

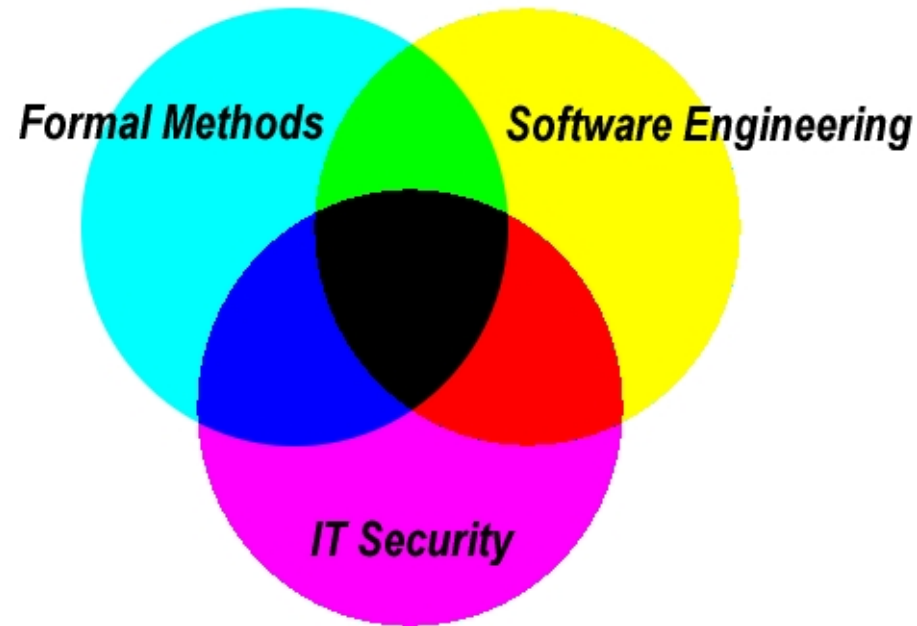
E-Government

From Policies to Mechanisms

David Basin
ETH Zürich



About myself



Since 1.1.2003

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Chair for Information Security and director of ZISC

About this talk

- This is a **nontechnical** talk about **privacy**
 - ▶ What is privacy?
 - ▶ Why is it so difficult to ensure?
 - ▶ What are relevant research issues?
- Discussion of privacy will be general.

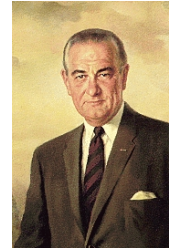
But research problems and solutions will leverage key features of E-Government context.

Road map

Privacy: a fundamental good?

- What is privacy?
- Privacy: requirements, policy, and mechanisms.
- Research highlights.
- Conclusions.

Privacy, a fundamental good?



- Lyndon B. Johnson, President of the USA, 1963-1969.

Every man should know that his conversations, his correspondence, and his personal life are private.



- Directive 95/46/EC of the European Parliament

Whereas data-processing systems are designed to serve man; whereas they must respect their fundamental rights and freedoms, notably the right to privacy ... In accordance with this Directive, Member States shall protect the fundamental rights and freedoms of natural persons, and in particular their right to privacy with respect to the processing of personal data.

- Scott McNealy, Sun Microsystems

You have no privacy — get over it.



Is your data worth protecting?

- Your personal data is interesting.

Shopping habits, family status, religion, political party, criminal record, vita/career, health, finances, sports/hobbies, ...

- Your data is everywhere and computers are good at collecting it.

Bank: transfers, investments, credit card purchases, taxes.

Telephone: source, time, location.

Shopping/travel: from (online) shops, loyalty programs.

Entertainment: movies watched in hotels (also < 2 minutes).

- Valuable to sales departments, (future) employers, agencies, etc.

Valuable for you?

Example: Public companies, public data



- The situation at hand

The latest dot-com casualty, Voter.com, announced plans to sell its list of 170,000 e-mail correspondents, complete with political party affiliations, issues they're concerned about and demographic information, such as home zip-codes and their gender.

— *Center for Individual Freedom Newsletter*, 2000

- And what might await us

Google has created an information repository of a sort that the CIA would envy. It is reported that Google has maintained a record of essentially every search (including the IP address information, time, etc.) done on their systems, and has developed tools to mine this vast storehouse. There is no reason to suspect that Google has evil intentions. But rosy motives don't provide immunity from the ways in which their vast machine could someday become an instrument of genuine repression despite Google's best intentions today.

— Lauren Weinstein, *The Privacy Forum*, 2004

Example: Government agencies, government data



Activists are demanding that the government halt the program, which links municipal computer systems and gives each Japanese citizen an 11-digit identification number. The new database stores personal data — names, addresses, date of birth, and the new ID numbers — for each of Japan's 127 million citizens, making it easier for them to obtain documents for a variety of public services and benefits.

At least five municipalities have refused to join the system. Critics say that ID numbers can act as keys to personal data stored at different locations, making it easy for hackers to create mischief. And doubts have emerged over the technical aspects after several municipalities reported computer glitches.

Reuters Limited, 2002

Road map

- Privacy: a fundamental good?

What is privacy?

How does it compare to traditional security objectives?

- Privacy: requirements, policy, and mechanisms.
- Research highlights.
- Conclusions.

Privacy and its friends

Confidentiality: No unauthorized access to information.

Examples: mail, bank balance, e-votes.

Anonymity: Communication partners are secret.

Examples: E-complaint boxes, most surfing, criminal transactions.

Privacy: Collected data is only used for limited, predefined purposes.

Example: Customer address used only for sending software updates.

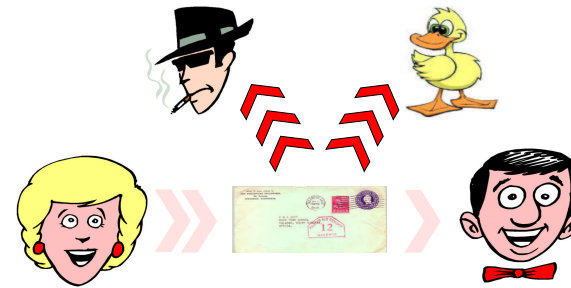
Privacy is different!

It concerns **how** data is used, rather than by **whom**.



"On the Internet, nobody knows you're a dog."

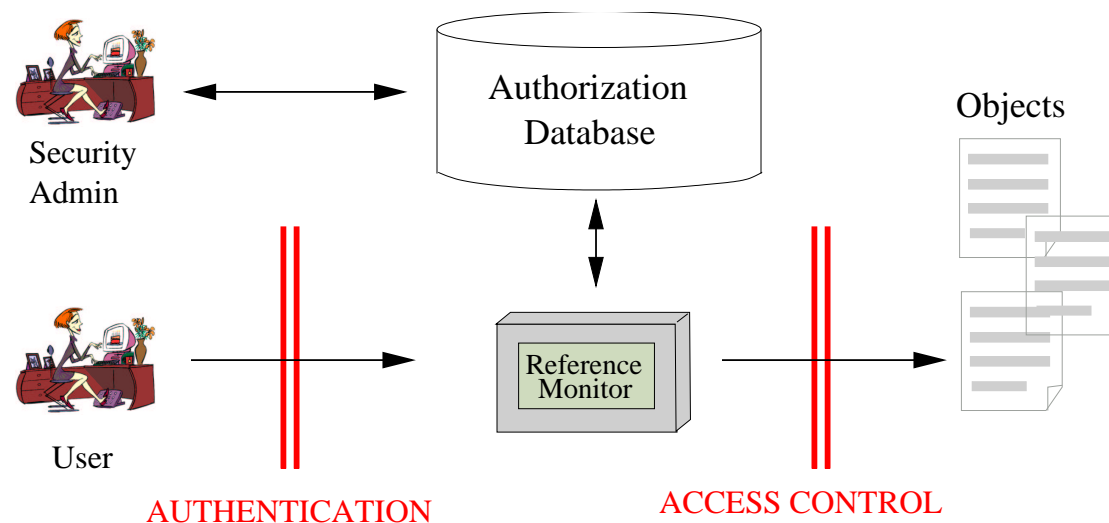
Confidentiality



- Requirements specify **who** can access **what** information.

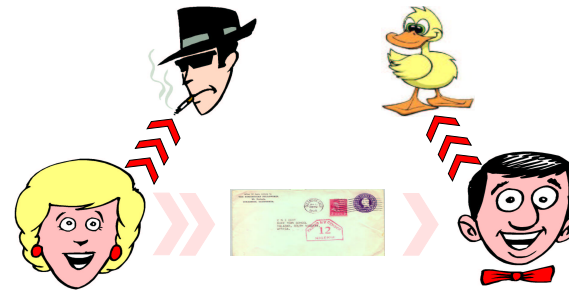
Example: mail must be unreadable in transit.

- Standard mechanisms may be used. E.g., **encrypt** communication and **control access** through a **server-side reference monitor**.



Here database formalizes policy, e.g., lattice-based, RBAC, ...

Anonymity



- **Requirements:** Anonymize the sender and/or the receiver.
Provide confidentiality of principals' identities.
- **Pseudonyms** as a lightweight **mechanism**.



Mouse !!!
Come back – I love you!



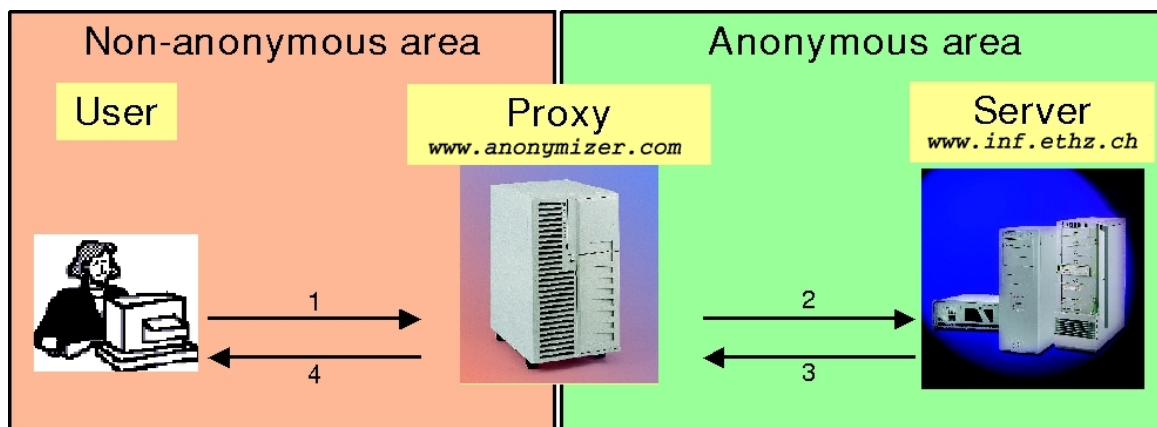
Your kissy-bear

Linkage to actual identity only in restricted cases.

- IT equivalent: mail and surf from an ISP account, Hotmail, ...

Sender/recipient anonymity: proxies

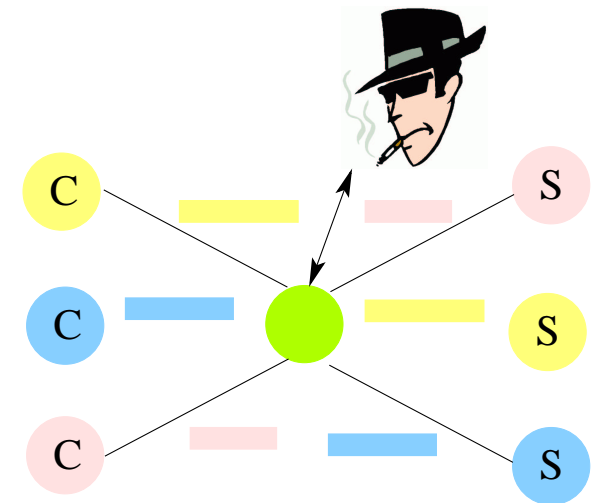
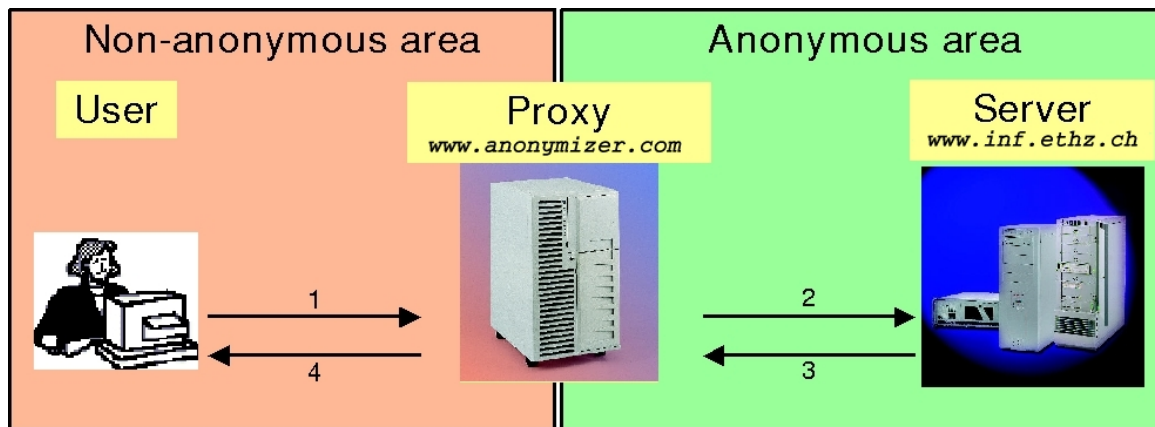
- Packets, e.g., HTTP requests are anonymized by a proxy



- Weaknesses?

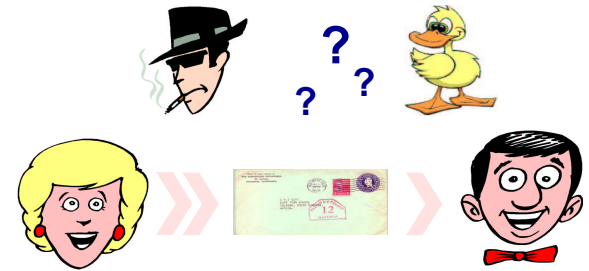
Sender/recipient anonymity: proxies

- Packets, e.g., HTTP requests are anonymized by a proxy



- Weaknesses?
 - ▶ The proxy knows everything.
 - ▶ Traffic analysis is also possible.
- Solutions: cascaded proxy chains, mix networks, onion routing, ...

Privacy



- According to Oxford English Dictionary
A state in which one is **not observed** or **disturbed** by others.
- In IT context this can be specialized as follows:
Anonymity: nonobservability of our actions when they occur.
Data Protection: ensuring that our collected data is not distributed and used in undesired ways.
- As anonymity is more-or-less understood, we focus on **data protection**, **controlling the dissemination and usage of sensitive personal data**.

Acceptable usage? How do we formulate this? Enforcement?

Road map

- Privacy: a fundamental good?
- What is privacy?
- 👉 **Privacy: requirements, policy, and mechanisms**
 - Research highlights.
 - Conclusions.

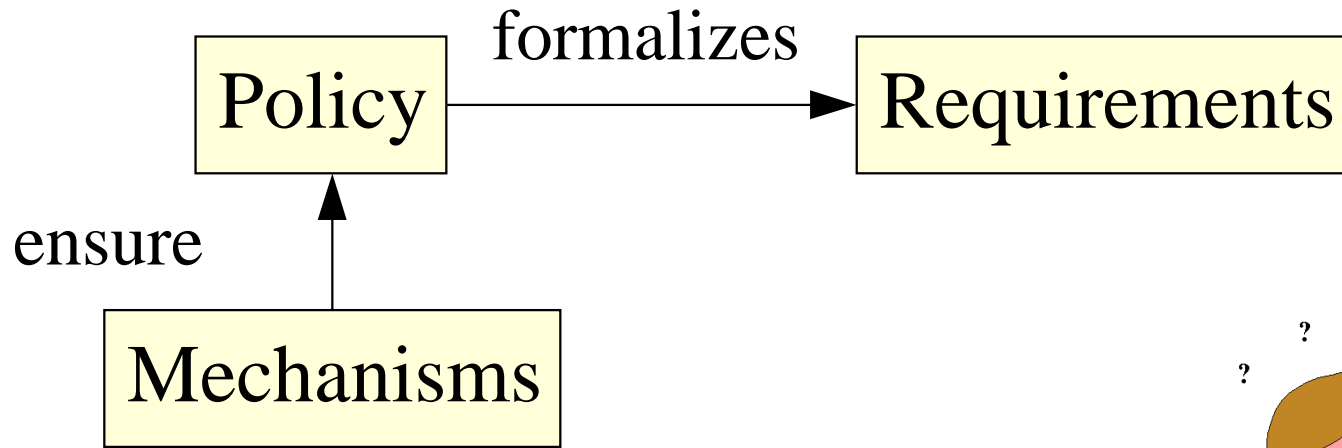
Privacy requirements and policies

- Privacy requires that data **usage respects purpose**.
 - A privacy policy specifies **how** data may be used, under which **conditions**, and what **obligations** this entails.
- ▶ We use personal information that you provide to register you in programs, create and maintain accounts, ship products, ...
 - ▶ We share your personal information with certain business affiliates.
 - ▶ We and our subsidiaries may send you email with offers about products **unless you indicate you do not want to receive them**.
 - ▶ Before **processing a minor's order**, **parental consent must be given**.
 - ▶ **Data associated with inactive accounts will be deleted after one year**.

Examples of privacy requirements

- Access control requirements. Only X may access Y .
- Actions required before the access. Gathering owner consent.
- Actions that must be performed within a certain time period. Informing data owner whenever the data is used.
- Restrictions on the further distribution of the data. Own use only.
- Restriction of purposes for which data may be used. Statistical purposes only.
- Limitations on retention time. Delete after 7 days.
- Mandatory use of protection mechanisms. Encrypt backups.
- Duties of keeping the data up-to-date. Update every 30 days.

What are the problems?



How do we formalize requirements?

How do we enforce them?



Let's briefly consider current standards and limitations.

Privacy policy languages

- Normally just unstructured text.
- Machine-friendly forms being standardized.

Example: Platform for Privacy Preferences (P3P) of W3C.

- P3P allows web sites to communicate their privacy policies in computer readable format (XML).

Data collected, purpose, how long it is retained, etc.

- Enables tools (in browsers or stand-alone) to be developed that:
 - ▶ summarize privacy policies,
 - ▶ compare policies with user preferences, and
 - ▶ advise and alert users.

An example

We do not collect any information from site visitors except the information contained in **standard web server logs** (your IP address, browser information, etc.). This log information will be used only by us for **web site administration**. It will not be disclosed unless required by law. We may retain these log files **indefinitely**. Please direct questions about this privacy policy to **privacy@company.com**.

```
<POLICY discuri="http://company.com/privacy.html" name="policy">
  <ENTITY>
    <DATA-GROUP>
      <DATA ref="#business.contact-info.online.email">privacy@company.com</DATA>
      <DATA ref="#business.contact-info.online.uri">http://company.com/ </DATA>
    </DATA-GROUP>
  </ENTITY>
  <STATEMENT>
    <CONSEQUENCE>We keep standard web server logs.</CONSEQUENCE>
    <PURPOSE><admin/><current/><develop/></PURPOSE>
    <RECIPIENT><ours/></RECIPIENT>
    <RETENTION><indefinitely/></RETENTION>
  </STATEMENT>
</POLICY>
```

P3P and other languages

- There are other “privacy” languages. E.g.,
 - ▶ Security Assertion Markup Language (SAML)
 - ▶ Microsoft security metadata to SOAP.
 - ▶ IBM’s Enterprise Privacy Authorization Language (EPAL)
 - ▶ Permission-based attributes in Liberty Alliance
- But relationship to privacy is often poorly understood.
In most cases, privacy = anonymity (via pseudonyms).
- P3P attempts to do more here.

Semantics? Enforcement?

Privacy enforcement mechanisms



- Current mechanisms based on company/legal processes.
- Example: “US Online Privacy Alliance”.
 - ▶ 50+ large (“global players”) US companies.
 - ▶ Set of **guidelines** addressing collection, use and disclosure of individually identifiable information and data, as well as special measure to protect privacy of children on-line.
 - ▶ **Self-enforcing**, with third-party **monitoring** and **audit**.
- While a good first step, such best efforts have limited effectiveness.
 - ▶ Analogous to self restraint in conventional (CIA) security.
 - ▶ Policies prone to misunderstanding, misuse, and abuse.

Is more possible here?

Road map

- Privacy: a fundamental good?
- What is privacy?
- Privacy: requirements, policy, and mechanisms.

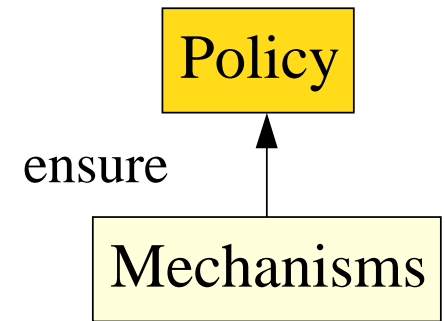
Research highlights.

Policy languages

- ▶ Privacy mechanisms
- ▶ Document security
- Conclusions.

Privacy policies

Formal language and semantics



Semantics: what do privacy policies mean?

Not merely an academic question. Answer needed for:

- Specifying/combining complex (even contradictory) policies.
 - ▶ You would like to protect your financial (banking) data.
 - ▶ The banks have their own enterprise privacy policies.
 - ▶ Governments regulate disclosures, e.g., if tax fraud suspected.
 - Providing a basis for enforcement.
 - ▶ Must bridge the gap from policy to IT mechanisms.
 - ▶ Analogy to AC: good policy description formalisms exist.
- ⇒ Research in (precise, logical) specification formalisms.

Consider some examples

Owner consent is required prior to access.
Server maintains log files for at most 30 days.
Data may only be used for statistical purposes.

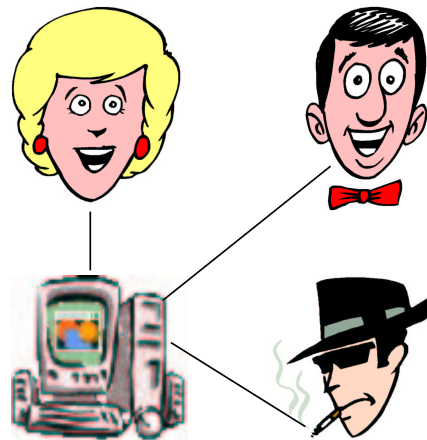
Consider some examples

Owner consent is required prior to access.
Server maintains log files for at most 30 days.
Data may only be used for statistical purposes.

Such requirements combine propositions about

time: prior to access, at most 30 days, ...,

distribution: distributed entities like owner, server, ...



Best formalized in a language that combines these two dimensions.

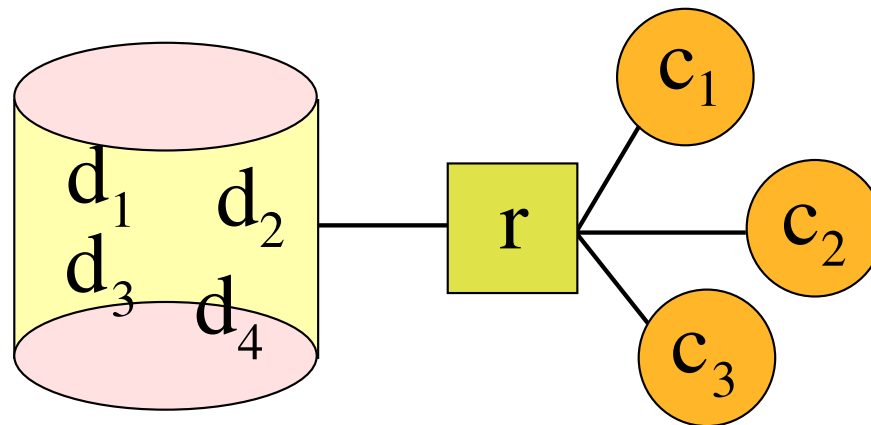
Consider some examples

Owner consent is required prior to access.
Server maintains log files for at most 30 days.
Data may only be used for statistical purposes.

Such requirements combine propositions about

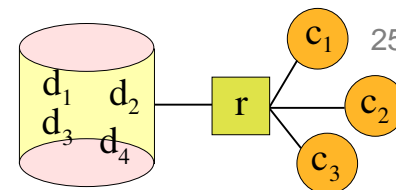
time: prior to access, at most 30 days, ...,

distribution: distributed entities like owner, server, ...



Best formalized in a language that combines these two dimensions.

Examples (cont.)



Owner consent is required prior to access

- Formalized using **Distributed Temporal Logic**
 - ▶ Logic formalizes **temporal** properties, e.g., $\diamond \textit{Consent}$
 - ▶ relative to view of **distributed** principals, e.g., $@_r \diamond \textit{Consent}$
 - ▶ as well as communication $@_r \diamond @_{c_2} \textit{Consent}$
- Example can be formalized as


$$\bigwedge_{c_i, c_j, d_k, purpose} \cdot @_r ((\textit{access}(c_i, d_k, \textit{purpose}) \wedge \textit{owner}(c_j, d_k)) \\ \implies \\ \diamond @_{c_j} \textit{Consent}(c_i, d_k, \textit{purpose})))$$

- Provides a **semantics** for privacy policies as well as means to rigorously **reason** about them.

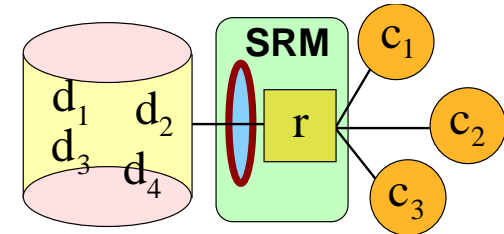
Road map

- Privacy: a fundamental good?
- What is privacy?
- Privacy: requirements, policy, and mechanisms.

Research highlights.

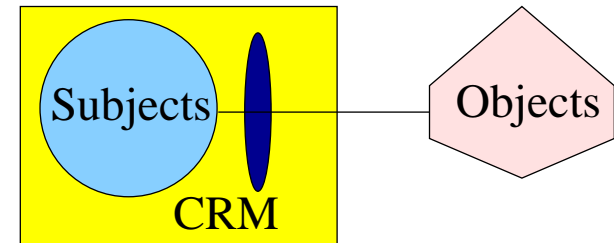
- ▶ Policy languages
-  **Privacy mechanisms**
- ▶ Document security
- Conclusions.

The crux of the problem



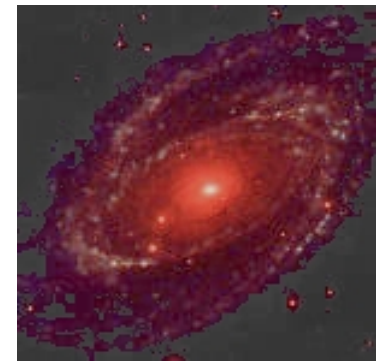
- State-of-the art builds upon **reference monitors** for **access control**.
 - ▶ **Example:** IBM Tivoli Privacy Manager integrates identity management, with access control, auditing, and reporting for J2EE applications over LDAP data sets.
 - ▶ Can handle conditions and limited obligations (e.g., logging) for P3P specified policies.
- Control is via a **server-side reference monitor**.
- Privacy however requires **client-side restrictions**, e.g., against
 - ▶ users **careless** in their use and distribution of data,
 - ▶ **negligent** users, e.g., their computers may be compromised,
 - ▶ or even **dishonest** users.

Mechanism research



28

- Investigate tradeoffs between **requirements** (including **trust**), and **mechanisms**. Contrast, e.g.,
 - ▶ **B-to-B e-commerce**: exchanging customer data, and
 - ▶ **A-to-A e-government**: exchanging citizen records.
- Mechanism tradeoffs
 - ▶ How far can one go with server-side architectures?
 - ▶ How light-weight can client-side controls be?
 - ▶ Avoid the Trusted Computing/DRM black-hole?
- Related architectural questions:
 - ▶ How do we communicate and enforce distributed obligations?
 - ▶ How do we formally establish adequacy of mechanisms and correctness of concrete architectures?



Road map

- Privacy: a fundamental good?
- What is privacy?
- Privacy: requirements, policy, and mechanisms.

Research highlights.

- ▶ Policy languages
- ▶ Privacy mechanisms

Document security

- Conclusions.

Document Security

Problem: protection of semi-structured documents within and outwith institutions.

Approach taken

- **Formal specification** of document model, document processing operations, policy language, and enforcement mechanisms
- **Refinement** to a document processing architecture providing both client-side and server-side control

Aim is to provide foundations for design and validation of “privacy architectures”.

```
<addresses>
  <address>
    <name>David Basin</name>
    <email>basin@ethz.ch</email>
  </address>
  <address>
    <name>Paul Sevinc</name>
    <email>ps@ethz.ch</email>
  </address>
</addresses>
```

Road map

- Privacy: a fundamental good?
- What is privacy?
- Privacy: policy and mechanisms.
- Research issues.

 **Conclusions.**

Conclusions

- Ensuring privacy is of the utmost importance.
- It is fundamentally different than conventional (CIA) security.
- Solutions depend on context and trust assumptions.
- It is a fascinating topic for research.
 - ▶ It is of central importance, e.g., in e-government.
 - ▶ Problems are still open.

For more on our research see:

www.infsec.ethz.ch and www.zisc.ethz.ch