Security and Privacy for Geospatial Data: Concepts and Research Directions

Inaugural Paper for the ACM SPRINGL Workshop

Elisa Bertino
Department of Computer Science
Purdue University, U.S.A.
bertino@cs.purdue.edu

Michael Gertz
Institute of Computer Science
University of Heidelberg, Germany
gertz@informatik.uni-heidelberg.de

Bhavani Thuraisingham
Department of Computer Science
University of Texas at Dallas, U.S.A.
Bhavani.thuraisingham@utdallas.edu

Maria Luisa Damiani
Dipartimento di Informatica e Comunicazione
Università degli Studi di Milano, Italy
damiani@dico.unimi.it

ABSTRACT

Geospatial data play a key role in a wide spectrum of critical data management applications, such as disaster and emergency management, environmental monitoring, land and city planning, and military operations, often requiring the coordination among diverse organizations, their data repositories, and users with different responsibilities. Although a variety of models and techniques are available to manage, access and share geospatial data, very little attention has been paid to addressing security concerns, such as access control, security and privacy policies, and the development of secure and in particular interoperable GIS applications. The objective of this paper is to discuss the technical challenges raised by the unique requirements of secure geospatial data management and to suggest a comprehensive framework for security and privacy for geospatial data and GIS. Such a framework is the first coherent architectural approach to the problem of security and privacy for geospatial data.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Spatial Databases and GIS*; K.6.5 [**Management of Computing and Information Systems**]: Security and Protection; K.4.1 [**Computers and Society**]: Public Policy Issues—*Privacy*

General Terms

Management, Security, Theory, Legal Aspects

Keywords

GIS, Geospatial Data, Security, Privacy

SPRINGL 2008 Irvine, CA, USA

1. INTRODUCTION

Advancements in sensor technology, satellite imagery, and field surveys have made it possible to generate and collect large amounts of geospatial data, at an ever increasing level of temporal coverage and spatial resolution. For example, thematic and topographical maps in support of disaster and emergency management, homeland security, and environmental crises provide geospatial data for various features of locations and facilities at very fine-grained levels of detail. These advancements have recently raised many data security, privacy, and safeguarding concerns, not only by the public but also by federal, state, and local government organizations, which are concerned that publicly available geospatial information can be exploited by attackers for corrupting critical infrastructures and compromising the security and privacy of people, property, and systems.

National and international efforts, such as the National Spatial Data Infrastructure [49], the Global Earth Observation System of Systems [8], and the Critical Infrastructure Protection Initiative [7], primarily focus on the coordinated development, use, sharing, and dissemination of geospatial data among the wide range of federal government agencies and offices. While consistent and reliable means to manage, share, and access geospatial data are available, very little attention has been given to addressing security concerns, such as access control, security and privacy policies, and the development of secure interoperable GIS applications. This is despite many reports, such as the RAND vulnerability report [15] and the NSDI Guidelines for Providing Appropriate Access to Geospatial Data in Response to Security Concerns [48], which identify the various potential risks that arise due to the production and (public) dissemination of geospatial information. These reports in particular call for mechanisms to safeguard the increasing amount of geospatial data.

To date, however, the problem of security and privacy for geospatial data and GIS has not been much investigated and there is not even a comprehensive understanding of all the issues that need to be addressed. The goal of this paper is to identify and explore such issues and develop a comprehensive framework for security and privacy for geospatial data and GIS.

The paper is organized as follows: Section 2 discusses

three real world scenarios; such scenarios provide concrete examples motivating the needs for research in the area of security and privacy for geospatial data and GIS. The examples also show that in most cases GIS and geospatial applications need interoperable access to heterogeneous data repositories, which poses interesting security challenges. Section 3 discusses security issues that are unique to geospatial data and argues that current security and privacy solutions developed for conventional data are not adequate. We then proposes a comprehensive framework that is the first coherent architectural approach to the problem of security and privacy for geospatial data. Any actual solution will likely implement most of the components of the suggested framework. Based on such a framework, in Section 6, the paper explores research issues by discussing in detail requirements, shortcomings of current approaches, and possible solutions. Relevant issues include access control models for geospatial data, privacy, and trust. Many relevant open issues are identified throughout the discussion. The paper concludes by outlining suitable evaluation metrics for the envisioned solutions and presenting concluding remarks.

2. CURRENT PRACTICES AND MOTIVATIONAL SCENARIOS

To better illustrate the diverse and complex security and privacy requirements of geospatial data repositories and applications, we outline some practically relevant GIS application scenarios increasingly complex in nature. The main theme of the scenarios is to demonstrate the need for a security and privacy management framework in support of GIS applications typically used in disaster and emergency preparation, mitigation, and response. The basis for our scenarios is a collection of GIS repositories available for Northern California, directly accessible through the California Spatial Information Library (CaSIL)[1].

As illustrated in Fig.1, for particular geographic regions, several thematic layers of geospatial data are available. A collection of layers is typically managed by a single federal, state, or local organization; there is no single organization that manages all geospatial data of interest for all regions. For example, water resource thematic layers, managed by an organization such as the U.S. Geological Survey (USGS), comprise information about streams, drainage areas, surface terrains, and rainfall responses for particular regions. Census unit and boundary layers manage information about administrative units (e.g., blocks and tracts), streets and addresses as well as municipal, county, state, and federal census boundaries.

A base layer can be either an image, i.e., field-based data in the form of satellite imagery or a digitized map, or some vector data that represent particular geographic features of interest, such as roads, buildings, and particular public and private areas. Vector data may include fine-grained type information, such as types of roads, buildings etc. In general, all data managed in a GIS repository is geo-referenced, thus providing an easy way to stack layers in order to obtain new (derived) GIS layers and data products. The hybrid-view feature in Web map servers such as Google Maps, Google Earth, NASA World Wind, and Yahoo! Maps is a prominent example of the overlay of satellite imagery with vector data, such as streets, boarders, and services.

[1]http://casil.ucdavis.edu or http://gis.ca.gov

Scenario I.

Assume a single GIS data repository that manages information about parcels (as basic units of geography for local government) and cadastre, including land use and zoning, environmental areas, and municipal utility services. Such type of repository is typically used by public sector staff at the municipal level to assist property owners and to support emergency, fire, and police operations. The latter type of usage includes identifying property structures and owners. Parcel maps in particular can be useful to do damage assessment after a disaster. They are also an important access point during emergencies for linking data from different GIS repositories. While such types of geospatial are used to serve the public, e.g., through Web-based interfaces, not all data layers are made publicly available. For example, fine-grained property owner demographic information is not publicly accessible. A similar separation of public and private GIS data can be made for other types of themes. For example, environmental theme layers do not make information about locations of endangered species or nesting sites public.

Analysis: Based on this type of separation of GIS data, the following question arises: *What security mechanisms are used to specify and enforce different types of access to data in a single GIS repository?* In particular, what provisions do GIS data managers have to (1) give public sector staff only access to GIS data relevant to their function, and (2) ensure that no sensitive geospatial data (e.g., parcel owner information) is made public?

In current GIS practice, access control is typically realized at the theme layer level, not taking into account complex geospatial features, the composition of features through theme overlays and in particular the overlay of satellite imagery with vector data. How can one *blend out* or *obfuscate* particular geospatial features at one theme layer so that sensitive features at another layer cannot be *derived* by a simple overlay? Regarding satellite imagery, they are provided by Google Maps at a resolution of a few inches for some U.S. states. That is, fine-grained satellite imagery and aerial photography is already publicly available. Therefore, while access control may not be an issue here, the more interesting security aspect is the study of potential privacy threats that may occur when high-resolution satellite imagery is combined with vector data layers, in particular when the vector data represent demographic (and possibly aggregated) data of people living in an area. Such aspects are key to geo-marketing. In general, privacy in GIS is not well understood (e.g., [31, 32, 50]), and thus requires a more thorough investigation in the context of disaster and emergency management.

Scenario II.

Suppose California is preparing for a potential natural disaster, say a major flooding in the Sacramento/San Joaquin area. Geospatial data is clearly key to disaster preparation and response, which is conducted by various county and municipal emergency service teams. Information from autonomous GIS repositories, each managing geospatial data for a particular thematic aspect such as hydrology, parcels, land use, health facilities etc., needs to be integrated to simulate events over time and to plan the coordination of emergency mitigation tasks. Simulations address aspects such as assessment of potential damages to critical utility and trans-

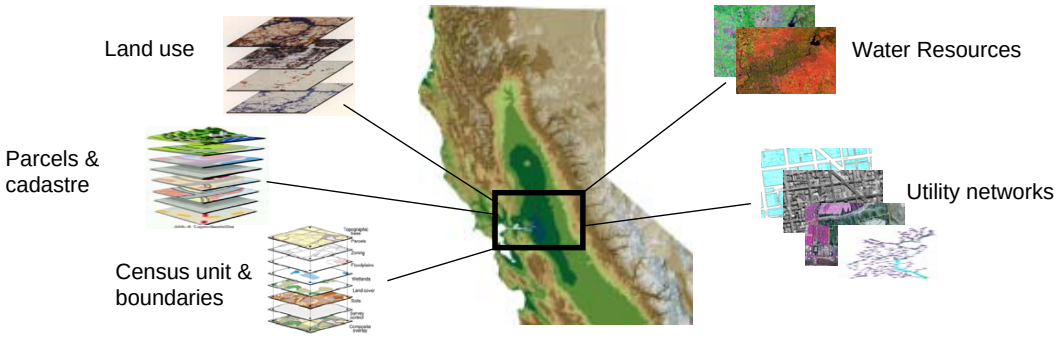

Figure 1: Example of several GIS repositories and GIS themes/layers for Northern California

portation infrastructures or planning for the evacuation of the public in endangered areas, to name a few.

Analysis: If the entire body of geospatial data would be made available by simply integrating the data from the different repositories, there is clearly a potential for data misuse and privacy violations. Sensitive information such as building ownerships might be revealed or information about critical infrastructure could become publicly accessible. Given the number of people and organizations involved in a disaster preparation scenario, security measures must be taken to provide users and applications only with data on a need-to-know basis. Building on the security and privacy requirements indicated in Scenario I, the question is how such security mechanisms, if present in individual should GIS, be extended to be applicable to interoperable GIS repositories and applications? Assuming individual GIS repositories have been equipped with access controls and security mechanisms, security policies now spanning several repositories need to be integrated and analyzed to discover potential security and privacy violations. Role-based access controls for interoperable GIS repositories seem a desirable mechanism to address these important security requirements. However, none of these provisions are made in such types of GIS applications, despite being strongly emphasized in the recent NSDI report [48]. A key question is *How to set up a secure interoperable GIS framework?*

Scenario III.

The above scenario presumably takes place in a static setting. That is, geographic data sets and associated security policies are integrated in a controlled and stepwise manner to satisfy the diverse information requirements of applications employed in the simulation. This changes once a disaster or emergency actually occurs. No simulation can anticipate all possible scenarios, required geographic data sets, and supporting GIS applications. Thus, in the event of an emergency, it is very likely that new data sets need to be integrated in an ad-hoc fashion, data requirements of applications change over time, GIS data sets need to be updated to account for changes in the real-world, and in particular access and privacy policies need to be adjusted.

Analysis: While geospatial data integration approaches likely suffice to integrate heterogeneous forms of geo-referenced data, it is not clear how security and privacy management should occur. Many security principles are often abandoned in favor of timely access to mission critical data. An important aspect often neglected in this approach is that a disaster causes changes to the transportation and utility infrastructure, building, geological properties of areas, and location of potentially large parts of the population in affected areas. A key question then is *How the geographic data used by GIS applications can be kept up-to-date, reliable, and trustworthy while still enforcing basic security and privacy needs?* Ideally, security policies and mechanisms should be more dynamic, context-dependent, and reactive. All this should happen with little intervention from GIS data managers and disaster response teams. For example, consider a team of first responders that reports a road being damaged. This information needs to be fed back into the interoperable GIS infrastructure in a trustworthy manner. Also, teams might require immediate access to previously (because of security and privacy policies) restricted information, for example, in case another team is requesting help in evacuating elderly people from some buildings. How can one still support appropriate levels of data privacy? How can one minimize the risk of misuse in such dynamic settings? Virtually no work has been carried out to address such pressing security and privacy needs. Respective security models, techniques, and architectures are not yet established for even individual GIS data repositories.

3. SECURITY ISSUES UNIQUE TO GEOSPATIAL DATA

Security issues for geospatial data are different and in many ways more complex than security issues for relational data. These differences concern both the data organization and structures, and in particular the ways the data are used. In a GIS, data is typically organized in different thematic layers; these layers, which can be large in number, represent different aspects of an application domain. Also, the same spatial region can be represented by either field-based data, i.e., satellite imagery or map data, or by vector-based data, i.e., a collection of possibly complex geographic features. Because of the organization in layers, the same geo-references feature, e.g., a building or road, can be represented in different layers and ways.

Geospatial data can be characterized as complex data objects with complex relationships among them. Securing this type of data poses challenges that are yet to be fully understood and addressed. For example, integrity techniques (such as Merkle hash tree) have been developed for simpler

data structures, but not for complex objects with complex relationships. Access control and privacy pose many issues, such as the unit of protection. Is such unit a layer, a portion of a layer, a geographic feature presented at a layer, a component of a feature, an application entity (independent from the layers in which it appears)? Are relationships also to be protected by access control? How does one specify and enforce content-based access control? The conventional view mechanisms developed for relational databases would not work here since diverse context information may not be readily accessible in a view. In terms of data usage, many applications generating and using geospatial data are dynamic as the set of subjects and geographic features may dynamically and rapidly change, as in the case of dynamic GIS coalitions for emergency response. Moreover, in such context, one may need to combine data from several sources that are independently administered and therefore characterized by heterogeneous security policies. Such usage requires different approaches to architecting the data security solutions. For example, one may need to perform dynamic security policy reconfiguration and merge possibly large numbers of heterogeneous security policies due to the complexity and diversity of geospatial features and operations on such features.

Several questions need answers in order to appropriately secure geospatial data. Since geospatial data come in diverse formats, such as thematic maps or satellite imagery, what is an appropriate data model underlying a security framework to specify semantic-rich security policies and to reason about such policies? What constitutes the notion of a geospatial feature, collections and compositions of features, and operations on these features? What types of security policies for geospatial data are necessary and how can such policies be composed in a meaningful and consistent manner? How do we integrate active aspects into policies, such as blending out or obfuscating particular features? How do reasoning techniques for security policies take advantage of topological and spatio-temporal properties of geospatial data? What constitutes a modular approach for a service-oriented security architecture to manage geospatial data, and how does such an infrastructure interact with diverse types of GIS data repositories and applications? How will such an infrastructure change the way practitioners and researchers manage the security and privacy of geospatial data? A better understanding of these issues will provide insights to design and implement modular, scalable, and service-oriented systems for the secure management of geospatial data.

4. RELATED WORK

There has been considerable work in building security infrastructures for relational data and, more recently, semi-structured data. Several access control models and security policy frameworks have been developed (see, e.g., [23, 24, 37, 36, 57]). Many of these techniques have been successfully deployed as part of database management systems. Relevant models also take into account organizational functions such as role-based access control (e.g., [19, 55, 56]) and credential-based access control [22, 61]. Techniques for composing and reasoning about security policies (e.g., [20, 39, 42, 53, 58]) as well as concepts for managing trust and privacy have been developed. All these developments have led to principles and paradigms for the secure management of data in the traditional realm of relational and semi-structured data.

One approach would be to adopt and extend such security models and techniques to geospatial data and GIS applications. However, the complexity and heterogeneity of geospatial data as well as the various non-traditional (compared to relational) operations on geospatial data pose several challenges, requiring research on the theory and the engineering of geospatial security models, techniques, and tools. Some work on securing geospatial data systems has been reported. However, existing access control models and techniques for geospatial data primarily consider only one type of data, either field-based data (e.g., [13, 14]) or feature-based data (e.g., [16, 21]). There is no framework that investigates the (dynamic) combination of field- and feature-based data, the required functionality of a uniform and comprehensive security and privacy policy specification language, implications on reasoning about policies, and realizing policies in practical geographical application settings, including security mechanisms for GIS, service-oriented architectures, and Web services. On the other hand, the emerging application and research areas of *spatially-aware* or *location-based access control models* [28, 41] do not address the issue of geospatial data protection, but rather focus on the use of position information for strong access control, which is a different objective.

5. A COMPREHENSIVE FRAMEWORK FOR GEOSPATIAL DATA SECURITY

The development of a comprehensive and coherent framework for managing the security of interoperable GIS data repositories and applications (see Fig. 2) requires investigating several issues and novel specialized tools, including the following.

Security Policy Specification and Reasoning. Most geospatial applications require fine-grained and flexible access to geospatial data. Building ad-hoc data sets for each type of access and application is not suitable when the user community is large and dynamic, and access control policies change in the case of certain events, such as those dealt with in emergency situations. Relevant components of a security policy specification and reasoning framework include:

- A comprehensive geospatial data model that is able to express different types of geospatial and spatio-temporal data (geographic features and field-based data), and that provides a rich set of typical operations on geospatial data (image operations and spatial transforms).
- A security policy specification tool supporting multiple geospatial data representations and granularities, feature and policy grouping mechanisms, diverse data access and modification rights, active policies, and dynamic application contexts and user populations.
- A security policy reasoning tool able to determine inconsistent and redundant policies at policy compile time and/or data access time. One approach toward the development of such tool is to extend existing logic-based reasoning approaches to incorporate specifics of geospatial data, such as topological and temporal properties.
- Access control modules organized as service to GIS repositories, applications, and Web services, to verify

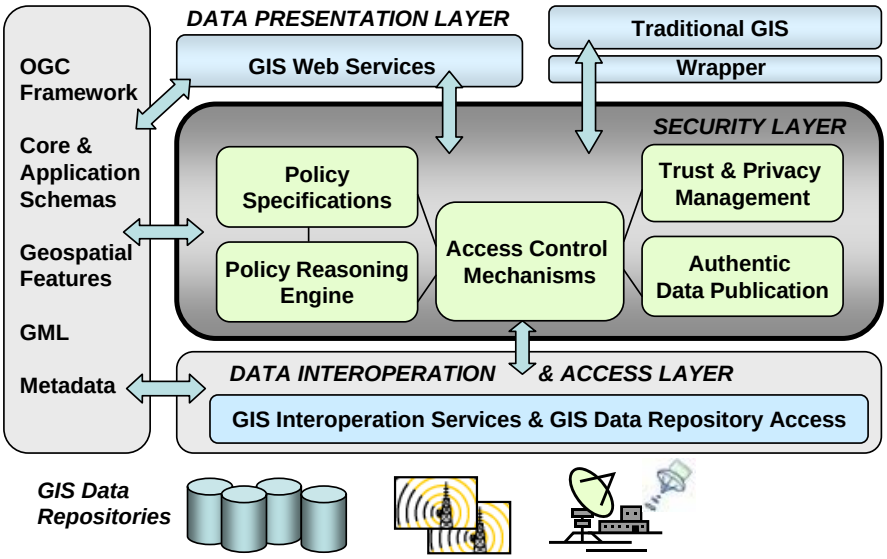

Figure 2: Components of the proposed security and privacy framework

accesses to and operations on (possibly distributed and heterogeneous) GIS repositories.

Interoperability of Security Policies. Many emerging GIS applications and services rely on standards developed by the Open Geospatial Consortium (OGC) [3]. In order to provide a feasible security and privacy approach that can be applied in practical settings and real-world applications, one should:

- Incorporate components of the adopted geospatial data model and security policy specification language into standards such as the Geography Markup Language (GML) [9] based application schemas, Web Feature/ Coverage Services, and Web Map Services.
- Develop mapping techniques specialized for the integration and translation of security policies and among heterogeneous and distributed GIS repositories.

Trust, Privacy and Integrity. Data privacy and trust play a crucial role in GIS applications that deal with private and mission critical data. It is important to have mechanisms that detect tampering with the data or the release of private data, thus undermining their integrity and privacy. To address such concerns, services need to be developed that provide a modular extension of the policy specification and reasoning framework. More specifically, activities should:

- Design and implement a trust management component to enforce trust-based policies that describe what level of trust should be placed on users and geospatial data sets in dynamic and context-based scenarios, in particular those in the context of dynamic GIS repository coalitions.
- Develop a privacy-centric security policy and enforcement component that will handle both trusted and untrusted GIS applications that require exact geospatial data for a service.
- Develop mechanisms and techniques that allow users to verify the trustworthiness of geospatial data through authentic data publication schemes, ideally in combination with GIS Web Services.

6. RESEARCH DIRECTIONS

We now discuss research directions concerning individual components of the security framework proposed in the previous section, and we illustrate how to integrate them into a scalable, coherent and flexible architecture. For each of the components, we outline some of the requirements, current state-of-the-art approaches and their limitations, and some possible approaches.

6.1 Geospatial Data Model: Conceptual Framework and Realization

Requirements. A fundamental requirement for the specification of security and privacy policies for different representations of geospatial data and for reasoning about such policies is a geospatial data model (GDM). Such a model defines different types of spatially referenced objects and a precise semantics for operations on these objects. In particular, it provides the basic vocabulary for specifying security policies.

Current approaches and limitations. Existing GDMs either deal only with feature-based (also known as object-based) data, such as vector data, maps, or thematic layers [52], or field-based data, such as satellite imagery and aerial photography [40, 45]. There is no model that combines feature-based and field-based data in a single coherent and semantic rich framework as basis for the security requirements outlined in Section 5. What does it mean to discuss geospatial features in a layer that is a combination of a satellite imagery overlaid with some vector data? A comprehensive GDM is crucial to specify meaningful security policies for geospatial data and to reason about such poli-

cies. For example, if field-based data can easily be overlaid with vector data in an application but there is no provision in terms of security policy and access control for restricting such operations, sensitive information might be revealed to users, thus potentially violating privacy policies. These types of operations are typically supported by most GIS and Web Services, and several security threat scenarios are conceivable that cannot be dealt with using existing technology and security mechanisms.

Approach: The Core Geospatial Data Model. In order to address the pressing needs for a comprehensive GDM as basis for our proposed geospatial security framework, a suitable approach is to extend the Layered Spatial Data Model (LSDM) [17], which is a reference framework for multiple representations of geographic maps and supports point, line, and region data. To be viable for security and privacy, such an extension, that we refer to as Generalized LSDM, has to integrate field-based data into LSDM and provide a rich set of operations on feature- and field-based data as well as combinations thereof. Fig. 3 illustrates the layering. In addition, such a model should support temporal features (time-stamped data, events, and moving objects), feature composition, grouping of features - the latter two aspects being an important functionality for specifying security policies at different levels of object granularity. Operations on features in GLSDM should include: feature selections (based on non-spatial feature properties), spatial and temporal selections, spatial transforms for both feature- and field-based data (map and spatial re-projections, zooming, general affine transformations [34, 45], and map algebra operations [59], neighborhood operations, traditional GIS theme operations (theme selection, overlay, union, merger), and insertions, deletions, and modifications of features.

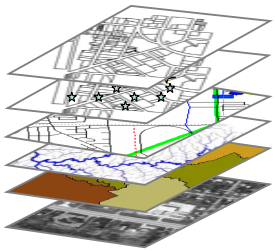

Figure 3: GIS data layers (themes)

Conceptually, GLSDM can be designed as a logic-based language, primarily to allow for effective policy reasoning techniques and tools. That is, a logic-based language can be used to describe geospatial features and operations on the features in the context of the security layer (see Fig. 2). In such context, the OGC framework and the Geographic Markup Language (GML) [6, 43] can be used to draw feature types and operations on (collections of) features from GML core and application schemas. These schemas serve as main vocabularies to specify security and privacy policies. GML offers several components for modeling geometry, topology, temporal elements, dynamic features, and coordinate reference systems. These core components thus need to be mapped into GLSDM; approaches to perform a one-to-one mapping between a GML application schema and an instance of GLSDM thus need to be investigated. Furthermore, in order to integrate field-based data into such a mapping, one possibility is to investigate the recently proposed JPEG 2000 standard[2], which can be used as reference format for satellite imagery and aerial photography. In fact, OGC recently initiated the support of GML metadata encoding in JPEG 2000 Image files, thus providing the proposed framework and model with a coherent framework to draw upon both feature- and field-based data and object types. Thus, by mapping GML features and metadata specifications into GLSDM, one can describe security policies that are tightly integrated with GML application schemas and services using these schemas, such as Web Feature Services, Web Coverage Services, and Web Map Services.

6.2 Policy Specification and Reasoning Framework

Requirements. The development of a policy specification and reasoning framework for geospatial data and applications poses a number of challenging requirements, arising from the richness and multiplicity of data representations, the dynamic and mobile user populations, and dynamic application contexts as illustrated in the scenarios. Relevant requirements include:

(i) the specification of authorizations against geospatial objects at flexible granularity levels, and for the various spatial representations, object dimensions, and resolution;

(ii) a variety of object grouping mechanisms, based not only on feature types, but also on object content and metadata, and spatial position;

(iii) a variety of access rights corresponding to operations that can be executed on data;

(iv) mechanisms for dynamically generating modified representations of protection objects, through, for example, obfuscation and blurring, depending of the privileges of the subject accessing the objects and the contents of the objects; we refer to such mechanisms as *active authorizations*;

(v) attribute-based and profile-based user specification and authorization;

(vi) event-based activation/deactivation of policies.

Current approaches and limitations. Despite the relevance of access control in the context of high-assurance security for geospatial data, no comprehensive approach exists addressing these requirements. Current approaches can be categorized in two broad classes, based on whether they are more focused on geospatial representation of data, as in spatial data infrastructures [14, 16, 21, 46], or user mobility, as in location-based services [25, 27, 38, 41]. These approaches suffer from several shortcomings, such as lack of integration with current standards for geospatial data, limited set of access rights, limited support in access control rules for the rich and complex structures, no support for dynamic policies, active authorization and modularization, and no reasoning mechanisms. Notably the GeoXACML model [46] has recently acquired the status of an OGC standard [1]. Yet the

[2]http://www.jpeg.org/jpeg2000/

model only provides some basic functionality for integrating spatial data and functions into XACML.

Approach: The policy model and language. The core of any suitable approach to security for geospatial data is represented by a semantically rich logical model for access control; this model is the basis on which access control decisions are made and reasoning is performed. The core model, however, is not intended to be used by users or applications and a concrete syntax of the model would also need to be developed based on GML and other relevant standards, such as XACML. The policy model must support a broad spectrum of dimensions in access control policies, including:

- *Deny/allow policies with flexible granularity, grouping mechanisms for protected objects, and space-related access restrictions.* Deny/allow policies can be supported through the use of positive/negative authorizations; negative authorizations are crucial in order to support exceptions by which, for example, an authorization is assigned to all objects in a set but one. In our context, this paradigm is complicated by the larger options that we provide for denoting protected objects and by the presence of different object representations and dimensions. A possible main mechanism that we propose to support flexible grouping is based on the notions of object-locator and spatial window. An object-locator is a query expression that may include predicates against properties of feature types, metadata and provenance data. Predicates may also refer to topological relationships holding among the data objects, such as *Within* and *Touches*. An example of a policy using Touches is the one allowing a subject, which has access to information on a particular land parcel, to access information about all adjacent land parcels. The query expression may also include a projection component to specify an object representation and components. A spatial window is simply a spatial region in the reference space and denotes the set of objects that are inside the boundary of the region. By combining two such mechanisms, one can specify sets of objects such as *all shelters occupying an area greater than 3000sf in Yolo County*; in this case *Yolo County* represents the authorization window. The use of spatial windows is particularly important to give/deny authorizations to insert or modify features in a given spatial region.

- *Active policies.* These are policies that, when applied to a protected object, perform certain transformations on the object before returning it to the requester. Two relevant classes are filtering policies and obfuscating policies. Filtering policies refer to policies that filter out some portions of the objects before returning them to the users. These policies are directly supported by the suggested object locator mechanisms. Obfuscating policies act like filter policies except that they do not simply select objects but perform possibly complex computations on the feature(s) to be returned. Typical examples include computing a lower resolution image, and distorting some vector data (but preserving topological relationships). In our proposed model these policies are supported by the projection component, suitably extended with the possibility of invoking functions of the object locator. We expect that such functions can be provided as part of a pre-defined library.

- *Context-dependent access control policies.* Under such policies, information from the environment is taken into account by the access control module when taking decisions about access requests. Typical contextual information includes time and subject location. Subject location information is used to specify policies allowing a subject to access a resource only if the current location of the subject verifies certain spatial constraints. Context-dependent access policies can be supported by the introduction of a context component, as part of authorization rules, and by attribute-based specification of subjects in authorization rules.

- *Event-based access control policies.* Event-based access control policies are novel and are based on the idea that policies can be enabled/disabled depending on the occurrence of specified events. Events can include data modifications, very much like in database triggers, or application-dependent events, such as an emergency. We notice that current sensor networks and intelligent appliances make it very easy for a computer system to detect events arising in the environments. It is important that such capabilities be exploited as part of a model like the one envisioned here.

Authorization Model. The basic element of the envisioned logical access control model is represented by the notion of authorization, which is a 7-tuple of the form

$$\{Ev, Co, Sj, Op, Ol, Sw, Ef\}$$

whose components represent the **Event (Ev)**, the **Context (Co)**, the **Subject (Sj)**, the **Operation (Op)**, the **Object Locator (Ol)**, the **Spatial Window (Sw)**, and the **Effect (Ef)**.

The **Effect** component specifies the intended effect of the authorization, that is, **Permit** to give access, **Deny** to deny access, **Partial Permit** to give access under the condition that the object is modified according to the functions specified in the *Object Locator* component of the authorization. Authorizations having as effect Partial Permit are active authorizations in that they dynamically modify the objects returned to the application. Different active authorizations can be specified for the same protected objects and different subjects, thus allowing different subjects to see different representations of the protected objects, according to their need to know. The **Object Locator** component of an authorization thus consists of the specification of a protected object, or set of protected objects, according to the GLSDM language (or to the language used for spatial data representation), and of an optional transformation function.

In addition to what we have already discussed, it is important to point out that the **Subject** can be not only a user id or a (possibly spatial) role, but also a Boolean combination of predicates against subject attributes; in such case, the authorization applies to all users verifying such Boolean combination. The **Operation** component of the authorization specifies the type of operation that can be executed on the protected object; such specification may also include a dimension specification to denote that the operation has to be restricted only to the specified dimension. For example, a user can view objects of feature type residential buildings in

maps where they appear but only with dimension 0 (these objects will be indicated as points in the maps that such users will view). It is important to note that not all operations can be applied to all objects. Therefore, the envisioned authorization model needs to be complemented by a set of conditions ensuring that each authorization be consistent; we refer to this form of consistency as *well-formedness*.

The authorization model needs also to include some derivation rules supporting the automatic inference of additional authorizations, referred to as *derived authorizations*, from other authorizations. Derivation rules can be further categorized as: *structural derivation rules*, defined over object relationships, and *application-defined derivation rules*. The set of authorizations and derivation rules comprises the *geospatial authorization base*, whereas the set of authorizations and all authorizations that can be derived from this set according to the derivation rules comprises the *entailed geospatial authorization base*. The mechanism supporting the various consistency checks, such as well-formedness and authorization derivation, is referred to as *policy reasoner*. An important question is related to the consistency of the geospatial authorization base and the entailed authorization base; because of negative authorizations and the fact that different administrative authorities may exist, such bases may contain conflicting authorizations. It is, however, important to notice that the notion of consistency, due to the various representations of objects and conflict resolution techniques for geospatial authorizations, may be far more complex than in the case of conventional authorizations. For example, solving a conflict between two authorizations may require modifying, through filtering/obfuscating techniques, the representation of the objects being accessed.

Reasoning. A spatial policy reasoner typically exploits techniques proposed for spatial reasoning, such as reasoning on direction relations [51] and on topological relations [18]. Such techniques exploit properties typical of spatial objects, for example, if a certain direction relation holds among two regions, then it holds for all the points inside of these regions. This type of knowledge could be expressed as structural derivation rules defined as part of the GLSDM language. In order to provide support for additional application-dependent knowledge, it is however crucial to identity a general knowledge representation tool and develop a mapping from the knowledge encoded in terms of the GLSDM structural derivation rules onto the language of the identified knowledge representation tool. Suitable candidates for such tools are those based on descriptions in that these logics directly support the representation of complex objects and would directly support features that have complex structures.

The reasoner is crucial not only for checking the consistency of authorizations, but also to prevent possible inferences, which is important in order to achieve strong protection. In particular when dealing with diverse data sets, describing different aspects of the same set of entities, controlling access to one data set only is not enough, in that by accessing information from the other datasets, a subject may still be able to infer the content of the protected dataset. As an example suppose that there are small airfields inside a forest; suppose that a subject is given access to a map of the area in order to see which are the available roads but he is not supposed the know the location of these airfields; then one can provide him with a view of this map where the airfields are not shown. Suppose now that a thematic map reporting the types of trees in the area is available. This is useful in the context of fire emergencies. If this map shows that the area covered by the forest has large holes in which there are no trees, the subject may infer the presence of the airfields and may easily locate them. This means that the thematic map should not show those holes. Similar examples can be found concerning underground objects. In order to detect such possible inferences, an approach is to introduce the notion of strongly protected objects; protecting these objects would require making sure that all their features be hidden from all possible layers in which the objects may appear. It would also require determining whether their presence can be inferred from the presence or absence of other spatial objects; in our example, the presence of an area without trees inside a forest may denote the presence of an airfield in this area. Being able to support such types of inference requires encoding domain-dependent knowledge, in addition to basic knowledge about direction and topological relations. This domain-dependent knowledge can be typically encoded as a set of application-defined derivation rules, which is part of the proposed framework.

The authorization model suggested in this paper also needs to include a suitable administration model, supported by a high level tool, stating which subjects can grant, revoke, and modify authorizations. Because geospatial applications may consist of diverse data and support a variety of applications, a suitable administration model should support multiple administration authorities. In some cases, for a user to access a complex geospatial object, several authorizations by possibly different authorities may be required. Mechanisms are needed to streamline the authorization acquisition process by the users, without undermining the autonomy of the various administration authorities involved.

Architectural Organization. The proposed logical model can be represented as a GML application schema. An important question is whether the GML access control application schema must be integrated with the application schema of the actual application or can the two schemas be kept separated. In general, the latter approach has advantages, such as modularity. However, because the proposed access control system will support fine access control granularities, authorizations may directly reference specific objects. An important open issue is *How to support object references across different GML application schemas?*. A respective architecture should include a tool for authorization administration, the reasoner, and functions related to authorization enforcement.

A possible approach to integrate the access control system with other components of the data management infrastructure is based on an extension of the conceptual approach defined in the context of the XACML standard. Such an approach decomposes the functions related to the management of access control into four components: *Policy Enforcement Point* (PEP) that receives the access request and returns authorization decisions to the external environment; *Policy Decision Point* (PDP) that evaluates the applicable policies and renders an authorization decision to the PEP; the *Policy Information Point* that serves as source of data required for policy evaluation, for example location of users; the *Policy Administration Point* that supports the policy definitions and stores them in the appropriate repository. All the data required by these components, such as authorizations, can be represented according to the GML access control applica-

tion schema. A major difference with respect to the case of conventional applications for which XACML was designed is that in geospatial applications objects and subjects may be mobile and data and applications have strong integrity and privacy requirements; access control may thus have to be performed not only before the access takes place but also while the access is being executed and possibly after it has been completed. This calls for a continuous interaction of the PEP with rest of the data management infrastructure and requires the PDP to keep track how the status of the execution of the various access requests. Any suitable solution thus requires revisiting and extending the conceptual architecture of XACML.

6.3 Policy Interoperability

Requirements. As illustrated in the scenarios, geospatial data from heterogeneous sources often have to be integrated in a secure fashion. Industry and federal geospatial clearinghouses are collaboratively standardizing the geospatial data integration process (see [2, 5, 33]). This process does not address security. While there are many aspects to integrating heterogeneous GIS repositories, in the following, we will focus on policy integration [11, 12]. Since the individual agencies implement their own security policies to protect the data, some issues arise during policy integration.

The first issue is the mismatch of policy rule semantics. When policies are integrated, attributes and targets of the policies should be interpreted consistently. For example, if two policies from separate agencies use *manager* and *supervisor* respectively, to specify the same role attribute, the integration algorithm should interpret this equivalence.

The second issue is rules mismatch where attribute set and targets of separate policies have to be matched properly. Note that in our envisioned 7-tuple logical access model, the attributes are represented by the subset $\{Ev, Co, Sj\}$ (i.e., $Effect, Contex, Subject\}$ and targets are specified by $\{Ol, Sw\}$ (i.e., $\{ObjectLocator, SpatialWindow\}$.

Current approaches and limitations. Current policy integration algorithms (e.g., [47, 44]) attempt to solve the rules mismatch problem. However, there are several shortcomings in these approaches with respect to geospatial data integration. First, they require prior coordination among the agencies. Geospatial integrated environments are usually based on autonomous agencies, which manage their security policies independently of each other. As a result, prior coordination to integrate security policies may not be feasible. Second, existing models assume that target resources (i.e., $\{Ol, Sw\}$) are the same for the policies to be integrated. However, one or more subjects in the $\{Sj\}$ element can request a hybrid-view of some data where the target from one agency is field-based and the other is vector-based.

An approach: Geo-referencing. A suitable integration and interoperation approach is based on the notion of geo-referencing, in that it is often the case that geospatial datasets to be integrated contain data items that are geo-referenced with respect to a coordinate system. This way one can tie differing targets from separate policies using $\{Ol, Sw\}$. In the hybrid-view raster/vector example, if a subject does not have access to part of the vector dataset, the coordinates representing the restricted part allows the integration algorithm to prevent access to corresponding map areas. A possible approach is based on an algorithm consisting of two stages: disambiguating naming heterogeneity and resolving policy mismatch.

Naming heterogeneity can be addressed by utilizing a common vocabulary that will include rules such as *administrator and manager are different roles*. This vocabulary may be generated from the GML schemas and the GIS repositories (see Fig. 2). Policy mismatch heterogeneity problem can be addressed by measuring rule similarity and policy unification. We illustrate the envisioned approach using the following example.

Consider two geospatial repositories A and B and user John who is associated with B. A carries raster data whereas B contains thematic data about census units for a particular city. Suppose A and B enforce respective policies $P1$ and $P2$:

P1: $IF\ \{Subject\}\ !=\ Admin\ \&\ Event(Resolution < 6)\ \&$
$Event(Alert\_level < 5)\ \&$
$Context(Location(33.41\ \ -82.91,\ \ 33.43\ \ -82.904))$
$THEN$ Effect=Deny on
SpatialWindow(RegionA(33.40 −82.91, 33.43 −82.9101,
33.4301 −82.9103, 33.43 −82.9103))

P2: $IF\ \{Subject\} = Manager\ \&\ Event(Resolution < 5)\ \&$
$Event(Alert\_level < 4)\ \&$
$Context(Location(33.39\ -82.91,\ \ 33.43\ -82.904))$
$THEN\ \ Effect = Allow\ on$
$ObjectLocator(All\_Features\_in\_Zip(79900))$

$P1$ specifies that only an Administrator with a certain resolution at or above specific alert level can access the map region represented by an array of latitude/longitude coordinate pairs. $P2$ specifies that if a manager is at a specific location, she can access all census data including highways, streets and addresses for zip code 79900. Now, if we match these two policies, we will observe two types of heterogeneity. The first one is naming heterogeneity. In P1, the Subject value (i.e., Admin) is different from P2's Subject (i.e., Manager). Once we address this naming heterogeneity, rule mismatch heterogeneity may occur. Here, the second attribute of the rule in $P1$ (i.e., Event) is different from $P2s$ attributes.

Disambiguating naming heterogeneity. Naming heterogeneity in policies can be resolved by exploiting a concept-based model using domain-dependent common vocabularies (from GML schemas and repositories) that may be part of GIS repositories. In our envisioned approach, for {**Subject**}, we define a set of abstract roles with various high and low level concepts that will help to resolve naming heterogeneity. Therefore, if there is an ambiguity in **{Subject}**, it can be resolved through the correct interpretation provided by the vocabulary based on semantic distance. Examples of **{Subject}** concepts on the first level are *Admin*, *Manager*, *Facility Personnel* etc. We assume agency B creates a role called *Emergency Operation Commander* (EOC) by taking the union of *Manager* and *Facility Personnel* and assigns it to John. While processing $P1$ and $P2$, a vocabulary manager would detect that EOC is made up of roles that are siblings of Admin. Therefore, the condition {**Subject**}! = *Admin* in $P1$ is satisfied for John. In other cases, a new role such as EOC could be made up of concepts below the Admin level, and we then consider the separation between the roles to resolve naming heterogeneity.

Resolving policy rule mismatches. To address policy

rule mismatch, one possible approach is to first measure the similarity of rules across policies and next to apply policy unification by aligning similar policies.

Policy Similarity Evaluation: Once the semantic heterogeneity of attribute terms (e.g., subjects) across policies is resolved, policies have to be aligned. We consider each 7-tuple entry as a rule and we calculate the similarity of the rules across policies. As a result, some rule may be extended, restricted, etc. by other rules. Consider the above policies, $P1$ and $P2$ consisting of rules $r1$ and $r2$, respectively. $P1$ belongs to the resource holder and $P2$ belongs to the agency the requester belongs to. We illustrate the integration of $P1$ and $P2$ with geospatial data. The integration cases reflect the unique characteristics for geospatial data discussed in section 1. For example, rules have attributes that correspond to geospatial objects such as features or layers. There can be one attribute whose value represents a bridge feature, while another attribute value represents the center-line of the bridge. Although the attributes are separately specified, from their geo-references we can determine their geospatial relationship. This helps us to determine which integration case the attributes falls under. We have identified five integration cases including:

Case C1 : If attribute values of $r1$ are matched by $r2$ and $r1 > r2$ ($r1$ has more attributes than $r2$) then $r1$ is said to *extend* $r2$. If the attributes of $r1$ and $r2$ are different, then $r1$ and $r2$ are said to *diverge*.

Case C2 : If only some of the attributes of $r1$ and $r2$ match, but not all, then $r1$ and $r2$ are said to *intersect*.

In $P1$ and $P2$, the attributes are as follows: a $\{Subject\}$, two *events* and a $\{Context\}$. The *resolution* event refers to a geospatial constraint on a particular map or image. The $\{Context\}$ as defined in Section 6.2 introduces an extent constraint by specifying the location coordinates that a user request has to satisfy. Even if $P1$ and $P2$ have same set of policies, the attribute values are different. Suppose $P1$ allows a higher resolution than $P2$ and $P1$'s $\{Context\}$ contains *P2*'s $\{Context\}$ location. To illustrate the use of the integration cases we note that the values for two event attributes in $P1$ subsume those in $P2$. Also, since the geospatial extent of $\{Context\}$ in $P1$ confines that in $P2$, we say $P1$ extends $P2$, which falls under case $C1$. If $\{Context\}$ of $P1$ were contained in $P2s$ $\{Context\}$, they would have intersected (i.e., case *C2*). Therefore, geospatial data determines the specific case a rule mismatch resolution algorithm has to follow.

Policy Unification: We factor in the policy combination algorithms. We consider action/consequent of similar rules and exploit deny-override and permit-override strategies. The deny or permit is referred to as policy {**Effect**} that implies the final decision on the rule. Since there could be multiple policies protecting a particular target (i.e., map resource), one could run into a conflicting scenario where some effects correspond to deny and others to permit. At this stage, anomaly/conflict in policies will be detected. Manually detecting and resolving these anomalies is a critical but tedious and error prone task. For firewall policy rules, prior research has focused on the analysis and detection of anomalies. For example, possible relations between rules as well as anomalies are defined in terms of the relations and algorithms to detect the anomalies by analyzing the rules have been given. Approaches proposed for the analysis of firewall rules [10] can be extended to support to account for aspects specific to geospatial data.

6.4 Trust and Privacy Management

As illustrated in Fig. 2, the suggested security framework also includes components for trust and privacy management, because these are important goals of any data security solution.

Requirements. The trust and privacy problems in geospatial databases require comprehensive solutions that can handle the potential richness of the data types, nature of the dynamic access and mobile user populations. Since geospatial data is used for critical applications such as emergency response, it is important that users can trust the retrieved geospatial data. On the other hand, the unauthorized access and misuse of geospatial data can lead to privacy violations. In order to address these challenges, systems are needed with the following capabilities: (i) mechanisms for dynamic verification of the source of the data, and (ii) privacy policies that address the dynamic and complex nature of potential GIS applications as illustrated in the scenarios.

Current approaches and limitations. In terms of trust, current XML standards provide mechanisms for attaching digital signatures to the documents. Unfortunately, current signature standards do not provide semantics to make trust decisions given the digital signatures [4]. Also, the new JPEG standard provides mechanisms for attaching digital signatures to the images but this may not be directly used for supporting different geospatial data types. Regarding privacy, although approaches have been proposed to enforce privacy policies (e.g., [26]), there is no comprehensive solution covering all types of geospatial data and their use. More importantly, an understanding of privacy in geospatial data is needed before we propose solutions.

Proposed approach I: Trust management. Maps, satellite images and other geospatial data used during an emergency must be trusted. In other words, the users must be confident about the freshness and the correctness of the geospatial data. Even worse, in some cases, an adversary may try to deliberately submit incorrect data to the system. For example, a developer may want to modify demographic information related to some county to coerce city planers into giving more construction permits. To protect against such malicious intent or honest mistakes, we propose to add two components to the suggested security framework. The first component will provide services to maintain logs to verify the source and the submission date of the geospatial data and other provenance data. In a typical database system, the database may create and store such logs. In this framework, the geospatial data may need to be combined on demand by using various sources. Therefore, logs maintained by a single geospatial database may not be sufficient. Instead, an alternative approach is to extend LSDM, or whichever model is used to represent the spatial data of interest, to automatically incorporate the necessary metadata and semantics to reason about the trust associated with metadata and XML digital signatures. For example, an emergency responder who reaches an accident scene first, may update some geospatial information about the roads. If another team approaching the scene requires the road information, the system must ensure that this team receives the

latest information. By adding necessary digital signatures and attributes about the trusted system logs, the second team may verify the source, freshness and correctness of data.

After the verification, trust must be evaluated. For example, trust could be calculated by taking into account the past feedback related to the source (i.e., how accurate was the geospatial data provided in the past), the creation time (i.e., how fresh is the geospatial data) and the contents of the data (i.e., how likely is it that someone may tamper with the geospatial data). The applicability of different trust calculation techniques for GIS applications [60] must certainly be investigated. After the trust calculation, based on the estimated trust value, the system typically needs to validate the correctness of the data through other sources. Since geospatial data such as satellite images and maps may be retrieved through other sources, one has to exploit the availability of such sources and specify verification policies based on the underlying trust of the data. If the verification time is longer than what can be tolerated by an emergency, first responders may be better off with no validation. Cost effective validation procedure for critical geospatial data must be investigated to take such conflicting requirements into account.

Proposed approach II: Privacy management. As discussed in the scenarios, geospatial repositories are used to store high resolution satellite images, the location of individuals, their demographic information, resulting in privacy concerns. However, the definition of privacy in the context of geospatial data is highly subjective [62]. For example, a person may not want the image of his house accessible. However, these images are already publicly available. Therefore, unless effective laws are enacted, this situation cannot be a privacy concern. Due to this subjectivity, it is crucial to first develop a better understanding of privacy issues in geospatial data based for example on the scenarios described in Section 2 as well as reviewing various privacy related documents for geospatial data. Next one has to explore how to implement the common requirements from different privacy definitions. These requirements may include: collecting only the necessary information, giving control over information about one's self and preventing misuse of collected data [62].

To address the *collecting only the necessary information requirement*, a possible approach is to develop filters for different geospatial data types. For example, to store high resolution satellite images for city planning purposes, one should not store the portions of the images that accurately show the communication infrastructure on top of buildings as such an infrastructure could be used to identify say military buildings. Filtering tools need to be developed that automatically remove security and privacy sensitive geospatial data.

To address the *giving control over information about one's self* requirement, an approach is to extend the envisioned access control tool to support privacy policies also. Each individual, when providing privacy sensitive geospatial data, may be notified about the privacy policies of the GIS data repository. Such preferences specify who can access the data under what conditions and for what purposes. Suppose that an elderly person enrolls in a location-based geospatial service that can send an ambulance automatically to the location of the individual during an emergency. Therefore, privacy policy for such a service can state that in case of an emergency, only first responders can access his address information for sending help. Clearly, this approach assumes that the user trusts the system with respect to policy enforcement.

To address the *preventing misuse of collected data*, a possible approach is to develop techniques to enforce privacy policies. For example, for emergency planning in the case of an E.coli outbreak in a city water system, geospatial data related to water pipe system could be combined with demographic and health-care related data. Such integration of different geospatial data sources could create privacy problems. First, we should enforce the privacy policies of different sites during the integration. If a GIS repository has a privacy policy that states that individual's exact addresses can be only released during an emergency, then for planning purposes, we may need to use some carefully aggregated demographic data. Note that different sites may enforce different privacy policies. Therefore, we need to integrate these policies using the integration techniques discussed in Section 6.3. In addition, access to certain combinations of geospatial data may be restricted. For example, an individual may allow his demographic and health-care data to be accessed separately for city planning purposes. However, the same individual may not allow his health care and demographic data to be combined unless it is an emergency. This requires extending the privacy policies to include separation of duties for different geospatial data types. In some situations, queried geospatial data may need to be altered for preventing privacy/security violations. For example, a city planning expert may be given access to high resolution satellite images that do not show say the military facilities. Therefore, we need to filter out the parts of the satellite images that have the sensitive data. Such filtering of geospatial data should be carried out *on the fly* based on the privacy or security requirements.

6.5 Integrity: Authentic Data Publication Schemes for Geospatial Data

Requirements. Often state/county organizations employ third-party data publishers to provide GIS data consumers with access to large amounts of selected geospatial data sets and products collected by the organization (e.g., the CaSIL repository mentioned in Section 2 is such a third-party service). These publishers receive periodic copies and/or updates of the organization's GIS data sets and answer queries from GIS applications and Web services on behalf of the data owner. Publishers relieve owners from maintaining a secure, fault-tolerant and often expensive data management infrastructure, an important aspect of mission critical geospatial data that need to be replicated.

Current approaches and limitations. For authentic data publications, much of the work has focused on relational data and XML data. These data publication schemes allow data consumers to efficiently verify the correctness and completeness of the data provided by the data publisher (see, e.g., [22, 29, 30, 35, 63]).

That is, data owners can verify whether the query result provided by the publisher is the same as the data owner would have provided. These approaches are based on hashing schemes over the owner's data to compute a digital signature to be (periodically) distributed to data consumers. Existing approaches have limitations, because they do not

take into consideration the complexity of GIS data.

Proposed approach: Authentication schemes. A possible approach is to develop authentic data publication schemes for diverse types of geospatial data. These schemes should allow data consumers to verify the correctness and completeness of both geographic features (e.g., vector data describing thematic maps) and field-based data (e.g., satellite imagery and aerial photography). The key challenge will be to develop space efficient hash structures that allow for small signature objects to be computed by the owner and publisher, and to be evaluated by the data consumer. For vector data, a possible approach to develop and evaluate hashing schemes on commonly used spatial index structures, such as R-trees that index point, line, and region data (see, e.g., [52]). For image data in GeoTIFF, JPEG, or PNG formats, one should investigate Quadtree-like structures [54] in combination with digital signature schemes and the Digital Rights Management (GeoDRM) framework developed by OGC. In particular, compression schemes for raster image data in combination with Quadtrees seem to be good candidates for hashing schemes that allow scaling in terms of precision (e.g., based on the spatial resolution). An authentic data publication module should be developed that allows GIS Web Service to request verification objects in addition to the geospatial data results, provided that an owner/publisher setting has been established at the data access layer. A challenge here is a suitable extension of GML/GIS Web Service components to embed respective protocol information that allows the immediate verification of query results by the application and ensures the maintenance of digital signatures periodically distributed by data owners.

7. EVALUATION

The evaluation of techniques for securing geospatial data, like those outlined in the previous section, requires devising suitable metrics. A suitable set of metrics should include: (1) cost and utility of the method, (2) integration into existing GIS infrastructures, and (3) security and privacy threats.

(1) Cost and Utility of the Method. As the primary focus of the outlined research directions is on developing novel types of security and privacy specifications for diverse types of geospatial data, an important metric is the expressiveness of policy specifications. While an expressive language itself is essential, it is equally important to balance expressiveness of the language with the cost of performing reasoning (e.g., to what extent can soundness and completeness properties be assured?). The expressiveness of the specification language also has an impact on its utility. That is, how easy is it for GIS security managers to specify security and privacy policies, not using a formal logic-based language but a language that resembles concepts of GML application schemas? A similar approach to balancing expressiveness, cost of reasoning, and utility of the method must be taken for security policies in the context of GIS interoperations, trust management approaches, and data authentication.

(2) Integration into Existing GIS Infrastructures. Rather than developing security components in an idealized environment, any realistic solution must integrate these components into existing GIS infrastructures and applications. Such infrastructures include GRASS, PostgreSQL, ArcGIS, and typical GIS Web Services, such as Web map services and Web feature services. The effort to add security components such as policy-driven access control modules or data authentication components to a system (e.g., in the form of a wrapper or integral system component, Fig. 2), should be minimal, requiring only minor extensions (not modifications) to the system. Note that this metric needs to be considered in the context of the cost and utility metrics, as more functionality and better utility likely requires more integration efforts.

(3) Security and Privacy Threats. The incremental development of security components and their integration into GIS infrastructures is accompanied with the development of security, privacy, and trust threat scenarios. For each type of policy, one needs develop threat scenarios and security attacks. The availability of such scenarios will allow one to evaluate the correctness of the security policies and investigate threats to guide the development of more sophisticated policy specifications, in particular those that deal with the filtering and obfuscation of geospatial features. The operations included in the data model (Section 6.1) can provide the basis for constructing such threat scenarios, as these operations will be the basic means to select and modify geospatial data from GIS layers, combinations of GIS layers, and from different GIS repositories in the context of GIS repository coalitions.

8. CONCLUDING REMARKS

Protecting geospatial data includes several tasks that raise a variety of technical and organizational problems and challenges. In this paper, we have presented a comprehensive view of such requirements and discussed major technical issues concerning policy specification, policy interoperability, and privacy and trust. We furthermore outlined important directions of research to address the numerous security aspects of (interoperable) geospatial data and GIS applications. In order to move from theory to practice, the availability of standards for geospatial data representation and data security is fundamental. Standards make the development of geospatial data protection techniques affordable and effective for the deployment of those solutions. In that perspective, contributing to the development of standards for advanced geospatial data protection is another major goal to address.

9. REFERENCES

[1] GeoXACML Implementation Specification, http://www.opengeospatial.org/standards/geoxacml.

[2] Open GIS Consortium Interoperability Demonstration Focuses on Emergency Response Situations, http://xml.coverpages.org/ogc-wsinterop.html.

[3] The Open Geospatial Consortium (OGC). http://www.opengeospatial.org.

[4] XML Signature Syntax and Processing, W3C Recommendation, June 2008. http://www.w3.org/TR/xmldsig-core/, 2008.

[5] Geospatial Interoperability Reference Model (GIRM, V 1.1). http://gai.fgdc.gov/, 2003.

[6] GML3.1 ISO/TC 211/WG 4/PT 19136 Geographic information, Geography Markup Language (GML), Committee Draft. http://portal.opengeospatial.org/files/?artifact_id=4700, 2004.

[7] OGC Critical Infrastructure Protection Initiative (CIPI),http://ip.opengis.org/cipi/, 2006.

[8] Global Earth Observation System of Systems (GEOSS). http://www.epa.gov/geoss/, 2006.

[9] OpenGIS Geography Language (GML) Encoding Specification, version 3.1.1, http://www.opengeospatial.org/standards/gml, 2007.

[10] M. Abedin, S. Nessa, L. Khan, and B. M. Thuraisingham. Detection and resolution of anomalies in firewall policy rules. In *20th Annual IFIP WG 11.3 Working Conference on Data and Applications Security (DBSec)*, 2006.

[11] A. Alam, G. Subbiah, and B. Thuraisingham. Reasoning with semantics-aware access control policies for geospatial Web services. In *ACM Workshop on Secure Web Services (SWS), George Mason University, Fairfax VA, USA*, 2006.

[12] A. Alam and B. Thuraisingham. Geography resource description framework (GRDF) and secure GRDF (S-GRDF). Technical report, The University of Texas at Dallas, 2006.

[13] V. Atluri and S. A. Chun. An authorization model for geospatial data. *IEEE Transactions on Dependable and Secure Computing*, 1(4):238–254, 2004.

[14] V. Atluri and P. Mazzoleni. Uniform indexing for geospatial data and authorizations. In *Research Directions in Data and Applications Security, IFIP WG 11.3 Sixteenth International Conference on Data and Applications Security*, 2002.

[15] J. C. Baker, B. E. Lachman, D. R. Frelinger, K. M. O'Connell, and A. Hou. Mapping the risks: Assessing the homeland security implications of publicly available geospatial information. Technical Report, RAND National Defense Research Institute, 2004.

[16] A. Belussi, E. Bertino, B. Catania, M. L. Damiani, and A. Nucita. An authorization model for geographical maps. In *12th ACM International Workshop on Geographic Information Systems, (ACM-GIS)*, 2004.

[17] A. Belussi, B. Catania, and E. Bertino. A reference framework for integrating multiple representations of geographical maps. In *Proceedings of the Eleventh ACM International Symposium on Advances in Geographic Information Systems (ACM GIS)*, 2003.

[18] F. L. Ber and A. Napoli. Design and comparison of lattices of topological relations for spatial representation and reasoning. *Journal of Experimental & Theoretical Artificial Intelligence*, 15(3):331–371, 2003.

[19] E. Bertino, P. A. Bonatti, and E. Ferrari. Trbac: A temporal role-based access control model. *ACM Transactions on Information and System Security*, 4(3):191–233, 2001.

[20] E. Bertino, B. Catania, E. Ferrari, and P. Perlasca. A logical framework for reasoning about access control models. *ACM Transactions on Information and System Security*, 6(1):71–127, 2003.

[21] E. Bertino, M. L. Damiani, and D. Momini. An access control system for a web map management service. In *14th International Workshop on Research Issues in Data Engineering (RIDE-WS-ECEG 2004), Web Services for E-Commerce and E-Government Applications*, 2004.

[22] E. Bertino and E. Ferrari. Secure and selective dissemination of xml documents. *ACM Transactions on Information and System Security*, 5(3):290–331, 2002.

[23] E. Bertino, S. Jajodia, and P. Samarati. A flexible authorization mechanism for relational database systems. *ACM Transactions on Information Systems*, 17(2):101–140, 1999.

[24] E. Bertino and R. S. Sandhu. Database security-concepts, approaches, and challenges. *IEEE Transactions on Dependable and Secure Computing*, 2(1):2–19, 2005.

[25] R. Bhatti, A. Ghafoor, E. Bertino, and J. Joshi. X-GTRBAC: an XML-based policy specification framework and architecture for enterprise-wide access control. *ACM Transactions on Information and System Security*, 8(2):187–227, 2005.

[26] S. A. Chun and V. Atluri. Protecting privacy from continuous high-resolution satellite surveillance. In *Data and Application Security, Development and Directions, IFIP TC11/ WG11.3 Fourteenth Annual Working Conference on Database Security*, 2000.

[27] M. J. Covington, W. Long, S. Srinivasan, M. A. Anind K. Dey, and G. D. Abowd. Securing context-aware applications using environment roles. In *6th ACM Symposium on Access Control Models and Technologies*, 2001.

[28] M. Damiani, E. Bertino, B.Catania, and P. Perlasca. Geo-RBAC: a spatially-aware RBAC. *ACM Transactions on Information and System Security*, 10(1):2, 2007.

[29] P. Devanbu, M. Gertz, A. Kwong, C. Martel, and S. Stubblebine. Flexible authentication of XML documents. *Journal of Computer Security*, 12(6):841–864, 2004.

[30] P. Devanbu, M. Gertz, C. Martel, and S. Stubblebine. Authentic data publication over the Internet. *Journal of Computer Security*, 11(3):291–314, 2003.

[31] J. Dobson. Is GIS a privacy threat? GIS World, 1198.

[32] A. Entchev. GIS and privacy. Directions Magazine, 2005.

[33] ESRI. OpenGIS Interoperability Add-ons for ArcGIS, http://www.esri.com/software/standards/ogc-download.html,2005.

[34] M. Gertz, Q. Hart, C. Rueda, S. Singhal, and J. Zhang. A data and query model for streaming geospatial image data. In *11th International Workshop on Foundations of Models and Languages for Data and Objects (Query Languages and Query Processing - QLQP)*, Revised Selected Papers. LNCS 4254, Springer, 687–699. 2006.

[35] M. Gertz, A. Kwong, C. Martel, G. Nuckolls, P. Devanbu, and S. Stubblebine. Databases that tell the truth: Authentic data publication. *Bulletin of the Technical Committee on Data Engineering*, 7(1):21–41, 2004.

[36] M. Gertz and S. Jajodia (Editors). *The Handbook of*

Database Security: Applications and Trends. Springer, 2007.

[37] M. Gertz and A. M. Rosenthal. *Database Security*. In Bidgoli, H. (editor) Handbook of Information Security, Threats, Vulnerabilities, Prevention, Detection and Management, pages 380–395, Wiley. 2006.

[38] F. Hansen and V. A. Oleshchuk. Spatial role-based access control model for wireless networks. In *IEEE Vehicular Technology Conference VTC2003-Fall*, 2003.

[39] S. Jajodia, P. Samarati, M. L. Sapino, and V. S. Subrahmanian. Flexible support for multiple access control policies. *ACM Transactions on Database Systems*, 26(2):214–260, 2001.

[40] J. R. Jensen. *Introductory Digital Image Processing*. Third Edition, Prentice Hall, 2004.

[41] J. Joshi, E. Bertino, U. Latif, and A. Ghafoor. A generalized temporal role-based access control model. *IEEE Transactions on Knowledge Data Eng*, 17(1):4–23, 2005.

[42] M. Koch, L. V. Mancini, and F. Parisi-Presicce. On the specification and evolution of access control policies. In *6th ACM Symposium on Access Control Models and Technologies (SACMAT 2001)*, 2001.

[43] R. Lake, D. S. Burggraf, M. Trninic, and L. Rae. *Geography Markup Language-Foundation for the Geo-Web*. Wiley, 2004.

[44] M. Lorch, S. Proctor, R. Lepro, D. Kafura, and S. Shah. Access control: First experiences using XACML for access control in distributed systems. In *ACM workshop on XML security*, 2003.

[45] P. M. Mather. *Computer Processing of Remotely-Sensed Images*. Wiley, 2004.

[46] A. Matheus. Declaration and enforcement of fine-grained access restrictions for a service-based geospatial data infrastructure. In *10th ACM Symposium on Access Control Models and Technologies (SACMAT 2005)*, 2005.

[47] P. Mazzoleni, E. Bertino, and B. Crispo. XACML policy integration algorithms: not to be confused with XACML policy combination algorithms! In *SACMAT*, 2006.

[48] FGDC. Guidelines for providing appropriate access to geospatial data in response to security concerns. http://www.fgdc.gov/policyandplanning/Access_Guidelines.pdf, June 2005.

[49] FGDC. National Spatial Data Infrastructure (NSDI). http://www.fgdc.gov/nsdi/nsdi.html.

[50] H. J. Onsrud, J. P. Johnson, and X. Lopez. Protecting personal privacy in using geographic information systems. *Photogrammetic Engineering and Image Processing*, 60(9):1083–1095, 1994.

[51] D. Papadias, M. J. Egenhofer, and J. Sharma. Hierarchical reasoning about direction relations. In *Proceedings of the 4th ACM international workshop on Advances in geographic information systems*, 1996.

[52] P. Rigaux, M. Scholl, and A. Voisard. *Spatial Databases: With Application to GIS*. Morgan Kaufmann, 2002.

[53] P. Samarati and S. D. C. di Vimercati. *Foundations of Security Analysis and Design, Tutorial Lectures (FOSAD 2000)*, chapter Access Control: Policies, Models, and Mechanisms, pages 137–196. LNCS 2171, Springer, 2001.

[54] H. Samet. *Applications of Spatial Data Structures: Computer Graphics, Image Processing and GIS*. Addison-Wesley, 1989.

[55] R. S. Sandhu, V. Bhamidipati, and Q. Munawer. The ARBAC97 model for role-based administration of roles. *ACM Transactions on Information and System Security*, 2(1):105–135, 1999.

[56] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman. Role-based access control models. *IEEE Computer*, 29(2):38–47, 1996.

[57] B. Thuraisingham. *Database and Applications Security, Integrating Data Management and Applications Security*. CRC Press/Auerbach, 2005.

[58] B. Thuraisingham and W. Ford. Security constraint processing in a multilevel secure distributed database system. *IEEE Transactions on Knowledge and Data Engineering*, 7(2):274–293, 1995.

[59] C.D. Tomlin. Geographic Information Systems and Cartographic Modeling. Prentice-Hall, 1990.

[60] Y. Wang and J. Vassileva. Bayesian network trust model in peer-to-peer networks. In *Second International Workshop on Agents and Peer-to-Peer Computing (AP2PC 2003)*, 2003.

[61] M. Winslett, N. Ching, V. E. Jones, and I. Slepchin. Using digital credentials on the world wide web. *Journal of Computer Security*, 5(3):255–266, 1997.

[62] T. Wright. Geographic information systems. Ontario Office of Information and Privacy Commissioner, 1997.

[63] Y. Yang, S. Papadopoulos, D. Papadias, and G. Kollios. Spatial outsourcing for location-based services. In *Proceedings of the 24th International Conference on Data Engineering, ICDE 2008*, 1082–1091, 2008.