CrySP

Government of Canada
Vanier Canada Graduate Scholarships

Gouvernement du Canada
Bourses d'études supérieures du Canada Vanier

Canada

One (Block) Size Fits All

PIR and SPIR with Variable-Length Records via Multi-Block Queries

WATERLOO
CHERITON SCHOOL OF COMPUTER SCIENCE
cs.uwaterloo.ca

Ryan Henry
Yizhou Huang
Ian Goldberg

The 20th Annual
Network & Distributed System Security
Symposium (NDSS 2013)

NSERC CRSNG

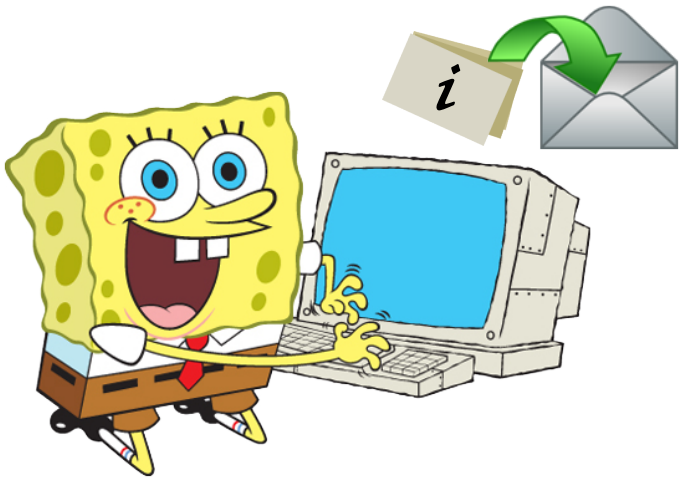
Ontario Research Fund

Private Information Retrieval

pri·vate in·for·ma·tion re·triev·al [**prahy**-vit in-fer-**mey**-sh*uh*n re-**tree**-v*uh*l]

Noun. A database technology that ***protects its users' privacy*** by ***hiding their interests*** from the database server operator(s).

Abbreviation: **PIR**



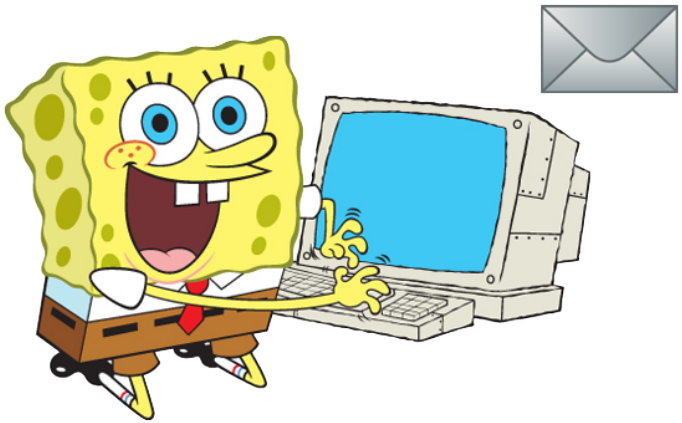
Bob queries for record i . . .

Private Information Retrieval

pri·vate in·for·ma·tion re·triev·al [**prahy**-vit in-fer-**mey**-sh*uh*n re-**tree**-v*uh*l]

Noun. A database technology that ***protects its users' privacy*** by ***hiding their interests*** from the database server operator(s).

Abbreviation: **PIR**



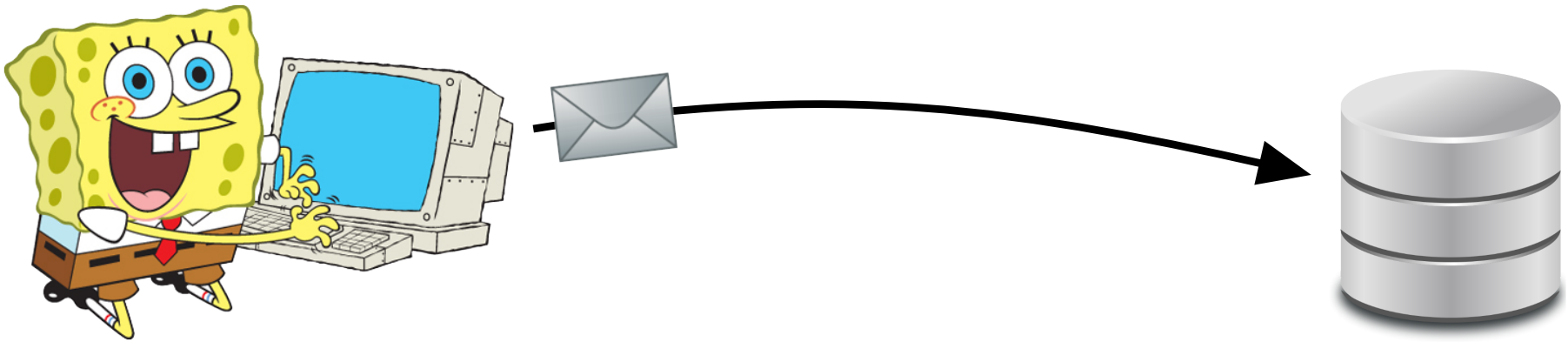
Bob queries for record i . . .

Private Information Retrieval

pri·vate in·for·ma·tion re·triev·al [**prahy**-vit in-fer-**mey**-sh*uh*n re-**tree**-v*uh*l]

Noun. A database technology that ***protects its users' privacy*** by ***hiding their interests*** from the database server operator(s).

Abbreviation: **PIR**



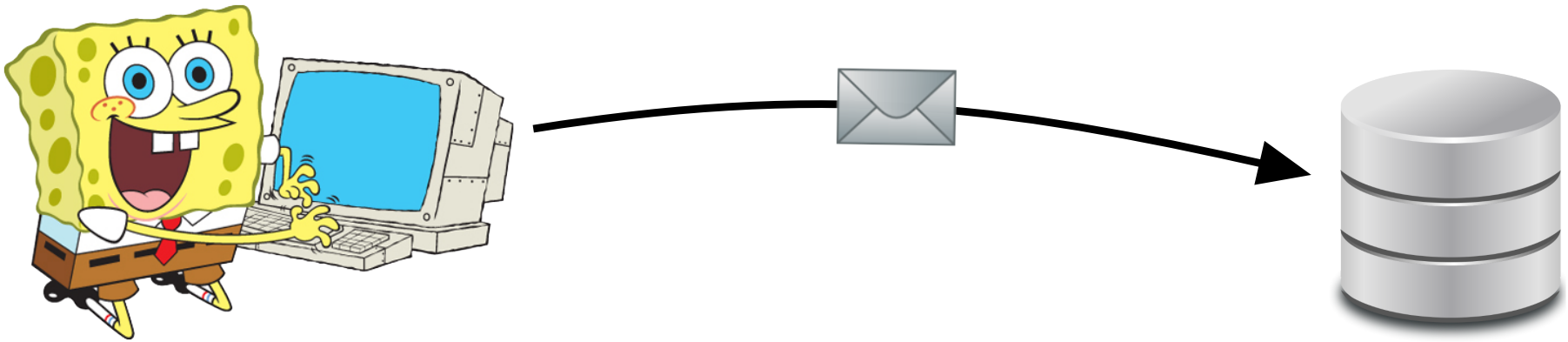
Bob queries for record i . . .

Private Information Retrieval

pri·vate in·for·ma·tion re·triev·al [**prahy**-vit in-fer-**mey**-sh*uh*n re-**tree**-v*uh*l]

Noun. A database technology that ***protects its users' privacy*** by ***hiding their interests*** from the database server operator(s).

Abbreviation: **PIR**



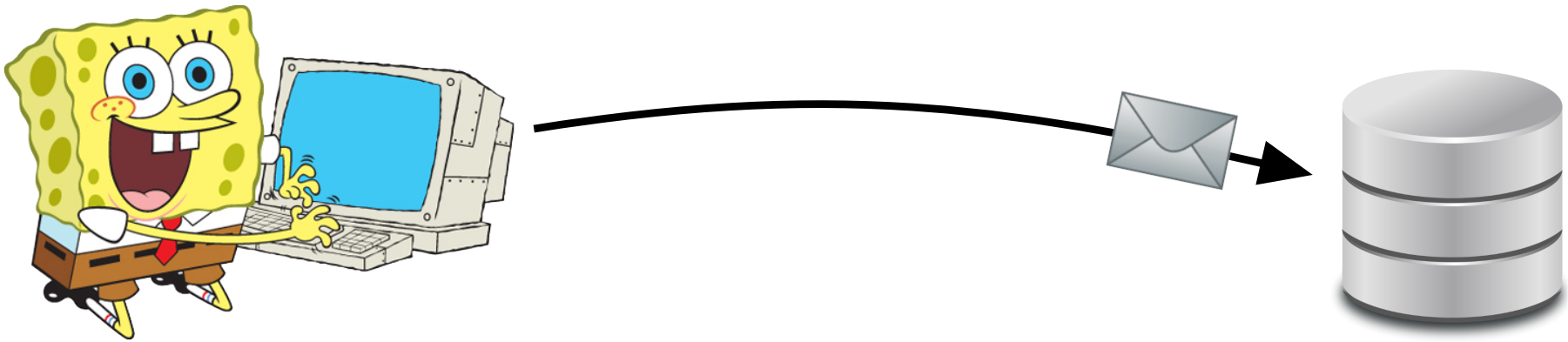
Bob queries for record i . . .

Private Information Retrieval

pri·vate in·for·ma·tion re·triev·al [**prahy**-vit in-fer-**mey**-sh*uh*n re-**tree**-v*uh*l]

Noun. A database technology that ***protects its users' privacy*** by ***hiding their interests*** from the database server operator(s).

Abbreviation: **PIR**



Bob queries for record i . . .

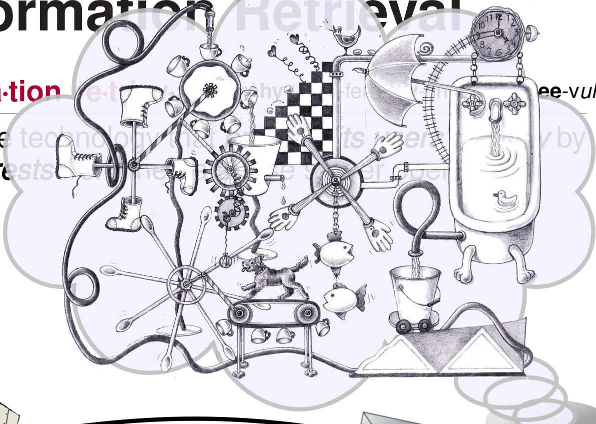
Source: http://wordkeepers.blogspot.ca/2012/06/

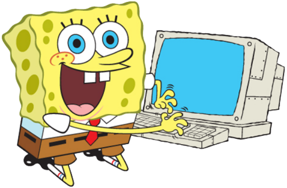
Private Information Retrieval

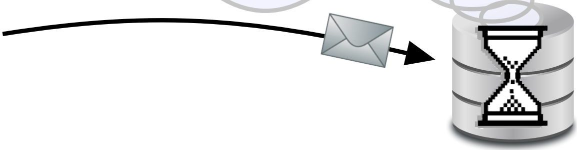
pri·vate in·for·ma·tion ... ee-vuhl]

Noun. A database technology that ... by
hiding their interests ...

Abbreviation: **PIR**







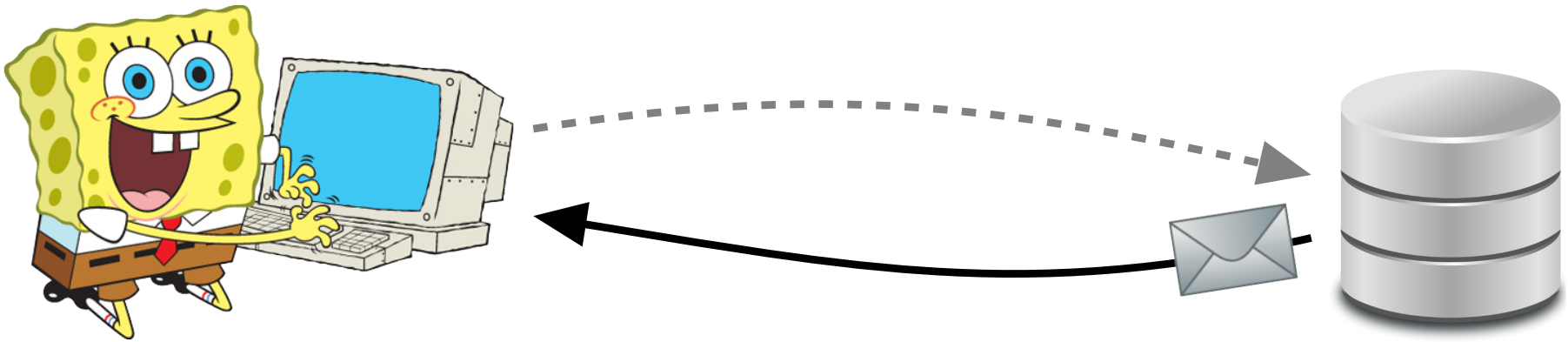
Bob queries for record i . . .

Private Information Retrieval

pri·vate in·for·ma·tion re·triev·al [**prahy**-vit in-fer-**mey**-sh*uh*n re-**tree**-v*uh*l]

Noun. A database technology that ***protects its users' privacy*** by ***hiding their interests*** from the database server operator(s).

Abbreviation: **PIR**



Bob queries for record i . . .

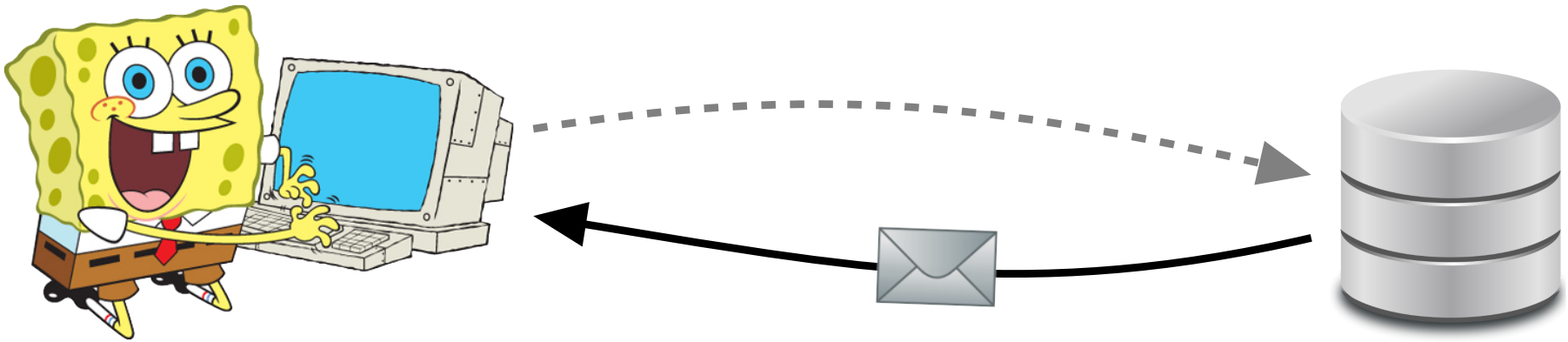
. . . and the database responds

Private Information Retrieval

pri·vate in·for·ma·tion re·triev·al [**prahy**-vit in-fer-**mey**-sh*uh*n re-**tree**-v*uh*l]

Noun. A database technology that ***protects its users' privacy*** by ***hiding their interests*** from the database server operator(s).

Abbreviation: **PIR**



Bob queries for record i . . .

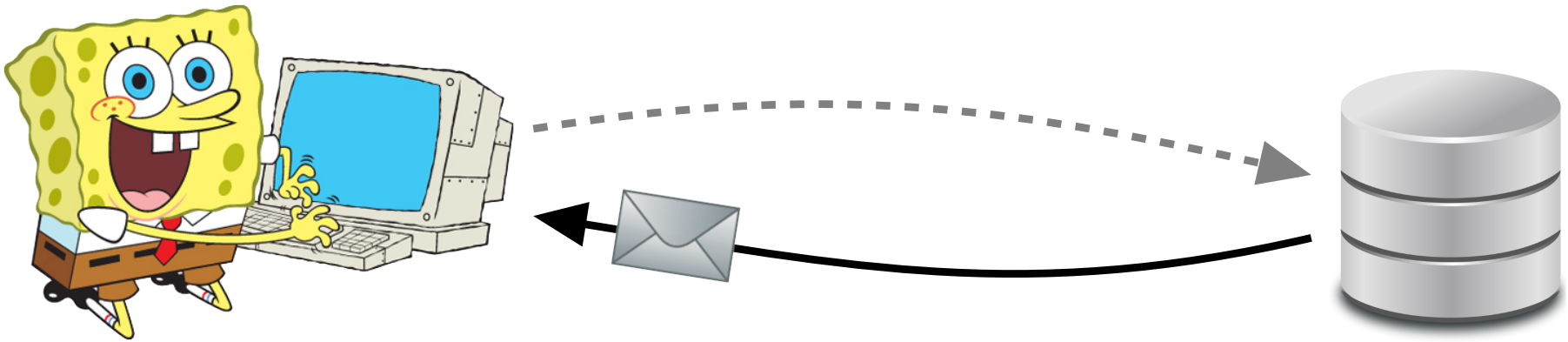
. . . and the database responds

Private Information Retrieval

pri·vate in·for·ma·tion re·triev·al [**prahy**-vit in-fer-**mey**-sh*uh*n re-**tree**-v*uh*l]

Noun. A database technology that ***protects its users' privacy*** by ***hiding their interests*** from the database server operator(s).

Abbreviation: **PIR**



Bob queries for record i . . .

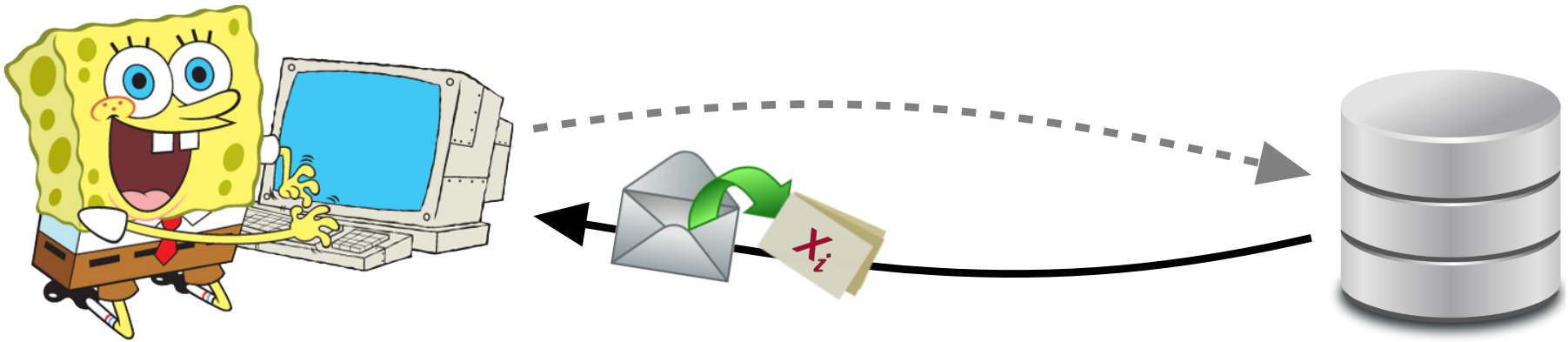
. . . and the database responds

Private Information Retrieval

pri·vate in·for·ma·tion re·triev·al [**prahy**-vit in-fer-**mey**-sh*uh*n re-**tree**-v*uh*l]

Noun. A database technology that ***protects its users' privacy*** by ***hiding their interests*** from the database server operator(s).

Abbreviation: **PIR**



Bob queries for record i . . .

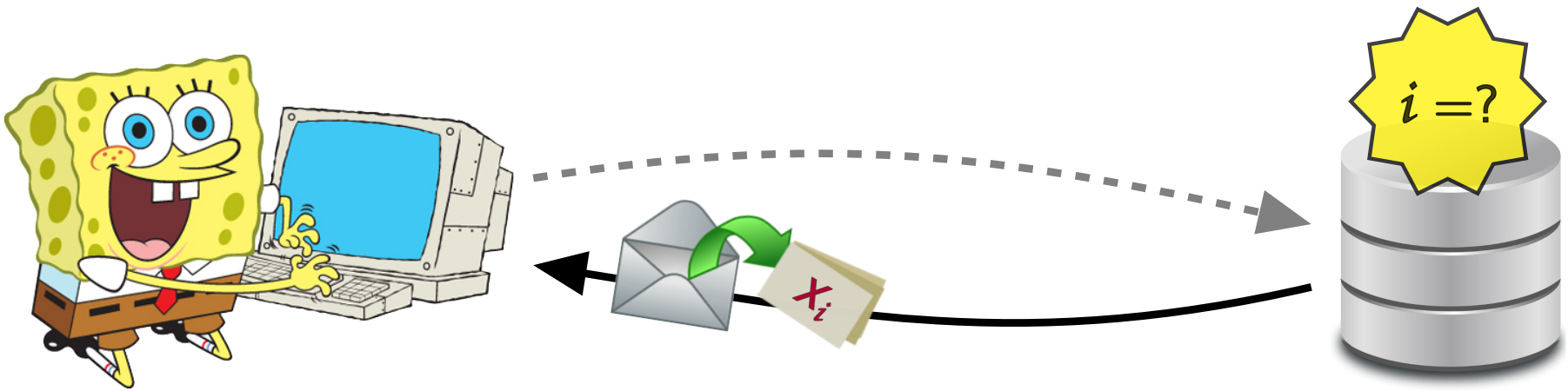
. . . and the database responds

Private Information Retrieval

pri·vate in·for·ma·tion re·triev·al [**prahy**-vit in-fer-**mey**-sh*uh*n re-**tree**-v*uh*l]

Noun. A database technology that ***protects its users' privacy*** by ***hiding their interests*** from the database server operator(s).

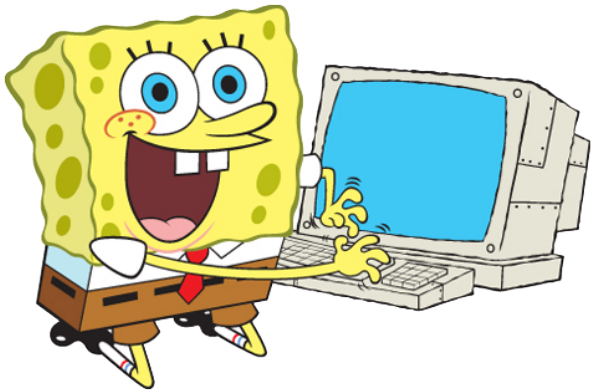
Abbreviation: **PIR**



Bob queries for record i . . .

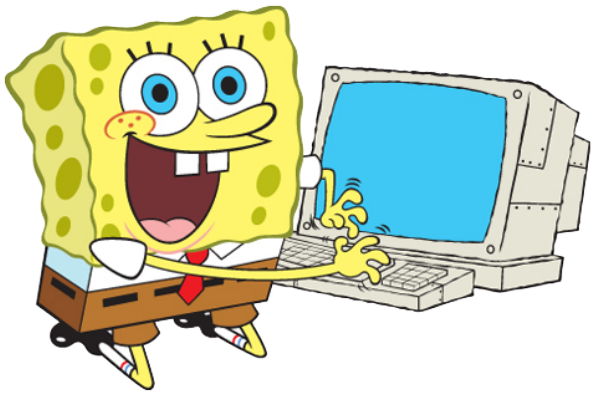
. . . and the database responds ***without learning i!***

The *trivial* PIR construction



I seek a record from your database.

The *trivial* PIR construction



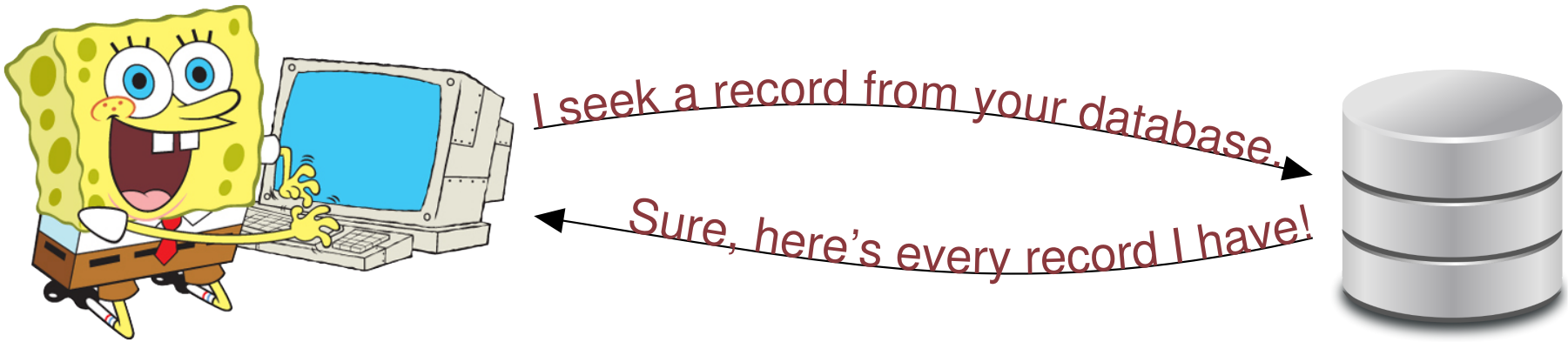
I seek a record from your database.

Sure, here's every record I have!

The *trivial* PIR construction

- Database sends *everything* to Bob

  $\Longrightarrow$ communication cost is $\{\backslash\mathbf{Huge}\ldots\}$



The *trivial* PIR construction

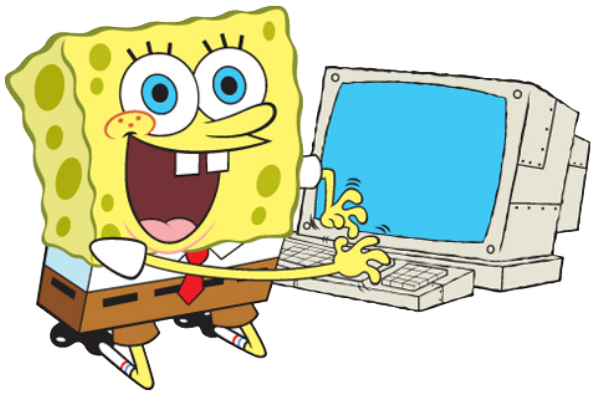
- Database sends *everything* to Bob

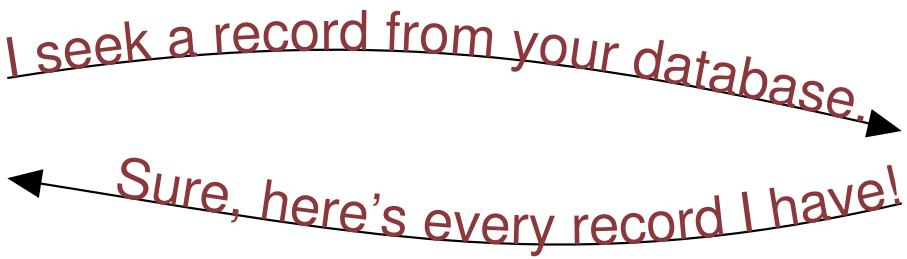
  $\Longrightarrow$ communication cost is $\{\backslash\mathbf{Huge}\ldots\}$

- **Non-triviality:** communication cost must be $o\left(\text{database size}\right)$

Three approaches:

1. cryptography
2. replication + noncollusion
3. trusted hardware





The *trivial* PIR construction

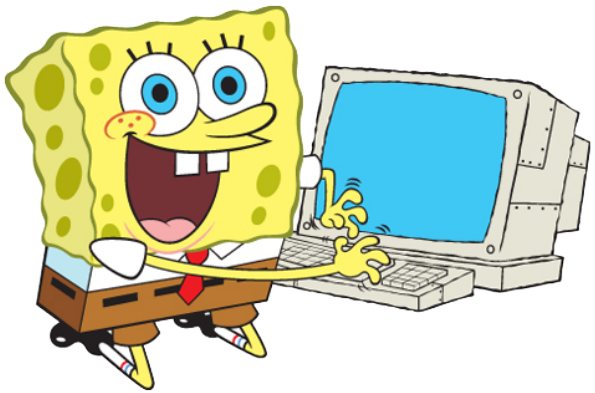
- Database sends *everything* to Bob

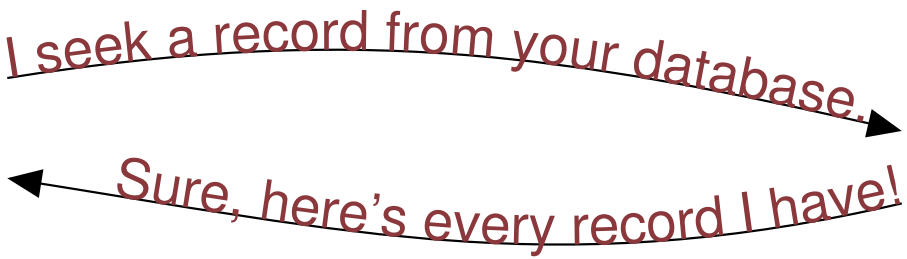
  $\Longrightarrow$ communication cost is $\{\backslash\mathbf{Huge}\ldots\}$

- **Non-triviality:** communication cost must be $o\left(\text{database size}\right)$

Three approaches:

1. cryptography
2. **replication + noncollusion**
3. trusted hardware

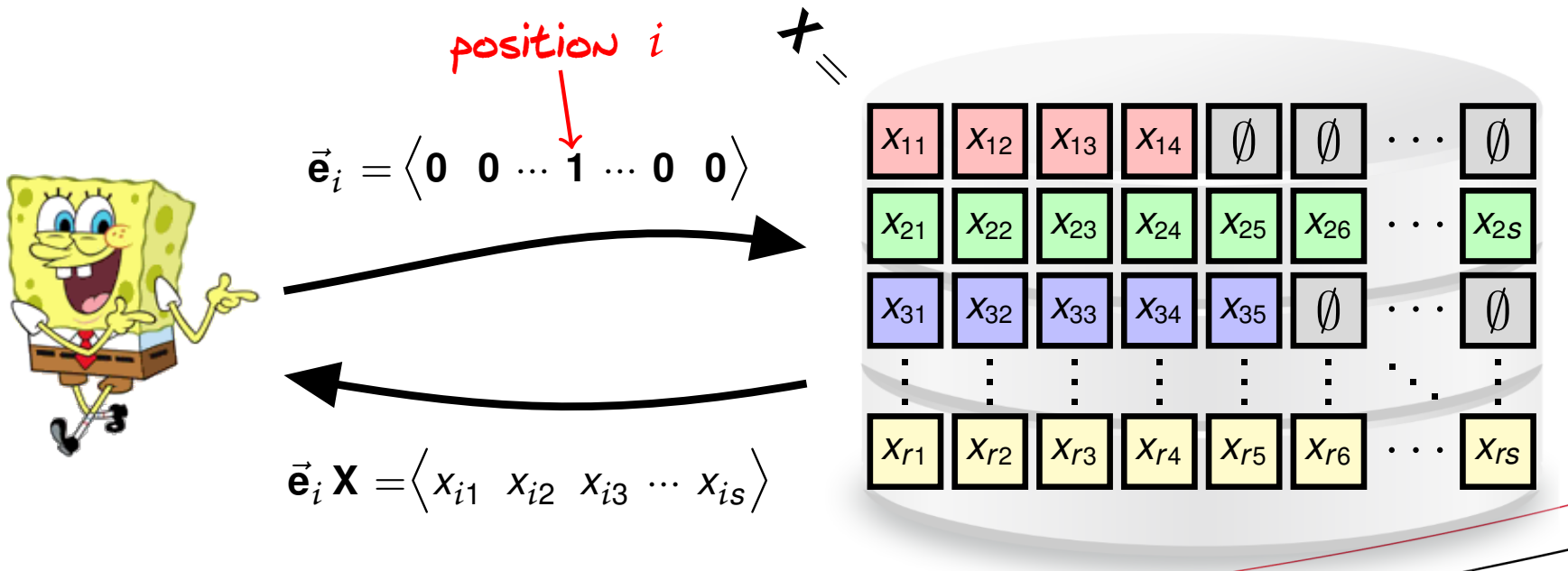




The *database* as a matrix

high school linear algebra $\Rightarrow$ Non-private database queries

- Encode each record as string of "words" from $\mathbb{F}$
- Pad all records to fixed length (s words)
- Arrange as rows of a matrix in $\mathbb{F}^{r \times s}$

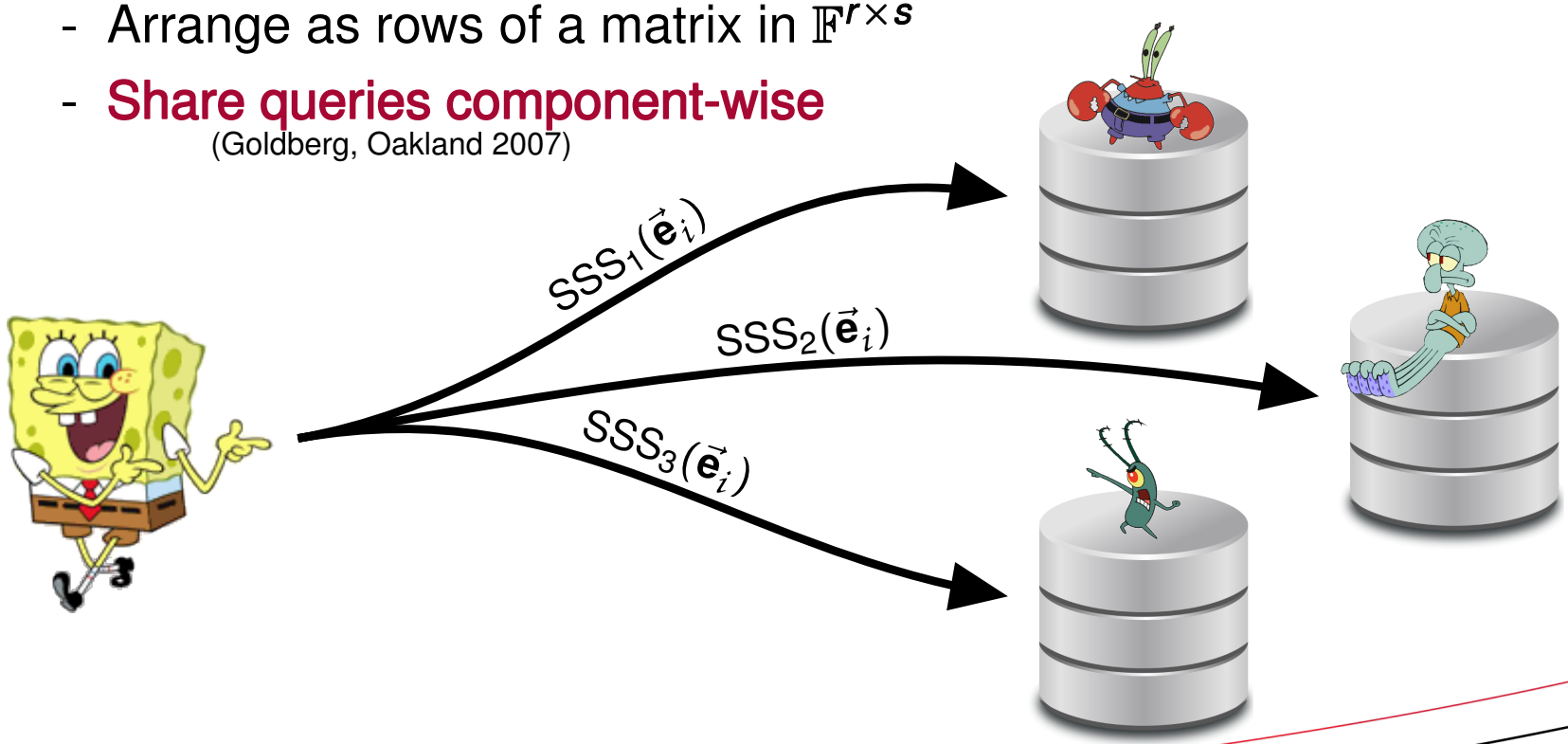


The *database* as a matrix

high school linear algebra $\Rightarrow$ **private database queries**

+Shamir secret sharing

- Encode each record as string of "words" from $\mathbb{F}$
- Pad all records to fixed length (s words)
- Arrange as rows of a matrix in $\mathbb{F}^{r \times s}$
- Share queries component-wise
  (Goldberg, Oakland 2007)

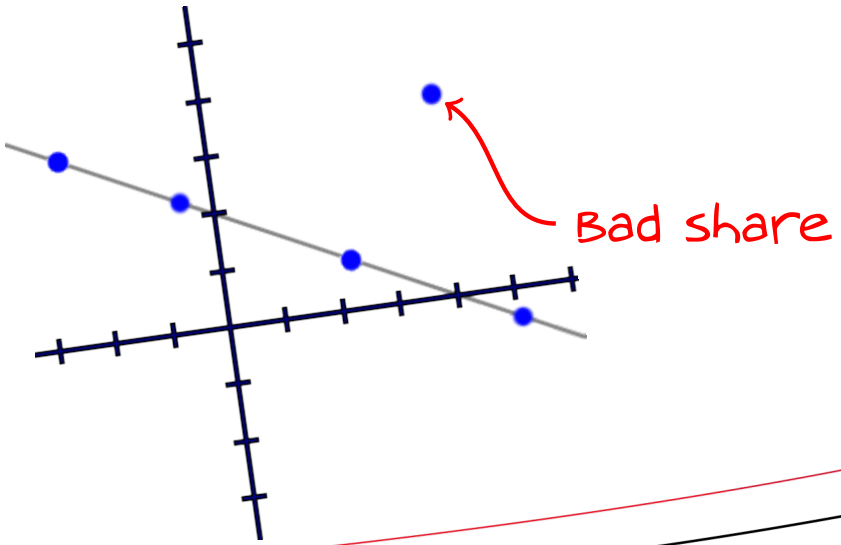


The *database* as a matrix

high school linear algebra $\Rightarrow$ **private database queries**

+Shamir secret sharing

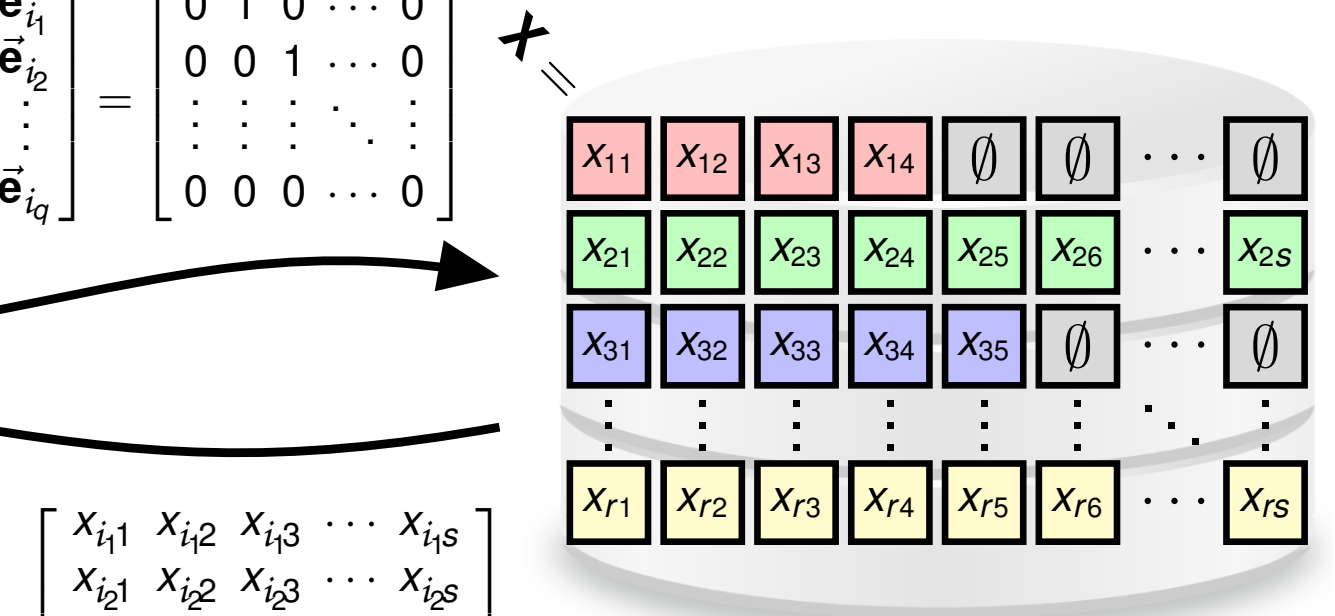
- Encode each record as string of "words" from $\mathbb{F}$
- Pad all records to fixed length (s words)
- Arrange as rows of a matrix in $\mathbb{F}^{r \times s}$
- Share queries component-wise
- **Robustness:** query succeeds despite some Byzantine servers



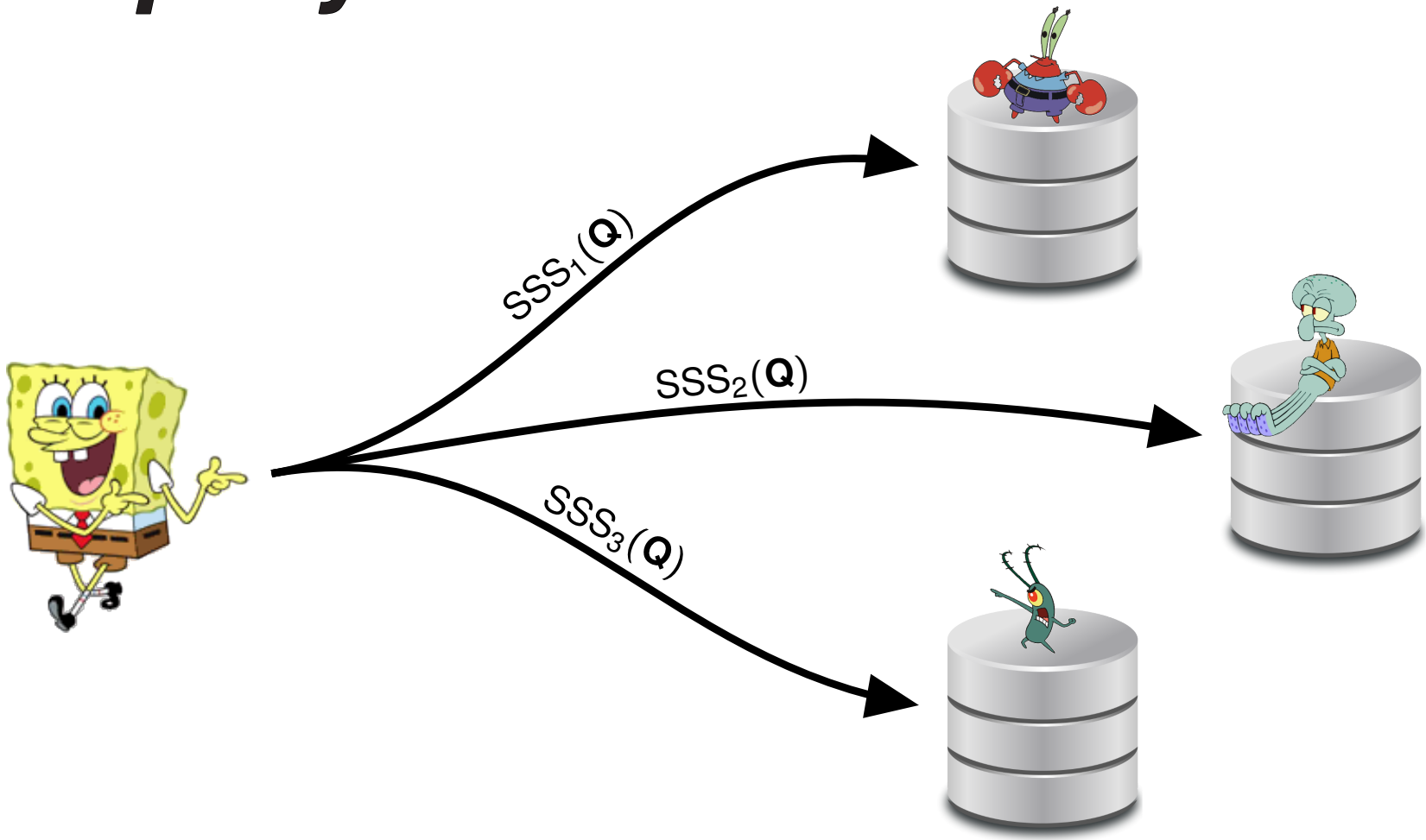
The *query* as a matrix

$$\mathbf{Q} = \begin{bmatrix} \vec{\mathbf{e}}_{i_1} \\ \vec{\mathbf{e}}_{i_2} \\ \vdots \\ \vec{\mathbf{e}}_{i_q} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

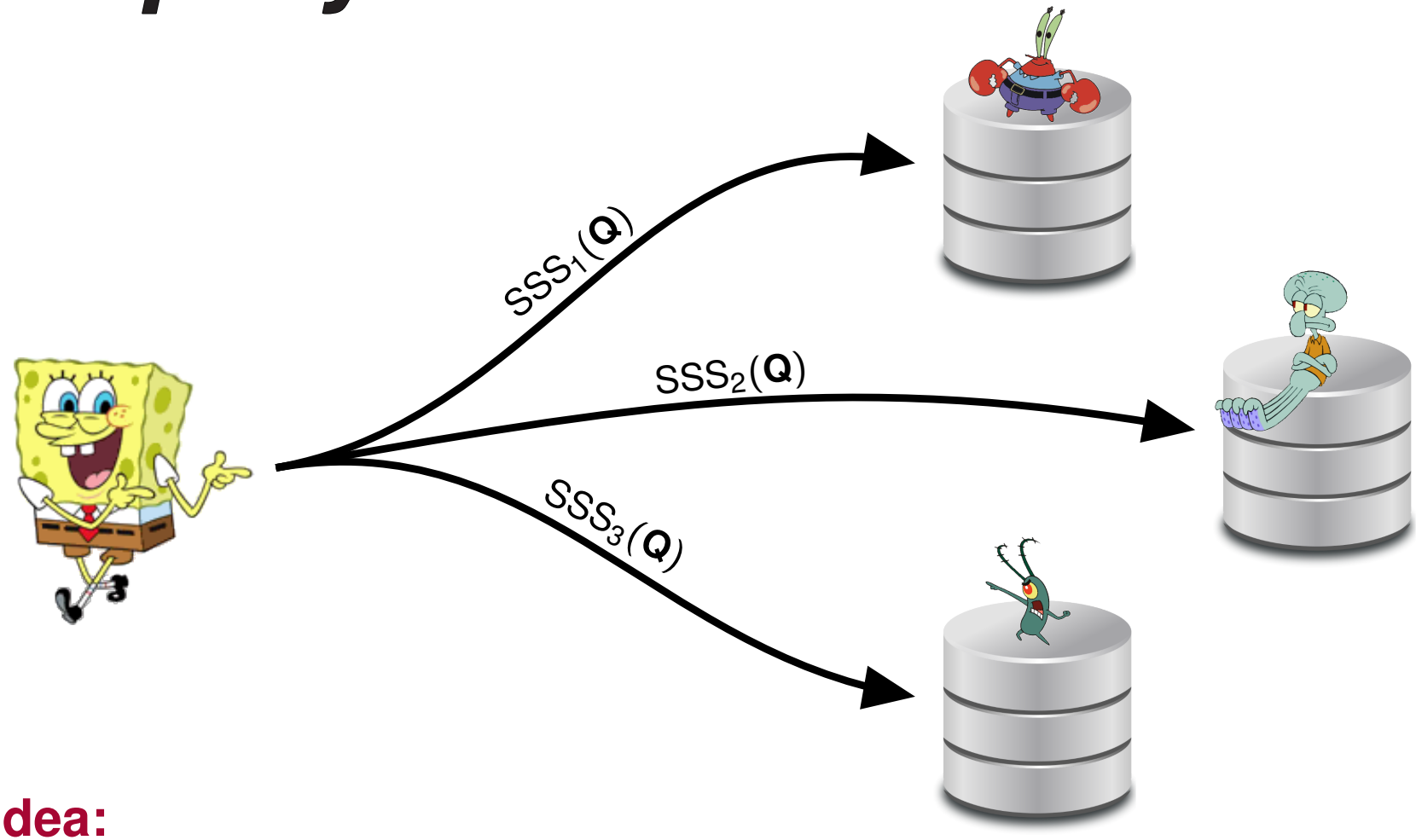
$$\mathbf{Q}\,\mathbf{X} = \begin{bmatrix} x_{i_1 1} & x_{i_1 2} & x_{i_1 3} & \cdots & x_{i_1 s} \\ x_{i_2 1} & x_{i_2 2} & x_{i_2 3} & \cdots & x_{i_2 s} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{i_q 1} & x_{i_q 2} & x_{i_q 3} & \cdots & x_{i_q s} \end{bmatrix}$$



The *query* as a matrix



The *query* as a matrix

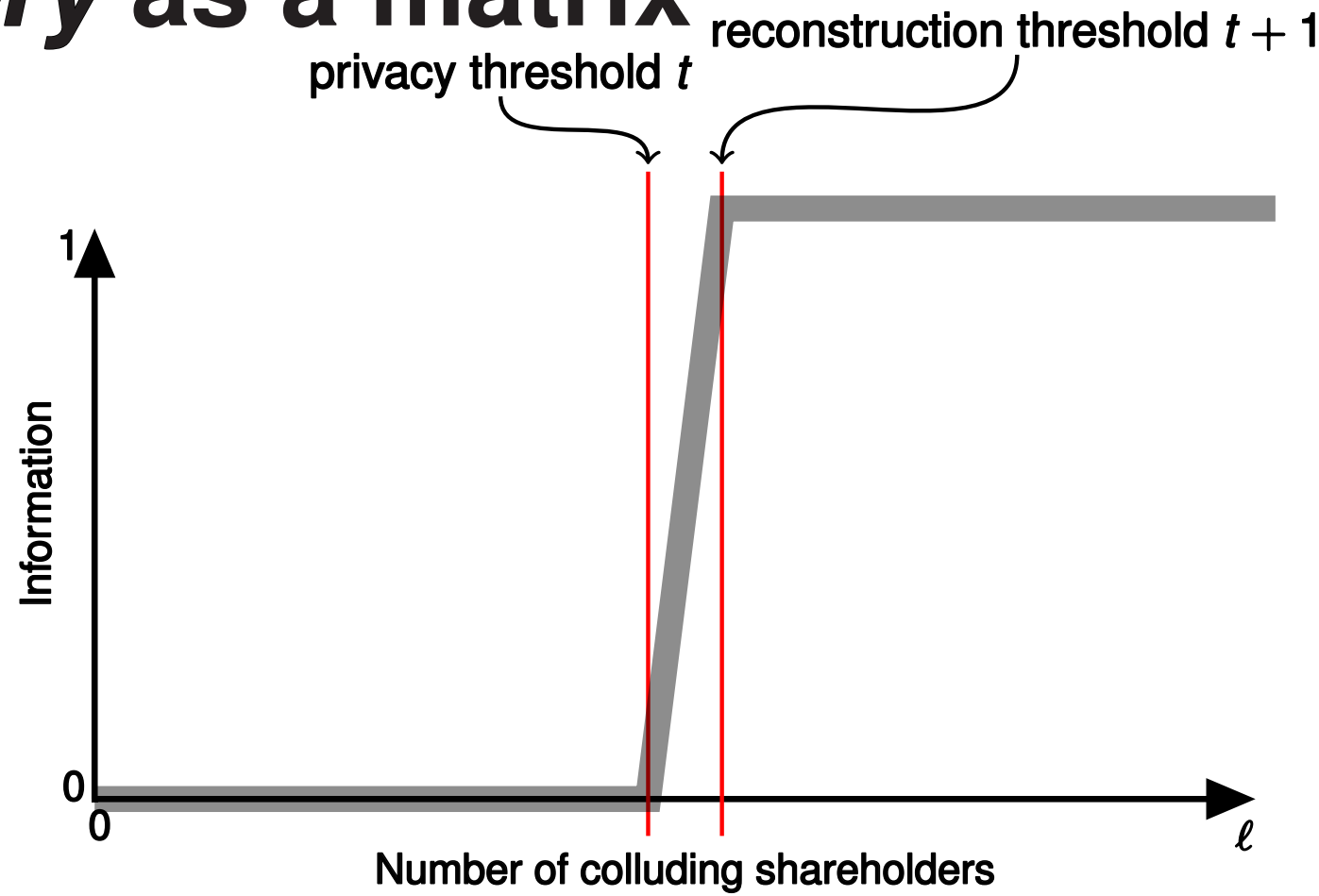


Idea:

share **Q** column-wise with a **ramp scheme**

$\Longrightarrow$ $SSS_j(\mathbf{Q})$ is still a ***vector*** of length r

The *query* as a matrix

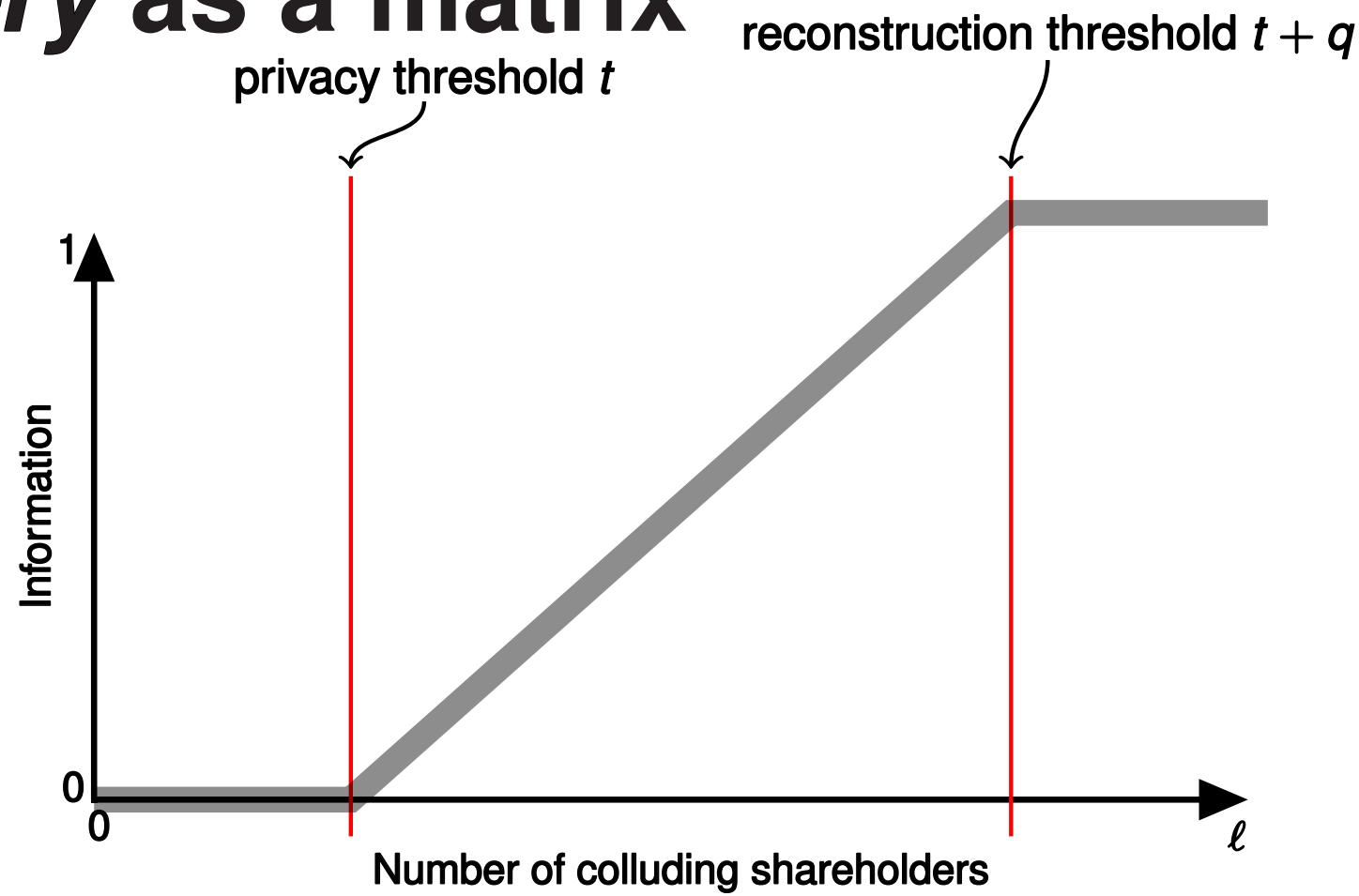


Idea:

share **Q** column-wise with a **ramp scheme**

$\Longrightarrow$ $\mathrm{SSS}_j(\mathbf{Q})$ is still a ***vector*** of length r

The *query* as a matrix



Idea:

share **Q** column-wise with a **ramp scheme**

$\Longrightarrow$ $\mathrm{SSS}_j(\mathbf{Q})$ is still a ***vector*** of length r

Impact to privacy & robustness

$$\ell = t + q + v + 1$$

where

- ℓ = number of servers
- t = collusion threshold (for perfect privacy)
- q = rows retrieved per query
- v = Byzantine robustness bound

$$\therefore \text{ fixed } (\ell, t) \ \wedge \ \uparrow q \ \Rightarrow \ \downarrow v$$

Sacrificing robustness

2007: (Goldberg, Oakland 2007)

$$\ell \text{ servers } \& \ t \text{ privacy} \implies v < \ell - \lfloor \sqrt{\ell t} \rfloor \text{ robust}$$

Sacrificing robustness

2007: (Goldberg, Oakland 2007)

$$\ell \text{ servers } \& \ t \text{ privacy} \implies v < \ell - \lfloor \sqrt{\ell t} \rfloor \text{ robust}$$

2012: (Devet, Goldberg, and Heninger, USENIX Security 2012)

$$\ell \text{ servers } \& \ t \text{ privacy} \implies v < \ell - t - 1 \text{ robust}$$

Sacrificing robustness

2007: (Goldberg, Oakland 2007)

$$\ell \text{ servers } \& \ t \text{ privacy} \implies v < \ell - \lfloor \sqrt{\ell t} \rfloor \text{ robust}$$

2012: (Devet, Goldberg, and Heninger, USENIX Security 2012)

$$\ell \text{ servers } \& \ t \text{ privacy} \implies v < \ell - t - 1 \text{ robust}$$

There is no shortage of robustness to sacrifice.

Impact to throughput & overhead

Assume fixed: number of servers ℓ, privacy threshold t

Fact 1

Coalitions of t or fewer servers cannot distinguish between single-block and multi-block queries

Impact to throughput & overhead

Assume fixed: number of servers ℓ, privacy threshold t

Fact 1

Coalitions of t or fewer servers cannot distinguish between single-block and multi-block queries

$\Rightarrow$ **no change in per-server communication/computation cost**

Impact to throughput & overhead

Assume fixed: number of servers ℓ, privacy threshold t

Fact 1

Coalitions of t or fewer servers cannot distinguish between single-block and multi-block queries

$\Rightarrow$ **no change in per-server communication/computation cost**

Fact 2

The user obtains a factor q more bits of useable data with a multi-block query as compared to with a single-block query

Impact to throughput & overhead

Assume fixed: number of servers ℓ, privacy threshold t

Fact 1

Coalitions of t or fewer servers cannot distinguish between single-block and multi-block queries

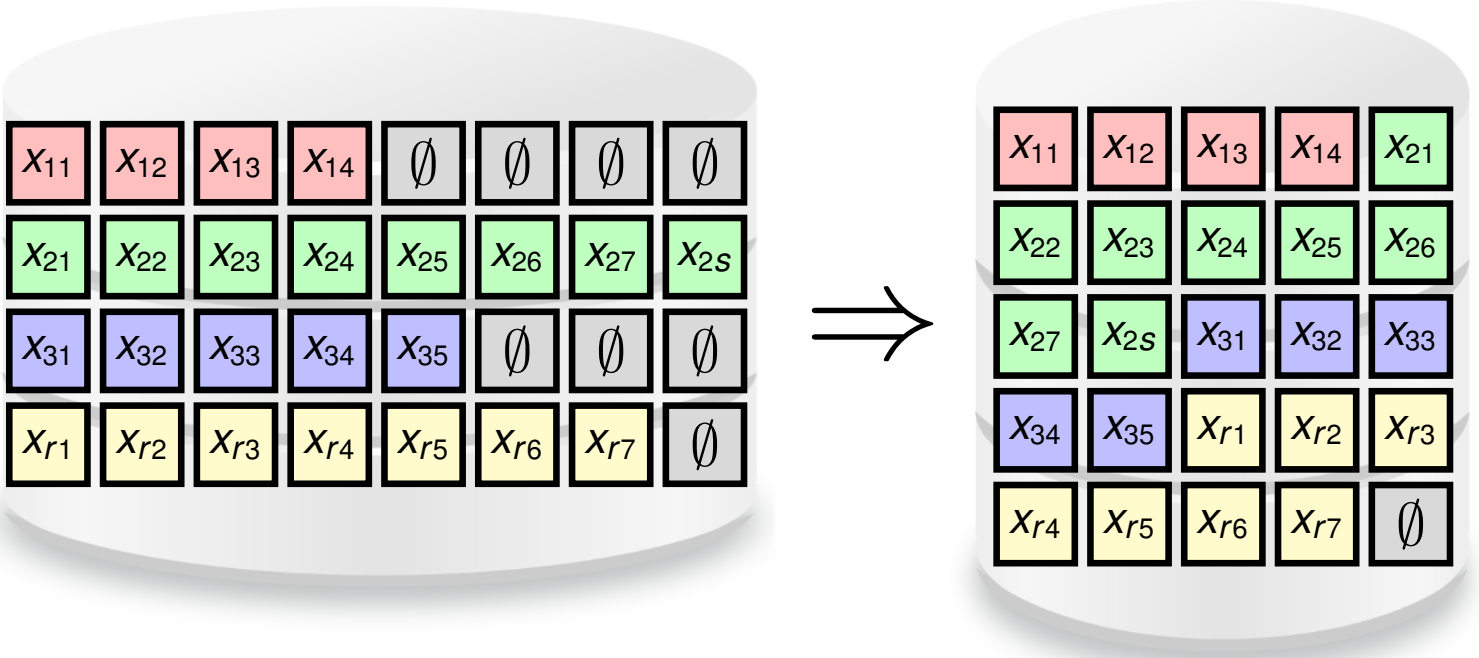
$\Rightarrow$ **no change in per-server communication/computation cost**

Fact 2

The user obtains a factor q more bits of useable data with a multi-block query as compared to with a single-block query

$\Rightarrow$ **factor q increase in throughput/decrease in overhead**

PIR with variable record lengths



Records padded to fixed length:

response size
$\wr$
longest record size

No padding:

response size
$\wr$
$\frac{1}{q}$(longest record size)

Also in the paper

- Rationality-based argument to support trading off some robustness

Also in the paper

- Rationality-based argument to support trading off some robustness
- Derivations for optimal parameter

Also in the paper

- Rationality-based argument to support trading off some robustness
- Derivations for optimal parameter
- Extensions to *symmetric PIR* (SPIR) and *priced symmetric PIR* (PSPIR)

Also in the paper

- Rationality-based argument to support trading off some robustness
- Derivations for optimal parameter
- Extensions to *symmetric PIR* (SPIR) and *priced symmetric PIR* (PSPIR)
- Performance measurements / a pretty graph

Takeaways

1. Multi-block PIR queries: retrieve q blocks for cost of one

Takeaways

1. Multi-block PIR queries: retrieve q blocks for cost of one
2. There is *plenty of robustness* to go around

Takeaways

1. Multi-block PIR queries: retrieve q blocks for cost of one
2. There is *plenty of robustness* to go around
3. PIR can model real databases

Takeaways

1. Multi-block PIR queries: retrieve q blocks for cost of one
2. There is *plenty of robustness* to go around
3. PIR can model real databases
4. PIR-based applications may be on the horizon

The aforementioned “pretty graph”