Privacy Preserving Data Dissemination

Bharat Bhargava
CERIAS and CS Department
Purdue University

www.cs.purdue.edu/homes/bb
What needs to be done

• Protect the Data
• Protect the Communication
• Protect the Identity of sender and requester
• Evaluate Trust, Privacy Policies, Privacy Metrics, Deploy Apoptosis, Filtering
Research Topics

• Privacy is needed to protect source of information, the destination of information, the route of information transmission of dissemination and the information content itself.

• Trusted Router and Protection Against Collaborative Attacks. Identify malicious activity and collusion.

• Can we protect disseminated data? Active Bundle approach for controlled dissemination or apoptosis. Managed Information Object in Cross-Domain Information Exchange.

• Privacy of Sender and Requester in Data Sharing.
Northrop Grumman TechFest

Privacy Preserving Cross-Domain Data Dissemination
(with adaptable service selection)

PI: Prof. Bharat Bhargava

Technical Co-Chairs: Dr. Sunil Lingayat, Mr. Jason C. Kobes

June 2015
End-to-End Security in Trusted and Untrusted SOA and Cloud

1. Service request
   - Browser

2. Analysis of request source based on W3C standards
   - Insecure browser
     - Send active bundle to cloud for execution
   - Secure browser
     - CAC authentication
     - PIN authentication
     - Trust = X + Y
     - Low trust
     - High trust

3. Execution of active bundle in cloud
   - Active Bundle
     - Data access request
     - Encrypted search
     - Filtered/lower quality data
     - Filtered search results

4. Service Domain
   - Service X
     - Data request
     - Data Service X is authorized to access
     - Active Bundle

UNTRUSTED

TRUSTED
End-to-End Information Flow

1. Client sends request to the service using browser and shares data by means of Active Bundle (AB)
2. Service checks the request source (secure or insecure browser)
   • Based on W3C Crypto standards
3. Service executes AB in Cloud if created by an insecure browser
4. Service interacts with AB and requests data
5. AB behaves differently under different contexts
   • Full data dissemination based on service authorization/trust level
   • Context-based partial data dissemination based on insufficient authorization level
   • No data dissemination for unauthorized access/attacks
6. Cross-domain information exchange with trustworthy/untrustworthy subscribers
   • Data dissemination is done on a “need to know” basis by limiting the disclosure of decryption keys
   • Incremental disclosure of keys based on increase in the “need”
Active Bundle

- Active Bundle (AB) approach for secure data dissemination
  - Self-protecting data encapsulation mechanism
  - Provides secure cross-domain information exchange
- Sensitive data
  - Encrypted data items
- Metadata
  - Access control and operational policies
- Virtual Machine
  - Protection mechanism (self-integrity check)
  - Policy evaluation, enforcement and data dissemination
• Each UAV creates active bundle with captured image data and trust-based dissemination policies
• When active bundle is sent to a UAV, image data is filtered (blurred) based on trust level
• Trust levels of UAVs change based on context, distance, communication bandwidth
P2P Secure Data Sharing in UAV Network

Dissemination Policy:

- If $2.5 > \text{trust} \geq 2.2 \Rightarrow \text{blur level} = 20$
- If $2.2 > \text{trust} \geq 2.0 \Rightarrow \text{blur level} = 40$
- If $2.0 > \text{trust} \Rightarrow \text{blur level} = 80$

trust = 2.4

trust = 2.1

trust = 1.5
P2P Secure Data Sharing in UAV Network

Dissemination Policy:

If context = emergency => contrast = 0.4
If 2.5 > trust ≥ 2.2 => contrast = 0.2
If 1.8 > trust => contrast = 0.1

trust = 2.0
context = emergency

trust = 2.4

trust = 1.7
Example of Controlled Data Filtering of Electronic Healthcare Record (EHR)

- EHRs stored in a database and filtered for different data consumers using SQL queries run in

<table>
<thead>
<tr>
<th>PAT-ID</th>
<th>NAME</th>
<th>Mobile</th>
<th>Test Date</th>
<th>HEIGHT</th>
<th>WEIGHT</th>
<th>CHOLESTEROL</th>
<th>BLOOD-SUGAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>99999</td>
<td>MNP</td>
<td>11111</td>
<td>11/11/2010</td>
<td>175</td>
<td>190</td>
<td>198</td>
<td>95</td>
</tr>
<tr>
<td>99998</td>
<td>ABC</td>
<td>22222</td>
<td>12/02/2010</td>
<td>170</td>
<td>180</td>
<td>192</td>
<td>91</td>
</tr>
<tr>
<td>99997</td>
<td>XYZ</td>
<td>33333</td>
<td>13/03/2010</td>
<td>180</td>
<td>201</td>
<td>199</td>
<td>98</td>
</tr>
</tbody>
</table>

a. Data consumer verified as doctor at the hospital can get all patient data

<table>
<thead>
<tr>
<th>PAT-ID</th>
<th>NAME</th>
<th>Mobile</th>
<th>Test Date</th>
<th>HEIGHT</th>
<th>WEIGHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>99999</td>
<td>MNP</td>
<td>11111</td>
<td>11/11/2010</td>
<td>175</td>
<td>190</td>
</tr>
<tr>
<td>99998</td>
<td>ABC</td>
<td>22222</td>
<td>12/02/2010</td>
<td>170</td>
<td>180</td>
</tr>
<tr>
<td>99997</td>
<td>XYZ</td>
<td>33333</td>
<td>13/03/2010</td>
<td>180</td>
<td>201</td>
</tr>
</tbody>
</table>

b. Hospital Receptionist gets filtered data

c. Insurance company gets only the minimal required data
Context-based Data Dissemination (disclosure during emergency context)

- **Paramedic’s App**
  - Active Bundle
  - Medical data
  - Medical history
  - E(Patient ID)
  - E(Insurance ID)
  - E(Medical test prescription)
  - E(Prescription)
  - E(Treatment code)

- **Doctor’s App**
  - Active Bundle
  - Patient ID
  - Medical data
  - E(Insurance ID)
  - E(Medical test prescription)
  - Prescription
  - E(Treatment code)

- **Laboratory’s App**
  - Active Bundle
  - Patient ID
  - Medical test prescription
  - E(Medical data)
  - E(Insurance ID)
  - E(Medical history)
  - E(Prescription)
  - E(Treatment code)

- **Pharmacy’s App**
  - Active Bundle
  - Patient ID
  - Prescription
  - E(Medical data)
  - E(Insurance ID)
  - E(Medical history)
  - E(Medical test prescription)
  - E(Treatment code)

- **Insurance’s App**
  - Active Bundle
  - Patient ID
  - Insurance ID
  - Treatment code
  - E(Medical data)
  - E(Medical history)
  - E(Medical test prescription)
  - E(Prescription)

**Medical Information System**

- Emergency context
- Data access: GRANTED

14
Trust-based Data Dissemination

- Name
- Email
- Credit card type
- Credit card
- Shipping preference
- Mailing address

SERVICE MONITOR

1. Order request + Active Bundle
2. Verify request + Active Bundle

Trust Request

Trust level: 5

Trust level: 1

Low trust level of Seller
Data access: DENIED

Shopping Service

Payment Service

Seller Service

Shipping Service
• UAV on search and rescue mission for a fire hazard
• When invoked from secure browser, active bundle is sent to each involved service domain (Fire Control, Ambulance, Weather), so they access data they are authorized for.
Trusted vs. Untrusted Service Invocation (cont.)

- When invoked from insecure browser, active bundle is sent to the cloud and each service (Fire Control, Ambulance, Weather), interacts with active bundle in the cloud for data access.
Service monitor intercepts all client-service/service-service interactions.

The approach aims to provide a unified security architecture for SOA and cloud by integrating components for:

- **Service trust management**
- **Interaction authorization** between different services
- **Anomaly detection** based on service behavior
- **Dynamic service composition**
- **Secure data dissemination** using active bundles
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  
/home address

Salisbury Street  
West Lafayette, IN

somewhere in  
West Lafayette, IN

250 N. Salisbury Street  
West Lafayette, IN

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
Another Example of Controlled Data Dissemination

Example: Imagine you send an email and the following will be enforced:

1) Receiver can read it
2) Receiver can not copy it
3) Receiver can not change it
4) Receiver can forward it only to destinations allowed by the owner (originator) or a guardian (owner’s delegate)
5) Receiver can access it only if authorized by the owner/guardian
Introduction

• Privacy is fundamental to trusted collaboration and interactions to protect against malicious users and fraudulent activities.
Distance Metrics for Trust

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

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

---

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
IDEAS

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
IDEAS

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

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
IDEAS

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
IDEAS

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
Privacy Metrics

A. Anonymity set size metrics
B. Entropy-based metrics
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)
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
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
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
Can we protect disseminated data?

**Data dissemination** is the process of forwarding sensitive data from any guardian to other guardians:
- A guardian is an entity (either human or not) that accesses data or disseminates them.
- The owner loses control of her data when it is disseminated.
- Send Data encapsulated with policies for Access Control.
- Trust level of requester determines the Content of Data Disclosure.
- Monitor Obligation Compliance of requester.
Problem Statement

• Research Problem
  • How to securely share sensitive data, control its dissemination, minimize unnecessary disclosure and protect them throughout their lifecycle?

• Research Goal
  • Propose a solution to control data disclosure and minimize the risk of unauthorized disclosure
Current Solutions for Data Dissemination

Traditional approaches:

- Data are considered passive entities unable to protect themselves
- Require another active and trusted entity to protect them – a trusted processor, a trusted memory module, a trusted application or a trusted third party
Active Bundle Approach

• **Active bundle (AB)**
  - An encapsulating mechanism **protecting data** carried within it
  - Includes **sensitive data**
  - Includes **metadata** used for managing the privacy
  - Includes Virtual Machine (VM)
    • performing a set of **operations**
    • protecting data **privacy**
  - VM is dependent on the data and metadata
    • Unlike client applications, it is independent of the application that uses the data

• **AB is used to provide data privacy protection**
Active Bundle Scheme

- **Metadata:**
  - Access control policies
  - Data integrity checks
  - Dissemination policies
  - Life duration
  - ID of a trust server
  - ID of a security server
  - App-dependent information
  - ...

- **Sensitive Data:**
  - Identity Information
  - ...

- **Virtual Machine (algorithm):**
  - Interprets metadata
  - Checks active bundle integrity
  - Enforces access and dissemination control policies
  - ...

* E( ) - Encrypted Information
Active Bundle Operations

• Privacy policy enforcement
  • Enforces privacy policies included in metadata of the data

• Self-Integrity check
  • Using a hash function

• Evaporation
  • Evaporates (a part of) sensitive data when threatened with disclosure

• Apoptosis
  • Self-destruction of AB’s sensitive data and metadata
Components of Active Bundle

- **Identity data**: Data used during authentication, getting service, using service (i.e. SSN, Date of Birth).
- **Disclosure policy**: A set of rules for choosing Identity data from a set of identities in IDM Wallet.
- **Disclosure history**: Used for logging and auditing purposes.
- **Negotiation policy**: This is Anonymous Identification, based on the Zero Knowledge Proofing.
- **Virtual Machine**: Code for protecting data on untrusted hosts. It enforces the disclosure policies.
Enabling AB

1. Get decryption information and host’s trust level from SSA

2. Is host’s trust level lower than AB’s trust level?
   - True: Apoptosis
   - False: Proceed to step 4

4. Is integrity check successful?
   - False: Apoptosis
   - True: Proceed to step 6

5. Apoptosis

6. Decrypt AB

7. Enforce AB

8. Provide output to the host

9. Update audit information
TTP-based Implementation of Active Bundle Approach

• Can use Trusted Third Party (TTP) to enable dissemination of ABs
• The role of TTP is:
  – Generate and store cryptographic keys for Abs
  – Trust management of requestors
  – Authorize hosts wishing to access ABs
  – Provide the cryptographic keys to ABs so that ABs can deliver contents to their authorized hosts
TTP-based Implementation of Active Bundle Approach—Cont.

• Tools used to develop the prototype
  ➢ Programming language Java 1.6.01
  ➢ Development environment Eclipse 3.5.2
  ➢ Cryptographic libraries: Java cryptographic libraries
    • Java Cryptography Architecture (JCA)
    • Java Cryptography Extension (JCE)
  ➢ JADE
    • A software framework that provides basic MA middleware functionalities
Improvements in the AB Scheme

- Policy-based selective dissemination
- Decrease dependence on Trusted Third Party (TTP)
- Resilience against malicious receivers
Improving the AB Scheme

- **Policy-based selective dissemination**
  
  - Organize data in AB into separate items
  
  - Encrypt each item with a separate key
  
  - Specify XACML based access policies for authorization for an item or a set of items
Improving the AB Scheme

- Decreasing dependence on TTP
  - Use Shamir’s threshold secret sharing technique
    - Split decryption keys into n shares
    - Set a threshold t such that t (<n) shares are required for key reconstruction
  - Use a public Distributed Hash Table (DHT) to store key shares at random nodes
Advantages of using DHT

• Decentralized storage with no single point of trust

• Huge scale with millions of geographically distributed nodes
  • Hard to deduce key shares
  • Difficult to compromise nodes that store key shares

• Refresh key shares periodically
  • Ensures availability when nodes crash or leave
  • Protects against attacks
Improving the AB Scheme

• Protection Against Malicious Hosts
  • Use TPM (trusted platform module) to ensure that host is not already compromised
  • Intertwine code and data together – hide data within the code to make it incomprehensible
  • Use polymorphic code – code changes itself each time it runs but its semantics don't change
  • Can store the control flow information in random DHT nodes
  • Perform code obfuscation - hide data and real program code within a scrambled code
Active Bundle Features

• Controlled and Selective Dissemination: Control the dissemination and selectively share the data based on the policies

• Quantifiable Data Dissemination: Track the amount of data disclosed to a particular host and decide to further disclose or deny data requests

• Contextual Dissemination: Context-aware dissemination

• Dynamic Metadata Adjustment: Update the policies based on a context, host, history of interactions, trust level etc.

• Do not require hosts to have a policy enforcement engine or a trusted component

• Doesn’t rely on a TTP

• No trusted destination host assumption – works on unknown hosts
Advantages of the Active Bundle Approach

- Ensure protection mechanisms are available for ABs when needed
- No need for a client application to enforce policies associated with the received data
- No need to trust the client application
  - The VM is provided by the data owner
- Reduced severity of the risk of code tampering
  - VM is dependent on data and input; e.g., encrypt a set of instructions of VM such that the decryption key is a valid input.
- Attacker who accesses one AB does not learn how to access other ABs
Active Bundle Applications

A. Identity Management in SOA
B. Mobile-Cloud Pedestrian Crossing Guide Application for the Blind
C. A Trust-based Approach for Secure Data Dissemination in a Mobile Peer-to-Peer Network of UAVs
Policy Enforcement at Data Source

• Requires source visibility and availability
• E.g. client-server paradigm
  • Server authenticates client requests
• Source becomes a single point of failure
• Scalability issues
  • Source becomes bottleneck
• Messages
  • Source: O(n)
  • Receiver: O(1)
Policy Enforcement at Data Mediator

- Trusted Third Party (TTP) becomes a single point of trust and failure
- E.g. Pub-Sub system
- Attacks on TTP lead to data leakage
- TTP can aggregate private information
  - Disclosure for profit
  - Disclosure for subpoenas
- Messages
  - Source: $O(1)$
  - Receiver: $O(1)$
  - TTP: $O(n)$
Policy Enforcement at Data Receiver

• Requires presence of a trusted EM on the receiver
  • Trusted hardware
  • Trusted software
• Requires advance knowledge of receivers
• Distribution of trusted EM to external domains is a challenge
• Messages
  • Source: $O(n)$
  • Receiver: $O(1)$
Proposed Solution

• Mobile (portable) Execution Monitor (EM)
  • Transmitted along with data and policies to each endpoint in interaction
  • Monitors data access requests at each end point
  • Enables policy evaluation and enforcement
  • Ensures controlled data disclosure

• Messages
  • Source: O(1)
  • Receiver: O(1) – O(n) depending on the service composition topology
if (service = authorized)

\[ \text{EM}(k_i, P, C) = d_i \]
Proposed Solution

• Solution is realized using Active Bundle (AB) scheme
  • Self-protecting data encapsulation mechanism
  • Provides secure cross-domain data disclosure

• Sensitive data
  • Encrypted data items

• Metadata
  • Access control policies

• Monitor
  • Self-integrity check
  • Policy evaluation
  • Policy enforcement
  • Data dissemination

- Identity information
- Operational information
- Access control policies
- Operational policies
- Life duration

- Encrypted data items

- Interprets metadata
- Checks bundle integrity
- Evaluates requests
- Enforces policies
- Discloses data
Classification of AB

• Immutable AB
  • Static data and policies that cannot be modified after creation
  • Prevents inconsistencies and offers better security

• Mutable AB
  • Data and policies can be updated after creation
  • Local storage of state information and interaction logs

• Stateless AB
  • Self-sufficient execution using inherent knowledge
  • No external information exchange and state maintenance

• Stateful AB
  • Communicates with a third party to exchange state information and store interaction logs
  • Uses external information for policy evaluation and data disclosure decisions
  • Data provenance and context-aware disclosure capabilities
Solution Components

• Service authentication can be based on
  • Password, Certificate, Biometric, PKI

• Service request authorization is based on policy evaluation
  • Flexible policy specification based on access control models such as Attribute/Role-based

• Key management for data disclosure
  • Key inclusion (prone to attacks)
  • Centralized key management service (use of trusted third party for key storage and distribution)
  • Distributed key management that splits the keys into shares using threshold secret sharing and uses a Distributed Hash Table (DHT) to store the shares and reconstruct the key by retrieving minimum threshold number of shares (unsuitable for real-time interactions in a service environment)
  • Dynamic key derivation based on the unique information generated in AB execution control flow steps only if the service is authenticated and authorized

• Tamper Resistance
  • Correct data dissemination depends on the correct execution of AB control flow steps
  • Verify the integrity of the execution steps to ensure there is no difference from the original code (using secure one-way hash function)
  • Derive keys based on digests of AB execution steps and their resources
  • Any modification of AB changes the digest resulting in incorrect key derivation
Implementation
Implementation

• Active Bundle (AB) implemented an executable JAR file
• AB API implementation using Apache Thrift RPC framework
• Policy specification using XACML-based XML or JSON
• Policy evaluation using WSO2 Balana or Java conditions
• AB transfer by means of REST message (HTTP body)
• Web services implemented as Node.js applications
• SOA composition using RESTful Service APIs
AB API

- `getSLA()`
  - Public information regarding service provided by AB
    - Defined as operation policies, e.g. expiration time, maximum active time, maximum session requests
- `getValue(req, sigReq, cert)`
  - Authentication of service using signature and certificate
  - Authorization of data request using applicable policies
  - Key derivation is based on digests of self-integrity checks
  - Data decryption and disclosure
- `getSecureValue(req, sigReq, cert)`
  - In addition to `getValue()`, disclosed data is encrypted with service’s public key
AB Data

- Name: Bob
- Email: bob55@purdue.edu
- Address: 305 N Univ St West Lafayette 47906
- Payment: 4321 1235 7890 0145
- AB data file (cipher.json)

```json
{
  "ab.user55.name": "aI/vXZyavQ8gCK5yyjn9tg==",
  "ab.user55.email": "tX/clGBR5AN0jyF3P95Q==",
  "ab.user55.street": "0FpW9FhT6MjVvEL0AjLuA==",
  "ab.user55.city": "s/jRmjejcdAH1lGkzzPELQ==",
  "ab.user55.zip": "EKk65cXy29gBUVkpOJw==",
  "ab.user55.payment": "+ZlnFwX2oAOHtMy+eR3aPA=="
}
```
### AB Access Policy Examples

Credit card access policy (Attribute-based):

<table>
<thead>
<tr>
<th>Deny</th>
<th>Resource</th>
<th>Visa payment services</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subject: CA signed certificate</td>
<td>Credit card</td>
</tr>
<tr>
<td></td>
<td>Action</td>
<td>READ</td>
</tr>
<tr>
<td></td>
<td>Environment: Service trust</td>
<td>&lt; 9</td>
</tr>
</tbody>
</table>

Medical data access policy (Role-based):

<table>
<thead>
<tr>
<th>Deny</th>
<th>Resource</th>
<th>Electronic Healthcare Record</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subject: Role</td>
<td>Doctor</td>
</tr>
<tr>
<td></td>
<td>Action</td>
<td>READ</td>
</tr>
<tr>
<td></td>
<td>Environment: Subject rating</td>
<td>&lt; 4 star</td>
</tr>
</tbody>
</table>
## AB Access Policy Examples

Cross-domain data access policy (Context-based):

<table>
<thead>
<tr>
<th>Deny</th>
<th>Intelligence reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resource</td>
<td>Electronic Healthcare Record</td>
</tr>
<tr>
<td>Subject: Classification</td>
<td>Top secret</td>
</tr>
<tr>
<td>Action</td>
<td>READ</td>
</tr>
<tr>
<td>Environment: Attack context</td>
<td>Secret, Confidential</td>
</tr>
</tbody>
</table>

Medical data access policy (Context-based):

<table>
<thead>
<tr>
<th>Deny</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resource</td>
<td>Dr. Alice</td>
</tr>
<tr>
<td>Subject: Role</td>
<td>Electronic Healthcare Record</td>
</tr>
<tr>
<td>Action</td>
<td>READ</td>
</tr>
<tr>
<td>Environment: Emergency context</td>
<td>Paramedics, Doctors</td>
</tr>
</tbody>
</table>
Example AB XACML Policy

```xml
<Policy xmlns="urn:oasis:names:tc:xacml:3.0:core:schema:wd-17" PolicyId="policy-user55-payment1"
      RuleCombiningAlgId="urn:oasis:names:tc:xacml:3.0:rule-combining-algorithm:deny-overrides" Version="1.0">
  <Target>
    <AnyOf>
      <AllOf>
        <Match MatchId="urn:oasis:names:tc:xacml:1.0:function:string-equal">
          <AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">ab.user55.payment</AttributeValue>
          <AttributeDesignator AttributeId="urn:oasis:names:tc:xacml:1.0:resource:resource-id"
                                 Category="urn:oasis:names:tc:xacml:3.0:attribute-category:resource"
                                 DataType="http://www.w3.org/2001/XMLSchema#string" MustBePresent="true"/>
        </Match>
      </AllOf>
    </AnyOf>
    <AnyOf>
      <AllOf>
        <Match MatchId="urn:oasis:names:tc:xacml:1.0:function:string-equal">
          <AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">Payment</AttributeValue>
          <AttributeDesignator AttributeId="http://absoa.cs.purdue.edu/policy/payment"
                                 Category="urn:oasis:names:tc:xacml:1.0:subject-category:access-subject"
                                 DataType="http://www.w3.org/2001/XMLSchema#string" MustBePresent="true"/>
        </Match>
      </AllOf>
    </AnyOf>
  </Target>
</Policy>
```
Example AB XACML Policy cont.

```xml
<Match>
  </AllOf>
  </AnyOf>
</Target>

<Rule Effect="Deny" RuleId="rating">
  <Condition>
    <Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:integer-less-than">
      <Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:integer-one-and-only">
      </Apply>
      <AttributeValue DataType="http://www.w3.org/2001/XMLSchema#integer">5</AttributeValue>
    </Apply>
  </Condition>
</Rule>

<Rule Effect="Permit" RuleId="permit-rule"/>

</Policy>
```
Example AB JSON Policy

• 78% reduction in policy size (0.79 KB) using JSON

```json
{
    "ab.user55.payment" : {
        "subject" : ["Payment"],
        "rating" : "5-star"
    }
}
```
AB Creation

• AB Generator Application (ABGen) is used to create an Active Bundle
  • Frontend: Node.js web application (pre AB creation)
    • Provides a web interface to users to specify data and policies
    • Parses user input and submits it to the backend
  • Backend: Java application
    • Generates keys and encrypts data
    • Embeds encrypted data and policies in an AB template
    • Generates the AB as a JAR file
  • Frontend (post AB creation)
    • Serializes the AB file using Base64 encoding
    • Sends the encoded AB file to the specified service by appending it to the HTTP message
AB-Service Interaction

- AB is sent as part of the request message (HTTP body)
- AB Handler intercepts the message containing AB
AB-Service Interaction

- AB Handler extracts AB from the message
  - Executes AB that starts the AB process
  - Forwards the request to the service
AB-Service Interaction

- Service interacts with AB to process the request
  - Sends data access requests to AB
AB Actions on Data Request

• Step 1: Authentication
  • Validate data request and verify service signature
  • Validate certificate and verify CA signature
  • Calculate digest for authentication module
• Step 2: Authorization
  • Select and evaluate applicable policies
  • Calculate digest for authorization module
• Step 3: Key derivation
  • Aggregate digests and module outputs
  • Use aggregated value as a secret to derive key
• Step 4: Data disclosure
  • Decrypt data item using the derived key
Demonstration
Applications
Online Shopping using AB

- Name
- Email
- Credit card type
- Credit card
- Shipping preference
- Mailing address

1. Order request
2. Seller Service
3. Shipping Service
4. Payment Service
Online Shopping using AB

- Name
- Email
- Credit card type
- Credit card
- Shipping preference
- Mailing address

1. Order request
   - Active Bundle
   - Shopping Service

2. Verify request +
   - Active Bundle
   - Seller Service

3. Shipping request +
   - Active Bundle
   - Shipping Service

4. Payment request
   - Active Bundle
   - Payment Service

- E(Name)
- E(Email)
- E(Payment type)
- E(Credit card)
- Shipping preference
- E(Mailing address)

- Name
- Email
- Payment type
- E(Credit card)
- E(Shipping preference)
- E(Mailing address)
Unauthorized Data Access

- Name
- Email
- Credit card type
- Credit card
- Shipping preference
- Mailing address

Shopping Service

1. order request +
   Active Bundle

2. verify request +
   Active Bundle

Seller Service

3. shipment request +
   Active Bundle

Shipping Service

4. payment request +
   Active Bundle

Payment Service

Unauthorized access to credit card Data access: DENIED

- E(Name)
- E(Email)
- E(Payment type)
- E(Credit card)
- Shipping preference
- E(Mailing address)

- Name
- E(Email)
- E(Payment type)
- Credit card
- E(Shipping preference)
- E(Mailing address)
Context-based Data Dissemination

- Medical data
- Medical history
- E(Patient ID)
- E(Insurance ID)
- E(Medical test prescription)
- E(Dispensation)
- E(Treatment code)

Paramedic’s App

Emergency context
Data access: GRANTED

Active Bundle

Doctor’s App

- Patient ID
- Medical data
- Medical history
- Medical test prescription
- Prescription
- E(Insurance ID)
- E(Treatment code)

Active Bundle

Laboratory’s App

- Patient ID
- Medical test prescription
- E(Medical data)
- E(Insurance ID)
- E(Medical history)
- E(Dispensation)
- E(Treatment code)

Active Bundle

Pharmacy’s App

- Patient ID
- Prescription
- E(Medical data)
- E(Insurance ID)
- E(Medical history)
- E(Medical test prescription)
- E(Treatment code)

Active Bundle

Insurance’s App

- Patient ID
- Insurance ID
- Treatment code
- E(Medical data)
- E(Medical history)
- E(Medical test prescription)
- E(Dispensation)
Trust-based Data Dissemination

- Name
- Email
- Credit card type
- Credit card
- Shipping preference
- Mailing address

1. Shopping Service
   - Name
   - Email
   - Payment type
   - E(Credit card)
   - E(Shipping preference)
   - E(Mailing address)

2. Seller Service
   - Trust level: 5
   - Trust level: 1
   - Data access: DENIED

- Active Bundle
- Order request
- Verify request

SERVICE MONITOR

Low trust level of Seller
Data Dissemination under Tamper attack

**Paramedic’s App**
- Active Bundle
- Patient ID
- Medical test prescription
- E(Medical data)
- E(Insurance ID)
- E(Medical history)
- E(Prescription)
- E(Treatment code)

**Laboratory’s App**
- Active Bundle
- Patient ID
- Medical test prescription
- E(Medical data)
- E(Insurance ID)
- E(Medical history)
- E(Medical test prescription)
- E(Prescription)
- E(Treatment code)

**Pharmacy’s App**
- Active Bundle
- Patient ID
- Prescription
- E(Medical data)
- E(Insurance ID)
- E(Medical history)
- E(Medical test prescription)
- E(Prescription)
- E(Treatment code)

**Medical Information System**
- Active Bundle
- Patient ID
- Insurance ID
- Medical data
- Medical history
- Medical test prescription
- Prescription
- Treatment code

**Attacker’s App**
- Active Bundle
- Patient ID
- Insurance ID
- Treatment code
- E(Medical data)
- E(Medical history)
- E(Medical test prescription)
- E(Prescription)

Tampering attack on AB
Data access: DENIED
Resilience
Attack Resilience

• Type of Attacks on the framework
  • Repudiation attack
    • Attacker sends a malicious AB
  • Man in the middle attack during AB transfer
    • Attacker intercepts the AB and tries to gain unauthorized access
  • Man in the middle attack during AB-Service interaction
    • Attacker eavesdrops and alters the communication to gain unauthorized access
  • Tamper attack
    • Attacker compromises the AB by modifying its policies and code
  • Execution Hijack attack
    • Attacker reverse engineers the execution control flow of AB to gain unauthorized access
Attack Resilience

• Trusting the AB and its execution
  • Digital signatures to validate the authenticity and integrity of AB
  • Isolated execution of AB using containers e.g. Docker
  • Cloud-based execution

• Protecting AB against attacks
  • Secure communication e.g. HTTPS
    • During AB transfer
    • During AB-Service interaction (`getSecureValue()` method)
  • Tamper resistant code (code integrity checks)
    • Correct secret key generation only in case of correct execution
  • Execution Hijack Avoidance
    • Use of type safe language e.g. Java
    • Digitally signed code in AB JAR file
      • JVM validates signatures at runtime and prevents execution of malicious code
    • Secure hardware based execution (e.g. TPM, Encrypted write protected memory)
  • Cloud-based execution
    • Third party code execution platform that does not broker any data or policies
Performance
Experimental Setup

• Measurements
  • Experiment 1: Growth in AB size with increase in the number of policies
  • Experiment 2: Growth in AB and Service interaction time with increase in the number of policies
  • Experiment 3: Tamper Resistance overhead in AB execution

• Variations
  • AB versions
    • ABx – XACML-based policies and WSO2 Balana-based policy evaluation
    • ABxt – ABx with tamper resistance capabilities
    • ABc – JSON-based policies and JAVA-based policy evaluation
    • ABct – ABc with tamper resistance capabilities
  • Number of AB policies

• Environment
  • Amazon EC2 C3 Large and XLarge instances

• Data collection
  • 5 runs of each experiment
  • 100 requests per run
Experiment 1: AB Size vs Number of policies

- Observations
  - Tamper resistance adds a slight overhead to AB size (< 2 KB)
Observations

- 78% reduction per policy (0.79 KB) with JSON-based policies
  - Additional one-time reduction of 8.5 KB with Java-based policy engine

Experiment 1: AB Size vs Number of policies
Experiment 2.1: AB-Service Interaction
Time vs Number of policies (EC2 Large)
Experiment 2.1: AB-Service Interaction
Time vs Number of policies (EC2 Large)

• Observations
  – Significant reduction in time with the use of JSON-based policies and Java-based evaluation
    • Evaluation of XACML policies involves the traversal of XML policy and request trees
    • Evaluation of JSON policies involves execution of highly optimized Java code
Experiment 2.2: AB-Service Interaction Time vs Number of policies (EC2 XLarge)
Experiment 3.1: AB Tamper Resistance Overhead (EC2 Large)

- Observations
  - Tamper resistance has higher overhead for XACML policies
    - Digest calculation of XACML policies involves the traversal of XML policy and request trees
    - Digest calculation of JSON policies takes less time due to smaller policy size
Experiment 3.2: AB Tamper Resistance Overhead (EC2 XLarge)
Experiment 4.1: Scenario Time
Experiment 4.1: Scenario Time (EC2 Large)

• Observations
  • Less than 1.7 sec for XACML policies
  • Less than 1 sec for JSON policies
Experiment 4.2: Scenario Time (EC2 XLarge)

- Observations
  - Less than 1.3 sec for XACML policies
  - Less than 0.8 sec for JSON policies
Framework Capabilities

• Policy-based access control
• Privacy-preserving selective data dissemination
• Context-based adaptable data dissemination
• Independent of third party data and policy management
• Independent of source availability after initial AB transfer
• Ability to operate in external environment
• Reduced service liability for protecting PII
• Compatible with standard service infrastructure
• Agnostic to policy language and evaluation engine
Cost-Analysis of Framework

• Costs:
  • Message size and network overhead due to AB
    • Monitor adds a small constant overhead to every message containing AB
  • Service response delay due to interaction between AB and service
    • Interaction with EM at data owner, mediator, or receiver also has this overhead
  • Increased resource usage in service domain
    • Low interaction time does not have a significant impact on service performance