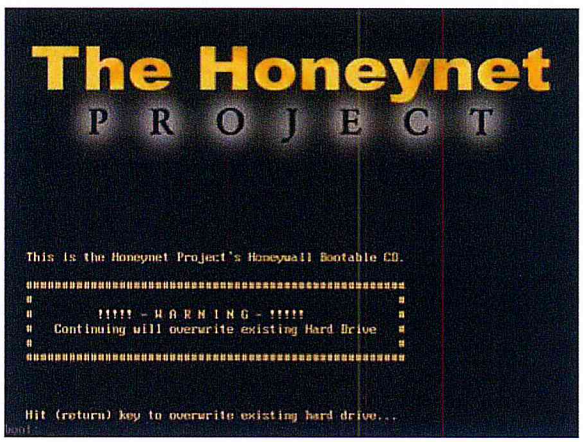

FIGURE 3.3 – Interface d'installation de honeywall Roo 1.4

– La configuration du Honeywall est faite via un menu de dialogue(figure 3.4)ou via l'interface web Walleye(figure 3.5).

* Menu de Dialogue :

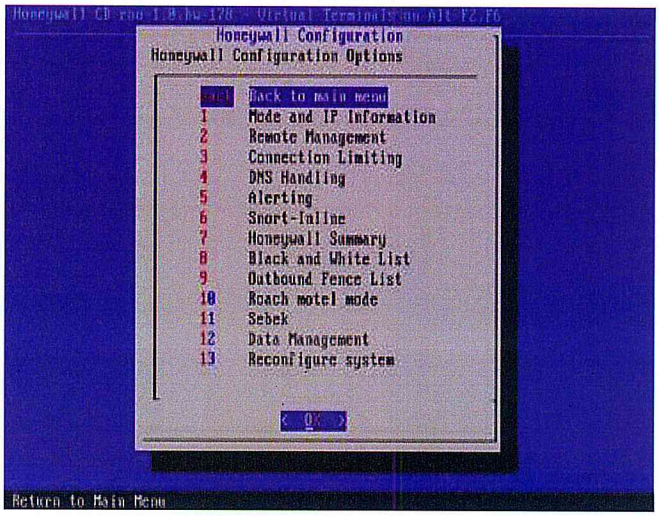

FIGURE 3.4 – Menu de configuration

* Interface Walleye :

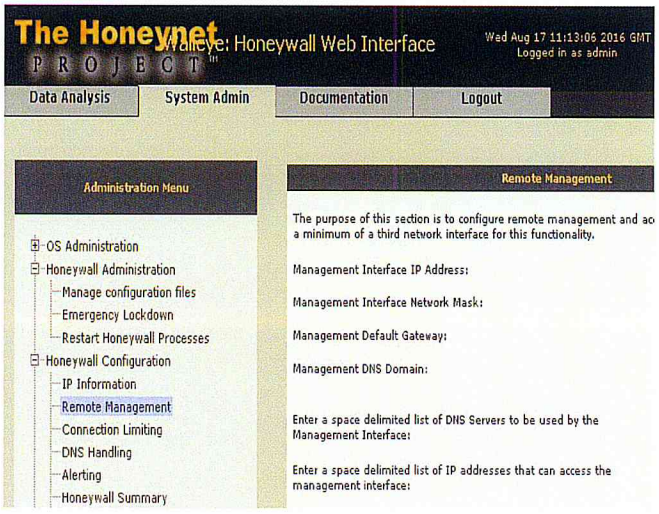

FIGURE 3.5 – Interface de configuration Walleye

3.2.4 Architecture de Honeywall

Honeywall collecte les données, analyse et enregistre toutes les communications réseaux. A fin de mieux les comprendre, nous présentons dans cette section son architecture et les différentes éléments qui la compose, comme montré la figure 3.6 :

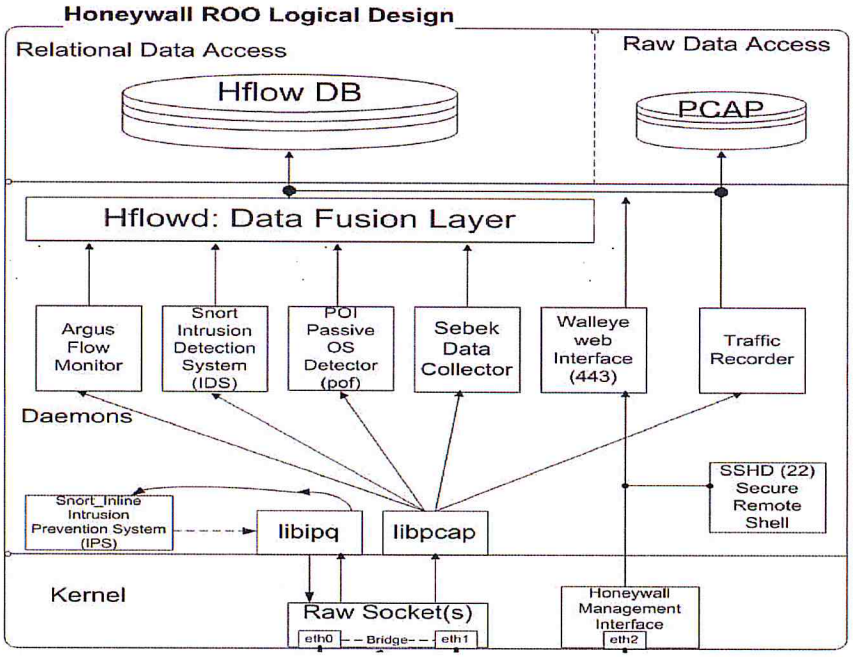

FIGURE 3.6 – Architecture de système honeywall [24]

- Librairie de capture de trames réseaux (libpcap)
- Un logiciel de détection/prévention d'intrusions réseaux (Snort/SnortInline)
- Un outil de capture et analyse des flux applicatifs (Sebek)
- Un outil d'analyse de netflow (Argus)
- Un outil de détection de prise d'empreinte passif (p0f)
- Un outil de fusion des données (Hflow)
- un outil de détection des entêtes des paquets (Tcpdump)
- un interface web pour l'administration du honeywall (walleye) comme montré dans les figures suivantes :

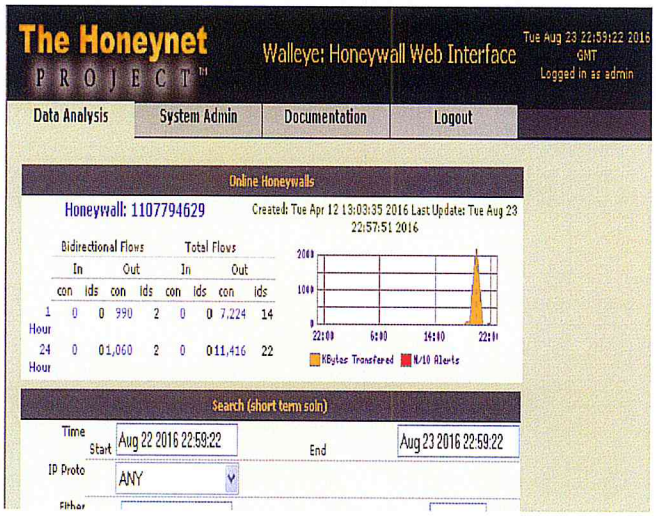

FIGURE 3.7 – Interface principal de walleye

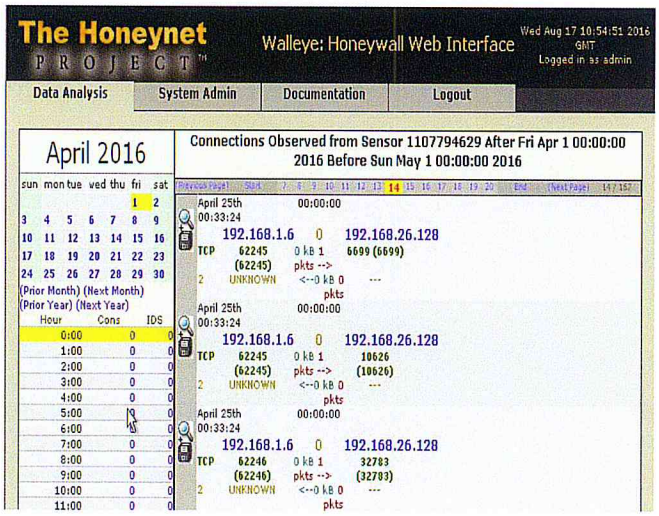

FIGURE 3.8 – Les attaques enregistrées

3.3 L'entrepôt de données de sécurité (Module d'analyse de données)

3.3.1 Contexte

L'entrepôt de données[3] de sécurité est un outil d'analyse de données a pour objectif de centraliser et faire converger les grandes quantités des données de sécurité collectées via les honeypots dans le but de faciliter les taches des administrateurs de sécurité et de réseau en focalisant sur l'accès à l'information pertinente, l'analyse et la prise de décision.

Ce module consiste à mettre en place une solution décisionnelle complète qui traite, analyse et mesure les attaques commises contre les honeypots déployés afin de permettre d'effectuer des statistiques, des analyses efficaces et la rédaction des rapports d'actualité.

3.3.2 Modélisation et Conception du Data Warehouse

Pour la conception de notre entrepôt de données de sécurité , nous avons eu recours à la modélisation dimensionnelle qui est souvent associée aux entrepôts de données compte tenu de ses avantages dont l'un des avantage est de présenter les données sous forme standardisé intuitive et qui permet les accès performants.

Elle consiste en deux nouveaux concepts tels que les faits et les dimensions. Chaque modèle multidimensionnel est composé d'une table contenant une clé, la table des faits qui permettent de mesurer l'activité et d'un ensemble de tables dimensionnelles (axes d'analyses) qui contiennent les informations contextuelles faisant varier les mesures de l'activité en question. Chaque table de faits possède une clé qui la relie avec la clé primaire de chaque table de dimension.

L'interrogation des tables de faits à travers les tables de dimensions produit des rapports agrégés qui sont théoriquement capables de répondre à l'ensemble des besoins en information des administrateurs.

3. voir l'annexe A : le datawarhouse

Pour réaliser notre modèle dimensionnel et construire notre entrepôt de données a partir des données (alertes et événements) enregistrées par le honeywall qui a été décrit dans la section précédente, nous avons suivi les quatre étapes suivantes :

Étape 1 : Choix de l'activité à modéliser

L'étape de choix de l'activité à modéliser est fondamentale et doit se faire en fonction de son importance au sein de l'organisation. Dans notre cas, La détection, l'enregistrement et l'analyse d'attaque , demeure comme l'activité principale des administrateurs de sécurité pour assurer l'efficacité de la politique de sécurité.

La grande quantité des données de sécurité capturées par les honeypots déployés, présente une source idéale pour attirer des informations essentielle afin de rendre l'analyse de sécurité plus performante, de comprendre la situation sécuritaire et les différents nouvelles attaques. Ainsi la disponibilité de ces informations s'avère indispensable aux administrateurs pour générer des rapports de sécurité d'actualité.

Définition de l'activité « Attaque »

– L'attaque : faute d'intéraction malveillante, à travers laquelle un attaquant cherche à délibérement violer une ou plusieurs propriétés de sécurité. Il s'agit d'une tentative d'intrusion .

– Chaque attaque enregistrée par le système honeywall dans notre plateforme comme indiqué dans la figure 3.4 comprend :

* l'adresse IP source et port source qui représentent «Qui» (attaquant) ;

* l'adresse IP de destination et le port de destination qui représentent "à qui" (cible).

* les actions et le protocole utilisé dans l'attaque qui représente "Quelle / quoi".

* Le temps est inclus et unique est représente "Quand".

Étape 2 : Granularité des données

Le choix de degré de granularité le plus fin donne une capacité d'analyse élevée en terme de croisement entre les différents axes d'analyses. Dans le cas des

attaques le degré de granularité le plus fin , ou le niveau de détail de l'information le plus bas , correspond à une opération d'enregistrement d'une attaque, d'où une ligne de table de fait correspondant à :

Suivi de nombre et de type d'attaque par, ip machine source, ip machine destination, port source, port destination, protocole utilisé dans un intervalle de temps donnée.

Étape 3 : Les dimensions

Le but de cette étape est de choisir quels sont les dimensions (axes d'analyse) adéquates pour le processus en question. Elles ont pour objectif de décrire le fait, donc on essaye de recenser toutes les informations qui décrivent une attaque et qui peuvent intéresser les administrateurs de sécurité.

1. **Dimension Temps**

   La dimension temps est « la seule dimension qui figure systématiquement dans tout entrepôt de données, car en pratique tout entrepôt de données est une série temporelle. Le temps est le plus souvent la première dimension dans le classement sous jacent de la base de données » [28].

   La dimension temps se présente comme suit :

Table Dim temps	
PK	temps attaque
	Minute
	Heure
	jour
	mois
	annee

TABLE 3.1 – Table Dimension Temps.

Dans cette dimension, il est utilisé une clé artificielle comme clé primaire, cette clé sert à faciliter la manipulation de la dimension.

2. **Dimension Machine Source**

La dimension Machine Source représente l'attaquant et définit la source de l'attaque, elle contient les informations liés a la machine qui est initié l'attaque comme l'adresse IP de la machine source, le port utilisé et la classe de la machine.

La dimension Machine Source se présente comme suit :

Table Dim Machine Source	
PK	ip source
PK	port source
	description machine

TABLE 3.2 – Table Dimension Machine Source.

Dans cette dimension, il est utilisé une clé primaire composée de deux attribut (ip source, port source).

3. **Dimension Machine Destination**

La dimension Machine Destination représente la machine cible, elle définit la cible de l'attaque, elle contient les informations liés a la machine destination comme l'adresse IP de la machine destination , le port visé et la classe de la machine.

La dimension Machine Destination se présente comme suit :

Table Dim Machine Destination	
PK	ip destination
PK	port destination
	description machine

TABLE 3.3 – Table Dimension Machine Destination.

Dans cette dimension, il est utilisé une clé primaire composée de deux attribut (ip Destination, port Destination).

4. **Dimension Protocole**

La dimension protocole représente le nom du protocole attaqué (visé) et la classe de protocole et sa description, elle se présente comme suit :

Table Dim Protocole	
PK	id-protocole
	protocole
	classe

TABLE 3.4 – Table Dimension Protocole.

5. **Dimension Action**

La dimension action représente les informations sur les codes d'actions et leurs description et classement, elle se présente comme suit :

Table Dim Action	
PK	id-action
	action
	description

TABLE 3.5 – Table Dimension Action.

Étape 4 : Quelles mesures (fait)

Dans cette étape, Pour identifier le fait, il faut répondre à la question : Qu'est ce qu'on mesure ?

Les mesurables qui correspondent à l'activité d'enregistrements des attaques et qui permettent de mesurer les performances de cette activité, sont :

- le nombre d'attaques,
- les types d'attaques,
- la durée d'attaque.

La table Fait attaque se présente comme suit :

Table Fait Attaque	
FK	ip source
FK	port source
FK	ip destination
FK	port destination
FK	id-protocole
FK	id-action
	nombre attaque
	type attaque
	duree attaque

TABLE 3.6 – Table Fait Attaque.

le Modèle en étoile de l'entrepôt de données de sécurité Après la réalisation des quatre étapes citées avant, nous avons obtenu notre modèle en étoile de l'entrepôt de donné de sécurité suivant :

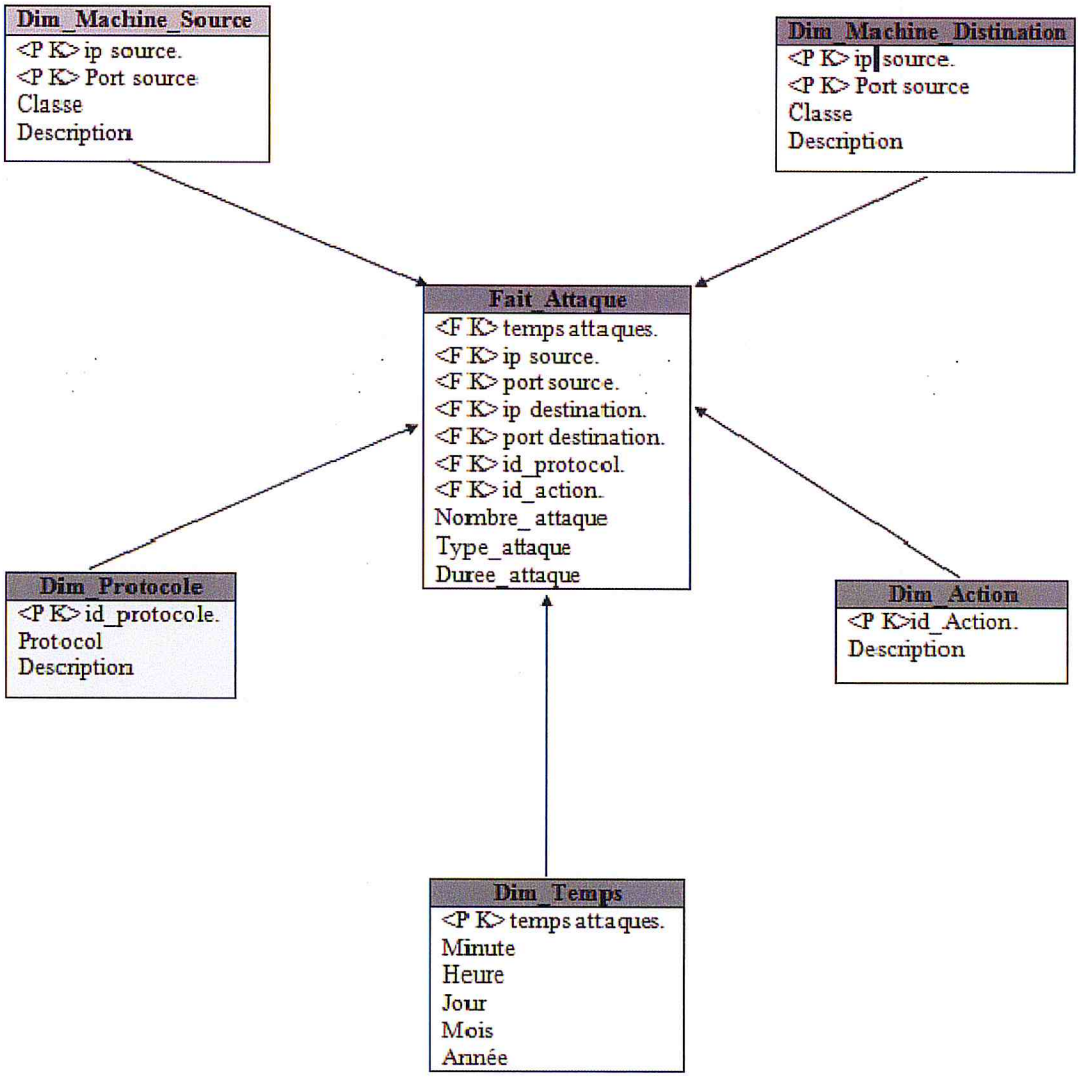

FIGURE 3.9 – Le Modèle en étoile de l'entrepôt de données de sécurité

3.3.3 Alimentation du Data Warehouse

Une fois le Data Warehouse conçu, nous passerons à l'alimentation et le chargement des données collectées par le système honeywall. Cette phase alimentation se déroule en plusieurs étapes : extraction, transformation, chargement et rafraîchissement des données. Ces étapes sont prises en charge par le processus ETL (Extracting, Transforming and Loading) qui constitue la phase de migration de données collectées via les honeypots dans l'entrepôt de données après qu'elles ont subi des opérations de sélection, de nettoyage et de reformatage dans le but de les homogénéiser.

Afin de réalisé cette phase nous avons utilisé l'outil ETL de SQL Server Intégration Services comme montré dans la figure suivante :

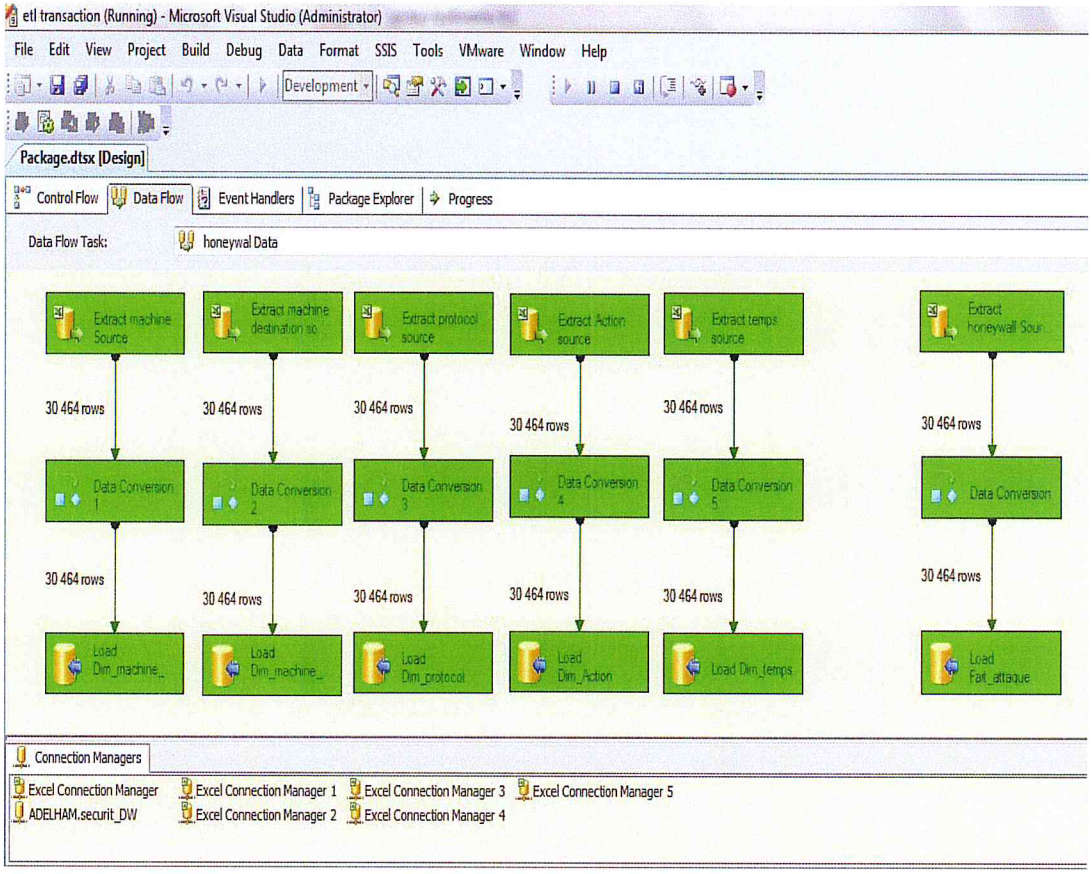

FIGURE 3.10 – Processus ETL

Ensuite après l'alimentation de l'entrepôt de données via le processus ETL précédent, nous avons obtenu le view de données sources suivant :

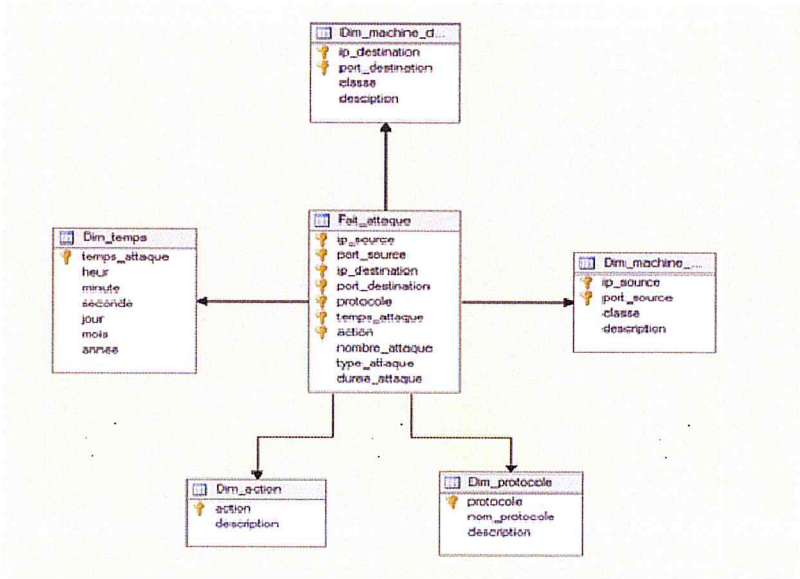

FIGURE 3.11 – View de données sources

3.3.4 Résumé

Dans ce chapitre, nous avons présenté la démarche de mis en place de notre solution en s'appuyant sur les technologies des honeypots et de datawarehousing.

Dans le chapitre suivant nous passerons à l'expérimentation de notre plate forme par la réalisation des études de cas des attaques et d'analyses des données.

Chapitre 4

Expérimentation et Analyse de données

Dans ce chapitre, nous évaluons notre solution à base des honeypots des attaques informatiques et d'analyse de données de sécurité mis en place. A cet égard, nous commençons par présenter une étude de cas d'une attaque d'ingénierie sociale et de phishing (hameçonnage). Ensuite, nous décrivons la série d'expérimentations exécutés sur la plateforme afin de réaliser des attaques et analyser le trafic enregistré par le système Honeywall. Une analyse forensic des traces d'attaques a également été effectuée. Ensuite, Nous présentons l'analyse de données des data-warehouse de sécurité qui est déjà conçu et alimenté par les données collectés via les honeypots déployés pendant la période d'expérimentation.

4.1 Expérimentation

Pour expérimenter notre plateforme d'investigation des crimes informatiques, des attaques réelles doivent être mises en œuvre et des expériences doivent être élaborées pour évaluer et analyser les traces des attaques commises et à identifier et distinguer entre plusieurs types d'attaques.

4.1.1 Étude de cas

Vu le grand nombre des techniques des attaques existes, nous avons choisi l'attaque de phishing et de social ingénierie qui est utilisé fréquemment pour l'escroquerie sur internet.

- Social engineering (Ingénierie sociale) est une méthode de manipulation des personnes afin de contourner des dispositifs de sécurité. Il s'agit ainsi d'une technique consistant à obtenir des informations sur les potentiels victimes par téléphone, courrier électronique, courrier traditionnel ou contact direct. L'ingénierie sociale repose sur les points faibles des personnes qui sont en relation avec un système informatique. Le but est de piéger les gens en leur faisant révéler leur mot de passe ou toute autre information qui pourrait compromettre la sécurité du système informatique.

- Le phishing (hameçonnage), est une technique frauduleuse d'ingénierie sociale, utilisée par les pirates informatiques pour récupérer des informations (généralement bancaires) auprès d'internautes. Grâce à ces informations, les pirates sont capables de transférer directement l'argent sur un autre compte ou bien d'obtenir les données nécessaires pour pirater des comptes sociaux [3].

Durant la période d' expérimentation de notre palteforme , nous avons exécuté et analysé plusieurs techniques d'attaque,mais nous détaillons seulement un aperçu sur l'attaque de phishing et de social ingeneering choisi pour hacker un compte de réseaux sociaux :

Phase d'attaque : l'attaquant exécute des commandes de social engineering toolskit comme montré les figures suivants [1]

1. Une démonstration détaillée sera exposée

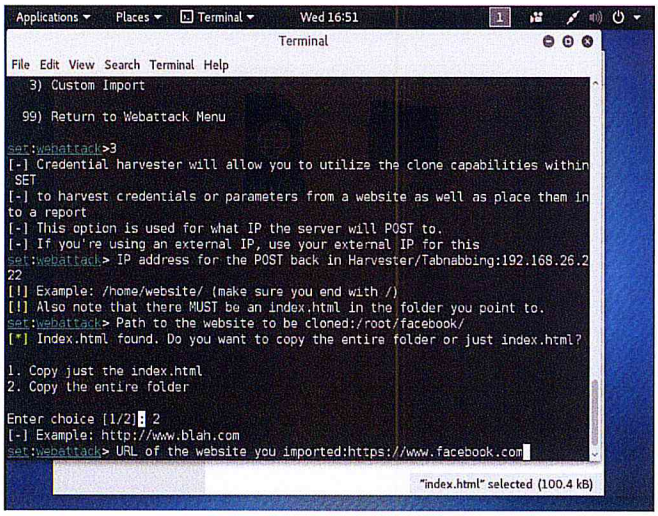

FIGURE 4.1 – Exécution des commandes d'attaque

FIGURE 4.2 – La page de site altérée.

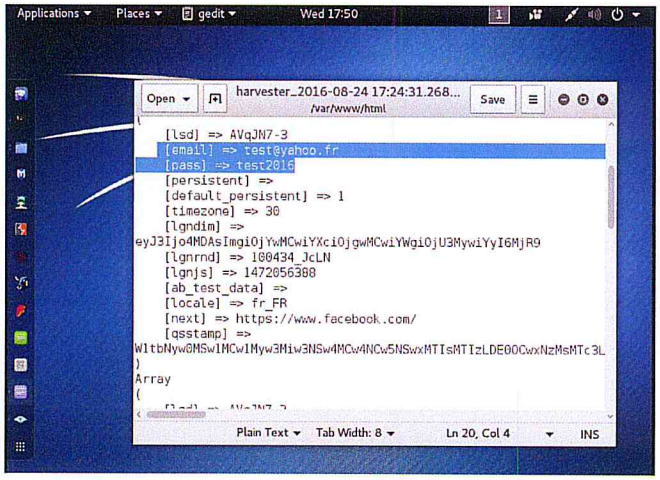

FIGURE 4.3 – Récupération de compte et mot de passe du victime

Phase d'analyse : L'analyse de fichier pcap enrigistré par le système honeywall, nous avons permis d'identifier les traces de l'attaque (figure 4.4) :

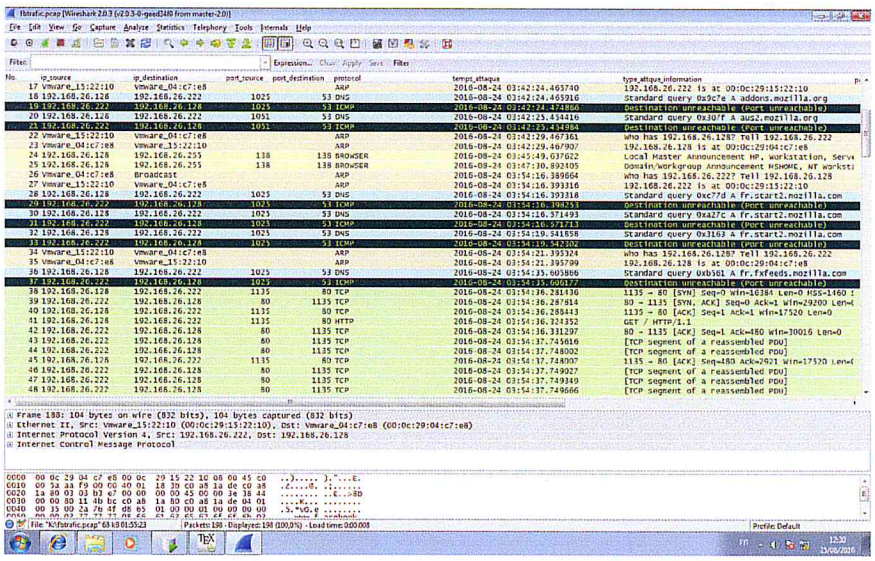

FIGURE 4.4 – Analyse de trafic capturé par le Honeywall

4.2 Analyse de données et présentation

L'analyse de données ou la phase d'exploitation de l'entrepôt de données de sécurité se fait par le biais d'un ensemble d'outils analytiques développés autour du Data Warehouse. Pour cela nous avons utilisé les logiciels : Microsoft Visuel Studio et SQL Server Business Intelligence Development Studio.

4.2.1 Analyse dimensionnelle des données

L'analyse dimensionnelle exploite et fait ressortir au mieux les capacités de l'entrepôt des données, dans Le but d'offrir la possibilité d'analyser les données selon différents critères afin de calculer les indicateurs de performances. Cette analyse se fait selon le principe OLAP, qui offrent les possibilités de recourir à différentes opérations facilitant la navigation dans les données.

La mise en place de ces outils est une option très intéressante dans la mesure où les données seront accessibles en analyses instantanées. Plusieurs fournisseurs de solution OLAP existent sur le marché et offrent des solutions construites sur des méthodes et technologies différentes. Pour cela nous avons utilisé le logiciel Microsoft Visuel Studio 2008.

Création de Cube

Dans cette partie nous expliquons la démarche de la création du cube. En effet nous avons utilisé comme outil MS Analysis Services.

La première étape de la création du cube consiste à définir la source de données, ensuite sélectionner les tables des dimensions et la table de faits.

La deuxième étape est la création du cube à l'aide de l'assistant. Cette étape nous permettons de vérifier et éventuellement modifier les dimensions ainsi que les mesures attribués.

Une fois le cube est créé, il nous reste l'étape de déploiement de cube, pour ce faire l'outil Analysis Services traite le cube construit et vérifier s'il n'y pas d'erreur lors de la création, ensuite il le déploie sur le serveur.

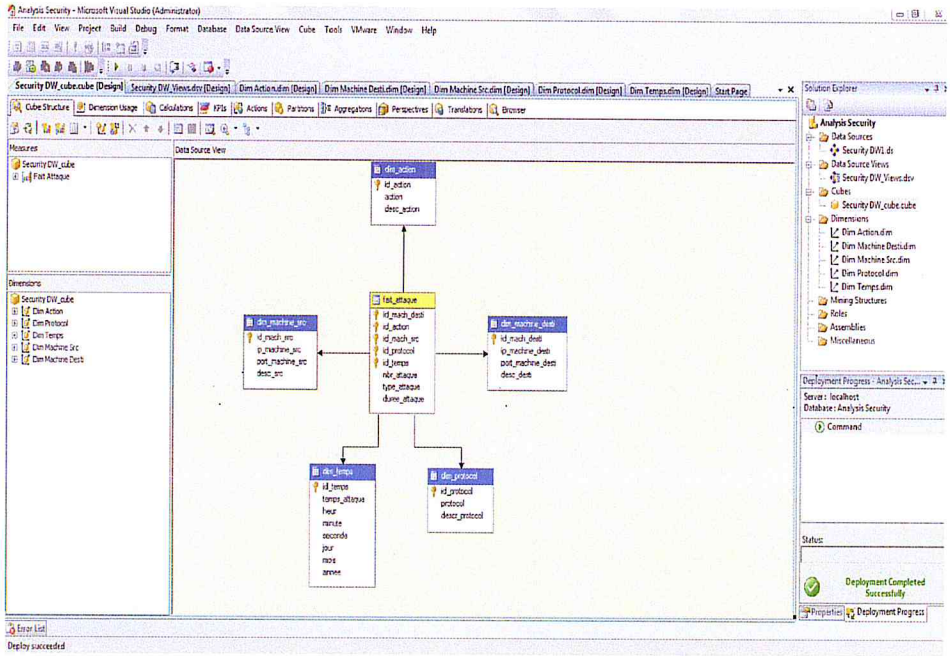

FIGURE 4.5 – Déploiement de cube créé

Par conséquence il est possible d'accéder au Reporting Services pour traiter tout ce qui concerne le reporting.

4.2.2 Présentation des données (Reporting)

Dans cette section, nous précédons à la création des rapports par le biais de SQL Server Reporting Services. Les requêtes sont constituées lors de l'élaboration des rapports qui seront ensuite diffusés périodiquement ou à la demande, tout en basant sur l'agrégation des données contenant dans la table des faits détaillées. Nous choisissons les agrégats suivants qui nous semblent les plus pertinents et susceptibles de faire l'objet d'accès fréquents :

• Nombre d'attaques par protocole par jour :

On observe une répartition égale entre les protocoles ICMP et TCP. Ceci laisse penser que, dans la plupart des cas, une attaque est toujours précédée par une phase de reconnaissance utilisant le protocole ICMP.

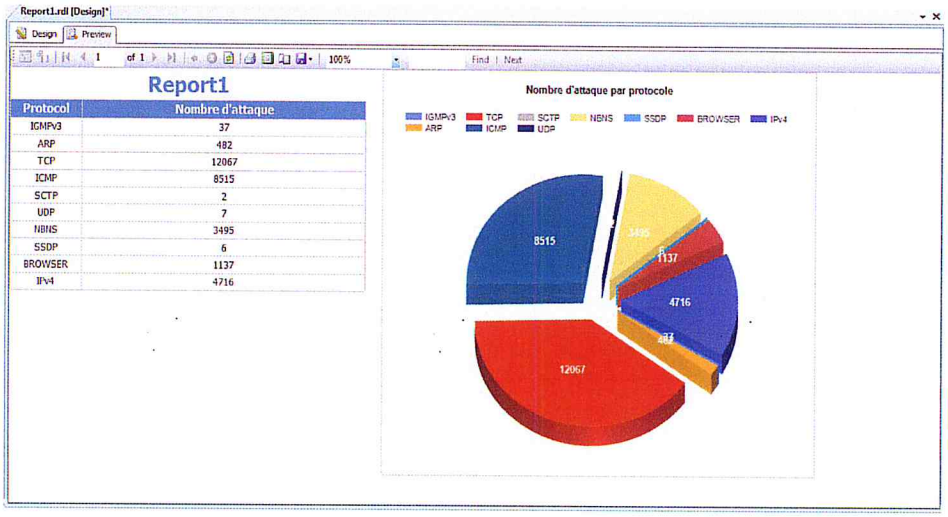

FIGURE 4.6 – Nombre d'attaque par protocole par jour

• Nombre d'attaques par adresse ip source vers la même adresse ip destination a eu lieu le même protocole en même date.

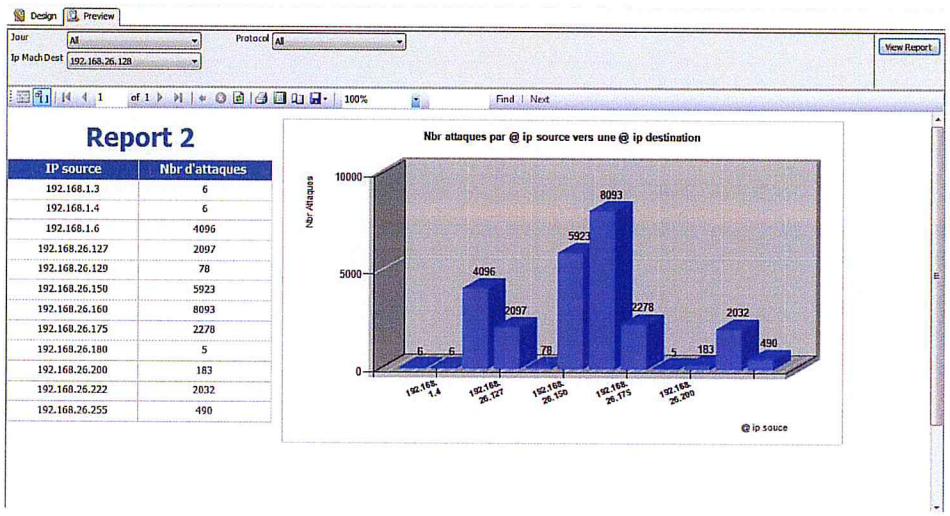

FIGURE 4.7 – Nombre d'attaques par @ ip source vers la même @ ip destination

• Nombre d'attaques par jour selon les critères de sélection : @ ip source, @ ip destination et protocole.

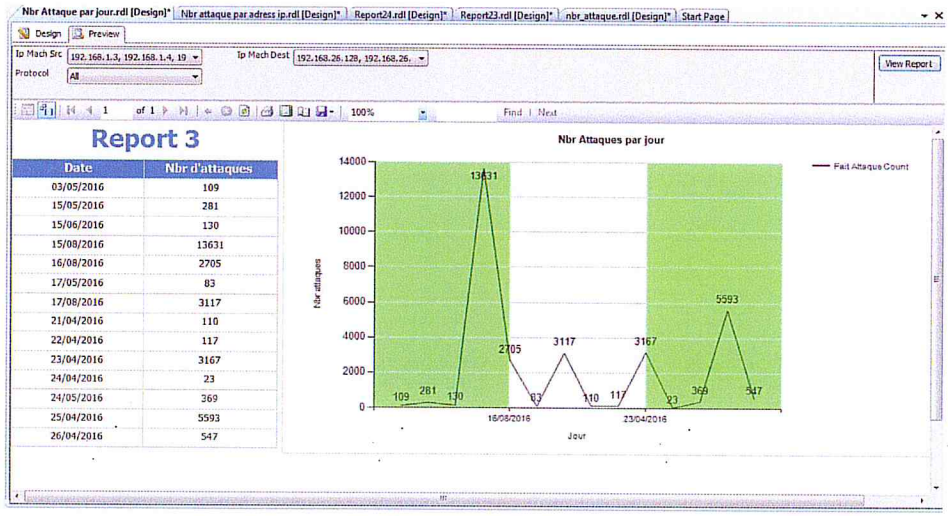

FIGURE 4.8 – Nombre d'attaque par jour selon des critères de sélection

4.3 Résumé

Dans ce chapitre, nous avons expérimenté notre plateforme des attaques informatiques en utilisant les outils open source pour la réalisation des attaques réels et les analysés ensuite. Les expérimentations ont montré une efficacité du systéme Honeywall en matière d'enregistrements d'attaques, mais ils reste difficile d'analyser une grande quantité des données enregistrées, pour cette raison nous avons implémenter l'entrepôt de données de sécurité afin de faciliter la tache d'analyse.

Conclusion

Avec la croissance continue des crimes informatiques, des mécanismes de lutte, de contre mesure et de gestion de sécurité deviennent de plus en plus une nécessité. Les honeypots sont un des moyens idéaux pour étudier le comportement, les tactiques et les motivations des pirates informatiques et ainsi se prémunir efficacement des attaques de ces derniers. Ils représentent une technologie très prometteuse dans la lutte contre la cybercriminalité.

Dans le même sillage, notre objectif de ce mémoire est de concevoir et d'implémenter une solution à base des honeypots pour l'investigation des crimes informatiques qui réunit les différentes fonctionnalités nécessaire à l'expérimentation, à l'étude de différents modes opératoires des attaques et aussi un module d'analyse de données de sécurité collectées via les honeypots, afin de mesurer, détecter, analyser et prévenir les activités illicites, dont l'utilisation permet de leurrer et tromper les attaquants pour mieux les appréhender. Pour cela il convient d'analyser et étudier leurs nouvelles techniques afin d'anticiper au mieux les prochaines attaques, d'élaborer des nouveaux mécanismes de protection adaptés et d'améliorer les techniques d'investigations des crimes informatiques.

Dans ce mémoire, nous avons voulu donner un aperçu sur les différentes phases de hacking, l'évolution de la criminalité informatique et les mécanismes de sécurité.Puis nous avons donné un résumé de fonctionnement et des objectifs des honeypots à travers quelques exemples pour dresser un rapide état de l'art. Nous nous sommes ensuite focalisés sur l'élaboration du notre solution à base des honeypots forte interaction qui sera utilisé durant notre stage, en spécifiant la conception et l'implémentation de l'architecture de la plateforme ainsi que la modélisation dimensionnelle et l'alimentation de l'entrepôt des données de

sécurité afin de faciliter l'analyse du grande masse de données capturés par le système Honeywall.

L'Expérimentation de notre solution, nous a permis de connaitre l'activité et le fonctionnement des honeypots, ainsi que la réalisation des attaques réels contre les honeypots installés comme social engineering et le phishing qui sont utilisés généralement pour commettre le crime d'escroquerie sur internet. Ensuite l'investigation et l'analyse des traces d'attaques capturées par le honeypots à l'aide des outils disponibles.

L'expérimentation a montré l'efficacité du système Honeywall en matière d'enregistrements d'attaques, par contre une difficulté de la tâche d'analyse des grandes quantités de données capturées a été enregistrée. A cet égard, nous avons implémenté un entrepôt de données de sécurité qui nous a permet d'attirer les informations pertinentes, de faciliter l'analyse des indices d'attaques et de réaliser des rapports de sécurité attendus.

Par ailleurs, Plusieurs directions sont intéressantes et importantes pour le futur :

- Déployer les honeypots (distributed honeypots) sur plusieurs sites pour détecter les différentes attaques sur des emplacements différents afin d'analyser et comparer les techniques d'attaques et les résultats obtenus de différents sites.

- Étudier des attaques comme Botnet, Sql injection et Deni de servie distribuée (DDOS) à l'aide des honeypots.

- Améliorer les mécanismes de capture de données dans les honeypots avec l' utilisation du système sebek qui capture le trafic crypté.

Annexe A

Le Data Warehouse

A.1 Qu'est ce qu'un Data Warehouse

Bill Inmon définit le Data Warehouse, dans son livre considéré comme étant la référence dans le domaine "Building the Data Warehouse" [26] comme suit : « **Le Data Warehouse est une collection de données orientées sujet, intégrées, non volatiles et évolutives dans le temps, organisées pour le support d'un processus d'aide à la décision.** »

Les paragraphes suivants illustrent les caractéristiques citées dans la définition d'Inmon.

- Orientation sujet : le Data Warehouse est organisé autour des sujets majeurs de l'entreprise, destinées à un processus analytique.

- Intégration : l'intégration des données désigne un processus qui rassemble différentes sources de données pour les concentrer dans un seul espace. Ce principe permet aux clients du data warhouse, les applications OLAP de bénéficier d'une source de données qui soit homogène, fiable et unique. Evitant ainsi les conflits entre des données issues de différentes sources.

- Historisation et atomicité Une donnée rentrant dans le data warehouse doit toujours correspond à une date et par conséquent à l'échelle temporelle. De plus, elle doit toujours être disponible dans la granularité la plus fine qu'il soit possible d'avoir pour pouvoir répondre à des requêtes inattendues des utilisateurs, soit sous sa forme atomique.

- Non volatilité : La non volatilité est un principe qui stipule que les données du système doivent être stables, non modifiables et en lecture seule. Ce qui a pour conséquence qu'une fois une information rentrée dans le data warehouse , elle ne change pas de teneur, ni d'emplacement et que les utilisateurs du système ne peuvent pas le modifier. Ce principe permet non seulement de garantir l'archivage et ainsi de servir de mémoire de l'organisation , mais il garantit l'exhaustivité des données et leur disponibilité.

- Organisées pour le support d'un processus d'aide à la décision : Les données du Data Warehouse sont organisées de manière à permettre l'exécution des processus d'aide à la décision (Reporting, Data Mining).

A.2 Historique des Data Warehouse

L'origine du concept « *Data Warehouse* » D.W (entrepôt de données en français) remonte aux années 80, durant lesquelles un intérêt croissant au système décisionnel a vu le jour, dû essentiellement à l'émergence des SGBD relationnel et la simplicité du modèle relationnel et la puissance offerte par le langage SQL. Au début, le Data Warehouse n'était rien d'autre qu'une copie des données du système opérationnel prise de façon périodique, dédiée à un environnement de support à la prise de décision. Ainsi, les données étaient extraites du système opérationnel, stockées dans une nouvelle base de données «concept d'infocentre », le motif principal étant de répondre aux requêtes des décideurs sans pour autant altérer les performances des systèmes opérationnels. Le Data Warehouse, tel qu'on le connaît actuellement, n'est plus vu comme une copie -ou un cumul de copies prises de façon périodique- des données du système opérationnel. Il est devenu une nouvelle source d'information, alimenté avec des données recueillies et consolidées des différentes sources internes et externes.

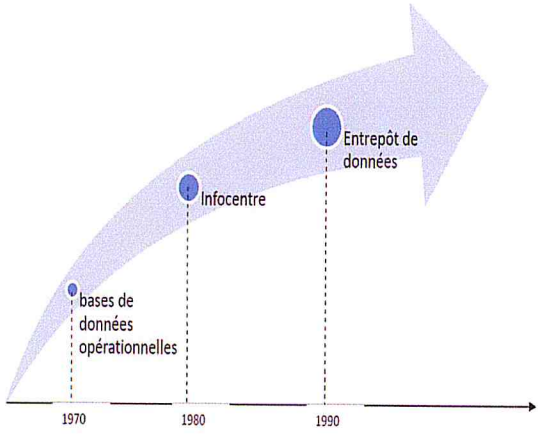

FIGURE A.1 – Évolution des bases de données décisionnelles.

A.3 Structure des données d'un Data Warehouse

Le Data Warehouse a une structure bien définie, selon différents niveaux d'agrégation et de détail des données. Cette structure est définie par Inmon [27] comme suit :

- **Données détaillées :** ce sont les données qui reflètent les événements les plus récents, Fréquemment consultées, généralement volumineuses car elles sont d'un niveau détaillé.

- **Données détaillées archivées :** anciennes données rarement sollicitées, généralement stockées dans un disque de stockage de masse, peu coûteux, à un même niveau de détail que les données détaillées.

- **Données agrégées :**données agrégées à partir des données détaillées.

- **Données fortement agrégées :** données agrégées à partir des données détaillées, à un niveau d'agrégation plus élevé que les données agrégées.

- **Meta données :** ce sont les informations relatives à la structure des données, les méthodes d'agrégation et le lien entre les données opérationnelles et celles du Data Warehouse.

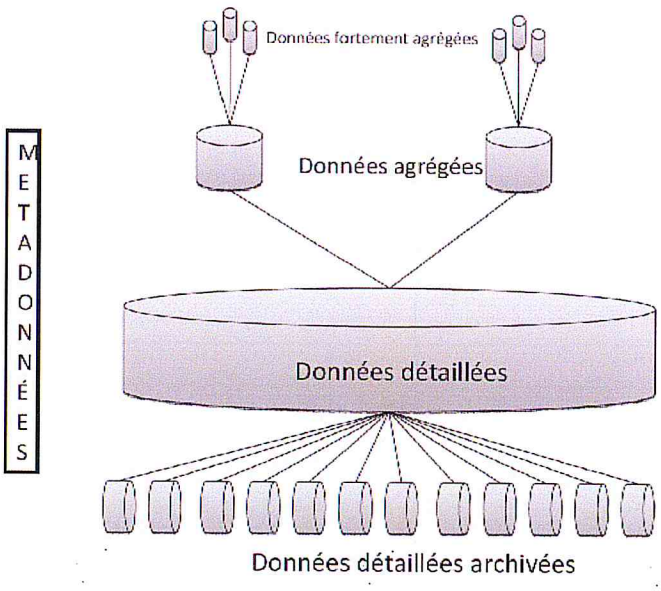

FIGURE A.2 – Structure des données d'un Data Warehouse.

A.4 Les éléments d'un Data Warehouse

L'environnement du Data Warehouse est constitué essentiellement de quatre composantes : les applications opérationnelles, la zone de préparation des données, la présentation des données et les outils d'accès aux données.

- Les applications opérationnelles : ce sont les applications du système opérationnel de l'entreprise et dont la priorité est d'assurer le fonctionnement de ce dernier et sa performance.Ces applications sont extérieures au Data Warehouse.

- Préparation des données : la préparation englobe tout ce qu'il y a entre les applications opérationnelles et la présentation des données. Elle est constituée d'un ensemble de processus appelé ETL, « Extract, transform and Load », les données sont extraites et stockées pour subir les transformations nécessaires avant leur chargement[29].

- Présentation des données :c'est l'entrepôt où les données sont organisées et stockées. Si les données de la zone de préparation sont interdites aux utilisateurs, la zone de présentation est tout ce que l'utilisateur voit et touche par le biais des outils d'accès.

A.5 Architecture d'un Data Warehouse

Après avoir exposé et défini chacun des éléments constituant l'environnement d'un Data Warehouse, il serait intéressant de connaitre le positionnement de ces éléments dans une architecture globale d'un Data Warehouse :

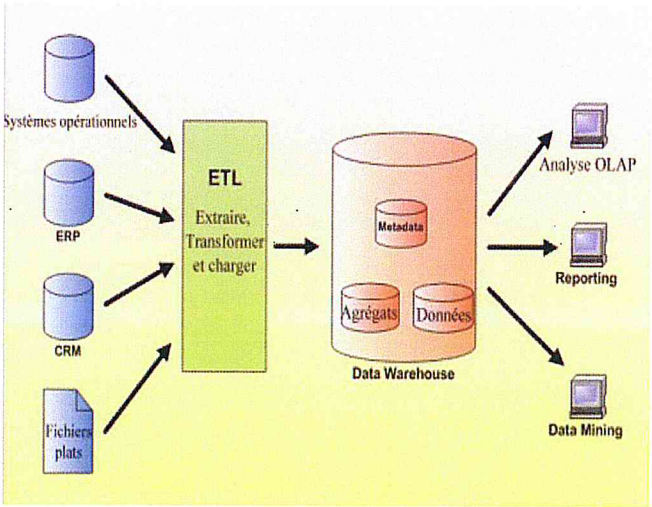

FIGURE A.3 – Architecture globale d'un Data Warehouse [1].

A.6 Modélisation des données de l'entrepôt

- La modélisation dimensionnelle et ses concepts :

Les Data Warehouse sont destinés à la mise en place de systèmes décisionnels. Ces systèmes, devant répondre à des objectifs différents des systèmes transactionnels, ont fait ressortir très vite la nécessité de recourir à un modèle de données simplifié et aisément compréhensible. La modélisation dimensionnelle permet cela. Elle consiste à considérer un sujet d'analyse comme un cube à plusieurs dimensions, offrant des vues en tranches ou des analyses selon différents axes.

En plus de la perception intuitive qu'offre la modélisation dimensionnelle, celle-ci est réputée pour ses performances élevées.

La nomination « schéma des jointures en étoile » a longtemps été adoptée pour décrire un modèle dimensionnel. Cette nomination est due au fait que le diagramme qui représente un modèle dimensionnel ressemble à une étoile, avec une grande table centrale et un jeu de petites tables auxiliaires disposées en étoile autour de la table centrale. Celle-ci est appelée table de faits et les autres tables sont appelées tables de dimensions.

- Concept de fait : Une table de faits est la table centrale d'un modèle dimensionnel, où les mesures de performances sont stockées. Une ligne d'une table de faits correspond à une mesure. Ces mesures sont généralement des valeurs numériques, additives ; cependant des mesures textuelles peuvent exister mais sont rares. Le concepteur doit faire son possible pour faire des mesures textuelles des dimensions, car elles peuvent êtres corrélées efficacement avec les autres attributs textuels de dimensions.

Une table de faits assure les liens plusieurs à plusieurs entre les dimensions. Elles comportent des clés étrangères, qui ne sont autres que les clés primaires des tables de dimension.

- Concept de dimension : Les tables de dimension sont les tables qui raccompagnent une table de faits, elles contiennent les descriptions textuelles de l'activité. Une table de dimension est constituée de nombreuses colonnes qui décrivent une ligne. C'est grâce à cette table que l'entrepôt de données est compréhensible et utilisable ; elles permettent des analyses en tranches et en dés. Une dimension est généralement constituée : d'une clé artificielle, une clé naturelle et des attributs. « *Une table de dimension établit l'interface homme / entrepôt, elle comporte une clé primaire* » [28]

- Différents modèles de la modélisation dimensionnelle :

*** Modèle en étoile :** comme indiqué précédemment, ce modèle se présente comme une étoile dont le centre n'est autre que la table des faits et les branches sont les tables de dimension. La force de ce type de modélisation est sa lisibilité et sa performance.

*** Modèle en flocon :** identique au modèle en étoile, sauf que ses branches sont éclatées en hiérarchies. Cette modélisation est généralement justifiée par l'économie d'espace de stockage, cependant elle peut s'avérer moins compréhensible pour l'utilisateur final, et très couteuse en terme de performances.

*** Modèle en constellation :** C'est un plusieurs modèles en étoile liés entre eux par des dimensions communes.

A.7 Le concept OLAP :

Le terme OLAP (On-Line Analytical Processing) désigne une classe de technologies conçue pour l'accès aux données et pour une analyse instantanée de ces dernières, dans le but de répondre aux besoins de Reporting et d'analyse.

R. Kimball définit le concept « OLAP » comme «*Activité globale de requêtage et de présentation de données textuelles et numériques contenues dans l'entrepôt de données ; Style d'interrogation spécifiquement dimensionnel* »[29].

Bibliographie

[1] S.Ghernaouti-Hélie , Sécurité Informatique et Réseaux - 2eme édition, Edition ,Aout 2008.

[2] Eric Filiol – Philippe Richard, Cybercriminalité : Enquête sur les mafias qui envahissent le web , 2006.

[3] Jean-François PILLOU, Introduction à la sécurité informatique ,article ,www.commentcamarche.net, sept 2015.

[4] Andrew Whitaker, Keatron Evans et Jack B.Voth , Chaines d'exploits : scénarios de hacking avancé et prévention, 2011.

[5] Arnaud Jacques, Initiation à la sécurité, article , www.secuiteinfo.com , Janvier 2002.

[6] Jon Ericson, Techniques de hacking , Pearson edition , 2011.

[7] Sébastian Baudru, Jérome Hennetcart and all, Sécurité Informatique :Ethical Hakhing 2eme édition ,Eni edition ,2011.

[8] Salvatore J.Stolfo and all, Insider Attack and Cyber Security : beyond the hacker, Springer, 2008.

[9] Kimberly Graves, CEH Officiel Certified Ethical Hacker Review Guide - Version6, Wiley Publishing Inc, 2007.

[10] Eric Charton, Hackers Guide 4eme edition, Pearson edition, 2011.

[11] Eric freyssinet , La cybercriminalité en mouvement , edition lavoisier, 2013.

[12] Master Professionnel « droit des activités dans le cyberespace ». Patrick Haeyaert - Direction interrégionale de la police judiciaire de Lille, juin 2007

[13] Ludovic Mé : Sécurité des systèmes d'information. Hermès, 2006.

[14] David Powell et Robert Stroud : Conceptual Model and Architecture of MAFTIA. Technical Report Series-University of Newcastle Upon Tyne Computing Science, 2003.

[15] Eric Alata, Observation, caractérisation et modélisation de processus d'attaques sur Internet. Thèse de doctorat, Institut National des Sciences Appliquées de Toulouse, 2007.

[16] Gérard WAGENER, Self-Adaptive Honeypots Coercing and Assessing Attacker Behaviour, Thèse de doctorat ,Université de Luxembourg , juin 2011.

[17] Lance Spitzner, Honeypots, definitions and value of honeynets, Addison Wesley Proffessional mai 2002.

[18] The Honeynet Project, Papers , https ://www.honeynet.org/papers , 2015

[19] Lance Spitzner, Honeypots :Tracking Hackers, Addison Wesley Proffessional , Septembre 2002.

[20] Yosra Ben Mustapha, Alert correlation towards an efecient response decision Support, Thèse de doctorat, Université Pierre et Marie Curie , juin 2015

[21] L. Spitzner, "The Honeynet Project : Trapping the Hackers," IEEE Security Privacy Magazine, Volume 1, Issue 2, Mar-Apr 2003, pp. 15-23.

[22] A. Chuvakin : "Honeypot essentials," Information Systems Security, Volume 11, Number 6, Jan-Feb 2003, pp. 15-20.

[23] The Honeynet Project, "Know Your Enemy : GenII Honeynets," http ://www.honeynet.org/papers/gen2/index.html, 2003.

[24] Experiences with a Generation III Virtual Honeynet, Fahim H. Abbasi and R. J. Harris, Massey University New Zealand

[25] Thibaut PROBST, Evaluation Et Analyse Des Mécanismes De Sécurité Des Réseaux Dans Les Infrastructures Virtuelles De Cloud Computing, Thèse de doctorat,L'université fédérale Toulouse, Septembre 2015

[26] B. Inmon , What is a Data Warehouse, Article, http ://www.billinmon.com, 2000.

[27] W. Inmon , Building the Data Warehouse Third Edition,Wiley Computer Publishing,2002.

[28] R. Kimball et M. Ross, Entrepôts de Données : Guide Pratique de Modélisation Dimensionnelle 2ème édition , Vuibert, 2002.

[29] R. Kimball et J. Caserta , The Data warehouse ETL Toolkit , Wiley Publishing Inc, 2004.