La confidentialité des entrepôts de données dans le Cloud Computing à base de profil utilisateur

Amina El Ouaazzani, Nouria Harbi, Hassan Badir

La confidentialité des entrepôts de données dans le Cloud Computing à base de profil utilisateur

Amina EL Ouaazzani[1], Nouria Harbi[2], Hassan Badir[3]

[1 3]Laboratoire LabTIC, Ecole Nationale des Sciences Appliquées Tanger, Maroc
{a.elouazzani2000,hbadir}@gmail.com
[2] Laboratoire Eric, Université Lyon II, France
nouria.harbi@univ-lyon2.fr

Résumé.
Un entrepôt de données (ED) présente un facteur primordial de l'entreprise qui donne une vue clair sur ses activités et une source riche pour les décideurs. Il contient les données sensibles sur l'entreprisse et ses clients, et par conséquence elles ne doivent pas être accessibles sans contrôle d'accès.(**El ouazzani 2018**) La solution de l'hébergement de l'ED dans le CC (Cloud Computing) gagne progressivement plus de popularité dans les entreprises, car elle permet de surmonter l'expansion sans fin des données et bénéficier de sa capacité de traitement et le stockage de ces données. Cependant la confidentialité de ces EDs dans le CC a besoin de nombreuses améliorations et de la mise en place des normes précises, en raison de l'évolutivité et l'élasticité du paradigme CC, car il n'y a pas un protocole standard pour gérer la connectivité des utilisateurs du CC aux ressources hébergés en prenant compte la performance d'exécution des requêtes. L'objectif de nos travaux est de proposer un cadre garantissant la confidentialité des EDs hébergés dans le CC à base de profil utilisateur.

1 Introduction

Un ED représente est un facteur primordial pour la haute direction qui cherche à prendre les bonnes décisions stratégiques. Il présente une source des données cruciales de l'entreprise et la vie privée de ses clients telle que les données médicales, financières protégées par des lois, parmi ces lois, HIPAA[1] (Health Insurance Portability and Accountability Act HHS (1996)). Et par conséquence elles ne doivent pas être accessibles sans contrôle d'accès.

A un certain stade, et malgré l'excellent utilitaire, Le maintien de ces EDs de grande quantité de données au sein de l'entreprise implique un grand investissement matériels et ressources humaines afin de les gérer. Aujourd'hui, la solution de l'hébergement de l'ED dans le CC gagne progressivement plus de popularité dans les entreprises, car elle permet de surmonter l'expansion sans fin des données et bénéficier de sa capacité de traitement et le stockage de ces données.

Le contrôle d'accès est l'un des mécanismes de sécurité les plus importants des services de CC qui garantit la confidentialité des données, cependant le service CC ne peut pas appli-

[1] http://www.hhs.gov/hipaa/

quer 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 CC aux ressources hébergés en prenant compte la performance d'exécution des requêtes **(Blanco, et al., 2015)**.

Dans cet article, nous présentons nos travaux qui se composent de trois parties :

- La première partie décrit un mécanisme de contrôle d'accès qui aide le propriétaire à bien définir les permissions de chaque profil utilisateur selon son rôle dans l'entreprise, et de calculer le niveau de sensibilité de chaque élément de l'ED.
- Notre deuxième contribution permet de détecter des cas d'inférence d'un utilisateur qui occupe un ou plusieurs rôles en analysant la totalité des permissions en se basant sur cinq règles proposées.
- La troisième partie, traite la confidentialité des EDs hébergés dans le CC à base de profil usage d'un utilisateur, en augmentant le coefficient de confidentialité/performance afin de contrôler l'accès à l'ED contenant une grande quantité de données en maintenant l'évolution du système et l'optimisation de temps de réponse.

Après la présentation de la problématique dans la section 1, le reste de cet article est structuré comme suit. La section 2 présente une vue d'ensemble des travaux connexes. La Section 3 décrit l'architecture proposée dont l'accès est basé sur les profils des utilisateurs. La section 4 présente la mise en œuvre et le test de notre solution. Enfin, la section 5 présente nos conclusions et perspectives.

2 Etat de l'art et synthèse

2.1 Etude de l'existant

Nous avons organisé les travaux selon deux parties, la première est consacrée à la présentation des modèles de contrôle d'accès intégrant la confidentialité dans le processus de modélisation des EDs (niveau conception), Dans la deuxième partie nous traitons les modèles de contrôle d'accès pour un ED déjà mise en place (niveau exploitation) **(El ouazzani, et al.,2015).**

2.1.1 La confidentialité dans le processus de modélisation des EDs

Afin de garantir la confidentialité d'ED au niveau conception, certains auteurs **(Rosenthal, 2000), (Saltor, et al., 2002)** ont proposé l'utilisation des autorisations définit au niveau des sources de l'ED. Alors que d'autres auteurs **(Trujillo, et al., 2009) (Soler, et al., 2008)** ont considèrent cette proposition non performante puisque l'ED à ces propres caractéristiques. Dans des travaux recent, le langage UML (**U**nified **M**odeling **L**anguage) présente un standard afin de modéliser les regles de sécurité d'un ED. Dans ce sens, nous pouvons citer le travail **(Blanco, et al, 2015)** qui présente une architecture MDA (Model Driven Architecture) automatique pour sécuriser un ED, cette architecture est composé d'un modèle logique et ses transformations depuis le modèle conceptuel en utilisant l'extension de UML et le package CWM (**C**ommon **W**arehouse **M**etamodel). Ainsi que le travail **(Arora, et al., 2016)** qui modélise le contrôle d'accès à des données sensibles de l'ED dans la phase analyse

et conception, à l'aide des diagrammes UML et la programmation orientée objet est la dernière tendance dans l'industrie du logiciel en raison de ses différentes fonctionnalités.

A noter également que la gestion des inférences s'est inspirée et s'inspire encore aujourd'hui des travaux réalisés dans le domaine des ED ou les bases de données en général. On retrouve dans la littérature le travail de **(Triki, et al., 2013)** qui propose un modèle pour sécuriser les données multidimensionnelles contre les inférences précises et partielles. Cette approche consiste à identifier les éléments sensibles à protéger en interrogeant le concepteur de l'ED. Ensuite, le propriétaire construit un graphe permettant de détecter les combinaisons sensibles. Par contre le travail de **(Blanco, et al., 2010)** traite 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. Ce travail indique que la combinaison de plusieurs permissions peut être plus sensible, ce qui est approuvé dans le travail de **(Sweeney, 2002).** Ce travail décrit un cas réel de l'inférence des données sensibles, par une démonstration d'identification du nom de l'ancien gouverneur « William Weld » et ses dossiers médicaux en se basant sur le croisement des données d'un groupe d'assurance, et une liste d'inscription des électeurs. Nous trouvons également, le travail de **(Accorsi, et al., 2013)** qui propose une approche dans laquelle les règles d'inférences sont connues par le moteur d'inférence sans mentionner comment les préciser. Le diagramme proposé montre un processus de détection des inférences qui se compose d'une politique composée dans lequel l'utilisateur compose la politique et les règles de confidentialité. Ensuite, le moteur d'inférence prend cette politique qui calcule à son tour toutes les fermetures d'inférence possibles de la politique entrée en se basant sur un algorithme. Et Le noyau teste pour chaque élément non noyau s'il est obtenu à partir d'un élément noyau.

2.1.2 La confidentialité des EDs au niveau exploitation

Le contrôle d'accès à un ED déjà mis en place prend en compte l'emplacement de l'ED qui peut être le site de l'entreprise ou chez un fournisseur CC :

- **La confidentialité d'un ED sur le site de l'entreprise :** la plupart des méthodes de contrôle d'accès à l'ED déjà proposées se basent sur le profil utilisateur. Ce dernier souffre toujours du problème de l'efficacité dans la gestion de l'intégrité. Dans ce sens les auteures **(Thangaraju, et al., 2016)** ont proposé un profil multi-utilisateurs orienté vers la gestion de l'intégrité basée sur la mesure des profondeurs d'accès en fonction du niveau des objets appelés et du niveau d'accès autorisé à l'utilisateur et du nombre d'objets auxquels l'utilisateur a accès. Alors que dans un autre travail **(Kechar, et al., 2015),** qui propose un système de contrôle d'accès basé sur les rôles en exploitant l'architecture de la norme XACML, afin de garder les performances des requêtes d'aide à la décision.
- **La confidentialité d'un ED hébergé dans le CC :** Malgré les avantages de la solution de l'hébergement d'un ED dans le CC, la confidentialité des données dans cet environnement reste un risque à traiter. Parmi les travaux qui traitent cette problématique on trouve (**Al-Aqrabi, et al., 2015)** qui se focalise sur la sécurité des systèmes décisionnels hébergés dans le CC et il décrit deux modèles de gestion des accès en tenant en compte le rapport temps de réponse/sécurité. Alors que d'autres travaux **(Bensaidi, et al., 2012) (Ray, et al., 2014)** se basent sur la notion de la confiance, on se focalisant sur la diminution du niveau de confiance affecté à l'utilisateur lors d'une tentative de violation des droits fixés. Nous trouvons également le travail **(Naushahi, 2016)** qui utilise le concept de liste de contrôle d'accès ACLs en intégrant la notion de Profile en définis-

sant des règles pour chaque profil afin d'accorder l'accès à un système et à des ressources hébergé dans le CC, Les résultats de la simulation montrent que cette solution offre un temps d'accès aux données réduit en diminuant les demandes d'authentification

2.2 Limitation des solutions existantes

La protection des ED contre les accès illégaux s'est fait sentir et traiter d'une manière incontestable dans plusieurs travaux **(Fernandez-Medina, et al., 2006) , (Soler, et al., 2008), (Trujillo, et al., 2009)**. Suite à l'étude des travaux existants, nous avons constaté les points suivants :

- La confidentialité des ED a été traditionnellement considérée dans la mise en œuvre définitive d'un ED **(Villarroel, et al., 2006) (Eavis, et al., 2012)**, par contre les travaux les plus récents **(Blanco, et al., 2015) (Rodriguez, et al., 2011)** considèrent son inclusion dans les stades de développement ce qui peut produire des solutions de qualité plus robustes, ainsi le système peut accueillir ces exigences de sécurité d'une façon plus naturelle.
- La majorité des travaux de recherche, surtout ceux intervenant dans la phase de modélisation conceptuelle, se sont appuyés sur le méta modèle CWM, dans le but de concevoir un ED sécurisé. Sachant que le modèle CWM est basé sur trois standards, à savoir UML, MOF et XMI. pour représenter correctement toutes les règles de sécurité et d'audit définies dans la modélisation conceptuelle des ED.
- Une architecture MDA pour une conception automatique, sécurisé d'un ED est appliquée dans **(Blanco, et al., 2015), (Inmon, 1991)**, mais les deux approches ont été incapables de comprendre des règles de sécurité qui sont complexes.
- La plupart des travaux modélisent le contrôle d'accès à base des politiques RBAC (**R**ole **b**ased **A**ccess **C**ontrol) et MAC (**M**andatory **A**ccess **C**ontrol), alors que le profil utilisateur est considéré comme une table isolée qui regroupe les données nécessaires pour l'accès d'un utilisateur d'une façon statique sans la prise en compte des priorités de l'utilisateur authentifié.
- Bien que les autorisations présentent l'axe principal pour garantir la confidentialité de l'accès à l'ED, 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, certains auteurs **(Rosenthal, 2000) , (Saltor, et al., 2002)** ont proposé de tirer le modèle de contrôle d'accès à l'ED, à partir des sources de données, tandis que d'autres auteurs **(Priebe, et al., 2001) (Fernández-Medina, et al., 2007)** ont considéré cette proposition difficile puisque les données sources proviennent de différents systèmes (avec des politiques différentes). Ainsi que les systèmes opérationnels utilisent le modèle relationnel alors que les systèmes OLAP utilisent le modèle multidimensionnel.
- A noter également, que la notion d'inférence a été citée dans plusieurs travaux en tant qu'élément essentiel pour garantir la confidentialité, et dont la maîtrise est cruciale. Néanmoins, malgré les risques élevés d'inférences, il n'est pas suffisamment pris en compte dans la phase conceptuelle.
- Aucun travail ne propose une méthode qui permet d'assurer la cohérence des permissions d'un utilisateur selon son profil.
- Aucun travail ne propose une méthode conviviale pour détecter les combinaisons sensibles qui peuvent provoquer des inférences.

Nous constatons que la plupart des travaux affecte la tâche de la classification des données selon leur niveau de sensibilité (Très sensible, sensible, confidentiel) au propriétaire de données. Sachant que selon le rôle de l'utilisateur, le propriétaire de données lui affecte un niveau de sensibilité des données pour accéder à des données possédant le même niveau de sensibilité ou inférieure. Le propriétaire de l'ED peut alors attribuer un niveau de sensibilité moins important à une donnée cruciale. Il en résulte cependant un problème de perte de confidentialité de l'information. De plus, les permissions définit au niveau des sources ne sont pas suffisamment exploitées pour aider le propriétaire à bien déterminer les permissions d'un utilisateur de l'ED.

A noter également que la solution de l'hébergement de l'ED dans le CC prend de plus en plus sa place dans les entreprises. Afin de bénéficier de ses avantages, des nouveaux défis concernant la sécurité des données hébergés ont été posés par la multi-location, l'élasticité et l'évolutivité de ce paradigme. Suite à l'étude des travaux existants dans ce sens, nous avons constaté aussi les points suivants :

- D'après le travail de **(Moussa, et al., 2013) et (Al-Aqrabi, et al., 2013)**, le mécanisme de contrôle d'accès ne doit pas influencer l'évolutivité et la performance de l'ED 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. Alors qu'ils n'ont pas proposé un mécanisme dans ce sens.
- les auteurs **(Naushahi, 2016)** ont utilisé le concept de liste de contrôle d'accès ACLs afin d'accorder l'accès à un système et à des ressources hébergées dans le CC, Les résultats de la simulation montrent que cette solution offre un temps d'accès aux données réduit. Par contre ils n'ont pas utilisé l'historique des accès qui peut minimiser le trafic et par conséquence réduire le temps de réponse.
- Plusieurs travaux **(Bensaidi, et al., 2012), (Ray, et al., 2014)** proposent d'utiliser la notion de confiance qui consiste à attribuer un niveau de confiance à chaque utilisateur, chaque tentative de violation provoque sa diminution, Après un nombre bien défini des tentatives malveillantes, le connecté perd tous ses privilèges au sein de l'entreprise. Ce qui peut dégrader la performance, augmenter la charge de traitement en recalculant le niveau de confiance à chaque tentative de violation, et retarder le travail d'un utilisateur en lui retirant ses autorisations initiales d'accès.

Donc, la migration des ED 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 al., 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'ED 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 **(Moussa, et al., 2013).** Dans la section suivante, nous présentons l'architecture de notre proposition qui pallie à ces limites.

3 Architecture globale proposée

D'après la synthèse des travaux réalisés qui traitent la confidentialité des ED, nous avons constaté que l'implémentation et l'administration des permissions en utilisant les modèles MAC, DAC et RBAC d'une façon manuelle est difficile et insuffisante, ce qui a motivé la création du modèle de contrôle d'accès dynamique à base de profil utilisateur. Ce modèle se compose de parties montrées dans notre architecture qui sont :

- **Interface Propriétaires de données** : l'interface administrateur permettant au propriétaire de données, d'accéder à la couche contrôle d'accès avec ses différents modules.
- **ED dans le CC** : présente la partie CC qui contient les données multidimensionnelles hébergées.
- **Couche contrôle d'accès à base de profils utilisateur** : Cette couche contient les trois modules proposés afin de contrôler l'accès à l'ED hébergés dans le CC, qui sont (Figure 1:
  - La classification dynamique des niveaux de sensibilité basée sur les profils utilisateur (1).
  - La détection des inférences par la combinaison de plusieurs profils (2).
  - La gestion des profils à partir des usages (3).

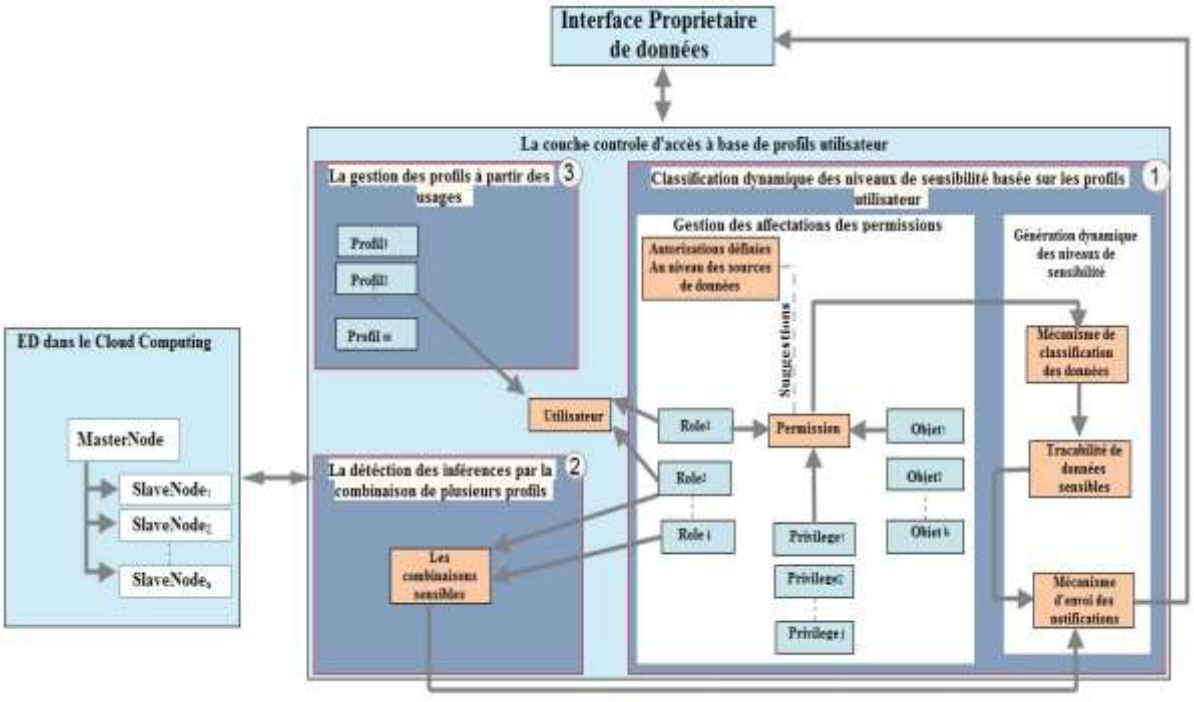

Figure 1 Architecture globale proposée

3.1 Description des modules de l'architecture proposée

Selon les travaux étudiés, et en se basant sur leur synthèse nous proposons deux contribution présentés dans les deux parties :

- La confidentialité des ED : consiste à définir les permissions ou les droits d'accès de chaque utilisateur selon son rôle, à générer le niveau de sensibilité des données, et à détecter les inférences en utilisant des combinaisons des données autorisées.
- L'implémentation de notre approche dans un environnement CC : consiste à gérer l'accès à l'DW hébergé dans le CC selon le profil usage de l'utilisateur.

3.1.1 La classification dynamique des niveaux de sensibilité basée sur les profils utilisateur

Ce module permet de générer automatiquement le niveau de sensibilité de chaque objet de l'ED à base des profils utilisateurs. Il se compose de deux phases (**El ouazzani, et al.** D**écembre 2016**) (**El Ouazzani.et al. 2018**) :

Phase 1 : Il s'agit de définir les permissions d'un rôle sur une donnée avec un privilège précisé, en prenant compte les autorisations définies au niveau des sources de données. Il s'agit de suggérer ces derniers à l'administrateur propriétaire des données de l'ED, lors de l'affectation des permissions.

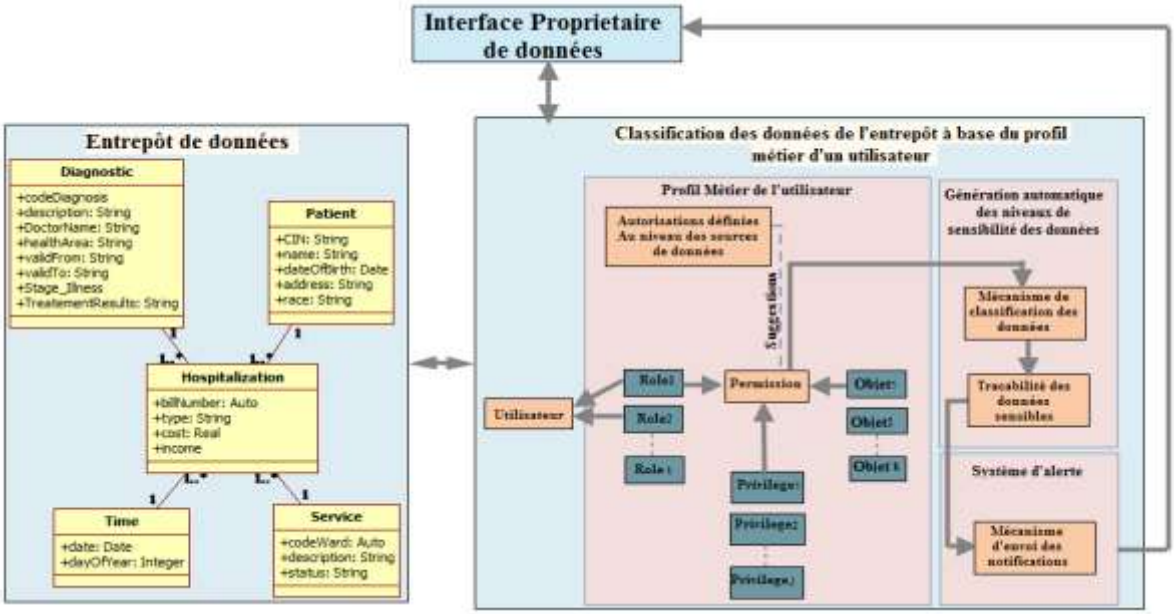

Figure 2 Classification des données de l'entrepôt à base de profil métier d'un utilisateur

Phase 2: génération automatiquement les niveaux de sensibilité des données de l'ED en se basant sur les permissions définies dans la phase 1. Ensuite le mécanisme **de traçabilité des données sensibles** permet de tracer les actions des utilisateurs sur les données qui ont un niveau de sensibilité élevé. Selon le niveau de sécurité détecté par le mécanisme de génération automatique des niveaux de sensibilité des données, notre système d'alerte permet d'envoyer des notifications à l'administrateur lors d'une tentative de violation d'une permission sur une donnée sensible.

3.1.2 La détection des inférences par la combinaison de plusieurs profils

Afin de permettre une extraction automatique des inférences à partir des permissions autorisées, nous proposons un modèle informatique visuel avec des règles à vérifier en se basant sur la présentation graphique des profils accordés à un utilisateur et les liens entre les données en utilisant le diagramme de classe source. Ce module est la suite du travail de (Triki, et al., 2013) qui propose une méthode de détection des inférences précises et partielles, cependant notre proposition consiste à détecter les combinaisons sensibles.

Sachant qu'un utilisateur peut avoir un ou plusieurs rôles au sein de l'entreprise, ce dernier accède à l'ED avec un ou plusieurs profils. Le but de notre système de détection des inférences est de détecter si l'utilisateur peut déduire indirectement des informations non autorisées en utilisant deux ou plusieurs permissions depuis un ou plusieurs profils différents.

Le module 2 de notre architecture globale que nous allons détailler dans cette partie, se focalise sur la détection visuel des inférences par la combinaison de plusieurs permissions par un même utilisateur. Afin d'envoyer les combinaisons sensibles détectées au propriétaire de données.

La figure 3 présente l'architecture détaillée de notre module de détection des inférences, qui permet d'analyser les permissions de chaque profil, afin de détecter les combinaisons sensibles, il se base sur deux entrées qui sont (**El ouazzani, et al. 2017**)

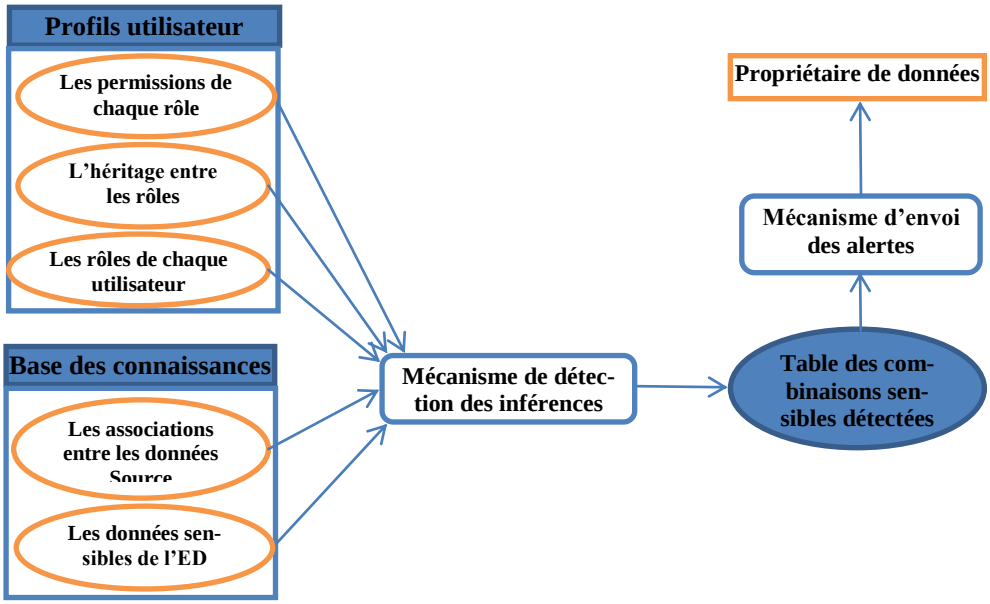

Figure 3 la détection des inférences par la combinaison de plusieurs profils

- **Les profils utilisateur :** présentent les utilisateurs qui endossent des rôles qui leurs sont attribués par le propriétaire de données. Ces rôles sont organisés en hiérarchie, et ils disposent des permissions qui ne peuvent pas être accordées directement aux utilisateurs.
- **La base des connaissances :** ce sont les données sensibles de l'ED à protéger contre les inférences, et les associations entre les données selon le diagramme de classe de la base de données source.

.

3.1.3 La gestion des profils à partir des usages

Notre but dans cette partie est de proposer une politique de contrôle d'accès performante à base des profils usage des utilisateurs PACUP (Performing Access Control based on Use Profiles) qui présente le module 3 de notre contribution. Cette politique économique combine entre le profil métier et le profil usage d'un utilisateur. Sachant que le profil métier était l'objet du module 1 (El Ouazzani, et al., 2016) de notre contribution qui définit le rôle et les permissions d'accès autorisées à un utilisateur et qui classifie les données de l'ED selon leur niveau de sensibilité. Donc PACUP permet de sécuriser l'accès avec le profil métier de l'utilisateur, et minimiser en même temps le trafic et l'échange de données entre le CC et l'organisation, en affectant un profil usage pour chaque utilisateur.

Nous cherchons à améliorer le premier scénario de (Al-Aqrabi, et al., 2015) qui se comporte d'une manière hautement sécurisé et qui garantit la facilité de gestion de l'application de la confidentialité. Nous proposons un nouveau modèle (figure 4) dont la politique de la gestion des profils utilisateurs PACUP est centralisé ce qui dit la facilité de la gestion, et en même temps cette politique optimise le trafic entre le CC et l'organisation. Cette solution minimise la charge de traitement et le temps de réponse par l'ajout d'un profil usage pour chaque utilisateur, qui gère l'utilisation et l'accès optimal à l'ED. Notre modèle comprend deux parties :

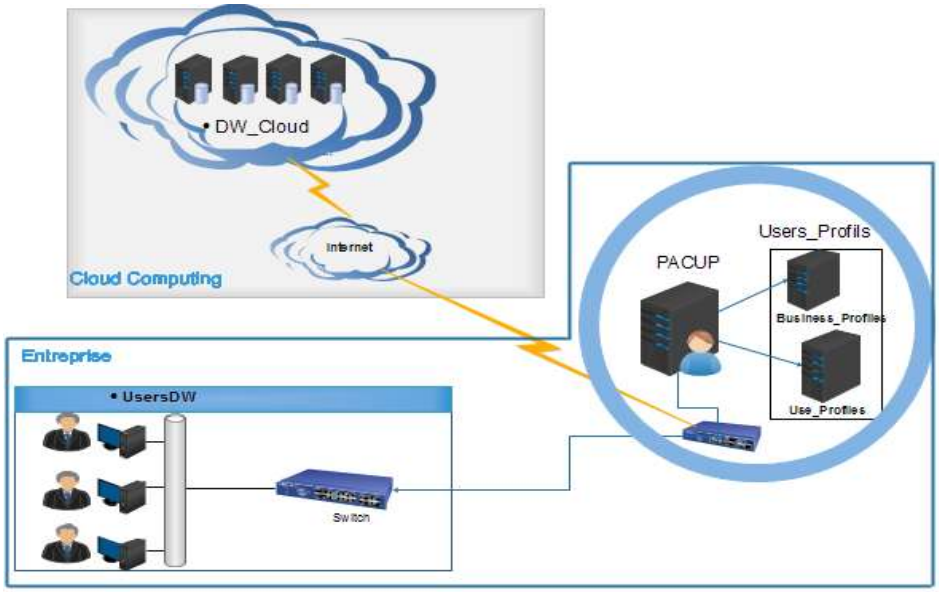

Figure 4 Architecture globale de PACUP

- **DW_Cloud :** présente les données multidimensionnelles hébergées dans le CC.
- **Entreprise :**
  - **usersDW** : c'est le réseau local de l'entreprise regroupant les utilisateurs.
  - **PACUP** : une politique qui permet de gérer l'accès à l'ED hébergé dans le CC en combinant entre le profil usage et les profils métiers d'un utilisateur. Elle permet également de gérer le cache. Donc les clients ont besoin de vérifier eux-mêmes au cours de l'accès aux services de CC en analysant les requêtes demandée selon :
    - **Profil Métier de l'utilisateur :** Il s'agit de définir les permissions d'un utilisateur selon son rôle sur une donnée avec un privilège précis, en prenant en compte les autorisations définies au niveau des sources de données (El Ouazzani, et al., 2016).
    - **Profil usage de l'utilisateur :** c'est un profil regroupant les droits d'usage de chaque utilisateur. Nous détaillons les composants de ce profil dans la partie suivante.

L'idée principale de l'approche est de créer pour chaque utilisateur un profil usage selon son rôle, et un cache partagé entre l'ensemble des utilisateurs. Ce cache se base sur les préférences et les usages précédents ou historiques des accès des profils. L'objectif est de faire face aux problèmes de performance d'accès à un ED dans le CC en réduisant le temps d'accès et en minimisant le trafic avec le CC.

Un ED de données hébergé dans le CC serve un grand nombre des utilisateurs. Le système de cache des requêtes dans les bases de données est un axe de recherche prometteur, en particulier dans les systèmes OLAP où l'utilisateur navigue interactivement dans un cube en lançant une séquence de requêtes. Les utilisateurs peuvent avoir des résultats qui ne les satisfont pas à cause du temps de réponse.

En outre, nous sommes devant la problématique de trouver les requêtes à garder en cache afin d'améliorer la performance de notre modèle de contrôle d'accès à l'ED hébergé dans le CC et optimiser le temps de réponse.

4 Implémentation

4.1 Outils et environnement de développement

Pour la mise en œuvre de notre contribution et la réalisation des expérimentations, nous avons utilisé une machine DELL PRECISION T1700 sous Windows 7 professionnel 64 bits. cette machine est dotée d'un processeur Intel Core i7-4770 de 3.40 Ghz et 8 Go de RAM. Nous avons utilisé les outils logiciels suivants :

- Le langage Java avec l'environnement de développement Eclipse. Ce choix a été motivé par les avantages qu'offre ce langage en termes de portabilité, de robustesse ainsi que la disponibilité de nombreuses bibliothèques ;
- Le SGBD Oracle version 12c pour concevoir l'entrepôt de données de tests. Ce choix a été principalement argumenté par la disponibilité de l'option OLAP offrant en particulier un moteur analytique OLAP, des espaces de travail et un gestionnaire d'espace de travail analytique (Analytic Workspace Manager-AWM) ;

- Le SGBD relationnelles MySql pour concevoir notre méta-modèle (Figure 7). Ce choix a été principalement argumenté par sa simplicité d'utilisation et ses interfaces pour effectuer diverses opérations.

4.2 Interface d'affectation des permissions

L'affectation des permissions est une tâche que seul l'administrateur s'en charge. Dans cette interface on choisit le rôle qu'on va affecter à un utilisateur. Le programme nous affiche les informations concernant cet utilisateur ainsi que ses objets autorisés au niveau des bases de données source, ce qu'on a déjà traité dans (El Ouazzani, et al., 2016), afin d'aider l'administrateur à bien définir les permissions. Cette interface permet de préciser le type d'objet (table/colonne/valeur), et les privilèges à autoriser.

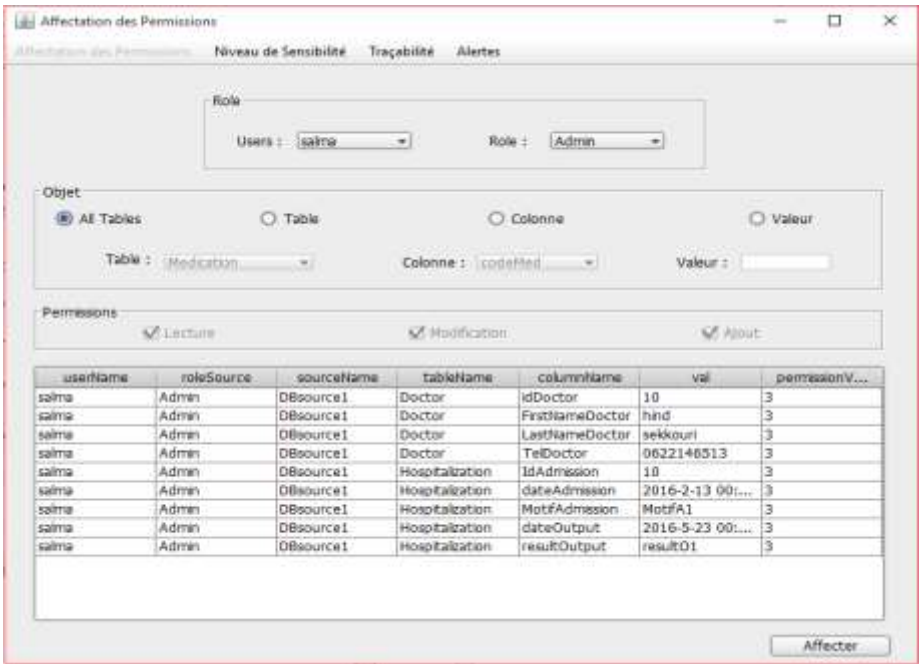

Figure 5 Interface d'affectation des permissions

4.3 Interface de détection des inférences

Cette interface permet d'afficher les profils affectés à l'utilisateur choisi. Ensuite l'administrateur peut vérifier l'existence d'une inférence selon les 5 règles proposées dans cet article. (**El Ouazzani.et al. 2018**)

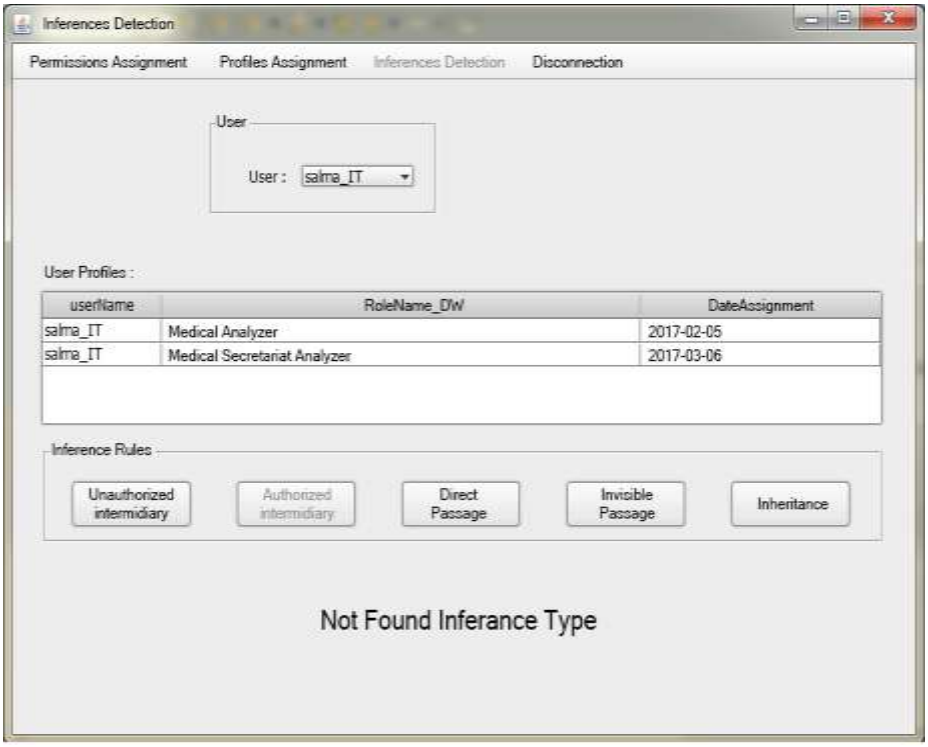

Figure 6 Interface de détection des inférences «Intermédiaire autorisé»

Dans cette capture d'écran (Figure 7), notre programme a détecté une inférence de type « Passage direct » entre les deux rôles « Analyseur Médical » et « Analyseur Secrétariat Médicale ». (**El Ouazzani.et al. 2017**)

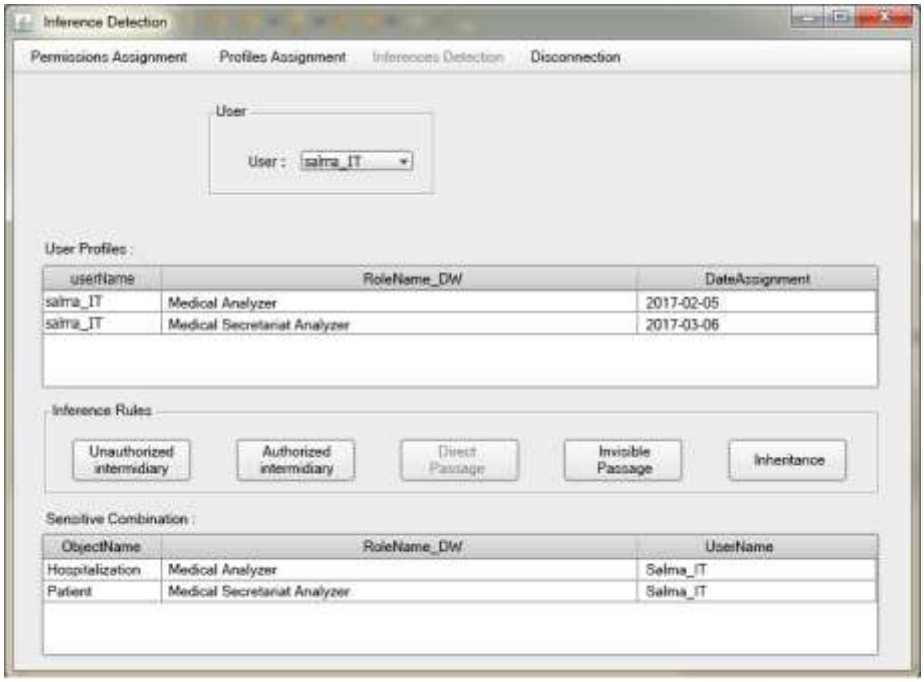

Figure 7 Interface de détection des inférences « Passage direct»

5 Conclusion

Nos travaux de recherche se situent dans le cadre de la confidentialité des ED hébergés dans le CC à base de profil utilisateur. Nous avons décomposé notre sujet en deux parties. La première partie consiste à garantir la confidentialité de l'ED contre les accès illégaux et les inférences, en se basant sur le profil métier de l'utilisateur. La deuxième partie consiste à adapter la première partie de notre contribution dont l'objectif est de contrôler l'accès d'une façon performante à un ED hébergé dans le CC. Nos propositions s'articulent autour de trois modules :

- Une méthode de classification des données de l'ED selon leur niveau de sensibilité.
- Une méthode de détection des inférences entres les profils utilisateurs.
- Une méthode de contrôle d'accès aux ED hébergés dans le CC à base des profils usage des utilisateurs.

Références

Accorsi, R. and Müller, G. 2013. Preventive inference control in data-centric business models. s.l. : Security and Privacy Workshops (SPW), 2013 IEEE (pp. 28-33). IEEE. , 2013

Al-Aqrabi, H., et al. 2013. Business intelligence security on the clouds: challenges, solutions and future directions. s.l. : Service Oriented System Engineering (SOSE) IEEE 7th International Symposium on (pp. 137-144). I, 2013.

Al-Aqrabi, H., et al. 2015. Business intelligence security on the clouds: challenges, solutions and future directions. . s.l. : Service Oriented System Engineering (SOSE), IEEE 7th International Symposium on (pp. 137-144). I, 2015.

Arora, D. and Kumar, U. 2016. Protecting Sensitive Warehouse Data through UML based Modeling. s.l. : Proceedings of the International Conference on Informatics and Analytics (p. 31). ACM., 2016.

Bensaidi, M., Aboulkalam, A. and Marzouk, A. 2012. Politique de contrôle d'accès au cloudcomputing: Recommandation à base de confiance. s.l. : Network Security and Systems (JNS2), National Days of (pp. 90-96). IEEE., 2012.

Blanco, C., et al. 2010. Towards the secure modelling of olap users behaviour. 2010.

Blanco, C., et al. 2015. An architecture for automatically developing secure OLAP applications from models. . s.l. : Information and Software Technology, 59, 1-16, 2015.

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

El Ouazzani, A., Harbi. N., Badir. H. 2015. Confidentialité des entrepôts de données dans le Cloud Computing: Etat de l'art et Perspectives.. 9ème édition de la conférence sur les Avancées des Systèmes Décisionnels. ASD, 2015.

El Ouazzani, A., Rhazlane, S., Harbi. N., Badir. H. 2016 . Dynamic management of data warehouse security levels based on user profiles. s.l. : Information Science and Technology (CiSt). 4th IEEE International Colloquium on (pp. 59-64). IEEE., 2016 .

El Ouazzani, A., Harbi. N., Badir. H. December 2016 . DYNAMIC CLASSIFICATION OF SENSITIVITY LEVELS OF DATA WAREHOUSE BASED ON USER PROFILES. s.l. : International Journal of Database Management Systems (IJDMS) on Volume 8, Number 6, 2016 .

El Ouazzani, A., Harbi. N., Badir. H. 2017. LA DETECTION DES INFERENCES PAR LA COMBINAISON DE PLUSIEURS PROFILS.. INTIS, 2017.

EL OUAZZANI, A., Harbi, N., & Badir, H. (2018). User Profile Management to protect sensitive data in Warehouses. *INTERNATIONAL JOURNAL OF NEXT-GENERATION COMPUTING*, *9*(1) 2018.

Fernandez-Medina, E., et al. 2006. Access control and audit model for the multidimensional modeling of dws. . s.l. : Decision Support Systems, 1270–1289., 2006.

Fernández-Medina, E., et al. 2007. Developing secure data warehouses with a UML extension. s.l. : Information Systems, 32(6), 826-856., 2007.

Inmon. 1991 . Building the data warehouse. 1991 .

Kechar, M. and &Bahloul, S. N. (2015, November). An Access Control System Architecture for XML Data Warehouse Using XACML. s.l. : Proceedings of the International Conference on Intelligent Information Processing, Security and Advanced Communication (p. 15). ACM., (2015, November).

Moussa, R and Badir, H. 2013. Data Warehouse Systems in the Cloud: Rise to the Benchmarking Challenge. s.l. : IJ Comput. Appl 20(4), 245-254, 2013. Vols. 20(4), 245-254.

Naushahi, U. M. A. 2016. Profile-Based Access Control in Cloud Computing Environments with applications in Health Care Systems. 2016.

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

Ray, I. and Ray, I. 2014. Trust-based access control for secure cloud computing. . s.l. : High PerformanceCloud Auditing and Applications (pp. 189-213). Springer New York, 2014.

Rodriguez, A., et al. 2011. Secure business process model specification through a uml 2.0 activity diagram profile. 2011.

Rosenthal, A. et S. Sciore. 2000. . View security as the basis for data warehouse security. . s.l. : In DMDW (p. 8)., 2000.

Saltor, F, et al. 2002. Building secure data warehouse schemasfrom federated information systems. 2002.

Soler, E., Stefanov, V. and Mazon, N.J. 2008. Towards Comprehensive Requirement Analysis for Data Warehouses: Considering Security Requirements, pp. 104–111. IEEE,. Los Alamitos : s.n., 2008.

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

Thangaraju, G. and Rani, X. A. K. 2016. Multi User Profile Orient Access Control based Integrity Management for Security Management in Data Warehouse. . s.l. : Indian Journal of Science and Technology, 9(22)., 2016.

Triki, S., et al. 2013. Sécurisation des entrepôts de données: de la conception à l'exploitation. Lumiere II Lyon : Rapport de thèse., 2013.

Trujillo, J., et al. 2009. A UML 2.0 profile to define security requirements for Data Warehouses. . s.l. : Computer Standards & Interfaces, 31(5), 969-983., 2009.

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

Summary

A data warehouse (DW) is a critical business factor that gives a clear view of its operations and a rich source for decision-makers. It contains sensitive data about the company and its customers, and therefore they must not be accessible without access control. The hosting solution of ED in the CC (Cloud Computing) is gradually gaining more popularity in companies because it helps to overcome the endless expansion of data and benefit from its ability to process and store these data. However the confidentiality of these EDs in the CC needs many improvements and the setting up of precise standards, due to the scalability and elasticity of the CC paradigm, since there is not a standard protocol to handle connectivity of CC users to hosted resources by taking into account query execution performance. The objective of our work is to propose a framework guaranteeing the confidentiality of the EDs hosted in the CC based on user profile.