Dynamic management of data warehouse security levels based on user profiles

Amina El Ouazzani, Sara Rhazlane, Nouria Harbi, Hassan Badir

Dynamic management of data warehouse security levels based on user profiles

Amina El Ouazzani[1], Sara Rhazlane[2], Nouria Harbi[3], Hassan Badir[4]
[1 2 4]LabTIC Laboratory, ENSA Tangier, Morocco
[3] ERIC Laboratory, Lyon II University, France

Abstract– **Respect for privacy and data confidentiality in a company are two fundamentals that must be protected. However, a Data Warehouse can be used as a very powerful mechanism for discovering crucial information, hence the importance of implementing security measures which guarantee the data confidentiality by establishing an access control policy. In this direction, several propositions were made, however, none is considered as a standard of access management to data warehouses. In this article, we will present our approach that allows first to exploit the permissions defined in the data sources in order to help the administrator to define access permissions to the data warehouse, and then our system will automatically generate the sensitivity level of each data warehouse element according to the permissions granted to an object in the data warehouse.**

Index Terms—Data Warehouse, Confidentiality, Profession profile, Crucial information, Traceability.

I. INTRODUCTION

A Data Warehouse (DW) [13] is defined as a complex system that meets the information needs for the strategic decisions making of an organization, this system consists of the extraction and integration of data from multiple heterogeneous sources [7] [12], its data is organized in a multidimensional model, which presents a modeling approach to store data in facts and dimensions. The facts are the topics of interest to the organization and the dimensions represent the context in which the facts are analyzed.

Because of the critical information stored in the DW, it is important to check its privacy. According to [6], confidentiality in the context of DW is considered an important requirement, which must be ensured by an authorization management mechanism which is the specification and execution of access rights in the databases in general and more specifically in data warehouses.

In general, conventional security requirements are summarized by the acronym CIA (confidentiality, integrity and avail-ability). All other security requirements such as authentication, authorization, access control, etc., can be assigned to these three basic properties. Confidentiality is defined as the absence of disclosure of unauthorized information. Integrity is defined as the absence of the unauthorized modification of information, and availability that ensures the continuity of service [8].

Our goal is to ensure the confidentiality of the DW by access control, since there is no standard that manages this important aspect, and mechanisms of confidentiality specified for OLTP systems cannot be used for the DW because in operational systems, access control is defined on the tables, rows, columns, etc. While in an DW, we have a large number of users with different analysis needs, seeking access to the multidimensional DW [10,11].

In this paper, we focus on the conceptual modeling process of the data warehouse by offering an approach providing access control, based on the job profile of a user who de-scribes its access rights. We use the RBAC access control policy (Role-based Access Control) that focuses on grouping users according to their professions or their roles. With this approach, we are able to classify the data warehouse automatically by generating their sensitivity levels, in order to trace user actions on sensitive data.

II. STATE OF THE ART AND SYNTHESIS

Recently, a number of DW security models have been proposed. In this section we will organize these research works according to two approaches, the integration of security in the modeling process of the DW, and the DW access control models already in place.

A. Security at the conceptual level of the DW

1) The data warehouses security based on permissions defined at the sources

[19] Rosenthal and Sciore (2000) propose a theoretical approach that exploits the access permissions defined in the data source, rather than creating new access mechanisms. They use the rewrite queries to verify that they comply with the restrictions defined in the data sources, as well as creating relational views in order to minimize the risk to infer sensitive data.

[20] Saltor and al. (2002) since there are similarities between the architecture of a federated database and the architecture of a data warehouse, the authors proposed the use of the multi-level access permissions pattern defined for the federated database without being modified to construct a secure data warehouse, this authorization scheme describes the multi-level access rules.

2) Confidentiality of the data in a data warehouse during the modelization

Among the works that have been developed on integrating security into the modelization of warehouses, we find:

[10] Fernandez-Medina and al. (2006) have developed an access control and auditing model (ACA) specific for data warehouses, based on two access management policies: MAC and RBAC. They specify the security rules during the mod-elization process of a conceptual model, by incorporating the concept of ”user profile”, which consists of an isolated table containing all user information (identity, classification level: top secret, secret, confidential or

978-1-5090-0751-6/16/$31.00 ©2016 IEEE

unknown). This model remains a purely theoretical model since no solution for its implementation has been proposed.

[22] Villarroel and al. (2006) have defined an OCL extension Object Constraint Language using UML2.0 extension mechanisms to resolve issues of confidentiality; this extension specifies the security requirements of the elements during the conceptual modeling of data warehouses.

[21] Soler and al. (2008) have used extension mechanisms provided by the CWM (Common Warehouse Metamodel) to extend the relational package and build a star schema, which represents security and verification rules captured during the conceptual phase of the data warehouse.

[11] Trujillo and al. (2009) have developed a methodology consisting of four phases: analysis, modeling, implementation and validation, which covers the five levels of abstraction: requirements analysis, conceptual level, logic level, the physical level and the post-development review, the latter being a new discipline introduced by Lujan and Trujillo (2004). This methodology offers all the security requirements throughout the life cycle of the data warehouse.

[23] Rodriguez and al. (2011) presented an UML 2.0 extension of the activity diagram. This proposition, called BPSec (Business Security Process), allows to define a set of security requirements (access control, detection of attack risks, non-repudiation, integrity, confidentiality and security audit), which improves the expressiveness of the business processes models, and enables to secure a data warehouse during its development, taking into account this requirement.

[14] Blanco and al. (2015) have developed an automatic MDA architecture to secure a data warehouse; this architecture is composed of a logic model and its transformations from the conceptual model using the UML extension and the CWM package. They defined these constraints in the metadata layer that connects the data warehouse with the OLAP tools. This proposition consists of models and transformations.

3) Inferences Management

[1] Triki and al. (2013) proposed a model to secure multidimensional data against the inferences in the conceptual phase, this approach assumes that the DW scheme is already designed. It allows to detect both types of inferences:

- Precise Inference: where the values of the derived data are accurate.
- Partial Inference: where the data values are partially disclosed, meaning that the user can have an idea of the value of data. This approach consists of three steps:
  - Step 1. A domain expert identifies the sensitive elements to protect by interrogating the data warehouse designer.
  - Step 2. Build the graph inferences from the class diagram, specifying the elements that are specific or partial inferences.
  - Step 3. Present the data warehouse with UML annotations highlighting both types of inferences.

[4] Blanco and al. (2010) proposed an approach based on the state diagram to detect inferences on the design level. This proposition focuses on sensitive requests and its evolutions, but they do not take into account to infer the data from the accessible data. The approach is presented as an OLAP security model of 3 states:

- Static model: presents the Fernandez Medina 2007 UML profile specific to data warehouses, adding a new kind of rule named Joint Rules, which presents the necessary privileges to certain combinations according to a specific grammar.
- Dynamic model: or the transactions-state model that has the objective of enhancing the static model, in order to ensure that confidentiality is not compromised by treating the evolutions of the combinations defined with JR through the application of OLAP operations.
- Session Control: This step makes it of interest to the user sessions to analyze them by checking each event in order to detect any possible inference.

[24] Sweeney (2002) Describes a real data inference case, using a demonstration of the identification of sensitive data based on data crossing of an insurance group, those data supposed anonymous, as well as a voter registration list, which allowed the detection of the name of the former governor William Weld and his medical records, by linking shared attributes. He defined a protection model called k-anonymity including support policies, that allows avoiding inferences, which are:

- Sorting rows of the tables to be published in a random manner in order to not disclose sensitive information.
- Avoid having a row with a unique value in the table to be published.
- Take into account the old version of the data during the construction of the new table.

B. The security in the operating level of the DW

Online Analytical Processing (OLAP) has become increasingly an important component in decision support systems. The OLAP server is supposed to provide access based on authorizations defined for each user. He may refuse the access to data of measure, a dimension, and/ or beyond a level in a hierarchy. Access rights can be explicitly specified on tables / columns of the tables of the data warehouse. However, the OLAP server alone cannot protect access to prohibited data. Research works has been done to strengthen the access rights and authorizations of users, and to prevent any malicious user to infer prohibited data from data which he has access to.

[9] Kirkgoze and al. (1997) defined a secure model for data warehouses which consists in the elaboration of a custom cube with its own dimensions and hierarchies. This model is based on the AMAC management policy. It is an extension of the MAC model that specifies the tasks that the user can perform according to its role within the organization. The advantage of this model is the flexibility in assigning roles to different virtual cubes.

[15] Priebe and Pernul (2000) explored the security problems in the Goal Project, a project that aims to study the integration of a distributed information system. During this research, they have developed a proper method to the OLAP world. This method follows the traditional methodology in the elaboration of databases (analysis of needs, conceptual model, logical and physical), while

incorporating the multi-dimensional aspect during the conceptual phase.

[16] Priebe and Pernul (2001) continue their research on the creation of access control mechanisms to ensure data confidentiality, and they have created an access control mecha-nism in the form of a language expressing the security-related constraints, during the conceptual phase. This is a language based on MDX MulDimensionnale Xpression. [17] Eavis and Althamimi (2012) presented an authentication framework build on algebra especially designed for OLAP. It is object-oriented and uses query rewrite rules to ensure access to consistent data across all levels of the conceptual model. The process is essentially transparent to the user, a notification is provided in the case where a subset of the original request is returned. The final result is an intuitive and powerful approach for the database authentication that is uniquely adapted to the OLAP field.

[1] Triki and al. (2013) proposed an approach that does not require additional processing after each alimentation phase of the DW. It is based on Bayesian networks in order to protect a data warehouse against inferences; they use a control module, which seeks to prohibit a user to infer protected data from the accessible data by using Min and Max aggregations functions.

C. Comparison and Synthesis of existing work

We presented in the previous section, research works that propose solving problems related to the confidentiality of access to the data warehouse. Some approaches seem to be relevant and provide an acceptable level of privacy but still insufficient. In this section, we present a comparative table of these works based on criteria, and followed by a synthesis.

1) Comparison

The evaluation of some important approaches that address the confidentiality of DW for the development of a secured multidimensional model, is presented in Table 1 according to several criteria are:

- Used approach: the name of the proposed approach.
- Used technique: the technologies used in the proposed approach.
- Transformation between models: development of a logic model of ED from a conceptual model taking account security in every model.
- Inferences management: detection of feasible conclusions from the accessible data.
- Validation: implementation of the proposed solution.
- Data dynamic sensitivity: automatic determination of data sensitivity level.
- Traceability of access: to trace the user transactions on the DW for analysis and decision making.

2) Synthesis

DW protection against illegal access was felt beyond reasonable doubt since several years [10] [16] [17]. According to the authors [25], modeling the access control of the DW is the process of building an abstract model that needs to be stored in the DW. This model is a representation of the reality or a part of the reality. Following the study of existing works, we found the following:

- The confidentiality of the DW has traditionally been considered in the definitive implementation of a DW[16] [17], however the most recent works [14] [18] consider its inclusion in the development stages which can produce more robust quality solutions, and the system can accommodate these security requirements in a more natural way.
- In the research work [5], the authors proposed an automatic MDA architecture based on the UML extension to secure a data warehouse, but the question that arises in this case is: What is the use of precising the security level in the user profile as well as for the elements of the data warehouse since the permissions are specified.
- The majority of research works, especially those involved in the conceptual modeling phase relied on the CWM meta-model, in order to develop a secure data warehouse. Knowing that the CWM model is based on three standards, namely UML, MOF and XMI to properly represent all security and audit rules defined in the conceptual modeling of data warehouses.
- Most of the works are modeling the access control based on MAC and RBAC policies, while the user profile is considered an isolated table that contains the necessary data for a static access of a user without taking account the priorities of the user authenticated.
- An MDA architecture for an automatic secure conception of a DW is applied in [14, 13], but the two approaches were unable to understand the safety rules that are complex.
- Although the permissions present the main axis to en-sure confidential access to the warehouse, however, the absence of a standard that supports the accuracy of these permissions can cause inconsistencies and inferences as consequences. In this sense, we find the work of [2], which proposed the use of the permissions schema de-fined for federated databases without any modification to build a secure data warehouse, and [3] that propose rewriting requests in order to verify that they comply with the restrictions defined in the sources.
- Some authors [19] [20] proposed to make the access control model in the DW, from data sources, while others [16] [26] have considered this proposition as difficult since the source data arise from different systems (with different policies). And operational systems use the relational model while the OLAP systems use the multidimensional model.

Citation	Used approach	Used Technique	Transformation between models	Inferences management	Validation	Data Dynamic sensitivity	Traceability of access
[10] Fernandez 2006	ACA	UML 2.0 OCL	No	No	No	No	No
[22] Villaroel 2006	Extension UML 2.0	UML 2.0	OCL	No	No	No	No
[21] Soler2008	MDA	Framework MDA	No	No	No	No	No
[4] Blanco 2010	UML	Transaction State diagram	No	yes	Yes	No	No
[1] Triki 2013	Semiautomatic Approach for Inference	Graphs	No	Yes	Yes	No	No
[14] Blanco2015	MDA	MDA + UML 2.0	QVT	No	Yes	No	No

Tableau 1. Research works comparaison

- Also note that the concept of inference was cited in several works as an essential element to ensure confidentiality, and whose mastery is crucial. In this sense, there is the work of [1] which proposed an approach that can detect partial and accurate inferences, [4] propose an approach based on the state-diagram to detect inferences in the conceptual phase without considering the possibility of inferences from accessible data. Nevertheless, despite the high risk of inferences, it is not sufficiently taken into account in the conceptual phase.

We find that most of the research works affects the task of classifying data according to their level of sensitivity (very sensitive, sensitive, and confidential) to the data owner. Knowing that according to the role of the user, the data owner assigns a level of data sensitivity in order to access data having the same level of sensitivity or lower. The owner of the DW may then assign a lower level of sensitivity to a critical data. This results, however, a problem of information confidentiality loss. In addition, the permissions defined in the sources are not sufficiently exploited to help the owner well determine the permissions of a DW user. In the next section, we propose an approach that overcomes these limitations.

III. DYNAMIC MANAGEMENT OF SECURITY LEVELS

A. Motivation

Confidentiality of a DW is based on the access control model that specifies the permissions granted to each user. According to studies cited in the state of the art section, access permission to an object of the DW is granted to a user according to his role. The sensitivity level makes data accessible to the user, with a sensitivity level defined on the object of DW according to its role. Such situation is difficult to manage by the data owner, who can assign a not proper level of sensitivity to an object of DW. What may abuse its confidentiality. For this we propose a different way of defining user permissions by using the permissions defined in the data sources such as suggestions that can help the data owner. Then our access control model can generate the sensitivity level of each object in the DW based on these permissions. This classification will help us to trace user actions on sensitive data. In the remainder of this section we present the architecture of our contribution, then we detail our contribution which consists of two stages.

B. Architecture

In this section we present in details the architecture of our proposition which consists in establishing a DAC model (Dynamic Access Control) based on the PPU (Profession Profile of a User) constituted of several mechanisms (Figure 1): **1. Profession profile of the user**: This is to define the permissions of a role on a data with a given specified privilege, taking into account the permissions defined on the data sources level. And suggest them to the administrator, data owner of the ED, when as-signing permissions**. 2. Data classification mechanism:** Automatic generation of data sensitivity levels based on defined permissions. **3. Traceability of sensitive data**: Is a mechanism to trace user actions on the data with a high level of sensitivity. **4. Sending alerts Mechanism:** Depending on user access traceability mechanism, our alert system can send notifications to the administrator during an attempted violation of permission on a sensitive data.

C. Proposition

Our solution is structured following two steps:

1) Step 1: The inclusion of permissions defined at the sources:

The permission is a primordial axis in an access control mechanism, the management of these permissions is a difficult task for the administrator. In this phase, we propose to use the permissions defined in the data sources as suggestions that will help the data owner to well define the permission of a role on an object in the data warehouse according to a given privilege **P(Ri,Prj,Ok)** where :

- **P :** Permission (0,1).
- **Ri :** Role of the user.
- **Prj:**Privilege permission(Read,Write, Modification).

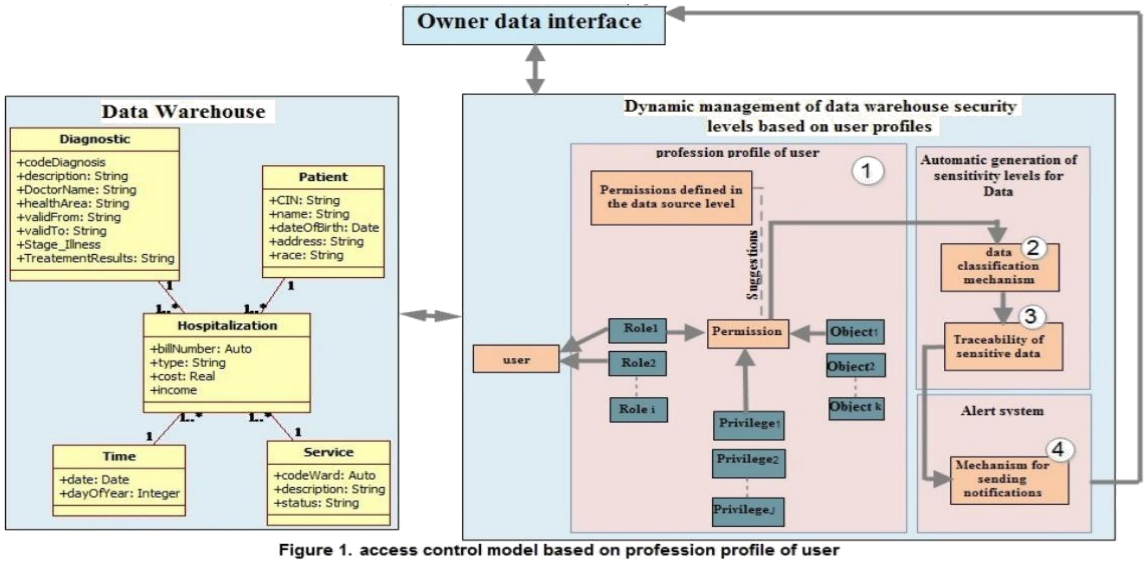


Figure 1. access control model based on profession profile of user

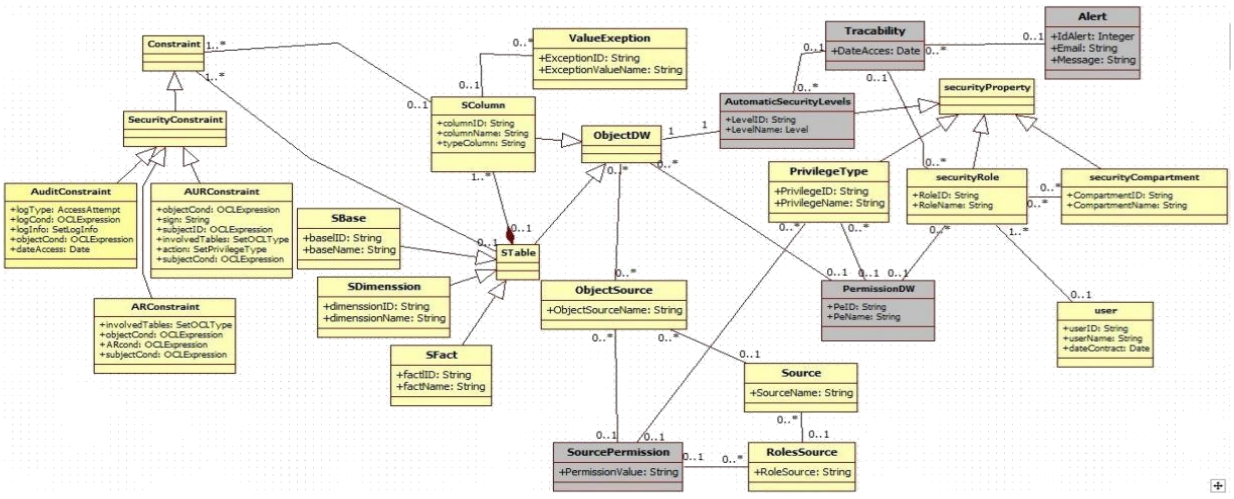


Figure 2. Meta-Model of profession profile

- Ok : Object of the data warehouse (Table fact, dimension, column, column value).

The meta-model proposed is an extension of the standard metamodel CWM (Common Warehouse Metamodel) and presents the user's profession profile. It contains five classes considered the core of our contribution:

- **PermissionDW :** present the permission of access to a role with a specific privilege on an object of the data warehouse.
- **SourcePermission :** for each object in the data warehouse, we precise the object and the source permission.
- AutomaticSecurityLevels : generation of the level of sensitivity of an object according to the defined permissions (Phase 2).
- **Traceability :** according to the sensitivity level of the object, we trace the actions made on sensitive data (Phase 2) .
- **Alert :** sending alerts allows to claim an attempt of unauthorized access to a sensitive data by a user(Phase 2).

2) *Step 2 : Amutomatic generation of data sensitivity levels*:

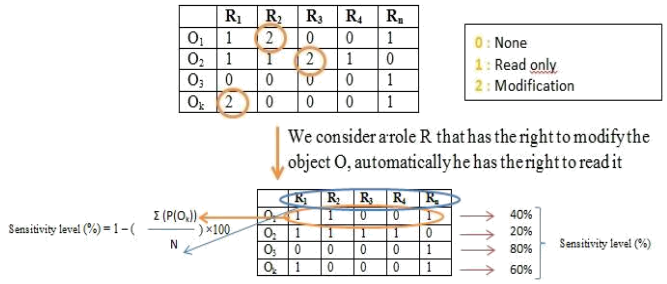
To trace user actions on sensitive data, and send alerts to the owner of data to inform him of permissions violation attempts, we must first determine the level of sensitivity of the DW objects. In our approach, the sensitivity level of an object is automatically generated as follows

- A user can have one or multiple roles.

	R_1	R_2	R_3	R_n
U_1	✖			
U_2		✖	✖	
U_3			✖	✖
U_j		✖		

U : User
R : Role
O : Objet

- The role that can modify an object hasthe right to read it.
- The object consulted by several roles is less sensitive than the object accessible by a low number of roles.



The sensitivity level is calculated in percentage using the equation:

$$1 - \left(\frac{\Sigma(\mathrm{P}(O_k))}{2a}\right) * 100 \qquad (1)$$

With

$$\Sigma(\mathrm{P}(O_k)) \qquad (2)$$

The sum of permissions of an object on the set of all the roles.

N

The total number of the roles in our system. (3)

So the sensitivity of a subject o is the number of roles having the permission to read it, compared to total number of roles in our system, which allows obtaining a percentage presenting the sensitivity of the object o.

However traceability allows tracing user actions on sensitive data, which their sensitivity level exceeds the threshold specified by the data owner. Then our system sends an alert to the administrator if the action done is an attempt to breach permissions.

IV. CONCLUSION

This study allowed us to see the major problems of the confidentiality of the data in the warehouse. To reduce these risks, we have proposed a solution based on the user profile, which consists in the definition of access permissions according to the user role using the access rights defined in the sources, generate the level of sensitivity of each object in the DW according to these permissions, trace the access and detect violation attempts of access rights on a sensitive data (a data with a high level of sensitivity). The objective of this solution is to reduce the vulnerability of the data in a DW, and help the owner of the DW to well manage the access control
of the users.

V. PERSPECTIVES

The results show that the research prospects in this direction are numerous, however we have fixed three perspectives that we see interesting in the context of this work, which are:

- Implement our approach in a CC environment (Cloud Computing) : propose an approach that allows the management of the access to the warehouse data hosted in the CC, based on the profession profile and the usage profile of a user, taking into account the specificities of the CC environment.
- Detect inferences by combining accessible data: these combinations that can infer sensitive data detected by this proposition. Note that the combination of many multidimensional elements tends to be more sensitive than data separately.
- Make in practice our approach: to develop an application example.

REFERENCES

[1] H. Kopka and P. W. Daly, A Guide to LATEX, 3rd ed. Harlow, England: Addison-Wesley, 1999.

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

[3] Rosenthal, A. et S. Sciore (2000). View security as the basis for data warehouse security. In DMDW (p. 8).

[4] Blanco, C., E. Fernandez-Medina, J. Trujillo, et J. Jurjens (2010). Towards the secure modelling of olap users behaviour.

[5] 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.

[6] Devbandu, P., Stubblebine, S.: Software Engineering for Security: a roadmap. In: Finkelstein, A. (ed.) The Future of Software Engineering, pp. 227239. ACM Press, New York (2000)

[7] Inmon, H.: Building the Data Warehouse, 3rd edn. John Wiley and Sons, USA (2002)

[8] Landwehr, C.E.: Computer security. Int. Journal of Information Security 1(1), 13 (2001)

[9] 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 (DEXA97), 619626.

[10] 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, 12701289.

[11] Trujillo, J., Soler, E., Fernndez-Medina, E., and Piattini, M. (2009). A UML 2.0 profile to define security requirements for Data Warehouses. Computer Standards and Interfaces, 31(5), 969-983.

[12] Abraham, A., Mauri, J. L., Buford, J., Suzuki, J., and Thampi, S. M. (2011).Advances in Computing and Communications, Part I: First International Conference, ACC 2011, Kochi, India, July 22-24, 2011. Proceedings (Vol. 1). Springer Science and Business Media.

[13] Inmon , 1991 Building the data warehouse

[14] Blanco, C., de Guzmn, I. G. R., Fernndez-Medina, E., and Trujillo, J. (2015). An architecture for automatically developing secure OLAP applications from models. Information and Software Technology, 59, 1-16.

[15] Priebe, T. et G. Pernul (2000). Towards olap security design - sur-vey and research issues. Proceedings of the 3rd ACM International Workshop on Data Warehousing and OLAP (DOLAP00), 3340.

[16] Priebe, T. et G. Pernul (2001). A pragmatic approach to concep-tual modeling of olap security. Proceedings of the 20th Interna-tional Conference on Conceptual Modeling (ER01) 2224, 311324.

[17] Eavis, T. et A. Althamimi (2012). Olap authentication and autho-rization via queryre-writing. The Fourth International Conference on Advances in Databases, Knowledge, and Data Applications, 130139

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

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

[20] Saltor, F., M. Oliva, A. Abello, et J. Samos (2002). Building se-cure data warehouse schemas from federated information systems.

[21] Soler, E., Stefanov, V., Mazon, N.J.: Towards Comprehensive Requirement Analysis for Data Warehouses: Considering Security Requirements, pp. 104111. IEEE, Los Alamitos (2008)

[22] 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, 3143. 23

[23] Rodriguez, A., E. Fernandez-Medina, J. Trujillo, et M. Piattini (2011). Secure business process model specification through a uml 2.0 activity diagram profile. 24

[24] Sweeney, L. (2002). k-anonymity: A model for protecting privacy. 25

[25] Khajaria, K., and Kumar, M. (2011). Evaluation of Approaches for Modeling of Security in Data Warehouses. In Advances in Com-puting and Communications(pp. 9-18). Springer Berlin Heidelberg. 26

[26] Fernndez-Medina, E., Trujillo, J., Villarroel, R., and Piattini, M. (2007). Developing secure data warehouses with a UML extension. Information Systems, 32(6), 826-856