Innovative Ideas in Privacy Research

Prof. Bharat Bhargava
Department of Computer Sciences, Purdue University, West Lafayette, IN 47907
bb@cs.purdue.edu
http://www.cs.purdue.edu/homes/bb
Sept 2006

This research is supported by Cisco, Motorola, NSF grants, ANI 0219110, CCR-0001788, IIS-0209059 and 0242840

Introduction

- Privacy is fundamental to trusted collaboration and interactions to protect against malicious users and fraudulent activities.
- Privacy is needed to protect source of information, the destination of information, the route of information transmission of dissemination and the information content itself

Introduction

A. Basis for idea: The semantic of information changes with time, context and interpretation by humans

Ideas for privacy:

- Replication and Equivalence and Similarity
- Aggregation and Generalization
- Exaggeration and Mutilation
- Anonymity and Crowds
- Access Permissions, Authentication, Views

Introduction

B. Basis for Idea: The exact address may only be known in the neighborhood of a peer (node)

Idea for Privacy:

- Request is forwarded towards an approximate direction and position
- Granularity of location can be changed
- Remove association between the content of the information and the identity of the source of information
- Somebody may know the source while others may know the content but not both
- Timely position reports are needed to keep a node traceable but this leads to the disclosure of the trajectory of node movement
- Enhanced algorithm(AO2P) can use the position of an abstract reference point instead of the position of destination
- Anonymity as a measure of privacy can be based on probability of matching a position of a node to its id and the number of nodes in a particular area representing a position
- Use trusted proxies to protect privacy

Introduction

C. Basis for idea: Some people or sites can be trusted more than others due to evidence, credibility , past interactions and recommendations

Ideas for privacy:

- Develop measures of trust and privacy
- Trade privacy for trust
- Offer private information in increments over a period of time

Introduction

D. Basis for idea: It is hard to specify the policies for privacy preservation in a legal, precise, and correct manner. It is even harder to enforce the privacy policies

Ideas for privacy:

- Develop languages to specify policies
- Bundle data with policy constraints
- Use obligations and penalties
- Specify when, who, and how many times the private information can be disseminated
- Use Apoptosis to destroy private information

"To Report or Not To Report": Tension between Personal Privacy and Public Responsibility

An info tech company will typically lose between ten and one hundred times more money from shaken consumer confidence than the hack attack itself represents if they decide to prosecute the case.

Mike Rasch, VP Global Security, testimony before the Senate Appropriations Subcommittee, February 2000 reported in The Register and online testimony transcript

Further Reluctance to Report

- One common fear is that a crucial piece of equipment, like a main server, say, might be impounded for evidence by over-zealous investigators, thereby shutting the company down.

- Estimate: fewer than one in ten serious intrusions are ever reported to the authorities.

Mike Rasch, VP Global Security, testimony before the Senate Appropriations Subcommittee, February 2000

reported in The Register and online testimony transcript

Methods of Defense

- Five basic approaches to defense of computing systems
  - Prevent attack
    - Block attack / Close vulnerability
  - Deter attack
    - Make attack harder (can't make it impossible ☹)
  - Deflect attack
    - Make another target more attractive than this target
  - Detect attack
    - During or after
  - Recover from attack

A) Controls

- Castle in Middle Ages
  - Location with natural obstacles
  - Surrounding moat
  - Drawbridge
  - Heavy walls
    - Arrow slits
    - Crenellations
  - Strong gate
    - Tower
  - Guards / passwords
- Computers Today
  - Encryption
  - Software controls
  - Hardware controls
  - Policies and procedures
  - Physical controls

- Medieval castles
  - location (steep hill, island, etc.)
  - moat / drawbridge / walls / gate / guards /passwords
  - another wall / gate / guards /passwords
  - yet another wall / gate / guards /passwords
  - tower / ladders up
- Multiple controls in computing systems can include:
  - system perimeter – defines „inside/outside”
  - preemption – attacker scared away
  - deterrence – attacker could not overcome defenses
  - faux environment (e.g. honeypot, sandbox) – attack deflected towards a worthless target (but the attacker doesn't know about it!)

→ Note layered defense /

multilevel defense / defense in depth (ideal!)

A.2) Controls: Policies and Procedures

- Policy vs. Procedure
  - Policy: *What* is/what is not allowed
  - Procedure: *How* you enforce policy
- Advantages of policy/procedure controls:
  - Can replace hardware/software controls
  - Can be least expensive
    - Be careful to consider *all* costs
      - E.g. help desk costs often ignored for for passwords (=> look cheap but migh be expensive)

- Policy - must consider:
  - Alignment with users' legal and ethical standards
  - Probability of use (e.g. due to inconvenience)

    Inconvenient: 200 character password,
    change password every week

    (Can be) good: biometrics replacing passwords
  - Periodic reviews
    - As people and systems, as well as their goals, change

A.3) Controls: Physical Controls

- Walls, locks
- Guards, security cameras
- Backup copies and archives
- Cables an locks (e.g., for notebooks)
- Natural and man-made disaster protection
  - Fire, flood, and earthquake protection
  - Accident and terrorism protection
- ...

B) Effectiveness of Controls

- Awareness of problem
  - People convined of the need for these controls
- Likelihood of use
  - Too complex/intrusive security tools are often disabled
- Overlapping controls
  - >1 control for a given vulnerability
    - To provide layered defense – the next layer compensates for a failure of the previous layer
- Periodic reviews
  - A given control usually becomess less effective with time
  - Need to replace ineffective/inefficient controls with better ones

2. Introduction to Privacy in Computing

Outline

1) Introduction (def., dimensions, basic principles, …)
2) Recognition of the need for privacy
3) Threats to privacy
4) Privacy Controls
    4.1) Technical privacy controls - Privacy-Enhancing Technologies (PETs)
        a) Protecting user identities
        b) Protecting usee identities
        c) Protecting confidentiality & integrity of personal data
    4.2) Legal privacy controls
        a) Legal World Views on Privacy
        b) International Privacy Laws: Comprehensive or Sectoral
        c) Privacy Law Conflict between European Union – USA
        d) A Common Approach: Privacy Impact Assessments (PIA)
        e) Observations & Conclusions
5) Selected Advanced Topics in Privacy
    5.1) Privacy in pervasive computing
    5.2) Using trust paradigm for privacy protection
    5.3) Privacy metrics
    5.4) Trading privacy for trust

1. Introduction (1)

[cf. Simone Fischer-Hübner]

- Def. of privacy [Alan Westin, Columbia University, 1967]

  = the claim of individuals, groups and institutions to determine for themselves, when, how and to what extent information about them is communicated to others

- 3 dimensions of privacy:

  1) Personal privacy

     Protecting a person against undue interference (such as physical searches) and information that violates his/her moral sense

  2) Territorial privacy

     Protecting a physical area surrounding a person that may not be violated without the acquiescence of the person

     - Safeguards: laws referring to trespassers search warrants

  3) Informational privacy

     Deals with the gathering, compilation and selective dissemination of information

1. Introduction (2)

[cf. Simone Fischer-Hübner]

- Basic privacy principles
  - Lawfulness and fairness
  - Necessity of data collection and processing
  - Purpose specification and purpose binding
    - There are no "non-sensitive" data
  - Transparency
    - Data subject´s right to information correction, erasure or blocking of incorrect/ illegally stored data
  - Supervision (= control by independent data protection authority) & sanctions
  - Adequate organizational and technical safeguards
- Privacy protection can be undertaken by:
  - Privacy and data protection laws promoted by government
  - Self-regulation for fair information practices by codes of conducts promoted by businesses
  - Privacy-enhancing technologies (PETs) adopted by individuals
  - Privacy education of consumers and IT professionals

2. Recognition of Need for Privacy Guarantees (1)

- By individuals [Cran *et al. '99*]
  - 99% unwilling to reveal their SSN
  - 18% unwilling to reveal their… favorite TV show
- By businesses
  - Online consumers worrying about revealing personal data held back $15 billion in online revenue in 2001
- By Federal government
  - Privacy Act of 1974 for Federal agencies
  - Health Insurance Portability and Accountability Act of 1996 (HIPAA)

2. Recognition of Need for Privacy Guarantees (2)

- By computer industry research (examples)
  - Microsoft Research
    - The biggest research challenges:

      According to Dr. Rick Rashid, Senior Vice President for Research
      - Reliability / Security / Privacy / Business Integrity
        - » Broader: application integrity (just "integrity?")

      => MS Trustworthy Computing Initiative

    - Topics include: DRM—digital rights management (incl. watermarking surviving photo editing attacks), software rights protection, intellectual property and content protection, database privacy and p.-p. data mining, anonymous e-cash, anti-spyware
  - IBM (incl. Privacy Research Institute)
    - Topics include: pseudonymity for e-commerce, EPA and EPAL—enterprise privacy architecture and language, RFID privacy, p.-p. video surveillance, federated identity management (for enterprise federations), p.-p. data mining and p.-p.mining of association rules, hippocratic (p.-p.) databases, online privacy monitoring

2. Recognition of Need for Privacy Guarantees (3)

- By academic researchers (examples from the U.S.A.)
  - CMU and Privacy Technology Center
    - Latanya Sweeney (k-anonymity, SOS—Surveillance of Surveillances, genomic privacy)
    - Mike Reiter (Crowds – anonymity)
  - Purdue University – CS and CERIAS
    - Elisa Bertino (trust negotiation languages and privacy)
    - Bharat Bhargava (privacy-trust tradeoff, privacy metrics, p.-p. data dissemination, p.-p. location-based routing and services in networks)
    - Chris Clifton (p.-p. data mining)
    - Leszek Lilien (p.-p. data disemination)
  - UIUC
    - Roy Campbell (Mist – preserving location privacy in pervasive computing)
    - Marianne Winslett (trust negotiation w/ controled release of private credentials)
  - U. of North Carolina Charlotte
    - Xintao Wu, Yongge Wang, Yuliang Zheng (p.-p. database testing and data mining)

3. Threat to Privacy (1)

[cf. Simone Fischer-Hübner]

1) Threats to privacy at application level

- Threats to collection / transmission of large quantities of personal data
  - Incl. projects for new applications on Information Highway, e.g.:
    - Health Networks / Public administration Networks
    - Research Networks / Electronic Commerce / Teleworking
    - Distance Learning / Private use
  - Example: Information infrastructure for a better healthcare
    [cf. Danish "INFO-Society 2000"- or Bangemann-Report]
    - National and European healthcare networks for the interchange of information
    - Interchange of (standardized) electronic patient case files
    - Systems for tele-diagnosing and clinical treatment

3. Threat to Privacy (2)

[cf. Simone Fischer-Hübner]

2) Threats to privacy at communication level

- Threats to anonymity of sender / forwarder / receiver
- Threats to anonymity of service provider
- Threats to privacy of communication
  - E.g., via monitoring / logging of transactional data
    - Extraction of user profiles & its long-term storage

3) Threats to privacy at system level

- E.g., threats at system access level

4) Threats to privacy in audit trails

3. Threat to Privacy (3)

[cf. Simone Fischer-Hübner]

- Identity theft – the most serious crime against privacy
- Threats to privacy – another view
  - Aggregation and data mining
  - Poor system security
  - Government threats
    - Gov't has a lot of people's most private data
      - Taxes / homeland security / etc.
    - People's privacy vs. homeland security concerns
  - The Internet as privacy threat
    - Unencrypted e-mail / web surfing / attacks
  - Corporate rights and private business
    - Companies may collect data that U.S. gov't is *not* allowed to
  - Privacy for sale - many traps
    - "Free" is not free…
      - E.g., accepting frequent-buyer cards reduces your privacy

4. Privacy Controls

1) Technical privacy controls - Privacy-Enhancing Technologies (PETs)

   a) Protecting user identities
   b) Protecting usee identities
   c) Protecting confidentiality & integrity of personal data

2) Legal privacy controls

4.1 Technical Privacy Controls (1)

[cf. Simone Fischer-Hübner]

- Technical controls - Privacy-Enhancing Technologies (PETs)

a) Protecting user identities via, e.g.:

  - Anonymity - a user may use a resource or service without disclosing her identity
  - Pseudonymity - a user acting under a pseudonym may use a resource or service without disclosing his identity
  - Unobservability - a user may use a resource or service without others being able to observe that the resource or service is being used
  - Unlinkability - sender and recipient cannot be identified as communicating with each other

4.1 Technical Privacy Controls (2)

[cf. Simone Fischer-Hübner]

- Taxonomies of pseudonyms
  - Taxonomy of pseudonyms w.r.t. their function
    - i) Personal pseudonyms
      - » Public personal pseudonyms / Nonpublic personal pseudonyms / Private personal pseudonyms
    - ii) Role pseudonyms
      - » Business pseudonyms / Transaction pseudonyms
  - Taxonomy of pseudonyms w.r.t. their generation
    - i) Self-generated pseudonyms
    - ii) Reference pseudonyms
    - iii) Cryptographic pseudonyms
    - iv) One-way pseudonyms

4.1 Technical Privacy Controls (3)

[cf. Simone Fischer-Hübner]

b) Protecting usee identities via, e.g.:
Depersonalization (anonymization) of data subjects

- Perfect depersonalization:
  - Data rendered anonymous in such a way that the data subject is no longer identifiable
- Practical depersonalization:
  - The modification of personal data so that the information concerning personal or material circumstances can no longer or only with a disproportionate amount of time, expense and labor be attributed to an identified or identifiable individual
- Controls for depersonalization include:
  - Inference controls for statistical databases
  - Privacy-preserving methods for data mining

4.1 Technical Privacy Controls (4)

[cf. Simone Fischer-Hübner]

- The risk of reidentification (a threat to anonymity)
  - Types of data in statistical records:
    - Identity data - e.g., name, address, personal number
    - Demographic data - e.g., sex, age, nationality
    - Analysis data - e.g., diseases, habits
  - The degree of anonymity of statistical data depends on:
    - Database size
    - The entropy of the demographic data attributes that can serve as supplementary knowledge for an attacker
  - The entropy of the demographic data attributes depends on:
    - The number of attributes
    - The number of possible values of each attribute
    - Frequency distribution of the values
    - Dependencies between attributes

4.1 Technical Privacy Controls (5)

[cf. Simone Fischer-Hübner]

c) Protecting confidentiality and integrity of personal data via, e.g.:

- Privacy-enhanced identity management
- Limiting access control
  - Incl. formal privacy models for access control
- Enterprise privacy policies
- Steganography
- Specific tools
  - Incl. P3P (Platform for Privacy Preferences)

4.2 Legal Privacy Controls (1)

[cf. A.M. Green, Yale, 2004]

- Outline

a) Legal World Views on Privacy
b) International Privacy Laws:
    - Comprehensive Privacy Laws
    - Sectoral Privacy Laws

c) Privacy Law Conflict European Union vs. USA
d) A Common Approach: Privacy Impact Assessments (PIA)
e) Observations & Conclusions

4.2 Legal Privacy Controls (2)

[cf. A.M. Green, Yale, 2004]

a) Legal World Views on Privacy (1)

- General belief: Privacy is a fundamental human right that has become one of the most important rights of the modern age

- Privacy also recognized and protected by individual countries
  - At a minimum each country has a provision for rights of inviolability of the home and secrecy of communications
  - Definitions of privacy vary according to context and environment

4.2 Legal Privacy Controls (3)

[cf. A.M. Green, Yale, 2004]

a) Legal World Views on Privacy (2)

United States: “Privacy is the right to be left alone” - Justice Louis Brandeis

UK: “the right of an individual to be protected against intrusion into his personal life or affairs by direct physical means or by publication of information

Australia: “Privacy is a basic human right and the reasonable expectation of every person”

4.2 Legal Privacy Controls (4)

[cf. A.M. Green, Yale, 2004]

b) International Privacy Laws

- Two types of privacy laws in various countries:

1) Comprehensive Laws
   - Def: General laws that govern the collection, use and dissemination of personal information by public & private sectors
   - Require commissioners or independent enforcement body
   - Difficulty: lack of resources for oversight and enforcement; agencies under government control
   - Examples: European Union, Australia, Canada and the UK

2) Sectoral Laws
   - Idea: Avoid general laws, focus on specific sectors instead
   - Advantage: enforcement through a range of mechanisms
   - Disadvantage: each new technology requires new legislation
   - Example: United States

4.2 Legal Privacy Controls (5)

[cf. A.M. Green, Yale, 2004]

b) International Privacy Laws
Comprehensive Laws – European Union

- European Union Council adopted the new Privacy Electronic Communications Directive [cf. A.M. Green, Yale, 2004]
  - Prohibits secondary uses of data without informed consent
  - No transfer of data to non EU countries unless there is adequate privacy protection
    - Consequences for the USA

- EU laws related to privacy include
  - 1994 — EU Data Protection Act
  - 1998 — EU Data Protection Act
    - Privacy protections stronger than in the U.S.

4.2 Legal Privacy Controls (6)

[cf. A.M. Green, Yale, 2004]

b) International Privacy Laws
Sectoral Laws – United States (1)

- No explicit right to privacy in the constitution
- Limited constitutional right to privacy implied in number of provisions in the Bill of Rights
- A patchwork of federal laws for specific categories of personal information
  - E.g., financial reports, credit reports, video rentals, etc.
- No legal protections, e.g., for individual's privacy on the internet are in place (as of Oct. 2003)
- White House and private sector believe that self-regulation is enough and that no new laws are needed (exception: medical records)
- Leads to conflicts with other countries' privacy policies

4.2 Legal Privacy Controls (7)

[cf. A.M. Green, Yale, 2004]

b) International Privacy Laws
Sectoral Laws – United States (2)

- American laws related to privacy include:
  - 1974 — US Privacy Act
    - Protects privacy of data collected by the executive branch of federal gov't
  - 1984 — US Computer Fraud and Abuse Act
    - Penalties: max{100K, stolen value} and/or 1 to 20 yrs
  - 1986 — US Electronic Communications Privacy Act
    - Protects against wiretapping
    - Exceptions: court order, ISPs
  - 1996 — US Economic Espionage Act
  - 1996 — HIPAA
    - Privacy of individuals' medical records
  - 1999 — Gramm-Leach-Bliley Act
    - Privacy of data for customers of financial institutions
  - 2001 — USA Patriot Act
  - — US Electronic Funds Transfer Act
  - — US Freedom of Information Act

4.2 Legal Privacy Controls (8)

[cf. A.M. Green, Yale, 2004]

c) Privacy Law Conflict: EU vs. The United States

- US lobbied EU for 2 years (1998-2000) to convince it that the US system is adequate

- Result was the "Safe Harbor Agreement" (July 2000):
  US companies would voluntarily self-certify to adhere to a set of privacy principles worked out by US Department of Commerce and Internal Market Directorate of the European Commission
  - Little enforcement: A self-regulatory system in which companies merely promise not to violate their declared privacy practices
  - Criticized by privacy advocates and consumer groups in both US and Europe

- Agreement re-evaluated in 2003
  - Main issue: European Commission doubted effectiveness of the sectoral/self-regulatory approach

4.2 Legal Privacy Controls (9)

[cf. A.M. Green, Yale, 2004]

d) A Common Approach: Privacy Impact Assessments (PIA)

- An evaluation conducted to assess how the adoption of new information policies, the procurement of new computer systems, or the initiation of new data collection programs will affect individual privacy
- The premise: Considering privacy issues at the early stages of a project cycle will reduce potential adverse impacts on privacy after it has been implemented
- Requirements:
  - PIA process should be independent
  - PIA performed by an independent entity (office and/or commissioner) not linked to the project under review
  - Participating countries: US, EU, Canada, etc.

4.2 Legal Privacy Controls (10)

[cf. A.M. Green, Yale, 2004]

d) A Common Approach: PIA (2)

- EU implemented PIAs
- Under the *European Union Data Protection Directive*, all EU members must have an independent privacy enforcement body
- PIAs soon to come to the United States (as of 2003)
- US passed the E-Government Act of 2002 which requires federal agencies to conduct privacy impact assessments before developing or procuring information technology

4.2 Legal Privacy Controls (11)

[cf. A.M. Green, Yale, 2004]

e) Observations and Conclusions (1)

- Observation 1: At present too many mechanisms seem to operate on a national or regional, rather than global level
  - E.g., by OECD
- Observation 2: Use of self-regulatory mechanisms for the protection of online activities seems somewhat haphazard and is concentrated in a few member countries
- Observation 3: Technological solutions to protect privacy are implemented to a limited extent only
- Observation 4: Not enough being done to encourage the implementation of technical solutions for privacy compliance and enforcement
  - Only a few member countries reported much activity in this area

4.2 Legal Privacy Controls (12)

[cf. A.M. Green, Yale, 2004]

e) Observations and Conclusions (2)

- Conclusions
  - Still work to be done to ensure the security of personal information for all individuals in all countries
  - Critical that privacy protection be viewed in a global perspective
    - Better than a purely national one –

      To better handle privacy violations that cross national borders

5 Selected Advanced Topics in Privacy (1)

[cf. A.M. Green, Yale, 2004]

Outline

5.1) Privacy in pervasive computing
5.2) Using trust paradigm for privacy protection
5.3) Privacy metrics
5.4) Trading privacy for trust

5.1 Privacy in Pervasive Computing (1)

- In pervasive computing environments, socially-based paradigms (incl. trust) will play a big role
- People surrounded by zillions of computing devices of all kinds, sizes, and aptitudes ["Sensor Nation: Special Report," *IEEE Spectrum*, vol. 41, no. 7, 2004]
  - Most with limited / rudimentary capabilities
    - Quite small, e.g., RFID tags, smart dust
  - Most embedded in artifacts for everyday use, or even human bodies
    - Possible both beneficial and detrimental (even apocalyptic) consequences
- Danger of malevolent *opportunistic* sensor networks
  — pervasive devices self-organizing into huge spy networks
  - Able to spy anywhere, anytime, on everybody and everything
  - Need means of detection & neutralization
    - To tell which and how many snoops are active, what data they collect, and who they work for
      - An advertiser? a nosy neighbor? Big Brother?
    - Questions such as "Can I trust my refrigerator?" will not be jokes
      - The refrigerator snitching on its owner's dietary misbehavior for her doctor

5.1 Privacy in Pervasive Computing (2)

- Will pervasive computing destroy privacy? (as we know it)
  - Will a cyberfly end privacy?
    - With high-resolution camera eyes and supersensitive microphone ears
  - If a cyberfly too clever drown in the soup, we'll build cyberspiders
  - But then opponents' cyberbirds might eat those up
  - So, we'll build a cybercat
  - And so on and so forth …
- Radically changed reality demands new approaches to privacy
  - Maybe need a new privacy category—namely, artifact privacy?
  - Our belief: Socially based paradigms (such as trust-based approaches) will play a big role in pervasive computing
    - Solutions will vary (as in social settings)
      - Heavyweighty solutions for entities of high intelligence and capabilities (such as humans and intelligent systems) interacting in complex and important matters
      - Lightweight solutions for less intelligent and capable entities interacting in simpler matters of lesser consequence

5.2 Using Trust for Privacy Protection (1)

- Privacy = entity's ability to control the availability and exposure of information about itself
  - We extended the subject of privacy from a person in the original definition ["Internet Security Glossary," *The Internet Society, Aug. 2004*] to an entity— including an organization or software
    - Controversial but stimulating
    - Important in pervasive computing
- Privacy and trust are closely related
  - Trust is a socially-based paradigm
  - Privacy-trust tradeoff: Entity can trade privacy for a corresponding gain in its partners' trust in it
  - The scope of an entity's privacy disclosure should be proportional to the benefits expected from the interaction
    - As in social interactions
    - E.g.: a customer applying for a mortgage must reveal much more personal data than someone buying a book

5.2 Using Trust for Privacy Protection (2)

- Optimize degree of privacy traded to gain trust
  - Disclose minimum needed for gaining partner's necessary trust level
- To optimize, need privacy & trust measures
  Once measures available:
  - Automate evaluations of the privacy loss and trust gain
  - Quantify the trade-off
  - Optimize it
- Privacy-for-trust trading requires privacy guarantees for further dissemination of private info
  - Disclosing party needs satisfactory limitations on further dissemination (or the lack of thereof) of traded private information
  - E.g., needs partner's solid privacy policies
    - Merely perceived danger of a partner's privacy violation can make the disclosing party reluctant to enter into a partnership
      - E.g., a user who learns that an ISP has carelessly revealed any customer's email will look for another ISP

5.2 Using Trust for Privacy Protection (3)

- Conclusions on Privacy and Trust
  - Without privacy guarantees, there can be no trust and trusted interactions
    - People will avoid *trust-building negotiations* if their privacy is threatened by the negotiations
    - W/o trust-building negotiations no *trust* can be established
    - W/o trust, there are no *trusted interactions*
  - Without privacy guarantees, lack of trust will cripple the promise of pervasive computing
    - Bec. people will avoid untrusted interactions with privacy-invading pervasive devices / systems
      - E.g., due to the fear of *opportunistic sensor networks*

        Self-organized by electronic devices around us – can *harm* people in their midst
  - Privacy must be guaranteed for trust-building negotiations

5.3 Privacy Metrics (1)

Outline

- Problem and Challenges
- Requirements for Privacy Metrics
- Related Work
- Proposed Metrics
  A. Anonymity set size metrics
  B. Entropy-based metrics

5.3 Privacy Metrics (2)

a) Problem and Challenges

- Problem
  - How to determine that certain degree of data privacy is provided?

- Challenges
  - Different privacy-preserving techniques or systems claim different degrees of data privacy
  - Metrics are usually ad hoc and customized
    - Customized for a user model
    - Customized for a specific technique/system
  - Need to develop uniform privacy metrics
    - To confidently compare different techniques/systems

5.3 Privacy Metrics (3a)

b) Requirements for Privacy Metrics

- Privacy metrics should account for:
  - Dynamics of legitimate users
    - How users interact with the system?

      E.g., repeated patterns of accessing the same data can leak information to a violator
  - Dynamics of violators
    - How much information a violator gains by watching the system for a period of time?
  - Associated costs
    - Storage, injected traffic, consumed CPU cycles, delay

5.3 Privacy Metrics (3b)

c) Related Work

- Anonymity set without accounting for probability distribution [Reiter and Rubin, 1999]
- An entropy metric to quantify privacy level, assuming static attacker model [Diaz *et al.,* 2002]
- Differential entropy to measure how well an attacker estimates an attribute value [Agrawal and Aggarwal 2001]

5.3 Privacy Metrics (4)

d) Proposed Metrics

A. Anonymity set size metrics

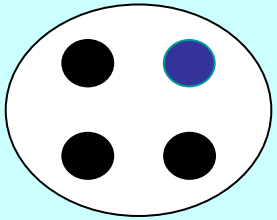
B. Entropy-based metrics

5.3 Privacy Metrics (5)

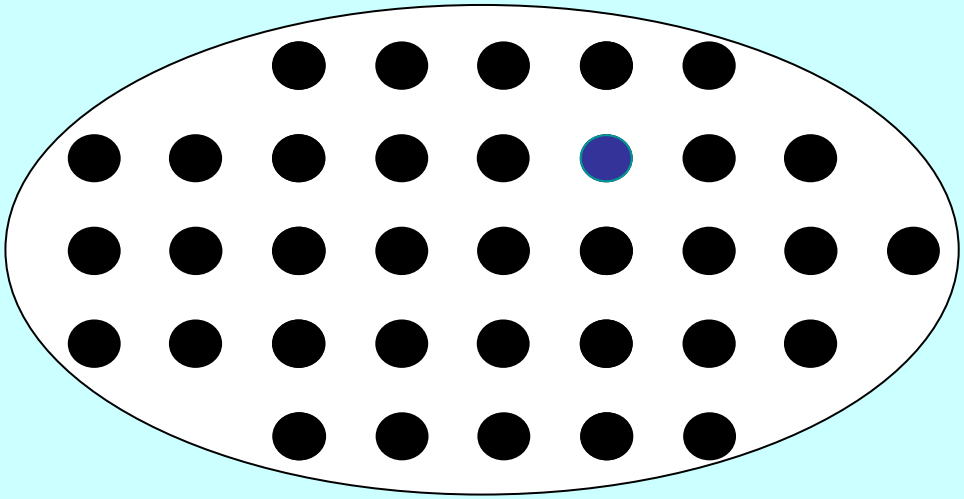
A. Anonymity Set Size Metrics

- The larger set of indistinguishable entities, the lower probability of identifying any one of them
  - Can use to ”anonymize” a selected private attribute value within the domain of its all possible values

“Hiding in a crowd”



“Less” anonymous (1/4)



“More” anonymous (1/*n*)

5.3 Privacy Metrics (6)

Anonymity Set

- Anonymity set A

  $A = \{(s_1, p_1), (s_2, p_2), \ldots, (s_n, p_n)\}$

  – s_i: subject *i* who might access private data

  or: *i*-th possible value for a private data attribute

  – p_i: probability that s_i accessed private data

  or: probability that the attribute assumes the *i*-th possible value

5.3 Privacy Metrics (7)

Effective Anonymity Set Size

- Effective anonymity set size is

$$L = |A| \sum_{i=1}^{|A|} \min(p_i, 1/|A|)$$

  - Maximum value of L is |A| iff all p_i's are equal to 1/|A|
  - L below maximum when distribution is skewed
    - skewed when p_i's have different values

- Deficiency:
  L does not consider violator's *learning* behavior

5.3 Privacy Metrics (8)

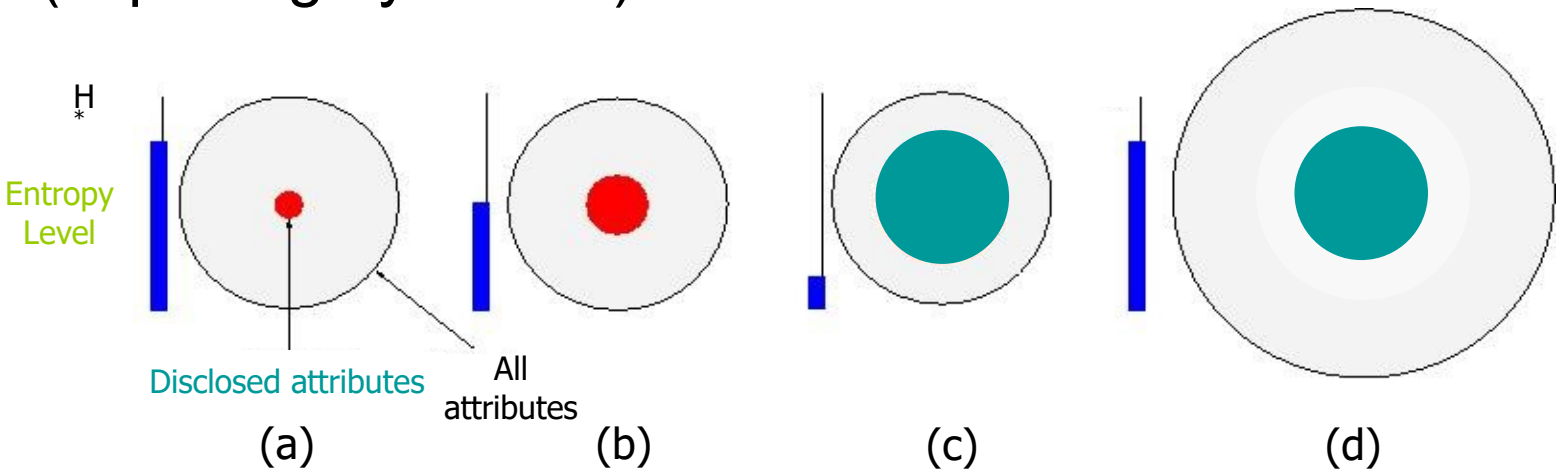
B. Entropy-based Metrics

- Entropy measures the randomness, or uncertainty, in private data
- When a violator gains more information, entropy decreases
- Metric: Compare the current entropy value with its maximum value
  - The difference shows how much information has been leaked

5.3 Privacy Metrics (9)

Dynamics of Entropy

- Decrease of system entropy with attribute disclosures (capturing dynamics)



  - When entropy reaches a threshold (b), *data evaporation* can be invoked to increase entropy by controlled data distortions
  - When entropy drops to a very low level (c), *apoptosis* can be triggered to destroy private data
  - Entropy increases (d) if the set of attributes grows or the disclosed attributes become less valuable – e.g., obsolete or more data now available

5.3 Privacy Metrics (10)

Quantifying Privacy Loss

- Privacy loss *D(A,t)* at time *t*, when a subset of attribute values *A* might have been disclosed:

$$D(A,t) = H^{*}(A) - H(A,t)$$

  - $H^{*}(A)$ – the maximum entropy
    - Computed when probability distribution of p_i's is uniform
  - *H(A,t)* is entropy at time *t*

$$H(A,t) = \sum_{j=1}^{|A|} w_j \left(\sum_{\forall i} \left(- p_i \log_2 (p_i) \right) \right)$$

  - w_j – weights capturing relative privacy "value" of attributes

5.3 Privacy Metrics (11)

Using Entropy in Data Dissemination

- Specify two thresholds for *D*
  - For triggering evaporation
  - For triggering apoptosis
- When private data is exchanged
  - Entropy is recomputed and compared to the thresholds
  - Evaporation or apoptosis may be invoked to enforce privacy

5.3 Privacy Metrics (12)

Entropy Example

- Consider a private phone number: $(a_1a_2a_3)\ a_4a_5\,a_6 - a_7a_8a_9\,a_{10}$
- Each digit is stored as a value of a separate attribute
- Assume:
  - Range of values for each attribute is [0—9]
  - All attributes are equally important, i.e., $w_j = 1$
- The maximum entropy – when violator has no information about the value of each attribute:
  - Violator assigns a *uniform* probability distribution to values of each attribute
    - e.g., $a_1 = i$ with probability of 0.10 for each i in [0—9]

$$H^*(A) = \sum_{j=0}^{9}\left(w_j \sum_{i=1}^{10}\left(-0.1\ \log_2(0.1)\right)\right) = 33.3$$

5.3 Privacy Metrics (13)

Entropy Example – cont.

- Suppose that after time *t*, violator can figure out the state of the phone number, which may allow him to learn the three leftmost digits
- Entropy at time *t* is given by:

$$H(A,t) = 0 + \sum_{j=4}^{10} w_j \left(\sum_{i=0}^{9} \left(-0.1 \ \log_2 (0.1) \right) \right) = 23.3$$

  - Attributes a_1, a_2, a_3 contribute 0 to the entropy value because violator knows their correct values
- Information loss at time *t* is:

$$D(A,t) = H^*(A) - H(A,t) = 10.0$$

5.3 Privacy Metrics (14)

Selected Publications

- "Private and Trusted Interactions," by B. Bhargava and L. Lilien.
- "On Security Study of Two Distance Vector Routing Protocols for Mobile Ad Hoc Networks," by W. Wang, Y. Lu and B. Bhargava, Proc. of IEEE Intl. Conf. on Pervasive Computing and Communications (PerCom 2003), Dallas-Fort Worth, TX, March 2003. http://www.cs.purdue.edu/homes/wangwc/PerCom03wangwc.pdf
- "Fraud Formalization and Detection," by B. Bhargava, Y. Zhong and Y. Lu, Proc. of 5th Intl. Conf. on Data Warehousing and Knowledge Discovery (DaWaK 2003), Prague, Czech Republic, September 2003. http://www.cs.purdue.edu/homes/zhong/papers/fraud.pdf
- "Trust, Privacy, and Security. Summary of a Workshop Breakout Session at the National Science Foundation Information and Data Management (IDM) Workshop held in Seattle, Washington, September 14 - 16, 2003" by B. Bhargava, C. Farkas, L. Lilien and F. Makedon, CERIAS Tech Report 2003-34, CERIAS, Purdue University, November 2003. http://www2.cs.washington.edu/nsf2003 or https://www.cerias.purdue.edu/tools_and_resources/bibtex_archive/archive/2003-34.pdf
- "e-Notebook Middleware for Accountability and Reputation Based Trust in Distributed Data Sharing Communities," by P. Ruth, D. Xu, B. Bhargava and F. Regnier, Proc. of the Second International Conference on Trust Management (iTrust 2004), Oxford, UK, March 2004. http://www.cs.purdue.edu/homes/dxu/pubs/iTrust04.pdf
- "Position-Based Receiver-Contention Private Communication in Wireless Ad Hoc Networks," by X. Wu and B. Bhargava, submitted to the Tenth Annual Intl. Conf. on Mobile Computing and Networking (MobiCom'04), Philadelphia, PA, September - October 2004. http://www.cs.purdue.edu/homes/wu/HTML/research.html/paper_purdue/mobi04.pdf

Introduction to Privacy in Computing

Reference & Bibliography (1)

Ashley Michele Green, “International Privacy Laws. Sensitive Information in a Wired World,” CS 457 Report, Dept. of Computer Science, Yale Univ., October 30, 2003.

Simone Fischer-Hübner, "IT-Security and Privacy-Design and Use of Privacy-Enhancing Security Mechanisms", Springer Scientific Publishers, Lecture Notes of Computer Science, LNCS 1958, May 2001, ISBN 3-540-42142-4.

Simone Fischer-Hübner, “Privacy Enhancing Technologies, PhD course,” Session 1 and 2, Department of Computer Science, Karlstad University, Winter/Spring 2003,
[available at: http://www.cs.kau.se/~simone/kau-phd-course.htm].

Introduction to Privacy in Computing

Reference & Bibliography (2)

- Slides based on BB+LL part of the paper:

  Bharat Bhargava, Leszek Lilien, Arnon Rosenthal, Marianne Winslett, "Pervasive Trust," *IEEE Intelligent Systems*, Sept./Oct. 2004, pp.74-77

- Paper References:

1. *The American Heritage Dictionary of the English Language*, 4th ed., Houghton Mifflin, 2000.
2. B. Bhargava et al., *Trust, Privacy, and Security: Summary of a Workshop Breakout Session at the National Science Foundation Information and Data Management (IDM) Workshop held in Seattle,Washington, Sep. 14–16, 2003*, tech. report 2003-34, Center for Education and Research in Information Assurance and Security, Purdue Univ., Dec. 2003; www.cerias.purdue.edu/tools_and_resources/bibtex_archive/archive/2003-34.pdf.
3. "Internet Security Glossary," *The Internet Society*, Aug. 2004; www.faqs.org/rfcs/rfc2828.html.
4. B. Bhargava and L. Lilien "Private and Trusted Collaborations," to appear in *Secure Knowledge Management* (SKM 2004)*: A Workshop*, 2004.
5. "Sensor Nation: Special Report," *IEEE Spectrum*, vol. 41, no. 7, 2004.
6. R. Khare and A. Rifkin, "Trust Management on the World Wide Web," *First Monday*, vol. 3, no. 6, 1998; www.firstmonday.dk/issues/issue3_6/khare.
7. M. Richardson, R. Agrawal, and P. Domingos,"Trust Management for the Semantic Web," *Proc. 2nd Int'l Semantic Web Conf.*, LNCS 2870, Springer-Verlag, 2003, pp. 351–368.
8. P. Schiegg et al., "Supply Chain Management Systems—A Survey of the State of the Art," *Collaborative Systems for Production Management: Proc. 8th Int'l Conf. Advances in Production Management Systems* (APMS 2002), IFIP Conf. Proc. 257, Kluwer, 2002.
9. N.C. Romano Jr. and J. Fjermestad, "Electronic Commerce Customer Relationship Management: A Research Agenda," *Information Technology and Management*, vol. 4, nos. 2–3, 2003, pp. 233–258.

6 Trust and Privacy (1)

- Privacy = entity's ability to control the availability and exposure of information about itself
  - We extended the subject of privacy from a person in the original definition ["Internet Security Glossary," *The Internet Society, Aug. 2004*] to an entity— including an organization or software
    - Maybe controversial but stimulating
- Privacy Problem
  - Consider computer-based interactions
    - From a simple transaction to a complex collaboration
  - Interactions always involve *dissemination of private data*
    - It is voluntary, "pseudo-voluntary," or compulsory
      - Compulsory - e.g., required by law
  - Threats of privacy violations result in lower trust
  - Lower trust leads to isolation and lack of collaboration

6 Trust and Privacy (2)

- Thus, privacy and trust are closely related
  - Privacy-trust tradeoff: Entity can trade privacy for a corresponding gain in its partners' trust in it
  - The scope of an entity's privacy disclosure should be proportional to the benefits expected from the interaction
    - As in social interactions
    - E.g.: a customer applying for a mortgage must reveal much more personal data than someone buying a book
- Trust must be established before a privacy disclosure
  - Data – provide quality an integrity
  - End-to-end communication – sender authentication, message integrity
  - Network routing algorithms – deal with malicious peers, intruders, security attacks

6 Trust and Privacy (3)

- Optimize degree of privacy traded to gain trust
  - Disclose minimum needed for gaining partner's necessary trust level
- To optimize, need privacy & trust measures

  Once measures available:
  - Automate evaluations of the privacy loss and trust gain
  - Quantify the trade-off
  - Optimize it
- Privacy-for-trust trading requires privacy guarantees for further dissemination of private info
  - Disclosing party needs satisfactory limitations on further dissemination (or the lack of thereof) of traded private information
  - E.g., needs partner's solid privacy policies
    - Merely perceived danger of a partner's privacy violation can make the disclosing party reluctant to enter into a partnership
      - E.g., a user who learns that an ISP has carelessly revealed any customer's email will look for another ISP

6 Trust and Privacy (4)

- Summary: Trading Information for Trust in Symmetric and Asymmetric Negotiations - When/how can partners trust each other?
  - Symmetric „disclosing:”
    - Initial degree of trust / stepwise trust growth / establishes mutual „full” trust
    - Trades info for trust (info is *private* or *not*)
  - Symmetric „preserving:” (from distrust to trust)
    - Initial distrust / no stepwise trust growth / establishes mutual „full” trust
    - No trading of info for trust (info is private or not)
  - Asymmetric:
    - Initial „full” trust of Weaker into Stronger and *no* trust of Stronger into Weaker / stepwise trust growth / establishes „full” trust of Stronger into Weaker
    - Trades *private* info for trust

6 Trust and Privacy (5)

- Privacy-Trust Tradeoff: Trading Privacy Loss for Trust Gain
- We're focusing on asymmetric trust negotiations:
  The weaker party trades a (degree of) privacy loss for (a degree of) a trust gain as perceived by the stronger party
- Approach to trading privacy for trust: [Zhong and Bhargava, Purdue]
  - Formalize the privacy-trust tradeoff problem
  - Estimate *privacy loss* due to disclosing a credential set
  - Estimate *trust gain* due to disclosing a credential set
  - Develop algorithms that *minimize privacy loss* for required trust gain
    - Because nobody likes loosing more privacy than necessary

More details later

7. Trading Privacy for Trust*

Trading Privacy Loss for Trust Gain

- We're focusing on asymmetric trust negotiations:
  Trading privacy for trust

- Approach to trading privacy for trust:

  [Zhong and Bhargava, Purdue]

  - Formalize the privacy-trust tradeoff problem
  - Estimate *privacy loss* due to disclosing a credential set
  - Estimate *trust gain* due to disclosing a credential set
  - Develop algorithms that *minimize privacy loss* for required trust gain
    - Bec. nobody likes loosing more privacy than necessary

More details available

Proposed Approach

A. Formulate the privacy-trust tradeoff problem
B. Estimate privacy loss due to disclosing a set of credentials
C. Estimate trust gain due to disclosing a set of credentials
D. Develop algorithms that minimize privacy loss for required trust gain

A. Formulate Tradeoff Problem

- Set of private attributes that user wants to conceal
- Set of credentials
  - Subset of *revealed* credentials *R*
  - Subset of *unrevealed* credentials *U*
- Choose a subset of credentials *NC* from *U* such that:
  - *NC* satisfies the requirements for trust building
  - PrivacyLoss(*NC*+*R*) – PrivacyLoss(*R*) is minimized

Steps B – D of the Approach

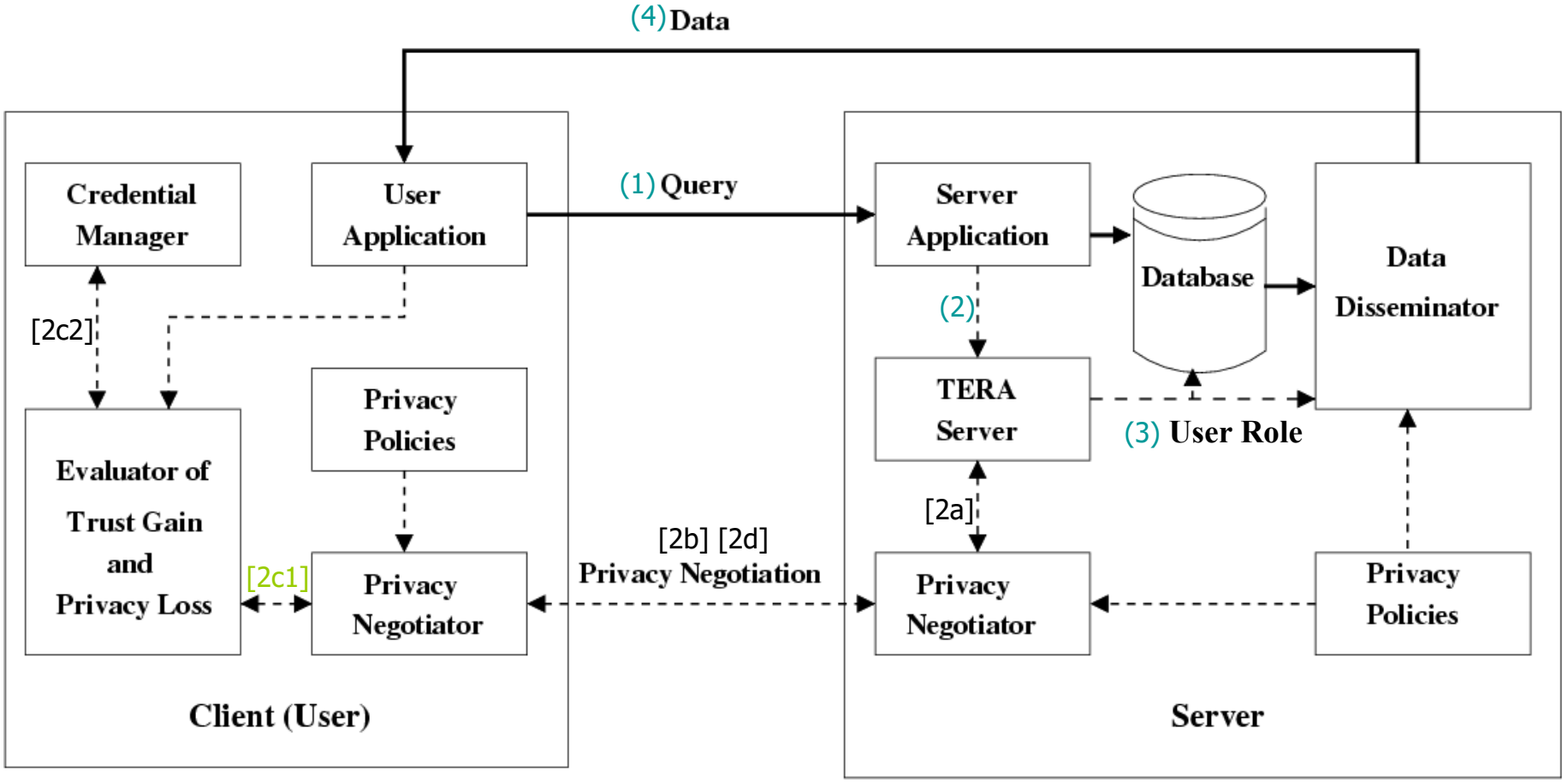
B. Estimate privacy loss due to disclosing a set of credentials
   - Requires defining privacy metrics

C. Estimate trust gain due to disclosing a set of credentials
   - Requires defining trust metrics

D. Develop algorithms that minimize privacy loss for required trust gain
   - Includes prototyping and experimentation

-- Details in another lecture of the series --

PRETTY Prototype for Experimental Studies



(<nr>) – unconditional path

[<nr>]– conditional path

TERA = Trust-Enhanced Role Assignment

Information Flow in PRETTY

1) User application sends query to server application.
2) Server application sends user information to TERA server for trust evaluation and role assignment.
   a) If a higher trust level is required for query, TERA server sends the request for more user's credentials to privacy negotiator.
   b) Based on server's privacy policies and the credential requirements, privacy negotiator interacts with user's privacy negotiator to build a higher level of trust.
   c) Trust gain and privacy loss evaluator selects credentials that will increase trust to the required level with the least privacy loss. Calculation considers credential requirements and credentials disclosed in previous interactions.
   d) According to privacy policies and calculated privacy loss, user's privacy negotiator decides whether or not to supply credentials to the server.
3) Once trust level meets the minimum requirements, appropriate roles are assigned to user for execution of his query.
4) Based on query results, user's trust level and privacy polices, data disseminator determines: (i) whether to distort data and if so to what degree, and (ii) what privacy enforcement metadata should be associated with it.

References

L. Lilien and B. Bhargava, "A scheme for privacy-preserving data dissemination," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, Vol. 36(3), May 2006, pp. 503-506.

Bharat Bhargava, Leszek Lilien, Arnon Rosenthal, Marianne Winslett, "Pervasive Trust," *IEEE Intelligent Systems*, Sept./Oct. 2004, pp.74-77

B. Bhargava and L. Lilien, "Private and Trusted Collaborations," *Secure Knowledge Management* (SKM 2004)*: A Workshop*, 2004.

B. Bhargava, C. Farkas, L. Lilien and F. Makedon, "Trust, Privacy, and Security. Summary of a Workshop Breakout Session at the National Science Foundation Information and Data Management (IDM) Workshop held in Seattle, Washington, September 14 - 16, 2003," CERIAS Tech Report 2003-34, CERIAS, Purdue University, Nov. 2003. http://www2.cs.washington.edu/nsf2003 or https://www.cerias.purdue.edu/tools_and_resources/bibtex_archive/archive/2003-34.pdf

"Internet Security Glossary," *The Internet Society*, Aug. 2004; www.faqs.org/rfcs/rfc2828.html.

"Sensor Nation: Special Report," *IEEE Spectrum*, vol. 41, no. 7, 2004.

R. Khare and A. Rifkin, "Trust Management on the World Wide Web," *First Monday*, vol. 3, no. 6, 1998; www.firstmonday.dk/issues/issue3_6/khare.

M. Richardson, R. Agrawal, and P. Domingos,"Trust Management for the Semantic Web," *Proc. 2nd Int'l Semantic Web Conf.*, LNCS 2870, Springer-Verlag, 2003, pp. 351–368.

P. Schiegg et al., "Supply Chain Management Systems—A Survey of the State of the Art," *Collaborative Systems for Production Management: Proc. 8th Int'l Conf. Advances in Production Management Systems* (APMS 2002), IFIP Conf. Proc. 257, Kluwer, 2002.

N.C. Romano Jr. and J. Fjermestad, "Electronic Commerce Customer Relationship Management: A Research Agenda," *Information Technology and Management*, vol. 4, nos. 2–3, 2003, pp. 233–258.

8. Using Entropy to Trade Privacy for Trust

Problem motivation

- Privacy and trust form an adversarial relationship
  - Internet users worry about revealing personal data. This fear held back $15 billion in online revenue in 2001
  - Users have to provide digital credentials that contain private information in order to build trust in open environments like Internet.
- Research is needed to quantify the tradeoff between privacy and trust

Subproblems

- How much privacy is lost by disclosing a piece of credential?
- How much does a user benefit from having a higher level of trust?
- How much privacy a user is willing to sacrifice for a certain amount of trust gain?

Proposed approach

- Formulate the privacy-trust tradeoff problem
- Design metrics and algorithms to evaluate the privacy loss. We consider:
  - Information receiver
  - Information usage
  - Information disclosed in the past
- Estimate trust gain due to disclosing a set of credentials
- Develop mechanisms empowering users to trade trust for privacy.
- Design prototype and conduct experimental study

Related work

- Privacy Metrics
  - Anonymity set without accounting for probability distribution [Reiter and Rubin, ’99]
  - Differential entropy to measure how well an attacker estimates an attribute value [Agrawal and Aggarwal ’01]
- Automated trust negotiation (ATN) [Yu, Winslett, and Seamons, ’03]
  - Tradeoff between the length of the negotiation, the amount of information disclosed, and the computation effort
- Trust-based decision making [Wegella *et al*. ’03]
  - Trust lifecycle management, with considerations of both trust and risk assessments
- Trading privacy for trust [Seigneur and Jensen, ’04]
  - Privacy as the linkability of pieces of evidence to a pseudonym; measured by using *nymity* [Goldberg, thesis, ’00]

Formulation of tradeoff problem (1)

- Set of private attributes that user wants to conceal
- Set of credentials
  - R(i): subset of credentials *revealed* to receiver *i*
  - U(i): credentials *unrevealed* to receiver *i*
- Credential set with minimal privacy loss
  - A subset of credentials *NC* from *U* (i)
  - *NC* satisfies the requirements for trust building
  - PrivacyLoss(*NC* ∪ *R(i)*) – PrivacyLoss(*R(i)*) is minimized

Formulation of tradeoff problem (2)

- Decision problem:
  - Decide whether trade trust for privacy or not
  - Determine minimal privacy damage
    - Minimal privacy damage is a function of minimal privacy loss, information usage and trustworthiness of information receiver.
  - Compute trust gain
  - Trade privacy for trust if trust gain > minimal privacy damage
- Selection problem:
  - Choose credential set with minimal privacy loss

Formulation of tradeoff problem (3)

- Collusion among information receivers
  - Use a global version R_g instead of *R(i)*
- Minimal privacy loss for multiple private attributes
  - nc_1 better for $attr_1$ but worse for $attr_2$ than nc_2
  - Weight vector $\{w_1, w_2, \ldots, w_m\}$ corresponds to the sensitivity of attributes
    - Salary is more sensitive than favorite TV show
  - Privacy loss can be evaluated using:
    - The weighted sum of privacy loss for all attributes
    - The privacy loss for the attribute with the highest weight

Two types of privacy loss (1)

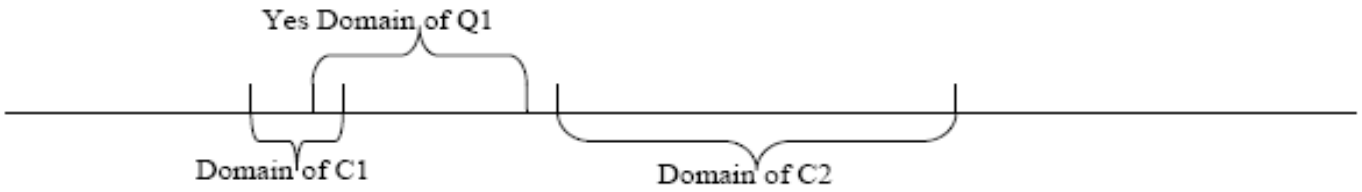
- Query-independent privacy loss
  - User determines her private attributes
  - Query-independent loss characterizes how helpful provided credentials for an adversarial to determine the probability density or probability mass function of a private attribute.

Two types of privacy loss (2)

- Query-dependent privacy loss
  - User determines a set of potential queries Q that she is reluctant to answer
  - Provided credentials reveal information of attribute set A. Q is a function of A.
  - Query-dependent loss characterizes how helpful provided credentials for an adversarial to determine the probability density or probability mass function of Q.

Observation 1

- High query-independent loss does not necessarily imply high query-dependent loss
  - An abstract example



Observation 2

- Privacy loss is affected by the order of disclosure
- Example:
  - Private attribute
    - age
  - Potential queries:

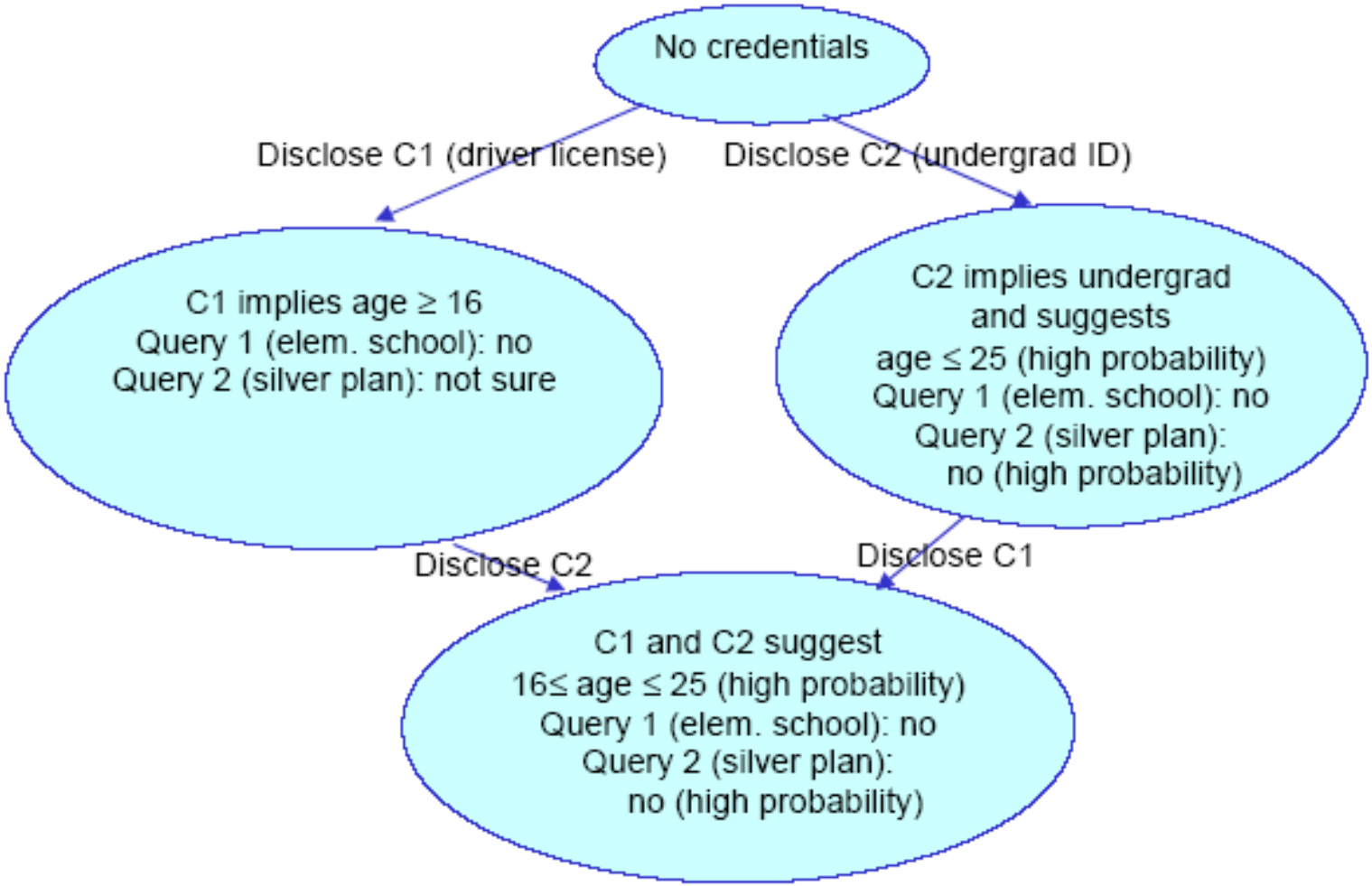
    (Q1) Is Alice an elementary school student?

    (Q2) Is Alice older than 50 to join a silver insurance plan?
  - Credentials

    (C1) Driver license

    (C2) Purdue undergraduate student ID

Example (1)


No credentials
Disclose C1 (driver license)
Disclose C2 (undergrad ID)
C1 implies age ≥ 16
Query 1 (elem. school): no
Query 2 (silver plan): not sure
C2 implies undergrad
and suggests
age ≤ 25 (high probability)
Query 1 (elem. school): no
Query 2 (silver plan):
no (high probability)
Disclose C2
Disclose C1
C1 and C2 suggest
16≤ age ≤ 25 (high probability)
Query 1 (elem. school): no
Query 2 (silver plan):
no (high probability)

Example (2)

- C1 → C2
  - Disclosing C1
    - low query-independent loss (wide range for age)
    - 100% loss for Query 1 (elem. school student)
    - low loss for Query 2 (silver plan)
  - Disclosing C2
    - high query-independent loss (narrow range for age)
    - zero loss for Query 1 (because privacy was lost by disclosing license)
    - high loss for Query 2 ("not sure" → "no - high probability"
- C2 → C1
  - Disclosing C2
    - low query-independent loss (wide range for age)
    - 100% loss for Query 1 (elem. school student)
    - high loss for Query 2 (silver plan)
  - Disclosing C1
    - high query-independent loss (narrow range of age)
    - zero loss for Query 1 (because privacy was lost by disclosing ID)
    - zero loss for Query 2

Entropy-based privacy loss

- Entropy measures the randomness, or uncertainty, in private data.
- When an adversarial gains more information, entropy decreases
- The difference shows how much information has been leaked
- Conditional probability is needed for entropy evaluation
  - Bayesian networks, kernel density estimation or subjective estimation can be adopted

Estimation of query-independent privacy loss

- Single attribute
  - Domain of attribute *a* : {v_1, v_2, …, v_k}

$$PrivacyLoss_a(nc \mid R) = \sum_{i=1}^{k} -P_i \log_2(P_i) - \sum_{i=1}^{k} -P_i^* \log_2(P_i^*)$$

$$where\ P_i = Prob(a = v_i \mid R) \text{ and } P_i^* = Prob(a = v_i \mid R \cup nc)$$

  - P_i and P^*_i are probability mass function before and after disclosing *NC* given revealed credential set R.
- Multiple attributes
  - Attribute set {a_1, a_2 …,a_n} with sensitivity vector {w_1, w_2, …, w_n}

$$PrivacyLoss_A(nc \mid R) = \sum_{i=1}^{n} W_i \times PrivacyLoss_{a_i}(nc \mid R)$$

Estimation of query-dependent privacy loss

- Single query Q
  - Q is the function *f* of attribute set A
  - Domain of *f(A)* : {qv_1, qv_2, …, qv_k}

$$PrivacyLoss_q(nc \mid R) = \sum_{i=1}^{k} -P_i \log_2(P_i) - \sum_{i=1}^{k} -P_i^* \log_2(P_i^*)$$

$$where\ P_i = Prob(f(A) = qv_i \mid R) \text{ and } P_i^* = Prob(f(A) = qv_i \mid R \cup nc)$$

- Multiple queries
  - Query set {q_1, q_2 …,q_n} with sensitivity vector {w_1, w_2, …, w_n}
  - Pr_i is the probability that q_i is asked

$$PrivacyLoss_Q(nc \mid R) = \sum_{i=1}^{n} (PrivacyLoss_{q_i}(nc \mid R) \times Pr_i \times W_i)$$

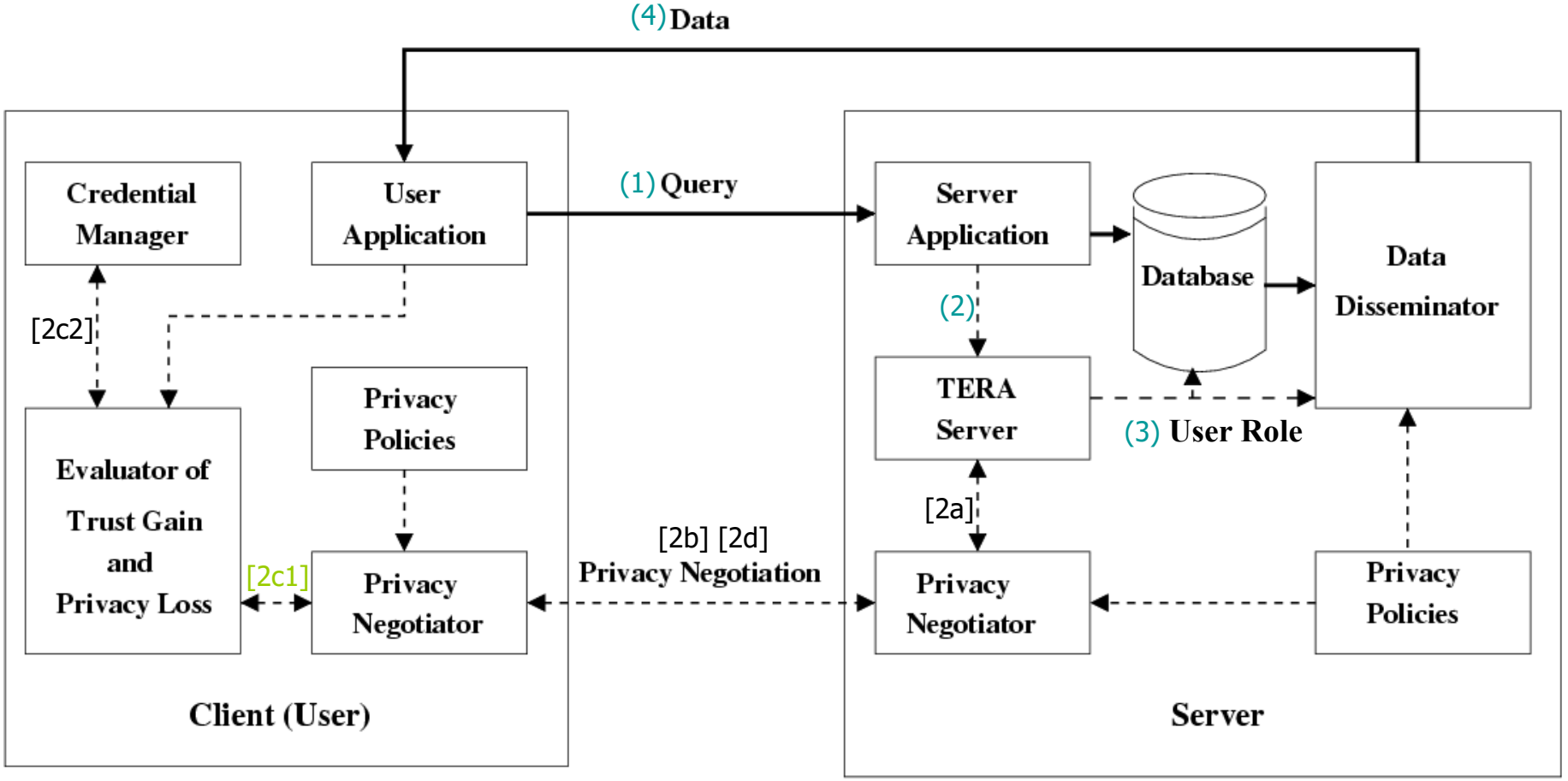
Estimate privacy damage

- Assume user provides one damage function $d_{usage}(PrivacyLoss)$ for each information usage
- PrivacyDamage(PrivacyLoss, Usage, Receiver) = $D_{max}(PrivacyLoss) \times (1-Trust_{receiver}) + d_{usage}(PrivacyLoss) \times Trust_{receiver}$
  - $Trust_{receiver}$ is a number Є [0,1] representing the trustworthy of information receiver
  - $D_{max}(PrivacyLoss) = Max(d_{usage}(PrivacyLoss)$ for all usage)

Estimate trust gain

- Increasing trust level
  - Adopt research on trust establishment and management
- Benefit function *TB(trust_level)*
  - Provided by service provider or derived from user's utility function
- Trust gain
  - $TB(trust\_level_{new}) - TB(tust\_level_{prev})$

PRETTY:
Prototype for Experimental Studies



(<nr>) – unconditional path

[<nr>]– conditional path

TERA = Trust-Enhanced Role Assignment

Information flow for PRETTY

1) User application sends query to server application.
2) Server application sends user information to TERA server for trust evaluation and role assignment.
   a) If a higher trust level is required for query, TERA server sends the request for more user's credentials to privacy negotiator.
   b) Based on server's privacy policies and the credential requirements, privacy negotiator interacts with user's privacy negotiator to build a higher level of trust.
   c) Trust gain and privacy loss evaluator selects credentials that will increase trust to the required level with the least privacy loss. Calculation considers credential requirements and credentials disclosed in previous interactions.
   d) According to privacy policies and calculated privacy loss, user's privacy negotiator decides whether or not to supply credentials to the server.
3) Once trust level meets the minimum requirements, appropriate roles are assigned to user for execution of his query.
4) Based on query results, user's trust level and privacy polices, data disseminator determines: (i) whether to distort data and if so to what degree, and (ii) what privacy enforcement metadata should be associated with it.

Conclusion

- This research addresses the tradeoff issues between privacy and trust.
- Tradeoff problems are formally defined.
- An entropy-based approach is proposed to estimate privacy loss.
- A prototype is under development for experimental study.

9. P2D2: A Mechanism for Privacy-Preserving Data Dissemination

P2D2 - Mechanism for Privacy-Preserving Data Dissemination

Outline

1) Introduction
   - 1.1) Interactions and Trust
   - 1.2) Building Trust
   - 1.3) Trading Weaker Partner's Privacy Loss for Stronger Partner's Trust Gain
   - 1.4) Privacy-Trust Tradeoff and Dissemination of Private Data
   - 1.5) Recognition of Need for Privacy Guarantees
2) Problem and Challenges
   - 2.1) The Problem
   - 2.2) Trust Model
   - 2.3) Challenges
3) Proposed Approach: Privacy-Preserving Data Dissemination (P2D2) Mechanism
   - 3.1) Self-descriptive Bundles
   - 3.2) Apoptosis of Bundles
   - 3.3) Context-sensitive Evaporation of Bundles
4) Prototype Implementation
5) Conclusions
6) Future Work

1) Introduction
1.1) Interactions and Trust

- Trust – new paradigm of security
  - Replaces/enhances CIA (confid./integr./availab.)
- Adequate degree of trust required in interactions
  - In social or computer-based interactions:
    - From a simple transaction to a complex collaboration
  - Must build up trust w.r.t. interaction partners
    - Human or artificial partners
    - Offline or online
- We focus on asymmetric trust relationships:
  One partner is “weaker,” another is “stronger”
  - Ignoring “same-strength” partners:
    - Individual to individual, most B2B,

1.2) Building Trust (1)
a) Building Trust By Weaker Partners

- Means of building trust by weaker partner in his stronger (often institutional) partner (offline and online):
  - Ask around
    - Family, friends, co-workers, …
  - Check partner's history and stated philosophy
    - Accomplishments, failures and associated recoveries, …
    - Mission, goals, policies (incl. privacy policies), …
  - Observe partner's behavior
    - Trustworthy or not, stable or not, …
    - Problem: Needs time for a fair judgment
  - Check reputation databases
    - Better Business Bureau, consumer advocacy groups, …
  - Verify partner's credentials
    - Certificates and awards, memberships in trust-building organizations (e.g., BBB), …
  - Protect yourself against partner's misbehavior
    - Trusted third-party, security deposit, prepayment,, buying insurance, …

1.2) Building Trust (2)
b) Building Trust by Stronger Partners

- Means of building trust by stronger partner in her weaker (often individual) partner (offline and online):
  - Business asks customer for a *payment* for goods or services
  - Bank asks for private information
  - Mortgage broker checks applicant's credit history
  - Authorization subsystem on a computer observes partner's behavior
    - Trustworthy or not, stable or not, …
    - Problem: Needs time for a fair judgment
  - Computerized trading system checks reputation databases
    - e-Bay, PayPal, …
  - Computer system verifies user's digital credentials
    - Passwords, magnetic and chip cards, biometrics, …
  - Business protects itself against customer's misbehavior
    - Trusted third-party, security deposit, prepayment,, buying insurance, …

1.3) Trading Weaker Partner's Privacy Loss for Stronger Partner's Trust Gain

- In all examples of Building Trust by Stronger Partners but the first (payments):

  Weaker partner trades his privacy loss for his trust gain as perceived by stronger partner

- Approach to trading privacy for trust:

  [Zhong and Bhargava, Purdue]

  - Formalize the privacy-trust tradeoff problem
  - Estimate *privacy loss* due to disclosing a credential set
  - Estimate *trust gain* due to disclosing a credential set
  - Develop algorithms that *minimize privacy loss* for required trust gain
    - Bec. nobody likes loosing more privacy than necessary

1.4) Privacy-Trust Tradeoff and Dissemination of Private Data

- Dissemination of private data
  - Related to trading privacy for trust:
    - Examples above
  - *Not* related to trading privacy for trust:
    - Medical records
    - Research data
    - Tax returns
    - …
- Private data dissemination can be:
  - Voluntary
    - When there's a sufficient competition for services or goods
  - Pseudo-voluntary
    - Free to decline… and loose service
      - E.g. a monopoly or demand exceeding supply)
  - Mandatory
    - Required by law, policies, bylaws, rules, etc.

Dissemination of Private Data is Critical

- Reasons:
  - Fears/threats of privacy violations reduce trust
  - Reduced trust leads to restrictions on interactions
    - In the extreme:
      refraining from interactions, even self-imposed isolation
    - Very high social costs of lost (offline and online) interaction opportunities
      - Lost business transactions, opportunities
      - Lost research collaborations
      - Lost social interactions
      - …

=> Without privacy guarantees, pervasive computing will never be realized

- People will avoid interactions with pervasive devices / systems
  - Fear of *opportunistic sensor networks* self-organized by electronic devices around them – can *help or harm* people in their midst

1.5) Recognition of Need for Privacy Guarantees (1)

[Ackerman *et al.* '99]

- By individuals
  - 99% unwilling to reveal their SSN
  - 18% unwilling to reveal their… favorite TV show
- By businesses
  - Online consumers worrying about revealing personal data

    held back $15 billion in online revenue in 2001
- By Federal government
  - Privacy Act of 1974 for Federal agencies
  - Health Insurance Portability and Accountability Act of 1996 (HIPAA)

1.5) Recognition of Need for Privacy Guarantees (2)

- By computer industry research
  - Microsoft Research
    - The biggest research challenges:

      According to Dr. Rick Rashid, Senior Vice President for Research
      - Reliability / Security / Privacy / Business Integrity
        - » Broader: application integrity (just “integrity?”)

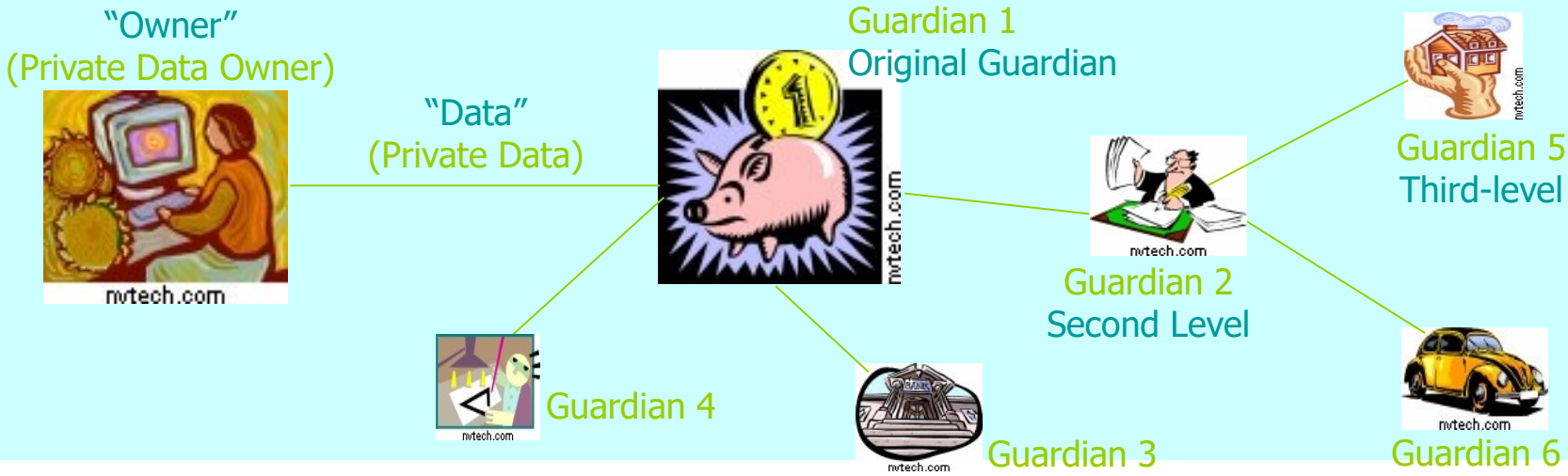
      => MS Trustworthy Computing Initiative

    - Topics include: DRM—digital rights management (incl. watermarking surviving photo editing attacks), software rights protection, intellectual property and content protection, database privacy and p.-p. data mining, anonymous e-cash, anti-spyware
  - IBM (incl. Privacy Research Institute)
    - Topics include: pseudonymity for e-commerce, EPA and EPAL—enterprise privacy architecture and language, RFID privacy, p.-p. video surveillance, federated identity management (for enterprise federations), p.-p. data mining and p.-p.mining of association rules, Hippocratic (p.-p.) databases, online privacy monitoring

1.5) Recognition of Need for Privacy Guarantees (3)

- By academic researchers
  - CMU and Privacy Technology Center
    - Latanya Sweeney (k-anonymity, SOS—Surveillance of Surveillances, genomic privacy)
    - Mike Reiter (Crowds – anonymity)
  - Purdue University – CS and CERIAS
    - Elisa Bertino (trust negotiation languages and privacy)
    - Bharat Bhargava (privacy-trust tradeoff, privacy metrics, p.-p. data dissemination, p.-p. location-based routing and services in networks)
    - Chris Clifton (p.-p. data mining)
  - UIUC
    - Roy Campbell (Mist – preserving location privacy in pervasive computing)
    - Marianne Winslett (trust negotiation w/ controled release of private credentials)
  - U. of North Carolina Charlotte
    - Xintao Wu, Yongge Wang, Yuliang Zheng (p.-p. database testing and data mining)

2) Problem and Challenges
2.1) The Problem (1)



- "Guardian:"
  Entity entrusted by private data owners with collection, processing, storage, or transfer of their data
  - owner can be an institution or a system
  - owner can be a guardian for her own private data
- Guardians allowed or required to share/disseminate private data
  - With owner's explicit consent
  - Without the consent as required by law
    - For research, by a court order, etc.

2.1) The Problem (2)

- Guardian passes private data to another guardian in a data dissemination chain
  - Chain within a graph (possibly cyclic)
- Sometimes owner privacy preferences *not* transmitted due to neglect or failure
  - Risk grows with chain length and milieu fallibility and hostility
- If preferences lost, even honest receiving guardian unable to honor them

2.2) Trust Model

- Owner builds trust in Primary Guardian (PG)
  - As shown in Building Trust by Weaker Partners
- Trusting PG means:
  - Trusting the integrity of PG data sharing policies and practices
  - Transitive trust in data-sharing partners of PG
    - PG provides owner with a list of partners for private data dissemination (incl. info which data PG plans to share, with which partner, and why)

    OR:

    - PG requests owner's permission before any private data dissemination (request must incl. the same info as required for the list)

    OR:

    - A hybrid of the above two

      E.g., PG provides list for next-level partners AND each second- and lower-level guardian requests owner's permission before any further private data dissemination

2.3) Challenges

- Ensuring that owner's metadata are never decoupled from his data
  - Metadata include owner's privacy preferences
- Efficient protection in a hostile milieu
  - Threats - examples
    - Uncontrolled data dissemination
    - Intentional or accidental data corruption, substitution, or disclosure
  - Detection of data or metadata loss
  - Efficient data and metadata recovery
    - Recovery by retransmission from the original guardian is most trustworthy

3) Proposed Approach: Privacy-Preserving Data Dissemination (P2D2) Mechanism

3.1) Design self-descriptive *bundles*

- bundle = private data + metadata
- self-descriptive bec. includes metadata

3.2) Construct a mechanism for *apoptosis* of bundles

- apoptosis = clean self-destruction

3.3) Develop context-sensitive *evaporation* of bundles

Related Work

- Self-descriptiveness (in diverse contexts)
  - Meta data model [Bowers and Delcambre, '03]
  - KIF — Knowledge Interchange Format [Gensereth and Fikes, '92]
  - Context-aware mobile infrastructure [Rakotonirainy, '99]
  - Flexible data types [Spreitzer and A. Begel, '99]
- Use of self-descriptiveness for data privacy
  - Idea mentioned in one sentence [Rezgui, Bouguettaya and Eltoweissy, '03]
- Term: apoptosis (clean self-destruction)
  - Using apoptosis to end life of a distributed services (esp. in 'strongly' active networks, where each data packet is replaced by a mobile program) [Tschudin, '99]
- Specification of privacy preferences and policies
  - Platform for Privacy Preferences [Cranor, '03]
  - AT&T Privacy Bird [AT&T, '04]

Bibliography for Related Work

AT&T Privacy Bird Tour: http://privacybird.com/tour/1 2 beta/tour.html. February 2004.

S. Bowers and L. Delcambre. The uni-level description: A uniform framework for representing information in multiple data models. *ER 2003-Intl. Conf. on Conceptual Modeling, I.-Y. Song, et al. (Eds.)*, pp. 45–58, Chicago, Oct. 2003.

L. Cranor. P3P: Making privacy policies more useful. *IEEE Security and Privacy*, pp. 50–55, Nov./Dec. 2003.

M. Gensereth and R. Fikes. Knowledge Interchange Format. Tech. Rep. Logic-92-1, Stanford Univ., 1992.

A. Rakotonirainy. Trends and future of mobile computing. *10th Intl. Workshop on Database and Expert Systems Applications*, Florence, Italy, Sept. 1999.

A. Rezgui, A. Bouguettaya, and M. Eltoweissy. Privacy on the Web: Facts, challenges, and solutions. *IEEE Security and Privacy*, pp. 40–49, Nov./Dec. 2003.

M. Spreitzer and A. Begel. More flexible data types. *Proc. IEEE 8th Workshop on Enabling Technologies (WETICE '99)*, pp. 319–324, Stanford, CA, June 1999.

C. Tschudin. Apoptosis - the programmed death of distributed services. In: J. Vitek and C. Jensen, eds., *Secure Internet Programming*. Springer-Verlag, 1999.

3.1) Self-descriptive Bundles

- Comprehensive metadata include:
  - owner's privacy preferences — How to read and write private data
  - owner's contact information — Needed to request owner's access permissions, or notify the owner of any accesses
  - guardian's privacy policies — For the original and/or subsequent data guardians
  - metadata access conditions — How to verify and modify metadata
  - enforcement specifications — How to enforce preferences and policies
  - data provenance — Who created, read, modified, or destroyed any portion of data
  - context-dependent and other components — Application-dependent elements; Customer trust levels for different contexts; Other metadata elements

Implementation Issues for Bundles

- Provide efficient and effective representation for bundles
  - Use XML – work in progress
- Ensure bundle atomicity
  — metadata can't be split from data
  - A simple atomicity solution using asymmetric encryption
    - Destination Guardian (DG) provides public key
    - Source Guardian (or owner) encrypts bundle with public key
      - Can re-bundle by encrypting different bundle elements with public keys from different DGs
    - DG applies its corresponding private key to decrypt received bundle
      - Or: decrypts just bundle elements — reveals data DG "needs to know"
  - Can use digital signature to assure non-repudiation
      - Extra key mgmt effort: requires Source Guardian to provide public key to DG
- Deal with insiders making and disseminating illegal copies of data they are authorized to access (but not copy)

  Considered below (taxonomy)

Notification in Bundles (1)

- Bundles simplify notifying owners or requesting their consent
  - Contact information in the *owner's contact information*
  - Included information
    - *notification* = [notif_sender, sender_t-stamp, accessor, access_t-stamp, access_justification, other_info]
    - *request* = [req_sender, sender_t-stamp, requestor, requestor_t-stamp, access_justification, other_info]
- Notifications / requests sent to owners *immediately*, *periodically*, or *on demand*
  - Via:
    - automatic pagers / text messaging (SMS) / email messages
    - automatic cellphone calls / stationary phone calls
    - mail
  - ACK from owner may be required for notifications
  - Messages may be encrypted or digitally signed for security

Notification in Bundles (2)

- If permission for a *request* or *request_type* is:
  - Granted in metadata
    => notify owner
  - Not granted in metadata
    => ask for owner's permission to access her data
- For very sensitive data — no default permissions for requestors are granted
  - Each request needs owner's permission

Optimization of Bundle Transmission

- Transmitting *complete* bundles between guardians is inefficient
  - They describe all foreseeable aspects of data privacy
    - For any application and environment
- Solution: prune transmitted bundles
  - Adaptively include only needed data and metadata
    - Maybe, needed “transitively” — for the whole down stream
  - Use short codes (standards needed)
  - Use application and environment semantics along the data dissemination chain

3.2) Apoptosis of Bundles

- Assuring privacy in data dissemination
  - Bundle apoptosis vs. private data apoptosis

    Bundle apoptosis is preferable – prevents inferences from metadata
  - In benevolent settings:

    use *atomic* bundles with recovery by retransmission
  - In malevolent settings:

    attacked bundle, threatened with disclosure, performs apoptosis

Implementation of Apoptosis

- Implementation
  - Detectors, triggers and code
    - Detectors – e.g. integrity assertions identifying potential attacks
      - E.g., recognize critical system and application events
  - Different kinds of detectors
    - Compare how well different detectors work
    - False positives
      - Result in superfluous bundle apoptosis
      - Recovery by bundle retransmission
      - Prevent DoS (Denial-of-service) attacks by limiting repetitions
    - False negatives
      - May result in disclosure – very high costs (monetary, goodwill loss, etc.)

Optimization of Apoptosis Implementation

- Consider alternative detection, trigerring and code implementations
- Determine division of labor between detectors, triggers and code
  - Code must include recovery from false positives
- Define measures for evaluation of apoptosis implementations
  - Effectiveness: false positives rate and false negatives rate
  - Costs of false positives (recovery) and false negatives (disclosures)
  - Efficiency: speed of apoptosis, speed of recovery
  - Robustness (against failures and attacks)
- Analyze detectors, triggers and code
- Select a few candidate implementation techniques for detectors, triggers and code
- Evaluation of candidate techniques vis simulate experiments
- Prototyping and experimentation in our testbed for investigating trading privacy for trust

3.3) Context-sensitive Evaporation of Bundles

- Perfect data dissemination not always desirable
  - Example: Confidential business data shared within an office but *not outside*
- Idea:

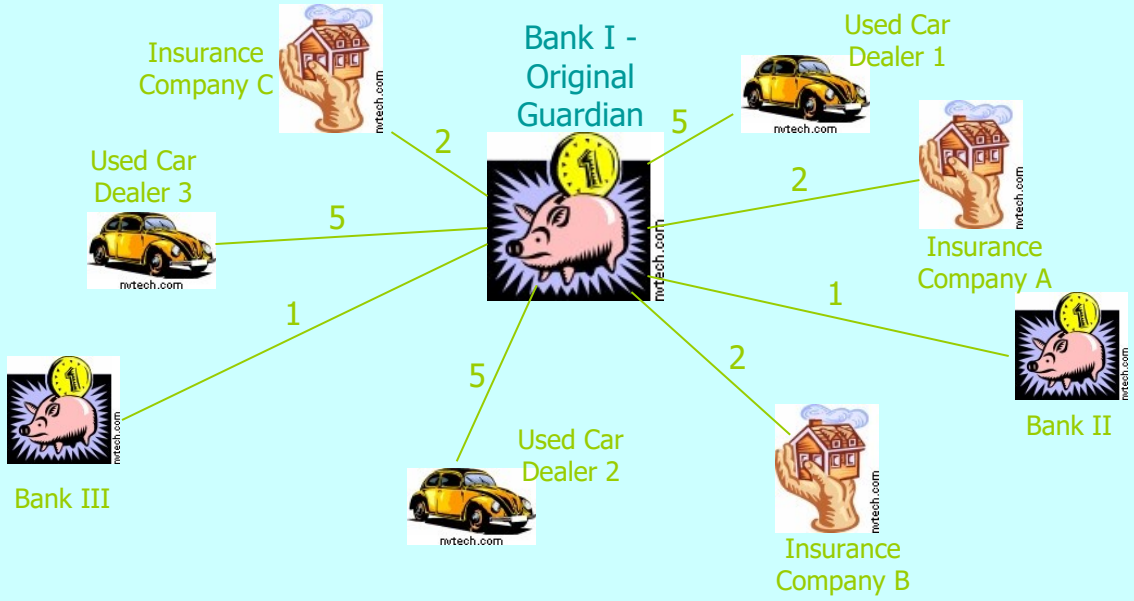
  Context-sensitive bundle *evaporation*

Proximity-based Evaporation of Bundles

- Simple case: Bundles *evaporate* in proportion to their "distance" from their owner
  - Bundle evaporation prevents inferences from metadata
  - "Closer" guardians trusted more than "distant" ones
  - Illegitimate disclosures more probable at less trusted "distant" guardians
  - Different distance metrics
    - Context-dependent

Examples of Distance Metrics

- Examples of one-dimensional distance metrics
  - Distance ~ business type



If a bank is the original guardian, then:
-- any other *bank* is "closer" than any *insurance company*
-- any *insurance company* is "closer" than any *used car dealer*

  - Distance ~ distrust level: more trusted entities are "closer"

- Multi-dimensional distance metrics
  - Security/reliability as one of dimensions

Evaporation Implemented as Controlled Data Distortion

- Distorted data reveal less, protects privacy
- Examples:

accurate data		more and more distorted data		
250 N. Salisbury Street West Lafayette, IN	→	Salisbury Street West Lafayette, IN	→	somewhere in West Lafayette, IN
250 N. Salisbury Street West Lafayette, IN [home address]	→	250 N. University Street West Lafayette, IN [office address]	→	P.O. Box 1234 West Lafayette, IN [P.O. box]
765-123-4567 [home phone]	→	765-987-6543 [office phone]	→	765-987-4321 [office fax]

Evaporation Implemented as Controlled Data Distortion

- Distorted data reveal less, protects privacy
- Examples:

accurate data	more and more distorted data	
250 N. Salisbury Street West Lafayette, IN	Salisbury Street West Lafayette, IN	somewhere in West Lafayette, IN
250 N. Salisbury Street West Lafayette, IN [home address]	250 N. University Street West Lafayette, IN [office address]	P.O. Box 1234 West Lafayette, IN [P.O. box]
765-123-4567 [home phone]	765-987-6543 [office phone]	765-987-4321 [office fax]

Evaporation as Generalization of Apoptosis

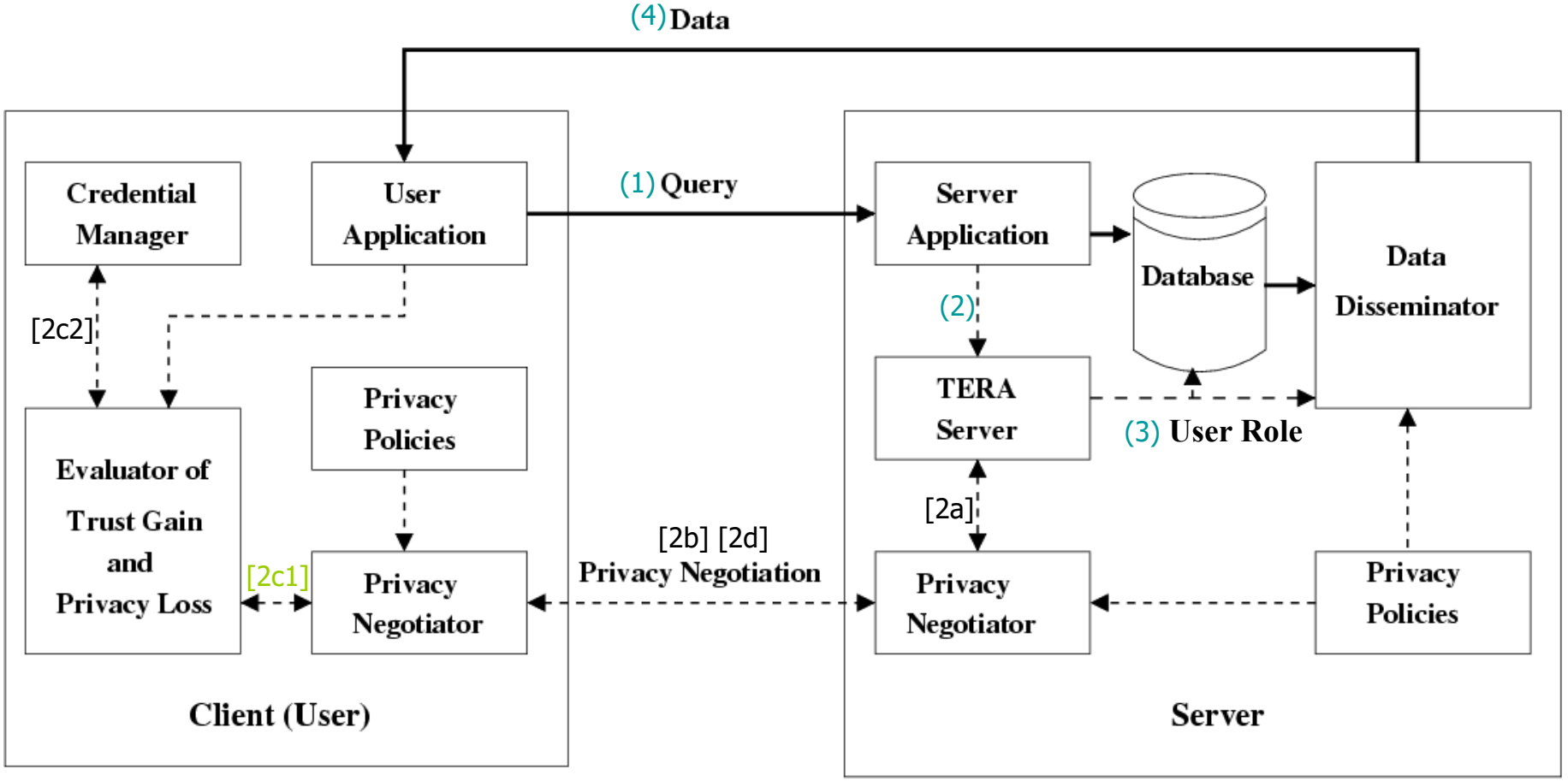
- Context-dependent apoptosis for implementing evaporation
  - Apoptosis detectors, triggers, and code enable context exploitation
- Conventional apoptosis as a simple case of data evaporation
  - Evaporation follows a step function
    - Bundle self-destructs when proximity metric exceeds predefined threshold value

Application of Evaporation for DRM

- Evaporation could be used for “active” DRM (digital rights management)
  - Bundles with protected contents evaporate when copied onto ”foreign” media or storage device

4) Prototype Implementation

- Our experimental system named PRETTY (PRivatE and TrusTed sYstems)
  - Trust mechanisms already implemented



(<nr>) – unconditional path

[<nr>]– conditional path

TERA = Trust-Enhanced Role Assignment

Information Flow in PRETTY

1) User application sends query to server application.
2) Server application sends user information to TERA server for trust evaluation and role assignment.
   a) If a higher trust level is required for query, TERA server sends the request for more user's credentials to privacy negotiator.
   b) Based on server's privacy policies and the credential requirements, privacy negotiator interacts with user's privacy negotiator to build a higher level of trust.
   c) Trust gain and privacy loss evaluator selects credentials that will increase trust to the required level with the least privacy loss. Calculation considers credential requirements and credentials disclosed in previous interactions.
   d) According to privacy policies and calculated privacy loss, user's privacy negotiator decides whether or not to supply credentials to the server.
3) Once trust level meets the minimum requirements, appropriate roles are assigned to user for execution of his query.
4) Based on query results, user's trust level and privacy polices, data disseminator determines: (i) whether to distort data and if so to what degree, and (ii) what privacy enforcement metadata should be associated with it.

5) Conclusions

- Intellectual merit
  - A mechanism for preserving privacy in data dissemination (bundling, apoptosis, evaporation)
- Broader impact
  - Educational and research impact: student projects, faculty collaborations
  - Practical (social, economic, legal, etc.) impact:
    - Enabling more collaborations
    - Enabling “more pervasive” computing
      - By reducing fears of privacy invasions
    - Showing new venues for privacy research
  - Applications
    - Collaboration in medical practice, business, research, military…
    - Location-based services
  - Future impact:
    - Potential for extensions enabling “pervasive computing”
      - Must adapt to privacy preservation, e.g., in *opportunistic* sensor networks (self-organize to help/harm)

6) Future Work

- Provide efficient and effective representation for bundles (XML for metadata?)
- Run experiments on the PRETTY system
  - Build a complete prototype of proposed mechanism for private data dissemination
    - Implement
    - Examine implementation impacts:
      - Measures: Cost, efficiency, trustworthiness, other
    - Optimize bundling, apoptosis and evaporation techniques
- Focus on selected application areas
  - Sensor networks for infrastructure monitoring (NSF IGERT proposal)
  - Healthcare enginering (work for RCHE - Regenstrief Center for Healthcare Engineering at Purdue)

Future Work - Extensions

- Adopting proposed mechanism for DRM, IRM (intellectual rights managenment) and proprietary/confidential data
  - Privacy:
    Private data – owned by an individual
  - Intellectual property, trade/diplomatic/military secrets:
    Proprietary/confidential data – owned by an organization
- Custimizing proposed mechanismm for selected pervasive environments, including:
  - Wireless / Mobile / Sensor networks
    - Incl. *opportunistic* sens. networks
- Impact of proposed mechanism on data quality
- L.Lilien and B. Bhargava, A scheme for Privacy Preserving Data Dissemination, IEEE SMC, May 2006, 502-506

10. Position-based Private Routing in Ad Hoc Networks

- Problem statement
  - To hide the identities of the nodes who are involved in routing in mobile wireless ad hoc networks.
- Challenges
  - Traditional ad hoc routing algorithms depend on private information (e.g., ID) exposure in the network.
  - Privacy solutions for P2P networks are not suitable in ad hoc networks.

Weak Privacy for Traditional Position-based Ad Hoc Routing Algorithm

- Position information of each node has to be locally broadcast *periodically*.
- Adversaries are able to obtain node trajectory based on the position report.
- Adversaries can estimate network topology.
- Once a match between a node position and its real ID is found, a tracer can always stay close to this node and monitor its behavior.

AO2P: *A*d Hoc *O*n-Demand *P*osition-based *P*rivate Routing

- Position of destination is the information exposed in the network for routing discovery.
- A receiver-contention scheme is designed to determine the next hop in a route.
- Pseudo IDs are used instead of real IDs for data packet delivery after a route is built up.
- Route with a smaller number of hops will be used for better end-to-end throughput.

AO2P Routing Privacy and Accuracy

- Only the position of destination is revealed in the network for routing discovery. The privacy of the destination relies on the difficulty of matching a position to a node ID.
- Node mobility enhances destination privacy because a match between a position to a node ID is temporary.
- The privacy for the source and the intermediate forwarders is well preserved.
- Routing accuracy relies on the fact that at a specific time, only one node can be at a position. Since the pseudo ID for a node is generated from its position and the time it is at that position, the probability that more than one node have the same pseudo ID is negligible.

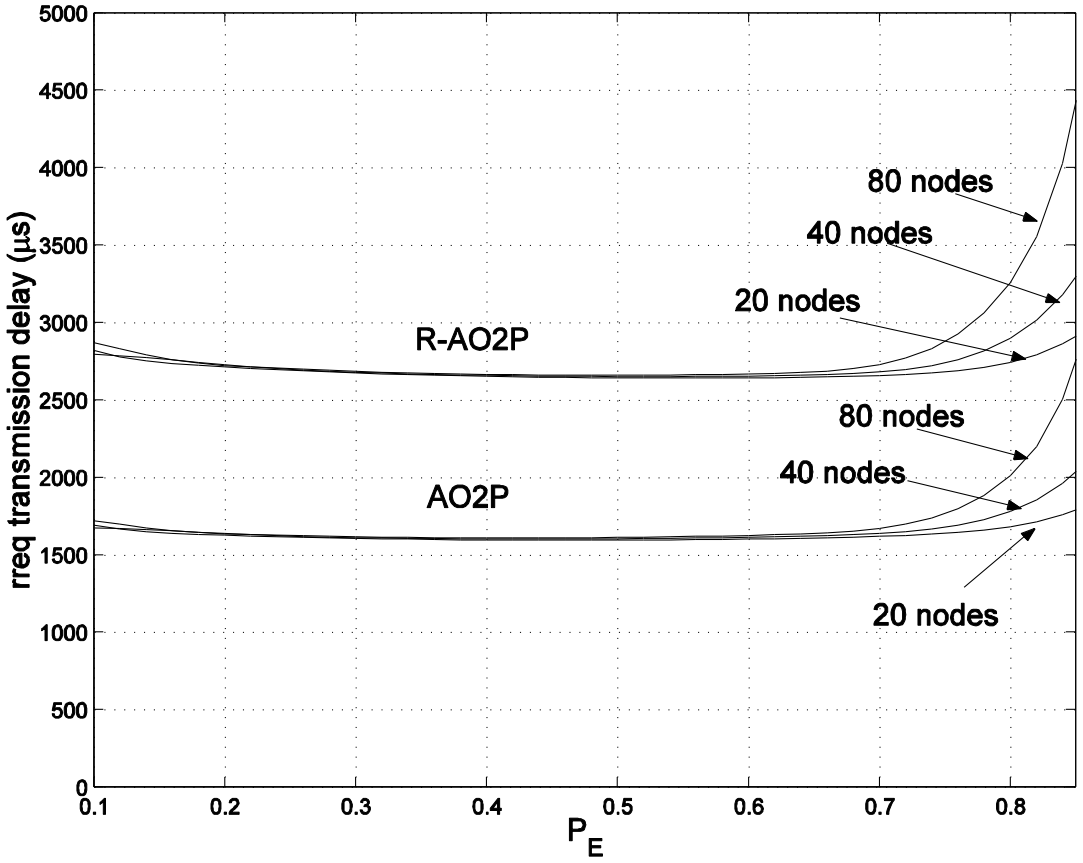
Privacy Enhancement: R-AO2P

- The position of reference point is carried in *rreq* instead of the position of the destination.
- The reference point is on the extended line from the sender to the destination. It can be used for routing discovery because generally, a node that processes the *rreq* closer to the reference point will also process the *rreq* closer to the destination.
- The position of the destination is only disclosed to the nodes who are involved in routing.

Reference point in R-AO2P

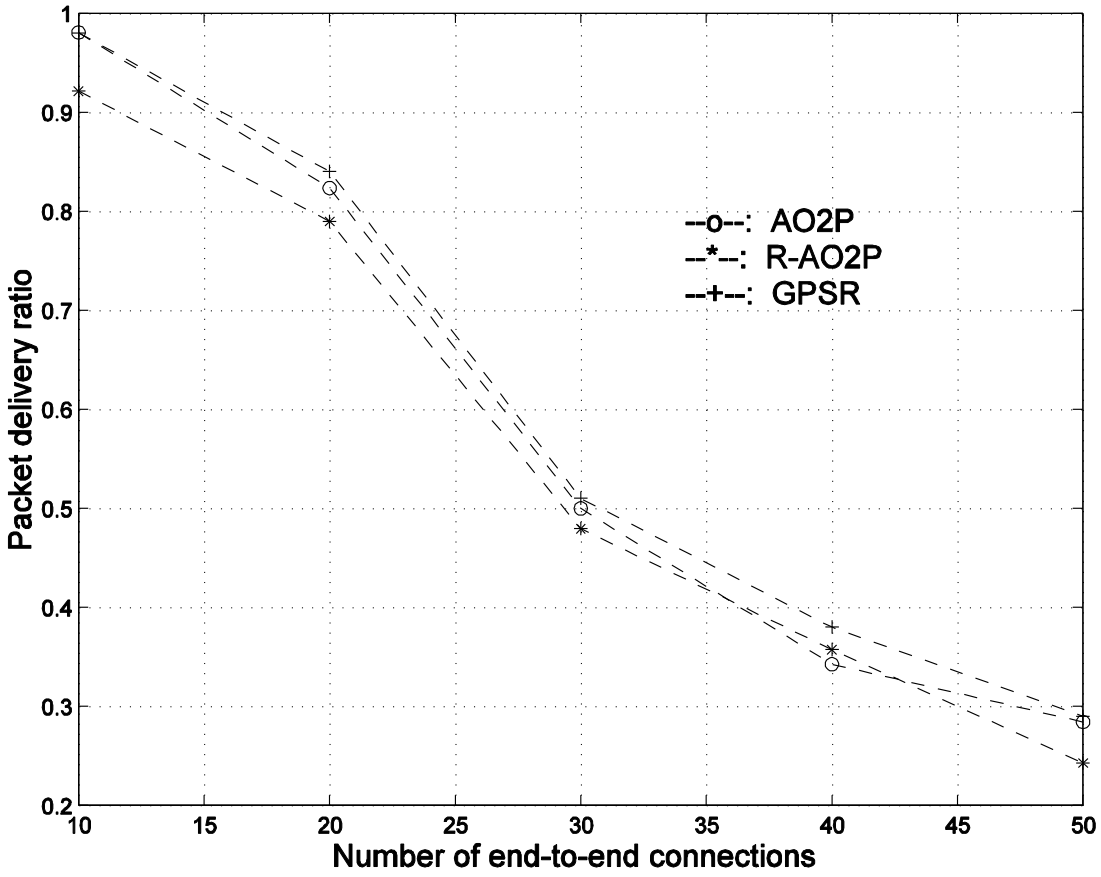
Illustrated Results

- Average delay for next hop determination



Illustrated Results

- Packet delivery ratio



Conclusions

- AO2P preserves node privacy in mobile ad hoc networks.
- AO2P has low next hop determination delay.
- Compared to other position-based ad hoc routing algorithm, AO2P has little routing performance degradation.
- X.Wu and B.Bhargava, AO2P, IEEE TMC, Vol. 4, No.4, 2006 pp 325-348.

7. Trust-based Privacy Preservation for Peer-to-Peer Data Sharing

Problem statement

- Privacy in peer-to-peer systems is different from the anonymity problem
- Preserve privacy of requester
- A mechanism is needed to remove the association between the identity of the requester and the data needed

Proposed solution

- A mechanism is proposed that allows the peers to acquire data through trusted proxies to preserve privacy of requester
  - The data request is handled through the peer's proxies
  - The proxy can become a supplier later and mask the original requester

Related work (1)

- Trust in privacy preservation
  - Authorization based on evidence and trust, [Bhargava and Zhong, DaWaK'02]
  - Developing pervasive trust [Lilien, CGW'03]
- Hiding the subject in a crowd
  - K-anonymity [Sweeney, UFKS'02]
  - Broadcast and multicast [Scarlata *et al*, INCP'01]

Related work (2)

- Fixed servers and proxies
  - Publius [Waldman *et al*, USENIX'00]
- Building a multi-hop path to hide the real source and destination
  - FreeNet [Clarke *et al*, IC'02]
  - Crowds [Reiter and Rubin, ACM TISS'98]
  - Onion routing [Goldschlag *et al*, ACM Commu.'99]

Related work (3)

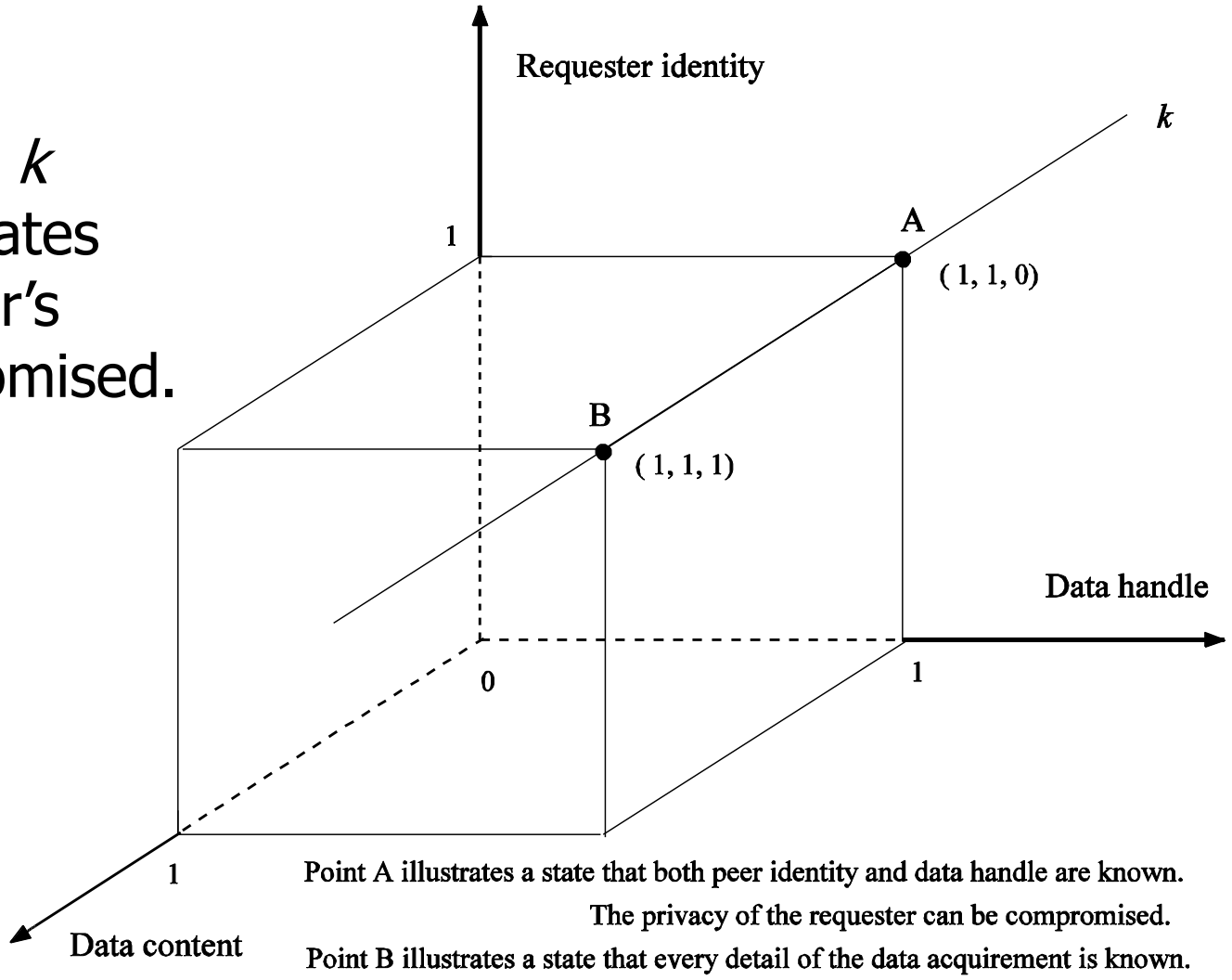
- p^5 [Sherwood *et al*, IEEE SSP'02]
  - p^5 provides sender-receiver anonymity by transmitting packets to a broadcast group
- Herbivore [Goel *et al*, Cornell Univ Tech Report'03]
  - Provides provable anonymity in peer-to-peer communication systems by adopting dining cryptographer networks

Privacy measurement (1)

- A tuple <requester ID, data handle, data content> is defined to describe a data acquirement.
- For each element, “0” means that the peer knows nothing, while “1” means that it knows everything.
- A state in which the requester’s privacy is compromised can be represented as a vector <1, 1, y>, (y Є [0,1]) from which one can link the ID of the requester to the data that it is interested in.

Privacy measurement (2)

For example, line *k* represents the states that the requester's privacy is compromised.



Mitigating collusion

- An operation "*" is defined as:

$$< c_1, c_2, c_3 >=< a_1, a_2, a_3 > * < b_1, b_2, b_3 >$$

$$c_i = \begin{cases} \max(a_i, b_i), & a_i \neq 0 \quad and \quad b_i \neq 0; \\ 0, & otherwise. \end{cases}$$

- This operation describes the revealed information after a collusion of two peers when each peer knows a part of the "secret".
- The number of collusions required to compromise the secret can be used to evaluate the achieved privacy

Trust based privacy preservation scheme

- The requester asks one proxy to look up the data on its behalf. Once the supplier is located, the proxy will get the data and deliver it to the requester
  - Advantage: other peers, including the supplier, do not know the real requester
  - Disadvantage: The privacy solely depends on the trustworthiness and reliability of the proxy

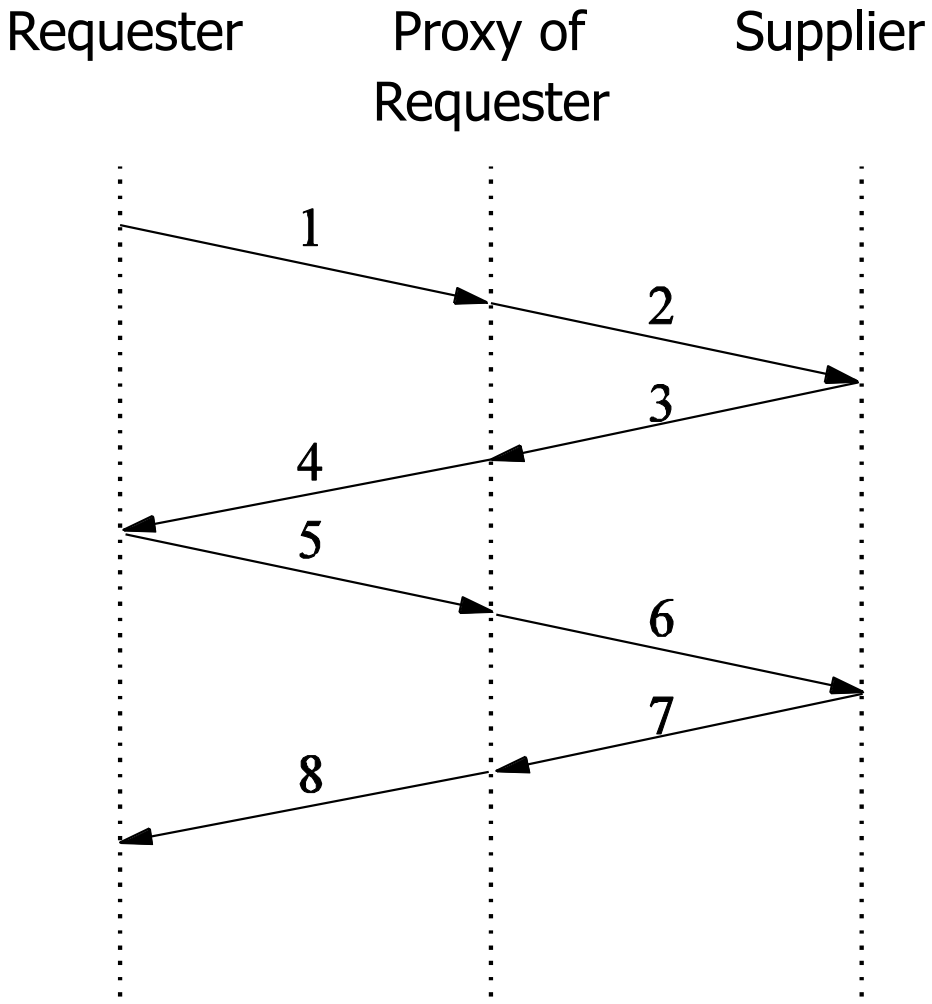
Trust based scheme – Improvement 1

- To avoid specifying the data handle in plain text, the requester calculates the hash code and only reveals a part of it to the proxy.
- The proxy sends it to possible suppliers.
- Receiving the partial hash code, the supplier compares it to the hash codes of the data handles that it holds. Depending on the revealed part, multiple matches may be found.
- The suppliers then construct a bloom filter based on the remaining parts of the matched hash codes and send it back. They also send back their public key certificates.

Trust based scheme – Improvement 1 (cont)

- Examining the filters, the requester can eliminate some candidate suppliers and finds some who may have the data.
- It then encrypts the full data handle and a data transfer key k_{Data} with the public key.
- The supplier sends the data back using k_{Data} through the proxy
- Advantages:
  - It is difficult to infer the data handle through the partial hash code
  - The proxy alone cannot compromise the privacy
  - Through adjusting the revealed hash code, the allowable error of the bloom filter can be determined

Data transfer procedure after improvement 1



R: requester *S*: supplier

Step 1, 2: *R* sends out the partial hash code of the data handle

Step 3, 4: *S* sends the bloom filter of the handles and the public key certificates

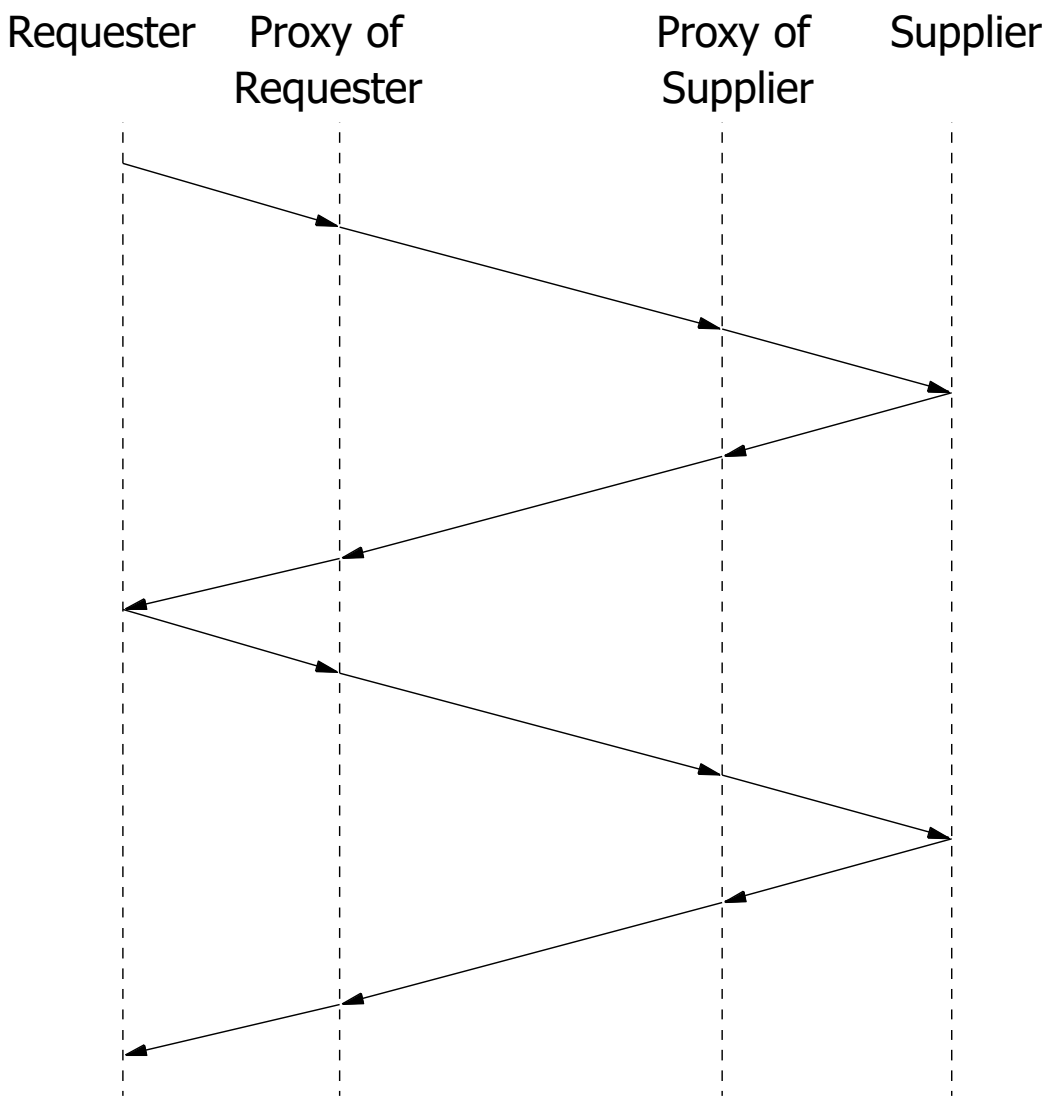
Step 5, 6: *R* sends the data handle and k_{Data} encrypted by the public key

Step 7, 8: *S* sends the required data encrypted by k_{Data}

Trust based scheme – Improvement 2

- The above scheme does not protect the privacy of the supplier
- To address this problem, the supplier can respond to a request via its own proxy

Trust based scheme – Improvement 2



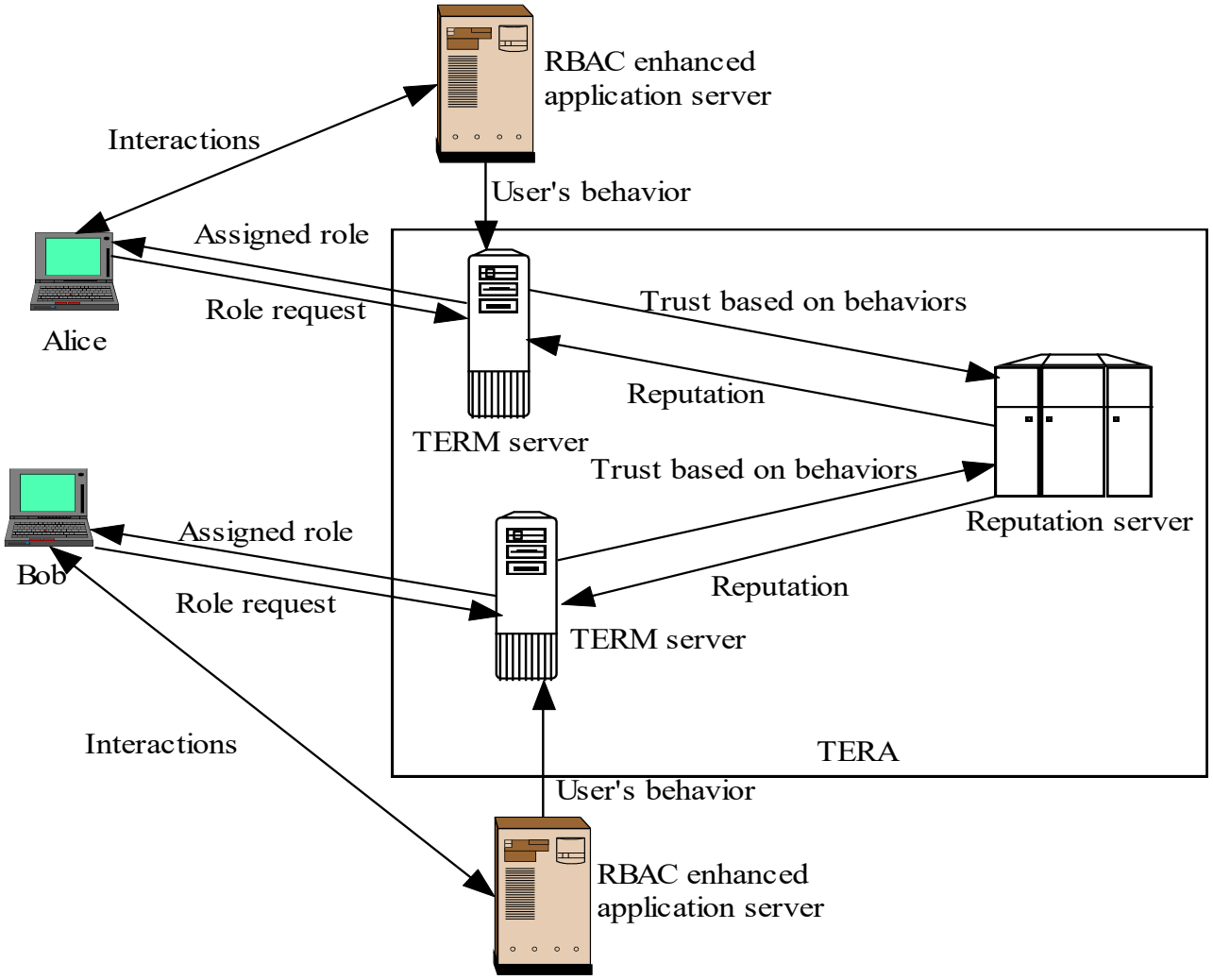
Trustworthiness of peers

- The trust value of a proxy is assessed based on its behaviors and other peers' recommendations
- Using Kalman filtering, the trust model can be built as a multivariate, time-varying state vector

Experimental platform - TERA

- Trust enhanced role mapping (TERM) server assigns roles to users based on
  - Uncertain & subjective evidences
  - Dynamic trust
- Reputation server
  - Dynamic trust information repository
  - Evaluate reputation from trust information by using algorithms specified by TERM server

Trust enhanced role assignment architecture (TERA)



Conclusion

- A trust based privacy preservation method for peer-to-peer data sharing is proposed
- It adopts the proxy scheme during the data acquirement
- Extensions
  - Solid analysis and experiments on large scale networks are required
  - A security analysis of the proposed mechanism is required

- More information may be found at
  http://raidlab.cs.purdue.edu

- Our papers and tech reports

W. Wang, Y. Lu, B. Bhargava, *On vulnerability and protection of AODV*, CERIAS Tech Report TR-02-18.

W. Wang, Y. Lu, B. Bharagav, “On vulnerability and protection of AODV”, in proceedings of ICT 2003.

W. Wang, Y. Lu, B. Bhargava, “On security study of two distance vector routing protocols for two mobile ad hoc networks”, in proceedings of PerCOm 2003.

Y. Lu, W.Wang, D. Xu, B. Bhargava Trust-based Privacy Preservation for p2p data sharing, IEEE SMC, May 2006,498-502