THE VALUE OF PERFORMANCE.

NORTHROP GRUMMAN

Northrop Grumman Cybersecurity Research Consortium (NGCRC)

2016 Fall Symposium

CYBERSECURITY RESEARCH Consortium

Privacy-Preserving Data Dissemination and Adaptable Service Compositions in Trusted and Untrusted Cloud

04 November 2016

Bharat Bhargava
Purdue University

Technical Champion(s): Leon Li, Jason Kobes, Sunil Lingayat, Donald Steiner

COLLABORATION WITH NGC

“WaxedPrune” Project:

Web-based Access to Encrypted Data - Processing in Untrusted Environments

Researchers at NGC

Leon Li

Donald Steiner

Sunil Lingayat

Jason C Kobes

COLLABORATION WITH NGC

Weekly meetings to:

- **Advance research based on vision of Donald Steiner, Leon Li, Jason Kobes**
- **Install and configure software at MIT side**
- **Integrate work with MIT (Harry Halpin)**

Researchers at Purdue

Bharat Bhargava
Denis Ulybyshev
Pelin Angin
Miguel Villarreal
Byungchan An
Rohit Ranchal
Tim Vincent
Leszek Lilien

Outline

- Problem Statement
- Benefits of Proposed Research
- Prototype Demo
- Impact
- State of the Art
- Year 7 (2015-2016) Final Report
  - Methodology
  - Results
- Year 8 (2016-2017) Proposal

Problem Statement

Focus: Secure Data Dissemination in Cloud

- Authorized service can only access data items for which it is authorized
- Unauthorized service denied
- Provide data dissemination based on cryptographic capabilities of client’s browser and authentication methods
- Support different authentication methods for client service
- *Adaptable service compositions in cloud

Benefits of Proposed Research

- Independent of data owner's (source) availability
- Dissemination is based on access control policies and client's attributes:
  - Browser's cryptographic capabilities
  - Authentication method (password- vs. hardware-based vs. fingerprint)
  - Source network (corporate vs. unknown)
  - Type of the device (mobile vs. desktop)
  - Trust level (is continuously monitored)
- Context-based dissemination supported
- Different authentication methods supported
- Ability to operate in untrusted environments
- Reduced host liability for data

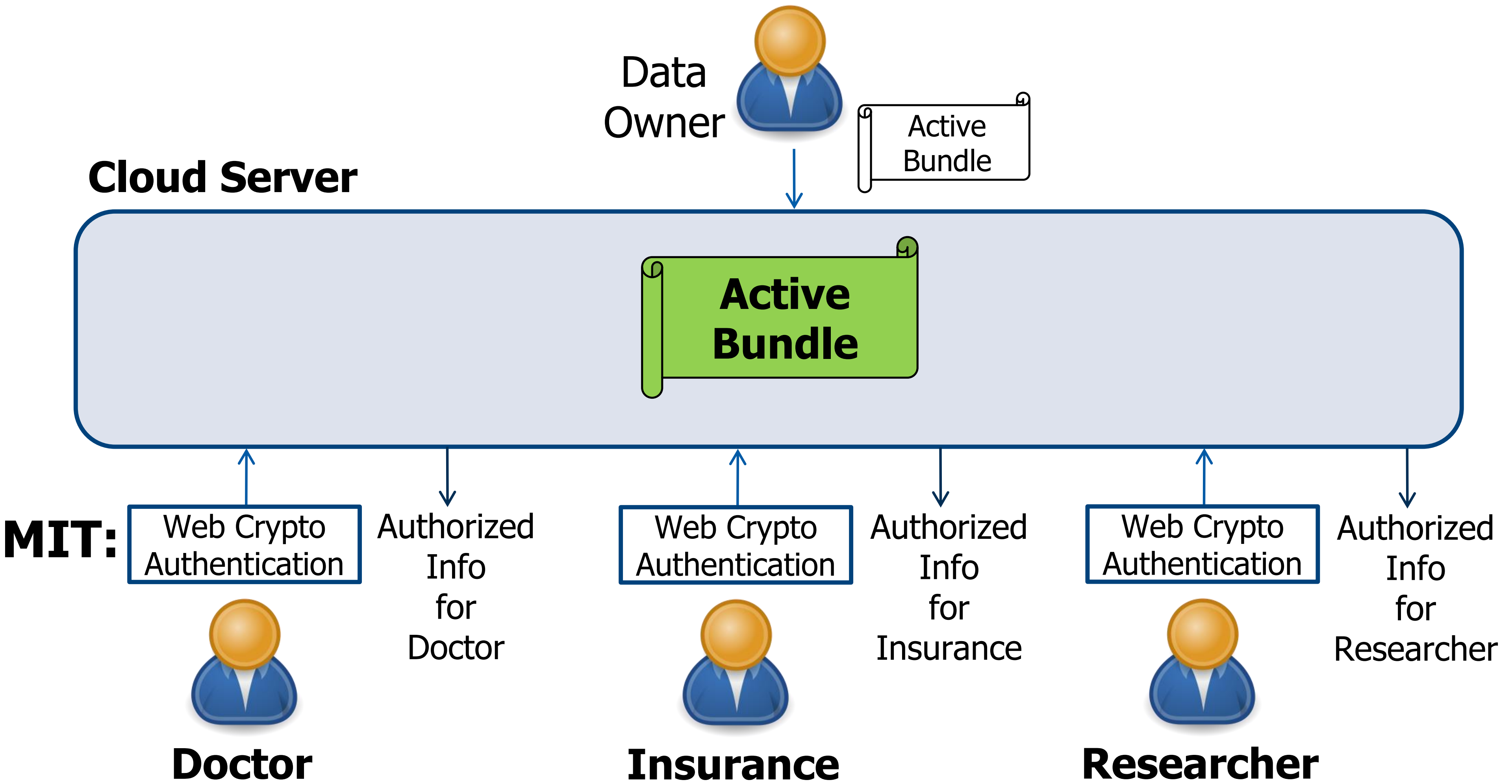
Applications and Demonstration of Prototypes

- Electronic Health Records (EHRs) dissemination in untrusted cloud

- Dynamic service composition and trust management

Prototype for TechFest'16: Electronic Health Record Dissemination in Cloud

NORTHROP GRUMMAN



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

Impact

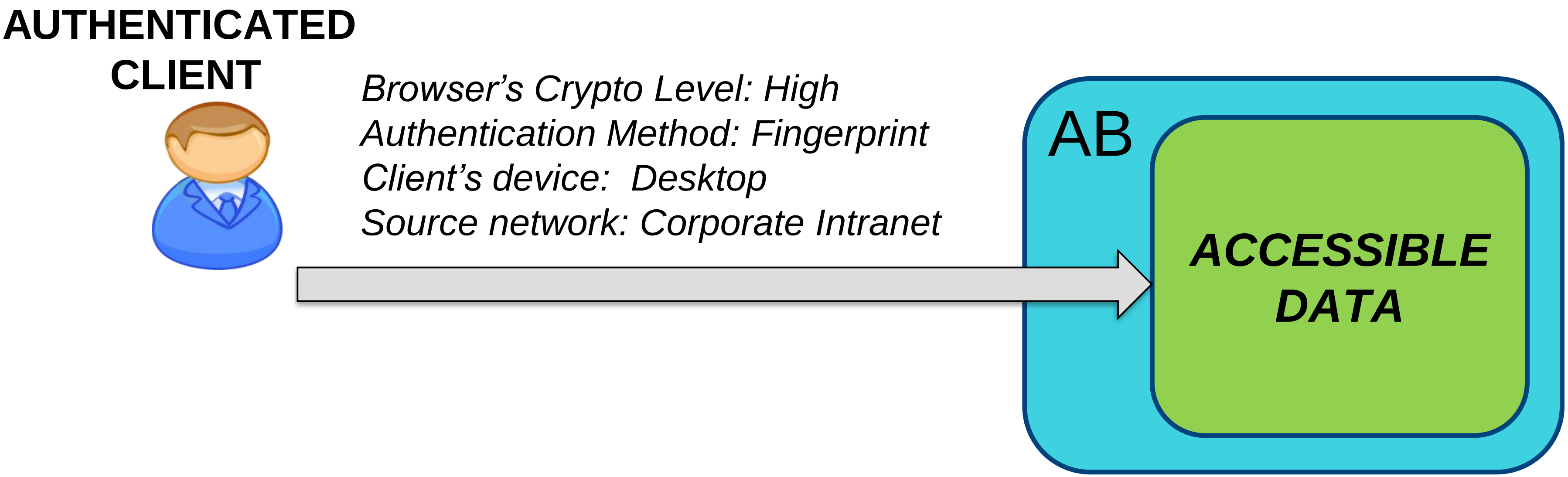
Comprehensive security and privacy auditing and enforcement architecture for trusted and untrusted cloud

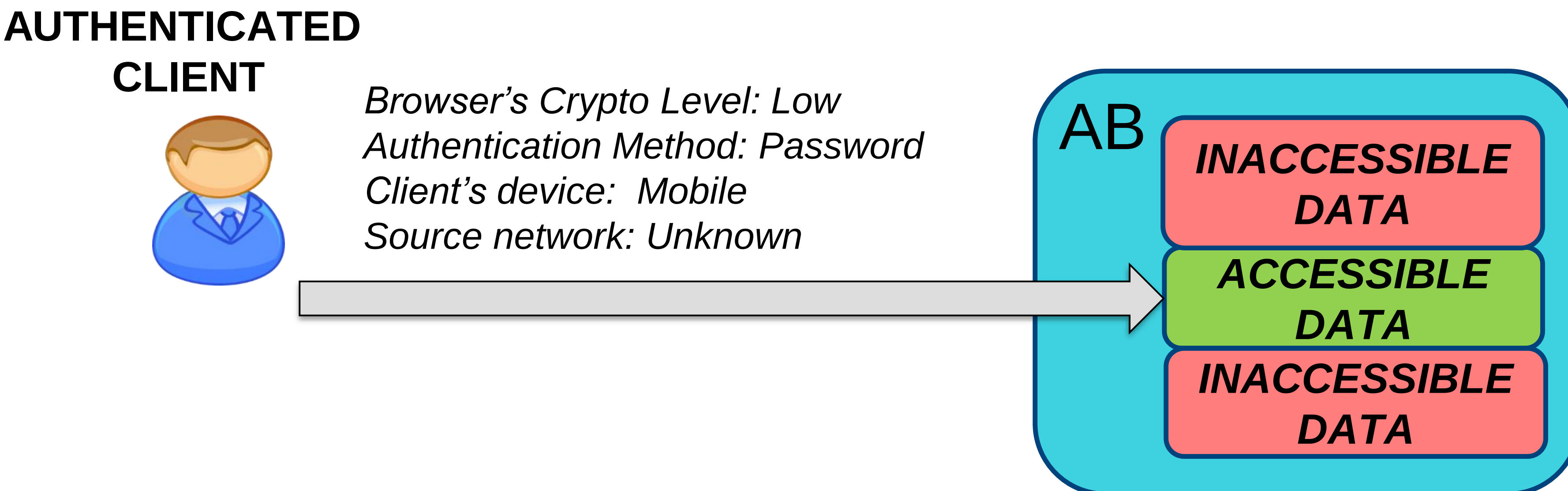
- Privacy-preserving data sharing approach for client-to-service and service-to-service interactions
- Independence of data owner's (source) availability
- Continuous monitoring of SLA and policy compliance
- Swift detection of failures and attacks in the system
- Efficient mechanism to dynamically reconfigure service composition based on the system context/state (failed, attacked, compromised) and resiliency requirements
- Resilient architecture to ensure continuous service availability under failures and attacks
- Compatible with industry-standard SOA/cloud frameworks

State of the Art

- **EnCoRe**: Sticky policies to manage privacy of shared data across domains
  – Prone to TTP related issues
  – Sticky policies vulnerable to attacks from malicious recipients
- **DataSafe**: Software-hardware architecture supporting confidentiality throughout data lifecycle
  – Require special architecture limited to well-known hosts
- **CloudWatch:** Coarse-grain monitoring capabilities of industry-standard cloud systems (such as Amazon EC2)
- **Splunk** (log management and analysis tool), **GrayLog, Kibana**
  – provide storage, search and analysis of big data, but require human intelligence for detection and action for resiliency

2015/2016 Methodology





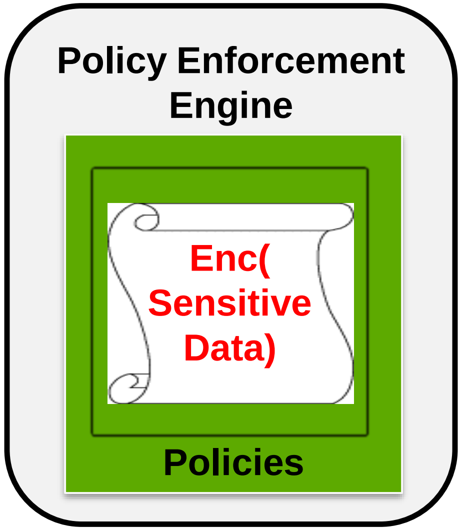
2015/2016 Methodology

NORTHROP GRUMMAN

- Redirect unauthenticated client's request from Cloud Provider to Authentication Server (AS)
- Selective Data dissemination based on:
  - Role-based access control policies
  - Security level of client's browser (crypto capabilities)
  - Authentication method (password-based, fingerprint etc)
  - Source network (secure intranet vs. unknown network)
  - Type of client's device: desktop vs. mobile (detected by Authentication Server)

Active Bundle Solution Components

NORTHROP GRUMMAN



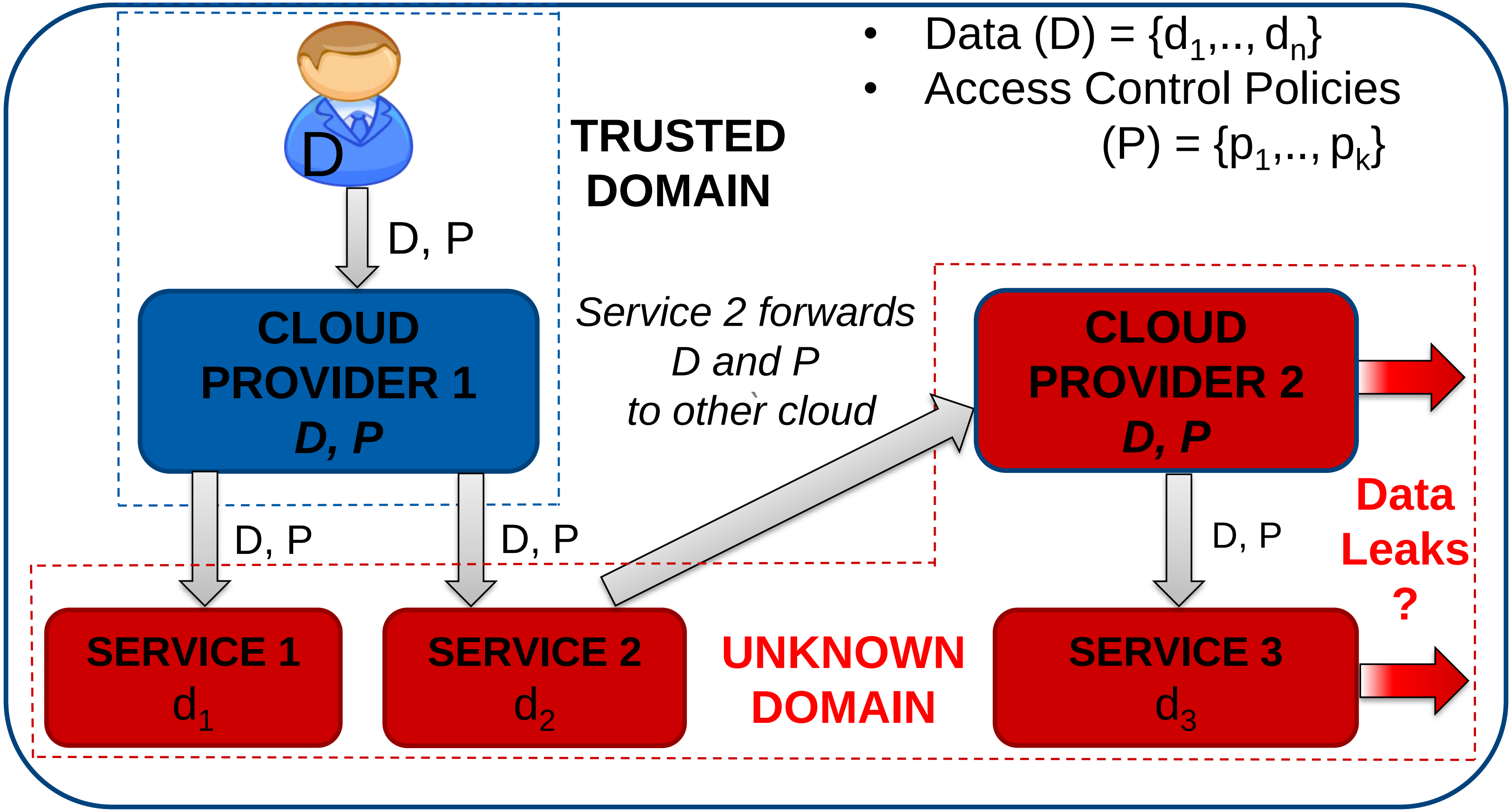
- **Sensitive data**
  - Encrypted data items, each value encrypted with its own key: { “ab.patientPhone” : “Enc(1234567890)” }
- **Metadata**
  - Access control and operational policies
- **Virtual Machine (Policy Enforcement Engine)**
  - Protection mechanism (self-integrity check)
  - Policy evaluation, enforcement; data dissemination

ACTIVE BUNDLE IMPLEMENTATION

- Executable JAR file
- Apache-thrift based API
- JSON-based policies
- WSO2 Balana-based policy engine
- Node.js-based SOA architecture

Data Dissemination and Leakage Detection in Untrusted Cloud

NORTHROP GRUMMAN

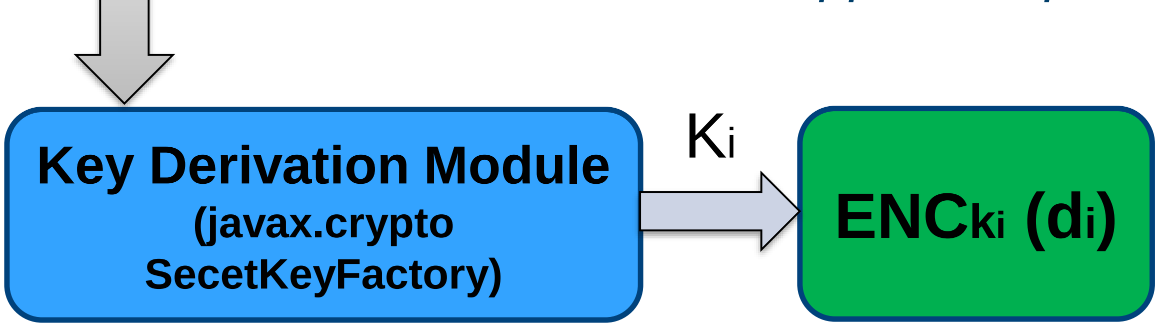


"Authentication of User's Device and Browser for Data Access in Untrusted Cloud," D. Ulybyshev, B. Bhargava, L. Li, J. Kobes, D. Steiner, H. Halpin, B. An, M. Villarreal, R. Ranchal. *CERIAS Security Symposium, April 2016*.

* This work is used in PhD Thesis Proposal of Denis Ulybyshev, Purdue University

Key Generation during AB Creation

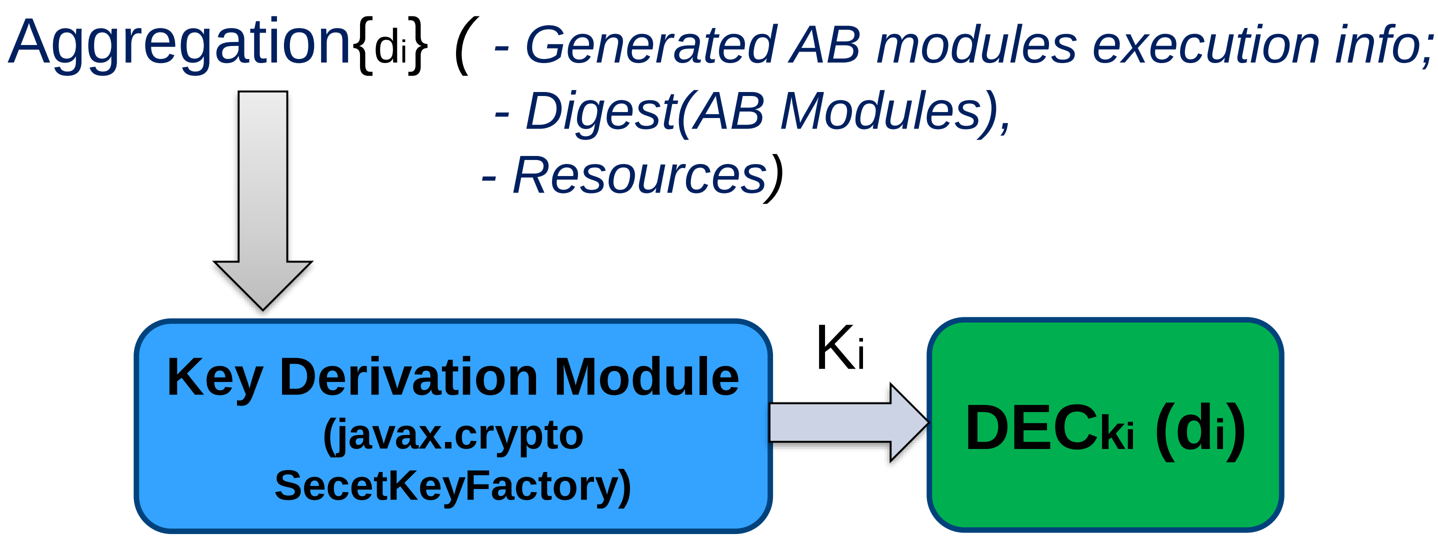
Aggregation{d_i} (*- Generated AB modules execution info;*
- Digest(AB Modules),
- Resources: authentication code + certificate,
authorization code, applicable policies with evaluation code)



- AB Template is used to generate new ABs with data and policies (specified by data owner)
- Template includes implementation of invariant parts (monitor) and placeholders for customized parts (data and policies)
- Template is executed to simulate interaction between AB and service requesting access to each data item of AB

Key Derivation during AB Execution

NORTHROP GRUMMAN



- AB receives access request to data item from service
- AB authenticates the service and authorizes its request
- If any module fails (i.e. service is not authentic or the request is not authorized) or is tampered: derived decryption key K_i is incorrect => data is not decrypted

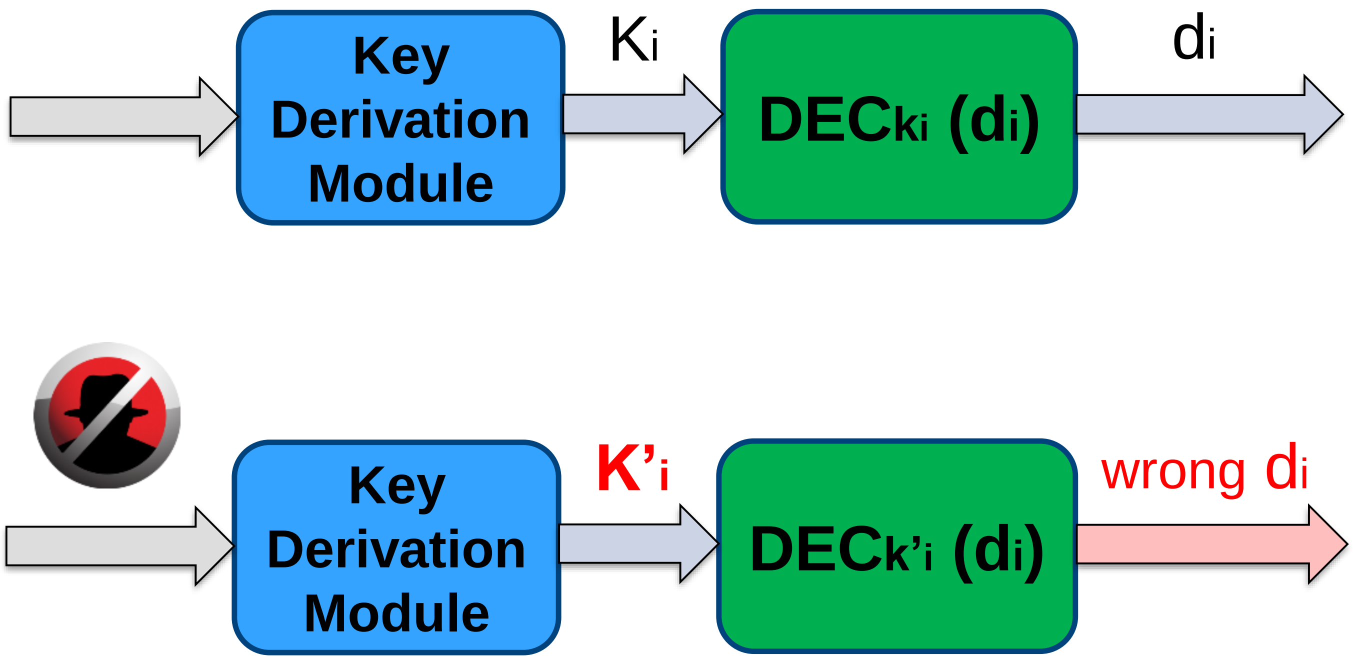
"Cross-Domain Data Dissemination and Policy Enforcement", R. Ranchal, PhD Thesis, Purdue University, Jun. 2015.

Key Management in AB

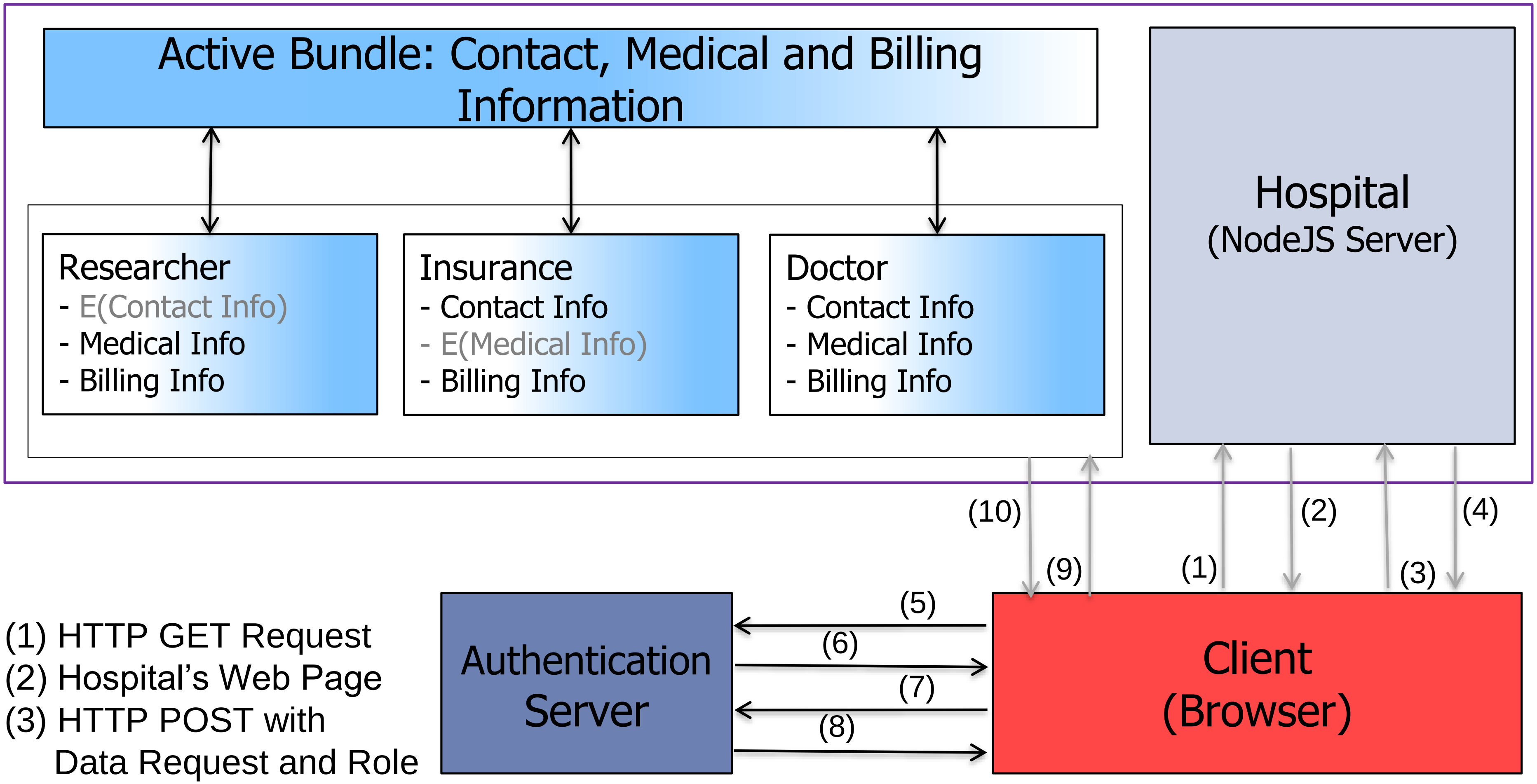
- Key is not stored inside AB
- Separate symmetric key is used for each separate data value
- Ensure protection against tampering attacks (discussed with Jason Kobes)

Aggregation{d_i} (*Execution info; Digest(AB Modules); Resources*)

Aggregation{d_i} (***Tampered (*** *Execution info; Digest(AB Modules); Resources* ***)***)



TechFest'16 Demo: Electronic Health Record Dissemination in Cloud



(1) HTTP GET Request
(2) Hospital's Web Page
(3) HTTP POST with Data Request and Role
(4) HTTP 302 with AB Request and Role
(5) HTTP Get Request
(6) AS Web Page
(7) HTTP POST with Credentials
(8) HTTP 302 with Ticket
(9) HTTP Get Request with Ticket
(10) Data provided by AB

Authentication Ticket Creation and Validation

NORTHROP GRUMMAN

Authentication Server:

- Knows shared secret K and Private Key PrivKey
- Ticket_Info = (Auth_Level, Expiration_Time, Client_ID, Client_Role, Request_Field)
- Enc_Ticket_Info = $EncAES256_K$(Ticket_Info)
- Ticket_Signature = $Enc_{PrivKey}$(SHA512(Enc_Ticket_Info))
- Ticket = <Enc_Ticket_Info, Ticket_Signature>

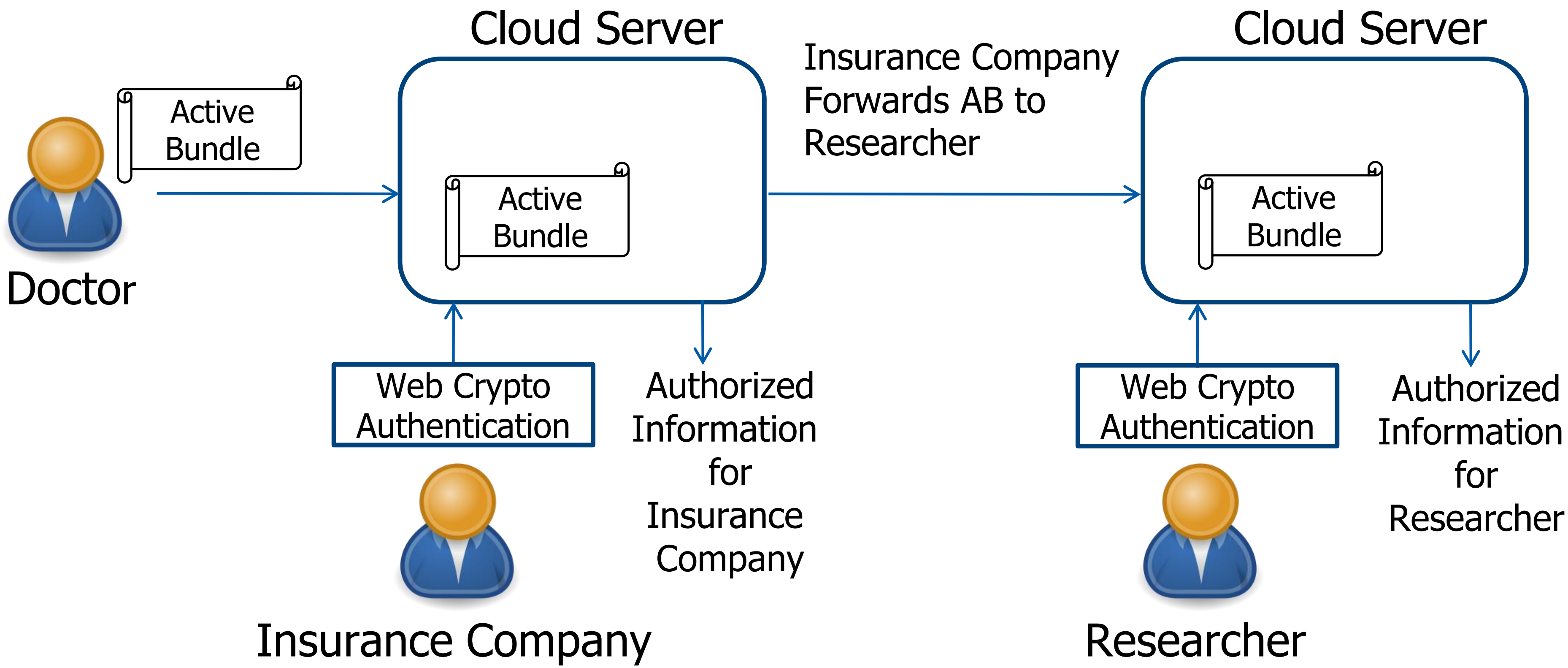
Doctor, Insurance or Researcher Service:

- Knows shared secret K and Public Key PubKey
- Receives Ticket = <Enc_Ticket_Info, Ticket_Signature>
- Checks: Dec_{PubKey}(Ticket_Signature) = SHA512(Enc_Ticket_Info)
- Gets data: $DecAES256_K$(Enc_Ticket_Info)

- Doctor is authorized for Contact, Medical and Billing Info of either "*only her own patients*" or "*her own and other doctor's patients*"
- Clients are directed to the corresponding services according to their role (researcher, insurance or doctor)
- *Authorization level (included in ticket) and role define the data accessible by requester*

Scenario 2: Electronic Health Record Dissemination in Cloud

NORTHROP GRUMMAN



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

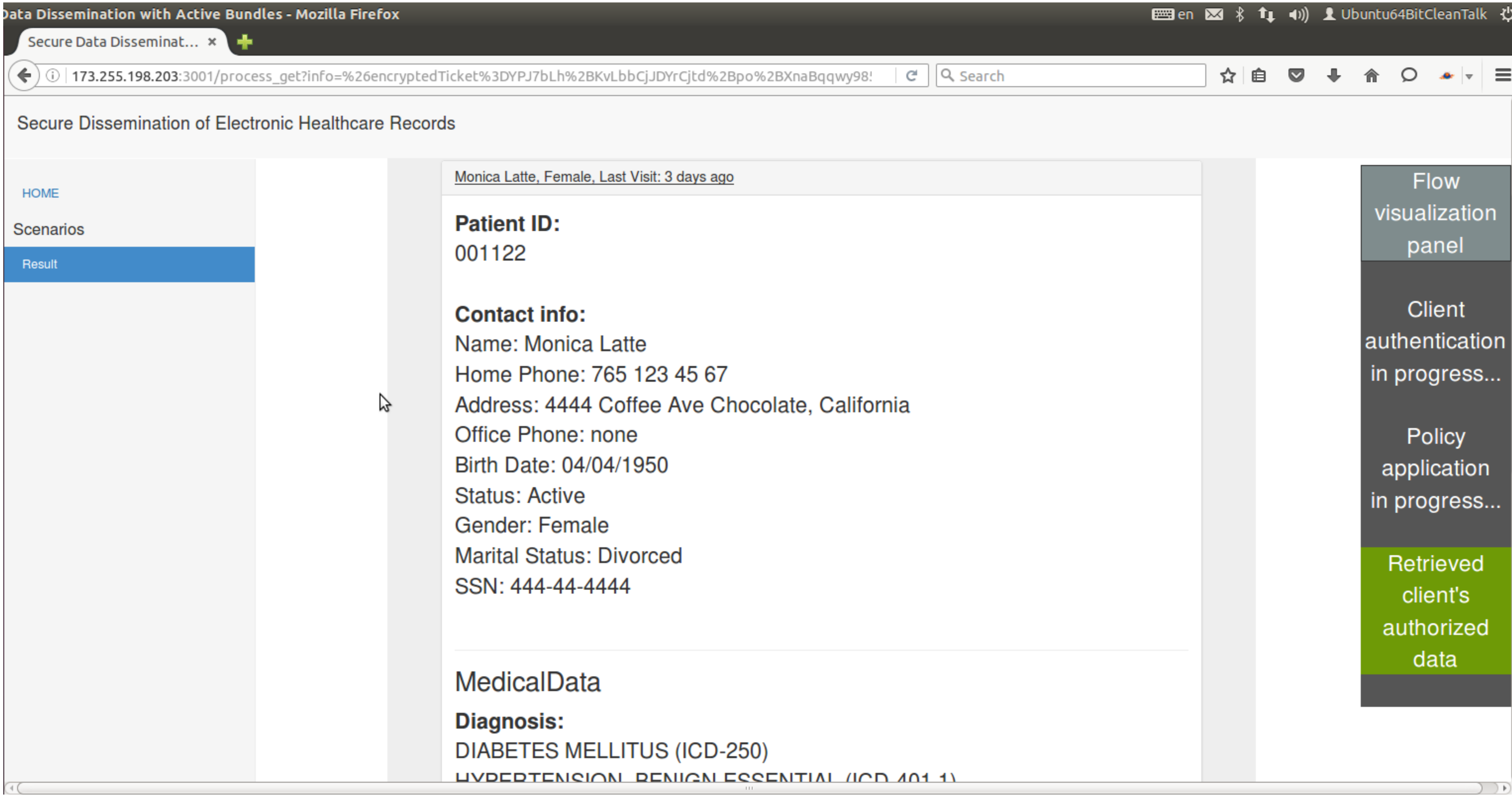
2015/2016 Results Overview

NORTHROP GRUMMAN

- Prototype for EHR dissemination in cloud (collaboration with W3C/MIT)
  - Demonstrated at TechFest 2016

- “Privacy-Preserving Data Dissemination and Data Leakage Detection in SOA” , D. Ulybyshev, B. Bhargava, L. Li, D. Steiner, J. Kobes, H. Halpin, M. Villarreal and R. Ranchal. *Submitted for ICDE-2017*
- “Authentication of User’s Device and Browser for Data Access in Untrusted Cloud,” D. Ulybyshev, B. Bhargava, L. Li, J. Kobes, D. Steiner, H. Halpin, B. An, M. Villarreal, R. Ranchal. *CERIAS Security Symposium, April 2016*.
- "Cross-Domain Data Dissemination and Policy Enforcement", R. Ranchal, PhD Thesis, Purdue University, June 2015.
- “End-to-End Security in Service-Oriented Architecture,” Mehdi Azarmi. *PhD Thesis*, Purdue University, April 2016.

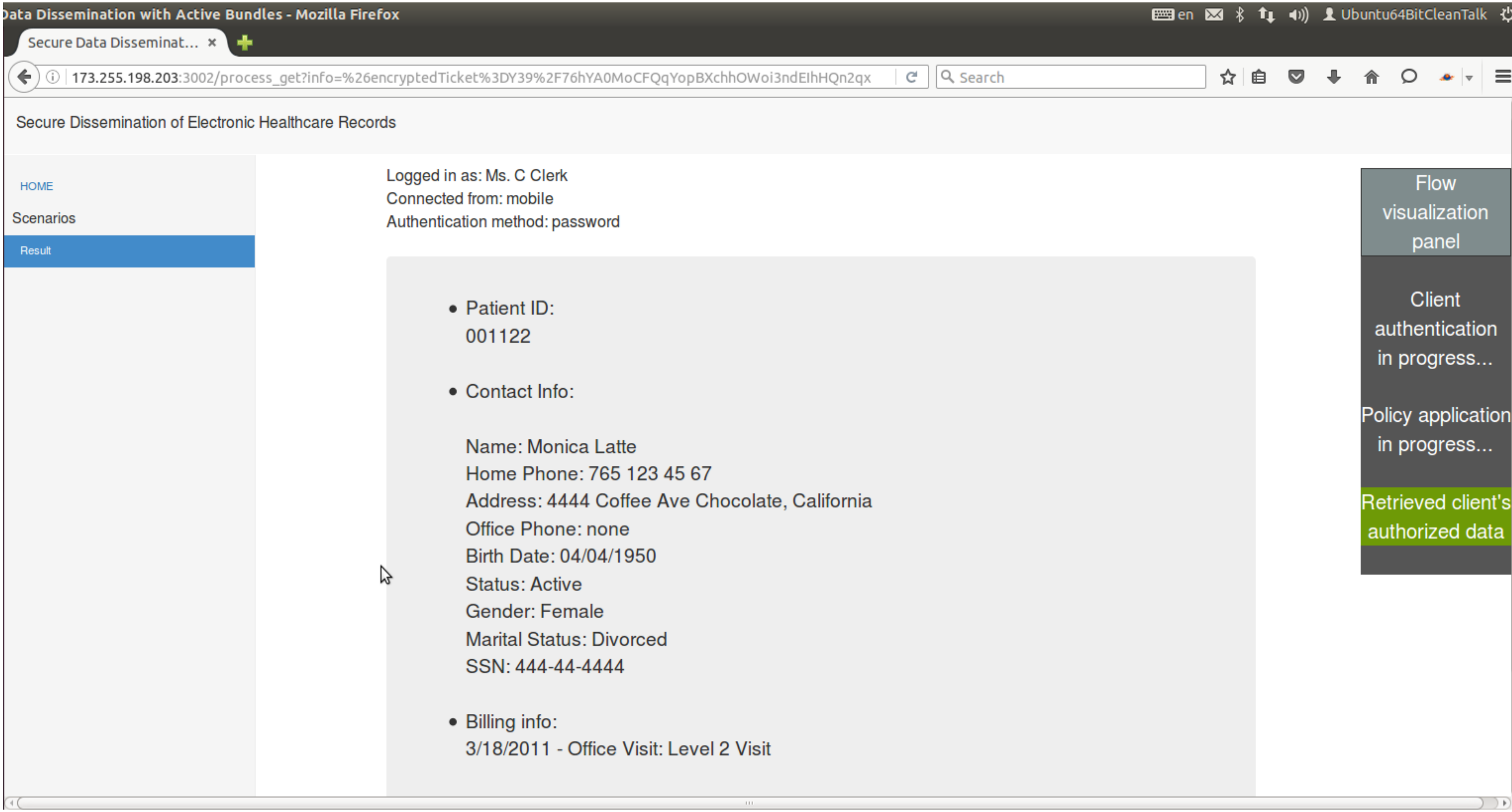
TechFest’16 Scenario: Data Available for Doctor



Doctor can access Contact, Medical and Billing Info of a patient

TechFest'16 Scenario: Data Available for Insurance

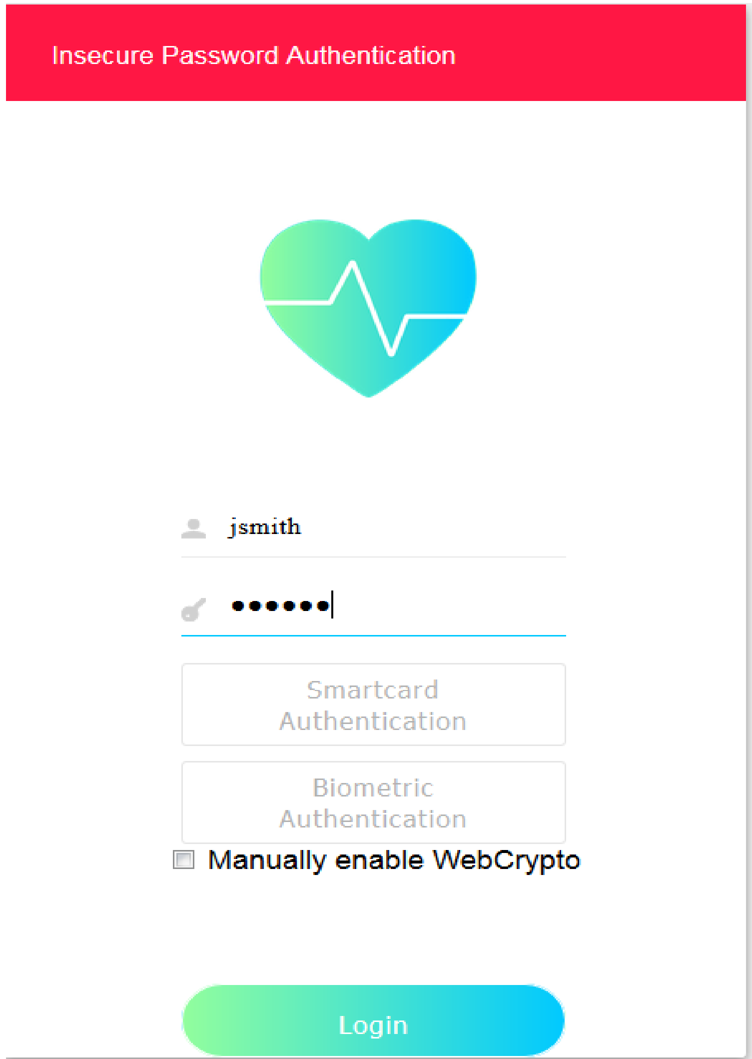
NORTHROP GRUMMAN



Insurance can get access to Contact and Billing Info of a patient (not to Medical)

TechFest'16 Scenario: Client Authentication

NORTHROP GRUMMAN



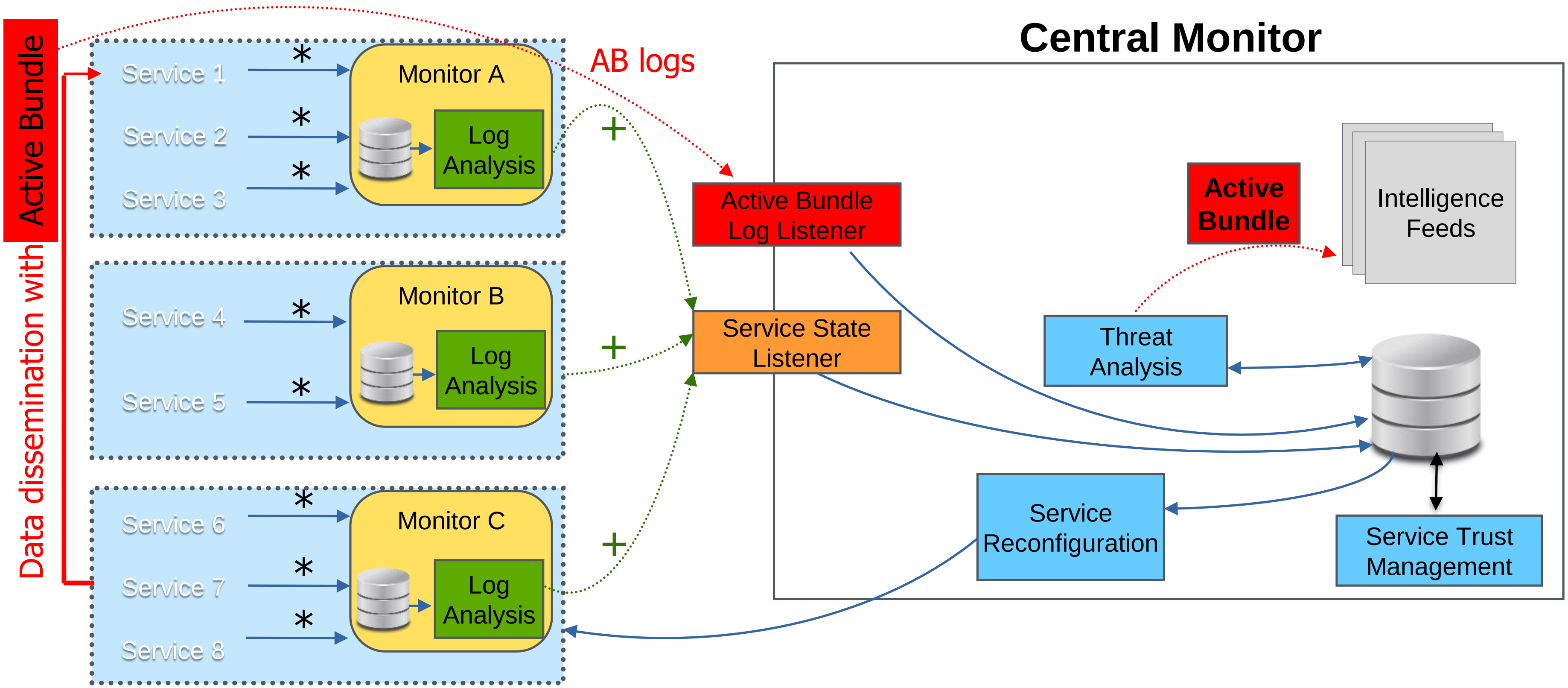
Password-based authentication of a client at Authentication Server

Focus: Adaptable Service Compositions in Cloud

Need systematic monitoring of service operations and data dissemination for:

- **Resiliency** (withstand cyber-attacks, sustain and recover critical function)
- **Antifragility** (increase in resilience and robustness as a result of failures)
- **Adaptability** (swiftly adapt to changes in context, choose services in orchestration to comply with QoS requirements)

System Overview



*: service performance & security parameter values

+: summary service health data

Core Technology & Benefits

- A novel distributed monitoring tool to:
  - Audit and detect service behavior and performance changes*
  - Gather service trust data and share them securely in various domains
  - Dynamically reconfigure service orchestrations based on security context and QoS requirements**

- A secure and adaptable data dissemination technology to deal with:
  - Context changes (e.g. crypto capabilities of web browser, user location/device),
  - Trust
  - Sharing policies of data owner

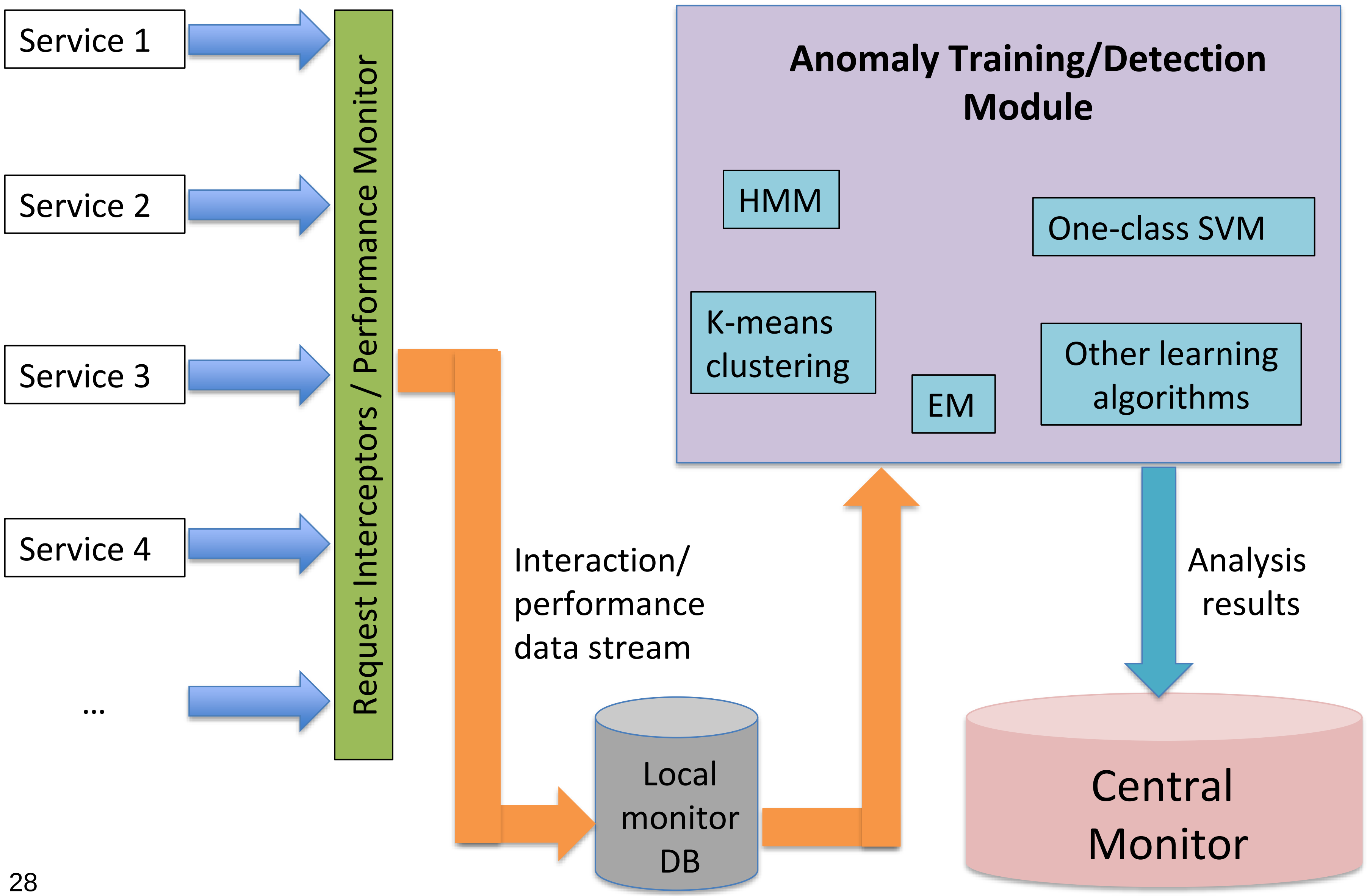
- System modules for *service anomaly detection*, *service performance monitoring* and *trust management* can be easily integrated into NGC cybersecurity software.

- The **modular architecture** and use of **standard software** in the monitoring framework allows for **easy plugin** to any system.

* " A Distributed Monitoring and Reconfiguration Approach for Adaptive Network Computing," Bharat Bhargava, Pelin Angin, Rohit Ranchal, Sunil Lingayat. *DNCMS in conjunction with SRDS 2015* ***(Best paper award)***

** "End-to-End Security in Service-Oriented Architecture," Mehdi Azarmi. *PhD Thesis*, Purdue University, April 2016.

Service Monitoring / Anomaly Detection System Details



Core Machine Learning Technique Used for Service Anomaly Detection in System

NORTHROP GRUMMAN

- No class labels needed for training.
- Training data: input vectors without any corresponding target values
- Unsupervised learning with security and performance parameters used to find clusters
- Values outside clusters will be detected as outliers and help detection of anomalies.
- Algorithms: K-means clustering, one class support vector machines (SVM)

Model training parameters used:

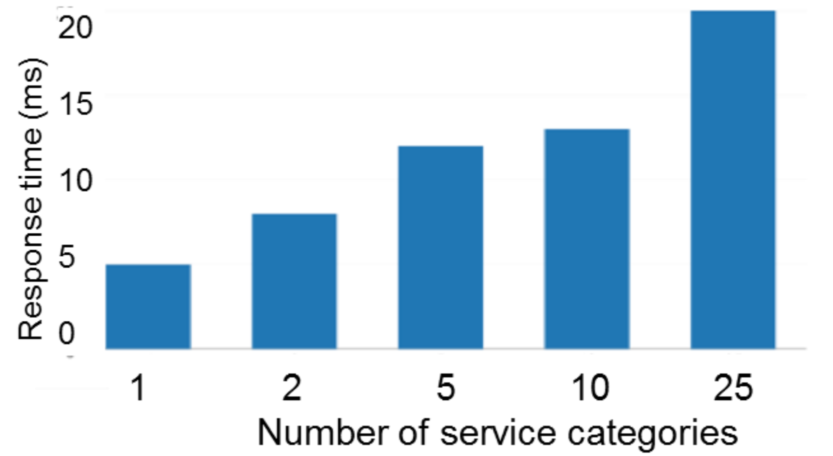
Parameter	Cloud services	Cloud data services
Number of requests/sec	X	
Bytes downloaded/sec	X	X
Bytes uploaded/sec	X	X
Total error rate	X	
CPU utilization	X	
Memory utilization	X	
Number of authentication failures	X	
Number of connections	X	
Number of connection failures	X	
Number of disk reads/writes	X	X
Network latency	X	
Service response time	X	
Disk space usage	X	X
Throughput	X	
Number of database connections	X	X
Service/cluster health status	X	

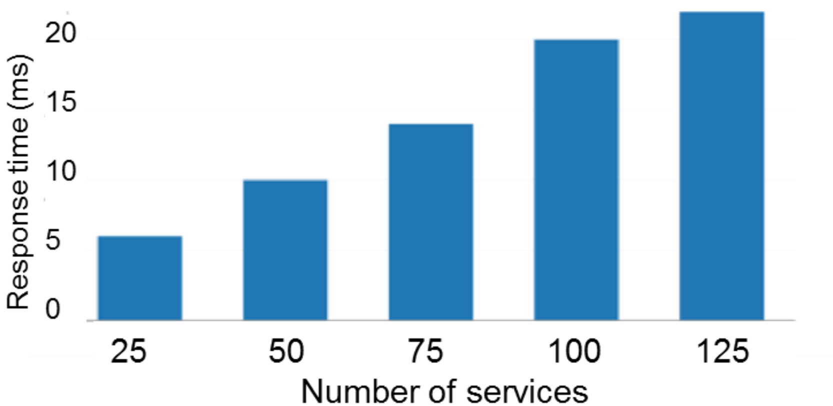
Dynamic Service Composition Experiments

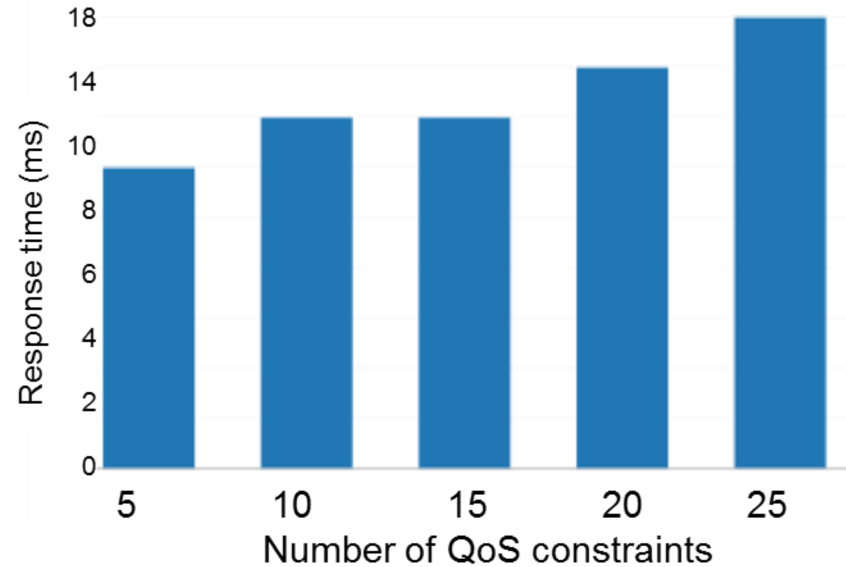
Experiment settings:

Service instance	t2.small (1 vCPU, 2GB memory, and EBS storage)
PE and TM instances	t2.small (1 vCPU, 2GB memory, and EBS storage)
Client instance	t2.micro (1 vCPU, 1GB memory, and EBS storage)
Operating system	Amazon Linux 2015.03 64-bit OS
Geographical region	US-w2 (Oregon region)

- Overhead evalution for three cases:
  - Different number of service categories in composition
  - Different number of services to choose from for each category
  - Different number of QoS contraints
- Composition time not affected significantly by the number of QoS constraints. Number of services and service categories have more visible effect, with still reasonable overhead.







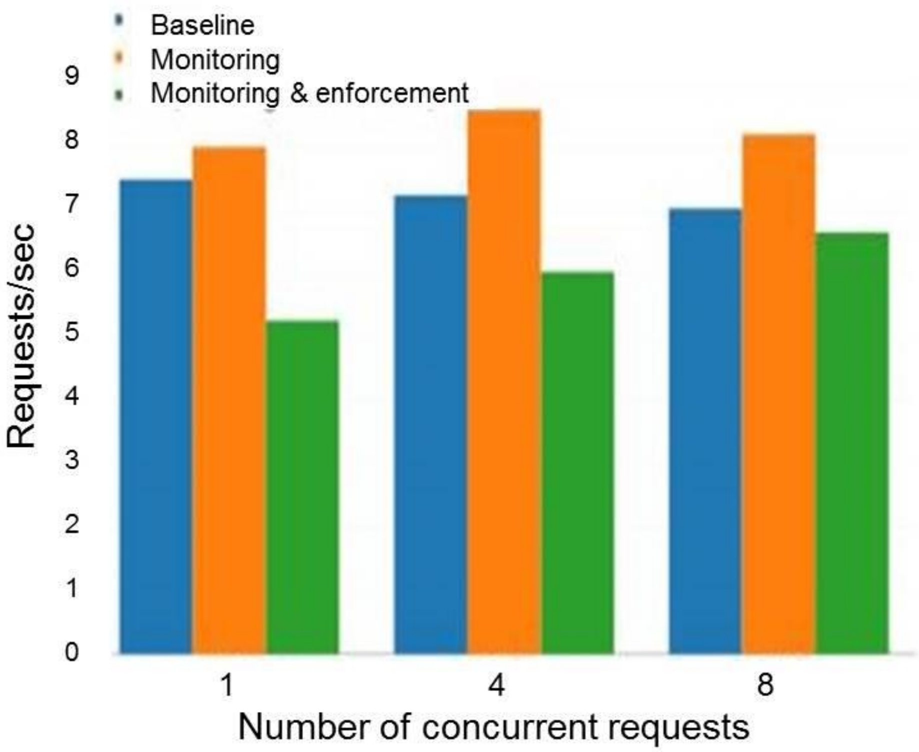
Overhead of Service Monitoring / Request Interception

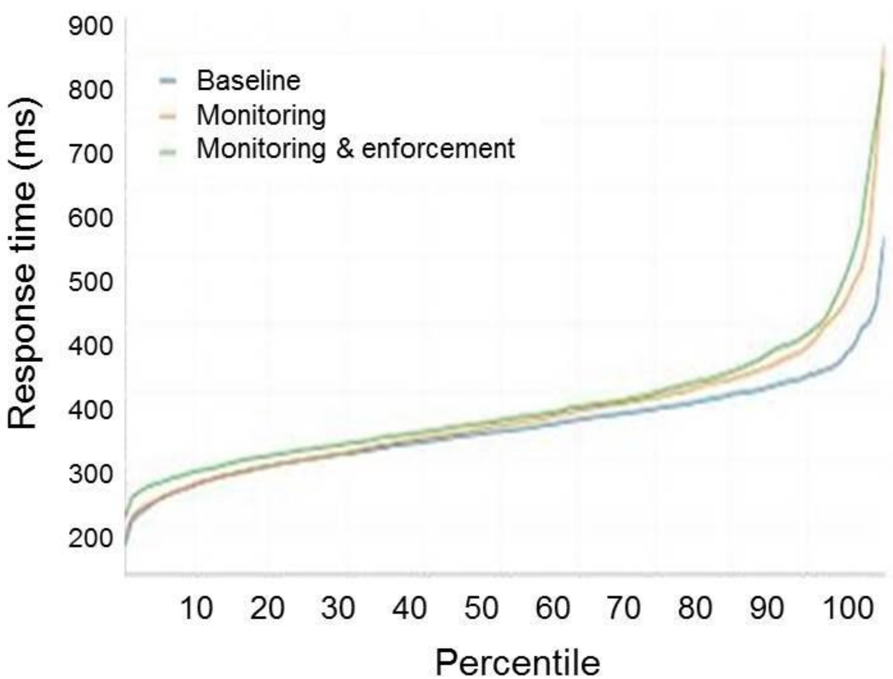
NORTHROP GRUMMAN

- Performance evaluation of service domain in terms of throughput and service response time
- Negligible overhead incurred by monitoring

Experiment settings:

Service instance	t2.small (1 vCPU, 2GB memory, and EBS storage)
PE and TM instances	t2.small (1 vCPU, 2GB memory, and EBS storage)
Client instance	t2.micro (1 vCPU, 1GB memory, and EBS storage)
Operating system	Amazon Linux 2015.03 64-bit OS
Geographical region	US-w2 (Oregon region)





Demonstration

NORTHROP GRUMMAN

- Secure data dissemination in untrusted cloud

https://www.dropbox.com/s/30scw1srqsmyq6d/BhargavaTeam_DemoVideo_Spring16.wmv?dl=0
https://www.youtube.com/watch?v=SIUupq5V6zk&feature=youtu.be

- Dynamic service composition and trust management
  - Composite trust algorithms
    https://www.youtube.com/watch?v=6uHEfoxjEgs
  - Trust update mechanisms
    https://www.youtube.com/watch?v=xnm0-MzGBzw
  - Policy enforcement in service interactions
    https://www.youtube.com/watch?v=ePtAM0N7jdY
  - Service redirection
    https://www.youtube.com/watch?v=e8xkCgZcQ1s
  - Adaptive and secure service composition
    https://www.youtube.com/watch?v=VQDbPD2q9-8

Deliverables

- **Prototype implementation of monitoring framework:**
  - Active Bundle Module
    - AB implementation as an executable JAR file
    - AB API implementation using Apache Thrift RPC framework
    - Policy specification in XACML/JSON and evaluation using WSO2 Balana
    - Healthcare scenario with services running on a remote host with functionality to communicate with remote Authentication Server
  - Local (domain-level) Service Monitor (Apache Axis2 valves for interception, MySQL database for logging)
  - Central Monitor (as Web service on Amazon EC2)
  - Anomaly Detection Module (with pluggable algorithms)
  - Dynamic Service Composition Module (algorithm)

Deliverables (cont.)

- **Documentation**
  - Source code
    http://github.com/Denis-Ulybysh/absoa16
  - Deployment and user manual
  - Final report for 2015 – 2016
    https://www.cs.purdue.edu/homes/bb/NGC2016/NGCRC_Final_Report-Sept-2016.pdf

Presentations and Publications

- “Policy-based Distributed Data Dissemination,” R. Ranchal, D. Ulybyshev, P. Angin, and B. Bhargava. *CERIAS Security Symposium, April 2015* ***(Best poster award)***
- “A Distributed Monitoring and Reconfiguration Approach for Adaptive Network Computing,” B. Bhargava, P. Angin, R. Ranchal, S. Lingayat. *DNCMS in conjunction with SRDS 2015* ***(Best paper award)***
- “Privacy-Preserving Data Dissemination and Data Leakage Detection in SOA” , D. Ulybyshev, B. Bhargava, L. Li, D. Steiner, J. Kobes, H. Halpin, M. Villarreal and R. Ranchal. *Submitted for ICDE-2017*
- “Authentication of User’s Device and Browser for Data Access in Untrusted Cloud,” D. Ulybyshev, B. Bhargava, L. Li, J. Kobes, D. Steiner, H. Halpin, B. An, M. Villarreal, R. Ranchal. *CERIAS Security Symposium, April 2016*.
- "Cross-Domain Data Dissemination and Policy Enforcement", R. Ranchal, PhD Thesis, Purdue University, June 2015.
- “End-to-End Security in Service-Oriented Architecture,” Mehdi Azarmi. *PhD Thesis*, Purdue University, April 2016.
- “Consumer Oriented Privacy Preserving Access Control for Electronic Health Records in the Cloud,” R. Fernando, R. Ranchal. B. An, L. Ben Othmane, B. Bhargava. Submitted to *IEEE CLOUD 2016*.
- “A Self-Cloning Agents-based Model for High Performance Mobile-Cloud Computing,” P. Angin, B. Bhargava, and Z. Jin. *IEEE CLOUD 2015*.

CYBERSECURITY RESEARCH Consortium

Back-up Slides

Active Bundle Experiments

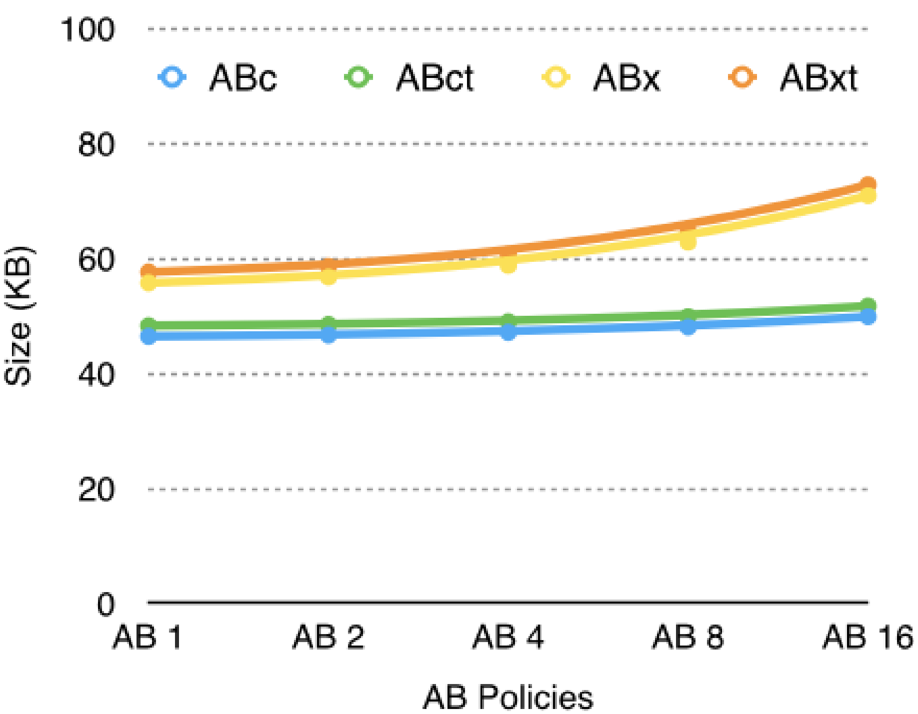
NORTHROP GRUMMAN

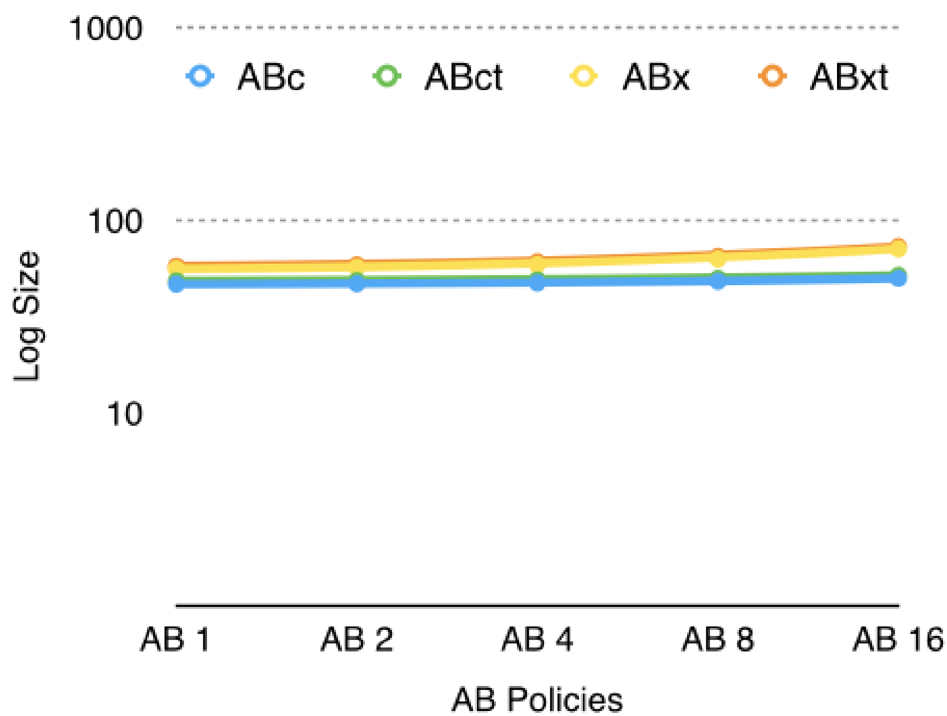
- Measurements
  - Experiment 1: Growth in AB size with increase in the number of policies
  - Experiment 2: Growth in AB and Service interaction time with increase in # of policies
  - Experiment 3: Tamper Resistance overhead in AB execution
- Variations
  - AB versions
    - ABx – XACML-based policies 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

* Experiments were conducted by Rohit Ranchal

Experiment 1: AB Size vs. Number of policies

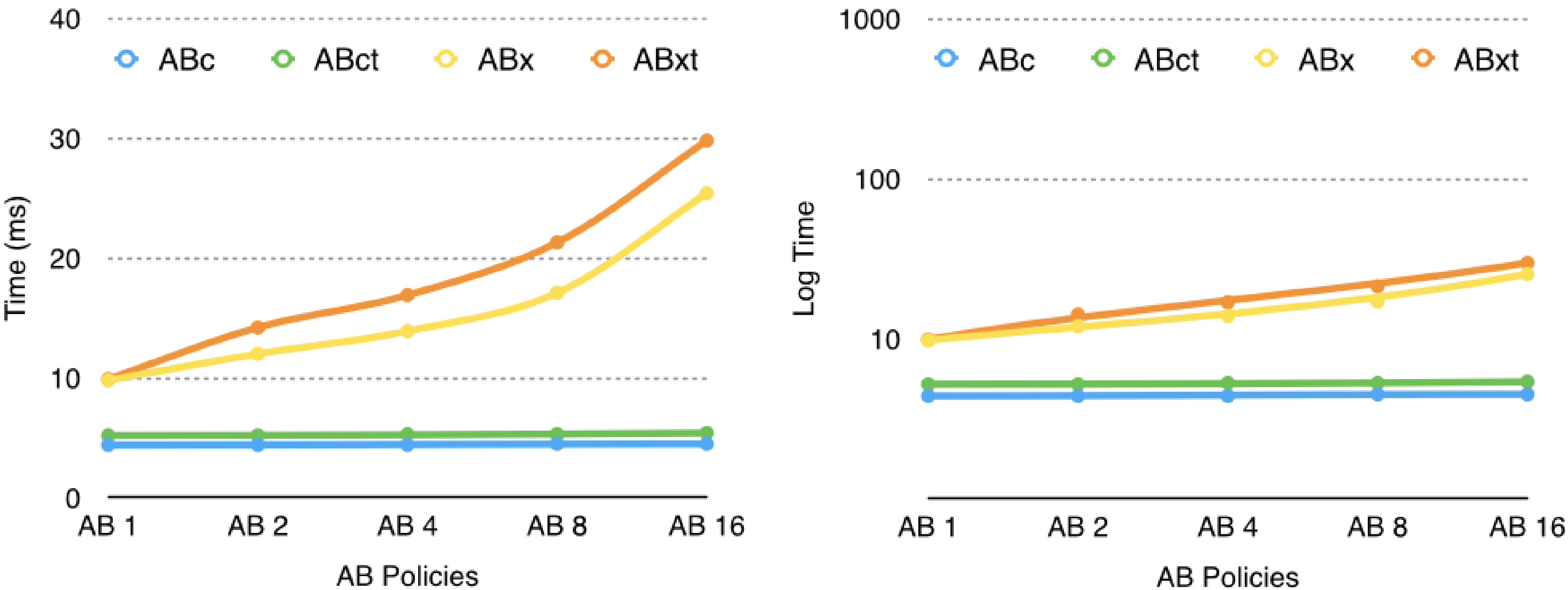
NORTHROP GRUMMAN





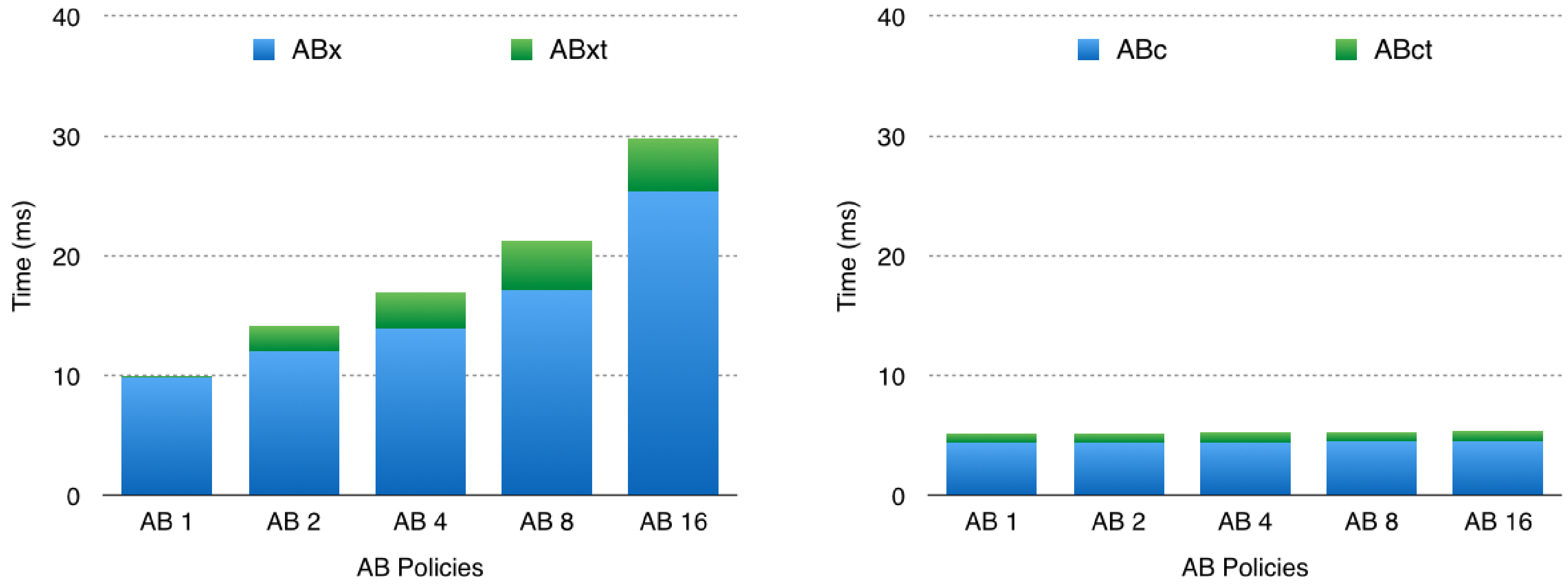
- Observations
  - Linear growth in AB size with increase in number of policies for all versions
  - Tamper resistance adds a slight overhead to AB size (< 2 KB)
  - 79% reduction in policy size (0.79 KB) with JSON-based policies
    - Additional reduction of 8.5 KB with Java-based policy engine

Experiment 2: AB-Service Interaction Time vs. Number of policies



- Observations
  - Linear growth in interaction time with increase in policies for ABx and ABxt
    - Use of XACML-based policies and external library (WSO2 Balana) for policy evaluation
    - Evaluation of XACML policies involve the traversal of XML policy and request trees
  - Constant growth in interaction time with increase in policies for ABc and ABct
    - Use of JSON-based policies and Java code for policy evaluation
    - Highly optimized Java code evaluation

Experiment 3: Tamper Resistance Overhead



- 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

Isolated AB Execution

- Recipient may be reluctant to execute AB => we support the isolated execution of AB by means of Docker.
  – Docker is based on Linux container which is light-weight virtual machine
  – When AB arrives at recipient machine, one virtual machine is created and AB is copied into that virtual machine.
  – AB can be executed inside virtual machine. Only the result returns to host machine

Key Generation during AB Creation

NORTHROP GRUMMAN

- An AB Template is used to generate new ABs with data and policies specified by a user
  - An AB Template includes the implementation of the invariant parts (monitor) and placeholders for customized parts (data and policies)
- User specified data and policies are included in the AB Template
- AB Template is executed to simulate the interaction process between an AB and a service requesting access to each data item of AB
- The information generated during the execution of different AB modules and the digests of these modules and their resources (such as authentication (authentication code, CA certificate that it uses), authorization (authorization code, applicable policies, policy evaluation code)) are collected and aggregated into a single value for each data item
- The value for each data item is input into a Key Derivation module (such as SecretKeyFactory, PBEKeySpec, SecretKeySpec provided by javax.crypto library)
- The Key Derivation module outputs the specific key relevant to the data item
- This key is used encrypt the related data item

Key Derivation during AB Execution

NORTHROP GRUMMAN

- AB receives access request to a data item from a service
- AB authenticates the service and authorizes its request
- The information generated during the execution of different AB modules and the digests of these modules and their resources (such as authentication (authentication code, CA certificate that it uses), authorization (authorization code, applicable policies, policy evaluation code)) are collected and aggregated into a single value for each data item
- The value for each data item is input into the Key Derivation module (such as SecretKeyFactory, PBEKeySpec, SecretKeySpec provided by javax.crypto library)
- The Key Derivation module outputs the specific key relevant to the data item
- This key is used decrypt the requested data item
- If any module fails (i.e. service is not authentic or the request is not authorized) or is tampered, the derived is incorrect and the data is not decrypted

Context-Sensitive Data Disclosure

- Perfect data dissemination not always desirable
  - Example: Confidential business data shared within an office but not outside
- Context-sensitive AB evaporation
  - AB evaporates in proportion to their “distance” from their owner
- “Closer” subscribers trusted more than “distant” ones
- Illegitimate disclosures more probable at less trusted “distant” guardians
- Different distance metrics
  - Context-dependent