MONDRAGON UNIBERTSITATEA
GOI ESKOLA POLITEKNIKOA
FACULTY OF ENGINEERING

Cifrado CP-ABE para el ecosistema Apache Hadoop

– Pinceladas sobre un trabajo en desarrollo –

Aitor Osa, Iñaki Garitano, Iñaki Arenaza,
Urko Zurutuza, Mikel Iturbe

JNIC2017

Presentación

– Sistemas Inteligentes para Sistemas Industriales –

- Nos dedicamos
  - Seguridad informática
    - Sobre todo relacionado con el mundo industrial
  - Análisis de datos
- En este trabajo participamos


Aitor Osa


Oier Saizar


Iñaki Garitano


Iñaki Arenaza


Urko Zurutuza


Mikel Iturbe

Antecedentes

– Sistemas Inteligentes para Sistemas Industriales –

- Comienzo -> curso 2016/2017

- Pasos
  1. Análisis de la seguridad de Apache Hadoop (done)
  2. Análisis de CP-ABE (done)
  3. Diseño de integración de CP-ABE sobre Apache Hadoop y Apache Ranger (doing)
  4. Integración de CP-ABE sobre Apache Hadoop y Apache Ranger (todo)

Estado actual

– Sistemas Inteligentes para Sistemas Industriales –

- Diseñando la integración de CP-ABE sobre Apache Hadoop y Apache Ranger

Estado actual

– Sistemas Inteligentes para Sistemas Industriales –

- Diseñando la integración de CP-ABE sobre Apache Hadoop y Apache Ranger
  - Problema
    - Apache Hadoop >= 3M líneas de código
    - Apache Ranger >= 700K líneas de código

- Por poner en contexto
  - Versión 2.4 del kernel de Linux <= 2.5M líneas de código
  - Windows NT 3.1 -> entre 4 y 5M de líneas de código

Índice

1. Decapando Apache Hadoop
2. Un poco de color sobre Apache Ranger
3. Últimas pinceladas
   - CP-ABE

Índice

1. Decapando Apache Hadoop
2. Un poco de color sobre Apache Ranger
3. Últimas pinceladas
   - CP-ABE

Mecanismos de Seguridad
– Decapando Apache Hadoop –

- Apache Hadoop
  - es
    - framework de software para el almacenamiento y procesamiento distribuido de grandes conjuntos de datos
  - puede
    - Trabajar en “modo seguro”
      - “modo seguro” implica
        - Autenticación mediante Kerberos tanto los usuarios, como los servicios y acceso web
        - Confidencialidad de los datos mediante cifrado durante la transmisión (TLS)
    - Emplear ACLs para controlar el acceso a los servicios
    - Utilizar *Transparent Data Encryption (TDE)*
    - Aislar la ejecución para los contenedores de YARN
    - …

Transparent Data Encryption
– Decapando Apache Hadoop –

- Características:
  - Permite almacenar los datos cifrados en disco
  - Transparente para el usuario y las aplicaciones
  - HDFS no tiene acceso a los datos descifrados
    - *At-rest encryption*
      - mientras los datos se almacenan en el disco
    - *In-transit encryption*
      - mientras los datos se envían por la red

Índice

1. Decapando Apache Hadoop
2. Un poco de color sobre Apache Ranger
3. Últimas pinceladas
   - CP-ABE

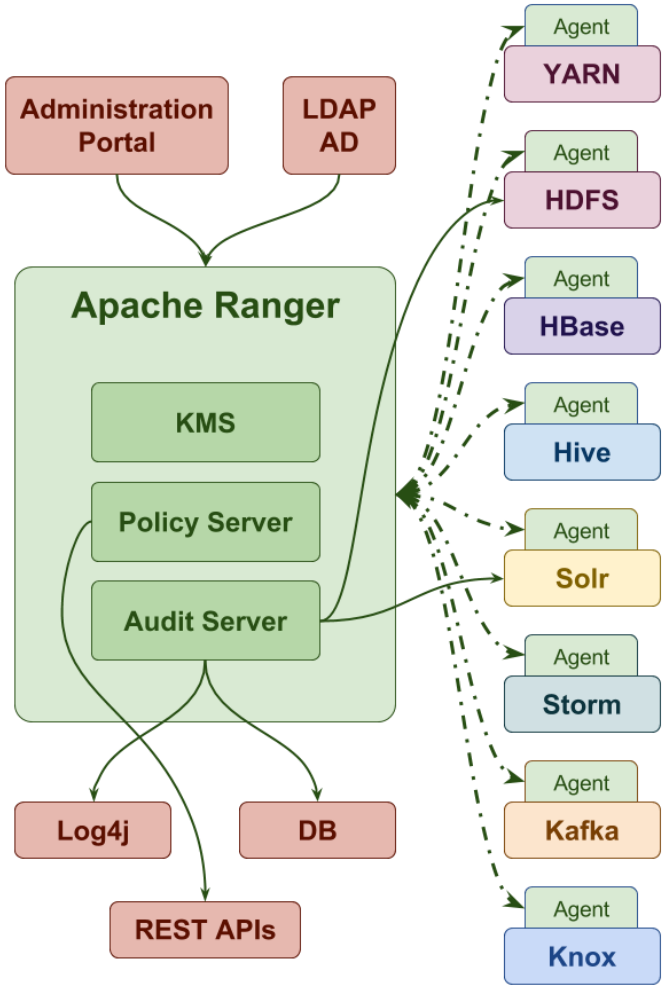
Descripción

– Un poco de color sobre Apache Ranger –

- Apache Ranger es un framework de seguridad centralizada
- Permite centralizar la administración de la seguridad mediante el uso de una interfaz centralizada o un API REST
- Ofrece autorización de grano fino mediante diferentes métodos de autorización

Arquitectura

– Un poco de color sobre Apache Ranger –



- Apache Ranger
  - Se compone de:
    - Key Management Server
    - Policy Server
    - Audit Server
    - Diversos agentes para cada uno de los componentes
  - Ofrece:
    - Un portal de administración
    - Un API Rest
  - Permite
    - Generar logs en
      - Log4j
      - Base de datos

Índice

1. Decapando Apache Hadoop
2. Un poco de color sobre Apache Ranger
3. Últimas pinceladas
   - CP-ABE

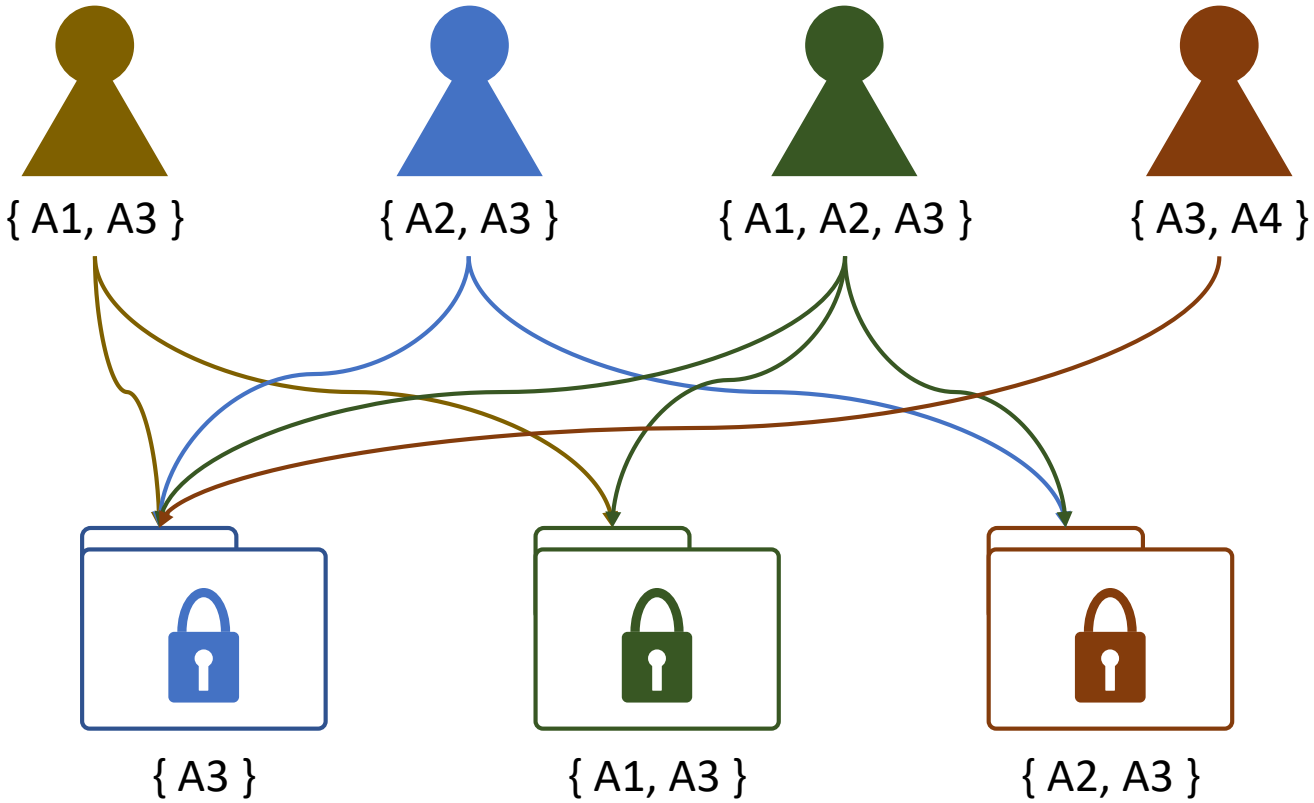
Ciphertext-Policy Attribute-Based Encryption

– Últimas pinceladas –

- Trabajo realizado por John Bethencourt (Carnegie Mellon University), Amit Sahai (UCLA) y Brent Waters (SRI International)

- Es un sistema de control de acceso cifrado donde las claves privadas de los usuarios se especifican en base a atributos
- Así mismo, los usuarios capaces de descifrar se especifican en base a atributos
- Este sistema es capaz de aguantar ataques de conspiración, *collusion*, donde un atacante obtenga más de una clave privada

CP-ABE – ejemplo sencillo
– Últimas pinceladas –



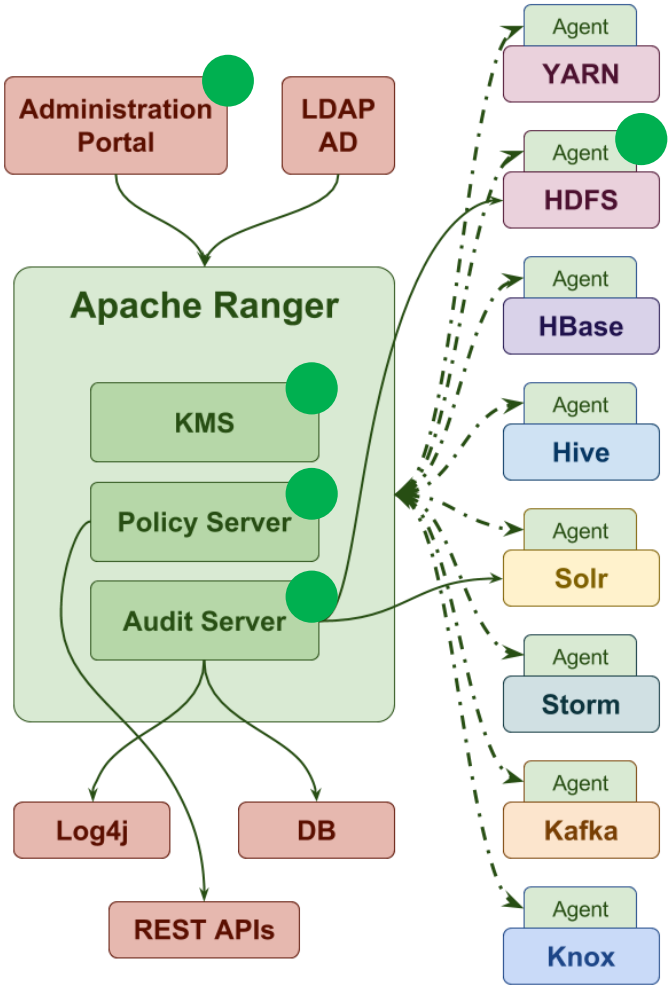
Situación actual
– Últimas pinceladas –

Situación actual
– Últimas pinceladas –

- Java Native Interface (JNI)
  - Nos permite interactuar con funciones escritas en otros lenguajes
- Diseñando la gestión de las claves de cifrado y descifrado
  - Tanto a nivel de Apache Hadoop,
    - Apache Hadoop utiliza el algoritmo AES -> claves simétricas
  - como a nivel de Apache Ranger

CP-ABE & Apache Ranger

– Últimas pinceladas –



- El sistema permitirá definir atributos
- A los usuarios se les podrá asignar uno o varios atributos

● Partes donde se está actuando

CP-ABE – Bibliografía
– Últimas pinceladas –

- Para más información:
  - http://ieeexplore.ieee.org/document/4223236/

- Software disponible en:
  - http://hms.isi.jhu.edu/acsc/cpabe/

Muchas gracias

Iñaki Garitano
Mondragon Unibertsitatea
igaritano@mondragon.edu