Security of Shared Data in Large Systems: State of the Art and Research Directions

Arnon Rosenthal
The MITRE Corporation
202 Burlington Road /k308
Bedford MA 01730 USA
+1-781-271-7577
arnie@mitre.org

Marianne Winslett
University of Illinois
Dept. of Computer Science, 201 N. Goodwin Ave
Urbana, IL 61820 USA
+1-217-333-3536
winslett@cs.uiuc.edu

ABSTRACT

The target audience for this tutorial is the entire SIGMOD research community. The goals of the tutorial are to enlighten the SIGMOD research community about the state of the art in *data* security, especially for enterprise or larger systems, and to engage the community's interest in improving the state of the art.

1. INTRODUCTION

Security is increasingly recognized as a key impediment to sharing data in enterprise systems, virtual enterprises, and the semantic web. Yet the topic has not been a focus for mainstream database research, industrial progress in data security has been slow, and (too) much security enforcement is in application code, or else is coarse grained and insensitive to data contents.

The SIGMOD community is in an excellent position to make significant improvements in the way people think about security policies, due to the community's experience with declarative and logic-based specifications, automated compilation and physical design, and both semantic and efficiency issues for federated systems. These strengths provide a foundation for improving both theory and practice.

This tutorial aims to enlighten the SIGMOD research community about the state of the art in *data* security, especially for enterprise or larger systems, and to engage the community's interest in improving the state of the art. Thus after a very brief look at security basics, the tutorial focuses on the following questions:

- What is the current state of the art in the real world, with respect to data security?
- What sorts of additional research results are needed to improve the current state of affairs?
- What frameworks can be helpful for architects and researchers tackling these problems?

For researchers, we will present many open research problems (some grad-student ready, others requiring more formalization) as we move through the sections of the tutorial. For architects, we will suggest unifying concepts and distinctions that they may wish to support, even before the research matures.

Our overall goal for the tutorial is to present material that is not found in any textbook---to show the audience how *they* can help improve the state of the art in data security. Thus the first section of the tutorial, entitled *Basics*, is **not** intended as a replacement for a security textbook. While a tutorial focused entirely on security basics might be helpful to the SIGMOD community, that is not the goal of this proposal.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
SIGMOD 2004, June 13–18, 2004, Paris, France.
Copyright 2004 ACM 1-58113-859-8/04/06 …$5.00.

2. TUTORIAL OUTLINE

Part I: Basics (0:30 total)

The tutorial includes this section to help frame how attendees think about the material in the following sections.

1. Basic terminology. 0:25.
    a. Data security is driven by data semantics and values, rather than protecting platforms and networks.
    b. Vital buzzwords: confidentiality, integrity, availability, authentication, authorization/access control, release policy, separation of duty, non-repudiation, linkability, audit trails, anomaly detection, accountability, Chinese wall policy, privacy, mandatory versus discretionary, Clark-Wilson, inference control, reference monitor, digital rights management, …
    c. Basic threats and attacks: repudiation, denial of service, man in the middle, impersonation, intrusions, revocation, Trojan horse, …
    d. Cryptography: what does it buy you?

2. What are the major themes that recur in data security research issues? (In other words, if you have a proposal to provide one of the services on the buzzword list, and you have managed to stave off the basic threats and attacks, what criteria do you use to evaluate your proposal?) 0:05.
    a. Scalability (to large numbers of objects, subjects, organization, sites, and fine-grained control of access to different parts of objects and different kinds of access)
    b. Usability (no CS degree required)
    c. Analyzability (what-if queries; determine effect of changes in advance; see what a given

subject/object can do/have done to it)

d. Manageability (including separation of management skills, availability of analysis tools, avoidance of redundancy)
e. Modularity/universality (need to create constructs and algorithms that contribute to *many* data models and security-management approaches, so that they are deployable in many different domains, applications, and OS/DBMS environments)
f. Orthogonality, composability, compatibility (need to be able to compose modules that control different aspects of security in a mix-and-match fashion, to the degree possible; need to be able to build security services out of reusable components, which is much harder than one would expect)
g. Flexibility and extensibility (because the world never stands still)
    i. What happens in an emergency?
    ii. What happens if rapid response is required to a new business opportunity with a new partner?

Part II: State of the Art and Open Problems (1:10)

3. What is the current real-world state of the art in security of shared data in large systems? (Research issues will be discussed as we encounter them.)0:35.
    a. SQL security and views, including efficiency issues and significant enrichments that seem feasible today if researchers will take them on:
        i. A formal basis, to simplify the statement of the SQL standard relating to security
        ii. Smaller refinements (view privileges, metadata controls, procedures like views, and more)
        iii. Other extensions that seem central, including factors and a negative authorization mechanism
    b. Role-based access control: what is it, what are its limits in data sharing scenarios, and what is its relationship to mandatory security? (no suggested enhancements)
    c. Oracle's (and now products') row-labeled tuples, and their relationship to mandatory security and to RBAC
    d. Kerberos (very briefly, what is it, what is it good for, and what doesn't it do, with no suggested enhancements)
    e. Middleware to manage trust
        i. Why is it needed?
        ii. Applications that require it (best known = supply chain management)
        iii. Companies that provide it
        iv. What it does and doesn't do

4. Where do things seem to be heading in the near term? 0:35.
    a. XML security proposals, and their relation to SQL
    b. Semantic web proposals
        i. The stack of semantic web formalisms
        ii. The inconsistent proposals for security at the different levels of the stack
        iii. Unmet needs: new abstractions that map across layers, and suggested research directions for supplying them (including property-of, nested-element, derived-object (and query language), and more)
    c. Trust management
        i. An alternative approach to 3c and 3d
        ii. Very short preview of section 15

5. Previous database security research. 0:05.
    a. What did it concentrate on? (MLS)
    b. Why didn't MLS have much impact (beyond inspiring row labels)? Lessons learned, and how we can avoid making the same mistakes.
    c. RBAC models and "role engineering"

Part III: Policies as a Unified Framework for Managing the Security of Shared Data (1:10) (0:35 for items 6-12, 0:35 for items 13-16))

6. What sort of policies do we need? (revisits section 1a)
    a. What other issues arise at run time? (e.g., the need for negotiation, conflict resolution, human intervention, and handling emergency situations)

7. Why might declarative policies be helpful?
    a. Uniform way of talking about security requirements for data in a wide variety of contexts and at different levels of the system: (distributed) databases, formatted messages, services/components, general documents, and higher (application) and lower (OS, network) levels
    b. Agreed-upon semantics
    c. Amenable to formal analysis

8. How can we express policies? (Strategy: Multiple formalisms, modularize the tractable parts)
    a. Delegation graphs and role hierarchies (as in SQL)
    b. Logic programming based languages
    c. Proposals from the AI community (Rei, KAoS)

9. How can we make policies amenable to administration?
10. What kinds of enforcement mechanisms can we use?
    a. Reference monitors, possibly delegating to policy engines
    b. Boundary guards
    c. Embedded code

11. How can we fill the gap between policy and enforcement semi-automatically?
    a. From very abstract statements (e.g., individually identifiable taxpayer data must be highly restricted), through more concrete policy statements, on down to policies that can be mapped directly into enforcement mechanisms.
        i. Can we map the policies up, down, and sideways? (Analogies: view update, view security, physical design)

   b. Relationship to query processing in heterogeneous systems (servers in a distributed system will differ in enforcement capabilities, trustworthiness, performance, load, quality of error messages, etc.)
   c. Need for bidirectionality. (Copernican view: it doesn't revolve around us)

12. Why the database community is in a good position to advance the state of the art
   a. Researchers: Declarative policies, compiled to executable form, over a heterogeneous set of servers (distributed query processing).
   b. Vendors can integrate with other metadata-driven aspects of database systems; should not be a separate stovepipe.

13. Example from the semantic web: using the XML as a basis for controlling message content. Just identify issues
   a. No global picture of all parties and all data, so policies must be specified in terms of their properties
   b. Local control only
   c. Multiple formalisms (XML, RDF, OWL, SQL, …)

14. Extended examples from enterprise-wide security
   a. Issues
      i. Multiple administrators: no single point of analysis or control
      ii. Semantic and physical heterogeneity of data and security systems
      iii. Need for enterprise-wide analysis
      iv. Need to react quickly to opportunities and emergencies
      v. Need to separate enforcement decisions made with respect to different aspects of security
   b. How to address each of these issues via policies
      i. In the more tractable case of benign but clueless servers and administrators
      ii. In the context of Digital Rights Management (some administrators and servers are hostile)

15. Distributed trust in dynamic coalitions
   a. Dynamic coalition examples and motivation (e-commerce, business virtual enterprises, military coalitions, disaster response)
   b. Issues
      i. Players unknown to one another
      ii. Sharing is unanticipated in advance
      iii. Need to be able to understand one another's policies
      iv. Need for quick adjustment of security policies
      v. Relevance of reputations as well as objective criteria
      vi. Run-time issues
   c. How to address the issues
      i. Example e-commerce policies
      ii. Run-time policy discovery and enforcement
      iii. The use of negotiation strategies
      iv. Zero-knowledge approaches

16. Summary of open problems in policy-based approaches to security management

3. ABOUT THE PRESENTERS

Arnie Rosenthal is a Principal Scientist at MITRE. He has broad interests in problems that arise when data is shared between communities, including a long-standing interest in the security issues that arise in data warehouses, federated databases, and enterprise information systems. He has also had a first-hand look at many security problems that arise in large government and military organizations.

Marianne Winslett has been a professor at the University of Illinois since 1987. She started working on database security issues in the early 1990s, focusing on semantic issues in MLS databases. Her interests soon shifted to issues of trust management for data on the web. Trust negotiation is her main current research focus.