MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE DJILLALI LIABES SIDI BEL ABBES

FACULTE DES SCIENCES DE L'INGENIEUR

RESUME DE MEMOIRE DE MAGISTER

Nom & Prénom(s)	CHAIEB yazid
E-mail (obligatoire)	chaiebyazid@yahoo.fr
Spécialité	Informatique.
Titre	Fouille de Données et Confidentialité des Données.
Date de soutenance	10 Mars 2010 (10 /03/10).
Nom, prénom(s) et grade de l'encadreur	Dr. MALKI Mimoun ; MC(A) ;Docteur D'Etat.

Résumé :
Récemment, l'habilité croissante en matière de sauvegarde des données ainsi que les avancées récentes dans le domaine de la fouille de données ont abouti à un souci croissant concernant la confidentialité de ces données.
Sachant que le sujet de la confidentialité a été étudié traditionnellement dans le cadre de la cryptographie et tout ce qui concerne la dissimulation de l'information, le progrès récent dans la fouille de données a conduit à un intérêt renaissant dans ce champ.
La fouille de données(Data Mining) a été vue comme une menace pour la confidentialité en raison de la prolifération des données électroniques, dans la plupart de temps sous le contrôle de sociétés, gouvernements ou de grands organismes.

Ceci a mène à un intérêt croissant porté au sujet de la confidentialité de telles données. Cependant, dans le cas d'ensembles de données se trouvant dans un milieu multiutilisateurs (multi-clients), les soucis de sécurité des données et ceux concernant leurs confidentialités peuvent empêcher les parties de partager leurs données de façon directe.

Des méthodes dites classiques ont été étudiées de manière intense, visant à garantir la confidentialité des données d'une base de données (ou un ensemble de BDD) se basant sur le principe de la séparation entre données confidentielles et celles ordinaires allant même jusqu'à modifier ou transformer les données de la base elle-même de manière à préserver leurs confidentialité. Cependant, ce type d'approche reste du point de vue domaine d'application très loin de la réalité surtout dans des environnements distribuées où les sources de données peuvent être hétérogènes et éparpillées comme c'est le cas dans les grilles de calcul.

L'objectif majeur de notre travail consiste à considérer le cas général où on n'a au préalable aucune information sur la base et de garantir par la suite qu'appliquer les techniques de fouille de données entre les nœuds communicants de la grille en vue d'arriver au résultat final préserve ces données en dépit de leurs type (confidentielles/non confidentielles), et en dépit aussi du type de nœuds de la grille(nœuds communicants, nœuds d'interconnexion).Donc le mécanisme de communication entre les nœuds doit garantir que leurs collaboration visant µa obtenir des résultats globaux et nouveaux, ne permette en aucun cas à une partie, partant des données qu'elle reçoit, ou celles qui y transitent, de pouvoir déduire des informations lors du transfert ou de la réception d'ensembles de données.

Mots clés :
confidentialité, information personnelle, fouille de données, préservation de la confidentialité, grilles de calcul, produit scalaire, crypto-système de paillier, motifs séquentiels.

Abstract

Recently, the increasing ability to store data and the recent advances in the data mining field have lead to increased concerns about privacy.

While the topic of privacy has been traditionally studied in the context of cryptography and data hiding, recent emphasis on data mining has lead to renewed interest

RESUME DE MEMOIRE DE MAGISTER

in the field.

Data mining has been viewed as a threat to privacy because of the widespread proliferation of electronic data maintained in most of time by corporations, governments or large organizations. This has lead to increased interest about the privacy of such data. However, in multi-client's data set, the privacy and security concerns may prevent the parties from directly sharing their data.

Traditional methods were studied, with principle goal to guarantee the data privacy of a single database (or a set of Databases) based on the separation between confidential data and those ordinary. Most methods modify or transform the data inside the database itself in order to preserve the data privacy.

However, this type of approach doesn't satisfy all privacy application domains, thus it remains far from the database reality especially in distributed environments where the data sources may be heterogeneous or distributed or both, as it could occur in grid computing.

Our work aims to consider the natural case where one has no information about the database, then be able to guarantee that after applying the data mining techniques between the communicating nodes of a computational grid, we preserve the whole database data privacy whatever their type is (private data/ordinary data), for all types of nodes of a grid (communication nodes, interconnection nodes). Thus the mechanism of communication between the nodes must guarantee that their collaboration aiming to obtain global and new results, does not allow, to any node, at any moment, to deduce other's nodes partial or total information during the data sets transfer.

Keywords :
privacy, personal information, privacy preserving data mining, grid computing, private scalar product, paillier encryption-system, sequential pattern.

ملخص

كلمات مفتاحيه