An Attribute-Based Access Control Model for Secure Big Data Processing in Hadoop Ecosystem

Maanak Gupta, Farhan Patwa, and Ravi Sandhu

Institute for Cyber Security,
Center for Security and Privacy Enhanced Cloud Computing,
Department of Computer Science
University of Texas at San Antonio

3rd ACM Workshop on Attribute-Based Access Control (ABAC'18)
Tempe, Arizona, USA, March 19-21, 2018

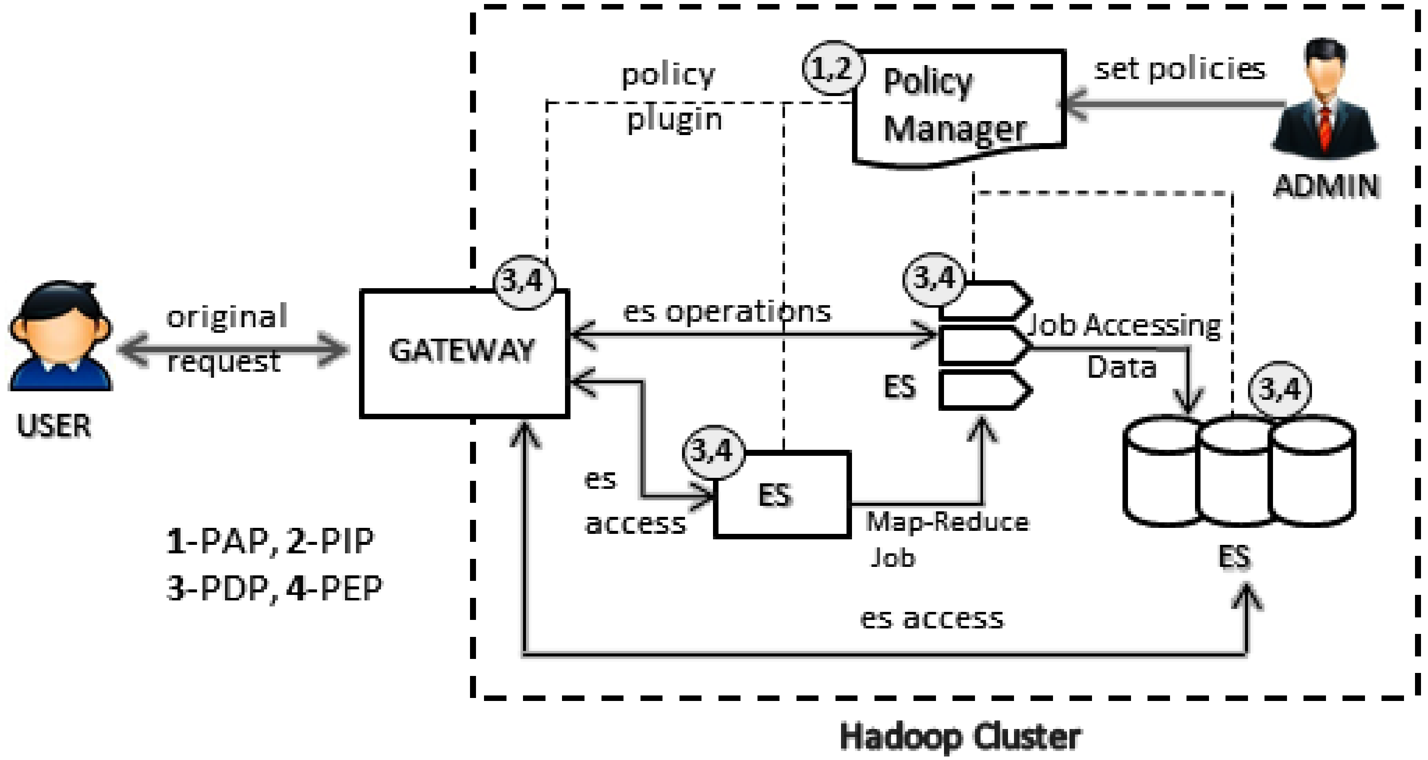
Outline

- Introduction and Motivation
- Multi-layer Hadoop Authorization Framework
- Object Tagged - RBAC Model
- HeABAC Model
- Implementation Approach
- Use Case

Big Data and Big Challenges

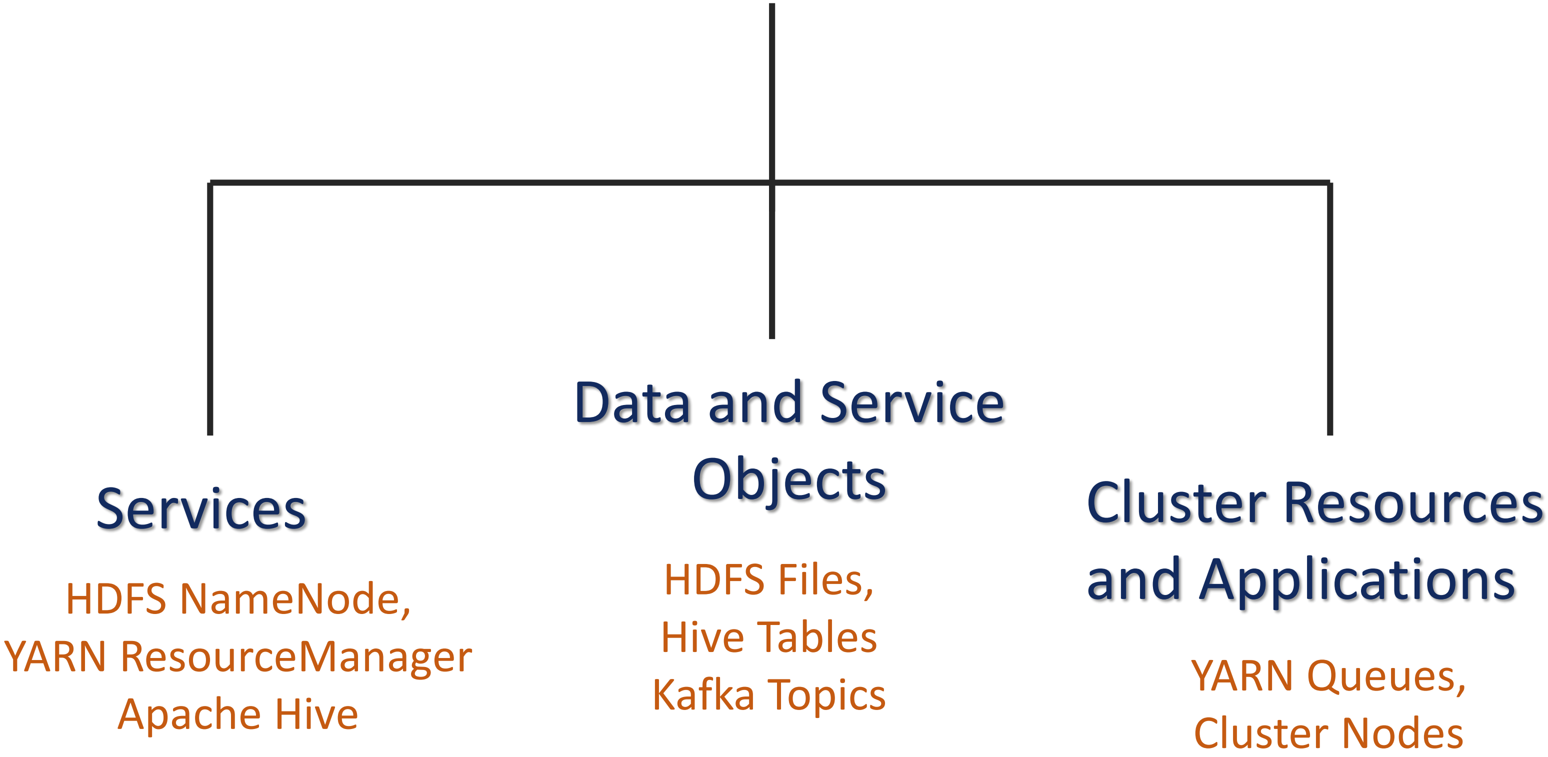
- IDC 2025 :
  - global "Datasphere" – 163 zettabytes
  - 10x than 2016
- Security:
  - Privacy Concerns (eg: HIPPA)
  - Fine granular access requirements
- Hadoop Ecosystem = Hadoop core + Open-Source Projects
- Hadoop Data Lake

Hadoop Ecosystem Authorization Architecture



Policy Manager : Apache Ranger, Apache Sentry
Gateway : Apache Knox
Ecosystem Service (ES) : Apache Hive, HDFS, Apache Storm, Apache Kafka, YARN

Multi-Layer Access Control



Services

HDFS NameNode,
YARN ResourceManager
Apache Hive

Data and Service Objects

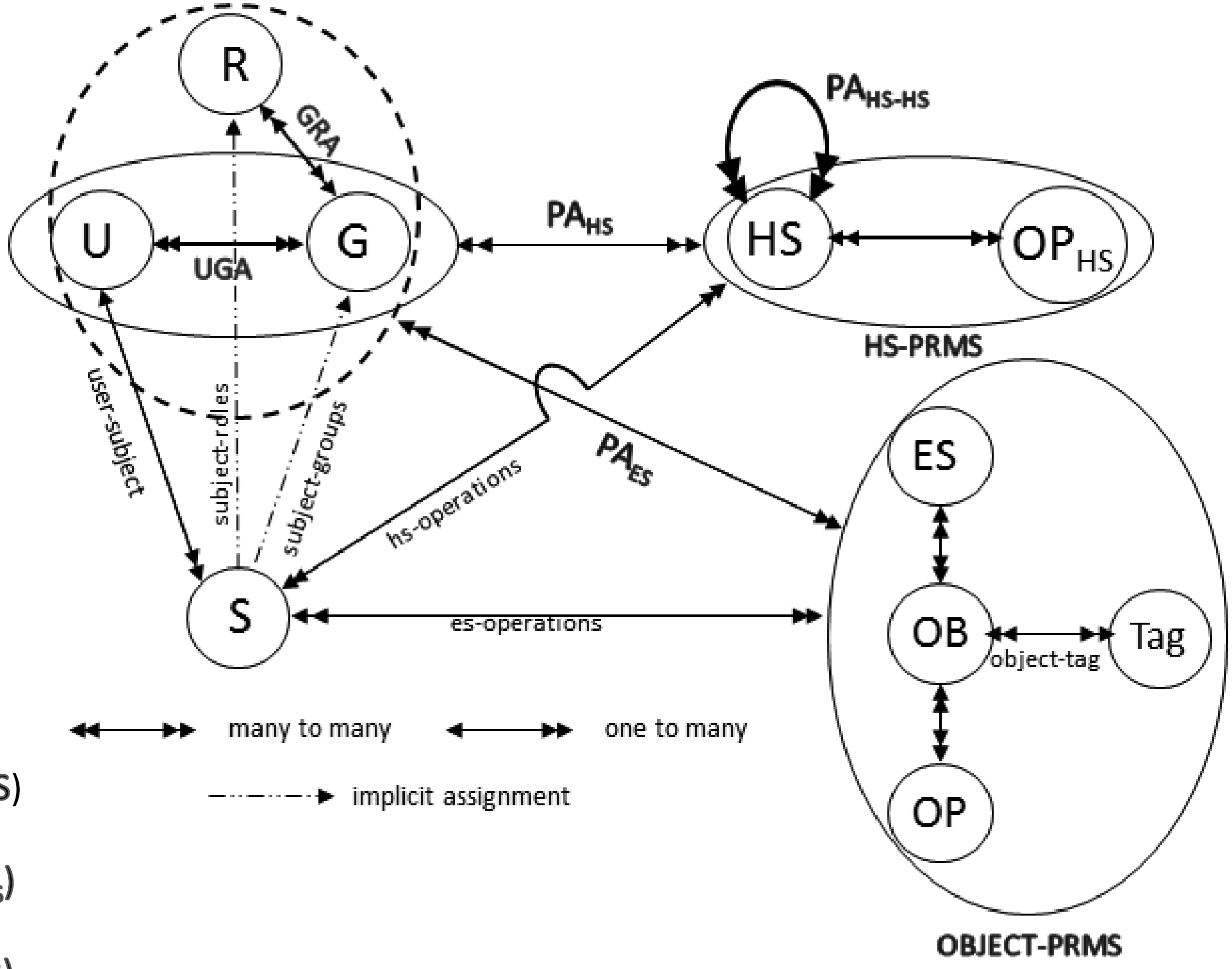
HDFS Files,
Hive Tables
Kafka Topics

Cluster Resources and Applications

YARN Queues,
Cluster Nodes

HeAC Model: Consolidated View

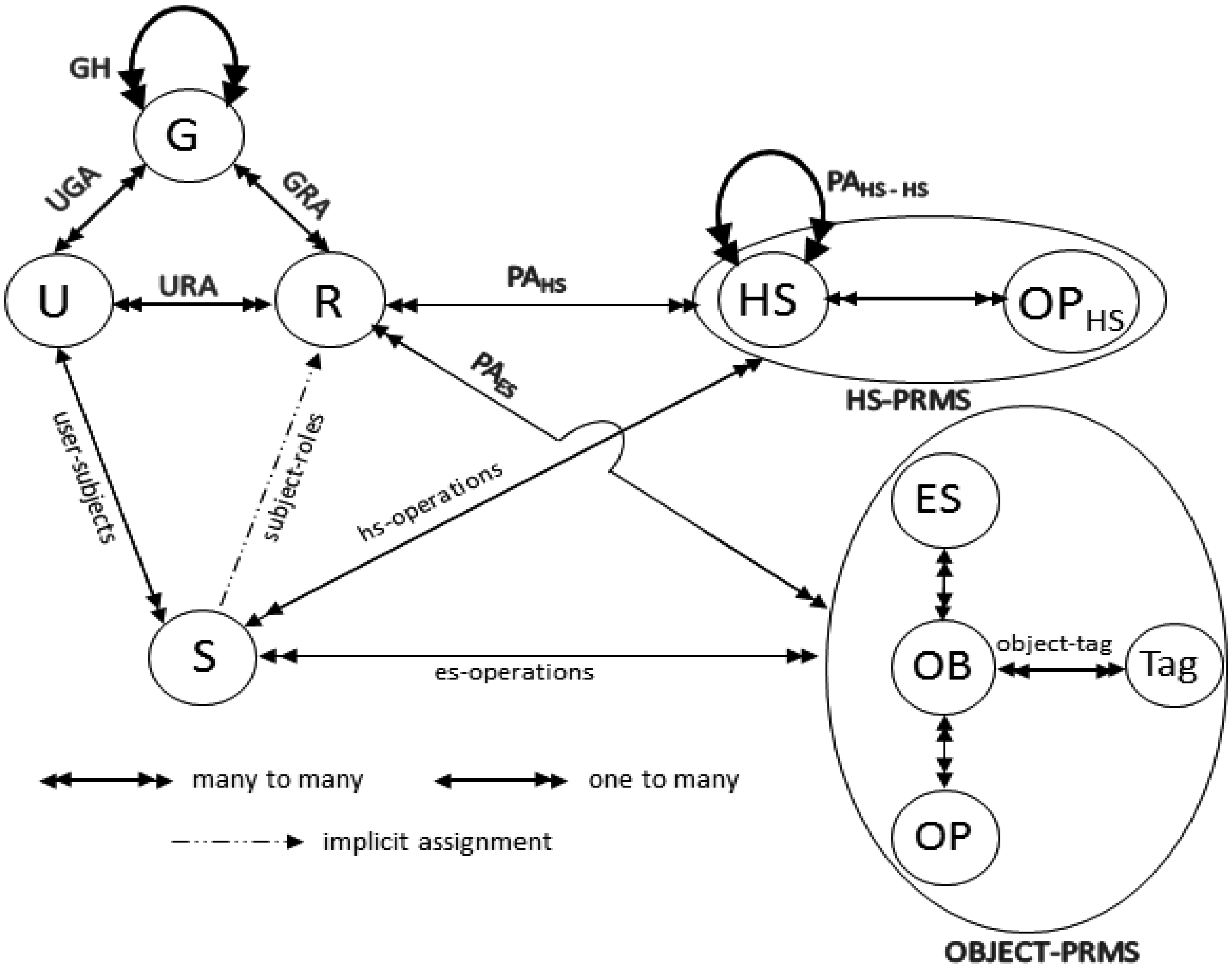
Hadoop Ecosystem Access Control Model



Users (U), Groups (G) , Subjects (S)
Hadoop Services (HS)
Hadoop Service Operations (OP_{HS})
Objects (OB), Operations (OP)
Ecosystem Service (ES), Objects (OB)
Operations (OP), Tag

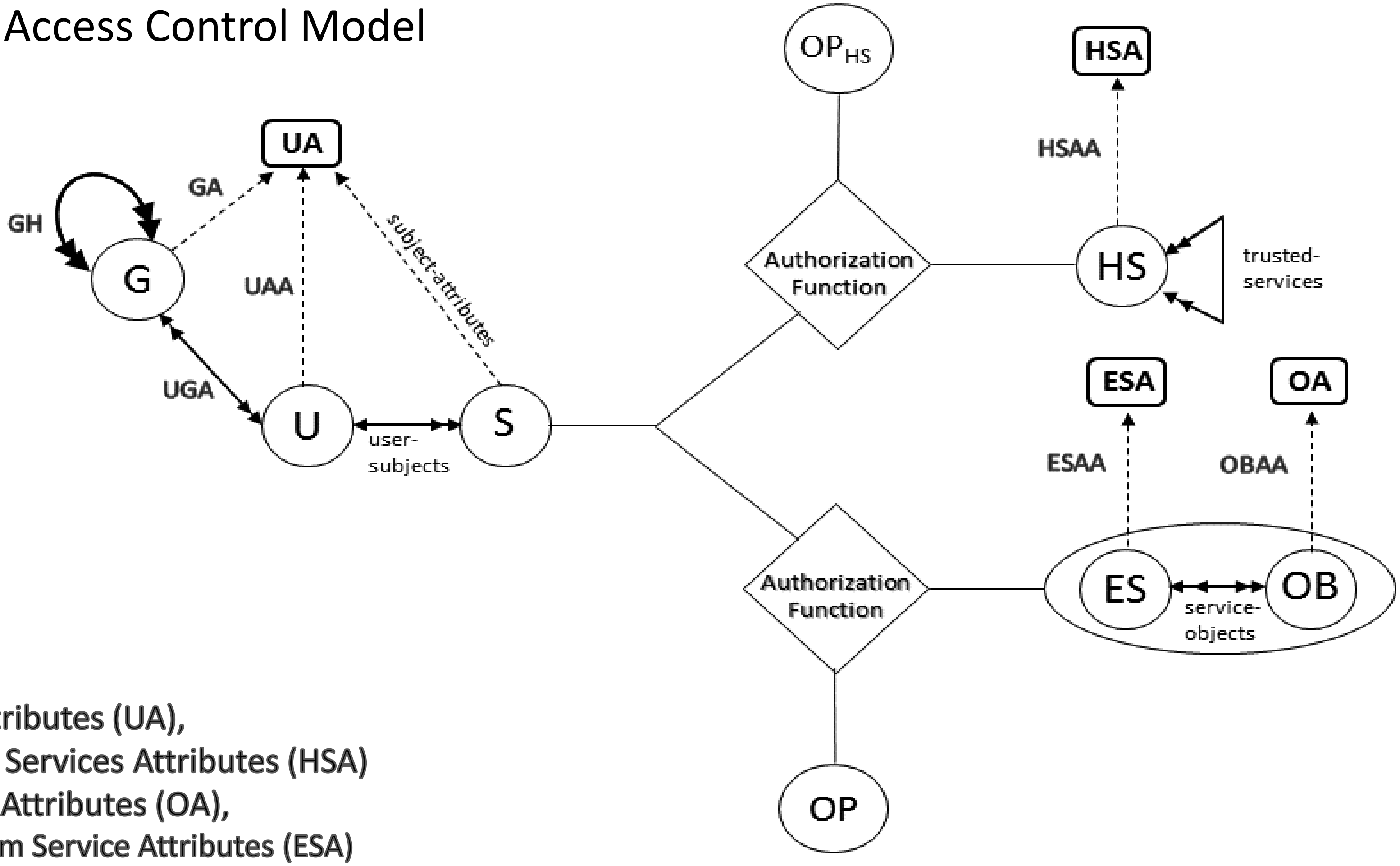
OT-RBAC Model

Object-Tagged RBAC



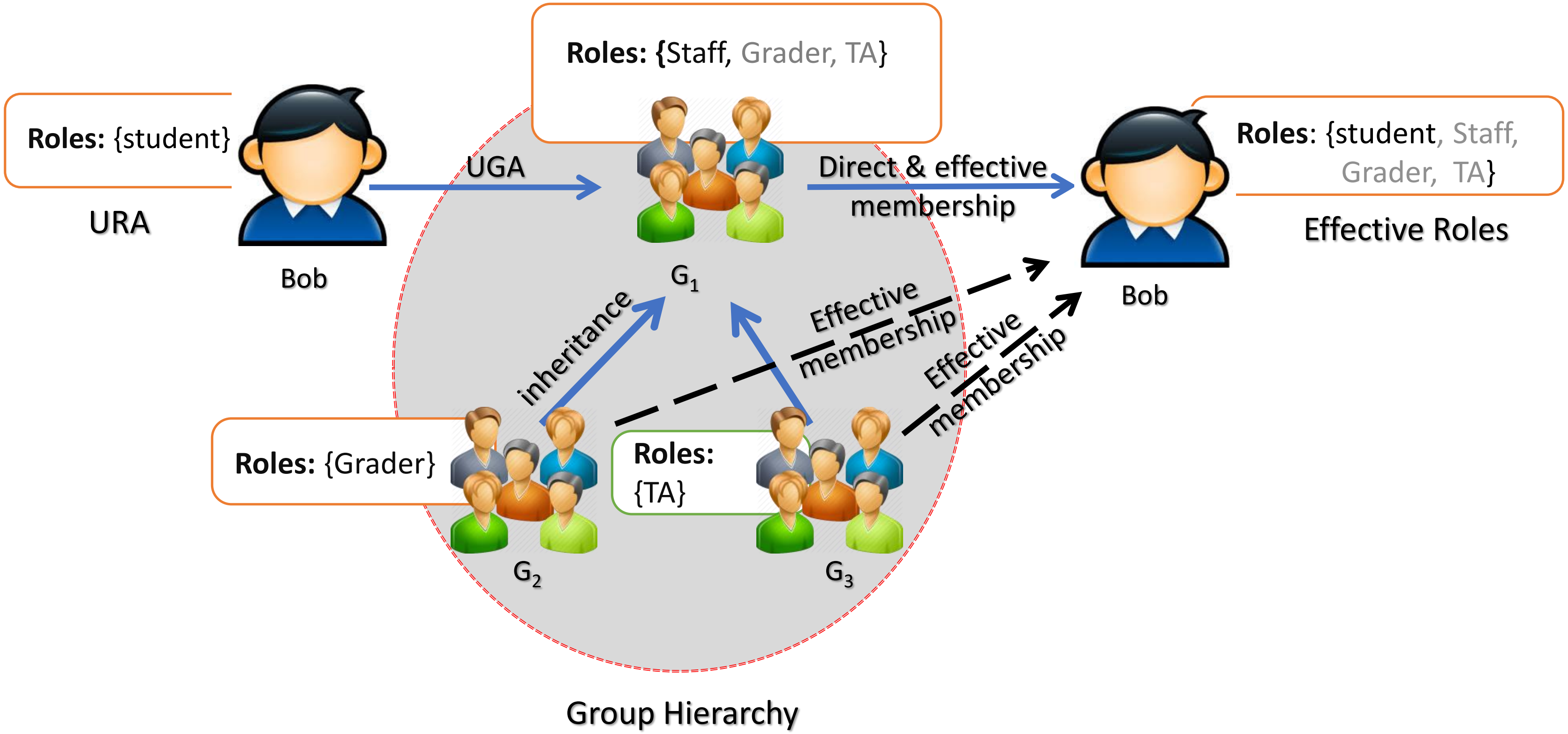
HeABAC Model

Hadoop Ecosystem Attribute-Based Access Control Model



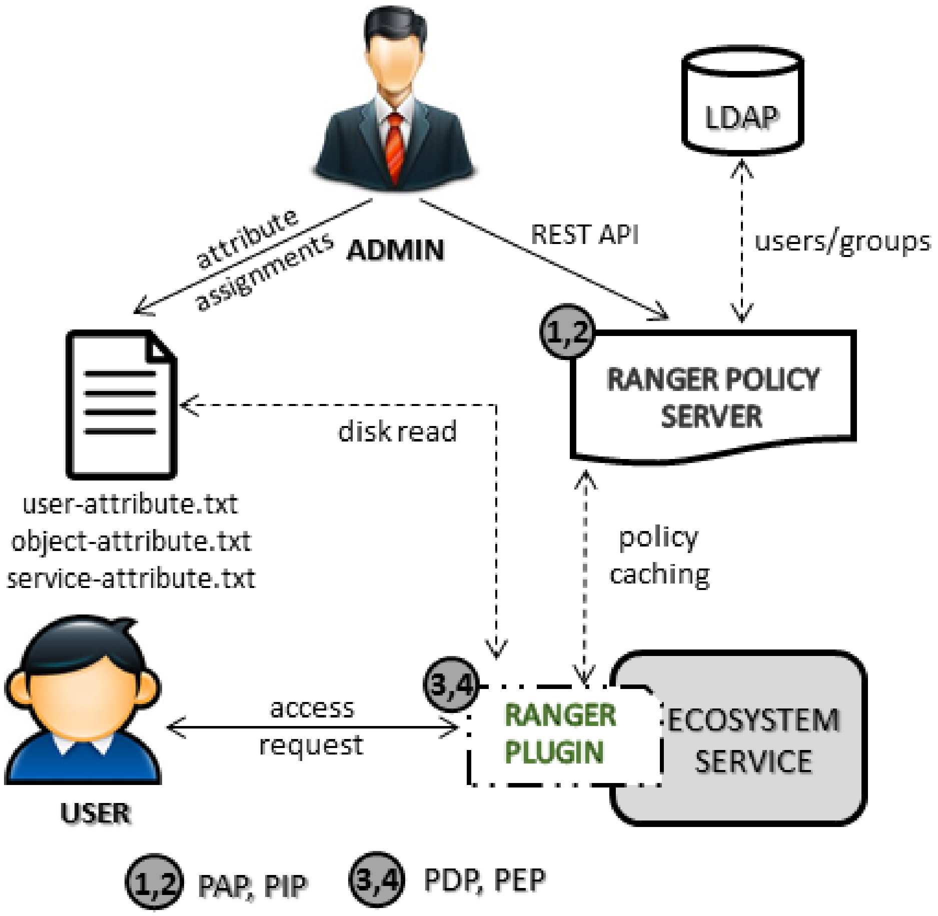
User Attributes (UA),
Hadoop Services Attributes (HSA)
Objects Attributes (OA),
Ecosystem Service Attributes (ESA)

Group Based Attribute Inheritance

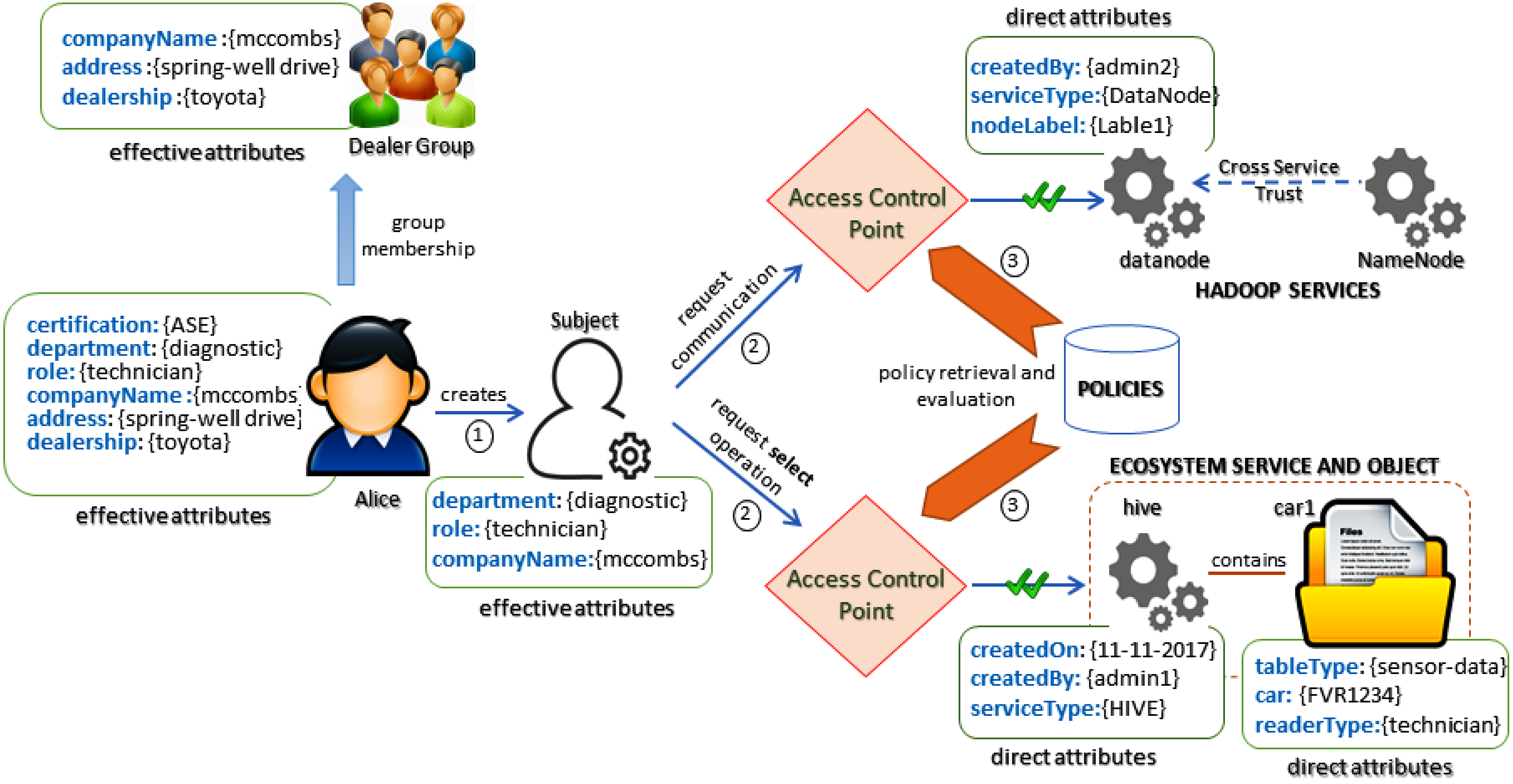


Major Benefits: Easy Administration where multiple roles can be assigned to user with single administrative operation.

Implementation Approach



HeABAC Use-Case



companyName :{mccombs}
address :{spring-well drive}
dealership :{toyota}

effective attributes

Dealer Group

group membership

certification: {ASE}
department: {diagnostic}
role: {technician}
companyName :{mccombs}
address: {spring-well drive}
dealership: {toyota}

effective attributes

Alice

creates

1

Subject

department: {diagnostic}
role: {technician}
companyName:{mccombs}

effective attributes

request communication

2

request select operation

2

Access Control Point

Access Control Point

policy retrieval and evaluation

3

3

POLICIES

direct attributes

createdBy: {admin2}
serviceType:{DataNode}
nodeLabel: {Lable1}

Cross Service Trust

datanode

NameNode

HADOOP SERVICES

ECOSYSTEM SERVICE AND OBJECT

hive

car1

contains

createdOn: {11-11-2017}
createdBy: {admin1}
serviceType:{HIVE}

direct attributes

tableType: {sensor-data}
car: {FVR1234}
readerType:{technician}

direct attributes

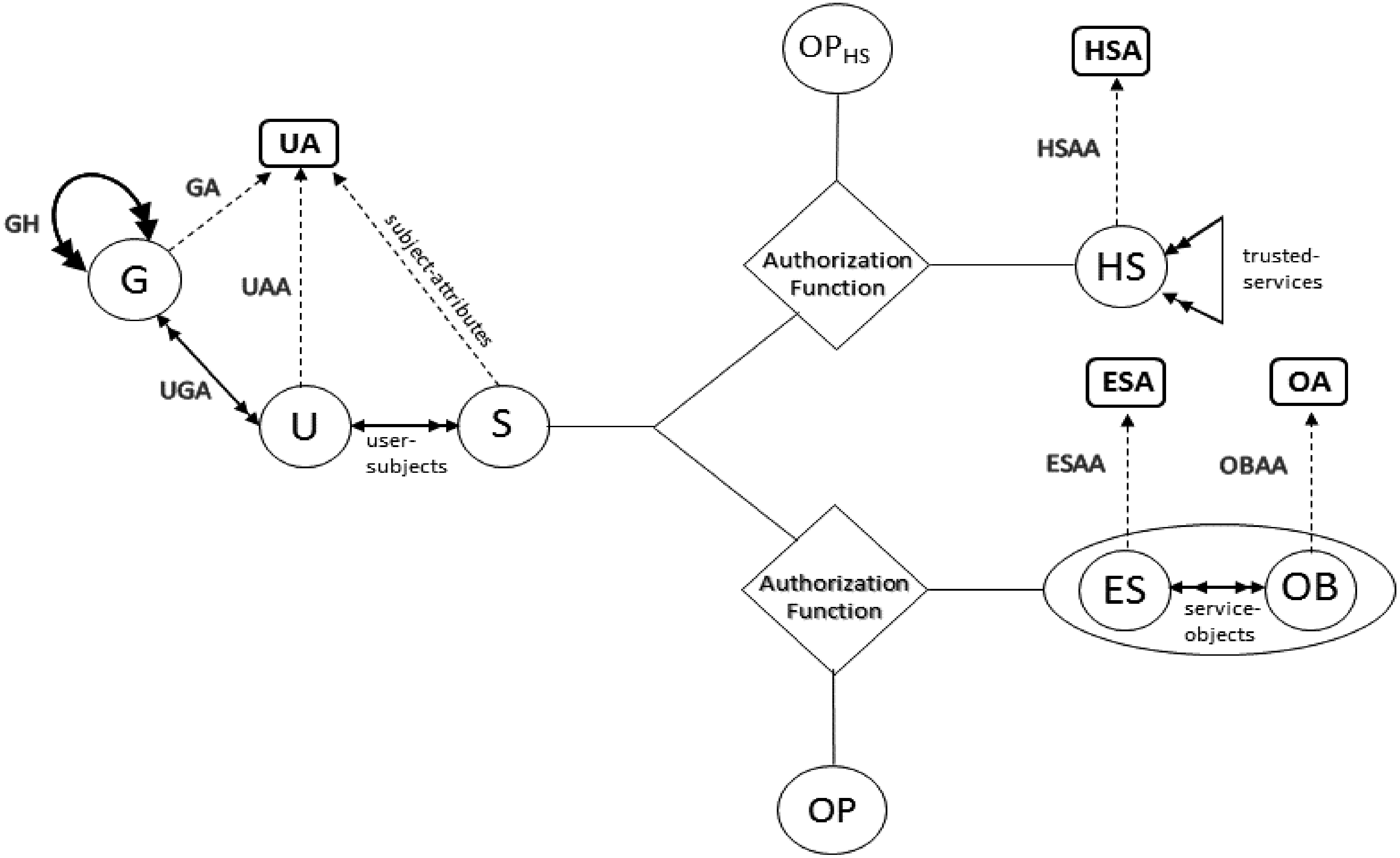
HeABAC Example Policies

1. $\text{Authorization}_{\text{access}}(\text{s:S}, \text{es:ES}) \equiv \text{diagnostic} \in \text{effective}_{\text{department}}(\text{s}) \wedge \text{technician} \in \text{effective}_{\text{role}}(\text{s}) \wedge \text{serviceType}(\text{es}) = \text{HIVE} \wedge \text{createdBy}(\text{es}) = \text{admin1}$.

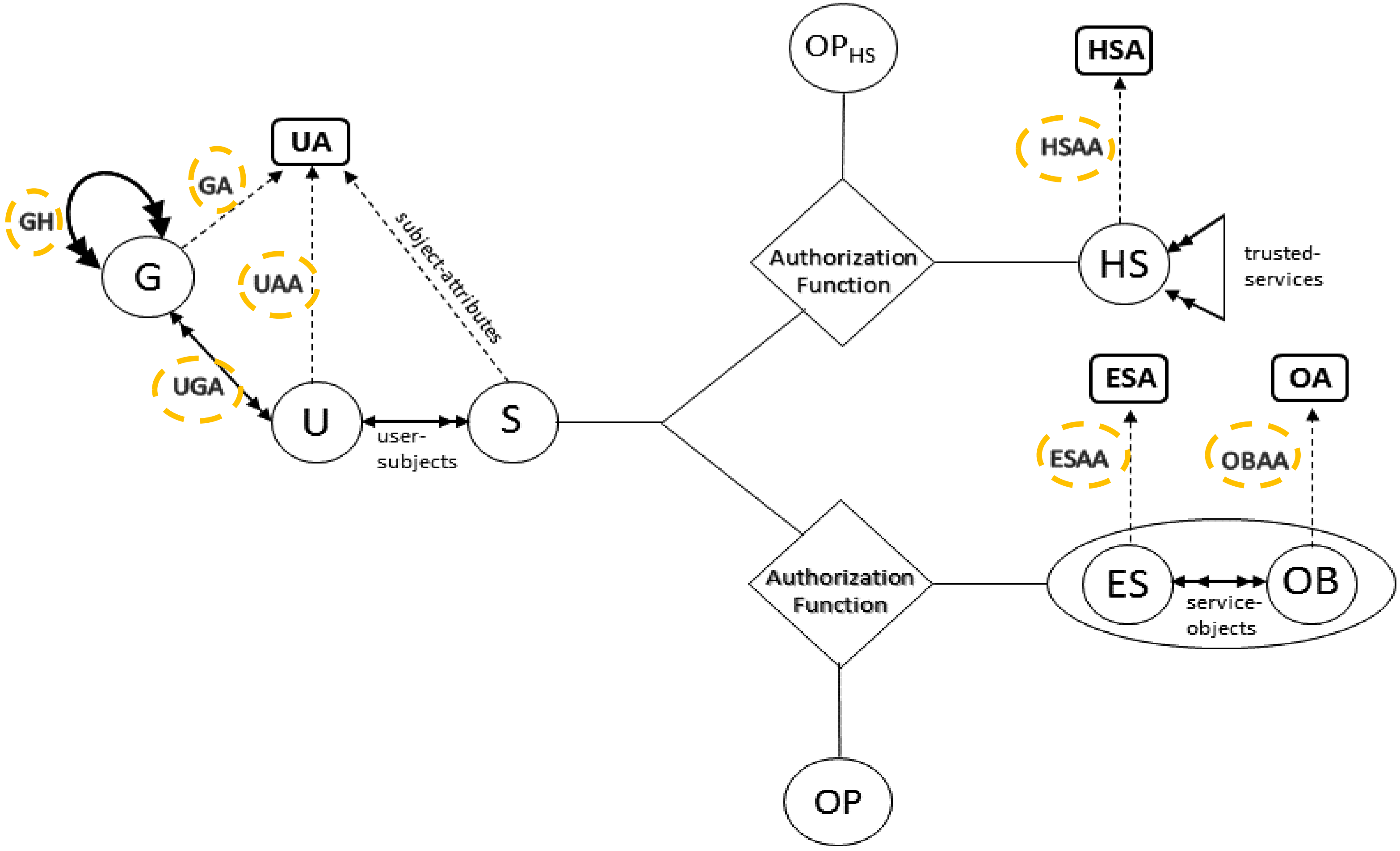
2. $\text{Authorization}_{\text{select}}(\text{s:S}, \text{es:ES}, \text{ob:OB}) \equiv \text{Authorization}_{\text{access}}(\text{s:S}, \text{es:ES}) = \text{True} \wedge \text{diagnostic} \in \text{effective}_{\text{department}}(\text{s}) \wedge \text{effective}_{\text{role}}(\text{s}) \in \text{readerType}(\text{ob}) \wedge \text{tableType}(\text{ob}) = \text{sensor-data} \wedge \text{car}(\text{ob}) = \text{FVR1234}$.

3. $\text{Authorization}_{\text{access}}(\text{s:S}, \text{hs:HS}) \equiv \text{diagnostic} \in \text{effective}_{\text{department}}(\text{s}) \wedge \text{technician} \in \text{effective}_{\text{role}}(\text{s}) \wedge \text{serviceType}(\text{hs}) = \text{DataNode} \wedge \text{createdBy}(\text{hs}) = \text{admin2}$

HeABAC Administrative Realms



HeABAC Administrative Realms



ARBAC inspired GURA, GURA$_G$ models are required.

Conclusion and Future Work

- Hadoop Authorization Layers
- Object-Tagged-RBAC Model
- Formalized Attributes based HeABAC Model

Some Future Goals:

- Introduce Data ingestion security
- Privacy concerns and finer grained approaches in multi-tenant Hadoop Lake