Cliptography: Clipping The Power Of Kleptographic Attacks

Qiang Tang
New Jersey Institute of Technology

Joint work with
Alexander Russell(UConn), Moti Yung(Snapchat & Columbia), and Hong-Sheng Zhou(VCU)

Modern Crypto

Modern Crypto

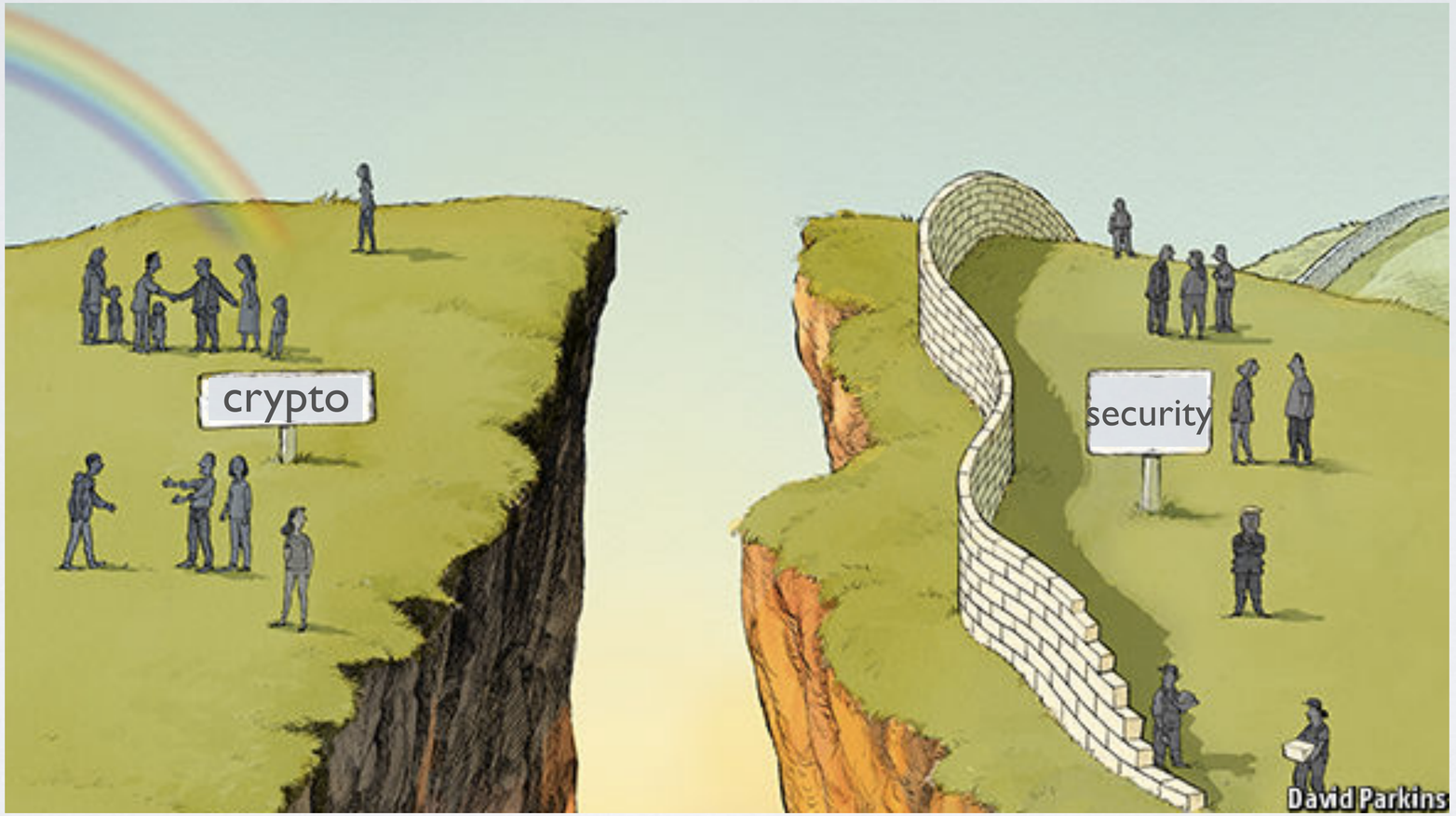
- *"Precise" models to capture attacks*
- *"Rigorous" proofs to establish security*

Modern Crypto

- *"Pre* *acks*
- *"Rigorous" proofs to establish security*

Still long way to go

The “Security Divide”



An Implicit Assumption

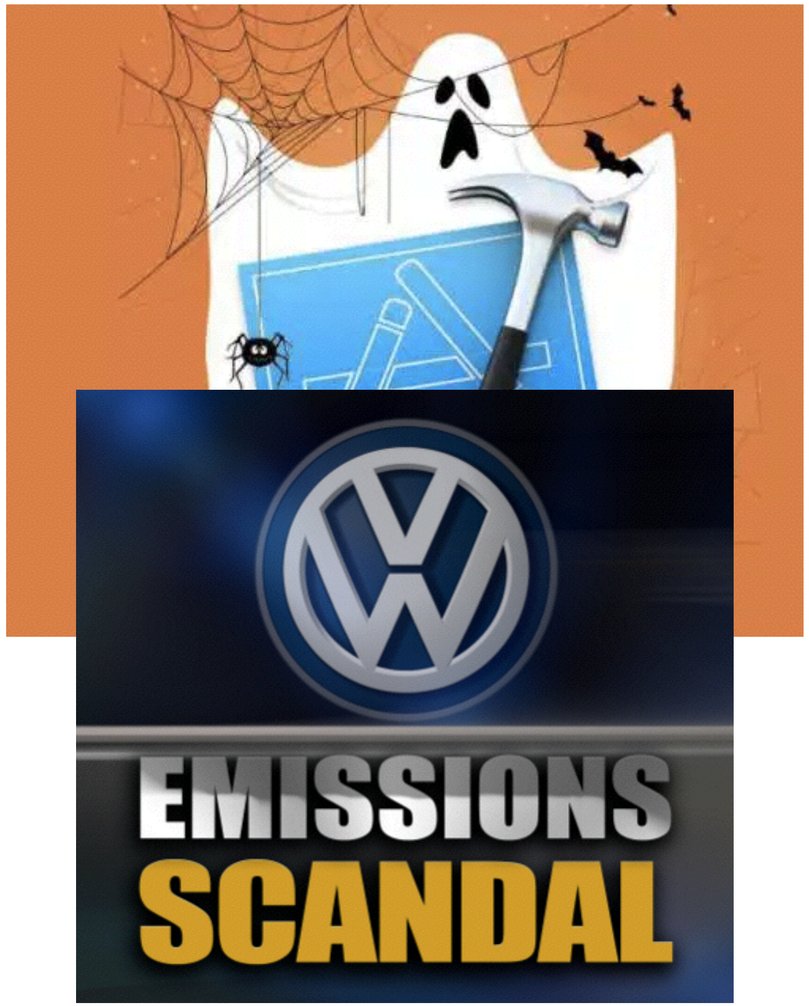
An Implicit Assumption

Tradition: after cryptographers design the crypto tools, someone will implement them correctly for use

Implementations are Untrustworthy

Implementations are Untrustworthy

Implementations are Untrustworthy



Implementations are Untrustworthy

Mac iPad iPhone Watch TV Music Support

February 16, 2016

A Message to Our Customers

The United States government has demanded that Apple take an unprecedented step which threatens the security of our customers. We oppose this order, which has implications far beyond the legal case at hand.

This moment calls for public discussion, and we want our customers and people around the country to understand what is at stake.

The Need for Encryption

Smartphones, led by iPhone, have become an essential part of our lives. People use them to store an incredible amount of personal information, from our private conversations to our photos, our music, our notes, our calendars and contacts, our financial information and health data, even where we have been and where we are going.

All that information needs to be protected from hackers and criminals who want to access it, steal it, and use it without our knowledge or permission. Customers expect Apple and other technology companies to do everything in our power to protect their personal information, and at Apple we are deeply committed to safeguarding their data.

Kleptography

- The science of stealing information *securely and subliminally* from black-box cryptographic implementations

Young & Yung ’96, ’97

RSA Key Generation

$$pk = (N, e) \quad sk = d$$

RSA KeyGen

$$\xrightarrow{N, e, d} \quad N = pq, \text{for random primes } p, q$$

$$\text{random } e, \text{ and } ed = 1 \mod \phi(N)$$

A Subverted Implementation

A Subverted Implementation

(A "backdoor")

The Attack:

$$pk = (N, e) \quad sk = d$$

$$\xrightarrow{N, e, d} \quad N = pq, \text{for random primes } p, q$$

$$e = e_1 || e_2, \text{ where } e_1 = \mathsf{SEnc}(z, p)$$

(A "backdoor")

The Attack:

$$pk = (N, e) \quad sk = d$$

$$\xrightarrow{N, e, d} \quad N = pq, \text{for random primes } p, q$$

$$e = e_1 || e_2, \text{ where } e_1 = \mathsf{SEnc}(z, p)$$

(A "backdoor")

Having the backdoor z, adversary can learn p from pk

The Attack:

$$pk = (N, e) \quad sk = d$$

$$\xrightarrow{N, e, d} \quad N = pq, \text{for random primes } p, q$$

$$e = e_1 || e_2, \text{ where } e_1 = \mathsf{SEnc}(z, p)$$

(A "backdoor")

Having the backdoor z, adversary can learn p from pk

Without z, e looks randomly distributed as in the SPEC

Two Decades Later

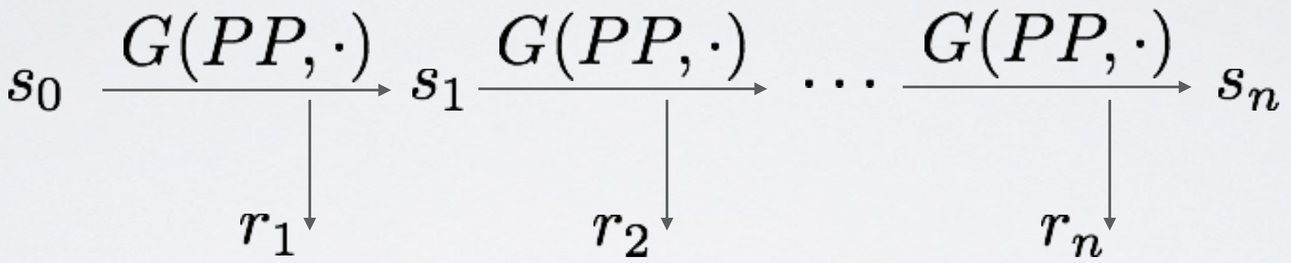
Two Decades Later

Two Decades Later

- Theory can go to practice!

Backdoored Dual EC

$$PP = (P, Q) \text{ for a random } Q, P = Q^z$$



$$s_i = P^{s_{i-1}}, r_i = Q^{s_i}$$

$$s_2 = P^{s_1} = Q^{zs_1} = r_1^z$$

Having z, r_2 can be computed from r_1

(P, Q) look as randomly generated

The Threat of Klepto Attacks

- Remarkably, an adversarially implemented cryptographic algorithm may…

The Threat of Klepto Attacks

- Remarkably, an adversarially implemented cryptographic algorithm may…
  - *leak private information* to the implementer

The Threat of Klepto Attacks

- Remarkably, an adversarially implemented cryptographic algorithm may…
  - *leak private information* to the implementer
  - while *adhering perfectly to the specification.*

Sudden Renewed Attention

Bellare-Paterson-Rogaway'14, Bellare-Hoang'15,
Dodis-Ganesh-Golovnev-Juels-Ristenpart'15,
Mironov-Stevens-Davidovitz'15, Degabriele-Farshim-Pottering'15,
Ateniese-Magri-Venturi'15,Bellare-Jaeger-Kane'15, Rogaway'15
Russell-T-Yung-Zhou'15A, Russell-T-Yung-Zhou'15B,
Dodis-Mironov-Davidovitz'16,Bellare-Kane-Rogaway'16
Degabriele-Paterson-Schult-Woodage'16
Russell-T-Yung-Zhou'16A,Russell-T-Yung-Zhou'16B

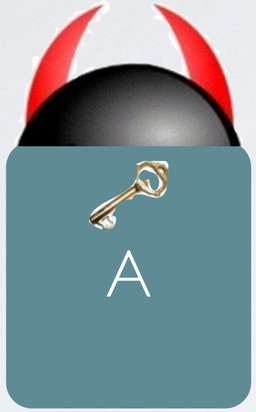
Sudden Renewed Attention

Bellare-Paterson-Rogaway'14, Bellare-Hoang'15,
Dodis-Ganesh-Golovnev-Juels-Ristenpart'15,
Mironov-Stevens-Davidovitz'15, Degabriele-Farshim-Pottering'15,
Ateniese-Magri-Venturi'15,Bellare-Jaeger-Kane'15, Rogaway'15
Russell-T-Yung-Zhou'15A, Russell-T-Yung-Zhou'15B,
Dodis-Mironov-Davidovitz'16,Bellare-Kane-Rogaway'16
Degabriele-Paterson-Schult-Woodage'16
Russell-T-Yung-Zhou'16A,Russell-T-Yung-Zhou'16B

Mostly depressing IMPOSSIBILITY results

Subliminal Channel Attack

[BPR14]



Secret s

Subverted implementation of randomized algorithm
can leak secrets exclusively to backdoor holder
via public communication channel
using steganography by doing rejection sampling

Status-of-the-Art for Defending

Status-of-the-Art for Defending

- Give up on randomized algorithms

Status-of-the-Art for Defending

- Give up on randomized algorithms
  - assume key generation algorithm is honest

Status-of-the-Art for Defending

- Give up on randomized algorithms
  - assume key generation algorithm is honest
  - consider deterministic encryption algorithm only

Status-of-the-Art for Defending

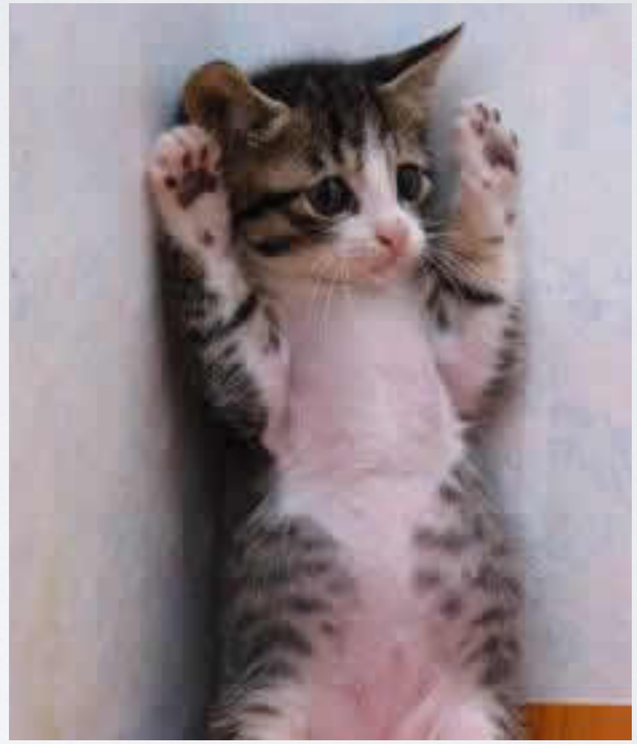
- Give up on randomized algorithms
  - assume key generation algorithm is honest
  - consider deterministic encryption algorithm only
- Assumed correctness

Status-of-the-Art for Defending

- Give up on randomized algorithms
  - assume key generation algorithm is honest
  - consider deterministic encryption algorithm only
- Assumed correctness
- Assuming trusted randomness (for re-randomizer)

Status-of-the-Art for Defending

- Give up on randomized algorithms
  - assume key generation algorithm is honest
  - consider deterministic encryption algorithm only
- Assumed correctness
- Assuming trusted randomness (for re-randomizer)



Current Status: *Wide Open*

Current Status: *Wide Open*

- No wide agreement on models

Current Status: *Wide Open*

- No wide agreement on models
- Very few defending mechanisms known: no idea what to do with randomized algorithms

Current Status: *Wide Open*

- No wide agreement on models
- Very few defending mechanisms known: no idea what to do with randomized algorithms
- Very few functionalities have been considered

Current Status: *Wide Open*

- No wide
- Very few n: no idea what to ns
- Very few functionalities have been considered

Far from being understood

Long Term Goal

- Revisit **cryptography**, build **cliptography**—clipping the power of kleptographic attacks

Our Initial Results

Our Initial Results

- Modeling: a general definitional framework, a hierarchy of definitions. all algorithms are subverted by the adversary;

Our Initial Results

- Modeling: a general definitional framework, a hierarchy of definitions. all algorithms are subverted by the adversary;

- Mitigating: properly control the public channel to salvage primitives even if subliminal channel exists—immediately deployable with minimal change of the specification

Our Defending Results

- Subversion resistant (TD)OWP
- Subversion resistant PRGs
- Subversion resistant signature with an online watchdog

Cliptographic Model

Cliptographic Model

Cliptographic Model

Cliptographic Model

Cliptographic Model

a4t*#f-1zd

f%5u7(bg@

Cliptographic Model

a4t*#f-1zd

f%5u7(bg@

Cliptographic Model

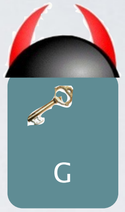
SPEC

a4t*#f-1zd

f%5u7(bg@

Cliptographic Model

SPEC



a4t*#f-1zd

f%5u7(bg@

Cliptographic Model

SPEC

Cliptographic Model

The Model(s)

Three participants:

The Model(s)

Three participants:

- **The Adversary**, who provides implementations of cryptographic algorithms, and later attempts to "break" them;

The Model(s)

Three participants:

- **The Adversary**, who provides implementations of cryptographic algorithms, and later attempts to "break" them;
- **The Challenger(User)**, who uses the subverted implementations.

The Model(s)

Three participants:

- **The Adversary**, who provides implementations of cryptographic algorithms, and later attempts to "break" them;
- **The Challenger(User)**, who uses the subverted implementations.
- **The Watchdog**, who tests the implementations against their specification;

The Model(s)

Three participants:

- **The Adversary**, who provides implementations of cryptographic algorithms, and later attempts to "break" them;
- **The Challenger(User)**, who uses the subverted implementations.
- **The Watchdog**, who tests the implementations against their specification;

The adversary is proud-but-malicious

The Basic Notion of Security

A primitive is **cliptographically secure/subversion resistant** if there exists a watchdog so that, for any efficient adversary,:

The Basic Notion of Security

A primitive is **cliptographically secure/subversion resistant** if there exists a watchdog so that, for any efficient adversary,:

- Either the watchdog can distinguish IMPL from SPEC, or

The Basic Notion of Security

A primitive is **cliptographically secure/subversion resistant** if there exists a watchdog so that, for any efficient adversary,:

- Either the watchdog can distinguish IMPL from SPEC, or
- The primitive is still secure according to the "adapted" security game.

The Basic Notion of Security

A primitive is **cliptographically secure/subversion resistant** if there exists a watchdog so that, for any efficient adversary,:

- Either the watchdog can distinguish IMPL from SPEC, or
- The primitive is still secure according to the "adapted" security game.

Several variants depending on the watchdog power, form of the implementation, etc

What Can the Watchdog Guarantee?

What Can the Watchdog Guarantee?

- W can guarantee that deterministic algorithms with public input distribution are (almost) consistent with the specification.

- W can guarantee the randomness generation algorithms produce unpredictable outputs.

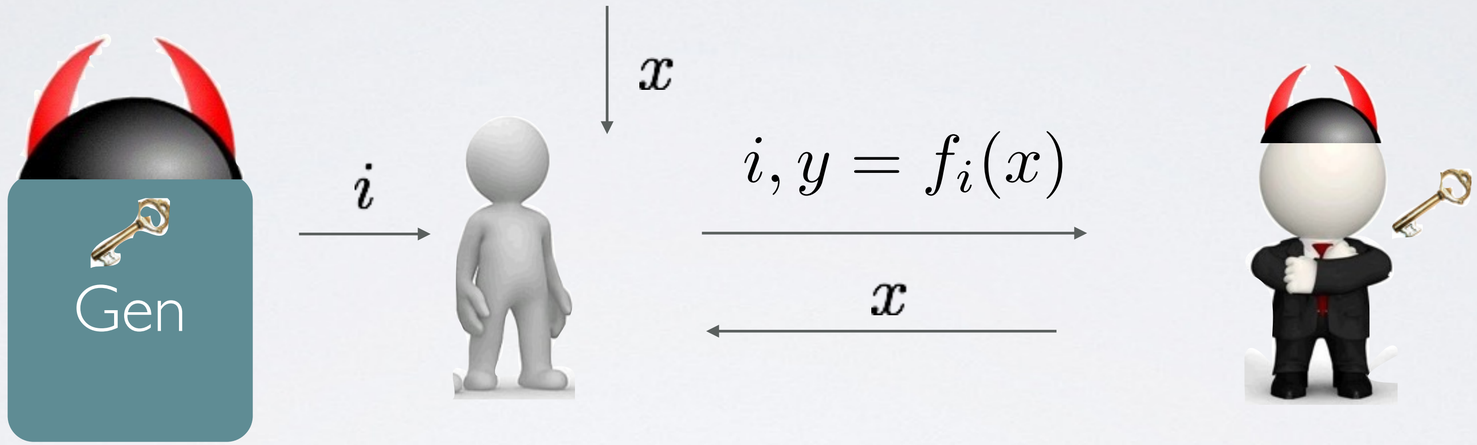
Mitigating Subliminal Channel

Key Generation must be randomized

One-Way Permutation

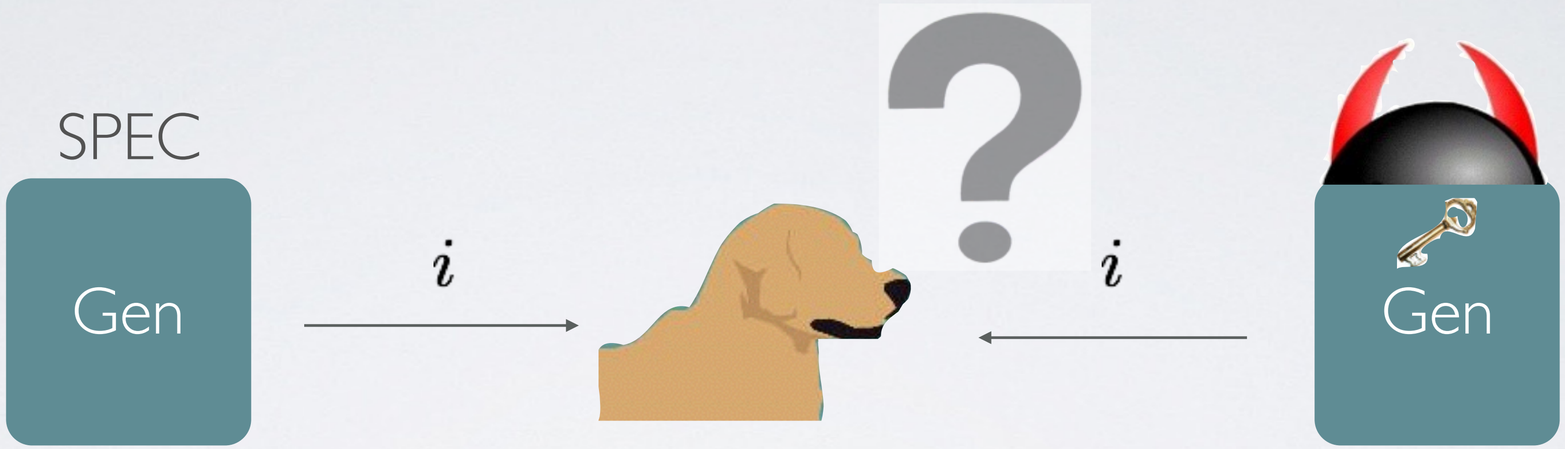
- A **one-way permutation**: A permutation that is
  - Easy to compute;
  - Hard to invert.
- Fundamental tool for constructing PRGs, symmetric encryption.

Subvertible OWPs:



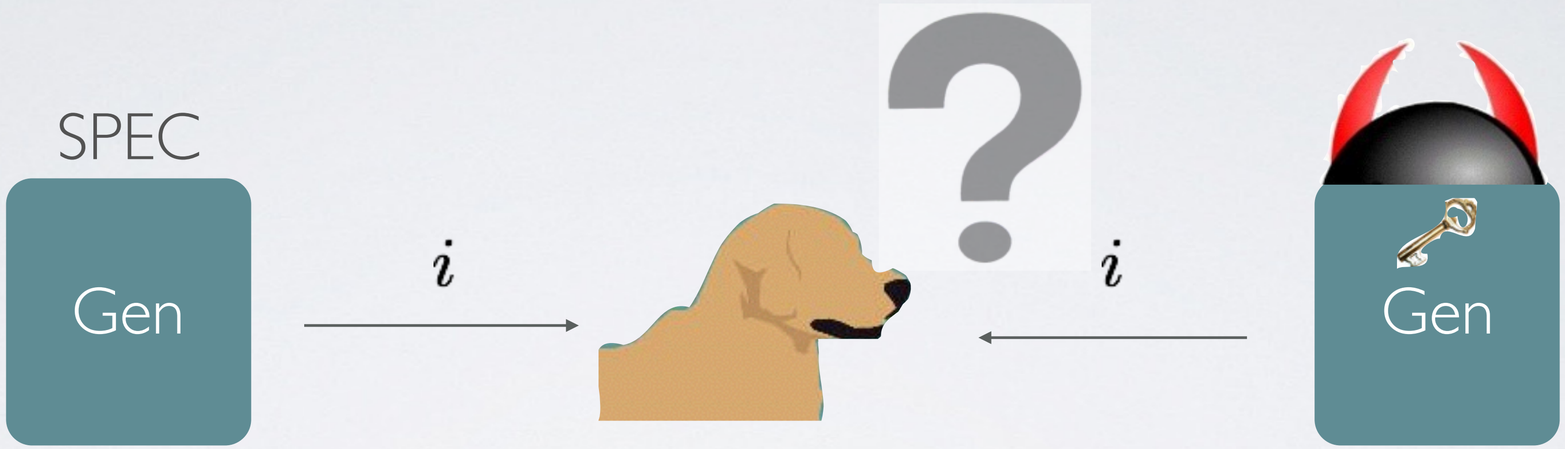
Adversary can win this game…and…

Subvertible OWPs



Two index distributions are indistinguishable

Subvertible OWPs



Two index distributions are indistinguishable

OK to ignore Eval as it is deterministic with
a public input distribution

Random Padding is Dangerous

$$f_{i,k}(x) = g_i(x)||k$$

Index

- SPEC: Outputs random i,k; here $\{g_i\}$ is a TDOWP.
- IMPL: (i,d) from a TDOWP, and k=SEnc(z,d); here d is the trapdoor.

Mitigating Subliminal Channel

Key Generation must be randomized

Conventional Wisdom

Conventional Wisdom

Nothing up my sleeve numbers

Conventional Wisdom

- π = 3.14159265358979323846264328327 95…… some bits of it were used as constants in some hash function (BLAKE), block cipher (Blowfish) and more

Nothing up my sleeve numbers

Conventional Wisdom

- π = 3.14159265358979323846264328327 95…… some bits of it were used as constants in some hash function (BLAKE), block cipher (Blowfish) and more
- e = 2.71828182845904523536028747 13527……some bits of it were used as constants in an AES candidate block cipher (RC5) and more

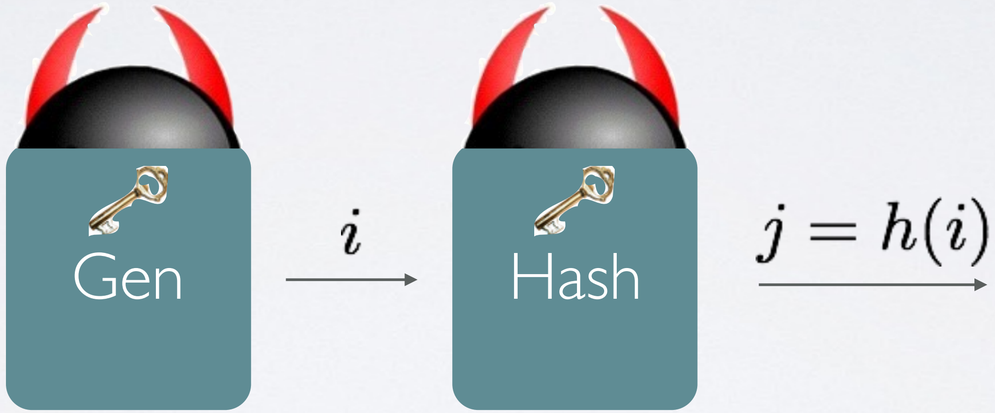
Nothing up my sleeve numbers

Mitigating Subverted KG

Nothing up my sleeve parameters/keys

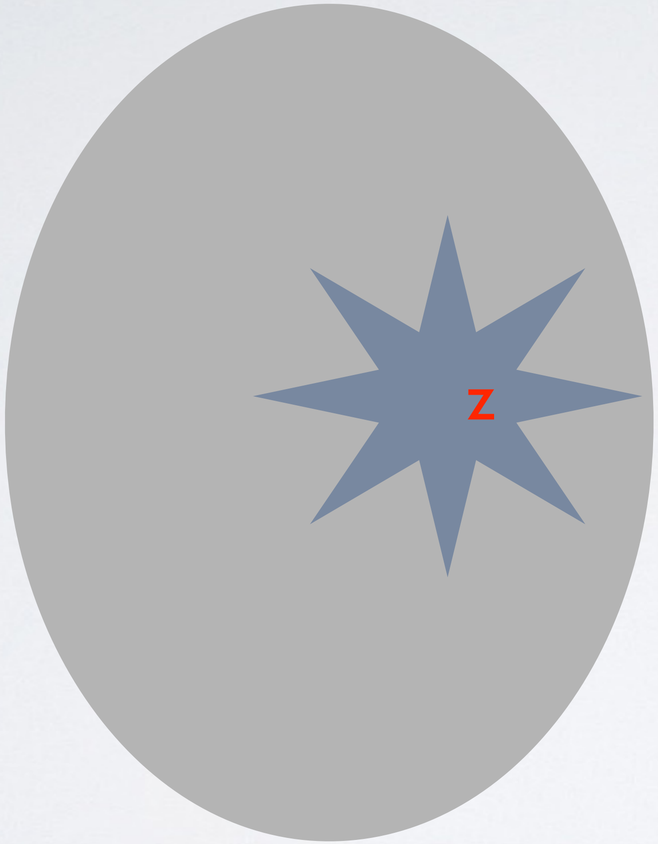
Mitigating Subverted KG

Nothing up my sleeve parameters/keys

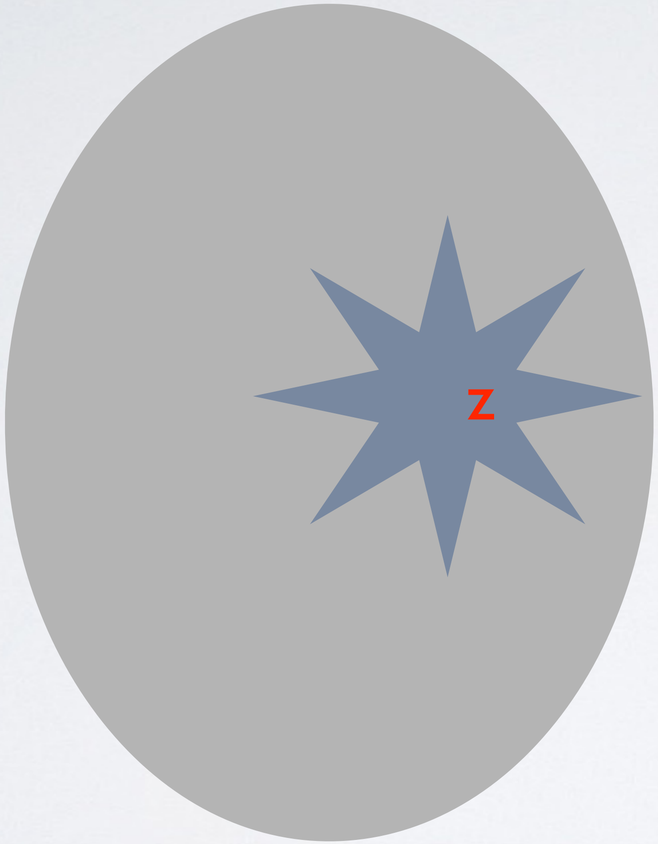


$$g_i(x) := f_{h(i)}(x)$$

Mitigating Subverted KG: Intuition

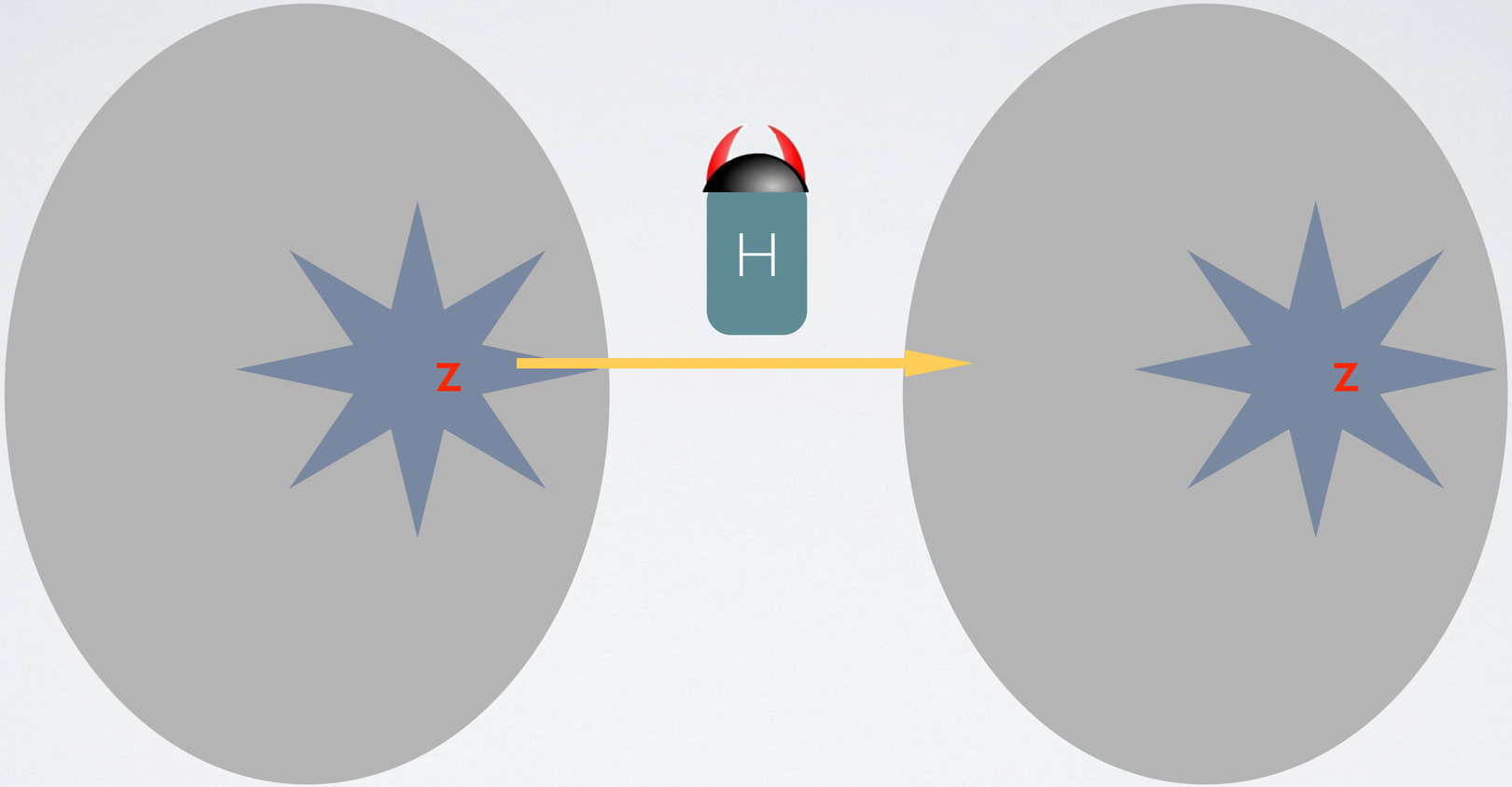


Mitigating Subverted KG: Intuition



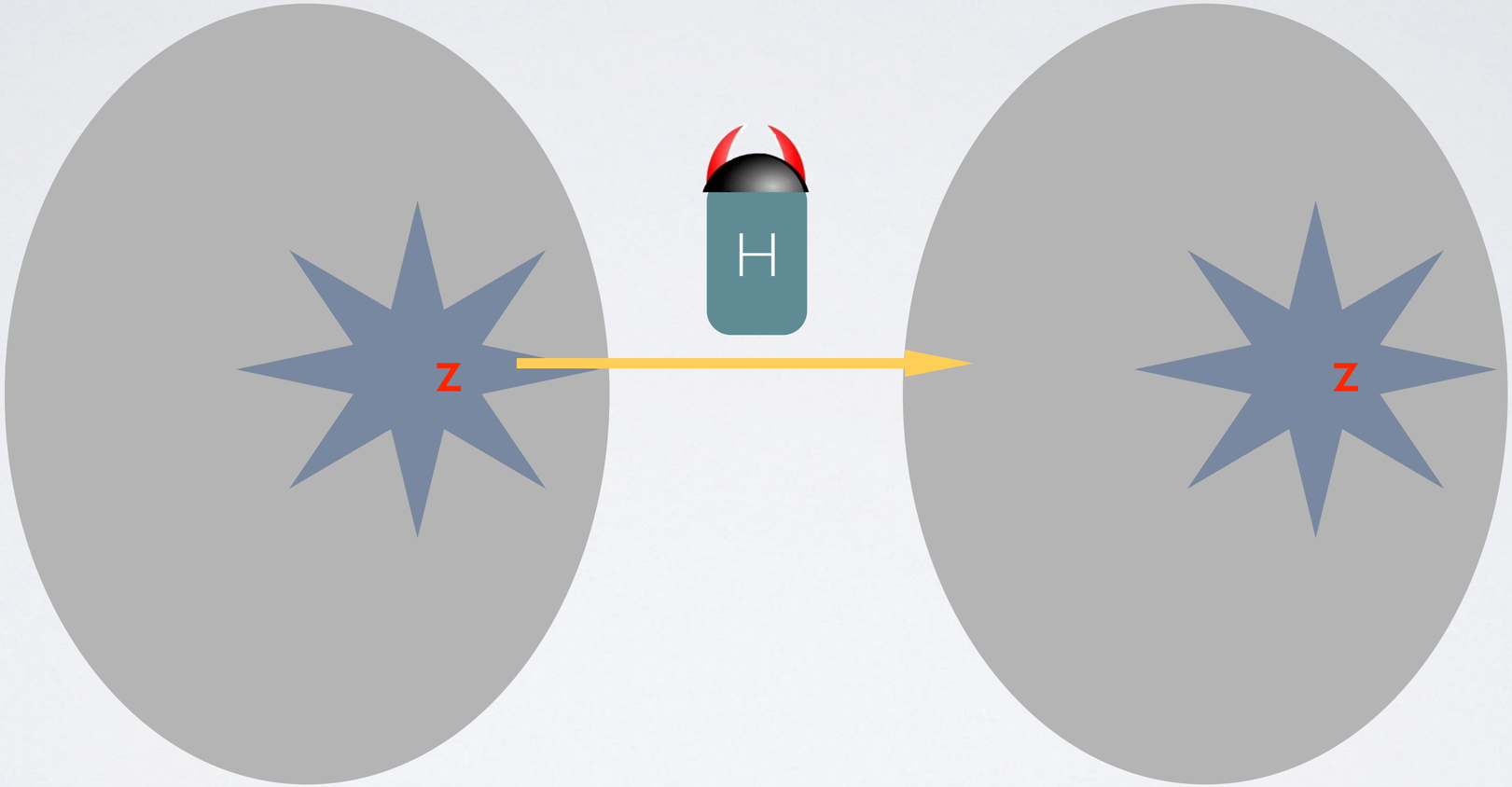
Any backdoor can be used to invert a **sparse** subset of functions, otherwise SPEC is insecure

Mitigating Subverted KG: Intuition



Any backdoor can be used to invert a **sparse** subset of functions, otherwise SPEC is insecure

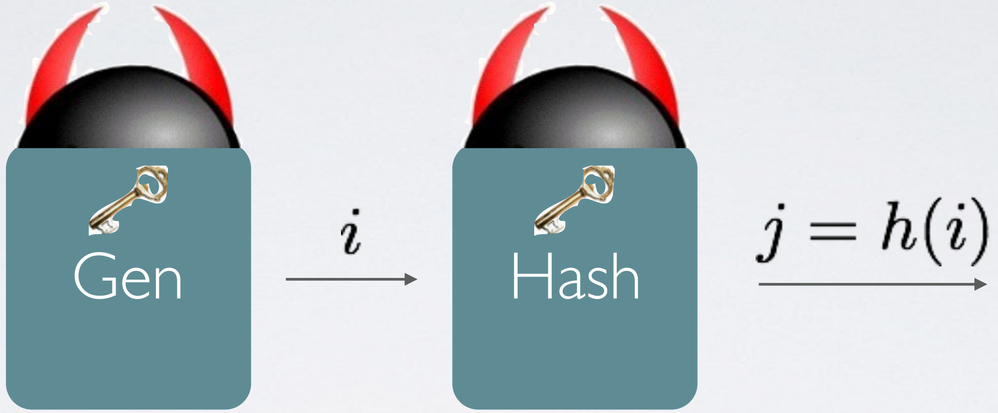
Mitigating Subverted KG: Intuition



Any backdoor can be used to invert a **sparse** subset of functions, otherwise SPEC is insecure

"Dispersing" the index to a "safe" place

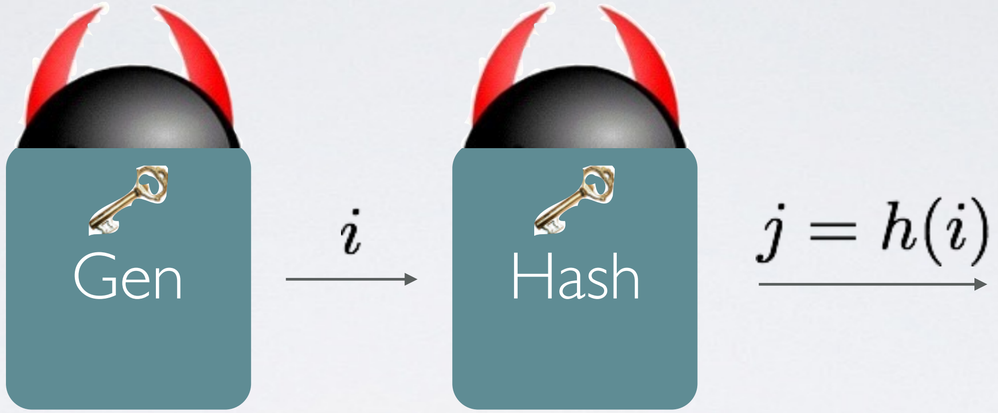
Mitigating Subverted KG



$$g_i(x) := f_{h(i)}(x)$$

Theorem: $\{g_i\}$ is a family of subversion resistant OWPs.

Mitigating Subverted KG



$$g_i(x) := f_{h(i)}(x)$$

Theorem: $\{g_i\}$ is a family of subversion resistant OWPs.

Assuming the SPEC of h is RO, and index domain is “simple”

Further Implications

Further Implications

- Similarly salvage Duel_EC PRNG: it was shown to be impossible to sanitize the output.

Further Implications

- Similarly salvage Duel_EC PRNG: it was shown to be impossible to sanitize the output.
- Similarly salvage trapdoor OWP, then further save the KG of the full domain hash digital signature scheme

Further Results

Further Results

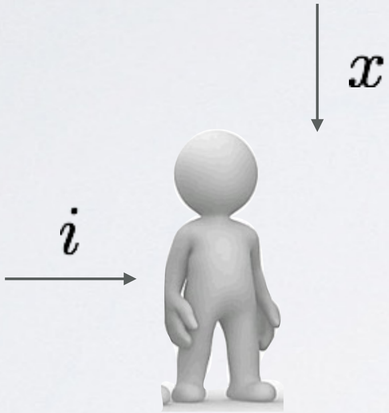
- Reduction of FDH does not go through, modification needed

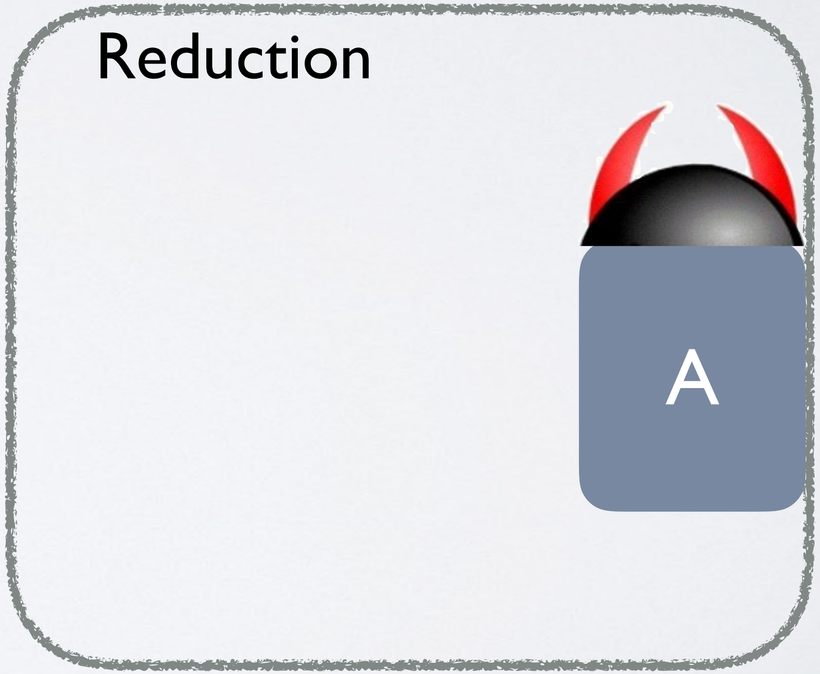
Further Results

- Reduction of FDH does not go through, modification needed
- Reduction from clipto-secure OWP to PRG preserves

Conventional FDH Proof

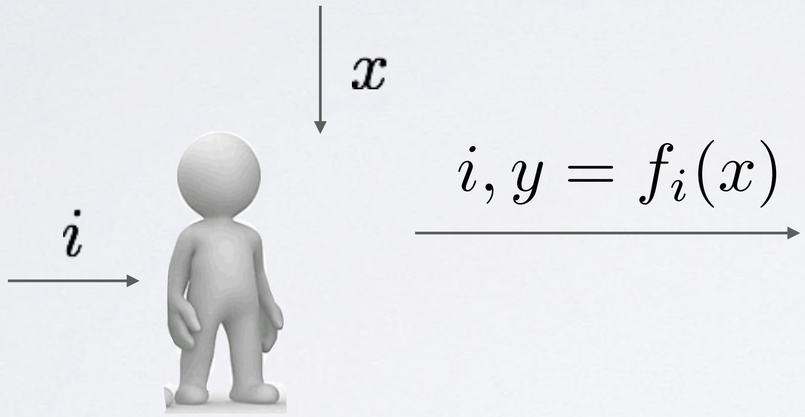
Embed the TDOWP challenge to one RO query answer:

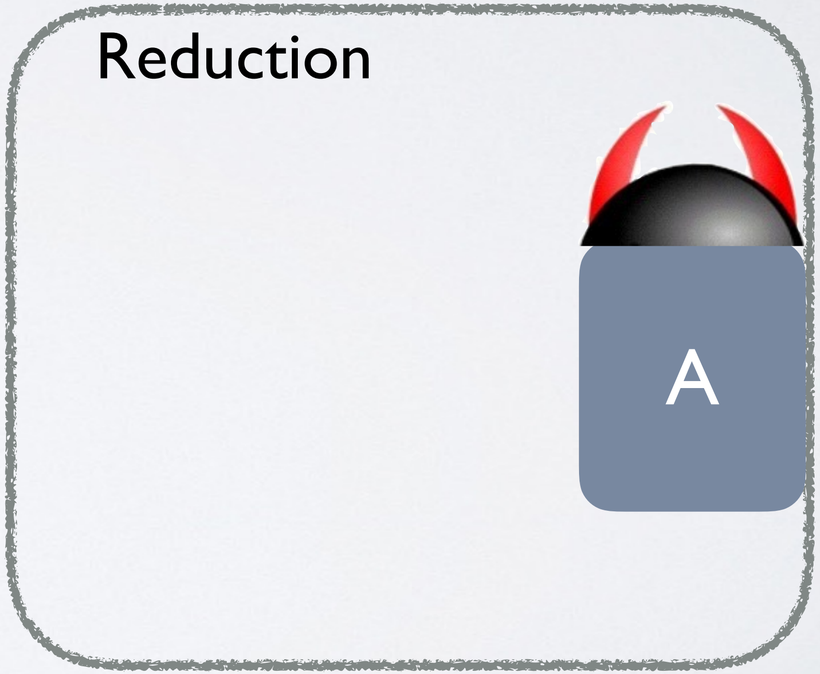




Conventional FDH Proof

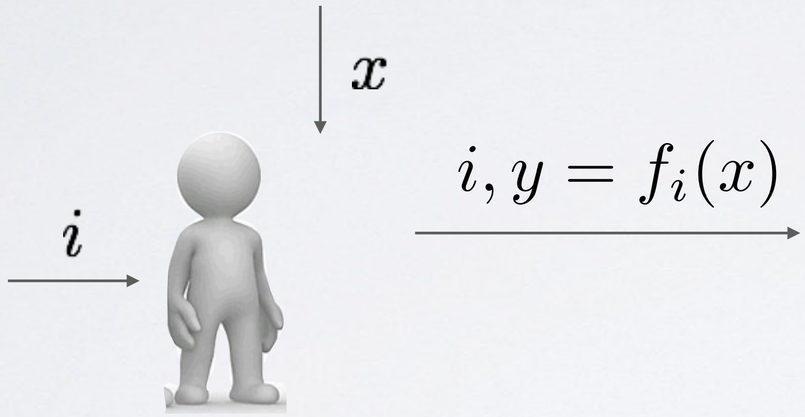
Embed the TDOWP challenge to one RO query answer:

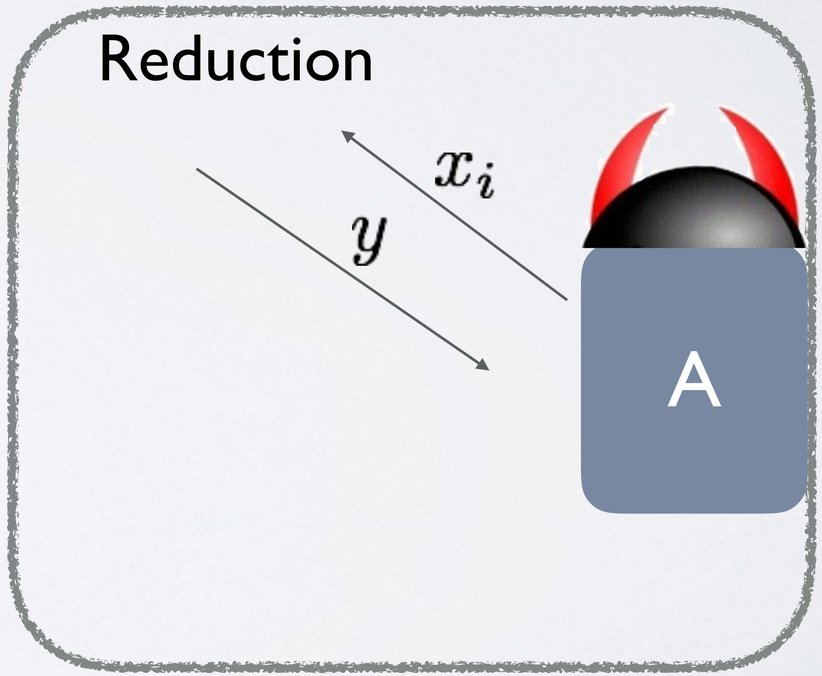




Conventional FDH Proof

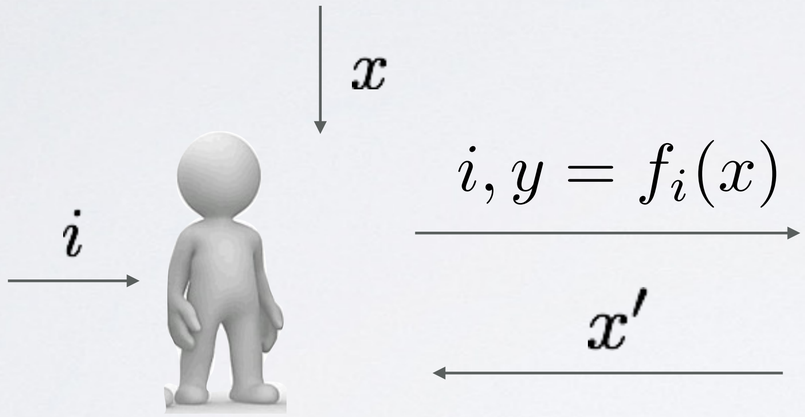
Embed the TDOWP challenge to one RO query answer:

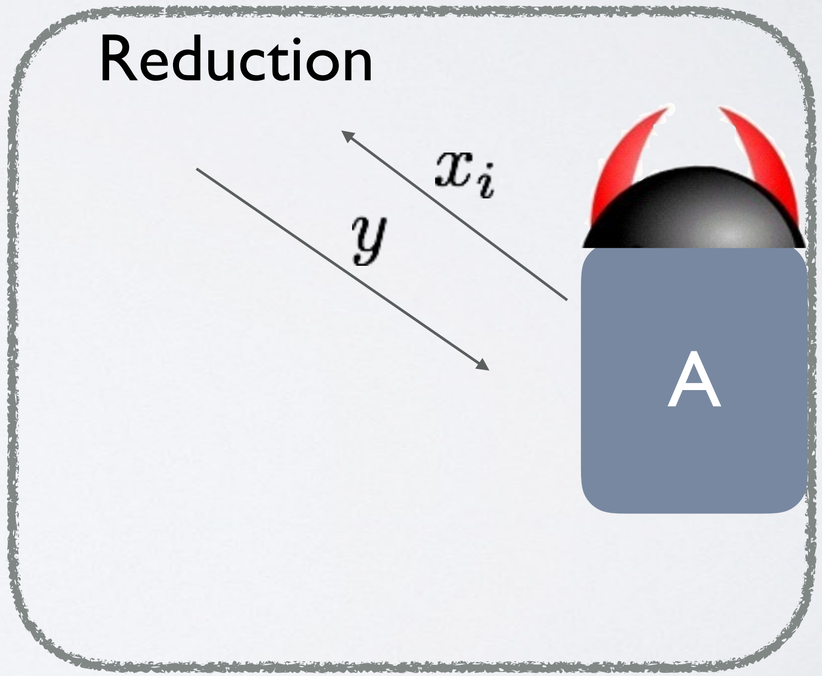




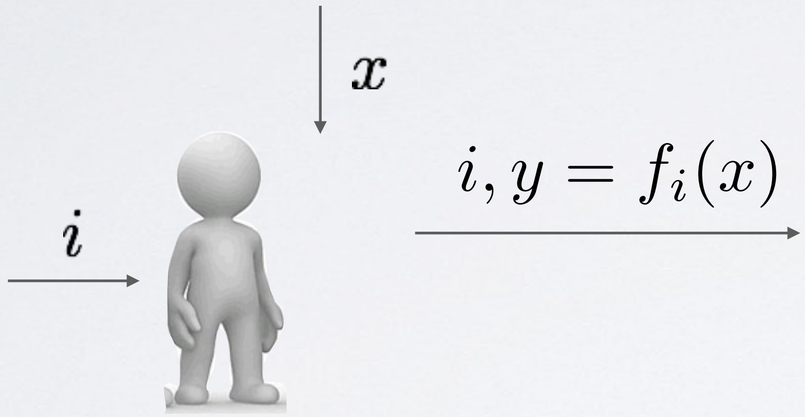
Conventional FDH Proof

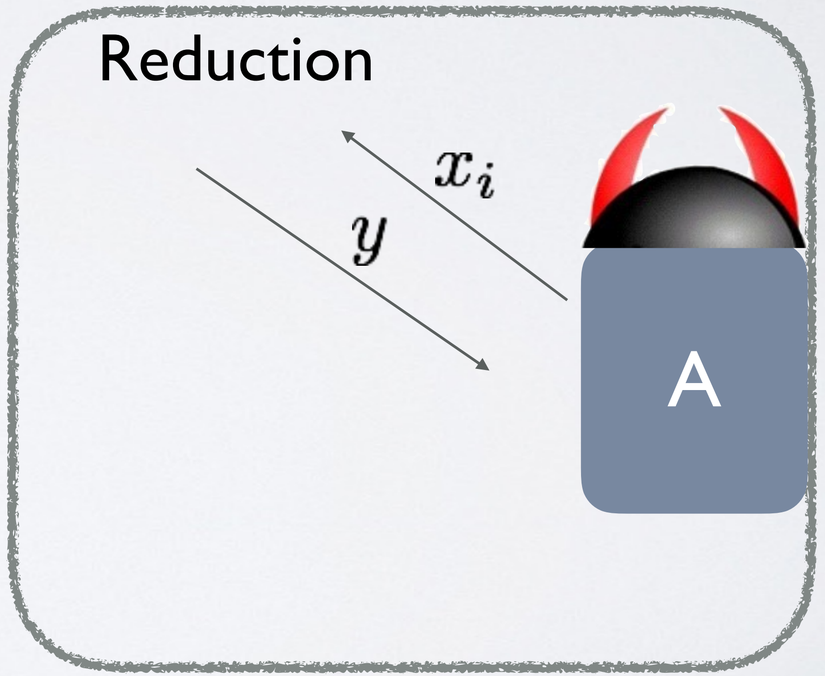
Embed the TDOWP challenge to one RO query answer:



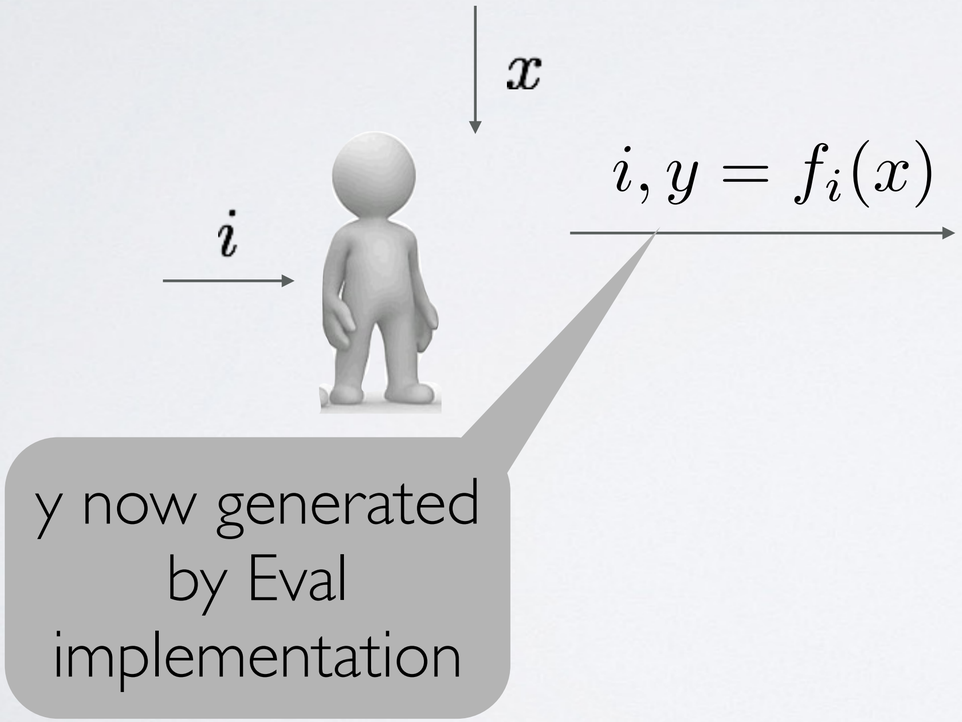


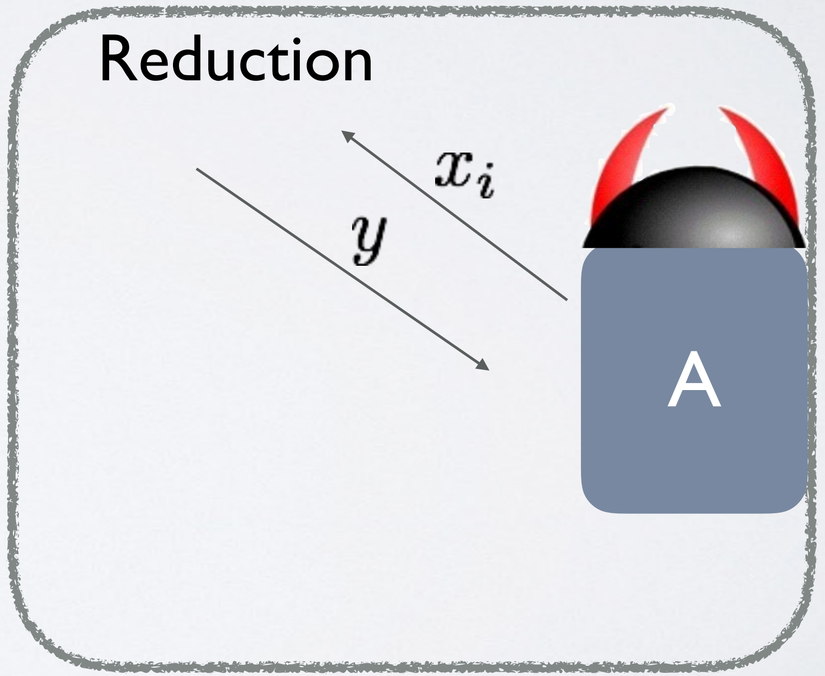
FDH in the Clipto Setting



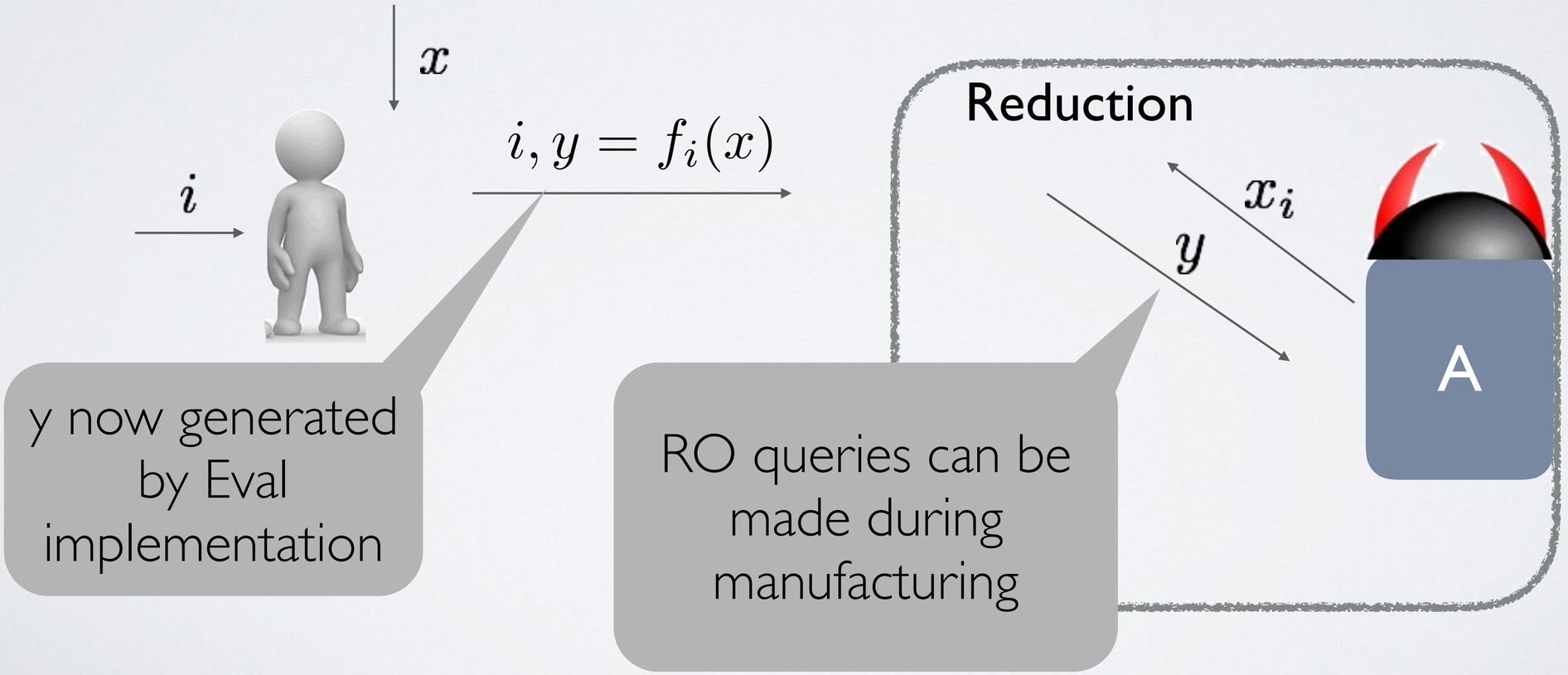


FDH in the Clipto Setting



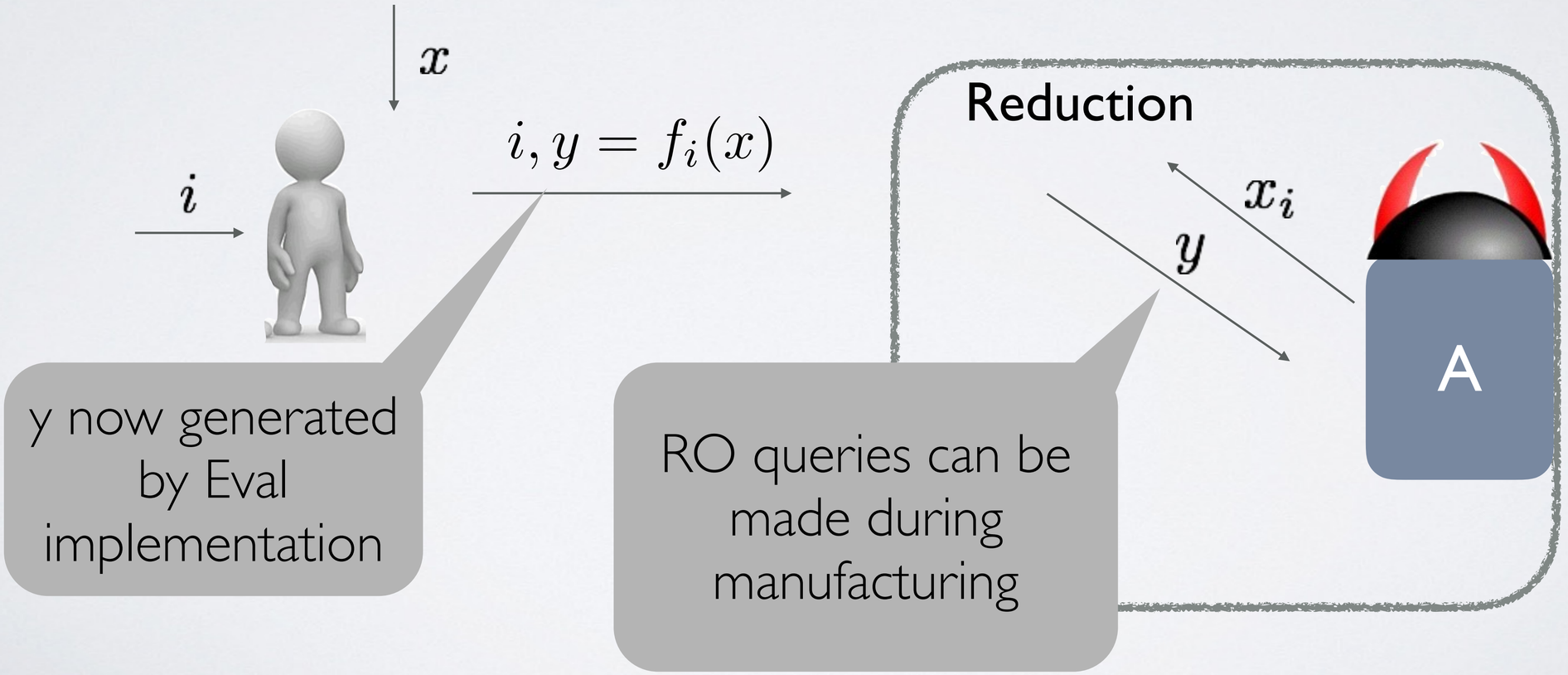


FDH in the Clipto Setting



FDH in the Clipto Setting

No way to embed TDOWP challenge



Revised FDH

Revised FDH

- Hash pk together with message

Revised FDH

- Hash pk together with message
  - RO queries have to be made after pk is generated which is after implementation is provided

Summary

Summary

- It is possible to save randomized algorithm from subversion with minimal trust via specification re-design

Summary

- It is possible to save randomized algorithm from subversion with minimal trust via specification re-design
- Landscape changes when adding one dimension, every piece of result worth revisiting

Open Problems

- Destroy subliminal channel
- Defend against hidden trigger attack
- Mitigating in the standard model
- Revisit cryptography, and build a robust cliptography theory
- Connection between correctness under subversion to self-correcting programs
- Many more…

Our Recent Progress: Destroying Subliminal Channel

General result of destroying subliminal channels and saving PKE to preserve IND-CPA security

Our Recent Progress: Signature with Offline Watchdog

Self-correcting random oracle and defend against hidden trigger attack for signatures

Cliptography: Clipping The Power Of Kleptographic Attacks

Alexander Russell, Qiang Tang, Moti Yung and Hong-Sheng Zhou

http://eprint.iacr.org/2015/695