Privacy-Enhancing Models and Mechanisms for Securing Provenance and its Use

October 2010

Lead PI: Ravi Sandhu (UT San Antonio)
PIs: Elisa Bertino (Purdue), Gabriel Ghinita (Purdue), Murat Kantarcioglu (UT Dallas), Bhavani Thuraisingham (UT Dallas), Shouhuai Xu (UT San Antonio)
Provenance Helps Enhance Security

Access and Usage Control of Data and its Provenance

Data Trustworthiness (e.g., sensor networks)

Data Forensics (e.g., SIM tools)

Data Privacy (e.g., track hospital records)

etcetera
But Provenance Itself Must Be Secured!

- Security Requirements for Provenance
  - Access and Usage Control
    - Only authorized users may access provenance, and data based on provenance for appropriate purpose
  - Privacy
    - Provenance disclosed at a level that preserves data and source privacy
  - Integrity
    - Ensure that provenance is authentic and not tampered with
  - Accountability
    - Subjects accountable for data changes, even when they are anonymous

- Contributions
  1. Privacy-enhancing framework (i.e., models and mechanisms) for securing provenance lifecycle
  2. Design and implementation of mechanisms for secure provenance management at OS layer and Data Layer
Provenance Life Cycle

- generation
- processing
- dissemination
- accountability
- compliance
Proposed Secure Provenance Core Layer

Application Layer

<table>
<thead>
<tr>
<th>Access/Usage Control</th>
<th>Privacy</th>
<th>Integrity</th>
<th>Accountability</th>
</tr>
</thead>
</table>

Proposed Core of Secure Provenance Systems

Logical Structure and Physical Storage of Provenance
Representation of Provenance

- Directed Acyclic Graph (DAG)
  - Nodes represent entities (sources) that forward/modify data
  - Node labels capture type of operation performed
    - E.g., concatenation, set difference, etc.
- Edges capture data flow
  - Logical and physical provenance
    - Logical provenance captures actual changes in data
    - Physical provenance models “forwarding only” cases
  - People and system/machine provenance
    - People attributes changes in data to a person or organization
    - System/machine tracks provenance wrt software/hardware
Provenance DAG Example

Provenance DAG Example

No_Op

A

B

C

D

E

Concatenate(A,B)

Set_Diff(D,E)

F

G

Z

Data Item DI

"Logical" Provenance

"Physical" Provenance

"Logical" Provenance

"Physical" Provenance

No_Op

No_Op

No_Op

Logical

Physical
Access/Usage Control: Challenges

- Traditional access control models not applicable
  - Existing techniques do not apply to DAGs*

- Complexity of authorization conditions
  - Authorization may depend on sequence of operations performed by sources
  - Changes in policy

- Conflict resolution
  - Sources that contribute to the same data object may have conflicting policy requirements
  - Reconciliation of source and recipient policies

Privacy Challenges and Techniques

- **Sanitization**
  - Release provenance at appropriate granularity levels
- **Generalize provenance**
  - E.g., instead of releasing employee and department name who modified document, only release organization name

- **Cryptography**
  - Employ advanced cryptographic techniques that allow private evaluation of conditions
  - E.g., private similarity evaluation of two provenance DAGs
Integrity & Accountability: Challenges and Techniques

- **Integrity:**
  - Authenticate source and content of provenance information

- **Accountability**
  - Non-repudiation of a source’s role in the provenance chain even if anonymized for privacy

- **Techniques**
  - Conventional digital signatures may not be suitable
    - Provenance is highly dynamic and may include multiple sources that may not know/trust each other
    - Sources may need to remain anonymous
  - Non-interactive Editable Signatures for Provenance
    - Novel cryptographic techniques
Prototypes

1. Secure Provenance Management for OS
   - Provenance protection is achieved through trusted VMMs running within a higher trust domain than the OS
   - OS-independent mechanism, where provenance is embedded as watermark in the data
     - Maintains compatibility with existing applications

2. XML Document Dissemination System
   - Provenance is maintained at XML element level (fine-grained)
   - Protection through cryptographic tokens
Summary

Application Layer

Access/Usage Control  Privacy  Integrity  Accountability

Proposed Core of Secure Provenance Systems

Logical Structure and Physical Storage of Provenance

Contributions
1. Privacy-enhancing framework (i.e., models and mechanisms) for securing provenance lifecycle
2. Design/implementation of mechanisms for secure provenance management at OS layer and Data Layer