Trust, Privacy, and Security

Summary of a Workshop Breakout Session at the
National Science Foundation Information and Data Management (IDM) Workshop
held in Seattle, Washington, September 14 - 16, 2003
(Version 2)

Bharat Bhargava[1] [*], Csilla Farkas[2], Leszek Lilien[1], Fillia Makedon[3] [†]

[1] Department of Computer Sciences and
Center for Education and Research in Information Assurance and Security (CERIAS)
Purdue University
West Lafayette, IN 47907

[2] Department of Computer Science and Engineering
University of South Carolina
Columbia, SC 29208

[3] Department of Computer Science
Dartmouth College
Hanover, NH 03755

[*] The moderator. E-mail: bb@cs.purdue.edu.

[†] With contributions from Sudarshan Chawathe (University of Maryland), Batya Friedman (University of Washington), Hichem Frigui (University of Memphis), Johannes Gehrke (Cornell University), Joseph Hellerstein (University of California, Berkeley), Hillol Kargupta (University of Maryland Baltimore County), Vipin Kumar (University of Minnesota), Sharad Mehrotra (University of Califorina at Irvine), Filippo Menczer (Indiana University), Sang H. Son (University of Virginia), Dan Suciu (University of Washington), and Marianne Winslett (University of Illinois at Urbana-Champaign).

1 Trust, Privacy, and Security Breakout Session - Background

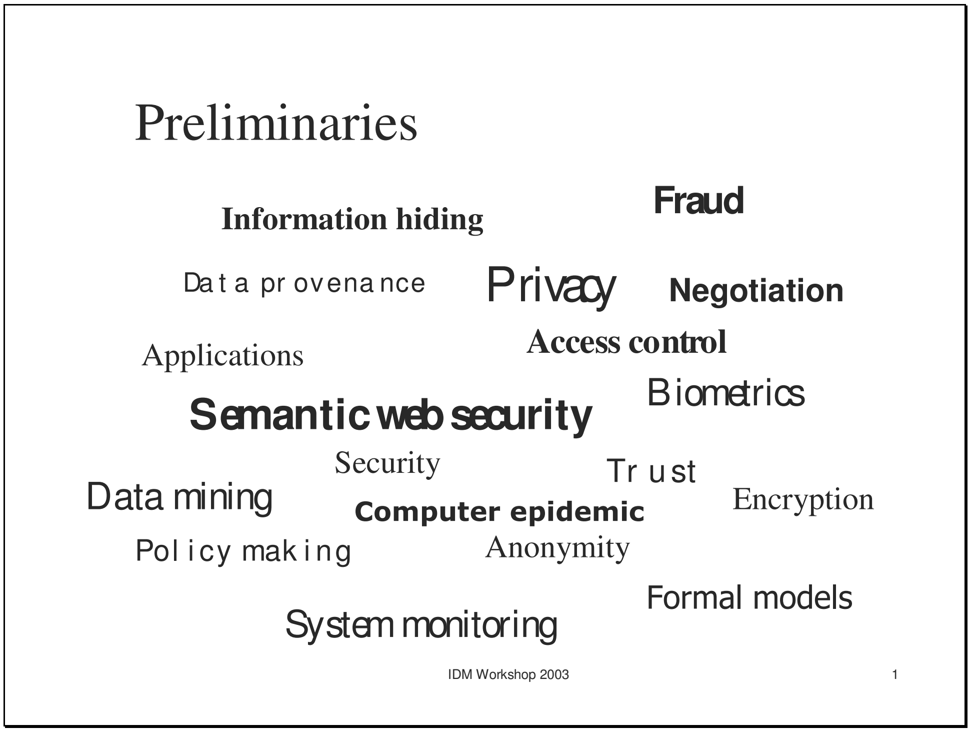
1.1 Summary

The session started with presentations of the interests of the participants in the areas of trust, privacy, and security. Then, the discussion continued with a listing of research challenges, directions, and opportunities for the IDM community.

Next, the participants attempted to categorize and prioritize the identified challenges, directions, and opportunities, as well as formulate recommendations for future research for the IDM research community. These results are presented below, divided into sections devoted in turn to trust, privacy, and security.

Research on trust, privacy and security in IDM can be viewed as a part of the Cyber Trust initiative of the National Science Foundation [Land03].

1.2 Preliminaries: The Realm of Trust, Privacy, and Security Terms



Preliminaries

Information hiding

Fraud

Data provenance

Privacy

Negotiation

Applications

Access control

Semantic web security

Biometrics

Security

Trust

Data mining

Computer epidemic

Encryption

Policy making

Anonymity

Formal models

System monitoring

IDM Workshop 2003

1

Fig. 1. The Realm of Trust, Privacy, and Security Terms

Fig. 1 shows a number of terms relevant to the discussion of Trust, Privacy, and Security issues. Defining and formalizing them is one of the research challenges.

2 Trust Problems

2.1 Trust Preliminaries

The essential question that needs to be raised first is:

Why do we need the notion of trust in computing systems?

The participants pointed out the following most important reasons:

- Trust enables collaboration and communication
  Both collaboration and communication without mutual trust are severely handicapped. (Lack of trust makes security solutions extremely difficult and expensive.)

- The significant role of trust as a social system paradigm: applicable to social systems from a small village to a big city
  Trust is used in social systems pervasively, sometimes implicitly (as in a small village with the social control where everybody knows everybody) and sometimes explicitly (as in a big city where explicit rules of behavior are needed to enter trust relationships, due to the scale and diversity of these relationships). If trust works so well in complex social systems, why not use it as a computing system paradigm?

- The growing role of trust in dynamic and open environments
  Trust use is limited or implicit in static and closed social systems (such as a small village), and explicit in dynamic and open social systems (such as a big city). By analogy, the growing role of trust in open *computing* systems should be recognized and trust-based approaches should be exploited.

2.2 Trust Challenges for IDM

Among the challenges to use of trust in IDM the participants identified the following:

- How to initiate and build trust?
  How to create formal models of trust, addressing the issues of different types of trust (e.g., trust towards data, or users, or system components)? How to define trust metrics to compare different trust models? How should trust models accommodate trust characteristics (such as context dependency, bi-directionality, and asymmetry)? How should the models of trust handle both direct evidence and second-hand recommendations related to the trusted subjects or objects? How trusted parties can be used to initiate and build trust? How timeliness, precision, and accuracy affect the process of trust building?

- How to maintain and evaluate trust?

How to collect and maintain trust data (e.g., credentials, evidence on the behavior of the trusted objects, recommendations)? How and when to evaluate trust data? How to discover betrayal of trust, and how to enforce accountability for damaging trust? How to prevent trust abuse, for example by means of access right revocation? How to motivate users to be good citizens and to contribute to trust maintenance?

- How to deal with fraud?
  How to create formal models of fraud? How to define metrics to compare different fraud models? How to design efficient methods and tools for fraud prevention and detection? How to prevent, detect, and trace fraud? When to tolerate fraud? How to use trust assessment, threat avoidance and threat tolerance to prevent fraud?

- How to guarantee scalability, performance, and economic parameters for trust solutions?
  How to scale up trust models and solutions? What is the impact of trust solutions on system performance and economics? How and what economic incentives and penalties can be used?

- How to engineer trust-based applications and systems?
  How to experiment with and implement trust-based applications and systems for e-government, e-commerce, and other applications? How to enhance system performance, security, economics, etc. with trust-based ideas (e.g., like enhancing role-based access control with trust-based mappings)? How to use incentives and penalties for building trust and preventing fraud?

2.3 Trust Recommendations for IDM

Trust research is a relatively new and fast growing area of IDM research, and members of this breakout group consider it a very promising and exciting approach.

We recommend research in the following IDM-related areas:

- *Social paradigm of trust.* Utilization of the powerful social paradigm of trust, based on the analogies to uses of the notion of trust in social systems, should be explored in many ways. We realize that finding out what makes trust work in existing social systems, and transferring this to a computing world is a big challenge. This work calls for strong cooperation with social scientists to produce viable results.

- *Liability of trust.* Trust may be considered a form of limited liability with an associated confidence factor. Broadly, we need to provide methods, algorithms, and tools to identify which components and processes of the system depend on trust, to which extent and how the security of a system may be compromised if any one of these trust-dependent components fails. As an example, the role of data provenance explanations in trust-based systems needs be investigated.

- *Scalable and adaptable trust infrastructure.* A high priority should be given to building scalable and adaptable trust infrastructure, including infrastructure for trust management and trust-based negotiations. In particular, we believe that investigations should include gaining insights from different applications, exploring the issue of dynamic trust, building interoperable tools for the trust infrastructure, developing flexible and extensible standards, and facilitating trust-based negotiations.

- *Benchmarks, testbeds, and development of trust-based applications*. Research is needed on development of benchmarks and testbeds for experimenting with diverse roles of trust in computing systems. The experiments should form a strong basis for the development of prototype trust-based applications, such as ones for crisis and emergency management for homeland security, broad collaborations among researchers or government agencies, or medical information sharing between healthcare providers. Trust-based IDM solutions for new and emerging technologies should be an especially active investigation area. An example is using trust for ensuring data integrity and privacy in sensor networks deployed in trustless environments.

- *Fraud prevention and detection.* Research on fraud should counteract growing knowledge and capabilities of computer fraudsters. This work should result in modeling fraud, defining its metrics, determining efficient methods and algorithms for fraud prevention and detection, and building tools.

- *Trust-related interdisciplinary research.* We would like to see more trust-related interdisciplinary research outside of the realm of computer science and engineering. In addition to already mentioned interdisciplinary work on the social paradigm of trust, it should include research on ethical, social, and legal issues, both human-centered and system-centered. Another important interdisciplinary work should focus on economic incentives for building trust, and disincentives and penalties for committing fraud.

3 Privacy Problems

3.1 Privacy Preliminaries

The essential question that needs to be raised first is:

What is privacy? Why do we need it?

3.2 Privacy Challenges for IDM

Among the challenges in IDM privacy research, the participants identified the following:

- How to define and measure privacy and its multifaceted aspects?

How to define and assess quality, safety, and privacy of personal data? How to define metrics for this assessment?

- How to define, analyze, and manage privacy policies?
  How to define privacy policies? How to best perform privacy requirement analysis and stakeholder analysis? How to address privacy of primary and secondary uses of information? How to optimize digital rights management (DRM)?

- What technologies (or system components) endanger privacy in IDM environments, and how to prevent this?
  As an example, how to prevent pervasive computing from illegitimate monitoring and controlling people? How to assure anonymity in more and more pervasive computing environment? How to balance anonymity with accountability under these circumstances?

- What technologies can be utilized or exploited to provide privacy, and how to use them to this end?
  For example, is there a way to insert "privacy monitors" or tools that provide alerts when privacy is endangered due to inference or careless transactions? What are the best ways of privacy-preserving data mining and querying? How to monitor Web privacy and prevent privacy invasions by undesirable inferences? How to address the issue of "monitoring the monitor," including identification and prevention of situations when incorrect monitor data result in a personal harm?

3.3 Privacy Recommendations for IDM

We recommend the following areas of privacy research for IDM:

- *Privacy metrics.* Issues of privacy of users or applications, on the one hand, and privacy (secrecy, confidentiality) of data, on the other hand, intertwine. We recommend more research on metrics for personal and confidential data usage. The metrics should include measures of who and how accesses data, what data are accessed, and for how long. Research is needed on the development of metrics and methods for measurements of privacy-related aspects of data quality. Researchers should also propose measures of accuracy in information extraction with respect to privacy, since inaccurate information can obstruct accountability or harm privacy (like in a case of a wrongly identified individual). One example application is to consider the HIPAA compliance rules set by NIH for medical systems, and evaluate their benefits and shortcomings.

- *Privacy policy monitoring and validation.* We need to better understand how to monitor and validate privacy policies. We need to develop technology that ensures the correct enforcement of privacy policies. This research should include addressing the issues of monitoring and validating privacy aspects of data integration, separation, warehousing, and aggregation. An interesting issue, related to validation, is licensing of personal data by their owners for specific uses (an example is Ms. Smith agreeing to receive real property advertising by licensing her e-mail rights to a real estate advertiser).

- *Information hiding, obfuscation, anonymity, and accountability.* Research should address different ways of assuring anonymity via information hiding and obfuscation, ranging from steganography through location security and hiding message source and destination from intermediate nodes to approaches used for digital elections. At the same time, for accountability, we need to investigate how to prevent illegitimate or improper information hiding. We need models supporting accountable anonymity that do not depend on a trusted third party. As an example, accountability suffers when data provenance obfuscation or user anonymity hinder intruder identification. Other interesting issues are information hiding and anonymity preservation in negotiations among parties with variable degrees of mutual trust.

- *New privacy-enabling and privacy-disabling technologies.* We need more research on the impact of new technologies on preserving privacy. In particular research on privacy for pervasive computing is needed, since pervasive computing results in an easy information flow. Unless proper access control is provided, this flow threatens to ruin anonymity with perfect accountability (e.g., not only GPS-enabled devices but even cell phones and RFID tags on purchased products introduce the risk of monitoring of location of individuals). Similarly, permanent availability (or "always-on" connectivity) complicates protection against denial-of-service attacks. Interesting aspects of trust-related privacy are raised by data-sharing peers, including limiting data disclosures on the as-needed basis, and avoiding sharing irrelevant or highly sensitive data (such as trade secrets). Another important issue is privacy-preserving data mining on massive datasets.

- *Interdisciplinary privacy research.* More interdisciplinary research on privacy is needed. One important direction of interdisciplinary work is proposing comprehensive and rich privacy models based on social and ethical privacy paradigms. Another direction is considering public acceptance of privacy requirements and rules, and their enforcement.

4 Security Problems

4.1 Security Preliminaries

The essential question that needs to be raised first is:

How to define security research for the IDM community?

We might start with dividing IDM research into the following two areas:

1. *Information security:* how to prevent, detect, and deter improper disclosures or modifications of information, or denials of access to information.

2. *Security information management*: how to collect, maintain, and analyze data relevant to security.

4.2 Security Challenges for IDM

The participants identified the following research directions as raising security-related challenges for the IDM community:

1. Challenges for information security:

- How to protect information in demanding computing environment?
  What are the security needs in pervasive environments? How to develop access control models for semantic-aware, large-scale data integration? What are the security needs of mobile, wireless systems? How to ensure data integrity and availability in decentralized and dynamic environments? How to trust a data source and perform data quality assessment (data freshness, correctness, etc.)?

- How to manage security policies?
  How to delegate efficiently and accurately security metadata, i.e., how to "translate" a global policy into local requirements?

- How to protect data in emerging technologies?
  For example, how to protect data for Semantic Web, sensor networks, multimedia systems? How to develop proper access control models, including direct and indirect data accesses and semantic constructs? How to develop efficient and correct inference and aggregation control?

2. Challenges for security information management:

- How to manage data for vulnerability and threat analysis?
  How to collect, maintain, and analyze large amount of such data? How to manage data efficiently, assess its quality, and provide integrated analysis framework?

- How to detect attacks and identify intruders?
  How to provide data management support for attack detection and intruder identification? What are the legal aspects of data collection and use of the results of the identification?

- How can information management aid physical security?

- What are the tradeoffs (such as performance vs. security) and how to balance the conflicting requirements?

4.3 Security Recommendations for IDM

We recommend research in the following areas of security investigation for IDM:

- *Security implications of emerging technologies.* We see the importance of research on security implications of data management for emerging technologies, including the following:
  - Pervasive computing
  - Semantic Web, and context and semantic-aware applications
  - Multimedia
  - Data and Web mining
  - Peer-to-peer (P2P) computing
  - Sensor networks
  - Mobile, wireless, and ad hoc computing

- *Modeling and design for security.* Investigation of modeling and design for security is needed. Examples of specific issues include considerations of levels of security as well as security aspects of data quality, QoS/SLA (quality of service/service level agreement), real-time constraints, inference control, data provenance, metadata management, information flow models, negotiations, and direct and indirect data disclosure analysis.

- *Vulnerability analysis and threat assessment.* Investigation of vulnerabilities and assessment of threats in IDM systems should result in creating new models, metrics, and tools. These developments should, in turn, limit impact of vulnerabilities, and contribute to threat avoidance and tolerance in IDM systems. This should include research on attacks on data and metadata quality, including software attacks (such as malicious transactions) and physical attacks (such as destroying a disk drive).

- *Security aspects of information integration or separation, and data warehousing.* Information integration needs careful study of two security aspects: on the one hand the security issues in information integration, and on the other hand integration of security-relevant data. Both raise the issues of data quality (incl. its provenance), heterogeneous integration framework, task-specific integration, metadata management, security in and for integration, and correctness of integration methods. Work is also needed on security aspects of information separation, when an integrated system is split into two or more separate systems due, for example, to splitting a company into two or more separate entities. Data warehousing is another area that requires a study of its security aspects.

5. Concluding Remarks

We have specified separate recommendations for research on trust, privacy, and security (TPS). It is, however, worth noting that there are common threads running through all these research areas. The first common thread are the tradeoffs, including the performance vs. TPS, cost and functionality vs. TPS, and data monitoring and mining vs. TPS.

The second common thread contains policies, regulations, and technologies for TPS. This includes creation of flexible TPS policies, appropriate TPS data management (including collection, usage, dissemination, and sharing of TPS data), and development of domain- and application-specific TPS approaches (such as TPS solutions for commercial, government, medical, and e-commerce fields).

The third and the fourth threads are development of economic models for TPS, and investigation of legal and social TPS aspects.

We would like to conclude with the motto:

Any data management issue is a security issue!

Acknowledgements

B. Bhargava and L. Lilien gratefully acknowledge ideas derived and insights gained from participation in the NSF Cyber Trust Point Meeting [Nati03], lead by Carl Landwehr with Helen Gill, Bhavani Thuraisingham, and Ty Znati. The meeting included panels on *Trust and Usability*, chaired by Mike Reiter; *Privacy Policies and Mechanisms*, chaired by Giuseppe Ateniese; *Pervasive Security*, chaired by Srini Devadas; and *Trust and Economics*, co-chaired by Joan Feigenbaum and Michael Smith.

L. Lilien is also happy to express his thanks for some of the ideas included in this report to the moderator and the participants of the IICIS 2003 Panel Session on *Data Quality* [Thur03].

References

[Nati03] National Science Foundation Cyber Trust Point Meeting Information, Johns Hopkins University Information Security Institute, Baltimore, Maryland, August 13-15, 2003, http://www.jhuisi.jhu.edu/institute/cybertrust.html.

[Land03] C. Landwehr, T. Znati, B. Thuraisingham, and H. Gill, “Introduction,” in: [Nati03], August 14, 2003.

[Thur03] B. Thuraisingham (moderator), *Data Quality*, Panel Session, *Sixth IFIP TC-11 WG 11.5 Working Conference on Integrity and Internal Control in Information Systems (IICIS 2003)*, Lausanne, Switzerland, November 13-14, 2003.