

Security with Privacy*

ISAT 2002 Study

13 Dec 2002

ISAT Security With Privacy/13Dec02/Final

*See "Notes Page" in this PowerPoint presentation for full annotation of slides

Table of Contents

I	Executive Summary	2
II	The Privacy Problem	2
III	Research Agenda:	
	Selective Revelation	10
	Strong Audit	13
	Rule Processing	16
IV	Concluding Comments	23

Executive Summary

Privacy of personal data is an absolutely essential element of any information system that carries information about American citizens. But the challenge of privacy sharply increases as the use of information aggregation systems continues to grow -- both in commercial and government spheres. This study examines specific technological agendas for increasing privacy.

This study examined privacy within the context of national security, particularly from the viewpoint of data aggregation systems. The study recommendations are technology oriented rather than policy oriented, keeping within its charter. Our thesis is that technology can allow us to make substantial progress towards supporting *both* privacy *and* national security in information aggregation systems. Technology can also assist in providing information that will support policy decisions about how to handle private data.

This study recommends three key technical strategies: Selective Revelation, a method for minimizing exposure of individual information while enabling continuous analysis of potentially interconnected data; Strong Audit, a tamper-resistant method that identifies where data goes and who has seen it; and Rule Processing Technologies that guide how data from multiple sources with potentially different privacy constraints can be processed.

Privacy is a key issue for DARPA and for our society at large. We urge DARPA to pursue research in this important area.

Introduction

This is the report of an ISAT study conducted in 2002.

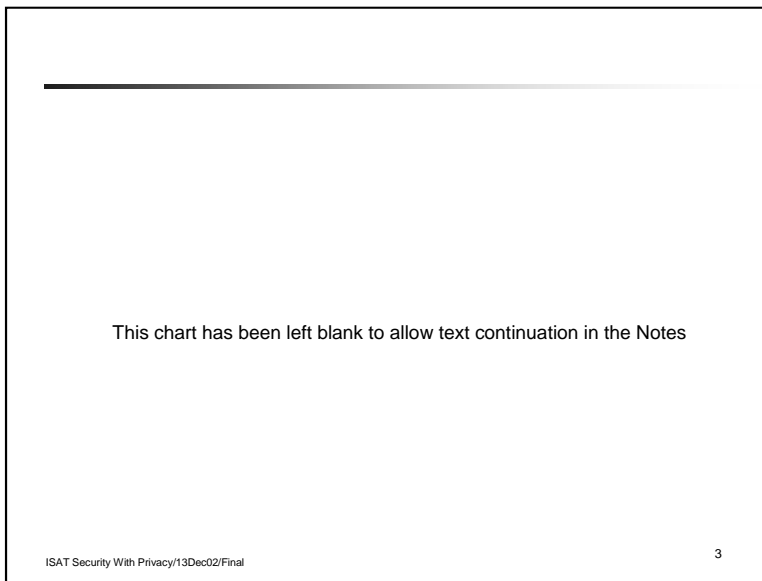
The study examined privacy -- in particular in the context of data aggregation systems. Such systems are ubiquitous in the private sphere (consider, for example, the stunning amount of personal data collected by the private firm Acxiom.) But although the most serious privacy questions exist in private systems, concerns are also raised in government systems -- both existing and proposed. A large number of government agencies, ranging from the IRS to the FBI to the Social Security Administration to the Immigration and Nationalization Service, today collect large amounts of personal information.

A number of continuing trends (such as the spread of the Internet, the increase in electronic payment methods, near-universal use of cellular phone communications, ubiquitous computation, sensor webs, etc.) mean that private organizations will have increased ability to build detailed profiles of individual Americans. And on the government side, there have been widespread calls in the media in the wake of the September 11, 2001 terrorist attacks for government organizations to share information and "connect the dots."

Privacy poses significant, challenging problems. These are not merely hypothetical problems: commercial use of data means that these problems are present, important, and pressing today.

DARPA is a particularly appropriate agency to tackle research in privacy. First, DARPA has a history of solving apparently "unsolvable" problems. Rather than relying on limited "incremental" research initiatives often found in non-DARPA programs, DARPA has a history of tackling the hardest problems and coming up with powerful, useful solutions. DARPA is a true research organization, and can rally the abilities of leading scientists to tackle these important problems. Second, DARPA currently has a number of programs in its "Information" offices: the Information Processing Technology Office, the Information Awareness Office, and the Information Exploitation

(continued on next page)



(continued from previous page)

Office, which involve the potential use of information derived from sensors, distributed systems, and government and private databases. Third, although the mission of DARPA is to support technologies for the Department of Defense and National Security community, DARPA programs have a long history of technology transfer to the general public. Thus, if DARPA tackles and makes progress on the privacy problem, there is substantial reason to believe that these technologies will disseminate into common use in the commercial sphere.

The ISAT study considered these problems in the true sense of research -- we can not prejudge their success. And the challenges are enormous. However, given DARPA's history of regularly solving very hard problems, we have decided to examine privacy issues even though many researchers will likely agree that the questions are hard.

Privacy is an issue that includes both technical and policy elements, and in many of our discussions both issues came up. However, the expertise of ISAT (and of DARPA) is in technology, and so our primary recommendations are on privacy research topics. While we may touch on related policy concerns, our policy discussion should be understood with caution -- we have no special expertise in policy.

A few comments on what this study is *not*: This study is *not* a critique or endorsement of any particular DARPA program (including programs in IPTO, IAO, and IXO.) This study is *not* an attempt at policy recommendations. This study is *not* a review of Total Information Awareness (although we did at times consider TIA's Genisys as an example of a system with ambitious privacy goals.)

This study *is* an attempt at a high-level research agenda for Privacy with Security. This agenda is appropriate for several DARPA programs, and we believe that the importance of these problems (and recent technical advances) emphasize the privacy problem. Again, these are *research* problems -- although we believe substantial progress can be made on this problem, no definitive statement can be made until the research is actually attempted.

The ISAT study did not attempt to reach consensus on all issues. The study drew on the widest possible range of views, ranging from opinions from well-known and widely quoted heads of major privacy organizations to opinions from individuals with substantial law-enforcement and operational intelligence experience. Rather than attempting to reconcile these various views, we used the input from the people to try to identify *technical* areas where research dollars could make a difference in end-level privacy. Since we do not try to reach consensus, this document should not be interpreted as necessarily reflecting the opinion of any single person involved with the study. Rather, it is an anthology of interesting technical ideas raised by various individuals who contributed to the study.

Why privacy? Why now?

- Ripe area -- these problems are ready to move
- Central problem for both commercial and government spheres.
- Example of system with strong privacy goals:
Total Information Awareness
- Substantial spill-off technologies:
 - improved computer security
 - privacy in commercial sphere
 - better intelligence processing

Study Discussion and Observations

If we can make progress on the privacy problem, we will benefit from many powerful spin-off technologies. Better privacy means better handling of sensitive information, which can directly lead to better computer security. If we can deny attackers access to sensitive information, we raise the bar on the ability of attackers to successfully attack systems. This point lies behind much of the concerns raised by the President's Critical Infrastructure Protection Board's September 18 *Draft National Strategy to Secure Cyberspace*.

In addition, we can also hope that this technology will see use in the commercial sphere, as mentioned above.

Finally, in the wake of the terrorist attacks of September 11, 2001, intelligence handling by several government agencies, including the FBI, INS, and NSA, were widely criticized in the media. Part of this poor handling came from awkward handling of "private" information (that is, information that was about private individuals). If we had better privacy controls, we could hope that critical information could reach the right people while protecting irrelevant personal information from being disclosed, yielding improved intelligence.

Security **with** Privacy

- Traditional: national security or privacy: **choose one**
- It doesn't have to be an either/or choice
- Difficult issue because of strong opinions
 - National security: 9/11
 - Privacy: Clipper, Carnivore, TIPS
- Charter: Technologies for national security **with** privacy
 - Technology focus

We are interpreting "Security with Privacy" to mean "National Security" (more particularly, Homeland Defense) with Privacy.

As government information systems supporting intelligence and law enforcement counter-terrorism activities are put into place, and as these systems increasingly draw on outside sources of information such as commercial entities and other governments, we face the challenge of reconciling the increased capabilities of these systems with the need to protect individual privacy.

In any case, the deployment of powerful distributed information systems, as envisioned by the Transportation Security Agency, the FBI, and TIA will need powerful privacy mechanisms or else the American people (rightly so) will refuse to accept deployment.

An additional point that was raised is that in many cases, intelligence agencies are reluctant to share data because of concern that other organizations may not be scrupulous in respecting both the privacy and secrecy of sensitive data. While we cannot directly comment on this concern, to the extent that it exists, it will benefit from improved privacy technologies.

Interesting questions **outside** scope

- We are **not** studying
 - Weighing national security vs. privacy
 - Issues relating *only* to national security
 - How to do data mining better
 - Any specific DARPA program or office

Privacy is a controversial, “hot button” topic, and people have strong opinions on the topic. We don’t want this study misunderstood, and we want to clearly explain what this study does not attempt to cover. The disclaimer on this slide is meant to help prevent several possible misinterpretations. This study is not attempting to balance concerns against improving privacy or national security or topics relating purely to national security. Neither is it a manifesto on data mining -- that field has its own substantial research challenges, but these are outside the scope of this study.

And, as mentioned above, this study does not attempt to critique any given program. (Although we would like to thank representatives of both IPTO and IAO for several useful high-level briefings on their research programs.)

Why now?

- **Need**
 - Driven by homeland defense
 - Explosion in commercial collection, use, & sharing of personal data
 - Rapid expansion of government information systems
 - Increased commercial surveillance technologies
- **Opportunity to leverage recent tech advances**
 - Tamper-evident logging/timestamping
 - Search on encrypted data without revealing query
 - Automated synthesis & verification of crypto protocols
 - Dynamic coalitions
 - Proof carrying code & related semantic analysis
 - Static security analysis of code

ISAT Security With Privacy/13Dec02/Final

7

Of course, privacy has been and will continue to be a central concern. So why study these issues now? There are four key reasons:

- (1) This is an important area, and will likely benefit from DARPA's attention
- (2) Commercial collection and exploitation of data is exploding.
- (3) Homeland defense is likely to need improved information handling, and thus raise serious privacy questions.
- (4) From a technical perspective, this is a ripe time to build on recent scientific advances.

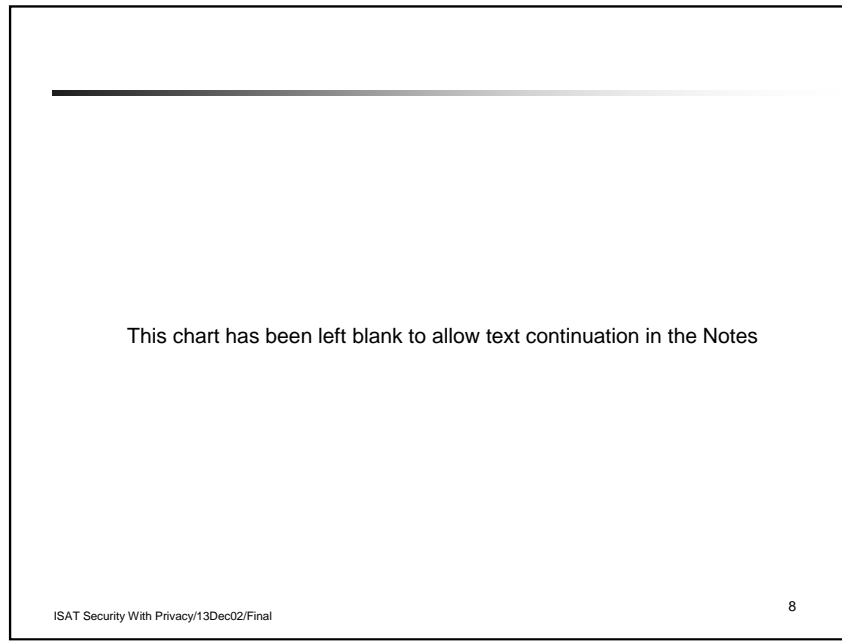
There are opportunities to leverage recent advances in technology:

- Starting with research by Haber and Stornetta¹, researchers have built powerful techniques for building tamper-evident logs. Given the power of this technology, it is surprising that private industry has not pursued this area. DARPA's attention will help us build logging techniques that can record misuse of data.
- In recent work by Dawn Song, Adrian Perrig, and David Wagner², powerful techniques have been presented that allow information to be searched in data repositories, without revealing the nature of the queries or the results either to eavesdroppers or the data repositories itself. This work, which has since been built on by a number of researchers, offers a surprising result -- we can build databases that are secure both in conventional senses and in the sense of distributed computation. This suggests that a number of extensions may be possible: (a) government agencies may be able to use data from private organizations such as Acxiom without revealing the nature of inquiries or searches to Acxiom. Since private commercial organizations often have appalling security (example: the widely reported recent "identity theft" of highly sensitive private information using poor security at private credit agencies and their clients), protecting the nature of queries is a central concern for effective government use of private information. (b) To the extent that privacy handling rules can be expressed in computer readable format, it may be possible to enforce

1- Ref: <http://www.surety.com/solutions/DN/presentation.html>

2- Ref: <http://paris.cs.berkeley.edu/~dawnsong/papers/se.pdf>

(continued on next page)



(continued from previous page)

privacy restrictions within the framework of an automated processing system. Now, the study does not mean to imply that this technology is ready to use “off the shelf” -- it does not fully support the functionality listed above and has efficiency issues. But the theoretical success of the Song/Perrig/Wagner approach suggests that we will be able to make real progress on an ideal system that will be efficient and support the above goals.

- Recent research at a number of institutions, including UC Berkeley, Stanford, the Naval Research Laboratories, and elsewhere has suggested that we are making progress on the cryptographic protocol verification problem -- we can find errors in real cryptographic protocols. In fact, the Athena system developed at Carnegie Mellon University and UC Berkeley has even suggested that strong cryptographic protocols can be efficiently synthesized given a formal description of the properties desired. This suggests that we may be able to extend this technology to examine and synthesize the complex cryptographic protocols needed by information systems that attempt to allow information to be shared across different organizations -- each with their own important and central privacy policies.
- The DARPA Information Processing Technology Office (and its predecessor, ITO) have pioneered the effective use of *dynamic coalitions* that allow different organizations to work together and share information. This reflects the way that many organizations are likely to work in case of a crisis -- different countries and organizations may choose or decline to cooperate in the face of specific incidents. Again, this technology is evocative, but will need specific work to be adapted to specific government and Defense programs.
- Building on the research of Peter Lee and George Necula³ in DARPA sponsored work, a number of researchers have built systems for proof carrying code that allow programs to carry certain types of specifications and proofs that those specifications are met. Extensions of this work may allow queries or mobile code to be transferred among various high-privacy data repositories while respecting privacy constraints.
- Similarly, static semantic analysis has rapidly advanced in the last five years allowing both many security bugs to be identified and properties to be verified. This may be adaptable to privacy concerns.

It is notable that most of the work cited above is a direct result of focused DARPA investment in core computer security, language, and distributed systems research. DARPA's investment has already paid off substantially, and DARPA should consider the number of technologies mentioned here that would not today exist without DARPA's direct involvement.

3 - Ref: <http://raw.cs.berkeley.edu/pcc.html>

Key strategies

- Selective revelation
- Strong audit
- Rule processing technologies

The technical approaches we explored are Selective Revelation, Strong Audit, and Rule Processing Technologies. They are discussed in the following charts.

In an earlier version of this report, we had a different categorization of fundamental technologies. For the sake of completeness, here is that earlier breakdown of our technical recommendations (these points are elaborated in this report in a different format):

Technology challenges

- Accurate labels for derived data
- Formal language for expressing privacy rules
- Simulator for testing policy alternatives
- Privacy toolbar
- Tamper-evident distributed audit

Fundamental research topics

- Privacy & human factors
- Distributed information flow security
- Advanced crypto protocols
- Adaptation

We also suggested some policy recommendations in a previous draft (although policy is outside the scope and expertise of the committee):

First, we suggested a citizen advisory board to inform and shape policy on privacy rules. Such a board could help inform technical directions for privacy research. Second, we encouraged DARPA or some other government organization to support examination and a voluminous list of research on privacy laws -- even lawyers often informed us that current US privacy law is a maze and it is often hard to understand what privacy constraints exist.

The following charts focus on key technical challenges.

Strategy: Selective revelation

- Architecture based on *selective revelation*
- **Goal:** minimize revelation of personal data while supporting analysis
- **Approach:** partial, incremental revelation of personal data
- **Procedure:**
Initial revelation by statistics & categories
Subsequent revelation as justified by earlier results
- Supports both “standing” & real-time queries

ISAT Security With Privacy/13Dec02/Final

10

Selective Revelation is a method for minimizing exposure of individual information while supporting continuous analysis of all data. The challenge in doing this is that much of the data is private information about people which cannot necessarily be revealed to a person. Even data that is derived via an analysis algorithm may be private, depending on the status of the data from which it was derived, and on the nature of the algorithm.

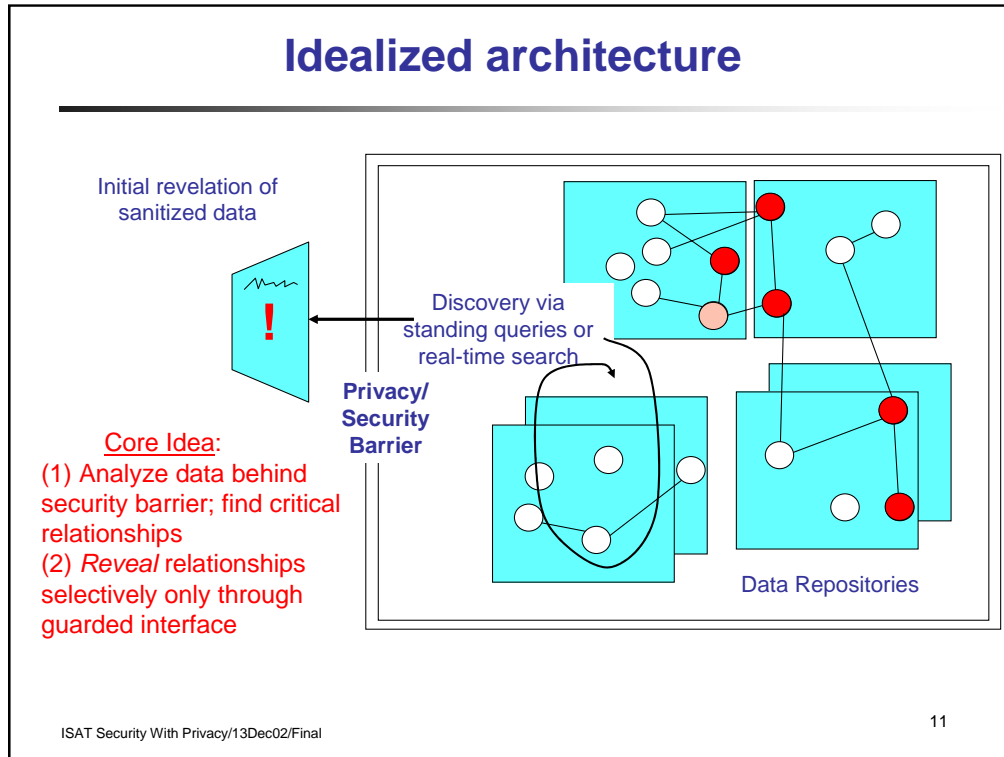
The idea of selective revelation is that initially we reveal information to the analyst only in sanitized form, that is, in terms of statistics and categories that do not reveal (directly or indirectly) anyone’s private information. If the analyst sees reason for concern he or she can follow up by seeking permission to get more precise information. This permission would be granted if the initial information provides sufficient cause to allow the revelation of more information, under appropriate legal and policy guidelines.

For example, an analyst might issue a query asking whether there is any individual who has recently bought unusual quantities of certain chemicals, and has rented a large truck. The algorithm could respond by saying yes or no, rather than revealing the identity of an individual. The analyst might then take that information to a judge or other appropriate body, seeking permission to learn the individual’s name, or other information about the individual. By revealing information iteratively, we prevent the disclosure of private information except when a sufficient showing has been made to justify that revelation.

Selective revelation works by putting a security barrier between the private data and the analyst, and controlling what information can flow across that barrier to the analyst. The analyst injects a query that uses the private data to determine a result, which is a high-level sanitized description of the query result. That result must not leak any private information to the analyst.

Selective revelation must accommodate multiple data sources, all of which lie behind the (conceptual) security barrier. Private information is not made available directly to the analyst, but only through the security barrier.

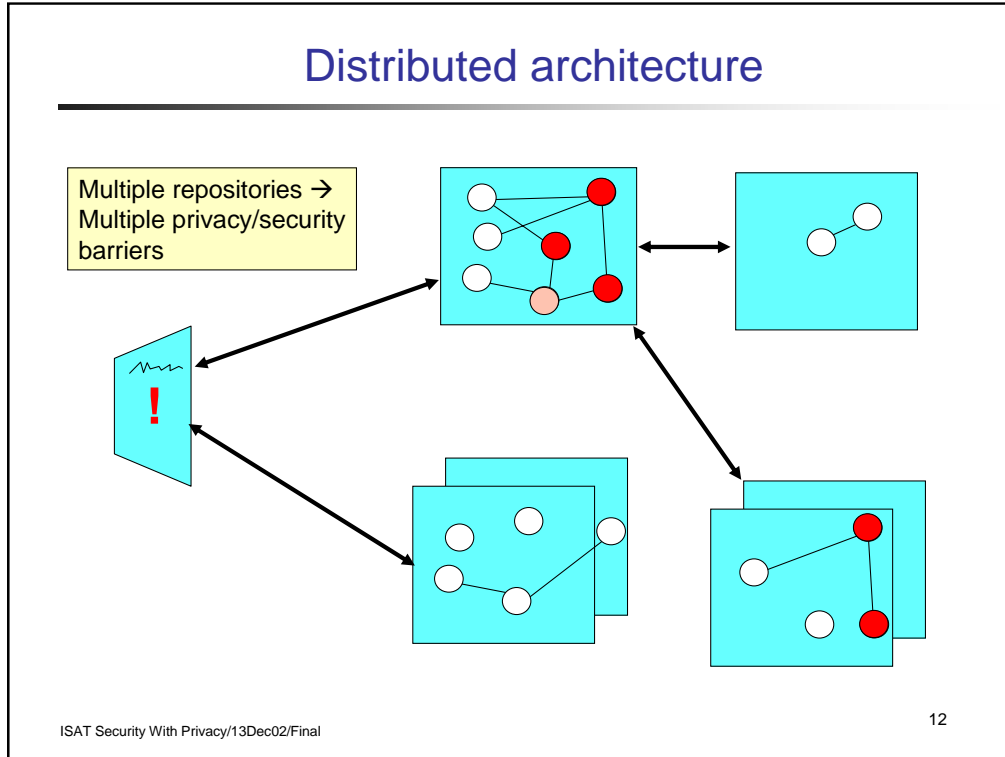
A key technology challenge is making data relationships successively refinable, and thus selectively revealable. One way to address this challenge is to develop data ontologies that provide structured and logical ways for the analyst to make appropriate queries.



It is easiest to think of a protective privacy/security barrier as existing outside a single monolithic repository, as shown in Slide 11. However, the single monolithic repository will not exist. In the sort of systems we envision, a key feature is cooperation across multiple repositories.

Slide 12 illustrates this issue. In this sort of system, there can be no central privacy/security barrier -- each repository must have its own barriers, and those barriers must be coordinated to support privacy restrictions of the system as a whole and of the individual repositories.

Such a distributed privacy barrier has some resemblance to the "privacy appliances" mentioned in a November 11, 2002 interview with Admiral Poindexter that appeared in the *Washington Post*. However, Admiral Poindexter was quoted as speaking about the Total Information Awareness system, whereas we have a much broader goal here -- the general protection of information shared across organizational boundaries whether in commercial or government systems. In any case, research is merited in finding ways to protect privacy shared across organizations, and we suggest that DARPA actively pursue this topic.



(Comments about this Slide on preceding page.)

Audit

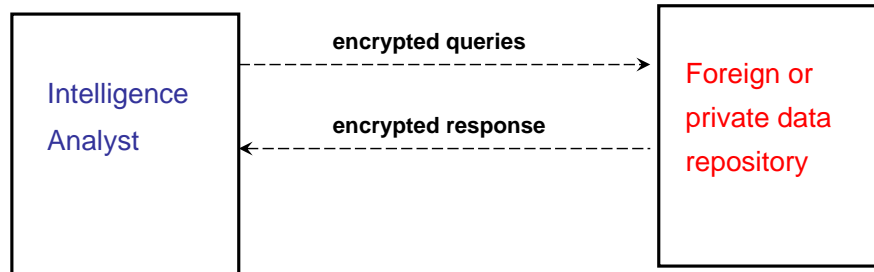
- Protect against abuse by “watching the watchers”
- Design goals: Distributed audit
 - Everyone subject to audit
 - Cross-organizational audit
 - Measure accuracy of auditors by cross-validation
 - Usage records are tamper-evident
- Hall of Mirrors: Audit is recursive to original data mining problem
 - Data sets are voluminous
 - Usage records are sensitive

Perhaps the strongest protection against abuse of information systems is **Strong Audit** mechanisms. We need to “watch the watchers.” These audit systems must be tamper-evident or tamper-resistant, and since repositories span different organizations, must themselves span different organizations. If such audit mechanisms exist, we will realize substantial advantages. (For example, a strong audit mechanism would have been likely to identify a spy such as Aldrich Ames or Jonathan Pollard very early on.)

However, these audit systems themselves pose a substantial challenge. Audit data will be voluminous and highly sensitive (certainly, foreign intelligence agents would be very interested in finding out what sorts of queries are run through US commercial or governmental information systems.) How can we find instances of inappropriate queries?

In many ways, this is a recursive instance of the general intelligence data mining problem, and should probably be considered in conjunction with that problem. This hall of mirrors presents a number of technical challenges, and would benefit from DARPA’s attention.

Example technology: encrypted search



- Queries are sent encrypted
- Queries are processed but **not decrypted** by repository
- Repository prepares response but does not know what search was or whether it was successful.

This example is discussed in the commentary on Slide 7 relating to the Song/Perrig/Wagner work.

Crypto protocols

- Need new classes of protocols to better support
 - Audit compliance
 - Selective revelation
- Types of protocols:
 - Searching on encrypted data
 - Oblivious transfer and extensions
 - Negotiation
 - Escrow
 - Distributed audit
- Current protocols limited by
 - Functionality (special cases)
 - Scalability
 - Efficiency

ISAT Security With Privacy/13Dec02/Final

15

At the heart of contemporary security technology is work on cryptographic protocols. These protocols allow us to build systems with a wide variety of properties -- and it seems promising for research in privacy. (Much of the fundamental work on cryptographic protocols in distributed systems arose from basic research funded by DARPA.)

Unfortunately, some current cryptographic protocols are currently of greater theoretical than applied interest. Current protocols are often limited in functionality to special cases, are limited in the number of parties they can support (scalability), and use excessive amounts of computation resources. DARPA can make a wise investment by supporting fundamental and applied research in cryptographic protocols that appear likely to support privacy properties, such as protocols that support audit compliance or selective revelation.

In addition, DARPA should consider supporting general research on protocols, including work on verifying and synthesizing a variety of cryptographic protocols.

Better cryptographic protocols will not only result in better privacy, but also better computer security in general.

Rule processing technologies

- Labels record attributes of data, e.g.:
 - US person / foreign person / nationality unknown
 - Origin of data (e.g. credit card companies, INS)
 - Reliability/freshness of information
 - Data subject to specific law or agreement

- Label + policy rules → limitations on use

- Challenges:
 - Derived data
 - Legacy data

Rule Processing Technologies. Distributed information systems combine data from diverse sources. Their privacy systems must support privacy constraints: both systemic privacy constraints and privacy constraints specific to a particular set of information repositories. (For example, information derived from a foreign source, such as country X's information repositories, may come with specific privacy concerns attached.) Since computers in general cannot understand the underlying representation of private information, it is necessary to label data with information that will allow it to be properly processed, both with respect to privacy constraints but also with respect to general constraints.

Information varies tremendously in quality as well. For example, substantial anecdotal evidence supports the claim that significant data appearing in commercial credit bureau sources is not always accurate. Information from foreign sources may be tampered with. Government agencies vary in the degree of scrutiny they apply to keep data accurate.

All of this poses issues for accurate labeling. Further concerns arise because even a new information system will likely build on substantial amounts of (unlabeled or inaccurately labeled) previously existing "legacy data."

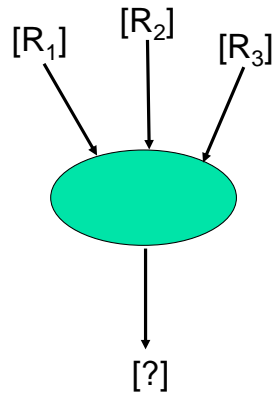
And when data is combined to produce "derived data," how should the derived data be labeled? A conservative approach would suggest labeling the derived data with the most conservative labels possible. However, in many cases, this will be inappropriate -- derived data is often sanitized and poses less privacy restrictions than the original source data used. On the other hand, in some cases derived data may actually be more sensitive than the original source data.

Data labeling is actually an old idea – it dates back to some of the "pre-Orange Book" discussions in the security community. And it is being widely used in a variety of new systems today, including digital rights management systems (DRMs). Indeed, data labeling is purportedly a key feature of several commercial DRMs.

Labeling Derived Data

- Conservative approach: output inherits all restrictions of inputs
 - Often too restrictive
 - Sometimes too liberal
- Hard problem
- Seek semi-automated solution to minimize human overhead
 - Idea: use recent work on program semantics

Example: derived restrictions



This slide illustrates the challenges associated with labeling derived data. See the commentary on Slide 16 for further details.

Privacy rules

- Need language for expressing rules
 - Related technology: Digital Rights Management
 - Translate English → agent based language
- Rules differ based on data
 - Types of data (foreign vs. domestic, video vs. textual)
 - Contents of data
- Need tools for compliance checking
 - Both automated and human in the loop

At the heart of any privacy system will be the ability to express rules for handling private information. These rules must be readable both by machine (so that they can be electronically enforced) and by humans (who can check the rules for accuracy).

Similarly, compliance to these rules must be checked (and checkable) both automatically and by people.

Privacy Toolbar

- Rules are highly complex
- Need real-time toolbar to guide user through the maze
 - Helps users produce required documentation to support actions
 - Show privacy status of information
 - Highlight compliance requirements
 - Support audit functions
- Help analyst understand
 - Why the system said “no”
 - What to change to get to “yes”
 - What laws/rules apply to a situation
 - How rules interact
 - How to ask permission for more access

Several lawyers told the ISAT study group that current US privacy law and practice was so complicated that no single person fully understood all the issues. Different types of data have different sort of privacy rules.

This poses risks in multiple directions. On the one hand, there is the risk that current complexity of US privacy laws and rules may result in inappropriate disclosure of information. But, the ISAT study heard a report from a former Justice Department official that in many cases intelligence analysts and law enforcement personnel miss the opportunity to use essential intelligence information: in their desire to comply with privacy rules, the government officials fail to use material that they are legally entitled to use. In other words, the haze of privacy law makes officials go beyond legislative and regulatory privacy restrictions and means that the government misses the chance to “connect the dots.”

Clearly, we have a significant challenge in allowing users of databases (whether employees of companies such as Acxiom or law enforcement officials or intelligence analysts) to reasonably understand what the real privacy restrictions are. Here is a place where technology can help -- if we can develop a “privacy toolbar” that helps inform users of privacy restrictions, we can help eliminate mistakes caused by human misunderstanding of the United State’s currently complex privacy law. This especially applies when rules interact. If a privacy restriction is reached, such a system can help explain the procedures necessary to legally access data.

Furthermore, such a system can help record annotations on how and why data is being used, which will make audit logs richer and help reconstruct use of data. Such reconstructions of data use will be invaluable: both for helping to improve information systems and for helping to detect misuse.

Information Policy Simulator

- “What if” tool for privacy rules
 - Generates synthetic data & runs queries against them
 - Allows development, science, and manipulation of abstract privacy models
- Benchmarks system behavior
 - Resource use, false positives/negatives, effect of data errors
 - Probe potential tradeoffs: gains in accuracy vs. losses in privacy
 - Testbed for developing search strategies
- Research issues
 - Design of simulator
 - How to generate synthetic data
 - Verification/validation

ISAT Security With Privacy/13Dec02/Final

20

Moving towards more advanced and speculative research, we envision a system which can simulate different information handling policies. Such a system would use synthetic data and run queries against them. By comparing different privacy policies, we aspire to find examples that will help illustrate respective advantages and disadvantages of a variety of privacy policies.

This could help inform debate by policy makers as they consider different privacy policies. (This stands in marked contrast to contemporary approaches to privacy policy making, which is often marked by political rhetoric and vague sociological characterizations.)

However such a simulator faces substantial challenges: we need to design the simulator itself, we must find a way to generate meaningful synthetic data, and we must find ways to verify or validate the accuracy of reports from the simulator. These are all hard problems, and their solution is far from obvious. This is an example of “high-risk” (the challenges are real), “high-payoff” (even partial solutions can help shed considerable light on policy making) research.

DARPA is the pre-eminent sponsor of high-risk, high-payoff strategy. DARPA is likely to find this is an exciting direction for further consideration.

Distributed information flow security

- When can information be sent from A to B?
- What promises must B make?
- How will B's system enforce these promises?
- Does A trust B to keep his promises?

- Research issues
 - Translating high-level (ambiguous) policies into concrete rules
 - Possible approach: Formal language + compiler for privacy/access
 - More than just DRM
 - Measure amount of private information, impact on privacy
 - Fine-grained adaptive access control

The sort of distributed information systems we envision in this study will combine data from a variety of sources. Each source will have its own restrictions, and each of those sources needs to be sure that users will respect its privacy considerations.

These concerns are specific examples of the more general problem of distributed information flow security. How can A trust B to handle information correctly.

Another well known instance of this problem arises in digital rights management (DRM) systems. In the last year, these have been widely deployed in commercial software, and with notably mixed success. However, the problems raised by privacy go far beyond what existing digital rights management systems use.

However, DARPA has some advantages in considering these problems. For example, DARPA can consider distributed information flow security in environments that are significantly more heavily supervised and controlled than in commercial operating system deployment. DARPA can consider the use of specialized hardware (such as the previously discussed privacy appliances) that can help enforce restrictions.

This is a fascinating research topic, and one that will have considerable spin-off value. DARPA would benefit highly from investments in distributed information flow security research.

Challenge for Privacy: Adaptation

- Understanding adaptation is an open problem
 - Bad guys change behavior to avoid getting caught
 - Good guys change behavior to avoid hassle, protect privacy
- **Vicious circle leads to loss of privacy and denigration of data**
 - Bad guys learn rules, share info, adapt behavior, reduce signal
 - Analysts counter: make rules more complex, dig deeper into private data
- Constructs from game theory & economics to break out of cycle
 - Understand limits of privacy policies in face of adaptation
- Potential testbeds
 - Gaming simulations, multiplayer worlds

Consider the problem of adaptation. As people realize that certain data is subject to surveillance or response, they change their behavior.

Here is an example familiar to any frequent flier: prior to the terrorist attacks of September 11, 2001, many experienced airline passengers angled to be among the very first in line to board commercial airplanes -- in that way, the passengers could place their carry-on luggage and settle in before the rest of the passengers boarded.

In the wake of the the September 11, 2001 attacks, authorities instituted thorough searches of some passengers (and their carry-on luggage) boarding flights. While these checks are ostensibly random, they in fact tend to select more highly from the first people to board a flight. Now experienced travelers angle to be the tenth in line instead of the first in line.

In the same way, information systems designed to identify certain groups of people are likely to result in different behavior from both the people they are intended to track as well as innocent people who for personal reasons desire to evade surveillance (for example, observe the "arms race" between telephone caller-ID systems, those who desire to make anonymous calls, those who desire to reject anonymous calls, etc.) Failure to correctly anticipate this sort of adaptation is likely to lead to unexpected and (often undesirable) results.

In the worst case, this pattern of adaptation could lead to widespread evasion of surveillance, followed by a counter-move of analysts who need to dig deeper into private data, leading to a spiral resulting in markedly decreased privacy in practice. To some degree, simulation as mentioned on Slide 20 may help prevent this. But beyond that, it makes sense to attempt to use highly distributed gaming testbeds to try different policies.

Conclusions: Security with Privacy

- Safe, private handling of information enhances freedom, democracy, and national security
- DARPA-hard technology problems
 - Behind the curve in understanding these issues
 - US now faced with significant security problem
- Potential R&D and solutions promise many benefits
 - In both government and commercial systems
 - Effective information systems enhance homeland defense
- Our aspiration must be security with privacy
- Key enabling technologies:
 - Selective revelation
 - Strong audit
 - Privacy rule processing

The American people need systems that protect national security together with privacy. Just security or privacy by itself is unthinkable.

DARPA can play an important role in furthering technical methods for privacy. The problem is difficult (in computer science slang, it is “DARPA-hard” -- as hard as any challenge that DARPA faces) but recent technical developments suggest that full or substantial partial solutions may be possible. Moreover, both rapid development of information aggregation in the commercial sphere and post-9/11 government activities make the privacy problem especially pressing.

Concluding Comments

This study focuses on technical recommendations. It has benefited from the input of many people, but it does not necessarily reflect opinions of each of the study members. DARPA may wish to consider commissioning a more far-reaching study that can consider policy recommendations in detail.

Some media sources have indirectly referred to this study as a review of DARPA Total Information Awareness program. These reports are not accurate. The ISAT study group did not attempt to review TIA or any other DAPRA program. While these recommendations may help inform TIA (and other programs), the recommendations both go beyond the scope of TIA (for example, in considering commercial aggregation of data) and also do not address significant portions of TIA (for example, we do not consider data mining.)

These recommendations in this study are primarily *research* recommendations. Since they are research, we expect that if DARPA sponsors research in all of the areas we recommend, some approaches will be highly effective and others will be less effective. Without pursuing the research, we can not say definitively which approaches will work. These recommendations are not *development* recommendations.

But the committee unanimously feels that privacy is a key issue for DARPA and for society at large. We urge DARPA to pursue research in privacy.