Searchable Symmetric Encryption

Seny Kamara

Advanced Topics in Network Security
Spring 2006

Yesterday

- Motivation for searchable encryption
- First SSE scheme [SWP00]
- Attacks on [SWP00]
- Conjunctive SSE [GSW04,PKL04,BKM05]

Today

- Limitations of Song et al.'s security model
- More formal work on SSE [Goh03,CM05]
- New definitions

Practical Techniques[SWP00]

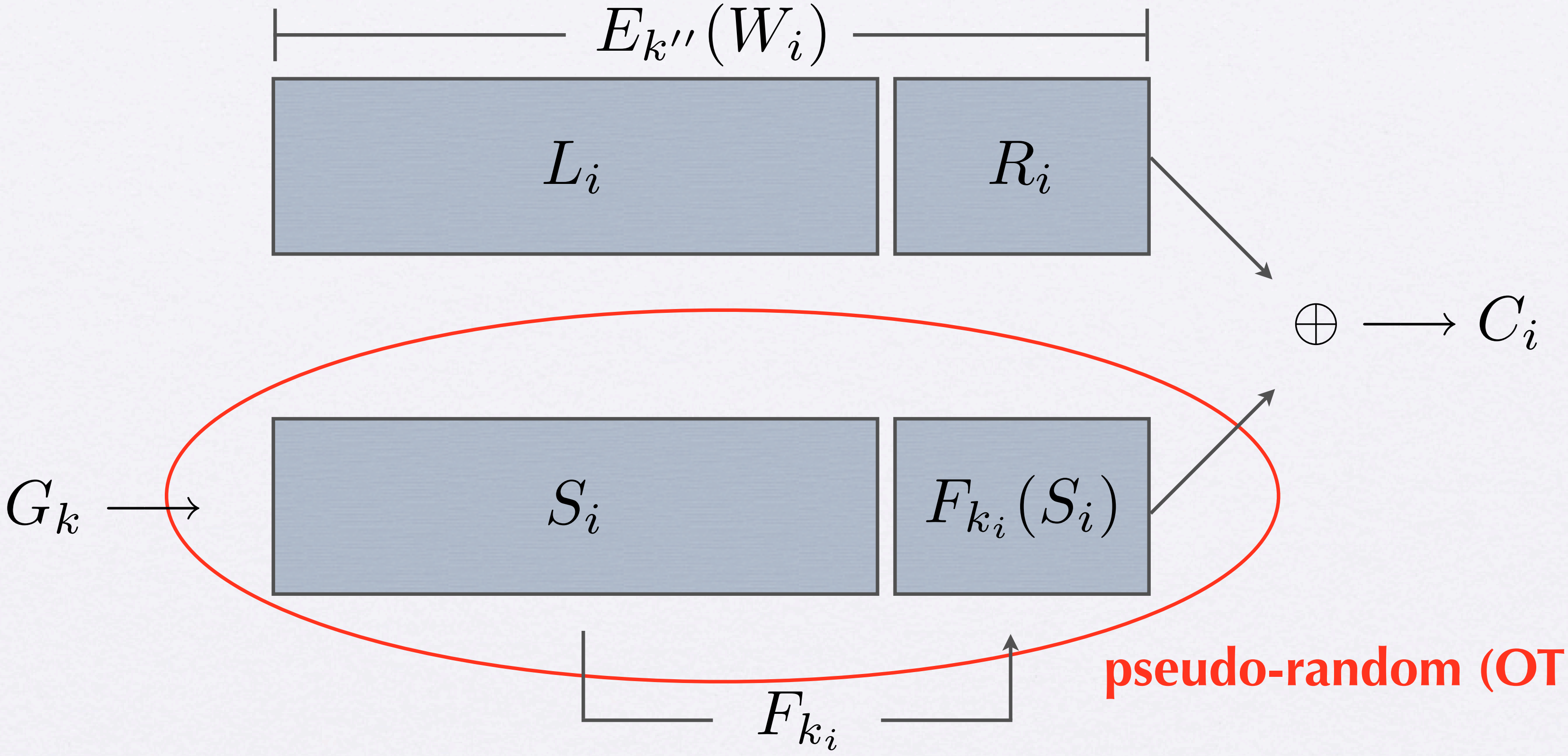
- Song et al. provide proofs of security
  - **"Our techniques are provably secure"** (p. 1)
- Yet
  - there are statistical attacks
  - leaks location of words

What's Going on?

- Are the proofs wrong?
- What are they proving?
- Is it meaningful?

What are they Proving?

$$k_i \leftarrow f_{k'}(L_i)$$



$E_{k''}(W_i)$

L_i

R_i

$\oplus \longrightarrow C_i$

$G_k \longrightarrow$

S_i

$F_{k_i}(S_i)$

F_{k_i}

pseudo-random (OTP)

Is it Meaningful?

- Is proving that the key stream is pseudo-random useful?
- **Depends on the adversarial model!**

Adversarial Model

- **Who are we protecting against?**

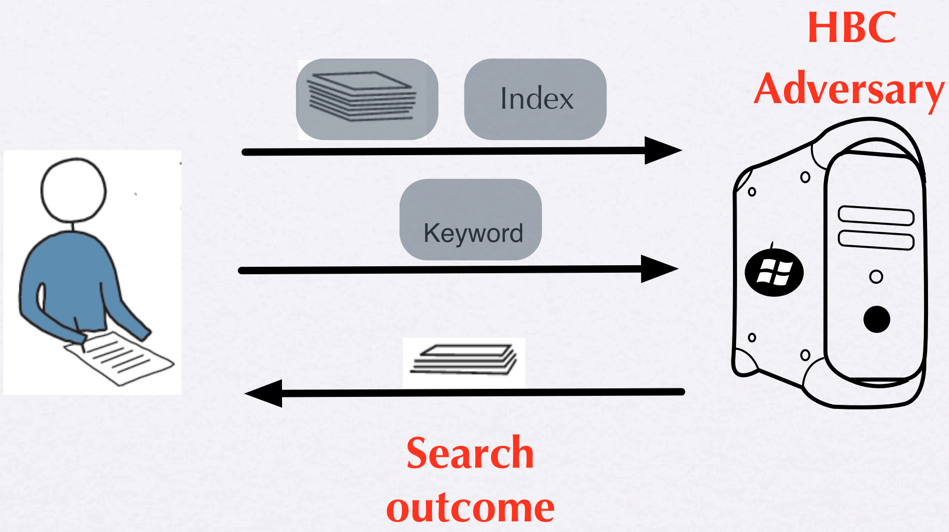
  the server

- **What are its goal?**

  info. about documents and keywords

- **How much power does it have?**

  it can search!



Index

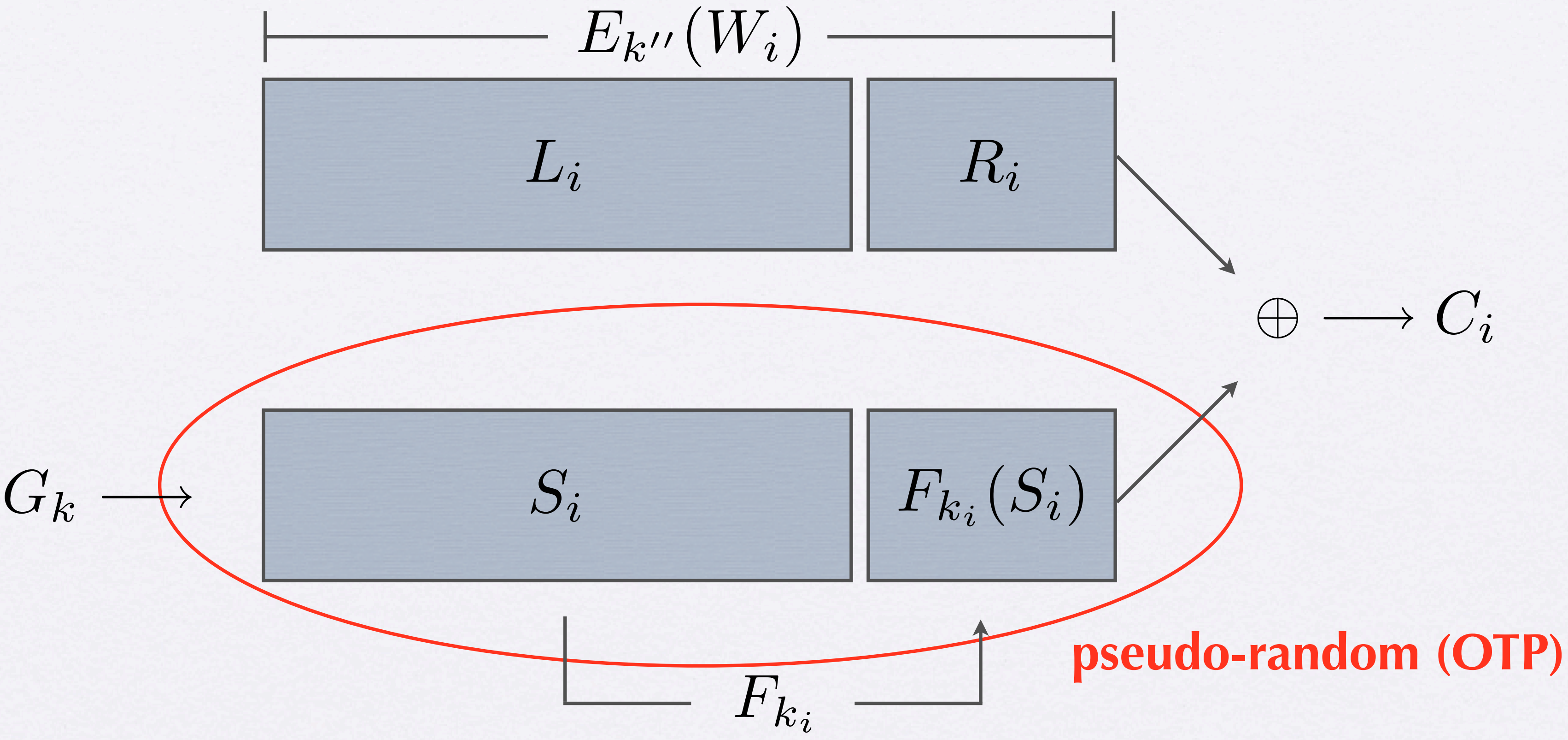
Keyword

HBC Adversary

Search outcome

What are they Proving?

$$k_i \leftarrow f_{k'}(L_i)$$



Is it Meaningful?

	Ideal model	**[SWP00]**
Adversary	server	server
Adv.'s Goal	recovering documents & keywords	recovering documents & keywords
Adv.'s Power	it can search	none
Meaning	documents and keywords are secure against server that can search	documents are secure against server that cannot search

Secure Indexes [Goh03]

- Introduces a stronger (better) security model
  - **IND2-CKA**: security against chosen-keyword attacks
- Provides **provably secure** and efficient construction
  - separates index from ciphertext
  - one index per document
  - based on pseudo-random functions & Bloom filters

Adversarial Model

- **Who are we protecting against?**

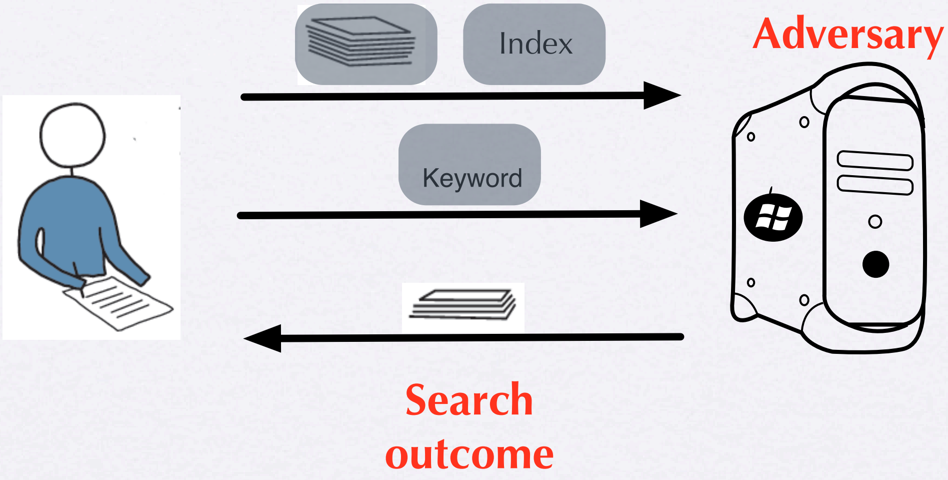
  the server

- **What are its goals?**

  info. about documents and keywords

- **How much power does it have?**

  it can search!



Formalizing the Adversarial Model

- How exactly do we capture the adversarial model formally?

Adversarial Model

- **Who are we protecting against?**

  the server

- **What are its goals?**

  info. about documents and keywords

- **How much power does it have?**

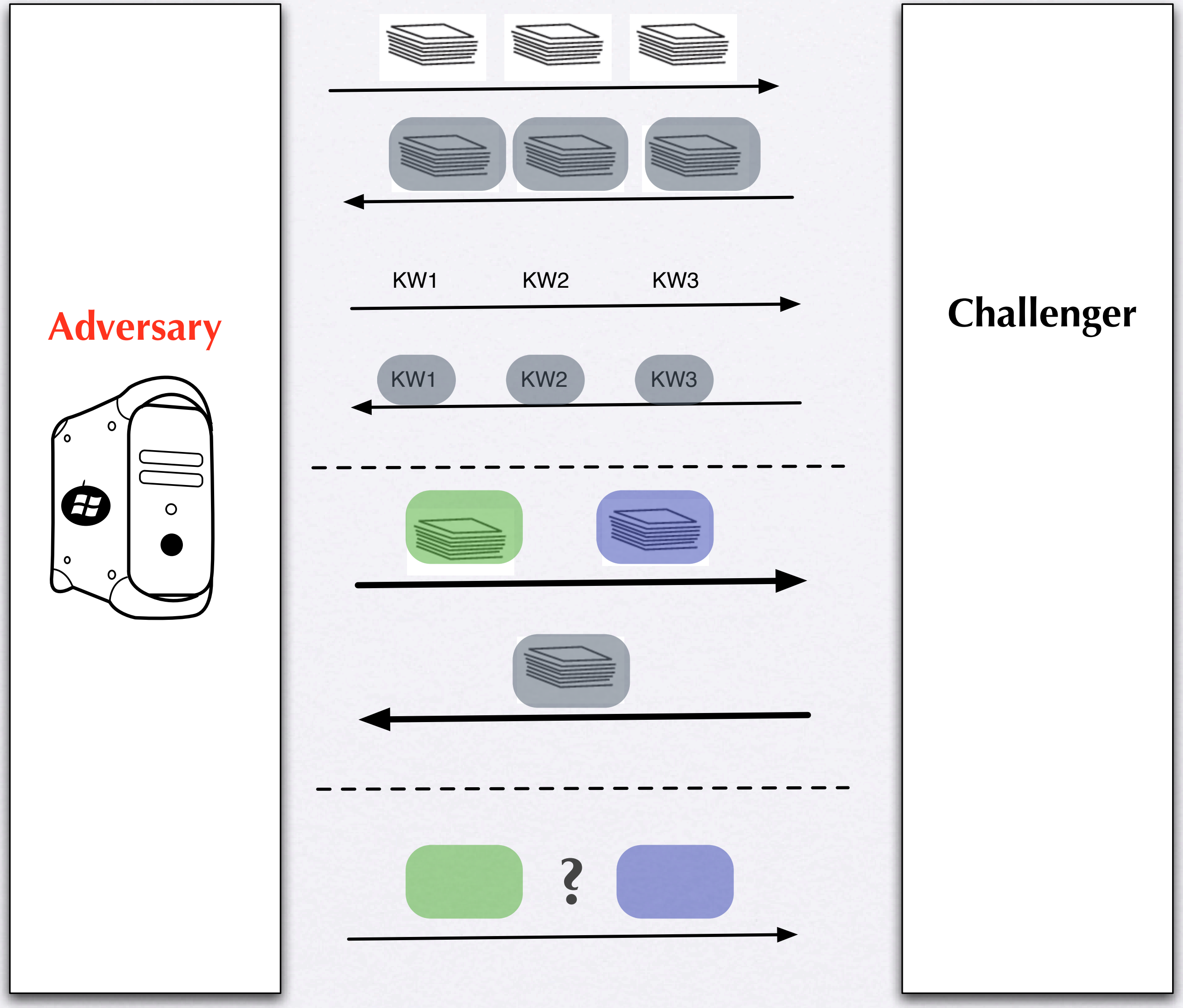
  it can search!

Probabilistic polynomial-time (PPT) algorithm

indistinguishability

allow adversary to generate and search many documents and keywords

IND2-CKA



Is it Meaningful?

	Ideal model	[SWP00]	IND2-CKA
Adversary	server	server	server
Adv.'s Goal	recovering documents & keywords	recovering documents & keywords	recovering documents
Adv.'s Power	it can search	none	it can search
Meaning	documents and keywords are secure against server that can search	documents are secure against server that cannot search	documents are secure against server that can search

Secure Indexes [Goh03]

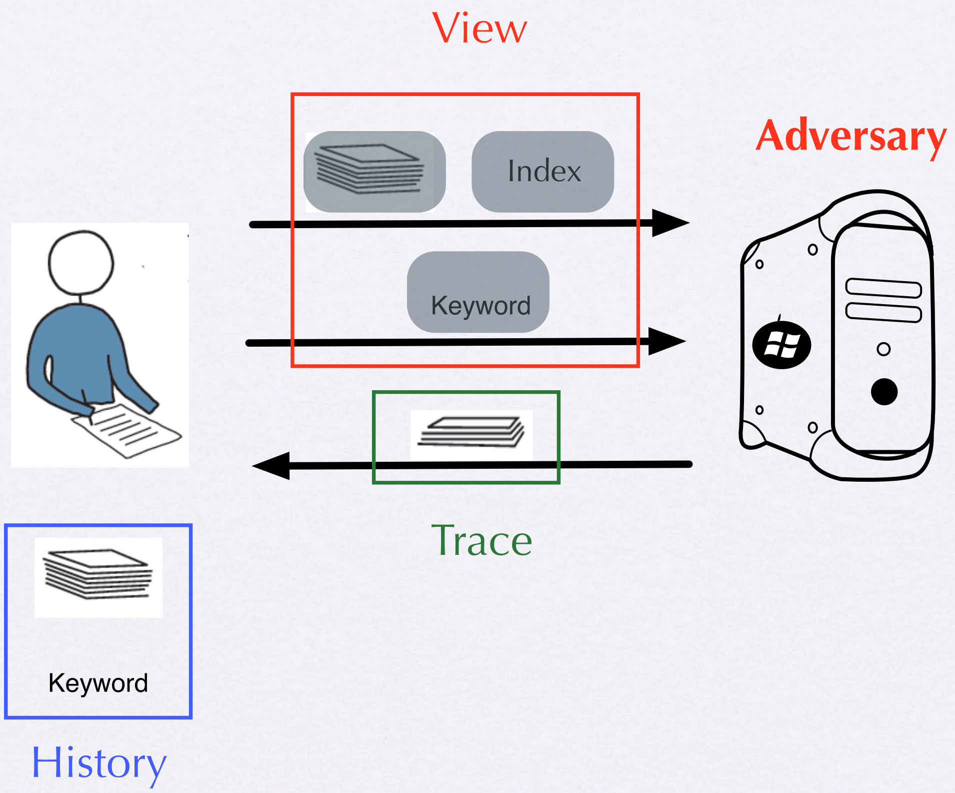
- Limitations:
    - IND2-CKA says nothing about trapdoors
    - One has to prove IND2-CKA + security of trapdoors

Privacy Preserving [CM05]

- Introduces a stronger security model than IND2-CKA
    - CM: security of index *and trapdoors* against chosen-keyword attacks
- Provides provably secure constructions
    - separates index from ciphertext
    - one index per document
    - Pseudo-random functions

CM Security [CM05]

- **History**: documents and words queried
- **View**: what the server sees
- **Trace**: minimum information leaked



View

Index

Keyword

Adversary

Trace

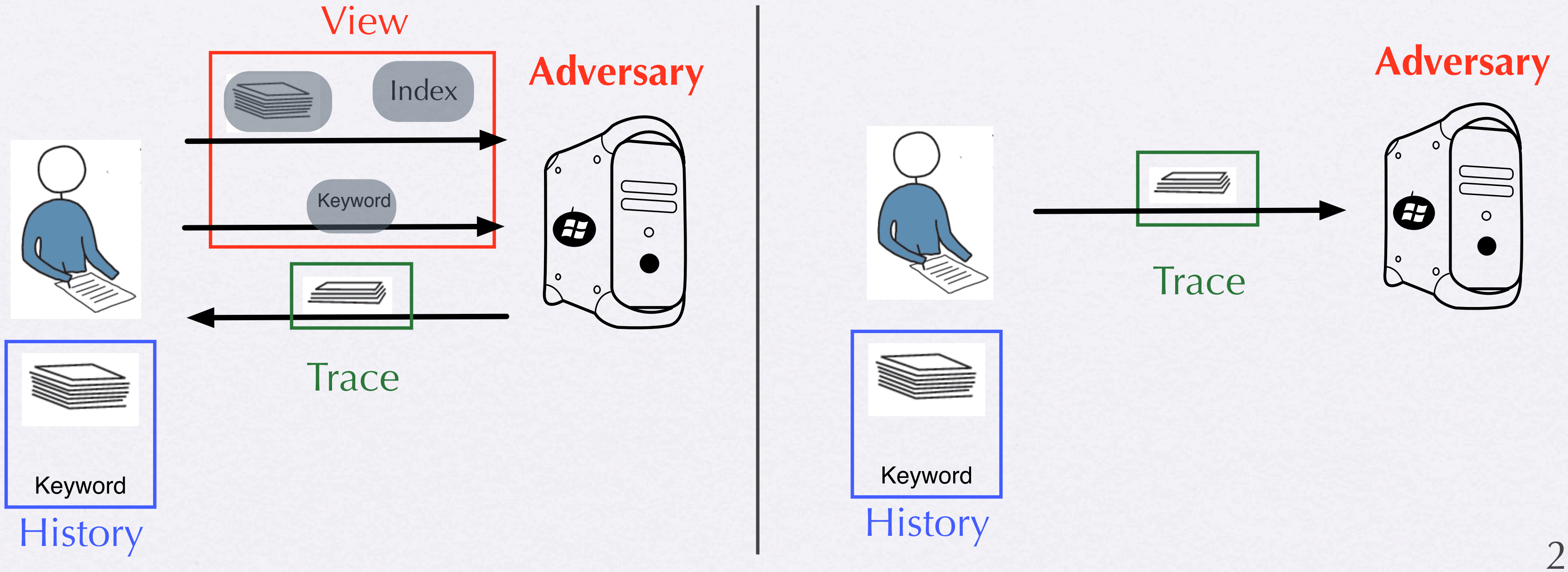
Keyword

History

CM Security [CM05]

- for all q, for all adversaries, for any function f, there exists a simulator such that for all histories

$$\left| \Pr\left[\begin{array}{c} \mathcal{A}(\mathsf{View_q}) = \\ f(\mathsf{History_q}) \end{array} \right] - \Pr\left[\begin{array}{c} \mathcal{S}(\mathsf{Trace_q}) = \\ f(\mathsf{History_q}) \end{array} \right] \right| \leq \mathsf{negl}(k)$$



CM Security [CM05]

- **Intuition**: anything the adversary can recover about the history from the view, can be recovered from the trace
- **Implication**: no adversary can recover any information about the documents or word queries that he is not supposed to

Is it Meaningful?

	Ideal model	**[SWP00]**	**IND2-CKA**	**CM**
Adversary	server	server	server	server
Adv.'s Goal	recovering documents & keywords	recovering documents & keywords	recovering documents	recovering documents & keywords
Adv.'s Power	it can search	none	it can search	it can search
Meaning	documents and keywords are secure against server that can search	documents are secure against server that ***cannot*** search	documents are secure against server that ***can search***	documents ***and keywords*** are secure against server that ***can search***

Is it Meaningful?

- So did Chang and Mitzenmacher finally get it right?
- Not exactly...

Is it Meaningful?

	Ideal model	[SWP00]	IND2-CKA	CM
Adversary	server	server	server	server
Adv.'s Goal	recovering documents & keywords	recovering documents & keywords	recovering documents	recovering documents & keywords
Adv.'s Power	it can search	none	it can search	it can search
Meaning	documents and keywords are secure against server that can search	documents are secure against server that ***cannot*** search	documents are secure against server that ***can search***	documents ***and keywords*** are secure against server that ***can search***