“Efficient”, “Expressive”, “Private” Information Retrieval from Indexes of Queries

Ryan Henry
Research talk @ Boise State
2017-11-16
Information Retrieval

REQUEST

RESPONSE

Ryan Henry
Information Retrieval

(For Justin Bieber videos)

REQUEST

RESPONSE
Anonymous Information Retrieval

Embarrassing Bieber videos, please!

Who just searched for Bieber?
Private Information Retrieval (PIR)

(for super-secret patent idea)

REQUEST

RESPONSE

What did Alice just fetch?

Ryan Henry
**Defn:** PIR

**Private information retrieval**

IPA: /praɪvət ɪnfərˈmeɪʃn ɪˈtrɪvəl/

**Noun.** A database technique that uses ideas from cryptography to protect the privacy of users by hiding the details of their queries from the operators of the database.

**Initialism:** PIR

sometimes treated as an acronym; pronounced "peer" or "purr"
“the 5 cafés nearest to Boise State”
"the best Italian in Manhattan"
“the most recent episode of Daredevil”
“the 3 most popular songs by P!nk”

1. Rockstar
   Post Malone Featuring 21 Savage
   Last Week: 1

2. Bodak Yellow (Money Moves)
   Cardi B
   Last Week: 2

3. 1-800-273-8255
   Logic Featuring Alessia Cara & Khalid
   Last Week: 3

4. Feel It Still
   Portugal. The Man
   Last Week: 4

5. Thunder
   Imagine Dragons
   Last Week: 5

6. Sorry Not Sorry
   Demi Lovato
   Last Week: 6
Justin Trudeau • @JustinTrudeau • Oct 11
Today, I conveyed Canada's sympathies to President Trump regarding the devastating fires in California. We stand in support of our friends.

Barack Obama • @BarackObama • Sep 11
We remember everyone we lost on 9/11 and honor all who defend our country and our ideals. No act of terror will ever change who we are.

Metallica • @Metallica • Sep 18
Switzerland! Additional tickets have been released for next April's show in Geneva! ticketcomer.ch/tickets.html??

Adele • @Adele • Feb 22
Thank you @BRETS, I actually am on the other side of the world! I'm in Australia, that's why I'm not there. Have a great night x

Barack Obama • @BarackObama • Oct 2
Michelle & I are praying for the victims in Las Vegas. Our thoughts are with their families & everyone enduring another senseless tragedy.

Cancer Support Comm Retweeted
BreakawayfromCancer • @BreakawayCancer • Oct 16
You are stronger than you know... More brave than you think... More loved than you can imagine. #meditationmonday

Rafa Nadal • @RafaelNadal • Oct 10
On my way to practice. Court 5 #Shanghai #Tennis #practice pics on my IG account later :)
"Recent news for #LGBTQ"
“Latest tweets by @NRA”
Trending tweets for #MAGA

"Thank you @BTS. I actually am on the other side of the world! I'm in Australia, that's why I'm not there. Have a great night x"

"Today, I conveyed Canada's sympathies to President Trump regarding the devastating fires in California. We stand in support of our friends."

"We remember everyone we lost on 9/11 and honor all who defend our country and our ideals. No act of terror will ever change who we are."

"Additional tickets have been released for next April's show in Switzerland! ticketscomer.ch/tickets.html?..."

"Michelle & I are praying for the victims in Las Vegas. Our thoughts are with their families & everyone enduring another senseless tragedy."

"You are stronger than you know... More brave than you think... More loved than you can imagine. #meditationmonday"

"On my way to practice. Court 5 #ShanghaiTennisPractice pics on my IG account later ;-)"

"Chavez HS bringing awareness about gender norms. Wear whatever you want to express yourself! #ROSA #LGBTRights #RevelsLove #loboHOCO2017"
“Most retweeted from @NORML”
Most liked tweets of @CancerSupportHQ
I'm sold! Private queries for these applications sounds great!!

But, like...

HOW!?
- Database server sends *everything* to client

😊 Privacy protection is perfect...but

😢 Communication overhead is \{\Huge...\}
The database as a matrix

- Total of $r$ rows
- Each row holds one $s$-word "block" of data
- Each word is an element of some finite field $\mathbb{F}$
The database as a matrix

The request as a vector

Bright idea: (non-private) requests are just a basis vector from the "standard orthonormal basis" of $F^n$
The database as a matrix

The request as a vector

\[ \begin{bmatrix} 0 & 0 & \cdots & 1 & \cdots & 0 \end{bmatrix} \cdot \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} & \cdots & x_{1s} \\ x_{21} & x_{22} & x_{23} & x_{24} & \cdots & x_{2s} \\ x_{31} & x_{32} & x_{33} & x_{34} & \cdots & x_{3s} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{r1} & x_{r2} & x_{r3} & x_{r4} & \cdots & x_{rs} \end{bmatrix} \]

\[ \hat{e}_j \cdot X = \begin{bmatrix} 0 \cdot x_{11} + 0 \cdot x_{21} + \cdots + 1 \cdot x_{j1} + \cdots + 0 \cdot x_{r1} \\ 0 \cdot x_{12} + 0 \cdot x_{22} + \cdots + 1 \cdot x_{j2} + \cdots + 0 \cdot x_{r2} \\ 0 \cdot x_{13} + 0 \cdot x_{23} + \cdots + 1 \cdot x_{j3} + \cdots + 0 \cdot x_{r3} \\ \vdots & \vdots & \vdots \vdots & \ddots & \vdots \\ 0 \cdot x_{1s} + 0 \cdot x_{2s} + \cdots + 1 \cdot x_{js} + \cdots + 0 \cdot x_{rs} \end{bmatrix}^T \]

\[ \in \mathbb{F}^{r \times s} \]

jth position

(transposed)
"Vector-matrix" PIR

• (Non-private) query for row $j$ is a vector; namely,
  $$\vec{e}_j := \langle 0 \ldots 1 \ldots 0 \rangle$$

Observation 1: Encrypting $\vec{e}_j$ component-wise using IND-CPA secure, additively homomorphic encryption gives computationally private PIR!

Observation 2: Sharing $\vec{e}_j$ component-wise using some linear, $(t, \ell)$-threshold secret sharing gives a information-theoretically private PIR!
Goldberg'07: Robust IT-PIR

- Database is an $r \times s$ matrix over $\mathbb{F}$; record is a single row.
- (Non-)private query via vector-matrix product with a standard basis vector:

$$\langle 0 \ 0 \ \cdots \ 1 \ \cdots \ 0 \rangle \cdot X = \vec{X}_j$$

Idea: Split query vector component-wise with Shamir's $(t+1, \ell)$-threshold secret sharing in $\mathbb{F}$.

Improving the Robustness of Private Information Retrieval
Ian Goldberg
replicate the database to $\ell$ servers...

...and secret share the request!

$e_j = \langle 0 \ 0 \ldots \ 1 \ldots \ 0 \rangle$
$\bar{e}_j = \langle 0 \ 0 \ \ldots \ 1 \ \ldots \ 0 \rangle$

$\text{Share}_1(\bar{e}_j)$

$\text{Share}_2(\bar{e}_j)$

$\text{Share}_3(\bar{e}_j)$

$\vdots$

$\vdots$

$\text{Share}_n(\bar{e}_j)$
Shamir secret sharing

2-out-of-$l$

Random slope
+ 1 point
⇒ all secrets are equally likely!

Any 2 points
⇒ only 1 secret is possible!
Shamir secret sharing

2-out-of-\(l\)

secret

share \((x_2, y_2)\)

share \((x_3, y_3)\)
Shamir secret sharing

4-out-of-\( \ell \)

cubic polynomial

secret
Shamir’s \((t+1, \ell)\)-secret sharing

**Goal:** Share a secret \(S\) among \(\ell\) shareholders so that

1. Any subset of \(k \geq t+1\) shareholders can recover \(S\)
2. No subset of \(k \leq t\) shareholders learns anything about \(S\)

- Choose \(f \in \mathbb{F}[x]\) subject to \(f(0) = S\) and \(\deg f \leq t\)
- Shareholder \(i\) receives the share \((x_i, f(x_i))\)

**Fact:** Any \(k \geq t+1\) shareholders can compute \(f(0)\) via polynomial interpolation:

\[
f(0) = \sum_{i=1}^{k} f(x_i) \cdot \prod_{j=1, j \neq i}^{k} \frac{x_j}{(x_j - x_i)^{-1}}
\]
Goldberg's IT-PIR

\[ X_j = \sum_{k=1}^{\ell} \text{Shamir}_k(\tilde{e}_j) \cdot \prod_{i=1 \neq k} x_i / (x_i - x_k)^{-1} \]
“Most recent tweet of @BarackObama”

Shamir_k(ψ₁,ψ₂,ψ₃,ψ₄,ψ₅,ψ₆,ψ₇,ψ₈,ψ₉,ψ₁₀)
PIR-by-keywords

Private Information Retrieval by Keywords
Benny Chor, Niv Gilboa, and Moni Naor
Technion-Israel Institute of Technology
Tech. Rep. CS 0917
PIR-by-keywords

- a-c
- e-g
- g-i
- a-i
- i-o
- o-t
- t-z
- a-o
- o-z
- i-k
- k-m
- m-n
- n-o
- o-q
- q-t
- t-w
- w-z

Ryan Henry
Replicate & reorganize data into many database "views"
<table>
<thead>
<tr>
<th>tweet</th>
<th>username</th>
<th>date</th>
<th>time</th>
<th>❤️</th>
<th>Retweets</th>
<th>Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thanks for...</td>
<td>@Barack</td>
<td>2017-11-05</td>
<td>14:45:13Z</td>
<td>903</td>
<td>87</td>
<td>#worstgiftever</td>
</tr>
<tr>
<td>Playing golf...</td>
<td>@Donald</td>
<td>2017-11-05</td>
<td>03:27:12Z</td>
<td>112</td>
<td>22</td>
<td>#MAGA, #borderwall</td>
</tr>
<tr>
<td>I could use...</td>
<td>@Alice</td>
<td>2017-11-05</td>
<td>14:00:10Z</td>
<td>6</td>
<td>0</td>
<td>#covfefe</td>
</tr>
<tr>
<td>Fake MSNBC...</td>
<td>@Donald</td>
<td>2017-11-06</td>
<td>04:01:58Z</td>
<td>87</td>
<td>34</td>
<td>#USA, #fakenews</td>
</tr>
<tr>
<td>Ryan Henry...</td>
<td>@Carol</td>
<td>2017-11-06</td>
<td>18:00:13Z</td>
<td>51</td>
<td>4</td>
<td>#UIUC</td>
</tr>
<tr>
<td>Puerto Rico...</td>
<td>@Barack</td>
<td>2017-11-06</td>
<td>20:16:02Z</td>
<td>151</td>
<td>101</td>
<td>#hurricane</td>
</tr>
</tbody>
</table>

Replicate & reorganize data into many database "views"
<table>
<thead>
<tr>
<th>tweet</th>
<th>username</th>
<th>date</th>
<th>time</th>
<th>❤️</th>
<th>❯</th>
<th>#tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thanks for...</td>
<td>@Barack</td>
<td>2017-11-05</td>
<td>14:45:13Z</td>
<td>903</td>
<td>87</td>
<td>#worstgiftever</td>
</tr>
<tr>
<td>Playing golf...</td>
<td>@Donald</td>
<td>2017-11-05</td>
<td>03:27:12Z</td>
<td>112</td>
<td>22</td>
<td>#MAGA, #borderwall</td>
</tr>
<tr>
<td>I could use...</td>
<td>@Alice</td>
<td>2017-11-05</td>
<td>14:00:10Z</td>
<td>6</td>
<td>0</td>
<td>#covfefe</td>
</tr>
<tr>
<td>Fake MSNBC...</td>
<td>@Donald</td>
<td>2017-11-06</td>
<td>04:01:58Z</td>
<td>87</td>
<td>34</td>
<td>#USA, #fakenews</td>
</tr>
<tr>
<td>Ryan Henry...</td>
<td>@Carol</td>
<td>2017-11-06</td>
<td>18:00:13Z</td>
<td>51</td>
<td>4</td>
<td>#UIUC</td>
</tr>
<tr>
<td>Puerto Rico...</td>
<td>@Barack</td>
<td>2017-11-06</td>
<td>20:16:02Z</td>
<td>151</td>
<td>101</td>
<td>#hurricane</td>
</tr>
<tr>
<td>tweet</td>
<td>username</td>
<td>date</td>
<td>time</td>
<td>💖</td>
<td>⏱️</td>
<td>#tags</td>
</tr>
<tr>
<td>----------------</td>
<td>----------</td>
<td>------------</td>
<td>------------</td>
<td>----</td>
<td>----</td>
<td>------------------------------</td>
</tr>
<tr>
<td>Thanks for...</td>
<td>@Barack</td>
<td>2017-11-05</td>
<td>14:45:13Z</td>
<td>903</td>
<td>87</td>
<td>#worstgiftever</td>
</tr>
<tr>
<td>Playing golf...</td>
<td>@Donald</td>
<td>2017-11-05</td>
<td>03:27:12Z</td>
<td>112</td>
<td>22</td>
<td>#MAGA, #borderwall</td>
</tr>
<tr>
<td>I could use...</td>
<td>@Alice</td>
<td>2017-11-05</td>
<td>14:00:10Z</td>
<td>6</td>
<td>0</td>
<td>#covfefe</td>
</tr>
<tr>
<td>Fake MSNBC...</td>
<td>@Donald</td>
<td>2017-11-06</td>
<td>04:01:58Z</td>
<td>87</td>
<td>34</td>
<td>#USA, #fakenews</td>
</tr>
<tr>
<td>Puerto Rico...</td>
<td>@Barack</td>
<td>2017-11-06</td>
<td>18:00:13Z</td>
<td>51</td>
<td>4</td>
<td>#UIUC</td>
</tr>
<tr>
<td>Ryan Henry...</td>
<td>@Carol</td>
<td>2017-11-06</td>
<td>20:16:02Z</td>
<td>151</td>
<td>101</td>
<td>#hurricane</td>
</tr>
<tr>
<td>I could use...</td>
<td>@Alice</td>
<td>2017-11-05</td>
<td>14:00:10Z</td>
<td>6</td>
<td>0</td>
<td>#covfefe</td>
</tr>
<tr>
<td>Thanks for...</td>
<td>@Barack</td>
<td>2017-11-05</td>
<td>14:45:13Z</td>
<td>903</td>
<td>87</td>
<td>#worstgiftever</td>
</tr>
<tr>
<td>Ryan Henry...</td>
<td>@Carol</td>
<td>2017-11-06</td>
<td>18:00:13Z</td>
<td>51</td>
<td>4</td>
<td>#UIUC</td>
</tr>
<tr>
<td>Playing golf...</td>
<td>@Donald</td>
<td>2017-11-05</td>
<td>03:27:12Z</td>
<td>112</td>
<td>22</td>
<td>#MAGA, #borderwall</td>
</tr>
<tr>
<td>tweet</td>
<td>username</td>
<td>date</td>
<td>time</td>
<td>likes</td>
<td>Retweets</td>
<td>#tags</td>
</tr>
<tr>
<td>----------------------------</td>
<td>----------</td>
<td>-----------</td>
<td>--------------</td>
<td>-------</td>
<td>----------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Thanks for…</td>
<td>@Barack</td>
<td>2017-11-05</td>
<td>14:45:13Z</td>
<td>903</td>
<td>87</td>
<td>#worstgiftever</td>
</tr>
<tr>
<td>Playing golf…</td>
<td>@Donald</td>
<td>2017-11-05</td>
<td>03:27:12Z</td>
<td>112</td>
<td>22</td>
<td>#MAGA, #borderwall</td>
</tr>
<tr>
<td>I could use…</td>
<td>@Alice</td>
<td>2017-11-05</td>
<td>14:00:10Z</td>
<td>6</td>
<td>0</td>
<td>#covfefe</td>
</tr>
<tr>
<td>Fake MSNBC…</td>
<td>@Donald</td>
<td>2017-11-06</td>
<td>04:01:58Z</td>
<td>87</td>
<td>34</td>
<td>#USA, #fakenews</td>
</tr>
<tr>
<td>Puerto Rico…</td>
<td>@Barack</td>
<td>2017-11-06</td>
<td>20:16:02Z</td>
<td>151</td>
<td>101</td>
<td>#hurricane</td>
</tr>
<tr>
<td>Puerto Rico…</td>
<td>@Barack</td>
<td>2017-11-06</td>
<td>20:16:02Z</td>
<td>151</td>
<td>101</td>
<td>#hurricane</td>
</tr>
<tr>
<td>Ryan Henry…</td>
<td>@Carol</td>
<td>2017-11-05</td>
<td>14:00:10Z</td>
<td>6</td>
<td>0</td>
<td>#covfefe</td>
</tr>
<tr>
<td>Fake MSNBC…</td>
<td>@Donald</td>
<td>2017-11-06</td>
<td>04:01:58Z</td>
<td>87</td>
<td>34</td>
<td>#USA, #fakenews</td>
</tr>
<tr>
<td>I could use…</td>
<td>@Alice</td>
<td>2017-11-05</td>
<td>14:00:10Z</td>
<td>6</td>
<td>0</td>
<td>#covfefe</td>
</tr>
<tr>
<td>Puerto Rico…</td>
<td>@Barack</td>
<td>2017-11-06</td>
<td>20:16:02Z</td>
<td>151</td>
<td>101</td>
<td>#hurricane</td>
</tr>
<tr>
<td>Ryan Henry…</td>
<td>@Carol</td>
<td>2017-11-06</td>
<td>18:00:13Z</td>
<td>51</td>
<td>4</td>
<td>#UIUC</td>
</tr>
<tr>
<td>tweet</td>
<td>username</td>
<td>date</td>
<td>time</td>
<td>like</td>
<td>retweet</td>
<td>#tags</td>
</tr>
<tr>
<td>--------------------------</td>
<td>----------</td>
<td>----------</td>
<td>---------------</td>
<td>------</td>
<td>---------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>Thanks for...</td>
<td>@Barack</td>
<td>2017-11-05</td>
<td>14:45:13Z</td>
<td>903</td>
<td>87</td>
<td>#worstgiftever</td>
</tr>
<tr>
<td>Playing golf...</td>
<td>@Donald</td>
<td>2017-11-05</td>
<td>03:27:12Z</td>
<td>112</td>
<td>22</td>
<td>#MAGA, #borderwall</td>
</tr>
<tr>
<td>I could use...</td>
<td>@Alice</td>
<td>2017-11-05</td>
<td>14:00:10Z</td>
<td>6</td>
<td>0</td>
<td>#covfefe</td>
</tr>
<tr>
<td>Fake MSNBC...</td>
<td>@Donald</td>
<td>2017-11-06</td>
<td>04:01:58Z</td>
<td>87</td>
<td>34</td>
<td>#USA, #fakenews</td>
</tr>
<tr>
<td>Ryan Henry...</td>
<td>@Carol</td>
<td>2017-11-06</td>
<td>18:00:13Z</td>
<td>51</td>
<td>4</td>
<td>#UIUC</td>
</tr>
<tr>
<td>Puerto Rico...</td>
<td>@Barack</td>
<td>2017-11-06</td>
<td>20:16:02Z</td>
<td>151</td>
<td>101</td>
<td>#hurricane</td>
</tr>
</tbody>
</table>

\[
\pi_{\heartsuit} \cdot X = X_{\heartsuit} = \pi_{\heartsuit} \cdot X \\
\pi_{\#} \cdot X = X_{\#} = \pi_{\#} \cdot X
\]
<table>
<thead>
<tr>
<th>tweet</th>
<th>username</th>
<th>date</th>
<th>time</th>
<th>likes</th>
<th>retweets</th>
<th>#tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thanks for…</td>
<td>@Barack</td>
<td>2017-11-05</td>
<td>14:45:13Z</td>
<td>903</td>
<td>87</td>
<td>#worstgiftever</td>
</tr>
<tr>
<td>Playing golf…</td>
<td>@Donald</td>
<td>2017-11-05</td>
<td>03:27:12Z</td>
<td>112</td>
<td>22</td>
<td>#MAGA, #borderwall</td>
</tr>
<tr>
<td>I could use…</td>
<td>@Alice</td>
<td>2017-11-05</td>
<td>14:00:10Z</td>
<td>6</td>
<td>0</td>
<td>#covfefe</td>
</tr>
<tr>
<td>Fake MSNBC…</td>
<td>@Donald</td>
<td>2017-11-06</td>
<td>04:01:58Z</td>
<td>87</td>
<td>34</td>
<td>#USA, #fakenews</td>
</tr>
<tr>
<td>Ryan Henry…</td>
<td>@Carol</td>
<td>2017-11-06</td>
<td>18:00:13Z</td>
<td>51</td>
<td>4</td>
<td>#UIUC</td>
</tr>
<tr>
<td>Puerto Rico…</td>
<td>@Barack</td>
<td>2017-11-06</td>
<td>20:16:02Z</td>
<td>151</td>
<td>101</td>
<td>#hurricane</td>
</tr>
</tbody>
</table>
\[ \bar{e} \begin{bmatrix} \pi \cdot \begin{pmatrix} X_{11} & \cdots & X_{1s} \\ \vdots & \ddots & \vdots \\ X_{r1} & \cdots & X_{rs} \end{pmatrix} \end{bmatrix} \]
\[ \begin{vmatrix} \vec{e} \cdot \Pi \end{vmatrix} = \left( \begin{array}{ccc} x_{11} & \cdots & x_{1s} \\ \vdots & \ddots & \vdots \\ x_{r1} & \cdots & x_{rs} \end{array} \right) \]
"Simple" indexes of queries

rows can repeat

query in

columns can be all zeros

can "stretch" or "compress" queries

query out

each row is a (non-private, positional) query

\[
\langle 1000 \rangle \cdot \begin{pmatrix}
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0
\end{pmatrix} = \langle 00010 \rangle
\]
Simple indexes of queries

- All-zero columns $\Rightarrow$ some rows of $D$ are unfetchable!
- Repeated rows $\Rightarrow$ some rows of $D$ are fetchable in multiple ways!

Simple indexes of queries leak information about the query...
It's not a bug...

information leakage
It's not a bug...

...it's a feature!

information leakage
### "Simple" Indexes of Queries

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Query in:** \( \langle 1000 \rangle \)

**Columns can be all zeros**

Rows repeat

**Can "stretch" or "compress" queries**

\[ \langle 1000 \rangle \cdot \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} = \langle 00010 \rangle \]

**Query out**
"Batch" indexes of queries

<table>
<thead>
<tr>
<th>tweet</th>
<th>username</th>
<th>date</th>
<th>time</th>
<th>like</th>
<th>retweet</th>
<th>#tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thanks for...</td>
<td>@Barack</td>
<td>2017-11-05</td>
<td>14:45:13Z</td>
<td>903</td>
<td>87</td>
<td>#worstgiftever</td>
</tr>
<tr>
<td>Playing golf...</td>
<td>@Donald</td>
<td>2017-11-05</td>
<td>03:27:12Z</td>
<td>112</td>
<td>22</td>
<td>#MAGA, #borderwall</td>
</tr>
<tr>
<td>I could use...</td>
<td>@Alice</td>
<td>2017-11-05</td>
<td>14:00:10Z</td>
<td>6</td>
<td>0</td>
<td>#covfefe</td>
</tr>
<tr>
<td>Fake MSNBC...</td>
<td>@Donald</td>
<td>2017-11-06</td>
<td>04:01:58Z</td>
<td>87</td>
<td>34</td>
<td>#USA, #fakenews</td>
</tr>
<tr>
<td>Ryan Henry...</td>
<td>@Carol</td>
<td>2017-11-06</td>
<td>18:00:13Z</td>
<td>51</td>
<td>4</td>
<td>#UIUC</td>
</tr>
<tr>
<td>Puerto Rico...</td>
<td>@Barack</td>
<td>2017-11-06</td>
<td>20:16:02Z</td>
<td>151</td>
<td>101</td>
<td>#hurricane</td>
</tr>
</tbody>
</table>

23
\[
\begin{pmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0
\end{pmatrix} \cdot x
\]

\[
\begin{pmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0
\end{pmatrix} \cdot x
\]

\[
\begin{pmatrix}
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & x
\end{pmatrix} \cdot x
\]

\[
\begin{pmatrix}
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 1-x & 0 & 0 & 0 \\
0 & 0 & 0 & 1-x & 0
\end{pmatrix} \cdot x
\]
\[ X_{\varnothing,\varnothing}(2) \quad X_{\varnothing,\varnothing}(3) \quad X_{\varnothing,\varnothing}(4) \quad X_{\varnothing,\varnothing}(\ell+2) \]

\[ \begin{pmatrix}
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1-x & 0 & 0 & 0 & 0 & 0 & x \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & x & 0 & 1-x & 0 & 0 & 0
\end{pmatrix} \cdot X \]
Shamir\(_2(\varepsilon_j)\) → \(X_{\varepsilon_j}(2)\)

Shamir\(_3(\varepsilon_j)\) → \(X_{\varepsilon_j}(3)\)

Shamir\(_4(\varepsilon_j)\) → \(X_{\varepsilon_j}(4)\)

Shamir\(_{\ell+2}(\varepsilon_j)\) → \(X_{\varepsilon_j}(\ell+2)\)
Indexes of (batch) queries

positional query

x=0

x=1

x=k-1

@username

rank (1..k)
4-batch index of queries with $p = 2^{14}$ rows

(CUDA implementation; running on an Nvidia Tesla K20 GPU Accelerator)
IACR Cryptology ePrint Archive

- Scraped 10,181 papers + metadata
- Pruned 56 outliers (large files)
- 10,125 papers survived pruning
- Largest surviving file: 4.69 MiB
- Resulting DB: 46.35 GiB
- PIR over $GF(2^8)$ using open-source percy++ library
### IACRR Cryptology ePrint Archive

<table>
<thead>
<tr>
<th>Search criteria</th>
<th>Sort criteria</th>
<th>Simple index creation</th>
<th>Bucket creation</th>
<th>Bucket size</th>
<th>Throughput (GPU)</th>
<th>Throughput (CPU)</th>
<th>Throughput (PIR)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>keyword</strong></td>
<td>recency</td>
<td>1.5 s</td>
<td>54 ms</td>
<td>71.76 KiB</td>
<td>49 100</td>
<td>32 800</td>
<td>19.1 s</td>
</tr>
<tr>
<td></td>
<td>citations</td>
<td>1.3 s</td>
<td>52 ms</td>
<td>70.17 KiB</td>
<td>49 100</td>
<td>30 700</td>
<td>18.8 s</td>
</tr>
<tr>
<td><strong>author</strong></td>
<td>recency</td>
<td>3.1 s</td>
<td>63 ms</td>
<td>92.38 KiB</td>
<td>49 000</td>
<td>22 700</td>
<td>32.6 s</td>
</tr>
<tr>
<td></td>
<td>citations</td>
<td>3.1 s</td>
<td>63 ms</td>
<td>91.69 KiB</td>
<td>49 000</td>
<td>20 100</td>
<td>32.4 s</td>
</tr>
</tbody>
</table>

Standard positional query takes ~70 seconds!
This material is based upon work supported by the National Science Foundation under Grant No. CNS-1718475.

Disclaimer: Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
That's all for today, folks!