Searchable Encryption

Cloud Computing

Data breaches

Data on nearly 200 million US voters exposed in huge GOP contractor leak

By **James Rogers** • Published June 20, 2

Picture illustration. (REUTERS/Pawel Kopczynski)

The personal details of nearly 20(
exposed in the largest U.S. leak c

Information on more than 198 mill
the Internet by a firm working on t
Committee (RNC) in their efforts t
Vickery, a cyber-risk researcher a

FedEx Customer Data Left Publicly Exposed on Cloud Storage Server

By: Sean Michael Kerner | February 16, 2018

s time it's 119,000 documents with
FedEx-owned company Bongo

List of data breaches

From Wikipedia, the free encyclopedia

For broader coverage of this topic, see Data breach.

This is a list of **data breaches**, using data compiled from various sources, including press reports, government news releases, and mainstream news articles. The list includes those involving the
smaller breaches occur continually. Breaches of large organizations where the number of records is still unknown are also listed. The various methods used in the breaches are also listed, with ha

Most breaches occur in North America. It is estimated that the average cost of a data breach will be over $150 million by 2020, with the global annual cost forecast to be $2.1 trillion.[1][2] It is estim
exposed as a result of data breaches.[3] In 2019, a collection of 2.7 billion identity records, consisting of 774 million unique email addresses and 21 million unique passwords, was posted on the w

Entity	Year	Records	Organization type	Method	Sources
2019 Bulgarian revenue agency hack	2019	over 5,000,000	government	hacked	[54]
Canva	2019	140,000,000	web	hacked	[56][57][58]
Capital One	2019	106,000,000	financial	hacked	[59][60]
Desjardins	2019	2,900,000	financial	inside job	[89]
Facebook	2019	540,000,000	social network	poor security	[121]
Facebook	2019	1,500,000	social network	accidentally uploaded	[122]
First American Corporation	2019	885,000,000	financial service company	poor security	[124]
Health Sciences Authority (Singapore)	2019	808,000	healthcare	poor security	[151]
Justdial	2019	100,000,000	local search	unprotected api	[170]
Ministry of Health (Singapore)	2019	14,200	healthcare	poor security/inside job	[196][197]
Quest Diagnostics	2019	11,900,000	Clinical Laboratory	poor security	[223]
StockX	2019	6,800,000	retail	hacked	[255]
Truecaller	2019	299,055,819	Telephone directory	unknown	[277][278]
Woodruff Arts Center	2019	unknown	arts group	poor security	[314]
Westpac	2019	98,000	financial	hacked	[329]
Australian National University	2019	19 years of data	academic	hacked	[330]
AerServ (subsidiary of InMobi)	2018	75,000	advertising	hacked	[13]
Air Canada	2018	20,000	transport	hacked	[15]
Bell Canada	2018	100,000	telecoms	hacked	[42]
Bethesda Game Studios	2018		gaming	accidentally published	[44]
Blank Media Games	2018	7,633,234	gaming	hacked	[45][46]
BMO and Simplii	2018	90,000	banking	poor security	[50]
British Airways	2018	380,000	transport	hacked	[51][52]
Cathay Pacific Airways	2018	9,400,000	transport	hacked	[63]

AMERICA

Watchdog: Hillary Clinton Violated State Dept. Policies By Using Private Email

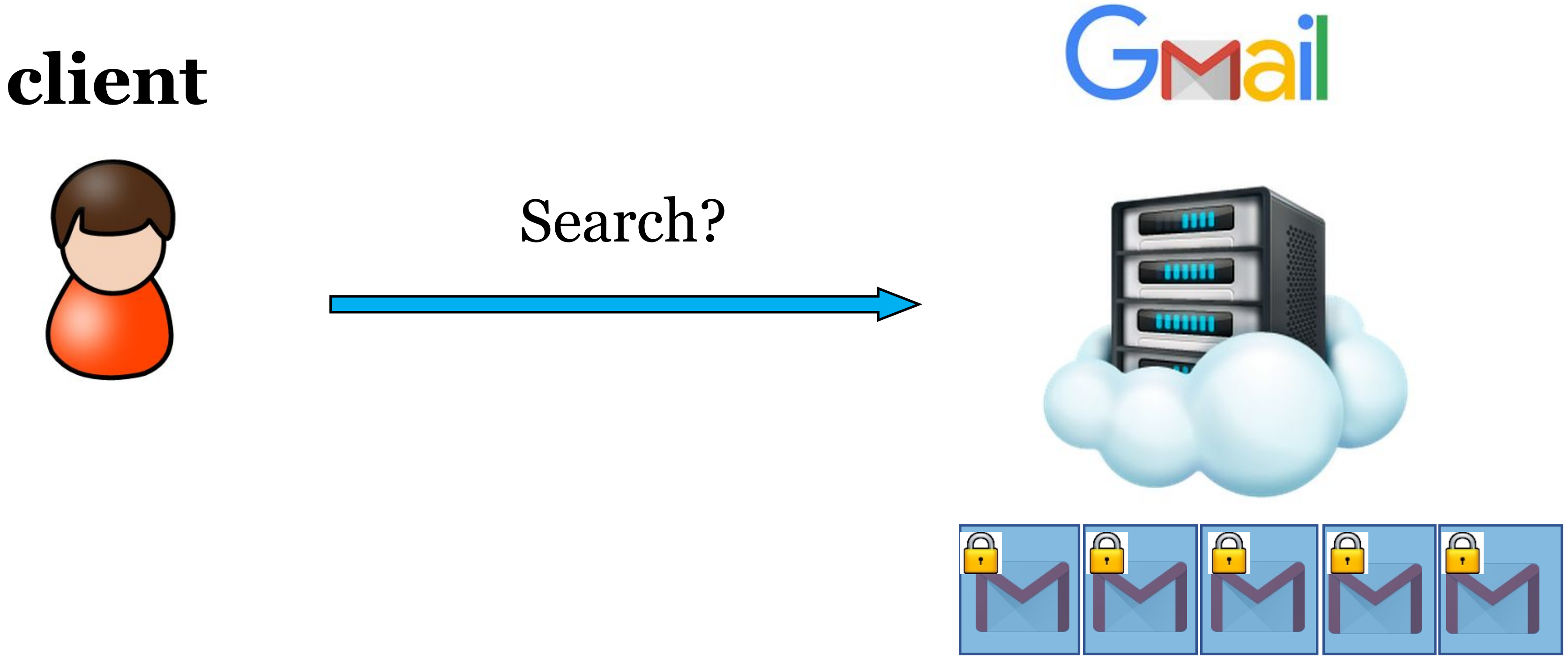
May 25, 2016 · 1:09 PM ET

EYDER PERALTA

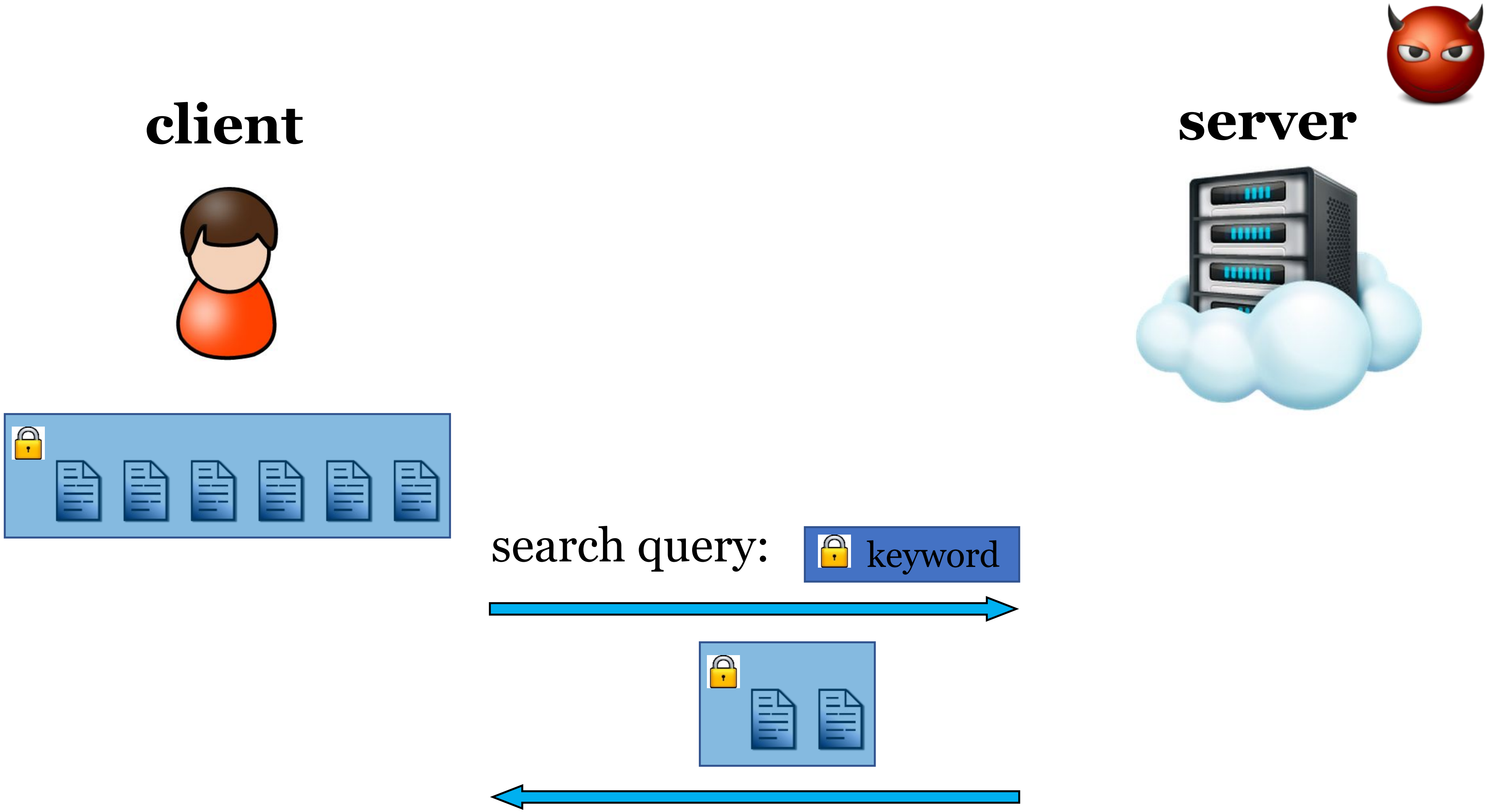
Democratic presidential candidate Hillary Clinton speaks at an International Brotherhood of Electrical Workers training center on Tuesday in Commerce, Calif.

John Locher/AP

Is encryption enough?



Searchable symmetric encryption (SSE)

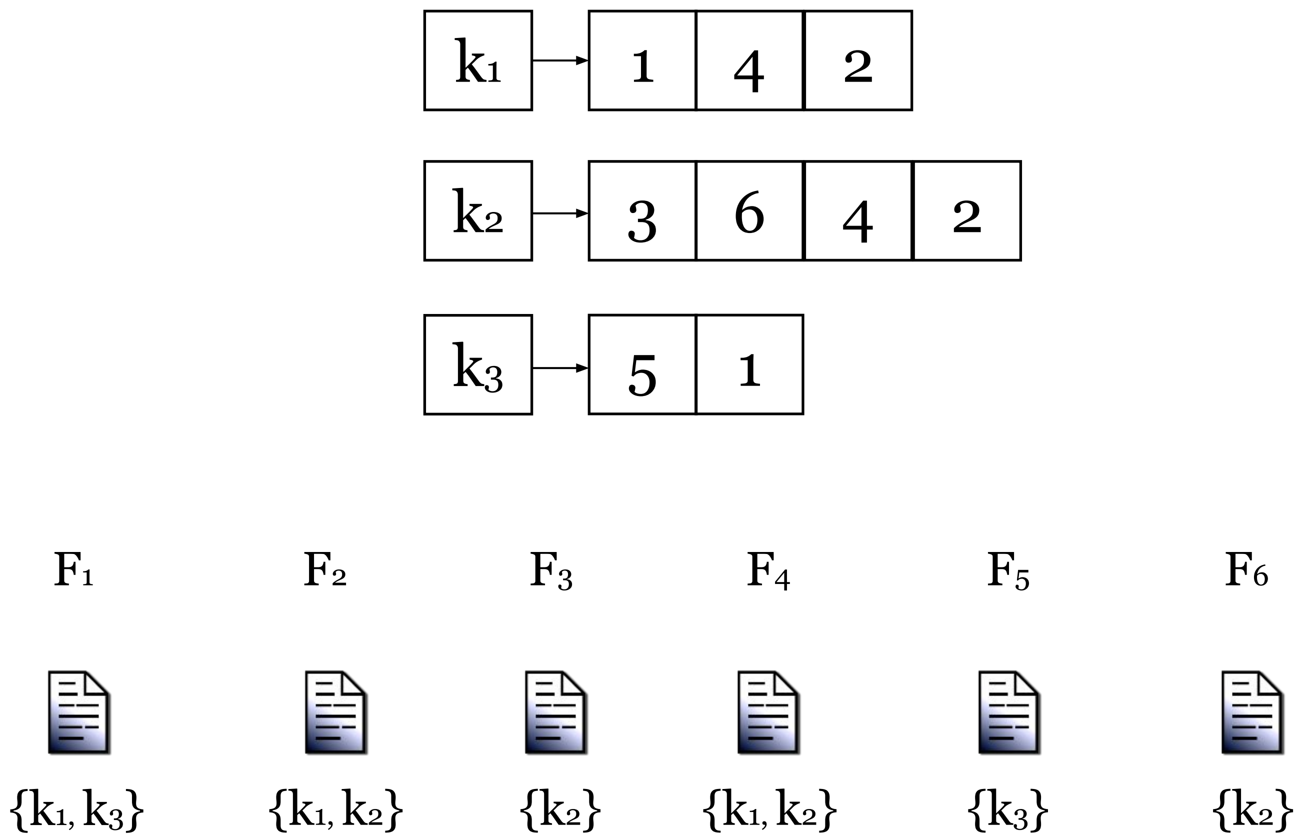


MPC, fully homomorphic encryption (FHE) and oblivious RAM (ORAM)

Solves the problem in theory, but too slow

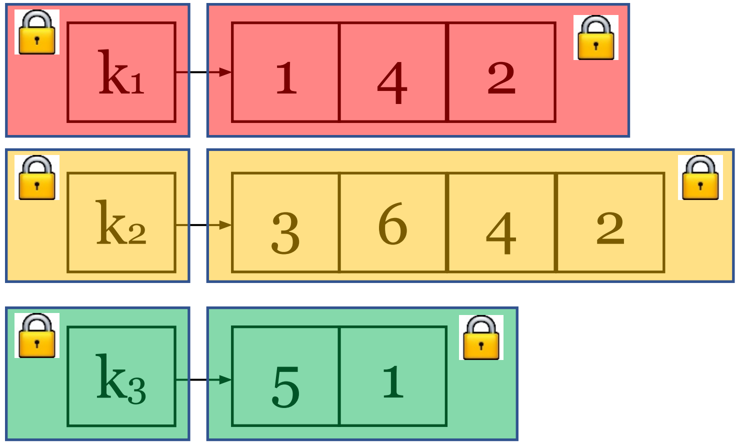
Static SSE

Inverted index

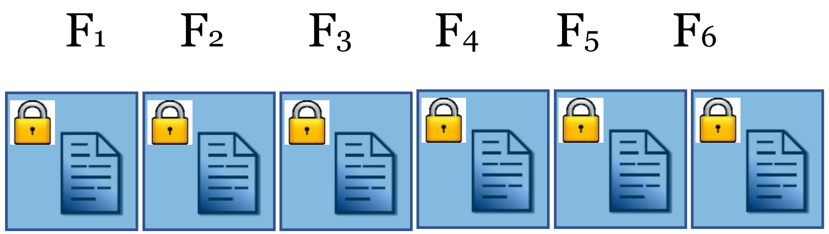


Encrypted index

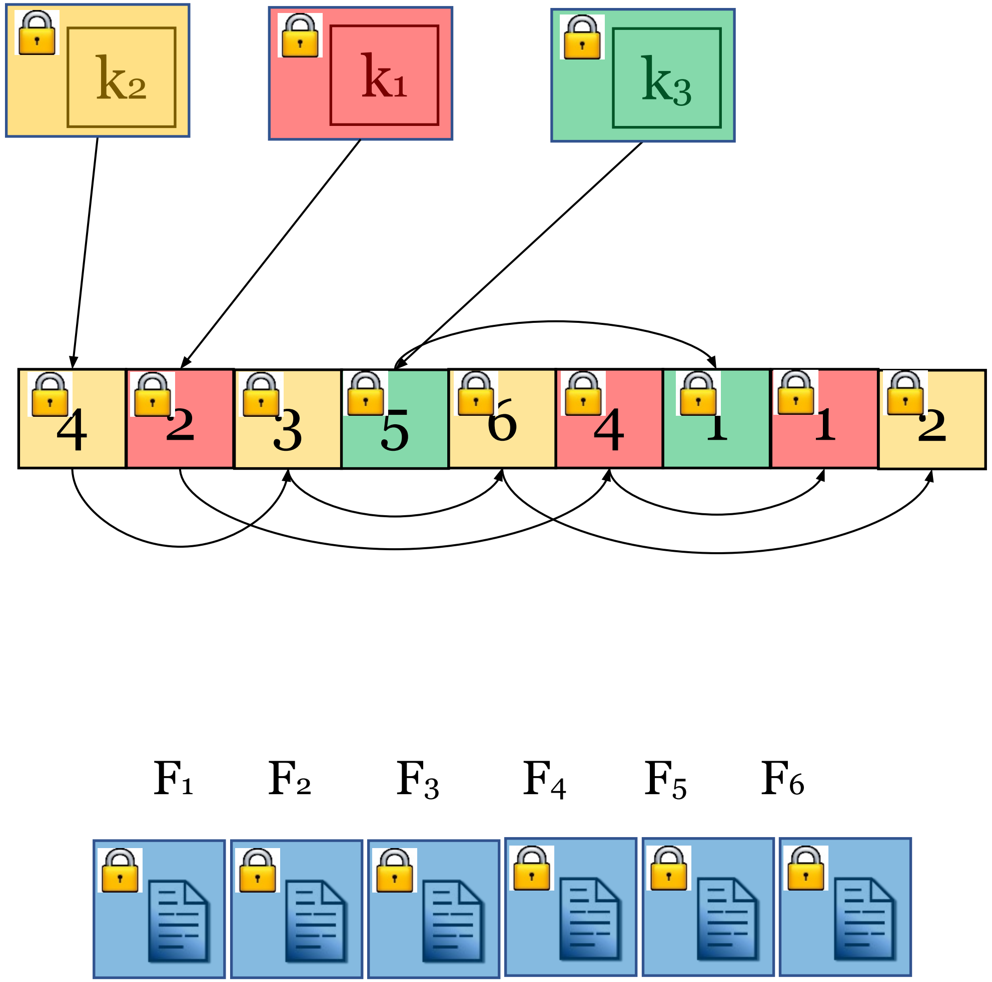
Pseudo random function:



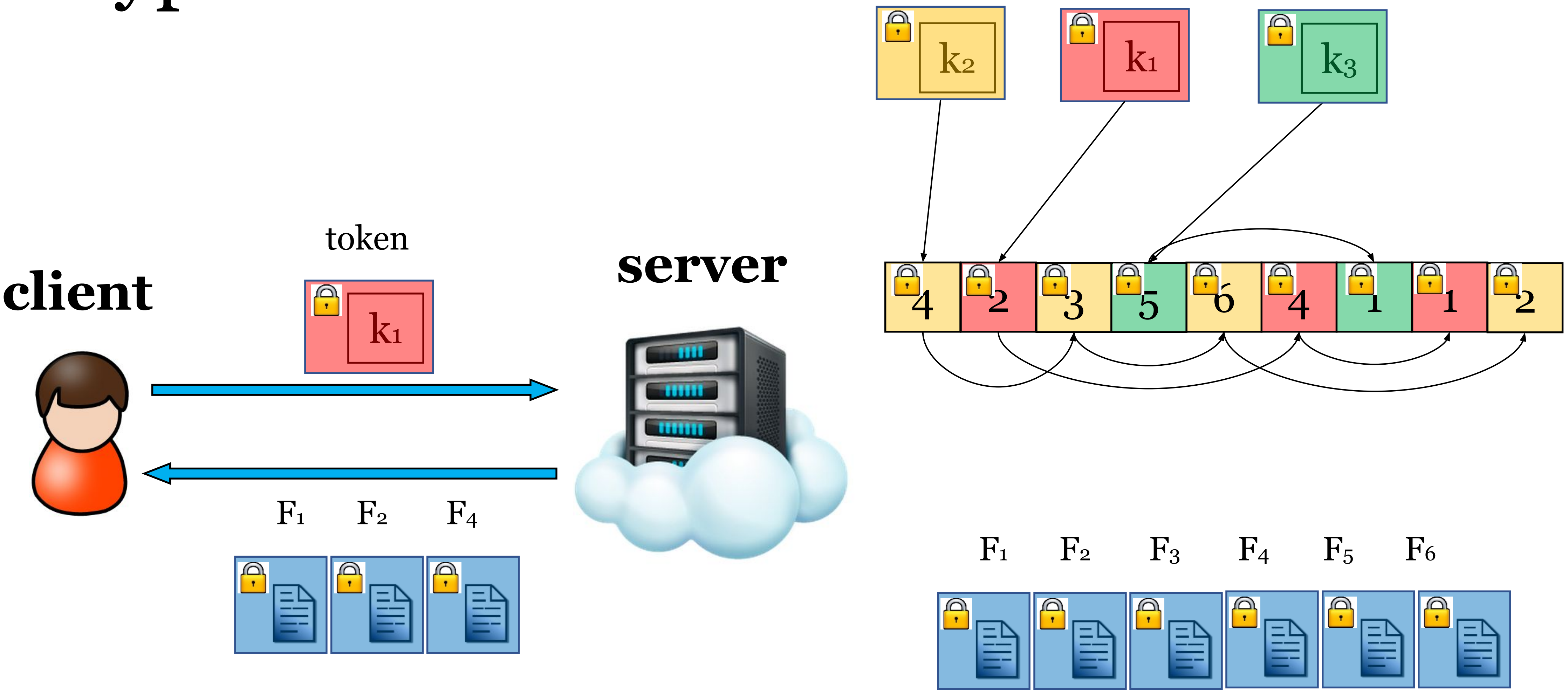
Encryption:



Encrypted index



Encrypted index



Advantage

- Super efficient in practice. Encrypt and PRF are not bottleneck.

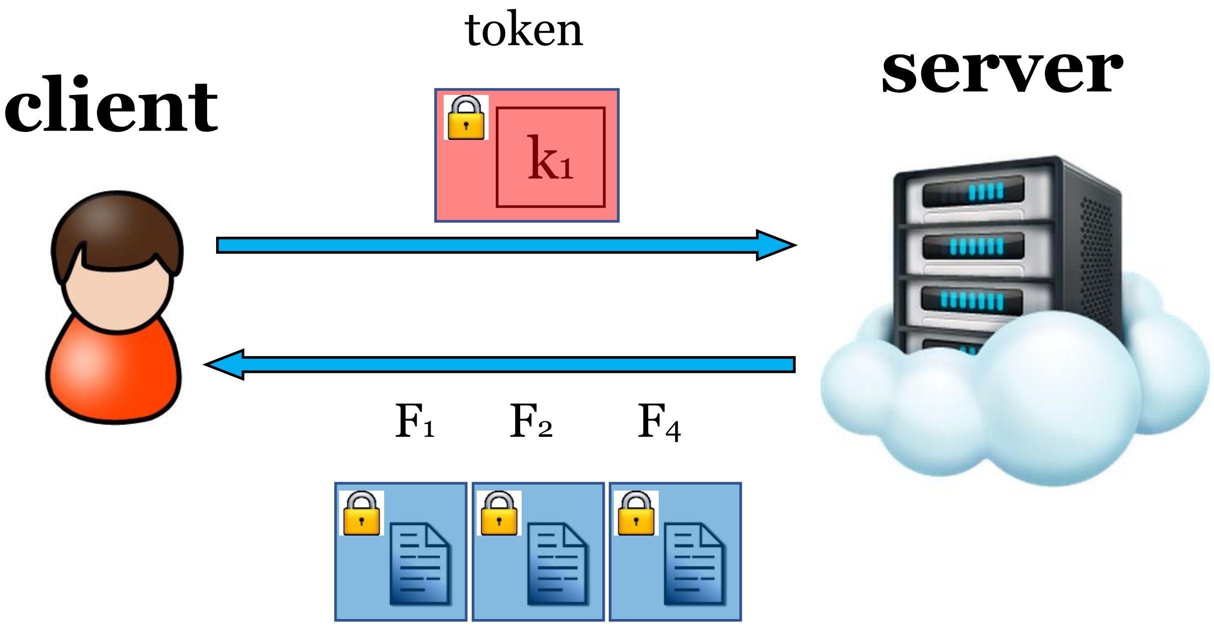
- People are looking at locality

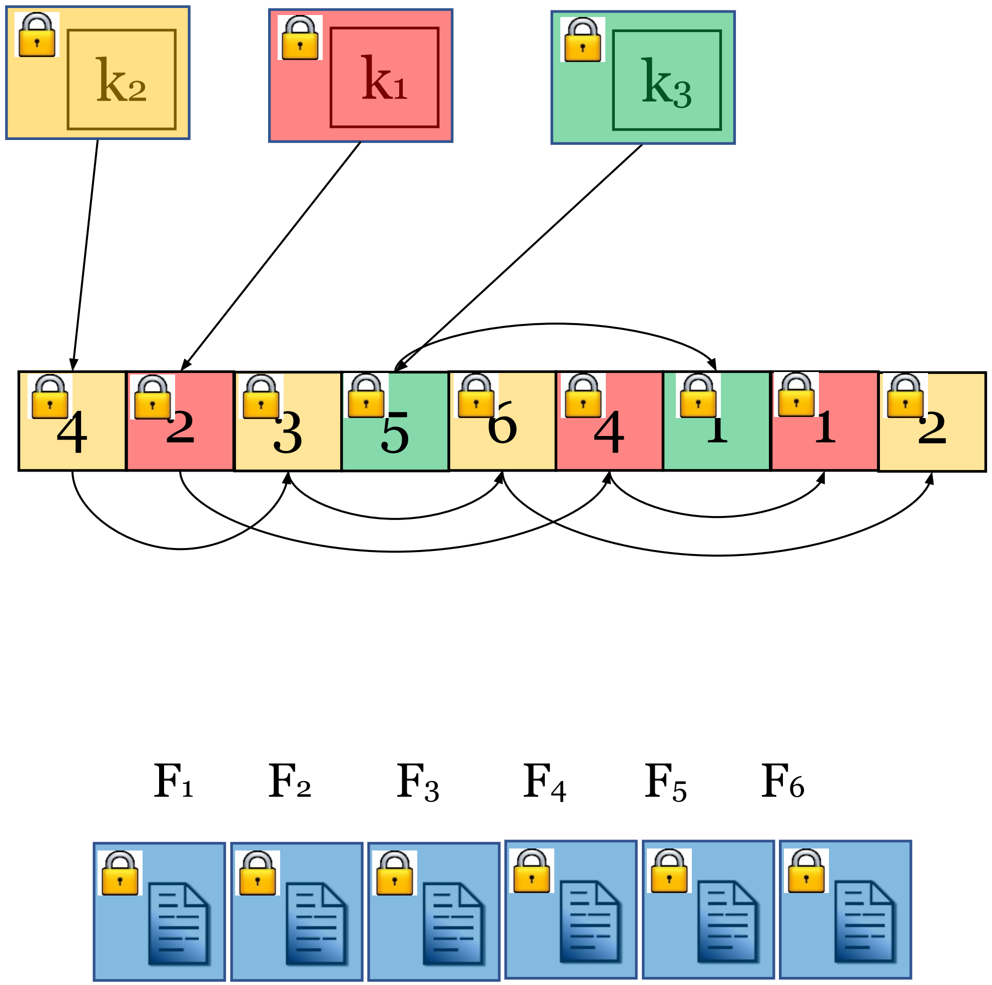
Leakage

- Search pattern leakage: token is generated from keyword deterministically

Encrypted index

deterministic!



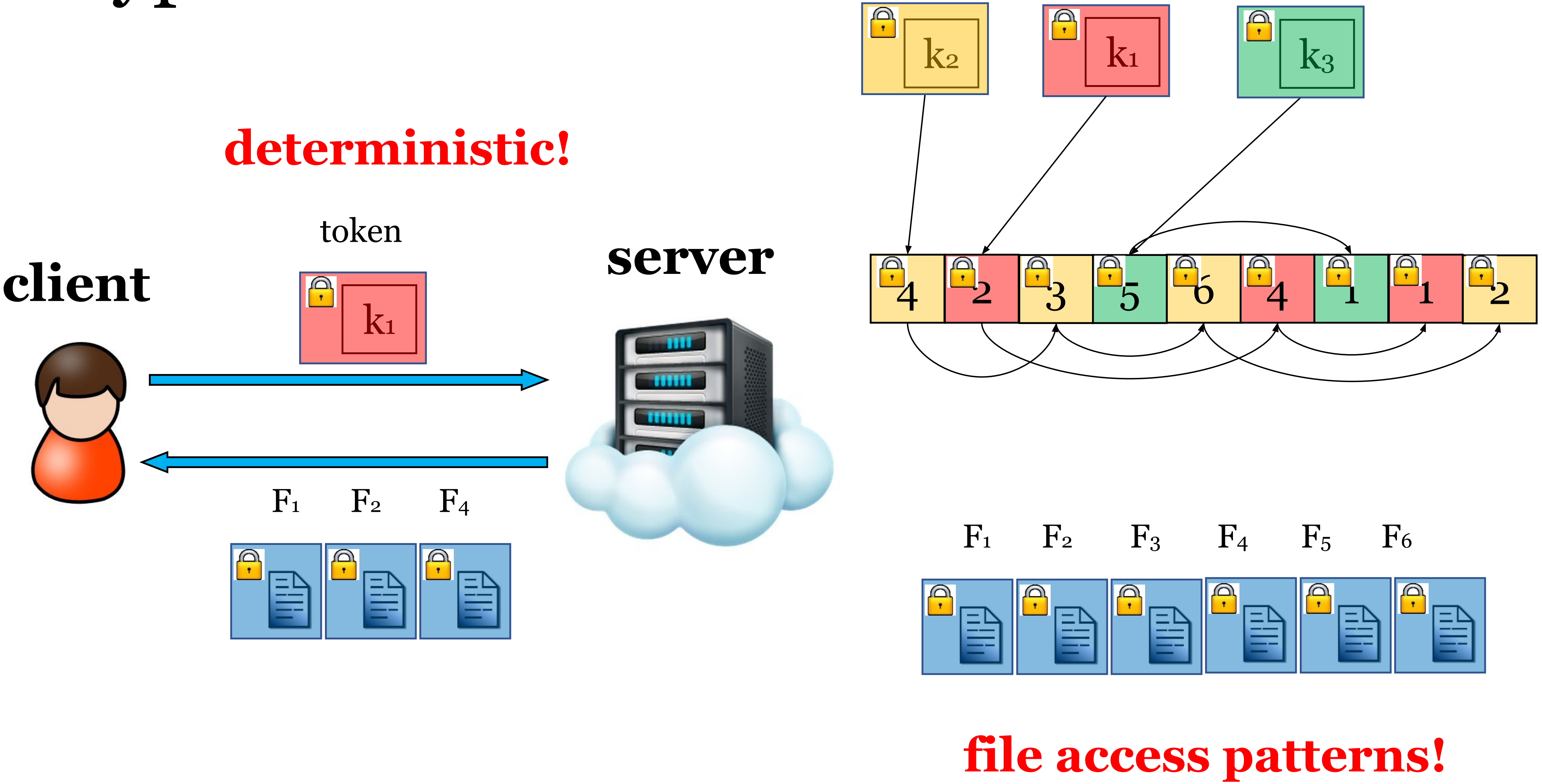


Leakage

- Search pattern leakage: token is generated from keyword deterministically

- Access pattern leakage: whether a file is returned

Encrypted index



Leakage

- Search pattern leakage: token is generated from keyword deterministically

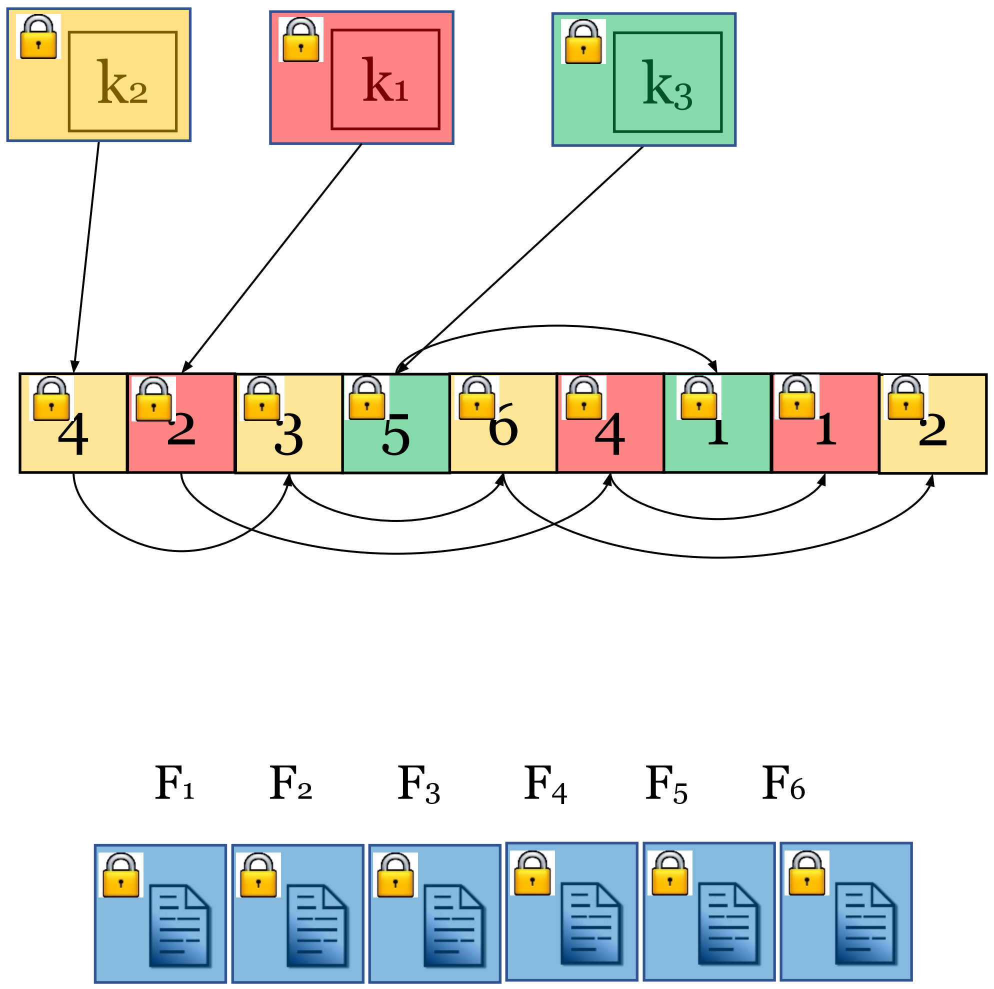
- Access pattern leakage: whether a file is returned

Different from MPC!

Security definitions

- $\Delta = (w_1, \ldots, w_d)$: keywords set
- $\mathbf{D} = (D_1, \ldots, D_n)$: documents set
- $\mathbf{D}(w)$: search result for w

- $\mathrm{K} \leftarrow \mathrm{Gen}(1^k)$
- $(\mathrm{I}, \mathrm{c}) \leftarrow \mathrm{Enc}(\mathrm{K}, \mathbf{D})$
- $t \leftarrow \mathrm{Trpdr}(\mathrm{K}, w)$
- $\mathrm{X} \leftarrow \mathrm{Search}(\mathrm{I}, \mathrm{t})$
- $D_i \leftarrow \mathrm{Dec}(\mathrm{K}, c_i)$



Correctness

An index-based SSE scheme is correct if for all $k \in \mathbb{N}$*, for all* K *output by* $\mathsf{Gen}(1^k)$*, for all* $\mathbf{D} \subseteq 2^{\Delta}$*, for all* $(I, \mathbf{c})$ *output by* $\mathsf{Enc}_K(\mathbf{D})$*, for all* $w \in \Delta$*,*

$$\mathsf{Search}\big(I, \mathsf{Trpdr}_K(w)\big) = \mathbf{D}(w) \bigwedge \mathsf{Dec}_K(c_i) = D_i, \ \text{for } 1 \leq i \leq n.$$

- $\Delta = (w_1, \dots, w_d)$: keywords set
- $\mathbf{D} = (D_1, \dots, D_n)$: documents set
- $\mathbf{D}(w)$: search result for w

- K ← Gen(1^k)
- (I, c) ← Enc(K, **D**)
- t ← Trpdr(K, w)
- X ← Search(I, t)
- D_i ← Dec(K, c_i)

Soundness/security

Definition 4.2 (History). *Let* Δ *be a dictionary and* $\mathbf{D} \subseteq 2^{\Delta}$ *be a document collection over* Δ. *A* q*-query* history *over* $\mathbf{D}$ *is a tuple* $H = (\mathbf{D}, \boldsymbol{w})$ *that includes the document collection* $\mathbf{D}$ *and a vector of* q *keywords* $\boldsymbol{w} = (w_1, \ldots, w_q)$.

Definition 4.3 (Access Pattern). *Let* Δ *be a dictionary and* $\mathbf{D} \subseteq 2^{\Delta}$ *be a document collection over* Δ. *The access pattern induced by a* q*-query history* $H = (\mathbf{D}, \boldsymbol{w})$, *is the tuple* $\alpha(H) = (\mathbf{D}(w_1), \ldots, \mathbf{D}(w_q))$.

Definition 4.4 (Search Pattern). *Let* Δ *be a dictionary and* $\mathbf{D} \subseteq 2^{\Delta}$ *be a document collection over* Δ. *The search pattern induced by a* q*-query history* $H = (\mathbf{D}, \boldsymbol{w})$, *is a symmetric binary matrix* $\sigma(H)$ *such that for* $1 \leq i, j \leq q$, *the element in the* i^{th} *row and* j^{th} *column is* 1 *if* $w_i = w_j$, *and* 0 *otherwise.*

- $\Delta = (w_1, \ldots, w_d)$: keywords set
- $\mathbf{D} = (D_1, \ldots, D_n)$: documents set
- $\mathbf{D}(w)$: search result for w

- K ← Gen(1^k)
- (I, c) ← Enc(K, $\mathbf{D}$)
- t ← Trpdr(K, w)
- X ← Search(I, t)
- D_i ← Dec(K, c_i)

Soundness/security

Definition 4.5 (Trace). *Let Δ be a dictionary and $\mathbf{D} \subseteq 2^{\Delta}$ be a document collection over Δ. The trace induced by a q-query history $H = (\mathbf{D}, \boldsymbol{w})$, is a sequence $\tau(H) = (|D_1|, \ldots, |D_n|, \alpha(H), \sigma(H))$ comprised of the lengths of the documents in $\mathbf{D}$, and the access and search patterns induced by H.*

Throughout this work, we will assume that the dictionary Δ and the trace are such that all histories H over Δ are *non-singular* as defined below.

Definition 4.6 (Non-singular history). *We say that a history H is non-singular if (1) there exists at least one history $H' \neq H$ such that $\tau(H) = \tau(H')$; and if (2) such a history can be found in polynomial-time given $\tau(H)$.*

- $\Delta = (w_1, \ldots, w_d)$: keywords set
- $\mathbf{D} = (D_1, \ldots, D_n)$: documents set
- $\mathbf{D}(w)$: search result for w

- K ← Gen(1^k)
- (I, c) ← Enc(K, **D**)
- t ← Trpdr(K, w)
- X ← Search(I, t)
- D_i ← Dec(K, c_i)

Soundness/security

Definition 4.7 (Non-adaptive indistinguishability). *Let* $\mathsf{SSE} = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Trpdr}, \mathsf{Search}, \mathsf{Dec})$ *be an index-based* SSE *scheme over a dictionary* Δ, $k \in \mathbb{N}$ *be the security parameter, and* $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ *be a non-uniform adversary and consider the following probabilistic experiment* $\mathbf{Ind}_{\mathsf{SSE},\mathcal{A}}(k)$*:*

$$
\begin{array}{l}
\mathbf{Ind}_{\mathsf{SSE},\mathcal{A}}(k) \\
\quad K \leftarrow \mathsf{Gen}(1^k) \\
\quad (st_{\mathcal{A}}, H_0, H_1) \leftarrow \mathcal{A}_1(1^k) \\
\quad b \xleftarrow{\$} \{0,1\} \\
\quad \textit{parse } H_b \textit{ as } (\mathbf{D}_b, \boldsymbol{w}_b) \\
\quad (I_b, \boldsymbol{c}_b) \leftarrow \mathsf{Enc}_K(\mathbf{D}_b) \\
\quad \textit{for } 1 \le i \le q, \\
\qquad t_{b,i} \leftarrow \mathsf{Trpdr}_K(w_{b,i}) \\
\quad \textit{let } \boldsymbol{t}_b = (t_{b,1}, \ldots, t_{b,q}) \\
\quad b' \leftarrow \mathcal{A}_2(st_{\mathcal{A}}, I_b, \boldsymbol{c}_b, \boldsymbol{t}_b) \\
\quad \textit{if } b' = b, \textit{ output } 1 \\
\quad \textit{otherwise output } 0
\end{array}
$$

with the restriction that $\tau(H_0) = \tau(H_1)$*, and where* $st_{\mathcal{A}}$ *is a string that captures* $\mathcal{A}_1$*'s state. We say that* SSE *is secure in the sense of* non-adaptive indistinguishability *if for all polynomial-size adversaries* $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$,

$$
\Pr\left[\,\mathbf{Ind}_{\mathsf{SSE},\mathcal{A}}(k) = 1\,\right] \le \frac{1}{2} + \mathsf{negl}(k),
$$

where the probability is taken over the choice of b *and the coins of* Gen *and* Enc.

- $\Delta = (w_1, \ldots, w_d)$: keywords set
- $\mathbf{D} = (D_1, \ldots, D_n)$: documents set
- $\mathbf{D}(w)$: search result for w

- K ← Gen(1^k)
- (I, c) ← Enc(K, **D**)
- t ← Trpdr(K, w)
- X ← Search(I, t)
- D_i ← Dec(K, c_i)

Simulation-based security definition

Definition 4.8 (Non-adaptive semantic security). *Let* SSE = (Gen, Enc, Trpdr, Search, Dec) *be an index-based* SSE *scheme,* $k \in \mathbb{N}$ *be the security parameter,* $\mathcal{A}$ *be an adversary,* $\mathcal{S}$ *be a simulator and consider the following probabilistic experiments:*

$$
\begin{array}{ll}
\mathbf{Real}_{\mathsf{SSE},\mathcal{A}}(k) & \mathbf{Sim}_{\mathsf{SSE},\mathcal{A},\mathcal{S}}(k) \\
\quad K \leftarrow \mathsf{Gen}(1^k) & \quad (H, st_{\mathcal{A}}) \leftarrow \mathcal{A}(1^k) \\
\quad (st_{\mathcal{A}}, H) \leftarrow \mathcal{A}(1^k) & \quad \mathrm{v} \leftarrow \mathcal{S}(\tau(H)) \\
\quad \textit{parse } H \textit{ as } (\mathbf{D}, \boldsymbol{w}) & \quad \textit{output } \mathrm{v} \textit{ and } st_{\mathcal{A}} \\
\quad (I, \mathbf{c}) \leftarrow \mathsf{Enc}_K(\mathbf{D}) & \\
\quad \textit{for } 1 \le i \le q, & \\
\qquad t_i \leftarrow \mathsf{Trpdr}_K(w_i) & \\
\quad \textit{let } \boldsymbol{t} = (t_1, \ldots, t_q) & \\
\quad \textit{output } \mathrm{v} = (I, \mathbf{c}, \mathbf{t}) \textit{ and } st_{\mathcal{A}} &
\end{array}
$$

We say that SSE *is semantically secure if for all polynomial-size adversaries* $\mathcal{A}$*, there exists a polynomial-size simulator* $\mathcal{S}$ *such that for all polynomial-size distinguishers* $\mathcal{D}$*,*

$$\left|\Pr\left[\mathcal{D}(\mathrm{v}, st_{\mathcal{A}}) = 1 : (\mathrm{v}, st_{\mathcal{A}}) \leftarrow \mathbf{Real}_{\mathsf{SSE},\mathcal{A}}(k)\right] - \Pr\left[\mathcal{D}(\mathrm{v}, st_{\mathcal{A}}) = 1 : (\mathrm{v}, st_{\mathcal{A}}) \leftarrow \mathbf{Sim}_{\mathsf{SSE},\mathcal{A},\mathcal{S}}(k)\right]\right| \le \mathsf{negl}(k),$$

where the probabilities are over the coins of Gen *and* Enc.

- $\Delta = (w_1, \dots, w_d)$: keywords set
- $\mathbf{D} = (D_1, \dots, D_n)$: documents set
- $\mathbf{D}(w)$: search result for w

- $\mathrm{K} \leftarrow \mathrm{Gen}(1^k)$
- $(\mathrm{I}, \mathrm{c}) \leftarrow \mathrm{Enc}(\mathrm{K}, \mathbf{D})$
- $t \leftarrow \mathrm{Trpdr}(\mathrm{K}, w)$
- $\mathrm{X} \leftarrow \mathrm{Search}(\mathrm{I}, \mathrm{t})$
- $D_i \leftarrow \mathrm{Dec}(\mathrm{K}, c_i)$

Non-adaptive vs. adaptive

Definition 4.7 (Non-adaptive indistinguishability). *Let* $\mathsf{SSE} = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Trpdr}, \mathsf{Search}, \mathsf{Dec})$ *be an index-based* SSE *scheme over a dictionary* Δ, $k \in \mathbb{N}$ *be the security parameter, and* $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ *be a non-uniform adversary and consider the following probabilistic experiment* $\mathbf{Ind}_{\mathsf{SSE},\mathcal{A}}(k)$*:*

$$
\begin{aligned}
&\mathbf{Ind}_{\mathsf{SSE},\mathcal{A}}(k)\\
&\quad K \leftarrow \mathsf{Gen}(1^k)\\
&\quad (st_{\mathcal{A}}, H_0, H_1) \leftarrow \mathcal{A}_1(1^k)\\
&\quad b \xleftarrow{\$} \{0,1\}\\
&\quad \textit{parse } H_b \textit{ as } (\mathbf{D}_b, \boldsymbol{w}_b)\\
&\quad (I_b, \mathbf{c}_b) \leftarrow \mathsf{Enc}_K(\mathbf{D}_b)\\
&\quad \textit{for } 1 \le i \le q,\\
&\quad\quad t_{b,i} \leftarrow \mathsf{Trpdr}_K(w_{b,i})\\
&\quad \textit{let } \mathbf{t}_b = (t_{b,1}, \ldots, t_{b,q})\\
&\quad b' \leftarrow \mathcal{A}_2(st_{\mathcal{A}}, I_b, \mathbf{c}_b, \mathbf{t}_b)\\
&\quad \textit{if } b' = b, \textit{ output } 1\\
&\quad \textit{otherwise output } 0
\end{aligned}
$$

with the restriction that $\tau(H_0) = \tau(H_1)$, *and where* $st_{\mathcal{A}}$ *is a string that captures* $\mathcal{A}_1$*'s state. We say that* SSE *is secure in the sense of* non-adaptive indistinguishability *if for all polynomial-size adversaries* $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$,

$$\Pr\left[\,\mathbf{Ind}_{\mathsf{SSE},\mathcal{A}}(k) = 1\,\right] \le \frac{1}{2} + \mathsf{negl}(k),$$

where the probability is taken over the choice of b *and the coins of* Gen *and* Enc.

Definition 4.10 (Adaptive indistinguishability security for SSE). *Let* $\mathsf{SSE} = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Trpdr}, \mathsf{Search}, \mathsf{Dec})$ *be an index-based* SSE *scheme,* $k \in \mathbb{N}$ *be a security parameter,* $\mathcal{A} = (\mathcal{A}_0, \ldots, \mathcal{A}_{q+1})$ *be such that* $q \in \mathbb{N}$ *and consider the following probabilistic experiment* $\mathbf{Ind}^{\star}{}_{\mathcal{A},\mathsf{SSE}}(k)$*:*

$$
\begin{aligned}
&\mathbf{Ind}^{\star}{}_{\mathsf{SSE},\mathcal{A}}(k)\\
&\quad K \leftarrow \mathsf{Gen}(1^k)\\
&\quad b \xleftarrow{\$} \{0,1\}\\
&\quad (st_{\mathcal{A}}, \mathbf{D}_0, \mathbf{D}_1) \leftarrow \mathcal{A}_0(1^k)\\
&\quad (I_b, \mathbf{c}_b) \leftarrow \mathsf{Enc}_K(\mathbf{D}_b)\\
&\quad (st_{\mathcal{A}}, w_{0,1}, w_{1,1}) \leftarrow \mathcal{A}_1(st_{\mathcal{A}}, I_b)\\
&\quad t_{b,1} \leftarrow \mathsf{Trpdr}_K(w_{b,1})\\
&\quad \textit{for } 2 \le i \le q,\\
&\quad\quad (st_{\mathcal{A}}, w_{0,i}, w_{1,i}) \leftarrow \mathcal{A}_i(st_{\mathcal{A}}, I_b, \mathbf{c}_b, t_{b,1}, \ldots, t_{b,i-1})\\
&\quad\quad t_{b,i} \leftarrow \mathsf{Trpdr}_K(w_{b,i})\\
&\quad \textit{let } \mathbf{t}_b = (t_{b,1}, \ldots, t_{b,q})\\
&\quad b' \leftarrow \mathcal{A}_{q+1}(st_{\mathcal{A}}, I_b, \mathbf{c}_b, \mathbf{t}_b)\\
&\quad \textit{if } b' = b, \textit{ output } 1\\
&\quad \textit{otherwise output } 0
\end{aligned}
$$

with the restriction that $\tau(\mathbf{D}_0, w_{0,1}, \ldots, w_{0,q}) = \tau(\mathbf{D}_1, w_{1,1}, \ldots, w_{1,q})$ *and where* $st_{\mathcal{A}}$ *is a string that captures* $\mathcal{A}$*'s state. We say that* SSE *is secure in the sense of* adaptive indistinguishability *if for all polynomial-size adversaries* $\mathcal{A} = (\mathcal{A}_0, \ldots, \mathcal{A}_{q+1})$ *such that* $q = \mathsf{poly}(k)$,

$$\Pr\left[\,\mathbf{Ind}^{\star}{}_{\mathsf{SSE},\mathcal{A}}(k) = 1\,\right] \le \frac{1}{2} + \mathsf{negl}(k),$$

where the probability is over the choice of b, *and the coins of* Gen *and* Enc.

SSE with adaptive security

I: lookup table for k,j instead of k