An Access Control Model for Geographic Data in an XML-based Framework

Bat-Odon Purevjii[1], Toshiyuki Amagasa[2], Sayaka Imai[1], and Yoshinari Kanamori[1]

[1]Department of Computer Science
Faculty of Engineering, Gunma University
1-5-1 Tenjin-cho, Kiryu, Gunma 376-8515, Japan

[2]Graduate School of Information Science
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara 630-0192, Japan

Abstract. XML is accepted as a standard format for data exchange on the Web. XML security research has mainly been focused on textual documents since its origin. The previous works are not appropriate in the context of geographic data, which comprises spatial and non-spatial data. Controlling access to those data was a troublesome task because of proprietary and composite data formats of GISs and complex data structures of spatial database systems. Besides allowing flexible integration of geographic data, XML offers us methods to deploy access control for these systems. In this paper we propose an access control model for Web GIS, in which controlled access depends on spatial extents of geographic data, by employing XML, and XML-based 2D vector graphic format SVG. We show its utility in the domain of Internet Mapping into the application of public safety and disaster management for national and local governments.

1 Introduction

Recent Web and XML technologies allow us to transfer from conventional Geographic Information System (GIS) to Web/Human GIS. As a result, every user on the Internet can retrieve geographic information using communication devices, such as PCs, mobile phones, and PDAs, independent of his/her location. The GIS paradigm is shifting from expert-oriented GIS to low-cost consumer oriented GIS which provides an open and rich information container for typical users on the Internet.

Internet Mapping is the major integral part of Web GIS, which can be implemented either within GIS or separately by using XML technologies as glue. The W3C's Scalable Vector Graphics (SVG) [15] has become a new visualization format of such geographic and map data on the Internet[1] [6, 13].

[1] In fact, major GIS vendors started to support the format in their product families.

This work is partially supported by JSPS Grant-in-Aid for Scientific Research (B) No. 15300029.

Purevjii B., Amagasa T., Imai S. and Kanamori Y. (2004).
An Access Control Model for Geographic Data in an XML-based Framework.
In *Proceedings of the 2nd International Workshop on Security in Information Systems*, pages 251-260
DOI: 10.5220/0002677402510260

Such powerful Web GIS and online interactive maps support facilities for collecting, manipulating, sharing, visualizing, and analyzing geo-related information and human related information. Since there is no geographical boundary for accessing such private information, there arises specific security concerns in this context.

Applications for ensuring public safety [7, 14], such as disaster management, homeland security, emergency medical services, crime and terror analysis are emerging and have been under the attention of nations for the last few years. When we face a big disaster, it is important to be able to make quick access to geographic and personal data at any level of government, non-government, and private entities for minimizing damage and saving lives [7, 14]. Thus, offering a large amount of personal information with real world spatial positions and addresses, privacy protection is inevitable in this incorporated environment. Such open and rich information sources may be used by various kinds of users for aggressive purposes.

In this paper we introduce a technique to protect sensitive information in a Web GIS environment by developing an XML-based geographic access control model. In particular, we consider spatial extents and levels-of-details (LoD) of geographic data to regulate access.

The remainder of this paper is structured as follows: Section 2 introduces related work and Section 3 provides a demonstrative scenario that sketches the access control policies and requirements. Section 4 presents a brief introduction to SVG vector format and XML foundations. The underlying XML-based geographic data model is presented in Section 5 and the access control model formulation is introduced in Section 6, respectively. Section 7 describes spatial access control enforcement algorithm. Finally, section 8 gives concluding remarks and future improvements in this area.

2 Related Work

Controlling access to XML documents in the context of text data have received notable efforts and the data security community has proposed several models and systems.

Damiani, et al. [3] and Bertino, et al. [1] have developed access control systems and prototypes in which fine-grained and content-dependent control of XML document fragments are realized. Damiani, et al. [4] presented a model for selectively controlling access to fragments of SVG graphics. Authorization objects are defined by explicit object IDs and implicit conditions on the objects. Kodali, et al. [8] have proposed a model for enforcing control to SMIL movies. S. De Capitani di Vimercati [5] has considered temporal aspects of XML and proposed an authorization model for Temporal XML documents.

Since these previous proposals are considered to be inappropriate in a geographic context, we extend them and propose an XML-based access control model by taking into account the peculiarities of geographic data. Specifically, we consider spatial extents and LoD of XML-based geographic data to regulate access.

Chun, et al. [2] have proposed a novel model for protecting a geo-spatial image database by developing an indexing structure for geo-spatial data. However, they did not consider XML formats and map data.

3 Motivation and Demonstrative Scenario

Within a framework of a disaster management system for local and national governments, various kinds of users, (organized as teams, such as administrative, rescue, medical, research and inhabitants etc.,) interact with the system dynamically depending on their needs and responsibilities[2]. The teams are allowed to access only relevant information to execute their job functions, e.g. a medical team has access to the detailed information of invalids and/or people who need special care during a disaster, such as concrete address, current position of invalids, physical condition, etc. On the other hand, a rescue team has access to detailed information of all houses and buildings within the danger area, such as the number of people, the house condition, pictures and detailed information of surrounding buildings, and personal data of members of households etc.

The members of a team need to be classified into levels depending on their administrative units: from local to national (such as area, district, town, city, prefecture and nation). Furthermore, members in the same team at the same administrative unit level are divided into distinct administrative regions, depending on their locations.

Before introducing requirements for access control in such a working environment, we describe some geographic data concepts briefly. In Figure 1 we can see a map fragment depicting layers of geographic information with an instance of descriptive data of a geographic feature (e.g. a building) in a Web GIS environment.

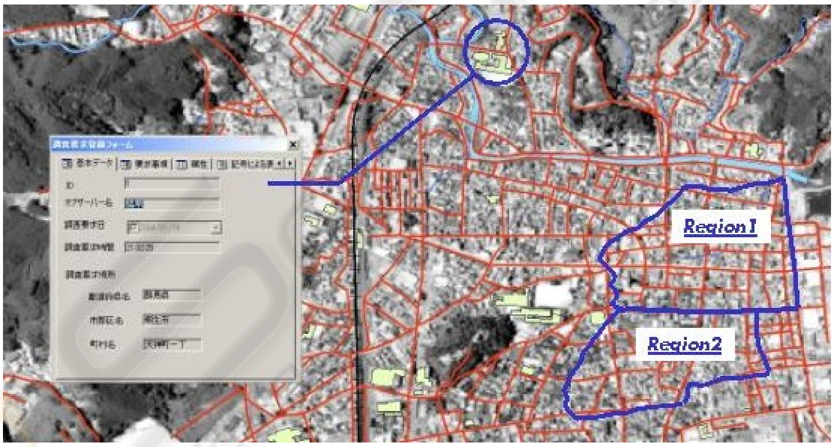


Fig. 1. Fragment of a map in a Web GIS environment.

The descriptive data handles alphanumeric properties of features, and is called the non-spatial component of geographic data. In our case, it can be detailed information of the building. They are stored in a database behind the Website. Besides having descriptions, features have spatial extents, geometry and topology, and are called the

[2] The intention here is to introduce conceivable security requirements in our model but not to illustrate a robust emergency system.

spatial component of geographic data. At the conceptual level, an instance of geographic data is called a geographic object[3] [11].

When a user clicks on a feature, the database query will be issued and related descriptive data is rendered as shown in Figure 1. In a conventional geographic context, users are allowed to access both spatial and non-spatial components of geographic data without any types of security limitations. However, according to security policies in the beginning of this section, only users who belong administratively to *Area1* (on the right side of Figure 1) might be allowed to access geographic objects within this region only, but not in *Area2* or in any of the remaining regions. *Area1* and *Area2* become a bigger administrative region, and the principle should function at this level and for all higher levels up to the national level.

To summarize, the following requirements are derived:

- Flexible organization of various kinds of users by employing any possible combinations of role-based, ID-based and profile-based paradigms;
- Controlling access to geographic objects based on spatial extents;
- Permissions to access geographic data with several LoD;

4 Basic Concepts

4.1 XML basics

To derive a data model of an XML document, graphs and trees are widely employed. XML expresses information using four basic components: elements, attributes, data values, and hierarchy/graph. XML elements are tagged and they contain data values or other elements recursively. Elements may have one or more attributes and the attributes define properties of the elements. In this paper we use a directed graph to structure components of XML because of its expressive power. Vertices of the graph represent elements and attributes, and edges represent relationships between them. We define an XML document as follows:

Definition 1 (XML document). An XML document is a tuple d=(V^e, V^a, E, Tg, Vl, r, $\phi_E{}'$), where:

- V^e is a set of vertices representing elements,
- V^a is a set of vertices representing attributes,
- E is a set of edges,
- Tg is a set of element and attribute names,
- Vl is a set of element and attribute values,
- r is the root of an XML document,
- ϕ_E is a function that associates edges and vertices.

Here we use an ordered model of an XML document. Each vertex representing an element contains the graph-wide unique identifier.

[3] We use terms geographic data and geographic object interchangeably later on.

4.2 Overview of SVG

Scalable Vector Graphics (SVG) [15] is an XML-based markup language used to describe and integrate two-dimensional vector graphics, raster images and text. Basic geometric/graphics elements are: rectangle, circle, ellipse, line, polyline, polygon, and finally the path object. Graphics in SVG format can be semantically reached and are highly structured. Objects in this structure can be grouped, styled and composed into higher-level objects. SVG offers a number of important advantages over raster formats, especially when it comes to displaying map graphics. The advantages include:

- SVG works well across platforms, output resolutions, and a range of bandwidths
- SVG fully supports the DOM (Document Object Model) and is fully scriptable
- SVG offers greater structural control than other raster and vector graphic formats

5 Geographic Data Model based on XML

The geometric attributes of geographic data are used to effectively store and retrieve such data. In order words, the attributes define only the physical properties of those data. Hence, we need to specify a data model which is capable of expressing geographic objects on a high level before proceeding to the definitions of the access control model.

In GIS, information is organized into layers and a *layer* consists of homogeneous geographic objects/*features* (i.e., objects having the same structure). To model geographic data in XML, we employ XML and XML-based graphic format SVG.

Generally, SVG has five kinds of elements, including graphics/geometric elements, non-rendered text elements, rendered elements by references, reference elements, and container elements. The geometric elements (listed in section 4.2) are the main building blocks of *features* of geographic data. Thus, we assume classifications of SVG elements as only two, geometric and non-geometric, for formal definitions. IDs of *layers*, *features* and SVG elements are unique and can be the graph-wide identifiers or can be defined by the ID attributes of elements. Let *LR* be a set of layers and *FT* a set of *features*.

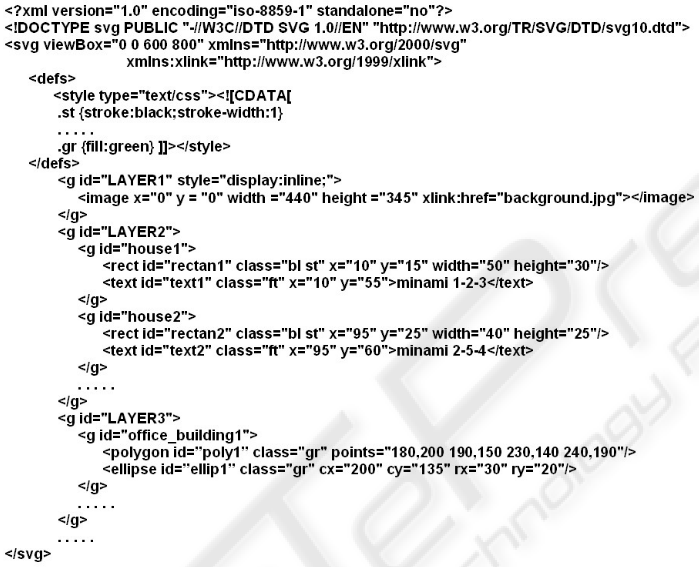
Definition 2 (SVG map). A SVG map is a tuple $sm=(V^{e}_{sm}, V^{a}_{sm}, E_{sm}, Tg_{sm}, Vl_{sm}, r_{sm}, \phi_{Esm}, G_{sm})$, where:

- $V^{e}_{sm} = V^{geo}_{sm} \cup V^{nongeo}_{sm}$ is a set of vertices, where V^{geo}_{sm} are vertices representing geometric elements and V^{nongeo}_{sm} are vertices representing non-geometric elements, respectively.
- V^{a}_{sm} is a set of vertices representing attributes of all kinds of elements,
- E_{sm} is a set of edges representing relationships among elements;
- Tg_{sm} is a set of element and attribute names,
- Vl_{sm} is a set of element and attribute values.

- r_{sm} is a SVG root element,
- ϕ_{Esm} is a function that associates edges and vertices,

A *feature* can be denoted by listing identifiers of geometric elements in SVG.

Definition 3 (A feature in a SVG map). A *feature* in a SVG map, denoted by *ft*, is a set of identifiers of geometric element in *sm*, that is, $ft=\{id_1, id_2, \ldots , id_n\}$, with $id_i \in \{id_v \mid v \in V_{sm}^{geo}\}$.



```
<?xml version="1.0" encoding="iso-8859-1" standalone="no"?>
<!DOCTYPE svg PUBLIC "-//W3C//DTD SVG 1.0//EN" "http://www.w3.org/TR/SVG/DTD/svg10.dtd">
<svg viewBox="0 0 600 800" xmlns="http://www.w3.org/2000/svg"
                 xmlns:xlink="http://www.w3.org/1999/xlink">
    <defs>
        <style type="text/css"><![CDATA[
         .st {stroke:black;stroke-width:1}
         . . . . .
         .gr {fill:green} ]]></style>
    </defs>
         <g id="LAYER1" style="display:inline;">
             <image x="0" y = "0" width ="440" height ="345" xlink:href="background.jpg"></image>
         </g>
         <g id="LAYER2">
             <g id="house1">
                 <rect id="rectan1" class="bl st" x="10" y="15" width="50" height="30"/>
                 <text id="text1" class="ft" x="10" y="55">minami 1-2-3</text>
             </g>
             <g id="house2">
                 <rect id="rectan2" class="bl st" x="95" y="25" width="40" height="25"/>
                 <text id="text2" class="ft" x="95" y="60">minami 2-5-4</text>
             </g>
             . . . . .
         </g>
         <g id="LAYER3">
             <g id="office_building1">
                 <polygon id="poly1" class="gr" points="180,200 190,150 230,140 240,190"/>
                 <ellipse id="ellip1" class="gr" cx="200" cy="135" rx="30" ry="20"/>
             </g>
             . . . . .
         </g>
         . . . . .
</svg>
```

Fig. 2. An example of SVG, presenting a layered map including features.

Consequently, a *layer* can be denoted by listing identifiers of *features* or directly by its own unique ID. A very simple SVG map example is shown in Figure 2.

For definitions of the components of descriptive data of geographic objects, XML formulations are employed. We denote the set of descriptive data as XML description database.

Definition 4 (XML description database). An *XML description database* denoted by *db* is a XML document formulated by Definition 1.

The definitions of *records* and *tables* can be done by listing XML element identifiers of the XML description database. Thus, we denote a set of *records* as *RD* and a set of *tables* as *TB,* respectively.

Definition 5 (Feature-description mapping). Feature-description mapping (depicted in Figure 3) is a one-to-one relation from feature set *FT* to record set *RD*. In order words, there is a function *fdm*: $FT \rightarrow RD$ where each ordered pair *(a, b)*, $a \in FT$ and $b \in RD$, is unique and is connected by an ID/key.

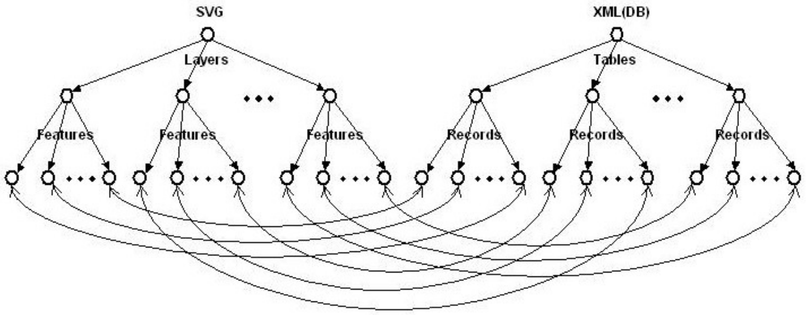


Fig. 3. Feature-description mapping

6 XML-based Geographic Access Control Model

6.1 Authorization object

We define authorization objects directly in terms of XML element identifiers and/or by XPath and XQuery expressions [1, 3] in our model. However, in order not to lose semantics and structural organizations of geographic data, we should define authorization objects by identifiers of features and layers or spatial extents of them. In addition, defining authorizations on each feature is impractical since geographic data hold huge amounts of geographic information that leads the authorizations base to become too large and unorganized. Instead, we define authorization objects at a layer level in authorizations and use spatial extents to filter protection objects within a region in a layer.

6.2 Authorization subject

We employ the integrated approach of user ID, profile and role-based paradigms in our system. Thus, an authorization subject represented as a tuple, (*uid, profile, role-set*), where *uid* is the unique ID of a user, *profile* is a set of user properties, and *role-set* is a set of role names. *Profile* in turn is a tuple (*name, address, work*), where *name* is a user name, *address* is an affiliation address and *work* is an affiliation function, respectively. The users can be modeled and organized by role hierarchies like any other role-based models. In our current proposal we utilize a role-graph structure such as [9] and will develop more flexible structures in further works.

Definition 6. (Authorization subject) An authorization subject specification is the form, structured as XML document:

```
<ELEMENT  subject          (uid, profile, role+)>
<ELEMENT> uid              (#PCDATA)
<ELEMENT> role             (#PCDATA)
<ELEMENT  profile          (name, address, work)>
```

```
<ELEMENT> name              (#PCDATA)
<ELEMENT> address           (#PCDATA)
<ELEMENT> work              (#PCDATA)
```

6.3 Authorization

An authorization in the model is specified as follows:

Definition 7 (Authorization). An authorization a is a tuple (*auth-subj, auth-obj, oper_reg, LoD, md*), where: *auth-subj* is the authorization subject specification as defined in Definition 6, *auth-obj* is the authorization object specification defined as layers, *md* is a set of action modes like view, insert, update, delete, all, etc., *oper_reg* and *LoD* are the values which define a set of operative regions and *LoD* the subject allowed to access. Currently, the *LoD* is defined only on the descriptive database part of objects and the values of *LoD* range from 3 to 1, i.e. from fine-detailed to coarse-detailed.

Example 1: A_1 = ((*su123, null, adm*), {$layer_k$, $layer_{k+1}$, $layer_{k+4}$}, *regionL2-1, 1, {all}*): This authorization allows the subject, who is included in role *administrative* and *uid* = *su123* to execute all action modes (view, insert, update and delete) on objects inside *regionL2-1* on $layer_k$, $layer_{k+1}$ and $layer_{k+4}$ with detail level 1.

Example 2: A_2 = ((*su456, null, medical*), {$layer_{k+1}$, $layer_{k+2}$}, *regionL1-1, 2, {view}*): This authorization allows the subject, who is included in role *medical* and *uid* = *su456* to execute view mode on objects inside *regionL1-1* on $layer_{k+1}$ and $layer_{k+2}$ with detail level 2.

In the following section we introduce a spatial access control enforcement algorithm.

7 Spatial Access Control Enforcement

Although access control enforcement is quite similar in many systems [12], in order to manage access control spatially a spatial indexing structure on authorization objects is needed [2]. Chun, et al. [2] have chosen a variant of quadtree, which belongs to the space-driven approach and the main memory access method. To manage access control spatially and effectively for geographic data, however, we need a secondary memory structure and the data-driven approach with LoD support [10]. Since, typically, geographic databases occupy several gigabytes of storage and the distribution of geographic objects on a plane is rambling. Due to space limitation, we will present such an indexing structure in another paper.

Here, we describe spatial access control enforcement in an easier way by employing R-tree based indexes, such as R*-tree. Data values of x, y coordinates of SVG geometric elements express positions of features in a SVG map. Consider features in the SVG map as indexed by R-tree on those values and the leaf nodes of the tree contain IDs of features of the SVG map.

To describe our algorithm we have used literals, because our intention is to give a clear explanation rather to be efficient or provide well-formalized listing. The enforcement algorithm comprises the following main stages:

Step 1. *Authorizations evaluation*: Determine the set of suitable authorizations for the user's request.

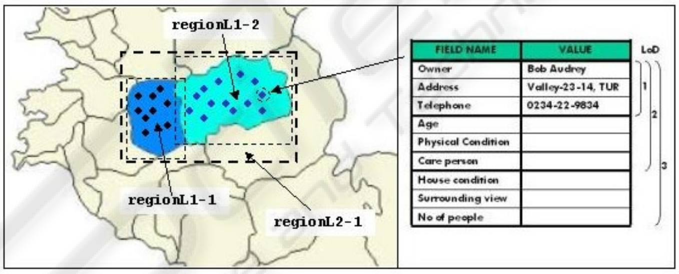
Step 2. *Spatially authorized objects calculation*. Retrieve IDs of features by traversing R-tree indices of authorized layers by window queries on the operative region set values of the requestor.

Step 3. *Elimination of unnecessary IDs*. By employing a complex elimination algorithm or using an address-matching function, check and eliminate IDs of features, which do not reside in authorized administrative regions of the requestor.

Step 4. *Retrieval of corresponding records*. By the extracted IDs, parse and retrieve corresponding records with appropriate LoD from the *XML description database*. Retrieval of records with appropriate LoD can be done by XQuery expressions.

Step 5. *Document rendering*. Form a final view of the SVG map only with spatially authorized features and render for the requestor.

During navigation with the map, the requestor can access/see only authorized features. The descriptive data of the features will be depicted on the allowed LoD as well. In Figure 4 we can see an instance of a user (who is allowed to access features in *regionL2-1* with LoD 1) view, which rendered after considering corresponding authorizations.



FIELD NAME	VALUE
Owner	Bob Audrey
Address	Valley-23-14, TUR
Telephone	0234-22-9834
Age	
Physical Condition	
Care person	
House condition	
Surrounding view	
No of people	

Fig. 4. An instance of a user view.

8 Conclusions and Future Work

In this paper we have presented an XML-based access control model for regulating access to geographic data considering spatial extents and LoD. We have shown its utility in the domain of Internet Mapping into the application of a disaster management framework for national and local governments.

Since the work is derived from our preliminary results, a number of improvements and investigations need to be done. The performance issues will be evaluated and efficiency improvements will be conducted. Temporal constraints of access control will also be integrated into the model.

References

1. E. Bertino and E. Ferrari, "Secure and selective dissemination of XML documents", *ACM Transactions on Information and System Security (TISSEC)*, Vol. 5, No. 3, pp. 290-331, August 2002.
2. S. A. Chun and V. Atluri, "Protecting Privacy from Continuous High-resolution Sattellite Surveillance", *Data and Application Security: Developments and Directions,* (eds.) Bhavani Thuraisingham, R. Van de Riet, K. R. Dittrich and Z. Tari, Kluwer Academic Publishers, 2001.
3. E. Damiani, S. De Capitani di Vimercati, S. Paraboschi, and P. Samarati, "A Fine-Grained Access Control System for XML Documents" *ACM Transactions on Information and System Security (TISSEC)*, Vol. 5, No. 2, pp. 169-202, May 2002.
4. E. Damiani, S. De Capitani di Vimercati, E. Fernandez-Medina, and P. Samarati, "An Access Control Sytem for SVG Documents", *16-th IFIP Conference on Data and Application Security*, University of Cambridge, UK, July 2002.
5. S. De Capitani di Vimercati, "An Authorization Model for Temporal XML Documents"**,** *Proc. of the 17th ACM Symposium on Applied Computing,* Madrid, Spain, March 2002.
6. R. George, " GIS meets XML: SVG - Scalable Vector Graphics", http://www.web-maps.com/svg-gis.html, 2001.
7. S. Kakumoto, Y.Kosugi, M.Hatayama and H.Kameda, "Development of Spatial Temporal Geographic Information System", *Technical Report of the Geographical Survey Institute* / 1-No.275-2, March 2002.
8. N. Kodali and D.Wijeseker, "Regulating Access to SMIL formatted Pay-per-view Movies", *ACM Workshop on XML Security*, George Mason University, Fairfax VA, USA November 2002.
9. S. Osborn and Y. Guo, "Modeling Users in Role-Based Access Control" *Fifth ACM Workshop on Role-Based Access Control*, Berlin, July 2000.
10. P. van Oosterom. "The Reactive-tree: A Storage Structure for a Seamless, Scaleless Geographic Database", *In proceedings Auto-Carto 10*, Baltimore, March 1991.
11. Ph. Rigaux, M. Scholl, and A. Voisard, "Spatial Databases - with applications to GIS", Morgan Kaufmann, 2002.
12. P. Samarati and S. De Capitani di Vimercati, "Access Control: Policies, Models, and Mechanisms" in *Foundations of Security Analysis and Design*, R. Focardi and R. Gorrieri (eds), LNCS 2171, Springer-Verlag 2001.
13. DBx Geomatics, SVG mapping, http://www.dbxgeomatics.com/svg.asp
14. FGDC "Homeland Security & GIS" http://www.fgdc.gov/publications/homeland.htm
15. W3C, Scalable Vector Graphics, http://www.w3.org/Graphics/SVG/