Security In the Cloud

Giovanni Russello
g.russello@auckland.ac.nz

What will you learn?

- What cloud computing is
- Which are the security shortcomings
- An Encrypted Search Scheme supporting SQL-like encrypted queries

Computing as an Utility

In ***John McCarthy*** 1960 opinioned:

"computation may someday be organized as a ***public utility***."

What is Cloud Computing?

- Appearance of infinite resources on demand
  - No need to plan ahead for load surges
- Outsourcing is more convenient
  - One hour on 1000 servers = 1000 hours in one server
- Flexible Pay-as-you-go model
  - Processing by the hour
- No need for up-front commitments
  - Small and Medium Companies can get very reliable IT infrastructures

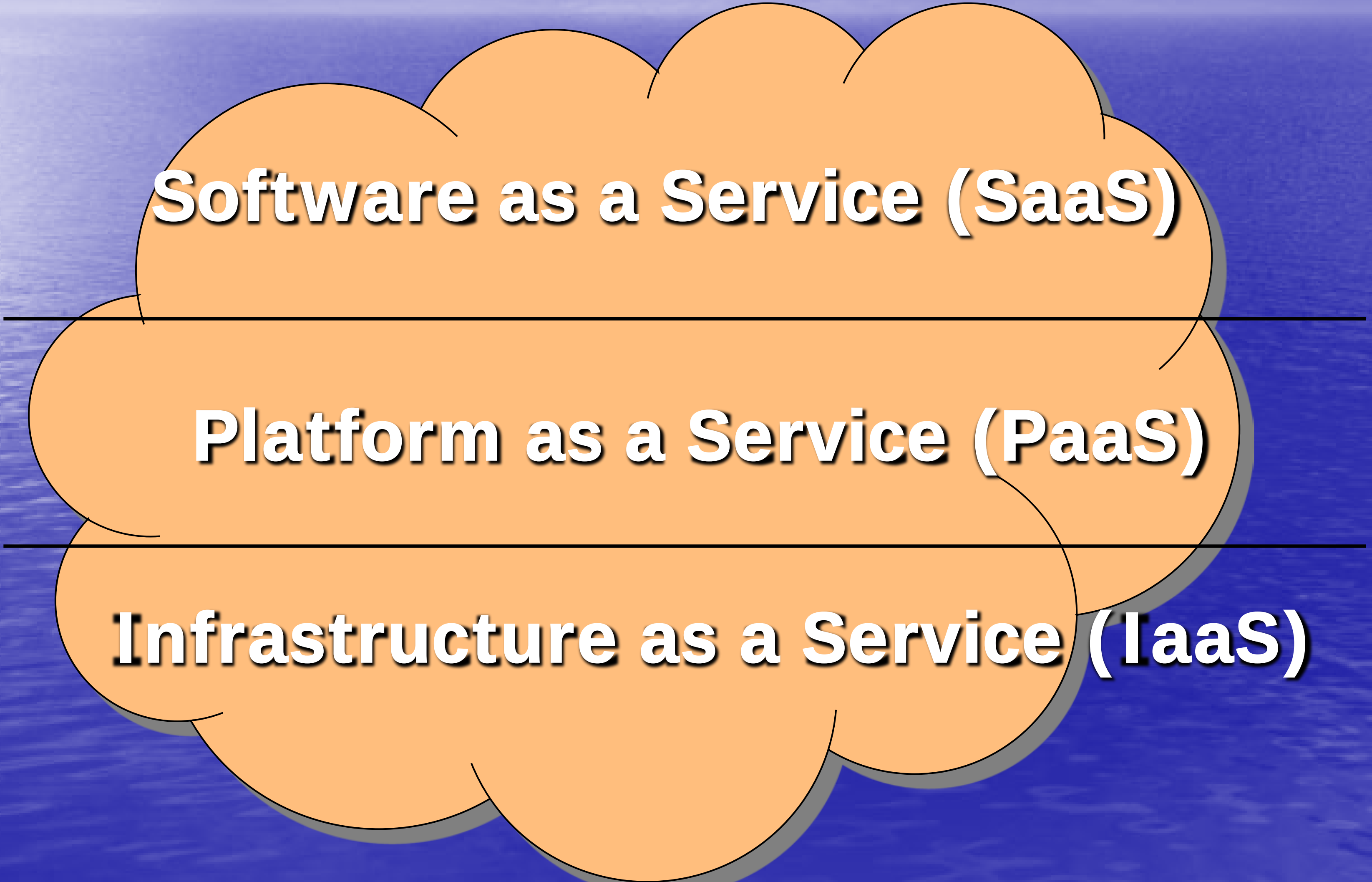
Based on Armbrust et al. - Communications April 2010

Some Definitions

Cloud Computing refers to:

- The software offered as Internet services
- The hardware and system software used for providing the services

Cloud Layers



Software as a Service (SaaS)

Platform as a Service (PaaS)

Infrastructure as a Service (IaaS)

Private vs Public Cloud

- Private cloud refers to internal datacentres of an organisation
- Public cloud refers to datacentres made available to the general public with a pay-as-you-go model

As an analogy

- A semiconductor fabrication line costs over $3 Billions
- Only big players in the market could afford one (Intel, Samsung)
- Then came companies that build chips for others
- Small companies, like nVidia, can capitalise on the chip design without the needs of buying the fab-lines

Who are the big Cloud players?

- Amazon
- Google
- Ebay
- Microsoft

Big datacentres + large-scale software already available

The Cloud Economics

- ***Elasticity***: Shifting the risk to the Cloud provider

- ***Pay-as-you-go*** model avoids:
  - Underprovisioning
  - Overprovisioning

Careless Computing?

According to ***Richard Stallman***:

“It’s stupidity. It’s worst than stupidity”

“I think that marketers like cloud computing because it is devoid of substantive meaning. [] it’s an attitude: ‘*Let any Tom, Dick and Harry **hold your data***, let any Tom, Dick and Harry do your computing for you (and ***control it***).’ Perhaps the term ‘***careless computing***’ would suit it better.”

Storm in the Clouds

Major Security Challenges:

- Availability of Service
- Data Lock-In
- Data Confidentiality

Service Availability

- A cloud computing service by a single provider represents a **Single Point of Failure**
- The provider can go out of business

Data Lock-In

- Cloud Computing API are still proprietary
- Not possible to move from one provider to another

Confidentiality in the Cloud

- Data Confidentiality represents the main obstacle to the adoption of cloud computing
- It is all about trusting valuable data to the cloud
- This data can be strictly regulated (HIPAA, SOX) for auditability

Data Confidentiality Today

- No cloud providers offer data confidentiality as a service
- Amazon Simple Storage Service (S3)
  - "Data stored within Amazon S3 is not encrypted at rest by AWS. However, users can **encrypt their data before it is uploaded** to Amazon S3 " [http://aws.amazon.com/articles/1697?_encoding=UTF8&jiveRedirect=1]

What are the Threats

- User-to-user threat
- User-to-infrastructure threat
- Provider-to-user threat

Protection Mechanisms

The main security mechanism in today cloud is virtualisation

This is effective for user-to-user and user-to-infrastructure threats.

Virtualisation Shortcomings

- However, not all virtualisation software is bug free and
- It is possible to user Cartography to map on which physical server an instance is running

[Ristenpart, et al. Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds. CCS09]

Protection from Providers

- Virtualisation is no effective means for **provider-to-user** threat
- Access control mechanisms are not effective when the infrastructure is not fully trusted
- Moreover there is always the problem of hard drivers “left around”

Some Successful Stories

- TC3 use case for the a HIPAA-compliant application to AWS
  - Reduction/elimination of protected health information (PHI) from the data stored and processed in the cloud
- AWS GovCloud (US): a specialised regional cloud where only restricted personnel as access to its facilities

What about Encryption?

- Traditional Encryption can help to protect the data confidentiality. But it is not practical because:
  - No computation is possible on the ciphertext
  - Ciphertext cannot be searched

We loose the initial benefits of Cloud Computing

Homomorphic Encryption

- Enables computation on encrypted data
- In 2009 Craig Gentry showed that fully homomorphic encryption was possible (but not practical)
- Recent work at Microsoft (Lauter et al) provides some practical breakthrough
  - Adds 100 numbers (128 bit) in 20 millisecs
  - Lots of statistical analysis can be done (i.e. predict when a person is going to have a heart attack)

Encrypted Search

- Performing of search and matching operations on fully encrypted data
- Several schemes exist
  - Single-user
  - Semi-fledged multi-user
  - Full-fledged multi-user

Single-user Searchable Encryption

- Crypto-components are divide between the user and the server
- The user performs encryption/decryption
- The server is responsible for search without learning information about the query and the data

Single-user Searchable Encryption

However

- It is only based on keyword match
- Only a single user can do insert and retrieve operations
- The key can be shared but this complicates key management

Semi-fledged multi-user

- Multiple users can perform search operations
- However, only one single user can do insert operations

Full-fledged multi-user

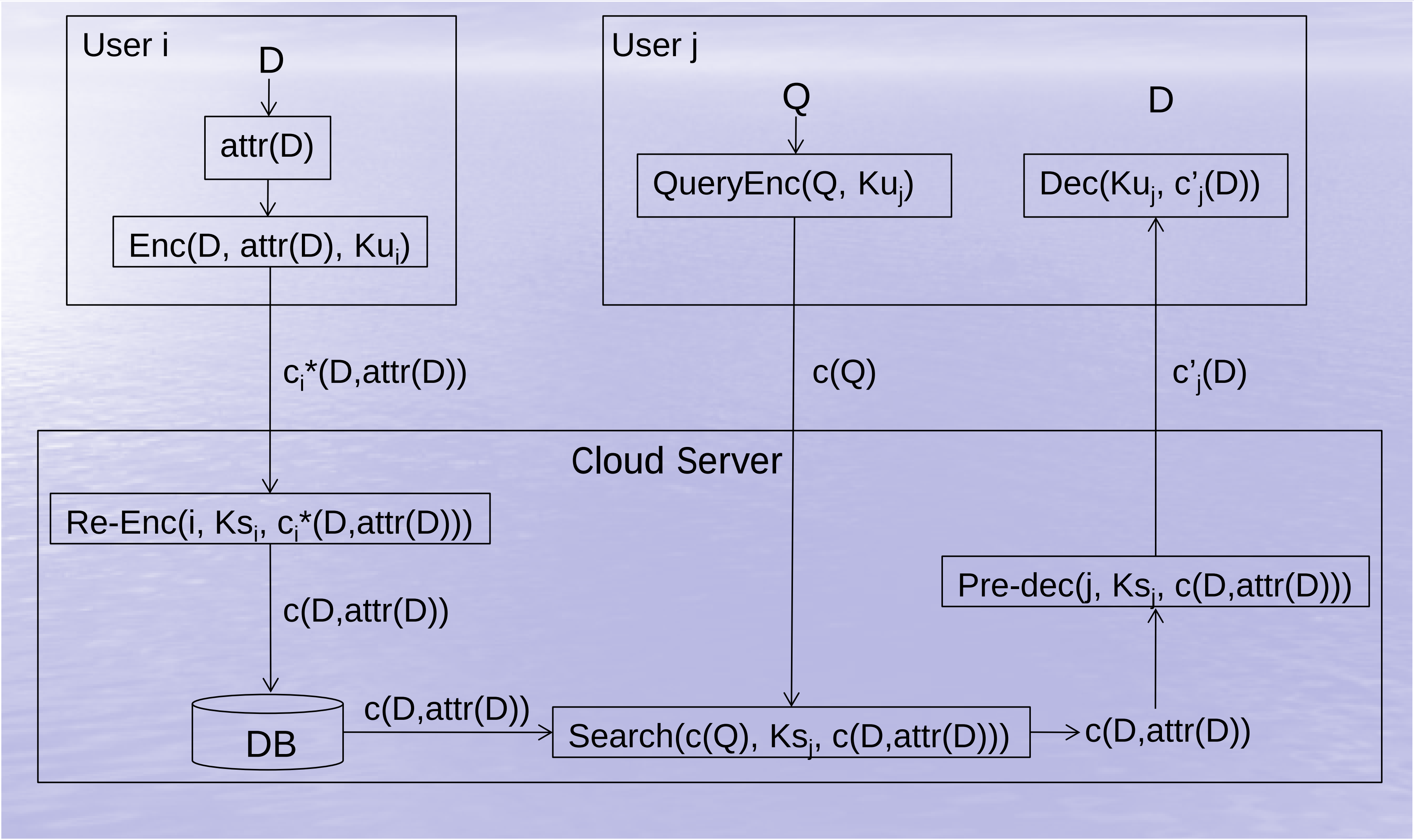
- Each authorised users can do insert and retrieve operations
- Users do not need to share keys

A Concrete Multi-User Scheme

System Model

- Database Owner
  - Organisation buying storage from a provider
- Key Management Authority (KMA)
  - Key generation and revocation
- User
  - Authorised entity to write and read the db
- Cloud Server
  - Stores and retrieves encrypted data for the users

Overview



User i

D

attr(D)

Enc(D, attr(D), Ku_i)

c_i*(D,attr(D))

User j

Q

QueryEnc(Q, Ku_j)

D

Dec(Ku_j, c'_j(D))

c(Q)

c'_j(D)

Cloud Server

Re-Enc(i, Ks_i, c_i*(D,attr(D)))

c(D,attr(D))

DB

c(D,attr(D))

Search(c(Q), Ks_j, c(D,attr(D)))

c(D,attr(D))

Pre-dec(j, Ks_j, c(D,attr(D)))

Init Algorithm

$Init(1^k)$:

on input $1^K \rightarrow p, q : q = (p-1)/2 \ and \ |q| = k$

g generator of $G : unique \ order \ q \ subgroup \ of \ Z_p^*$

choose x random from Z_q^*

output $h = g^x, H, f, s$

$PubParams(G, g, q, h, H, f)$

$MSK(x, s)$

Key Generation

$KeyGen(MSK, i)$

For each user i choose a random x_{i1} from Z_q^*

Compute $x_{i2} = x - x_{i1}$

Transmit $Ku_i = (x_{i1}, s)$ to user i

Transmit $Ks_i = (i, x_{i2})$ to cloud server

Performing an INSERT operation

User u_i

$$INSERT\ \ INTO$$

$$table\_name(attr\_name_1,\ldots,attr\_name_n)$$

$$VALUES(value_1,\ldots,value_n)$$

Preparing Record for Encryption

$$D = (a_1, \ldots, a_n)$$

$$a_k = (attr\_name_k, value_k)$$

$$attr(a_k) \rightarrow \{v_1, \ldots, v_m\}$$

if $value_k$ is a string

then $attr(D)$ outputs $\{v_k = value_k\}$

Preparing Record for Encryption

If $value_k$ is a numerical value

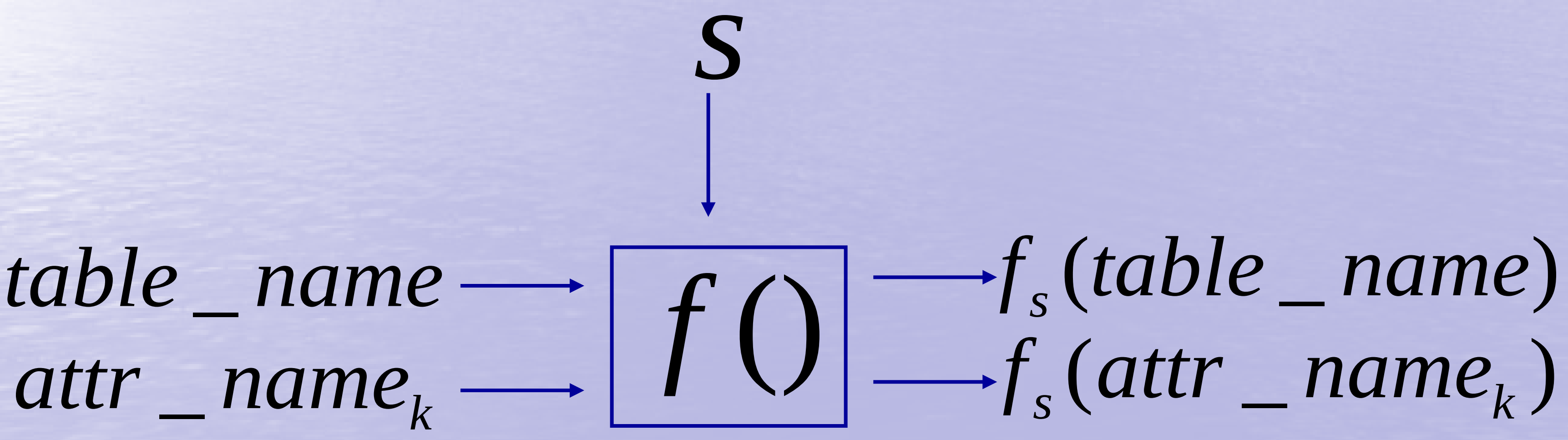
then it creates a bit representation

For instance, $(age,18)$ in 6 - bit is 010010

$$attr((age,18)) \rightarrow \quad \{v_{age_1} = 0*****, v_{age_2} = *1****, v_{age_3} = **0***, v_{age_4} = ***0**, v_{age_5} = ****1*, v_{age_6} = *****0\}$$

User-side Encryption

User u_i $Ku_i(x_{i1}, s)$



s

$table\_name$

$attr\_name_k$

$f()$

$f_s(table\_name)$

$f_s(attr\_name_k)$

User-side Encryption

User u_i $\quad Ku_i(x_{i1}, s)$

$$\begin{tikzcd}
 & x_{i1} \arrow[d] & \\
a_k \arrow[r] & \boxed{PE} \arrow[r] & C^*_{PE}(a_k) = (g^{r_k}, g^{r_k x_{i1}} a_k)
\end{tikzcd}$$

User-side Encryption

$$attr(a_k) = \{v_{k_1}, \ldots, v_{k_m}\}$$

$$\begin{tikzcd}
 & x_{i1} \arrow[d] & \\
v_{k_i} \arrow[r] & \boxed{SDE} \arrow[r] & C^*_{SDE}(v_{k_i}) = \begin{cases} \hat{c}_1 = g^{r_{k_i} + \sigma_{k_i}} \\ \hat{c}_2 = \hat{c}_1^{\,x_{i1}} \\ \hat{c}_3 = H(h^{r_{k_i}}) \end{cases}
\end{tikzcd}$$

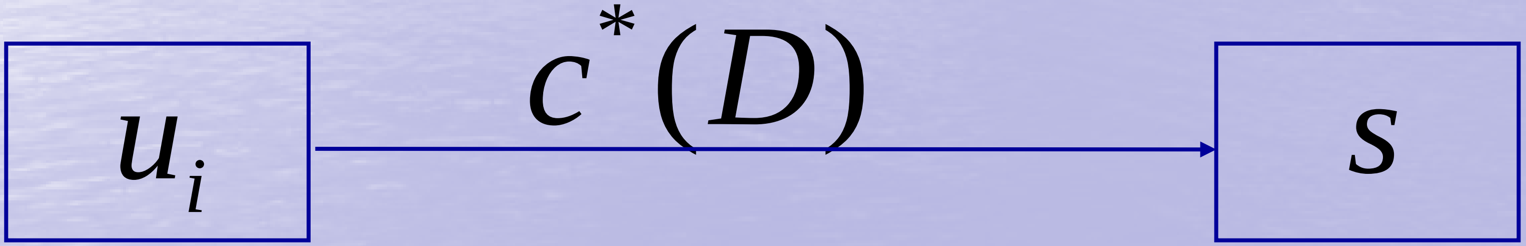
$$\sigma_{k_i} = f_s(v_{k_i})$$

User-side Encryption

$$c^*(D) = (c^*(a_1), \ldots, c^*(a_n))$$

$$\begin{aligned} c^*(a_k) = (& f_s(attr\_name_k), \\ & c^*_{PE}(a_k), \\ & c^*_{SDE}(v_{k_1}), \ldots, c^*_{SDE}(v_{k_m})) \end{aligned}$$

Sending the data to the Cloud



u_i

$c^*(D)$

s

Server-side PE-Re-Encryption

$$Ks_i(i, x_{i2})$$

$$\begin{tikzcd}
 & x_{i2} \arrow[d] & \\
C^*_{PE}(a_k) \arrow[r] & \boxed{PE-REncr} \arrow[r] & C_{PE}(a_k)
\end{tikzcd}$$

Server-side PE-Re-Encryption

$$C^{*}_{PE}(a_k) = (g^{r_k}, g^{r_k x_{i1}} a_k)$$

$$g^{r_k}$$

Server-side PE-Re-Encryption

$$C^*_{PE}(a_k) = (g^{r_k}, g^{r_k x_{i1}} a_k)$$

$$(g^{r_k})^{x_{i2}}$$

Server-side PE-Re-Encryption

$$C_{PE}^{*}(a_k) = (g^{r_k}, g^{r_k x_{i1}} a_k)$$

$$(g^{r_k})^{x_{i2}} \cdot g^{r_k x_{i1}} a_k$$

Server-side PE-Re-Encryption

$$C^{*}_{PE}(a_k) = (g^{r_k}, g^{r_k x_{i1}} a_k)$$

$$(g^{r_k})^{x_{i2}} \cdot g^{r_k x_{i1}} a_k = g^{r_k x_{i2} + r_k x_{i1}} a_k = g^{r_k x} a_k$$

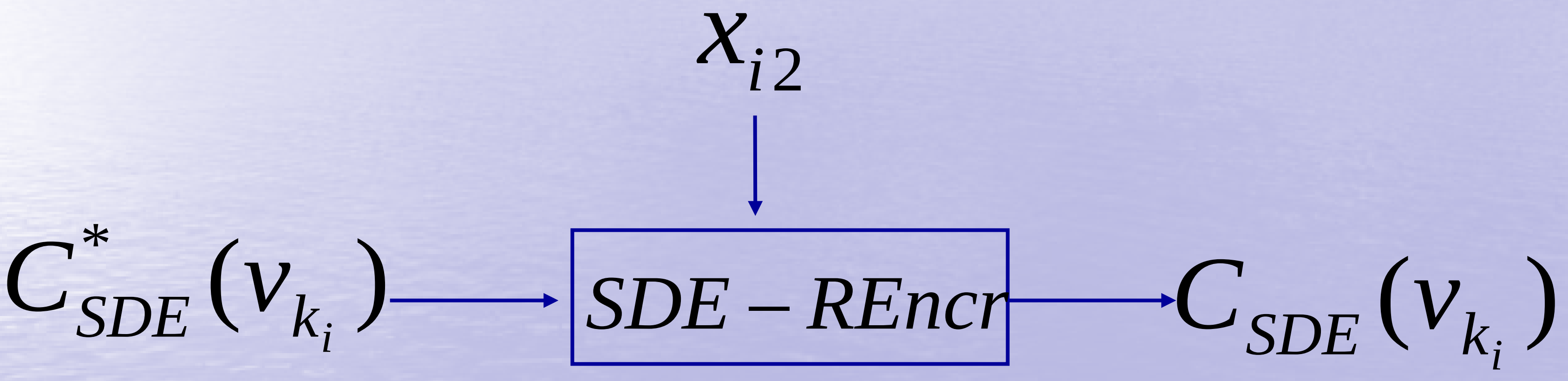
Server-side PE-Re-Encryption

$$C^{*}_{PE}(a_k) = (g^{r_k}, g^{r_k x_{i1}} a_k)$$

$$(g^{r_k})^{x_{i2}} \cdot g^{r_k x_{i1}} a_k = g^{r_k x_{i2} + r_k x_{i1}} a_k = g^{r_k x} a_k$$

$$C_{PE}(a_k) = (g^{r_k}, g^{r_k x} a_k)$$

Server-side SDE Re-Encryption



x_{i2}

$C^{*}_{SDE}(v_{k_i})$

$SDE-REncr$

$C_{SDE}(v_{k_i})$

Server-side SDE Re-Encryption

$$C^{*}_{SDE}(v_{k_i}) = \begin{cases} \hat{c}_1 = g^{r_{k_i}+\sigma_{k_i}} \\ \hat{c}_2 = \hat{c}_1^{\,x_{i1}} \\ \hat{c}_3 = H(h^{r_{k_i}}) \end{cases}$$

Server-side SDE Re-Encryption

$$c_1 = (\hat{c}_1)^{x_{i2}} \cdot \hat{c}_2 = (\hat{c}_1)^{x_{i2}+x_{i1}} = \hat{c}_1{}^{x} =$$

$$= (g^{r_{k_i}+\sigma_{k_i}})^{x} = h^{r_{k_i}+\sigma_{k_i}}$$

$$c_2 = \hat{c}_3 = H(h^{r_{k_i}})$$

$$C_{SDE}(v_{k_i}) = (c_1, c_2)$$

Storing in the Encrypted DB

id	$f_s(attr\_name_1)$	…	$f_s(attr\_name_n)$
1	$\{a_1\}_{PE}$ $\{v_{11}\}_{SDE},\ldots,\{v_{1m}\}_{SDE}$	…	$\{a_n\}_{PE}$ $\{v_{n1}\}_{SDE},\ldots,\{v_{nl}\}_{SDE}$

Performing a Query

User u_j

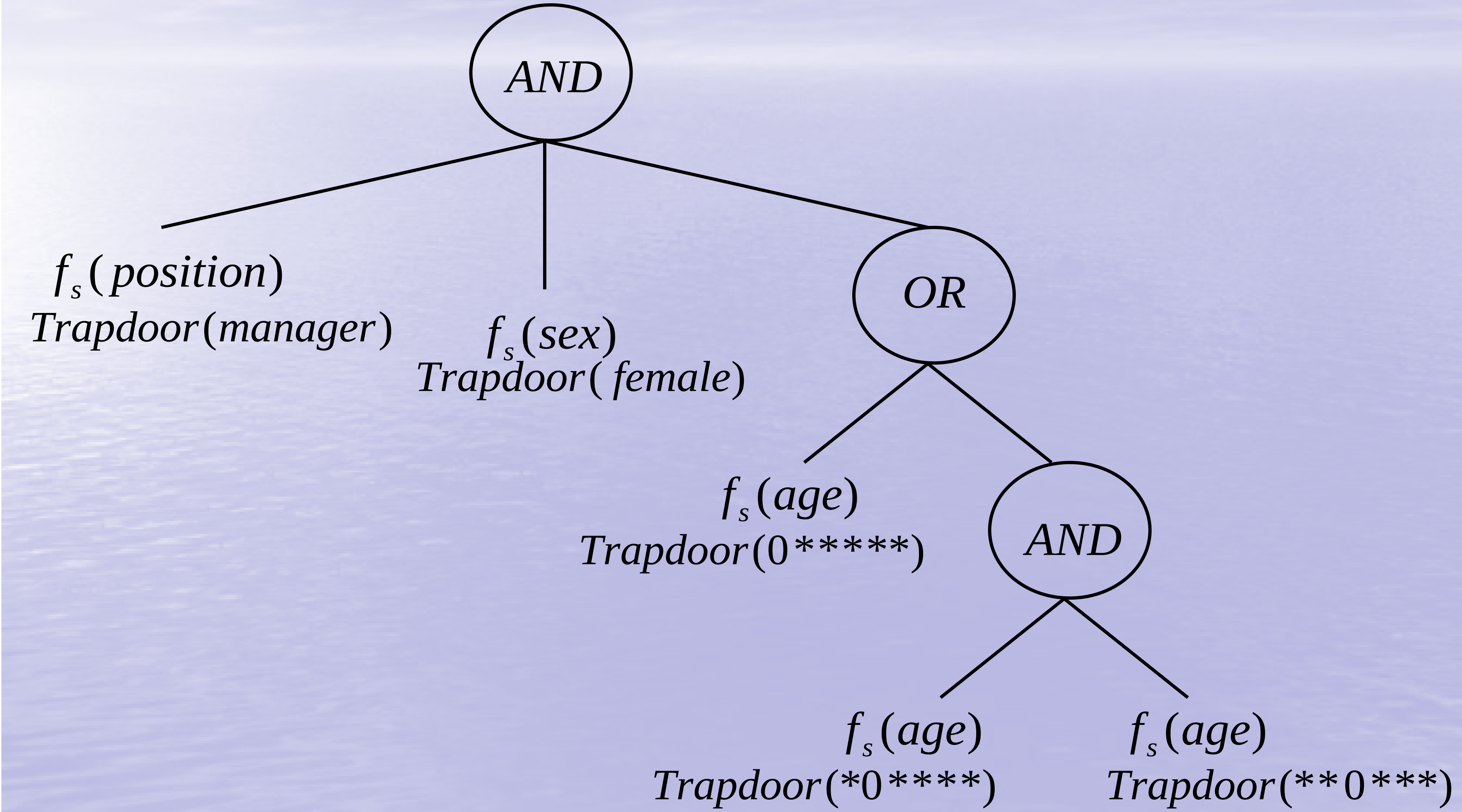
$$Q: SELECT\ \ name\ FROM\ \ Personnel$$

$$WHERE\ \ position = manger\ and$$

$$sex = \ female\ and$$

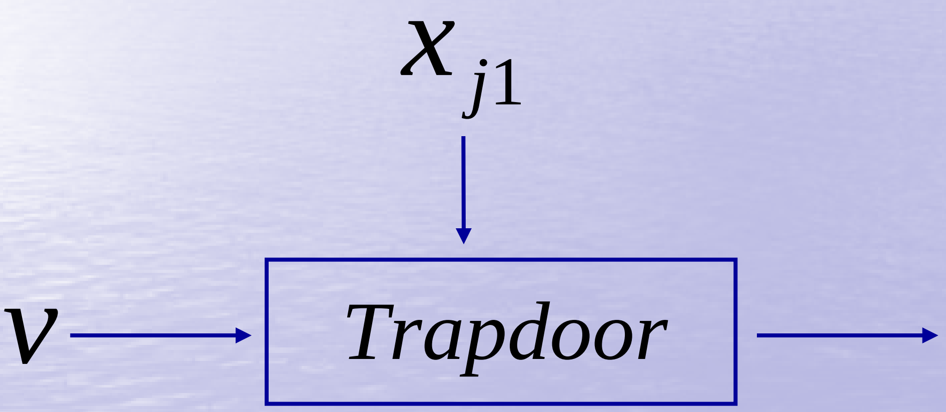
$$age < 40$$

Tree Representation



Generation of Trapdoors

User u_j $\quad$ $Ku_j(x_{j1}, s)$



x_{j1}

v

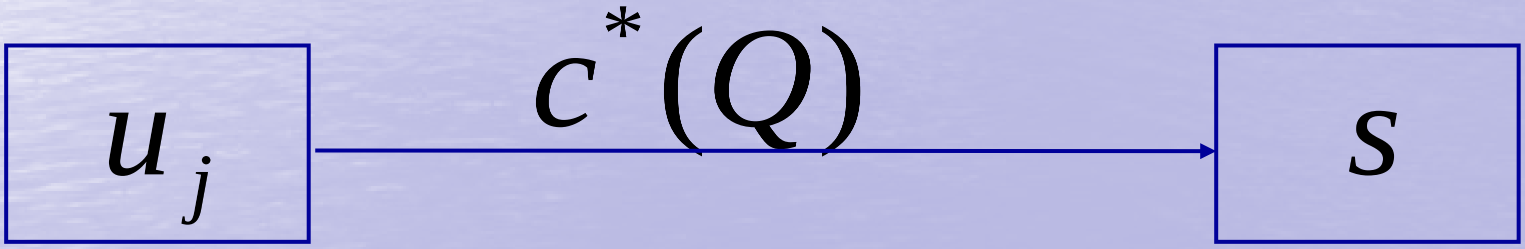
Trapdoor

$$T_j(v) = (t_1, t_2)$$

$$t_1 = g^{-r_v} g^{\sigma_v}$$

$$t_2 = h^{r_v} g^{-x_{j1} r_v} g^{x_{j1}\sigma_v} = g^{x_{j2} r_v} g^{x_{j1}\sigma_v}$$

Sending the query to the Cloud

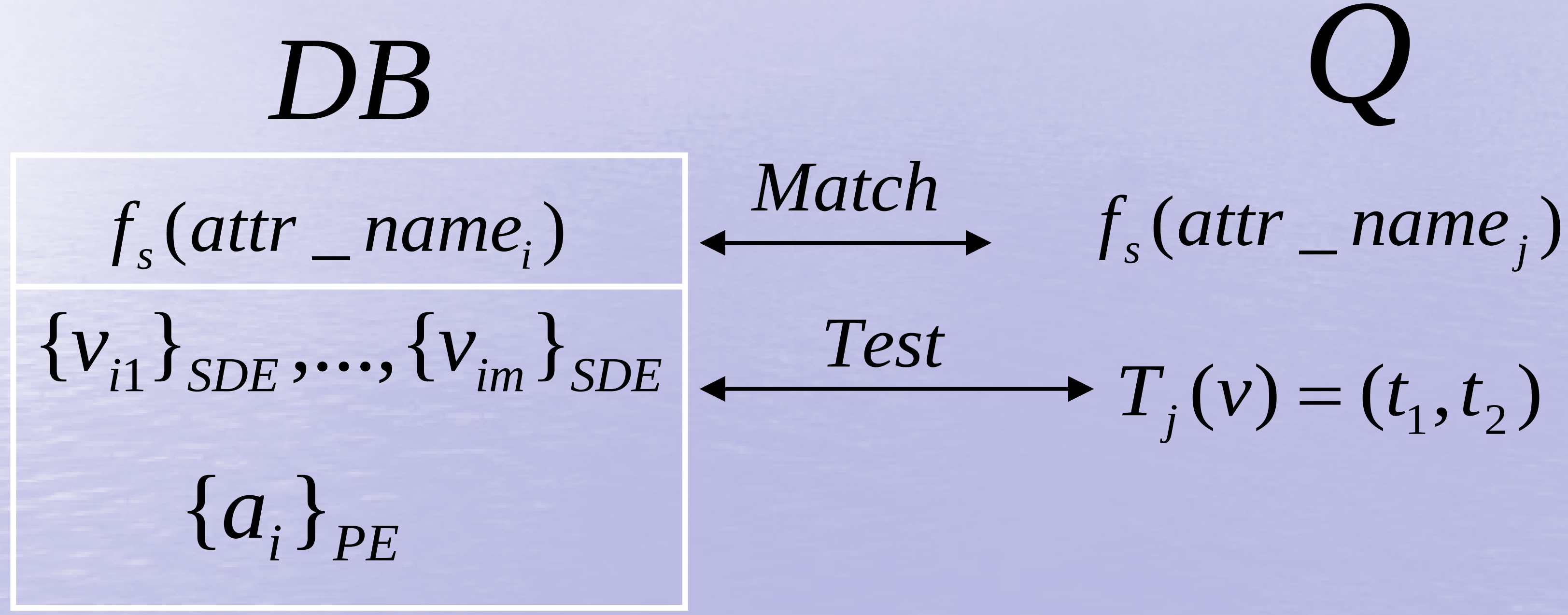


u_j

$c^*(Q)$

s

Performing the Search on the Cloud



DB

Q

$f_s(attr\_name_i)$

Match

$f_s(attr\_name_j)$

$\{v_{i1}\}_{SDE},\ldots,\{v_{im}\}_{SDE}$

Test

$T_j(v)=(t_1,t_2)$

$\{a_i\}_{PE}$

Performing the Test

$$Ks_j(j, x_{j2})$$

$$\begin{tikzcd}
 & x_{j2} \arrow[d] & \\
T_j(v) \arrow[r] & \boxed{Test} \arrow[r] & T = g^{x\sigma_v}
\end{tikzcd}$$

$$T = t_1^{x_{j2}} \cdot t_2 = (g^{-r_v} g^{\sigma_v})^{x_{j2}} \cdot g^{x_{j2} r_v} g^{x_{j1}\sigma_v} =$$

$$= g^{-r_v x_{j2}} g^{\sigma_v x_{j2}} \cdot g^{x_{j2} r_v} g^{x_{j1}\sigma_v} = g^{x\sigma_v}$$

Performing the Test

$$C_{SDE}(v_{k_i}) = (c_1, c_2)$$

$$c_1 = h^{r_{k_i} + \sigma_{k_i}}$$

$$c_2 = H(h^{r_{k_i}})$$

Performing the Test

$$c_2 \overset{?}{=} H(c_1 \cdot T^{-1})$$

$$H(h^{r_{k_i}}) \overset{?}{=} H(h^{r_{k_i}+\sigma_{k_i}} \cdot g^{-x\sigma_v})$$

$$H(h^{r_{k_i}}) \overset{?}{=} H(h^{r_{k_i}} \cdot g^{x\sigma_{k_i}} \cdot g^{-x\sigma_v})$$

$$\sigma_{k_i} \overset{?}{=} \sigma_v$$

$$\sigma_{k_i} = f_s(value_{k_i}) \qquad \sigma_v = f_s(value)$$

PE-Pre-Decryption

$$C_{PE}(a_k) = (g^{r_k}, g^{r_k x}\; a_k) \qquad Ks_j(j, x_{j2})$$

$$\begin{tikzcd}
 & x_{j2} \arrow[d] & \\
C_{PE}(a_k) \arrow[r] & \boxed{PE - RDecr} \arrow[r] & C'(a_k)\begin{cases} g^{r_k} \\ g^{r_k x_{j1}} a_k \end{cases}
\end{tikzcd}$$

$$g^{r_k x}\; a_k \cdot (g^{r_k})^{-x_{j2}} = g^{r_k x}\; a_k \cdot g^{-r_k x_{j2}} =$$

$$= g^{r_k(x\; - x_{j2})} a_k = g^{r_k x_{j1}} a_k$$

PE-Pre-Decryption

$$C_{PE}(a_k) = (g^{r_k}, g^{r_k x} a_k)$$

$$(g^{r_k})^{-x_{j2}}$$

PE-Pre-Decryption

$$C_{PE}(a_k) = (g^{r_k}, g^{r_k x} a_k)$$

$$(g^{r_k})^{-x_{j2}} \cdot g^{r_k x} a_k$$

PE-Pre-Decryption

$$C_{PE}(a_k) = (g^{r_k}, g^{r_k x}\ a_k)$$

$$(g^{r_k})^{-x_{j2}} \cdot g^{r_k x}\ a_k = g^{-r_k x_{j2}} \cdot g^{r_k x}\ a_k =$$

$$= g^{r_k(x\ -x_{j2})} a_k = g^{r_k x_{j1}} a_k$$

PE-Pre-Decryption

$$C_{PE}(a_k) = (g^{r_k}, g^{r_k x} a_k)$$

$$(g^{r_k})^{-x_{j2}} \cdot g^{r_k x} a_k = g^{-r_k x_{j2}} \cdot g^{r_k x} a_k =$$

$$= g^{r_k(x - x_{j2})} a_k = g^{r_k x_{j1}} a_k$$

$$\Big\downarrow$$

$$C'(a_k) = (g^{r_k}, g^{r_k x_{j1}} a_k)$$

PE-Pre-Decryption

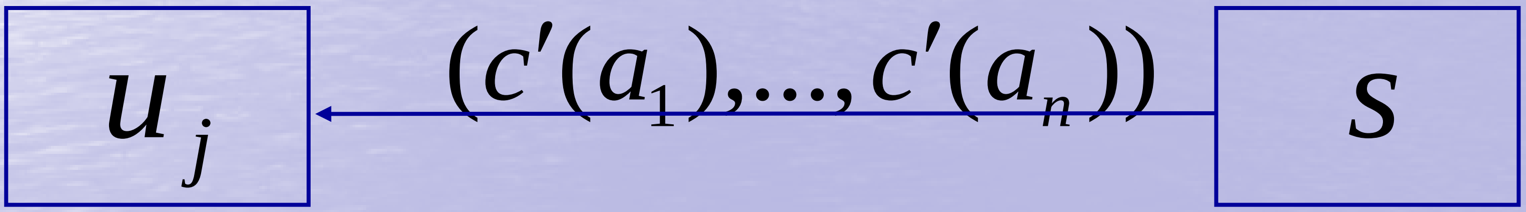
$$C_{PE}(a_k) = (g^{r_k}, g^{r_k x} a_k)$$

$$(g^{r_k})^{-x_{j2}} \cdot g^{r_k x} a_k = g^{-r_k x_{j2}} \cdot g^{r_k x} a_k =$$

$$= g^{r_k(x - x_{j2})} a_k = g^{r_k x_{j1}} a_k$$

$$C'(a_k) = (g^{r_k}, g^{r_k x_{j1}} a_k)$$

Sending the result to the User



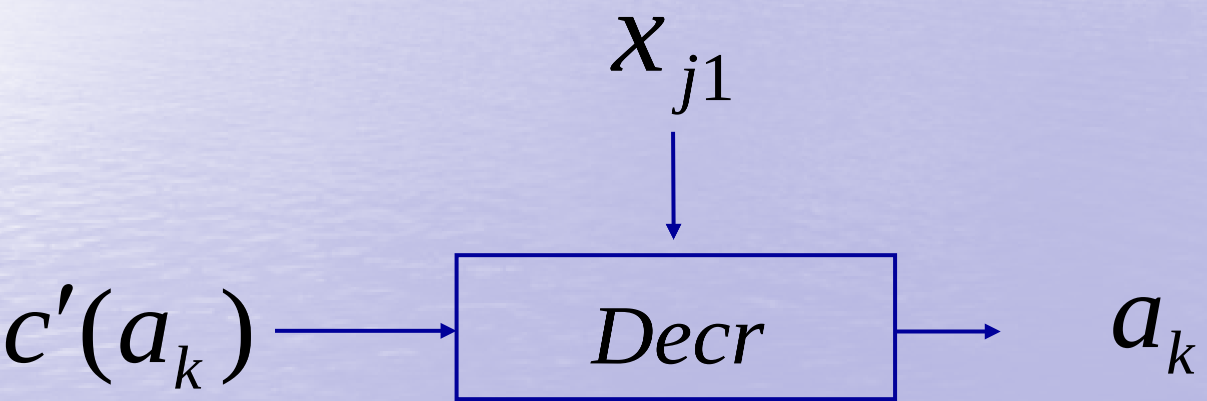
u_j

$(c'(a_1),\ldots,c'(a_n))$

s

Final Decrypt

User u_j $\qquad Ku_j(x_{j1}, s)$



x_{j1}

$c'(a_k)$

$Decr$

a_k

$$g^{r_k x_{j1}} a_k \cdot (g^{r_k})^{-x_{j1}} = a_k$$

Final Remarks

- Is this The Solution for Cloud Confidentiality?
  - Efficiency: Decently fast although search on numerical values can take seconds
  - Collusions
  - Fine-grained access: Instead of PE we could use other encryption schemes (ABE)
  - Query structure in clear

Take Away

- Confidentiality solutions exist but still more needs to be done
- More effort from the cloud providers towards security solutions