Privacy Issues

(NSF's Cybertrust panel '08)

Moti Yung
Google Inc. (NY Office)

Cryptography, Security & Privacy Research

(1) Is a humbling experience

(2) Often in protocols research: the definitions are hard, solutions are hard, proofs are tough.

(3) Crypto protocol: When all is finally done, we can say something like: “This cryptographic protocol under the formulated adversary (standard or weakened version thereof) is secure in the sense of reduction to a hard question with this factor (perhaps assuming some idealization).”

Then… leading this Into Industrial Settings

- Many problems are not well formulated; people can find
  - Analogies to theoretical problems in suggesting solutions to sub-problem (this justifies a lot of the work on secure components, provably secure crypto, etc.)
  - Relaxation of the problem in a theoretical setting/ demo-ware setting.
  - Totally different issues that are hard to formulate and need solution under quite hard constraints.

- At times (but not often) two way exchange from industry to research can be found.
- Can this specific (tech transfer) area be funded? Hard to say… all aspects have relevance (theory, engineering, experiments, etc.)
- Different issues dominate practice than those of theoretical constructions… there is a challenge there!

- Above is true for cryptography and any theoretical area of security research

When it comes to Privacy (among other areas of security research)

- This is one of these areas where even the experts lack expertise!
- Various theories about keeping data private. No good practice!
- → (Playing a visionary is hard as well)…Can give some ideas:

- Example: Definition of secure multi-party computation: users do not know more than they can conclude from knowing their inputs when the output is revealed. BUT: revealing the output may violate privacy; ALSO: the fact that the computation took place may violate other type of privacy.

- Privacy is a social, legal, technical, psychological issue, examples:
  - I do not think I mind wearing gloves that leave fingerprints myself, but some privacy advocates may object…..
  - I do not mind being traced pseudonymously when I shop at the supermarket, …others with more unique taste may object…

Stages of computing– increased function/connectivity → increased chances to break systems, to violate privacy,… etc.

Evolution of computing (partial view):

- Stand alone special computer
- Time sharing (safety and protection, access control)
- Enterprise Computers
- Networked (with middleware to adjust applications)
- Inter-networked
- Hosted (scalable distributed systems)
- Mobile, etc.
- Etc.

In last years there is acceleration of change in technology of computing/ communication. Thus: We suggest solutions for one setting and the setting changes!

The changing nature of Computing

- Implies that past solutions may not work (e.g., password on the clear in local environment in late 80's did not scale to wide area network that the architecture evolved into)
- Security/ Authentication models as we know them do not apply!

- While winning all the battles: defining and implementing efficiently (say, secure encryption) and much better integration of theoretical constructions and implementations than in many other areas, the changed environment may make us lose the actual wars (being always late, working on yesterday's problem)…

- Currently: Many privacy and security issues. Very new problems arise as attackers learn the Internet/ mobile networks.

Issues I: arbitrary worries at times…

- Users are afraid:… their IP address is stored at some service provider and.. that they carry the evil cookies…
- Users usually do not complain about the mobile phone providers knowing almost exactly where they are.
- WHY?: the functionality of getting the phone call to the place where you are is understood as necessary! The need to record IP's to prevent attacks, say and to use cookies to improve the shopper/user experience is not well understood!

Issues II

- A web services company ideally, would have liked to give maximum privacy and security to users storing data at the service, etc. BUT needs to balance it against protection of service and quality of the service. All modern businesses hold information of various degree of privacy!

- This is not an easy challenge, but the realization that privacy is important is one of the leading factors. Providers try to assure users they try to maintain clients' privacy.

- Given the world where services are hosted, web (cloud) computing is dominant, mobile and regular computing connect to the physical world, the issues are getting harder and harder.

We Live in a World of “Privacy Tradeoffs”:

- The issues are hard, many impossibilities in various avenues of privacy research (cryptography, statistics, etc.)
- Even possibilities, in practice are impossible: theoretically implied distortion of data may be beyond any usability.
- Lack definitions, Models and security proofs are probably harder (many agents) but we have many constructions with no model and proof or non serious foundational work.
- For applied research: hard to demonstrate “privacy” without first defining it and showing certain metrics according to which the system is evaluated.
- The various tradeoffs and engineering designs beyond the theory and models is not well understood either.
- We are very far away from understanding, in all fronts: theory, engineering, methodologies and implementations/ testing/ certification.

Problems for Investigation (examples)

- **I would like to have a tool where a hosting service treats user data in a way that is careful, safe, private… and**
- **There is accountability, and we need additional properties…**

(Preferably: This is based on well-defined set of what privacy-preserving goals are… that will be determined as needed in this "trust infrastructure" (as part of the investigation itself). Then what I would like to have…)

What we need? Example of problems:

- **(problem 1) Service provider may “prove to users” or to public representatives that they act in a privacy-preserving ways only. This implies privacy assurances of some kinds.**
- **(problem 2) Detection of privacy violation is automated/ repairing (containment of violation), or some issues like this (self detection?)**
- **Efficient workable solutions under some reasonable architectural constraints/ assumptions, and demonstration beyond the theoretical constructions when possible.**
- **Platforms and solution based on new trust infrastructure (not necessarily as we see it today, not only human users involved but certain agents as well, and so on…)**

This can be tried at all levels of investigation

- Theoretical; definitions, sub-problems, constructions. Better understanding of issues
- Engineering methodologies: how to really apply this (avoid using “multiparty computation and be concrete).
- Actual metrics, system designs, etc.
- Experiments.. Example of systems trying conceptually novel aspects of dealing with privacy

- Combinations of the above + feedback (between the levels of investigation, from industry, etc.)…

Conclusions:

- In security: there is a lot to do in theory of actual crypto/ security mechanisms, in systems/ applications research, in bridging gaps to actual systems, and developing actual systems…
- When it comes to privacy mechanisms: the issues are even more obscured/ harder to understand/ derived from disciplines that are hard to formalize.
- Over many years: privacy was no one's interest, (governments, companies, etc.). However– this is a mistake since modern social/ political structures are based on balance of powers and good privacy is an assurance of balance (in many ways).

Therefore

- Funding Agencies: Fund the next best research achievements!! ☺

- Researchers: Do not try to solve Industry's problem (you will never get it totally right), better look for problems from industry and solve relevant issues that the industrial labs have strong interest in! Learn industrial needs, but apply free thinking and high creativity at all levels of research !!! (this is harder than "me too" research, but more "in need.")