Journal intime d'un auditeur

Julien VAUBOURG

UNIVERSITÉ DE REIMS CHAMPAGNE-ARDENNE

Stage de fin de DUT effectué du 6 avril au 12 juin 2009
Entreprise : **CS One** (24 rue Philippe 51100 REIMS)
Tuteur : David REINA (consultant)
Maître de stage : Jean-Michel Nourrit

Merci.

1. David pour sa confiance, sa patience, son aide et ses conseils
2. Raphaël pour son accueil chaleureux au sein de CS One
3. Pierre pour sa compagnie et sa bonne humeur
4. Adrien pour sa sympathie
5. Google pour ses nombreux conseils
6. Wikipédia pour ses définitions
7. Tous ceux qui font vivre le logiciel libre, sans qui rien ne serait aussi passionnant

Synthèse

Création d'une sonde permettant la capture des données d'un réseau et l'édition de rapports détaillés sur la direction et le type des flux rencontrés.

Annotations

Table des matières

1 Introduction

Lolica est une association de promotion des logiciels libres sur Reims. En quête d'un stage dans une entreprise qui privilégiait le logiciel libre, c'est vers eux que je me suis tourné pour trouver des entreprises qui pouvaient correspondre à ce critére. C'est ainsi que j'en suis venu à contacter ARESSI, une société de recherche et développement établie sur Reims depuis plusieurs années. Je découvre par la suite que la société n'existe plus, mais que c'est CS One, une jeune entreprise fleurissante et du même créateur que ARESSI qui m'accueillera durant deux mois.

Si on prend notre société au rang d'un royaume, il nous faut une monarchie. Pour faire une monarchie, il nous faut une famille royale. Les rôles qui s'en découlent reviennent alors de droit aux membres de la famille REINA, qui constituent pour l'instant quasiment l'intégralité de l'équipage de la petite société. Il y a des gueux aussi, une place qui revient aux trois stagiaires, dont je fais partie. Mais il faut également un nom et un blason à ce royaume : nous le nommerons donc CS One (le blason est en couverture). Enfin, satisfaits d'avoir notre royaume avec tout son équipement, pour raconter une histoire il nous faut une quête. Cette quête sera la conquête de la mairie K, et elle sera racontée en détails dans la suite de cette histoire.

Installez-vous chaudement, je m'en vais vous la conter.

2 Le royaume de CS One

2.1 ARESSI, un premier royaume

2.1.1 La pierre philosophale : la NS Appliance

Ares est un dieu grec, le dieu de la guerre et de la destruction dans la mythologie grecque (il est assimilé à Mars chez les romains). Un nom paradoxal pour une société qui décida en l'an de grâce 2004 (fin 2003) de se lancer dans la recherche et le développement d'un concept novateur : les solutions `U.T.M.`[1] associées à un système `S.S.M.`[2].

Ce type de solution propose une gestion de la sécurité d'un réseau informatique complète, entiérement automatisée et regroupée dans un seul boitier. ARESSI avait nommé sa solution NS Appliance[3]. Elle se présentait sous la forme d'un boîtier de la taille de quatre annuaires empilés deux à deux, qui embarquait une mémoire vive allant de deux-cent-cinquante-six méga-octets à deux giga-octets (selon la taille et l'utilisation du réseau), un disque dur de quarante à quatre-vingt gigas-octets et trois interfaces réseaux. Côté logiciel, l'indispensable de la sécurité informatique, à savoir : un `antivirus`[4], un `anti-spams`[5], un `anti-phishing`[6],

[1] Unified Threat Management, soit « Gestion des menaces unifiées » en presque bon gaulois.

[2] Service de Sécurité Managée.

[3] NetSecure Appliance, le NetSecure s'étant réduit à NS suite à un conflit de propriété intellectuelle.

[4] L'`antivirus` contrôle des logiciels malveillants qui pourraient arriver sur le réseau local depuis Internet, par courriel ou depuis une page Internet.

[5] L'`anti-spams` intercepte des courriels jugés indésirables qui arrivent sur le réseau.

[6] L'`anti-phising` prévient des courriels qui semblent ne pas provenir réellement de la personne qui tente de s'indentifier en tant

un proxy[7], un pare-feu[8], et quelques autres services.

Ce boitîer se branche entre le routeur[9] qui se charge de fournir la connexion Internet et le commutateur[10] principal du réseau. Ainsi, aucune entrée ou sortie entre le réseau et Internet ne passe sans être controlé attentivement par la NS Appliance.

Les solutions U.T.M. offrent la possibilité à des sociétés de type [11]Petites et Moyennes Entreprises, ce sont des entreprises dont la taille, définie à partir du nombre d'employés, du bilan ou du chiffre d'affaires, ne dépasse pas certaines limites ; les définitions de ces limites diffèrent selon les pays. de s'assurer de la sécurité de leur réseau informatique sans avoir à se soucier d'installer et de configurer toutes les briques logicielles dont elles ont besoin. Dans la mesure où c'est un produit automatisé il est facilement accessible et s'installe simplement. Le concept était d'autant plus intéressant que ARESSI se lance à l'époque dans un concept nouveau et sans concurrents : les solutions S.S.M..

Service de Sécurité Managée La principale force de cette solution résidait dans la sécurité qu'elle apportait au

---

qu'expéditeur.

[7]Le proxy s'occupe de télécharger les pages Internet qui sont demandées et les renvois au commanditaire si elles ne correspondent pas à certaines catégories de pages interdites.

[8]Le pare-feu a pour charge de veiller à ce qu'on ne puisse pas se connecter n'importe où et n'importe comment sur le réseau local, il surveille les entrées sorties, et plus précisément les ports

[9]Unité qui permet d'interconnecter deux ou plusieurs réseaux.

[10]Dispositif permettant d'établir ou de faire cesser des connexions (circuits) temporaires entre plusieurs points quelconques d'un réseau.

[11]P.M.E.

client : aucun accès ne devait être possible depuis l'extérieur sur la machine. Pour autant, le client est assuré que la NS Appliance est toujours totalement opérationnelle, l'astuce résidant dans le simple fait que la NS envoie régulièrement d'elle-même des rapports sur son état, contrôle la présence éventuelle de mises à jour, et les installe si nécessaire. Ainsi, des établissements sensibles comme le secteur bancaire ne sont pas contraints d'ouvrir des accès qui pourraient être violés par des pirates informatiques, tout en ayant la certitude que leur équipement est contrôlé et mis à jour. Ce système fonctionne grâce à des agents de surveillance qui ont pour mission de régulièrement contrôler l'état des différents services et les relancent éventuellement, sans oublier d'envoyer une alerte au serveur ARESSI. A l'époque, seules des offres d'infogérance existent : elles consistent à gérer un réseau à distance en ayant un accès autorisé. ARESSI se rend alors vite compte que cette solution ne peut pas être fiable pour des systèmes de sécurité, les conséquences d'un piratage de l'entreprise autorisée à accéder aux différents réseaux des entreprises pouvant être dramatiques.

2.1.2 Une entreprise en pleine crise

Après plusieurs années de recherches et développement, des investissements collossaux de la part des associés, du Conseil Régional et de grosses sociétés comme OSEO, le produit arrive enfin à terme. Il faut alors penser à passer à la phase commerciale de l'aventure, et c'est à ce moment précis que les problèmes commencent.

Deux solutions s'offrent alors à ARESSI :

1. Trouver des investisseurs, embaucher un directeur commercial, une équipe de commerciaux, et se construire

un réseau commercial.

2. Trouver des partenaires et exploiter leurs propres réseaux commerciaux.

Les deux solutions sont envisagées et la première tombe rapidement à l'eau : nous sommes quelque temps avant que les médias ne parlent de crise, et pourtant elle semble déjà présente. Les investisseurs se font donc rares et surtout frileux, il sera impossible d'exploiter cette solution dans ce contexte. La seconde est donc exploitée, et ARESSI parvient à intéresser quatre entreprises, réunies par un cabinet parisien. C'est alors qu'un audit de la société est effectué : le concept est novateur et la solution de qualité, celui-ci se passe très bien. Malgré ça et le temps passant, la crise fait rage et les quatre sociétés abandonnent finalement l'idée de risquer l'aventure. ARESSI continuera quelque temps, avant de décider en mars 2009 de cesser l'aventure, faute de trouver suffisamment de moyens pour commercialiser le fruit de son travail.

2.2 CS One, une aventure plus modeste

2.2.1 Sa création

CS One a été créée par Raphaël REINA, le fondateur de ARESSI, en avril 2008. C'est dans la période florissante et pleine d'avenir pour ARESSI qu'il décide de monter cette petite société, au capital et à l'ambition plus modeste que sa précédente oeuvre. CS One, CS pour Consulting et Security et One pour l'aspect minimaliste de la structure de la société, a pour vocation de faire du conseil. Les audits réseaux et téléphonie sont le coeur de métier de CS One.

Sa principale originalité par rapport à ses concurrents est de réunir sous un même drapeau la possibilité d'offrir du conseil en matière de réseaux, de sécurité et de téléphonie. Pour les missions nécessitant d'autres qualifications, l'entreprise choisit de sous-traiter en faisant appel à des consultants indépendants.

Malgré le désir bien marqué de se détacher complétement de l'ancienne société ARESSI, la NS Appliance, produit phare de la disparue, continuera d'être supportée par CS One pour encore deux années. Pour l'occasion elle sera renommée PSM Box et une série de problémes constatés sur la NS Appliance seront corrigés. C'est d'ailleurs la mission de mon collégue stagiaire, Pierre, issu aussi de l'I.U.T. de Reims. Le troisième stagiaire, étudiant à SUPINFO et présent deux jours par semaines sur la période où nous étions là, s'occupe quant à lui de concevoir un outil de supervision à distance des PSM Box afin de pouvoir déterminer sur un seul écran l'état matériel des machines (occupation du processeur, température, charge de la mémoire, etc) qui sont déployées chez les clients.

2.2.2 Les audits réseaux

Les audits ont plusieurs intérêts et peuvent intervenir dans plusieurs cas. Ils peuvent être demandés par la direction d'une entreprise ou par le propre chef du responsable informatique. Lorsqu'une société est controntée à une migration ou un choix de solution à acheter, elle prend le risque de se faire installer et de payer des choses qui ne lui sont pas utiles. Ainsi, l'audit est là pour dresser une liste clairement définie des besoins constatés, sans aucun intérêt marchand

derrière. Il peut encore être sollicité pour définir correctement les besoins des utilisateurs du réseau informatique, ou conseiller en pensant à prévoir les futures évolutions du réseau.

Il peut également intervenir pour justifier la refonte d'un réseau, et être demandé par le responsable informatique pour trouver les arguments nécessaires et convaincants. De manière plus générale, il est là pour apporter un oeil extérieur sur des organisations qui en ont parfois besoin pour réaliser des abbérations qui pourraient paraitre évidentes. Ainsi, il est déjà arrivé de découvrir des sociétés basées sur plusieurs sites qui payaient des offres `A.D.S.L.`[12] pour chacun des sites, en ayant un site qui bénéficiait de la fibre optique, sans penser à mutualiser cette fibre optique pour centraliser l'accès à Internet et économiser des frais de connexion. Dans d'autres cas, ce sont les éléments du réseau qui convergent tous vers un même équipement de sept ans d'âge et qui n'est plus produit par le constructeur. Le jour où la machine empirique rend l'âme, ce genre de détails peut ainsi paralyser une entreprise pendant des heures si il n'est pas prévu au préalable. On peut encore citer des administrateurs qui oublient de sauvegarder la configuration de leurs `commutateurs`, s'offrant ainsi le plaisir de perdre un précieux temps à reconfigurer une machine qui a pourtant été remplacée dans l'heure. Les exemples que j'ai pu entendre sont nombreux, bien assez nombreux pour comprendre que le conseil informatique a de l'avenir et que CS One semble marcher sur une branche encore bien épaisse et reliée solidement à un tronc qui continue de pousser.

[12] Asymmetrical Digital Subscriber Line, technologie capable de transporter plusieurs mégabits par seconde sur les deux fils de cuivre du téléphone.

Le métier de consultant réside donc principalement dans l'expérience et la bonne connaissance du marché, afin d'être à même de conseiller en toutes connaissances de cause. Lorsqu'un audit est réalisé, CS One accompagne souvent leurs clients jusqu'au bout des démarches : de l'audit pour déterminer les besoins de l'entreprise, jusqu'à l'organisation des différentes tâches à effectuer en passant par le choix des prestataires.

Plusieurs types d'audit-type sont proposés par CS One :

1. Audit d'architecture réseau et sécurité : Il s'agit de conseils concernant la structure du réseau et la pertinence de la politique de sécurité.
2. Audit de l'architecture de téléphonie : Le plus souvent il s'agit d'entreprises qui expriment la volonté de passer à un système de téléphonie sur `I.P.`[13] (le réseau informatique est utilisé pour transmettre les données téléphoniques mais les téléphones restent analogiques) ou un système de voix sur `I.P.` (les téléphones sont directement reliés au réseau et possédent directement une adresse `I.P.`, tout est informatisé).
3. Audit de sécurité des flux réseau et Internet : Il s'agit de surveiller le trafic réseau pour déterminer quelle machine communique avec quelle machine, de quell façon, et dans quelles proportions. Il permet de faire un bilan sur la sécurité et de vérifier qu'il ne se passe rien d'anormal sur le réseau. C'est directement pour ce travail que le projet sur lequel j'ai passé deux mois de stage sera utilisé.

[13] Internet Protocol, protocole de transmission de l'Internet, décrivant aussi les adresses du réseau.

4. Audit de sécurité organisationnelle et technique : Il arrive parfois qu'un réseau soit parfaitement sécurisé, à jour et impénétrable, mais que les clés de la salle des serveurs soient accrochées à un clou à côté de la porte blindée. Ce genre de détails fait parti des conseils qui sont parfois à apporter. On peut aussi citer les cas où les extincteurs sont absents, où les serveurs de sauvegardes sont entreposés dans la même piéce que les autres machines, et encore bien d'autres.
5. Audit de sécurité externe (tests d'intrusions) : Véritables jeux de *hackers*, le but est de tenter de parvenir à introduire le système du client et de récupérer un maximum d'informations sans qu'on on ait eu la moindre information sur son réseau. Les moyens peuvent aller des prouesses informatiques au simple coup de téléphone mensonger.

L'une des difficultés d'un audit réside dans la communication nécessaire avec l'équipe de la gestion informatique comme avec les utilisateurs du réseau, pour que les uns ne se sentent pas jugés et contrôlés, et les autres acceptent de changer leurs habitudes (qui sont parfois si dures à quitter).

Outre les audits, CS One propose des formations dans le cadre d'une veille technologique.

2.3 Une entreprise sur trois sites

2.3.1 L'équipe de CS One

Malgré la petite structure de l'entreprise, elle n'est pas pour autant dispensée d'être éclatée en plusieurs sites. Ainsi, si Reims reste le principal centre des activités, un bureau

est basé à Rennes et un autre à Fresne. Enfin, le siége social est à Paris, et se résume à la prestation de service d'une société chargée de relayer les appels téléphonique et fournir des salles de réunion en cas de besoin. Il semblerait qu'avoir une adresse parisienne est primordiale et qu'un consultant rémois ne peut pas prétendre aux même parts de marché qu'une société équivalente de la capitale.

Je l'ai évoqué dans mon introduction moyenâgeuse, CS One est pour l'instant une entreprise centrée sur quelques membres d'une même famille. Ainsi, j'ai pu côtoyer entre autres David REINA, fils de Raphaël REINA, qui a été mon maître de stage durant ces deux mois. Après avoir été consultant avant-vente chez Rétis Communication, il est arrivé chez CS One en 2009, en tant que consultant. Il posséde différentes qualifications dans le domaine de téléphonie `I.P.` (notamment des certifications Alcatel et Cisco) et c'est principalement pour cette raison qu'il a trouvé sa place chez CS One, aux côtés de Raphaël avec qui les connaissances sont complémentaires. Il est passé par une classe prépa, par l'ENSASAT de Lannion, pour enfin faire un master à la Staffordshire University en Angleterre. Son stage de fin d'étude s'est déroulé chez ARESSI, pour travailler sur la première version de la NS Appliance.

Lucie REINA, fille de Raphaël et donc soeur de David, est chargée de la communication, en contrat professionel. Elle a fait l'ISCOM en marketing publicité et sort de l'ESP (formation de chef de publicité pour valider une licence en communication). Elle travaillera théoriquement chez CS One encore jusqu'à fin juillet pour aider à définir la communication interne et dynamiser les actions marketing auprès des clients.

Enfin, nous finiront par Espérance REINA, qui s'occupe de la comptabilité à mi-temps depuis le site de Fresne.

A terme, il est prévu que de nouveaux consultants soient embauchés.

2.3.2 Les moyens techniques

Lorsque nous sommes arrivés en début de stage, David s'occupait de virtualiser différents serveurs du local technique. L'intérêt de la virtualisation est de pouvoir rassembler plusieurs machines en une seule et d'économiser ainsi autant d'énergie que de machines fusionnées, l'impact étant aussi bien économique qu'écologique.

Le bilan technologique se résume donc à :

1. Un serveur Xen Server : C'est lui accueille les différents serveurs virtualisés. On peut y retrouver un serveur de développement (celui sur lequel nous avons travaillé), un serveur de courriels (c'est un simple relai, afin qu'on ne puisse pas identifier l'adresse `I.P.` de la société par son simple biais), un serveur de supervision de l'état des machines (actuellement en développement par Adrien, le stagiaire de SUPINFO évoqué auparavant, et utilisé certainement dans le futur en production), et enfin le serveur qui possède la seule adresse `I.P.` autorisée à accéder, de façon exceptionnelle et sur demande du client, aux NS Appliances et PSM Box. Sans oublier le serveur applicatif Windows qui accueille le `C.R.M.`[14] et le logiciel de

[14] Customer Relationship Management (Gestion de la Relation

comptabilité.

2. Un serveur `N.A.S.`[15] : Il reçoit quotidiennement toutes les données des différents serveurs.
3. Un serveur `F.T.P.`[16] : Il contient en réalité deux serveurs `F.T.P.` sur une même machine, il sert pour les mises à jour et les sauvegardes des NS et PSM.
4. Un `P.A.B.X.`[17] : Il sert à gérer la téléphonie interne et externe.
5. Une PSM Box : Inutile de revenir en détails dessus, elle est là pour sécuriser le réseau informatique.

Un serveur de terminaux a été installé sur le Xen Server afin de servir aux terminaux que nous avons utilisés en tant que stagiaires. Les terminaux sont de petites machines très silencieuses qui chargent un système Linux au démarrage et se connectent sur un serveur distant pour fonctionner.

Outre ces machines, on peut citer des Zyxell Zywall qui sont des équipements qui font fonction de `routeur` et de `pare-feu` sur les différents sites, un `onduleur`[18], un système de vidéo surveillance, et deux lignes téléphoniques,

Client), ensemble des processus visant à gérer les activités d'avant et d'après vente au client.

[15] Network Attached Storage, désigne un périphérique de stockage relié à un réseau dont la principale fonction est le stockage de données en un gros volume centralisé pour des clients-réseau hétérogènes.

[16] File Transfer Protocol, protocole d'échange de fichiers.

[17] Private Automatic Branch eXchange, sert principalement à relier les postes téléphoniques d'un établissement (lignes internes) avec le réseau téléphonique public (lignes externes).

[18] Appareil destiné à protéger un équipement électronique contre les baisses de tension et les micro-coupures du réseau électrique et qui dispose d'une batterie permettant le relais du réseau électrique en cas de coupure.

une pour Internet et l'autre pour la liaison entre les sites. Les sites sont interconnectés grâce au protocole V.P.N.[19] (un passage à de l'S.D.S.L.[20] étant prévu dans le but de faire fonctionner de la visioconférence). Un commutateur HP est utilisé sur le site de Reims.

La politique de sécurité consiste à envoyer une sauvegarde du serveur de données, le N.A.S., toutes les nuits vers le site de Fresnes. Au niveau du serveur de développement, un système de R.A.I.D.[21] des disques dur a été mis en place (deux disques dur sont écrits en même temps pour plus de sécurité) à l'aide d'une carte physique qui assure son fonctionnement. Cette installation a fait suite à un orage qui est intervenu au cours du stage et qui a été fatal à l'onduleur qui était chargé de contrôler la tension délivrée aux serveur ainsi qu'au disque dur du serveur de développement. Un système de R.A.I.D. logique (logiciel) avait pourtant été mis en place, mais c'est à ce moment précis que David et Raphaël ont découvert qu'il n'avait jamais fonctionné correctement. Les priorités avaient été mal définies, son fonctionnement n'avait jamais été vérifié. Ainsi aucun service de sauvegardes sur site distant n'avait encore été mis en place (il était prévu pour les jours à venir), et l'intégralité de nos projets de stage se trouvaient sur le disque

[19] Virtual Private Network, architecture logicielle permettant de créer un réseau virtuel entre des machines connectées par Internet. Cela peut permettre d'utiliser Internet comme support de communication entre une société et ses filiales dans le monde, en assurant la confidentialité des échanges.

[20] Symetric Digital Suscriber Line, technologie qui, comme l'A.D.S.L., permet d'augmenter la capacité des lignes téléphoniques traditionnelles en offrant des vitesses de transmission jusqu'à trente fois plus élevées qu'une connexion classique.

[21] Redundant Array of Independent Disks.

dur. On peut également noter qu'un travail de Raphaël a été perdu ainsi que de nombreuses plaquettes de publicité de l'entreprise. Les deux autres stagiaires ont pu récupérer la partie de leur travail qu'ils avaient envoyé sur leur PSM Box, et pour ma part je n'ai rien pu sauver (bien que mon travail était déjà très avancé, rien n'était encore passé en production). Ironie du sort, je m'étais préparé à cette éventualité et en copiant l'intégralité de mes fichier sur le serveur des terminaux, une semaine auparavant. Malheureusement les joies de la virtualisation m'ont été fatales, je ne savais pas que les deux serveurs étaient virtualisés sur le même équipement, et il nous était strictement interdit de copier nos codes-sources sur un support externe ou encore de les envoyer par Internet. Malgré des semaines entières de travail perdues, nous pouvons nous satisfaire d'avoir eu une démonstration convaincante de l'importance de ne pas négliger la sécurité des données, et de ne pas la bafouer au détriments d'autres impératifs.

3 Projet de stage

3.1 La mairie K

3.1.1 Une commande problématique

Alors que CS One fête bientôt ses un an d'existence juridique, une commande d'audit des flux réseaux émane d'une mairie située en banlieue parisienne. Nous dénominerons cette mairie en tant que « mairie K », faute d'avoir les autorisations nécessaires pour la citer au sein de ce rapport. Et si c'est un *K* qui a été choisi pour ce pseudonyme, c'est parce qu'elle s'est vite avérée être un *cas* aussi intéressant que problématique. En effet la mairie n'est pas petite :

le réseau pour lequel elle demande un audit est déployé à travers une quarantaine de sites différents (hôtel de ville, crêches, écoles, etc), soit environ six-cent machines. Mais surtout il est curieusement organisé, et c'est précisément la surprise qu'auront les consultants de la société. Un empilement de matériels, de `routeurs` et de `commutateurs` est constaté, rendant ainsi le suivi des données et la détermination de leur provenance particulièrement difficile.

3.1.2 Des lenteurs sur le réseau sont constatées

L'audit est commandé par le service informatique de la mairie, des lenteurs sur le réseau ont été constatées et on soupçonne des sites de ne pas utiliser le réseau à bon escient (pour le téléchargement, par exemple). Aussi, tous les flux Internet sont censés passer par le `proxy` de la mairie, et personne n'est censé se connecter directement sur Internet en outrepassant la surveillance du réseau. L'objectif étant également de déterminer si certaines personnes parviennent à déroger à la régle. Enfin, et de manière plus générale, nous ferons un bilan complet des activités enregistrées sur le réseau.

3.1.3 Un premier échec

Un réseau complexe est aussi un réseau dans lequel il n'est pas facile de s'insérer. A l'époque, CS One décide d'utiliser une PSM Box pour faire ce travail d'automatisation de l'audit, en pensant que le `proxy` qu'elle intégre et le système de rapports qui lui est associé suffira. Le boîtier est alors branché en coupure sur le réseau, un maximum de trafic devait passer au travers de son système d'analyse. Un

premier probléme survient alors : une fois le boitier mis en coupure sur le réseau, celui-ci ne répond plus. La machine est prévue pour faire fonction de passerelle pour des données qui transitent via le `port 80`[22], or le `proxy` propre au réseau de la ville fait transiter les données sur le `port` exotique 8080. Le problème est détecté puis corrigé dans la foulée. Second essai : un autre problème est alors mis en évidence, la PSM Box ne posséde pas l'adresse `I.P.` autorisée à sortir sur Internet, ses données sont donc contrôlées par le `proxy` duquel elle vient d'intercepter les données. La boucle est bouclée et la conclusion tombe comme une évidence, les PSM Box ne sont pas des machines destinées à ce genre de tâche et c'est une erreur d'essayer de les utiliser en tant que tel sur un réseau aussi complexe. En janvier 2009, CS One promet donc de revenir sur cette commande, une fois qu'un outil capable de répondre aux attentes de la mairie aura été développé.

3.2 Le sujet de stage

3.2.1 Préambule

Objectifs et état des lieux Dès mon arrivée au sein de CS One et dès les premières discussions au sujet de mon travail, on me parle de cette commande problématique. Le sujet de mon stage aura donc immédiatement un objectif, parvenir à permettre à CS One de réaliser cet outil, en développant un logiciel capable de répondre aux exigences liées à ce cas (car si cet audit est réalisable avec l'outil, la plupart des audits le seront).

[22] Nom de la sortie habituelle des données à destination de Internet.

Le rapport devra être un rapport P.D.F.[23], et permettra de visualiser le trafic qui circule entre les différents sites[24] qui utilisent le réseau. Il devra également permettre de savoir qui sort sur Internet et comment. Le bilan des observations entre chaque site (ou entre chaque site et Internet, entre chaque site et le réseau, entre deux sites, ou entre le réseau et Internet) devra rendre compte des protocoles concernés (typiquement T.C.P.[25] ou U.D.P.[26]) et des applications utilisées (S.S.H.[27], S.M.T.P.[28], WWW[29], etc). Ceci en indiquant les volumes transférés (sur la base des octets), leur fréquence dans le temps (graphiques temporels) et les principales adresses I.P. qui émettent et reçoivent pour chacune des deux zones. On souhaitera également pouvoir observer les proportions des protocoles ou des applications à l'aide de graphiques de type camemberts.

Dans un premier temps, David souhaite que j'ajoute à la PSM Box la possibilité d'être, cette fois-ci, réellement utilisée en audit. Je devais donc utiliser les outils déjà existants pour parvenir à réaliser ce rapport. Il m'oriente alors vers un outil qui s'appelle Ntop. Ntop est un utilitaire open-

[23] Portable Document Format, un langage de description de pages d'impression créé par Adobe Systems.

[24] La compréhension de site devra souvent se faire en tant qu'établissement, et non la réduction de « site Internet ».

[25] Transmission Control Protocol, protocole de transport fiable, en mode connecté.

[26] User Datagram Protocol, protocole de transport en mode non-connecté.

[27] Secure SHell, à la fois un programme informatique et un protocole de communication sécurisé.

[28] Simple Mail Transfer Protocol, protocole de communication utilisé pour transférer les courriels vers les serveurs de messagerie électronique.

[29] World Wide Web, désigne ici la navigation classique par Internet.

source qui s'occupe de capturer les données du réseau, et de fournir presque en temps réel un bilan détaillé de l'analyse des données observées et cumulées. Ce bilan est présenté sur une interface *web* et illustré à l'aide de graphiques de temps et de tableaux de synthétisation. Globalement l'analyse est très poussée et l'outil performant puisqu'il capture le trafic et propose directement une organisation des données. Je signale donc à David que j'utiliserai effectivement ce logiciel, qui réalise déjà une bonne partie de mes objectifs.

Les zones La notion de zones intervient dans le souhait de vouloir observer les interactions entre différents sites. Sur le réseau tout est adresse `I.P.`, et nous il faut un moyen de savoir à quel site appartient chaque adresse pour pouvoir évoquer ces différents sites dans le rapport final. La répartition des adresses sera définie dans ce que nous appellerons des zones.

Une zone regroupe un ensemble d'adresses `I.P.` et de sous-réseaux (un sous-réseau étant lui-même tout un ensemble d'adresses `I.P.`) et posséde un nom (exemple : « Hôtel de ville »). Il est également possible d'indiquer des intervalles d'adresses `I.P.` (sous-ensembles d'un sous-réseau) ou des intervalles de sous-réseaux (plusieurs sous-réseaux pour lesquels un octet de l'adresse est consécutif et le masque de sous-réseau identique).

Nous pourrons également créer des zones de zones, j'ai ainsi implémenté cette fonctionnalité en créant les GZones. Une GZone particulière « réseau » sera créée automatiquement, puisqu'elle rassemblera toutes les zones définies par l'utilisateur et fera office de représentation de l'ensemble

du réseau. Toutes les adresses qui n'appartiennent pas à la GZone réseau seront donc vues comme des adresses d'une autre zone particulière et créée spontanément : Internet.

3.2.2 La capture du réseau

Capturer un réseau, c'est intercepter les données qui passent entre deux éléments essentiels de l'architecture. Lorsqu'elles circulent sur le réseau, ces données sont appelées des paquets. Admettons qu'un réseau possède un `commutateur` qui constitue le noyau dur de l'architecture : sur celui-ci passe plusieurs sous-réseaux, ainsi qu'un `proxy` qui permet une sortie vers Internet. Si une machine, quelle se trouve n'importe où sur le réseau, souhaite sortir sur Internet, elle passera donc obligatoirement par la connectique qui relie le `proxy` au `routeur` qui permet l'accès à Internet. Pour observer ce qui sort et ce qui rentre d'Internet pour l'ensemble du réseau, il suffit alors de se placer entre ce `commutateur` et ce `proxy`. Ces flots de paquets qui transitent d'un point à un autre constituent ce qu'on appelle le trafic.

La notion de paquets Chaque fois qu'on demande une information sur le réseau (que ce soit en relation avec Internet ou localement), des paquets sont envoyés à travers les différents composants de celui-ci. Un paquet est une information qui contient un corps et des entêtes. Le corps c'est l'information qu'on envoie et les entêtes ce sont les informations liées à la distribution du paquet (à qui, comment, et aussi à qui envoyer la réponse (donc l'expéditeur)). En lisant l'entête d'un paquet qui transite sur le réseau, on peut donc savoir d'où il vient, où il va, et comment il a été envoyé.

Ce « comment » se résume en trois questions :

1. Est-ce un paquet `I.P.` ?
2. Quel protocole utilise-t-il ?
3. Quels sont les ports qui ont été définis pour chacune des deux machines pour qu'il soit envoyé et bien reçu ?

Ces informations sont toutes les trois directement fournies en dur dans les entêtes, et les expéditeurs et destinataires sont pour leur part identifiés à l'aide de leur adresse `I.P.` (qui sert d'identifiant unique des machines sur le réseau local).

Le moyen de les capturer Il existe deux façons d'intercepter les paquets d'un réseau filaire : le mode coupure et le mode transparent. Le mode coupure est celui qui a été utilisé lors du premier essai avec la PSM Box : on place une machine entre deux équipements et elle transmet les informations de l'un à l'autre après avoir copié les informations intéressantes. Ce mode est dangereux car il risque d'influer directement sur le réseau. Le cas de la mairie est flagrant, filtrer les paquets ainsi ne peut pas convenir dans une architecture complexe. De plus, si la machine d'analyse venait à avoir un problème logiciel ou physique, tout le réseau du client serait alors paralysé.

Le mode transparent, quant à lui, a pour objectif de se greffer sur un réseau sans aucun risque de le perturber. Cette méthode d'analyse est possible grâce à un système de *`port mirroir`* que proposent les `commutateurs` récents. Ce `port` renverra une copie brute de tous les paquets que le

commutateur reçoit et relaie. Ainsi, on récupère toutes les informations d'un commutateur plutôt que simplement celles qui transitent entre deux équipements, et il n'y a plus aucun risque de venir perturber le réseau.

Malheureusement, les équipemenents trop anciens ne disposent pas de cette fonctionnalité, et dans beaucoup de cas les responsables techniques ne sont pas capables de déterminer eux-même si leur matériel est adapté. L'idée a alors été de faire l'acquisition d'un switch[30] qui offre la possibilité de configurer un *port miroir* et de l'utiliser systèmatiquement chez le client. Un switch est un matériel très simple et fiable, et ne risque pas de perturber le réseau. Avec la configuration de vlan[31], il est possible de regrouper plusieurs switchs en un seul et d'observer encore plus de trafic par le *port mirroir*. Un switch ving-quatre ports de la marque Xyzell a alors été choisi.

3.2.3 La collecte et l'exploitation

Ntop, un choix approprié ? La mission de mon stage apparait alors en deux parties bien distinctes : la récupération des données (la capture des paquets) et leur utilisation pour en extraire les informations et ordonner celles-ci afin de pouvoir tirer des observations claires et synthétisées de leur analyse.

Avec Ntop, les deux rôles sont remplis, il s'occupe à lui tout-seul de capturer les paquets et propose cette pre-

[30] Matériel chargé de mettre en relation plusieurs postes d'un même sous-réseau.

[31] Définition de plusieurs sous-réseaux différents au sein d'un même switch.

mière analyse sous forme d'interface *web*. Je découvrirai assez rapidement que Ntop donne des informations sur beaucoup de choses, qu'il est capable de dire précisemment le taux d'utilisation par protocoles ou applications identifiés pour chacune des adresses `I.P.` rencontrées, mais qu'il ne donne pourtant aucune information sur la provenance des paquets, ni même de leur destination. On sait ce que chacun reçoit, ce que chacun envoie, mais on ne sait rien des sources et destinations. Dans ces conditions, il est alors impossible de satisfaire l'exigence liée à l'obligation de pouvoir définir des zones et établir un rapport sur les interactions observées entre elles.

Après beaucoup d'heures de recherche et d'étude de Ntop, l'essai d'options ou de composants complémentaires, j'en suis arrivé à la conclusion que Ntop n'était pas destiné à l'usage que je voulais en faire et que de ce fait il m'était impossible de l'utiliser dans la mesure où il ne stockait nul part des informations liées à la direction des paquets. Chercher à modifier son comportement était vain et consommateur de temps, j'ai donc trouvé une solution alternative. La partie collecte des paquets de Ntop n'étant pas satisfaisante, il fallait que je trouve un outil qui les capture à sa place. Dans le jargon informatique on appelle ça un `sniffeur`, et je me suis finalement rabattu sur le `sniffeur` le plus brut et probablement le plus connu du monde open-source, Tcpdump.

Une alternative : Tcpdump Tcpdump est un outil en ligne de commande, qui permet de capturer les paquets du réseau et de les écrire dans un fichier binaire. Il permet ensuite de ressortir les informations des paquets, et certaines

options nous permettront de ne lui faire ressortir que les entêtes limités aux informations qui nous intéressent. De plus, son système de filtres nous permet de ne pas polluer nos analyses avec des protocoles qui peuvent se négliger comme I.C.M.P.[32], I.G.M.P.[33] ou encore A.R.P.[34]. Nous sommes alors libres de récupérer toutes les informations qui nous intéressent, et particulièrement les adresses I.P. sources et cibles de chacun des paquets.

Tcpdump remplacera donc Ntop dans sa fonction de collecte des données. Concernant la fonction d'exploitation des données de Ntop, celle-ci n'étant que sous forme d'une interface *web*, qu'il est fastidieux de récupérer les informations et que ces informations ne sont de toutes manières pas traitées par zones comme nous le souhaiterions, l'idée de Ntop est tout simplement abandonnée.

4 Les choix techniques

4.1 Préambule

Le sujet de stage étant maintenant clairement défini et les grands axes corrigés, nous pouvons passer à un exposé des choix techniques qui ont été effectués au fur et à mesure de la résolution de la problématique générale.

[32] Internet Control Message Protocol, un protocole notamment utilisé par ping.

[33] Internet Group Management Protocol, protocole notamment utilisé dans le cadre du multicast.

[34] Address resolution protocol, protocole utilisé pour associer une adresse I.P. à une adresse physique (M.A.C.).

Les grands axes pour les outils d'exploitation des données sont donc :

1. Une application qui transforme les données binaires fournies par Tcpdump en un système d'information exploitable.
2. Une autre qui traite les informations qui en ressortent pour dresser le rapport final.
3. Une interface *web* pour permettre de saisir la définition des zones, l'organisation souhaitée du rapport, les couples de zones qui nous intéressent d'observer et différents éléments comme les dates de début et de fin d'observation du réseau, ainsi que le nom du client à intégrer dans le rapport.

Pour utiliser Tcpdump afin de capturer les paquets, il nous faut aussi un système de collecte. Il faudra donc également concevoir un moyen d'automatiser la collecte des données sur le réseau, au travers d'un boîtier, qu'il soit effectivement équipé d'un système de PSM Box ou non. Nous appellerons ce boîtier la sonde[35].

4.2 La sonde

4.2.1 Debian

Les PSM Box embarquent un système Linux hérité des NS Appliance. Une multitude de services et de tâches automatisées ont été installés dessus afin de répondre aux attentes des logiciels de sécurité installés dessus.

[35] J'ai bien tenté de la faire appeler la CSonde mais CS One n'a finalement pas souhaité lui attribuer de nom particulier, celle-ci n'étant présentée que comme un outil interne.

Lorsque la PSM Box sera en mode audit, elle devra récolter les données sur le réseau en désactivant son `proxy`, son `pare-feu` et tous les services pour lesquels son usage est destiné. Une fois rentrée à la maison mère, il faudra traiter ces données, configurer les différents éléments du rapport via l'interface *web* et générer le rapport. Et puis si la PSM Box venait à redémarrer (à cause d'une coupure de courant par exemple), il faudrait qu'elle se remette d'elle-même en mode audit et qu'elle bloque les lancements des différents services avant tout. Tout ça le plus rapidement possible, puisque le temps perdu à effectuer toutes ces tâches c'est autant de paquets qui ne seront pas visibles dans les rapports.

La conclusion qu'on pourrait tirer de ce bilan est que le système des PSM ne nous est d'aucune utilité, qu'il ralentit les redémarrages, et qu'il demande une modification du système assez conséquente et dangereuse pour les autres services. C'est effectivement ce que j'ai conclu et j'ai donc proposé à David d'installer un système déchargé de toutes les contraintes liées aux PSM Box. Ce dernier a approuvé ma vision des choses, il ne restait plus qu'à choisir le système à installer.

David comme moi-même étant accoutumés à Debian plus que d'autres distributions Linux, nous sommes rapidement tombés d'accord. J'ai donc installé un système Debian sur une machine de type PSM Box, en choisissant de n'installer que les programmes fondamentaux, afin d'être à même de composer moi-même un système léger et rapide à démarrer.

Tcpdump utilise la librairie Pcap pour capturer les pa-

quets, tout comme Ntop le faisait. A l'époque de Ntop j'avais étudié une optimisation que David m'avait soumise pour accélérer cette capture. Il s'agit d'un *patch* du noyau du système Linux pour ajouter des fonctionnalités qui permettent d'être plus efficace dans la capture des données du réseau (PF_RING). J'ai donc effectué cette optimisation en recompilant le noyau avec PF_RING et j'ai recompilé la librairie Pcap pour qu'elle utilise ces nouvelles fonctionnalités plus efficaces.

Des optimisations du `B.I.O.S.`[36] et du système ont été effectuées, notamment pour que la sonde puisse démarrer sans clavier et que le gestionnaire de démarrage ne perde pas de temps à demander quel noyau Linux utiliser.

4.2.2 Création d'un service

La sonde doit répondre à plusieurs impératifs :

1. Etre automatisée : il suffit d'appuyer sur le bouton de la machine, et la capture démarre dans les secondes qui suivent.
2. Etre capable de reprendre une capture : si la machine s'éteint, la capture doit reprendre normalement et il ne doit y avoir aucune autre conséquence que la perte des quelques paquets qui n'ont pas pus être capturés pendant le redémarrage.
3. Etre capable de s'auto-analyser : régulièrement, la sonde doit vérifier que son fonctionnement est toujours correct et que son travail ne s'est pas arrêté, et

[36] Basic Input Output System, ensemble de fonctions contenu dans la mémoire morte de la carte mère d'un ordinateur lui permettant d'effectuer des opérations élémentaires lors de sa mise sous tension.

si c'est le cas, faire reprendre la capture.

L'automatisation s'est faite grâce à un service. Les services sont de petits scripts qui se lancent au démarrage de la machine, et qui ont pour propriétés de pouvoir s'arrêter ou se redémarrer sur ordre. Ce service peut donc lancer Tcpdump, l'arrêter, le redémarrer, dire si il est actif, et purger ses captures.

La reprise de la capture s'est faite grâce à Tcpdump et aux noms des fichiers de captures, qu'il demande à son lancement. Une option intéressante de Tcpdump est de permettre de redémarrer un nouveau fichier chaque fois que le fichier en cours atteint une certaine taille (nous avons exploité cette fonction en définissant une taille maximale de un giga). Tcpdump ajoute alors un numéro incrémentable à chaque nouveau fichier, en prenant pour base le nom de fichier passé en paramétre. Si Tcpdump démarre une capture avec un nom de fichier et qu'une série de fichiers ayant pour base ce même nom existe déjà, il écrasera les fichiers existants, ce qui est problématique concernant la reprise de la capture en cas de redémarrage. Le probléme s'est alors solutionné par la définition du nom de fichier à l'aide d'un timestamp (nombre de secondes depuis le premier janvier 1970), qui assure un nom de fichier différent à chaque redémarrage et qui permet une indexation alphanumérique fiable des fichiers dans le temps.

L'auto-analyse quant à elle s'est résolue par une tâche CRON[37] qui demande régulièrement au service l'état de la capture, et qui demande à ce même service de (re)démarrer

[37] Un logiciel qui effectue la même tâche dans des intervalles de temps réguliers.

la capture si elle s'est arrêtée.

4.3 Exploitation des données

4.3.1 Des bases R.R.D. pour stocker les données

A nouveau un reste de mon étude de Ntop, j'ai choisi d'utiliser des fichiers de type R.R.D.[38] pour stocker les informations lues dans les entêtes des paquets des fichiers produits par Tcpdump.

Les fichiers R.R.D. Les R.R.D. sont des fichiers qui permettent la sauvegarde de données chronologiques. Ils sont créés avec des sources qui récupérent les données lorsqu'on envoie une mise à jour au fichier, et des archives qui permettent de visualiser ces données. Les sources permettent par exemple de ne pas prendre en compte une donnée si il y eu un espace de temps trop important par rapport à la dernière donnée indiquée et que la moyenne des données pour l'intervalle de temps dans laquelle elle essaie de s'insérer ne saurait alors être représentative. Nous utiliserons les R.R.D. en mode moyenne, et nous pourrons par exemple leur indiquer que les données doivent être fournies sous forme de moyennes de cinq minutes. La mise à jour d'un fichier R.R.D. se fait en lui injectant un couple temps-valeur. Un fichier R.R.D. a une taille fixe : à sa création sa taille est indirectement définie, et il la conservera en élargissant les moyennes des valeurs les plus anciennes, qui deviendront alors de plus en plus imprécises.

[38] Round-Robin Database.

R.R.D.Tool C'est un bref aperçu des possibilités du format R.R.D., mais ce qui est vraiment intéressant c'est l'outil de gestion quoi leur est associé : R.R.D.Tool. Afin d'exploiter ces bases de données, Tobi Oetiker, a créé cet outil capable de ressortir les moyennes en fonction du temps et surtout de tracer des graphiques temporels.

Exploitation de cette technologie Les bases R.R.D. semblent donc très intéressantes pour notre projet. L'idée est alors de créer une arborescence de dossiers à deux niveaux. Un premier niveau de dossiers auront pour nom des I.P. sources. Et chacun de ces dossiers contiendra une séries de dossiers qui consitueront le second niveau, et qui auront pour nom une I.P. destination. Enfin, dans chacun de ces répertoires destination, se trouveront les bases R.R.D.. Elles auront pour nom le nom du protocole ou du port concerné par les transferts entre l'I.P. source et l'I.P. destination des dossiers dans lesquels ils se trouvent. A chaque paquet que nous traiterons, nous irons mettre à jour les bases R.R.D. concernées avec la taille du paquet et la date à laquelle il a été capturé (indiqué dans les entêtes).

Par exemple : si nous traitons un paquet dont la taille indiquée est de vingt-quatre octets, qui va de A vers B en T.C.P. et depuis le port 80 vers le port 1298 le 10 juin à 12h45, nous ajouterons le couple [10/06/08 12h45]/24 aux fichiers suivants :

1. A/B/TCP.rrd
2. A/B/80.rrd

Ainsi, en demandant à R.R.D.Tool de tracer un graphique du fichier A/B/TCP.rrd pour le 10 juin, nous constaterons

que vers 12h45 il y a eu vingt-quatre octets qui ont transité de A vers B.

A partir de ce concept, il est possible de tout déterminer. Ainsi, pour savoir quel volume de données A a reçu de B en `T.C.P.`, il suffit de consulter le fichier `R.R.D.` TCP.rrd de ce que B a envoyé à A. Et pour savoir quel volume de données au total A a envoyé à B, il suffit de demander à afficher la courbe qui réunit tous les fichiers `R.R.D.` de tous les protocoles qui sont enregistrés dans A/B/. Et puisque l'addition est possible, on peut alors travailler par zones. Puisqu'une zone est constituée d'adresses `I.P.`, de sous-réseaux et d'intervalles, il suffit de tout transformer en adresses `I.P.`, de repérer tous les fichiers `R.R.D.` du protocole qui nous intéresse qui sont dans des répertoires correspondants au motif I.P.deZoneA/I.P.deZoneB/PROTOCOLE.rrd et de les afficher en une seule courbe.

Je n'irai pas plus loin dans les explications, mais ce principe est le coeur technique de mon projet, et justifie l'exploitation du format `R.R.D.`, qui permet la souplesse que je désirais et la possibilité de tracer des graphiques de qualité simplement.

Le temps de traitement Un gros probléme rencontré a été le temps de traitement nécessaire pour passer des fichiers de Tcpdump à cette arborescence de fichiers `R.R.D.`. En effet, Tcpdump capture parfois jusqu'à dix ou vingt paquets pour une même seconde. En traitant ensuite les paquets un par un pour mettre à jour les bases `R.R.D.`, mon logiciel prenait quasiment autant de temps que la capture des données. Pour des fichiers Tcpdump résultant d'une semaine

de capture, il aurait alors fallu une semaine supplémentaire juste pour les transformer dans un format exploitable. La problématique était lourde et s'est résolue par un système de mise en cache des paquets lors du traitement des fichiers. Ce cache se base sur le fait que les graphiques présenteront une semaine de données, et que la précision sera donc relativement grossière. Inutile alors de savoir ce qui a été envoyé à chaque dixième de seconde de la semaine. Ni même à chaque seconde, ni même à chaque minute. Plutôt que de mettre à jour les `R.R.D.` à chaque paquet, le cache accumule la taille des paquets jusqu'à ce qu'un paquet posséde une date de plus de cinq minutes minimum que la date du premier paquet de l'addition en cours. A ce moment là, le fichier `R.R.D.` correspondant est mis à jour. Ainsi, à titre d'exemple, plutôt que d'indiquer au `R.R.D.` qu'il y a eu deux kilo-octets de reçus toutes les dix secondes pendant une minute, on lui dira qu'il y a eu douze kilo-octets de reçu au bout d'une minute. Comptant qu'il y a plusieurs paquets par seconde, et que la part la plus importante dans le traitement d'un paquet est le temps consacré à dire au `R.R.D.` de se mettre à jour, ce système de cache a permis de réduire le temps de traitement de plusieurs milliers de fois. De plus, puisque les `R.R.D.` peuvent faire plusieurs mises à jour à la fois, les couples temps-valeur sont accumulés jusqu'à atteindre un nombre critique, moment auquel les `R.R.D.` prennent toutes les mises à jour à la fois. Ceci limitant encore d'autant le nombre d'accès aux fichiers `R.R.D.` (et donc au disque dur).

La problématique du temps de traitement a aussi été décisive dans le choix de l'arborescence des bases `R.R.D.`. Plutôt que d'aller additionner les bases `R.R.D.` des différentes `I.P.` d'une zone, il aurait été bien plus simple de

prendre directement en considération la définition des zones et de créer une série de `R.R.D.` directement par zones (pour les sources et destinations). Cette solution avait été implémentée après la perte de toutes les données suite à l'orage, mais finalement abandonnée à cause du nombre de fichiers que cela nécessitait de mettre à jour pour chaque paquet traité, et donc du temps nécessaire qui en découlait. De plus, traiter la définition des zones a posteriori permet de pouvoir modifier les zones sans avoir à devoir prendre le temps de regénérer toute l'arborescence des `R.R.D.`.

4.3.2 Les camemberts : appellation contrôlée Java

Alors que R.R.D.Tool génére directement les graphiques temporels, il n'en est pas de même pour les graphiques de type camemberts. Plusieurs solutions ont été explorées, notamment des librairies Perl permettant de créer des camemberts facilement. Mais chaque fois le constat est le même : technologiquement la solution est intéressante, mais les camemberts sont esthétiquement très laids. Ayant pour objectif personnel de concevoir des rapports agréables à consulter, j'ai cherché une solution me permettant de générer des camemberts un minimum stylisés.

Mon choix s'est finalement arrêté sur une librairie Java qui permet de faire de jolis camemberts en trois dimensions et avec de la transparence. A défaut de plus simple, j'ai donc conçu une petite application Java en ligne de commandes qui généres des camemberts en `PNG`[39] à partir de paramétres permettant de le définir.

[39] Portable Network Graphics, un format ouvert d'images numériques.

Mon regret est de ne pas avoir réussi à sortir ces camemberts en vectoriel plutôt qu'en PNG (contrairement aux graphiques temporels qui le sont), ce qui permettrait une utilisation extrêmement libre des images et d'assurer au rapport de ne jamais présenter un contenu flou. Cette solution oblige également d'installer une machine virtuelle Java sur la machine. Les contraintes sont donc un peu lourdes, mais correspondent précisément à la limite que j'ai observée entre la volonté de la performance technologique et l'obligation de ne pas oublier que le client final se moque de la technique, tout en restant très regardant sur le résultat.

4.3.3 Un ciment pour le projet : Perl

Le logiciel que j'ai développé est constitué à 99% de codes Perl. Perl est un rétro-acronyme (rétro parce que sa signification initiale est plutôt d'ordre biblique) signifiant « *Practical Extraction and Report Language* » (Langage Pratique d'Extraction et de Rapport). La simple lecture des mots constituant cet acronyme indique que Perl est le langage approprié pour la mission que j'ai à effectuer. Au delà des considérations littéraires du langage, Perl est un langage système rapide, performant et facilement manipulable.

Toujours pour des problématiques de temps de traitement, il aurait pu être préférable de privilégier le C à Perl. C'est effectivement la question que je me suis posé, et nous nous sommes amusés, avec Pierre, mon collègue stagiaire, à effectuer des comparaisons de traitement entre des scripts C et Perl. Il s'est avéré que malgré que Perl soit un langage interprêté (et donc compilé à la volée), il est plus rapide dès lors que le script en question demande un minimum

de traitement en relation avec le système, notamment pour ce qui est de la gestion des fichiers. Si on considère que l'aspect interprêté permet de gagner du temps, au niveau du développement cette fois-ci, et que la simplicité du langage par rapport au langage C permet également de gagner énomément de temps, on peut conclure qu'utiliser Perl au détriment de C est un choix finalement judicieux (d'autant plus que j'étais curieux d'apprendre ce langage si souvent adulé par les administrateurs systèmes). On peut également ajouter que pour les traitement de bases, les commandes Perl utilisent souvent des fonctions écrites en ... C.

Malgré l'aspect souvent procédural des scripts Perl, j'ai choisi de traiter mon projet en programmation orientée objet. L'objectif étant d'obtenir un système souple et propre qui pourra évoluer par la suite. La construction des rapports s'effectue donc à partir d'un script Perl qui utilise toute une collection de classes Perl créées pour l'occasion. Ces classes permettent notamment de traiter les fichiers `R.R.D.` et de manipuler les zones très facilement. Un système de requêtes a été mis en place et permet en quelques lignes d'avoir toutes les informations nécessaires concernant les échanges entre deux zones selon un type de donnée précis et des paramétres souples.

4.3.4 La mise en forme du rapport : LaTeX

A ce stade, nous savons comment créer les graphiques temporels, les camemberts, et comment traiter et récupérer les données. Il reste à assembler le tout est le rendre présentable.

Nous avons vu dans les objectifs que David avait évo-

qué le format PDF. J'ai donc dans un premier temps orienté les recherches sur des librairies Perl qui permettaient de générer des documents dans ce format. Finalement, la méthode ne m'a pas semblée adaptée, et surtout dangereuse. Dangereuse parce que la génération du rapport dépendait du format `P.D.F.` et qu'il devenait alors compliqué de s'en détacher si on souhaitait un format différent (un format modifiable, type `R.T.F.`[40] par exemple). La solution s'appelle alors LaTeX.

LaTeX est un « *système logiciel de composition de documents créé par Leslie Lamport* » (source Wikipédia), en 1985. Concrétement, il permet de concevoir des documents mis en forme en se passant d'outils tels que OpenOffice.org-Writer ou encore Microsoft Word, directement depuis un simple éditeur de texte brut. La principe est alors de se concentrer sur la structure logique du document plutôt que sa structure esthétique. Ainsi, plutôt que de décider de mettre une ligne en gras de taille seize pour indiquer que c'est un titre, on préférera utiliser la macro-commande section qui entourera le titre. Le principe est le même que la notion de sémantique lorsqu'on crée des documents `H.T.M.L.`[41] : on préférera utiliser des balises de titre plutôt que des balises de mises en forme (pour des problématiques d'accessibilité). Ici la problématique n'est pas l'accessibilité mais l'universalité. Ainsi, en disant à LaTeX quels sont les titres et en lui indiquant dans le préambule comment les présenter, il pourra convertir ce qu'on lui a demandé d'une façon universelle en une façon spécifique au format de sortie du fichier qu'on lui demandera. Ce format est indiqué lors de la compilation

[40] Rich Text Format, un format de fichier descriptif non compressé est reconnu par la plupart des logiciels de traitement de texte.

[41] HyperText Markup Language, langage utilisé pour la création de pages Internet et interprêté par des navigateurs.

du document LaTeX à l'aide des outils qui sont disponibles. Le langage passe par un format D.V.I.[42] qui est indépendant du système, avant de le convertir en P.D.F., R.T.F., H.T.M.L., et d'autres. L'avantage est également le même que l'utilisation de CSS[43] dans une page Internet, à savoir que la possibilité d'indiquer dans le préambule du document le style d'un élément particulier offre le loisir de partager ce préambule entre différentes sources LaTeX et ainsi obtenir un ensemble homogéne et dont la charte graphique est simple à mettre à jour.

Dans notre cas, LaTeX nous offre deux avantages :

1. Inutile d'utiliser des librairies, il suffit de créer un fichier texte qui sera compilable par LaTeX.
2. Inutile de se soucier du format de sortie : LaTeX nous offre la liberté de le choisir une fois que le rapport est édité.

Un autre avantage, qui prouve la puissance du langage, c'est que la table des matières par exemple, se crée en un appel de commande, en fonction des titres et sous-titres définis dans le document. J'ai d'ailleurs été tellement convaincu par ce langage (que je connaissais mais que je n'avais jamais réellement utilisé) que j'ai réalisé toutes mes documentations en LaTeX ainsi que le rapport que vous avez actuellement sous les yeux et que je suis en train de taper

[42] DeVice-Independent, c'est un format de fichier ouvert utilisé par le système de composition de texte TeX qui posséde la faculté de pouvoir être imprimé sur presque n'importe quel type d'appareil de sortie typographique.

[43] Cascading Style Sheets, langage de mise en forme utilisé pour la conception de pages Internet.

sous `vi`[44].

5 La communication

5.1 La documentation

Un des aspects les plus importants et pourtant totalement absent de notre formation d'étudiant est peut-être bien la documentation. A notre arrivée, notre première mission a été d'installer un système de NS Appliance, de la récupération sur un serveur FTP distant jusqu'à la configuration du `pare-feu` avec Iptables. Une documentation claire et précise a pu nous guider, fruit du travail de nos prédécesseurs.

Malheureusement, il n'en a pas été de même pour l'ensemble des scripts de la NS Appliance, rendue ainsi difficile à exploiter. C'est ainsi que Pierre, par exemple, qui travaillait dessus, a perdu énormément de temps à modifier l'existant, faute d'instructions écrites sur le fonctionnement du système. Pour ma part, le système de rapports déjà existant aurait pu être exploité pour éviter de tout réécrire d'un bout à l'autre. Or je ne l'ai pas fait, considérant que le temps consacré à la compréhension du code existant aurait été supérieur au temps pris pour refaire ce qui existe déjà, en en profitant pour mieux l'adapter à mon système global.

Ainsi, une documentation minitieuse compléte précise et fournie permet aux développeurs qui nous succédent de gagner du temps en ne cherchant pas à comprendre les scripts,

[44] Editeur de texte présent d'office sur la majorité des distributions Unix.

et d'une façon plus concréte, de pouvoir tout simplement réexploiter ce code déjà existant. C'est pourquoi David a fortement insisté pour que nous tenions des documentations à jour sur tout ce que nous faisions. La documentation signifiant aussi bien des commentaires généreux intraséques au code que des documents externes qui expliquent le mécanisme de certains algorithmes.

Il m'a été reproché durant ce stage de ne pas avoir été à même de produire la documentation régulièrement, et de ne pas avoir toujours été très pertinent dans le choix du nom de mes fonctions et variables. Bien que j'ai finalement produit des documentations sur l'ensemble du système, un diagramme de classes ainsi que des organigrammes (cf. Annexes A et B) sur différents points sensibles du code, il aurait été souhaitable que je m'y prenne plus tôt, afin de prouver que j'en étais capable et d'avoir le temps de produire des documents exempts d'oublis. Mesurant bien l'importance de cette documentation et la possibilité que le fruit de plusieurs mois de travail puisse finalement être relégué et refait, faute de compréhension, j'ai finalement bien compris que j'avais tout intérêt à respecter cet impératif, et on ne m'y reprendra plus à le laisser en attente indéfiniment.

5.2 La vie en entreprise

Notre maître de stage, David, étant basé sur Rennes, nous ne l'avons vu que très peu physiquement. Malgré cela, un téléphone et un courriel se tenait à notre disposition en cas de problémes. David faisait régulièrement le point avec nous par téléphone et chaque début de semaine nous lui faisions un compte-rendu par courriel du travail ou des

recherches que nous avions effectués la semaine passée. En cas de probléme technique, la puissance des réseaux lui permettait de résoudre nos problémes à distance. Le métier d'informaticien est un métier compatible avec la vie de nomade et je donne raison aux publicités IBM qui se demandent si se déplacer est réellement utile. Ainsi, le stage s'est parfaitement déroulé dans ces conditions, sans que nous ne nous sentions lésés par cette absence. David était tout de même présent certaines semaines.

CS One nous aura finalement fait très vite confiance. Tant au niveau sécurité qu'au niveau responsabilité. Ainsi, l'entreprise s'était engagée auprès de la mairie K à leur refournir un audit dans les mois qui venaient, alors que je n'avais encore jamais produit de résultat concret de mon travail. Au niveau sécurité, on nous a très vite confié un badge pour déverrouiller la porte d'entrée, puis la clé de cette porte et enfin la clé de la porte du batîment. Nous pouvions ainsi rester éventuellement le midi, ou certains jours entre stagiaires quand personne ne travaillait sur place. A noter tout de même la présence d'une caméra de surveillance au sommet de la pièce principale qui pouvait être consultée à tout instant. Toutefois, cette liberté fut plaisante et agréable à vivre.

Une expérience intéressante de la vie d'entreprise, et particulièrement sur le sujet de la communication, aura été un accident mineur survenu avec Raphaël, le responsable. Alors que je préparais mon rapport de stage, je tenais à avoir le plus d'informations possibles sur CS One comme sur ARESSI, pour laquelle l'histoire ne peut pas être complétement passée sous silence. J'ai ainsi posé énormément de questions à David au sujet de CS One, et je me suis

tourné vers Raphaël pour les questions concernant ARESSI, je lui ai alors proposé de lui envoyer un courriel. Et bien que l'ambiance était conviviale au quotidien, je n'ai pas réussi à adapter mon courriel aux exigences protocolaires qui semblaient m'être imposées implicitement. Ainsi, réellement intéressé par l'ancienne entreprise, j'ai posé énormément de questions très précises, et le tout sur un ton très détendu, trop détendu apparemment. Raphaël a donc été surpris de mes questions qu'il a jugé trop curieuses et choqué de ce ton trop détendu donné dans le courriel. Nous nous sommes ensuite expliqués et je me suis excusé et cette histoire s'est finalement réglée. Mais je retiens pour leçon que les écrits restent et qu'ils demandent de suivre ces exigences protocolaires dues au monde de l'entreprise et au rang du responsable. La communication par courriels avec David étant beaucoup plus fréquente, il nous avait permis d'emblée un mode de communication plus détendu, et je l'en remercie. Cette expérience aura en tous cas confirmé mon désir de me tourner vers des entreprises d'une très faible structure lorsque je serai en quête d'un emploi, afin d'éviter au maximum ce genre d'exigences qui ne sont pas obligatoirement requises, et qui ne semblent pas correspondre au cadre de travail auquel j'aspire.

Nos horaires étaient fixés à 8h30 le matin et 18h00 le soir, avec deux heures de pause le midi. Je suis resté très fréquemment le soir, parfois jusqu'à 20h00 pour faire avancer mon projet, et je suis resté chaque midi pour travailler.

6 Conclusion

6.1 Mais où en est la mairie K

Au bout de plusieurs semaines de stage, la sonde était fin prête à partir en direction de la ville de K dans le but d'effectuer sa mission de collecte des informations qui circulent sur le réseau. Elle a donc été branchée à un endroit stratégique du réseau, et la configuration de `vlan` sur le `switch` que nous avons utilisé a permis de capturer un maximum d'informations. Après une semaine à écouter le réseau, elle est revenue à la maison mère ... chargée d'une quarantaine de gigas de données. La sonde a parfaitement effectué son travail, c'est une première victoire pour cette première étape de l'audit à réaliser.

Les fichiers binaires de capture des paquets créés par Tcpdump ont ensuite été analysés par un premier script de la collection Perl développée, pour devenir l'arborescence des fichiers `R.R.D.` nécessaire dans la suite du processus. Aucun problème n'a été rencontré sur cette étape de plusieurs heures, si ça n'est le temps qui reste encore un peu excessif.

L'étape suivante étant de définir les zones à définir en fonction des adresses `I.P.` du réseau, le commanditaire de chez K nous a demandé de lui fournir la liste de toutes adresses locales que nous avons rencontrés. Un second script s'est facilement chargé d'effectuer cette tâche. La liste a donc été envoyée à la mairie K pour qu'elle classe ces adresses par zone (et donc logiquement par site, dans ce cas présent), afin que nous puissions passer à la dernière de génération du rapport.

Le retard dû notamment aux pertes des données du serveur au moment de l'orage n'aura pas permis à cet audit d'arriver au terme de sa mission dans les délais du stage, CS One n'ayant pas encore reçu la configuration des zones. J'assisterai donc à la réussite (j'espère) de cette aventure dans le cadre de la prolongation du stage.

Après une analyse rapide des résultats collectés, il semblerait qu'il y ai des machines qui ne se connectent pas à Internet par le biais du `proxy` de la mairie et qu'une centaine de sites Internet pornographiques aient été visités (en ne se limitant qu'aux adresses des sites Internet qui contiennent les mots *sex* ou *porn*). Les résultats de l'audit s'avérent donc intéressants.

Un extrait de rapport d'audit de test produit est disponible en annexe C.

6.2 Un premier pas dans le monde de l'entreprise

Cette première approche concrète du métier d'informaticien aura été l'occasion de confirmer et d'infirmer certaines appréhensions.

En effet, j'ai découvert avec ce stage que je pouvais me passionner pour un projet que je ne développe pas de mon propre chef et pour lequel je suis restreint à des obligations. Cet aspect me rassure et me conforte dans mon désir de travailler dans cette branche.

Mais j'ai confirmé mon appréhension concernant l'aspect

sédentaire de ce métier, et je reste dans mon objectif de ne pas travailler en tant que développeur. Je poursuis donc mon ambition de m'orienter vers les métiers systèmes et réseaux qui sont parfois plus sociables (avec si possible une large tendance pour l'open-source), et ma poursuite dans ce domaine en licence professionnelle aura pour objectif de me faciliter l'accès aux métiers envisagés.

La vie en entreprise aura été agréable, mais elle aura confirmé mon souhait d'intégrer une petite structure sans les exigences protocolaires parfois superflues, à l'instar de ce qu'on peut observer dans certaines petites sociétés. J'espère avoir la chance d'être en mesure de choisir.

6.3 Un stage qui a répondu à mes attentes

J'ai choisi d'effectuer mon stage chez CS One pour me donner un maximum de chances de travailler avec du libre : je n'aurais manipulé que du libre. J'ai également choisi ce type de stage pour me rapprocher des applications systèmes/réseaux : mon sujet de stage n'aura été qu'en rapport avec le réseau et j'ai eu l'occasion de faire du système grâce à la sonde.

Je suis donc parfaitement satisfait d'avoir effectué mon stage de fin de D.U.T. chez CS One et particuliérement satisfait de la tâche qui m'a été confiée. Cette tâche aura évolué au fur et à mesure des semaines et, grâce aux propositions de David et sa régulière approbation face à mes propres initiatives, a pris bien plus d'ampleur que la mission qui avait été envisagée initialement.

Le projet aura d'ailleurs pris tellement d'ampleur, que CS One me propose de prolonger le stage afin d'arriver au terme du projet. La dernière évolution sera la partie *web* du rapport, à savoir un bilan complet des sites Internet qui sont consultés par chaque zone accompagné d'une étude statistique basée sur un classement des sites Internet par catégories définies à l'aide des listes que la NS Appliance utilisait déjà. L'interface *web* de configuration des rapports est déjà en chantier, mais nécessite encore du temps pour devenir compléte et fonctionnelle, c'est le second objectif de cette prolongation.

Nomenclature

A.D.S.L. Asymmetrical Digital Subscriber Line, technologie capable de transporter plusieurs mégabits par seconde sur les deux fils de cuivre du téléphone.

A.R.P. Address resolution protocol, protocole utilisé pour associer une adresse `I.P.` à une adresse physique (`M.A.C.`).

anti − phishing L'`anti-phising` prévient des courriels qui semblent ne pas provenir réellement de la personne qui tente de s'indentifier en tant qu'expéditeur.

anti − spams L'`anti-spams` intercepte des courriels jugés indésirables qui arrivent sur le réseau.

antivirus L'`antivirus` contrôle des logiciels malveillants qui pourraient arriver sur le réseau local depuis Internet, par courriel ou depuis une page Internet.

B.I.O.S. Basic Input Output System, ensemble de fonctions contenu dans la mémoire morte de la carte mère d'un ordinateur lui permettant d'effectuer des opérations élémentaires lors de sa mise sous tension.

C.R.M. Customer Relationship Management (Gestion de la Relation Client), ensemble des processus visant à gérer les activités d'avant et d'après vente au client.

commutateur Dispositif permettant d'établir ou de faire cesser des connexions (circuits) temporaires entre plusieurs points quelconques d'un réseau.

CRON Un logiciel qui effectue la même tâche dans des intervalles de temps réguliers.

CSS Cascading Style Sheets, langage de mise en forme utilisé pour la conception de pages Internet.

D.V.I. DeVice-Independent, c'est un format de fichier ouvert utilisé par le système de composition de texte TEX qui posséde la faculté de pouvoir être imprimé sur presque n'importe quel type d'appareil de sortie typographique.

F.T.P. File Transfer Protocol, protocole d'échange de fichiers.

H.T.M.L. HyperText Markup Language, langage utilisé pour la création de pages Internet et interprêté par des navigateurs.

I.C.M.P. Internet Control Message Protocol, un protocole notamment utilisé par `ping`.

I.G.M.P. Internet Group Management Protocol, protocole notamment utilisé dans le cadre du `multicast`.

I.P. Internet Protocol, protocole de transmission de l'Internet, décrivant aussi les adresses du réseau.

N.A.S. Network Attached Storage, désigne un périphérique de stockage relié à un réseau dont la principale fonction est le stockage de données en un gros volume centralisé pour des clients-réseau hétérogènes.

onduleur Appareil destiné à protéger un équipement électronique contre les baisses de tension et les micro-coupures du réseau électrique et qui dispose d'une batterie permettant le relais du réseau électrique en cas de coupure.

P.A.B.X. Private Automatic Branch eXchange, sert principalement à relier les postes téléphoniques d'un établissement (lignes internes) avec le réseau téléphonique public (lignes externes).

P.D.F. Portable Document Format, un langage de description de pages d'impression créé par Adobe Systems.

pare − feu Le `pare-feu` a pour charge de veiller à ce qu'on ne puisse pas se connecter n'importe où et n'importe comment sur le réseau local, il surveille les entrées sorties, et plus précisément les `ports`

PNG Portable Network Graphics, un format ouvert d'images numériques.

*port*80 Nom de la sortie habituelle des données à destination de Internet.

proxy Le `proxy` s'occupe de télécharger les pages Internet qui sont demandées et les renvois au commanditaire si elles ne correspondent pas à certaines catégories de pages interdites.

R.A.I.D. Redundant Array of Independent Disks.

R.R.D. Round-Robin Database.

R.T.F. Rich Text Format, un format de fichier descriptif non compressé est reconnu par la plupart des logiciels de traitement de texte.

routeur Unité qui permet d'interconnecter deux ou plusieurs réseaux.

S.D.S.L. Symetric Digital Suscriber Line, technologie qui, comme l'`A.D.S.L.`, permet d'augmenter la capacité des lignes téléphoniques traditionnelles en offrant des vitesses de transmission jusqu'à trente fois plus élevées qu'une connexion classique.

S.M.T.P. Simple Mail Transfer Protocol, protocole de communication utilisé pour transférer les courriels vers les serveurs de messagerie électronique.

S.S.H. Secure SHell, à la fois un programme informatique et un protocole de communication sécurisé.

S.S.M. Service de Sécurité Managée.

switch Matériel chargé de mettre en relation plusieurs postes d'un même sous-réseau.

T.C.P. Transmission Control Protocol, protocole de transport fiable, en mode connecté.

U.D.P. User Datagram Protocol, protocole de transport en mode non-connecté.

U.T.M. Unified Threat Management, soit « Gestion des menaces unifiées » en presque bon gaulois.

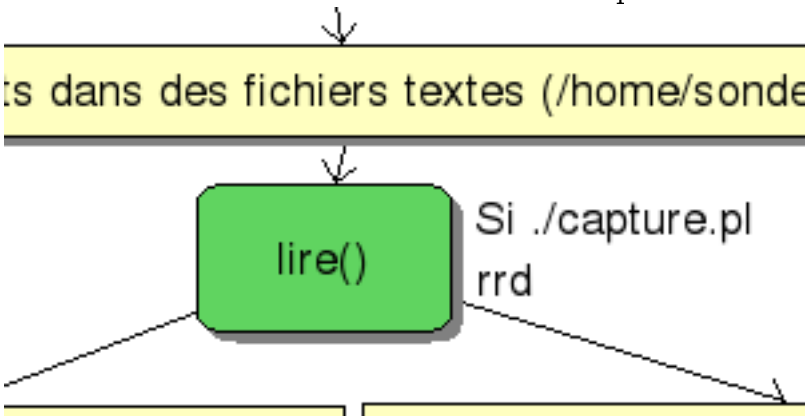
V.P.N. Virtual Private Network, architecture logicielle permettant de créer un réseau virtuel entre des machines connectées par Internet. Cela peut permettre d'utiliser Internet comme support de communication entre une société et ses filiales dans le monde, en assurant la confidentialité des échanges.

vi Editeur de texte présent d'office sur la majorité des distributions Unix.

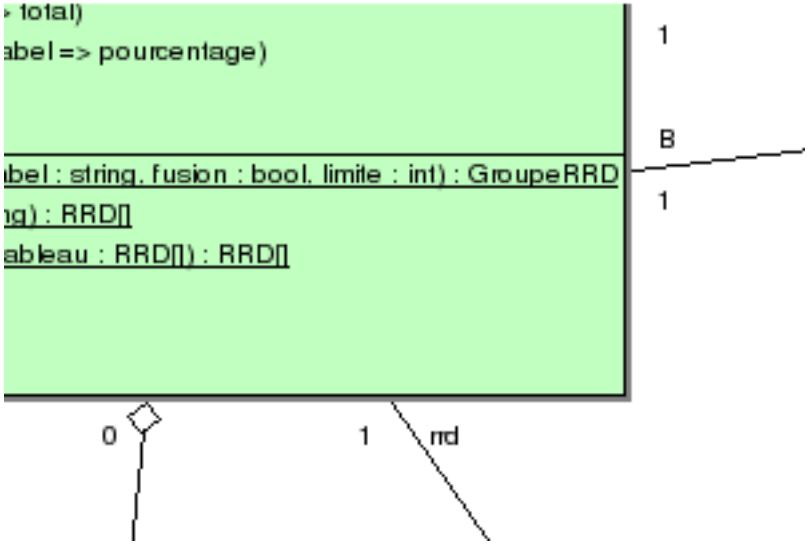
vlan Définition de plusieurs sous-réseaux différents au sein d'un même `switch`.

WWW World Wide Web, désigne ici la navigation classique par Internet.

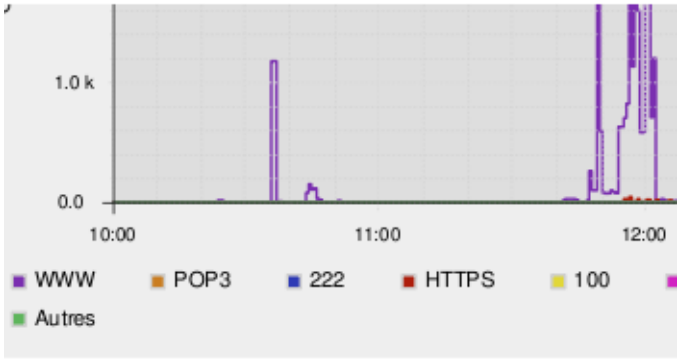
Annexe A : Extrait d'un schéma séquentiel

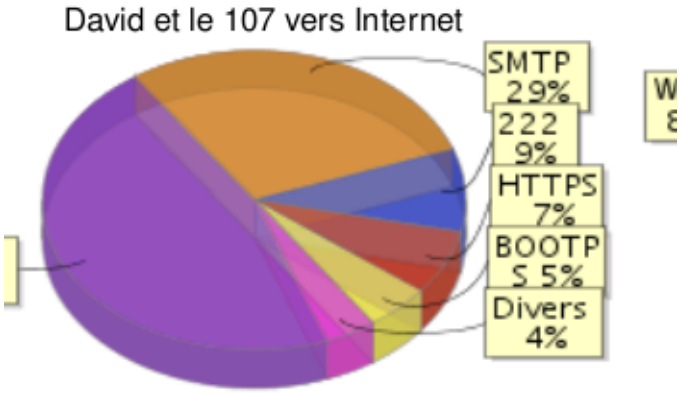


Annexe B : Extrait du diagramme de classes



Annexe C : Extrait d'un rapport d'audit





	David et le 107 vers Interne
WWW	511.25 ko
SMTP	316.83 ko
222	96.28 ko
HTTPS	72.27 ko

A tous les manchots qui tentent d'apprendre à voler.

Journal intime d'un auditeur

Editions Plume