DP5: Privacy-preserving Presence Protocols

Ian Goldberg

joint work with Nikita Borisov, George Danezis

Cryptography, Security, and Privacy Research Lab
University of Waterloo

Real World Cryptography
9 January 2015

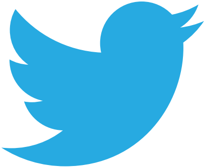
WATERLOO CHERITON SCHOOL OF COMPUTER SCIENCE

CrySP

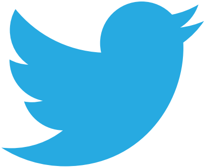
NSERC CRSNG

Social applications

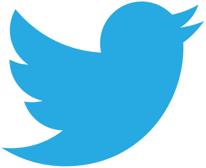
Social applications



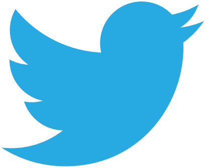
Social applications



Social applications

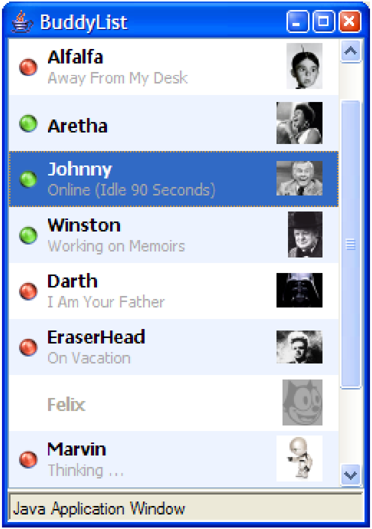


Social applications



Online presence

Online presence



BuddyList

Alfalfa
Away From My Desk

Aretha

Johnny
Online (Idle 90 Seconds)

Winston
Working on Memoirs

Darth
I Am Your Father

EraserHead
On Vacation

Felix

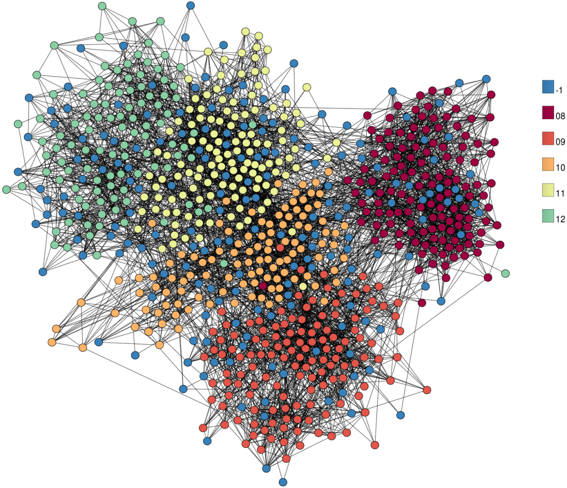
Marvin
Thinking ...

Java Application Window

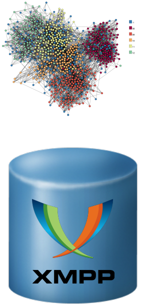
How it typically works

How it typically works

How it typically works

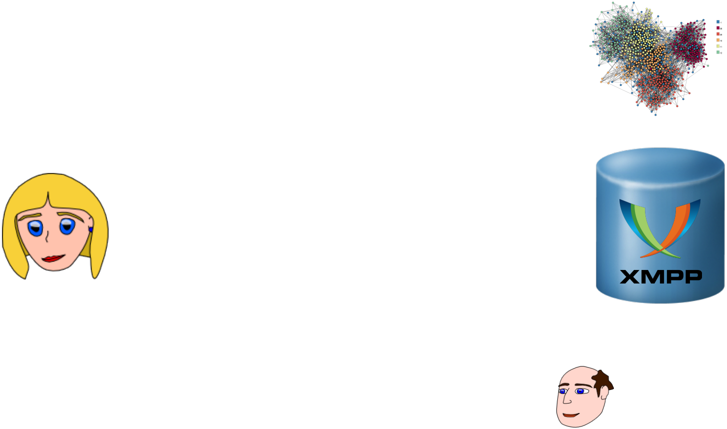


How it typically works

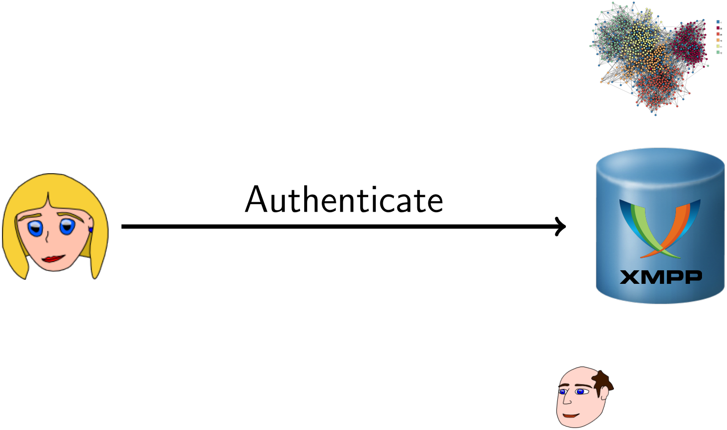


How it typically works

How it typically works

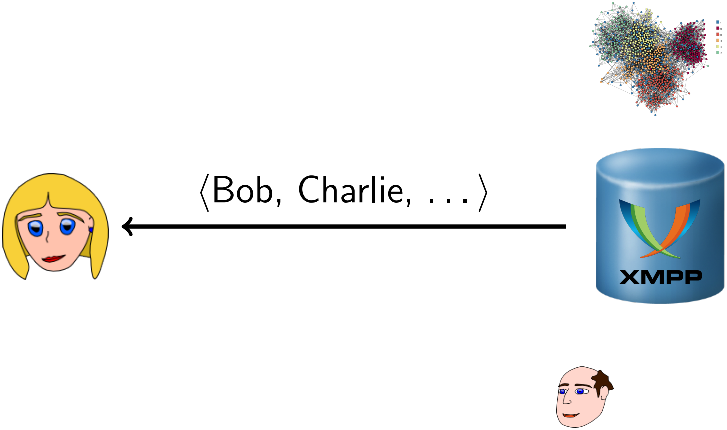


How it typically works



Authenticate

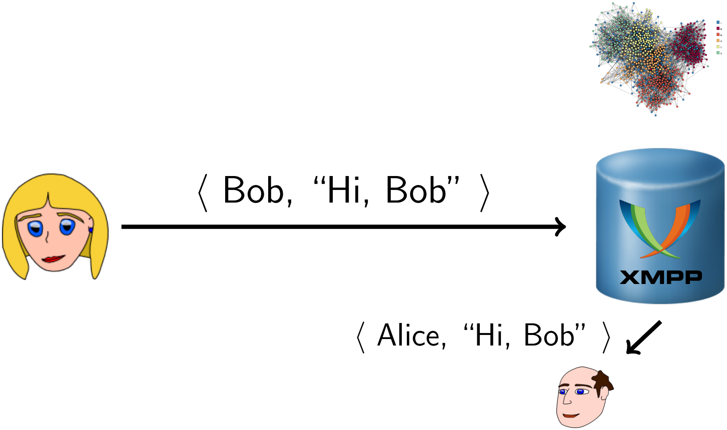
How it typically works



$\langle \text{Bob, Charlie, } \ldots \rangle$

How it typically works

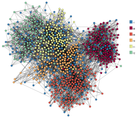
How it typically works



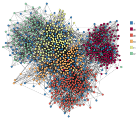
⟨ Bob, “Hi, Bob” ⟩

⟨ Alice, “Hi, Bob” ⟩

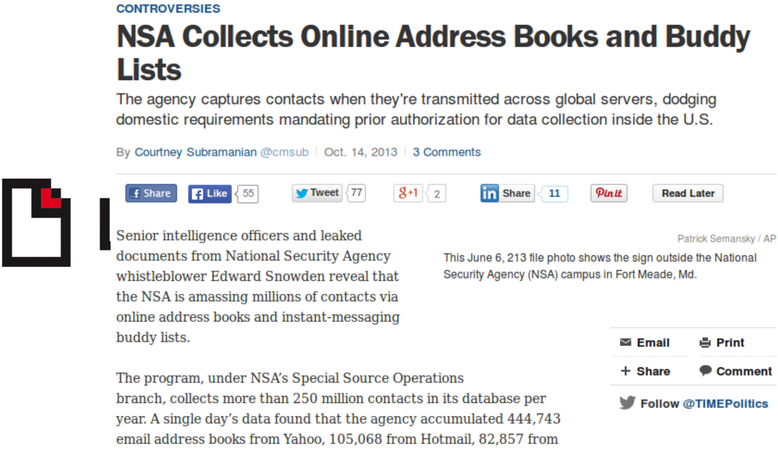
The problem



The problem



The problem



CONTROVERSIES

NSA Collects Online Address Books and Buddy Lists

The agency captures contacts when they're transmitted across global servers, dodging domestic requirements mandating prior authorization for data collection inside the U.S.

By Courtney Subramanian @cmsub | Oct. 14, 2013 | 3 Comments

Senior intelligence officers and leaked documents from National Security Agency whistleblower Edward Snowden reveal that the NSA is amassing millions of contacts via online address books and instant-messaging buddy lists.

Patrick Semansky / AP

This June 6, 213 file photo shows the sign outside the National Security Agency (NSA) campus in Fort Meade, Md.

The program, under NSA's Special Source Operations branch, collects more than 250 million contacts in its database per year. A single day's data found that the agency accumulated 444,743 email address books from Yahoo, 105,068 from Hotmail, 82,857 from

“We kill people based on metadata”

General Michael Hayden, former Director of NSA

http://www.youtube.com/watch?v=UdQiz0Vavmc

Want: private presence

Presence features

Threat model

Security goals

Want: private presence

Presence features

Threat model

Security goals

- Friend registration
- Presence registration
- Presence status query
- Friend suspension / revocation

Want: private presence

Presence features

Threat model

Security goals

- Secure end hosts
- Global passive adversary
- Dishonest users
- Threshold of honest infrastructure servers
- Can't break strong crypto

Want: private presence

Presence features

Threat model

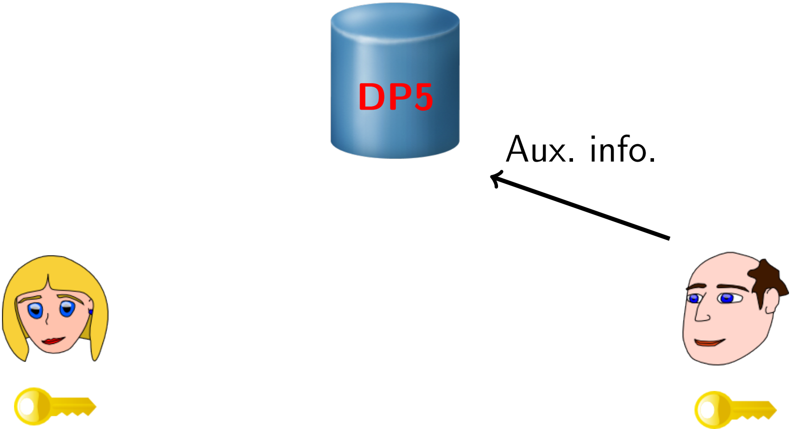
Security goals

- Privacy, integrity of presence and auxiliary data
- Privacy of social network
- Unlinkability
- Suspension / revocation indistinguishable from offline
- Forward and backward secrecy
- Auditability

Introducing DP5 (High level idea)

Introducing DP5 (High level idea)

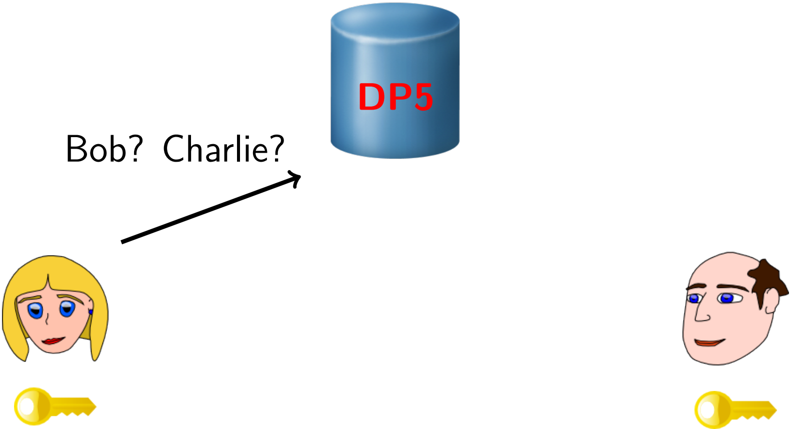
Introducing DP5 (High level idea)



DP5

Aux. info.

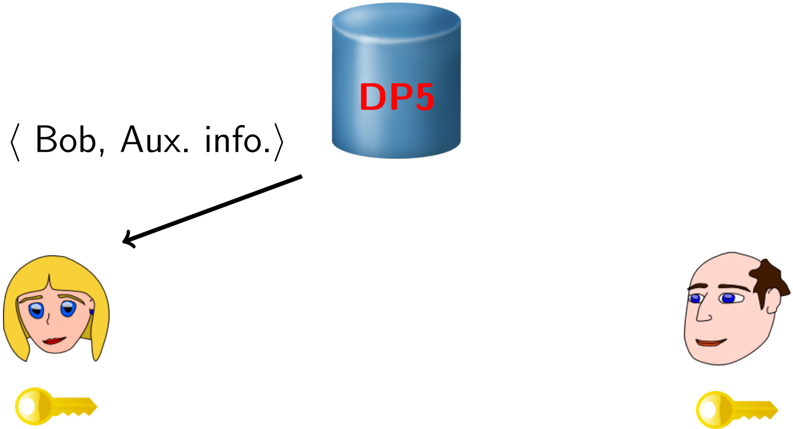
Introducing DP5 (High level idea)



DP5

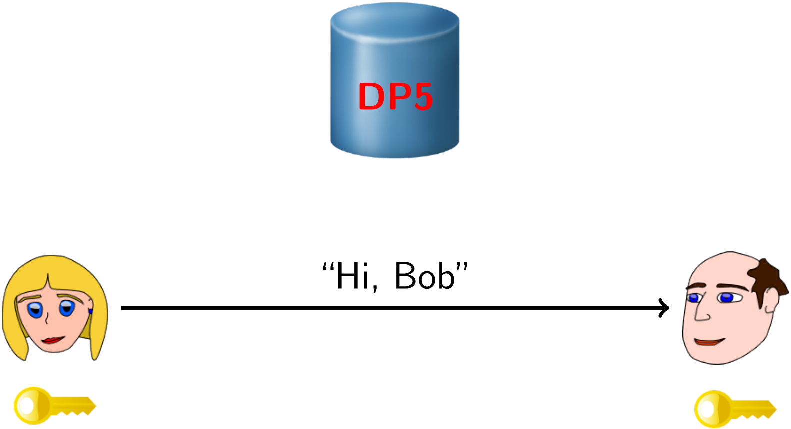
Bob? Charlie?

Introducing DP5 (High level idea)

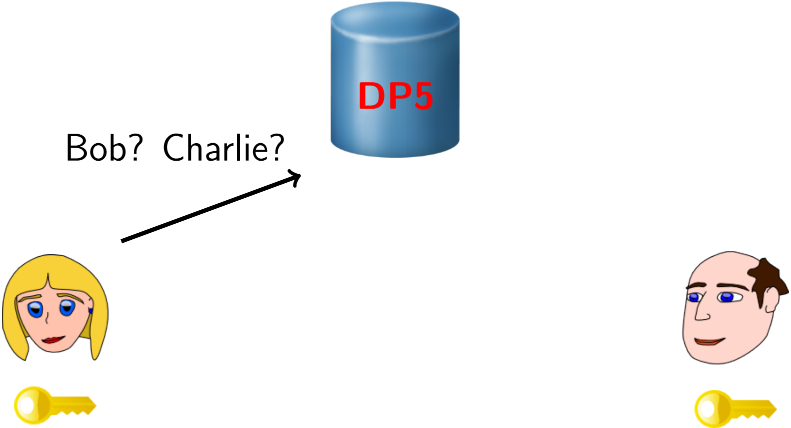


⟨ Bob, Aux. info.⟩

Introducing DP5 (High level idea)



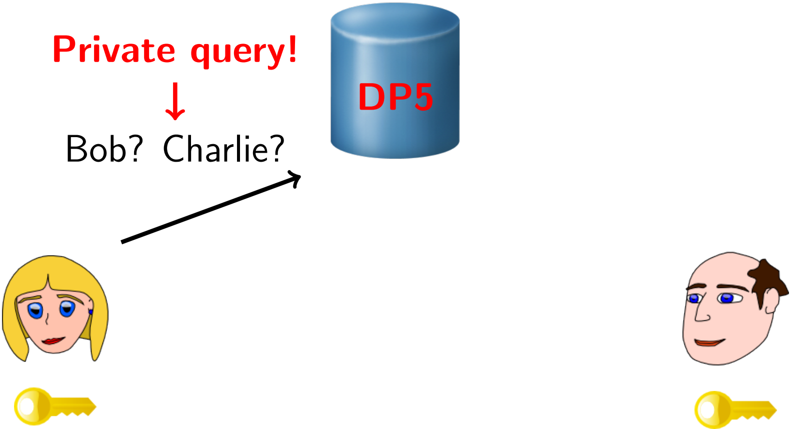
Introducing DP5 (High level idea)



DP5

Bob? Charlie?

Introducing DP5 (High level idea)



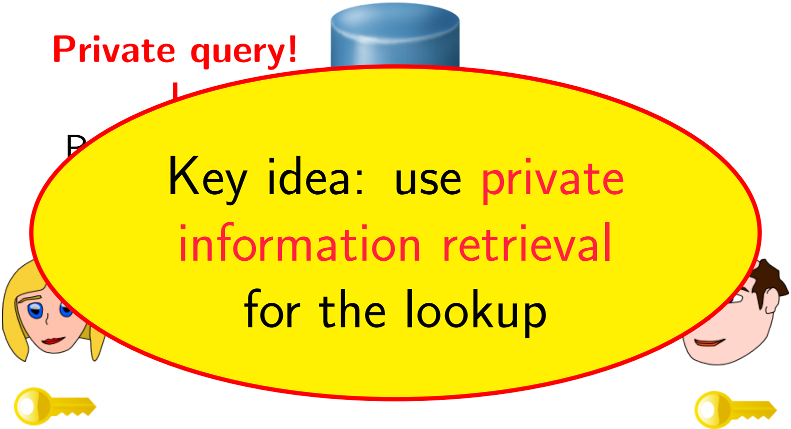
Private query!

↓

Bob? Charlie?

DP5

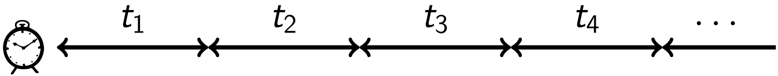
Introducing DP5 (High level idea)



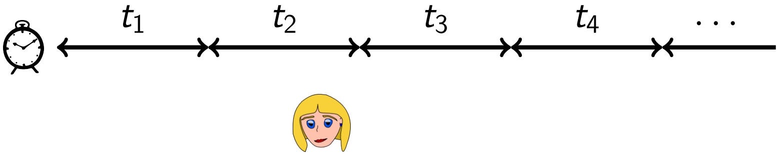
Key idea: use private information retrieval for the lookup

DP5: Strawman version

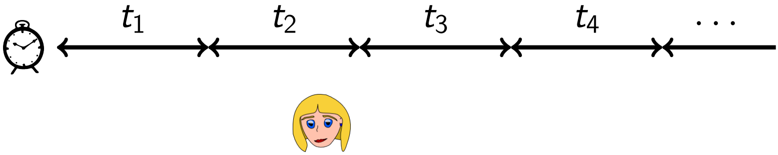
DP5: Strawman version



DP5: Strawman version

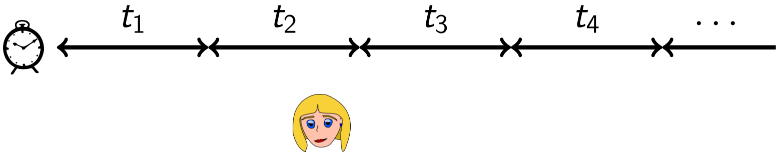


DP5: Strawman version



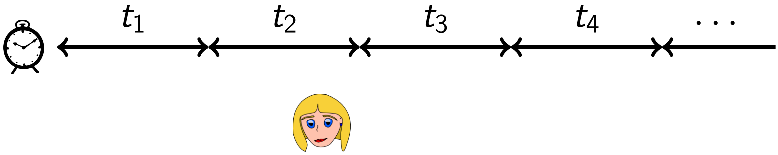
$$\mathrm{PRF}_{K_{ab}}(t_i)$$

DP5: Strawman version



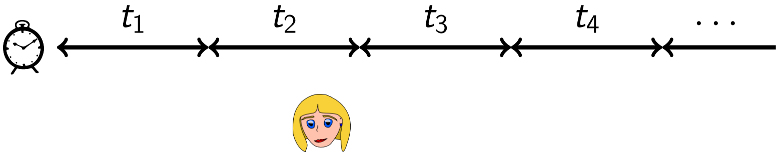
$$\begin{tikzcd}
 & K \\
\mathrm{PRF}_{K_{ab}}(t_i) \arrow[ur] \arrow[dr] & \\
 & ID
\end{tikzcd}$$

DP5: Strawman version



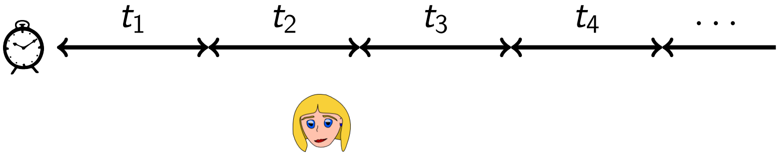
$$\begin{tikzcd}
 & \mathrm{AEAD}_{K}(aux) \\
\mathrm{PRF}_{K_{ab}}(t_i) \arrow[ur] \arrow[dr] & \\
 & \mathrm{ID}
\end{tikzcd}$$

DP5: Strawman version

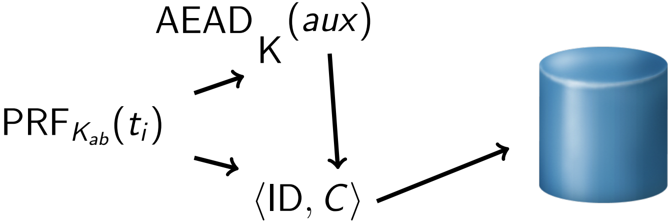


$$\begin{tikzcd}
 & \mathrm{AEAD}_{\mathrm{K}}(aux) \arrow[dd] \\
\mathrm{PRF}_{K_{ab}}(t_i) \arrow[ur] \arrow[dr] & \\
 & \langle \mathrm{ID}, C \rangle
\end{tikzcd}$$

DP5: Strawman version



t_1 t_2 t_3 t_4 . . .

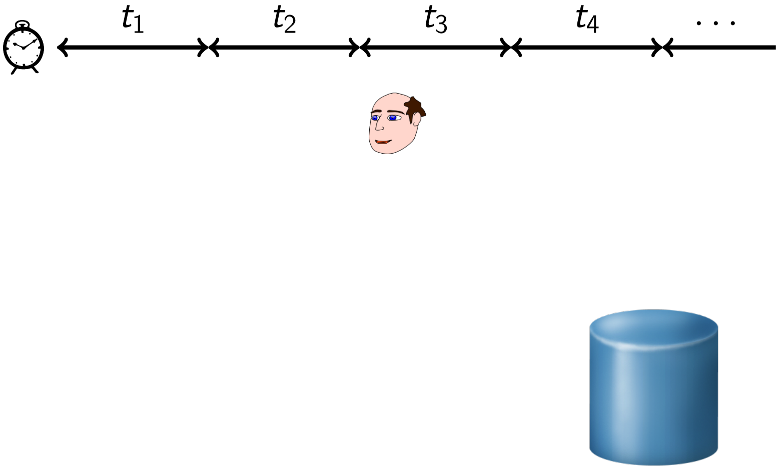


$\mathrm{AEAD}_{\mathrm{K}}(aux)$

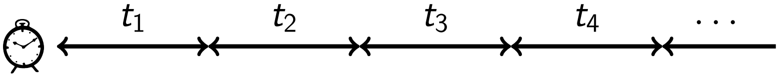
$\mathrm{PRF}_{K_{ab}}(t_i)$

$\langle \mathrm{ID}, C \rangle$

DP5: Strawman version

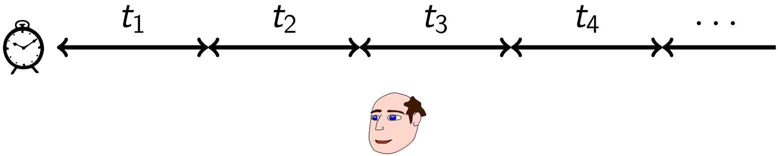


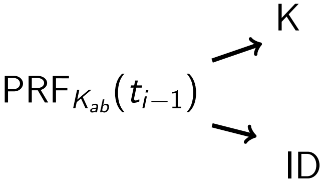
DP5: Strawman version



$\mathrm{PRF}_{K_{ab}}(t_{i-1})$

DP5: Strawman version

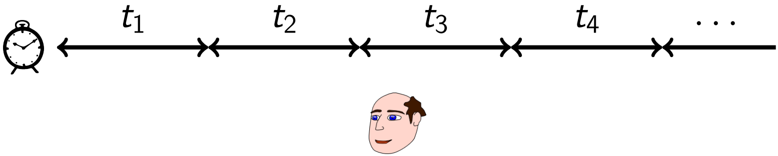




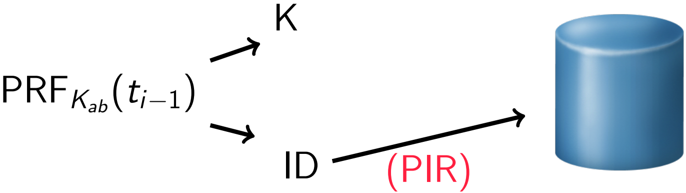
$$\mathrm{PRF}_{K_{ab}}(t_{i-1}) \to \mathrm{K}$$

$$\mathrm{PRF}_{K_{ab}}(t_{i-1}) \to \mathrm{ID}$$

DP5: Strawman version



t_1 t_2 t_3 t_4 $\ldots$



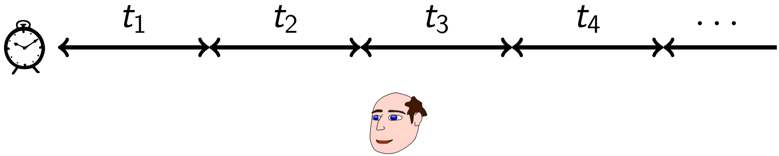
$\mathrm{PRF}_{K_{ab}}(t_{i-1})$

K

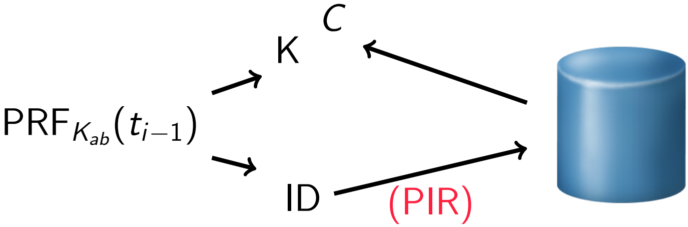
ID

(PIR)

DP5: Strawman version



t_1 t_2 t_3 t_4 $\ldots$



$\mathrm{PRF}_{K_{ab}}(t_{i-1})$

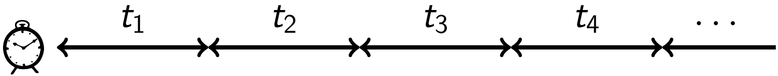
K

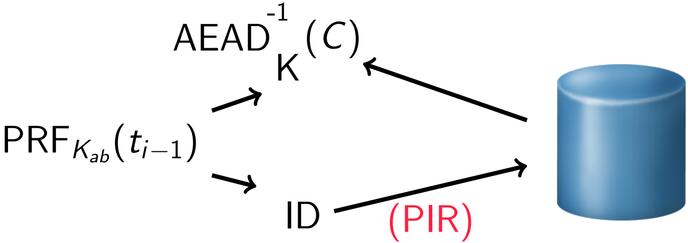
C

ID

(PIR)

DP5: Strawman version





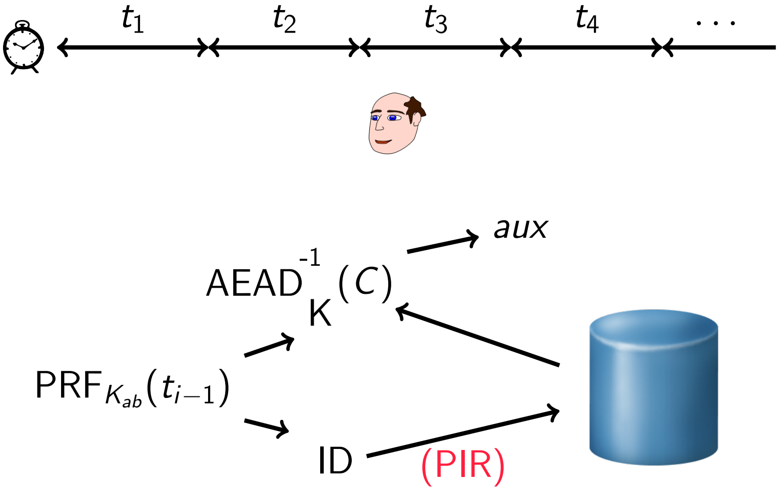
$\text{AEAD}^{\text{-1}}_{\text{K}}(C)$

$\text{PRF}_{K_{ab}}(t_{i-1})$

ID

(PIR)

DP5: Strawman version



t_1 t_2 t_3 t_4 $\ldots$

aux

$\mathrm{AEAD}^{-1}_{K}(C)$

$\mathrm{PRF}_{K_{ab}}(t_{i-1})$

ID

(PIR)

The problem of the large database

The problem of the large database

David Wheeler

The problem of the large database

David Wheeler

Any problem in computer science can be solved with another layer of indirection.

The problem of the large database

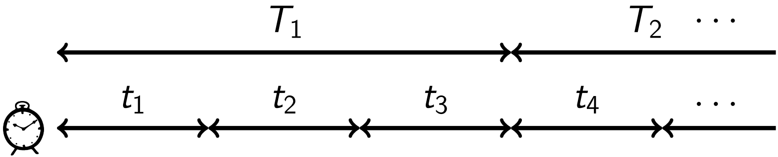
David Wheeler

Any problem in computer science can be solved with another layer of indirection.

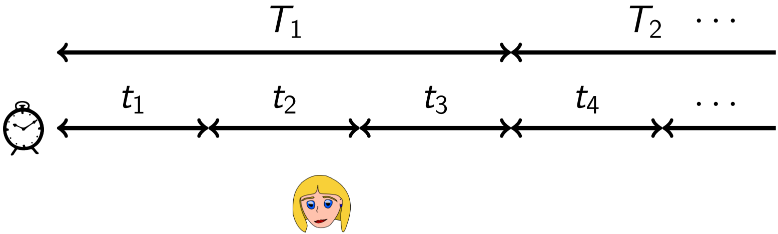
But that will usually create another problem.

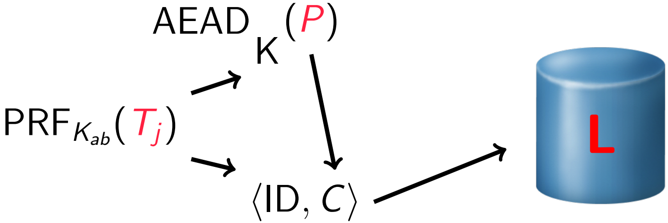
Two timescales, two databases

Two timescales, two databases

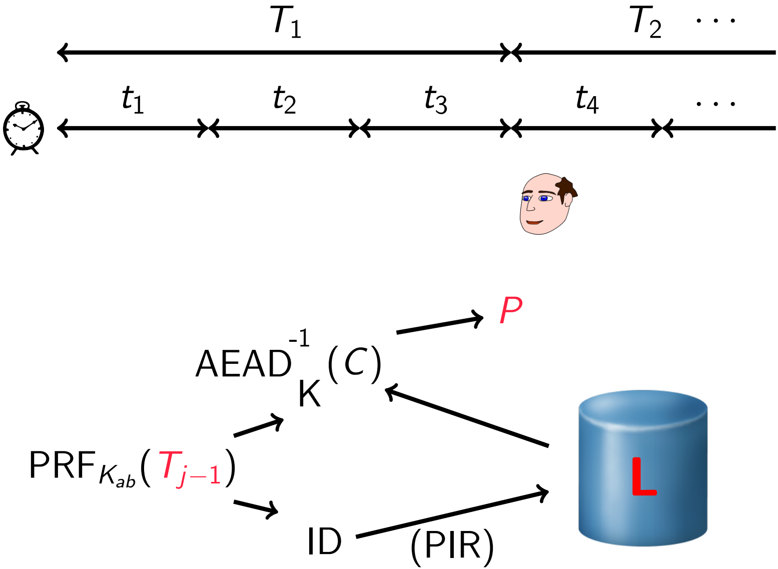


Two timescales, two databases




$\text{AEAD}_{\text{K}}(P)$
$\text{PRF}_{K_{ab}}(T_j)$
$\langle \text{ID}, C \rangle$
L

Two timescales, two databases



T_1 T_2 $\cdots$

t_1 t_2 t_3 t_4 $\cdots$

P

$\text{AEAD}^{-1}_{\text{K}}(C)$

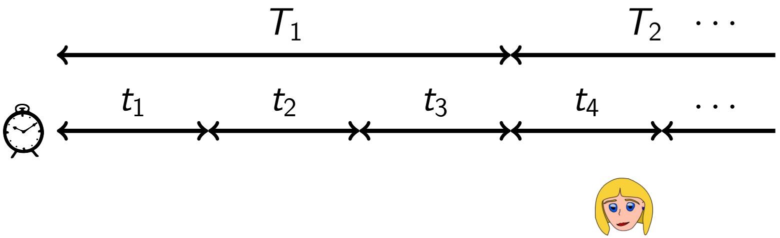
$\text{PRF}_{K_{ab}}(T_{j-1})$

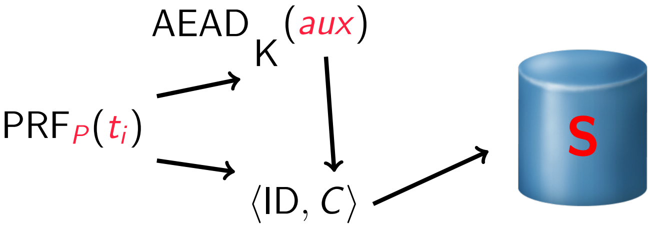
ID

(PIR)

L

Two timescales, two databases

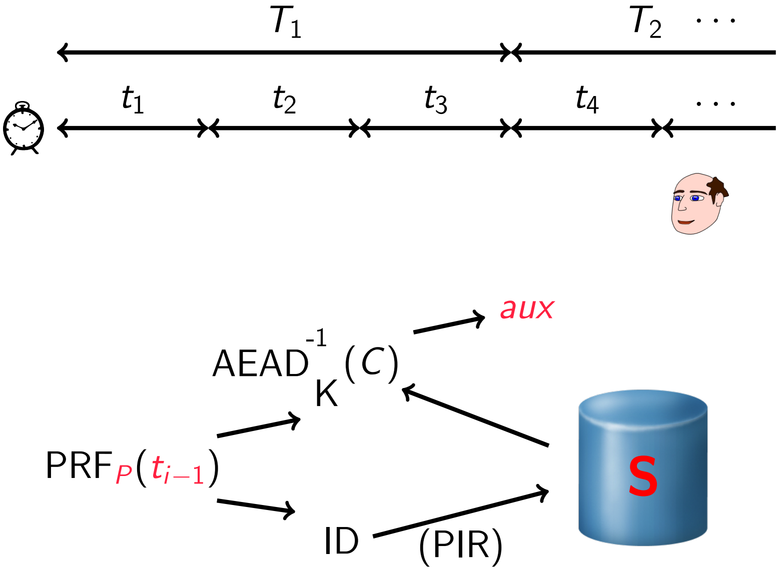



$\mathrm{AEAD}_{\mathrm{K}}(aux)$

$\mathrm{PRF}_P(t_i)$

$\langle \mathrm{ID}, C\rangle$

Two timescales, two databases



T_1 T_2 $\cdots$

t_1 t_2 t_3 t_4 $\cdots$

aux

$\mathrm{AEAD}^{-1}_{K}(C)$

$\mathrm{PRF}_P(t_{i-1})$

ID (PIR)

S

Implementation

PIR: Percy++ PIR library (C++)

DP5 core: C++

Networking: Cherrypy framework (Python)

git://git-crysp.uwaterloo.ca/percy
git://git-crysp.uwaterloo.ca/dp5

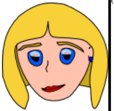
Takeaways

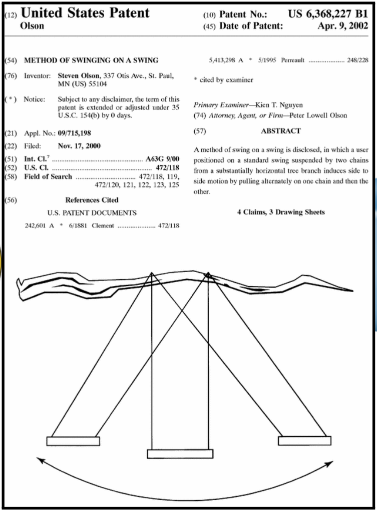
- Metadata in social communication is being targeted
- Private information retrieval (PIR) allows database lookups without revealing the query to the database servers themselves
- DP5 uses PIR to achieve private presence—people learn when their friends are online (and how to contact them securely) without any server ever learning who is friends with whom

Private information retrieval

Private information retrieval

Private information retrieval





(12) **United States Patent**
Olson

(10) **Patent No.: US 6,368,227 B1**
(45) **Date of Patent: Apr. 9, 2002**

(54) **METHOD OF SWINGING ON A SWING**

(76) Inventor: **Steven Olson**, 337 Otis Ave., St. Paul, MN (US) 55104

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/715,198**

(22) Filed: **Nov. 17, 2000**

(51) **Int. Cl.^{7}** **A63G 9/00**
(52) **U.S. Cl.** **472/118**
(58) **Field of Search** 472/118, 119, 472/120, 121, 122, 123, 125

(56) **References Cited**

U.S. PATENT DOCUMENTS

242,601 A * 6/1881 Clement 472/118
5,413,298 A * 5/1995 Perreault 248/228

* cited by examiner

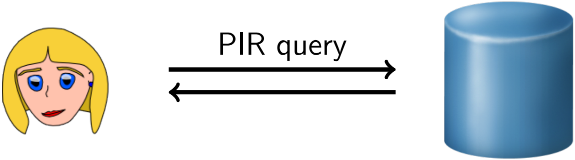
Primary Examiner—Kien T. Nguyen
(74) *Attorney, Agent, or Firm*—Peter Lowell Olson

(57) **ABSTRACT**

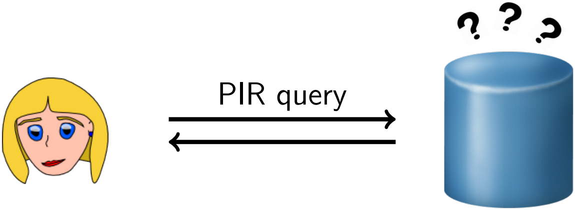
A method of swing on a swing is disclosed, in which a user positioned on a standard swing suspended by two chains from a substantially horizontal tree branch induces side to side motion by pulling alternately on one chain and then the other.

4 Claims, 3 Drawing Sheets

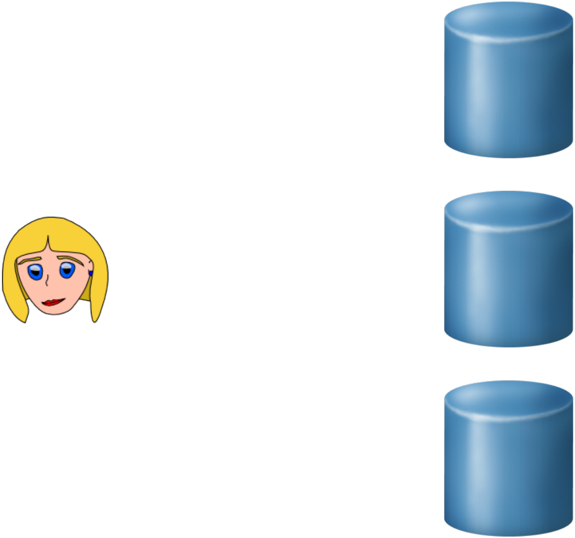
Private information retrieval



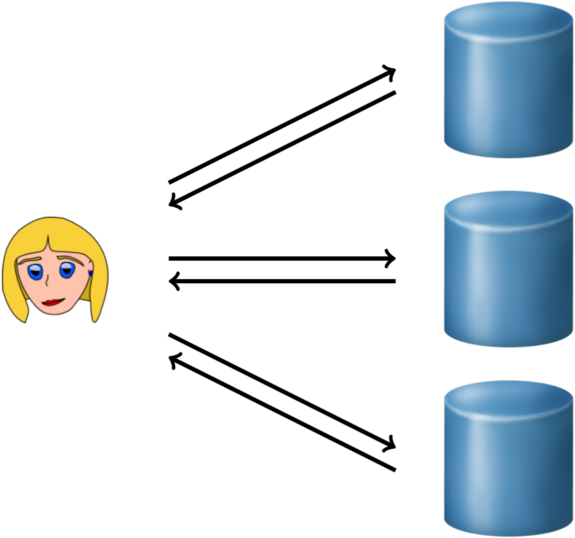
Private information retrieval



IT-PIR



IT-PIR



A simple PIR protocol

$$D = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & \dots & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & \dots & 0 \\ & & & & \vdots & & & \ddots & \vdots \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

A simple PIR protocol

$$D = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & \dots & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & \dots & 0 \\ & & & & \vdots & & & \ddots & \vdots \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

- If $\mathbf{e}_i = [0\ 0\ 1\ 0\ \ \dots\ \ 0]$, then $\mathbf{e}_i \cdot D = \text{Block } i$
- $\mathbf{v}_1 \cdot D + \mathbf{v}_2 \cdot D + \cdots + \mathbf{v}_\ell \cdot D = (\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_\ell) \cdot D$

A simple PIR protocol

$$D = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & \dots & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & \dots & 0 \\ & & & & \vdots & & & \ddots & \vdots \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

- If $\mathbf{e}_i = [0\ 0\ 1\ 0\ \dots\ 0]$, then $\mathbf{e}_i \cdot D =$ Block i
- $\mathbf{v}_1 \cdot D + \mathbf{v}_2 \cdot D + \cdots + \mathbf{v}_\ell \cdot D = (\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_\ell) \cdot D$

A simple PIR protocol

$$D = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & \dots & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & & & 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

Robustness issue!

- If $\mathbf{e}_i = [0\ 0\ 1\ 0\ \dots\ 0]$, then $\mathbf{e}_i \cdot D =$ Block i
- $\mathbf{v}_1 \cdot D + \mathbf{v}_2 \cdot D + \cdots + \mathbf{v}_\ell \cdot D = (\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_\ell) \cdot D$

A simple PIR protocol

$$D = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & \dots & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & \dots & 0 \\ & & & & \vdots & & & \ddots & \vdots \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

- If $\mathbf{e}_i = [0\ 0\ 1\ 0\ \dots\ 0]$, then $\mathbf{e}_i \cdot D =$ Block i
- $\mathbf{v}_1 \cdot D + \mathbf{v}_2 \cdot D + \cdots + \mathbf{v}_\ell \cdot D = (\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_\ell) \cdot D$

A simple PIR protocol

$$D = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & \dots & 0 \\ \color{red}{1} & \color{red}{1} & \color{red}{0} & \color{red}{0} & \color{red}{1} & \color{red}{1} & \color{red}{0} & \color{red}{\dots} & \color{red}{1} \\ 1 & 0 & & 0 & 1 & 1 & 0 & \dots & 0 \\ & & & & \vdots & & & \ddots & \vdots \\ 0 & 1 & 1 & & & 0 & 0 & \dots & 1 \end{bmatrix}$$

Previous work:
variable-sized records

- If $\mathbf{e}$
- $\mathbf{v}_1 \cdot D + \mathbf{v}_2$ $_\ell) \cdot D$

A simple PIR protocol

$$D = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & \dots & 1 \\ 1 & 0 & & 0 & 1 & 1 & 0 & \dots & 0 \\ & & & & \vdots & & & \ddots & \vdots \\ 0 & 1 & 1 & & & 0 & 0 & \dots & 1 \end{bmatrix}$$

Previous work:

~~variable-sized records~~

- If **e**
- $\mathbf{v}_1 \cdot D + \mathbf{v}_2$ $\ell) \cdot D$

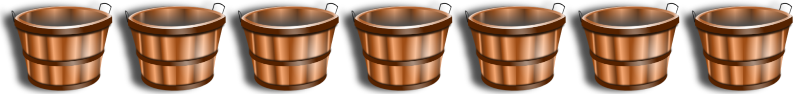
A simple PIR protocol

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 1 & \dots & 0 \\ & & & & & & & \vdots & \vdots \\ 0 & & 1 & 1 & 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

Previous work: lookups by keyword or SQL

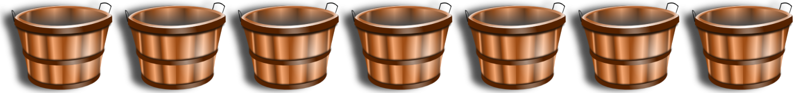
- If $\mathbf{e}_i = [0\ 0\ 1\ 0\ \dots\ 0]$, then $\mathbf{e}_i \cdot D =$ Block i
- $\mathbf{v}_1 \cdot D + \mathbf{v}_2 \cdot D + \cdots + \mathbf{v}_\ell \cdot D = (\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_\ell) \cdot D$

(Key,value) pair PIR lookups



(Key,value) pair PIR lookups

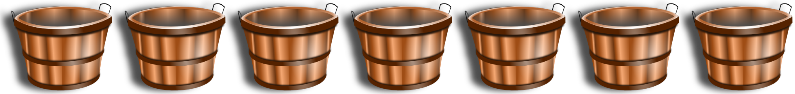
$(key_1, value_1)$
$(key_2, value_2)$
$(key_3, value_3)$
$\ldots$



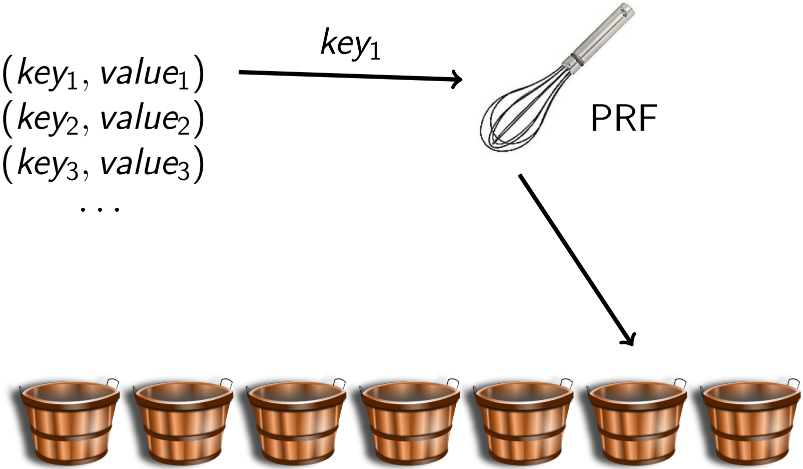
(Key,value) pair PIR lookups

$(key_1, value_1)$
$(key_2, value_2)$
$(key_3, value_3)$
$\ldots$

PRF

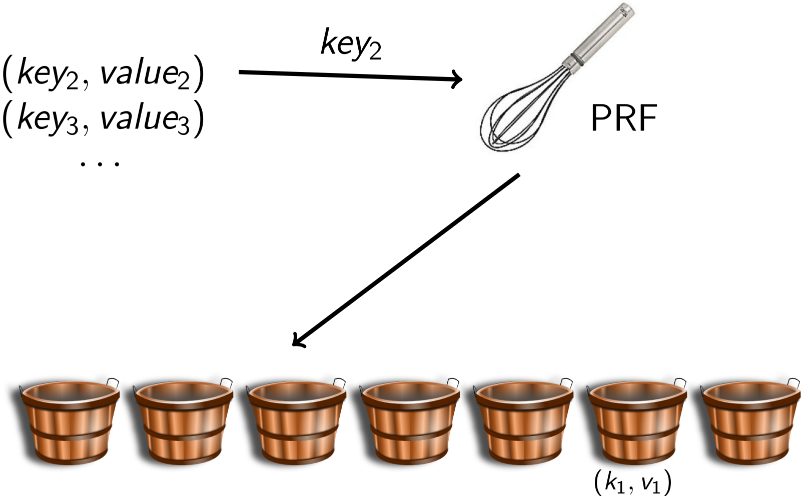


(Key,value) pair PIR lookups



$(key_1, value_1)$
$(key_2, value_2)$
$(key_3, value_3)$
$\ldots$

(Key,value) pair PIR lookups



$(key_2, value_2)$

$(key_3, value_3)$

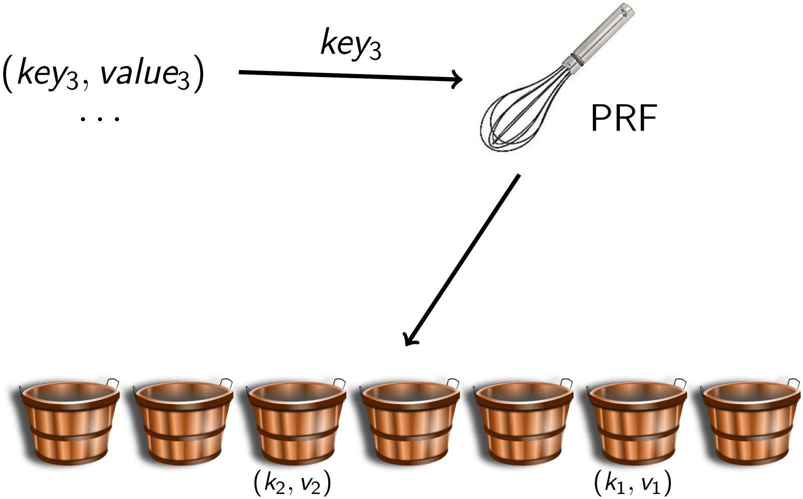
$\ldots$

key_2

PRF

(k_1, v_1)

(Key,value) pair PIR lookups



$(key_3, value_3)$

$\ldots$

key_3

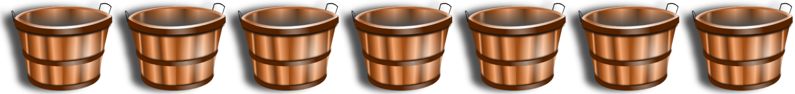
PRF

(k_2, v_2)

(k_1, v_1)

(Key,value) pair PIR lookups

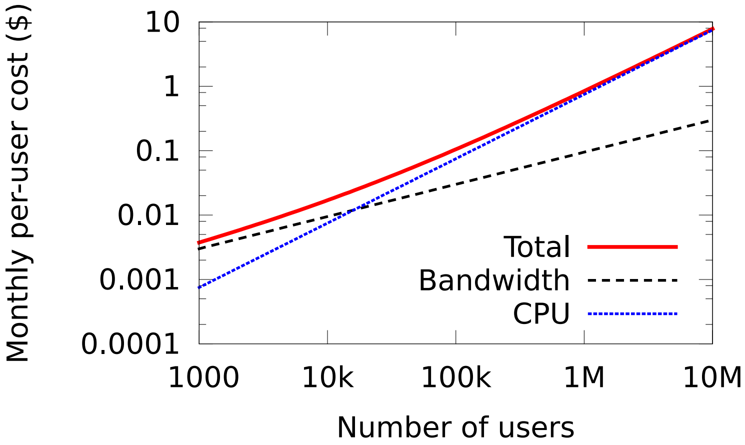
PRF



(k_6, v_6)	(k_{11}, v_{11})	(k_2, v_2)	(k_3, v_3)	(k_4, v_4)	(k_1, v_1)	(k_7, v_7)
(k_{10}, v_{10})	(k_{14}, v_{14})	(k_5, v_5)	(k_8, v_8)	(k_{12}, v_{12})	(k_{13}, v_{13})	(k_9, v_9)
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Cost of running a DP5 PIR server

(Long-term database, 24-hour epoch)


Monthly per-user cost ($)
10k
100k
1M
10M
Number of users
Total
Bandwidth
CPU