Distributed Information Processing

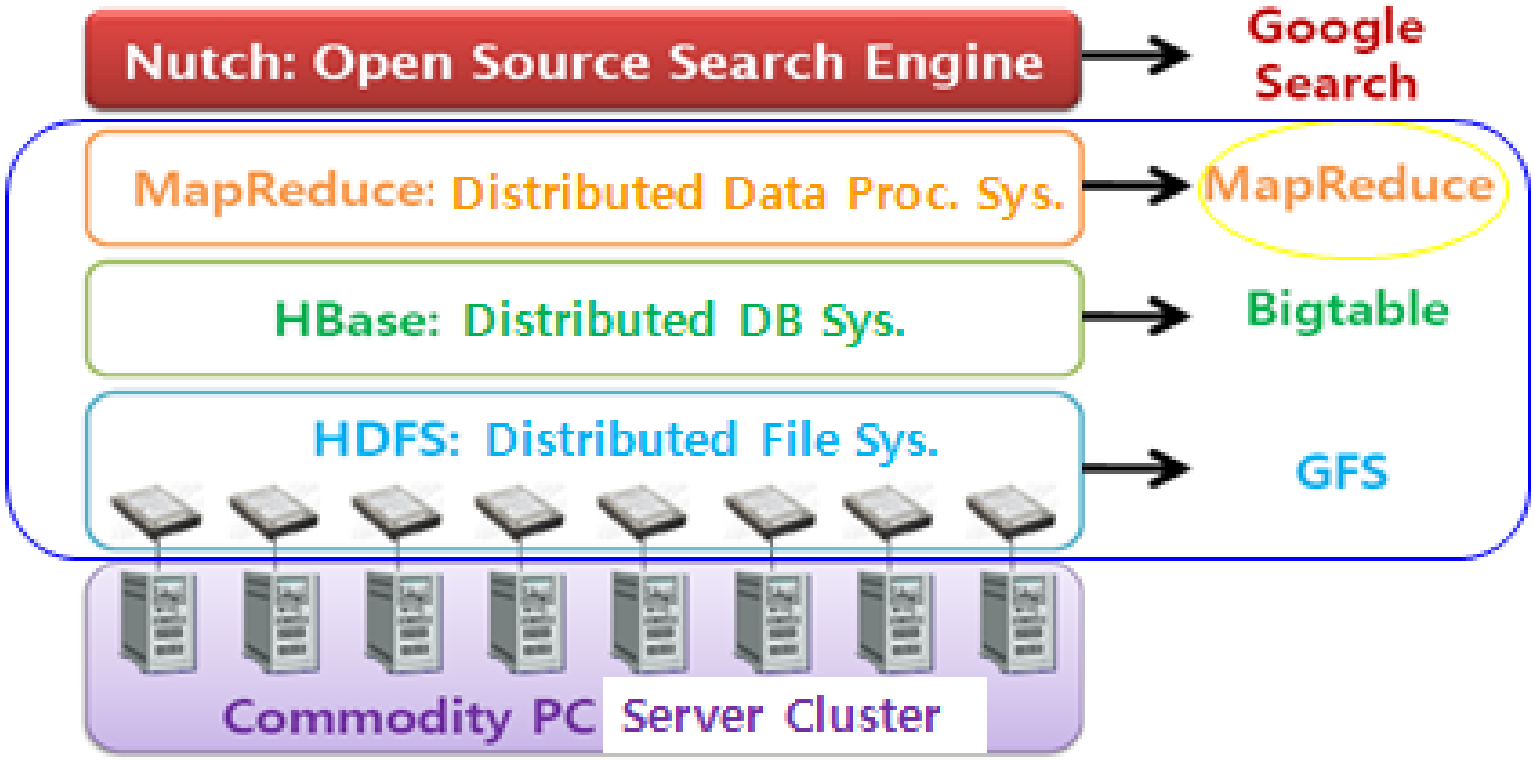
23rd Lecture

Eom, Hyeonsang (엄현상)
Department of Computer Science & Engineering
Seoul National University

Outline

- Hadoop & MapReduce Example
- Review
  - Information Protection
  - PKI Example
  - DRM Example
- Q&A

Hadoop Architecture



MapReduce Overview

- MapReduce: a Framework for processing huge datasets stored in a filesystem or database using a large number of computers (nodes)
  - Map step: The master node takes the input, partitions it up into smaller sub-problems, and distributes those to worker nodes. The worker node processes that smaller problem, and passes the answer back to its master node
  - Reduce step: The master node then takes the answers to all the sub-problems and combines them in some way to get the output

Wikipedia

MapReduce Overview Cont'd

- Map: Map(k1, v1)->list(k2, v2)
  - All independent maps performed in parallel
    - Limitation: data source and the number of CPUs
- Reduce: Reduce(k2, list(v2))->list(v3)
  - Each reducer presented outputs of the map operation sharing the same key
- Parallelism leading to fault tolerance
- Connecting the processes
  - Distributed file system
  - Direct streaming

Wikipedia

MapReduce Overview Cont'd

- Data flow
  - Input reader
    - Input from stable storage -> splits (possibly of key/value pairs)
  - Map function
    - Splits -> key/value pairs
  - Partition function
    - Key and # of reducers -> index of the desired reduce

Wikipedia

Map Reduce Overview Cont'd

- Data flow cont'd
  - Shuffle
    - Parallel sort & exchange
      - Transient data usually stored on local disk and fetched remotely by the reducers
  - Comparison function
    - Sorting the input for each reduce using the func.
  - Reduce function
    - Sorted keys with values -> those with reduced values
  - Output writer
    - Output to stable storage

Wikipedia

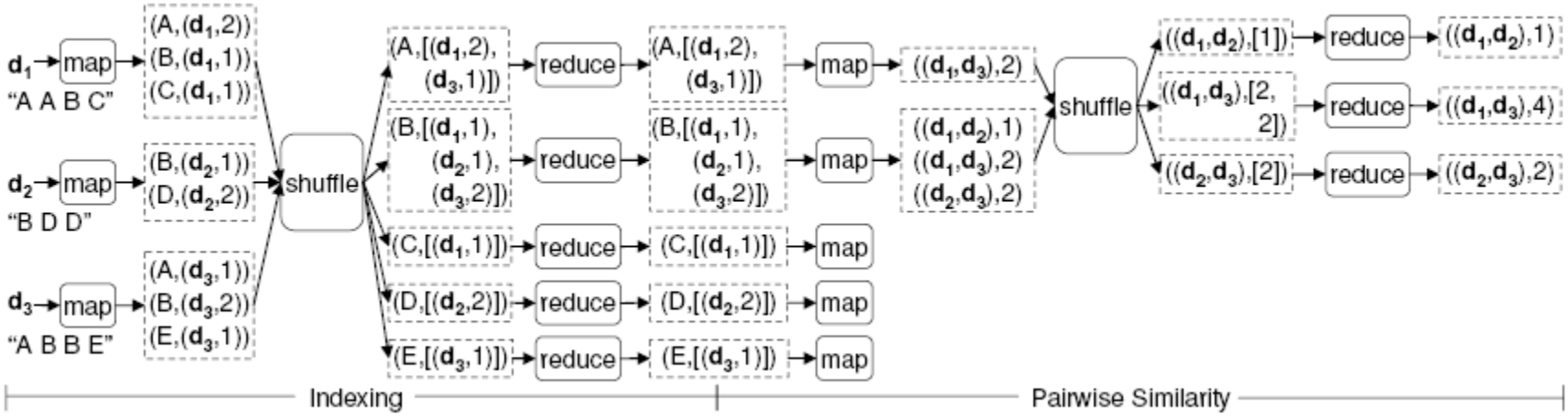
Document Clustering

- Pairwise Document Similarity

$$sim(d_i, d_j) = \sum_{t \in V} w_{t,d_i} \cdot w_{t,d_j}$$

where $sim(d_i, d_j)$ is the similarity between the documents, and $w_{t,d}$ indicates the importance of each term t in the documents

MapReduce Example: Document Clustering



Pairwise Document Similarity in Large Collections with MapReduce

Review: Information Protection

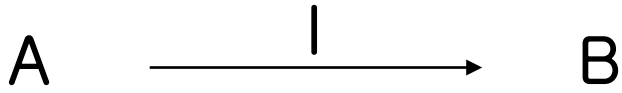
- ATM PIN Security
  - Splitting of a Customer's PIN into Two Parts and Storing Them Separately
    - PIN Offset in the ATM server
    - Natural PIN derived with the PIN key in the PIN machine

    CustomerPIN =(?) f(Acct#, PINOffset, PINKey)

    Natural PIN Is Not Stored Anywhere in the Entire Process

Review: PKI Example

- Basics
  - Digital Signature (DS)
    - DS(I,pr) for Information I and a private key pr
  - Certificate C (Containing a Public Key and DS)
    - C(pu,pr0) for a public key pu and pr0 from CA



Used to Make DS in C (by Encrypting the Rest of C)

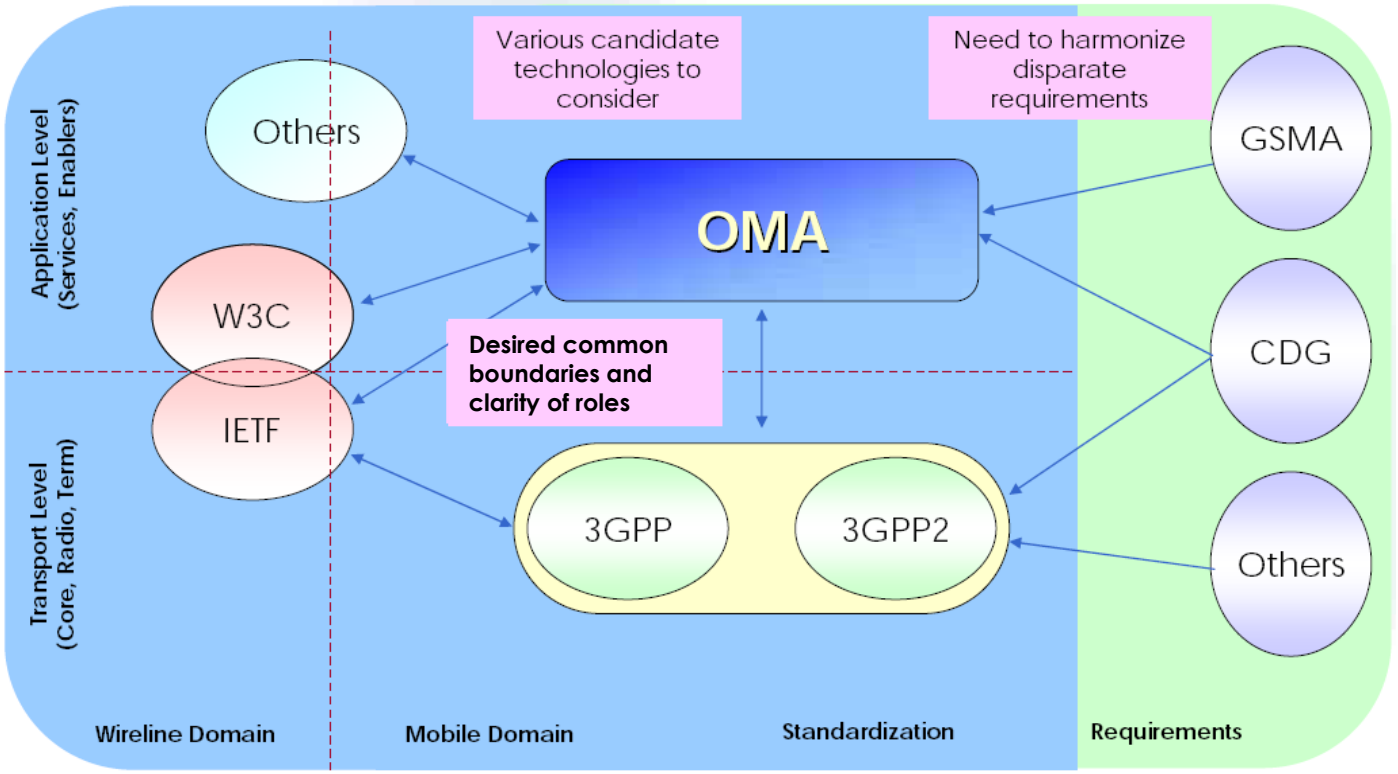
- Question
  - Is This Secure?
    - A sends B I+DS(I,pr1)+C1(pu1,pr0)
    - B verifies C1(pu1,pr0) by obtaining C0(pu0,pr0) from CA
      - Verification with DS of C1(pu1,pr0), and pu0
    - B verifies DS(I,pr1) with pu1

Self-Signed

Outline: DRM Example

- DRM Example: OMA (Open Mobile Alliance) DRM
  - Open Mobile Alliance
  - Overview of OMA DRM V1.0
  - Overview of OMA DRM V2.0
  - DRM Architecture
  - Domains
- Summary

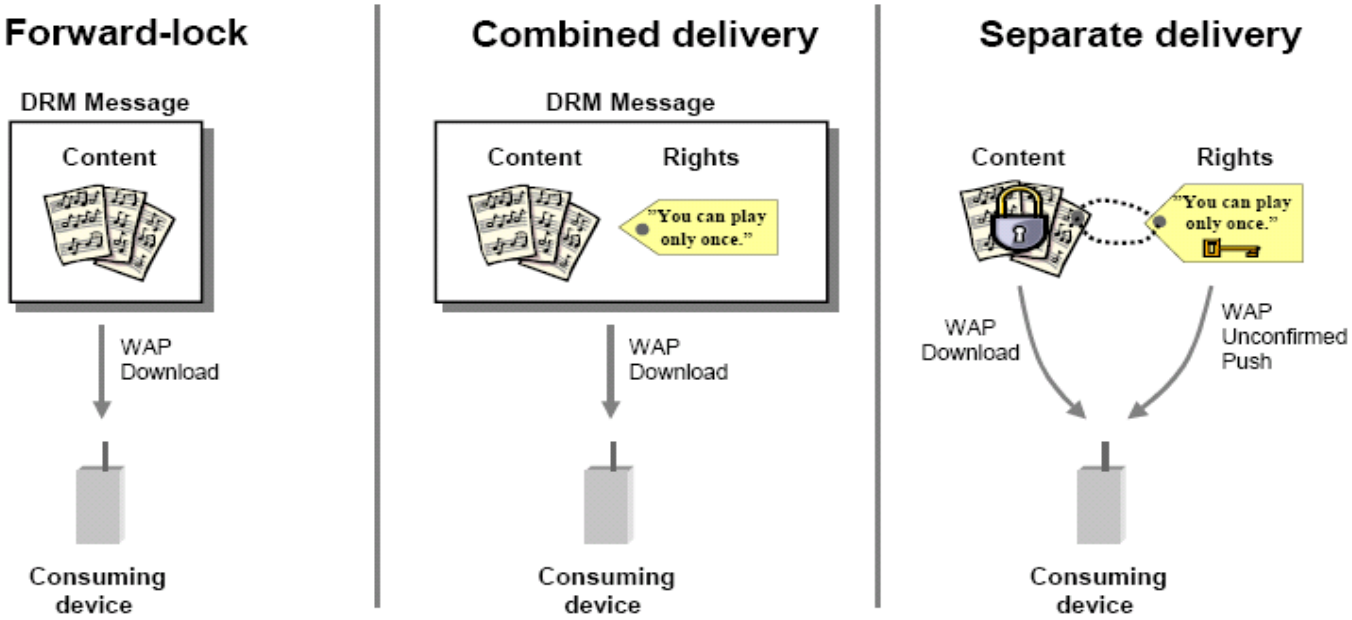
Open Mobile Alliance



GSMA : Global System for Mobile communication Association
CDG : CDMA Development Group
3GPP : 3rd Generation Partnership Project
W3C : World Wide Web Consortium
IETF : Internet Engineering Task Force

Overview of OMA DRM V1.0

- To



- Prevent peer-to-peer distribution of low-value content
- Prohibit device from forwarding content to other devices
- Consider only one media object

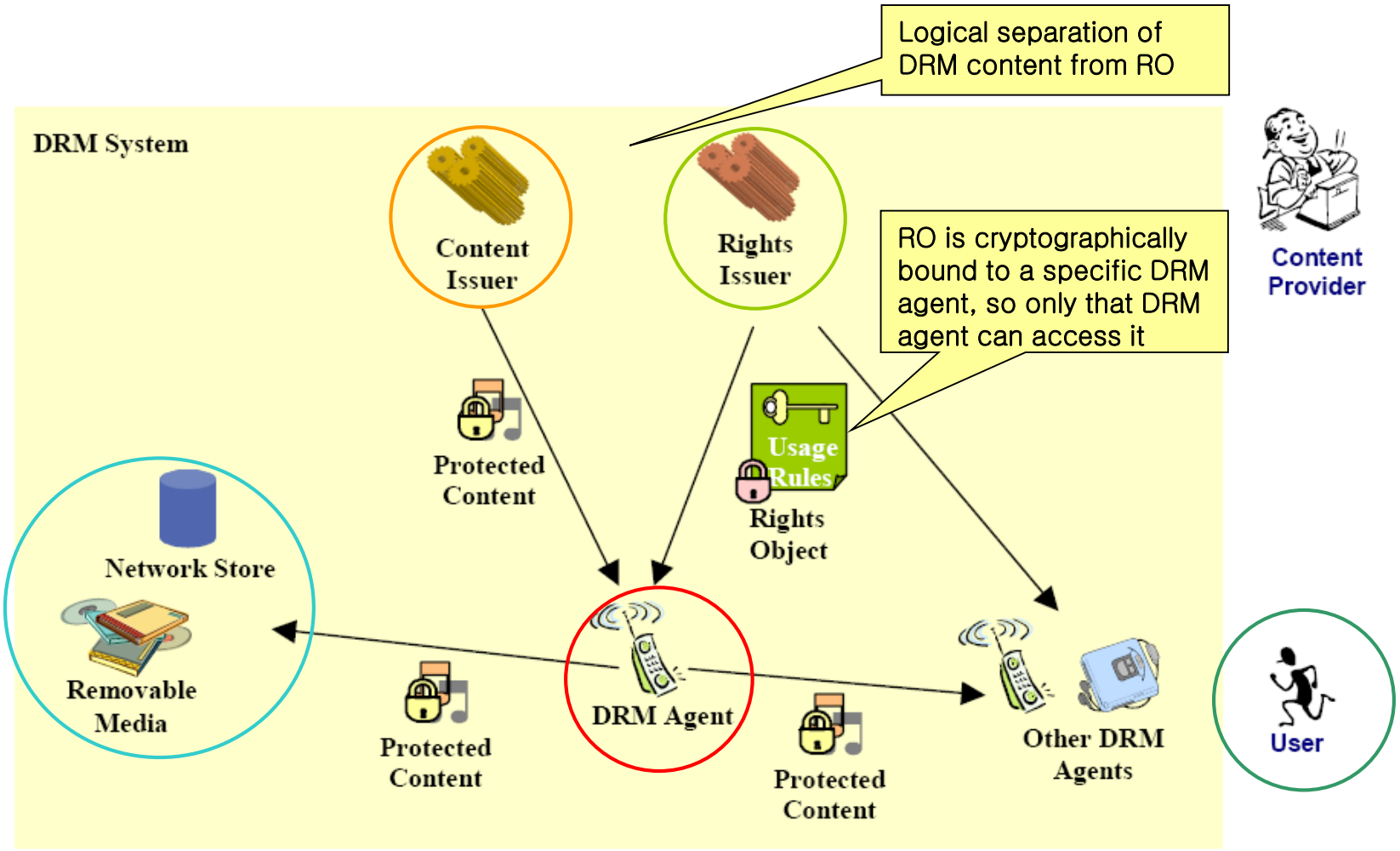
- Define a rights object containing permissions and constraints
- Package both content and a rights object in a DRM message

- Protect higher value content using encryption
- Separate content and a rights object
  - Protected content delivered over any medium
  - Rights object delivered via WAP push

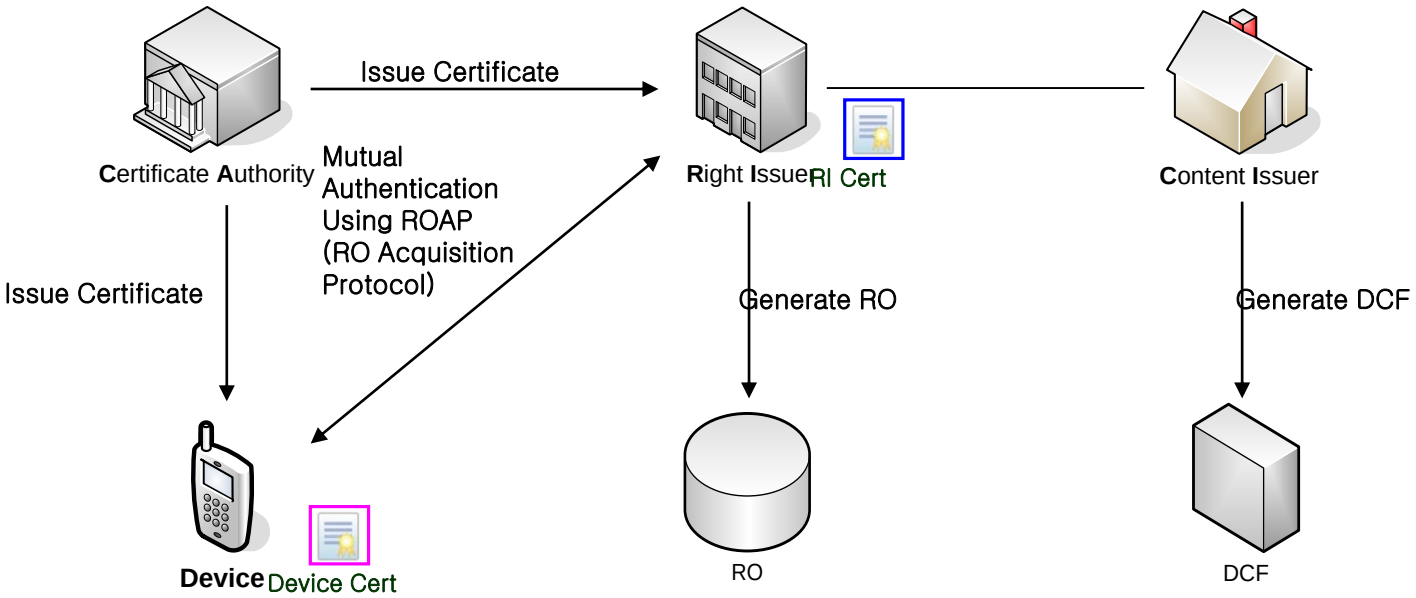
Overview of OMA DRM V2.0

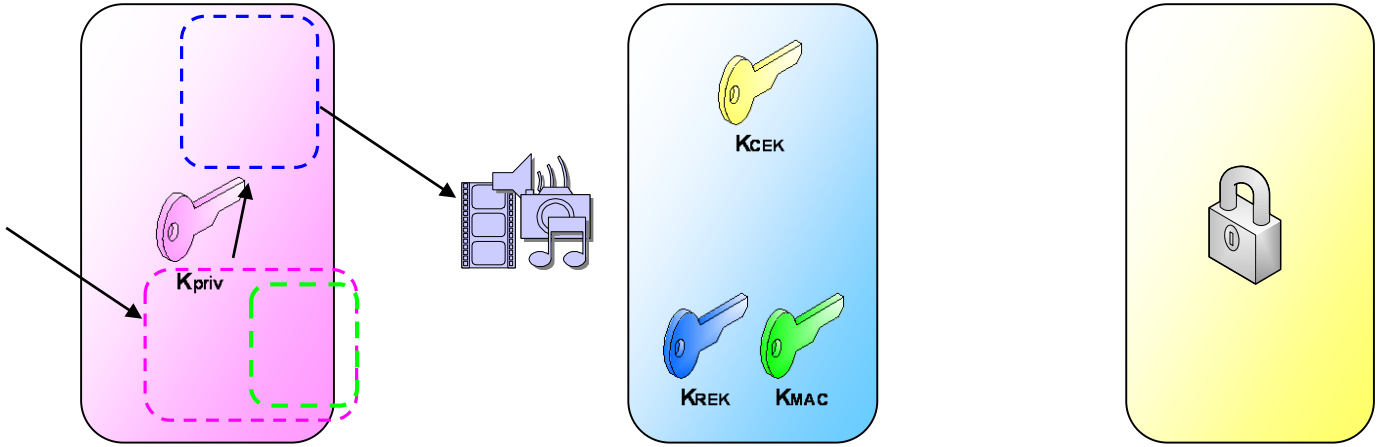
- Target
  - Enhanced Protection of Premium Content
    - Basic Pull Model
    - Push of DRM Content
    - **Streaming of DRM Content** *(Added)*
    - **Domains** *(Added)*
    - **Backup** *(Added)*
    - **Superdistribution** *(Added)*
    - **Export** *(Added)*
    - **Unconnected Device Support** *(Added)*

DRM Architecture



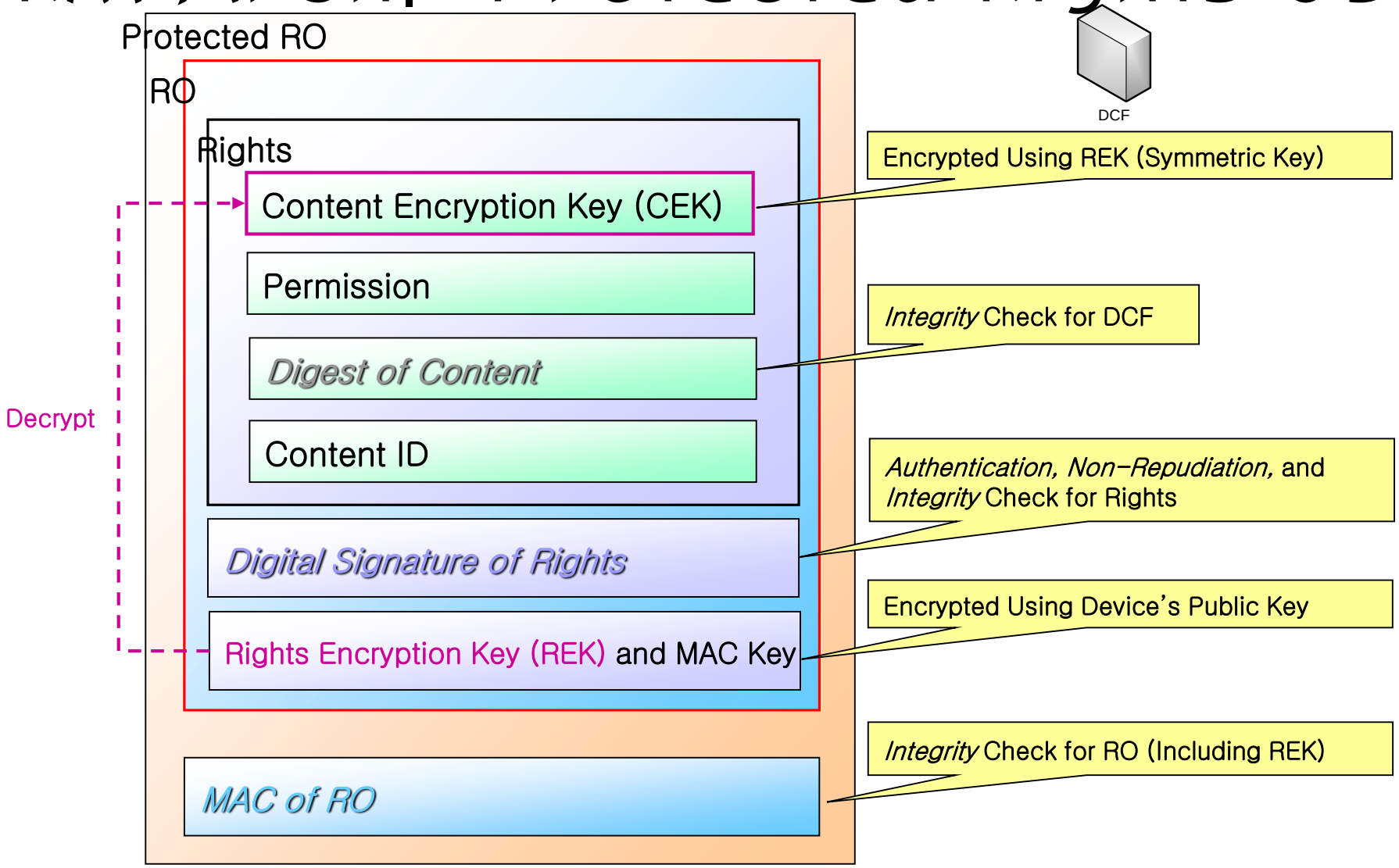
DRM Arch.: Cryptographic Chain



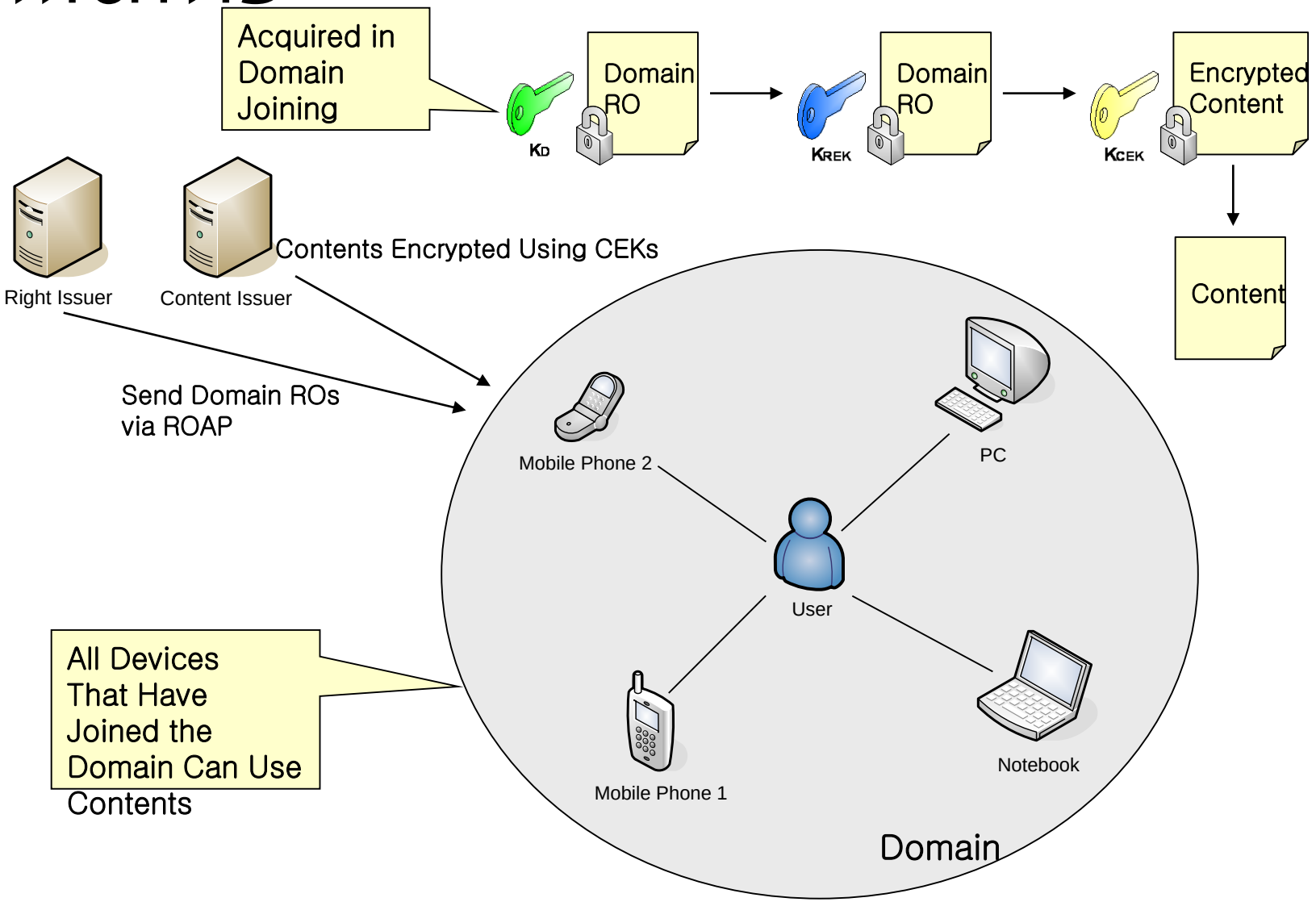


REK Confirmation

DRM Arch.: Protected Rights Obj.



Domains


Acquired in Domain Joining
KD
Domain RO
KREK
Domain RO
KCEK
Encrypted Content
Content
Right Issuer
Content Issuer
Contents Encrypted Using CEKs
Send Domain ROs via ROAP
Mobile Phone 2
PC
User
Mobile Phone 1
Notebook
Domain
All Devices That Have Joined the Domain Can Use Contents