Application de la théorie de graphe pour sécuriser un système d'information

Razafy N. R.[1], Rakotomiraho S.[2], Randriamaroson R. M [3]

Laboratoire de Recherche Systèmes Embarqués, Instrumentation et Modélisation des Systèmes et Dispositifs Electroniques
(LR-SE-I-MSDE)
Ecole Doctorale en Sciences et Techniques de l'Ingénierie et de l'Innovation
(ED-STII)

[1]r.razafy@bfm.mg, [2]rakmiraho@gmail.com, [3]rivomahandrisoa@gmail.com

Résume

Pour modéliser un réseau, un administrateur commence toujours à prendre un papier blanc et dessine les principaux composants. Tous les équipements vont être présentés par des points et les liaisons par des traits. Sans savoir la théorie de graphe, il en utilise une partie, qui est un domaine d'étude mathématique. En ayant connaissance de l'existence de la richesse de cette théorie, l'objectif de cette étude est d'appliquer les différentes représentations, définitions, algorithmes et méthodes pour analyser et pour mettre en place les contre-mesures.

Mots-clés : Théorie de graphe, sécurité réseau, système d'information, modélisation réseau

Abstract

To modelize a network, an administrator always starts with the basics which is white paper and starts to draw the key components. All equipment will be presented as "dots" and the link symbolized by "straight lines". Without having a specific graph theory knowledge, the administrator is using some part of it as a science, which basically is a mathematic field of study. With a proper knowledge of the existence of what one can get from the exploitation of the Graph Theory, this etude aims to apply all different representations, definitions, algorithms, and method to analyze and deploy all counter measures.

Keywords: Graph theory, network security, computer science, network modelling

1. Introduction et présentation du modèle d'étude

Normalement toute entreprise dispose d'un minimum de système d'information pour gérer l'entreprise. L'automatisation n'est plus une question de décision mais une nécessité. Généralement, un système d'information est composé d'un logiciel de gestion et d'un logiciel de comptabilité. L'introduction de multi-utilisateur mène au stockage des informations serveurs pour que chacun manipule simultanément les informations. Une entreprise peut avoir par exemple des serveurs de domaine, des serveurs de base de données, des serveurs d'application, des serveurs de messageries.

La révolution technologique a changé les modes de communication et les modes de transaction. L'aire numérique oblige les gens à se connecter à internet et à utiliser des courriers électroniques, donc « connecter ». Les systèmes d'information évoluent suivant cette tendance. L'interconnexion n'est plus une mode mais une obligation.

Les entreprises dans son évolution ont besoin de partager des informations. Isoler pour mieux protéger n'est plus valable. Dans le cas pratique, elles ont besoins de diffuser des informations sur Internet. Pour les cas d'entreprise de prestataire

de service en ligne, les opérations autorisées pour les clients externes ne se limitent pas seulement pour une diffusion d'information mais aussi des transactions. Des informations confidentielles comme des numéros de carte de crédit peuvent être échangées. Parfois ces informations sont partagées en interne pour la gestion des opérations.

Pour cette étude, un réseau simplifié représentant un système d'information comprenant une simulation des cas illustrés ci-dessous va être pris comme exemple. Il prend l'hypothèse d'un système d'information composé des serveurs en interne, des machines clients internes pour plusieurs départements, des serveurs WEB pour des applications servant des clients internes et des clients externes. En plus, les utilisateurs peuvent aussi consulter des informations sur Internet. Pour éviter de se noyer dans les représentations, seules des échantillonnages vont être faites. Mais dans la réalité, les nombres de chaque composant peuvent être nombreux selon le cas.

L'architecture de ce système d'information est schématisée par la ***figure 1.***

2. Analyse du modèle

La ***figure 2*** est la transformation modèle avec la théorie de graphe

Le modèle est un graphe complet. Chaque sommet peut atteindre tous les autres et le sens des communications est bidirectionnel [1]. Dans ce cas, les risques sont élevés : attaque par saturation, vol d'information et attaque de non-répudiation. La prise de control d'un des postes de travail implique un contrôle total de l'ensemble du réseau.

La matrice suivante montre les relations entre chaque sommet [2] :

$$\begin{array}{c|ccccccccccc}
 & SI_1 & SI_2 & CI_1 & CI_2 & CI_3 & CI_4 & SWI_1 & SWI_2 & CE_1 & CE_2 & W_1 \\
\hline
SI_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
SI_2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
CI_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
CI_2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
CI_3 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
CI_4 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
SWI_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
SWI_2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
CE_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
CE_2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
W_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{array}$$

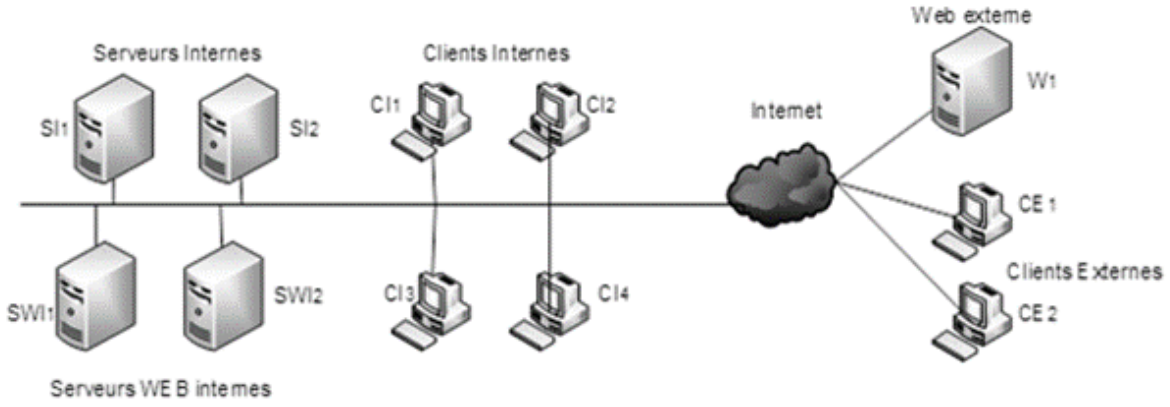

Figure 1 : Modèle réseau simplifié

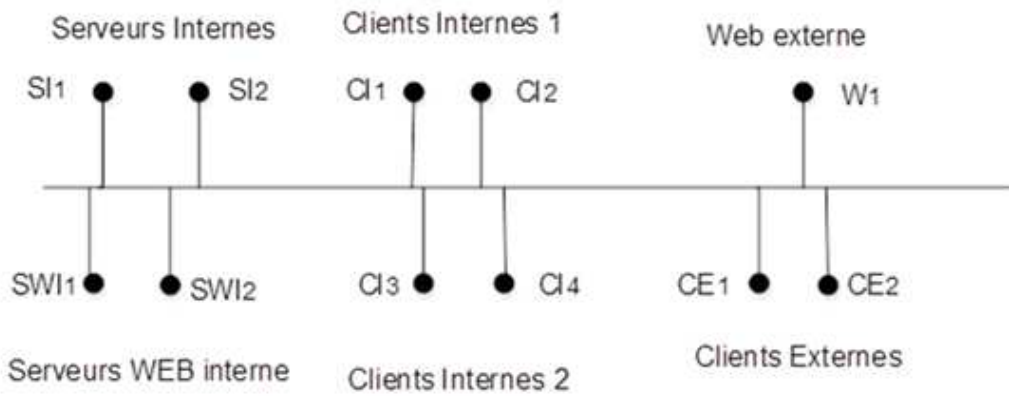

Figure 2 : Représentation par théorie de graphe du modèle

En analyse des risques, les actifs à protéger sont les données stockées dans les serveurs et les informations circulant sur le réseau. Ils peuvent faire l'objet des attaques en interne ou venant de l'externe. La sécurité se résume à la disponibilité, intégrité, confidentialité et la non-répudiation. En regardant la matrice d'adjacences, ils sont vulnérables. La charge de lien de tous les serveurs est maximale et tous les postes de travail peuvent recevoir de tous les informations circulant sur le réseau.

Comme n'importe quel poste peut atteindre tous les autres postes, un pirate peut effectuer une analyse de vulnérabilité. Il peut utiliser par exemple l'utilitaire « *nmap* » pour explorer l'ensemble de l'équipement. Ci-après un exemple de résultat d'un recensement avec l'utilitaire :

```
# nmap -A -T4 scanme.nmap.org
playground

Starting nmap (
http://www.insecure.org/nmap/ )
Interesting ports on scanme.nmap.org
(205.217.153.62):
(The 1663 ports scanned but not shown
below are in state: filtered)
PORT    STATE  SERVICE VERSION
22/tcp  open   ssh     OpenSSH 3.9p1
(protocol 1.99)
53/tcp  open   domain
70/tcp  closed gopher
80/tcp  open   http    Apache httpd
2.0.52 ((Fedora))
113/tcp closed auth
Device type: general purpose
Running: Linux 2.4.X|2.5.X|2.6.X
OS details: Linux 2.4.7 - 2.6.11,
Linux 2.6.0 - 2.6.11
Uptime 33.908 days (since Thu Jul 21
03:38:03 2005)

Interesting ports on
playground.nmap.org (192.168.0.40):
(The 1659 ports scanned but not shown
below are in state: closed)
PORT     STATE SERVICE       VERSION
135/tcp  open  msrpc         Microsoft
Windows RPC
139/tcp  open  netbios-ssn
389/tcp  open  ldap?
445/tcp  open  microsoft-ds  Microsoft
Windows XP microsoft-ds
1002/tcp open  windows-icfw?
1025/tcp open  msrpc         Microsoft
Windows RPC
1720/tcp open  H.323/Q.931   CompTek
AquaGateKeeper
5800/tcp open  vnc-http      RealVNC
4.0 (Resolution 400x250; VNC TCP port:
5900)
5900/tcp open  vnc           VNC
(protocol 3.8)
MAC Address: 00:A0:CC:63:85:4B (Lite-
on Communications)
Device type: general purpose
Running: Microsoft Windows NT/2K/XP
OS details: Microsoft Windows XP Pro
RC1+ through final release
Service Info: OSs: Windows, Windows XP

Nmap finished: 2 IP addresses (2 hosts
up) scanned in 88.392 seconds
```

Il obtient alors facilement la cartographie de l'ensemble du réseau, les ports ouverts et la version de tout l'équipement. La matrice de flux

montre qu'il peut faire cette action même depuis un client externe CE_1 ou CE_2. Il peut combiner l'attaque avec un *sniffing* par les utilitaires *WireShark* ou *Sniffpass* par exemple pour capter des informations utiles. L'absence des blocs de contrôle le favorise aussi et rend difficile à détecter les deux analyses.

Un autre type d'attaque consiste à prendre possession d'un des postes de travail du réseau en utilisant un cheval de Troie. A partir d'un minimum de privilège, le pirate essayera d'augmenter son privilège pour prendre un contrôle total d'un serveur ou d'un poste de travail. A partir de ce moment, l'ensemble des mesures de sécurité sont réduit à néant. Le pirate peut soustraire les informations confidentielles directement sur le poste contrôlé ou sur d'autre poste en utilisant la technique de DNS Spoofing ou en déployant des logiciels espions de type spyware, cookie, phishing et hoax [3]. Il peut étendre son attaque à faire aussi de détournement. En effet, une attaque sophistiquée exploite le maillon faible du système pour prendre contrôle et pour contourner les mesures de sécurité. La ***figure 3*** illustre les étapes effectuées lors de l'attaque de la Banque Centrale de Bangladesh [4]. Le réseau de cette banque est semblable à cette modèle. Elle n'a pas utilisé des firewalls et n'a pas mis en place des concepts de VLAN.

En absence d'équipements de contrôle pour gérer les communications et sans les dispositifs permettant de faire une authentification matérielle, un utilisateur peut masquer son identité en utilisant l'IP Spoofing. L'objectif est d'éviter d'être repéré. Il peut faire aussi des attaques par rebond en pivotant les attaques sur un des autres postes de travail.

Dans le cas des actions de sabotage, il est facile d'inonder les équipements du réseau. En effet, le poids de chaque nœud est élevé. En ayant un poste de travail assez puissant, un utilisateur peut attaquer tous les ordinateurs par les techniques de débordement de tampon, de « ping of Death », « SYN Flooding » et de « smurf ». Ces attaques se passent au niveau des couches transports. Plus loin un utilisateur peut exploiter tous les ordinateurs en tant que « zombie » pour élaborer un DDoS sur une cible donnée [3].

Figure 3 : Schéma de l'attaque de la Banque Centrale de Bangladesh

3. Résultat de l'application des théories de sécurité basées sur la théorie de graphe

Pour réduire la charge de chaque sommet et le surface d'attaque, il est nécessaire de faire des regroupements en sous-graphes. Le regroupement par communauté identifie les groupes de sommet ayant les mêmes règles de gestion et encapsule les sommets nécessitant des protections. En appliquant les études théoriques, le groupement serait comme suit (***figure 4***) : les serveurs internes, les clients internes (pour la modélisation, on va prendre en compte deux VLAN), les serveurs WEB internes (DMZ), les clients externes, et les serveurs web externe (internet). Les serveurs internes devront être protèges car ils ont de niveau de risque haut.

Le premier objectif est de bloquer toutes communications entre les composants pour éviter les attaques en interne et externe. Avec la configuration actuelle, la prise de contrôle d'un des sommets implique une prise de contrôle du système d'information. La bonne pratique suggère de bloquer tous les flux puis d'ouvrir seulement les flux utiles.

Pour cela, la matrice cible est de la suivante :

	SI_1	SI_2	CI_1	CI_2	CI_3	CI_4	SWI_1	SWI_2	CE_1	CE_2	W_1
SI_1	1	1	0	0	0	0	0	0	0	0	
SI_2	1	1	0	0	0	0	0	0	0	0	
CI_1	0	0	1	1	0	0	0	0	0	0	
CI_2	0	0	1	1	0	0	0	0	0	0	
CI_3	0	0	0	0	1	1	0	0	0	0	
CI_4	0	0	0	0	1	1	0	0	0	0	
SWI_1	0	0	0	0	0	0	1	1	0	0	
SWI_2	0	0	0	0	0	0	1	1	0	0	
CE_1	0	0	0	0	0	0	0	0	1	1	
CE_2	0	0	0	0	0	0	0	0	1	1	
W_1	0	0	0	0	0	0	0	0	0	0	1

L'interconnexion entre les sous-graphes doit être régie par des règles de sécurité. Elles sont composées de filtrage ou de blocage de flux. L'étape suivante consiste à mettre en place les sommets pour les interconnexions de sous-groupes. Ils serviront comme des points d'entrée/sortie et créeront aussi après les blocs de contrôle.

La mise en place de ce cloisonnement empêche l'analyse du réseau avec les scanners de vulnérabilité et les utilitaires de sniffing. En effet, les utilisateurs peuvent accéder seulement aux ordinateurs dans son sous–réseau. De même pour les attaques de saturation, ils ne peuvent inonder que sous réseau et ne peuvent pas aussi exploiter les autres ordinateurs pour des actions malveillantes. Le cloisonnement permet aussi d'alléger les flux globaux échangés dans l'ensemble du réseau.

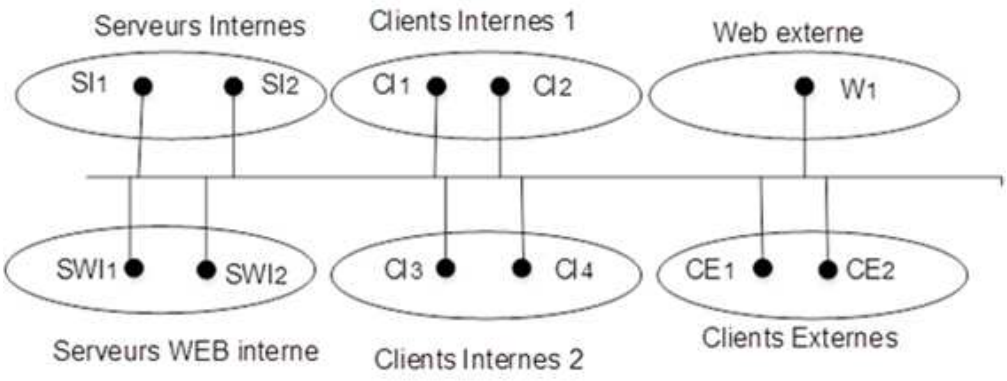

Figure 4 : Regroupement par sous-graphe

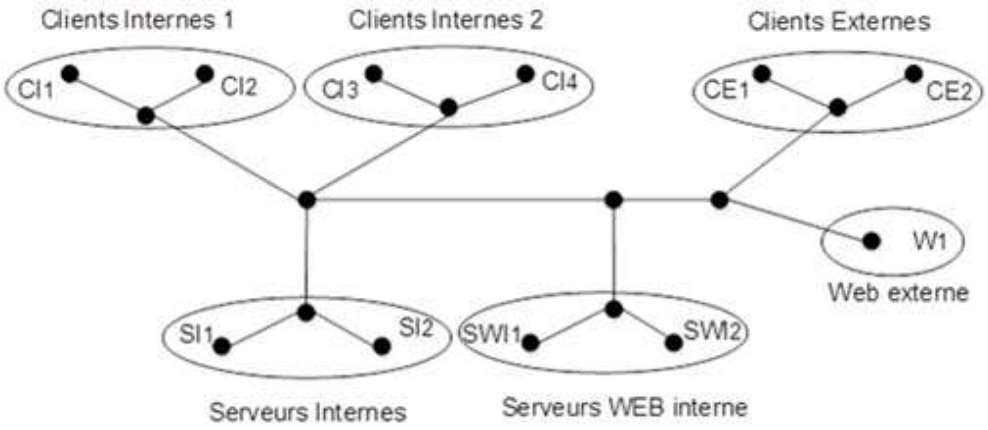

Figure 5: Mise en place de la sécurité

La communication entre deux sommets quelconques de groupe sera effectuée avec des algorithmes de routage. Elles forceront les flux interconnexions à passer sur les blocs de contrôle. L'étape suivante est donc de mettre en place les nouveaux sommets pour établir les connexions. Pour minimiser les risques, la séparation en interne se fera par des VLAN et la communication avec les mondes externes doit se faire à travers un firewall. Il est donc nécessaire de mettre deux sommets différents pour protéger les équipements internes.

La ***figure 5*** est l'application du résultat de toutes les théories.

Le graphe devient un graphe multi-bloc composé des sous-graphes définissant les communautés et les sous-graphes considérés comme des blocs de control. Ces derniers vont définir les sous réseaux virtuels. A partir de ce graphe, les sous-réseaux sont le « VLAN Interne 1 », le « VLAN Interne 2 », le « VLAN Serveur », le « VLAN DMZ » et l'internet [5]. La mise en place de ces blocs de contrôle permet d'autoriser avec des mesures de control les communications entre les clients de blocs différents. Elle assure aussi le blocage des flux autorisés.

A partir de ces résultats, les administrateurs peuvent modéliser l'architecture logique. La méthode est d'inverse les principes de transformation de la première étape.

4. Mise en place de control de flux

L'objectif principal de la mise en place de contrôle de flux est d'autoriser les flux de communication entre le composant. Les filtrages sont basés sur les protocoles qui sont associés aussi au numéro de port. En se basant sur la bonne pratique, il est toujours judicieux de refuser tous les flux puis d'autoriser seulement les flux utiles. En cas de nécessité d'ouvrir des ports vulnérables à des attaques comme le « RDP » (Remote Desktop Control), il est indispensable de les surveiller.

Les communications dépendent à ce stade de la configuration des flux dans les équipements. La matrice d'adjacences peut être utilisée pour représenter la matrice de flux. Voici quelques règles simples pour montre un exemple :

- Les clients internes CI_1 et CI_2 (premier VLAN) peuvent accéder aux serveurs internes en utilisant seulement le port de base de données (condition c1) ;
- Les clients internes CI_1 et CI_2 (premier VLAN) peuvent consulter des informations

sur internet avec des filtrages pour éviter les virus et en passant par un firewall (c2) ;

- Les clients internes CI_3 et CI_4 (deuxième VLAN2) peuvent accéder aux serveurs internes en utilisant les ports de base de données et le transfert de fichier (c3) ;
- Les serveurs internes envoient des mises à jour des données aux serveurs WEB externes seulement via un lien de base de données (c4) ;
- Les clients externes peuvent consulter et modifier les informations sur les serveurs WEB internes en passant par le firewall et seulement avec des protocoles sécurisés (c5).

La matrice de représentation sera comme suite :

	SI_1	SI_2	CI_1	CI_2	CI_3	CI_4	SWI_1	SWI_2	CE_1	CE_2	W_1
SI_1	1	1	0	0	0	0	$c4$	$c4$	0	0	0
SI_2	1	1	0	0	0	0	$c4$	$c4$	0	0	0
CI_1	$c1$	$c1$	1	1	0	0	0	0	0	0	$c2$
CI_2	$c1$	$c1$	1	1	0	0	0	0	0	0	$c2$
CI_3	$c3$	$c3$	0	0	1	1	1	1	0	0	0
CI_4	$c3$	$c3$	0	0	1	1	1	1	0	0	0
SWI_1	0	0	0	0	0	0	1	1	0	0	0
SWI_2	0	0	0	0	0	0	1	1	0	0	0
CE_1	0	0	0	0	0	0	$c5$	$c5$	1	1	0
CE_2	0	0	0	0	0	0	$c5$	$c5$	1	1	0
W_1	0	0	0	0	0	0	0	0	0	0	1

Les contrôles $c1, c2, c3, c4$ et $c5$ peuvent être de filtrage de paquet ou de contrôle de flux. Le poids de chaque lien dépend du nombre de contrôle et la latence qui est le nombre des sommets visités pour arriver à la destination. Les orientations naissent lors de la mise en place de ces conditions. Le graphe devient graphe orienté (***figure 6***). Les liens auront de sens et représenteront des protocoles bien spécifiés. La matrice changera avec les conditions et les valeurs de chaque condition dépendent de la destination finale du document. Dans la pratique, les conditions sont plus complexes et il est nécessaire de mettre en évidence aussi toutes les protocoles.

La ***figure 6*** est une représentation en graphe du modèle avec les protocoles autorisés pour chaque lien. Elle a été allège en enlevant les liens en interne des VLAN pour une meilleure lisibilité.

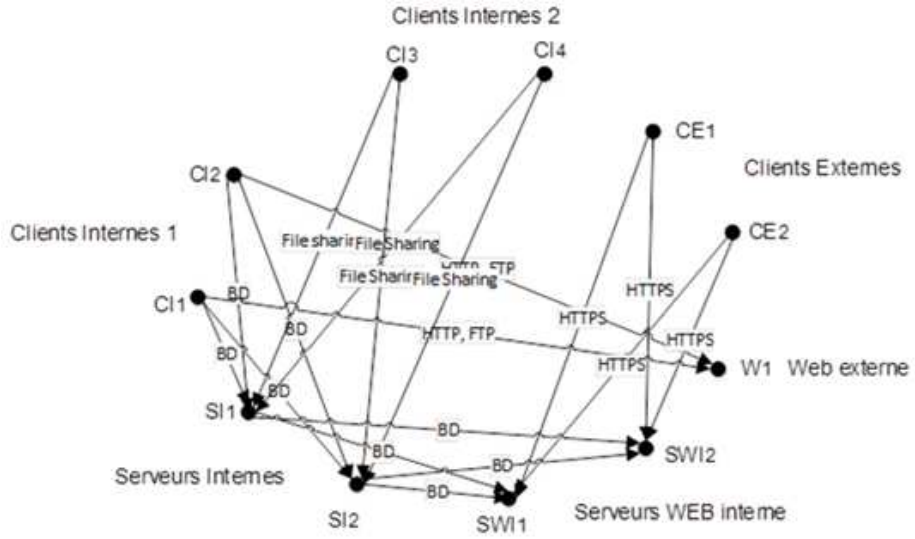

Figure 6 : Diagramme de flux

En partant sur la définition des poids de nœud définis précédemment, la matrice devient comme suit :

	SI_1	SI_2	CI_1	CI_2	CI_3	CI_4	SWI_1	SWI_2	CE_1	CE_2	W_1
SI_1	1	1	0	0	0	0	4	4	0	0	0
SI_2	1	1	0	0	0	0	4	4	0	0	0
CI_1	3	3	1	1	0	0	0	0	0	0	4
CI_2	3	3	1	1	0	0	0	0	0	0	4
CI_3	3	3	0	0	1	1	1	1	0	0	0
CI_4	3	3	0	0	1	1	1	1	0	0	0
SWI_1	0	0	0	0	0	0	1	1	0	0	0
SWI_2	0	0	0	0	0	0	1	1	0	0	0
CE_1	0	0	0	0	0	0	4	4	1	1	0
CE_2	0	0	0	0	0	0	4	4	1	1	0
W_1	0	0	0	0	0	0	0	0	0	0	1

Les valeurs « 1 » dans la matrice indique un libre échange entre les sommets. Il favorise les temps de réponse en donnant l'opportunité d'utiliser en toute sécurité tous les protocoles comme le partage de fichier. Les « 3 » et « 4 » illustrent l'existence des passages de communication dans les blocs de contrôle. Dans ce modèle, il n'y a pas de « 2 » car on n'a pas utilisé des switch programmables pour établir les VLAN. Ce cas est possible pour des réseaux avec un nombre d'ordinateurs limité. Dans une grande entreprise, il est toujours nécessaire de faire le déploiement avec des switch d'accès et switch cœur.

L'utilisation des protocoles sécurisés améliore l'architecture en empêchant les vols d'information par sniffing. Elle renforce la sécurité mais augmente le poids de chaque lien. Le renforcement de la sécurité aura un impact la performance. Il est nécessaire d'équilibrer performance et sécurité. La notion de protocole sécurisé se base sur la notion de cryptage et de décryptage. Elle aura un coût sur les performances.

Dans la pratique, un VLAN dédié pour les administrateurs est nécessaire. Il aura un accès à tous les équipements de réseau. L'objectif est que les administrateurs peuvent contrôler l'ensemble du système depuis leur poste de travail au lieu de mettre en place les outils dans des postes de chaque VLAN. Mais il faut prendre en compte que l'infection d'un des postes des administrateurs met en danger l'ensemble du système.

5. Architecture physique du modèle

A partir de la représentation logique, les administrateurs choisissent les équipements nécessaires pour le déploiement. Il est nécessaire de bien choisir les équipements pour pouvoir appliquer les règles de sécurité. Il existe plusieurs constructeurs qui fournissent ces normes. Dans cette étude l'exemple (***figure 7***) va être effectué à partir des équipements CISCO. Ils sont destinés pour les entreprises de grande envergure ou les entreprises ayant des informations sensibles à gérer ou à protéger. Pour les petites entreprises, il est possible de paramétrer tout simplement des switch programmables pour établir les différents sous-réseaux et pour configurer les flux.

Avec cette configuration, les points de contrôle pour générer les journaux d'analyse et de surveillance au niveau couche réseau sont le firewall et le switch core. Ils ne sont pas suffisants mais nécessaires. Il faut les compléter avec les journaux des applications hébergées par les serveurs. Le paragraphe suivant illustre la méthode de détection et de prévention d'intrusion.

Conclusion

La théorie de graphe est une étude complète pouvant être utilisée dans le cycle d'une sécurisation d'un système d'information. Depuis la conception, les représentations à partir des points et des traits sont plus faciles à dessiner sur papier et elles sont utilisées par de nombreux administrateurs. Ces petits dessins font partis de la théorie des graphes mais utilisés inconsciemment. La théorie des graphes est un outil complet utilisable dans le cas de cette étude. Elle possède les composants nécessaires pour mener à un résultat fiable.

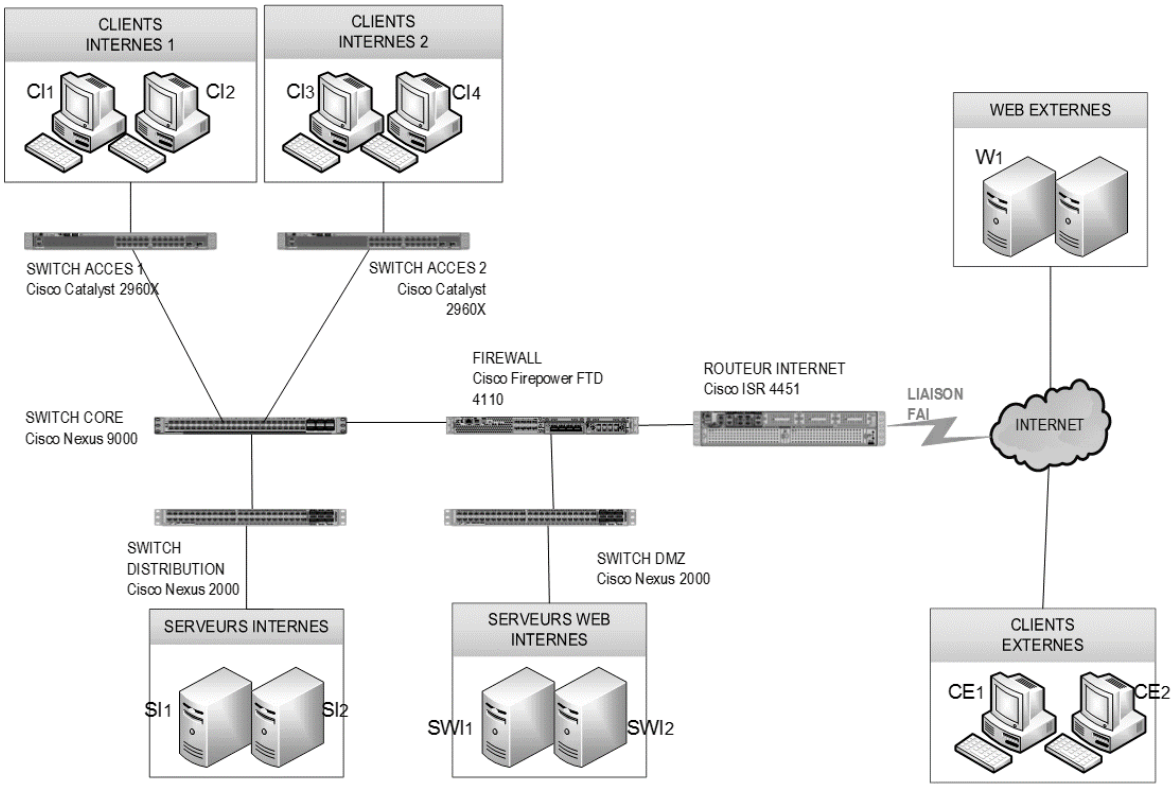

Figure 7 : Représentation physique de l'architecture

Bibliographies

[1] Paul Erdös et Alfréd Rényl, *« On random graphs »*, Publicationes Mathematicae, 1959

[2] Didier Müller, *« Introduction à la théorie des graphes »*, Commission Romande de Mathématique, 2012

[3] Jean-Olivier GERPHAGNON, *« Attaques informatiques »*, 2010

[4] *http://thehackernews.com/2016/04/swift-bank-hack.html*, 2016

[5] *http://resources.infosecinstitute.com/vlan-network-chapter-5/*, 2016