Privacy-Enhancing Overlay Networks

Overview of preliminary research

Matthijs Koot
(mkoot@science.uva.nl)

Faculteit van Natuurwetenschappen, Wiskunde en Informatica
Universiteit van Amsterdam

2007-06-06 / ISOC meeting, Amsterdam

Outline

Outline

Questions.

Inception of my research:

- Questions:
  - “What is the problem to be addressed?"
  - “What is the greater context?"
  - “What is privacy? What is an overlay network?"
  - “What research opportunities are out there?"
- Answers found w/colleagues and through literature study.

Security in VL-e.

Security in Virtual Laboratories for e-Science, or *VL-e*:

- Job-centric model, as proposed in 2005 by Demchenko et al in *"Job-centric Security model for Open Collaborative Environment"*
- AuthN, AuthZ, auditing are present
- Privacy, confidentiality aren't
- Why? Does it matter?
  - SP1.2: trade secrets in food industry (Unilever).
  - SP1.3: privacy of medical data (AMC).

Outline

Why should anyone care?

Statement

Privacy is a trust promoter for the ‘global information society’ [Bangemann94].

Why should anyone care?

Assuming that grid technology will become widespread:

- e-Science has privacy issues for medical science (e.g. toxicogenetics), CBRN research, and particular other disciplines
- But moreover, when the grid transcends the scientific context:
  - Hospitals: image processing for daily healthcare
  - Government: image processing for public safety
  - Insurance: increasingly complex risk analysis
  - Commercial: business process simulation

Who should care?

Privacy is a matter of concern to:

- Designers of grid technology
  - Provide privacy-aware data storage, transport and usage
  - E.g. privacy-aware programming models, crypto services
- Prospective users of grid technology
  - Be involved in requirements engineering
- Legislators, lawyers
  - For ‘code is law’
  - Ref.: HealthGrid (lawyer-intensive backtrack)

Outline

Exploration of ‘privacy’.

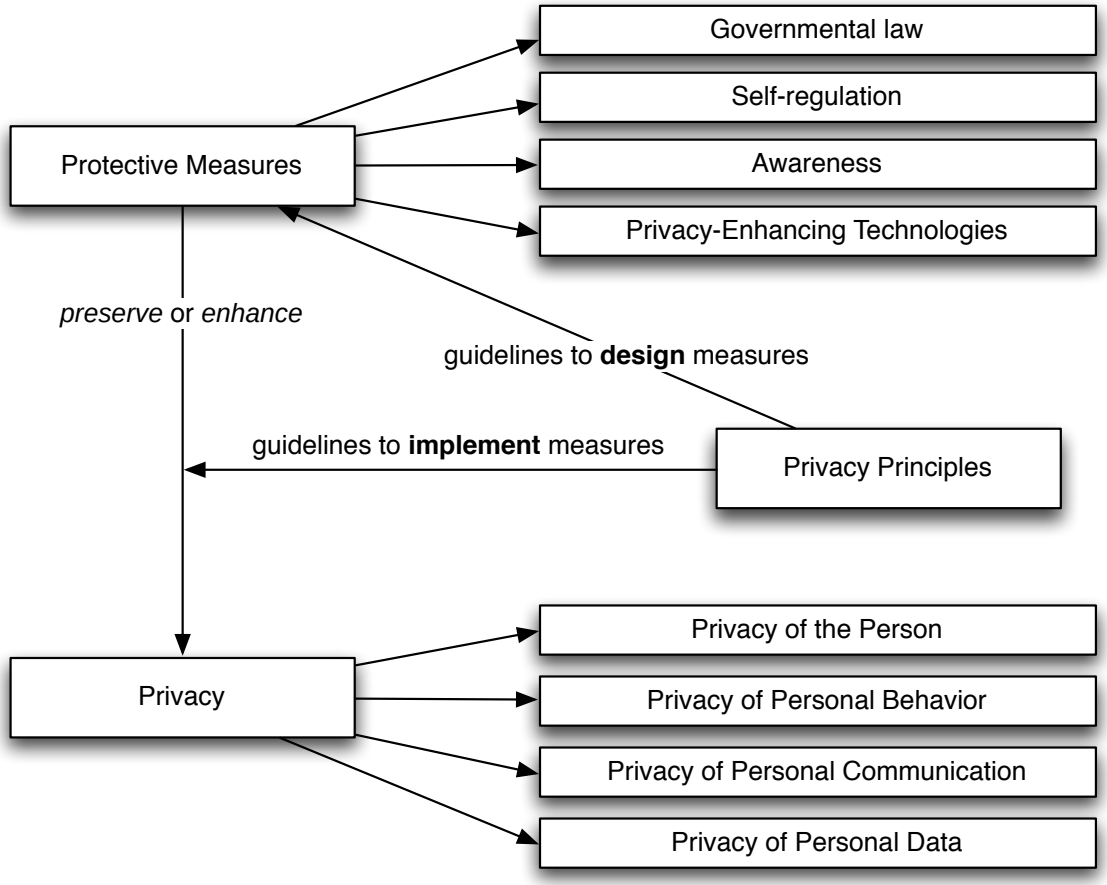
Statement

Privacy is about IDENTITY, CONFIDENTIALITY and TRUST.

Exploration of ‘privacy’.

- Privacy . . .
  - is a sociological notion;
  - is a psychological notion;
  - is an economical notion;
  - is a political notion;
  - . . . and therefore a cultural idiosyncrasy.
- It is context-sensitive, thus not an absolute notion.
- It is (still) lacking understanding, despite lots of research in privacy law, technology and the aforementioned areas.
- It is said to be often confused with ‘security’ or ‘anonymity’.

Exploration of ‘privacy’.

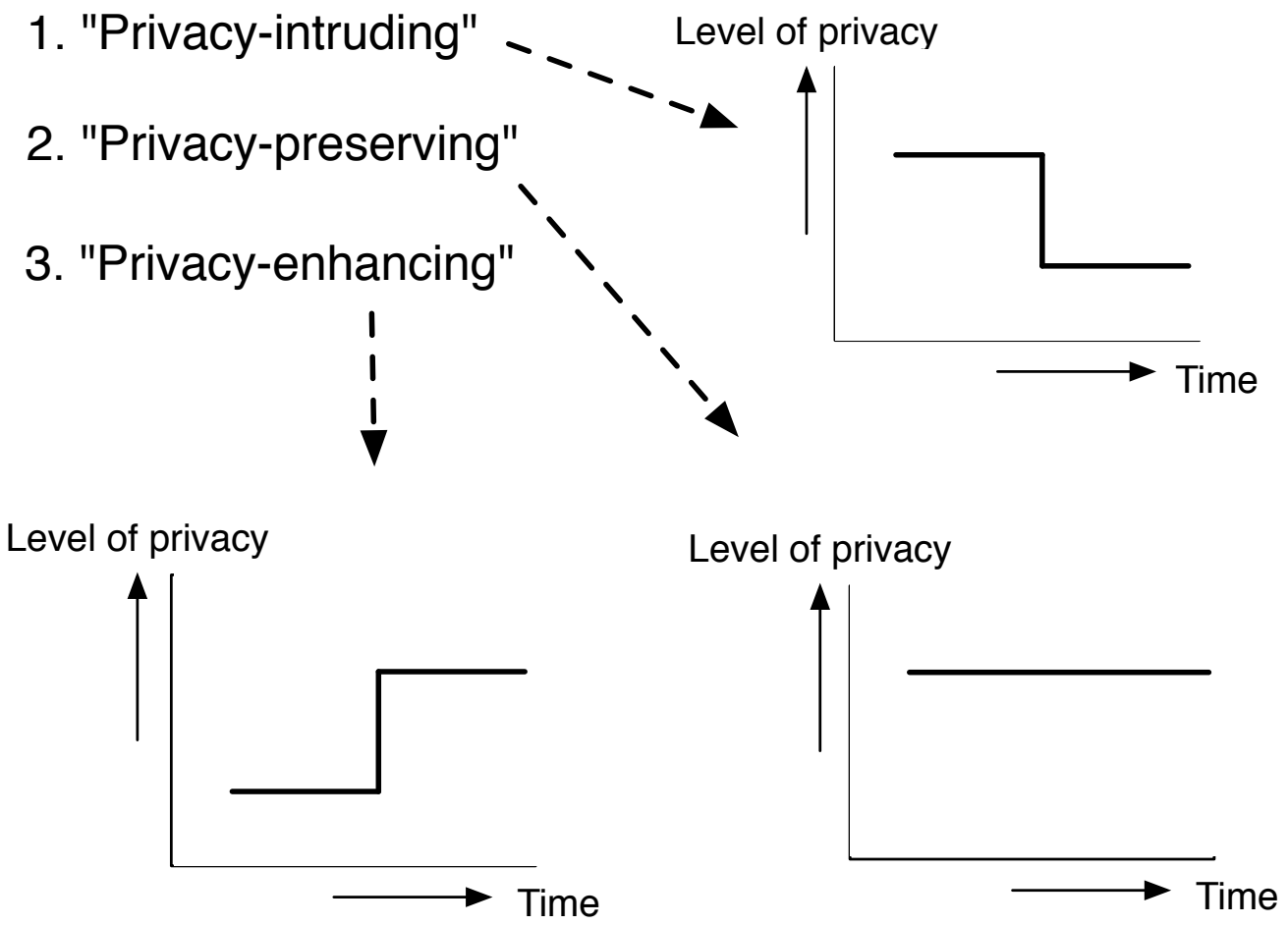


Exploration of ‘privacy’.

Some concepts:

- Anonymity
  - Sender anonymity
  - Receiver anonymity
  - Relationship anonymity (unlinkability)
  - Location anonymity
- Pseudonymity
- Confidentiality
- Privacy-enhancing, privacy-preserving, privacy-intruding

Exploration of ‘privacy’.



Exploration of ‘privacy’.

- Research subtopics include . . .
  - Mix networks.
  - Cryptography.
  - Policy-enforcement systems.
  - Filters, blockers, erasers.
  - Hippocratic databases.
- So what?
  - What/who are we trying to protect, anyway?
  - Or is our goal to establish *trust*?

Outline

Exploration of ‘overlay networks’.

- Overlay networks . . .
  - Are ‘networks on top of other networks’.
  - P2P, grid computing, et cetera.
  - Indeed, this broad definition includes IRC, SMTP, et cetera.

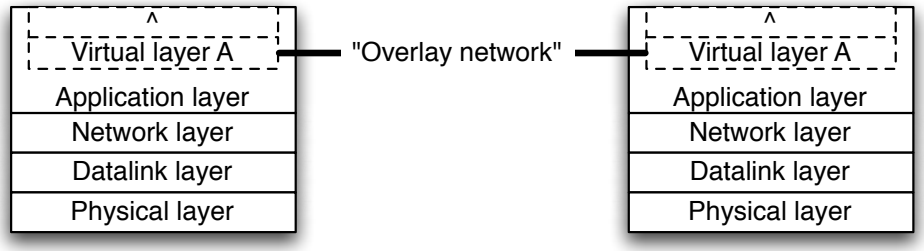
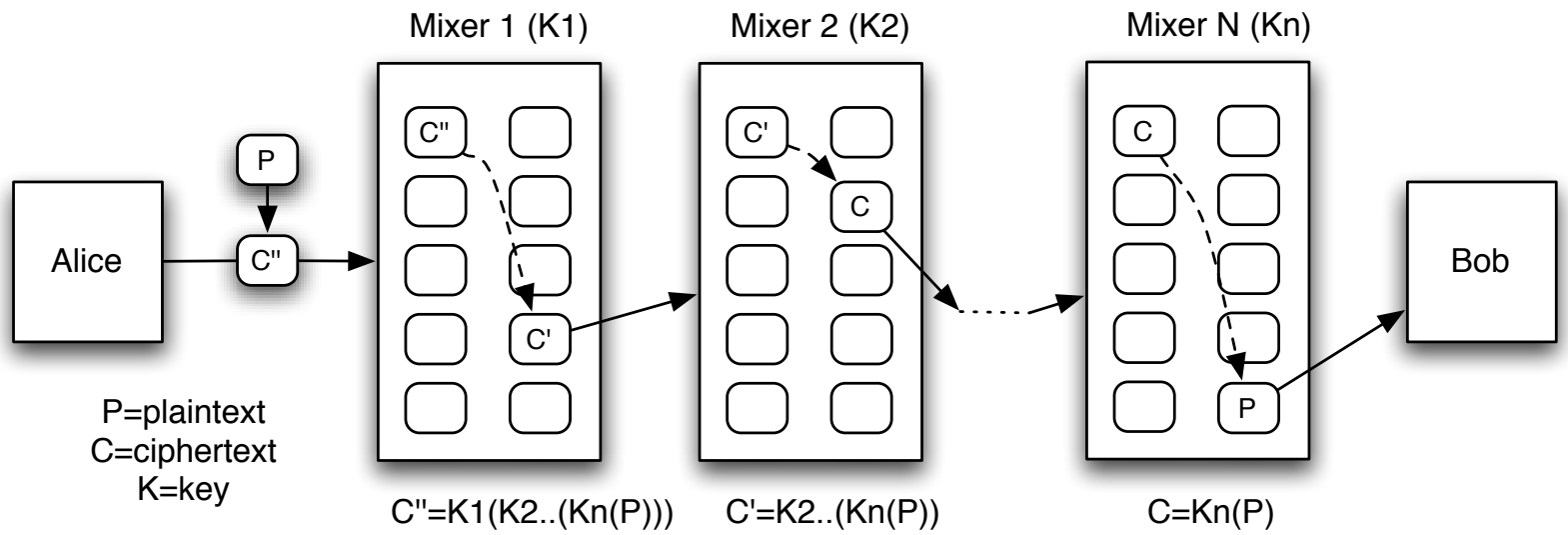


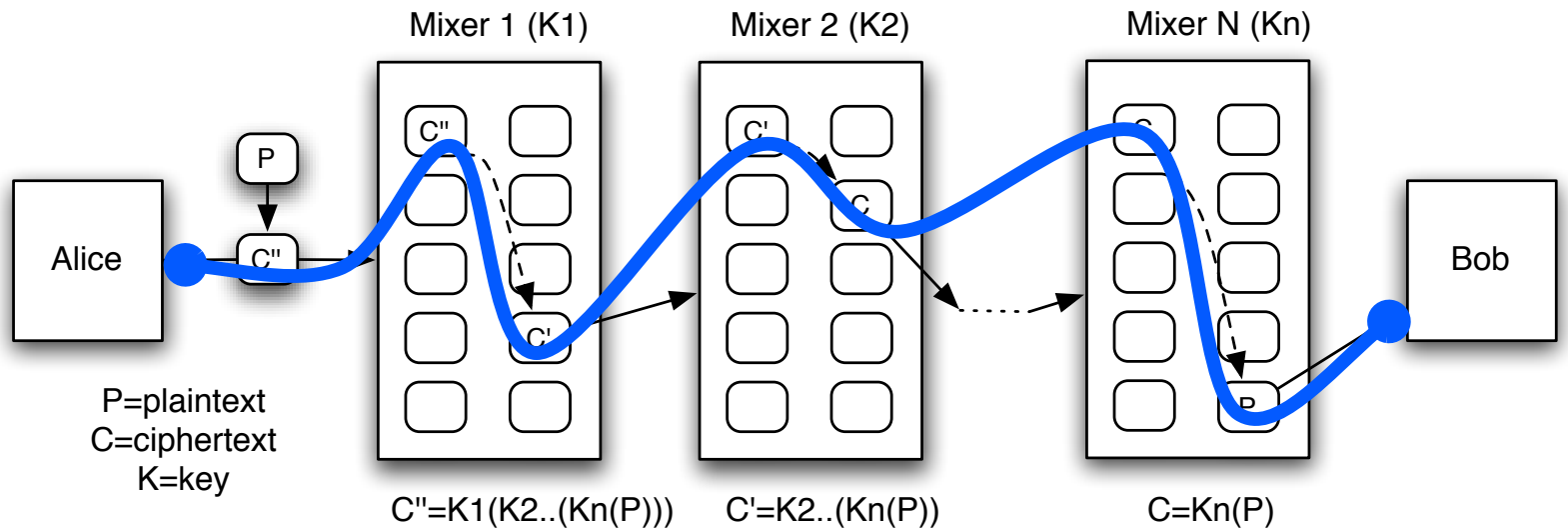
Figure: Today’s view on overlay networks.

Exploration of 'overlay networks': mixnets.



- Traditional message-based Chaumian mixnets offer sender and relationship anonymity.
- Broadcasting *p* will also offer some receiver anonymity.

Exploration of ‘overlay networks’: mixnets.



- *Circuit*-based mixnets: Tor, MorphMix, Tarzan, I2P, ...
- TCP proxy (L4), HTTP proxy (L5+), ...

Exploration of ‘overlay networks’.

- Research subtopics include . . .
  - Network establishment and routing (e.g. self-*).
    - Enable multicast on backbones w/o replacing routers!
  - Application services (e.g. storage).
  - Management services (e.g. monitoring).
  - Non-functional aspects (e.g. security, privacy).
- So what?
  - How about anonymous sharing of grid data?
  - How about an overlay for secure multiparty computation?
  - How about a (privacy)policy enforcing overlay?
  - How about MLS-aware resource sharing?

Approaches / opportunities.

Question

How do IDENTITY, CONFIDENTIALITY and TRUST map onto grid technology?

Privacy in the grid.

- IDENTITY is represented by VOs (organizational identity) and their members (individual identity)
- When TRUST isn't high, CONFIDENTIALITY becomes a (business) requirement:
  - VOs aren't created for fun, but to achieve some (business) goal.
  - Goals are threatened by risks, which are mitigated through (security) measures.
  - One possible measure is CONFIDENTIALITY; privacy from other VOs, privacy from other VO-members, privacy from grid operators.
- TRUST is a relation between VOs (organizations), VO-members (people) and the grid (technology), and is present in every workflow (process).

Privacy in the grid.

Privacy of VOs:

- VOs need privacy from other VOs
  - Example: insurance-VO should not be able to learn anything from healthcare-VO
  - AuthN+AuthZ cover primary concerns
    - Data privacy in storage: AuthN+AuthZ + **glite-data-hydra-service** (AES, IDEA or Blowfish)
    - Data privacy in transport: AuthN+AuthZ + TLS
    - Data privacy in processing: n/a to VO-level
- Suppose hospitals would like to contribute MRI-scans (DICOM) for research purposes, but they want to be unlinkable to that data?
  - This requires sender pseudonymity or sender anonymity
  - This could be facilitated through an overlay network on grid level

Privacy in the grid.

Privacy of VO-members:

- VO-members need privacy from other VO-members
  - Example: VO-members have private datasets which they don't want to disclose to other VO-members, but which they *do* want to include in the computation
  - AuthN+AuthZ don't suffice
    - Solution #1: agent-based secure computation
    - Solution #2: secure multiparty computation, or *SMC*
  - But what if VO-members want an VO-specific anonymous data store?
    - Overlay network within context of single VO?

Considerations.

- How does one *gather* privacy requirements, anyway?
  - Who should be interviewed? A VO? Every VO-member? The grid operator? A lawyer?
  - Is it OK to limit it to common sense, like a ‘privacy mindset’, or should someone establish official guidelines for privacy risk analysis?
  - How should(n’t) privacy technology facilitate this? And refined insights in the future?
- How to *specify*?
- How to *verify*?
- How does one know one’s *coverage* of privacy issues, and if a spot is missed? What are the implications, and for whom?

Addressing privacy.

- Crypto for message privacy
- Mixnets for delivery privacy
- Data minimization
  - 'Avoid the problem': preferred if possible, but often not enough
  - User-centric identity management
- Metadata models, formal methods
  - (E)P3P
  - Hippocratic databases
- Privacy-aware system engineering
  - Adhere to purpose binding, accountability, ...
  - Security models: Bell-LaPadula, Clark-Wilson, Biba
  - EU PRIME

Relevance of other research.

Relevance of other research:

- Workflows
- Grid computing
- Streaming media
- Distributed filesystems
- Autonomic {computing,communication}
- Law enforcement, surveillance
- Data mining

Summary

Topics discussed:

- Justification of research on (grid) privacy.
- Exploration of privacy, overlay networks.
- Approaches to solving privacy problems.
- Approaches to creating new problems.

Feedback!

Thanks for your attention

Are there any questions?