CERIAS

The Center for Education and Research in Information Assurance and Security

PURDUE UNIVERSITY

Fine-Grained Encryption-Based Access Control for Big Data

Mohamed Nabeel, Elisa Bertino - *Purdue University*

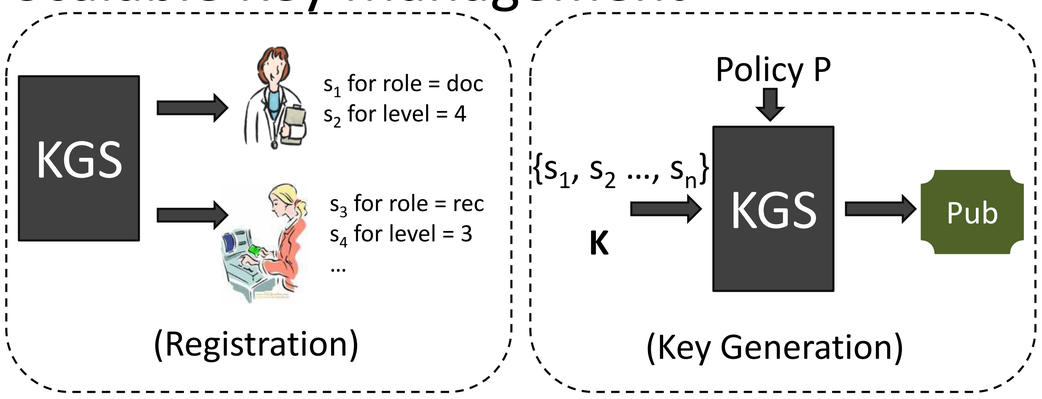
Problem Statement

- Big Data technologies (e.g. Hadoop) are increasingly used to store/analyze sensitive data
- Such data needs to be encrypted and access to it controlled
- Current schemes provide only coarse-grained access control, are vulnerable to key leakage, and inefficient
- There is a timely need to support fine-grained access control

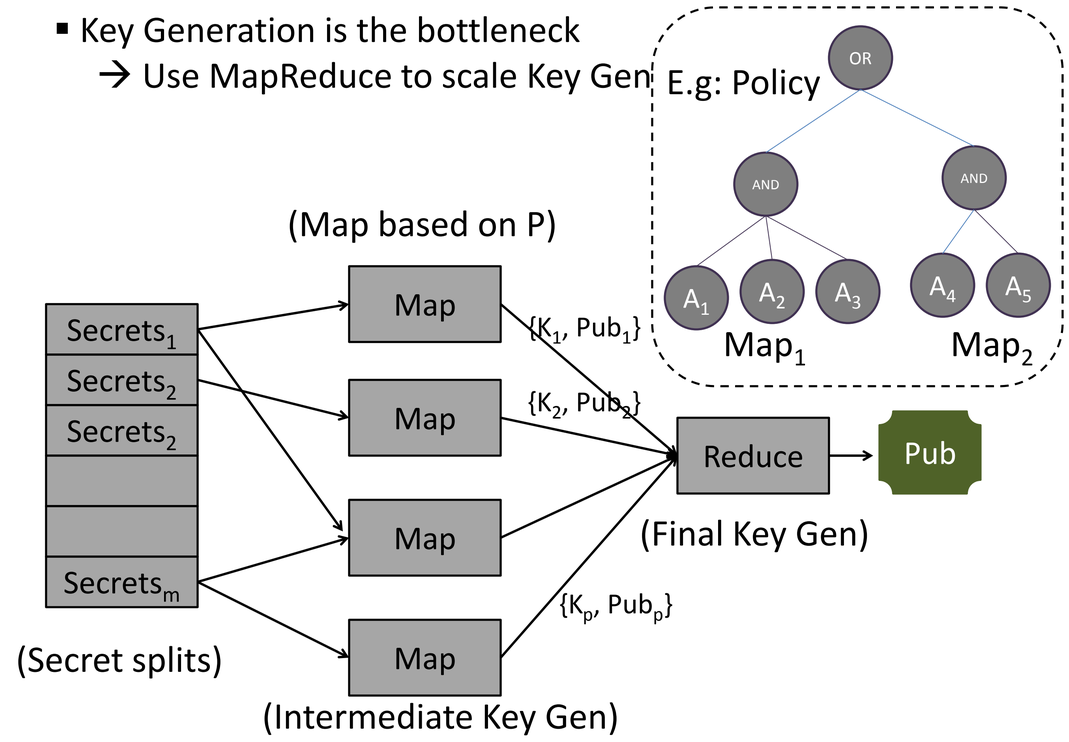
Requirements for an access control mechanism

- Data *at rest* should be encrypted
- Should support *attribute based access control*
- Should be able to support a *large dynamic user base*
- Should support *backward secrecy*
- Should be able to handle a *large amount of data*
- Should not degrade the *performance* of the system

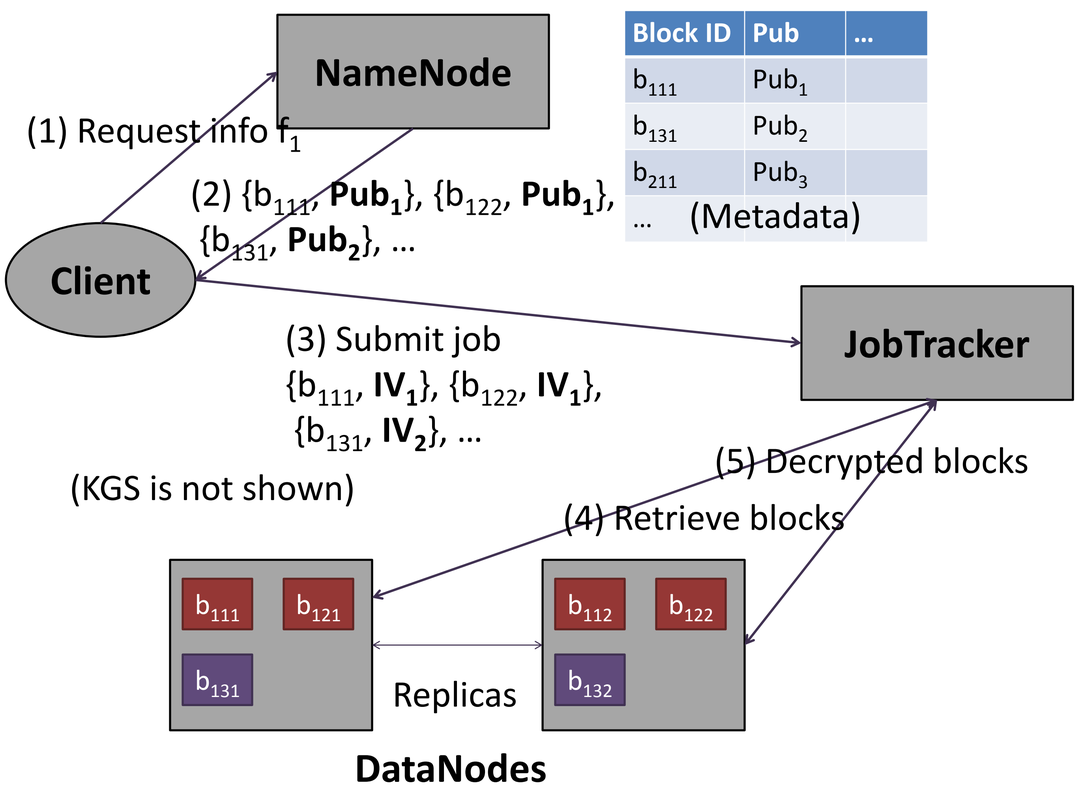
Scalable Key Management



- Key Generation is the bottleneck
  → Use MapReduce to scale Key Gen

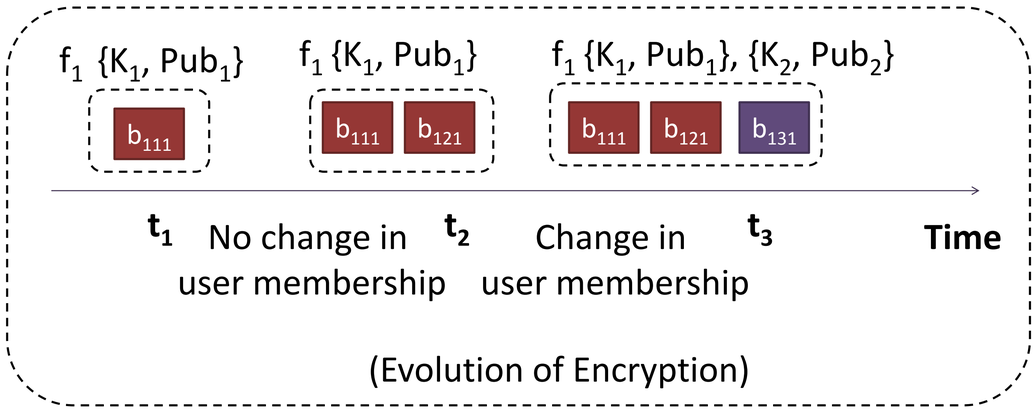


Fine-Grained Encryption on Hadoop



Block ID	Pub	...
b_{111}	Pub_1	
b_{131}	Pub_2	
b_{211}	Pub_3	
...	(Metadata)	

- Each file consists of a set of blocks
  → Blocks are encrypted and replicated across DataNodes
- Hadoop never stores actual symmetric keys
- Clients never send their secrets
  → An adversary cannot derive the secrets from IVs
- Clients can produce correct IVs if and only if their attributes satisfy the policy
- Encryption/Decryption is transparent to Clients/JobTracker



CERIAS

Discovery Park e-Enterprise Center