CERIAS

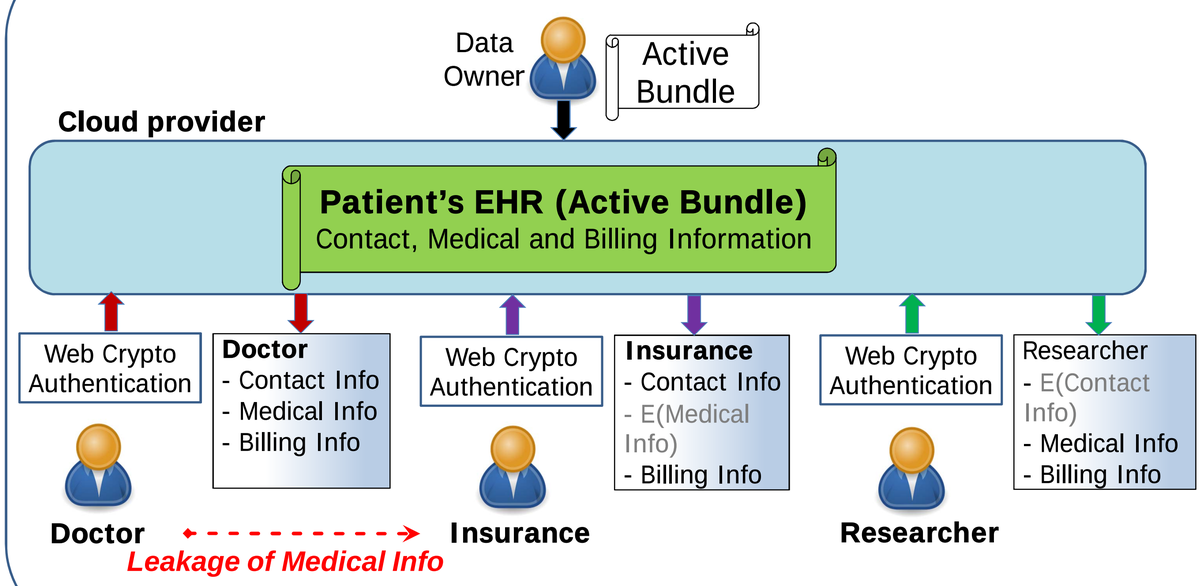
The Center for Education and Research in Information Assurance and Security

Data Leakage Detection and Privacy

Denis Ulybyshev[1], Bharat Bhargava[1], Aala Alsalem[1], Harry Halpin[3], Mai Elkady[1], Rohit Ranchal[2]

[1]Computer Science and CERIAS, Purdue University; [2]IBM Watson Health Cloud; [3] MIT

MOTIVATION



Scenario of EHR Dissemination in Cloud (suggested by Dr. Leon Li, NGC)

PROBLEMS

- Opaque data sharing
- Undetected data leakages

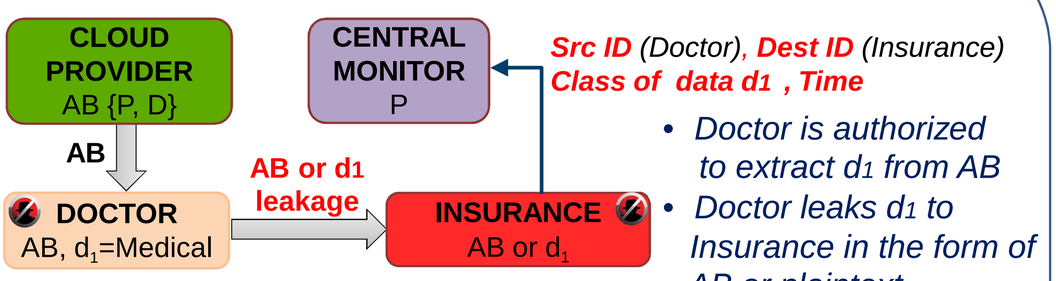
OBJECTIVES

- Detect data leakage to unauthorized services
- Authorized service should only be able to access data items for which it is authorized
- Provide data dissemination based on cryptographic capabilities of client's browser
- Support different authentication methods for client service

PROPOSED SOLUTION

- Use Active Bundles (AB) to ensure that authorized service is only able to access data items for which it is authorized
- Redirect unauthenticated client's request from Cloud Server to Authentication Server (AS)
- Detect crypto capabilities of client's browser, authentication method, type of the device, source network
- Based on that and access control policies retrieve data from AB

PROPOSED SOLUTION

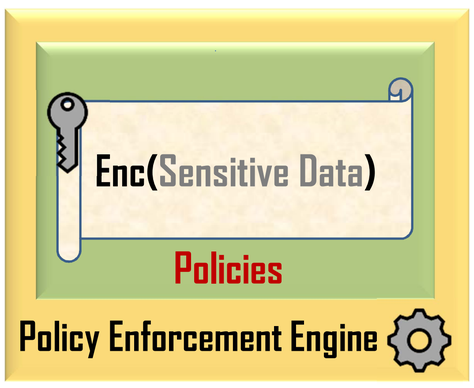


- *Doctor is authorized to extract d_1 from AB*
- *Doctor leaks d_1 to Insurance in the form of AB or plaintext*

- Insurance tries to decrypt d_1 => CM is notified
- CM checks whether d_1 is supposed to be at Insurance
- In addition, web crawler can scan folders, verify digital watermarks of files to detect copyright violations
- Visual watermarks are embedded into data
- At first, give less data to service and monitor its trust level
- Anti – fragility: once leak is detected system gets stronger
- **Damage = P (Data is Sensitive) * P (Service is Malicious) * P(t)** ,
  P(t) – probability function for data sensitivity in time

ACTIVE BUNDLE (AB)

- Self-integrity check
- Policy evaluation and enforcement
- Selective data dissemination



- Access control policies
- Operational policies

FEATURES

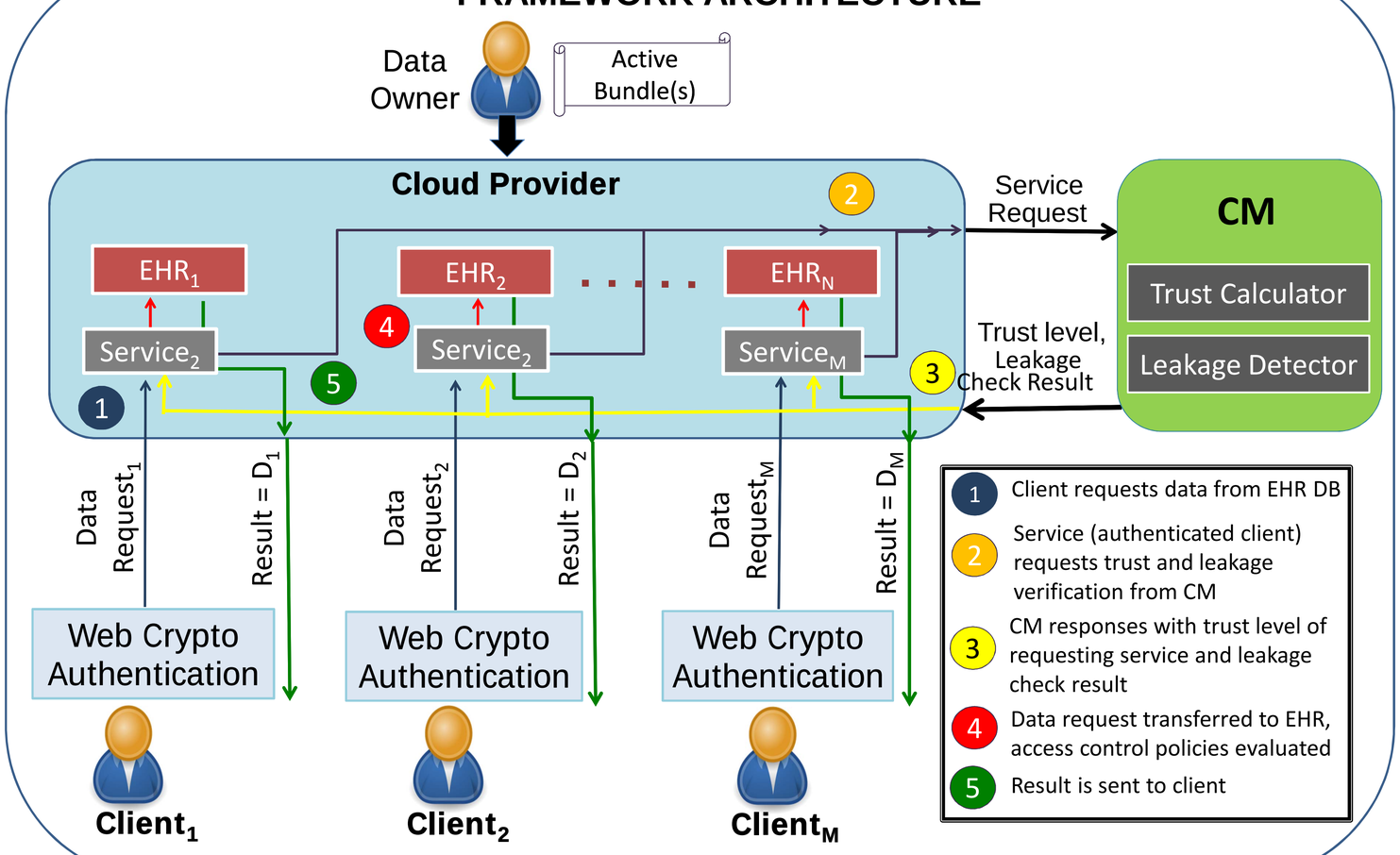
- Selective dissemination based on access control policies, browser crypto capabilities, authentication method, source network
- Independent of source availability
- Independent of trusted third parties
- Can operate in untrusted environment
- Supports data leakage detection

IMPLEMENTATION

- AB implemented as an executable JAR file
- Apache-thrift based API
- JSON-based policies
- WSO2 Balana-based policy engine
- Node.js-based SOA architecture
- RESTful web-services

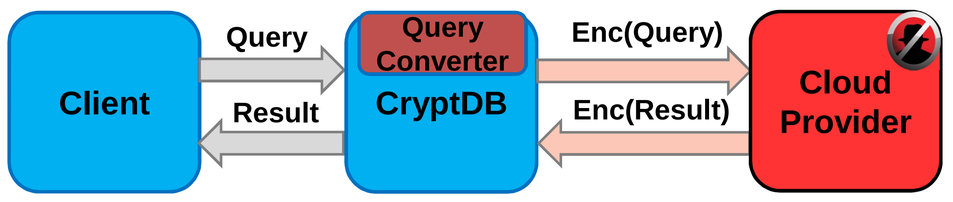
https://github.com/Denis-Ulybysh/absoa16

FRAMEWORK ARCHITECTURE



ENCRYPTED SEARCH

- Cloud provider hosts database of ABs
- AB contains data in encrypted form
- AB has extra-attribute used for indexing and search



- **Query example:**
  select prescription from EHR_DB where diagnosis LIKE '%insomnia%';
- **Converted query:**
  select c1 from Alias1
  where ESRCH (Enc(diagnosis), Enc(insomnia));

FUTURE WORK

- Support database of ABs in Hospital Information System
- Comprehensive performance and scalability evaluation
- Support isolated AB execution (Linux Docker Container)

ACKNOWLEDGEMENT: This research is supported by NGC Research Consortium. We collaborated with Donald Steiner, Leon Li and Jason Kobes

CERIAS

PURDUE UNIVERSITY