



## Big Data is Useful

- Cloud computing
- Efficiencies in cost, productivity, innovation
- Development: fighting poverty
  - Scientists combined satellite imagery and machine learning to predict poverty
- SOCIAL YEAR SPROCESS WOLUME SHARES SALVES SALVES ANALYSIS OF STORAGE BIG STORAGE BIG SALVES S

- Smart Cities & Meters
- Healthcare

Can we use all this technology without violating individual privacy?

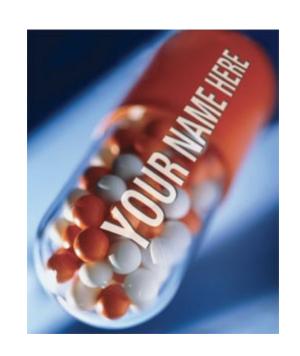


Paranoia about Privacy -> Reluctance to share data Conflict with Big Data Benefits



#### Personalized Medicine

"The dream for tomorrow's medicine is to understand the links between DNA and disease — and to tailor therapies accordingly. But scientists have a problem: how to keep genetic data and medical records secure while still enabling the massive, cloud-based analyses needed to make meaningful associations."



Erika Check Hayden, Nature, 2015

# Case Study: Cloud Computing



#### Cloud Computing

Having secure access to all your applications and data from any network device

#### Functionality:

- Store big data on cloud
- Cloud can perform expensive computations for us

#### Plus Security:

- Access control on encrypted data
- Compute on encrypted data (eg medical research)

#### Plus Efficiency:

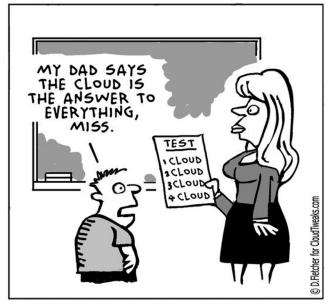
- Running time should not depend on database size
- Support inputs of variable size

### What we want from a database in the cloud?

#### **SECURITY**



#### **FUNCTIONALITY**



#### **EFFICIENCY**



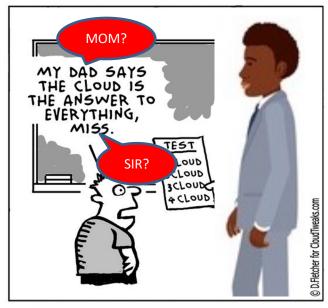
Sources: cloudtweaks.com, officeguycartoons.com

### What we want from a database in the cloud?

#### **SECURITY**



#### **FUNCTIONALITY**



#### **EFFICIENCY**

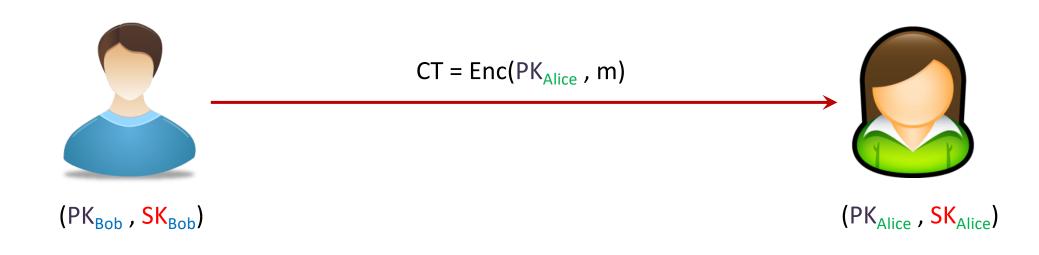


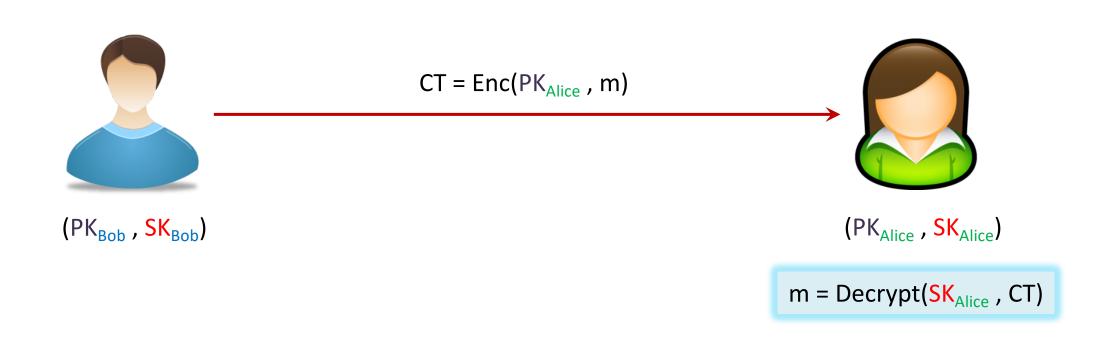
Sources: cloudtweaks.com, officeguycartoons.com

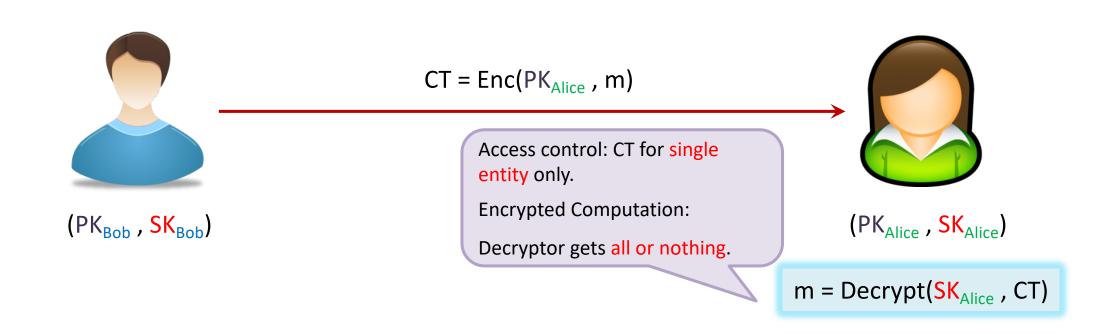








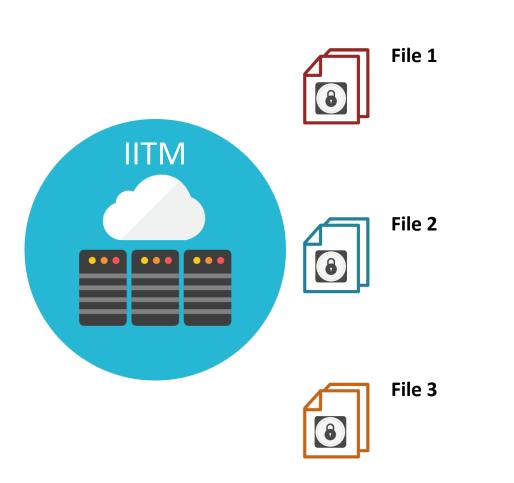




#### **Basic Requirement**

## Access control over encrypted data

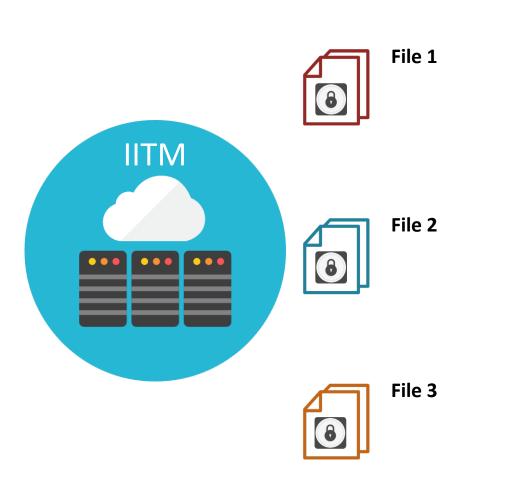
**Attribute Based Encryption!** 

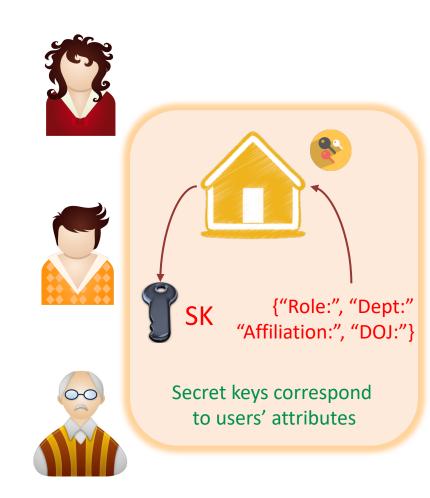


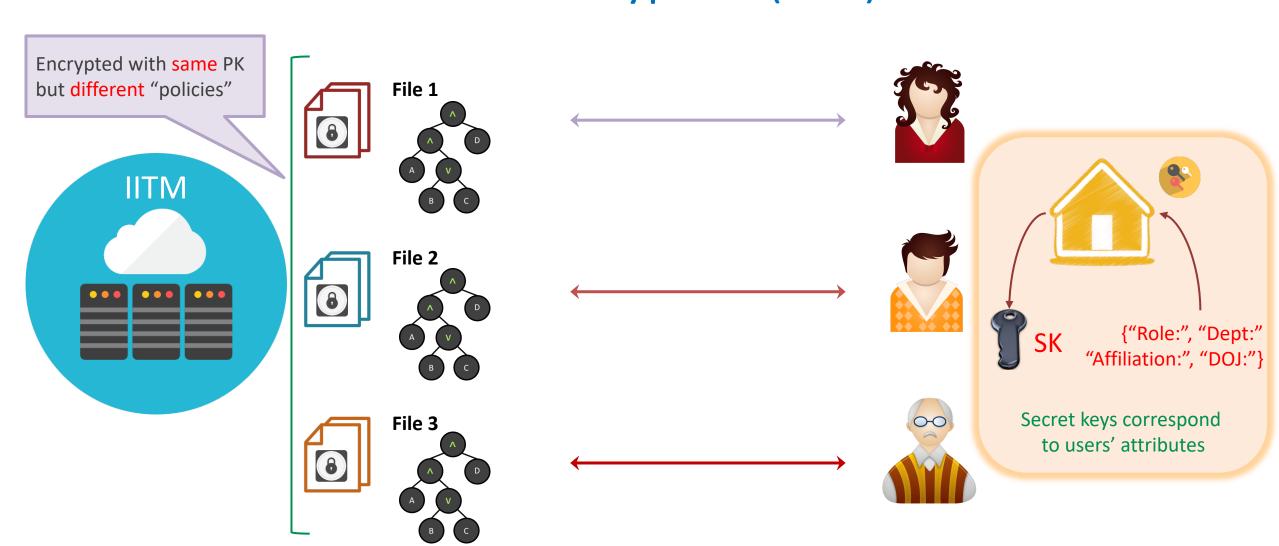


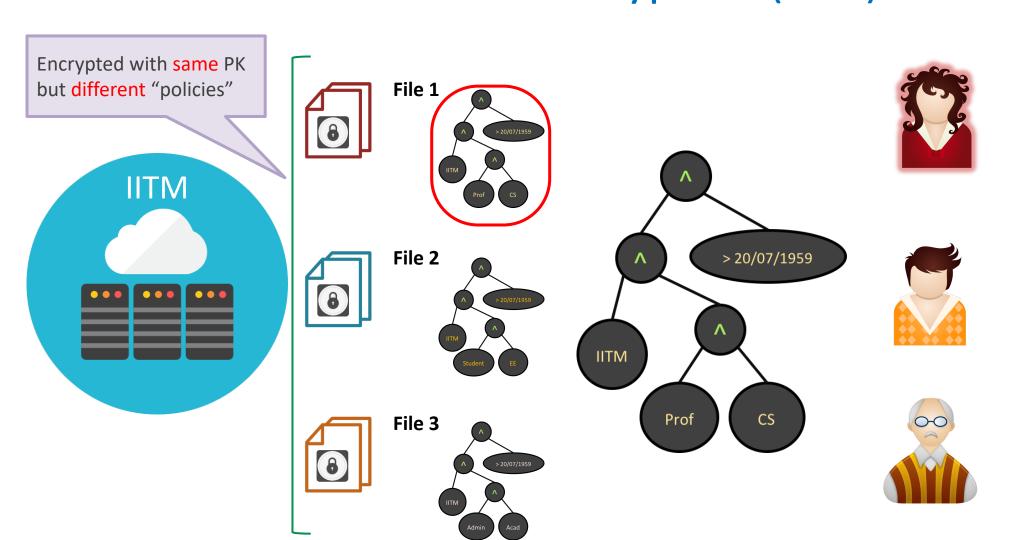












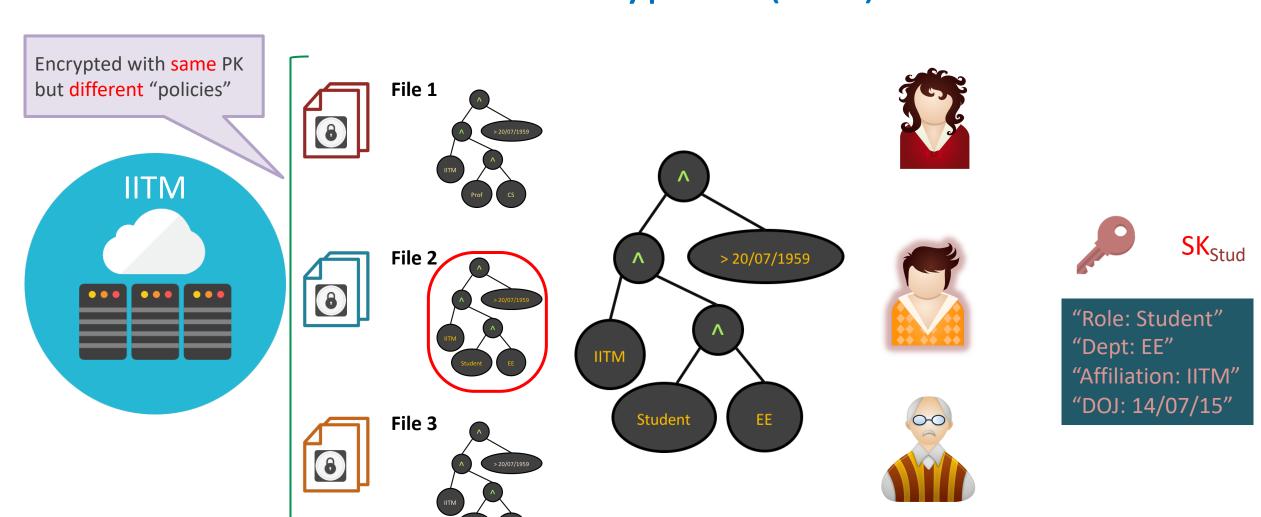


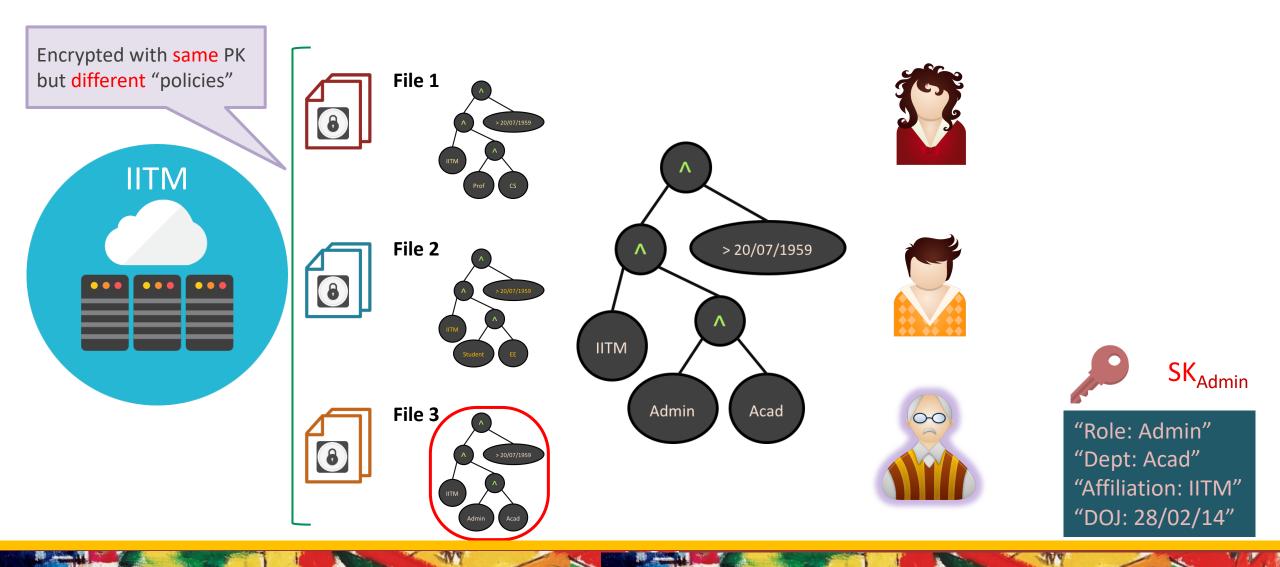
 $SK_{Prof}$ 

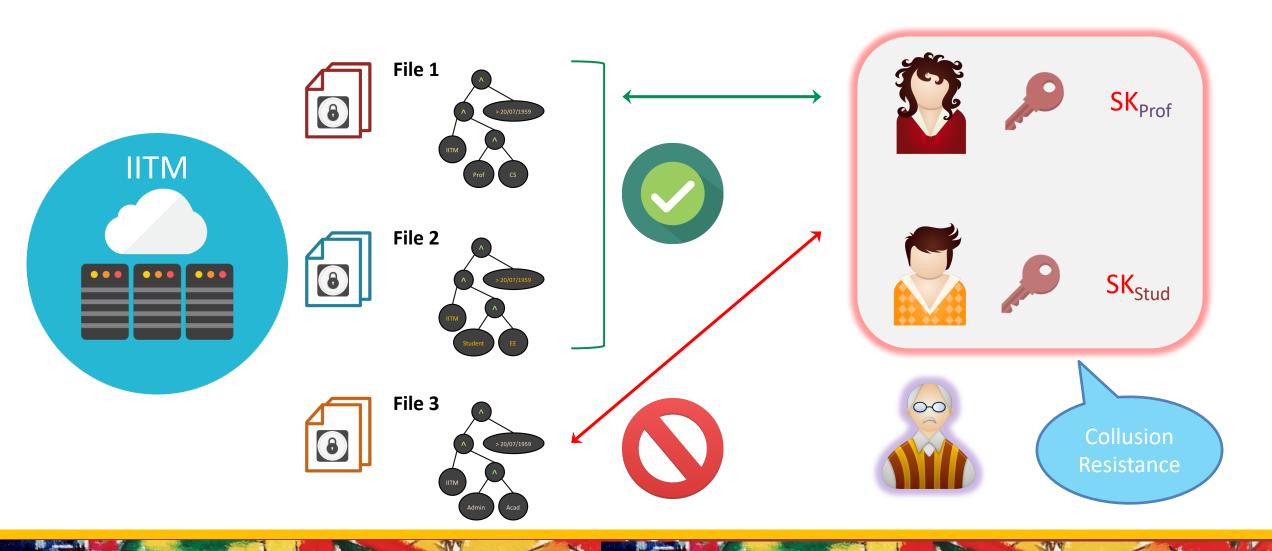
"Role: Professor"

"Dept: CS"
"Affiliation: IITM"

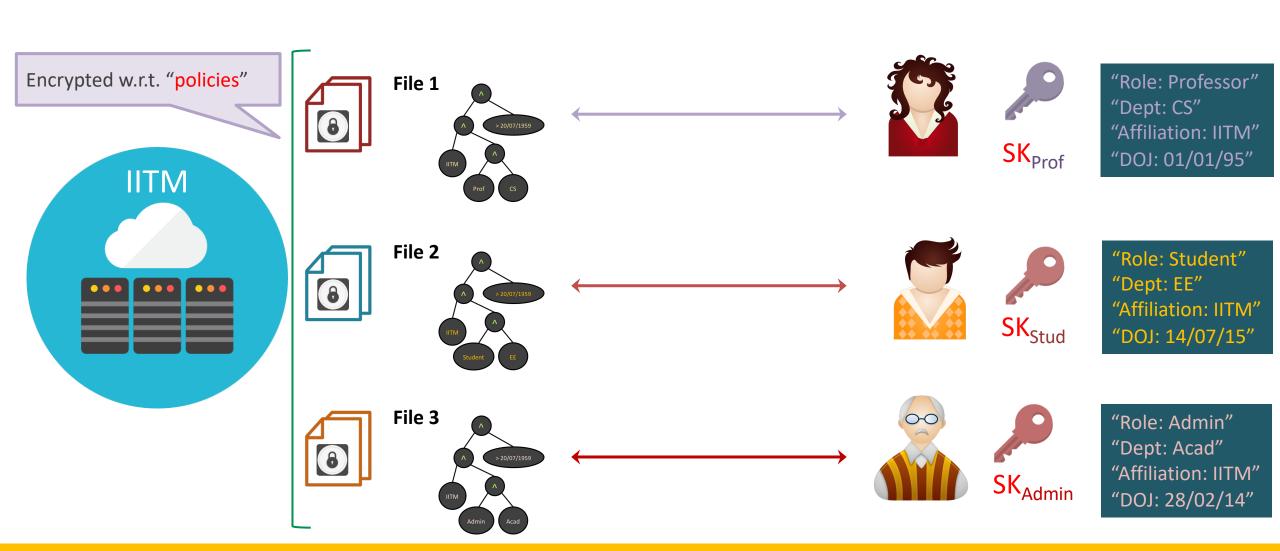
"DOJ: 01/01/95"



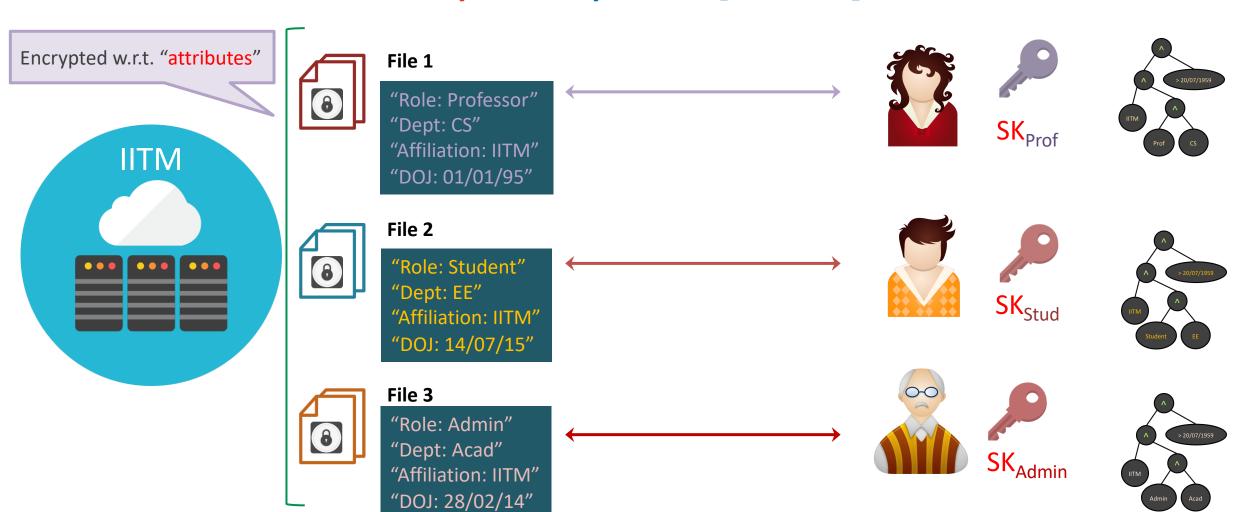




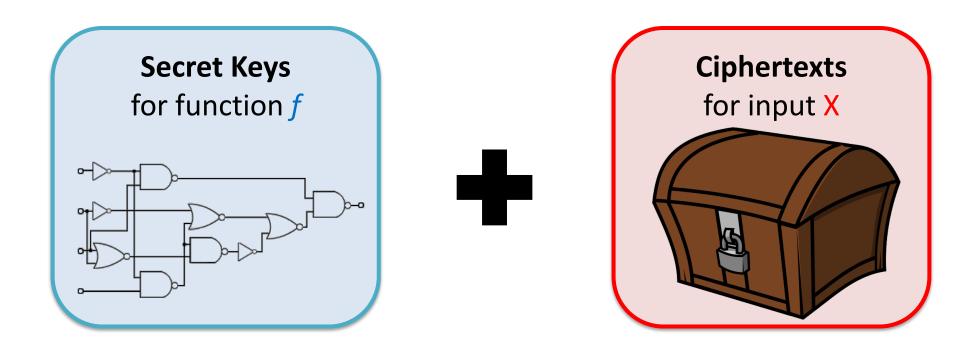
### Ciphertext-Policy ABE



### Key-Policy ABE [SW05]

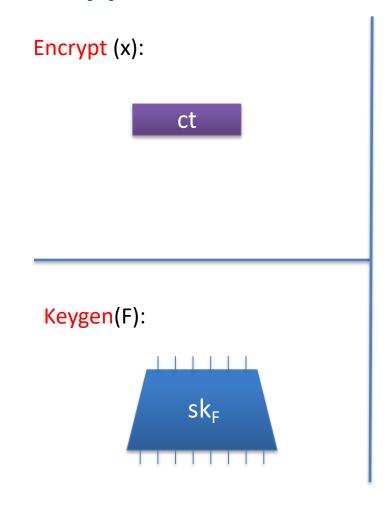


## Generalization: Functional Encryption [BSW11, O'N10]

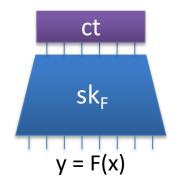


Decrypt to learn f(X)

### **Encryption with Partial Decryption Keys**



Decrypt (sk<sub>F</sub>, ct):



Security:

Adversary possessing keys for multiple circuits  $F_i$  cannot distinguish  $Enc(x_0)$  from  $Enc(x_1)$  unless  $F_i(x_0) \neq F_i(x_1)$ 

#### Personalized Medicine?

Encrypt(x) x = genomic data of users

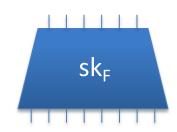
ct(Maggie)

ct(Chris)

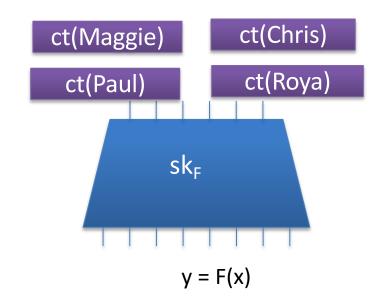
ct(Paul)

ct(Roya)

Keygen(F) F: some medical research algo



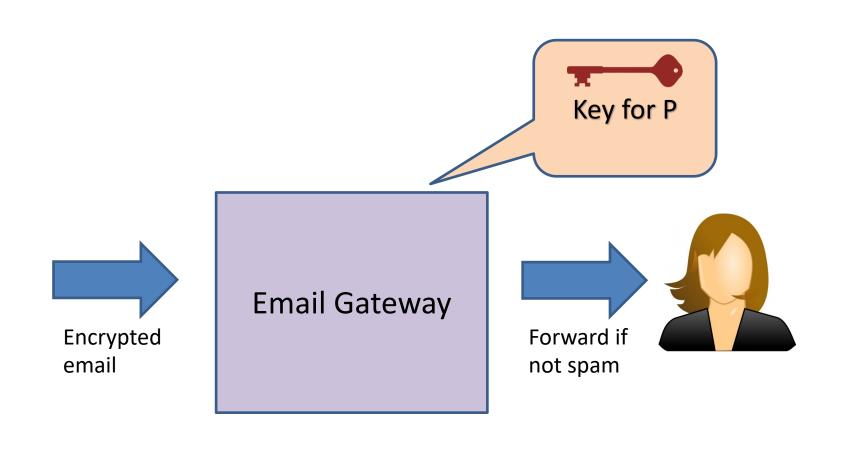
Decrypt (sk<sub>F</sub>, ct):



Security: No one's personal genomic data is leaked!

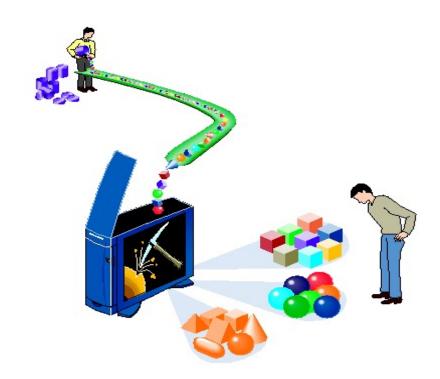
# Spam Detection on Encrypted Email

Say we have a program P to detect spam on unencrypted email.

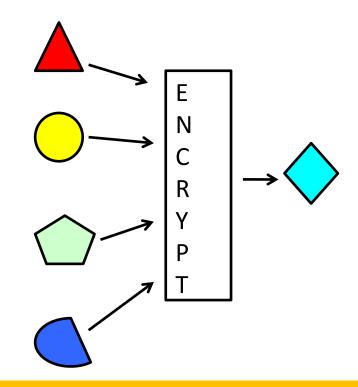


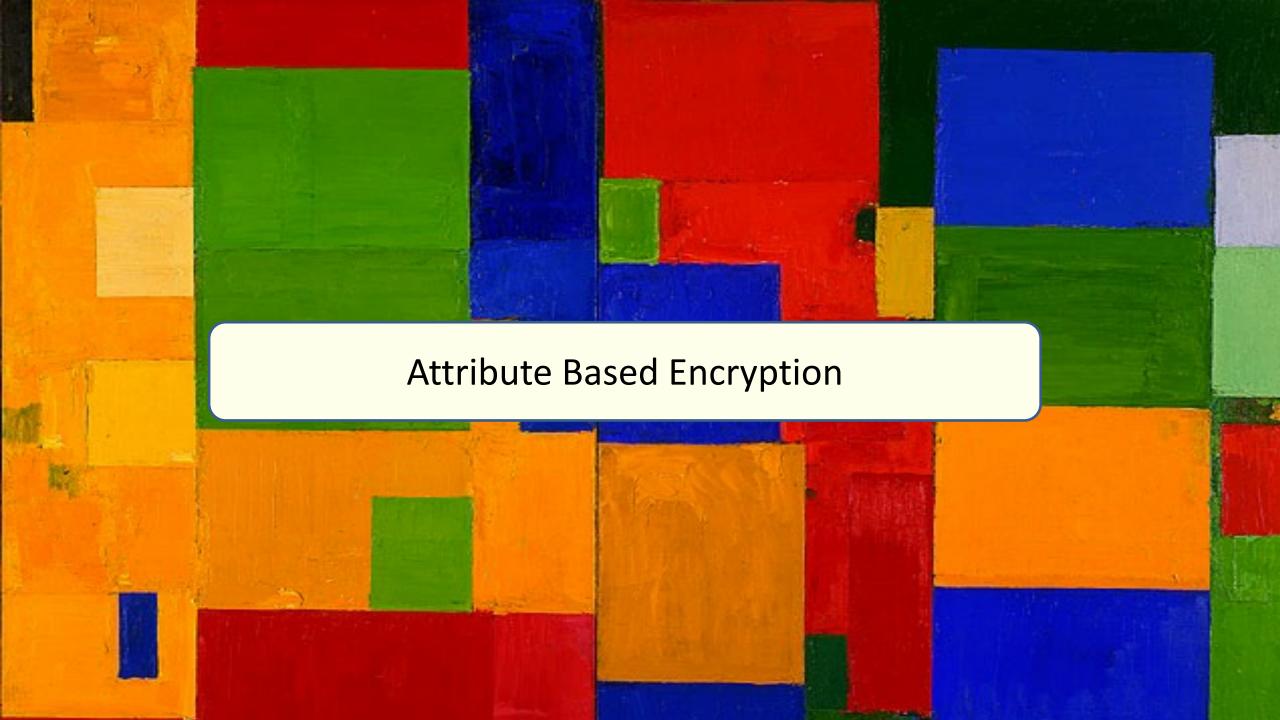
## Challenges: Walking the Fine Line

Want to compute on encrypted data so as to be useful, e.g. Data Mining



All encrypted data should look "the same" so as to hide underlying plaintext





#### **Prior Work**

#### **Key Policy**

- Restricted circuit classes (point functions, threshold functions, NC<sub>1</sub> circuits...): [SW05, GPSW06, BW07, KSW08, LOS+10, OT10, OT12, CW14, AFV11, LW11, LW12, Wat12, Wee14, Att14, GV15, AF18, AMY19a, AMY19b
- Polynomial Sized Circuits: GVW13, BGG+14, GVW15, BV16

#### **Ciphertext Policy**

- Restricted circuit classes (point functions, threshold functions, NC<sub>1</sub> circuits...): BSW07, Wat11, LOS+10, OT10, LW12, RW13, Att14, Wee14, AHY15, CGW15, AC17, KW19, AMY19b, Tsa19,KNYY20,AY20a]
- Polynomial Sized Circuits: ?

#### Our Results

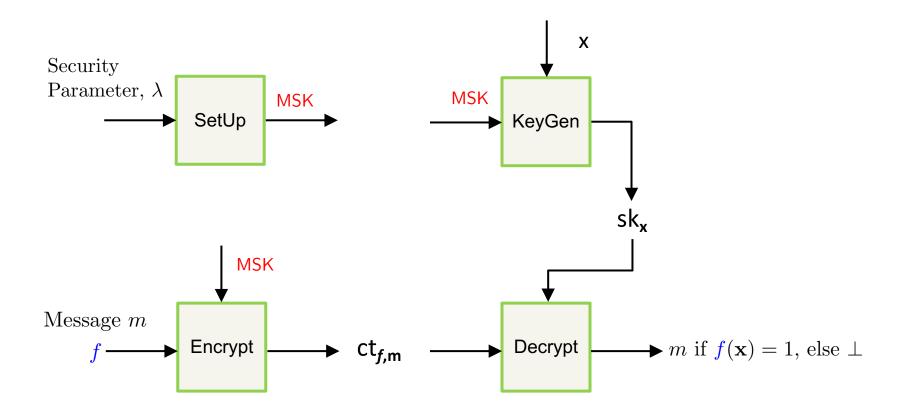
#### **Key Policy**

- Restricted circuit classes (point functions, threshold functions, NC<sub>1</sub> circuits...): [SW05, GPSW06, BW07, KSW08, LOS+10, OT10, OT12, CW14, AFV11, LW11, LW12, Wat12, Wee14, Att14, GV15, AF18, AMY19a, AMY19b
- Polynomial Sized Circuits: GVW13, BGG+14, GVW15, BV16

#### **Ciphertext Policy**

- Restricted circuit classes (point functions, threshold functions, NC<sub>1</sub> circuits...): BSW07, Wat11, LOS+10, OT10, LW12, RW13, Att14, Wee14, AHY15, CGW15, AC17, KW19, AMY19b, Tsa19,KNYY20,AY20a]
- Polynomial Sized Circuits: AY20 (symmetric key, from Learning With Errors)

## SK Ciphertext-Policy ABE [swo5, GPSW06]



### Folklore Approach: Via Universal Circuits

Universal Circuit U: Takes as input f, x and computes U(f, x) = f(x).



Input

**Folklore Transformation** 



**Function** 









Use key policy ABE for circuits (GVW13,BGG+14) to build ciphertext policy ABE.

- CP-ABE Encryption creates KP-ABE ciphertext for (f, m)
- CP-ABE KeyGen creates KP-ABE key for function U<sub>x</sub>
- CP-ABE Decryption is same as KP-ABE decryption

#### Drawbacks:

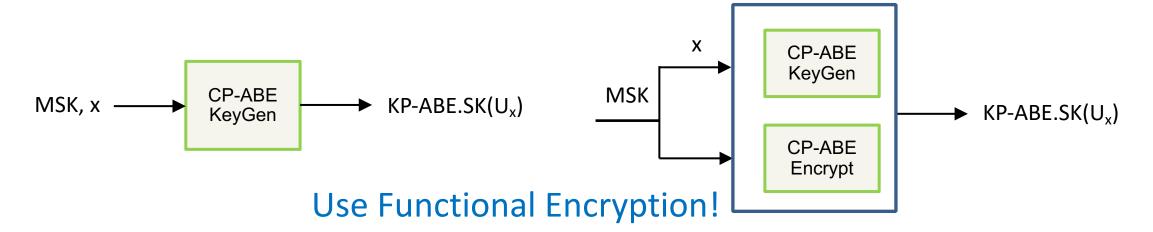
- 1. Size of PK, CT, runtime of setup, Kgen, Enc, Dec grow with max function size  $f_{max}$
- 2. Cannot support unbounded size circuits

### Redistributing Computation [AMY19]

To support unbounded circuits: Only encrypt and decrypt should depend on circuit size

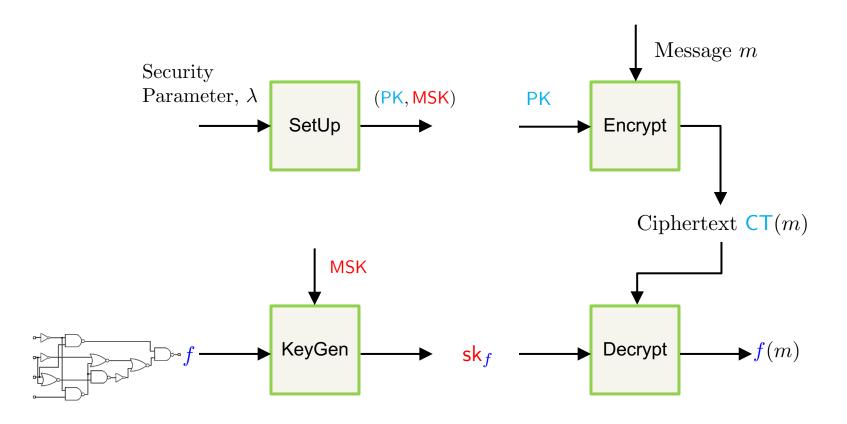
Problem : Computing KP-ABE keygen for function  $U_x$  forces CP-ABE keygen to depend on  $f_{max}$  since  $U_x$  must support inputs of size  $f_{max}$ 

Idea: Distribute computation of KP-ABE keygen for function  $U_x$  between CP-ABE encrypt and CP-ABE keygen so that each respects efficiency requirement of CP-ABE



## (Key Policy) Functional Encryption (FE)

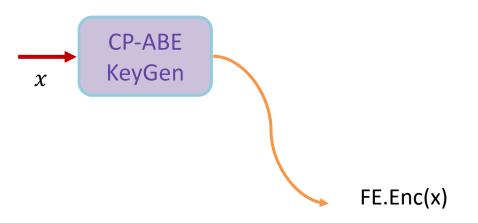
SW05, BSW11, O'N10



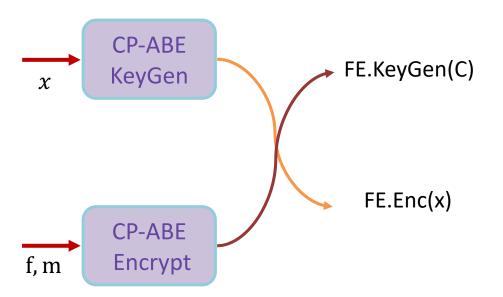
For single key security, can be based on LWE (GKPVZ13, Agr17)

## **Redistributing Computation**

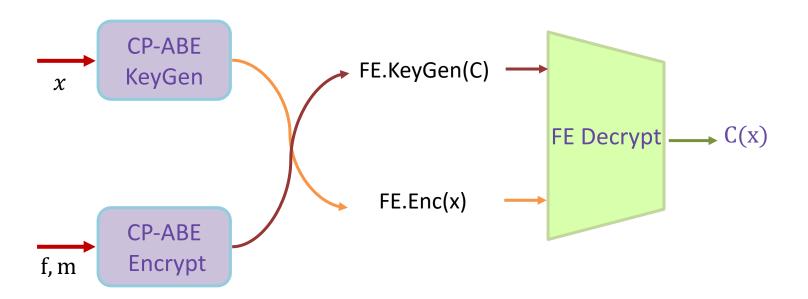
Solution: FE helps to defer the computation of KP-ABE KeyGen



Solution: FE helps to defer the computation of KP-ABE KeyGen

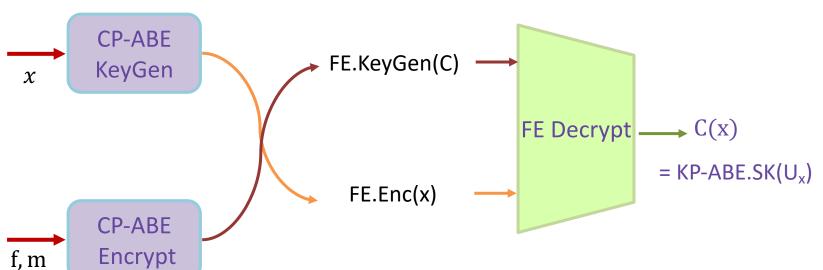


Solution: FE helps to defer the computation of KP-ABE KeyGen



Solution: FE helps to defer the computation of KP-ABE KeyGen

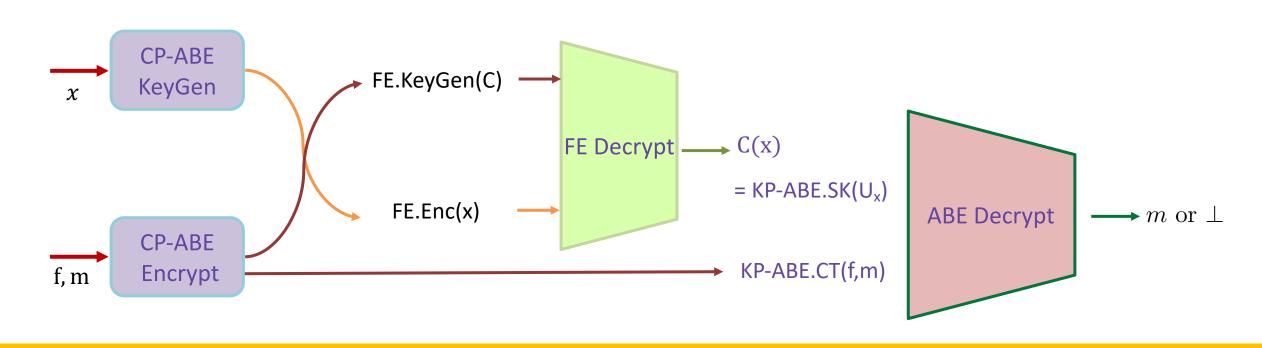
C(x): KP-ABE.KeyGen(MSK,  $U_x$ )

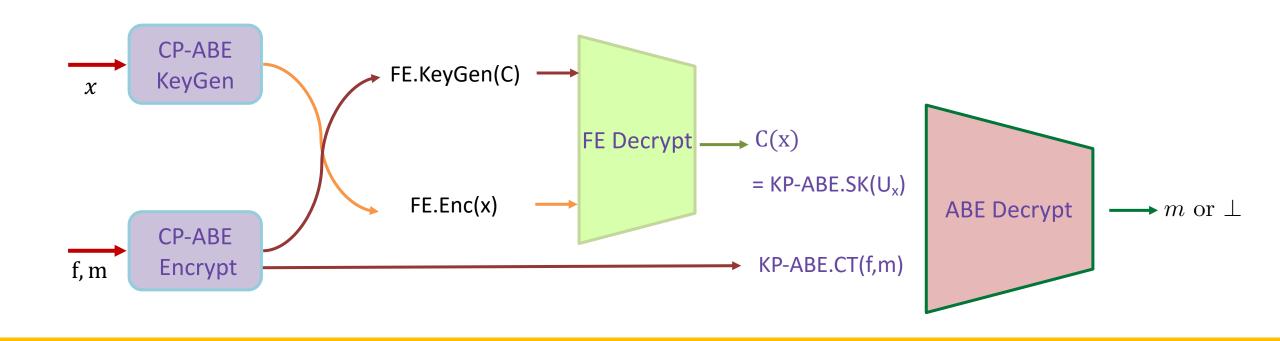


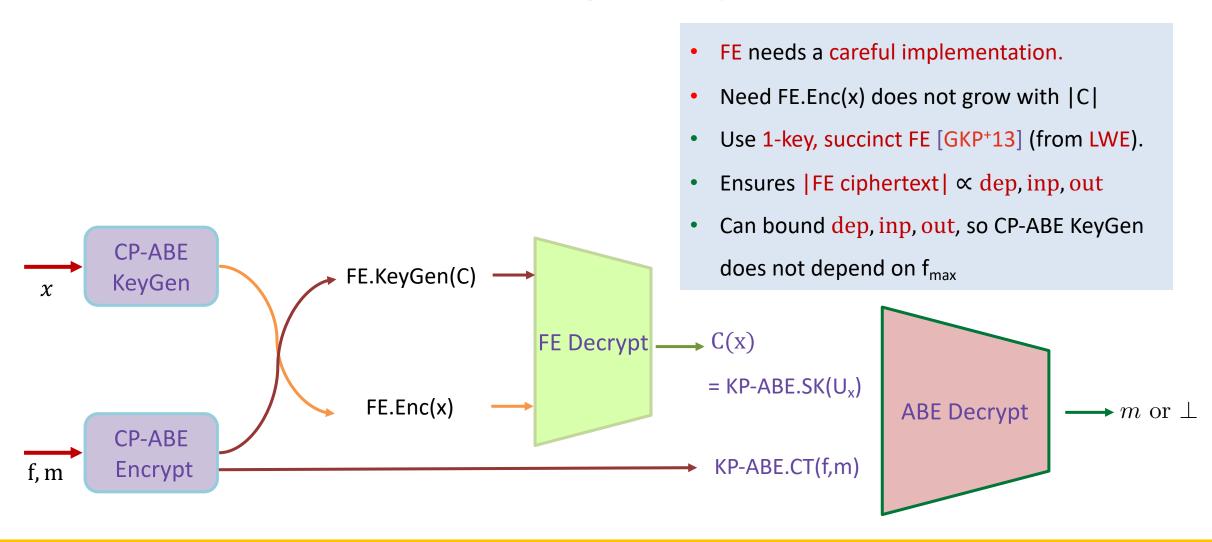
Observe: C depends on  $f_{max}$  but C(x) is short by BGG+14

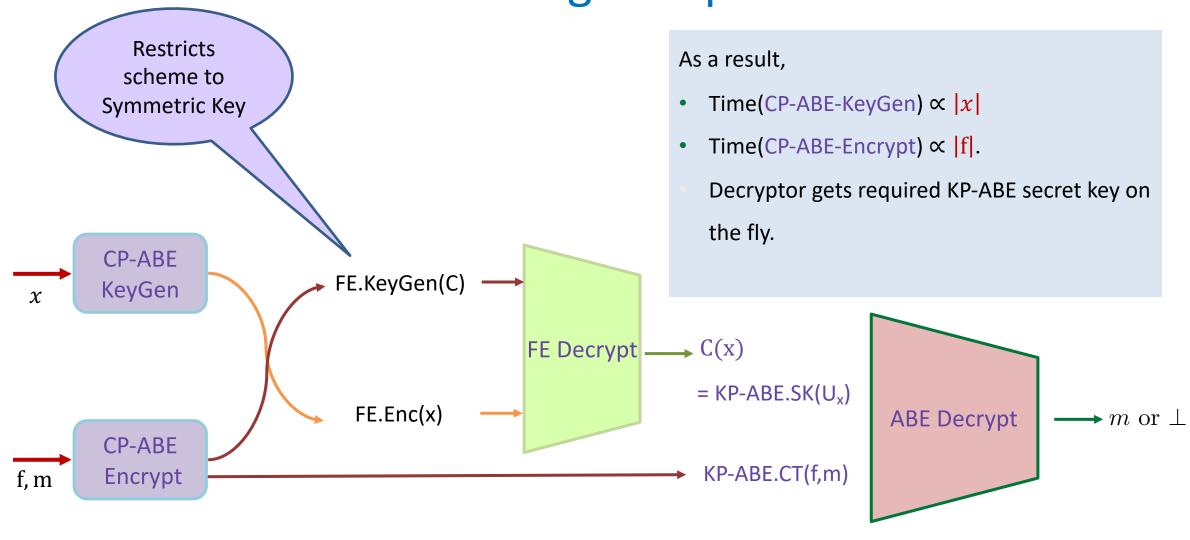
Solution: FE helps to defer the computation of KP-ABE KeyGen

C(x): KP-ABE.KeyGen(MSK,  $U_x$ )







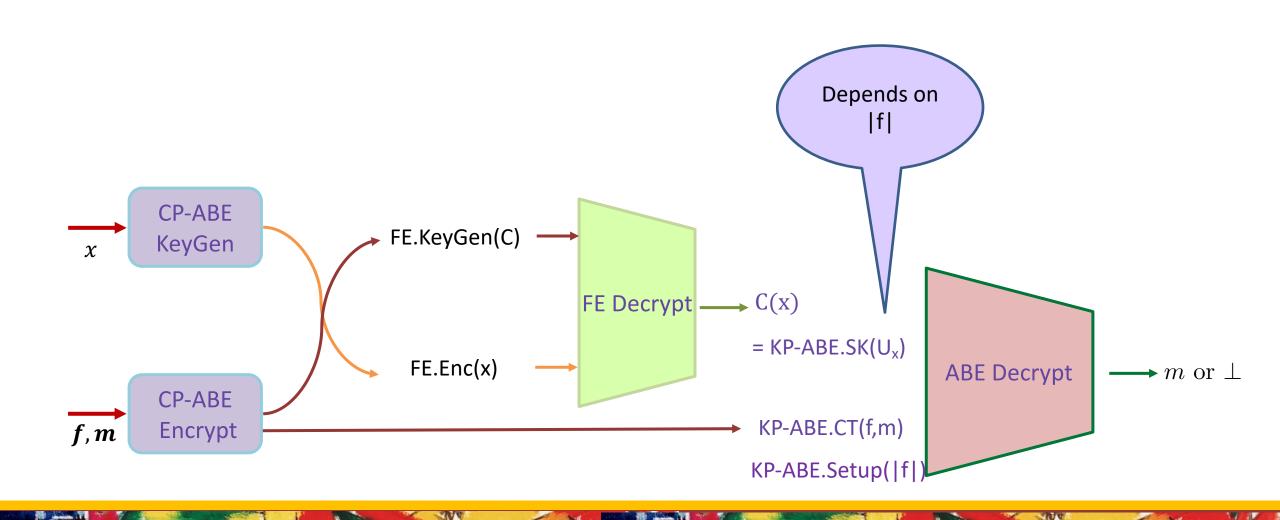


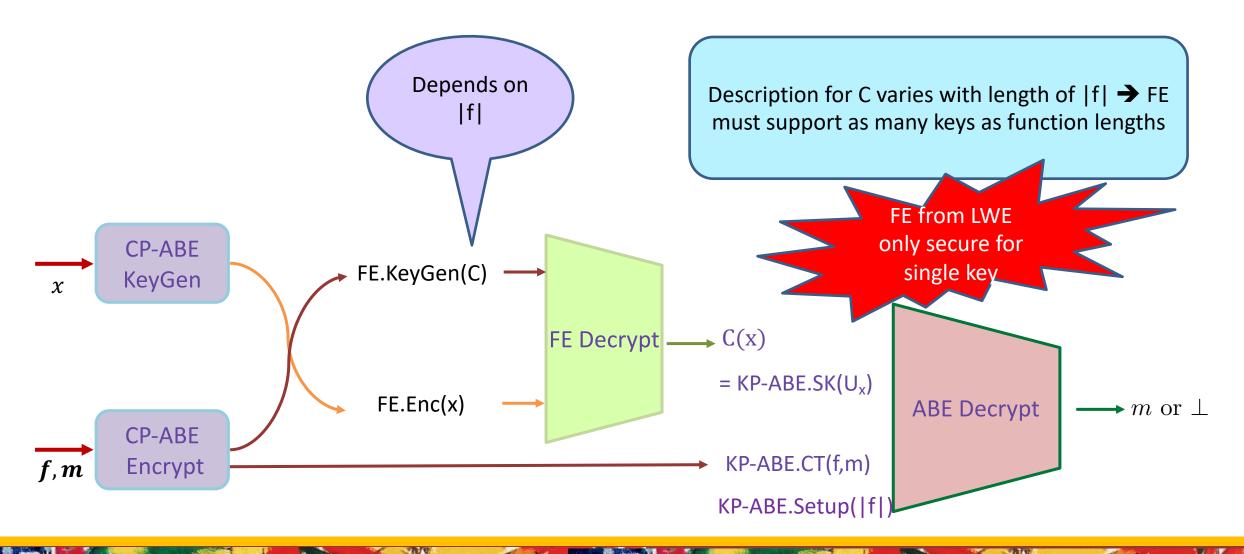
So far, only bounded size circuits....

Problem: If KP-ABE must encrypt f, KP-ABE setup must be initialized with bound f<sub>max</sub>

only supports bounded circuits.

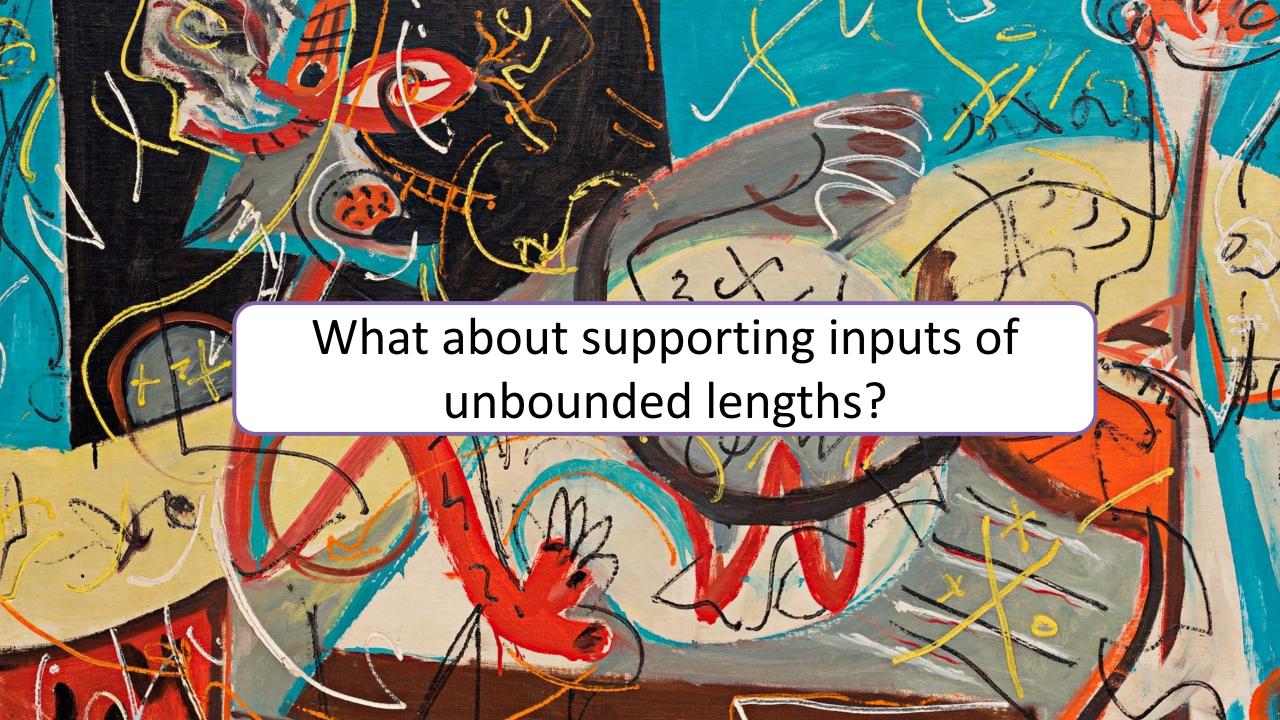
Idea: Let CP-ABE Encrypt, which knows |f|, run KP-ABE setup.
Possible since symmetric Key





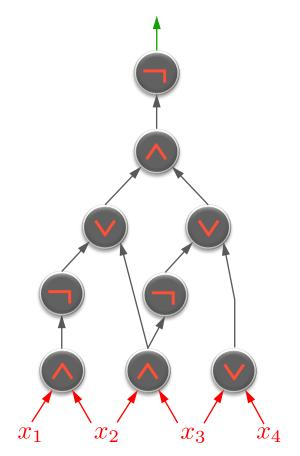
Solution: Pad circuit size to power of 2, and run  $\lambda$  instances of single key FE

CP-ABE KeyGen does not know |f| so computes Fe.Enc for all  $\lambda$  FEs



# Circuits are powerful, but...

- Non uniform model of computation
  - Support fixed length inputs
  - Function description changes based on input length
  - Incurs worst-case runtime over all inputs of a certain length



#### **Uniform Models of Computation**

#### Finite automata, Turing Machines, RAM:

- Supports arbitrary length inputs
- Fixed description for all inputs
- Input-specific runtime

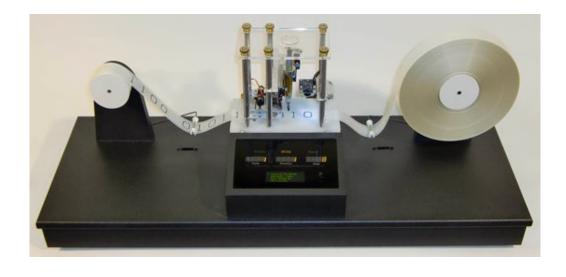


Image Courtesy: <a href="http:/aturingmachine.com/">http:/aturingmachine.com/</a>

# What is known (unbounded inputs)

Construction	Model	KP / CP	Number of Keys	Assumption
[Wat12]	DFA	KP	Unbounded	q-type assumption on bilinear maps
[Att14]	DFA	KP & CP	Unbounded	q-type assumption on bilinear maps
[AS17]	DFA	KP	Single	LWE
[AMY19a]	NFA	KP	Unbounded	LWE
[AMY19b]	DFA	KP & CP	Unbounded	DLIN
[AMSY21]	TM	KP	Bounded	LWE

### Goal: ABE for uniform models of computation

- Encryptor chooses attribute of unbounded (poly) length
- Key corresponds to machine (DFA/NFA/TM/...)
  - Keygen does not know input length.
  - Same key for any input length
- Decryption succeeds iff machine accepts input

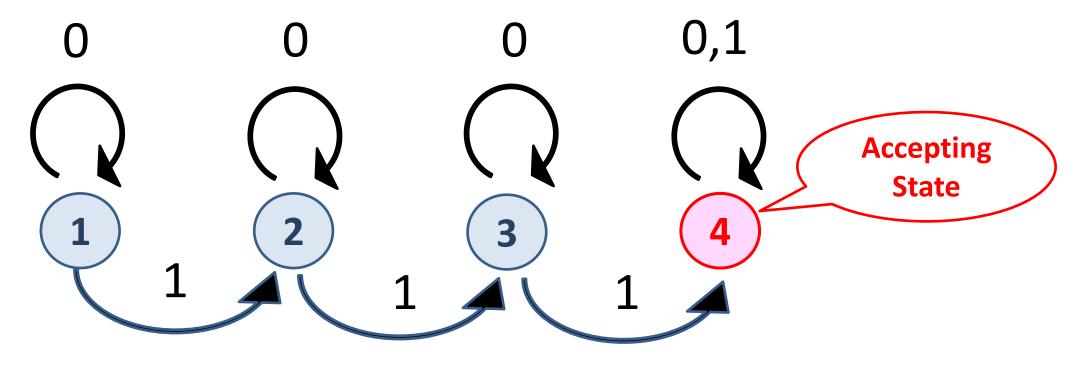
#### Main challenges?

- Previously, encryptor and key generator could agree on input length n
- Assign common "PK component" PK<sub>i</sub> (matrix or group element) for each index i ∈[n], use to compute both CT and SK

#### **Deterministic Finite Automata**

Alphabet:  $\Sigma = \{0,1\}$ ,

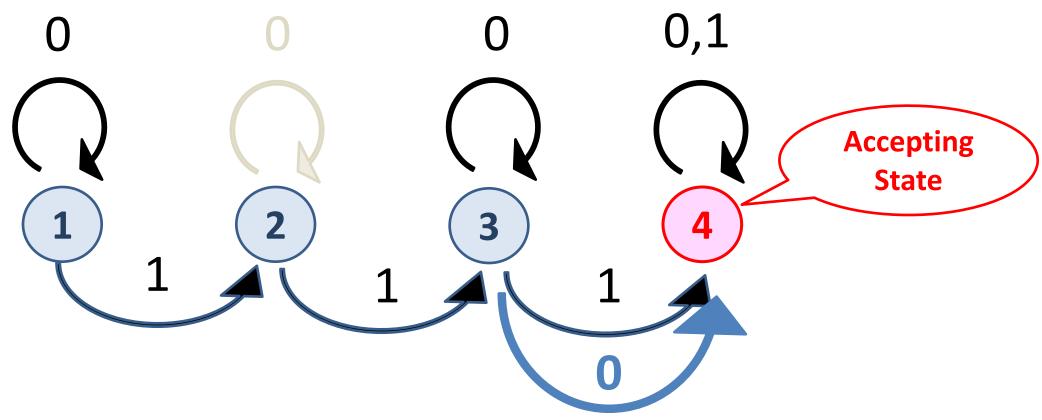
Set of States =  $\{1,2,3,4\}$ , Accepting state =  $\{4\}$ ,



Single outgoing edge from a state w.r.t an alphabet.

$$111000 \Rightarrow Accept$$
  
010100 ⇒ Reject

#### Non Deterministic Finite Automata



- Number of outgoing edges from a state w.r.t an alphabet can be arbitrary.
- The input is accepted if there is an accepting path.

#### ABE for NFA: A Two Step Solution [AMY19a]

NfaABE<sub>u,b</sub>

(unbounded length inputs, bounded NFA machines)

NfaABE<sub>b.u</sub>

(bounded length inputs, unbounded NFA machines)

uNfaABE

(unbounded length inputs, unbounded NFA machines)

#### Naive solution:

• For security parameter  $\lambda$ , input (attribute) length  $|\mathbf{x}| \leq 2^{\lambda}$ . MSK for NfaABE<sub>u,b</sub> is a PRF key K.

#### Naive solution:

- For security parameter  $\lambda$ , input (attribute) length  $|\mathbf{x}| \leq 2^{\lambda}$ . MSK for NfaABE<sub>u,b</sub> is a PRF key K.
- Use an ABE scheme for circuits to sample  $2^{\lambda}$  key pairs (using K) for each input length.

 $(ABE.PK_1, ABE.MSK_1)$ 

 $(\mathsf{ABE}.\mathsf{PK}_2,\mathsf{ABE}.\mathsf{MSK}_2)$ 

. . .

 $(\mathsf{ABE}.\mathsf{PK}_{2^\lambda},\mathsf{ABE}.\mathsf{MSK}_{2^\lambda})$ 

#### Naive solution:

- For security parameter  $\lambda$ , input (attribute) length  $|\mathbf{x}| \leq 2^{\lambda}$ . MSK for NfaABE<sub>u,b</sub> is a PRF key K.
- Use an ABE scheme for circuits to sample  $2^{\lambda}$  key pairs (using K) for each input length.

```
(\mathsf{ABE}.\mathsf{PK}_1,\mathsf{ABE}.\mathsf{MSK}_1) \qquad (\mathsf{ABE}.\mathsf{PK}_2,\mathsf{ABE}.\mathsf{MSK}_2) \qquad \bullet \bullet \qquad (\mathsf{ABE}.\mathsf{PK}_{2^\lambda},\mathsf{ABE}.\mathsf{MSK}_{2^\lambda})
```

• Encrypt message m under attribute  ${\bf x}$  with ABE.PK $_{|{\bf x}|}$  .

#### Naive solution:

- For security parameter  $\lambda$ , input (attribute) length  $|\mathbf{x}| \leq 2^{\lambda}$ . MSK for NfaABE<sub>u,b</sub> is a PRF key K.
- Use an ABE scheme for circuits to sample  $2^{\lambda}$  key pairs (using K) for each input length.

```
(\mathsf{ABE}.\mathsf{PK}_1,\mathsf{ABE}.\mathsf{MSK}_1) \qquad (\mathsf{ABE}.\mathsf{PK}_2,\mathsf{ABE}.\mathsf{MSK}_2) \qquad \bullet \bullet \bullet \qquad (\mathsf{ABE}.\mathsf{PK}_{2^\lambda},\mathsf{ABE}.\mathsf{MSK}_{2^\lambda})
```

- Encrypt message m under attribute  ${f x}$  with ABE.PK $_{|{f x}|}$  .
- Convert NFA M to a set of  $2^{\lambda}$  circuits

#### Naive solution:

- For security parameter  $\lambda$ , input (attribute) length  $|\mathbf{x}| \leq 2^{\lambda}$ . MSK for NfaABE<sub>u,b</sub> is a PRF key K.
- Use an ABE scheme for circuits to sample  $2^{\lambda}$  key pairs (using K) for each input length.

 $(\mathsf{ABE}.\mathsf{PK}_1,\mathsf{ABE}.\mathsf{MSK}_1) \qquad (\mathsf{ABE}.\mathsf{PK}_2,\mathsf{ABE}.\mathsf{MSK}_2) \qquad \bullet \bullet \qquad (\mathsf{ABE}.\mathsf{PK}_{2^\lambda},\mathsf{ABE}.\mathsf{MSK}_{2^\lambda})$ 

- Encrypt message m under attribute  ${f x}$  with ABE.PK $_{|{f x}|}$
- Convert NFA M to a set of  $2^{\lambda}$  circuits, one for each input length.

$$\widehat{M}_1$$

$$\widehat{M}_2$$

$$\widehat{M}_{2^{\lambda}}$$

#### Naive solution:

- For security parameter  $\lambda$ , input (attribute) length  $|\mathbf{x}| \leq 2^{\lambda}$ . MSK for NfaABE<sub>u,b</sub> is a PRF key K.
- Use an ABE scheme for circuits to sample  $2^{\lambda}$  key pairs (using K) for each input length.

$$(\mathsf{ABE}.\mathsf{PK}_1,\mathsf{ABE}.\mathsf{MSK}_1) \qquad (\mathsf{ABE}.\mathsf{PK}_2,\mathsf{ABE}.\mathsf{MSK}_2) \qquad \bullet \bullet \bullet \qquad (\mathsf{ABE}.\mathsf{PK}_{2^\lambda},\mathsf{ABE}.\mathsf{MSK}_{2^\lambda})$$

- Encrypt message m under attribute  ${f x}$  with ABE.PK $_{|{f x}|}$
- Convert NFA M to a set of  $2^{\lambda}$  circuits, one for each input length. Generate a secret key for each circuit.

$$\widehat{M}_1 \longrightarrow \mathsf{ABE.sk}_{\widehat{M}_1} \qquad \widehat{M}_2 \longrightarrow \mathsf{ABE.sk}_{\widehat{M}_2} \qquad \dots \qquad \widehat{M}_{2^{\lambda}} \longrightarrow \mathsf{ABE.sk}_{\widehat{M}_{2^{\lambda}}}$$

#### Naive solution:

- Note  $|\mathbf{x}| \leq 2^{\lambda}$ . Let MSK: PRF key K.
- Use an ABE scheme for circuits to sample  $2^{\lambda}$  key pairs (using K) for each input length.

$$(\mathsf{ABE}.\mathsf{PK}_1,\mathsf{ABE}.\mathsf{MSK}_1) \qquad (\mathsf{ABE}.\mathsf{PK}_2,\mathsf{ABE}.\mathsf{MSK}_2) \qquad \bullet \bullet \bullet \qquad (\mathsf{ABE}.\mathsf{PK}_{2^\lambda},\mathsf{ABE}.\mathsf{MSK}_{2^\lambda})$$

- Encrypt message m under attribute  ${f x}$  with ABE.PK $_{|{f x}|}$
- Convert NFA M to a set of  $2^{\lambda}$  circuits, one for each input length. Generate a secret key for each circuit.

$$\widehat{M}_1 \longrightarrow \mathsf{ABE.sk}_{\widehat{M}_1} \qquad \widehat{M}_2 \longrightarrow \mathsf{ABE.sk}_{\widehat{M}_2} \qquad \dots \qquad \widehat{M}_{2^{\lambda}} \longrightarrow \mathsf{ABE.sk}_{\widehat{M}_{2^{\lambda}}}$$

• Decryptor knows  $i = |\mathbf{x}|$ , chooses. ABE.sk $_{\widehat{M}_i}$  accordingly; decrypts to get m, if M accepts  $\mathbf{x}$ .

#### Problem #1: Too many keys

• Key size exponential, via  $2^{\lambda}$  instances of ABE.

#### Problem #1: Too many keys

• Key size exponential, via  $2^{\lambda}$  instances of ABE.

Solution: Handle only inputs of length  $2^i$ ,  $\forall i \in [0, \lambda]$ .

#### Problem #1: Too many keys

• Key size exponential, via  $2^{\lambda}$  instances of ABE.

Solution: Handle only inputs of length  $2^i$ ,  $\forall i \in [0, \lambda]$ . Generate ABE keys for  $\lambda + 1$  instances only.

Pad **x** (such that  $2^{i-1} < |\mathbf{x}| \le 2^i$ ) with  $\perp$  to make  $\mathbf{x'} = (\mathbf{x}, \perp, ..., \perp)$  of length  $2^i$ .

#### Problem #1: Too many keys

• Key size exponential, via  $2^{\lambda}$  instances of ABE.

Solution: Handle only inputs of length  $2^i$ ,  $\forall i \in [0, \lambda]$ . Generate ABE keys for  $\lambda + 1$  instances only.

- Pad  $\mathbf{x}$  (such that  $2^{i-1} < |\mathbf{x}| \le 2^i$ ) with  $\bot$  to make  $\mathbf{x}' = (\mathbf{x}, \bot, ..., \bot)$  of length  $2^i$ .
- Encrypt message m under attribute  $\mathbf{x'}$  with ABE.PK $_{|\mathbf{x'}|}$

#### Problem #1: Too many keys

• Key size exponential, via  $2^{\lambda}$  instances of ABE.

Solution: Handle only inputs of length  $2^i$ ,  $\forall i \in [0, \lambda]$ . Generate ABE keys for  $\lambda + 1$  instances only.

- Pad  $\mathbf{x}$  (such that  $2^{i-1} < |\mathbf{x}| \le 2^i$ ) with  $\bot$  to make  $\mathbf{x}' = (\mathbf{x}, \bot, ..., \bot)$  of length  $2^i$ .
- Encrypt message m under attribute  $\mathbf{x'}$  with ABE.PK $_{|\mathbf{x'}|}$
- Convert NFA M to a set of  $\lambda + 1$  circuits.

$$\widehat{M}_{2^0}$$

$$\widehat{M}_{2^1}$$

$$\widehat{M}_2$$

#### Problem #1: Too many keys

• Key size exponential, via  $2^{\lambda}$  instances of ABE.

Solution: Handle only inputs of length  $2^i$ ,  $\forall i \in [0, \lambda]$ . Generate ABE keys for  $\lambda + 1$  instances only.

- Pad x (such that  $2^{i-1} < |\mathbf{x}| \le 2^i$ ) with  $\perp$  to make  $\mathbf{x}' = (\mathbf{x}, \perp, ..., \perp)$  of length  $2^i$ .
- Encrypt message m under attribute  $\mathbf{x'}$  with ABE.PK $_{|\mathbf{x'}|}$
- Convert NFA M to a set of  $\lambda + 1$  circuits. Generate  $\lambda + 1$  ABE secret keys.

$$\widehat{M}_{2^0} \longrightarrow \mathsf{ABE.sk}_{\widehat{M}_{2^0}} \qquad \widehat{M}_{2^1} \longrightarrow \mathsf{ABE.sk}_{\widehat{M}_{2^1}} \qquad \dots \qquad \widehat{M}_{2^{\lambda}} \longrightarrow \mathsf{ABE.sk}_{\widehat{M}_{2^{\lambda}}}$$

#### Problem #2: Too large keys

• Secret key size  $\left|\mathsf{ABE.sk}_{\widehat{M}_{2i}}\right|$  may be exponential

#### Possible way out:

• Use suitable ABE [BGG+14]. SK sizes short : poly( $\lambda$ , d) for depth d circuits

#### How to bound circuit depth?

- Naïve conversion of M to circuit results in circuits of depth 2<sup>i</sup>
- Use divide-and-conquer technique to evaluate *M*.
- Ensures circuit depth =  $\operatorname{polylog}(|\mathbf{x}|) \Rightarrow \left| \operatorname{ABE.sk}_{\widehat{M}_{2i}} \right|$  is  $\operatorname{poly}(\lambda)$ .

#### Problem #3: Inefficient keygen

- $ullet \left| \widehat{M}_{2^i} 
  ight|$  too large to even be read
- Keygen cannot know input length
- Even if it did, runtime must be independent of this

#### Solution: Redistribute computation

- Encryptor and decryptor know input length, and can run in time O(|x|)
- Delegate inefficient part of keygen to enc and dec!
- Key Tool: Functional Encryption



# Is Science Objective?

"Whenever the subject of women in science comes up, there are people fiercely committed to the idea that sexism does not exist. They will point to everything and anything else to explain differences while becoming angry and condescending if you even suggest that discrimination could be a factor. But these people are wrong. This data shows they are wrong."

[Ilana Yurkiewicz, Scientific American 2012]

Can we be open to this idea?

## Society has biases!

A girl receives literally thousands of suggestions over time that tell her what her place/role is...

- My earliest memory: Blessings received when touching feet of elders
- Today's experience: Society accepts women who need, not women who lead
- Enormous pressure felt by (esp.) MS/PhD students about "own desires" versus family/expectations. Seen many bright young girls giving up or compromising heavily on career

Sometimes, discards/rebels.

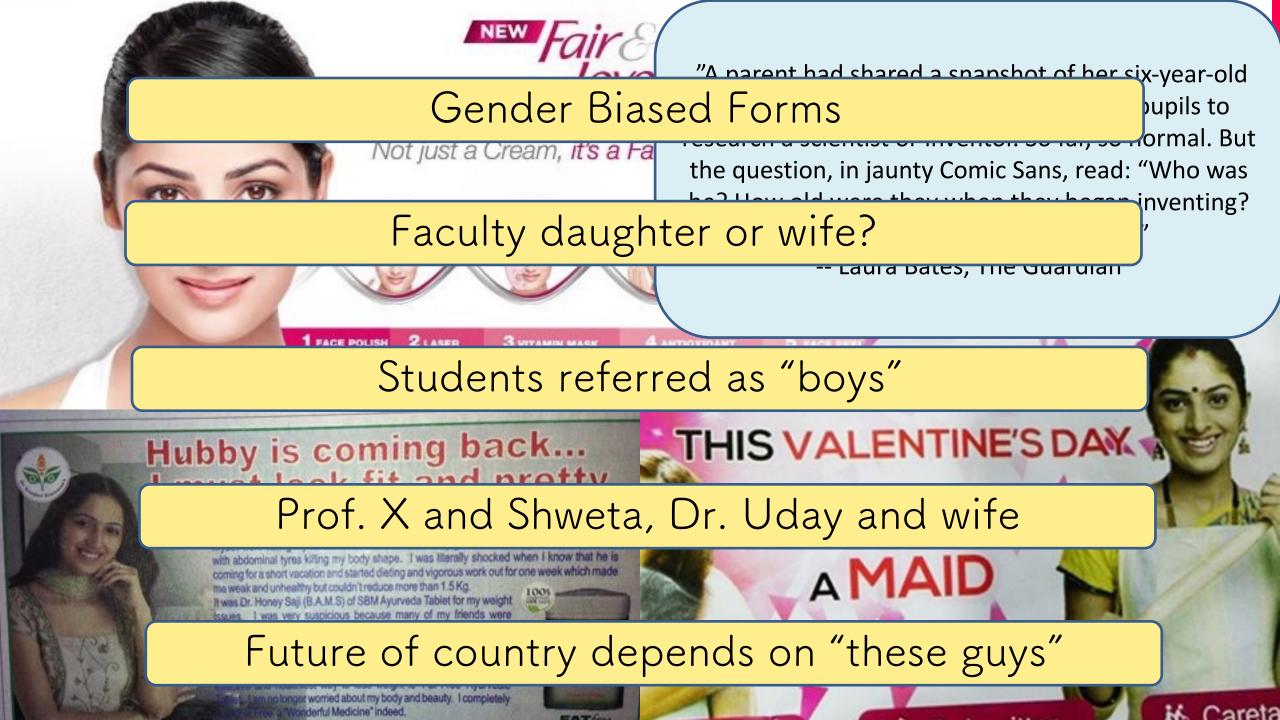
Often internalizes/compromises.



Let's talk marriage. Let's talk certainties.

Insurance plans for the certainties of life.

Life isn't full of accidents waiting to happen. In fact, it's full of certainties like getting married,



# Society has biases! And Science?

Scientists are supposed to be objective, able to evaluate data and results without being swayed by emotions or biases. This is a fundamental tenet of science. What this extensive literature shows is, in fact, scientists are people, subject to the same cultural norms and beliefs as the rest of society.

[Prof. Alison Coil, UCSD]

# - Women can't do math

- Women are good at rote learning not reasoning
- It is not feminine to argue
- Why would Google pay so much to hire a woman whose just going to go on maternity leave
- I saw a very beautiful woman on our floor today and I wondered, what she is doing on the science floor?
- Your paper has better chances since it will get sympathy, being an all woman paper
- You left your parents to follow your desires? Our daughters would never do that
- If you study so much, who will marry you?
- Women need to put family first

#### So... what next?

Is this daunting/depressing? Seeing it is overcoming it!

#### It only matters if you let it!

- Reject these suggestions and they cannot touch you.
- Calling it out? Take a call.
- Preserve creative energy: results talk loudest!
- Cultivate support system
- Personally: Follow(ed) gut even when no support. Willing to accept consequences.

# No Looking Back!

Life is not easy for any of us. But what of that? We must have perseverance and above all confidence in ourselves. We must believe that we are gifted for something and that this thing must be attained.

--- Marie Curie (first scientist to be awarded a Nobel Prize in two different categories)





Thank You

Images Credit: MF Hussain, Hans Hoffman