The ARES Project: Privacy Technologies in the Information Society

Josep Domingo-Ferrer
Rovira i Virgili University

ARES
Advanced Research on Information Security and Privacy
CONSOLIDER INGENIO 2010

http://crises-deim.urv.cat/ares

May 15, 2008

Introduction

◇ New security-centered services for the Information Society are appearing at an ever faster rate.

◇ Corporate profits are often the main driving force for those services.

⟹ Corporate security is always focused but consumer security and, especially, consumer privacy tend to be disregarded.

Objective and structure of this talk

An overview of privacy problems and solutions for the following “hot” technologies:

- Critical infrastructure protection
- Ubiquitous computing
- Electronic transactions
- Digital rights management
- Data mining, data warehousing and search engines

⟹ These are the topics of the CONSOLIDER INGENIO 2010 CSD2007-0004 “ARES” project (“Advanced REsearch team in information Security and privacy”, http://crises-deim.urv.cat/ares).

Critical infrastructure protection

- The protection of critical infrastructures (airports, power plants, financial facilities, hospitals, defense systems, etc.) is a priority for homeland and corporate security.
- The protection of such infrastructures increasingly depends on the safe operation of the information systems that control them (Critical Information Infrastructures or CIIs).

⟹ CIIs should be dependable.

Dependability of CIIs

- Dependability vs accidental faults is a matter of reliability and it can be solved without attacking the privacy of citizens.
- Dependability vs intentional faults typically is privacy-unfriendly.
  - Intrusion detection systems (IDS) are an example: individuals are profiled to detect whether they deviate from their standard behavior.
  - How to collect IDS without jeopardizing individual privacy is a technological challenge.

Privacy challenges in CIIs

- Data ought to be collected at the lowest possible granularity level compatible with the security of the critical infrastructure (*e.g.* cloak passenger locations in an airport into cells if exact locations are not needed).
- Strict access control policies to personal data should be enforced. *E.g.*, devise schemes whereby low-clearance employees can operate the CII with minimum confidential information[1].
- If data on individuals is released outside the CII, proper anonymization procedures must be used[2].

[1] Sebé, Domingo-Ferrer, Martinez, Deswarte and Quisquater (2008), "Efficient remote data possession checking in critical information infrastructures", *IEEE Trans. on Knowledge and Data Engineering* (to appear), show how the integrity of a data vault can be checked by a low-clearance operator.

[2] Hundepool, Domingo-Ferrer et al. (2006), *Handbook on Statistical Disclosure Control*, Eurostat.

Ubiquitous computing

- Ubiquitous computing has become real with the expansion of wireless communications and mobile devices (cellphones, GPS devices, RFID tags, handheld devices, vehicle-embedded computers, etc.)
- Location-based services are an attractive possibility opened by ubiquitous computing, although they raise the issue of location privacy.
- Even cheap RFID tags can be used to track individuals without their consent.

Example scenarios in ubiquitous computing

Low-end An RFID-tagged shirt can be linked to their buyer's identity by a retailer chain, who could send unwanted SMS ads to the buyer every time his/her shirt is spotted near a shop of the chain.

High-end Car-to-car communication[3] is an application where authentication and confidentiality must not impair privacy (location, driving habits, etc.), because a car conveys a lot of information on its driver.

$\Longrightarrow$ Ubiquitous computing must be made compatible with privacy preservation

[3] Berg (2007), "Standards for car talk", *The Institute*, IEEE, March 2007

Privacy challenges in ubiquitous computing: location privacy

♠ Tracking the location of mobile users of location-based services leaks a lot of confidential information: places visited tell about people's lives, driving habits might be informative of the driver's emotional state, etc.

⟹ Location privacy is to be preserved in location-based services

Approaches for this include:

- Suppression of user identifiers (this is weak because queries themselves are quite identifying)
- Location cloaking to keep locations confidential[4]
- More generally, private information retrieval (PIR)[5] keeps confidential which location-based information is retrieved by the user.

[4] Domingo-Ferrer (2006) "Microaggregation for database and location privacy", LNCS 4032, pp. 106-116.

[5] Chor et al. (1995), "Private information retrieval", IEEE FOCS, pp. 41-50.

Privacy challenges in ubiquitous computing: MANETs and VANETs

- In a mobile ad hoc network (MANET), peer untrusted nodes act as router and the privacy of the routed information must be guaranteed.
- In a vehicular ad hoc network (VANET), it might be necessary for a car to cast a vote to confirm an alarm notification (icy road, traffic jam, etc.) sent by another car to filter out false alarms. Voting should be anonymous and preserve location privacy.

Electronic transactions

◇ Electronic transactions usually entail a loss of privacy for the buyer

◇ Unlike for cash transactions, information on who buys what is automatically collected

◇ Anonymous electronic payment systems exist [6] [7] to emulate the anonymity of cash payments, but they are seldom used partly due to

- Lack of consumer privacy awareness
- Implementation cost
- Corporate wish to conduct market analysis
- Government reluctance (money laundering, etc.)

$\Longrightarrow$ Banks and companies automatically collect huge amounts of information on the customers and their consumption habits.

[6] http://www.ecash.net

[7] Chaum (1989) "Privacy protected payments: unconditional payer and/or payee untraceability", in Smart Card 2000, North-Holland, pp. 69-93.

Privacy challenges in e-transactions: payments

The anonymous e-payment technology is relatively mature, but there is room for improvement in

- Low-value and revocable anonymous payments
- Practical demonstrators offering inexpensive deployment (*e.g.* synergetic with the deployment of electronic ID cards).

Privacy challenges in e-transactions: transaction data

- Transaction data are exploited in data warehouses, often by third parties, so transaction records must be anonymized.
- Beyond suppression of direct identifiers, records should be masked so that the buyers they correspond to cannot be re-identified (*e.g.* if civil state and age are recorded, a record corresponding to an 18-year old widow is easy to re-identify).
- Masking methods for anonymization draw on Statistical Disclosure Control or SDC[8]
- For on-line databases and search engines, the privacy of queries submitted by the users is also an issue [9].
- Query privacy can be handled by PIR protocols[10].

[8] Hundepool, Domingo-Ferrer *et al.* (2006), *op. cit.*

[9] Cohen (2005), "Google needs a privacy upgrade", *International Herald Tribune*, Nov. 29, 2005

[10] In August 2006, the AOL search engine "took the liberty" of disclosing 658000 queries "for research", a lot of which were very identifying.

Digital rights management

- Digital rights management (DRM) has the legitimate goal of protecting the intellectual property (IP) of digital content.
- Imperceptible watermarks embedded in the content are a usual technique to detect illegal copies.
- If each copy sold carries a unique watermark to trace illegal redistributors, the watermark is called fingerprint
- A fingerprint is like a serial number which the vendor can link to the identity of the copy buyer.

Owner’s IP and buyer’s privacy

♣ On behalf of the content owner, the vendor wishes to identify the buyer to link the latter’s identity to the fingerprint embedded in the copy sold.

$\Longrightarrow$ The vendor knows who is buying what content, which is a violation of the buyer’s privacy

♣ If electronic payment can be anonymous, it is unacceptable for the buyer to sacrifice her privacy to the content owner’s IP protection.

♣ Privacy loss should be limited to dishonest buyers who illegally redistribute the content they have purchased.

Privacy challenges in DRM: anonymous fingerprinting

- Anonymous fingerprinting is a theoretical solution to combine DRM and privacy of honest buyers[11].
- The merchant fingerprints the content sold to a buyer without knowing the identity of the buyer nor seeing the fingerprinted copy.
- Finding a (redistributed) fingerprinted copy enables the merchant to find out and prove to third parties whose copy it was.
- However, anonymous fingerprinting protocols proposed so far rely on secure multiparty computation, and they are completely impractical.
- Coming up with practical anonymous fingerprinting protocols is a major challenge.

[11] Pfitzmann and Waidner (1997) "Anonymous fingerprinting", LNCS 1233, 88-102.

Data mining, data warehousing and search engines

- As noted above, privacy in databases underpins a lot of applications (critical information infrastructures, ubiquitous computing, electronic transactions, etc.)
- Database privacy has three dimensions:
  1. Respondent privacy. Preventing re-identification of the individuals/enterprises to which the records of a published database correspond.
  2. Database owner privacy. Allowing two or more autonomous entities to compute queries across their databases in such a way that only the results of the query are revealed.
  3. Database user privacy. Guaranteeing the privacy of user queries to prevent user profiling and re-identification by the database owner.

Database privacy technologies

- Respondent privacy is pursued by statistical disclosure control (SDC)[12]
- Database owner privacy is pursued by privacy preserving data mining (PPDM)[13][14]
- Database user privacy is pursued by private information retrieval (PIR) [15]

[12] Hundepool, Domingo-Ferrer *et al.* (2006), *op. cit*

[13] Aggrawal and Srikant (2000) "Privacy-preserving data mining", *Proc. of ACM SIGMOD*, pp. 439-450.

[14] Agrawal, Grandison, Johnson and Kiernan (2007) "Enabling the 21st century healthcare information technology revolution", *Communications of the ACM*, 50(2), pp. 35-42.

[15] Chor *et al.* (1995), *op. cit*

Privacy challenges in SDC

- Regarding SDC, make progress towards quantifying the disclosure risk inherent to publication of a masked database.
- Try to know what external public identified databases will be available to an intruder for matching with the anonymized transaction records.
- Develop SDC for on-line databases, that is, to protect respondent when the users can issue dynamic queries (tracker attacks).
- Improve the tradeoff between disclosure protection and analytical validity of masked data.

Privacy challenges in PPDM and PIR

- Current PPDM techniques only allow a restricted set of data analyses across several databases whose owners are unwilling to fully share.
- Come up with PPDM techniques allowing a broader range of joint analyses.
- Make PIR and SDC compatible: current SDC techniques used to protect the respondent privacy often need the database to know which precise queries are submitted by the user [16].

[16] Aguilar and Deswarte (2006) "Single database private information retrieval schemes", LNCS 4302, pp. 257-265

Conclusions

- The information society has to stay secure to survive...
- ...but it must respect privacy to stay human
- Security technologies will undoubtedly progress even without public support...
- ...but privacy technologies have less commercial appeal and their deployment must be promoted, enforced and partly sponsored by the administrations.
- This is our right as citizens!