Provenance-based Access Control in Cloud IaaS

August 23, 2013
Dissertation Proposal

Dang Nguyen
Institute for Cyber Security
University of Texas at San Antonio
Data Provenance in Computer Systems

“In computer systems, activities are carried out by processes that take input data, input state, input configuration, and produce output data and output state. Such processes are compositional by nature and can be the result of sophisticated compositions (sequential, parallel, conditional, etc) of simpler processes.” (Luc Moreau, “The Foundation for Provenance on the Web”)
Characteristics of Provenance Data

• Information of operations/transactions performed against data objects and versions
  – Actions that were performed against data
  – Acting Users/Subjects who performed actions on data
  – Data Objects used for actions
  – Data Objects generated from actions
  – Additional Contextual Information of the above entities

• Directed Acyclic Graph (DAG)
• Causality dependencies between entities (acting users / subjects, action processes and data objects)

• Dependency graph can be traced/traversed for the discovery of Origin, usage, versioning info, etc.
Provenance and Access Control

- Compared to traditional access control, Provenance-based Access Control (PBAC) provides richer access control mechanisms.
  - For example: dynamic separation of duties issues.
Provenance Data Model

Base PBAC Model

Contextual PBAC Model

Provenance data sharing approaches
Provenance-aware Systems

- Capturing provenance data
- Storing provenance data
- Querying provenance data
- Using provenance data
- Securing provenance data
Open Provenance Model (OPM)

- **3 Node Types**
  - Artifact (ellipse): Object
  - Process (Rectangle): Action
  - Agent (Octagon/Hexagon): User/Subject

- **5 Causality dependency edge Types** (not a dataflow)
  - U: Used(Role)
  - G: wasGeneratedBy(Role)
  - C: wasControlledBy(Role)
  - wasDerivedFrom
  - wasTriggeredBy
World-leading research with real-world impact!
Provenance Data Model

- 4 Node Types
  - Object (Artifact)
  - Action (Process)
  - Subject (Agent)
  - Attribute

- 5 Causality dependency edge Types (not a dataflow) and Attribute Edge

```
c wasControlledby
u used
g wasGeneratedBy
t hasAttributeOf
```

World-leading research with real-world impact!
Capturing Provenance Data

(Subject1, Grade1, HW1, GradedHW1, ContextualInfoSet-Grade1)

(Grade1, u, HW1)
(Grade1, c, Subject1)
(GradedHW1, g, Grade1)

(Grade1, t[actingUser], Alice)
(Grade1, t[activeRole], TA)
(Grade1, t[weight], 2)
(Grade1, t[object-size], 10MB)
Provenance Graph

Sub1

HW1

Grade1

HW1_G

Alice

TA

2

10MB

t(actUser) t(...) t(...) t(...)

u c g
Storing and Querying Provenance Data

- Resource Description Framework (RDF) provides natural representation of triples.

- RDF-format triples can be stored in databases.

- Utilizes SPARQL Protocol and RDF Query Language for extracting useful provenance information.
  - Starting Node: any entities (not attribute nodes)
  - A matching path pattern: combination of dependency edges
Provenance Graph

World-leading research with real-world impact!
Provenance Graph

SELECT ?agent WHERE { HW1_G [g:c] ?agent}

Sub1

Alice

TA

2

10MB

t(actUser)
t(...)
t(...)t(...)
Provenance Graph

SELECT ?user WHERE { HW1_G [g:t[actUser]] ?user }

Alice  TA  2  10MB
Provenance Graph

SELECT ?user
WHERE
{ HW1_G' [g:u:g:c] ?user }

{ HW1_G' [[g:u]*:g:c] ?user }
PBAC Model Components

World-leading research with real-world impact!
PBAC_C : PBAC_B + Contextual Info.
PBAC_C in Cloud IaaS

- Subjects (S) → Request (s,a,o) → Actions (A) → Action on O → Objects (O)
- Tenants (T) Association
- Access Evaluation
  - Dependency List (DLIST)
  - Policies (P)
- Provenance Data (PD)
  - Base PD
  - Attributes PD

Access
Decision Activity
Utilized by
Captured as
Associated with

World-leading research with real-world impact!
Capturing Provenance Data

(Subject1, Create1, VMI1, ContextualInfoSet-Create1)

Create1, c, Subject1)
(VMI1, g, Create1)

Create1, t[tenant], “Development”)

World-leading research with real-world impact!
Single- vs Multi-Cloud (IaaS)

• Most single-cloud CSP provides centralized service.
  – Facilitates data sharing (provenance).

• Multi-cloud CSPs require collaboration for sharing data.
Multi-cloud PBAC

Cloud 1

Cloud 2

Cloud 3

World-leading research with real-world impact!
Provenance Data Sharing

Integration Approaches:

- Using Cascading subquery
- Y/N Resulting Nodes
- Sticky PD

Response to cascading query:

Using Sticky Provenance Data (SPD)

Integrated Use of Provenance Data (PD)
Provenance Data Sharing

Integrated Use of Provenance Data (PD)

Integration Approaches:
- Using Cascading subquery
- Using Sticky Provenance Data (SPD)

Response to cascading query:
- Y/N
- Resulting Nodes
- Sticky PD

Centralized Provenance and PBAC Services
Provenance Service

CloudService

USR -- 1. request -- PEP

PEP -- 2. request -- PDC

2. evaluate decision

PDC -- 3. evaluate decision -- PDP

4. raw prov data

5. collected PD -- PDM

PDM -- 6. graph-format PD -- DB

ProvService

ProvAuthzService

World-leading research with real-world impact!
PBAC Service

CloudService

USR

PEP

1. request

2. request

11. evaluate decision

12. result

PDC

PDM

7. queries

8. result

DB

PIP

10. info req

9. prov info return

5. info req

4. policy

3. policy req

PDP

PAP

ProvService

ProvAuthzService
Cross-tenant PBAC

Tenant R

VMI

Tenant T

VMI

Tenant D

VMI

Resource movement

Provenance tracking

Policy utilization

Decision

PBAC Authz

Outgoing policy

Incoming policy

yes

no

World-leading research with real-world impact!
OpenStack Authz

1. creds
2. token
3. req msg
4. policy information
5. result

Keystone

Identification Service

OpenStack Services

Nova, Glance, etc.

policy.json

Requester
Nova Architecture
Thank you!!!

• Questions and Comments?