Confidentialité des entrepôts de données dans le Cloud Computing: Etat de l'art et Perspectives

Amina El Ouazzani, Nouria Harbi, Hassan Badir

Confidentialité des entrepôts de données dans le Cloud Computing: Etat de l'art et Perspectives

Amina El ouazzani*
Nouria Harbi **, Hassan Badir*

* Université Abdelmalek Essaadi ENSA LabTIC 1818 Tanger MAROC
{ a.elouazzani2000,h.badir}@gmail.com,
http://www.ensat.ac.ma/
**Université Lumière Lyon 2 Laboratoire ERIC 69635 Lyon, Cedex FRANCE
nouria.harbi@univ-lyon2.fr

Résumé. Un entrepôt de données est une base de données regroupant l'ensemble des données fonctionnelles et critiques employées par la haute direction des organisations pour la prise des décisions stratégiques. L'hébergement de ces entrepôts de données dans le Cloud Computing (CC) permet de surmonter l'expansion sans fin des données, en raison de sa capacité de traitement et de stockage des données. Cependant la confidentialité de ces entrepôts de données dans le CC a besoin de nombreuses améliorations et de la mise en place des normes précises dont l'objectif est d'adapter les méthodes traditionnelles de contrôle d'accès à ce nouveau paradigme CC, car ces données seront confiées à un prestataire externe. L'objectif de cet article est de donner un aperçu des aspects pertinents du contrôle d'accès aux entrepôts de données. Cet aspect qui présente un des mécanismes de confidentialité les plus importants de CC, puisque le service CC ne peut pas appliquer le modèle de contrôle d'accès traditionnel en raison de ses caractéristiques d'accès, ainsi qu'il n'y a pas un protocole standard pour gérer la connectivité des utilisateurs de CC aux ressources hébergés.

1 Introduction

Les entrepôts de données forment le socle des processus décisionnels qui constituent un support efficace pour avoir une vue clair pour les décideurs sur l'efficacité des différentes activités de l'entreprise et les aider à prendre des bonnes décisions afin d'augmenter leur profits.

D'autre part l'entrepôt de données est la concentration de toutes les bases de données d'une façon dénormalisée en accédant rapidement aux données sur mesure, restituer en différents endroits. Il regroupe les données sensibles et très pertinentes, et secrètes de l'entreprise, telle que les données médicales, financières qui ne doivent pas être accessibles sans contrôle d'accès. Dans ce contexte plusieurs gouvernements ont adopté des lois pour protéger la vie privée de leurs citoyens. Parmi ces lois, HIPAA (Health Insurance Portability and Accountability Act HHS (1996)) vise à protéger les données médicales des patients américains en obligeant les

établissements du secteur des soins de la santé de suivre des règles strictes de sécurité, de même GLBA (Gramm Leach Bliley Act GPO 1999) oblige les organisations financières américaines à protéger les données de leurs clients Fernandez-Medina et al. (2006).

Les entrepôts de données reçoivent des téraoctets des données stockées d'une façon historique, depuis les systèmes de production de l'entrepôt de données. A un certain stade, et malgré l'excellent utilitaire des entrepôts de données, le coût de maintien devient injustifié pour l'entreprise.

Aujourd'hui, la solution de l'hébergement de l'entrepôt de données dans le Cloud gagnent progressivement plus de popularité dans les entreprises, car de nombreuses entreprises se rendent compte de ses avantages. Par contre elle présente aussi des menaces de sécurité à l'égard des données des utilisateurs. Le contrôle d'accès est l'un des mécanismes de sécurité les plus importants des services de cloud computing qui garantie la confidentialité des données, cependant le service Cloud ne peut pas appliquer le modèle de contrôle d'accès traditionnel en raison de son évolutivité et son élasticité, car il n'y a pas un protocole standard pour gérer la connectivité des utilisateurs du Cloud aux ressources hébergés.

Après la présentation de la problématique dans la section 2, Le reste de cet article est structuré comme suit. La section 3 présente une vue d'ensemble des travaux connexes. La Section 4 décrit une synthèse des travaux. Enfin, la section 5 présente nos conclusions et perspectives.

2 Le contrôle d'accès aux données entreposées dans le Cloud Computing

L'analyse des risques conduit généralement à étudier cinq aspects de sécurité de l'information : la confidentialité, l'intégrité, la disponibilité, la traçabilité et la gestion des accès.

- Confidentialité : elle consiste à préciser les personnes qui ont le droit d'accès aux données sensibles, les privilèges de chaque utilisateur, les mécanismes de contrôle d'accès ainsi que le moment de chiffrement des données que ce soit lors du stockage ou bien lors de leurs mouvements.
- Intégrité : Pour faire des analyses stratégiques, l'entreprise aura besoin d'assurer l'intégrité des ces données en précisant qui peut modifier une information, de définir les mécanismes de vérification si l'information est changée, et de contrôler la cohérence des informations.
- Disponibilité : L'entrepôt de données doit être accessible aux utilisateurs autorisés à tout moment, pour garantir un service et une productivité efficace et sans interruption.
- Traçabilité : Dans un entrepôt de données, on doit garder la trace des actions effectuées sur les systèmes afin de savoir qui a effectué une tâche.
- Gestion des accès : La gestion des accès aide à protéger la confidentialité, l'intégrité et la disponibilité des actifs en s'assurant que seuls les utilisateurs autorisés peuvent y avoir accès et les modifier.

Le contrôle des accès est basé sur les informations d'identité pour permettre et contraindre l'accès. Les utilisateurs doivent avoir un accès limité aux données sensibles ce qui garantie la confidentialité de ces données, dans ce sens il faut mettre en place des procédures pour les identifier afin d'éviter les opportunités inutiles d'accès aux données des clients.

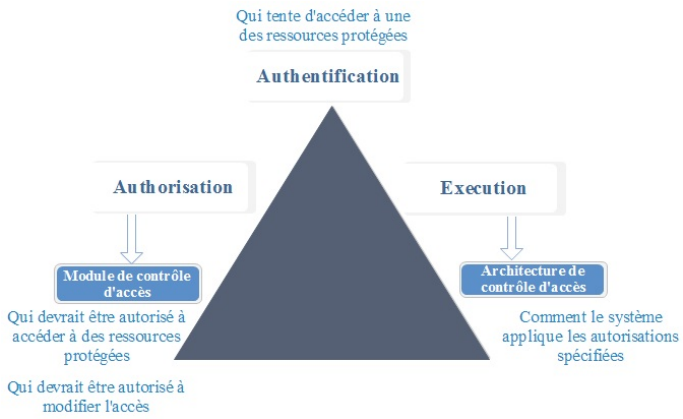

FIG. 1 – *La gestion d'accès aux données*

Lors de l'affectation des autorisations, il faut opter pour des contrôles d'accès fondés sur les rôles de manière à donner aux utilisateurs autorisés des accès qui dépendent de leurs fonctions. FIG 1 montre les trois étapes nécessaire pour l'accès aux données Ray et Ray (2014) :

- L'authentification : aborde le problème de déterminer l'identité de l'utilisateur qui essaye d'accéder à des ressources protégées.
- L'autorisation : s'exprime en terme d'un modèle de contrôle d'accès qui précise les ressources qui doivent être protégées, et qui est autorisé pour accéder à ces ressources.
- L'exécution : est un processus qui s'occupe d'appliquer les décisions prises d'autoriser l'accès ou de le refuser.

Aujourd'hui, la solution de l'hébergement de l'entrepôt de données dans le CC gagnent progressivement plus de popularité dans les entreprises, car de nombreuses entreprises se rendent compte de ses avantages. le contrôle d'accès dans un environnement CC a de nouveaux défis posés par la multi-location, l'élasticité et la dynamité de ce paradigme. Le mécanisme de contrôle d'accès doit prendre en compte les spécificités de déploiement d'un entrepôt de données dans le CC qui est très différent de celle de son déploiement sur site.

La migration des entrepôts de données vers le CC devrait améliorer la satisfaction de l'utilisateur final et induire une plus grande productivité de l'entreprise. Ce qui nécessite une haute performance qui peut être garanti par la mise en œuvre de l'intra-parallélisme de requête qui consiste à décomposer une requête complexe en sous-requêtes, et les traiter sur plusieurs processeurs, et enfin effectuer le post-traitement pour présenter une réponse à la requête principale Moussa et Badir (2013). Alors que la mise en place d'un mécanisme de contrôle d'accès ne doit pas augmenter la charge des traitements dont le but est d'avoir un système évolutif et productif avec des données qui sont bien protégées contre l'accès aux données interdites puisque ces données seront confiées à un prestataire externe.

Ce mécanisme de contrôle d'accès ne doit pas influencer l'évolutivité de l'entrepôt de données hébergé dans le CC en évaluant la charge des traitements sur l'échelle de temps, et en mesurant le nombre des requêtes traitées au cours d'un intervalle de temps. Un système

évolutif, devrait maintenir le même nombre Moussa et Badir (2013).

3 Etat de l'art

Récemment, un certain nombre de solutions de contrôle d'accès aux entrepôts de données ont été proposés, nous avons organisé les travaux selon trois parties, le contrôle d'accès aux entrepôts de données depuis les autorisations des sources de données, la deuxième présente le contrôle d'accès aux données de l'entrepôt lors de la modélisation. Dans la troisième partie traite le contrôle d'accès aux données entreposées dans les Cloud Computing et la dernière partie aborde le contrôle d'accès aux entrepôts de données existants (niveau exploitation)

3.1 Le contrôle d'accès aux entrepôts de données depuis les autorisations des sources de données

Certains chercheurs ont proposé l'exploitation des autorisations définit au niveau des sources de données afin de gérer l'accès aux entrepôts de données. Dans ce sens on trouve :

Rosenthal et Sciore (2000) Proposent une approche théorique qui exploite les permissions d'accès définit au niveau des sources de données, plutôt que de créer des nouveaux mécanismes d'accès. Ils utilisent la réécriture des requêtes pour vérifier que ces dernières respectent les restrictions définit dans les sources de données, et la création des vues relationnelles afin de minimiser le risque d'inférer les données sensibles.

Saltor et al. (2002) puisqu'il y a des similitudes entre l'architecture d'une base de données fédérée et l'architecture d'un entrepôt de données, les auteurs ont proposé l'utilisation du schéma des autorisations d'accès multi-niveaux défini pour la base de données fédérée sans être modifiée pour construire un entrepôt de données sécurisé, ce schéma des autorisations décrit les règles d'accès multi-niveaux.

3.2 Le contrôle d'accès aux données de l'entrepôt lors de la modélisation

Parmi les travaux qui ont été développés sur l'intégration du contrôle d'accès dans la phase conceptuelle des entrepôts, on trouve :

Sweeney (2002) Décrit un cas réel d'inférence de données par une démonstration d'identification de l'ancien gouverneur de l'état de Massachusetts, qui figure dans les 2 listes d'inscription des élections et les dossiers d'assurances maladie. Ils ont présenté une démonstration ou une façon didentification en se basant sur l'intersection des données d'un groupe d'assurance. Ces données supposé qu'ils sont anonymes, et une liste d'inscription des électeurs, ce qui permet de détecter le nom de l'ancien gouverneur William Weld et ses dossiers médicaux, en reliant les attributs partagés.

Fernandez-Medina et al. (2006) ont développé un modèle de contrôle d'accès et d'audit(ACA) spécifique aux entrepôts de données, qui repose sur deux politiques de gestion des accès : MAC et RBAC. Ils précisent des règles de contrôle d'accès lors de la modélisation d'un modèle conceptuel, en intégrant la notion de «profil utilisateur », qui est constitué d'une table isolée contenant toutes les informations des utilisateurs (identité, niveau de classification : top secret, secret, confidentiel ou inconnu). Ce modèle reste un modèle purement théorique car aucune solution concernant son implémentation n'a encore été proposée.

Villarroel et al. (2006) ont défini une extension OCL «Object Constraint Language» en utilisant les mécanismes d'extension UML2.0 pour résoudre les problèmes de la confidentialité, cette extension spécifie les contraintes de contrôle d'accès des éléments lors de la modélisation conceptuelle des entrepôts de données.

Soler et al. (2008) ont utilisé des mécanismes d'extension fournis par le CWM (Common Warehouse Metamodel) pour étendre le package relationnel et construire un schéma en étoile, qui représente les règles de contrôle d'accès et de vérification capturées pendant la phase conceptuelle de l'entrepôt de données.

Trujillo et al. (2009) ont développé une méthodologie comprenant quatre phases : analyse, modélisation, implémentation et validation, qui couvrent les cinq niveaux d'abstraction qui sont : analyse des besoins, niveau conceptuel, niveau logique, niveau physique et l'examen post- développement, ce dernier étant une nouvelle discipline introduite par Lujan et Trujillo (2004) . Cette méthodologie présente toutes les exigences de la contrôle d'accès tout au long du cycle de vie de l'entrepôt de données.

Blanco et al. (2010) ont Proposé une approche basée sur le diagramme états-transactions pour détecter les inférences au niveau de la conception. Cette proposition se focalise sur les requêtes sensibles et ses évolutions, mais ils ne tiennent pas en compte d'inférer les données à partir des données accessibles. L'approche est présentée sous forme d'un modèle de 3 états :

- Modèle statique : présente le profil UML spécifique aux entrepôts de données de Fernandez-médina 2007, en ajoutant un nouveau genre de règle nommée « Joint Rules », qui présente les privilèges nécessaires à certaines combinaisons.
- Modèle dynamique : ou bien le modèle états-transactions qui a l'objectif d'enrichir le modèle statique, en traitant les évolutions des combinaisons définis avec JR à travers l'application des opérations OLAP.
- Contrôle de session : cette étape rend plus d'intérêt aux sessions des utilisateurs afin de les analyser pour détecter toute possibilité d'inférence.

Rodriguez et al. (2011) ont présenté une extension d'UML 2.0 du diagramme d'activité. Cette proposition, libellée comme BPSec (Business Security Process), permet de définir un ensemble d'exigences de sécurité (contrôle d'accès, détection des risques d'attaques, non-répudiation, intégrité, confidentialité et vérification de la sécurité), ce qui améliore l'expressivité des modèles des processus métiers, et permet de sécuriser un entrepôt de données lors de son développement en prenant en considération cette exigence.

Triki et al. (2011) ont proposé un modèle pour sécuriser les données multidimensionnelles contre les inférences dans la phase conceptuelle, cette approche suppose que le schéma de DW est déjà conçu. Il permet de détecter les deux type d'inférences : Inférence précise : ou les valeurs des données déduites sont exactes. et inférence partielle : ou les valeurs des données sont partiellement divulguées, c'est-à-dire que l'utilisateur peut déduire une idée sur la valeur des données. Cette approche se compose de trois étapes :

- Etape 1. Un expert de domaine identifie les éléments sensibles à protéger en interrogeant le concepteur de l'entrepôt de données.
- Etape 2. Construire le graphe des inférences à partir du diagramme de classe, en précisant les éléments qui présentent les inférences précises ou partielles.
- Etape 3. Présenter l'entrepôt de données avec les annotations UML en mettant en évidence les deux types d'inférences.

3.3 Le contrôle d'accès aux données entreposées dans les Cloud Computing

Les chercheurs traitent le Cloud Computing comme un nouvel espoir pour les entrepôts de données, Malgré les avantages de cette solution, la confidentialité des données dans un environnement Cloud reste un risque à traiter, parmi les travaux qui traitent cette problématique on trouve :

Bensaidi et al. (2012) ont proposé un modèle de contrôle d'accès hybride basé sur la confiance Pour la sécurité des SID (Systèmes d'Information Distribués) , et sur le modèle de recommandation d'accès. Ils ont proposé également que des agents de confiance, appelés encore les tiers de confiances ou TTP pour «Trusted Third Party ». L'approche proposée se base sur la diminution de l'indice des permissions lors de la violation des droits fixés. Ainsi, Après un nombre bien défini des tentatives malveillantes, le connecté perd tous ses privilèges au sein de l'entreprise.

Al-Aqrabi et al. (2013) se focalisent sur la sécurité des systèmes décisionnels hébergés dans le Cloud Computing. Ils traitent la gestion des risques qu'apporte cette technologie du Cloud Computing, dans ce sens ils ont proposé deux modèles :

- UTM Unified Threat Management : ce modèle contient six réseaux locaux avec 500 postes de travail dans chaque réseau LAN, l'acheminement de tous les utilisateurs vers le BI Cloud est assuré par le réseau UTM_Cloud qui regroupe les applications qui s'occupent de la connectivité de tous les utilisateurs à BI Cloud qui regroupe les entrepôts de données et les serveurs d'applications OLAP contenant les tableaux de bord OLAP et les vues temporaires.
- Modèle basé sur la distribution de contrôle de sécurité entre plusieurs serveurs dans le Cloud Computing. Dans ce modèle, UTM Cloud est éliminé et les utilisateurs sont directement liés aux commutateurs de Cloud BI. L'exécution de la simulation du deuxième modèle, les auteurs ont observé que le temps de simulation a été réduit de deux minutes à une minute. Cependant, ce modèle est difficile de gérer car le mécanisme de la sécurité est réparti dans des milliers serveurs.

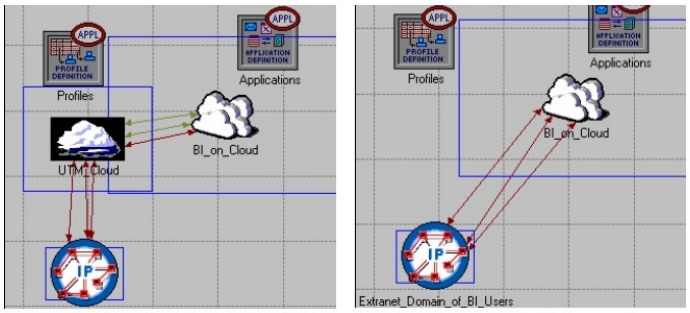

FIG. 2 – *comparaison entre les deux scénarios*

Ray et Ray (2014) ont proposé un nouveau modèle de contrôle d'accès dans le Cloud Computing, basé sur la confiance. La décision d'autoriser une demande d'accès est faite si les trois conditions suivantes sont remplies : (1) le rôle est autorisé à la permission demandée ; (2) l'utilisateur peut activer le rôle ; et (3) l'utilisateur est autorisé à la permission demandée. Chaque utilisateur a un niveau de confiance qui est nombre réel entre 0 et 1, plus la valeur de niveau de confiance est moins, l'utilisateur risque de perdre l'accès.

3.4 Le contrôle d'accès aux entrepôts de données existants (niveau exploitation)

Online Analytical Processing (OLAP) est devenue de plus en plus une composante importante et répandue des systèmes d'aide à la décision. Le serveur OLAP est censé d'assurer des accès en fonction des habilitations de chaque utilisateur. Il peut refuser les accès aux données d'une mesure, d'une dimension, et/ou au delà d'un niveau dans une hiérarchie. Les droits d'accès peuvent être explicitement spécifiés sur les tables/colonnes des tables de l'entrepôt de données. Cependant, le serveur OLAP tout seul ne peut pas protéger l'accès aux données interdites. Des travaux ont été réalisés pour renforcer les droits d'accès/habilitations des utilisateurs, et pour interdire tout utilisateur malicieux d'inférer des données qui lui sont interdites à partir des données auxquelles il a accès.

Kirkgoze et al. (1997) ont défini un modèle sécurisé pour les entrepôts de données qui consiste à élaborer un cube personnalisé possédant ses propres dimensions et hiérarchies. Ce modèle repose sur la politique de gestion AMAC. Il s'agit d'une extension du modèle MAC qui permet de spécifier les tâches que l'utilisateur peut exécuter selon son rôle au sein de l'organisation. L'intérêt d'un modèle comme celui-ci, est la flexibilité de l'assignation des rôles aux différents cubes virtuels.

Priebe et Pernul (2001) poursuivent leurs recherches concernant la création des mécanismes de contrôle des accès afin d'assurer la confidentialité des données, et ils ont créé un

mécanisme de contrôle d'accès sous forme d'un langage exprimant, au cours de la phase conceptuelle, les contraintes liées à la sécurité. Il s'agit d'un langage basé sur MDX «MulDimensionnale Xpression », celui-ci étant un langage de requête spécialisé dans l'interrogation et la manipulation des données multidimensionnelles. Il est comparable au langage SQL.

Triki et al. (2010) ont proposé une approche qui ne nécessite pas un traitement supplémentaire, après chaque phase d'alimentation de l'entrepôt de données. Elle est basée sur les réseaux Bayésiens afin de protéger un entrepôt de données contre les inférences, ils utilisent un module de contrôle, qui vise à interdire à un utilisateur d'inférer des données protégées à partir des données qui lui sont accessibles en utilisant les fonctions d'agrégations Min et Max.

Eavis et Althamimi (2012) ont présenté un cadre d'authentification qui s'appuie sur une algèbre spécialement conçue pour OLAP. Il est orienté objet et utilise des règles de réécriture de requêtes afin d'assurer l'accès aux données cohérentes à travers tous les niveaux du modèle conceptuel. Le processus est essentiellement transparent pour l'utilisateur, une notification est fournie dans le cas où un sous-ensemble de la demande initiale est renvoyé. Le résultat final est une approche intuitive et puissante pour l'authentification de base de données qui est uniquement adaptée au domaine OLAP.

4 Synthèse des travaux existants

Suite à l'étude des travaux existants, nous avons constaté les points suivants :

- Les auteurs Villarroel et al. (2006), Soler et al. (2008), Rodriguez et al. (2011), Priebe et Pernul (2001), Kirkgoze et al. (1997) ont réussi à bien définir les contraintes de la confidentialité des l'entrepôt de données, et proposer des approches intéressantes mais qui restent insuffisantes lors de l'hébergement de ces entrepôts de données dans le CC.
- La plupart de ces travaux qui traitent la confidentialité de l'entrepôt de données dans la phase conceptuelle sont appuyés sur le méta modèle CWM qui présente une norme pour l'échange et l'interopérabilité des métadonnées, cependant aucune proposition ne précise comment identifier le niveau de sensibilité des données, car la plupart des auteurs se basent sur le concepteur de l'entrepôt de données.
- Bien que les autorisations présentent l'axe principal pour garantir la confidentialité de l'accès à l'entrepôt, cependant l'absence d'une norme qui gère la précision de ces autorisations peut provoquer des incohérences et des inférences comme conséquences. Dans ce sens, on trouve le travail de Saltor et al. (2002) qui ont proposé l'utilisation du schéma des autorisations défini pour les bases de données fédérées sans aucune modification pour construire un entrepôt de données sécurisé, et Rosenthal et Sciore (2000) qui proposent la réécriture des requêtes afin de vérifier que ces dernières respectent les restrictions définit au niveau des sources.
- A noter également, que la notion d'inférence a été cité dans plusieurs travaux en tant qu'élément essentiel pour garantir la confidentialité, et dont la maîtrise est cruciale. Dans ce sens, on trouve le travail de Triki et al. (2011) qui ont proposé une approche qui permet de détecter les inférences partielles et précises, Blanco et al. (2010) qui proposent

une approche une approche basée sur le diagramme d'état-transactions permettant de détecter les inférences dans la phase conceptuelle. Néanmoins, malgré les risques élevés d'inférences, il n'est pas suffisamment pris en compte dans la phase conceptuelle.

	$Classification$	$Autorisation$	$Inference$
Kirkgoze et al. (1997)	Non	Non	Non
Rosenthal et Sciore (2000)	Non	Oui	Oui
Priebe et Pernul (2000)	Non	Non	Non
Saltor et al. (2002)	Non	Oui	Non
Sweeney (2002)	Non	Non	Oui
Fernandez-Medina et al. (2006)	Non	Non	Non
Villarroel et al. (2006)	Non	Non	Non
Soler et al. (2008)	Non	Non	Non
Blanco et al. (2010)	Non	Non	Oui
Triki et al. (2010)	Non	Non	Oui
Rodriguez et al. (2011)	Non	Non	Non
Triki et al. (2011)	Non	Non	Oui
Eavis et Althamimi (2012)	Non	Non	Non

TAB. 1 – *Comparaison des travaux sur la confidentialité au niveau modélisation et exploitation des entrepôts de données*

D'autre part, l'intégration d'un entrepôt de données dans un environnement Cloud Computing a pris beaucoup d'intérêt par les organisations et chercheurs, cela est dû aux avantages qu'elle offre, alors que la gestion d'accès reste un risque à traiter selon les spécificités de cet environnement. Dans ce sens on trouve le travail de Al-Aqrabi et al. (2013) ont proposé 2 modèles qui se focalisent sur la centralisation et la distribution des applications de la sécurité des systèmes décisionnels hébergés dans le Cloud, les auteurs ont observé que le temps de simulation a été réduit dans le modèle ou les applications de la sécurité sont distribuées mais il reste un modèle difficile à gérer, ainsi qu'ils n'ont pas précisé un mécanisme de contrôle d'accès. Bensaidi et al. (2012) ont proposé un modèle de contrôle d'accès hybride basé sur la confiance, et sur le modèle de recommandation d'accès dédiés aux systèmes d'information distribués. Le besoin d'un contrôle d'accès plus standardisé pour le Cloud devient plus qu'urgent, cependant l'inexistence à ce jour de normes reste une sérieuse limitation du Cloud.

5 Conclusion et perspectives

Dans cet article, nous avons défini la problématique de la confidentialité des entrepôts de données dans le CC. Ensuite, nous avons présenté un état de l'art sur la confidentialité des entrepôts de données en général, qui reposent sur le contrôle d'accès, et celle des entrepôts de données dans le CC. Nous avons aussi précisé les limites des travaux étudiés comme une base pour définir nos perspectives. Notre travail se compose de deux parties, La première consiste à traiter la confidentialité des données de l'entrepôt :

- Puisque la plupart des auteurs se basent sur le concepteur de l'entrepôt de données pour définir le niveau de sensibilité des données de l'entrepôt, ce qui peut provoquer des risques de divulgation des données sensibles en cas d'absence d'une approche. Nous avons l'intérêt de définir une approche qui consiste à classifier les données selon leur niveau de sensibilité(Sensible, Trop sensible, Confidentiel...).
- Les politiques de contrôle d'accès à partir des sources de données opérationnelles est un domaine attractif, qui doit être intégré dans la phase conceptuelle de l'entrepôt de données Fernandez-Medina et al. (2006). Ce domaine dans lequel certain chercheurs ont effectué des efforts tel que Saltor et al. (2002) et Rosenthal et Sciore (2000), mais qui restent insuffisants. Dans ce sens , nous avons l'intention de définir une approche qui permet de coordonner les droits d'accès à l'entrepôt avec ceux des sources de données d'une façon automatique, pour bien définir les autorisations d'accès à l'entrepôt de données.
- Les inférences constituent une menace grave sur la vie privée des utilisateurs.Néanmoins sauf le travail de Triki et al. (2011), l'état de l'art ne traite pas suffisamment ce menace. Dans ce sens nous avons l'intérêt de continuer sur les perspectives de Blanco et al. (2010) en détectant les inférences possibles d'une façon automatique en analysant les sessions des utilisateurs.
- Parmi nos perspectives est de mettre en place un profil d'usage d'un utilisateur, ce profil aidera l'administrateur à bien gérer l'utilisation et l'accès à l'entrepôt de données. Parmi les privilèges qu'il traitera sont : la gestion des mots de passes , le nombre des tentatives de violation des droits d'autorisations en précisant des réactions automatiques, la localisation d'un utilisateur lors d'une tentative de violation, traçabilité des actions, nombre des sessions possibles pour un utilisateur et la durée d'une session.

La deuxième partie de notre travail consiste à traiter la gestion d'accès aux données de l'entrepôt hébergées dans le Cloud Computing et proposer une approche qui permet de gérer de manière fine et précise qui a accès à quoi, quand comment et selon quelles conditions , pour un entrepôt de données hébergé dans le Cloud Computing, en se basant sur le profil utilisateur et le profil d'usage en tenant compte des spécificités de l'environnement Cloud Computing.

Références

Al-Aqrabi, H., L. Liu, R. Hill, Z. Ding, et N. Antonopoulos (2013). Business intelligence security on the clouds: challenges, solutions and future directions. *Service Oriented System Engineering (SOSE), IEEE 7th International Symposium on (pp. 137-144). IEEE*.

Bensaidi, M., A. Aboulkalam, et A. Marzouk (2012). Politique de contrÃ´le d'accÃ¨s au cloud computing: Recommandation Ã base de confiance. *Network Security and Systems (JNS2), 2012 National Days of (pp. 90-96). IEEE*.

Blanco, C., E. FernÃ¡ndez-Medina, J. Trujillo, et J. Jurjens (2010). Towards the secure modelling of olap users behaviour.

Eavis, T. et A. Althamimi (2012). Olap authentication and authorization via queryre-writing. *The Fourth International Conference on Advances in Databases, Knowledge, and Data Applications*, 130–139.

Fernandez-Medina, E., J. Trujillo, R. Villarroel, et M. Piattini (2006). Access control and audit model for the multidimensional modeling of dws. *Decision Support Systems*, 1270–1289.

Kirkgoze, R., N. Katic, M. Stolba, et A. Tjoa (1997). A security concept for olap. *Proceedings of the 8th International Workshop on Database and Expert System Applications (DEXA'97)*, 619–626.

Moussa, R. et H. Badir (2013). Data warehouse systems in the cloud: rise to the benchmarking challenge. *Journal International of Computers and Their Applications*, 245.

Priebe, T. et G. Pernul (2000). Towards olap security design - survey and research issues. *Proceedings of the 3rd ACM International Workshop on Data Warehousing and OLAP (DOLAP'00)*, 33–40.

Priebe, T. et G. Pernul (2001). A pragmatic approach to conceptual modeling of olap security. *Proceedings of the 20th International Conference on Conceptual Modeling (ER'01) 2224*, 311–324.

Ray, I. et I. Ray (2014). Trust-based access control for secure cloud computing. *High Performance Cloud Auditing and Applications (pp. 189-213). Springer New York*.

Rodriguez, A., E. Fernandez-Medina, J. Trujillo, et M. Piattini (2011). Secure business process model specification through a uml 2.0 activity diagram profile.

Rosenthal, A. et S. Sciore (2000). View security as the basis for data warehouse security.

Saltor, F., M. Oliva, A. Abello, et J. Samos (2002). Building secure data warehouse schemas from federated information systems.

Soler, E., J. Trujillo, E. Fernandez-Medina, et M. Piattini (2008). Building a secure star schema in data warehouses by an extension of the relational package from cwm. *Computer Standards and Interfaces 30*, 341–350.

Sweeney, L. (2002). k-anonymity: A model for protecting privacy.

Triki, S., H. Ben-Abdallah, J. Feki, et N. Harbi (2010). SÃ©curisation des entrepÃ´ts de donnÃ©es contre les infÃ©rences en utilisant les rÃ©seaux bayÃ©siens. *6Ã¨mes JournÃ©es francophones sur les EntrepÃ´ts de DonnÃ©es et l'Analyse en ligne*, 35.

Triki, S., H. Ben-Abdallah, N. Harbi, et O. Boussaid (2011). Securing data warehouses: A semi-automatic approach for inference prevention at the design level. *1st International Conference on Model and Data Engineering, Lecture Notes in Computer Science (LNCS) by Springer-Verlag*, 311–324.

Trujillo, J., E. Soler, C. Blanco, et E. Fernandez-Medina (2009). Designing secure data warehouse by using mda and qvt. *Journal of Universal Computer Science 8 15*, 1607–1641.

Villarroel, R., E. Fernandez-Medina, et M. Piattini (2006). A uml 2.0/ocl extension for designing secure data warehouses. *Journal of Research and Practice in Information Technology 38*, 31–43.

Confidentialité des entrepôts de données dans le Cloud Computing: Etat de l'art et Perspectives

Summary

A data warehouse is a database of all functional data and criticism employed by management organizations for taking strategic decisions. The hosting of these data warehouses in the Cloud Computing (CC) overcomes the endless expansion of the data, because of its processing capacity and data storage. However, the confidentiality of these data warehouses in the CC needs many improvements and implementing specific standards which aims to adapt traditional methods of access control to this new paradigm CC. The objective of this article is to provide an overview aspects of access control to data warehouses. this aspect, which that presents one of the most important confidentiality mechanisms cloud computing because the cloud service can't apply classical access control model because of its access characteristics, and there is no standard protocol for managing the connectivity of users to hosted cloud resources.