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

Bharat Bhargava
Department of Computer Sciences
Purdue University

With contributions from Prof. Leszek Lilien and Dr. Yuhui Zhong

Supported in part by NSF grants IIS-0209059 and IIS-0242840.

Outline

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 strongeer (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, …

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)

- By individuals [Ackerman *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)

1.5) Recognition of Need for Privacy Guarantees (2)

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

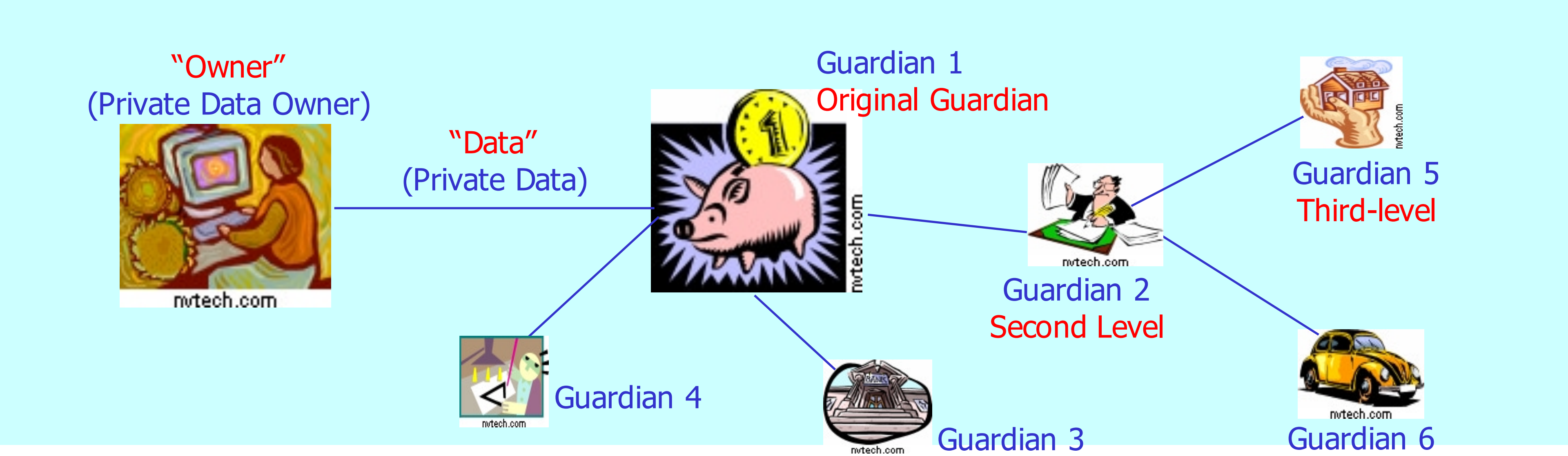
      According to Dr. Rick Rashid, Senior Vice President for Research
      - Reliability / Security / Privacy / Business Integrity
        - Broader: application integrity (just "integrity?")

      => 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:

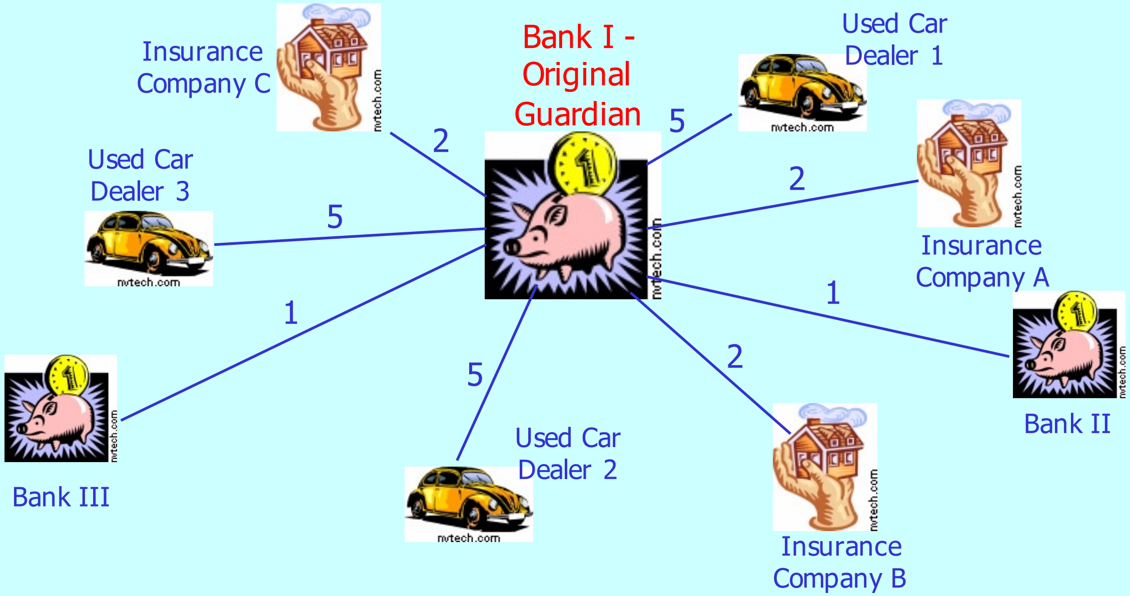
Context-sensitive bundle *evaporation*

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

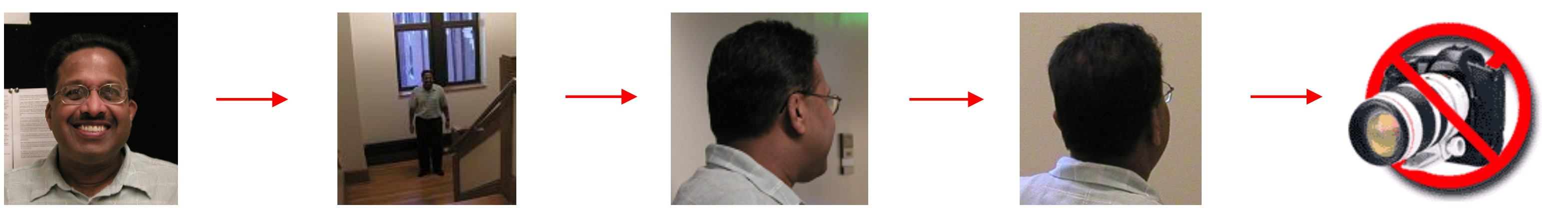
- Distorted data reveal less, protects privacy
- Examples:

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

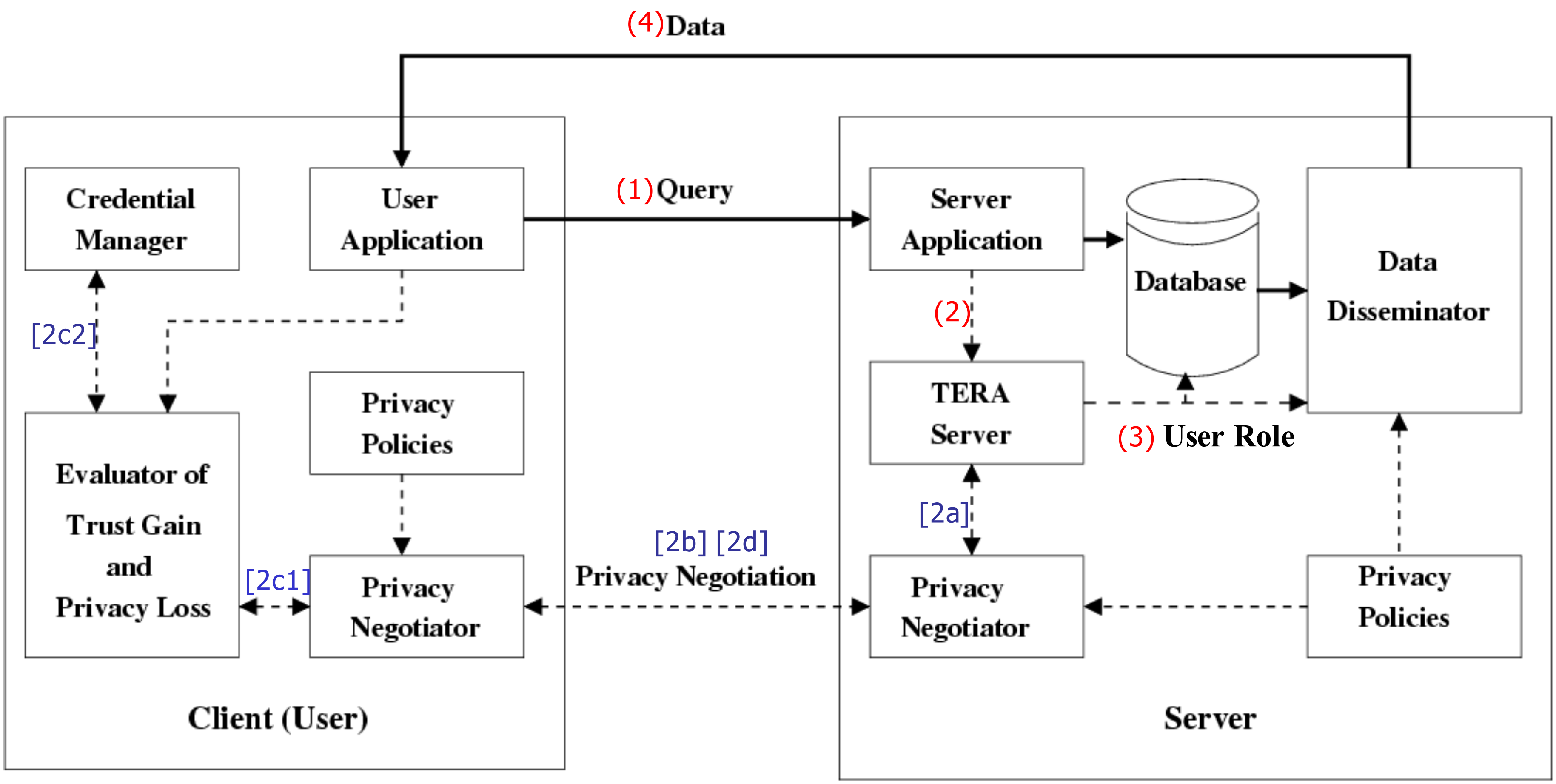
- Context-dependent apoptosis for implementing evaporation
  - Apoptosis detectors, triggers, and code enable context exploitation
- Conventional apoptosis as a simple case of data evaporation
  - Evaporation follows a step function
    - Bundle self-destructs when proximity metric exceeds predefined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