Results of Workshops on Privacy Protection Technologies

Carl Landwehr

IARPA
Office of the Director of National Intelligence
Washington, DC
landwehr@isr.umd.edu

Abstract. This talk summarizes the results of a series of workshops on privacy protecting technologies convened in the fall of 2006 by the Office of the Director of National Intelligence through its Civil Liberties Protection Office and the (then) Disruptive Technology Office (now part of the Intelligence Advanced Research Projects Activity, IARPA).

Keywords: privacy protection technologies, strategies, workshop, intelligence community.

1 Introduction

In the after lunch speaking spot, the organizers think I can keep you awake after lunch, I thought I would start with a joke, a cartoon indicating where data sharing may lead us which says "Oh, good. My complete sexual history is on tonight." As a couple is watching television[1].

I will review a series of workshops on privacy protecting technologies that were held in the fall of 2006, under the auspices of the Office of the Director of National Intelligence (ODNI) Civil Liberties Protection Office and the Disruptive Technology Office (DTO). In October 2007, DTO became part of the Intelligence Advanced Research Projects Activity (IARPA), which is continuing to pursue research opportunities in this area[2].

I believe, and I think a large part of the intelligence community agrees, that we are not trying to advance security at the cost of privacy but that we strive to advance both security *and* privacy together.

2 Privacy and Security

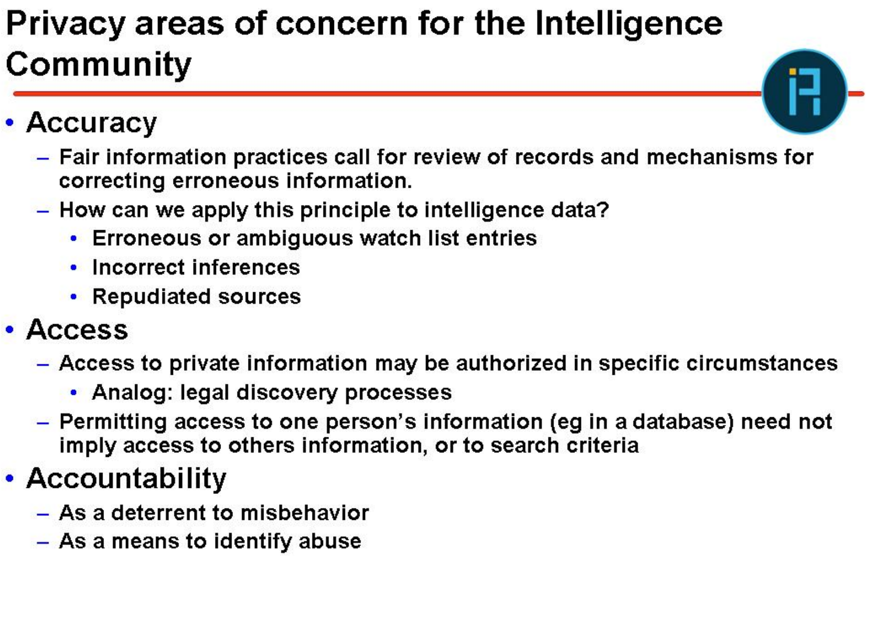
It's difficult sometimes to think about what privacy means for the intelligence community (IC) because, in some sense, the IC is interested in compromising the privacy of the "bad guys", the targets, in a big way. But the problem is to do that without compromising the privacy of everybody else.

[1] http://www.cartoonbank.com/product_details.asp?sitetype=2&sid=40949&pid=1003&did=4

[2] Toeffler Report, this report has not yet been published by the U.S. Government.

C.S. Gal, P.B. Kantor, and M.E. Lesk (Eds.): ISIPS 2008, LNCS 5661, pp. 45–56, 2009.

Figure 1 lists three primary areas of concern that relate to privacy and the intelligence community. One is *accuracy*, if you look at fair information practice principles, they typically call for review of records and mechanisms for correcting erroneous information. How can you apply that policy in the context of intelligence data? Certainly there are cases where you'd like to correct erroneous or ambiguous watch list entries, incorrect inferences that have been made, sources that have been repudiated. Those problems do come up and you'd like to have solutions for those.



Privacy areas of concern for the Intelligence Community

- Accuracy
  - Fair information practices call for review of records and mechanisms for correcting erroneous information.
  - How can we apply this principle to intelligence data?
    - Erroneous or ambiguous watch list entries
    - Incorrect inferences
    - Repudiated sources
- Access
  - Access to private information may be authorized in specific circumstances
    - Analog: legal discovery processes
  - Permitting access to one person's information (eg in a database) need not imply access to others information, or to search criteria
- Accountability
  - As a deterrent to misbehavior
  - As a means to identify abuse

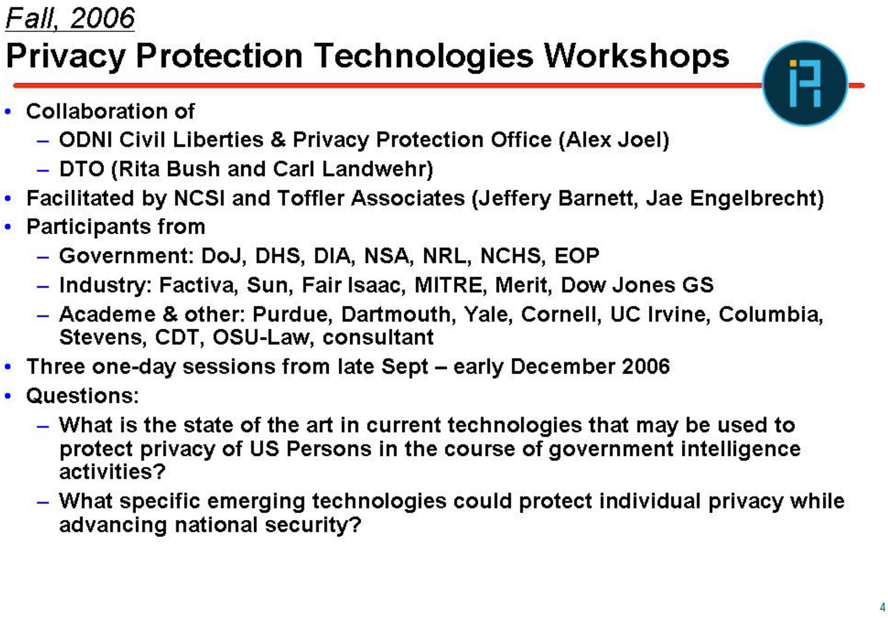
3

Fig. 1. The three A's

The second area is *access*. Access to private information can be difficult to arrange. Thinking about what's in an intelligence community database, are you likely to let the public review their record and correct it? That's going to be difficult. On the other hand, you don't want incorrect information in the records. One analog that's been suggested for this is the kinds of legal discovery processes that are in existence and that do have some steps that could enable individual access to private data under controlled conditions.

A second point under access, which I think Prof. Ostrovsky [1] has addressed in his research, is that permitting access to one person's information in a database doesn't have to imply granting access to everyone else's information, or to the criteria used in a search.

Accountability is the third general area of concern. We want *accountability* in these systems as a deterrent for user misbehavior and also as a means to identify where the system is being abused.



Fall, 2006

Privacy Protection Technologies Workshops

- Collaboration of
  - ODNI Civil Liberties & Privacy Protection Office (Alex Joel)
  - DTO (Rita Bush and Carl Landwehr)
- Facilitated by NCSI and Toffler Associates (Jeffery Barnett, Jae Engelbrecht)
- Participants from
  - Government: DoJ, DHS, DIA, NSA, NRL, NCHS, EOP
  - Industry: Factiva, Sun, Fair Isaac, MITRE, Merit, Dow Jones GS
  - Academe & other: Purdue, Dartmouth, Yale, Cornell, UC Irvine, Columbia, Stevens, CDT, OSU-Law, consultant
- Three one-day sessions from late Sept – early December 2006
- Questions:
  - What is the state of the art in current technologies that may be used to protect privacy of US Persons in the course of government intelligence activities?
  - What specific emerging technologies could protect individual privacy while advancing national security?

4

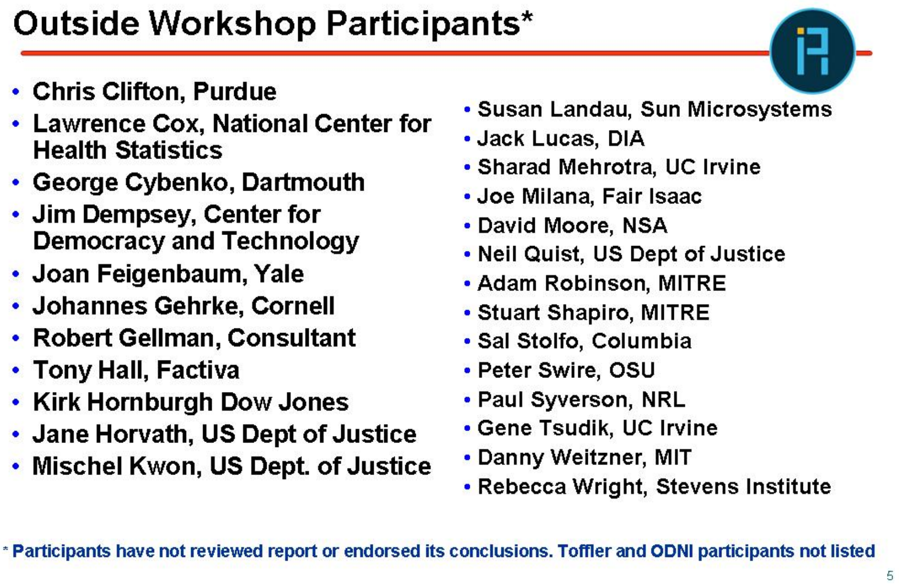
Fig. 2. CLPO-DTO workshops on privacy protection technologies

These three A's capture overarching concerns that came out of the workshops. Let me go into a little bit more detail about who was involved and what happened at the workshops. These workshops were a collaboration of the Office of the Director of National Intelligence, Civil Liberties Protection Office, including the chief, Alex Joel, and also Tim Edgar, who is here, and some others who are not here. And from the Disruptive Technology Office, Rita Bush, who was in charge of the information exploitation programs and myself, representing information assurance programs, were the primary organizers, but other members of the Disruptive Technolgy Office (DTO) participated as well. The workshops were facilitated by a conference organizer, National Conference Services, Inc. (NCSI), and also Toffler Associates, the people listed there. We had participants from a wide range of organizations as key players in the workshops. Figure 2 provides names of some of these people.

These ranged from government people including "technical" people but also people from the legal side of things and people from privacy concerns offices in different agencies. We had industry participation from people who work for companies with large commercial data mining activities and we had a fair number of academic researchers as well. We held three separate one-day sessions from late September through early December.

The primary questions we were trying to address in the workshop were first, (a) what's the state of the art of current technologies that might be used to protect US person's privacy, US person's data, privacy in the course of intelligence community activities, and second, (b) what specific emerging technologies might be brought forward in the next few years through some research funding, potentially, that could protect individual privacy while still advancing national security. So that’s what we were trying to do.

Figure 3 lists some of the people who were involved. Toffler Associates developed a report from this workshop which has had some limited distribution. Perhaps the workshop’s most important role to date has been in helping to develop language used to solicit new research proposals for privacy protecting technologies in the fall of 2007 under the NICECAP BAA.



Outside Workshop Participants*

- Chris Clifton, Purdue
- Lawrence Cox, National Center for Health Statistics
- George Cybenko, Dartmouth
- Jim Dempsey, Center for Democracy and Technology
- Joan Feigenbaum, Yale
- Johannes Gehrke, Cornell
- Robert Gellman, Consultant
- Tony Hall, Factiva
- Kirk Hornburgh Dow Jones
- Jane Horvath, US Dept of Justice
- Mischel Kwon, US Dept. of Justice
- Susan Landau, Sun Microsystems
- Jack Lucas, DIA
- Sharad Mehrotra, UC Irvine
- Joe Milana, Fair Isaac
- David Moore, NSA
- Neil Quist, US Dept of Justice
- Adam Robinson, MITRE
- Stuart Shapiro, MITRE
- Sal Stolfo, Columbia
- Peter Swire, OSU
- Paul Syverson, NRL
- Gene Tsudik, UC Irvine
- Danny Weitzner, MIT
- Rebecca Wright, Stevens Institute

* Participants have not reviewed report or endorsed its conclusions. Toffler and ODNI participants not listed

5

Fig. 3. Workshop participants (not a complete list)

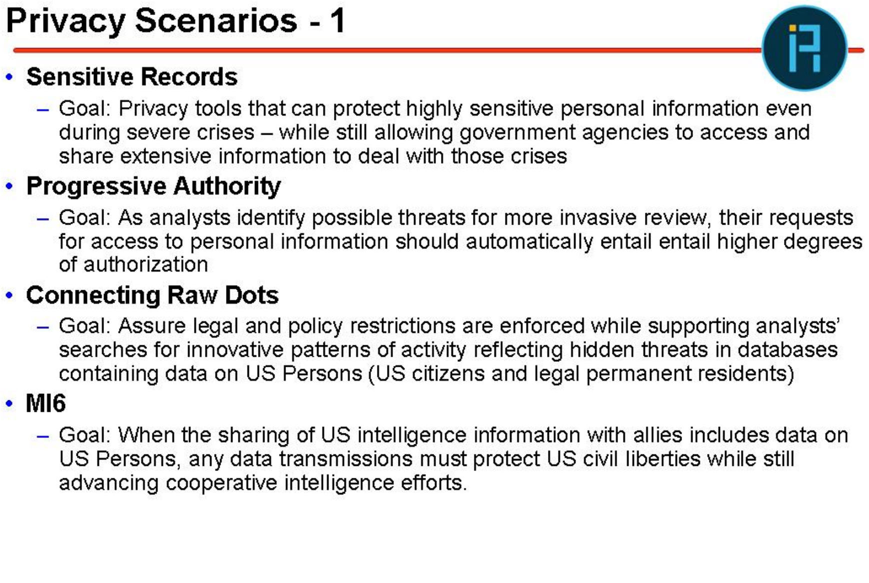
Our original intention had been to ask the participants in the workshop to provide comments on the draft report, and then to revise the report based on their comments. In the end, we were unable to do that and so we cannot represent the report as having been endorsed by the participants; it represents Toffler’s best effort to capture the sense of the discussions.

3 Privacy Scenarios

Now I will go through the scenarios that are in the report to give you an idea of how the workshop considered various situations that can arise in the intelligence community that may involve private information, potentially of United States persons.

3.1 Sensitivity and Authority

Sensitive Records. One concern is that we, that is, the government and businesses, have records that are sensitive, and we can expect that at some point in the future there's going to be another crisis of some sort. At that point there may be pressure to provide augmented access to information in certain situations. So it would be useful to have privacy tools that can function even in that sort of crisis mode and still provide some privacy guarantees. We don't want to have to say, "Well, it's a crisis, so all the privacy controls are going away". We would like to have controls that will work in that context.



Privacy Scenarios - 1

- **Sensitive Records**
  - Goal: Privacy tools that can protect highly sensitive personal information even during severe crises – while still allowing government agencies to access and share extensive information to deal with those crises
- **Progressive Authority**
  - Goal: As analysts identify possible threats for more invasive review, their requests for access to personal information should automatically entail entail higher degrees of authorization
- **Connecting Raw Dots**
  - Goal: Assure legal and policy restrictions are enforced while supporting analysts' searches for innovative patterns of activity reflecting hidden threats in databases containing data on US Persons (US citizens and legal permanent residents)
- **MI6**
  - Goal: When the sharing of US intelligence information with allies includes data on US Persons, any data transmissions must protect US civil liberties while still advancing cooperative intelligence efforts.

6

Fig. 4. Four privacy scenarios

Progressive Authority. A second scenario concerns operating, not in a crisis but in a normal situation in which you get some information about an activity or individual. As you get more information, you may be able to justify more access in relation to information about that particular topic or individual. And so, you would like to have some corresponding higher degrees of authorization required as you go up that scale.

I should say, as a general comment, that these scenarios were created to bring up representative problems and so in many cases the problems and concerns in the different scenarios overlap. They are examples that we used in the context of the workshop to try to help people understand how the intelligence community processes information and what sorts of situations can arise.

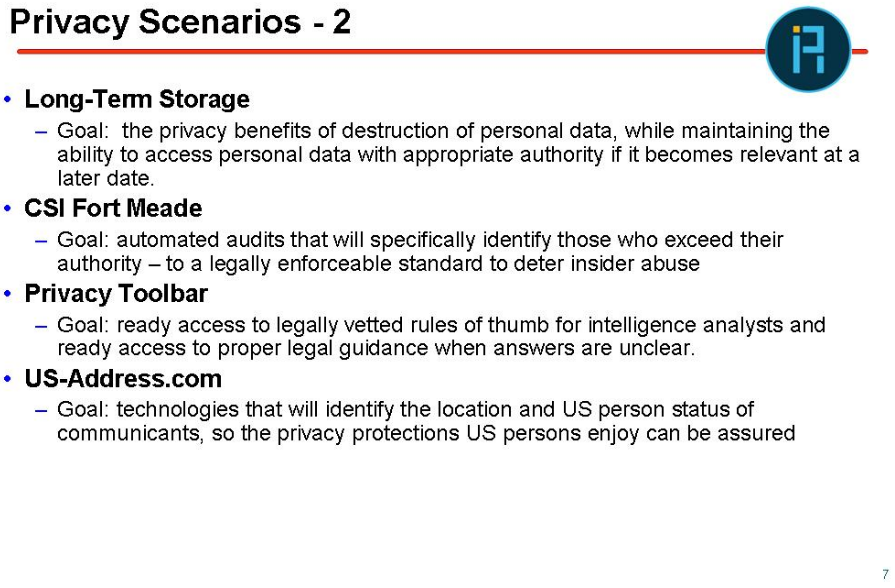
Connecting raw dots. As Maureen Baginski said this morning, a lot of people in the intelligence community believe that "connecting dots" is a terrible analogy for what they do because dot-to-dot figures are really well structured *a priori*. Perhaps a better analogy for the intelligence analysts' job would be that they have a lot of oddly shaped puzzle pieces, perhaps from several different puzzles, and they have to try to figure out which ones fit into which puzzle, and where.

Nevertheless the idea here is that we are searching for anomalous patterns of activity that may reflect hidden threats of some sort, and the databases holding this information may include some US person's information, but we want to protect the privacy of those US persons at the same time. So we would like to assure that the policy and restrictions are in place. And as Maureen Baginski said, this is done now in a very responsible way. But, it is done, in a way that demands a great deal of manual intervention, and consequently can't scale to problems of significant size.

The *"MI-6" scenario* refers to the UK intelligence organization, and captures the idea that we occasionally cooperate with foreign partners. If we want to get information from them, and they want to get information from us, we want to be sure that any information we provide in that context remains protected in their environment in the way that we ourselves are mandated to protect it. Tools to help with that kind of protection would be helpful.

3.2 Second Set

Long-term storage. This scenario has already come up in a discussion earlier at this conference. It's hard to say when information might no longer be useful for intelligence purposes. On the other hand, it's clear that the longer we keep personal information about individuals, in the present scheme of things; the more likely it is to get lost or exposed in some way. So there is a risk, in fact, of keeping information around indefinitely. So we seek to have some way of protecting it against exposure, at the same time keeping it in a way that we can, if necessary, somehow retrieve it.



Privacy Scenarios - 2

- **Long-Term Storage**
  - Goal: the privacy benefits of destruction of personal data, while maintaining the ability to access personal data with appropriate authority if it becomes relevant at a later date.
- **CSI Fort Meade**
  - Goal: automated audits that will specifically identify those who exceed their authority – to a legally enforceable standard to deter insider abuse
- **Privacy Toolbar**
  - Goal: ready access to legally vetted rules of thumb for intelligence analysts and ready access to proper legal guidance when answers are unclear.
- **US-Address.com**
  - Goal: technologies that will identify the location and US person status of communicants, so the privacy protections US persons enjoy can be assured

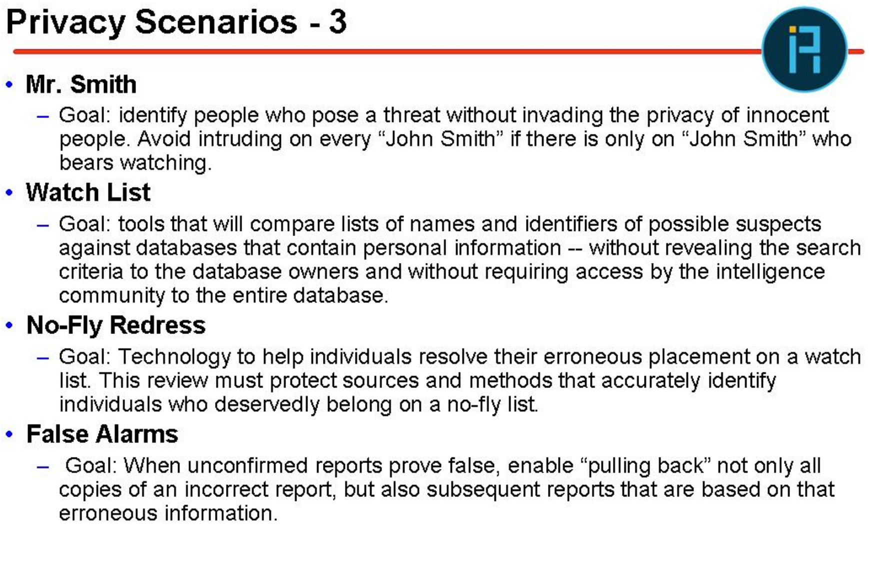
7

Fig. 5. Four more privacy scenarios

CSI Fort Meade. I am not enough of a fan of the TV show CSI to know exactly how this scenario was introduced. But the idea here is that we want to be able to detect whether people are actually exceeding their authority. We want to be able to provide assurances to the public that, if there are abuses going on inside government agencies, we have a way of detecting them and enforcing action against those who do abuse their authority. We've actually seen examples of this outside the intelligence community in recent months, for example, the abuse of the passport database that came up in the context of the presidential primary campaigns. We didn't prevent those instances of abuse, and we may never be able to prevent all those abuses, but, in fact, if we can detect them and somebody gets fired or disciplined as a result, that has, I think, a real effect on the workforce.

Privacy toolbar. The concept here is that we have, what is the other expression for stovepipes, "cylinders of excellence", at the present time. But as we try to bring together these cylinders into a funnel of excellence of some sort, new issues arise. Each of those cylinders have authorities and privacy rules associated with them and the rules are not necessarily the same. For an analyst working in some unified context, the issue becomes: which rules apply? It would be useful to provide clear guidance on this topic. It would also be useful to have better policy in place. But we have to deal with the situation as it is now.

US-address.com. This scenario refers to the issue of identifying the location and "US person" status of communicants, so that we can actually enforce the US person regulations appropriately. This is clearly a challenge that has been aggravated by technology. Technology has been very helpful in many respects but not for this purpose.



Privacy Scenarios - 3

- **Mr. Smith**
  – Goal: identify people who pose a threat without invading the privacy of innocent people. Avoid intruding on every "John Smith" if there is only on "John Smith" who bears watching.
- **Watch List**
  – Goal: tools that will compare lists of names and identifiers of possible suspects against databases that contain personal information -- without revealing the search criteria to the database owners and without requiring access by the intelligence community to the entire database.
- **No-Fly Redress**
  – Goal: Technology to help individuals resolve their erroneous placement on a watch list. This review must protect sources and methods that accurately identify individuals who deservedly belong on a no-fly list.
- **False Alarms**
  – Goal: When unconfirmed reports prove false, enable "pulling back" not only all copies of an incorrect report, but also subsequent reports that are based on that erroneous information.

8

Fig. 6. Final four privacy scenarios

3.3 Third Set

Mr. Smith: We've already discussed this scenario in some respects. The issue involves trying to disambiguate identities, so that, in fact, we don't have intrusions on every person named "John Smith" if there's just one John Smith causing problems. If they all share a common name, we would like to avoid intruding on the privacy of all the innocent parties.

Watch list: We've mentioned aspects of this scenario as well. In this context you don't want to reveal who is on the watch list to the people whose database you might want to search. This is an interesting challenge and I think Rafi Ostrovsky may discuss this as well.

No-fly redress. This one is commonly discussed. If you are erroneously placed on the no fly list, or, even worse if the algorithm keeps putting you back on it, how can you keep yourself off of it? We don't want to expose that list to the public, but we would like somehow to allow people to help us correct it. How can we deal with this kind of a problem?

False alarms. Sometimes we get information that turns out to be wrong. How do we retract that information, and not only the information itself, but all of the inferences that have been based on it? What if, the correction is, itself, wrong? How do we correct the (now) invalid inferences?

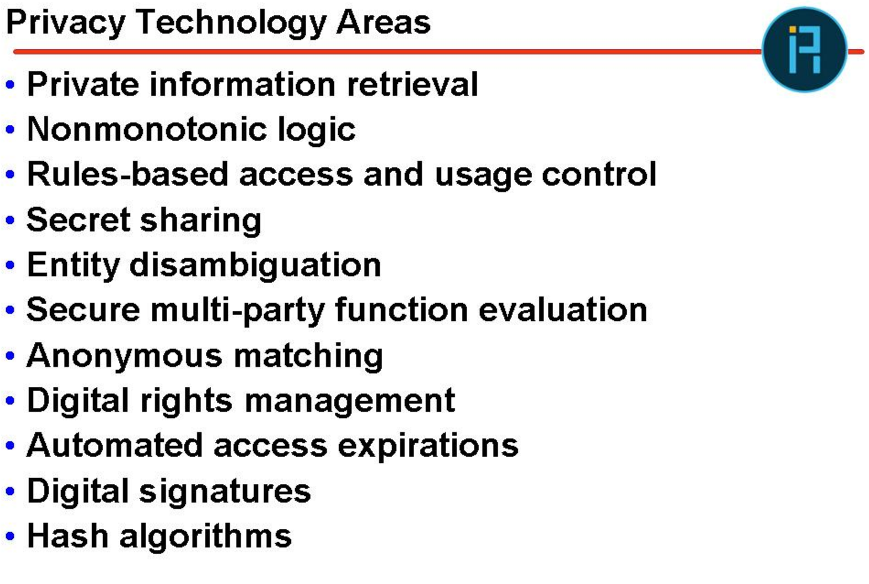
4 Privacy Technology Areas

These were the scenarios put in front of the group to stimulate discussion. Then we considered technology areas in relation to these. So what you are going to see next is a matrix matching the technology areas with those scenarios. I won't say too much about each of these individually.

Private information retrieval is a technique that's been around in the world of cryptography and computer science for a while, but has not seen very much use in practice. We're hoping to see some use of it in the future.

Nonmonotonic logic. For those of you from outside this field, the idea is that, in a typical logic exercise the more hypotheses we have, the more inferences we can make. In nonmonotonic logic, however, we may learn something new, which means we get a new hypothesis from an observation,, but it negates some hypotheses we had or inferences we made earlier. So we can actually infer fewer things after learning this new fact. This corresponds, in some ways, to the notion of a repudiated source, when we find out that this is actually unreliable information; we might need to retract some inferences made based upon it.

Rules based access and usage control. The idea here is access control that is based on rules, in particular the kinds of rules we have in privacy protection, and also on usage control. In many cases a privacy policy might authorize information to be used for one purpose but not another. Usage-based controls could be helpful in enforcing such policies.



Privacy Technology Areas

- Private information retrieval
- Nonmonotonic logic
- Rules-based access and usage control
- Secret sharing
- Entity disambiguation
- Secure multi-party function evaluation
- Anonymous matching
- Digital rights management
- Automated access expirations
- Digital signatures
- Hash algorithms

9

Fig. 7. Privacy technology areas considered by the workshop

Secret sharing is really a cryptographic technology. It provides a mathematical analog to the old tear-the-treasure-map-up-into-several-pieces-and-distribute-it scheme that prevents any subset of individuals from learning the full secret. The cryptographic form of that is more flexible than the paper based scheme, but the idea is similar.

Entity disambiguation, as we have discussed, this refers to techniques for trying to determine whether two or more records` refer to the same physical individual or not.

Secure multi-party function evaluation: Again, as I said, these are not all orthogonal categories of technologies since this technology underlies some kinds of private information retrieval. You may have already heard talks on this technology earlier today.

Anonymous matching: The idea here is to allow owners of two or more databases to determine whether they have records on the same individual or set of individuals without revealing data on other individuals. Another way to look at this technology is as anonymous set intersection. Again cryptographically based, this technology can be used to match sets without revealing any information outside of the intersection.

Digital rights management technology: I think we all have some familiarity with this technology, just from reading the daily press. But this is technology that might be brought to bear on some of these problems as well. We wanted the workshop to consider it.

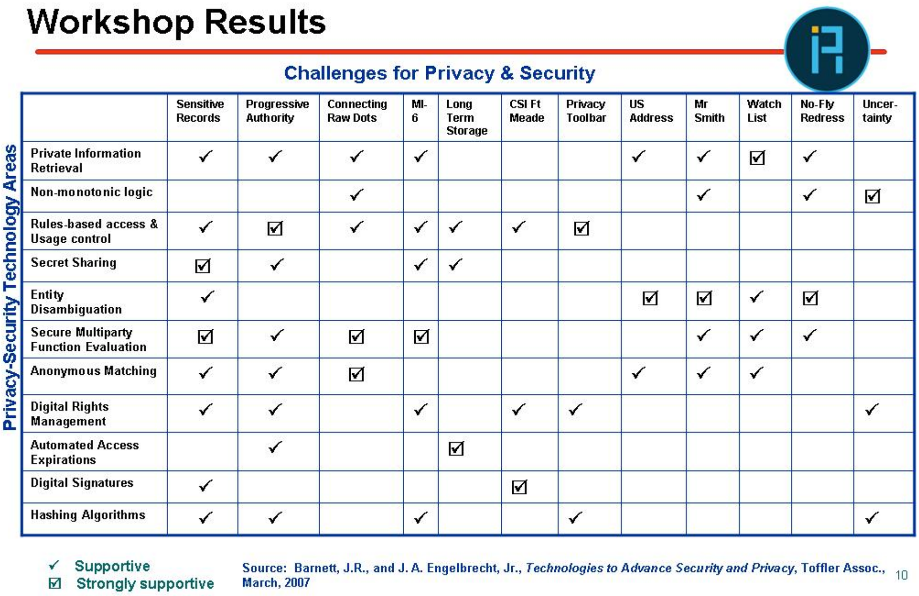
Automated access expirations: This technology could presumably deal with the retention of records. If you have a defined time horizon for record storage, you might be able to automatically "expire" records as they reach their horizon.

Digital signatures and hash algorithms are here as well. These are technologies already in use in many venues. But we wanted to keep them under consideration.

Figure 8 shows the result of a group activity, to identify which technologies might help most with which scenarios. A check mark indicates that the corresponding technology could contribute to the indicated scenario; a check mark with a box around it indicates strongly supportive technologies.

Figure 9 is a summary figure. The goal is to represent which technologies could advance both privacy and security and to provide a rough ordering of where the biggest potential gains might be made. I do not propose a logical argument for the contents for the figure; rather it represents the opinions of the workshop participants as reported by Toffler Associates. Private information retrieval emerged as the technology that supported both security and privacy most strongly The fact that some of these technologies are already in use also influenced the results -- the idea here was to identify technology areas where research investment might actually tip the scale sufficiently in a few years to bring some new technologies into use.

The final figure, Figure 10, shows how privacy technologies might fit into the intelligence information life cycle.



Workshop Results

Challenges for Privacy & Security

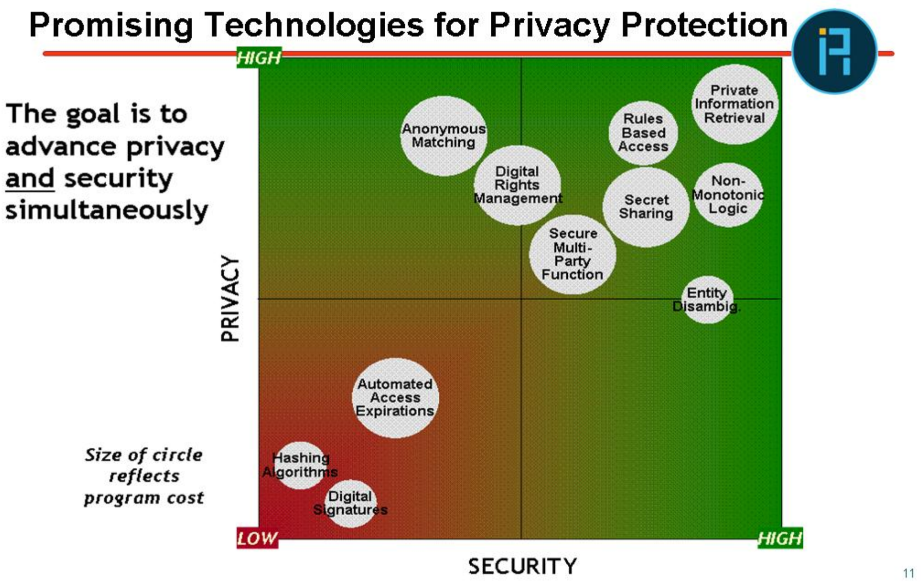
Privacy-Security Technology Areas	Sensitive Records	Progressive Authority	Connecting Raw Dots	MI-6	Long Term Storage	CSI Ft Meade	Privacy Toolbar	US Address	Mr Smith	Watch List	No-Fly Redress	Uncer-tainty
Private Information Retrieval	✓	✓	✓	✓				✓	✓	☑	✓	
Non-monotonic logic			✓						✓		✓	☑
Rules-based access & Usage control	✓	☑	✓	✓	✓	✓	☑					
Secret Sharing	☑	✓		✓	✓							
Entity Disambiguation	✓							☑	☑	✓	☑	
Secure Multiparty Function Evaluation	☑	✓	☑	☑					✓	✓	✓	
Anonymous Matching	✓	✓	☑					✓	✓	✓		
Digital Rights Management	✓	✓		✓		✓	✓					✓
Automated Access Expirations		✓			☑							
Digital Signatures	✓					☑						
Hashing Algorithms	✓	✓		✓			✓					✓

✓ Supportive
☑ Strongly supportive

Source: Barnett, J.R., and J. A. Engelbrecht, Jr., *Technologies to Advance Security and Privacy*, Toffler Assoc., March, 2007

10

Fig. 8. Relating privacy protecting technologies and scenarios

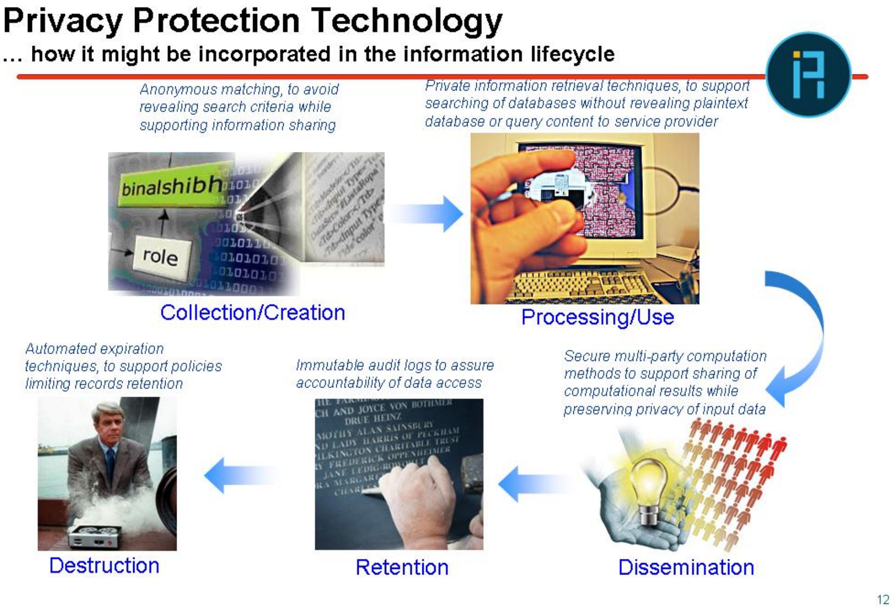


Promising Technologies for Privacy Protection

The goal is to advance privacy and security simultaneously

Size of circle reflects program cost

Fig. 9. Identifying technologies with the greatest security and privacy payoff



Privacy Protection Technology

… how it might be incorporated in the information lifecycle

Anonymous matching, to avoid revealing search criteria while supporting information sharing

Private information retrieval techniques, to support searching of databases without revealing plaintext database or query content to service provider

Secure multi-party computation methods to support sharing of computational results while preserving privacy of input data

Immutable audit logs to assure accountability of data access

Automated expiration techniques, to support policies limiting records retention

Fig. 10. Privacy Protection Technology

Anonymous matching might be of use at the collection/creation phase, depending on the notion of collection. Private information retrieval definitely seems like something hat could be used in the processing phase. Secure multi-party computation might be used in the dissemination phase. Immutable audit logs can certainly help assure accountability of access, though I am not sure research is required there. Finally, automated expiration techniques can be used at the end of the lifecycle of the information.

Acknowledgments. Thanks to all the PPT Workshop organizers and participants.

References

[1] Boneh, D., Kushilevitz, E., Ostrovsky, R., Skeith III, W.E.: Public Key Encryption That Allows PIR Queries. In: Menezes, A. (ed.) CRYPTO 2007. LNCS, vol. 4622, pp. 50–67. Springer, Heidelberg (2007)

[2] Toffler Report. Office of Disruptive Technologies, United States Government (unpublished report)