

Sécurisation des entrepôts de données contre les inférences en utilisant les réseaux Bayésiens

Salah TRIKI*, Hanene BEN-ABDALLAH*, Jamel FEKI*, Nouria HARBI**

* Laboratoire Mir@cl
Département d'Informatique,
Faculté des Sciences Economiques et de Gestion de Sfax,
Route de l'Aéroport Km 4 – 3018 Sfax, BP. 1088
{Salah.Triki, Hanene.BenAbdallah, Jamel.Feki}@Fsegs.rnu.tn
**Laboratoire ERIC
Université Lyon 2,
5 avenue P. Mendès France 69676 Bron, Cedex
Nouria.Harbi@univ-lyon2.fr

Résumé. Les entrepôts de données permettent aux analyste-décideurs d'établir des prévisions et de prendre des décisions stratégiques. La sécurisation de ces entrepôts est, par conséquent, importante afin de protéger les informations sensibles. Par ailleurs, cette sécurisation ne doit pas constituer une entrave à l'exploitation efficace et rapide de l'entrepôt, ni être trop souple induisant l'inférence des données interdites (i.e., données personnelles, confidentielles). Dans cet article, nous examinons la sécurisation des entrepôts de données à travers une approche basée sur les réseaux Bayésiens. Elle comporte deux avantages : d'une part, elle ne nécessite pas un traitement supplémentaire après chaque phase d'alimentation de l'entrepôt et, d'autre part, elle n'entraîne pas l'altération des données originales.

1 Introduction

Les entrepôts de données (ED) occupent une place importante dans les organisations. Ils permettent aux décideurs d'explorer les données de celles-ci et de prendre des décisions stratégiques. Les données sont agrégées selon des dimensions formant ainsi un cube. L'agrégation permet de ressortir les corrélations entre les données et de définir les tendances. L'utilisation des opérations de forage vers l'avant rend possible l'exploration de ces corrélations et de ces tendances, offrant de l'information qui aide à la prise de décision.

Evidemment, les données de l'ED d'une organisation sont sensibles et ne doivent pas être cédées sans contrôles. Dans ce cadre, plusieurs gouvernements ont promulgué des lois pour la protection des vies privées de leurs citoyens. Parmi ces lois, HIPPA (« 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 de sécurités strictes ; de même, GLBA (« Gramm-Leach-Bliley Act » GPO