Model Driven Development of Secure XML Data Warehouses: A Case Study

Belén Vela  
Rey Juan Carlos University of Cantabria  
C/ Tulipán s/n  
28933 Móstoles-Madrid, Spain  
+34 91 488.70.03  
belen.vela@urjc.es

Carlos Blanco  
University of Av. De los Castros s/n  
39071 Santander, Spain  
+34 942 206762  
Carlos.Blanco@unican.es

Eduardo Fernández-Medina  
University of Castilla-La Mancha  
Paseo de la Universidad, 4 – 13071 Ciudad Real, Spain  
+34 926 295300  
Eduardo.FdezMedina@uclm.es

Esperanza Marcos  
Rey Juan Carlos University  
C/ Tulipán s/n  
28933 Móstoles-Madrid, Spain  
+34 91 664.74.91  
esperanza.marcos@urjc.es

ABSTRACT
Data Warehouses (DWs) are currently considered to be the cornerstone of Business Intelligence (BI) systems. Security is a key issue in DWs since the business information that they manage is crucial and highly sensitive, and should be carefully protected. However, the increasing amount of data available on the Web signifies that more and more DW systems are considering the Web as the primary data source through which to populate their DWs. XML is therefore widely accepted as being the principal means through which to provide easier data and metadata interchange among heterogeneous data sources from the Web and the DW systems.

Although security issues have been considered during the whole development process of traditional DWs, current research lacks approaches with which to consider security when the target platform is based on the Web and XML technologies. The idiosyncrasy of the unstructured and semi-structured data available on the Web definitely requires particular security rules that are specifically tailored to these systems in order to permit their particularities to be captured correctly.

In order to tackle this situation, in this paper, we propose a methodological approach based on the Model Driven Architecture (MDA) for the development of Secure XML DWs. We therefore specify a set of transformation rules that are able to automatically generate not only the corresponding XML structure of the DW from secure conceptual DW models, but also the security rules specified within the DW XML structure, thus allowing us to implement both aspects simultaneously. A case study is provided at the end of the paper to show the benefits of our approach.

Keywords
Secure XML Data Warehouse, Model Driven Development, MDA, XML.

1. INTRODUCTION
Data Warehouse (DW) systems provide a Multidimensional (MD) view of huge amounts of historical data from heterogeneous operational sources, thus supplying useful and sensitive information which allows decision makers to improve business processes in organizations. The MD paradigm structures information into facts and dimensions. A fact contains the interesting measures (fact attributes) of a business process (sales, deliveries, etc.), whereas a dimension represents the context for analyzing a fact (product, customer, time, etc.) by means of hierarchically organized dimension attributes.

Traditional DW systems allow business people to acquire useful knowledge from their organization’s data by means of a variety of technologies, such as OLAP or data mining. However, in order to provide richer insights into the dynamics of today’s business, it is currently desired that the data inside the organization be combined with data from the outside, thus complementing the company’s internal data with value-adding information (e.g., retail prices of products sold by competitors). Since the amount of data available on the Web has been growing rapidly over the last decade, Web data prove to be more and more useful for this purpose. The main problem with data from the Web is that they are rather heterogeneous and complex. To overcome such drawbacks, designers of DW systems make use of this data by using XML technologies [19,23]. On the one hand, Web warehousing uses XML as a means of ameliorating the extraction and integration of heterogeneous Web data in the DW [10]. On the other hand, document warehousing requires XML to deal with unstructured data in DW systems [24]. In both cases XML is used to implement the MD model underlying the DW by defining the corresponding design artifacts (facts, dimensions, measures, hierarchies and so on) in order to facilitate the interchange of data and metadata among heterogeneous data sources and the DW system [29]. The design of XML DWs is therefore a cornerstone when it is necessary to use Web data in the decision making process, which is becoming more and more frequent.

Furthermore, every design issue should be considered in the development process of an XML DW. Specifically, one of the
most important design issues is security, which has, to date, been surprisingly overlooked when XML DWs are being developed. Considering that the information managed by DWs is frequently highly sensitive, and sometimes refers to personal data (protected under the law in most countries), DWs should be protected from unauthorized information accesses (whatever the implementation developed data management systems is a demand for adequate platform is). In fact, a key requirement underlying these recently developed data management systems is a demand for adequate security, and fine-grained flexible authorization models and access control mechanisms (since DWs deal mainly with read operations). Therefore, rather than considering security once the system has been completely built, we believe that security and privacy measures should be integrated in all layers of the DW design, from the early stages of its development as another relevant requirement, meaning that much more robust, secure and platform independent products will be produced [19,31].

In order to develop secure XML DWs considering confidentiality issues in the whole development process, from an early development stage to the final implementation, our proposal has been aligned with an MDA (Model Driven Architecture [20]) architecture in which security models are embedded in and scattered throughout the high level system models, which are transformed until their final implementation according to the MDA strategy. MDA can be used for this purpose, since it shares some similarities with the traditional MD modeling methods [27]: i) a conceptual design phase is carried out, whose output is an implementation-independent and expressive conceptual MD model for the DW (i.e. a Platform Independent Model, PIM), ii) a logical design phase aims to obtain a technology-dependent model (i.e. a Platform Specific Model, PSM) from the previously defined conceptual MD model, and iii) this logical model is then the basis for the implementation of the DW.

After presenting some related works in the following section, in Section 3 we will introduce the secure XML DW development approach. The PIM is the secure conceptual DW data model, which will be semi-automatically transformed into a secure XML DW, as a PSM, applying a set of transformation rules summarized in Section 4. In addition, in Section 5 we will present a case study to show our proposal. Finally, in Section 6, we will put forward our main conclusions and present our future work.

2. RELATED WORK
In this section the related work is organized according to the following topics: (1) MD modeling, (2) security integration into the design process and (3) security and access control models for DWs.

2.1 MD Modeling
Some MD data models are focused on the logical level [1,6]. However, the most interesting proposals consider the conceptual level in order to model multidimensional concepts (facts, dimensions, classification hierarchies and so on) in a platform independent manner by extending the classical EER model [30,32] by defining their own graphical notation [11,13]. On the other hand, other proposals use the object-oriented paradigm and are based on UML (YAM2 [2] or the object oriented metacube [5]).

Furthermore, when dealing with XML DWs, some works model and analyze this kind of systems by taking into account the fact that cubes and dimensions are stored in XML documents [7,22] and extend the XQuery language with OLAP capabilities [4,12,38]. All of these modeling proposals permit the definition of DWs’ structural aspects, but only some of them (ADAPTed UML [26] and SECDW [8]) consider security as an important issue to be considered in DW modeling, although they do not deal with XML data warehousing.

2.2 Secure Integration into the design process
Several relevant works can be found which concern a complete secure development but which focus on information systems in general. For instance, UMLSec [14] uses UML to define and evaluate security specifications using formal semantics. Moreover, Model Driven Security (MDS) [3] uses the MDA approach to include security properties in high-level system models and to automatically generate secure system architectures. Within the context of MDS, SecureUML [17] is proposed as an extension of UML for modeling a generalized role based access control. On the other hand, Mokum [33] is an active object oriented knowledge base system for modeling, which permits the specification of security and integrity constraints and automatic code generation. These are relevant contributions towards secure information systems development but are not specifically focused on DWs.

2.3 Security and Access Control Models for DWs
Since final users work with an MD model when querying a DW (facts, dimensions, classification hierarchies, etc.), security constraints must be defined in terms of MD modeling.

There are several interesting initiatives for the inclusion of security in DWs, but they are not conceived for integration into MD modeling as part of the DW design process, and, inconsistent security measures may consequently be defined. Katic et al. [15] present a security model based on metadata to define user groups and views. Rosenthal and Sciore [27] integrate security from the data sources and propagate it to DW design. Other proposals define authorization models and security for DWs [16,25,26,36,37] but they deal solely with OLAP operations (such as roll-up or drill-down).

3. SECURE XML DW MODELING
In this paper, we use the Model Driven Architecture (MDA) [20] to define security in the MD modeling of XML DWs. We concretely define security specifications in the Conceptual MD Data Model (PIM), independently of the target logical MD model. This Secure Conceptual MD Data Model will be used as a starting point and will be semi-automatically transformed into a Secure XML DW as a logical model (PSM) by applying Model to Model (M2M) Transformations. Finally, a Model to Text (M2T) transformation will generate the code for the Secure XML DW.

For the model driven development of a secure XML DW it is therefore necessary to perform the following tasks (see Figure 1):

- At the PIM level, the secure MD data model is carried out without considering the selected technology, since this model is independent of the platform. This MD PIM (described in more detail in the following subsection) is represented through an extended UML class diagram for DWs which furthermore permits the specification of security constraints on the model.
At the **PSM level**, the data logical design is performed, taking into account the selected target platform in which the DW will be implemented. In our case, XML technology will be used for the implementation of the DW in any secure commercial database management system. We shall start from the secure MD PIM obtained at the previous level and shall apply the mappings summarized in Section 4 to obtain an XML Schema, conforming to the XML Schema Metamodel [35].

### Figure 1. Development approach for Secure XML DW

#### 3.1 Secure MD PIM

As previously mentioned, our proposed development approach starts from the conceptual model of the secure MD PIM. In order to define this secure MD PIM, a secure UML profile denominated as SECDW has been developed (for more details, see [8]). SECDW (Figure 2) considers both specific aspects of DW modeling (such as facts, dimensions, base classes, measures, hierarchies, many-to-many relations, degenerated dimensions, multiple classifications or alternative paths of hierarchies) and security capabilities by using an Access Control and Audit (ACA) model [9].

The ACA model classifies authorization subjects and objects into security roles (“SRole” metaclass) which organize users into a hierarchical role structure according to the responsibilities of each type of work, levels (“SLevel” metaclass) which indicate the user’s clearance level, and compartments (“SCompartment” metaclass) which classify users into a set of horizontal compartments or groups.

The definition of several kinds of security rules related to the multidimensional elements of DWs is also permitted: Sensitive Information Assignment Rules (SIAR) (“SecurityRule” metaclass) which specify multilevel security policies and allow sensitive information to be defined for each element in the multidimensional model; Authorization Rules (AUR) (“AuthorizationRule” metaclass) which permit or deny access to certain objects by defining the subject that the rule applies to, the object that the authorization refers to, and the sign describing whether the rule permits or denies access; and Audit Rules (AR) (“AuditRule” metaclass) which ensure that authorized users do not misuse their privileges.
4. Mappings from PIM to PSM

In the same way that methodologies for relational or OR DBs propose certain rules for the transformation of a conceptual schema into a logical schema, in this section we propose the mappings from the secure MD PIM to the XML Schema for the Secure XML DW. The basis for this is the work of [34], in which the different mappings used to obtain the schema of a secure XML DB were defined but did not take into consideration security aspects for MD modeling aspects. Table 1 shows the mapping rules for the transformation of a secure multidimensional model (PIM level) into an XML Schema model (PSM level).

Table 1. PIM to PSM Transformation Rules

<table>
<thead>
<tr>
<th>Data PIM</th>
<th>Data PSM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Secure MD PIM</strong></td>
<td>XML Schema which includes the root element “SecureMDXML” of “SecureMDXML_Type”, which includes the Star Packages, Security Levels, Security Roles Hierarchy, Security Compartments and User Profile.</td>
</tr>
<tr>
<td><strong>Security Levels</strong></td>
<td>XML Element named “SecurityLevels” including a sequence complexType (SecurityLevel_Type) with all the defined Security Levels as subelements. Each subelement will contain the subelements: name, short name and order number (beginning with 1, as the highest security level).</td>
</tr>
<tr>
<td><strong>Security Roles Hierarchy</strong></td>
<td>XML Element named “SecurityRoles” including a sequence complexType (SecurityRole_Type) with all the defined Security Levels as subelements. Each subelement will contain the name and a reference element to its parent, that is, the security role which includes it.</td>
</tr>
<tr>
<td><strong>Security Compartments</strong></td>
<td>XML Element named “SecurityCompartments” including a sequence complexType (SecurityCompartments_Type) with all the defined Security Compartments as subelements.</td>
</tr>
<tr>
<td><strong>Secure Information Class</strong></td>
<td>XML ComplexType named “SecureInformationType” with three subelements:</td>
</tr>
<tr>
<td>SecurityLevels SecurityRoles securityCompartments</td>
<td>• SecurityLevels of SecurityLevel_Type with the corresponding attributes as XML subelements</td>
</tr>
<tr>
<td></td>
<td>• SecurityRoles of SecurityRoles_Type with the corresponding attributes as XML subelements</td>
</tr>
<tr>
<td></td>
<td>• securityCompartments of SecurityCompartments_Type with the corresponding attributes as XML subelements</td>
</tr>
<tr>
<td><strong>User Profile Class</strong></td>
<td>XML Element of a sequence complexType “UserProfile_Type” which includes as subelements its code, the name, the specific class attributes and the SecureInformation subelement of SecureInformation_Type.</td>
</tr>
<tr>
<td><strong>Secure Star Package</strong></td>
<td>XML Element named “StarPackage” including a complexType with the Fact, Dimension and Base XML Subelements, each of which contains the corresponding Security Information and Security Constraints as XML Subelements.</td>
</tr>
<tr>
<td><strong>Classes Attribute of PIM</strong></td>
<td>XML Element of complexType including references to all the elements transformed from the PIM classes XML Subelement of complexType</td>
</tr>
<tr>
<td><strong>OID</strong> classes Dimension**</td>
<td>XML Subelement (of the StarPackage) including the OID attribute, a sequence complexType with all class Descriptor and Dimension attributes</td>
</tr>
<tr>
<td><strong>OID</strong> Descriptor Dimension**</td>
<td>XML Subelement of complexType</td>
</tr>
<tr>
<td><strong>OID</strong> Fact**</td>
<td>XML Subelement of the complexType</td>
</tr>
<tr>
<td><strong>OID</strong> Dimension**</td>
<td>XML Subelement of the complexType</td>
</tr>
<tr>
<td><strong>Secure Class Attributes</strong></td>
<td>XML Subelement including the class attribute and the security properties</td>
</tr>
</tbody>
</table>
5. CASE STUDY
This section presents an example based on an airport DW that manages information concerning trips, flights and incidents, in which the XML secure model (PSM) is obtained from the secure conceptual MD data model (PIM).

Figure 3 shows a partial view of the multidimensional model focused on trip information (secure fact class “Trip”) involving passengers (secure dimension class “Passenger”) who take flights to reach their destinations (secure dimension class “Airport”) and a related secure base class “Airport”. Some attributes have been included to manage information about trips (price, purpose, seat, distance, flight time, and whether or not checking in and boarding have taken place), departure and arrival locations (gate, terminal and airport), and passengers (personal data: name and address; and security information: fingerprints, photo and whether or not the passenger is considered to be suspicious).

In order to establish security constraints, a security configuration has been defined by using security levels (SL) and security roles (SR). The security levels are top secret (TS), secret (S), confidential (C) and unclassified (U), and the hierarchy of security roles is composed of a main role “User” with three subroles “Airport Security”, “Administration” and “Passenger”. A set of sensitive information assignment rules (SIAR) has then been defined over some classes and attributes by using stereotypes. The “Trip” fact class and the “Place” dimension class can be accessed by users with the confidential (or upper) security level; the “Passenger” dimension class by those with the secret security level, and the “Airport” dimension class by those with the undefined security level. Several attributes also have fine grain security constraints associated with them, which permit the security role “Airport Security” to access the attributes: “purpose” (of the “Trip” fact), and “fingerprints”, “passportPhoto” and “suspicious” (of the “Passenger” dimension).

More complex security (SIAR) and authorization (AUR) rules have also been defined by using the metaclasses “SecurityRule” and “AuthorizationRule”. The “SIAR_TripPurpose” rule is associated with the “Trip” fact and involves the “Passenger” dimension. This rule increases the security requirements of the fact and the involved classes if the purpose of the trip is military (“purpose” attribute). In this case a security level of “Secret” and a security role of “Airport Security” will be required. The “AUR_Passenger” is a positive authorization rule which checks the user name (“name” attribute of “UserProfile”) and provides access to his/her passenger information.
We have used this Secure MD PIM as a starting point to apply the defined transformation rules and we have obtained the following XML Schema. Only a part of the generated XML PSM is shown owing to space constraints.
6. CONCLUSIONS AND FUTURE WORKS

In this work we have proposed an approach for the model driven development of Secure XML DWs. Our approach starts by defining the secure conceptual MD model (PIM) represented by means of the secure UML profile called SECDW, independently of the target logical MD model. This PIM is used as a starting point and is then semi-automatically transformed into a secure conceptual models of the DW, but also the security rules specified within the DW XML structure, thus allowing both aspects to be implemented simultaneously.

In order to validate our proposal we have carried out several case studies. One of these is shown in this paper to illustrate the benefit of our approach: an airport DW that manages information concerning trips, flights and incidents.

We are now working on several different lines, in an attempt to extend the proposal presented in this paper. One of these, on which we have already started to work, is the automation of the transformations of the constraints expressed in OCL at the PIM level, in order to convert them into XPATH language. Moreover, we are also working on the automation of the transformations between the metamodels and the corresponding models using the Query View Transformation (QVT) proposal [21]. A further goal is that of performing several case studies to detect new needs. These would also analyze the advantages of incorporating security aspects provided by the different XML DB administrators, and not only those which are native. The next step, will be to include our proposal in the case tool that we are developing for the semi-automatic development of Secure XML DW.

7. ACKNOWLEDGMENTS

This work has been carried out in the framework of the following projects: MODEL-CAOS (TIN2008-03582/TIN) and AGREEMENT-TECHNOLOGY (CSD2007-0022) financed by the Spanish Ministry of Education and Science, IDONEO (PAC08-0160-6141) financed by the “Consejería de Ciencia y Tecnología of the Junta de Comunidades de Castilla-La Mancha”, the Data Management network (TIN2008-04453-E), financed by the Spanish Ministry of Science and Innovation (MICI), the BUSINESS (PET2008-0136) financed by the MCI, the QUASIMODO (PAC08-0157-0668) financed by the “Viceministerio de Ciencia y Tecnología de la Junta de Comunidades de Castilla-La Mancha” and the SISTEMAS (PII209-0150-3135) financed by the “Consejería de Educación y Ciencia de la Junta de Comunidades de Castilla-La Mancha”.

8. REFERENCES

[12] Hachicha, M., Mahboubi, H. and Darmont, J. Expressing OLAP operators with the TAX XML algebra, in EDTB workshop on Database Technologies for Handling XML...


