Confidentialité et disponibilité des données entreposées dans les nuages

Kawthar Karkouda, Nouria Harbi, Jérôme Darmont, Gérald Gavin

Confidentialité et disponibilté des données entreposées dans les nuages

Kawthar Karkouda*, Nouria Harbi**
Jérôme Darmont ** Gérald Gavin ***

Laboratoire Eric, Université Lumière Lyon 2, 5 avenue Mendès France, Bât K, 69677 Bron Cedex
* kawthar.karkouda@gmail.com
** nouria.harbi,jerome.darmont@univ-lyon2.fr
*** gerald.gavin@univ-lyon1.fr

Résumé. Avec l'avènement de l'informatique dans les nuages (Cloud Computing) comme un nouveau modèle de déploiement des systèmes informatiques, les entrepôts de données profitent de ce nouveau paradigme. Dans ce contexte, il devient nécessaire de bien protéger ces entrepôts de données des différents risques et dangers qui sont nés avec l'informatique dans les nuages. En conséquence nous proposons dans ce travail une façon de limiter ces risques à travers l'algorithme le partage de clés secret de Shamir et nous mettons cette contribution en pratique (Karkouda 2011).

1 Introduction

L'informatique décisionnelle représente de nos jours un facteur primordial pour les grandes entreprises qui se trouvent face à un grand volume de données. En effet, ce grand volume des données accumulées au cours du temps est considéré comme une source riche pour les décideurs puisqu'il permet à ces derniers d'avoir une vue d'ensemble sur les différentes activités de l'entreprise et les aide à prendre des décisions. D'autre part, avec le développement du concept de l'informatique dans les nuages et les différents avantages qu'elle offre en termes de puissance de calcul, de temps de réponse et de réduction des coûts, les entreprises peuvent bénéficier pleinement de service qui permet de satisfaire leurs besoins à moindres coûts. En effet, la mise en œuvre d'un entrepôt de données dans les nuages représente pour chaque entreprise une bonne solution vue son efficacité et sa rentabilité. Toutefois, comme chaque avancée technologique, l'informatique dans les nuages apporte aussi son lot de risques notamment en terme de sécurité qu'il faut prendre en compte pour pouvoir bénéficier de tous les avantages apportés par cette solution.
D'après Mansfield-Devine, les systèmes traditionnels sont protégés par des firewalls et des passerelles d'où les cybercriminels doivent collecter des renseignements intensifs pour savoir qu'ils existent. Alors que dans le cloud computing les systèmes sont très visibles et sont conçus pour être accessibles de n'importe où. En plus, les applications avec ce type de dispositif sont accessibles à travers un navigateur ; il s'agit d'une interface dont les faiblesses sont bien connues (Steve 2008). Les risques de l'informatique dans les nuages s'accroissent avec

l'utilisation de la virtualisation, qui est l'une des techniques de base dans ce type de dispositif (Mladen et Vouk 2008). En effet, Zhou et son équipe ont montré qu'il existe une faille dans l'hyperviseur XEN puisqu'il permet aux machines virtuelles de consommer le temps CPU pour d'autres utilisateurs et permet le vol de service (Fangfei et al. 2011). Wei et al. ont aussi abordé les problèmes de la gestion de la sécurité des images de la machine virtuelle (Jinpeng et al. 2009).
A vrai dire, les dangers de l'informatique dans les nuages ne sont pas limités à ce stade puisqu'il faut s'assurer aussi que les services soient disponibles à tout moment ce qui n'est pas toujours le cas. Par exemple, en 2009 a eu lieu une coupure du courant dans le nuage d'Amazon dans les locaux abritant ses serveurs en Virginie (1). Une telle panne peut provoquer une grande perte pour les entreprises puisque l'activité de l'entreprise qui héberge son infrastructure est arrêtée. Plusieurs travaux ont noté que les entreprises sont réticentes à l'informatique dans les nuages parce qu'elle impose l'externalisation de leurs données. En fait, le problème qui revient toujours est celui de ne pas vouloir laisser leurs données entre les mains de la concurrence, chose plus difficile à contrôler lorsque les données sont confiées à un prestataire externe dont la localisation physique est souvent inconnue. Il ne faut donc pas penser à une solution utilisant l'informatique dans les nuages et particulièrement les entrepôts de données sans répondre aux questions suivantes : Est-il raisonnable de confier des données importantes et sensibles à un fournisseur de cloud computing ? Comment peut-on s'assurer que le fournisseur de cloud computing ne disparaisse pas un jour ? Nos données seront-elles effacées lorsqu'on veut changer de fournisseur ? Ya-t-il un risque de perte de données stockées dans les nuages ? Le transfert des données vers les nuages est-il sécurisé ?...
Dans cet article nous commençons par présenter un état de l'art sur la sécurité de l'informatique dans les nuages et plus particulièrement celle des entrepôts de données. Il s'agit d'une présentation synthétique et d'un examen critique des principaux travaux. Ensuite, nous décrirons la solution proposée qui permet de résoudre quelques problèmes liés à la sécurité des données, puis nous présenterons la concrétisation de cette solution à travers l'implémentation d'un prototype et nous terminerons par une conclusion et les perspectives.

2 Etat de l'art

L'informatique dans les nuages est un nouveau modèle de prestation de service informatique utilisant de nombreuses technologies existantes. Mais, comme toute nouvelle technologie elle a besoin de nombreuses améliorations et de la mise en place de normes précises (Sean et Kevin 2011) pour éviter les risques. La sécurité est souvent considérée comme le frein principal à l'adoption des services du cloud computing (Grange 2010). C'est ainsi que de nombreux travaux ont été consacrés à la recherche de solutions pour remédier à ce problème. Nous essayerons dans cette partie de présenter les principaux travaux de recherche qui ont proposé des solutions pour assurer la sécurité de l'informatique dans les nuages.

2.1 Sécurité de l'informatique dans les nuages

On peut classer les aspects de la sécurité dans les nuages en trois catégories qui sont la sécurité des données, la sécurité logique et la sécurité physique (Grange 2010). Nous ne nous intéressons ici à quelques travaux consacrés uniquement aux deux premières catégories :

2.1.1 Sécurité des accès et du stockage de données dans les nuages

Jensen et al. ont énuméré les différentes techniques utilisées dans cloud computing pour sécuriser les accès et ils ont dégagé les lacunes de ces techniques pour mettre en œuvre leur solution qui est basée sur le protocole TLS et la cryptographie XML (Jorg et al. 2009). Cette solution vient répondre au problème de navigateur web qui présente des lacunes au niveau sécurité. L'idée proposée consiste à utiliser le protocole TLS et à adapter le navigateur en intégrant la cryptographie XML. Cependant Wang et al. ont proposé une solution qui se base sur le code erasurecorrecting afin de permettre la redondance et garantir la fiabilité des données. Ils ont utilisé le jeton homomorphique pour l'exactitude du stockage et pour localiser les erreurs. La solution proposée est capable de détecter la corruption des données lors du stockage, elle peut garantir la localisation des données erronées et identifier le serveur qui a un mauvais comportement (Cong et al. 2009). Dans le cloud computing, aucune hypothèse sur la robustesse d'un nœud ne peut être faite. Divers facteurs imprévus peuvent tous entrainer une inaccessibilité temporaire de certains nœuds ou de l'inaccessibilité définitive des données. Dans un tel cas, les moyens traditionnels de protection des données sont souvent impuissants. Danwei Chen et Yanjun He ont proposé un algorithme de sécurité qui assure la restauration des données en cas d'échec de certains serveurs. Il s'agit d'un algorithme de séparation de données (Danwei et Yanjun 2010). Cet algorithme n'est qu'une extension du théorème fondamental de K équation en algèbre, l'algorithme du partage de la clé secrète de Shamir qui est un algorithme de cryptographie basé sur le partage du secret (2), l'algorithme de stockage de données en ligne de Abhishek (Parakh et Kak 2009) et la théorie du nombre. L'idée est de partager la donnée d en k parties d= d_1,d_2,d_3,..d_k. Ce partage est fait à l'aide de l'algorithme de séparation de données pour les stocker ultérieurement sur des serveurs choisis aléatoirement noté S= s_1, s_2,s_3,..s_m avec m>k. Le processus de stockage de données dans les nuages se fait donc sur deux étapes, la première consiste à diviser et stocker les données sur un serveur choisi arbitrairement et la deuxième consiste à pouvoir restituer ces données. A travers ces processus, les données sont prêtes à être transférées, stockées, traitées, en toute sécurité puisqu'elles sont cryptées. Les chercheurs ont déduit que la complexité temporelle de l'algorithme est la même pour générer k bloc de données et pour la restauration des données. Ils ont prouvé que même si un attaquant envahit un nœud de stockage, vole un bloc de données et essaie de rétablir la série de données, la complexité temporelle nécessaire pour faire les traitements ne peut pas être supportée par les environnements informatiques actuels. Parmi les autres avantages qui distinguent cette proposition on trouve la capacité de restaurer les données même si un ou plusieurs nœuds de stockage ne sont pas disponibles ce qui ne peut pas être le cas avec une solution traditionnelle de cryptographie.

2.1.2 Sécurité logique de l'informatique dans les nuages

Dans les nuages IAAS (Infrastructure as a service), les utilisateurs ont accès aux machines virtuels VM (3) sur lesquels ils peuvent installer et exécuter leurs logiciels. Ces machines virtuelles sont créées et gérées par un moniteur de machine virtuel VMM qui est une couche logicielle entre la machine physique et le système d'exploitation. Le VMM contrôle les ressources de la machine physique et crée plusieurs machines virtuelles qui partagent ces ressources. Les machines virtuelles ont des systèmes d'exploitation indépendants exécutant des applications indépendantes et sont isolées les unes des autres par le VMM. Ce type de disposi-

tif a provoqué beaucoup de problèmes de vulnérabilité de la machine virtuelle ce qui a poussé les auteurs à travailler dans ce domaine pour trouver des solutions efficaces. Zhou et al. ont proposé une solution pour éliminer la vulnérabilité de la machine virtuelle. La découverte des limites de l'hyperviseur XEN utilisé par AMAZON était leur point de départ. Ils ont proposé quatre approches pour améliorer la performance de cet hyperviseur qui sont basées sur la loi de Poisson, la loi de Bernoulli, la loi uniforme et enfin la loi exacte. Après une comparaison entre les quatre nouveaux modèles, ils ont déduit que la stratégie basée sur la loi de Poisson est la meilleure dans la pratique pour empêcher le vol de cycle (Fangfei et al. 2011). Wei et al. ont proposé un système de gestion des images de la machine virtuelle qui contrôle l'accès et la provenance de ces images à travers des filtres et des scanners qui permettent de détecter et réparer les violations en utilisant les techniques de fouille de données, ce système s'appelle Mirage (Jinpeng et al. 2009). S. Berger et al. ont aussi développé une technologie qui réponds aux problèmes rencontrés par la machine virtuelle. Cette technologie est appelée Trusted Virtual Data Center (TVDc), elle assure que les charges de travail ne peuvent être facturées qu'au client qui ont bénéficié du service. Elle assure aussi dans le cas de certains programmes malveillants comme les virus qu'ils ne peuvent pas se propager à d'autres nIJuds et elle permet également de prévenir les problèmes de mauvaise configuration. La TVDc utilise la politique d'isolement qui se base sur la séparation des ressources matérielles utilisées par les clients. Elle gère le centre de données, l'accès aux machines virtuelles et le passage d'une machine virtuelle à une autre (Fei et al. 2011).

2.2 Discussion

Nous avons présenté dans la partie précédente quelques travaux qui proposent de résoudre les problèmes liés à la sécurité dans le cloud computing. Quelques approches semblent être pertinentes et assurent un niveau de sécurité acceptable mais qui reste insuffisant. En plus, ces travaux ont mis en évidence de nouveaux problèmes : Danvei Chen et Yanjun He ont relevé la redondance des données. Mais ça n'empêche pas que l'idée proposée par ces chercheurs est très fiable pour sécuriser le transfert des données à travers le réseau et élimine le problème de non disponibilité du service en cas de panne de l'un des serveurs. Jensen et al ont proposé d'utiliser le protocole TLS et adapter le navigateur en intégrant la cryptographie XML. Une telle proposition n'est pas suffisante pour assurer la sécurité du transfert sur les réseaux puisqu'elle est basée sur une technologie dont ses faiblesses sont bien connues. On ce qui concerne l'idée proposée par Wang et ces collaborateurs, elle est basée sur le chiffrement homomorphique qui est la réponse à la question de confidentialité des données lors de transfert et lors de traitement dans le cloud. Néanmoins, le laboratoire de recherche en cryptographie dans le cloud de Microsoft a annoncé que cette nouvelle façons de chiffrer les données n'en est encore qu'à ces débuts et qu'ils sont loin de pouvoir exécuter sur les machines virtuelles des données cryptées avec le chiffrement homomorphique (4).
Généralement les solutions existantes pour sécuriser le transfert et les traitements des données sont basées sur la cryptographie des données qui n'est pas toujours une solution complète pour protéger les données, en plus le mécanisme de cryptage et de décryptage des données peut être intensive sur les processeurs ce qui engendre un gaspillage des ressources une chose que les fournisseurs des nuages ne veulent pas (Grange 2010).

3 Sécurisation des données par le secret sharing

3.1 Motivation

L'utilisation de l'informatique dans les nuages est basée sur la confiance qu'on peut accorder aux fournisseurs de ce type de service. Une telle situation est difficile à renforcée avec l'architecture traditionnelle du cloud qui repose sur un seul fournisseur. Cette dépendance menace la confidentialité des données des clients puisque ces dernières sont hébergées chez un seul prestataire externe qui risque de les exploiter. Pour cela nous proposons une autre façon d'hébergement des entrepôts de données afin d'éliminer la dépendance par rapport à un seul fournisseur en utilisant plusieurs fournisseurs. Cette solution rend les données hébergées chez chaque fournisseur non significatives et donc non exploitables.

3.2 Proposition

Notre proposition consiste à partager chaque donnée stockée dans l'entrepôt sur plusieurs fournisseurs des nuages à travers l'algorithme de secret sharing (Shamir 1979). Dans ce chapitre nous présentons en détail la solution pour sécuriser le stockage et l'exploitation d'un entrepôt de données dans les nuages, cette dernière est inspirée de l'idée proposée par Danwei Chen et Yanjun He dans leur article intitulé 'A Study on Secure Data Storage Strategy in Cloud Computing' (Danwei et Yanjun 2010). Il s'agit d'une solution basée sur l'algorithme de secret sharing qui partage les données en n-uplet et les stocke chez un seul fournisseur. Dans la solution que nous proposons, notre apport consiste à stoker le n-uplet chez plusieurs fournisseurs. Cette façon de répartir les données permet d'une part de stocker au niveau de chaque fournisseur une partie de l'information, celles-ci sont alors non compréhensibles et non exploitables par un utilisateur malveillant en cas d'intrusion et d'autre part de ne pas dépendre d'un seul fournisseur, ce qui minimise le risque de non disponibilité des données. Les étapes de la démarche sont :
-Chaque fournisseur des nuages possède une copie de l'architecture de l'entrepôt du client.
-Chaque donnée de l'entreprise est partagée et stockée chez les différents fournisseurs de manière à la rendre inexploitable par chaque fournisseur car non significative.
-Le nombre de fragments dépend du nombre de fournisseurs choisis par le client.
-Pour la restauration d'une donnée, le client doit récupérer les fragments stockés chez les différents fournisseurs pour reconstituer la donnée initiale (figure 1).

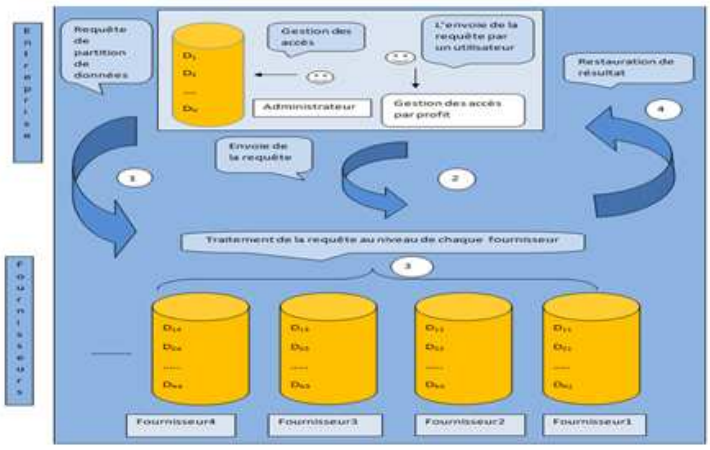

FIG. 1 – *Scénario d'un entrepôt de données partagé dans les nuages*

3.3 Le niveau de sécurité attendu

Notre solution assure trois niveaux de sécurité : - Capacité de restauration des données en cas de non disponibilité du service ou de la disparition d'un fournisseur puisque l'idée est basée sur l'algorithme de secret sharing qui est capable de reconstituer la donnée initiale à partir d'un nombre prédéfini de fragments qui peut être inférieur au nombre de fragments stockés chez les différents fournisseurs.
- Sécurité des transactions entre le client et les fournisseurs puisque les données qui transitent sur le réseau sont partielles et inexploitables.
- Sécurité des données stockées chez les différents fournisseurs puisque chacun d'eux n'a qu'une partie d'une donnée non significative.
Pour atteindre le niveau de sécurité attendu il faut respecter les deux règles énoncées dans les deux sous sections suivantes.

3.3.1 Choix des coefficients

Le choix des coefficients a une grande influence sur la sécurité des données. On suppose par exemple que les a_i 1<=i<=k-1 coefficients choisis aux hasards sont tous égaux à zéro, le polynôme f(x) =a_0+a_1x+a_2x_2+a_3x_3+...+$a_{K-1}x_{k-1}$ devient f(x)=a_0 ce qui n'assure pas la sécurité du transfert. Par conséquent l'algorithme nécessite que les coefficients ne doivent pas être nuls en même temps.

3.3.2 Complexité temporelle

On suppose qu'un attaquant envahit un nœud de stockage, et vole un bloc de donnée r_i et veut restaurer la donnée d'origine D avec les méthodes agressives basées sur r_i et décode les coefficients. Il a besoin de [P_{K-1}/(K-1) !] complexité temporelle, avec p» K» 2. Un tel rapport ne peut pas être calculé avec les capacités des traitements actuels des ordinateurs. En se

référant à cette théorie, nous pouvons remarquer que si on augmente le K on peut augmenter la complexité temporelle ce qui signifie la diminution des risques.
D'autre part le nombre K est un facteur qui dépend du nombre de pièces qu'on espère obtenir après la partition de la donnée d'origine D puisqu'il ne peut pas être plus grand que ce nombre. C'est ainsi qu'afin de diminuer les risques il faut augmenter le facteur K par conséquent augmenter le nombre de partitions N. Ce qui signifie qu'il faut augmenter le nombre de fournisseurs de l'informatique dans les nuages pour diminuer les risques de reconstitution de l'information.
La figure 2 présente l'augmentation de la complexité temporelle en fonction d'augmentation de facteur K qui est le nombre des fournisseurs dont on a besoin pour reconstituer l'information. Dans cet exemple le nombre entier p=9

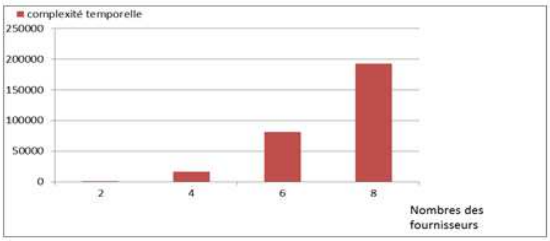

FIG. 2 – *L'augmentation de la complexité temporelle en fonction du nombre des Fournisseurs*

3.4 Résultats théoriques : Rapport coût/risque

L'avantage le plus important de l'informatique dans le nuage est que l'entreprise paye ce qu'elle consomme " pay as you go ", c'est-à-dire, le coût de consommation est en fonction de l'espace mémoire consommé, des accès sur le réseau et du temps de traitement des requêtes. D'où on peut résumer la fonction de coût par l'équation suivante :
D *C_S +T*C_T +T_{rq} *C_{Tr}= C_{tot} ; ou
D : les données stockées chez le fournisseur de nuage
C_s : Coût de stockage qui dépend de chaque fournisseur
T : Temps de traitement des requêtes
C_T : Coût de traitements de requêtes qui dépend de chaque fournisseur
T_{rq} : Taille de la requête et le résultat
C_{Tr} : Coût de transfert sur le réseau qui dépend de chaque fournisseur
Revenons maintenant sur la fonction de coût qui concerne notre proposition, elle change en fonction de nombre des fournisseurs avec lesquels l'entreprise s'est engagée d'où la fonction de coût se transforme en :
$\sum_{i=1}^{n}$ (D$_i$ *C_{Si} +T_i*C_{Ti} +T_{rqi} *C_{Tri})= C_{tot} ; où n est le nombre de fournisseurs des nuages
D'après cette formule on peut remarquer que le coût à payer à un seul fournisseur par l'entreprise avec notre proposition est n fois plus grand qu'une solution traditionnelle, mais elle assure aussi n fois plus la sécurité et réduit les risques. Le coût du risque correspond à l'impact que la perte de tout ou partie de l'entrepôt de données aura sur l'activité de l'entreprise, celle ci se calcule en temps de travail pour reconstituer les données. Plus précisément, notre

proposition est basée sur l'algorithme de secret sharing qui a besoin d'un nombre k de partie fixé dès le début pour la reconstitution de donnée. Ce nombre K représente dans notre proposition le nombre de fournisseurs du cloud qui doivent être disponibles pour la reconstitution des données. Néanmoins ce nombre de fournisseurs peut dépasser le nombre K nécessaire fixé dès le départ pour trouver la marge de sécurité en cas de non disponibilité de service de quelques fournisseurs par exemple et atteindre un nombre n qui dépend du choix de l'entreprise.
Pour diminuer le risque de non disponibilité, il faut alors augmenter le facteur n qui est le nombre de fournisseurs des nuages utilisé par l'entreprise. Cette dépendance influe sur le coût d'utilisation de l'informatique dans les nuages Ctot qui va augmenter proportionnellement au nombre de fournisseurs des nuages. Le rapport entre le coût et le risque reste alors un compromis qui dépend du niveau de sécurité que l'entreprise veut assurer : plus on diminue le risque de coalition par l'augmentation du nombre des fournisseurs plus on augmente le coût d'utilisation de l'informatique dans les nuages.
D'autres part la gestion des risques est devenue une partie intégrée dans les activités des entreprises. Elle consiste à gérer efficacement les risques auxquels le système informatique de l'entreprise est exposé et à préparer les contre-attaques pour dépasser ces risques. Une telle politique nécessite un budget financier important pour garantir la fiabilité et la continuité des services de l'entreprise puisqu'un accident peut provoquer des dégâts financiers importants (IBM 2008). En plus, en cas de faille de système informatique de l'entreprise elle risque de perdre ces clients et sa réputation sur le marché. Réduire ce type des risques est devenu une priorité stratégique pour les entreprises dans un contexte concurrentiel rude. En réalité la définition des risques associés au cloud computing est beaucoup plus large elle englobe les incertitudes, les dangers et les pertes (5) ce qui nécessite une gestion des risques plus efficace et un budget financier plus grand. Cette gestion des risques est assurée dans les entreprises par des moyens spécifiques comme la méthode Mehari etla méthode Marion. Ces méthodes aboutissent généralement à des résultats efficaces, qui consistent à estimer le coût des risques en cas de faille de système informatique. A travers ces deux facteurs qui sont le coût total de l'informatique dans les nuages calculé Ctot et le coût du risques estimé par les méthodes de gestion de risque, l'entreprise peut savoir à partir de quel moment une solution cloud computing devient coûteuse en tenant compte de la contrainte suivante : C_{tot} /coût des risques $<$ 1. Sachant que le coût des risques est une constante, l'entreprise peut diminuer ce rapport en diminuant le nombre de fournisseurs.

4 Mise en pratique de notre proposition

Dans ce chapitre nous allons présenter un prototype pour la mise en pratique de notre proposition. Ce prototype est constitué de trois fournisseurs de cloud computing chez lesquels nous hébergeons notre entrepôt de données contenant les données sur les ventes de produits dans plusieurs magasins, il s'agit d'une simulation d'un grand volume de données. Nous avons partagé les données en utilisant l'algorithme de secret sharing. La suite du chapitre contient les différentes fonctionnalités qui sont assurées par notre prototype.

4.1 Manipulation des données

4.1.1 Partition des données

Le partage des données se fait au niveau de l'entreprise à travers la première étape de l'algorithme de secret sharing. Cette première étape assure le partage de données en fonction de nombre des fournisseurs et inclus aussi l'envoie de chaque partie arbitrairement à un fournisseur (figure 3).

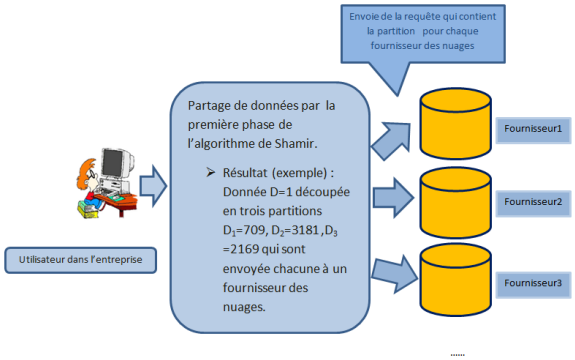

FIG. 3 – *La phase de partition de données*

4.1.2 Restitution de données

En fonction des différentes parties constituant l'information, la deuxième partie de l'algorithme de secret sharing est capable de restituer la donnée d'origine. Cette étape est faite au niveau de l'entreprise (figure 4).

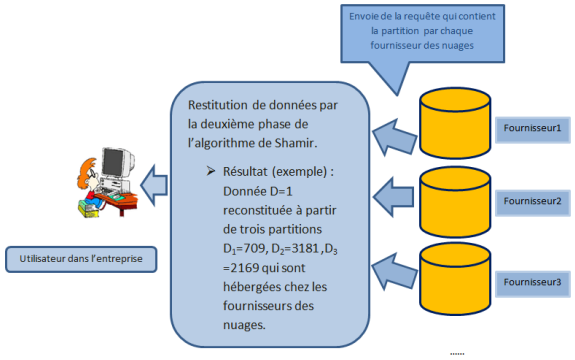

FIG. 4 – *La phase de restitution de données*

4.2 Autres opérateurs pour les analyses OLAP

On peut facilement faire des additions sur les données partagées, mais les analyses OLAP nécessitent d'autre types d'agrégation : moyenne, écart-type, min, max... qui nécessitent des adaptations pour qu'ils puissent fonctionner avec notre modèle. Dans notre prototype nous avons implémenté la variance et le maximum. Les deux sous sections qui suivent présentent le principe que nous avons utilisé pour l'implémenter.

4.2.1 La variance

Afin d'intégrer la variance dans notre travail nous avons proposé d'ajouter une colonne où on fait la multiplication de chaque valeur de xi au niveau de l'entreprise puis on partage la colonne au niveau des différents fournisseurs pour qu'on puisse se servir après pour la restitution de données. Ceci nous oblige à ajouter une autre colonne pour le stockage des xi aux carrés. La figure 5 explique le principe adopté pour assurer le calcul de la variance sur les données. L'exemple présenter consiste à calculer la variance sur les chiffres d'affaires par magasin, par département et par mois.

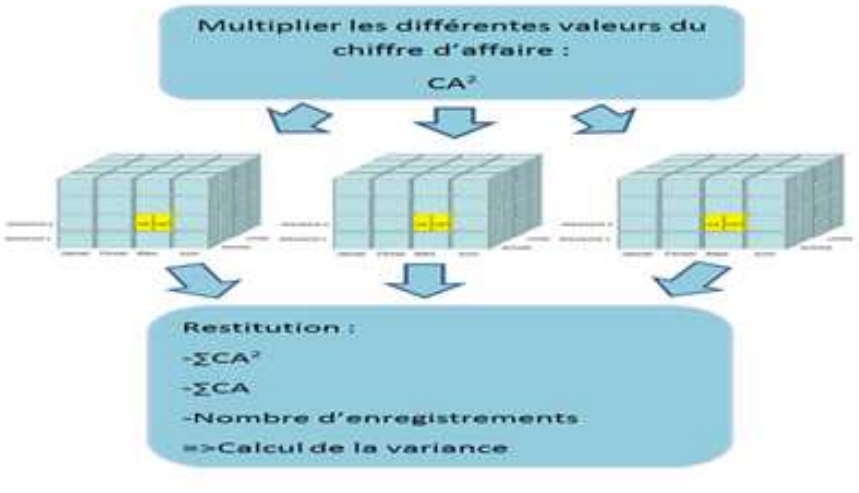

FIG. 5 – *Le processus de calcul de la variance*

4.2.2 Le maximum

L'idée que nous proposons est de calculer les valeurs des colonnes dont on veut estimer son maximum par la formule suivante : chaque enregistrement exposant un nombre très grand dans une première étape. Ensuite, on partage les nouvelles valeurs avec les différents fournisseurs. Une fois que nous voulons savoir le maximum des valeurs prix par la colonne on applique la fonction d'agrégation Max sur chaque entrepôt. En effet les résultats obtenues n'ont aucune

signification, pour cela il faut appliquer la méthode de restitution de donnée proposée par Shamir et ensuite multiplier le résultat par (1/le nombre que nous avons utilisé dans la première étape). Pour plus d'explication la figure 6 présente le processus de calcul du maximum sur le chiffre d'affaire.

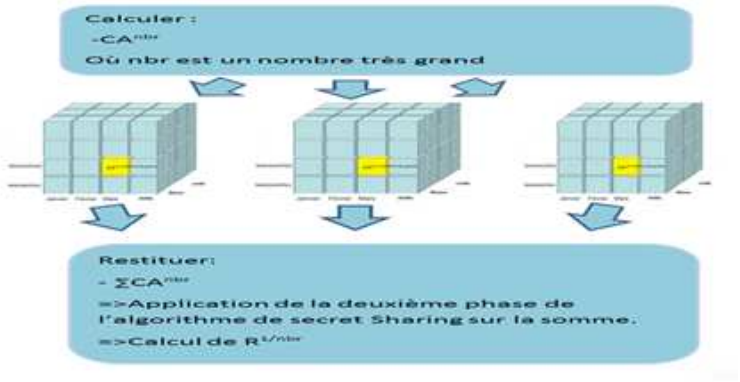

FIG. 6 – *Le processus de calcul de maximum*

4.3 Discussion

Comme il est déjà annoncé l'implémentation de la variance et du maximum nécessite l'ajout d'une colonne pour chacun. Alors que l'ajout des colonnes entraîne une nouvelle augmentation du volume de données, et donc du coût de la solution. Mais à priori, ça ne s'applique qu'aux mesures de l'entrepôt, donc à un ensemble limité d'attributs. Les opérateurs d'agrégation Max, variance, count, moyenne qui sont nécessaires pour les analyses OLAP sont mis en œuvre dans ce prototype.

5 Conclusion et perspectives

Cette étude nous a permis de constater que les problèmes majeurs d'hébergement d'un entrepôt de données dans les nuages est le manque de confiance accordée au fournisseur du cloud computing, de disponibilité de service et de sécurité de transfert des données vers le cloud computing. Pour diminuer ces risques, nous avons proposé une solution basée sur l'algorithme de secret Sharing, qui consiste à partager l'information chez plusieurs fournisseurs au lieu d'un seul. L'objectif de cette solution est de diminuer la vulnérabilité des données transmises. Chaque partie est stockée chez un des fournisseurs, ce qui les rend d'une part non

significatives pour chaque fournisseur et d'autre part résoud le problème de non disponibilité de service en cas de panne chez un des fournisseurs. Nous avons créé un prototype qui non seulement assure les traitements nécessaires pour le partage et la restitution des données mais aussi intègre quelques opérateurs d'agrégation (Max, Count, variance, moyenne) pour l'analyse OLAP.
Les résultats obtenus dans le cadre de ce travail montrent que les perspectives de recherche sur le sujet sont nombreuses :
- Etudier la possibilité d'intégrer la cryptographie traditionnelle pour sécuriser le transfert des requêtes sur les réseaux.
- Implémenter d'autres opérateur d'agrégation :exemple le Min qui est nécessaires pour les analyses OLAP.
- Appliquer une méthode de gestion des risques pour déterminer le coût de risque réel.
- Renforcer la sécurité des accès en intégrant dans la solution la gestion des accès à l'entrepôt dans le cloud computing en fonction des profits utilisateurs de l'entreprise.

6 Bibliography

Cong Wang, KuiRen ,Qian Wang ,Wenjing Lou (2009). Ensuring Data Storage Security in Cloud Computing, In the 17th IEEE International Workshop on Quality of Service (IWQoS'09), Charleston, South Carolina, July 13-15.

Danwei Chen, Yanjun He (September 2010). A Study on Secure Data Storage Strategy in Cloud Computing, Journal of Convergence Information Technology, Volume 5, Number 7.

Fangfei Zhou, Manish Goel, Peter Desnoyers, Ravi Sundaram (mars 2011). Scheduler Vulnerabilities and Attacks in Cloud Computing, College of Computer and Information Science Northeastern University, Boston, USA.

Fei Hu, MeikangQiu, Jiayin Li, Traavis Grant, Draw Tylor, Seth McCaleb, Lee Bulter and Richard Hamner (2011). A Review on cloud computing :Design challenges in Architecture and Security . Journal of Computing and Information Technology - CIT 19,1, 25-55, p. 17

Grange Philippe (quatrième trimestre 2010). Livre blanc sécurité du Cloud computing, analyse des risque, réponses et bonnes pratiques, Syntec numérique, p7.

IBM (Septembre 2008) :Méthodologie de gestion de risque informatique pour les directeurs des systémes d'information :un levier exceptionnel de création de valeur et de croissance,p2,p11.

Jinpeng Wei, Xiaolan Zhang, Glenn Ammons (2009). Managing Security of Virtual Machine Images in a Cloud Environment, Proceedings of the 2009 ACM workshop on Cloud computing security, New York.

Jorg Schwenk, Luigi Lo Lacono, Meiko Jensen, Nils Gruschka (2009). On Technical Security Issues in Cloud Computing, IEEE International Conference on Cloud Computing.

Karkouda K (septembre2011).Sécurité des entrepôts de données dans les nuages , mémoire de stage du master 2 ECD (extraction des connaissances à partir des données), encadré par Mme Nouria Harbi, Mr Jérome Darmont et Mr Gerald Gavin, laboratoire ERIC, université Lyon 2.

Mladen A. Vouk (September, 2008). Cloud Computing - Issues,Research and Implementations.Journal of Computing and Information Technology- CIT 16, 2008.Received : June, 2008 Accepted,p237

A.Parakh , S. Kak (2009). Online data storage using implicit security, Information Sciences, vol.179, no 3323-3331.

Sean Carlin et Kevin Curran (janvier-mars 2011). Cloud computing Security, International journal of Ambient Computing and Intelligence, vol.3, issue 9, p. 14.

A.Shamir (1979). How to Share a Secret, Communications of the ACM, vol.22, no.11, pp.612-613.

Steve Mansfield-Devine (December 2008). Danger in the clouds, Network Security, Volume 2008, Issue 12, Pages 9-11.

[1]www.presence-pc.com/actualite/Amazon-EC2-37511/

[2]http ://fr.wikipedia.org/wiki/partage de clé secrète de Shamir.

[3]Virtual Computing Lab. http ://vcl.ncsu.edu/.

[4]http ://blogs.orange-business.com/securite/2011/08/chiffrement-homomorphique-une-bete-concue-pour-le-cloud.html

[5]http ://www.cairn.info/resume.php ?ID-ARTICLE=RIGES-275-0063

Summary

With the rise of cloud computing as a new model for deploying computer systems, data warehouses benefit from this new paradigm. In this context, it becomes necessary to adequately protect these data warehouses of various risks and dangers that are born with cloud computing .Therefore, we propose in this work a way to limit these risks through the algorithm Shamir's secret Sharing and we make this contribution in practice .