UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ARCHITECTURE DE SÉCURITÉ POUR LES DONNÉES DANS UN CONTEXTE D'INFORMATIQUE EN NUAGE

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN INFORMATIQUE

PAR

MATHIEU SCHMITT

NOVEMBRE 2014

REMERCIEMENTS

La rédaction de ce mémoire s'est effectuée dans les meilleures conditions grâce au concours de personnes dévouées.

J'adresse particulièrement mes très sincères remerciements à mon directeur de recherche, Monsieur Guy Begin pour ses conseils précieux, son ouverture d'esprit et la grande marge de flexibilité qu'il m'a octroyée pour la réalisation de cet écrit.

Je tire ma révérence à mes parents pour leur patience, leur soutien dans les mauvais comme dans les bons moments tout au long de mes études.

Je pense à Marion, Lucas, Anthony, Benoît, Florent, Nicolas, mes co-locataires et amis à Montréal pour leur bonne humeur et convivialité.

À toute l'équipe pédagogique au sein de l'UQAM pour leur amabilité et services rendus.

TABLE DES MATIÈRES

TABLE DES FIGURES

LISTE DES TABLEAUX

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

AAA	Authentication, Authorization & Accounting. 26
API	Application Programming Interface. 70
BGP	Border Gateway Protocol. 14
CDN	Content Delivery Network. 14
CMMI	Capability Maturity Model + Integration. 44
COBIT	Control Objectives for Information and related Technology. 44
CP	Cloud Provider. Synonyme : Fournisseur d'informatique en nuage, fournisseur de services en nuage, fournisseur de cloud. 39
CSB	Cloud Security Broker. 83
DHT	Distributed Hash Table. 19
DSI	Direction des Systèmes d'Information. 103
DVD	Digital Versatile Disc. 55
FAI	Fournisseur d'Acces Internet. 14
FCoE	Fibre Channel over Ethernet. 13
HAIL	High Assurance & Integrity Layer. 64
HTTP	Hypertext Transfer Protocol. 16, 90
HTTPS	Hypertext Transfer Protocol Secure. 37, 80
IaaS	Infrastructure As A Service. 6

IEEE	Institute of Electrical and Electronics Engineers. 19
IGP	Interior Gateway Protocol. 14
ISAE3402	International Standards for Assurance Engagements No. 3402. 43, 81
IT	Information Technology. 8
ITIL	Information Technology Infrastructure Library. 43, 84
MIME	Multipurpose Internet Mail Extensions. 79
NAS	Network Attached Storage. 13
NFS	Network File System. 13
NIST	National Institute of Standards and Technology. 5
P2P	Peer-to-peer ou pair-à-pair. 16
PaaS	Plateforme As A Service. 6
PDP	Provable Data Possession. 62
PoR	Proof of Retrievability. 62
REST	REpresentational State Transfer. 91
ROI	Return On Investment. 9
ROM	Read Only Memory. 55
SaaS	Software As A Service. 6
SAN	Storage Area Network. 13
SAS70	Statement on Auditing Standards No. 70. 42
SE	Systeme d'exploitation. 49
SLA	Service Level Agreement. 1
SMB	Server Message Block. 13
SNIA	Storage Networking Industry Association. 70

SSAE16	Statement on Standards for Attestation Engagements No. 16. 44
TDC	Tiers de confiance. 67
WAN	Wide Area Network. 14
XMPP	Extensible Messaging and Presence Protocol. 20

RÉSUMÉ

L'infonuagique ou cloud computing propose différents modèles de déploiement et de distribution pour la délocalisation de l'infrastructure informatique (les applications et les données) d'une entreprise hors de ses frontières. La virtualisation et les réseaux de plus en plus performants agissent en catalyseurs de cette mouvance. L'adoption massive de ce concept n'exclut pas les dangers permanents et imprévisibles : l'entreprise perd le contrôle sur ses informations, de nombreuses questions sur la légalité ou la sécurité restent en suspens. Une multitude de menaces en provenance de divers horizons affluent et ciblent particulièrement les données. Nous considérons que l'intégrité et la confidentialité représentent les deux vecteurs concourant à la disponibilité de nos informations - crédo marketing de nos prestataires d'informatique en nuage. Une société aussi prospère soit-elle ne survit pas face à une perte de ses données.

La préservation de l'intégrité des données, en l'occurrence la propriété selon laquelle elles n'ont pas subi d'altération de manière non autorisée, sollicite des mécanismes de redondance, de correction et de détection d'erreurs performants et adaptés. Dans un contexte d'externalisation, de nouvelles méthodes émergent comme celles fondées sur la preuve de possession.

Cependant, ces mesures apparaissent comme insuffisantes. En l'absence de confiance, la sécurité est inexistante, la corollaire est tout aussi vrai. Nous nous employons dans ce mémoire à définir les facteurs tangibles et intangibles qui garantissent ce besoin de préservation. Nous argumentons de façon pertinente sur la nécessité d'un tiers de confiance et dressons une liste d'exigences qu'il est tenu de respecter.

Pour concrétiser ces mesures, nous proposons une architecture de sécurité fondée sur ce tiers. Elle se sert d'une topologie hybride composée de deux modules distincts : le module client qui place les informations sur les serveurs distants et le module en nuage. Ce dernier délivre les meilleures offres à sa clientèle en fonction de la criticité des données, monitore les SLA et les performances et audite les informations hébergées. Nous définissons également un modèle de stockage en peer-to-peer. Notre proposition élimine l'enfermement chez un fournisseur, fournit le chiffrement, écarte en conséquence la problématique sur la confidentialité ; simplifie par la même les services liés aux utilisateurs tout en délivrant les assurances attendues.

Mots-clés : Cloud computing, informatique en nuage, infonuagique, sécurité des données, tiers de confiance, courtage, peer-to-peer, architecture distribuée, intégrité, confidentialité

INTRODUCTION

0.1 Motivations

Toute entreprise qui ambitionne de s'implanter durablement sur le marché actuel est tenue d'assurer la bonne cohésion interne de ses processus. Nous trouvons dans les enceintes physiques d'une société des équipements performants tels que des serveurs informatiques, des réseaux LAN etc. qui traitent les informations nécessaires à la pérennité de sa mission. Statistiquement on constate que :

> *« 6 % des entreprises qui subissent une perte de données survivent, 43 % ferment définitivement et 51 % périclitent dans les deux années qui suivent. »*(2).

Malheureusement, ce cas de figure se présente très souvent. Les services informatiques ne sont pas toujours armés de toute l'expertise et de toute la compétence que requièrent la mise en place et le maintien d'une architecture de stockage efficace et sécurisée. Le cloud computing tombe à point nommé comme option pour résoudre cette problématique : ce concept simple propose - par essence - une externalisation des informations et des services d'une entreprise hors de ses propres limites dans des entrepôts de taille imposante appelés Datacenters. Des salariés expérimentés administrent ces sites physiques ; Amazon, Google, Microsoft présentent leur solution et offrent leur soutien pour le stockage des données. Soulignons que la mutualisation de ces ressources informatiques peut aussi occasionner une réduction des coûts pour une entreprise adhérente.

0.2 Problématiques

L'enseigne qui souscrit à ces solutions en nuage perd toutefois le contrôle physique sur ses informations. Elle accepte le principe du *Service Level Agreement* (SLA) (accord de niveau de service) qui se définit comme un contrat conclu entre un client et son fournisseur de cloud. Nous évoquerons plus largement ces SLA dans notre mémoire, sachons cependant qu'ils avantagent généralement le fournisseur au détriment du consommateur et offrent exclusivement des garanties de disponibilité de services. Toute entreprise qui s'implique est légitimement en droit de chercher les réponses aux questions suivantes :

- Que fait le prestataire de mes données ? Comment protège-t-il leur confidentialité ?
- Les exploite-t-il en dehors de mon champ de contrôle ? À des fins commerciales ?
- Quels sont les mécanismes que le prestataire met en place pour me donner la garantie que mes informations ne subiront pas une destruction ou une modification de manière accidentelle ?
- Suis-je correctement assuré en cas de litige ?
- Quelles sont mes options si un fournisseur de service dépose son bilan ou prend mes informations en otage ?

Le scandale lié à la fermeture de Mégaupload en janvier 2012 ou, plus récemment, celui impliquant la NSA et son programme de surveillance mondialisé PRISM [1] mettent en exergue nos interrogations et apportent des réponses bien concrètes mais inquiétantes.

0.3 Contributions

Nous contribuons à répondre à ces problématiques comme suit :

- **Identification des facteurs qui influent sur la confiance des usagers du cloud.** La confiance constitue la pièce maîtresse dans les échanges internet effectués entre un fournisseur de services et un consommateur. Quantifier cette confiance n'est pas chose aisée ; le cloud provider (via son SLA) n'offre que des garanties de haute-disponibilité ; le client ne se satisfait pas de si peu. Nous dressons une liste des facteurs tangibles et intangibles qui - selon nous - devraient garantir un niveau de confiance acceptable ;
- **Exigences liées à l'introduction d'un troisième acteur.** L'intervention d'un tiers de confiance/courtier de données nous paraît une solution pour offrir des garanties à un client. Nous proposons que cet acteur respecte une liste d'exigences qui s'inscrit à la fois dans une démarche technologique et contractuelle.
- **Proposition d'une architecture de stockage.** Nous définissons une architecture « à la demande », sécurisée et fondée sur un tiers de confiance. Cette structure permet au client de choisir la meilleure offre de stockage en fonction de ses besoins. Nous discutons de la mise en place d'une topologie hybride. Celle-ci fait intervenir deux modules : le premier nommé courtier/tiers en nuage propose un catalogue de ressources ainsi que la surveillance des données, alors que le second, appelé courtier interne à l'entreprise, vise à chiffrer les données, les placer sur les serveurs des fournisseurs de cloud, et à gérer les identifiants.

---

1. PRISM est un nom de code désignant un programme américain de surveillance de masse des communications internet à l'échelle mondiale

0.4 Plan du mémoire

Le mémoire se compose de cinq chapitres dont nous retraçons brièvement les grandes lignes. Le chapitre 1 définit les traits qui caractérisent le cloud computing : modèle d'affaires, technologie, avantages et inconvénients. Le chapitre 2 présente les principaux paradigmes relatifs à la sécurité informatique, détermine les risques encourus liés à l'adoption du cloud computing par les entreprises, en mettant l'emphase sur les données. Le chapitre 3 détaille les mécanismes qui permettent de conserver l'intégrité des informations dans un contexte interne et externe. Le chapitre 4 s'intéresse à la problématique de la confiance des utilisateurs envers le cloud computing, souligne la nécessité de l'intervention d'un troisième acteur - le tiers de confiance - et dresse la liste des exigences que ce dernier doit respecter pour garantir cette confiance. Dans le chapitre 5, nous proposons une architecture sécuritaire fondée sur un tiers de confiance. Le dernier chapitre présente nos conclusions et perspectives d'avenir.

CHAPITRE I

DÉFINITION ET ASPECTS DU CLOUD COMPUTING

Dans ce chapitre, nous allons définir plusieurs concepts et différentes technologies liés au cloud computing. La section 1.1 fait d'abord ressortir les traits saillants de l'infonuagique. La section 1.2 aborde les raisons qui font qu'une entreprise décide de s'orienter vers une prestation de services en infonuagique, étudie la manière adoptée par cette société pour y parvenir et en analyse les freins. La section 1.3 s'étend davantage sur le concept d'informatique en nuage ; en présentant les principales technologies. La section 1.4 poursuit dans cette logique d'exploration et analyse les multiples architectures qui composent un centre de données.

1.1 Définition du concept

Un nouveau paradigme informatique appelé le cloud computing représente le « buzzword » de ces trois dernières années. Le NIST définit ce concept récent comme « l'accès via le réseau, à la demande et en libre-service à des ressources informatiques virtualisées et mutualisées » (traduction de (38)). L'informatique en nuage permet donc d'externaliser les serveurs et les services traditionnellement localisés dans l'entreprise vers un fournisseur tiers doté de la compétence et des ressources requises au maintien de l'architecture et des services.

1.1.1 Principales caractéristiques

Prenons connaissance ci-dessous des principales caractéristiques de l'informatique en nuage déterminées par le NIST dans « *The Nist definition of cloud computing* » (38) :

- **Service à la demande :** les services sont fournis au client automatiquement et sans intervention humaine ;
- **Élasticité rapide :** le stockage, la puissance computationnelle peuvent être rapidement ajustés, parfois systématiquement en fonction des besoins immédiats de chaque client ;

- **Mise en commun des ressources :** les différents utilisateurs bénéficient de ressources de serveurs alloués de manière dynamique et rapide dans le cadre d'un modèle de cloud communautaire ;
- **Résilience :** le cloud computing se doit d'apporter des mécanismes hétérogènes où les technologies présentes supportent une multitude de clients légers (ordiphones, tablettes) et de clients lourds (ordinateurs) ;
- **Paiement à l'utilisation :** les services se facturent en fonction de l'utilisation des ressources et en toute transparence pour le client et le fournisseur de services en nuage.

Historiquement, le concept du cloud n'est pas nouveau puisqu'il agrège une multitude de services et de technologies établies. Prenons par exemple IBM qui dans les années 1960 apporte déjà des services à des clients par la création du Service Bureau (60). Les années 1990 voient l'explosion de l'Internet ; l'accès aux services peut s'effectuer via le réseau (Application Service Provider). Le cloud computing distribue à présent ces services sous la dénomination *Software As A service* (SaaS) ; le concept n'utilise plus de clients lourds mais un simple navigateur internet.

1.1.2 Modèles de distribution

Le cloud computing se décline selon trois modèles de distribution connus : le IaaS, le PaaS et le SaaS.

Dans la technologie **IaaS** (Infrastructure-as-a-Service) le matériel est l'unique élément à être décentralisé. Les clients s'abstraient donc de toutes contraintes de gestion du matériel physique.

La technologie **Paas** (Plateform-as-a-Service) offre aux entreprises un environnement de développement ou l'hébergement de leurs applications. L'entreprise cliente ne se soucie plus de l'infrastructure sous-jacente.

Le modèle le plus commun est le **SaaS** (Software-as-a-Service), qui apporte aux clients un logiciel à la demande. Les consommateurs du service ne se préoccupent plus de l'architecture matérielle (installation, maintien) et logicielle (mise à jour) sous-jacentes. Remarquons que l'entreprise ne contrôle plus directement ses données qui sont hébergées en totalité chez le prestataire de services en nuage.

1.1.3 Modèles de déploiement

Les quatre modèles de distribution cités en 1.1.2 sont déployés chez le client de manière traditionnelle via :

1. un **cloud public** géré par un prestataire de services privé par exemple, Google, Amazon, etc. ;
2. un **cloud privé** prévu pour les besoins de l'entreprise. Distinguons les clouds privés externes, où un fournisseur d'informatique en nuage réserve une partie de ses serveurs à l'usage exclusif d'un client et les clouds privés internes, où l'entreprise bâtit son propre centre de données pour ses propres besoins ;
3. un **cloud hybride** où l'entreprise peut gérer des ressources internes via un cloud privé et des ressources externes via un cloud public ;
4. un **cloud communautaire** où plusieurs entreprises partagent les mêmes données et services pour une meilleure collaboration.

La figure 1.1 montre l'interaction entre les différents services et modèles de déploiement.

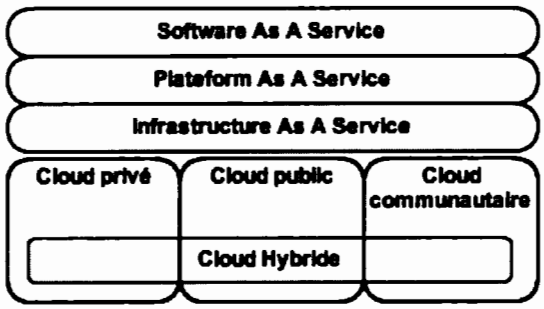


FIGURE 1.1: Modèles de déploiement et de services offerts via le cloud computing

1.2 Adoption du cloud par l'entreprise

1.2.1 Pourquoi migrer ?

Selon le cycle des tendances [1] élaboré par Gartner [2], le cloud computing entre en phase de désillusion et atteindra un plateau de productivité entre 2015 et 2020.

---

1. Le cycle des tendances ou « Hype Cycle » permet d'évaluer la maturité d'une technologie.

2. Voir : Gartner : Graphique des technologies émergentes [En ligne]. Disponible `http://blogs-images.forbes.com/gartnergroup/files/2012/09/2012Emerging-Technologies-Graphic4.gif` 2012. [Consulté le 19 novembre 2012]

Toutes les technologies passent inévitablement par cette phase de désenchantement ; l'explication réside dans le fait que les médias spécialisés délaissent un sujet devenu inintéressant. D'après le journal spécialisé ZDnet.fr, les industriels du secteur doivent fournir des réponses concrètes aux interrogations légitimes du client en terme de responsabilité et de processus interne, d'où cette phase de désillusion [3]. Dans leur étude rendue publique, Gartner affirme que le taux de croissance annuel moyen de 2011 à 2016 est et restera de l'ordre de 42 % pour l'IaaS et 18 % pour le SaaS (22), ce qui prouve effectivement que ces deux modèles de distribution représentent les catalyseurs du marché du cloud computing. Les entreprises, on le voit, investissent de plus en plus dans des solutions de services en nuage. Les attentes des petites et moyennes entreprises (PME) diffèrent cependant de celles des grandes structures : les PME migrent leur infrastructure en nuage car elles ne possèdent pas l'expertise adéquate pour maintenir efficacement un parc informatique. Elles nécessitent des conseils appropriés alors que les grandes entreprises ciblent de préférence une réduction du coût de leur IT.

En 2011, l'entreprise Forrester [4] montre que le stockage de données dans le cloud est moins onéreux qu'une solution en entreprise (47). De prime abord, l'auteur affirme que 100 Tb de données en interne coûtent à l'entreprise 400 000 $, sans compter la redondance des données, le coût d'administration, les coûts en infrastructure et en électricité. Une entreprise qui achète du stockage en interne doit prévoir un montant suffisant pour sa maintenance. De plus, les directeurs des systèmes d'information ont tendance à évaluer très approximativement le coût total du stockage, l'énergie requise étant peu quantifiable. Le niveau de redondance et le coût de migration représentent des paramètres à prendre également en considération. De surcroît, la mise en place obligatoire d'un plan de stockage qui permet de conserver des données peu exploitées ou des archives augmenterait sérieusement ce montant.

Toujours selon ce rapport, la facture finale pour une solution de stockage interne à l'entreprise s'élève donc à environ un million de $. Dans une solution en cloud, le prix est divisé par quatre pour atteindre environ 250 000 $! En conséquence, le stockage brut des données interne à l'entreprise qui affiche actuellement un montant de 400 000 $ baisse considérablement pour atteindre la somme de 12 000 $ avec la solution en nuage. Une redondance de base en cloud est

---

3. Voir : zdnet : Cloud Computing : c'est la désillusion [En ligne]. Disponible `http://www.zdnet.fr/actualites/cloud-computing-c-est-la-desillusion-39782366.htm` 2012. [Consulté le 19 juillet 2013]

4. D'après Wikipédia, « Forrester a pour rôle de fournir à ses clients des études de marché sur l'impact des nouvelles technologies dans le monde des affaires ».

proposée excluant un surplus de dépenses tel que le coût administratif, les frais en électricité, en infrastructure etc. Cette diminution est rendue possible du fait que toutes les ressources d'un fournisseur d'informatique en nuage sont mises en commun.

L'organisme KPMG confirme cette réduction de prix très avantageuse pour les entreprises. D'après leur étude statistique (29), 60 % des 180 répondants affirment que la principale raison à la migration en nuage de leurs services est la réduction de coût.

D'autres facteurs peuvent aussi être à l'origine d'une décision de migration : la rapidité de migrer, la transformation de la chaîne de valeur, la simplification et l'alignement des interactions des clients avec leurs partenaires d'affaires. Nous venons de définir les principales tendances de migration ; voyons à présent de quelle manière la migration peut se faire.

1.2.2 Quels indicateurs utiliser pour la migration ?

Nous assistons à un acharnement médiatique autour du concept d'informatique en nuage : en effet, les médias définissent le cloud computing de façon comme *la* solution des difficultés informatiques d'une entreprise. Nous devons faire abstraction de cette exaltation médiatique en mettant notamment en place des outils capables d'évaluer l'intérêt des services en nuage pour une entreprise. Par exemple, un retour sur investissement (ROI) peut être calculé. Il se fonde sur le coût d'investissement et sur le rendement de la future solution et permet de vérifier que l'entreprise fait le nécessaire pour atteindre ses objectifs de qualité.

D'autres stratégies de migration vers le cloud nous sont proposées telles que Real Option, l'analyse de la valeur, etc (1). Nous énumérons cependant ci-dessous les huit facteurs clés qui garantissent le succès du cloud dans une entreprise (50) :

- Choisir le modèle adapté en fonction des besoins de l'entreprise ;
- Évaluer le modèle de services pour acceptation ;
- Réaliser un audit de sécurité du système actuel et une analyse de sécurité du modèle de cloud ;
- Choisir le fournisseur adapté aux besoins avec un background solide ;
- Vérifier si le fournisseur applique les lois ;
- Négocier le *Service Level Agreement* (étudié en 2.2.2) ;
- Calculer le détail des bénéfices ;
- Se préparer au « big bang » à l'intérieur de toute l'entreprise (politique du changement, formation des utilisateurs, etc.).

1.2.3 Freins à l'adoption

Toutes les revues de littérature sur le sujet (12; 19; 26; 25) s'entendent pour dire que la sécurité reste la principale préoccupation de migration vers le cloud. Comme mentionné en 1.1.2, les données de l'entreprise sont déportées vers des datacenters hors de la société ou de chez soi. Les clients font parfois preuve de réticence vis-à-vis de cette migration, marqué par la disparition soudaine du contact physique avec leurs données (clé USB, disque dur, bande magnétique, etc.).

Le pare-feu jouait traditionnellement le rôle du garant dans la protection des données en assurant une sécurité à la périphérie de l'entreprise. Cette frontière n'existe plus, les clients qui confient leur données à l'extérieur sont désormais contraints d'accorder leur confiance à des fournisseurs en nuage. Enfin, de nombreux problèmes de confidentialité ne sont toujours pas résolus comme : la localisation des données, non communiquée au client ou les actions que le fournisseur de cloud réalise sur les données et à l'insu du dépositaire (datamining, revente à des tiers, etc.). Ainsi, la figure 1.2 illustre le niveau de contrôle de l'entreprise cliente en fonction du modèle de distribution.

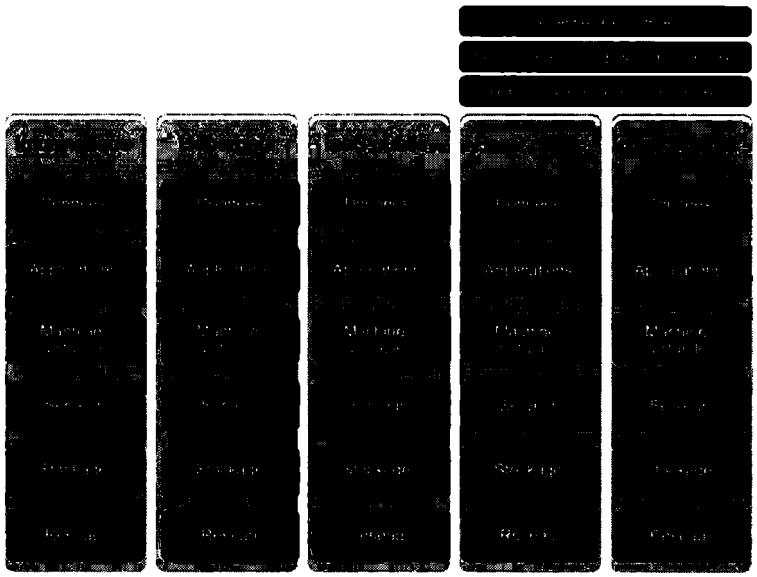


FIGURE 1.2: Niveau de contrôle de l'entreprise cliente sur son infrastructure en fonction du modèle de distribution des services (figure tirée de (24))

Lorsque le système d'information se gère en entreprise, la DSI a un monopole de contrôle sur l'ensemble de la pile, du réseau aux données en passant par les applications. Avec l'arrivée

du concept d'informatique en nuage, le niveau de contrôle de l'entreprise diminue en fonction du modèle de distribution. Le niveau de contrôle le plus élevé, effectué par la DSI, s'applique au service infrastructure qui utilise le système d'exploitation virtualisé. Le niveau de contrôle le plus bas se réalise sur les logiciels à la demande SaaS car seule une partie des données est administrée par l'entreprise cliente. De plus, le fournisseur d'informatique en nuage contrôle la totalité de l'infrastructure physique. Devant la perte partielle ou totale de la gestion de l'infrastructure, la direction des systèmes d'information gère et exploite son système d'information de manière bien moins transparente.

Par ailleurs, le manque de standard et d'interopérabilité entre les différents fournisseurs de services représentent un autre frein à l'adoption. Nous pouvons affirmer que les freins à l'adoption du cloud sont davantage psychologiques que technologiques.

1.3 Principales technologies employées dans les centres de données

Cette section présente les différentes technologies mises en œuvre dans les centres de données. La sous-section 1.3.1 présente le concept de virtualisation et la sous-section 1.3.2 illustre les notions de stockage et de distribution de contenu.

1.3.1 Virtualisation

La virtualisation permet de faire fonctionner plusieurs systèmes d'exploitation sur une seule machine physique (le plus souvent un serveur) via un hyperviseur. Un hyperviseur de type 1 (natif) fonctionne directement sur le matériel, à l'inverse d'un hyperviseur de type 2 qui se greffe sur un système d'exploitation standard. Microsoft Hyper-V [5] ou VMware vSphere ESX [6] sont des exemples d'hyperviseurs de type 1 et Oracle VM Virtualbox [7] un exemple d'hyperviseur de type 2.

Sur le site de Microsoft (40), nous constatons que les avantages de la virtualisation pour les datacenters par rapport à l'usage d'un serveur classique sont multiples :

---

5. Voir : Microsoft : Server and Cloud Platform [En ligne]. Disponible `http://www.microsoft.com/en-us/server-cloud/windows-server/server-virtualization.aspx` 2012. [Consulté le 15 mars 2013]

6. Voir : VMware : vSphere ESX and ESXi Info Center [En ligne]. Disponible `http://www.vmware.com/products/esxi-and-esx/overview` 2012. [Consulté le 15 mars 2013]

7. Voir : Oracle VirtualBox : Welcome to VirtualBox.org ! [En ligne]. Disponible `https://www.virtualbox.org/` 2012. [Consulté le 15 mars 2013]

- Le regroupement de plusieurs systèmes d'exploitation sur une même machine physique (serveur) permet de limiter la sous-utilisation des ressources de celle-ci (en comparaison d'un serveur physique qui n'héberge qu'un seul système d'exploitation). Cela suppose que les ressources de ces différents systèmes d'exploitation hébergés soient employées.
- Réduction de la consommation électrique, des besoins en climatisation, du nombre d'administrateurs, de la surface au sol utilisée.
- L'agrégation des deux points précédents permet de réaliser des économies d'argent.

La virtualisation favorise le développement durable. Les entreprises actuelles sont pourvues de postes clients à système d'exploitation installée communément appelés des clients lourds. Cette méthode manque de flexibilité et ne favorise pas par exemple une bonne itinérance du personnel, elle exclut le télétravail et l'emploi d'un parc hétérogène de périphériques. L'usage de bureaux virtuels autorise la centralisation de tous les environnements sur un même serveur ; ce procédé permet de réaliser des économies par la suppression d'un poste client gourmand en énergie. Un simple navigateur suffit pour exploiter le bureau et les applications virtualisés.

La virtualisation simplifie la sauvegarde (*snapshot*), le montage, le remplacement et la restauration d'un système d'exploitation à chaud (sans éteindre physiquement le serveur). Elle est garante — dans une certaine mesure — de la haute-disponibilité des services.

Par contre, la virtualisation procure une couche technologique supplémentaire et pose de nouveaux défis sécuritaires (49). Le principal défi qu'un hyperviseur ait à relever consiste à veiller sur les systèmes qu'il administre en les isolant les uns des autres. En effet, les machines virtuelles partagent les mêmes ressources physiques grâce à un partitionnement logique ou physique du système matériel géré par l'hyperviseur. De plus, si un rookit [8] venait à s'introduire dans un hyperviseur, il se créerait aisément un accès et pourrait observer toutes les actions générées par le système d'exploitation. Dans ce cas de figure, un attaquant pourrait s'approprier la clé de chiffrement des données (et obtenir les messages en clair), lire des mots de passe ou une portion de la mémoire RAM, etc.

1.3.2 Stockage et distribution de contenu

Quel que soit le modèle de déploiement, les centres de données nécessitent une architecture réseau qui puisse prendre en charge l'important volume de données en transit. Cette architecture

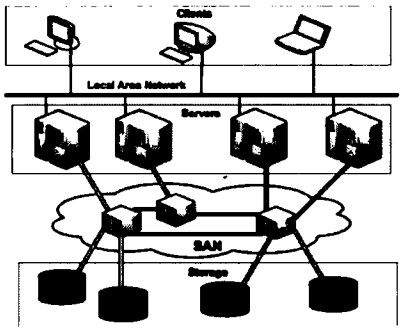


FIGURE 1.3: Topologie d'un SAN (image tirée de (18))

doit garantir des délais minimas qui assurent transitivement une haute-disponibilité des services associés [9].

Le concept de *Storage Area Network* abrégé SAN, décrit cette architecture complexe (figure 1.3) et prend ici toute son importance. Un SAN permet de mutualiser les ressources de stockage et interconnecte un ensemble de serveurs de stockage et de gestion, grâce à des protocoles réseaux comme FCoE ou Fibre Channel via des liens physiques en fibre optique (18). Ces serveurs de stockage appelés *Network Area Storage* (NAS) utilisent une multitude de technologies pour le partage des ressources (*Server Message Block* SMB, *Network File System* NFS [10]) vers les serveurs de traitement (par exemple, serveur de base de données, serveur de mails, serveur de fichiers, etc.). Du fait de la segmentation physique des rôles de traitement, de stockage, et de consommation des données, l'évolutivité de la capacité de stockage n'est plus un frein comparé à la gestion du stockage via un serveur intégrant tous ces rôles : il suffit dès lors de rajouter des disques durs dans la baie de stockage NAS. De plus, le nombre toujours plus important de données à traiter met à profit ce genre de réseau puisqu'il ne s'agit plus de traiter plusieurs fichiers sur un même disque dur mais un seul fichier réparti sur plusieurs disques. Dans un contexte d'informatique en nuage, ces serveurs de stockage sont très souvent mutualisés entre

8. Logiciel qui assure furtivement un accès, le plus souvent malveillant, en tout temps, vers un système.

9. Le délai se définit comme le temps entre l'envoi d'une requête cliente et la réponse du serveur. Plus ces temps sont longs, plus la (haute-)disponibilité du service sera impactée négativement.

10. SMB et NFS sont des protocoles réseau utilisant TCP/IP et permettant de partager des fichiers (lecture/écriture).

les divers clients. Ainsi, les données du client A peuvent se retrouver à côté (c'est-à-dire sur le même disque dur de la même baie de stockage) de celles du client B.

Les clients ne se trouvent pas nécessairement à proximité géographique du centre de données d'une entreprise qui assure un service. Pour assurer une distribution efficace des données, ces fournisseurs ont recourt à des serveurs de cache. Cette technique est connue sous le nom de *Content Delivery Network* (CDN). Ainsi, un serveur A (*origin server*) réplique ses données sur un serveur B appelé *edge server* (le serveur de cache) ; le contenu est livré au client depuis le serveur *edge* le plus proche géographiquement de ce dernier.

1.4 Considérations architecturales

1.4.1 Topologie réseau d'un datacenter

Multi-homing

Un réseau WAN nécessite les prestations d'un fournisseur d'accès internet (FAI) car les clients accèdent aux services gérés par le datacenter via ce réseau (par exemple, internet). Nous retrouvons ici un premier niveau de redondance pour assurer la disponibilité des offres, le balançage des liens qui supportent la charge et le basculement d'une connexion d'un FAI vers une autre (en cas de défaillance). Un datacenter doit souscrire à plusieurs lignes dont la gestion appartient à différents opérateurs internet pour assurer cette disponibilité [11].

Dans une topologie WAN, nous retenons deux familles de protocoles de routage :

1. Le protocole IGP concerne le routage des paquets IP au sein d'un même système autonome. Nous connaissons l'existence de nombreux protocoles de routage ouverts tel OSPF et propriétaires tel EIGRP. Un système autonome (AS) peut être perçu comme une zone gérée politiquement par un organisme (réseau universitaire) ou une grande entreprise (FAI) ;
2. Le protocole BGP permet le routage des informations entre les systèmes autonomes.

Chaque FAI administre son propre AS, il est donc tout à fait logique que les routeurs de bordure soient configurés à l'aide de ce protocole (voir la couche cœur de réseau de la figure 1.4).

---

11. Si le datacenter souscrit à une offre chez le FAI A et B et le FAI A n'assure plus le service le datacenter peut toujours assurer le sien grâce au FAI B.

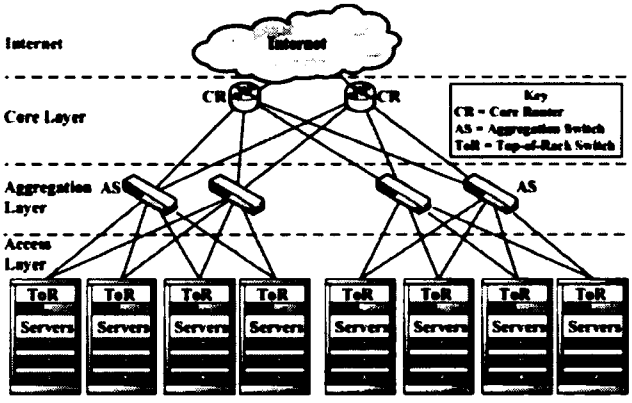


FIGURE 1.4: Une architecture réseau classique d'un datacenter (figure tirée de (5))

Topologie hiérarchique

Sur la figure 1.4, nous retrouvons le modèle hiérarchique mis de l'avant et diffusé par Cisco ; cette topologie est largement utilisée dans les datacenters.

Dans son cours CCNA (13), Cisco préconise l'utilisation d'une topologie hiérarchique en trois couches pour garantir une haute-disponibilité des services :

- La couche **cœur de réseau** « Core Layer » relie les différents segments du réseau ; les routeurs se veulent très rapides pour supporter la charge du réseau.
- La couche d'**agrégation** « Aggregation Layer » joue un triple rôle : filtrer, router et autoriser les paquets (le rôle d'un pare-feu en somme). Le premier niveau de segmentation de sécurité [12], intervient à cet endroit précis : dans un datacenter, nous pouvons segmenter un réseau par rapport au niveau de criticité, d'un client, etc. Cette couche fait la liaison entre le cœur de réseau et la couche d'accès.
- La couche d'**accès** « Access Layer » permet de connecter des commutateurs [13] aux périphériques de bout de réseau, dans le cas présent, les serveurs.

12. La segmentation implique, entre autres, la création de réseau virtuel, de domaine de diffusion, de règles de filtrage, de règles visant des objectifs de qualité de service, etc.

13. L'ensemble du travail de routage s'effectue par les deux premières couches, cette troisième couche nécessite uniquement des commutateurs de couche 2 (appelés commutateurs « top-of-rack »).

1.4.2 Architecture client / serveur

Dans le modèle client/serveur, les responsabilités de chaque partie sont nettement définies : le client sollicite des ressources sur un serveur, celui-ci répond à sa demande (35).

Le consommateur détient bien souvent une application lourde dotée d'une interface graphique pour accéder aux ressources à distance. La principale problématique qui découle de cette architecture concerne la mise à jour des applications clientes ; leur gestion est en réalité très fastidieuse (par exemple, le déploiement d'applications via un Active Directory couplé aux stratégies de groupe). Certains postes clients ne parviennent pas à effectuer ces installations en raison d'un problème de comptabilité matérielle ce qui implique la prise en charge d'une rétro-comptabilité. Ces applications lourdes complexifient considérablement la structure de l'entreprise (augmentation des coûts de l'administration système) et compliquent le maintien d'un parc hétérogène de périphériques.

Le duo cloud computing/web élimine ces préoccupations puisque :

1. les requêtes se réalisent le plus souvent via le protocole HTTP ;
2. les clients deviennent de simples navigateurs internet (par exemple, Mozilla Firefox, Google Chrome) : l'utilisation d'un seul logiciel (plutôt qu'une myriade) facilite l'administration et les mises à jour ;
3. la puissance computationnelle est déportée du côté serveur.

Certaines architectures client/serveur sont présentes sur plusieurs niveaux. La structure 3-niveaux par exemple se déploie sur trois briques dans un contexte d'application web et de SaaS, voyons le fonctionnement : les clients (première brique) envoient une requête (pas nécessairement de manière synchrone) à un serveur front-end (par exemple, un serveur HTTP Apache, la deuxième brique dans la figure 1.5) appelé fréquemment *middleware*. Ce serveur transmet la requête à un serveur de base de données (la troisième brique). Tour à tour, chacun de ces modules joue le double rôle de client et de serveur (exceptés les clients A/B/C qui sont toujours clients). Pour une meilleure compréhension, la figure 1.5 illustre ce principe.

Il est possible d'ajouter des briques en fonction de la segmentation de responsabilité à obtenir ; dès lors il faut parler d'architecture n-niveau. Notons toutefois que cette segmentation ne nécessite pas obligatoirement l'instauration d'un nouveau serveur physique ou virtualisé pour chaque brique. En effet, il est tout à fait possible d'avoir le service de serveur HTTP installé sur le même serveur — physique ou virtualisé — qui héberge la base de données c'est-à-dire que les requêtes ne transitent pas obligatoirement par le réseau. Cette architecture performante est pourtant confrontée au problème majeur de la centralisation des ressources sur un même serveur. Concrètement, si le serveur n'était plus disponible ou subissait trop de requêtes simultanément, le service ou les données seraient inexploitables. L'architecture pair-à-pair (P2P) que nous allons aborder évite les problèmes dus à la centralisation.

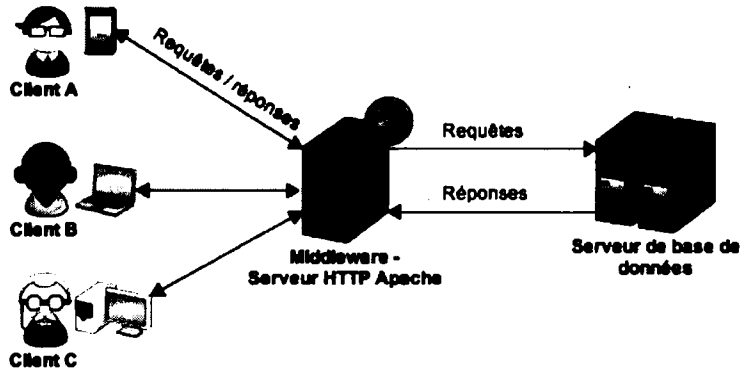


FIGURE 1.5: Architecture à 3 niveaux (3-tiers)

1.4.3 Architecture pair-à-pair

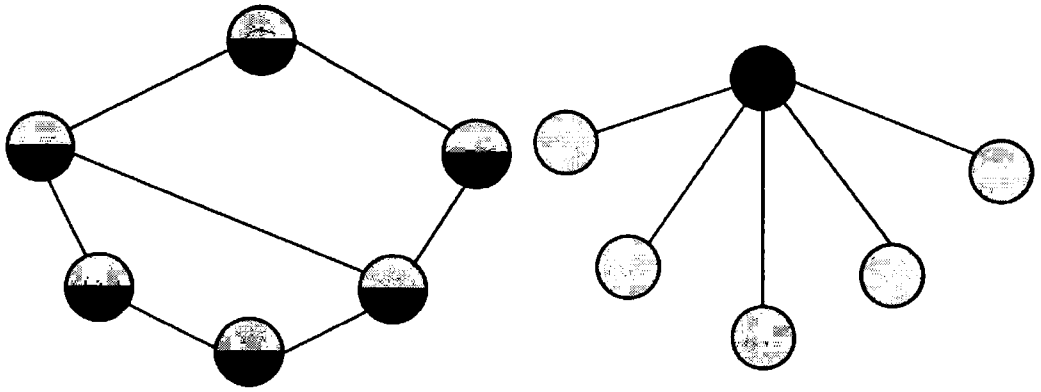
Contrairement à l'architecture client/serveur, dans une structure de type P2P les données et les ressources se répartissent sur plusieurs hôtes. Milojicic & al. (41) définissent le système P2P de la manière suivante : « Le modèle P2P se réfère à une classe de systèmes et d'applications qui emploie les ressources de manière distribuée pour effectuer une fonction de manière décentralisée (puissance de calcul, partage de données etc.) ». Ce modèle possède toutes les qualités pour en faire un partenaire privilégié dans les applications de partage de fichiers, de calculs distribués, ou de systèmes de fichiers répartis.

A l'opposé du modèle client/serveur, dans un système P2P chaque hôte/noeud joue le rôle soit de client soit de serveur. La méthode pair-à-pair se fixe trois objectifs (41) :

- **Réduction du prix de partage :** le coût de l'infrastructure est réparti entre tous les pairs car contrairement au modèle traditionnel client/serveur, la centralisation des ressources y est inexistante.
- **Anonymat / vie privée :** dans un modèle de type centralisé c'est-à-dire client/serveur, la totalité des ressources et des utilisations favorise une traçabilité des utilisateurs ; ne favorisant pas le respect de la vie privée. À contrario, dans un modèle P2P chaque hôte est l'égal de ses pairs ; ces derniers n'ont pas à s'identifier systématiquement à chaque ressource fournie.

- **Meilleure fiabilité / évolutivité :** la topologie d'un réseau P2P varie constamment puisque de nouveaux nœuds se connectent / déconnectent sans cesse. Par essence, un réseau de cette envergure apporte l'avantage d'une mise à l'échelle plus rapide qu'un traditionnel modèle client/serveur, où l'attribution des ressources doit s'effectuer via l'interaction d'un administrateur système.

Deux catégories d'architecture P2P sont possibles : l'une décentralisée (figure 1.6a), l'autre centralisée (figure 1.6b).



(a) Un réseau P2P de type décentralisé

(b) Un réseau P2P de type centralisé

FIGURE 1.6: Les deux types d'architecture pair-à-pair (figures tirées de (28))

Dans une architecture de type décentralisé, les nœuds se connectent de manière « *ad-hoc* » c'est-à-dire sans infrastructure qui organise le réseau. Les utilisateurs détiennent et contrôlent les données ainsi que les ressources, ce qui privilégie l'anonymat des usagers. Le réseau est très évolutif car il n'y a pas de serveur (par analogie au modèle client / serveur) qui constitue un goulot d'étranglement. Chaque hôte est à l'image de l'autre puisqu'aucun serveur centralisé ne dispose d'une vue globale de l'ensemble des pairs. Même si la taille du réseau est théoriquement infinie, l'accès aux ressources peut s'avérer complexe et engendrer des coûts supplémentaires considérables (route vers une ressource, mise à jour de la topologie) : ce type de réseau bâtit un graphe de manière aléatoire et utilise l'inondation du réseau dans le but de découvrir les services au niveau des nœuds. Ainsi, plus il y a d'hôtes sur le réseau, plus les performances sont mauvaises. SETI@Home [14] représente l'une des références les plus connues de ce type de topologie.

14. SETI@Home est un projet visant à démontrer la validité du calcul distribué et à trouver des formes de vie extraterrestre.

Dans une architecture de type centralisé, le serveur vérifie la structure du réseau et simplifie l'interaction entre les nœuds par exemple en indexant les fichiers que ces derniers se partagent. Néanmoins, dans cette solution, le serveur représente un point unique de défaillance ; son éventuelle indisponibilité à un moment donné se répercuterait sur l'ensemble du réseau. Ce serveur central implémente très fréquemment un index *Distributed Hash Table* (DHT) [15] qui s'abstrait de la sémantique des données et repose sur le concept de clé-valeur. D'après la thèse « Data Localization and Summarization Techniques in P2P Systems » (28), les problèmes de mise à l'échelle que rencontrent les architectures de type destructuré peuvent être minimisés par l'utilisation d'une table de hachage distribuée.

1.4.4 L'intercloud, un cloud de cloud

Les fournisseurs d'informatique en nuage possèdent différentes ressources en fonction de leurs secteurs d'affaires. Plus concrètement : le fournisseur A qui se spécialise dans le SaaS fera appel à davantage de puissance computationnelle pour effectuer ses traitements. Au contraire, le fournisseur B placera son cœur de métier dans l'externalisation de données ; il ne disposera pas d'un serveur très puissant mais détiendra une énorme capacité de stockage.

Les ressources computationnelles de ces deux prestataires ne sont pas infinies ; l'un pourrait donc avoir recours aux services de l'autre en cas de déficit en ressource physique. Par analogie, les réseaux cellulaires sont confrontés à l'itinérance des utilisateurs : la couverture géographique d'un opérateur téléphonique est incomplète. Nous pouvons donc utiliser dans une même région divers réseaux appartenant à différents fournisseurs de téléphonie mobile. Par ailleurs, le transport et la distribution dans le domaine de l'électricité permettent à quelques états de revendre leur surplus d'énergie aux pays limitrophes (par exemple, la France revend de l'électricité aux Belges, aux Allemands, aux Espagnols, etc.).

Quelques fournisseurs pourraient se servir d'un datacenter appartenant à la concurrence pour se fixer dans une zone géographique bien précise (contraintes de lois, distance des intervenants). L'introduction et la détermination d'une interopérabilité entre ces prestataires seront alors indispensables pour créer in fine une fédération de cloud qu'on appellera intercloud. Ainsi, l'article (8) nous montre que l'utilisation de plusieurs fournisseurs d'informatique en nuage est parfaitement possible. À ce sujet, l'organisme IEEE lance les bases d'une standardisation

15. Un DHT permet de contrôler la topologie et les données qui y résident. Il permet également de trouver un objet via sa clé dans un contexte où le nombre de nœud est important et très changeant (28).

dans leur document « Draft Standard for Intercloud Interoperability and Federation (SIIF) » (9). Cette organisation définit et justifie l'intercloud comme suit : « L'objectif de l'IEEE P2302 consiste à définir la topologie, les protocoles, les fonctionnalités et la gouvernance qui sont nécessaires au support d'une opérabilité entre cloud. Une analogie avec le système internet s'impose : dans un monde de TCP/IP et WWW, les données sont ubiquitaires et interopérables dans un réseau de réseaux comme l'internet ; dans l'univers du cloud computing, le contenu, le stockage et les ressources computationnelles sont ubiquitaires et interopérables dans un réseau de nuages appelé intercloud ». L'architecture d'un intercloud typique est illustrée à la figure 1.7 (d'après (9)).

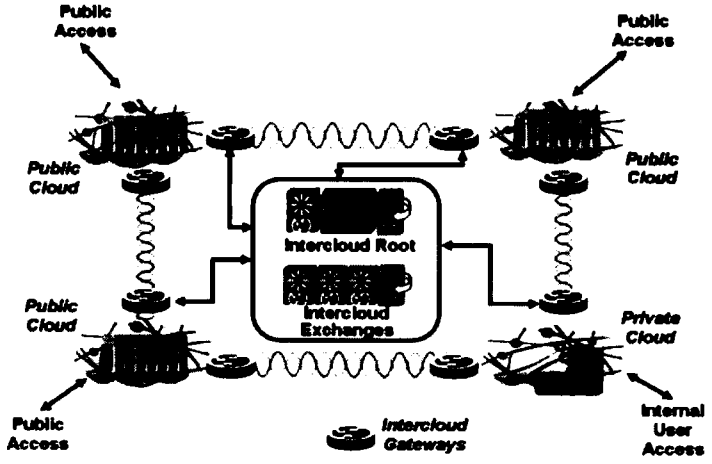


FIGURE 1.7: Topologie et éléments pour un réseau d'intercloud

L'élément « Intercloud Root » sur la figure 1.7 joue le rôle d'un fournisseur de ressources ; il héberge en effet le catalogue des ressources disponibles, des politiques et des services.

Nous qualifions de négociateur le fournisseur d'échange Intercloud (Intercloud Exchanges provider). Il simplifie les échanges et la collaboration avec un réseau hétérogène de cloud placé sous tutelle de l'élément « Intercloud root » sus-mentionné. Le protocole d'échange de l'intercloud serait l'*Extensible Messaging and Presence Protocol* (XMPP). Le réseau Intercloud fournit un environnement de confiance via différents mécanismes comme les infrastructures à clé publique ou la fédération d'identité.

Nous nous permettons à présent de porter un regard critique sur l'intercloud. Les auteurs suggèrent une architecture fédérant l'ensemble des fournisseurs de cloud ; elle les incite en effet

à adopter des standards. Cependant dans notre société capitaliste chaque entreprise tente de se démarquer. En l'état actuel des choses, le modèle proposé ne peut fonctionner puisque chaque fournisseur de cloud offre ses propres pratiques et technologies ; une collaboration entre chaque cloud provider serait extrêmement complexe. De plus, lorsqu'un fournisseur X loue des emplacements de stockage chez un fournisseur Y, un souci de non respect du contrat de service (SLA) peut apparaître si un client du CP X avait émis la condition que ses informations ne soient pas transférées dans un autre pays ou qu'elles restent chez le fournisseur X. Cette problématique n'est pas la seule à envisager dans le contexte du cloud. Dans une section ultérieure (voir 2.2) nous traiterons largement ce sujet.

1.5 Conclusion

Ce chapitre nous a permis de nous familiariser avec la définition du cloud computing et ses paradigmes, les divers modèles de déploiement et de distribution. Le recours aux technologies matures telles la virtualisation et le stockage en réseau permet de réaliser des gains substantiels ; ces procédés représentent de véritables catalyseurs de la mouvance du cloud. Grâce au débit toujours plus rapide offert par les FAI, les clients peuvent accéder à une multitude de services comme en témoigne la rapidité d'adoption de l'informatique en nuage. Bien que les bonnes pratiques de migration soient connues — via la mise en place d'un retour sur investissement ou l'identification de facteurs clés du succès — nous constatons qu'il subsiste malgré tout de nombreux freins et menaces à l'encontre d'une adoption massive du cloud computing. Il est maintenant nécessaire de porter notre réflexion sur la définition et la compréhension de ces menaces.

CHAPITRE II

MENACES GÉNÉRALES LIÉES À L'USAGE DU CLOUD COMPUTING

Ce chapitre qui poursuit notre état de l'art traite des menaces liées à l'usage de l'informatique en nuage. Dans la section 2.1 nous définissons les principaux paradigmes de sécurité et analysons leur portée dans un environnement externalisé. Nous poursuivons notre cheminement par l'analyse des considérations de localisation et de lois pour aboutir à l'étude du *Service Level Agreement*, pierre angulaire des relations client/fournisseur (2.2). La section 2.3 énumère les menaces générales de haut niveau et affiche des modèles d'attaque qui surviennent dans un environnement en nuage. Nous affinons notre recherche par la présentation des menaces liées aux données 2.4, principale problématique de notre mémoire. Nous terminons ce chapitre par l'introduction des mécanismes d'assurance existants en terme de standards, d'audits et de certifications (2.5).

2.1 Paradigmes sécuritaires liés au cloud computing

De quelle manière les acteurs du cloud computing gèrent les objectifs de la sécurité (haute-disponibilité, authentification, confidentialité, intégrité) ? Nous nous efforcerons d'élucider cette question dans cette partie et nous en débattrons amplement. Nous nous arrêterons un court instant sur certains axes de recherche. Nous examinerons les considérations légales pour aboutir en dernier lieu à la définition des principales menaces liées au concept.

2.1.1 Haute-disponibilité

Dans le livre blanc de Microsoft « Microsoft High Avaibility Overview » (39), les auteurs définissent la disponibilité en ces termes : « l'implémentation du design d'un système qui assure la continuité des activités durant une période de temps donnée ». L'expression « continuité des activités » prend une connotation subjective dès lors que les besoins d'une entreprise varient en

fonction de son cœur de métier. Les datacenters gérés par un prestataire de cloud se doivent de respecter des taux élevés de disponibilité.

L'utilisation de plusieurs mécanismes permet d'atteindre la haute disponibilité des données. Ainsi, la redondance permet une duplication de l'information qui peut ainsi rester disponible en cas de panne. Il est possible de faire de la redondance au niveau du stockage via la technologie RAID (Redundant Array of Independent/Inexpensive Disks) ; elle permet aussi d'améliorer les temps d'accès en dupliquant par exemple les données des disques durs physiques. D'autre part, le fait de répliquer un datacenter situé dans une zone géographique A vers une zone géographique B constitue également une forme de redondance. Un haut niveau de disponibilité ne s'obtient pas uniquement par le biais de l'informatique. Il est possible de faire appel à des équipements de secours comme les circuits électriques, les systèmes de climatisation ou les groupes électrogènes pour assurer la continuité du service. Ces mises en application impliquent une gestion rigoureuse du datacenter notamment une bonne gouvernance du système d'informations (élaboration d'un plan de continuité d'activités et d'un plan de reprise d'activités).

Le tableau 2.1 donne quelques exemples du pourcentage de disponibilité en fonction du temps d'indisponibilité observé par an.

Disponibilité en %	**Indisponibilité par année**
99,9 %	8,76 heures
99,99 %	52,56 minutes
99,999 %	5,26 minutes
99,9999 %	31,5 secondes

TABLE 2.1: Mesure de la disponibilité en pourcent (%)(59)

Comprenons que les pourcentages de disponibilité représentent un moteur de la concurrence puisque chaque Cloud Provider exhibe orgueilleusement ses taux dans le Service Level Agreement sans prendre la peine de s'attarder sur les autres paradigmes sécuritaires que sont la confidentialité et l'intégrité par exemple. Dans ce contexte, posons-la question suivante : Quelle est la différence fondamentale entre disponibilité et haute - disponibilité ?

Selon nous, il ne faut pas chercher la différence au niveau technologique mais dans le marketing puisque les mêmes objectifs doivent être atteints : un marchand de solutions ou un

fournisseur d'informatique en nuage capte notre attention par des adjectifs mélioratifs pour arriver à vendre un produit ou un service.

L'Uptime institute [1] délivre des certifications aux datacenters qui prétendent à de hauts taux de disponibilité. Les quatre niveaux détaillés ci-dessous ne traitent que des aspects de la topologie physique des datacenters qui influe directement sur les temps de disponibilité (51) :

- **Niveau Tiers I :** un datacenter Tiers 1 ne dispose pas de capacité de redondance (alimentation électrique et système de refroidissement) et peu ou pas de redondance matérielle. Ce niveau se note (n). Cette certification correspond généralement à un taux de disponibilité de 99,6 %, une mesure qui est considérée comme acceptable pour des applications peu critiques.
- **Niveau Tiers II :** cette certification accorde une redondance de niveau (n+1) ; en revanche, elle ne fournit pas de redondance d'ordre électrique ou du système de refroidissement. Par conséquent, le taux de disponibilité atteint environ 99,75 % ; ce pourcentage est satisfaisant pour la majorité des applications.
- **Niveau Tiers III :** ce niveau vise un taux de disponibilité de 99,98 % via la redondance (n+1) des équipements informatiques en plus d'une redondance des matériels électriques et de refroidissement. Cette certification à l'inverse des niveaux 1 et 2 n'a pas d'impact sur la disponibilité lors des opérations sur les données.
- **Niveau Tiers IV :** nous arrivons au plus haut niveau de certification avec un minimum de 99,99 % de disponibilité via la redondance [2(n+1)] des équipements (informatiques, électriques, et de refroidissement). La sagesse serait de recommander cette mesure aux applications ultra-critiques (bourses, finances, applications critiques temps réel, etc.).

Note : Nous pensons qu'il est possible de falsifier ces pourcentages de disponibilité. Démontrons-le par un scénario hypothétique : un hébergeur X possède cinq datacenters (dont quatre sont certifiés Tiers IV) et un taux de disponibilité à six neuf (99,9999 %). L'entreprise X refuse de certifier le dernier datacenter qui jouera le rôle de bouc émissaire. Par sa pratique de vente à la clientèle, l'hébergeur X, fait miroiter un taux de disponibilité à six neuf. À la survenue d'une panne l'hébergeur peut en rejeter la responsabilité sur le datacenter non-certifié, qui porte alors le blâme pour l'ensemble. La valeur de la disponibilité réelle est donc bien inférieure à la qualité du service annoncée ! Dans ce cas précis, nos propos rejoignent ceux de Monsieur

---

1. Voir : UptimeInstitute : Uptime Institute Tier Certifications [En ligne]. Disponible `http://uptimeinstitute.com/TierCertification/` 2013. [Consulté le 2 mai 2013]

Richard Stallman : *« It's [the cloud] stupidity. It's worse than stupidity : it's a marketing hype campaign.».*

- Des attaques par déni de service distribué contreviennent directement à la disponibilité des services. Ainsi, il est nécessaire d'affiner la définition de Microsoft, à savoir « l'implémentation du design d'un système qui assure la continuité des activités durant une période de temps donnée » en y ajoutant la dimension « dans les conditions normales d'utilisation ».

2.1.2 Authentification et identification

L'authentification consiste à s'assurer qu'une entité (personnes, périphériques) est bien celle dont nous parlons et que nous définissons comme telle [2]. De manière générale, des serveurs authentifient les entités pour leur accorder des autorisations (liste de contrôles d'accès) et offre ainsi une certaine forme d'imputabilité (journaux d'évènements). Ce type de service est communément appelé « AAA » c'est-à-dire *Authentication, Authorization & Accounting.*

Les serveurs authentifient les entités en se fondant sur trois types de facteurs d'authentification élémentaires :

1. **Ce qu'est cette entité :** une empreinte biométrique, un numéro identifiant unique, etc. ;
2. **Ce que possède cette entité :** une clé, un badge, une puce RFID, etc. ;
3. **Ce que connaît cette entité :** un mot de passe, NIP UQAM, etc.

L'authentification forte reconnaît une entité par l'utilisation d'au moins deux des facteurs cités précédemment (au-delà, l'utilisateur se heurte à la frontière utilisabilité versus sécurité).

Au dire des usagers, un mot de passe est contraignant pour plusieurs raisons. Microsoft offre un système d'exploitation Windows Server qui propose notamment la mise en place de GPO (Global Policies Object) via un Active Directory [3]. Ces stratégies conduisent à la création d'une politique de gestion vis-à-vis des ordinateurs et des utilisateurs. Tous les quarante jours une modification du mot de passe est nécessaire ainsi qu'une utilisation de mots de passe forts.

---

2. Pour atteindre un tel objectif, il est possible d'utiliser des identifiants formés d'un couple login/mot de passe. Par exemple, un client se voit remettre des identifiants de la part du fournisseur de cloud lorsqu'il souscrit à un de ces services.

3. Active directory est une implémentation Microsoft du protocole d'annuaire LDAP (Lightweight Directory Access Protocol) qui est un protocole de requêtage d'un service d'annuaire.

L'usager doit retenir un nouveau mot de passe à chaque renouvellement de période ; par expérience, il fera confiance à son post-it qui ne colle plus ou son papier brouillon qu'il ne retrouve plus. Par ailleurs, de nouveaux mots de passe à mémoriser affluent dans les services d'une entreprise ou sur les sites internet. Ces derniers suggèrent l'authentification unique pour la connexion (par ex. se connecter avec Google, se connecter avec Twitter, etc.). La mémoire de l'utilisateur n'est pas mise à rude épreuve puisqu'elle n'a qu'un seul mot de passe à retenir. Cependant, la sécurité des informations confiées par chaque individu sur ces différents sites et services n'est assurée que par un unique mot de passe (celui du compte Twitter ou Google). Une personne mal intentionnée mais habile peut s'emparer de ce mot de passe et accéder aux sites et services associés à ce compte.

Afin d'éviter toute confusion, nous sommes amenés à bien mesurer la différence entre une authentification et une identification. Cette dernière, comme son nom l'indique, identifie une entité, l'authentification suit le même cheminement en incluant des mécanismes supplémentaires de garantie. Ce rôle peut être assumé par un certificat électronique, qui a pour rôle d'identifier un objet (personne, périphérique). Il représente une carte d'identité numérique, contenant une clé publique, une signature et des informations d'identification sur l'objet en question.

2.1.3 Confidentialité

La confidentialité des données offre la garantie qu'aucune information dite sensible ne peut être divulguée à autrui, à un service ou un matériel non autorisé. Dans un contexte d'informatique décentralisé, les enjeux de la confidentialité sont bien plus importants puisque le datacenter qui héberge les données n'appartient pas nécessairement à l'entreprise cliente. Celle-ci doit pouvoir certifier que le cloud provider fait appel aux mécanismes d'authentification et d'autorisation adéquats car ces deux facteurs représentent l'assurance d'un certain degré de confidentialité. La pratique de la cryptographie offre des solutions pour assurer une bonne confidentialité des données. Étudions les deux grandes approches cryptographiques. La première, nommée **chiffrement symétrique**, consiste à crypter une donnée à l'aide d'un algorithme de chiffrement (par bloc ou par flux) et d'une clé. Le message crypté peut être recouvré par l'utilisation d'un algorithme de déchiffrage et de la même clé. La seconde approche, nommée **chiffrement asymétrique**, prévoit deux clés : l'une publique, l'autre privée. Le sujet A qui envisage d'envoyer un message crypté au sujet B, doit au préalable obtenir la clé publique de B [4] pour crypter le message. Le sujet B reçoit le cryptogramme et le déchiffre au moyen de sa clé privée.

---

4. La clé publique de B est, le plus souvent, obtenu via un certificat.

Le chiffrement des données présente des inconvénients pratiques : nous serions amenés à gérer autant de clés que de fichiers ou de hiérarchies de fichiers existants. Par ailleurs, la puissance computationnelle requise pour de tels chiffrements est considérable et le management de données cryptées bien plus complexe que le management de données en clair : il est nécessaire de gérer et d'utiliser ces clés en suivant de bonnes pratiques de sécurité [5]. Faisons toutefois preuve de modération dans nos propos. Une entreprise privilégiera de sécuriser ses secrets de fabrication qui lui assurent une position stratégique et s'inquiètera bien moins de la sécurité du courrier électronique de ses employés. Ainsi, si la protection par chiffrage des données est requise à la phase de stockage, il est nécessaire de savoir qui est garant de cette clé :

- Un cloud provider qui reçoit l'approbation d'une entreprise pour héberger cette clé, peut déchiffrer le contenu des données puisqu'il détient un accès physique aux salles hébergeant les serveurs. De plus, la divulgation d'informations dans un contexte de saisie juridique « *légal* », offrirait à un gouvernement l'accès aux données en clair (via cette clé). La véracité de nos propos se confirme suite à la parution d'un article publié par le quotidien Lemonde.fr qui soutient que le FBI et la NSA auraient accès aux serveurs de Google, Facebook, Microsoft, Yahoo!, etc. [6]. L'intervention du Président Obama à ce sujet est surprenante, il affirme « que le monitoring ne concerne pas les citoyens américains ». Logiquement, ces accès se rapporteraient au reste de la planète, et dans le cheminement de notre logique, puisque les entreprises étrangères (dans le sens où elles ne sont pas implantées sur le sol US) externalisent leurs données économiques dans les datacenters US, le gouvernement des États-Unis d'Amérique s'arroge un droit d'accès aux données du reste du monde !
- Un client qui possède la clé, doit pour la protéger, se doter d'une infrastructure informatique adaptée ; la conception du cloud pourtant n'abonde pas dans ce sens puisqu'elle préconise le déport de l'infrastructure informatique dans le nuage (donc transitivement, des données).

---

5. Voir : Elaine Barker & al : Recommendation for Key Management – Part 1 : General (Revision 3) [En ligne]. Disponible `http://csrc.nist.gov/publications/nistpubs/800-57/sp800-57_part1_rev3_general.pdf` 2012. [Consulté le 24 Janvier 2014]

6. Voir : LeMonde, AFP et Reuters : Le FBI aurait accès aux serveurs de Google, Facebook, Microsoft, Yahoo ! et d'autres géants d'Internet [En ligne]. Disponible `http://www.lemonde.fr/ameriques/article/2013/06/07/le-fbi-a-acces-aux-serveurs-des-geants-d-internet_3425810_3222.html` 2013. [Consulté le 7 juin 2013]

Les différents moyens d'encryption, qu'ils soient commerciaux, open-source ou intégrés, augmentent le niveau de complexité puisque la gestion des divers systèmes de cryptographie nécessite une bonne cohésion. Il est donc primordial d'assurer le stockage et l'accès aux données sur le long terme car par un effet boomerang une clé perdue ou corrompue conduit à la perte de toutes les données.

Le chiffrement des données présente un réel intérêt dans le cadre de leur transfert (de bout en bout), le déport s'effectue du client vers le datacenter et vice-versa. Par la création d'un tunnel virtuel privé (VPN) entre les deux protagonistes, la protection des informations est garantie. La suite de protocoles IPSec par exemple autorise la création d'un tunnel sécurisé qui atteste l'intégrité, la confidentialité et l'authentification des données.

2.1.4 Intégrité

L'intégrité des données est une composante majeure de notre travail de mémoire. Le chapitre 3 traite amplement de cet objectif sécuritaire.

2.2 Considérations légales

2.2.1 Localisation des données et lois

Connaître la localisation des données confiées est un droit légitime pour tout utilisateur d'un service informatique en nuage. Les sièges sociaux des plus grands acteurs du cloud (par ex. Amazon, Google, Salesforce, Microsoft, etc.) se situent aux USA. Les centres de données de ces différents protagonistes sont implantés aux USA, en Europe et au Canada. Chaque pays possède sa propre législation en matière de données ; les entreprises doivent se soumettre aux lois de l'état dans lequel leurs données résident. Le seul fait qu'un utilisateur ne sache pas systématiquement où se trouvent ses informations (dans un modèle de type cloud public ou cloud privé externe) constitue une atteinte à la confidentialité. Nous présentons ci-dessous quelques lois importantes en vigueur aux USA et nous poursuivrons par les directives européennes.

Le **USA Patriot Act** est une loi créée suite aux évènements du 11 septembre 2001 et amendée en 2005. Elle autorise l'accès au FBI (Federal Bureau of Investigation) à toutes les données avec l'accord préalable d'une juridiction spéciale. D'après le livre (19), cet Act demeure le plus controversé en ce qui concerne les aspects liés à la vie privée. En effet, la section 215 de l'USA Patriot Act affirme qu'un magistrat est accrédité pour exiger « [...] la production de tout élément tangible (incluant les livres, les enregistrements, les papiers, les documents et objets divers) dans le cadre d'une investigation sur la lutte contre le terrorisme international ou les activités clandestines [...] » (23). Comme les compagnies étrangères et les états extérieurs transfèrent leurs informations sur le sol américain, les agences gouvernementales des Etats-Unis possèdent la capacité légale et légitime d'inspecter les données hébergées pour prévenir les actes anti-terroristes. Dans ce contexte, les clients (entreprises et gouvernements) qui envisagent de faire appel au service en nuage sont face à une vraie problématique. Toujours d'après le livre (19), quelques provinces canadiennes dont la Colombie-Britannique et la Nouvelle-Ecosse ont d'ores et déjà interdit que les données de leur gouvernement soient stockées ou traitées aux USA.

Electronic Communications Privacy Management Act (ECPA) Cette directive datée de 1986 protège la confidentialité des informations dans un environnement électronique. Elle veille sur la vie privée des utilisateurs et interdit qu'une tierce personne intercepte et divulgue ces informations sans autorisation.

US Federal Information Security Act (FISMA) voit le jour en 2002 et d'après (12), cet Act « reconnaît l'importance de la protection de l'information pour les intérêts économiques et la sécurité nationale des États-Unis d'Amérique ». En conséquence, les agences fédérales américaines sont tenues d'évaluer périodiquement leurs systèmes d'information et ceux utilisés par les prestataires de services pour limiter notamment les actes de cyberguerre [7] en provenance d'autres gouvernements par exemple. Toujours d'après (12), il est vivement recommandé à ces agences d'effectuer une estimation périodique des risques, un test et une évaluation des contrôles de sécurité. Elles ont également pour mission d'organiser des formations du personnel afin d'éviter particulièrement l'ingéniérie sociale, de mettre en place un plan de continuité et de reprise d'activités et de réaliser des audits du système d'information de manière régulière. Le processus d'accréditation et de certification FISMA est définit par la NIST [8]. Quelques fournisseurs de services tels Google, Microsoft ou Amazon sont certifiés FISMA.

---

7. Une cyberguerre consiste à utiliser des outils informatiques pour créer un déni de service, un vol de données ou un sabotage contre un autre pays.

8. Voir : NIST : Guide for Applying the Risk Management Framework to Federal Information Systems [En ligne]. Disponible `http://csrc.nist.gov/publications/nistpubs/800-37-rev1/sp800-37-rev1-final.pdf` 2010. [Consulté le 24 Janvier 2014]

Health Insurance Portability and Accountability Act (HIPAA) est une loi promulguée en 1996 qui traite de la gestion des informations médicales de chaque citoyen américain. HIPAA fait ressortir le caractère privé de ces données visant donc à réguler leur usage en particulier et à limiter leur divulgation. Les points clés de cette loi concerne l'encryptage suffisant des données ainsi que l'audit du système d'information afin de garantir entre autres l'imputabilité.

Directives Européennes Le livre (19) stipule : « la différence fondamentale entre la législation européenne et américaine réside dans la notion de vie privée personnelle ». Cette dernière est considérée en Europe comme un droit humain inaliénable qu'on ne peut dissocier de la liberté personnelle. Une pratique de fouille de données sur les informations individuelles est interdite au même titre que leur transfert hors de la sphère européenne ; en effet plusieurs états qui se situent au-delà de cette frontière ne disposent pas de législations appropriées en matière de protection des données personnelles. Si nous nous référons au chapitre V des Directives Européennes (20), le transfert de données peut se réaliser sous certaines conditions : la personne à qui appartient ces informations autorise le transfert, l'objectif du déplacement vise une performance, d'un contrat par exemple, lorsque les intérêts vitaux d'un objet (personnes, entreprises, pays, etc.) sont engagés. Bien que les pratiques concernant les données soit radicalement différentes outre-atlantique, les entreprises américaines et européennes sont amenées à échanger des informations pour assurer une bonne économie de marché. C'est le rôle du Safe Harbor. Ce cadre de référence a pour rôle d'autoriser les entreprises américaines à rapatrier des données personnelles depuis l'espace économique européen - si et seulement si - elles sont certifiées Safe Harbor [9]. Dans un article paru dans le Monde (33), l'auteur Jamal Labed affirme que le Safe Harbor est une réaction américaine en trompe l'œil puisque cette sphère de sécurité n'offre aucune garantie sérieuse dans le sens où elle se base sur l'auto-certification [10] des entreprises américaines.

---

9. Voir : Wikipedia : Safe Harbor [En ligne]. Disponible `http://fr.wikipedia.org/wiki/Safe_Harbor` 2013. [Consulté le 11 juin 2013]

10. L'auto-certification est un principe qui permet à chaque entreprise de mettre en place des contrôles appropriés lui permettant de valider ou non l'ensemble des processus d'une certification donnée.

2.2.2 Service Level Agreement

Définition

D'après l'article de Pearson et Benameur (46), le SLA a pour objectif de :

- Formuler de manière explicite les devoirs à remplir par le fournisseur de cloud ; s'assurer de sa capacité à **gérer les données sensibles** et se soumettre aux lois **relatives sur la confidentialité** ;
- Préciser les responsabilités du fournisseur dans le cas d'une **manipulation incorrecte** des données sensibles ou de leur **perte** ;
- Préciser les règles de **gestion et de sécurité** des données auprès des clients du fournisseur ;
- Spécifier les régions où les données et les sauvegardes peuvent être **stockées**.

Le SLA nécessite des standards et des normes puisqu'il constitue la pierre angulaire qui garantit l'intégrité et la confidentialité des données. Néanmoins selon le rapport de la Société KPMG (29), seuls 59 % des prestataires de services en nuage proposent un SLA dans leur offre ! Nous jugeons cette pratique inacceptable.

Nous retrouvons ces SLAs sous différentes formes :

1. SLA unilatéral : la forme la plus usuelle de SLA est rédigé par le fournisseur de cloud qui favorise ses propres intérêts. Cette convention garantit uniquement un haut taux de disponibilité. Nous n'y trouverons cependant que très rarement des informations sur les mécanismes qui garantissent l'intégrité des données et leur confidentialité. Le SLA d'Amazon précise néanmoins : « Nous [...] n'effectuons aucune déclaration ou garantie d'aucune sorte, [...] que les offres de service ou celle des tierces parties sera ininterrompu, sans erreurs ou sans composants malveillants, ou qu'aucun contenu, incluant votre contenu ou celui de la tierce partie, sera sécurisé ou sans dommage ou sans perte. » [11]. Amazon affirme qu'il relève de la responsabilité de chaque client de réaliser des sauvegardes de façon régulière. Nous devons reconnaître et admettre qu'une entreprise qui prend les dispositions nécessaires pour échafauder un plan de gouvernance des informations relève du bon sens. De la sorte, le SLA mesure les obligations de chaque protagoniste.

---

11. Voir : Amazon AWS : AWS Customer Agreement (notre traduction) [En ligne]. Disponible `http://aws.amazon.com/fr/agreement/` 2012. [Consulté le 20 avril 2013]

2. SLA concerté : le SLA est rédigé suite à un travail concerté entre le client et le prestataire de cloud. Il vise à répondre aux besoins spécifiques du client. Nous apportons une remarque importante sur ce point : l'entreprise qui solidifie et rend acceptable l'accord écrit entre les deux protagonistes doit être de taille suffisante ou représenter un apport financier intéressant pour le fournisseur de cloud. Les moyennes et petites sociétés, majoritaires à travers le monde, se retrouvent sur le bord du chemin et sont contraintes d'accepter le SLA sans un droit de consultation du contenu.
3. SLA arbitré : la négociation et la gestion du SLA sont confiés à un tiers de confiance qui effectue des audits et veille au bon respect du contrat par l'ensemble des protagonistes. Ce type de SLA est écrit pour répondre aux exigences du client.

Il est clair que la taille de l'entreprise peut influer sur la capacité de négociation du SLA (58). Le remède palliatif consisterait très certainement à faire appel aux rouages de la compétition pour rendre ces SLA plus accessibles à toutes les structures, avec des accès à de meilleurs garanties de service.

SLA et sécurité des données

La sécurité des données lors de leur migration vers le cloud est primordiale. Un dépositaire de données avisé doit considérer les questions suivantes : qui a accès aux données une fois hébergées par le prestataire ? Quelles sont les mécanismes de sécurité physique mis en place et qui les gère ? Les réponses attendues sont consignées dans un SLA signé par le dépositaire et l'entreprise cliente ; cette dernière peut dès lors approuver un plan de recouvrement en cas de catastrophe naturelle par exemple.

Responsabilités du fournisseur Le client prudent dont l'objectif est de s'assurer que le fournisseur d'informatique en nuage respecte ses engagements peut faire appel à un organisme tiers qui délivre des certifications et effectue des audits externes. Les audits et les tests sont décrits dans la section 2.5. Le fournisseur prend donc à sa charge le stockage et la gestion des données en accord avec le Service Level Agreement et reste ouvert aux certifications et aux audits indépendants.

Responsabilités du client Le webcast relatif à la sécurité des données (15) décrit la nécessité pour un client de classifier ses informations en fonction de leur valeur. Si nous nous référons en effet à un bon vieil adage sur la sécurité informatique, les mesures de protection à appliquer devraient être proportionnelles à la valeur des éléments à protéger. Le client par ailleurs, identifie les exigences en matière de lois et de données et choisit in fine un fournisseur d'informatique en nuage en fonction de sa capacité à répondre à ces obligations.

2.3 Menaces générales liées au cloud computing

2.3.1 Définition des menaces

Le CSA (Cloud Security Alliance) organisation à but non lucratif, propose des bonnes pratiques sur la sécurisation du cloud computing par une énumération détaillée des concepts de sécurité et leur principe. En février 2013, cet organisme a établi une liste de neuf menaces qualifiées de sérieuses dans le cloud (26). Nous résumons cette liste par ordre décroissant d'importance :

1. **Divulgations des données :** ce point a été abordé dans la section 2.1.3.
2. **Perte de données :** une perte de données équivaut à une perte de compétitivité pour une entreprise, voire un manque à gagner. Le sujet n'est pas à prendre à la légère. Elle peut faire suite à une suppression accidentelle, une catastrophe naturelle ou un acte malveillant. Ainsi, ce risque peut être limité en mettant en place une gouvernance des données de concert avec un plan de résilience environnemental et d'équipement. Cependant ces deux moyens ne constituent pas une panacée.
3. **Vol de compte ou de trafic de services :** l'hameçonnage et la fraude correspondent à des attaques très répandues et anciennes. Elles ont vu le jour avec le commencement de l'internet et pénalisent lourdement les entreprises ou les particuliers. Un attaquant bien déterminé joue avec les informations qu'il dérobe par des suppressions, des manipulations etc. et parvient ainsi à aborder les applications critiques de l'entreprise. Des contre-mesures comme la définition des accès aux utilisateurs, la gestion des incidents, l'audit, le monitoring, la détection d'intrusion limitent le risque de vol de compte.
4. **Interfaces et API non sécurisées :** les fournisseurs d'informatique en nuage offrent de multiples services et interfaces de programmation dont la protection est rendue indispensable afin d'éviter des pertes d'intégrité, de disponibilité, de confidentialité et d'imputabilité. Ces interfaces et API sont vitales pour l'entreprise cliente et ses prestataires pour assurer une continuité des affaires. Ainsi, la mise en place de mécanismes comme l'architecture de sécurité des applications, des données ou encore la gestion des accès utilisateurs s'avère nécessaire.
5. **Déni de service :** une attaque par déni de service distribué est une activité à laquelle s'adonnent les hackers lorsqu'il s'agit de mettre un service web de l'entreprise hors ligne par exemple. Ils exploitent le plus souvent des vulnérabilités sur un serveur web, un système de gestion de base de données, etc. du fournisseur d'informatique en nuage. L'attaque sur le service en nuage utilisé par l'entreprise cliente engendre ainsi une consommation illimitée de ressources (chez le fournisseur de cloud du fait de l'élasticité de la puissance computationnelle) qui tend à ralentir tout le service.

6. **Attaque d'initiés internes :** n'importe quel employé qui figure sur la liste du personnel d'une entreprise est un attaquant potentiel : un administrateur ayant accès à des ressources critiques, un partenaire commercial ayant accès à ces mêmes ressources, etc. Des contrôles d'accès sur les clés de chiffrement éviteraient que des personnes mal intentionnées occasionnent des pertes de données. La gouvernance des données, un contrat avec les tiers associés, etc., réduisent les risques.
7. **Abus de service :** le cloud computing par essence offre une puissance computationnelle considérable et élastique qui donne à un client malveillant du cloud une véritable occasion pour lancer une attaque DDoS. La mise en application d'un SLA définit les responsabilités et une utilisation acceptable de tout à chacun.
8. **Ruée vers le cloud :** l'entreprise qui n'a pas pris les mesures adéquates pour la migration de ses services en nuage est confrontée à de nombreux problèmes (niveau de service attendu, méconnaissance des technologies, etc.). Nous avons débattu des bonnes pratiques de migration dans la section 1.2.2.
9. **Vulnérabilité des technologies partagées :** la virtualisation représente l'une des pierres fondatrices du concept de cloud computing, cependant, les entreprises qui proposent les logiciels correspondants sont peu nombreux (oligopoles). Une vulnérabilité critique dans un hyperviseur par exemple pourrait donc compromettre la sécurité des machines virtuelles chez les fournisseurs de cloud utilisant cet hyperviseur compromis.

Nous savons que les menaces les plus sérieuses concernent la sécurité des données, ce que confirme l'ENISA dans son rapport technique (12). Cette agence européenne considère la protection des données ainsi que la suppression incomplète ou non sécurisée des données comme étant critique pour les entreprises.

2.3.2 Modèles d'attaque dans le cloud

La liste ci-après précise les types d'attaques existantes en sécurité informatique (6) :

- **Modification** où un attaquant change une donnée ;
- **Interception** où un attaquant accède à une ressource de manière non-autorisé pour la copier, l'utiliser de manière illicite, etc. ;
- **Interruption** de service où un attaquant rend un élément ou un service inutilisable ;
- **Fabrication** où un attaquant crée un faux.

L'article « Attack Surfaces : A taxonomy for Attacks on cloud services » (27) de Gruscka et Jensen, définit la taxonomie et les critères de classification d'attaque dans le cloud (nommé Six attack surfaces). Trois acteurs se retrouvent dans les échanges en condition normale d'utilisation (voir la figure 2.1) :

1. L'utilisateur (le client) ;
2. Le centre de données qui fournit les ressources computationnelles nécessaires à la haute-disponibilité du service ;
3. Le service fourni.

Ainsi, la figure 2.1 illustre les conditions normales d'interaction de ces acteurs. Un attaquant potentiel dispose de plusieurs points d'attaque.

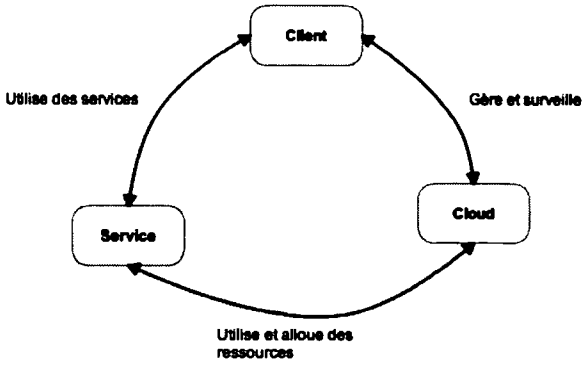


FIGURE 2.1: Interaction triangulaire entre l'utilisateur, le datacenter et le service dans des conditions normales d'utilisation

Dans la suite, nous décrivons les types d'attaque envisageables dans un environnement en nuage couplés aux interactions entre les différents protagonistes légitimes.

Attaques entre le client et le service Ce que nous appelons « l'interaction entre un client et le service » n'est autre que l'interaction traditionnelle exprimée par l'architecture Client/Serveur (vue en 1.4.2). Des attaques de type interception peuvent avoir lieu à l'instar de celle de « l'Homme du milieu » [12]. Elles permettraient à un agresseur de subtiliser un certificat utilisé entre le client et son service pour déchiffrer le contenu des échanges tout en usurpant l'identité du service en nuage.

12. En anglais « Man-in-the-middle ». C'est une attaque permettant à un individu malveillant de se placer en relais entre deux protoganistes légitimes dans le but d'intercepter leurs échanges.

Attaques entre un client et le centre de données (cloud) Dans l'entreprise cliente, la gestion et la surveillance des services sont permises au moyen du protocole HTTPS et d'un navigateur internet. Des attaques de type interruption, fabrication et modification peuvent avoir lieu. Par exemple, l'utilisation de chevaux de Troie [13] permettrait à un attaquant d'avoir un accès interne à l'entreprise. Grâce à cette ouverture, il pourrait falsifier des factures, demander l'acquisition de ressources supplémentaires au centre de données pour orchestrer d'autres attaques (par ex. une attaque DDoS sur une deuxième entreprise, etc.) ou simplement modifier des données du service utilisé tout en faussant leur surveillance pour que l'attaque passe inaperçue à la fois pour les administrateurs et les utilisateurs légitimes du service. Pire encore, un attaquant ayant un accès en interne de l'entreprise, pourrait demander des ressources illimitées au cloud (attaque sur l'élasticité). Il en résulterait une facture astronomique pour l'entreprise cliente puisque le mode de facturation de l'utilisation des services en nuage est de type paiement à l'utilisation.

Attaques entre le service et le cloud Au niveau du centre de données, des attaques de type interruption, interception et modification sont possibles. Par exemple, un attaquant qui exploite une faille dans un hyperviseur peut supprimer des machines virtuelles (VM) et affecter tous les protagonistes présents dans la figure 2.1, s'approprier les données hébergées et la clé de chiffrement, télécharger cette VM en vue d'une utilisation ultérieure, etc. Par ailleurs, un initié interne avec du pouvoir (un salarié du fournisseur de cloud avec un accès aux ressources) qui possèdent des droits d'administration pourrait supprimer une machine virtuelle par erreur et conduire le service à être interrompu.

Considérons à présent le schéma 2.2 qui illustre la configuration classique de l'utilisation de services en nuage.

Dans un environnement de ce genre, il est toujours très difficile de connaître les véritables intentions du consommateur, au même titre que celles d'un fournisseur d'informatique en nuage. Sont-elles honnêtes ou non ? Un client mal intentionné pourrait vouloir s'emparer des informations d'une société pour les revendre ou les redistribuer à des tiers. Nous considérons que les mécanismes de mise en commun des informations offerts par le prestataire en nuage doit faire l'objet de toutes les attentions en matière de sécurité. Dans la même perspective, un cloud provider malveillant peut s'approprier les données de ses entreprises clientes (hors de leur champ de contrôle) et marchander les informations à son avantage.

---

13. Un cheval de Troie est un programme contenant du code malveillant qui une fois exécuté permet d'effectuer des actions non autorisées contre l'utilisateur ciblé (suppression, modification, transfert de données, impact sur les performances de l'ordinateur visée ou du réseau contenant l'ordinateur infecté, etc.).

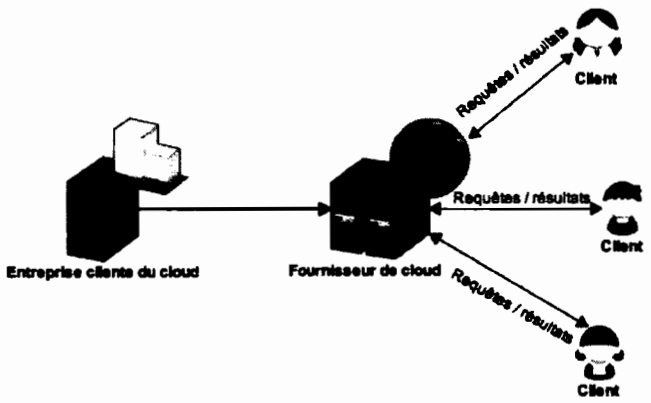


FIGURE 2.2: Configuration classique de l'utilisation de services en nuage

2.4 Menaces liées à l'externalisation des données dans le cloud computing

2.4.1 Cycle de vie et gestion des données

Nous jugeons opportun de présenter le cycle de vie tel qu'il est défini par les auteurs de la Cloud Security Alliance (25). Il comporte six étapes dont voici le résumé (ce cycle de vie est résumé par le diagramme de la figure 2.3) :

- **Création :** la création d'une donnée résulte de : la génération d'un ensemble de bits, la modification, la mise à jour ou l'altération d'un contenu existant. Nous devons savoir à qui **appartient** cette donnée et de quelle manière elle est maintenue par le prestataire d'informatique en nuage.
- **Stockage :** les étapes du stockage et de la création se réalisent simultanément. La création d'une donnée s'effectue sur un support de conservation quelconque. Notons qu'une donnée peut être restockée sans forcément passer par une nouvelle création de celle-ci. De nombreuses problématiques relatives au stockage restent en suspens : elles concernent la façon dont l'intégrité des données, leur disponibilité et leur confidentialité sont maintenues dans le cloud, d'une part, et d'autre part, la manière donc les données sont sauvegardées.
- **Utilisation :** ce que nous entendons par « utilisation des données » concerne uniquement l'exploitation de celles-ci dans le cadre d'un cloud privé/public mais n'affecte en aucun cas leur modification (étape de création). La manipulation de ces données crée quelques problèmes notamment celui de l'usage interne ou externe. Le maniement des informations s'opère dans un cadre légal, la question est de savoir si le fournisseur de cloud respecte un tant soit peu cette contrainte.

- **Partage :** les données sont disponibles à d'autres personnes avec certains droits associés. Il est nécessaire de savoir si le fournisseur d'informatique en nuage peut respecter les contrôles d'accès sur liste imposée par le client.
- **Archivage :** l'archivage est une action qui permet de gagner de la place et de conserver des données sur un long terme. Deux possibilités s'offrent à nous : soit ces informations sont engrangées et non accessibles, donc en mode hors ligne, soit elles restent dans un état actif c'est-à-dire en ligne. Nous devons nous soucier de la capacité du cloud provider (CP) à garder ces données à longue échéance car certaines lois imposent l'archivage sur une durée qui peut varier de trois ans à plusieurs décennies.
- **Destruction :** détruire une donnée consiste à la supprimer de tout média de stockage de manière définitive.

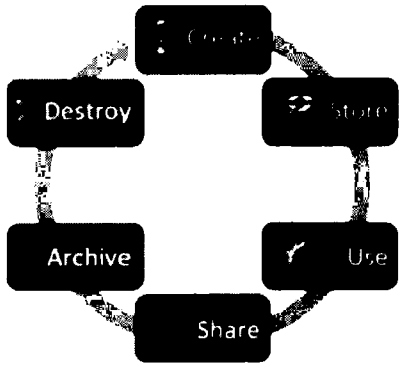


FIGURE 2.3: Cycle de vie des données (figure tirée de (25))

Notons qu'il peut y avoir des raccourcis dans ce cycle de vie : par exemple, une donnée peut simplement être détruite après son utilisation sans pour autant être partagée ou archivée.

Par ailleurs, chaque étape du cycle de vie est soumise à des contraintes sécuritaires. Puisque les données sont accueillies chez un fournisseur de cloud, le dépositaire doit en retour recevoir la garantie qu'elles suivent un traitement adéquat en relation avec de bonnes pratiques. Ainsi, une information qui est supprimée par la volonté d'un client nécessite pour la circonstance des procédures de destruction appropriées. Une question inquiétante cependant reste en suspens : une archive, une copie de données ou une sauvegarde que l'on vient de supprimer, l'est-elle réellement ?

Les systèmes d'exploitation actuels comportent une table de fichiers, par exemple Master File

Table pour le système de fichier *New Technology File System* de Windows [14], qui recense les liens vers les données et leur emplacement sur le disque (bloc). L'effacement typique d'une information se réalise par suppression de la relation entre la référence et la donnée sur le disque dur (sans pour autant détruire physiquement chaque bit de données). Nous sommes ici confrontés à un véritable souci en matière de sécurité (confidentialité des données) ; une multitude de logiciels proposent en effet la reconstruction d'une telle table de fichiers. Dans son rapport « Guidelines for Media Sanitization » (32), la NIST décrit les méthodes permettant de supprimer de manière efficace les données : les techniques consistent soit en l'écriture de bits sur les secteurs visés du disque dur, une démagnétisation [15] ou une destruction du disque. Cependant ces techniques devraient ici être mises en place par le fournisseur d'infonuagique et imposent logiquement un surcoût. Prouver de manière efficace ou démontrer la suppression d'une donnée s'avère complexe. L'article (45) propose une méthode de suppression efficace des données chez un fournisseur de cloud. Un algorithme de destruction modifiant le bit le plus significatif de chaque bloc de données chez le fournisseur de cloud rend les données non-recouvrables. De plus, le client peut utiliser des mécanismes lui permettant d'avoir l'assurance que la donnée a bien été rendu illisible chez le CP.

Par ailleurs, une nouvelle inquiétude vient s'ajouter à la précédente, relative aux nombreuses copies existantes d'un même fichier. Prenons un exemple concret : un fournisseur de cloud prévoit la migration d'une machine virtuelle qui stocke des archives de fichiers sur un serveur physique plus robuste. Le client décide durant cette migration de supprimer un fichier. Rationnellement, la donnée doit être supprimée de tous les serveurs (archive, back up, etc.) ; cependant qu'advient-il de la copie hébergée sur le serveur en cours de migration ? Dans ce sens, la preuve de suppression évoquée ci-dessus n'est plus effective. La préoccupation s'intensifie lorsque le fournisseur de cloud fait appel à d'autres prestataires en nuage (augmentation du niveau de redondance, de l'espace de stockage, etc.). Ainsi, il est impossible de prouver le principe de suppression.

---

14. Source : Microsoft Windows : Master File Table [En ligne]. Disponible `http://msdn.microsoft.com/en-us/library/windows/desktop/aa365230(v=vs.85).aspx` 2013. [Consulté le 6 Février 2014]

15. Les disques durs utilisant des plateaux se servent d'une tête de lecture pour magnétiser un emplacement physique de ce disque ; la démagnétisation de l'ensemble du disque implique donc la suppression des données. La même méthode est utilisé pour les bandes magnétiques. À l'inverse, les disques dur de type Solid State Drive (SSD) utilisent des mémoires flash non-volatiles (un contrôleur écrit dans des cellules) et ne peuvent pas être démagnétisés. Il est possible de garantir l'effacement par réécriture multiple mais la méthode la plus efficace consiste en une destruction du disque.

2.4.2 Menaces générales liées aux données

Le prestataire d'informatique en nuage impose sa politique sur les informations qu'il traite ; le client n'est plus le maître absolu de ses données en quelque sorte. D'après le rapport de l'ENISA « Cloud Computing Risk Assessment » (12), le risque d'enfermement que peut rencontrer le client chez un fournisseur de cloud met en péril les données confiées. En effet, ce fournisseur utilise des protocoles propriétaires et il n'existe, du moins pour l'instant, aucun standard d'interopérabilité des protocoles entre les fournisseurs. La fermeture d'un fournisseur d'informatique en nuage signifie pour l'utilisateur un risque de non-retour de ses informations. La presse internationale confirme nos propos car suite à la cessation d'activité du service Megaupload en janvier 2012, de nombreux adhérents se sont retrouvés privés de leurs données [16]. Des procédures juridiques sont en cours pour procéder à la récupération des fichiers. Plus récemment (janvier 2013), le fournisseur de cloud anglais 2e2 en cessation de paiement met en difficulté un très grand nombre d'entreprises [17]. La sauvegarde des données constitue la clé de voûte d'une entreprise qui se veut pérenne ; plus elle détient de données et les exploite efficacement, plus ses chances de se positionner en leader sur le marché augmentent. Il va de son intérêt de protéger le plus sérieusement possible l'intégralité de son capital informationnel.

2.5 Mécanismes d'assurance

2.5.1 Standards de sécurité

PCI DSS Le standard *Payment Card Industry Data Security Standard* offre un guide de douze bonnes pratiques qui assurent la sauvegarde des données bancaires pour prévenir, entre autres, la fraude et maintenir au niveau planétaire des mesures de sécurité homogènes (14). Ces douze conseils couvrent un grand nombre de domaines : la gestion sécuritaire d'un réseau via un pare-feu, la gestion appropriée des informations bancaires au repos et en transit, l'utilisation d'antivirus, la mise en place de contrôles d'accès, la surveillance des réseaux, la gestion d'une politique de sécurité des données. Ces normes s'adressent à toute entreprise qui possède un accès aux numéros de cartes bancaires donc à des banques, des marchands, etc.

---

16. Voir : Rémi Bricard : Fermeture du site Megaupload : quels enseignements pour le cloud computing ? [En ligne]. Disponible `http://www.journaldunet.com/ebusiness/expert/50873/fermeture-du-site-megaupload---quels-enseignements-pour-le-cloud-computing.shtml` 2012. [Consulté le 28 Janvier 2014]

17. Voir : Etienne Wery : Cloud : la perte totale des données est possible. La preuve par 2e2 et megauplaod. [En ligne]. Disponible `http://www.droit-technologie.org/actuality-1577/cloud-la-perte-totale-des-donnees-est-possible-la-preuve-par-2e2-et.html` 2013. [Consulté le 1 Mars 2013]

2.5.2 Audits et certifications

Définition L'audit consiste à vérifier les données, les enregistrements, les opérations et les performances (financières et autres) d'une entreprise [18]. Par exemple, un audit peut permettre de vérifier si le SLA est véritablement respecté.

Dans leur article « Auditing to Keep Online Storage Services Honest » (52), les auteurs argumentent sur les caractéristiques de l'audit et dressent en conséquence une liste des propriétés qui permettront à cet audit de se dérouler dans les meilleures conditions avec toutes les qualités attendues. Cet audit :

- ne doit pas provoquer de gouffre financier au fournisseur d'informatique en nuage ;
- doit éviter l'introduction de nouvelles vulnérabilités ;
- doit être effectué en toute impartialité c'est-à-dire, en neutralité vis-à-vis des fournisseurs en présence.

Notons l'existence de deux types d'audit. Nous connaissons l'audit interne pratiqué par l'entreprise elle-même qui vérifie si ses objectifs sont bien atteints par une série d'évaluations : management du risque, contrôle, gouvernance. Le deuxième type d'audit, externe, qui se réalise par un organisme tiers (le plus souvent accrédité), atteste que la société auditée s'est réellement conformée à une succession d'exigences. Le prestataire d'informatique en nuage ne doit pas s'opposer à une demande d'audit externe par un client qui veut obtenir des garanties sur la qualité des politiques de gouvernance, des bonnes pratiques utilisées, etc.

La norme américaine SAS70 très répandue sur le plan international, a été abrogée en juin 2011 car elle était à l'origine américaine ; or, les entreprises avaient besoin d'une norme internationale pour mieux répondre aux demandes du marché. Elle donne sa place à l'ISAE3402 que nous décrivons ci-dessous.

---

18. Voir : Wikipedia : Audit [En ligne]. Disponible `http://fr.wikipedia.org/wiki/Audit` 2013. [Consulté le 18 mars 2013]

La norme ISAE3402 a été introduite en juin 2011 avec comme but final de : « permettre aux clients de prestations externalisées d'obtenir un certain niveau d'assurance sur la fiabilité du dispositif de contrôle interne [19] de leur prestataire de services » (61). En effet, ces prestations de services ont un impact direct sur la présentation de l'information financière de l'entreprise cliente (7). Dans cette optique, le prestataire (par exemple, un fournisseur d'informatique en nuage) est tenu de répondre à certaines exigences telles que : (1) la sélection d'indicateurs pertinents pour évaluer l'efficacité des contrôles, (2) l'identification des risques qui porteraient atteinte aux objectifs de contrôles établis et la capacité du prestataire à faire face à ces difficultés ainsi que (3) la détermination et l'évaluation des contrôles en place pour assurer un certain degré d'assurance. Pour répondre à cette démarche de fiabilité du dispositif, nous retrouvons la traditionnelle roue de Deming ou PDCA (Plan-Do-Check-Act) [20] qui s'inscrit dans une démarche d'amélioration continue.

D'après (48), nous avons à notre disposition deux niveaux de contrôle :

1. Le premier - nommé **rapport de type I** - permet à l'auditeur d'évaluer les efforts de l'entreprise en matière de services au moment de l'audit pour prévenir, de manière ultime, les erreurs ou défaillances au niveau des processus financiers de l'entreprise dans les domaines suivants : gestion de paie, stocks, maintenance du système d'information. Ainsi, l'entreprise obtient un certain degré d'assurance sur l'existence d'un contrôle interne adapté.
2. Le deuxième - nommé **rapport de type II** - contient les mêmes éléments que le précédent rapport en plus du choix de l'auditeur à attester de l'efficacité des contrôles depuis leur implémentation. La principale différence avec le rapport de type I réside dans le fait que l'évaluation s'étend sur une période de six mois.

Cette certification universellement reconnue offre des mécanismes d'assurance robustes ; elle couvre en effet d'autres dimensions de gouvernance de l'entreprise :

- L'*Information Technology Infrastructure Library* (ITIL) vise à l'amélioration continue des services informatiques en proposant un référentiel de bonnes pratiques de management sur la façon de planifier, fournir et gérer des fonctions de service IT.

---

19. L'évaluation du contrôle interne concerne les dispositifs de haut niveau mises en place par la gouvernance de l'entreprise (contrôle de la finance, du stock, de la paie, du système d'information, etc.).

20. Voir : Fernandez : Définition Roue de Deming [En ligne]. Disponible `http://www.piloter.org/qualite/roue-de-deming-PDCA.htm` 2013. [Consulté le 8 octobre 2013]

- Le *Capability Maturity Model Integration* (CMMI) est un ensemble de bonnes pratiques de développement et de maintenance des systèmes et des applications informatiques visant notamment à réduire les coûts de développement et à augmenter la qualité des logiciels sans impact sur son coût.
- Le *Control OBjectives for Information and related Technology* (COBIT) est un référentiel général de la gouvernance des SI qui facilite cette gouvernance en permettant entre autres : « un meilleur alignement de l'informatique sur l'activité de l'entreprise du fait de son orientation métier, une vision compréhensible par le management de ce que fait l'informatique, une attribution claire de la propriété et des responsabilités, une bonne compréhension de toutes les parties prenantes grâce à un langage commun, etc. ». [21]

La norme ISAE3402 permet ainsi d'intégrer les principes fondateurs de ces modèles de gouvernance. Nous y voyons cependant des points négatifs : son coût, sa durée, le fait que l'auditeur peut être interne donc sans intervention d'organismes spécialisés et reconnus [22], que ce genre d'audit ne fournit aucune assurance au client final et risque ainsi de discréditer l'entreprise qui s'auto-certifie. Si la norme ISAE 3402 a une portée internationale, **la norme SSAE16** (dérivée de la certification SAS 70) est purement américaine; elle vise toutefois les mêmes objectifs.

La norme ISO27001 propose un système de gestion des SI assurant la sélection appropriée de contrôles de sécurité qui protègent les actifs de l'entreprise tout en donnant une confiance des parties (entreprise certifiée, clients, prestataires) en ce système [23]. ISO27001 se fonde également sur la roue de Deming (*Plan - Do - Check - Act*). La phase de planification (*Plan*) met l'accent sur les actions que l'entreprise va retenir pour sa sécurité, la phase *Do* met en pratique ces actions, la phase *Check* mesure les écarts entre ce que l'entreprise a retenu et mis en place. La dernière phase *Act* comble ces écarts. Les objectifs visés par cette norme sont :

- une amélioration de la sécurité;
- une bonne gouvernance;
- une conformité à la gouvernance de l'entreprise;
- une réduction des coûts;
- le marketing.

---

21. Source : AFAI-ISACA : Présentation de COBIT V4.1 [En ligne]. Disponible `http://www.afai.fr/index.php?m=29` 2008. [Consulté le 28 Janvier 2014]

22. Voir : Wikipedia : ISAE 3402 [En ligne]. Disponible `http://fr.wikipedia.org/wiki/ISAE_3402` 2013. [Consulté le 19 juin 2013]

23. Voir : American National Standard : Information technology - Security techniques - Information security management systems - Requirements [En ligne]. Disponible `http://www.accelerosblog.com/wp-content/uploads/2011/05/INCITS+ISO+IEC+27001-2005.pdf` 2006. [Consulté le 20 juin 2013]

2.6 Conclusion

Ce chapitre nous a permis de nous familiariser avec les paradigmes de sécurité et les principales menaces dans l'adoption de services en nuage. Nous avons poursuivi notre étude sur un cas plus spécifique qu'est l'externalisation des données. Les considérations légales comme les lois et les SLA offrent des garanties aux fournisseurs de cloud et aux états ; non aux clients du cloud. Il existe certes des mécanismes d'assurance pour ces clients : les standards de sécurité, les audits et les certifications. Nous estimons que ceux-ci sont nécessaires pour offrir un niveau de sécurisation appropriée des données ; mais insuffisants au regard de cette contrainte que représente l'externalisation des données. Nous poursuivons, dans ce sens, notre réflexion sur l'intégrité des données et les différents mécanismes attenants qui permettent d'obtenir une sécurisation efficace des données dans un environnement en nuage.

CHAPITRE III

INTÉGRITÉ DES DONNÉES DANS LE CLOUD

Ce chapitre présente en détail le concept de l'intégrité des données. Nous définissons dans un premier temps le concept, la manière de la mesurer et ses conséquences (3.1). Nous poursuivons par l'énumération des causes de violation d'intégrité (3.2). La section 3.3 établit une taxonomie des techniques de vérification d'intégrité. Enfin, nous étudions les mécanismes de vérification d'intégrité appropriés pour un environnement en nuage (3.4).

3.1 À propos de l'intégrité

Appliquée aux personnes, l'intégrité est une vertu qui représente la qualité de caractère d'un individu. Nous considérons également les divers aspects de la vie d'une personne ; nous parlons d'attributs intègres en terme de professionnalisme, d'intellect ou d'artistique. [1] Cependant agir avec intégrité ne signifie pas nécessairement agir avec moralité : deux individus peuvent avoir des avis opposés sur une personne dont l'un affirmera qu'elle est totalement intègre, l'autre qu'elle opère de façon immorale. Pour des objets, différentes définitions de l'intégrité s'emploient : l'intégrité d'un objet relève de sa capacité à être complet, intact ou pur. Ce cadre s'adapte aussi bien dans le domaine de l'écologie que celui de l'informatique. Si nous transposions cette définition à une donnée, nous n'aurions à retenir que la forme de cette donnée en l'occurrence sa capacité à être complète ou intacte et non son fond à savoir sa capacité à être pure. Le fond d'une donnée ou son caractère moral pur est dépendant du contexte et se situe en dehors du cadre de notre mémoire. Cette définition subjective pourrait constituer néanmoins la base d'un débat politique, philosophique ou religieux très passionnant.

---

1. Voir : Cox, Damian, La Caze, Marguerite and Levine, Michael : "Integrity", *The Stanford Encyclopedia of Philosophy* [En ligne]. Disponible `http://plato.stanford.edu/entries/integrity/` 2013. [Consulté le 26 avril 2013]

3.1.1 Définition généraliste de l'intégrité des données

Dans son rapport technique (55), la NIST définit l'intégrité des données de la manière suivante : *« L'intégrité des données est la propriété selon laquelle les données n'ont pas été altérées de manière non autorisée pendant leur phase de stockage, traitement ou transit. »*. Nous sommes conscients que bien d'autres définitions de l'intégrité des données auraient pu retenir notre attention [2], cependant celle de la NIST, plus récente, nous séduit par sa justesse. Les différentes définitions se rejoignent par ailleurs, seules quelques variations mineures sont à noter.

Toujours d'après la NIST, l'intégrité représente le deuxième facteur critique pour une entreprise derrière la disponibilité. Il est judicieux à ce stade d'introduire le concept d'intégrité du système de traitement (routeurs, ordinateurs, serveurs). L'intégrité même d'une donnée est liée directement à sa chaîne de traitement, de sa création à sa destruction. Lorsqu'un élément d'une chaîne d'action est défaillant — non en raison d'une panne mais parce qu'il fournit des informations incorrectes — la mise en place d'une redondance des données permet parfois de compenser grâce à un coordinateur, qui fournit une prise de décision qui permet transitivement de conserver les données intègres.

Le risque d'erreurs et le nombre de points d'attaque s'amplifie avec l'arrivée de l'informatique en nuage, d'autant plus que le client ne possèdent plus ses données en interne de l'entreprise. Dans ce contexte, il doit avoir l'assurance que ses données demeurent toujours intègres.

3.1.2 Mesures et niveaux d'intégrité d'une donnée

Considérons par analogie deux protocoles de couche 4 (transport) du modèle OSI, à savoir TCP et UDP qui se démarquent par leur approche radicalement différente en matière de transfert. TCP (*Transmission Controle protocol*) reste en mode connexion tant que cet état se conserve entre les deux initiateurs de l'échange (via un mécanisme de connexion en quatre étapes appelé *Handshake*). Par ailleurs, TCP permet un transport fiable du flux de données via divers mécanismes comme le contrôle de flux [3], les numéros de séquence et d'acquittement [4],

---

2. Voir : CCN-CERT : Data integrity [En ligne]. Disponible `https://www.ccn-cert.cni.es/publico/serieCCN-STIC401/en/d/data_integrity.htm` [Consulté le 7 juin 2013]

3. Mise en tampon qui évite la surcharge de l'un ou l'autre protagoniste.

4. L'ordre de transmission et l'ordre d'arrivée des paquets TCP est le même.

les sommes de contrôle[5], etc. TCP est donc utile lorsqu'un seul bit de donnée corrompu peut altèrer l'ensemble des données : par exemple, pour le transfert de fichiers.

Contrairement au protocole TCP, avec le protocole « User Datagram Protocol » (UDP) l'ordre d'arrivée des datagrammes n'est pas forcément celui de départ et il n'y a pas de mécanismes de détection d'erreurs, de pertes de datagrammes. Ce protocole convient lorsque la perte de données est acceptable. Quelque soit le protocole de couche 4, la perte d'un paquet peut survenir, par exemple, lorsque la file d'attente d'un routeur est saturée (mémoire tampon) : ce routeur prend alors la décision de supprimer le paquet. Les services de streaming de flux vidéos ou de téléphonie sur IP utilisent très largement le protocole UDP pour éviter une livraison d'images ou de bouts de phrases quelques instants après le moment attendu de leur présentation[6] ce qui serait inacceptable aux yeux du destinataire.

Nous souhaitons appliquer le même principe en matière d'intégrité de données à haut niveau (données exploitables), en définissant un seuil d'acceptabilité ou les objectifs à remplir par la donnée. Quelques normes comme JPEG se servent d'algorithmes de compression avec pertes. Le niveau de compression souhaité correspond donc à une perte d'intégrité volontaire de la donnée (image) initiale dans le but d'obtenir un certain degré de performance (par exemple, charger plus rapidement une page web).

Au vu de ces arguments, nous devons comprendre et retenir que la perte d'intégrité (au-dessus du seuil minimal) peut être acceptable dans la situation où interviennent les sens de l'être humain (essentiellement visuel et auditif). L'homme a en effet la capacité d'accepter ou de s'adapter à cette perte d'intégrité.

Les données à haut niveau sont généralement accompagnées de méta-données qui les décrivent afin qu'un programme ou un système d'exploitation (SE) sache comment les exploiter. La perte d'intégrité non contrôlée de méta-données risque d'entraîner l'incapacité d'exploiter l'ensemble des données associées.

Pour répondre in fine à la question « Existe-t-il des niveaux à partir desquels une donnée peut être qualifiée d'intègre ? », nous affirmons qu'une perte d'intégrité n'est tolérable que lorsqu'elle est sous contrôle de mécanismes bien définis comme l'échantillonnage (nous parlons

---

5. Une somme de contrôle vérifie que le paquet est intègre.

6. Ces désagréments peuvent être liés à une perte de paquets, une corruption de paquets entrainant sa retransmission ou une latence importante sur le réseau.

de taux) ou comme la compression avec pertes. Nous prenons en considération le coût de récupération versus l'acceptabilité d'une perte d'intégrité. Une perte d'intégrité non contrôlée peut conduire à une véritable catastrophe, elle n'est donc pas souhaitable, d'autant, qu'elle survient par définition de manière inopinée. Le mot catastrophe est ici bien approprié face à un fichier malheureusement illisible dans sa totalité. Nous devons cependant nuancer nos propos, prenons pour ce faire un cas concret : nous rédigeons notre mémoire de maîtrise mais perdons malencontreusement une certaine quantité de bits (pour l'une des raisons évoquées dans la section 3.2). Ces derniers correspondent par exemple au chapitre Introduction ; nous conservons toutefois intactes toutes les autres parties de notre rédaction. Nous sommes conscients de la perte d'intégrité de notre mémoire ; pour autant faut-il le disqualifier dans sa globalité, et par la même, l'ensemble de notre travail ?

Quel que soit le coût de récupération d'une donnée, il dépend intrinsèquement du niveau de sécurité / criticité que nous lui avons attribué au préalable (via une politique de gouvernance des données par exemple). Dans ce sens, la réécriture du chapitre de notre mémoire ne nous occasionnera pas une perte de temps considérable puisqu'il suffira de remettre en forme les idées développées une première fois. Par contre, qu'en est-il de la perte d'un album photo datant de dix ans ? Il est complètement impossible de refaire les photos ! Abordons un cas plus important à présent, à savoir, la perte pour une entreprise de la base de données qui contient la liste intégrale de sa clientèle ; la récupération d'un tel fichier est bien entendu capitale. Nous nous posons en toute logique la question qui suit : ***Sur quelles bases peut-on quantifier le coût d'une récupération des données ?***

3.1.3 Coût d'un recouvrement de données suite à une perte d'intégrité

Nous estimons que les coûts de recouvrement d'une donnée sont liés à :

1. Le coût de récupération : il est lié à la puissance computationnelle nécessaire pour recouvrer la donnée (en partant du principe que le recouvrement de cette donnée soit possible). Il est également lié au temps d'indisponibilité.
2. L'impact de cette perte pour le client.

Arrêtons-nous sur quelques exemples illustrés par la figure 3.1.

Un disque d'une grappe RAID-5 (présenté dans la suite, en 3.3.3), lorsqu'il tombe en panne, est remplacé et reconstruit. Le temps de reconstruction de la grappe varie en fonction de

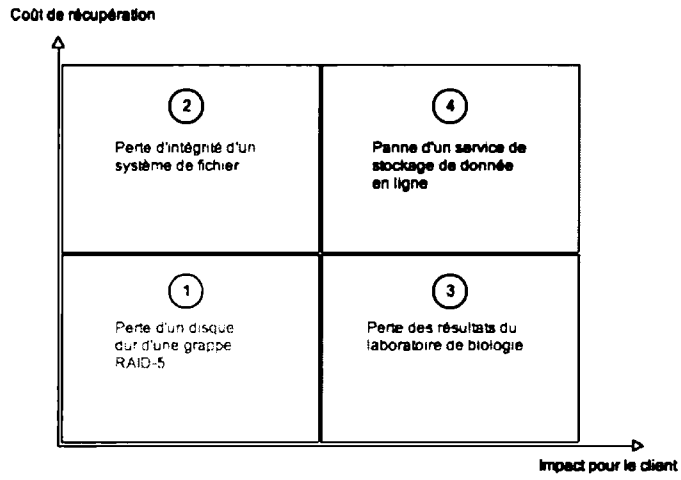


FIGURE 3.1: Graphique représentant le coût de récupération en fonction des effets lors de la perte d'une donnée.

la capacité du disque endommagé et n'excède pas quelques heures [7]. Le coût de récupération est donc faible et le temps d'indisponibilité minime. Il en va de même pour l'impact sur le client. Ce cas de figure est illustré par la pastille numéro 1 sur la figure 3.1.

Considérons la perte d'intégrité d'un système de fichiers (pastille numéro 2 sur la figure 3.1). Les administrateurs en charge du rétablissement de ce système sont contraints de l'interrompre, d'isoler le problème et éventuellement de le résoudre. Cet arrêt impacte directement sur la disponibilité du ou des services associés à ce système de fichiers, et engendre des coûts supplémentaires pour l'entreprise concernée.

Un laboratoire de biologie qui effectue des calculs parallèles (séquençage ADN, etc.) sur une vingtaine de jours et se trouve confronté par la suite à la perte de ses données est contraint de recommencer les calculs sur une nouvelle durée de vingt jours avec une même puissance computationnelle requise. Ce cas correspond à la pastille 3 dans le graphique 3.1.

Un fournisseur de cloud, qui propose à ses clients un services SaaS de stockage en ligne, est victime d'un bris d'eau dans sa salle serveur (pastille numéro 4 de la figure 3.1). Il n'a pas mis en place de politique de reprise d'activité et son serveur est fortement endommagé occasionnant

7. Voir : Memset : Raid Disk Failure Calculator [En ligne]. Disponible `https://www.memset.com/tools/raid-calculator/` 2014. [Consulté le 10 Février 2014]

une indisponibilité de services de deux jours. Le coût de remise en fonction du service est de deux jours et les données sont irrécupérables. Le client est largement impacté et il perd confiance en l'utilisation des services de ce fournisseur de cloud.

Par ailleurs, pour recouvrer la perte d'une clé privée dans un contexte de chiffrement asymétrique, nous devons compter en années voire en siècles. Le chiffrement asymétrique fait effectivement partie des problèmes NP-Dur (21). Ce cas peut s'apparenter à celui de la pastille numéro 4 de la figure 3.1. Autre exemple : une entreprise possédant des courriels chiffrés datant de plus de dix ans perd la clé privée. Cette clé peut être impossible à recouvrer dans temps acceptables ; cependant l'impact demeure sans grand intérêt pour ce client.

Lors d'un incendie par exemple, un disque dur peut subir un dommage partiel ou total. Pour se ré-approprier leurs données, les entreprises font appel à des sociétés comme Symantec qui possèdent des salles blanches [8], permettant de récupérer les données [9]. Cependant, les prix varient entre 800 $ et 2 000 $ pour un disque dur, en fonction de sa capacité. Quant au temps de récupération, il peut avoisiner un mois.

Ces exemples nous font prendre conscience qu'il est indispensable d'affecter un niveau de protection approprié à une donnée. Les informations critiques, telle une clé de chiffrement garante de la confidentialité ou un travail computationnel effectué sur plusieurs semaines sur un grand jeu de données, demandent une couverture selon des politiques de sécurité robuste et de mécanismes sécuritaires adaptés.

D'après l'établissement Kroll Ontrack spécialiste dans la récupération de données (43), une entreprise qui perd l'ensemble de ses données critiques est vouée à la catastrophe financière et par conséquent, à l'arrêt de ses activités. Kroll Ontrack fournit quelques estimations de coûts en matière de pertes de données (tableau 3.1).

Les cadres de haut rang réclament des informations fiables sur l'ensemble des processus de l'entreprise, sur son économie, etc., pour valider des prises de décisions stratégiques. Une corruption de données peut provoquer une ou plusieurs modifications de valeurs dans la base de données et porter préjudice aux décisions à venir et la pérennité de l'entreprise elle-même.

Le scénario le plus catastrophique nous est dévoilé dans le rapport (2) :

---

8. Salles dotées de technologies avancées pour effectuer, dans la mesure du possible, la lecture des données provenant de disques durs très endommagés. Ces pièces sont équipées pour ne contenir ni poussières, ni bactéries

9. Voir : Kroll Ontrack : Salle blanche & laboratoire [En ligne]. Disponible `http://www.ontrack.fr/salle-blanche/` 2013. [Consulté le 17 juin 2013]

Chiffre d'affaires sectoriel par heure	Perte de revenus par heure
Télécommunications	2 millions de $
Établissements financiers	1,4 million de $
Technologies de l'information	1,3 million de $

TABLE 3.1: Coûts liés à la perte de données

« 6 % des entreprises qui ont subi une perte de données jugées critiques survivent, 43 % ne rouvrent jamais et 51 % périclitent dans les deux ans ».

L'intégrité des données est donc une priorité pour toute entreprise soucieuse de son avenir ; nous insistons mais à juste titre pour réaffirmer qu'un niveau de sécurité doit être associé aux données critiques. Portons maintenant notre réflexion sur les causes de la violation d'intégrité, pour mieux s'en prémunir.

3.2 Causes de la violation d'intégrité

3.2.1 Identification et répartition des causes

Les causes de la violation d'intégrité de données sont diverses ; elles peuvent venir de différents niveaux. Selon (37), les causes d'une perte de données se répartissent en deux catégories :

1. les actes malveillants (vu en 2.3.2) ;
2. la suppression ou l'altération accidentelle de données (vue ci-après).

3.2.2 Suppression ou altération accidentelle de données

La suppression ou l'altération accidentelle de données peut résulter de différents types d'évènements.

Erreurs logicielles et matérielles Durant leur cycle de vie, les données traversent de nombreux médias et équipements réseaux et sont stockées et exploitées au travers de nombreux serveurs qui - tous - peuvent corrompre ces données. La loi de Moore stipule que la densité des processeurs double tous les dix huit mois, or selon (2) le nombre de défaillances augmente de façon exponentielle à mesure que cette densité croît (voir figure 3.2).

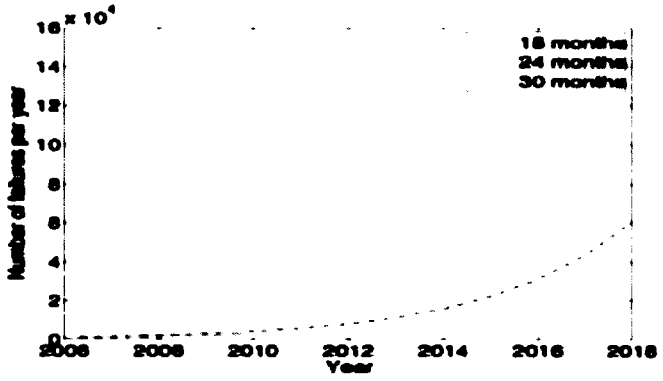


FIGURE 3.2: Projection du nombre de défaillance sur des CPU Multicœur (image tirée de (2))

Par ailleurs, une fonction d'un logiciel mal programmée risque de provoquer la corruption d'une donnée en cours de traitement, lors d'une levée d'exception ou du crash du logiciel par exemple, ou lorsqu'un pilote de périphérique réseau ou de disque dur est instable (53).

Du côté réseau : les équipements réseaux de mauvaise qualité ainsi que le bruit dans les médias de transmission peuvent altérer les données. Bien que les différents protocoles des couches TCP/IP soient capables — en théorie [10] — de détecter et de corriger ces erreurs (sauf UDP), celles-ci peuvent créer des complications au niveau des applications (couches hautes du modèle). Du point de vue matériel, la corruption silencieuse de données [11] est très sournoise et occasionne parfois de véritables ravages.

Erreur utilisateur Les utilisateurs qui possèdent les autorisations nécessaires à l'accès d'un système risquent de compromettre l'intégrité des données par des erreurs de manipulation. Ces inexactitudes se rencontrent au niveau logique la plupart du temps. Par exemple, la suppression accidentelle d'un fichier ou d'un enregistrement d'une base de données constituent une perte d'intégrité des données.

---

10. Notons que des erreurs non-détectées sont possibles.

11. Les problèmes de corruption de données silencieuses arrivent au cours du processus d'écriture sur le disque. Ces erreurs sont les plus dangereuses dans le sens où les données incorrectes ne font l'objet d'aucune erreur particulière au niveau des journaux d'évènements du SE.

Catastrophes naturelles Les évènements climatiques extrêmes, les incendies, etc. sont autant d'aléas qui peuvent porter préjudice à l'intégrité des données. L'ouragan Sandy en est l'exemple le plus récent : en octobre 2012 de nombreux centres de données de la côte Est des États-Unis ont été touchés. Ainsi, de nombreux sites internet et services en nuage ont été indisponibles pendant au moins deux jours suite aux inondations et pannes électriques occasionnées [12].

Nous venons de voir les causes de la perte d'intégrité d'une donnée, nous présentons ci-dessous les mécanismes à mettre en place ainsi que le niveau de sécurisation requis pour prévenir, détecter et vérifier l'intégrité d'une donnée.

3.3 Taxonomie des techniques de vérification d'intégrité

Sivathanu et al. (53) définissent une taxonomie des techniques qui vérifie l'intégrité d'une donnée lors de la phase de stockage. Nous allons faire la synthèse des différentes techniques de vérification d'intégrité (évitement, détection et correction) proposées par cet article.

3.3.1 Évitement

Quelques systèmes de fichiers fournissent par défaut un haut niveau de garantie sur l'intégrité des données. Par conséquent ajouter des mécanismes additionnels ne relève pas d'une absolue nécessité. Nous allons découvrir ci-après quelques-unes des techniques employées.

Stockage en lecture seule

Ce mode de stockage peut s'effectuer au niveau logique ou physique et nous éloigne des erreurs d'intégrité — en théorie — puisqu'aucune modification des informations n'est possible après écriture. Cependant, en pratique, cette solution ne constitue pas une panacée. Prenons l'exemple d'un DVD-ROM : le disque inscriptible qu'une seule fois, à une durée de vie limitée du fait de différents facteurs entrainant la détérioration du support (qualité des matériaux, stockage du disque, nombre de lecture, etc.) [13]. Il en va de même pour les autres médias de stockage

---

12. Voir : Geneste : Sandy met à mal les datacenters de la côte Est [En ligne]. Disponible `http://www.lemondeinformatique.fr/actualites/lire-sandy-met-a-mal-les-datacenters-de-la-cote-est-51081.html` 2012. [Consulté le 7 octobre 2013]

13. Voir : Optical Storage Technology Association : Understanding Recordable & Rewritable DVD [En ligne]. Disponible `http://www.osta.org/technology/dvdqa/dvdqa11.htm` [Consulté le 20 Mars 2014]

(Disque dur, bande magnétique, Compact Disk, etc.). Des modifications sur les données après stockage en lecture seule sont donc possibles.

Ce mode de stockage qui reçoit des données non modifiables reste très limitatif. De nombreux systèmes de fichiers comme NTFS, UFS, etc., ou le recours à des contrôles d'accès par liste permettent la mise en place de tels mécanismes. Même si ce mode de stockage est sécuritaire pour des données au repos, il ne propose aucune vérification d'erreurs lors de l'écriture c'est-à-dire au moment du transfert des données vers le média de stockage (par exemple du réseau vers le disque dur ou de la mémoire centrale RAM vers le disque dur). De la même manière, lorsqu'une donnée valide est lue depuis son média de stockage, des erreurs lors de la transmission subséquente de l'information peuvent survenir. Des mécanismes de détection et de correction (vu ci-dessous) sont capables de pallier ce genre de problème.

Journalisation d'évènements

La journalisation - au niveau du système de fichiers - consigne dans un journal toutes les transactions exécutées depuis et vers un disque. Ce mécanisme autorise éventuellement de rejouer ces transactions suite à un problème d'intégrité ou un crash système. Cependant, la journalisation ne fait pas systématiquement état d'une corruption silencieuse de données ou d'une intrusion malicieuse. Un utilisateur malintentionné pourrait tout à fait stopper la fonction de journalisation d'un système d'exploitation puis corrompre les données afin que celles-ci ne soient pas recouvrables. La très grande majorité des systèmes de fichiers actuels utilisent la journalisation : HFS+ pour Mac, NTFS pour Windows, ou encore Ext3 pour Unix.

La journalisation - au niveau applicatif - enregistre dans un journal toutes les actions effectuées par une application. Ces journaux permettent de rejouer des transactions non terminées au moment d'un crash système, des fonctions de débogage et des fonctions de non-répudiation (c'est-à-dire savoir : Qui ? A fait quoi ? Quand ? Comment ? Dans quelles circonstances ?). Par exemple, le système de gestion de base de données Microsoft SQL Server utilise un journal des transactions pour rejouer automatiquement les actions incomplètes au moment d'un crash (de l'application ou du système d'exploitation). Par ailleurs, il est possible de paramétrer différents degrés de journalisation. Ainsi, l'application Apache HTTP serveur propose huit niveaux de journalisation (de *emerg* qui indique que le système est inutilisable à *debug* qui enregistre toutes les actions effectuées).

Systèmes de fichiers cryptés

Ces systèmes chiffrent les fichiers dont ils ont la gestion. La confidentialité des données hébergées est le principal objectif que l'on cherche à atteindre, mais dans une certaine mesure l'intégrité y est également garantie. En effet, le remplacement de certains fichiers systèmes, le vol de données critiques ou bien de certificats, etc., par des actes malveillants (chevaux de Troie, attaques à distance, augmentation de privilèges, etc.) sont rendus plus difficiles. Toutefois, ce mode d'évitement reste totalement démuni face à la corruption silencieuse de données. De plus, le problème de l'intégrité des données est déporté à un autre niveau : si un utilisateur réussissait à s'emparer de la clé, tout l'intérêt du système de fichiers cryptés disparaîtrait.

Les systèmes de fichiers transactionnels

Les systèmes de fichiers transactionnels utilisent quatre propriétés, notées ACID, pour réaliser des transactions informatiques et livrer ainsi à leurs utilisateurs un niveau d'intégrité acceptable des données :

- **Atomicité :** la transaction doit se réaliser entièrement ou absolument pas. Il en résulte deux bénéfices : (1) une simplification (transaction entièrement annulée) de la gestion des erreurs et (2) l'absence de corruption de données puisque ces dernières n'ont pas été inscrites sur le média par le système de fichiers.
- **Cohérence :** un système de fichier reflète avec exactitude les données contenues sur le média de stockage c'est-à-dire que chaque action d'écriture sur le disque amène le système d'un état cohérent vers un autre état cohérent, ce qui garantit l'intégrité des données sur le disque ainsi qu'une gestion robuste des erreurs. L'utilisation de journaux d'évènements systèmes présentés plus haut permet d'atteindre un tel état de cohérence : en cas de crash du système d'exploitation, l'OS va lire le journal d'évènements, puis appliquer les changements. Ainsi, les transactions ne s'effectuent qu'après une exécution réussie de toutes les opérations (atomicité).
- **Isolation :** ce terme signifie qu'une transaction n'affecte pas l'exécution d'une transaction concurrente ; cette fonction n'est pas prise en charge par les systèmes de fichiers actuels.
- **Durabilité :** cette propriété garantit que lorsqu'une transaction a été correctement effectuée, c'est-à-dire sans erreurs, elle persiste même en cas d'erreurs systèmes. Toutes les applications devraient être capables de remplir cet objectif. Au sein d'un système de fichiers, la journalisation d'évènements abordée ci-dessus respecte cette contrainte.

3.3.2 Détection

La détection de violation d'intégrité fait appel à de nombreuses techniques que nous décrivons ci-après. Cependant, les méthodes employées ne corrigent pas les erreurs, elles sont destinées principalement à valider un message et ne demeurent toutefois pas infaillibles. La majeure partie de ces procédés s'utilisent largement dans la transmission réseau. Il paraît moins compliqué de solliciter la retransmission d'un paquet corrompu que de tenter d'en corriger les erreurs.

Somme de contrôle

La somme de contrôle appelée également *checksum* est une technique qui consiste à ajouter une empreinte à un bloc de données pour s'assurer que celui-ci n'a pas subi d' altération de la part d'un bit corrompu.

Bit de parité

Le bit de parité est un mécanisme simple qui inspecte notamment les erreurs de transmission. La parité d'un paquet est calculée sur l'ensemble des bits en appliquant une fonction ou-exclusif (XOR $\oplus$). Si le nombre de bits à la valeur un est pair, le bit de parité sera de zéro, sinon un. Le bit de parité calculé accompagne la donnée qu'il protège. Pour détection, le bit de parité recalculé par le récepteur est comparé avec celui qui a été reçu.

Contrôle de redondance cyclique

Le contrôle de redondance cyclique (CRC) appartient à la famille des sommes de contrôle et permet de détecter les erreurs lors du transfert des données (par exemple, réseau, d'un disque DVD vers l'application exploitant les données de ce média, etc.). Le but recherché est d'associer à chaque bloc de données un code de contrôle qui permet de vérifier si la donnée reste toujours intègre.

Hachage

Une fonction de hachage (nommée $h()$) permet de détecter des modifications sur une donnée. Elle transforme un message d'entrée de longueur variable m en une chaine de bits de sortie de taille fixe nommée empreinte (e). La technique de hachage permet effectivement de

vérifier après une opération (par exemple, après un transfert sur le réseau) l'état d'un message, c'est-à-dire la conservation de son intégrité. Ainsi, avant une opération, m est haché via le calcul de $e_{a\_priori} = h(m)$. Après l'opération, pour vérifier que le message résultant m' n'ait pas été modifié, on calcule via la même fonction de hachage : $e_{a\_posteriori} = h(m')$. Si $e_{a\_priori} = e_{a\_posteriori}$ alors $m = m'$ c'est-à-dire que le message n'a pas été altéré.

« Un petit changement du message donne un grand changement de haché » (6). Ainsi, le niveau d'intégrité du message vérifié est binaire : si aucun bit n'a été modifié, le message est intègre dans son ensemble, si au moins un bit de donnée est altéré, toute le message est considéré comme non-intègre.

Enfin, pour s'assurer de l'authenticité des données, nous pouvons employer une variante, la fonction de hachage avec une clé secrète k (nommée $h_k()$) de telle manière à ce que la vérification ultérieure requière la connaissance de cette clé k.

3.3.3 Correction

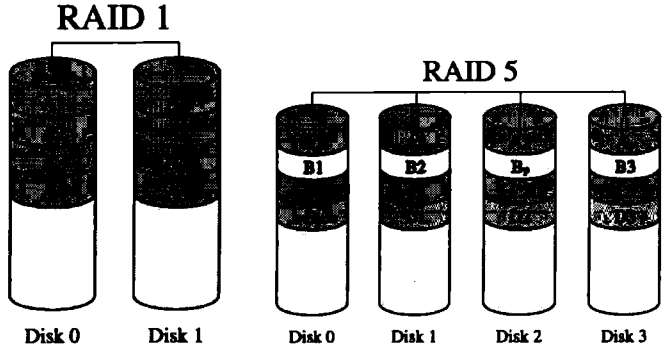
RAID

En 1986, Patterson et al introduisent le concept de RAID dans leur article « A Case for Redundant Arrays of Inexpensive Disks (RAID) » (44). Cette technologie autorise entre autres à combiner plusieurs disques durs physiques en une grappe de disques pour atteindre de meilleures performances ou une redondance des informations. Ces éléments concourent de la sorte à élargir la disponibilité des données et des services vivement sollicitée dans le domaine de l'infonuagique. Nous allons présenter plusieurs types de RAID :

1. **Le RAID 1 ou disques en miroir** contient pour chaque disque présent dans la grappe une copie identique à celui-ci (figure 3.3a). Une interruption sur l'un des disques de la grappe n'occasionne pas de perte des données. Cependant les performances sont amoindries puisque le système d'exploitation doit écrire la même donnée sur les deux disques de la grappe. Par ailleurs, la capacité effective de la grappe équivaut au volume de son plus petit disque : la capacité excédentaire de chaque disque n'est pas utilisée.
2. **Le RAID 5 ou volume agrégé par bandes à parité répartie** nécessite un minimum de trois disques durs ; il s'inspire du RAID 0 [14] pour obtenir des performances similaires

14. Le RAID 0 offre des performances en lecture/écriture puisque les données sont écrites au travers d'au moins deux disques et de façon parallèle.

et plus de sécurité du fait de l'ajout d'un bit de parité (figure 3.3b). Ces bits de parité sont issus du résultat d'un ou-exclusif $\oplus$ avec les blocs des différents disques de la grappe. Ils sont écrits circulairement sur chaque disque de telle manière qu'en cas de défectuosité de l'un de ces disques, ceux qui restent se chargent de la reconstruction de ce disque (et donc des données contenues) par la fonction ou exclusif. Revenons sur l'exemple de la figure 3.3b et considérons que le disque 2 est victime d'une panne : le système peut dès lors calculer $A1 \oplus A3 \oplus A_{parite} = A2$ puisque lors de l'écriture, le système a encodé $A1 \oplus A2 \oplus A3 = A_{parite}$. Cette opération se répète pour chaque bloc (A, B, C et D).
La capacité de stockage est égale au volume du plus petit disque multiplié par le nombre de disques moins le volume d'un disque (disque redondant) (63).



(a) Disques en miroir (b) Volume agrégé par bandes à parité répartie

FIGURE 3.3: Les principaux types de RAID qui garantissent l'intégrité des données (figures et titres tirés de (63))

Précisons qu'il est parfaitement possible de combiner ces différents types de RAID en fonction des objectifs de performance et de fiabilité que nous désirons atteindre. Par exemple, le RAID51 permet de construire un miroir de deux grappes RAID 5.

3.3.4 Correction d'erreurs

Code correcteur d'erreurs

Les codes correcteurs d'erreurs permettent, via l'ajout de bits de redondance, de détecter et de corriger les erreurs sur une donnée (53). Cette technique est utilisée lors du transfert de bits sur un canal peu fiable (la vérification se fait après la réception du message) ou lors du stockage de données (par exemple, sur un disque dur ou sur un média de stockage amovible). En ajoutant suffisamment de redondance, il est possible de corriger les patrons d'erreurs les plus probables.

3.3.5 Conclusion

Les techniques abordées dans cette section représentent des mécanismes qui peuvent ou doivent être mis en place par un fournisseur d'informatique en nuage pour apporter toute l'intégrité aux données souhaitée par l'ensemble de la clientèle. Nous réalisons également que cette intégrité se considère de bout-en-bout (transfert réseau, au repos) et durant tout le cycle de vie (voir 2.4.1). Le recours à l'infonuagique fait néanmoins intervenir - au minimum - deux familles d'acteurs : les prestataires de services et leurs clients. Or, les techniques actuelles ne favorisent pas les consommateurs : des vérifications à distance sur la consistance des fichiers devraient leur être proposées à travers divers mécanismes de surveillance. Différents documents issus de la recherche émergent dans ce sens ; ce sujet fera l'objet de la section suivante.

3.4 Vérification de l'intégrité dans le contexte du cloud computing

3.4.1 Hachage et vérification d'intégrité

Naïvement, ce type d'algorithme permet la vérification de l'intégrité des données hébergées chez le fournisseur en nuage.

Utilisation du hachage pour le cloud Soit $h_k()$ une fonction de hachage qui utilise la clé k générée de manière aléatoire [15]. F représente le fichier que le client envisage d'envoyer sur le service de stockage dans le cloud. Un client peut procéder au calcul de $r_{client} = h_k(F)$, stocker ce haché localement et ensuite faire parvenir le fichier F au prestataire de cloud. Si ce même client décidait de vérifier l'intégrité de son fichier, il divulguerait la clé k au fournisseur d'infonuagique. Ce fournisseur calcule : $r_{fournisseur} = h_k(F)$ et renvoie le résultat au client qui s'assurerait que $r_{client} = r_{fournisseur}$. De cette façon, le client obtient la preuve que le fournisseur de cloud dispose de F. Ce client peut faire de multiples vérifications d'intégrité, indépendantes les unes des autres, sur ce fichier F en changeant la clé k pour chaque nouvelle demande.

---

15. La notion d'aléatoire est nécessaire pour rendre le recouvrement de la clé générée plus complexe pour un attaquant utilisant la même fonction de génération.

Analyse de l'utilisation du hachage pour le cloud Il apparaît que cette implémentation constitue une solution inappropriée puisque :

1. Le calcul du haché représente pour le client un coût important en ressources informatiques (31) : sur des fichiers de taille importante, le client doit en effet appliquer la fonction de hachage sur l'intégralité du fichier. De plus, il doit posséder le fichier à vérifier à la fois en local et chez le fournisseur de cloud. Or, l'un des paradigmes prôné par le cloud consiste à déporter toutes ces ressources vers les datacenters.
2. Le client devrait recalculer un haché avec une clé différente pour chaque vérification : l'utilisation d'une même clé pour deux vérifications différentes pourrait conduire le fournisseur de cloud à anticiper ces vérifications et donc à fournir le résultat d'un calcul de haché antérieur à celui expressément demandé par le client. En jouant sur cet aspect de vérification temporel, le fournisseur de services en nuage peut cacher une perte d'intégrité sur ce fichier. L'utilisation de différentes clés est donc requise mais fastidieuse pour un client ; celui-ci serait tenu de gérer autant de clés que de vérifications à effectuer. Ceci va à l'encontre de l'essence même du concept de cloud computing.

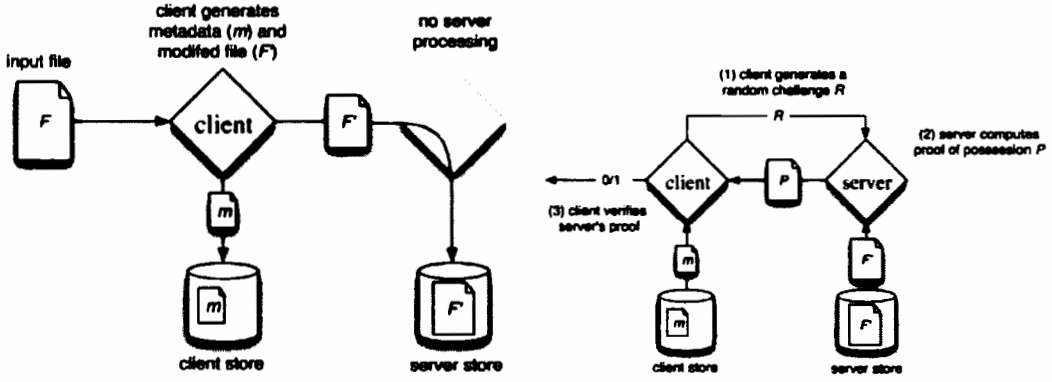
Pour remédier à cette problématique, des mécanismes fondés sur la preuve de possession ont été proposés.

3.4.2 Mécanismes qui se fondent sur la preuve de possession

Fonctionnement Nous retrouvons dans la littérature deux appellations qui gravitent autour de cette notion de preuve de possession, elles s'intitulent *Provable Data Possession* (PDP) et *Proof of Retrievability* (PoR) et se fondent sur deux phases bien distinctes : la première est la **phase de pré-traitement et de stockage**, la seconde concerne la **vérification** (4).

La première phase citée fonctionne généralement de la manière suivante (figure 3.4a) : un client envisage d'envoyer un fichier chez son fournisseur de cloud. Ce fichier (F) est composé de n blocs de données sur lequel le client vient greffer des sentinelles de manière aléatoire. Ces sentinelles sont des métadonnées générées aléatoirement et placées aléatoirement dans chaque bloc de données ; seul le client connait leurs positions.

Le client stocke la valeur et la position de chaque sentinelle localement : l'hébergeur ignore la position de ces sentinelles. En effet, ces dernières sont indissociables des blocs de données puisque le fichier F - avant d'être expédié et stocké chez le prestataire de cloud - est soumis à



(a) Phase de pré-traitement et stockage

(b) Phase de vérification

FIGURE 3.4: Mécanismes de preuve de possession (figures tirées de (4))

une procédure de chiffrement pour donner F'. Dès que le fichier est hébergé chez le fournisseur d'informatique en nuage, le client supprime sa copie locale.

Par la suite, si ce client projette de vérifier l'intégrité du fichier, il lui suffit de demander au fournisseur de cloud de retourner les informations spécifiques à une position dans le bloc qui correspond à celle d'une sentinelle (Preuve P). Ce type d'échange, nommé défi-réponse, permet à un client d'envoyer un défi au fournisseur d'informatique en nuage (par exemple, retourner les bits de la i^{eme} position à la j^{eme} position du bloc numéro n) ; ce dernier lui renvoie une réponse prouvant qu'il connait F. Ensuite, le client compare cette réponse à la valeur qu'il a stockée en local. Cette phase relève de la vérification (figure 3.4b) ; la mise en place d'un chiffrage asymétrique lors des différents échanges entre le client et le serveur sécurise cette étape contre des attaques de type « homme du milieu » où un individu malicieux pourrait, par exemple, falsifier la réponse du défi retourné par le fournisseur de cloud. Par l'ajout d'un nombre suffisant de sentinelles, la suppression ou l'altération des données peut facilement se détecter (en terme de probabilité). Par conséquent, lors d'un challenge de défi-réponse si la sentinelle retournée ne correspond pas à celle escomptée par le client, il est probable que le fichier soit corrompu.

La principale distinction entre les mécanismes PDP et PoR est à ce niveau. PDP permet uniquement la vérification d'un fichier pour constater son intégrité mais n'autorise pas de le recouvrer, ce qui inscrit cette fonctionnalité dans la catégorie « Détection » (étudiée en 3.3). Cette fonction est sollicitée lorsque le consommateur détient plusieurs copies de son fichier à divers endroits (datacenter du même fournisseur de cloud ou en environnement multifournisseurs) ou

s'il possède une politique de gouvernance de ses données robuste c'est-à-dire que les données doivent, par exemple, avoir un niveau de redondance suffisant pour ne pas entraver la continuité des affaires de l'entreprise (ce qui peut être dans certains cas transitif à l'argument précédent). PoR offre à la fois détection et correction d'erreurs.

Provable Data Possession (PDP) Ateniese et Al. ont introduit les premiers le concept de PDP « Provable Data Possession at Untrusted Stores » (4). D'après eux, la taille des requêtes réseau est moindre car seul le challenge de défi-réponse engendre une petite quantité de bande passante. Les performances sont donc meilleures, comparées à l'approche naïve que nous présentions ci-dessus. Cependant, l'article « Dynamic Provable Data Possession » (17), spécifie que ce contexte s'applique à des fichiers statiques consultés peu fréquemment, archivés ou en lecture seule. Or d'après l'article, les fichiers en question seraient fréquemment modifiés. Pour pallier ce manque de flexibilité, les auteurs étendent ce modèle au traitement des fichiers qui font l'objet de fréquentes manipulations comme la modification, l'ajout ou la suppression de blocs.

Proof of Retrievability (PoR) PoR effectue à la fois de la détection et de la correction d'erreurs. Selon (3), ce mécanisme est prévu pour des jeux de données statiques de grande taille. Le mécanisme comporte des codes correcteurs d'erreurs qui permettent - dans la mesure du possible - le retour à un état intègre des informations. Les auteurs adoptent le principe du PoR, y associent un code correcteur d'erreurs se limitant à un maximum de 15 % de redondance ; les sentinelles représentent quant à elles 2 % de données en sus. Ces chercheurs proposent un nouveau modèle de PoR nommé HAIL qui assure en même temps la vérification des informations et la haute-disponibilité des données (11). De plus, ce protocole détecte les plus petites erreurs d'intégrité (c'est-à-dire l'altération au niveau d'un bit) puisque le fichier détient suffisamment d'informations de redondance. Si le fichier est hébergé au sein de plusieurs datacenters (du même fournisseur ou non), HAIL autorise l'ajout de redondances additionnelles selon la technologie RAID-5 ; celles-ci font en sorte de récupérer le fichier corrompu en cas de défaut du fournisseur en nuage. La figure 3.5 explique un tel mécanisme.

Un client distribue un fichier F (composé de blocs de données n (*file*) et de sentinelles ainsi que d'un code correcteur d'erreurs (*encoding*)) chez différents fournisseurs d'informatique en nuage (de *P1* à *P3*). Comme pour la technologie RAID-5, les fournisseurs d'infonuagique *P4* et *P5* stockent des informations de redondance des blocs de données hébergées chez les fournisseurs *P1*, *P2* et *P3*. Par conséquent, si une sentinelle ou un code correcteur d'erreurs est corrompu (ce qui enlèverait toute possibilité de vérification pour un client), il est toujours possible de recouvrer les données affectées via *P4* et *P5*.

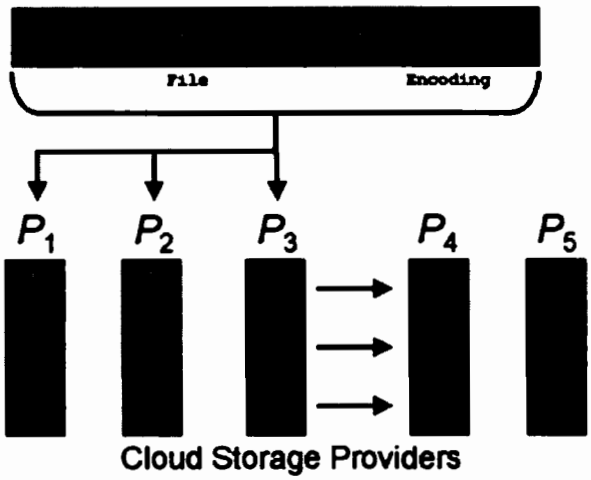


FIGURE 3.5: Transfert et ajout d'informations de redondance à plusieurs endroits (figure tirée de (34))

Intégrité des données dynamiques Dans l'article (57), les auteurs proposent une solution qui assure l'intégrité des données dynamiques et affirment que PoR et PDP ne sont satisfaisants que sur les fichiers statiques c'est-à-dire d'accès rares. Or, la plupart des données hébergées dans un environnement en nuage sont souvent modifiées, d'où la dénomination « données dynamiques ». Les auteurs organisent un fichier via une structure en arbre pour éviter de traiter l'ensemble du fichier à chaque modification comme le pratique PoR ou PDP par un ajout de codes correcteurs d'erreurs (opération trop longue qui affecte la disponibilité de la donnée). Un fichier peut être représenté comme étant une succession de n blocs de données. Par leur solution, chacun de ces blocs est placé puis haché dans un arbre de hachage de Merkle [16]. Comme illustré dans la figure 3.6, les feuilles constituent les blocs de données, les nœuds intermédiaires sont des hachés de ces blocs ou des nœuds en-dessous. Le nœud *top* ou *root* identifie tout le fichier.

D'après les auteurs, cet algorithme est plus performant que PoR et PDP car il traite uniquement les blocs de donnée modifiés, ajoutés ou supprimés et non tout le fichier. Pour une modification, l'algorithme recalcule le hashé du bloc de donnée concerné ; pour une suppression et une insertion, il met à jour uniquement les nœuds affectés. Concernant la possibilité d'un

16. « Un arbre de hachage de Merkle est une structure d'authentification qui prouve de manière efficace et sécuritaire qu'un certain nombre d'éléments de cette structure n'ait pas été endommagés » (57)

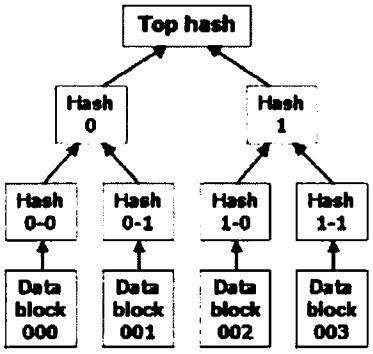


FIGURE 3.6: Arbre de Hachage de Merkle (figure tirée de (62))

client à vérifier à distance l'intégrité d'un fichier, les auteurs optent pour une authentification homomorphique des données [17] qui produit des sentinelles (comme PoR et PDP) infalsifiables. Ces sentinelles sont ensuite ajoutées, de manière aléatoire, aux blocs de données. Le mécanisme de vérification d'intégrité s'opère comme pour PDP et PoR via un mécanisme de défi-réponse.

Par ailleurs, notre étude réalisé dans le cadre du cours MGL7126 a mis en évidence que le système de fichiers Z-File System, prévu pour accueillir de très gros volumes de données (de l'ordre de 2^{60} octets), utilise le même arbre de hachage. Cette technique a été largement éprouvée lors de nos tests d'intégrité. Si on couple le principe de sentinelles au système d'arbre de hachage de ZFS, le client peut vérifier l'intégrité de l'ensemble d'un système de fichiers hébergé par un fournisseur de services en nuage (par exemple, IaaS).

Analyse des mécanismes Nous estimons que les mécanismes que nous venons de présenter suscitent l'intérêt pour plusieurs raisons. Selon les auteurs, ils assurent l'intégrité des informations de telle manière que le client n'ait à conserver qu'une quantité limitée de données en interne (la position et les valeurs des sentinelles, de l'ordre de 2 % de la taille du fichier) ; il existe peu d'accès réseau. Partant de ce principe, nous affirmons que ces mécanismes ne constituent pas une panacée puisque leur utilisation entraine des coûts en terme de bande passante et d'espace de stockage supplémentaires. Par ailleurs, ces mécanismes offrent une confiance supplémentaire au client vis-à-vis des services hébergés en nuage auxquels il souscrit. Cependant, toute entreprise qui opte pour ce genre de système devra posséder l'expertise interne adéquate ; à défaut, elle restera vulnérable et non assurée.

17. Le chiffrement homomorphique permet d'effectuer des traitements (calculs, etc.) directement sur des données chiffrées sans les déchiffrer au préalable.

CHAPITRE IV

DÉFINITION DES EXIGENCES POUR UN TIERS DE CONFIANCE

Ce chapitre détermine les exigences liées à l'utilisation d'un tiers de confiance (TDC) dans un contexte d'informatique en nuage. Ainsi, dans la section 4.1 on discute de la nécessité pour un client de recourir à un environnement multi-fournisseurs. La section 4.2 présente les problèmes de confiance qu'on accorde à un CP. La section 4.3 présente les différents rôles d'un tiers de confiance et met l'emphase sur l'identification des facteurs qui influent sur la confiance des usagers de l'informatique en nuage. La section 4.4 présente les exigences liées à l'introduction d'un troisième acteur. Enfin, la section 4.5 fait une revue du marché de la confiance dans le cloud en étudiant deux acteurs majeurs, CloudLock et CipherCloud.

Un tiers se définit comme étant une entité qui agit pour le compte de deux protagonistes à savoir le fournisseur de services en nuage et son client. À la différence d'un tiers, le tiers de confiance met en œuvre des mécanismes de sécurité visant les deux autres protagonistes à avoir confiance dans les services offerts par ce troisième acteur. Sur internet, nous retrouvons par exemple l'autorité de certification de clés publiques qui est un cas particulier de TDC ; elle fournit ces mécanismes via la signature électronique. Autre exemple : lors d'un paiement en ligne, la banque agit en tant que TDC pour valider la transaction entre un site d'e-commerce et le client (de la banque et du site).

Le terme « confiance » sera mentionné à plusieurs reprises dans ce chapitre. Arrêtons-nous un instant sur l'origine et la définition générale de ce mot. Étymologiquement, le terme confiance vient du latin *cum* qui signifie « avec » et *fidere* qui signifie « fier » [1]. Dans le cadre des affaires, elle peut se définir comme un concept qui vise à réduire les risques. Cependant, le dépositaire de la confiance peut se retrouver dans un état d'instabilité et de dépendance volontaire. Ainsi, un client qui envisage de déporter ses données vers le cloud doit accorder sa confiance à son fournisseur de services, sans oublier toutefois de faire preuve d'un minimum de prudence.

---

1. Voir : Marzano : Qu'est-ce que la confiance ? [En ligne]. Disponible `http://www.revue-etudes.com/Arts_et_philosophie/Qu_est-ce_que_la_confiance_/7498/12570` 2010. [Consulté le 20 juin 2013]

4.1 Sécurisation des données via l'utilisation de plusieurs fournisseurs d'informatique en nuage

4.1.1 Motivations d'un client à utiliser plusieurs CP

Les motivations des entreprises pour adhérer à un environnement multi-fournisseurs [2] sont diverses et variées. Aucune réponse préconçue pourrait nous inciter à privilégier un environnement uni-fournisseur ou multi-fournisseurs.

Notre propre analyse nous porte à dire que la principale motivation à utiliser un environnement multi-fournisseurs repose sur deux facteurs :

1. La réduction du risque d'une perte des données confiées ou la crainte d'une mauvaise gestion de celles-ci. Rappelons-nous qu'une perte des données reste la préoccupation majeure des clients du cloud (sous-section 2.3.1).
2. L'augmentation de la disponibilité des services.

Les données d'une entreprise peuvent être dupliquées à différents endroits. Prenons un exemple simple, où l'utilisation d'un environnement multi-fournisseurs présente l'avantage qu'une entreprise voit ses données répliqués :

- Au niveau du disque dur, un cloud provider peut faire appel aux technologies de redondance RAID (RAID 5 et RAID 1) ;
- Au niveau serveur (physique ou virtualisé), un CP peut en cloner les données vers plusieurs autres serveurs ;
- Sur le plan géographique, un seul fournisseur de cloud peut détenir différentes répliques du même datacenter ;
- Naïvement, nous pourrions conjecturer que l'utilisation combinée de n fournisseurs de cloud multiplie par n, c'est-à-dire proportionnellement, les deux niveaux de redondance sus-mentionnés. Cette situation peux s'avérer plus complexe selon le cas. Ainsi, imaginons deux fournisseurs de cloud : l'un possède deux datacenters répliqués, alors que l'autre un seul datacenter. L'effet n'est donc pas multiplié par deux. La situation deviens pire lorsque l'un de ces deux fournisseurs emploie des ressources de l'autre fournisseur. Si ce dernier périclite, alors l'utilisateur se retrouve sans ses données.

---

2. Par définition, un environnement uni-fournisseur signifie que le consommateur s'en remet à un seul prestataire pour la gestion et le traitement de ses données dans le cloud. Plusieurs organisations composent un environnement multi-fournisseurs (Amazon, Rackspace, Google, etc.) ; elles collaborent mutuellement (ou non) pour livrer le service à l'unique client.

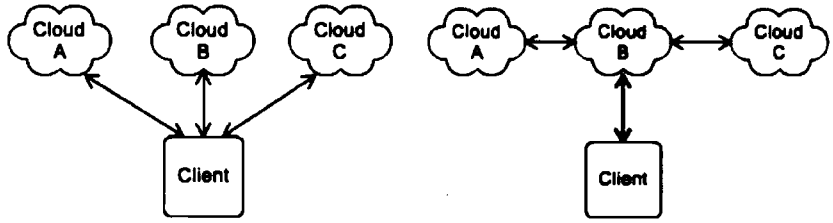
Réduction des risques Puisque plusieurs fournisseurs détiennent au minimum une copie de chaque donnée, une faillite de l'un d'entre eux ne désavantagerait pas directement le consommateur puisque celui-ci ne se trouverait pas dans un état d'enfermement chez son prestataire unique (*lock-in*). Il devra cependant revoir la distribution de ses données vers les autres fournisseurs de service.

Augmentation des temps de disponibilité Une entreprise cliente qui souhaite augmenter la disponibilité des services dans certaines zones géographiques souscrit, par exemple, à une offre de stockage chez le fournisseur B implanté dans cette zone puisque le fournisseur A ne l'est pas.

Limites D'après l'organisme SearchCloudComputing (56), de nouveaux problèmes logistiques se posent. D'une part, les fournisseurs de cloud n'offrent pas toujours un niveau de service identique (par exemple, en terme de disponibilité, de reprise d'activité, de standards, etc.). Ainsi, travailler avec plusieurs d'entre eux mènera l'entreprise cliente à utiliser uniquement les services supportés par tous ses fournisseurs. D'autre part, la négociation des contrats avec ces différents fournisseurs peut compliquer la tâche de l'entreprise. Enfin travailler de concert avec plusieurs fournisseurs apporte de nouveaux problèmes au niveau de l'accès aux ressources à distance. Cet aspect fera l'objet de la sous-section suivante.

4.1.2 Collaboration multi-fournisseurs en nuage

Cette collaboration peut exister à deux niveaux distincts : au niveau de l'entreprise cliente avec chacun de ses fournisseurs de cloud (figure 4.1a) ou à l'échelle des fournisseurs de services en nuage pour un même client (figure 4.1b).



(a) Collaboration du client avec chaque fournisseur de cloud

(b) Collaboration entre les fournisseurs de cloud pour un même client

FIGURE 4.1: Les différents niveaux de collaboration

Interopérabilité au niveau de l'entreprise cliente à l'égard ses fournisseurs de cloud L'entreprise cliente doit s'adapter aux différentes API et technologies fournies par les CP : elle doit mettre en place des outils d'interopérabilité vers ses différents fournisseurs d'informatique en nuage (figure 4.1a). Voyons un exemple concret. Pour récupérer un objet chez un fournisseur A, le client est contraint d'utiliser la méthode :

```
instance->getObject($key, $bucket)
```

alors que dans le cas du fournisseur B, le client se conforme à la méthode :

```
instance->get_object($key, $bucket)
```

Nous sommes face à un problème de nommage. Le remède consisterait donc à simplifier les échanges entre un client et ses nombreux prestataires d'informatique en nuage en adoptant un langage commun (norme) pour ces fonctions. La *Storage Networking Industry Association* (SNIA) a élaboré un rapport technique « Cloud Data Management Interface » (54) qui fournit aux applications utilisant des services de stockage en cloud des interfaces (au sens de la programmation) à prendre en considération pour créer, mettre à jour, retrouver et supprimer des informations. Cependant, un fournisseur de cloud qui n'utiliserait pas cette norme serait contraint de mettre à jour ses logiciels pour s'y conformer. Nous sommes confrontés à deux choix :

1. Le premier serait d'utiliser un métalangage, **chez le client**, et d'effectuer la conversion du nom des fonctions des fournisseurs de cloud. Par exemple, une norme impose que la fonction d'écriture d'une donnée soit nommée A et un fournisseur de cloud a nommé cette même fonction B. Le métalangage assurerait la traduction A -> B pour chaque appel du client pour la fonction A. Cette solution est coûteuse pour le client qui ne possède pas nécessairement l'expertise interne pour sa mise en œuvre.
2. Le second serait de modifier directement, **chez le fournisseur de cloud**, le nom des fonctions de la bibliothèque pour être conforme à la convention de nommage. Dans ce sens, chaque fournisseur d'informatique en nuage devrait réécrire sa bibliothèque. Cela implique de nombreux coûts supplémentaires pour ce fournisseur et une obsolescence de toutes les fonctions utilisées par leurs clients !

Notons que la fondation Apache a mis au point un projet « Libcloud » écrit en Python et qui vise à une harmonisation dans l'utilisation des différentes API de fournisseurs d'informatique en nuage [3]. Sur la page du projet, l'on apprend que de nombreux fournisseurs de services en nuage emploient « Libcloud ».

---

3. Voir : Apache : Apache Libcloud [En ligne]. Disponible `http://libcloud.apache.org/` 2013. [Consulté le 30 juin 2013]

Collaboration au niveau des fournisseurs d'infonuagique pour le client Dans ce cas, ces prestataires sont contraints d'offrir un seul point d'accès aux services (conformément à la figure 4.1b). Cependant, si ces quelques fournisseurs de cloud travaillent de concert, ils sont tenus d'adopter un langage commun pour les échanges en raison des mêmes problèmes de nommage rencontrés ci-dessus. L'article (8) s'attaque à cette problématique : ce langage commun passe par l'adoption de standards technologiques et de bonnes pratiques de management communes, ce qui est pratiquement impossible puisque chaque prestataire d'informatique en nuage dispose de ses propres technologies/API/modèles d'administration qui lui donne un avantage concurrentiel. De plus, en cas de litige, les différents fournisseurs de cloud pourraient se rejetter mutuellement la faute, au détriment du client. Dans ce contexte, la gestion des données qui est distribuée chez plusieurs fournisseurs d'informatique deviendrait encore plus opaque pour le client (perte de visibilité et de contrôle), ce qui n'est pas souhaitable.

4.2 Confiance et fournisseur d'informatique en nuage

Dans cette section, nous évaluons les facteurs mesurant la confiance envers un fournisseur d'infonuagique (4.2.1), puis nous dressons une liste d'incertitudes qu'a un client à utiliser des services offert par un fournisseur de cloud (4.2.2).

4.2.1 Mesure de la confiance

La confiance peut, selon nous, s'évaluer à l'aide de facteurs tangibles - qui sont mesurables, quantifiables et concrets — d'une part, et d'autre part via des facteurs intangibles - qui sont non mesurables, immatériels — c'est-à-dire uniquement qualifiables et voués à une certaine part de subjectivité de la part de celui qui émet le jugement.

Dans le monde des affaires, les entreprises d'un secteur particulier qui souhaitent se démarquer de la concurrence jouent évidemment sur ces deux aspects. Nous énumérons ci-après les facteurs tangibles qui peuvent, à notre avis, asseoir une relation de confiance vis-à-vis d'un prestataire de cloud :

- Certification ;
- Conformité aux lois ;
- Possibilité de négocier les SLA ;
- Audit externe régulier par des organismes de certification tiers indépendants.

L'ensemble de ces facteurs ont été définis en 2.2.2. Les facteurs intangibles relèvent du vécu de la clientèle ou de la culture d'entreprise. Ils concernent :

- l'**histoire** de l'entreprise qui souligne les faits marquants constatés tout au long de son existence ;
- La **réputation** de l'entreprise qui reflète l'opinion que se fait le consommateur de cette enseigne ou qui se forge à travers l'image de marque.

La confiance malheureusement peut se perdre rapidement. Prenons l'exemple de l'entreprise SONY [4] : crée en 1946, elle excelle sur de nombreux marchés comme celui de la musique, du jeu vidéo, de l'informatique, du cinéma (facteur intangible histoire et réputation) etc. De renommée mondiale, elle reçoit la confiance de nombreux clients (facteur intangible réputation). Cependant, elle fut victime en 2011 d'un acte de piraterie sur sa plateforme de jeux multijoueurs Play Station Network (PSN) par un groupe d'hacktiviste nommé Anonymous [5]. Ces pirates ont subtilisé des millions de comptes comportant entre autres les informations bancaires appartenant aux utilisateurs du service PSN, déstabilisant ainsi l'enseigne en ébranlant la confiance des utilisateurs pour une marque. Il s'en est suivi une baisse de l'action de Sony (liée à la confiance des actionnaires).

4.2.2 Évaluation de la confiance envers un fournisseur de cloud

La confiance accordé à un fournisseur d'informatique en nuage reste le sujet épineux lors de l'outsourcing des données de l'entreprise. Les différentes incertitudes et craintes conduisent à soulever les questions suivantes :

- Peut-on vérifier si le fournisseur de cloud ne rétrocède pas les données à des tiers dans un but marketing ? Les motivations et domaines marketing des CP sont nombreux : plus un prestataire de cloud détient d'information sur les activités de ses clients plus il sera équipé pour les cibler. Depuis quelques années, le paradigme du Big Data permet

4. Voir : Sony : Historique de l'entreprise [En ligne]. Disponible `http://www.sony.ch/lang/fr/article/id/1073396360844` 2012. [Consulté le 25 juin 2013]

5. Voir : Le monde, AFP : Des données personnelles piratées sur le PlayStation Network [En ligne]. Disponible `http://www.lemonde.fr/technologies/article/2011/04/27/des-donnees-personnelles-en-danger-sur-le-playstation-network_1513235_651865.html` 2011. [Consulté le 25 juin 2013]

via des techniques de forage de données d'extraire de la connaissance à partir de bases de données très volumineuses [6]. In fine, il paraît très difficile à une entreprise cliente de s'y retrouver et d'obtenir la certitude que le fournisseur de cloud fait preuve de transparence sur l'utilisation des données hébergées.

- Le prestataire d'informatique fait-il appel à des fournisseurs externes pour stocker les données du client ? Si un CP n'a plus les ressources matérielles pour héberger les données ou un autre prestataire lui propose une excellente offre d'hébergement alors il peut déporter les données du client hors de son champ de contrôle. Cependant, si la gestion des ressources alloués à ce prestataire reste dans son champs de contrôle alors le risque peut être considéré comme acceptable.
- En cas de perte d'une données, le CP responsable masque-t-il volontairement cette situation à sa clientèle dans l'espoir d'éviter de dégrader son image de marque ? Des mécanismes de vérification de possession de la donnée auprès du consommateur lèverait le doute.
- Un fournisseur peut-il supprimer une donnée rarement sollicitée sans en informer au préalable son client ? Effectivement, un CP peut opérer de cette manière pour économiser ou obtenir plus d'espace de stockage et garantir ainsi l'élasticité.

Un audit des pratiques du prestataire de cloud devrait se réaliser de façon régulière pour garantir le respect du SLA et par là même, améliorer les relations fournisseur/client. D'après Gartner et son Hype Cycle, à l'horizon 2015 20 % des services en nuage s'effectueront via des tiers en nuage, 46 % des répondants affirment que les tiers de confiance joueront un rôle de plus en plus important (30). Ainsi, un tiers en nuage semble nécessaire.

Cependant, si un client sceptique ne fait pas confiance à un ou plusieurs fournisseurs de cloud, pourquoi ferait-il confiance à un tiers de confiance en nuage ? Selon nous, il est impossible d'envisager un tiers en nuage sans confiance ; le client se retrouverait toujours sans assurance. De plus, un tiers sans confiance équivaut simplement à ajouter une couche inutile entre le client et le fournisseur d'informatique en nuage. La suite de ce chapitre s'attèle à formaliser, évaluer et définir les mécanismes qui assurent la confiance d'un client envers ce tiers.

---

6. Nous employons « très volumineuses » pour signifier que les bases de données relationnelles que nous connaissons (type SQL) ne sont plus assez performantes pour traiter efficacement ce volume de données. Ainsi, de nouveaux paradigmes en base de données émergent comme le modèle de base de données No-SQL.

4.3 Architecture fondée sur un tiers de confiance

4.3.1 Nécessité de l'utilisation d'un tiers de confiance

Notre analyse nous porte à constater que :

1. les fournisseurs d'informatique en nuage n'adoptent pas de façon systématique des standards d'interopérabilité ;
2. la gestion des différentes technologies et API des CP est un calvaire pour les entreprises clientes ;
3. il existe un réel besoin des entreprises à réduire les risques liés à leurs données ;
4. la négociation des SLA entre le client et ses divers fournisseurs est complexe ;
5. le client nécessite un arbitre en cas de litige.

Nous pensons donc qu'il est souhaitable de faire appel à un chef d'orchestre, le tiers de confiance. Si une entreprise envisageait d'utiliser un hébergement multi-fournisseurs, le TDC aurait notamment pour rôle de proposer une interface commune de gestion des différentes API offertes par les fournisseurs de services. Dans ce sens, le fardeau de l'interopérabilité ne se retrouverait ni chez le client ni chez le fournisseur d'infonuagique mais chez ce TDC : il offrirait ainsi une simplification indéniable de l'accès aux services.

4.3.2 Formalisation du concept de tiers de confiance

Le TDC propose des mécanismes garant de la confiance pour le client sceptique dans l'utilisation de services en nuage. Ainsi, le rapport technique SP500-292 du NIST (36) définit une architecture de référence pour les composants du cloud. On y indique qu'un TDC accepte de simplifier les nombreux services proposés par un prestataire d'informatique en nuage en adoptant un rôle intermédiaire efficace dans :

- **Le service médiateur :** le TDC fournit des mécanismes qui améliorent les services des différents fournisseurs d'infonuagique (meilleure sécurité, fiabilité, etc.).
- **L'agrégation de services :** le TDC combine et intègre divers services en un ou plusieurs nouveaux services. Dans cette situation précise, le tiers peut constituer un point d'accès unique pour le client dans le cas où les transactions effectuées entre ce dernier et le fournisseur de services en nuage transitent par ce tiers.
- **L'arbitrage des services :** le TDC est autorisé à choisir parmi une liste un fournisseur spécifique et à proposer au client final l'offre la mieux adaptée (alignement des affaires).

Une entreprise souhaitant migrer vers le cloud doit savoir si elle est capable de gérer ces différentes et si elle possède l'expertise interne adéquate à la gestion de l'ensemble des services en nuage auxquelles elle souscrit.

4.4 Établissement des obligations d'un tiers de confiance

4.4.1 Définition des exigences de bases

Nous confions des problèmes récurrents à un tiers de confiance capable de les résoudre. Pour réaliser l'ensemble de son travail dans des conditions optimales, ce tiers doit impérativement posséder les rôles que nous détaillons dans la liste ci-après :

Agrégagation des services Le tiers de confiance agrège divers services en nuage pour proposer une meilleure interopérabilité et portabilité des données. Nous reprenons ici l'une des préconisations de la NIST (SP500-292 (36)).

Négociation du SLA Le tiers de confiance s'exprimerait en tant que porte-parole d'une entreprise cliente : il négocierait le meilleur SLA possible avec le prestataire d'informatique en nuage de telle manière que le niveau de qualité soit le plus près des besoins et attentes du client.

Conformité du SLA et des actions du CP Le tiers de confiance vérifie si le fournisseur respecte ce SLA en ce qui concerne la localisation des données dans un pays ou une juridiction donnée.

Garantir l'intégrité des données Il s'assure de l'intégrité d'un fichier hébergé chez un prestataire de cloud ou il audite les échanges en analysant, par exemple, les journaux de transaction chez le fournisseur de cloud : en cas de problème d'intégrité, le tiers de confiance peut rejouer une transaction ce qui aura pour effet de recouvrer la donnée corrompue.

Garantir l'imputabilité des actions Le tiers de confiance fournit des mécanismes permettant de savoir qui est à l'origine de quelles actions, quand, comment et dans quelles circonstances.

Transparence et standardisation des politiques internes du tiers de confiance Le tiers de confiance applique une politique de gouvernance interne qui se fonde sur des standards éprouvés et communément admis.

Migration chez un fournisseur de cloud Le tiers de confiance offre des mécanismes qui autorisent la migration d'informations vers un autre prestataire de cloud. L'hébergement multi-fournisseurs est considéré comme une option efficace pour faciliter la migration. L'adoption d'une interface commune de stockage (au sens de la programmation) est un bon moyen pouvant aboutir à la standardisation. Nous évitons ainsi une situation d'enfermement chez un fournisseur donné. Puisqu'il existe bien des interfaces communes, la portabilité des données ne peut qu'être meilleure.

Indépendance vis à vis des fournisseurs de cloud Pour opérer en toute transparence et objectivité, le tiers de confiance ne doit pas montrer d'affiliation particulière envers l'un ou l'autre fournisseur d'informatique en nuage. Sa priorité consiste donc à faire preuve de neutralité dans les solutions qu'il suggère.

Indépendance des vérifications Il encourage fermement l'indépendance des agents de vérification (tests) et d'audit pour éviter par exemple la falsification des rapports. Il devrait proposer ses propres outils et ne pas accorder au fournisseur de services le soin d'auditer lui-même la sécurité des données ainsi que les privilèges consentis.

4.4.2 Définition des mécanismes optionnels

Le tiers de confiance peut, de manière optionnelle, proposer les mécanismes suivants :

Confidentialité des données Le tiers de confiance fournit aux clients des mécanismes qui permettent de protéger la confidentialité de leurs informations ; il est tenu de respecter les préconisations de la norme PCI DSS par exemple (vue en 2.5). Il suggère des solutions cryptographiques si elles ne sont pas proposées par le fournisseur d'informatique en nuage ou si le CP n'offre pas de sécurisation des échanges de bout en bout.

Anonymat du client Le tiers de confiance peut proposer des mécanismes qui garantissent l'anonymat du client auprès du fournisseur de services en nuage. Le tiers de confiance joue le rôle de mandataire pour le client c'est-à-dire qu'il souscrit à des offres de services dans le cloud pour le compte du client sans révéler l'identité de ce dernier. L'anonymat est assuré chez le fournisseur de cloud uniquement (non chez le TDC).

Recommandations et plan de gestion de données Il propose de part son expertise, le meilleur fournisseur de cloud en fonction des besoins du client en terme de redondance, localisation des données, performance, etc.

Positionnement géostratégique Si le tiers de confiance est un intermédiaire entre le fournisseur de cloud et le client dans la chaîne de flux de données, alors il doit pouvoir délivrer les données au plus près de l'utilisateur pour réduire les temps de latence et assurer une plus grande disponibilité des services.

4.4.3 Quels sont les inconvénients de l'utilisation d'un TDC ?

D'après les laboratoires RSA (10), le tiers de confiance en nuage tend à alourdir l'architecture. Par un mauvais design de l'architecture (réseau, serveurs, etc.), une perte de performance peut se produire par rapport à un modèle traditionnel d'informatique en nuage. À contrario EMC (acquéreur de RSA) suggère dans sa présentation que les entreprises doivent faire appel à un tiers de confiance (16) : ceci apporterait à toute entreprise une meilleure visibilité et un contrôle supérieur via l'agrégation et la simplification des services ; ces deux arguments constituent pour EMC les paradigmes de la confiance dans le cloud. Par ailleurs, le tiers de confiance est amené à effectuer de nombreux tests et audits ce qui laisse présager une augmentation de la bande passante.

Dépendamment de l'architecture employée, nous pourrions constater un goulot d'étranglement réseau si tous les flux passaient par le même point. De plus, le TDC constitue un unique point de défaillance. Nous devons donc envisager des mécanismes de redondance aussi bien au niveau des services proposés que des infrastructures (bâtiments, serveurs, etc.).

Un tiers de confiance qui aurait trop de pouvoir pourrait abuser de cette confiance : des fuites d'information sont possibles. Par exemple, le tiers de confiance chiffre les données pour le client avec une clé que ce client ne gère pas. Ainsi, ce TDC — s'il développe un partenariat privilégié avec un fournisseur de cloud — peut vendre les informations de ses clients sans leur approbation. Le problème de la confidentialité des données est déporté à un autre endroit et demeure non-résolu.

Les services offerts par le tiers de confiance constituent un frein à l'innovation. Par exemple, la mise en place de technologies et de protocoles par le tiers de confiance peut mener à une standardisation de par leur utilisation massive ; une entreprise concurrente n'aura pas forcément la volonté d'investir dans le développement de tels mécanismes.

4.5 Revue de cas de tiers de confiance

Plusieurs tiers de confiance sont actifs sur le marché du cloud computing. Citons CloudLock [7] ou CipherCloud [8]. Il serait intéressant de vérifier dans quelle mesure ces entités respectent la liste des exigences de base (énoncée en 4.4.1).

4.5.1 CloudLock

CloudLock est une société américaine basée dans le Massachusetts. Son cœur de métier est centré autour de la sécurisation des services SaaS offerts par Google comme GoogleApps ou Google Drive : elle réalise des audits de sécurité sur les fichiers et s'assure que ceux-ci sont bien en accord avec la loi américaine et suffisamment sécurisés sur Google Drive. Ainsi, CloudLock monitore les droits et conditions de partage sur les fichiers hébergés via GoogleApps puis affiche à l'attention du client un rapport d'alerte lorsqu'un ou plusieurs fichiers ne sont pas en adéquation avec les prérogatives édictés par PCI DSS et PII.

Ces services garantissent ainsi une forme d'imputabilité et proposent à la clientèle une interface simple de gestion et de vérification de la conformité. Elle ne propose pas de modèle de distribution IaaS ou PaaS puisqu'elle joue le rôle de service médiateur des produits SaaS de Google. De plus, via sa solution de pare-feu pour GoogleApps, CloudLock entreprend une revue de sécurité des diverses applications installées via le service Google Apps Marketplace [9] en fournissant au client des alertes lorsqu'une de ces applications présente un risque.

Par ailleurs, ses services sont certifiés SSAE16 (Type I) [10] par une grande organisation de certification. Toutefois on ne précise pas laquelle [11]. Nous supposons donc qu'il s'agit d'une auto-certification. Enfin, notons qu'un rapport SSAE16 de type I atteste que l'entreprise utilise des contrôles internes adaptés sans toutefois démontrer leur efficacité (prévue dans le rapport de SSAE16 - type II, l'efficacité des contrôles est évaluée sur une période de six mois).

---

7. Lien vers leurs solutions : `www.ciphercloud.com`

8. Lien vers leurs solutions : `www.cloudlock.com`

9. Google Apps Marketplace (`https://www.google.com/enterprise/marketplace/`) permet d'installer et de déployer des applications cloud spécialement conçues pour les entreprises.

10. Équivalent à ISAE3402 vu en 2.5.

11. Voir : Burke : CloudLock Completes SSAE 16 SOC 1 Audit [En ligne]. Disponible `http://blog.cloudlock.com/2012/05/31/cloudlock-completes-ssae-16-soc-1-audit/` 2012. [Consulté le 30 juin 2013]

D'après le site de la société, son infrastructure est implantée dans les centres de données de Google, rien d'étonnant à cela puisque tout le coeur de métier de CloudLock se fonde sur les produits Google : son logiciel s'appuie d'ailleurs sur un produit SaaS, Google Apps. Ainsi, nous supposons qu'expérimentalement les performances sont bien réelles vu la proximité entre le TDC et le fournisseur de services. De plus, la traditionnelle peur sur l'enfermement chez un fournisseur de cloud n'inquiète pas le tiers de confiance puisqu'il ne gère pas les données directement, il ne les crypte pas par exemple. Le lock-in se situe un niveau plus bas c'est-à-dire chez Google. L'architecture adoptée par CloudLock s'inspire très largement de celle proposée, plus loin dans le chapitre 5, par la figure 5.1. Ce point positif toutefois entraîne un point négatif : CloudLock n'offre pas véritablement de système de cryptage de données. Dans ce sens, la confidentialité des informations n'est pas garantie et le tiers de confiance peut accéder à toutes les données d'un client.

La société CloudLock affirme ne pas lire le contenu des documents sur Google Drive, cependant, elle collecte les méta-données de ces fichiers : « In providing the Service to you, CloudLock will analyze, map and collect meta-data only relating to the data stored on your Google Domain, and the manner in which data is stored and used on your Google Domain ("Customer Meta-Data") » [12]. D'après l'API développeur de Google Drive [13], les méta-données d'un document concernent au minimum :

- Le titre du fichier ;
- La description du fichier ;
- Le type MIME du fichier [14].

Nous savons que la société CloudLock peut traiter des données bancaires (PII) et médicales (HIPAA) ; l'accès que possède CloudLock vis-à-vis de ces méta-données est donc une possible atteinte à la confidentialité de ces données.

CloudLock s'est auto-certifié Safe Harbor [15] c'est-à-dire que des données privées appartenant à des entreprises européennes peuvent être traitées chez CloudLock. Cependant, comme

---

12. Source : `www.cloudlock.com/terms-of-service`

13. Voir : `https://developers.google.com/drive/v2/reference/files/insert`

14. Le type MIME, à l'origine utilisé pour typer les pièces jointes d'un mail, est aussi utilisé dans les transferts de document HTTP. On y retrouve le type de fichier ainsi que l'extension associée.

15. Voir : Burke : CloudLock Announces Safe Harbor Compliance [En ligne]. Disponible `http://blog.cloudlock.com/2012/01/19/cloudlock-announces-safe-harbor-compliance/` 2012. [Consulté le 3 Février 2014]

mentionné en 2.2, cette sphère de sécurité n'offre aucune garantie puisqu'elle se fonde sur l'auto-certification. Ainsi, CloudLock met en place le cadre de référence édicté par l'Union Européenne puis la société atteste par *elle-même* de sa propre conformité à ce cadre.

Les échanges entre le client et le TDC s'effectuent via HTTPS ce qui signifie que les données font l'objet d'une sécurisation durant leur transfert de bout-en-bout.

Une négociation du SLA est impossible et aucune indication contraire n'est relevée sur le site. À la lecture de ce SLA, [16] nous constatons que seule la haute-disponibilité du service est soulignée. La gestion de l'intégrité des données et la confidentialité ne sont absolument pas évoquées, ce qui apparaît comme paradoxal pour une société censée offrir de la sécurité et de la confiance dans le cloud, etc. Qui plus est, un lien vers ce SLA pointe vers Google Drive et affiche une erreur 404. Enfin, nous ne possédons aucune information sur le SLA négocié entre CloudLock et Google. Ce manque total de transparence impacte directement sur la confiance du client envers CloudLock.

4.5.2 CipherCloud

L'entreprise Ciphercloud située à San José aux Etats-Unis tient également un rôle de tiers de confiance. Son coeur de métier s'oriente autour du cryptage de données pour des solutions aussi diverses que Gmail, Microsoft Office 365, Amazon AWS, etc. CipherCloud ne stocke aucune donnée au sein de son infrastructure physique.

À première vue, Ciphercloud est capable d'offrir des garanties en matière de confidentialité des informations. Qui dit confidentialité, dit chiffrement des données. Deux choix sont disponibles : les clés de chiffrement des informations demeurent chez le client ou chez CipherCloud. Ciphercloud respecte bien la norme FIPS 140-2 qui préconise une utilisation de l'algorithme de chiffrement AES avec une clé de 256 bits. Notons ici l'excellence de la pratique puisque le tiers de confiance et le fournisseur de services ne possèdent que l'accès aux données chiffrées (non aux autres informations) si le client décide de gérer les clés de chiffrement en interne de son entreprise. L'entreprise cliente détient un plus large contrôle et décide de son propre chef de délivrer ses informations à des fins juridiques par exemple. La principale faiblesse de CloudLock est ainsi comblée.

---

16. Disponible en suivant ce lien : `www.cloudlock.com/cloudlock-sla`

Néanmoins, pour ce qui a trait aux courriels nous savons qu'ils sont chiffrés pour un usage interne-interne et pour des transactions externes-internes. Dans le cas d'un mail envoyé hors de l'entreprise, celui-ci est décrypté via un proxy inverse hébergé par CipherCloud. CipherCloud doit, pour effectuer l'opération, posséder les clés de chiffrement, elle a donc accès au contenu des courriels. Dans cette mesure, le client n'a pas d'autre choix que d'utiliser la plateforme de CipherCLoud pour l'hébergement des clés. Nous pouvons transposer ce modèle aux autres services proposés puisque nous sommes dans le cadre d'une collaboration externe à l'entreprise. À l'évidence, une atteinte à la confidentialité des données est possible. Par ailleurs, CipherCloud propose des outils pour l'audit des échanges entre le client et le fournisseur de services en nuage. Elle stocke ainsi les transactions sur un serveur de journalisation afin d'assurer l'imputabilité. Cependant, CipherCloud ne se prête pas au rôle de conseiller et ne fournit pas d'outils pour donner des indications sur le niveau de sécurité à adopter, etc.

Concernant la localisation de CipherCloud, nous apprenons qu'ils sont implantés en Europe (Allemagne, Grande-Bretagne, Lituanie) et au Canada. Le site de CipherCloud n'indique pas la possibilité de négocier un SLA avec un client [17]. De ceci, nous retenons ce que nous savions déjà : le SLA ne contient que les exigences liées à un haut taux de disponibilité et une faible latence des services. Mis à part fournir des données toujours intègres à la clientèle, aucune mention n'est faite sur les obligations du tiers de confiance par rapport à un défaut ou une perte d'informations. Si un tel évènement se produisait, le client serait pris en étau, par une bataille juridique, entre son TDC et son fournisseur de services. De plus, nous ne voyons aucune indication sur les métriques et technologies mises en place pour assurer l'intégrité des données. Globalement, nous constatons un manque de clarté dans la manière d'opérer puisqu'il est difficile de trouver des indications sur la manière dont les services sont livrés. Le constat est malheureusement sans appel car aucune mention n'indique que CipherCloud est certifiée ISAE3402. Le client n'a donc aucune assurance sur le fait que la société met en place des contrôles internes adaptés et adopte de bonnes pratiques de management.

17. SLA disponible sur ce lien : `http://www.cipherwave.co.za/CipherCloud%20Hosting%20SLA.php`

4.6 Conclusion

Ce chapitre apporte des réponses sur la nécessité de collaborer avec un tiers de confiance qui - par essence - vend de la confiance via des métriques et de bonnes pratiques. La confiance se fonde sur des facteurs quantifiables que nous qualifions de tangibles et des facteurs qualifiables que nous qualifions d'intangibles. Le manque de transparence et la complexité des services offerts par les principaux acteurs du cloud contraignent les entreprises à recourir à des tiers qui soient capables de respecter la liste des exigences que nous venons d'établir. La revue des tiers de confiance CloudLock et CipherCloud montre que chaque acteur amène son lot d'avantages et d'inconvénients : le SLA au centre de la collaboration reste toujours trop hermétique et ne concerne que la haute-disponibilité ainsi que les temps de latence des services. Cette constatation amène notre réflexion sur la définition d'un courtier en sécurisation des données dans le cloud.

CHAPITRE V

DÉFINITION D'UN MODÈLE DE TIERS DE CONFIANCE POUR LA SÉCURISATION DES DONNÉES DANS LE CLOUD

Ce chapitre propose une architecture de sécurité qui fait intervenir trois types d'acteurs : les fournisseurs d'informatique en nuage, leurs clients et le tiers de confiance. Ce chapitre introduit également la notion de *Cloud Security Broker* (CSB) : c'est une entité qui propose des mécanismes de sécurité au client. Le tiers de confiance a un rôle de monitoring et d'audit. Ce TDC/CSB représente le trait saillant de l'architecture puisqu'il va fournir un ensemble de nouveaux services à ses clients tout en offrant une sécurité des données. Nous utiliserons la liste établie en 4.1.1 pour guider notre réflexion et justifier l'intérêt d'une telle architecture.

La section 5.1 définit un catalogue critérié de services et propose une liste des fonctions que ce tiers devra remplir. La section 5.2 définit et débat sur les différentes topologies existantes pour notre troisième acteur. La section 5.3 détaille les composants du CSB en nuage et interne puis dresse un portrait des interactions entre ces deux parties et le fournisseur d'informatique en nuage. La section 5.4 détermine une architecture P2P en cloud qui utilise à bon escient ce catalogue de services ainsi qu'un environnement sécuritaire créé par notre CSB. La section 5.5 argumente sur les assurances supplémentaires qu'apporte notre solution aux clients du cloud computing et qui garantissent une meilleure transparence et confiance des usagers. Enfin, la section 5.6 présente deux études de cas : la première considère une entreprise faisant appel à une solution de stockage qui emploie un modèle client/serveur, la deuxième s'intéresse à une entreprise avec une solution de stockage qui utilise un modèle P2P.

5.1 Catalogue critérié de services et niveau de confiance

5.1.1 Définition d'un catalogue critérié de services

Comme nous le présentions dans la sous-section 1.4.4, l'intercloud autorise la création d'un cloud de cloud rendant possible le partage des ressources entre différents fournisseurs d'informatique en nuage pour répondre par exemple aux besoins d'élasticité. Dans le cadre de l'architecture proposée, nous ne visons pas les mêmes objectifs que les auteurs de l'article (8). Rappelons que ces derniers envisagent d'élaborer une fédération de clouds. Dans notre cas, nous pourrions nous inspirer de l'élément Intercloud Root pour construire notre catalogue de services. Ce dernier serait hébergé par le CSB. L'entité « Intercloud Exchange » ne nous intéresse guère ici : en effet, son unique rôle est de « faciliter le dialogue et la collaboration au travers d'un environnement hétérogène de fournisseurs d'informatique en nuage » (8), ce qui n'est pas utile ici.

5.1.2 Proposition d'un catalogue critérié de services

Nous suggérons de présenter un catalogue critérié de services, géré par le CSB, qui répertorierait entre autres :

- Les **ressources** dont disposent les fournisseurs d'informatique en nuage, incluant le niveau de sécurisation des données et de redondance proposé.
- Les **politiques de gestion** de chaque fournisseur d'informatique en nuage indiquant, si le prestataire d'informatique en nuage utilise une politique de gouvernance comme ITIL ou s'il propose un plan de reprise ou de continuité d'activités.
- Le **niveau de certification et de conformité normative** de chacun, en fonction de différents contextes : capacité de gestion de données critiques, certifications proposées, respect des lois, etc.
- Les **technologies** libres ou propriétaires adoptées par chaque fournisseur d'informatique en nuage.
- La **localisation** des centres de données et le cas échéant la possibilité de la spécifier : il est nécessaire de vérifier si le fournisseur d'informatique en nuage propose de choisir l'emplacement géographique où seront hébergées les données ; si ce n'est pas le cas, il faut simplement préciser où les données sont situées (raisons légales).
- Les **SLAs** proposés par chaque CP : pour aller plus loin, une négociation avec les fournisseurs de cloud sur les objectifs que ces SLAs doivent supporter (objectifs vus en 2.2.2) est préconisée.

- Les **performances** qu'offre chaque fournisseur d'informatique en nuage : il serait utile de faire des tests de latence, débit, disponibilité mesurée, etc.

Le CSB est ainsi capable d'obtenir des résultats qui lui permettront de choisir, en fonction de métriques, un fournisseur selon les besoins d'un client. Par ailleurs, il est possible de mesurer le niveau de confiance attribué à un fournisseur en particulier. Le CSB en nuage héberge le catalogue de services des centres de données lié à sa zone géographique.

5.1.3 Apports du catalogue critérié de services

Nous proposons une méthode qui fournit aux clients une simplification des services. En effet, aucune étude de marché complexe n'est à réaliser pour le client puisque tout ce travail aura été effectué en amont par le CSB. Le client doit simplement choisir quelle solution lui parait la plus appropriée en fonction de ses exigences en termes de sécurité des données, localisation, SLA, normes à respecter, services, etc.

L'utilisation d'un CSB comme celui que nous décrivons permettrait l'approbation de protocoles et technologies publics puisque c'est un critère qui entre directement dans la notation de confiance que nous proposons. Ainsi, l'adoption massive de notre CSB mènerait à une standardisation qui garantirait — transitivement — une mise en place plus simple d'un intercloud. *In fine*, la mise aux normes de l'ensemble de la topologie du cloud s'effectuerait via la technologie et la pression positive du marché.

5.2 Considération topologique pour la mise en place de la solution

5.2.1 Généralités sur les topologies disponibles

Deux types de topologie peuvent être envisagés : l'une triangulaire, l'autre en « L ».

Architecture triangulaire

La topologie dite triangulaire a été proposée dans l'article « Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing » (57). La figure 5.1 montre cette architecture.

Cette architecture fait intervenir trois protagonistes, à savoir, le client qui souscrit à un service en nuage, le fournisseur de services qui héberge les données du client et le tiers de confiance qui a pour rôle d'évaluer et d'exposer les risques du système de stockage à la demande du client. Le tiers de confiance audite les données hébergées chez le fournisseur de services en nuage sur demande de son client qui n'a pas nécessairement la capacité pour effectuer de tels audits.

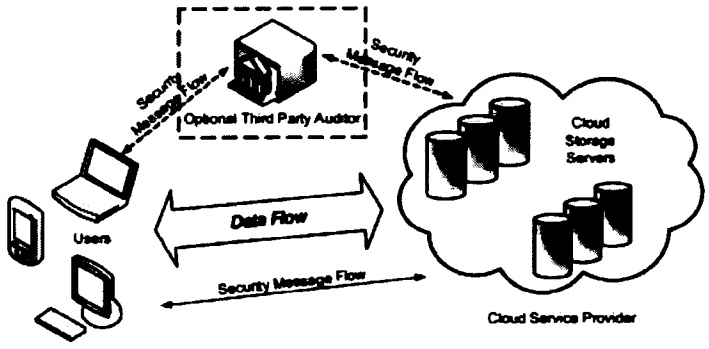


FIGURE 5.1: Architecture triangulaire qui utilise un tiers de confiance (figure tirée de (57))

Architecture en « L »

Elle utilise un CSB comme pierre angulaire de toutes les communications. La figure 5.2 montre ce principe.

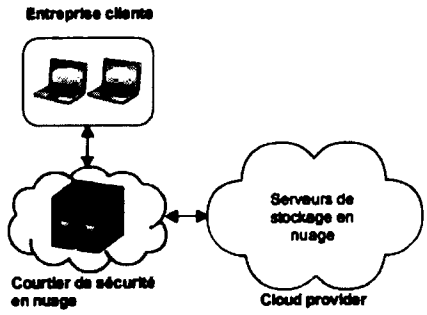


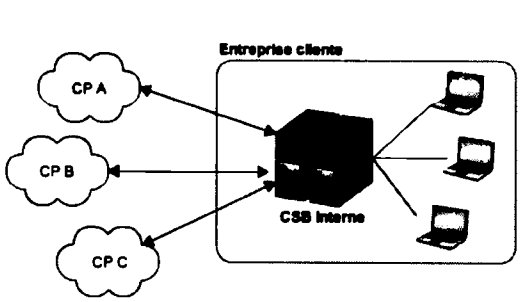
FIGURE 5.2: Architecture en « L » où toutes les données transitent via le CSB

La sous-section ci-après permet de définir de quelle manière ce CSB en « L » peut être implanté.

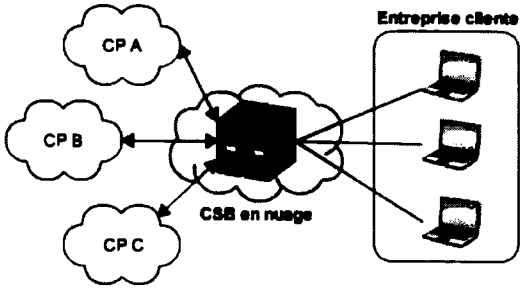
5.2.2 Implantation du CSB en « L »

Nous dressons à présent la liste des implantations en « L » réalisables par un courtier de sécurité (CSB) en nuage :

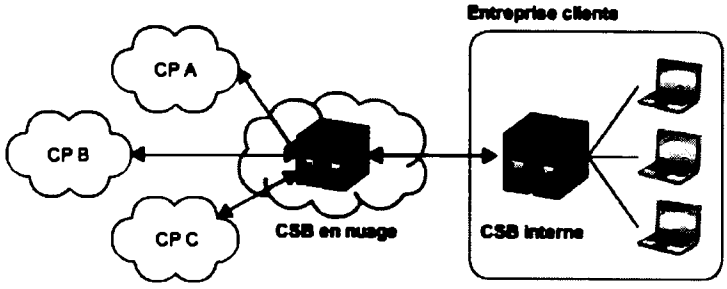
- **Une solution interne à l'entreprise** : l'application CSB est directement implantée dans l'infrastructure de l'entreprise et fait le lien entre les utilisateurs internes et les différents services en nuage (voir la figure 5.3a).
- **Une solution externe à l'entreprise** : le courtier en sécurité se situe à l'extérieur de l'entreprise (voir la figure 5.3b).
- **Une solution hybride** : la solution est présente dans l'entreprise interne et certaines actions sont effectuées depuis et vers le CSB en nuage (voir la figure 5.3c). Cette architecture hybride scinde le CSB en deux modules distincts.



(a) Cloud Security Broker interne



(b) Cloud Security Broker externe



(c) Topologie hybride

FIGURE 5.3: Les différents types de topologies disponibles pour un Cloud Security Broker

La topologie choisie doit respecter les exigences vues en 4.4.1.

5.2.3 Choix d'une topologie

Comme nous pouvons le constater, les topologies CSB interne (figure 5.3a) et CSB externe (figure 5.3b) s'apparentent à celle en « L » de la sous-section précédente. Les paragraphes ci-après présentent les avantages et inconvénients des différentes topologies.

Topologie utilisant un CSB interne à l'entreprise D'après nous, la topologie affichée en figure 5.3a où le Cloud Security Broker se retrouve entre les murs de l'entreprise n'est pas adaptée à notre situation. En effet, il n'est pas réaliste d'avoir un catalogue distribué dans ces conditions, ce qui contrevient à l'exigence « Proposer un système de classement des fournisseurs » édictée en 4.1.1. Il y aurait autant de catalogues critériés de services que de clients qui souscrivent au service que nous proposons. Dans ce sens, un nombre beaucoup trop important d'échanges réseaux (monitoring, audit, etc.) seraient requis entre tous les catalogues de services et les fournisseurs d'informatique en nuage. De plus, le CSB interne serait un point de défaillance unique et concentrerait plus de services que nécessaire. Enfin, dans la mesure où le nombre de services proposés est très important, le niveau d'expertise interne à l'entreprise indispensable au management du CSB le serait également. Cette solution ne convient donc pas dans le contexte qui vise à déporter des ressources et des compétences à l'externe. Tout l'aspect de la simplification des services, qui est une composante majeure des courtiers en nuage, serait perdue. Puisque les mécanismes habituellement offerts par un tiers de confiance (surveillance et audits des données) sont hébergés au sein de l'entreprise cliente, nous ne pouvons pas le considérer comme tel. Un audit impartial effectué par un tiers a plus de valeur que celui opéré par l'un des deux protagonistes du litige : n'oublions pas que l'un des rôles majeurs d'un TDC concerne l'arbitrage en cas de problème. L'un des points forts de cette topologie interne touche au respect de la confidentialité des données puisque le client gère lui même le chiffrement de ses informations. De plus les performances sont bonnes dans la mesure où les données ne transitent pas par un autre intermédiaire en nuage.

Topologie utilisant un CSB externe à l'entreprise Présentée en figure 5.3b, elle est plus intéressante que la précédente, car l'ensemble des services est géré par le Cloud Security Broker. Un nombre beaucoup plus restreint de catalogues critériés de services sont à gérer puisqu'il n'y a que quelques serveurs de catalogues à administrer (serveur principal plus redondance). Le CSB monitore et audite efficacement les données en transit et au repos. Cependant pour le client du CSB ce n'est pas une bonne solution pour plusieurs raisons :

1. Son catalogue de données est externe à l'entreprise. Pour chaque accès, le client doit contacter un serveur distant hébergé par le CSB en nuage, ce qui affecte la performance et donc la disponibilité.
2. Si le chiffrement des données était requis, les clés seraient hébergées chez le CSB en nuage, ce qui poserait des problèmes évidents de confidentialité puisque ce CSB aurait accès en lecture / écriture à toutes les ressources d'un client.

3. Si le CSB en nuage périclitait ou un différent apparaissait entre le CSB et le client, alors ce dernier ne pourrait récupérer efficacement ses données. Le problème de l'enfermement chez le fournisseur de services (ici, enfermement chez le CSB) n'est donc pas résolu.

Les points forts de la topologie externe sont les points faibles de la topologie interne et vice versa. Enfin, nous pensons que l'utilisation d'un catalogue de donnée interne à l'entreprise constitue une solution plus intéressante pour le client qui souhaite garder un contrôle sur ses données. La topologie hybride présentée ci-après propose une telle solution.

Topologie hybride La topologie hybride présentée par la figure 5.3c est composée de deux modules distincts :

1. Le module CSB interne : il a pour rôle de placer les données chez le fournisseur de cloud. Les données transitent directement du CSB interne vers le fournisseur de cloud sans passer par un intermédiaire. Les clés de chiffrement restent chez le client.
2. Le module CSB en nuage : il a un rôle d'audit, de conseiller, d'arbitre et d'agrégateur de services et n'a aucun accès aux données (chiffrées ou cryptées) du client ce qui garantit une confidentialité des informations du dépositaire. Si ce CSB en nuage périclite, le client a toujours accès à ses données puisque le module interne est distinct de ce CSB ce qui évite l'enfermement chez le CSB.

La figure 5.4 précise davantage la segmentation des rôles de chaque acteur.

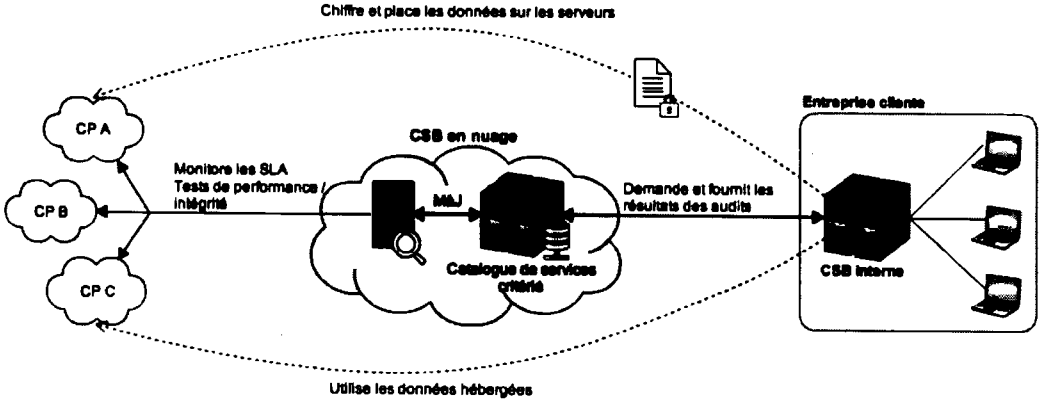


FIGURE 5.4: Mise en contexte de notre topologie hybride qui implique l'entreprise cliente, le CSB et les fournisseurs d'informatique en nuage

Cette architecture hybride correspond à l'architecture triangulaire illustrée par la figure 5.1 et semble répondre à certaines exigences vues dans la section 4.4.1. De par les différents rôles joués par le CSB en nuage, nous pouvons considérer que le CSB en nuage est un TDC. La section ci-après propose d'examiner plus en profondeur cette topologie hybride.

5.3 Solution hybride : composants et interactions

Dans cette section, nous détaillons respectivement les rôles et composants du CSB interne ainsi que ceux du CSB en nuage. Nous poursuivons notre étude sur les interactions entre le CSB interne et le CSB en nuage puis nous présentons celles entre le CSB interne et le cloud provider. Toutes les interactions entre le CSB interne et en nuage se font via des requêtes Web HTTP.

5.3.1 Rôles et composants du CSB interne

Le CSB interne est physiquement implanté dans l'entreprise cliente. Il peut prendre différentes formes :

1. Un serveur physique : il serait configuré par le CSB en nuage et livré au client, constituant ainsi une forme de solution clé en main. Les FAI adoptent la même pratique lorsqu'ils livrent leur modem/routeur à leur client.
2. Une machine virtuelle : elle serait pré-configurée par le CSB en nuage pour le client. Le client l'hébergerait sur un hyperviseur de son choix en interne de l'entreprise.
3. Un paquet logiciel disponible via un gestionnaire de dépôt (par ex., un paquet *dpkg* disponible via l'outil *aptitude* sous les distributions de type Debian) : dans ce cas de figure, le client doit avoir l'expertise nécessaire à la configuration des services.

Notons que si le client souhaite garantir la confidentialité de ses données, il devra lui-même générer la paire de clés symétriques. Si cette génération est opérée par le CSB en nuage, ce dernier à connaissance des clés : la contrainte de confidentialité ne demeure pas résolue. Par ailleurs, le CSB en nuage pourrait créer des portes dérobées [1] sur les différentes formes de CSB interne, ce qui lui garantirait un accès aux données de ses clients à leur insu.

Nous identifions plusieurs rôles pour le CSB interne, ils sont résumés à travers les points suivants :

1. Une porte dérobée est une fonction d'un logiciel, inconnue de l'utilisateur légitime, permettant l'accès à ce logiciel ou au système d'exploitation sous-jacent.

- Placer les données sur les serveurs en nuage.
- Récupérer les données depuis les serveurs distants.
- Gérer les identifiants des services fournis par le ou les cloud providers.
- Gérer et stocker les clés de chiffrement/déchiffrement des données.
- Mesurer et afficher des indicateurs de performance et de sécurité sur les données.
- Fournir des outils capables d'agréger des services de fournisseurs de cloud existants.
- Gérer la migration des données hébergées chez un fournisseur X vers un fournisseur Y en cas de problème.

Comme nous pouvons le voir, il existe un réel avantage en terme de confidentialité puisque seule l'entreprise peut accéder à ses données en clair : le CSB en nuage ne peut donc accéder aux données en clair ou aux méta-données. Nous répondons ici à l'exigence « Garantir la confidentialité des données ».

Le CSB interne qui permet de placer et de récupérer les données sur le serveur de stockage en nuage doit utiliser une interface WEB qui fait le lien entre l'entreprise et le ou les fournisseurs d'informatique en nuage. Cette interface pourrait prendre la forme d'une page Web disponible uniquement en interne dans l'entreprise cliente. Il est nécessaire de gérer l'interopérabilité des diverses solutions de stockage proposées par les fournisseurs de cloud ; le module Web jouerait ce rôle. Nous ne souhaitons pas définir davantage cet aspect puisqu'il fait partie du domaine des services Web orientés architecture (Representational State Transfer [2], etc.) ce qui dépasse le cadre de ce mémoire. L'annuaire des données pourrait être distribué en interne dans l'entreprise. Par ailleurs avec ce type de solution, il est beaucoup plus simple pour une société d'aligner sa politique d'entreprise avec un environnement hybride cloud. Si, par exemple, l'entreprise cliente ne souhaite pas héberger une partie de ses données en nuage (ultra-criticité, complexité de migration, etc.), elle pourrait, grâce à l'annuaire de ressources, relier ses données non-cloud aux données cloud afin de simplifier son modèle de gestion. Dans cette mesure, la solution permet une hétérogénéité des modèles de déploiement, une flexibilité et une intégration au patrimoine informatique existant de l'entreprise. Avant d'envoyer une donnée dans le cloud, le CSB interne y insère des sentinelles et des codes correcteurs d'erreurs (conformément à l'article (57) vu en 3.1.2). Dans cette mesure, le client est capable d'effectuer des tests d'intégrité de manière régulière. Il est ainsi persuadé que le fournisseur de cloud connaît de manière exacte

2. REST est une architecture qui sépare les responsabilités du client et du serveur, sans état, avec des mécanismes de mise en cache, etc.

la donnée. Nous pourrions envisager une vérification récurrente, par exemple via un cron [3], qui lance un script de vérification sur l'ensemble des données (ou sur un panel pré-établi par le client) quotidiennement à une heure fixée. Les résultats de ces tests seraient rendus disponibles sur une page Web gérée par le CSB interne et accessible en interne de l'entreprise.

L'utilisation d'un CSB interne va à l'encontre d'un des paradigme du cloud qui vise à déporter les ressources informatiques de l'entreprise à l'externe. Selon nous, cette solution interne est — certes une renoncement à ce paradigme — mais constitue néanmoins un compromis pour l'entreprise qui souhaite garder ses données dans son champs de contrôle.

5.3.2 Rôles et composants du CSB en nuage

Les rôles du CSB en nuage se résument de la manière suivante :

- Il met à jour continuellement le catalogue de services et propose les meilleures offres de stockage en cloud en fonction des besoins du client (voir section 5.1).
- Il monitore les changements ainsi que le respect du SLA. Il audite les performances des fournisseurs d'informatique en nuage. Nous nous inspirons ici de l'architecture triangulaire (voir 5.2.1) de la figure 5.1. Ceci répond à l'exigence « Respect du SLA ».
- Il s'occupe de la facturation des services et administre les différents SLA ce qui implique une simplification pour l'entreprise cliente finale.
- Il peut aussi jouer le rôle d'arbitre entre le client et le fournisseur d'informatique en nuage en cas de conflit.

Le service de respect des SLA passe en partie par la mise en place de tests de performance. Le CSB en nuage peut, à cet effet, souscrire à diverses offres de services chez les fournisseurs qu'il souhaite auditer (de manière anonyme ou en cachant son identité pour éviter que le fournisseur audité ne fraude et ne biaise les résultats). Un serveur de monitoring des performances qui vise à lancer les tests à distance chez le fournisseur de cloud jouerait ce rôle.

Les fournisseurs d'infonuagique changent régulièrement leur SLA et quelque fois sans en aviser leurs clients. Pour pallier ce problème, le rôle du CSB en nuage serait de vérifier si ce SLA n'a pas été modifié. Nous pourrions par exemple, proposer de télécharger régulièrement ce SLA dans le but de détecter les modifications (par ex., par hachage ou en utilisant un outil de comparaison de fichier comme la commande Unix ***diff***). Les clients seraient avertis via le panel de monitoring des SLAs du CSB interne (voir la figure 5.6).

Le serveur de monitoring serait couplé au serveur qui héberge le catalogue critérié de services.

---

3. Sur les systèmes de type Unix, un cron est une tâche planifiée qui peut s'exécuter de manière récurrente (par exemple, tous les lundis à 12h00).

5.3.3 Interactions entre le CSB interne et en nuage

Souscription d'une entreprise cliente aux services du CSB L'entreprise cliente souscrit aux services du tiers de confiance via un site Web créé et géré par le CSB en nuage. La figure 5.5 illustre de façon simple le cheminement de souscription aux services.

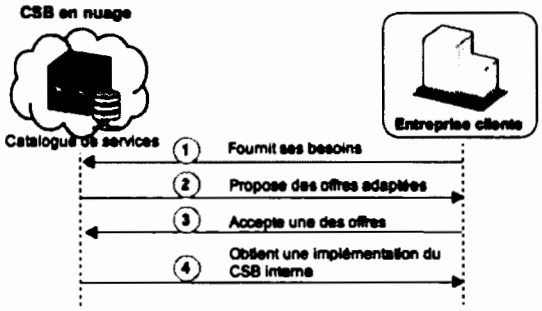


FIGURE 5.5: Une entreprise cliente qui souscrit aux services du CSB

Notons que l'étude de cas de la section présentée plus loin (5.6), donne de façon détaillée les diverses métriques employées pour fournir un résultat pertinent relatif aux exigences du client. Lorsque le client a défini ses besoins, par ex. via un formulaire Web fourni par le CSB en nuage, le CSB en nuage les analyse puis propose au client, via son catalogue critérié de services, un ou plusieurs fournisseurs de cloud. Le client obtient le CSB interne (serveur physique, machine virtuelle ou paquet logiciel) qu'il peut dès lors installer et utiliser (si le CSB interne est installé via un dépôt, le client devra configurer les services).

Audit Dès qu'une modification du SLA apparaît ou si les performances d'un CP baissent, le CSB en nuage en informe le client. Le diagramme 5.6 explique ce mécanisme.

Le CSB interne affiche le résultat de la requête envoyée par le CSB en nuage. L'administrateur de l'entreprise cliente peut dès lors prendre des mesures appropriées.

5.3.4 Interactions entre le CSB interne et le fournisseur d'informatique en nuage

L'un des principaux rôles du CSB interne consiste à placer et à récupérer les données sur les serveurs d'un fournisseur d'informatique en nuage. La figure 5.7 illustre de manière simplifiée

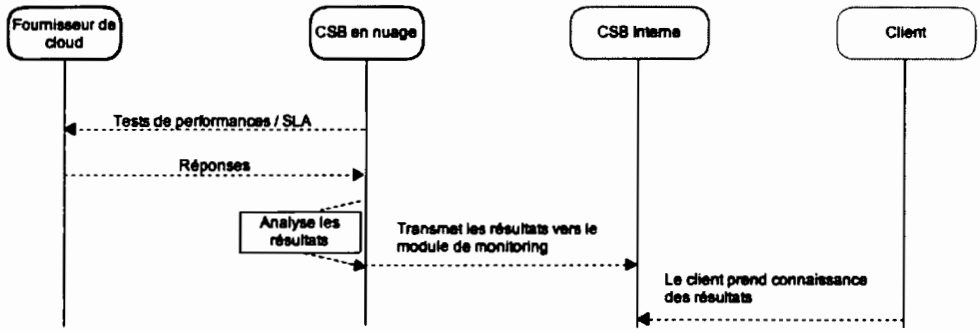


FIGURE 5.6: Diagramme d'interactions entre un CP, les modules du CSB (en nuage, interne) et un client lors d'un changement de performance ou de SLA chez ce CP.

ces actions (le CSB en nuage est volontairement masqué puisqu'il n'intervient pas directement dans les échanges entre le CP et le CSB interne).

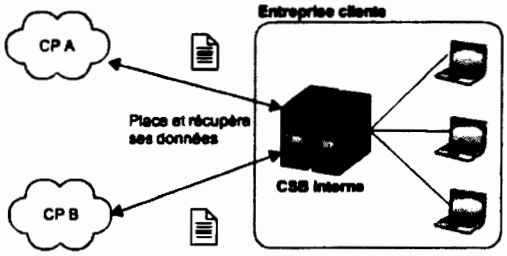


FIGURE 5.7: Une entreprise cliente qui utilise le CSB interne pour placer et récupérer ses données en cloud.

5.3.5 Apports

La topologie hybride que nous venons de décrire rallie les points forts des deux topologies 5.3a et 5.3b tout en supprimant leurs points faibles. Elle est d'après nous la plus intéressante et constitue donc notre choix. Par ailleurs, cette topologie se marierait bien à un modèle économique pour le CSB. Le CSB en nuage qui propose un catalogue de services fait payer aux clients ces mises à jour et il aligne les meilleures offres en fonction du niveau de sécurité des données de l'entreprise cliente. L'audit et le monitoring sont gérés par le CSB en nuage et le client est facturé pour ce service. Si ce consommateur ne souhaite plus faire appel aux services de ce

CSB en nuage, il pourra continuer à utiliser le CSB interne, mais ne pourra plus souscrire aux meilleures offres, avoir un audit efficace des données hébergées ce qui aura un impact direct sur la sécurité de ses données.

5.4 Modèle de stockage P2P en nuage

5.4.1 Définition du modèle

Nous proposons dans cette section, une topologie P2P garante de l'anonymat des clients. Nous partons du principe que :

- le CSB interne est disponible chez de nombreux clients ;
- toutes les ressources des serveurs hébergeant ce composant (mémoire vive, CPU, espace de stockage) ne sont pas complètement utilisées ;
- le CSB interne fait appel à des mécanismes cryptographiques, de surveillance, etc. utiles pour ce type de topologie,
- l'environnement est fédéré par le CSB en nuage et vise une même palette de services.

Pour mettre en oeuvre une topologie P2P, chaque CSB interne — pour les clients souhaitant utiliser un tel mécanisme — jouerait le rôle de nœud P2P. Chaque nœud offre un espace de stockage qui sert à héberger une partie des données des utilisateurs de ce réseau.

5.4.2 Topologie et rôles des composants CSB

La figure 5.8 (le CSB en nuage est volontairement écarté pour ne pas surcharger davantage la figure) montre différentes entreprises qui utilisent un modèle P2P pour stocker leurs données.

Rôles du CSB interne dans la topologie P2P Chaque CSB interne possède un annuaire des serveurs (nœuds) où sont stockées une partie des données (blocs) des différents clients. Avant d'envoyer un fichier sur le réseau P2P, le client va le chiffrer et y insérer des sentinelles qui lui permettront d'effectuer des vérifications d'intégrité. Le fichier est distribué sur les noeuds du réseau. Le CSB interne d'une entreprise ne peut pas accéder à l'ensemble des données d'une autre entreprise.

Rôles du CSB en nuage Le CSB en nuage garde son rôle d'auditeur : il teste la disponibilité des nœuds, leur performance, leur capacité de stockage et distribue ces résultats aux différents CSB internes qui peuvent alors choisir de placer les données sur les nœuds qui répondent à leurs exigences.

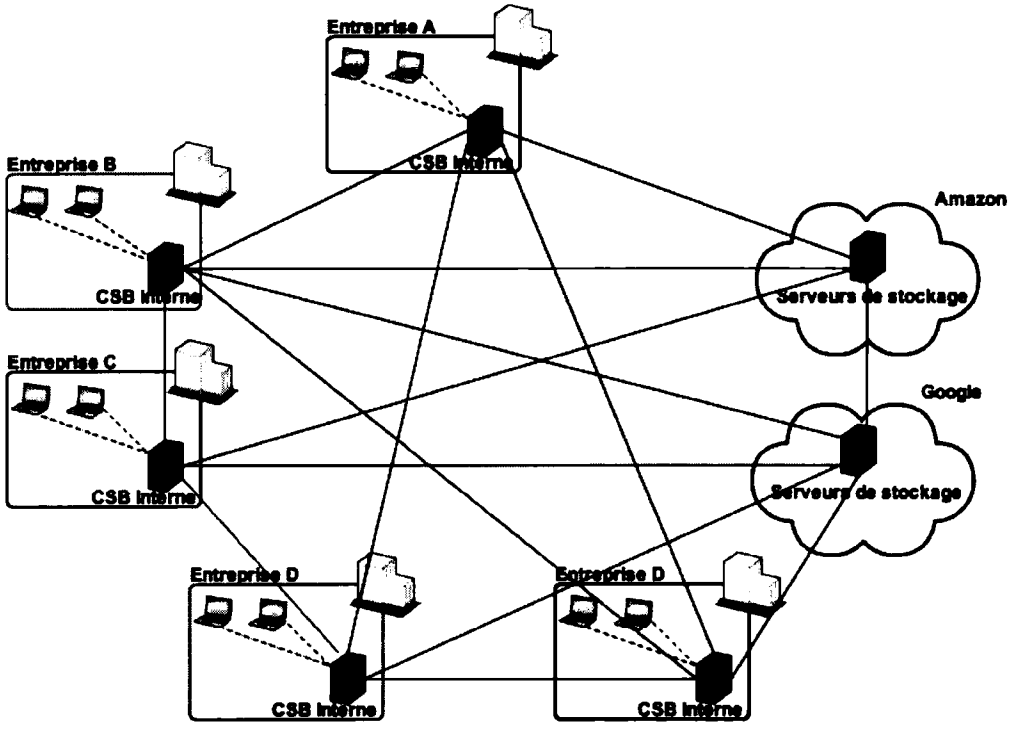


FIGURE 5.8: Topologie P2P où les nœuds sont soit des serveurs IaaS hébergés chez différents fournisseurs de cloud, soit le CSB interne d'une entreprise cliente.

5.4.3 Avantages et inconvénients d'un modèle P2P

Les avantages d'un modèle de stockage P2P en cloud sont nombreux. Comparativement à un modèle client / serveur, les nœuds ne subissent pas de goulot d'étranglement au cas où le serveur serait surchargé de demandes. Les protocoles P2P actuels permettent de reconstruire les données en cas de perte d'un nœud ce qui garantit dans une certaine mesure une intégrité des informations. Cette solution est intéressante en terme technologique pour un client qui souhaite un hébergement multi-fournisseurs ; il est simple de mettre en place une telle topologie puisque la seule contrainte concerne l'interconnexion des nœuds. Toujours dans un contexte multi-fournisseurs, les informations peuvent être placées sur les nœuds au plus proche des clients de telle manière à ce que le client dispose de meilleurs temps de disponibilité. Enfin, un client qui souhaite rester anonyme peut opter pour une solution P2P chez le CSB. Les problèmes du P2P sont d'ordre psychologique car les entreprises le perçoivent très largement comme un

moyen illégal de partager des données (films, musique, jeux, etc. piratés). Dans cette mesure, nous ne préconisons ni l'une ou l'autre solution mais nous préférons proposer deux études de cas (présentées dans la section 5.6). La première se concentre sur une entreprise qui opte pour un modèle client/serveur de stockage et la seconde s'oriente sur le modèle P2P.

5.5 Niveau de confiance du modèle proposé

Cette section reprend les différentes exigences définies dans la section 4.4. Pour chacune d'elles, nous évaluons les apports de la solution hybride proposée : nous argumentons sur les assurances supplémentaires qui garantissent une meilleure transparence et confiance des usagers.

5.5.1 Imputabilité

Le client, via le CSB interne, obtient le résultat des différents tests effectués par le CSB en nuage. Il acquiert ainsi des preuves sur l'intégrité de ses données, sur le respect du SLA et des performances, etc. qu'il pourra utiliser en cas de litige avec un fournisseur de cloud.

5.5.2 Recommandations et plan de gestion de données

Le tiers de confiance définit publiquement son système de notation sur les fournisseurs en nuage qu'il audite (SLAs, performances, etc.). Le niveau de service de chaque fournisseur est calculé par le CSB en nuage puis les résultats sont stockés dans son catalogue critérié de services.

5.5.3 Confidentialité des données

L'entreprise cliente doit être la seule à pouvoir effectuer les opérations cryptographiques et elle seule doit gérer les clés. Nous préconisons l'utilisation du chiffrement symétrique car le problème de distribution de clé ne se pose pas : les opérations de chiffrement et déchiffrement se font par le même sujet, au même endroit (au niveau du CSB interne). Enfin, par définition, le chiffrement symétrique est plus rapide que le chiffrement asymétrique (facteur 100) [4].

---

4. Source : Encryption, What is it and why is it necessary [En ligne]. Disponible `http://www.csee.umbc.edu/~wyvern/ta/encryption.html`. [Consulté le 26 Mars 2014]

5.5.4 Migration rapide

En cas de non-respect du SLA, le CSB en nuage doit proposer au client de migrer ses données vers un autre fournisseur de cloud via le composant CSB interne. Le CSB en nuage propose un nouveau fournisseur de cloud répondant aux besoins du client (si un tel fournisseur de cloud existe) via son catalogue critérié de services.

5.5.5 Enfermement chez un fournisseur

Au niveau du CSB en nuage Si le CSB en nuage n'était pas disponible ou périclitait, le client pourrait continuer à utiliser les services des fournisseurs de cloud puisqu'il possède l'application interne de gestion de ses données c'est-à-dire le CSB interne (valable pour le modèle client/serveur et P2P).

Au niveau du fournisseur de cloud Deux cas de figure peuvent se présenter :

1. **Le client a souscrit à une seule offre de service en nuage**. Dans ce cas, la résolution de l'enfermement chez le cloud provider est partielle puisque si ce fournisseur ferme ou retient les données, le client ne peut les récupérer même si notre CSB intervient en tant qu'entité tiers arbitre (en cas de litige). Ainsi, le seul cas possible de migration intervient lorsque le SLA du fournisseur en nuage change et le client n'est plus satisfait sur le niveau du service offert.
2. **Le client a souscrit à plusieurs offres de service en nuage** : l'un de ses fournisseurs ferme ou ne respecte plus le SLA. Pour que le client puisse rapidement retrouver le niveau de service souhaité, il peut facilement migrer ses données vers un autre fournisseur de services grâce au CSB interne, évitant ainsi la possibilité d'enfermement. Le CSB en nuage aurait ici pour rôle de proposer à son client un nouveau fournisseur de cloud.

5.5.6 Anonymat du client

Grâce au modèle P2P, un client peut rester anonyme car il ne divulgue pas son identité à un fournisseur de services en nuage. Rappelons que le CSB interne de la topologie P2P peut-être relié à un noeud hébergé par un fournisseur de cloud. L'anonymat est garanti puisque le CSB en nuage loue ces espaces de stockage en nuage en son nom.

5.5.7 Négociation du SLA

Le CSB en nuage négocie le SLA — en fonction des besoins de l'entreprise cliente — avec les différents fournisseurs de cloud.

5.5.8 Agrégation des services

Le CSB interne propose des mécanismes qui permettent à l'entreprise cliente d'utiliser des services en nuage de différents fournisseurs de cloud. Les contraintes liées à cette exigence ont été définies en 4.1.2.

5.5.9 Considération d'infrastructure pour le CSB en nuage

Nous appelons infrastructure les bâtiments physiques où s'effectuent les traitements informatiques du CSB en nuage. Nous nous posons ici la question de savoir si le tiers de confiance peut se servir d'un cloud privé interne ou d'un cloud privé externe. L'utilisation d'un cloud nous parait judicieux dans la mesure où le CSB se placerait dans la même configuration que celle des fournisseurs de services.

Un cloud privé interne est un centre de données qu'une entreprise bâtit pour ses propres besoins. Toute l'administration et la gestion de l'infrastructure (serveurs, réseaux, etc.) sont à la charge de cette entreprise. Dans un cloud privé externe, une entreprise réserve des serveurs chez un fournisseur d'informatique en nuage. L'entreprise se soucie uniquement de l'administration et du déploiement des services sur les serveurs qui sont mis à sa disposition. Tout l'aspect de gestion de l'infrastructure (serveurs, réseaux, etc.) est à la charge du fournisseur de cloud.

Dans la liste d'exigences présentée en 4.4.1, nous préconisons que le tiers de confiance soit indépendant vis-à-vis du fournisseur d'informatique en nuage. Or, si ce tiers souscrit à une offre de cloud privé externe chez un des fournisseurs qu'il audite (SLA, performance), il y a risque de conflit d'intérêt : le fournisseur d'informatique en nuage peut saisir les données du tiers de confiance par exemple, s'il est en désaccord avec le résultat de l'audit. De plus, ce fournisseur pourrait falsifier le résultat des tests de son infrastructure en les anticipant, voire même en réécrivant le rapport car il a un accès physique aux serveurs et donc aux données traitées et hébergées par le CSB en nuage.

Notre préférence va donc vers la création d'un cloud privé interne, non pas pour une raison de gestion ou de puissance computationnelle, mais pour une question de sécurité et d'isolement.

5.5.10 Redondance de l'infrastructure du CSB en nuage

Le recours à un datacenter unique pour le tiers de confiance n'est pas une bonne solution à retenir puisque le CSB en nuage représenterait l'unique point de défaillance (les causes ont été définies dans la section 3.2). Nous devons par conséquent opter pour une duplication et une localisation multiple de l'infrastructure en respectant les préconisations ci-après :

- **Distance géographique** : les datacenters sont suffisamment éloignés géographiquement pour éviter des répercussions d'une même catastrophe.
- **Proximité des clients** : les datacenters se placent au plus près des clients.
- **Continuité des activités** : les centres de données sont tenus de mettre en place des systèmes de balançage de liens pour assurer un bon niveau de disponibilité de services.

5.6 Études de cas

Nous présentons ici deux études de cas hypothétiques (l'une basée sur l'architecture client/serveur, l'autre basée sur l'architecture P2P) offrant un cadre pratique à la solution que nous proposons. Dans un premier temps, nous définissons le mode opératoire de l'étude. Ensuite, ce mode opératoire sera utilisé comme base de nos cas.

5.6.1 Mode opératoire

Dans le scénario hypothétique de l'étude de cas, le module CSB en nuage a pris le soin d'auditer les SLA et d'effectuer des tests de performances auprès de différents fournisseurs d'informatique en nuage, respectivement le CP A, B, C et D. Nous choisissons plusieurs critères d'appréciation :

- **Localisation des datacenters :** correspond à l'endroit géographique où sont installées les infrastructures des centres de données.
- **Choix de la localisation :** spécifie la possibilité ou non de choisir l'endroit géographique où seront hébergées les données.
- **Type de redondance :** si le paramètre est à « datacenter », cela signifie que chaque centre de données possède au moins un miroir, si le paramètre est défini à RAID, cela veut dire que les disques durs sont dupliqués en se fondant sur le système RAID 0 (miroir) ou RAID 5 (parité répartie).
- **Respect du SLA :** indique si le SLA est respecté.

- **Performance moyenne des E/S :** précise la bande passante réseau maximale offerte par le fournisseur.
- **Élasticité :** donne des indications sur les ressources physiques disponibles chez le fournisseur d'informatique en nuage. L'élasticité peut être mesurée en effectuant des tests de performance et de qualité de service sur la capacité qu'a un CP à fournir des ressources pour un service (par ex., sur la rapidité à fournir les ressources, si les ressources fournies ne sont pas sur-estimées ou sous-estimées [5] par rapport aux besoins réels du client, etc.).
- **Certifications :** liste les certifications que le fournisseur de cloud détient.
- **Technologies :** indique si les technologies employées par le fournisseur d'informatique en nuage sont publiquement définies, éprouvées, standardisées, si le choix des technologies s'inscrit dans des composantes publiquement définit et privé, etc.
- **Anonymat :** indique si le fournisseur de cloud propose de garantir l'anonymat de ses clients.

Nous ne souhaitons pas ajouter le critère « Confidentialité des données » chez le fournisseur de cloud puisque c'est une fonctionnalité en trompe l'œil. Le CP peut tout à fait déchiffrer les données puisqu'il possède la clé de chiffrement. La fonctionnalité de chiffrement des données du client devrait s'effectuer du côté du CSB interne.

Les résultats des audits et des tests sont présentés dans le tableau 5.1. Ces résultats sont hypothétiques et ne concernent aucunement les données collectées chez des fournisseurs d'informatique en nuage existants.

---

5. Ces estimations pourraient être le résultat de la mise en place de jeux de tests chez les hébergeurs. Les résultats (puissance requise, etc.) — qui sont connues par avance par le CSB en nuage — sont ensuite comparés aux résultats obtenues chez les différents fournisseurs de cloud. En proposant une marge de tolérance acceptable, le CSB en nuage peut conclure si les ressources fournies sont sous-estimées ou sur-estimées.

Nom du CP	Localisation des datacenters	Choix de la localisation	Niveau de redondance	Respect du SLA	Perf. moy. des E/S	Elasticité	Certifications	Technologies	Anonymat
A	France / USA	Oui	Datacenter / RAID	Inconnue	10 Gb/s	Bonne	SSAE16 Type I, Tiers I	Libres	Non
B	Allemagne / Irlande / USA	Non	Datacenter / RAID	Oui	5 Gb / s	Moyenne	ISO27001, Tiers I	Propriétaires	Non
C	Luxembourg / Allemagne	Oui	RAID	Non	1 Gb / s	Moyenne	Aucune	Propriétaires	Non
D	Hong-Kong / Allemagne / Irlande	Oui	Datacenter / RAID	Oui	10 Gb / s	Bonne	ISAE3402 Type II, ISO27001, Tiers III	Propriétaires	Non

TABLE 5.1: Résultats du monitoring des différents fournisseurs A, B, C, D

Ces résultats constituent les métriques dispensées dans le catalogue critérié de services hébergé par le CSB en nuage. Ainsi, lors de la souscription d'un client au service du CSB, ce client répond à une série de questions qu'il soumet au CSB.

Le choix d'un ou plusieurs fournisseurs se fait par l'intermédiaire d'une matrice de pondération. Nous nous inspirons ici de la méthode six sigma et plus particulièrement de la matrice de Pugh [6]. Le CSB affecte des coefficients aux différents critères en fonction des réponses du client. Plus le paramètre apparaît comme important aux yeux de ce client, plus le coefficient sera élevé. Si le critère est rempli, nous lui affectons 1, s'il l'est partiellement mais correspond au minimum requis nous lui affectons 0, s'il n'est pas du tout adapté, nous lui affectons la valeur -1. Ensuite, nous calculons pour chaque fournisseur d'informatique en nuage :

$$\text{Résultat}_{CP\quad X} = \sum (\text{Coefficient du critère} * \text{note de validation du critère})$$

Le prestataire d'infonuagique qui obtient le résultat le plus élevé répond le mieux aux attentes du client.

5.6.2 Cas 1 : stockage de données qui utilise une architecture client/serveur

Considérons l'entreprise Allemande « SauerKraut » qui d'un point de vue juridique, doit garder les enregistrements financiers de toutes les transactions effectuées et ce pour une durée de cinq ans minimum. Elle définit donc ses données comme étant critiques. Celles-ci sont décisives pour la pérennité de l'entreprise et constituent des preuves en cas d'action juridique. Du point de vue des lois, ces enregistrements doivent rester dans le pays où ils ont été produits. Ces informations sont actuellement hébergées en interne chez l'entreprise mais, après avoir perdu le mois dernier les données du service Marketing, la DSI décide de réagir. Leurs serveurs sont vieillissants et tombent régulièrement en panne. Pour couronner le tout, aucun informaticien ne possède les compétences pour la gestion de serveurs de fichiers en interne. De plus, l'entreprise y perd son latin en raison de la multitude des offres proposées par les fournisseurs de Cloud. Pour éviter la catastrophe, l'entreprise fait logiquement appel aux services du CSB et répond au questionnaire tel qu'indiqué en 5.2.

---

6. Voir : Pugh Matrix [En ligne]. Disponible `http://www.whatissixsigma.net/pugh-matrix/` [Consulté le 3 Février 2014]

Questions	Réponses du client	Analyse du CSB	Critères du CSB	Pondérations CSB
Type de données	Archives	Préconisation ISAE3402	Certifications	3
Besoin en élasticité	Faible	Volume de données augmente peu	Élasticité	1
Criticité	Haute	Besoin de redondance	Redondance	3 et 3
Fréquence d'accès	Rare	Peu d'exigences en I/O	Performance	1
Importance de la localisation	Oui	Localisation critique	Localisation	5
Pays de localisation	Allemagne uniquement	Localisation critique	Choix de la localisation	5
Anonymat	Non	Ne souhaite pas d'anonymat	Anonymat	0

TABLE 5.2: Réponse du client SauerKraut

Les résultats font preuve de logique puisque ces données de type archive sont très peu utilisées après leur archivage. Notre tiers de confiance préconise que le fournisseur de cloud ait une certification ISAE3402 Type I ou II attestant qu'il met bien en place ses processus internes et les valide.

Le CSB en nuage affecte des coefficients aux critères puis calcule les résultats. Les résultats sont disponibles sur la matrice de pondération (tableau 5.3).

Critère de décision	**Pondération des critères**	CP A	CP B	CP C	CP D
Localisation	5	-1	1	1	1
Choix de la localisation	5	1	-1	1	1
Redondance	3	1	1	0	1
Respect du SLA	3	-1	1	0	1
Performance	1	1	1	1	1
Elasticité	1	1	1	1	1
Certifications	3	1	0	-1	1
Technologies	1	1	1	1	1
Anonymat	0	-1	-1	-1	-1
Résultats		**6**	**9**	**10**	**22**

TABLE 5.3: Résultat de la pondération pour le client « SauerKraut »

Nous constatons que le CP D correspond le mieux aux besoins du client. Le rôle du CSB en nuage ne s'arrête pas là puisque « SauerKraut » envisage d'effectuer également des audits réguliers de l'intégrité de ses données.

5.6.3 Cas 2 : stockage de données qui utilise une architecture P2P

La DSI de la société européenne « Flugzeug », spécialiste en aéronautique, ne souhaite pas investir davantage dans son parc informatique actuel car le budget alloué par la direction n'est pas suffisant. Ses données sont sensibles et nécessitent un niveau de sécurisation adéquat. Elle décide donc de prospecter auprès des différents fournisseurs d'informatique en nuage. Elle ne s'y retrouve pas dans toutes les options proposées ; la direction de la société quant à elle presse ce projet informatique. Le directeur informatique exige que les données hébergés restent anonymes

et confidentielles. En effet, « Flugzeug » est concurrente d'une autre grande société aéronautique américaine « LockeedFranklin » et, avec l'affaire PRISM, elle craint qu'un gouvernement ne s'approprie les données de son entreprise et qu'un concurrent en profite pour obtenir un contrat important à son insu.

Pour y voir plus clair, elle fait appel aux services de notre CSB. L'entreprise « Flugzeug » répond au questionnaire du CSB :

Questions	Réponses du client	Analyse du CSB	Critères du CSB	Pondérations CSB
Type de données	Fichiers clients	Certifications nécessaires	Certifications	3
Besoin en élasticité	Important	Volume de données augmente	Élasticité	3
Criticité	Haute	Besoin de redondance	Redondance	3 et 3
Fréquence d'accès	Important	Peu d'exigences en I/O	Performance	1
Importance de la localisation	Non	Localisation peu importante	Localisation	0
Pays de localisation	Peu important	Localisation peu importante	Choix de la localisation	0
Anonymat	Oui	Exige l'anonymat	Anonymat	5

TABLE 5.4: Réponse du client « Flugzeug »

Le facteur anonymat est très important pour le client. Puisqu'aucun fournisseur d'infonuagique ne propose le respect d'un tel concept, il est inutile de choisir parmi le CP A, B, C ou D et donc de recalculer une matrice de pondération pour ce client. Dans ce sens, le CSB en nuage propose donc à son client d'opter pour une solution pair-à-pair.

5.6.4 Méthodologie de mise en pratique

Ces deux études de cas offrent un cadre théorique à la solution de sécurité proposée. L'implémentation pratique des architectures client/serveur et P2P dépasse le cadre de ce mémoire. Néanmoins, pour nos travaux futurs, nous fournissons - de manière sommaire - des pistes sur les technologies à adopter.

Tout ce qui a trait à la communication entre les différents modules de notre architecture et à la création d'un module assurant l'interopérabilité entre les API des différents fournisseurs de cloud implique une bonne compréhension des services web (SOA, REST, etc.). Ceux-ci offrent de bonnes pratiques aux développeurs. Toutes les transactions entre le client, le CSB en nuage et le fournisseur d'informatique en nuage se font sur le réseau internet. Nous proposons d'une part, des langages de programmation s'exécutant du coté client comme XML, HTML, CSS et Javascript et d'autre part, des langages comme PHP ou Python qui se lancent du côté serveur. Ces derniers permettent de traiter, stocker et afficher les informations aux différents protagonistes des échanges.

Si le CSB interne opte pour une solution de type client/serveur, les échanges se feront via ces mêmes technologies.

Si le CSB interne adopte une architecture pair-à-pair, il est nécessaire d'utiliser un protocole P2P garantissant chiffrement et anonymat des échanges (voir 5.4) comme Chord [7] ou Pastry [8]. D'autres caractéristiques attenantes à ces protocoles existent : nous citerons les protocoles de routage utilisés, la mise à jour de la topologie, les modes de recouvrement des données, le chiffrement des échanges, etc.

---

7. Voir : Github : The Chord Project [En ligne]. Disponible `https://github.com/sit/dht/wiki` 2013. [Consulté le 25 septembre 2014]

8. Voir : Rice University & Max Plank Institute for Software Systems : The Pastry Project [En ligne]. Disponible `http://www.freepastry.org/` 2009. [Consulté le 25 septembre 2014]

CONCLUSION

Le cloud computing représente sans conteste un concept qui transcende la manière de livrer des services aux différents clients que ce soit pour les entreprises ou le grand public. Cette idée, que nous jugeons noble technologiquement parlant, permet de réaliser des économies substantielles tout en supprimant le fardeau d'une gestion de l'infrastructure informatique complexe pour le client.

Néanmoins, de nombreuses problématiques de sécurité subsistent ; nous les avons relevées dans notre état de l'art. Nous regroupons ces problématiques en deux familles : l'une liée à la gestion des données par le prestataire d'infonuagique (technologique et contractuelle) ; l'autre à toutes les lois qui encadrent ces données. Les entreprises perdent en quelque sorte le contact et le contrôle physique avec leurs données qui sont déportées vers des serveurs en nuage et ne sont pas couvertes vis-à-vis de leur fournisseur d'infonuagique en cas de litige ou de la fermeture de l'un d'entre eux. Par ailleurs, des scandales comme l'affaire PRISM où des géants comme Google ou Facebook divulguent les informations de leurs clients - sous contrainte juridique - accentuent encore un peu plus l'appréhension des consommateurs à utiliser ces services. Même si des clients confiants souscrivent à des services en nuage et acceptent un contrat (le SLA), celui-ci n'offrira pas de garanties sérieuses sur la sécurité des informations : à juste titre, ce contrat est rédigé par le cloud provider pour se protéger juridiquement et fournit un minimum de garanties surtout de disponibilité.

Nous avons étudié par la suite les principales technologies et bonnes pratiques qui favorisent le maintient de l'intégrité des données ; mais elles tendent à se placer du côté du fournisseur de cloud, le client reste par conséquent avec peu de pouvoir. Il paraissait nécessaire et judicieux de rétablir le rapport de force entre le client et son fournisseur de services en nuage.

Principales contributions

Face à ce constat, nous avons apporté des réponses pertinentes sur la nécessité d'introduire un troisième acteur : le tiers de confiance. Nous avons défini les facteurs quantifiables et qualifiables qui garantissent la confiance des utilisateurs de services en nuage et établi une

liste d'exigences que ce tiers doit respecter. Nous avons proposé une architecture permettant de garantir ces exigences tout en utilisant un tiers de confiance couplé à un courtier de sécurité en nuage. L'architecture proposée simplifie la livraison de services au client et agrège plusieurs services en un seul. Nous avons dans un premier temps introduit la notion de catalogue critérié de services : le tiers de confiance audite les pratiques des fournisseurs d'infonuagique, dresse un bordereau des offres respectives, liste les ressources, etc. Dans un deuxième temps, une architecture a été proposée pour pouvoir offrir les services envisagés par ce tiers. Notre choix s'est arrêté sur une architecture hybride où chaque entreprise possède un module interne qui chiffre, dépose et récupère les données chez les fournisseurs de cloud tout en surveillant leur intégrité, etc. Le CSB en nuage propose à ses clients soucieux de rester anonyme, la possibilité d'utiliser un modèle de stockage P2P. Nous avons ensuite évalué le niveau de confiance du modèle proposé dans le but de vérifier si les exigences que nous avions établi pour un tiers de confiance étaient respectées. Enfin, nous avons mis en place deux scénarios distincts : une étude a été menée sur un modèle client/serveur ; une autre sur le modèle P2P.

Travaux futurs

Notre mémoire couvre de nombreux domaines de l'informatique ; certains aspects liés à notre proposition ont cependant été volontairement éclipsés. En effet, la mise en place d'une telle topologie nécessite une connaissance approfondie de plusieurs technologies de la programmation au réseau en passant par l'administration des systèmes.

La taxonomie que nous avons défini en annexe A permet de comprendre dans le détail ces diverses technologies et multiples domaines qui auraient couvert l'ensemble de notre proposition.

Dans le domaine des ressources humaines, il serait intéressant de dresser la liste des compétences (par exemple : administration systèmes, programmation, etc...) et spécialités (RH, informatique, juridique, etc...) exigées pour assurer une continuité et un certain niveau de qualité des services du Cloud Security Broker / tiers de confiance en nuage.

En ce qui concerne les tests de performance, nous avons prévu les différents types de tests sans toutefois préciser (techniquement parlant) le contenu de ces tests, leur fréquence, etc.

De plus, pour le CSB en nuage, certaines actions répétitives de la part des administrateurs systèmes pourraient être automatisées via l'utilisation de gestionnaires de configuration

comme les logiciels Chef[9] ou Puppet[10]. Ceux-ci permettent de déployer des configurations sur un ensemble de machines cibles.

9. Voir : Chef [En ligne]. Disponible `http://www.getchef.com/chef/` 2014. [Consulté le 6 Février 2014]

10. Voir : Puppet [En ligne]. Disponible `http://puppetlabs.com/` 2014. [Consulté le 6 Février 2014]

ANNEXE A

TAXONOMIE CSB

Nous présentons dans la figure A.1 une taxonomie reliée à notre CSB. Une taxonomie de base a été établie par la NIST (36). Nous nous inspirons de celle-ci pour l'étendre aux besoins propre du CSB en terme de mécanismes de sécurité, de services et d'assurances.

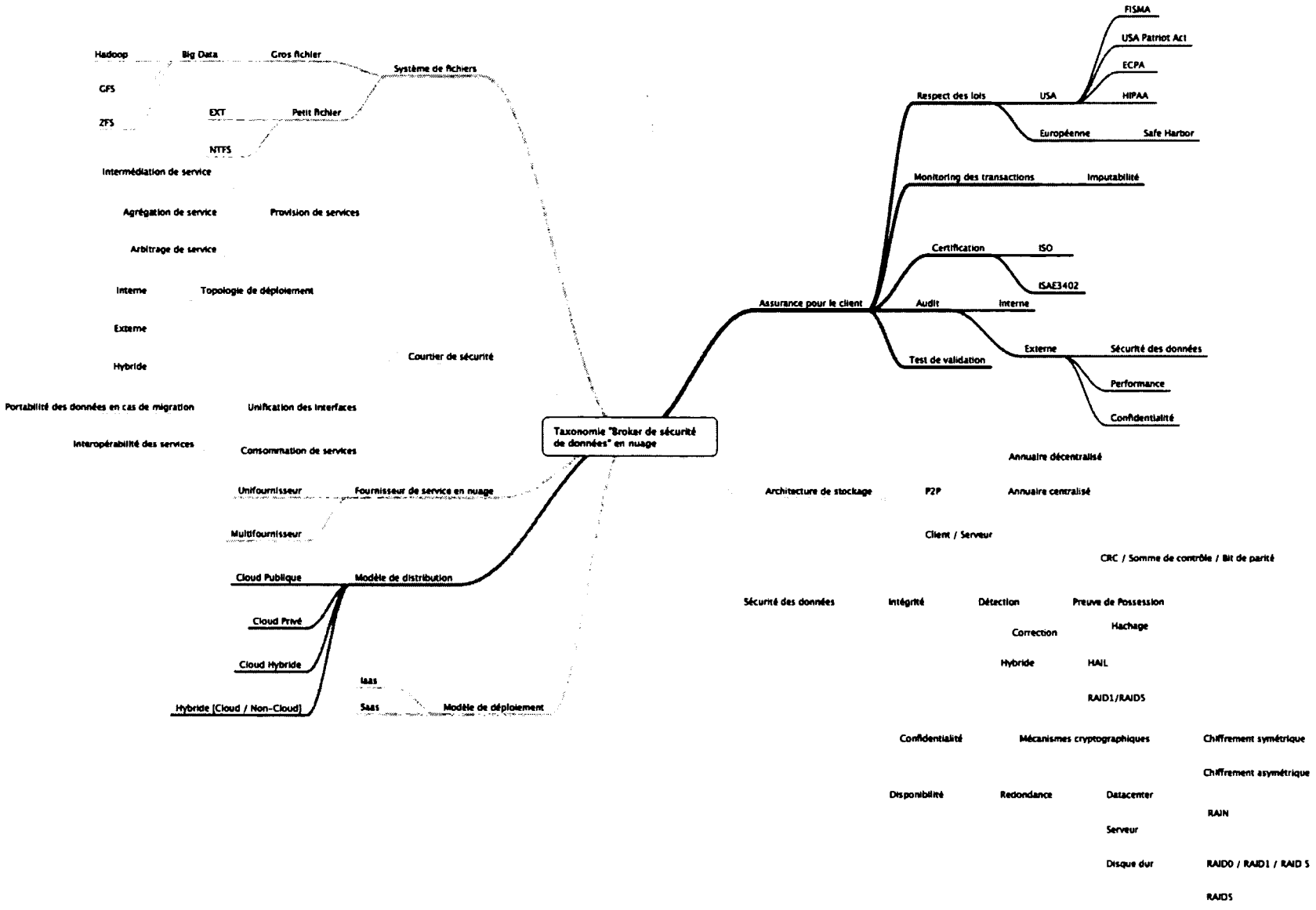


FIGURE A.1: Taxonomie d'un CSB

BIBLIOGRAPHIE

(1) *Michael Porter - L'avantage concurrentiel*, chapitre 5. InterEditions, 1986.

(2) Crespi & Al. : Why Data Integrity is Important to You - The Data Integrity Initiative. Rapport de projet, 2011.

(3) Khatri & al. : Survey on data integrity approaches used in the cloud computing. *International Journal of Engineering Research & Technology (IJERT)*, 2012.

(4) Giuseppe Ateniese, Randal Burns, Reza Curtmola, Joseph Herring, Lea Kissner, Zachary Peterson et Dawn Song : Provable data possession at untrusted stores. *In Proceedings of the 14th ACM conference on Computer and communications security*, CCS '07, pages 598–609, New York, NY, USA, 2007. ACM.

(5) Md. Faizul Bari, Raouf Boutaba, Rafael Esteves, Lisandro Zambenedetti Granville, Maxim Podlesny, Md Golam Rabbani, Qi Zhang et Mohamed Faten Zhani : Data Center Network Virtualization : A Survey. *IEEE Communications Surveys & Tutorials*, pages 1–20, 2012.

(6) Guy Begin : Notes de cours (INF8750). [Via Moodle]. Non disponible, 2011. [Consulté le 6 mai 2013].

(7) Blaise Berliner : Gouvernance des risques - adieu sas 70 et bienvenues isae 3402 et ssae 16. [En ligne]. Disponible : `http://www.deloitte.com/view/en_LU/lu/services/consulting/enterprise-risk-services/isae3402-ssae16/002a32d4379f0310VgnVCM3000001c56f00aRCRD.htm#`, 2013. [Consulté le 24 mars 2014].

(8) David Bernstein, Erik Ludvigson, Krishna Sankar, Steve Diamond et Monique Morrow : Blueprint for the intercloud - protocols and formats for cloud computing interoperability. *In Proceedings of the 2009 Fourth International Conference on Internet and Web Applications and Services*, ICIW '09, pages 328–336, Washington, DC, USA, 2009. IEEE Computer Society.

(9) Deepak K. Vij & David Bernstein : IEEE P2302^{TM}/D0.2 - Draft Standard for Intercloud - Interoperability and Federation (SIIF). [En ligne]. Disponible : `https://www.oasis-open.org/committees/download.php/46205/p2302-12-0002-00-DRFT-intercloud-p2302-draft-0-2.pdf`, 2012. [Consulté le 24 mai 2013].

(10) Kevin D. Bowers, Ari Juels et Alina Oprea : Proofs of retrievability : Theory and implementation. Cryptology ePrint Archive, Report 2008/175, 2008. `http://eprint.iacr.org/`.

(11) Kevin D. Bowers, Ari Juels et Alina Oprea : Hail : a high-availability and integrity layer for cloud storage. *In Proceedings of the 16th ACM conference on Computer and communications security*, CCS '09, pages 187–198, New York, NY, USA, 2009. ACM.

(12) Daniele Catteddu et Giles Hogben : Cloud computing - benefits, risks and recommendation for information security. Rapport de projet, ENISA, 2009.

(13) Cisco : Ccna 3. [En ligne]. Disponible sur inscription : http://www.cisco.com/web/learning/netacad/index.html, 2010. [Consulté le 12 juin 2011].

(14) PCI Standards Council : Pci ssc data security standards overview. [En ligne]. Disponible : https://www.pcisecuritystandards.org/security_standards/, 2012. [Consulté le 19 juin 2013].

(15) Mike Small & David Chapa Drew Amorosi : Don't gamble with your data : Critical questions for data security in the cloud. [En ligne]. Disponible : https://www.brighttalk.com/webcast/8325/65327, 2013. [Consulté le 7 mars 2013].

(16) RSA EMC : Rsa cloud trust authority, 2011. http://chucksblog.emc.com/content/EMC_RSA_Cloud_Trust_Authority.pdf.

(17) Chris Erway, Alptekin Küpçü, Charalampos Papamanthou et Roberto Tamassia : Dynamic provable data possession. *In Proceedings of the 16th ACM conference on Computer and communications security*, CCS '09, pages 213–222, New York, NY, USA, 2009. ACM.

(18) Marinescu et al. : *Cloud computing : Theory and Practice*. Morgan Kaufmann, 2013.

(19) Mather et al. : *Cloud Security and Privacy*. O'Reilly, 2011.

(20) Commission Européenne : Directive of the european parliament and of the council on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data. [En ligne]. Disponible : http://ec.europa.eu/home-affairs/doc_centre/police/docs/com_2012_10_en.pdf, 2012. [Consulté le 10 juin 2013].

(21) S. Goldwasser, S. Micali et C. Rackoff : The knowledge complexity of interactive proof systems. *SIAM J. Comput.*, 18(1):186–208, février 1989.

(22) Richard Gordon, Colleen Graham, Kathryn Hale, Jon Hardcastle, Peter Kjeldsen, George Shiffler et Ed Anderson : High-Tech Tuesday Webinar : Gartner Worldwide IT Spending Forecast , 2Q12 Update : Cloud Is the Silver Lining. Rapport de projet July, 2012.

(23) US Government : Uniting and strengthening america by providing appropriate tools required to intercept and obstruct terrorism (usa patriot act) act of 2001. [En ligne]. Disponible : http://www.gpo.gov/fdsys/pkg/PLAW-107publ56/pdf/PLAW-107publ56.pdf, 2001. [Consulté le 10 juin 2013].

(24) Philippe Grange : Livre Blanc - SÉCURITÉ DU CLOUD COMPUTING. *Computing*, page 24, 2010.

(25) CSA Working Group : Security guidance for critical areas of focus in cloud computing v3.0. Rapport de projet, CSA, 2011.

(26) CSA Working Group : The notorious nine, cloud computing top threats in 2013. Rapport de projet, CSA, 2013.

(27) N. Gruschka et M. Jensen : Attack surfaces : A taxonomy for attacks on cloud services. *In Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on*, pages 276–279, 2010.

(28) Rabab Hayek : *Data Localization and Summarization Techniques in P2P Systems*. Thèse de doctorat, UFR Sciences & Techniques, Universite de Nantes, 2009.

(29) KPMG International : Breaking through the cloud adoption barriers. Rapport de projet.

(30) Jim Reavis & Blake Dournaee John Messina : Solving cloud complexity with service brokers. [En ligne]. Disponible : http://www.brighttalk.com/webcast/5573/39861, 2010. [Consulté le 18 juin 2013].

(31) Ari Juels et Burton S. Kaliski, Jr. : Pors : proofs of retrievability for large files. *In Proceedings of the 14th ACM conference on Computer and communications security*, CCS '07, pages 584–597, New York, NY, USA, 2007. ACM.

(32) Skolochenko Kissel, Scholl et Li : Guidelines for media sanitization. Rapport de projet, NIST, 2006.

(33) Jamal Labed : Sa patriot act : un risque majeur pour la confidentialité des données dans le cloud. [En ligne]. Disponible : http://solutionsauxentreprises.lemonde.fr/cloud-computing/usa-patriot-act-un-risque-majeur-pour-la-confidentialite-des-donnees-dans-le-cloud_a-27-630.html, 2012. [Consulté le 11 juin 2013].

(34) RSA Laboratories : High assurance & integrity layer (hail). [En ligne]. Disponible : http://www.rsa.com/rsalabs/node.asp?id=3680, 2010. [Consulté le 2 février 2013].

(35) Scott M. Lewandowski : Frameworks for component-based client/server computing. *ACM Comput. Surv.*, 30(1):3–27, mars 1998.

(36) Fang Liu, Jin Tong, Jian Mao, Robert Bohn, John Messina, Lee Badger et Dawn Leaf : NIST Cloud Computing Reference Architecture Recommendations of the National Institute of Standards and. Rapport de projet, 2011.

(37) Simon Liu et R. Kuhn : Data loss prevention. *IT Professional*, 12(2):10–13, 2010.

(38) Peter Mell et Tim Grance : The nist definition of cloud computing. *National Institute of Standards and Technology*, 53(6):50, 2009.

(39) Microsoft : Microsoft high availability overview, white paper. Rapport de projet, Microsoft, 2008.

(40) Microsoft : Virtualisation - pourquoi virtualiser ? [En ligne]. Disponible : http://www.microsoft.com/france/serveur-cloud/virtualisation/pourquoi-virtualiser.aspx, 2013. [Consulté le 17 mai 2013].

(41) D Milojicic et al : Peer-to-peer computing. Rapport de projet, HP Lab, 2003.

(42) NIST : Fisma - detailed overview. [En ligne]. Disponible : http://csrc.nist.gov/groups/SMA/fisma/overview.html, 2012. [Consulté le 11 juin 2013].

(43) Kroll Ontrack : Mieux comprendre la perte de données. [En ligne]. Disponible : http://www.ontrack.fr/perte-de-donnees/, 2007. [Consulté le 22 mai 2013].

(44) David A. Patterson, Garth Gibson et Randy H. Katz : A case for redundant arrays of inexpensive disks (raid). *SIGMOD Rec.*, 17(3):109–116, juin 1988.

(45) Mithun Paul et Ashutosh Saxena : Proof of erasability for ensuring comprehensive data deletion in cloud computing. *In CNSA'10*, pages 340–348, 2010.

(46) S. Pearson et A. Benameur : Privacy, security and trust issues arising from cloud computing. *In Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on*, pages 693–702, 2010.

(47) Andrew Reichman : File Storage Costs Less In The Cloud Than In-House. Rapport de projet, 2011.

(48) Margaret Rouse : ISAE 3402. [En ligne]. Disponible : http://searchfinancialapplications.techtarget.com/definition/ISAE-3402, 2013. [Consulté le 19 juin 2013].

(49) Hoffman Scarfone, Souppaya : Guide to security for full virtualization technologies. *National Institute of Standards and Technology*, 2011.

(50) Mathieu Schmitt : Travail de mi-session (MGL7126) : Strategie d'adoption du cloud computing en entreprise. november 2012.

(51) Uptime Institut Professional Services : Data center site infrastructure tier standard : Topology. Rapport de projet, Uptime Institut, NY, 2012.

(52) Mehul A. Shah, Mary Baker, Jeffrey C. Mogul et Ram Swaminathan : Auditing to keep online storage services honest. *In Proceedings of the 11th USENIX workshop on Hot topics in operating systems*, HOTOS'07, pages 11 :1–11 :6, Berkeley, CA, USA, 2007. USENIX Association.

(53) Gopalan Sivathanu, Charles P. Wright et Erez Zadok : Ensuring data integrity in storage : techniques and applications. *In Proceedings of the 2005 ACM workshop on Storage security and survivability*, StorageSS '05, pages 26–36, New York, NY, USA, 2005. ACM.

(54) SNIA : Cloud data management interface. Rapport de projet, San Francisco, 2012.

(55) Gary Stoneburner : Sp 800-33. underlying technical models for information technology security. Rapport de projet, Gaithersburg, MD, United States, 2001.

(56) Dan Sullivan : Reducing risks with multiple cloud service providers. [En ligne]. Disponible sur inscription : http://searchcloudcomputing.techtarget.com/tip/Reducing-risks-with-multiple-cloud-service-providers, 2012. [Consulté le 23 avril 2013].

(57) Qian Wang, Cong Wang, Kui Ren, Wenjing Lou et Jin Li : Enabling public auditability and data dynamics for storage security in cloud computing. *Parallel and Distributed Systems, IEEE Transactions on*, 22(5):847 –859, may 2011.

(58) Andrew Watson : Making cloud standards customer-driven. Cryptology ePrint Archive, Report 2008/175, 2013. http://www.servicetechsymposium.com/dl/presentations/making_cloud_standards_customer-driven.pdf.

(59) Wikipédia : Haute disponibilité. [En ligne]. Disponible : http://fr.wikipedia.org/wiki/Haute_disponibilit%C3%A9, 2010. [Consulté le 17 mars 2013].

(60) Wikipédia : Cloud computing, 2013. http://fr.wikipedia.org/wiki/Cloud_computing.

(61) Wikipédia : Isae 3402. [En ligne]. Disponible : http://fr.wikipedia.org/wiki/ISAE_3402, 2013. [Consulté le 19 juin 2013].

(62) Wikipédia : Merkle Tree. [En ligne]. Disponible : http://en.wikipedia.org/wiki/Merkle_tree, 2013. [Consulté le 29 Janvier 2014].

(63) Wikipédia : RAID (informatique). [En ligne]. Disponible : http://fr.wikipedia.org/wiki/RAID_(informatique), 2013. [Consulté le 3 juin 2013].