Single-Server Private Information Retrieval With Side Information Under Arbitrary Popularity Profiles

Alejandro Gomez-Leos
(University of Texas at Austin)

Joint work with

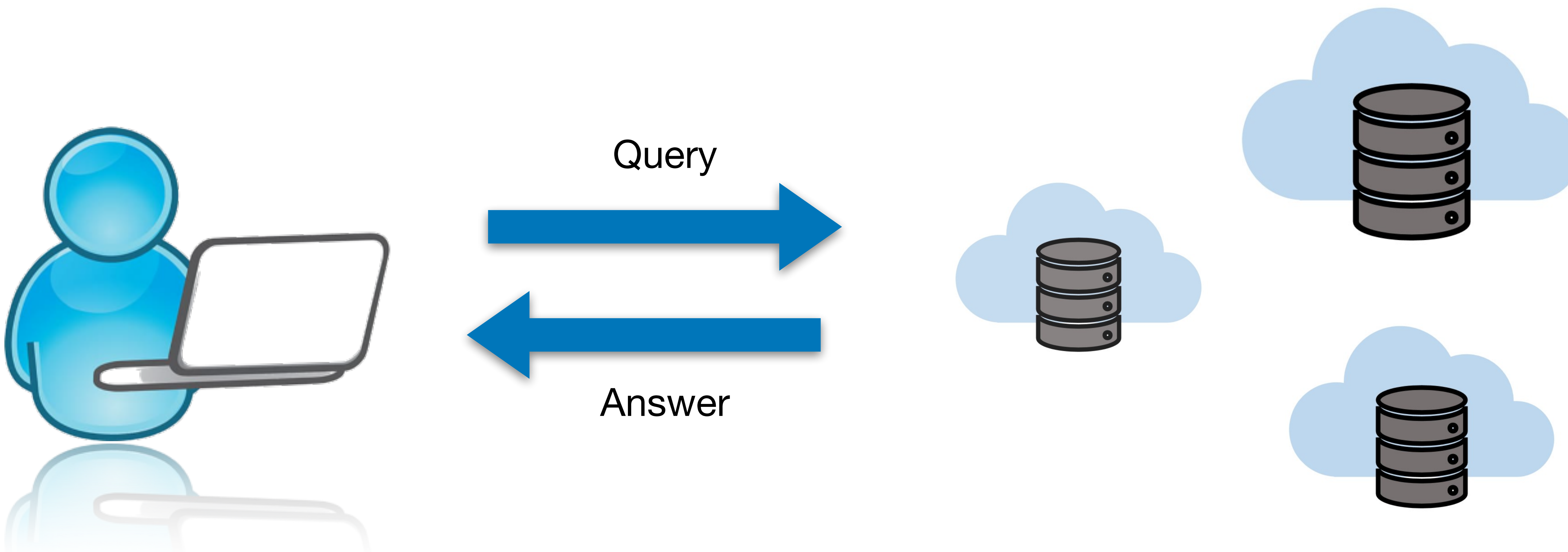
Anoosheh Heidarzadeh
(Santa Clara University)

This work was done while both authors were at Texas A&M University.

ITW 2022

Private Information Retrieval with Side Information (PIR-SI)

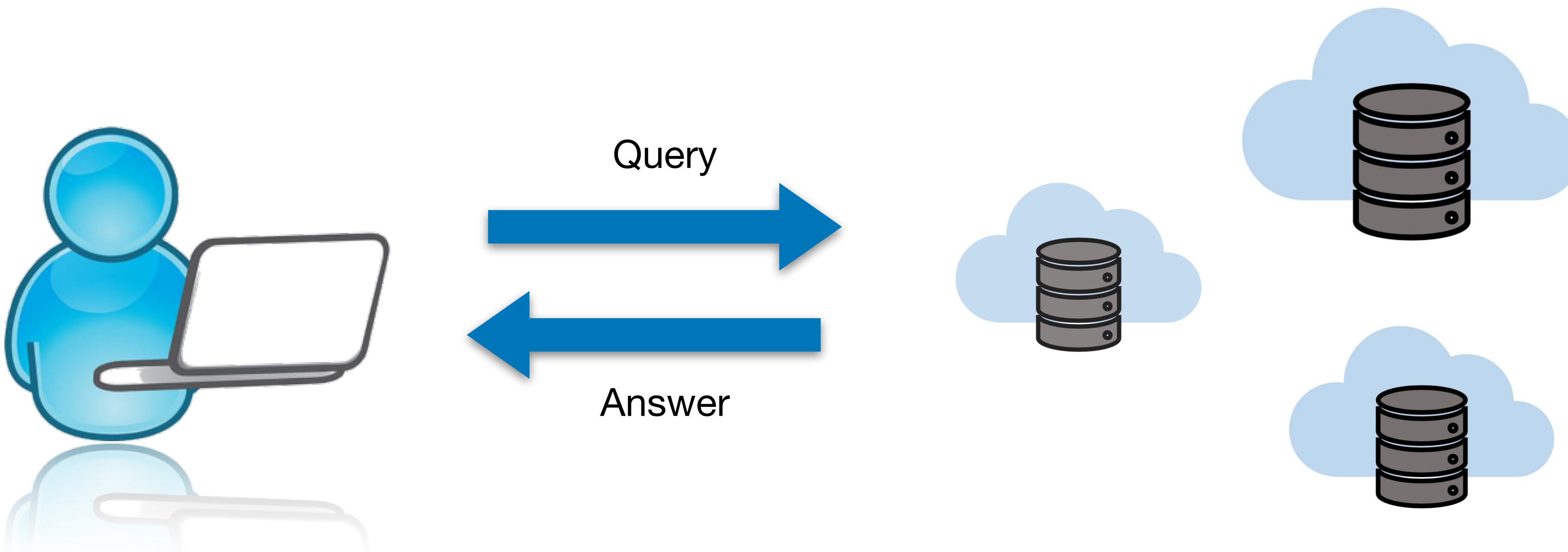
- A dataset is stored on one (or more) remote server(s).
- A user knows some subset of the dataset as side information, and desires a different subset of the dataset.



- Minimize download cost (i.e., total amount of information downloaded)
- Subject to leaking no information about the identities of the desired data

Private Information Retrieval with Side Information (PIR-SI)

- A dataset is stored on one (or more) remote server(s).
- A user knows some subset of the dataset as side information, and desires a different subset of the dataset.



- Minimize download cost (i.e., total amount of information downloaded)
- Subject to leaking no information about the identities of the desired data

Often, the PIR-SI formulations assume uniformly popular data.

Motivation

- From the servers' perspective, some data may be more popular than others.
    - E.g. any ranked dataset (video, image, forum messaging, etc.)
- Studies show the Zipf, Gamma, or Weibull distributions are more appropriate statistical models for online data access patterns*.

Cha-Kwak-Rodriguez-Ann-Moon, ***IEEE/ACM Trans. on Networking '09***
Cheng-Dale-Liu, ***IEEE/ACM Int. Workshop on QoS '08***
Breslau-Cao-Fan-Phillips-Shenker, ***INFOCOM '99***

Motivation

- From the servers' perspective, some data may be more popular than others.
  - E.g. any ranked dataset (video, image, forum messaging, etc.)
- Studies show the Zipf, Gamma, or Weibull distributions are more appropriate statistical models for online data access patterns*.

This work focuses on extending PIR-SI techniques to this more general setting.

Cha-Kwak-Rodriguez-Ann-Moon, ***IEEE/ACM Trans. on Networking '09***
Cheng-Dale-Liu, ***IEEE/ACM Int. Workshop on QoS '08***
Breslau-Cao-Fan-Phillips-Shenker, ***INFOCOM '99***

Related Work

	Data Popularity	# Servers	Side Info.
Sun-Jafar ‘17	No	Multiple	No
Banawan-Ulukus ‘17, ‘18	No	Multiple	No
Kadhe *et al*. ‘20	No	Multiple	Yes
Kadhe *et al.* ‘17	No	Single	Yes
Heidarzadeh *et al*. ‘18	No	Single	Yes
Li-Gastpar ‘18	No	Single	Yes
Heidarzadeh-Sprintson ‘22	No	Single	Yes
Vithana-Banawan-Ulukus ‘20	Yes	Multiple	No
This work	**Yes**	**Single**	**Yes**

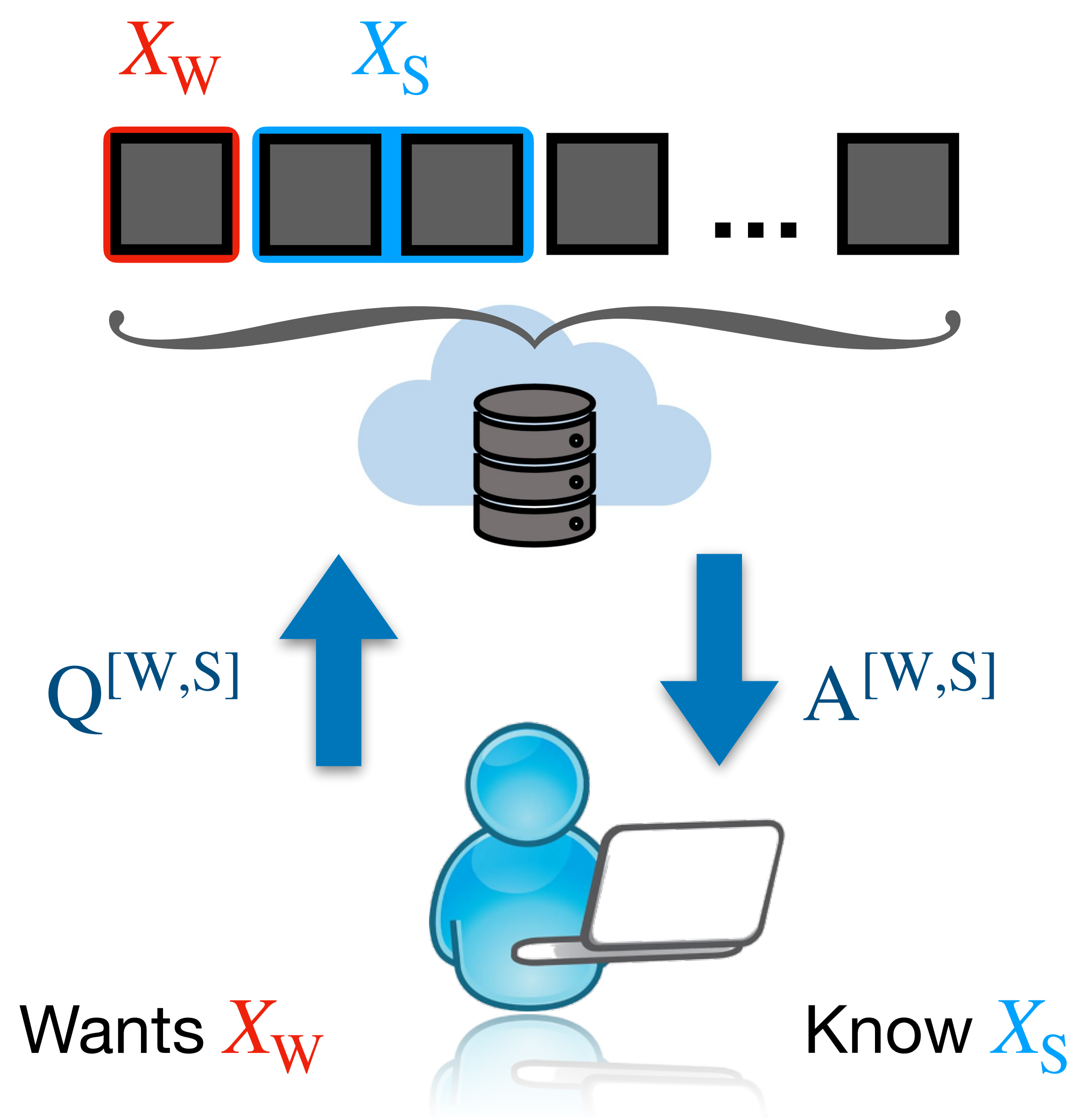
Outline

- Model + Assumptions
- A Motivating Example
- Main Results
- Simulations
- Summary and Open Problems

Popularity-Aware PIR-SI (PA-PIR-SI) Setting

- Server stores K messages $X_1, \ldots, X_K$ (independent and uniform over $\mathbb{F}_q^n$)

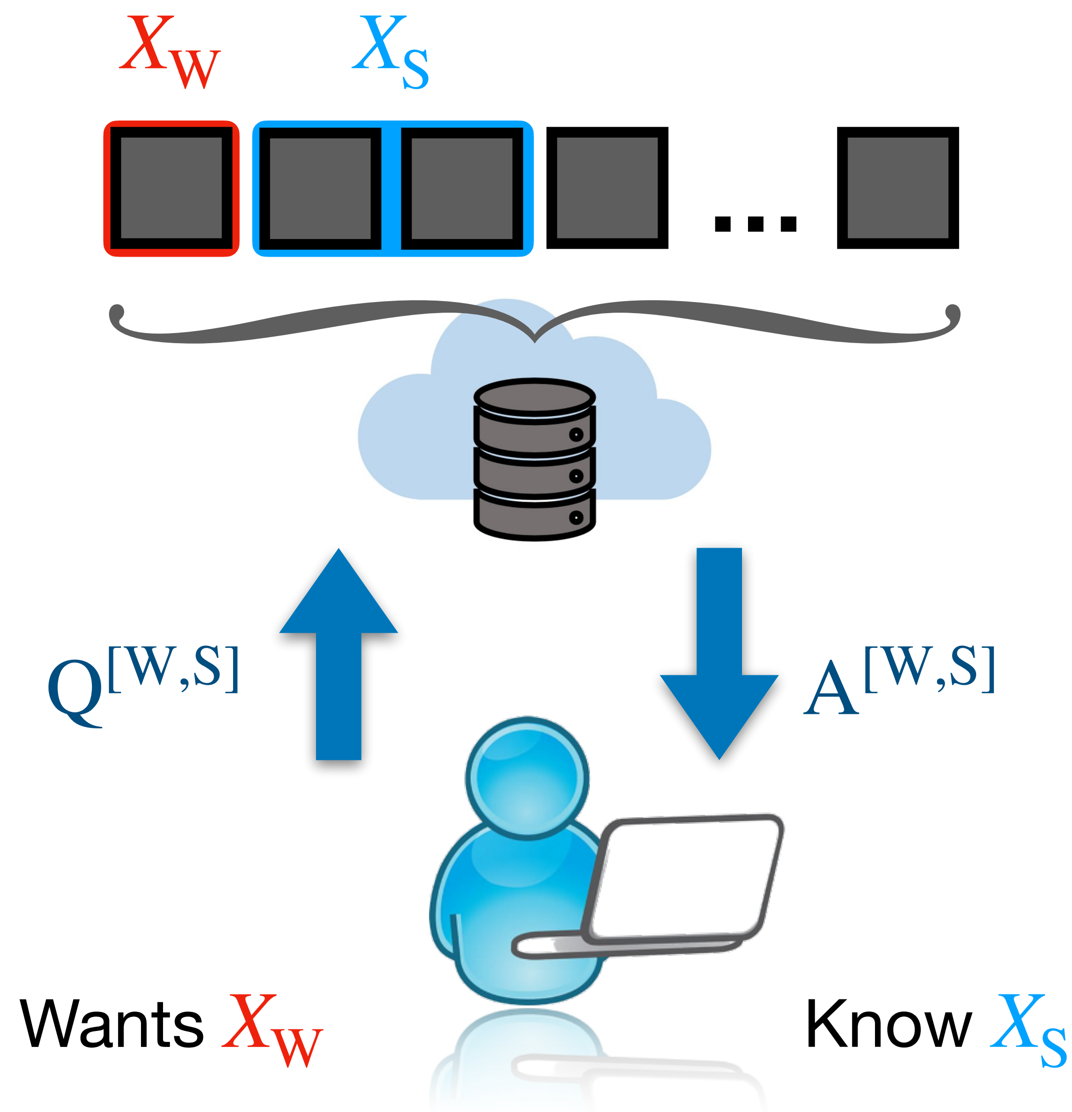
$$H(X_i) = n \log_2 q := B \quad \forall i \in \mathscr{K} \triangleq \{1, \ldots, K\}$$



Popularity-Aware PIR-SI (PA-PIR-SI) Setting

- Server stores K messages $X_1, \ldots, X_K$ (independent and uniform over $\mathbb{F}_q^n$)

$$H(X_i) = n \log_2 q := B \quad \forall i \in \mathscr{K} \triangleq \{1, \ldots, K\}$$



X_S: Side info. message(s)

X_W: Demand message(s)

S: Side info. index set

W: Demand index set

M: # side info. message(s)

Popularity-Aware PIR-SI (PA-PIR-SI) Setting

- Server stores K messages $X_1, \ldots, X_K$ (independent and uniform over $\mathbb{F}_q^n$)

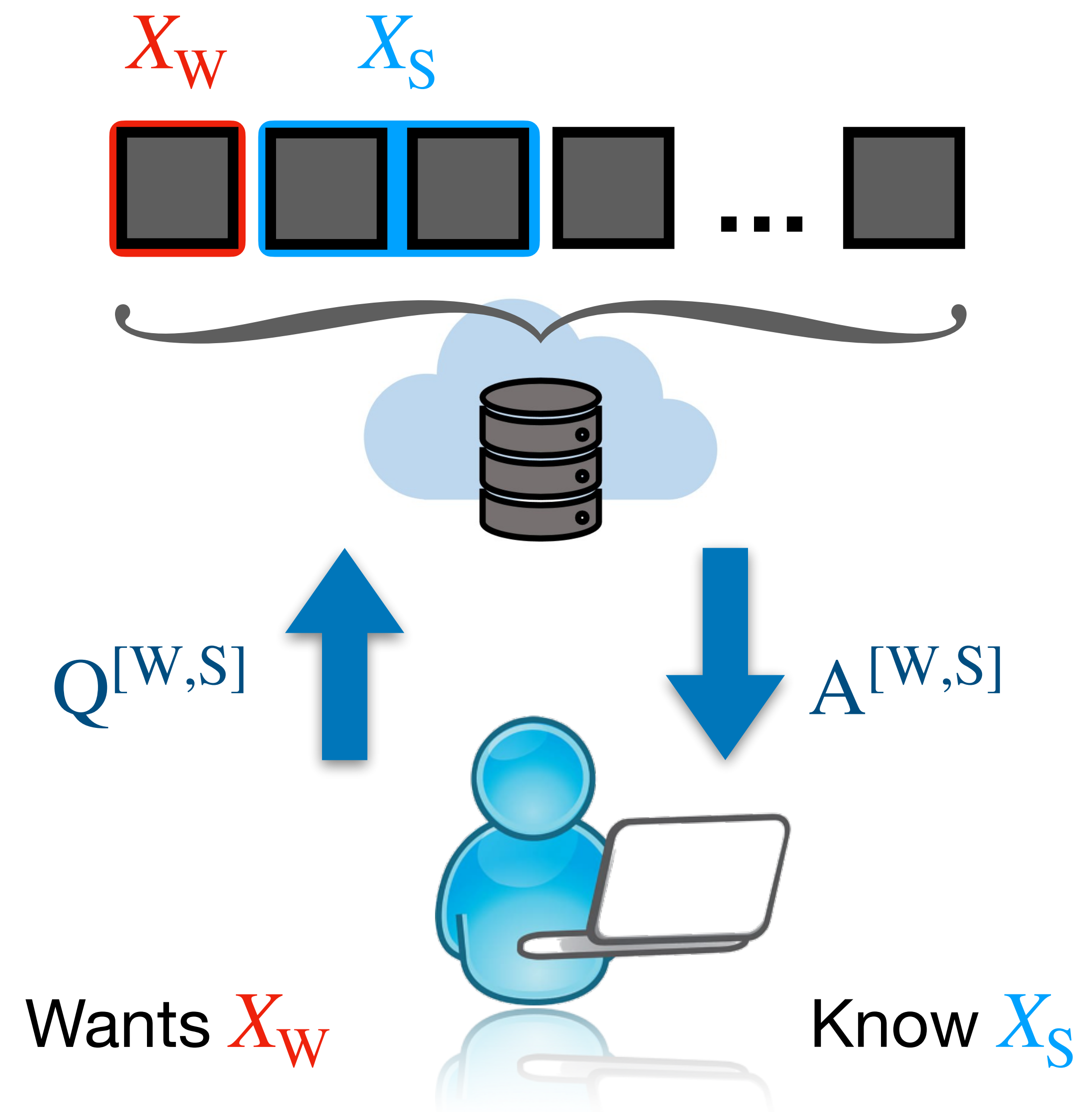
$$H(X_i) = n \log_2 q := B \quad \forall i \in \mathscr{K} \triangleq \{1, \ldots, K\}$$

- Message popularities

$$\lambda_i > 0 \quad \forall i \in \mathscr{K}$$

- Popularity Profile

$$\Lambda \triangleq (\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_K)$$



X_S: Side info. message(s)

X_W: Demand message(s)

S: Side info. index set

W: Demand index set

M: # side info. message(s)

Probability Model

- Side information index set is distributed uniformly.

$$p_{\mathbf{S}}(\mathrm{S}^*) \triangleq \frac{1}{\binom{K}{M}} \quad \forall \mathrm{S}^* \in [\mathscr{K}]^M$$

$[\mathscr{K}]^M$ denotes the set of all M-size subsets of $\mathscr{K}$.

Probability Model

- Side information index set is distributed uniformly.

$$p_{\mathbf{S}}(\mathrm{S}^*) \triangleq \frac{1}{\binom{K}{M}} \quad \forall \mathrm{S}^* \in [\mathscr{K}]^M$$

$[\mathscr{K}]^M$ denotes the set of all M-size subsets of $\mathscr{K}$.

Why is this a good assumption?

Probability Model

- Side information index set is distributed uniformly.

$$p_{\mathbf{S}}(\mathrm{S}^*) \triangleq \frac{1}{\binom{K}{M}} \quad \forall \mathrm{S}^* \in [\mathscr{K}]^M$$

$[\mathscr{K}]^M$ denotes the set of all M-size subsets of $\mathscr{K}$.

- Conditional distribution of demand index given side info. index set is a function of the popularity profile.

$$p_{\mathbf{W}|\mathbf{S}}(\mathrm{W}^* \,|\, \mathrm{S}^*) \triangleq \begin{cases} \frac{\lambda_{\mathrm{W}^*}}{\sum_{i \in \mathscr{K} \setminus \mathrm{S}^*} \lambda_i} & \forall \mathrm{W}^* \in \mathscr{K}, \forall \mathrm{S}^* \in [\mathscr{K} \setminus \mathrm{W}^*]^M, \\ 0 & \text{otherwise}. \end{cases}$$

Probability Model

- Conditional distribution of demand index given side info. index set is a function of the popularity profile.

$$p_{\mathbf{W}|\mathbf{S}}(\mathrm{W}^* \,|\, \mathrm{S}^*) \triangleq \begin{cases} \dfrac{\lambda_{\mathrm{W}^*}}{\sum_{i \in \mathscr{K} \setminus \mathrm{S}^*} \lambda_i} & \forall \mathrm{W}^* \in \mathscr{K}, \forall \mathrm{S}^* \in [\mathscr{K} \setminus \mathrm{W}^*]^M, \\ 0 & \text{otherwise} . \end{cases}$$

Probability Model

- Conditional distribution of demand index given side info. index set is a function of the popularity profile.

$$p_{\mathbf{W}|\mathbf{S}}(\mathrm{W}^* \mid \mathrm{S}^*) \triangleq \begin{cases} \dfrac{\lambda_{\mathrm{W}^*}}{\sum_{i \in \mathscr{K} \setminus \mathrm{S}^*} \lambda_i} & \forall \mathrm{W}^* \in \mathscr{K}, \forall \mathrm{S}^* \in [\mathscr{K} \setminus \mathrm{W}^*]^M, \\ 0 & \text{otherwise}. \end{cases}$$

E.g. $K = 4$**,** $M = 1$**,** $\Lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$

Probability Model

- Conditional distribution of demand index given side info. index set is a function of the popularity profile.

$$p_{\mathbf{W}|\mathbf{S}}(\mathrm{W}^* \mid \mathrm{S}^*) \triangleq \begin{cases} \dfrac{\lambda_{\mathrm{W}^*}}{\sum_{i \in \mathscr{K} \setminus \mathrm{S}^*} \lambda_i} & \forall \mathrm{W}^* \in \mathscr{K}, \forall \mathrm{S}^* \in [\mathscr{K} \setminus \mathrm{W}^*]^M, \\ 0 & \text{otherwise}. \end{cases}$$

E.g. $K = 4$, $M = 1$, $\Lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$

$$p_{\mathbf{W}|\mathbf{S}}(\{1\} \mid \{2\}) = \frac{\lambda_1}{\lambda_1 + \lambda_3 + \lambda_4}$$

Probability Model

- Conditional distribution of demand index given side info. index set is a function of the popularity profile.

$$p_{\mathbf{W}|\mathbf{S}}(\mathrm{W}^* \,|\, \mathrm{S}^*) \triangleq \begin{cases} \dfrac{\lambda_{\mathrm{W}^*}}{\sum_{i \in \mathscr{K} \setminus \mathrm{S}^*} \lambda_i} & \forall \mathrm{W}^* \in \mathscr{K}, \forall \mathrm{S}^* \in [\mathscr{K} \setminus \mathrm{W}^*]^M, \\ 0 & \text{otherwise}. \end{cases}$$

E.g. $K = 4$**,** $M = 1$**,** $\Lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$

$$p_{\mathbf{W}|\mathbf{S}}(\{1\} \,|\, \{2\}) = \frac{\lambda_1}{\lambda_1 + \lambda_3 + \lambda_4}$$

$$p_{\mathbf{W}|\mathbf{S}}(\{2\} \,|\, \{1\}) = \frac{\lambda_2}{\lambda_2 + \lambda_3 + \lambda_4}$$

Probability Model

- Conditional distribution of demand index given side info. index set is a function of the popularity profile.

$$p_{\mathbf{W}|\mathbf{S}}(\mathrm{W}^* \mid \mathrm{S}^*) \triangleq \begin{cases} \dfrac{\lambda_{\mathrm{W}^*}}{\sum_{i \in \mathscr{K} \setminus \mathrm{S}^*} \lambda_i} & \forall \mathrm{W}^* \in \mathscr{K}, \forall \mathrm{S}^* \in [\mathscr{K} \setminus \mathrm{W}^*]^M, \\ 0 & \text{otherwise} . \end{cases}$$

E.g. $K = 4$, $M = 1$, $\Lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$

$$p_{\mathbf{W}|\mathbf{S}}(\{1\} \mid \{2\}) = \frac{\lambda_1}{\lambda_1 + \lambda_3 + \lambda_4}$$

$$p_{\mathbf{W}|\mathbf{S}}(\{2\} \mid \{1\}) = \frac{\lambda_2}{\lambda_2 + \lambda_3 + \lambda_4}$$

$$p_{\mathbf{W}|\mathbf{S}}(\{1\} \mid \{1\}) = 0$$

Probability Model

- Side information index set is distributed uniformly.

$$p_{\mathbf{S}}(\mathrm{S}^*) \triangleq \frac{1}{\binom{K}{M}} \quad \forall \mathrm{S}^* \in [\mathscr{K}]^M$$

$[\mathscr{K}]^M$ denotes the set of all M-size subsets of $\mathscr{K}$.

- Conditional distribution of demand index given side info. index set is a function of the popularity profile.

$$p_{\mathbf{W}|\mathbf{S}}(\mathrm{W}^* \,|\, \mathrm{S}^*) \triangleq \begin{cases} \frac{\lambda_{\mathrm{W}^*}}{\sum_{i \in \mathscr{K} \setminus \mathrm{S}^*} \lambda_i} & \forall \mathrm{W}^* \in \mathscr{K}, \forall \mathrm{S}^* \in [\mathscr{K} \setminus \mathrm{W}^*]^M, \\ 0 & \text{otherwise}. \end{cases}$$

Probability Model

- Side information index set is distributed uniformly.

$$p_{\mathbf{S}}(\mathrm{S}^*) \triangleq \frac{1}{\binom{K}{M}} \quad \forall \mathrm{S}^* \in [\mathscr{K}]^M$$

$[\mathscr{K}]^M$ denotes the set of all M-size subsets of $\mathscr{K}$.

- Conditional distribution of demand index given side info. index set is a function of the popularity profile.

$$p_{\mathbf{W}|\mathbf{S}}(\mathrm{W}^* \mid \mathrm{S}^*) \triangleq \begin{cases} \frac{\lambda_{\mathrm{W}^*}}{\sum_{i \in \mathscr{K} \setminus \mathrm{S}^*} \lambda_i} & \forall \mathrm{W}^* \in \mathscr{K}, \forall \mathrm{S}^* \in [\mathscr{K} \setminus \mathrm{W}^*]^M, \\ 0 & \text{otherwise}. \end{cases}$$

- Joint distribution of demand index and side info. index set follows…

$$p_{\mathbf{W},\mathbf{S}}(\mathrm{W}^*, \mathrm{S}^*) = \frac{1}{\binom{K}{M}} \frac{\lambda_{\mathrm{W}^*}}{\sum_{i \in \mathscr{K} \setminus \mathrm{S}^*} \lambda_i} \quad \forall \mathrm{W}^* \in \mathscr{K}, \forall \mathrm{S}^* \in [\mathscr{K} \setminus \mathrm{W}^*]^M$$

Probability Model

- Side information index set is distributed uniformly.

$$p_{\mathbf{S}}(\mathrm{S}^*) \triangleq \frac{1}{\binom{K}{M}} \quad \forall \mathrm{S}^* \in [\mathscr{K}]^M$$

$[\mathscr{K}]^M$ denotes the set of all M-size subsets of $\mathscr{K}$.

- Conditional distribution of demand index given side info. index set is a function of the popularity profile.

$$p_{\mathbf{W}|\mathbf{S}}(\mathrm{W}^* \,|\, \mathrm{S}^*) \triangleq \begin{cases} \frac{\lambda_{\mathrm{W}^*}}{\sum_{i \in \mathscr{K} \setminus \mathrm{S}^*} \lambda_i} & \forall \mathrm{W}^* \in \mathscr{K}, \forall \mathrm{S}^* \in [\mathscr{K} \setminus \mathrm{W}^*]^M, \\ 0 & \text{otherwise}. \end{cases}$$

- Joint distribution of demand index and side info. index set follows…

$$p_{\mathbf{W},\mathbf{S}}(\mathrm{W}^*, \mathrm{S}^*) = \frac{1}{\binom{K}{M}} \frac{\lambda_{\mathrm{W}^*}}{\sum_{i \in \mathscr{K} \setminus \mathrm{S}^*} \lambda_i} \quad \forall \mathrm{W}^* \in \mathscr{K}, \forall \mathrm{S}^* \in [\mathscr{K} \setminus \mathrm{W}^*]^M$$

- Resulting marginal distribution of demand index…

$$p_{\mathbf{W}}(\mathrm{W}^*) = \frac{1}{\binom{K}{M}} \sum_{\mathrm{S}^* \in [\mathscr{K} \setminus \mathrm{W}^*]^M} \frac{\lambda_{\mathrm{W}^*}}{\sum_{i \in \mathscr{K} \setminus \mathrm{S}^*} \lambda_i} \quad \forall \mathrm{W}^* \in \mathscr{K}.$$

Requirements

- Feasibility: Answer must be a deterministic function of query and messages.

$$H(\mathbf{A}^{[\mathrm{W},\mathrm{S}]} \,|\, \mathbf{Q}^{[\mathrm{W},\mathrm{S}]}, \mathbf{X}_1, \ldots, \mathbf{X}_K) = 0$$

- Decodability: Demand must be recoverable from answer, query, and side info.

$$H(\mathbf{X}_{\mathrm{W}} \,|\, \mathbf{A}, \mathbf{Q}, \mathbf{X}_{\mathrm{S}}) = 0$$

- Privacy: Query must not reveal any information about the demand index.

$$\mathbb{P}(\mathbf{W} = \mathrm{W}^* \,|\, \mathbf{Q} = \mathrm{Q}) = \mathbb{P}(\mathbf{W} = \mathrm{W}^*) \quad \forall \mathrm{W}^* \in \mathscr{K}\,.$$

Requirements

- Feasibility: Answer must be a deterministic function of query and messages.

$$H(\mathbf{A}^{[\mathrm{W},\mathrm{S}]} \mid \mathbf{Q}^{[\mathrm{W},\mathrm{S}]}, \mathbf{X}_1, \ldots, \mathbf{X}_K) = 0$$

- Decodability: Demand must be recoverable from answer, query, and side info.

$$H(\mathbf{X}_{\mathrm{W}} \mid \mathbf{A}, \mathbf{Q}, \mathbf{X}_{\mathrm{S}}) = 0$$

- Privacy: Query must not reveal any information about the demand index.

$$\mathbb{P}(\mathbf{W} = \mathrm{W}^* \mid \mathbf{Q} = \mathrm{Q}) = \mathbb{P}(\mathbf{W} = \mathrm{W}^*) \quad \forall \mathrm{W}^* \in \mathscr{K} .$$

- Given the popularity profile Λ, the PA-PIR-SI problem is to design a protocol to generate Q and A, for any given (W, S), to satisfy these conditions.

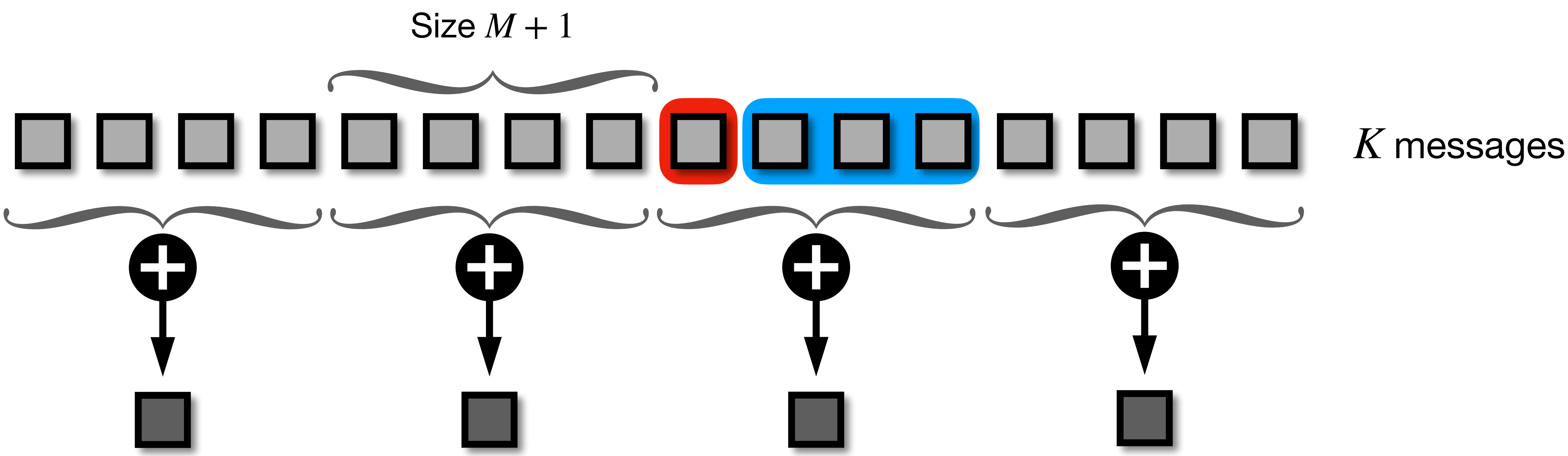
Characterizing Performance

- In particular, interested in the most efficient protocols.
- The **rate** of a PA-PIR-SI protocol given by $\dfrac{\text{Amount of info. demanded}}{\text{Expected amount of info. downloaded}}$

$$= \frac{B}{\sum_{\mathrm{W}^* \in \mathscr{K}} \sum_{\mathrm{S}^* \in [\mathscr{K} \setminus \mathrm{W}^*]^M} p_{\mathbf{W},\mathbf{S}}(\mathrm{W}^*, \mathrm{S}^*) H(\mathbf{A}^{[\mathrm{W}^*, \mathrm{S}^*]})}$$

- The **capacity** is the supremum of rates over all PA-PIR-SI protocols for Λ.
- Goal: Derive tight bounds on the capacity of the PA-PIR-SI problem
  - Upper bound (converse)
  - Lower bound (achievability)

Partition-and-Code Scheme (Kadhe *et al.* '17)



1. Assign $X_1, \ldots, X_K$ to disjoint sets, each of size $M+1$.
2. Assign the side info. messages and the demand message to one set.
3. Assign the rest of the messages to the remaining sets at random.
4. Query server for the sum of all messages in each set.

User decodes X_{W} by subtracting M side info. messages X_{S} off of the sum $\sum_{i \in \mathrm{W} \cup \mathrm{S}} X_i$

Download Rate $\dfrac{M+1}{K}$

MDS Code Scheme (Kadhe *et al*.)

1. Choose distinct $\omega_1, \ldots, \omega_K \in \mathbb{F}_q$
2. Given user knows M side-information, query $K - M$ linear combinations of form,

$$K - M \left\{ \underbrace{\begin{bmatrix} \omega_1^0 & \omega_2^0 & \cdots & \omega_K^0 \\ \omega_1^1 & \omega_2^1 & \cdots & \omega_K^1 \\ \vdots & \vdots & \ddots & \vdots \\ \omega_1^{K-M-1} & \omega_2^{K-M-1} & \cdots & \omega_K^{K-M-1} \end{bmatrix}}_{K} \right. \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_K \end{bmatrix}$$

User decodes $X_1, \ldots, X_K$ by subtracting off M side-information from each linear combination and solving resulting system of equations

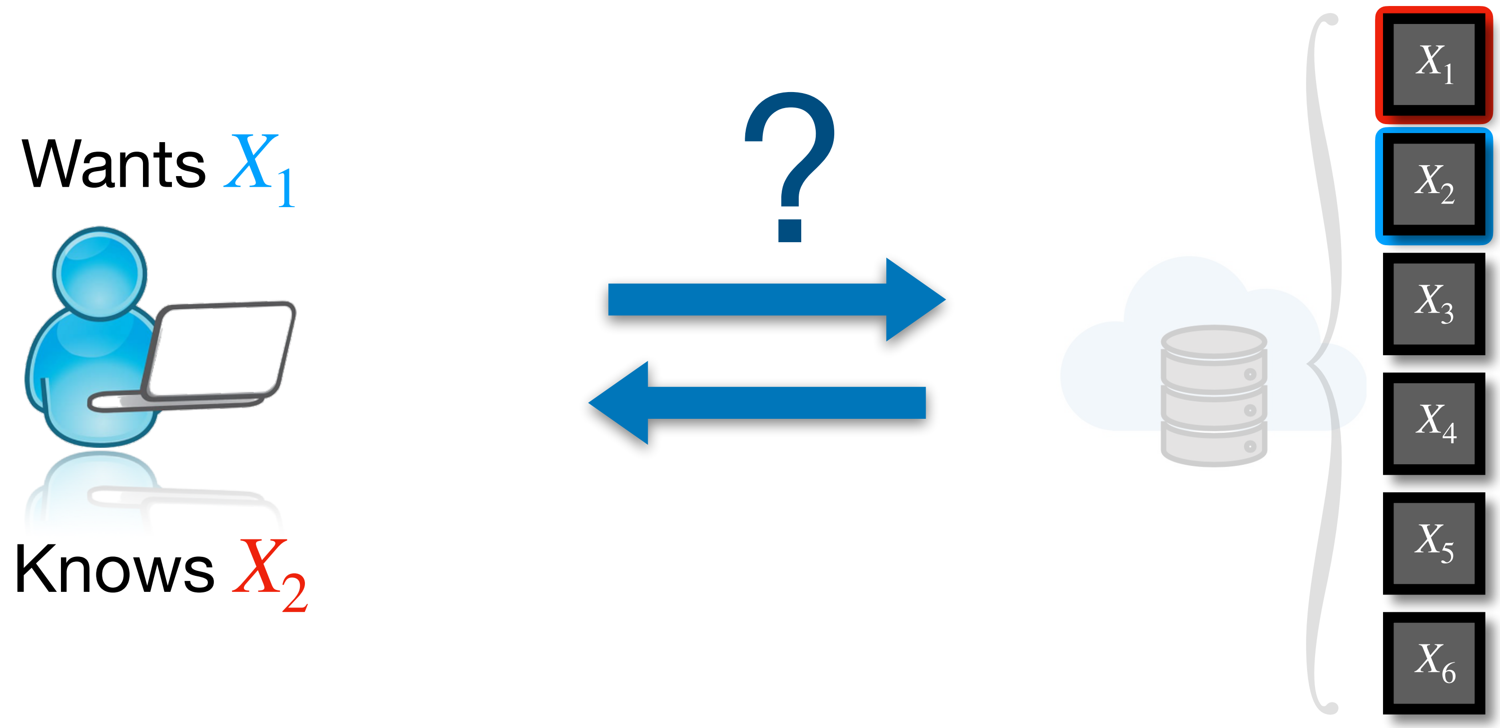
Download Rate $\dfrac{1}{K - M}$

Outline

- Model + Assumptions
- A Motivating Example
- Main Results
- Simulations
- Summary and Open Problems

$$K = 6, \quad \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6$$

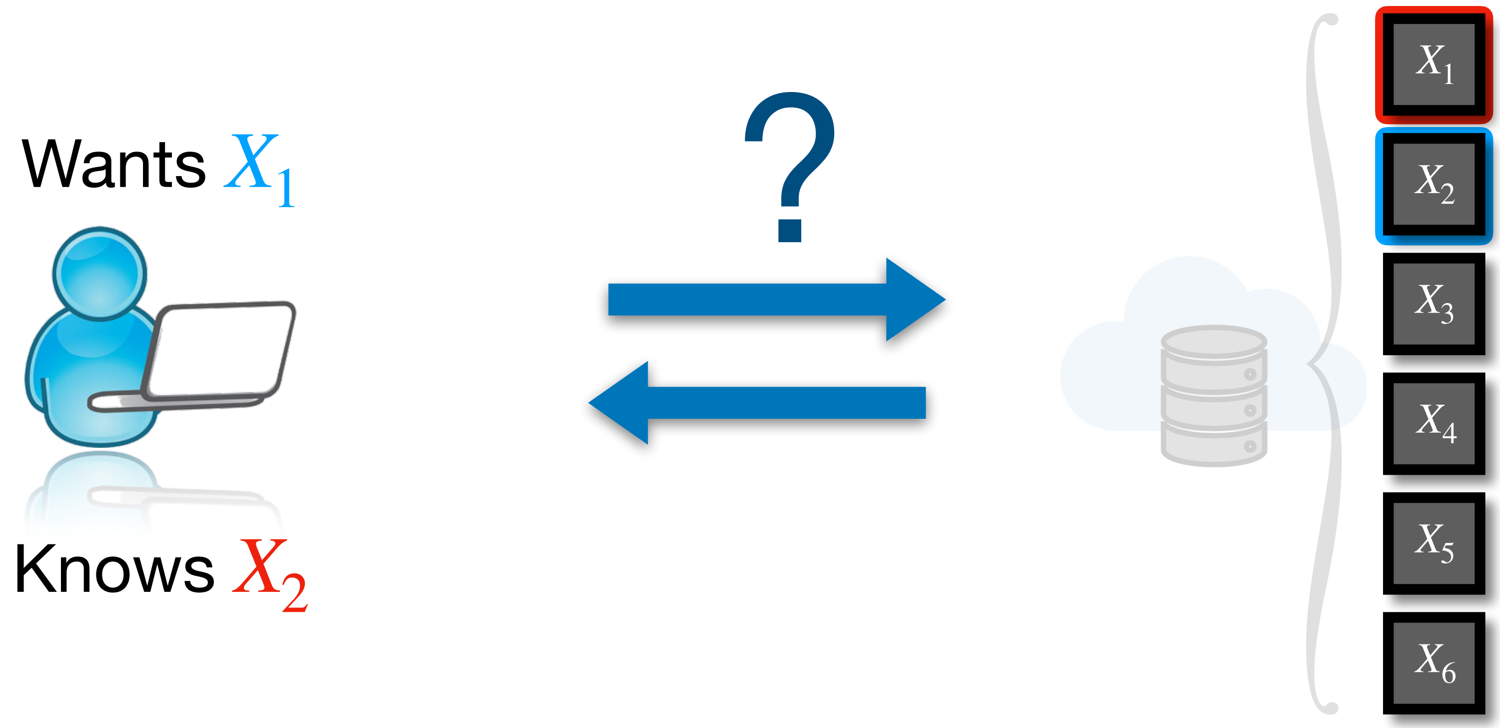
The "uniform popularities" case



What can we do with existing PIR-SI protocols in this setting?

$$K = 6, \quad \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6$$

The “uniform popularities” case



MDS Code Scheme

- Download rate $1/(K - M) = 1/5$
- Decodability satisfied by MDS property.
- Privacy satisfied, same query for all (W, S)

Partition-and-Code Scheme

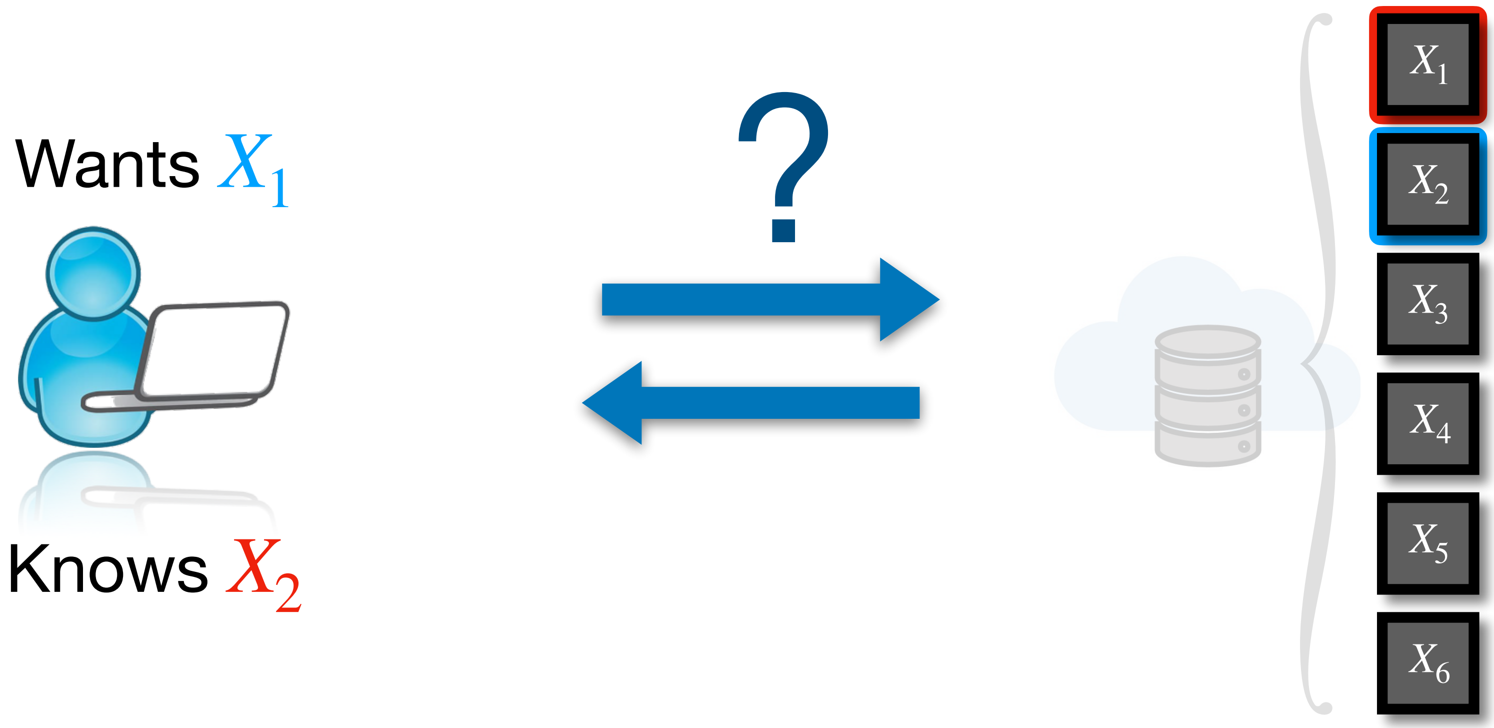
- Download rate $(M + 1)/K = 1/3$
- Decodability satisfied.
- Direct computation shows privacy satisfied.

Optimal scheme in this setting*

Heidarzadeh-Kadhe-Garcia-El Rouayheb-Sprintson, ***Allerton Conf. ‘18***

$$K = 6, \quad \Lambda = (2,1,1,1,1,1)$$

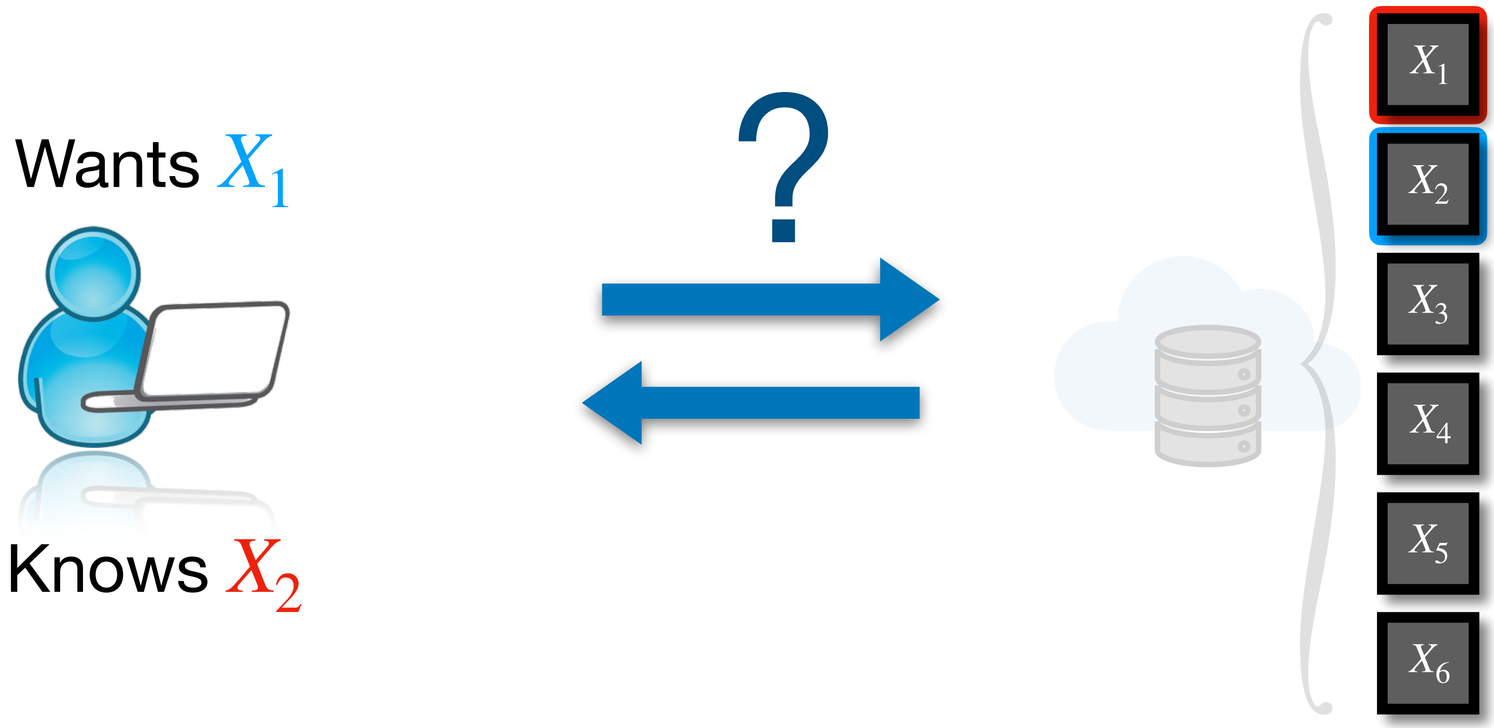
The “non-uniform popularities” case



What can we do with existing PIR-SI protocols in this new setting?

$$K = 6, \quad \Lambda = (2,1,1,1,1,1)$$

The "non-uniform popularities" case



MDS Code Scheme

- Download rate $1/(K - M) = 1/5$
- Decodability satisfied by MDS property.
- Privacy satisfied, same query for all (W, S)

Partition-and-Code Scheme

- Privacy condition does not hold.
- Cannot use this scheme in this setting.
- (Rate is immaterial).

Outline

- Model + Assumptions
- A Motivating Example
- Main Results
- Simulations
- Summary and Open Problems

Capacity of PA-PIR-SI

Theorem 1. *For PA-PIR-SI with K messages and M side info. messages such that $M+1$ is a divisor of K and strictly less than $\sqrt{K}$, under any popularity profile Λ, the capacity is upper bounded by*

$$R_{\text{UB}} = \frac{M+1}{K}$$

and is lower bounded by

$$R_{\text{LB}} = \left(K - M - \left(K - M - \frac{K}{M+1} \right) \times \Gamma_{\{1\},[2:M+1]} \frac{p_{\mathbf{W},\mathbf{S}}(\{1\},[2:M+1])}{p_{\mathbf{W}}(\{1\})} \binom{K-1}{M} \right)^{-1}$$

where

$$\Gamma_{\{1\},[2:M+1]} = \min_{i \in [K-M:K]} \left\{ 1, \frac{p_{\mathbf{W},\mathbf{S}}(\{i\},[K-M:K]\backslash\{i\}) p_{\mathbf{W}}(\{1\})}{p_{\mathbf{W},\mathbf{S}}(\{1\},[2:M+1]) p_{\mathbf{W}}(\{i\})} \right\}.$$

Capacity of PA-PIR-SI

Divisibility Condition

Theorem 1. *For PA-PIR-SI with K messages and M side info. messages such that $M+1$ is a divisor of K and strictly less than $\sqrt{K}$, under any popularity profile Λ, the capacity is upper bounded by*

$$R_{\text{UB}} = \frac{M+1}{K}$$

and is lower bounded by

$$R_{\text{LB}} = \left(K - M - \left(K - M - \frac{K}{M+1} \right) \times \Gamma_{\{1\},[2:M+1]} \frac{p_{\mathbf{W},\mathbf{S}}(\{1\},[2:M+1])}{p_{\mathbf{W}}(\{1\})} \binom{K-1}{M} \right)^{-1}$$

where

$$\Gamma_{\{1\},[2:M+1]} = \min_{i \in [K-M:K]} \left\{ 1, \frac{p_{\mathbf{W},\mathbf{S}}(\{i\},[K-M:K]\backslash\{i\}) p_{\mathbf{W}}(\{1\})}{p_{\mathbf{W},\mathbf{S}}(\{1\},[2:M+1]) p_{\mathbf{W}}(\{i\})} \right\}.$$

Achievability Scheme

Random Code Selection (RCS) Scheme

- Given realization $(\mathrm{W}*, S*)$

$$\Gamma_{\mathrm{W}*,\mathrm{S}*} \triangleq \Gamma_{\{1\},[2:M+1]} \frac{p_{\mathbf{W},\mathbf{S}}(\{1\},[2:M+1])p_{\mathbf{W}}(\mathrm{W}*)}{p_{\mathbf{W},\mathbf{S}}(\mathrm{W}*,\mathrm{S}*)p_{\mathbf{W}}(\{1\})}$$

Wants $X_{\mathrm{W}*}$

Know $X_{\mathrm{S}*}$

Achievability Scheme

Random Code Selection (RCS) Scheme

- Given realization (W^*, S^*)

$$\Gamma_{W^*,S^*} \triangleq \Gamma_{\{1\},[2:M+1]} \frac{p_{\mathbf{W},\mathbf{S}}(\{1\},[2:M+1])p_{\mathbf{W}}(W^*)}{p_{\mathbf{W},\mathbf{S}}(W^*,S^*)p_{\mathbf{W}}(\{1\})}$$

Wants X_{W^*}

Know X_{S^*}

w/ probability Γ_{W^*,S^*}

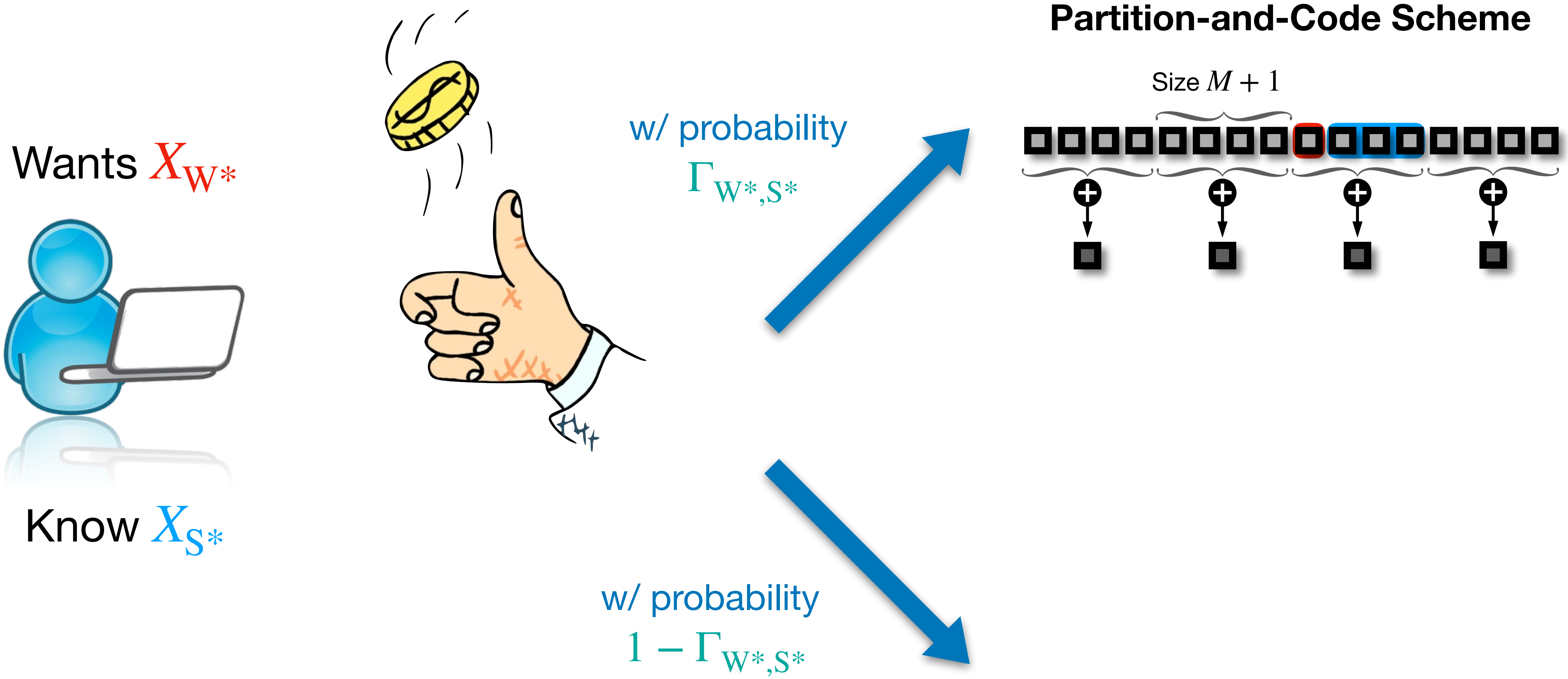
w/ probability $1-\Gamma_{W^*,S^*}$

Achievability Scheme

Random Code Selection (RCS) Scheme

- Given realization $(\mathrm{W}*, \mathrm{S}*)$

$$\Gamma_{\mathrm{W}*,\mathrm{S}*} \triangleq \Gamma_{\{1\},[2:M+1]} \frac{p_{\mathbf{W},\mathbf{S}}(\{1\},[2:M+1])p_{\mathbf{W}}(\mathrm{W}*)}{p_{\mathbf{W},\mathbf{S}}(\mathrm{W}*,\mathrm{S}*)p_{\mathbf{W}}(\{1\})}$$

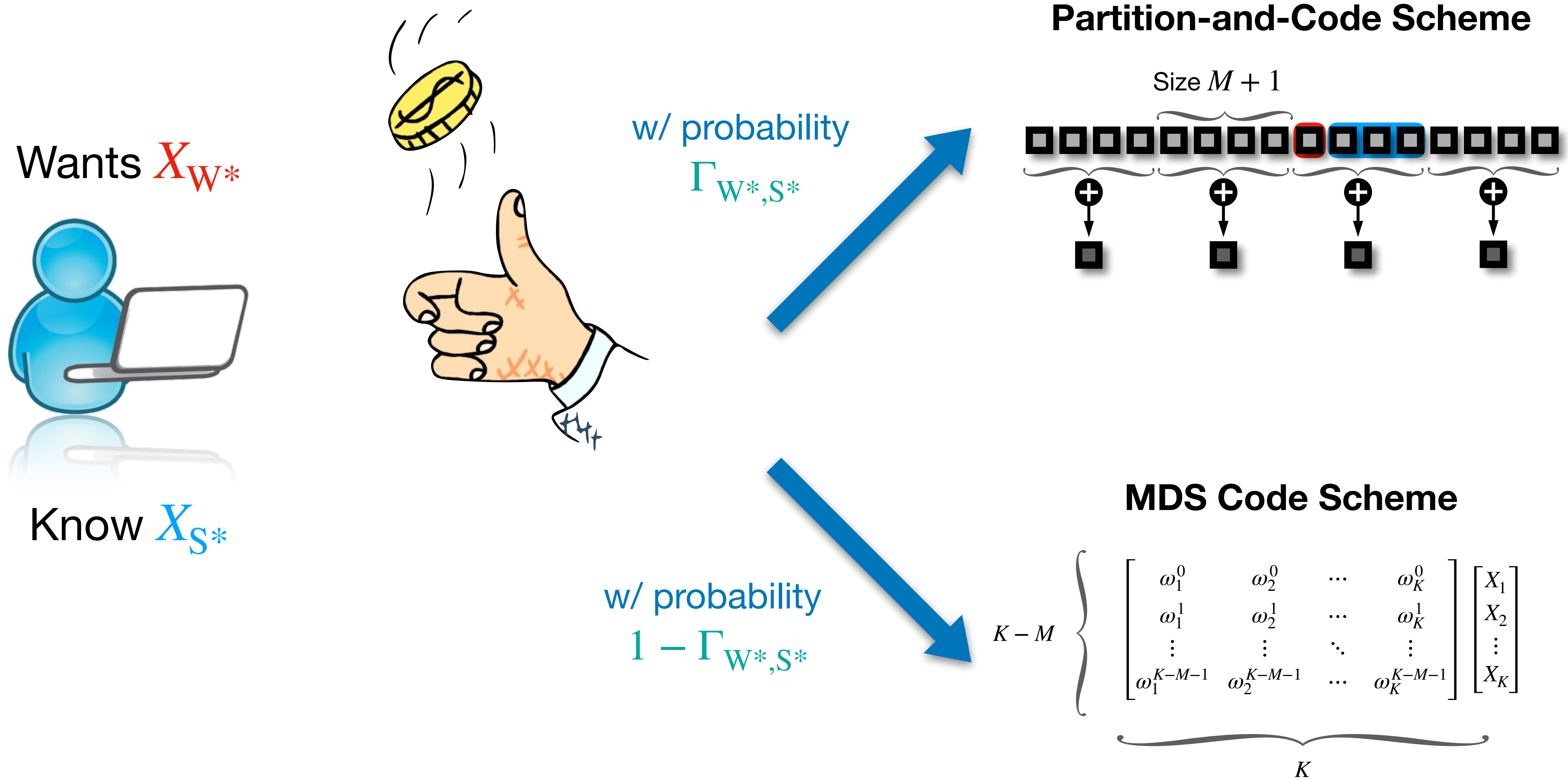


Achievability Scheme

Random Code Selection (RCS) Scheme

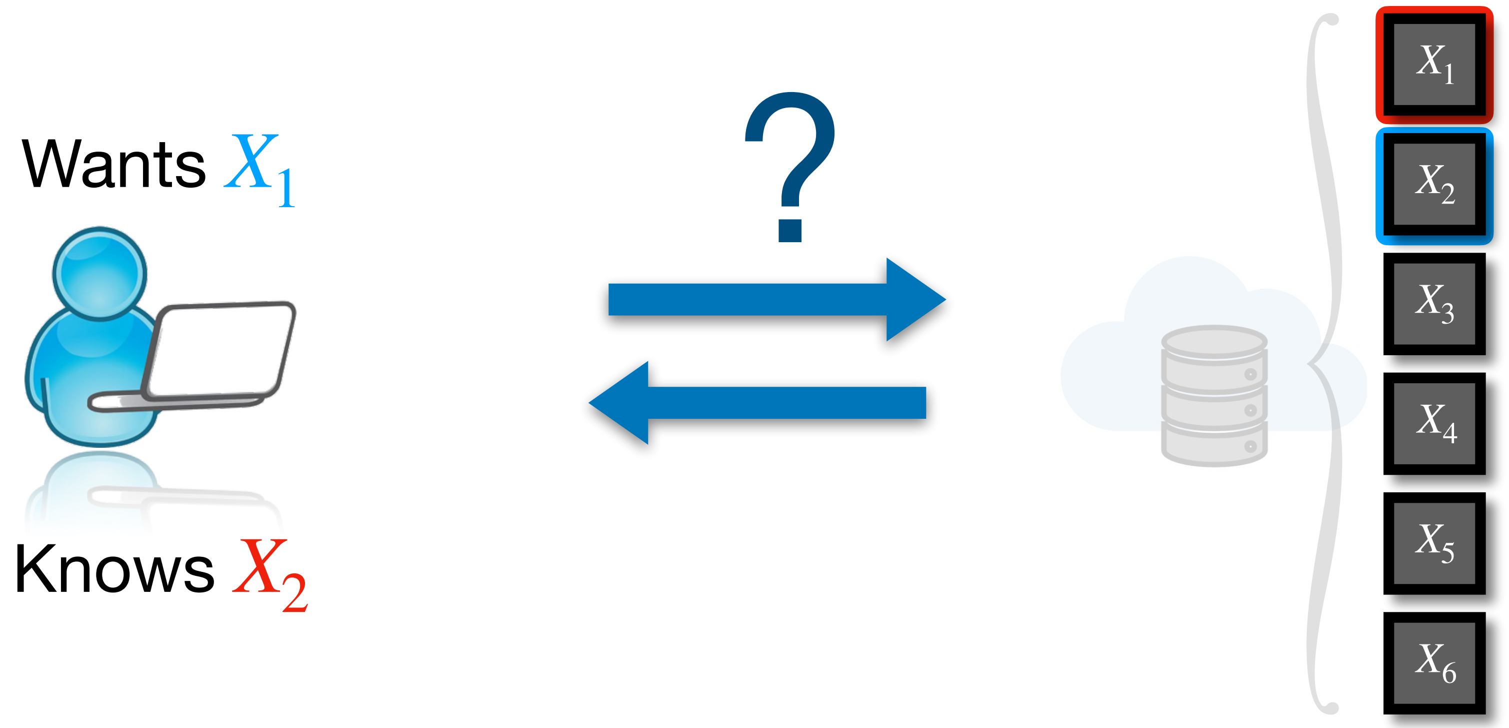
- Given realization (W^*, S^*)

$$\Gamma_{\mathrm{W}^*,\mathrm{S}^*} \triangleq \Gamma_{\{1\},[2:M+1]} \frac{p_{\mathrm{W},\mathrm{S}}(\{1\},[2:M+1])\,p_{\mathrm{W}}(\mathrm{W}^*)}{p_{\mathrm{W},\mathrm{S}}(\mathrm{W}^*,\mathrm{S}^*)\,p_{\mathrm{W}}(\{1\})}$$



$K = 6, \quad \Lambda = (2,1,1,1,1,1)$

The "non-uniform popularities" case



MDS Code Scheme

- Download rate $1/(K-M) = 1/5$
- Decodability and privacy satisfied.

Partition-and-Code Scheme

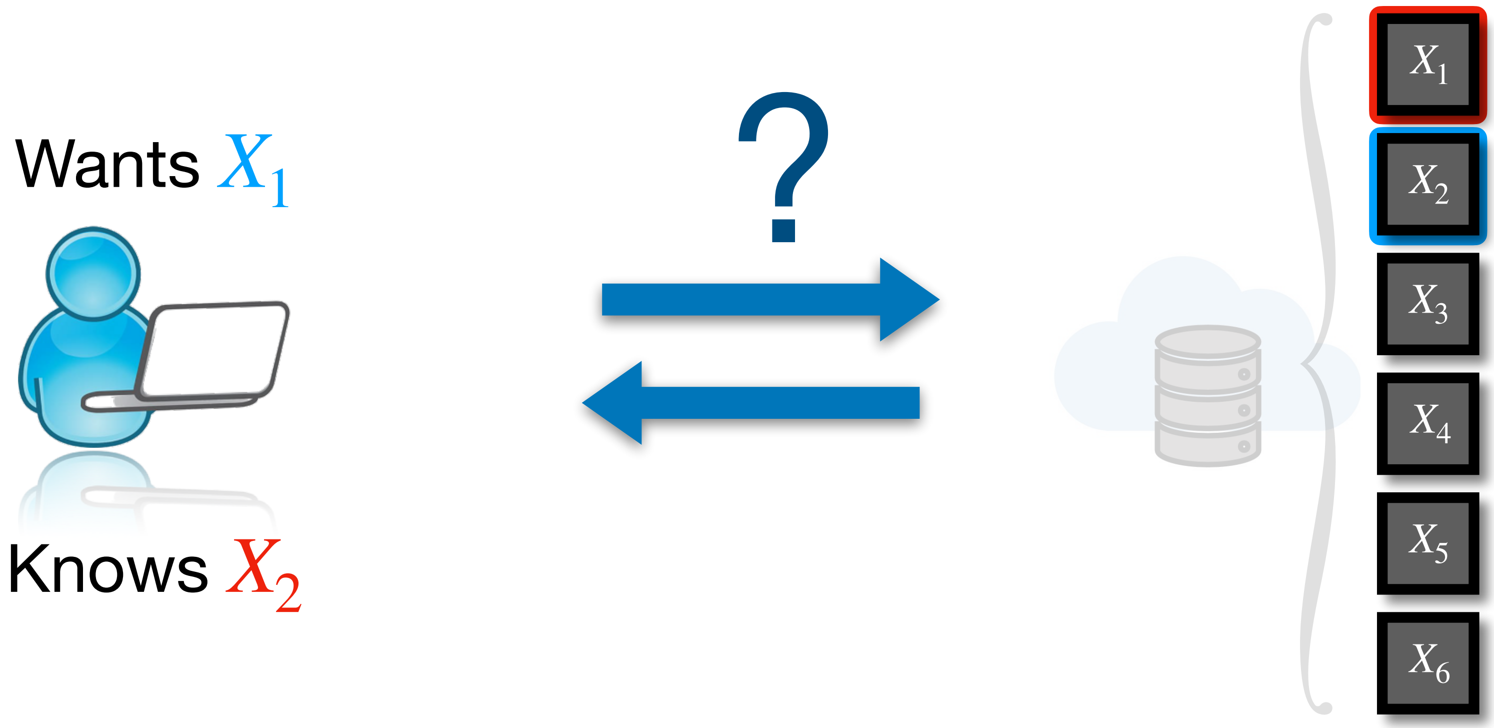
- Download rate N/A
- Cannot use this scheme in this setting.

RCS Scheme

- **Download rate** $13/40$ $(> 1/5)$
- **Decodability and privacy satisfied.**

$$K = 6, \quad \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6$$

The "uniform popularities" case



MDS Code Scheme

- Download rate $1/(K - M) = 1/5$
- Decodability and privacy satisfied.

Partition-and-Code Scheme

- Download rate $(M + 1)/K = 1/3$
- Decodability and privacy satisfied.

RCS Scheme

- **Download rate** $1/3$
- **Decodability and privacy satisfied.**

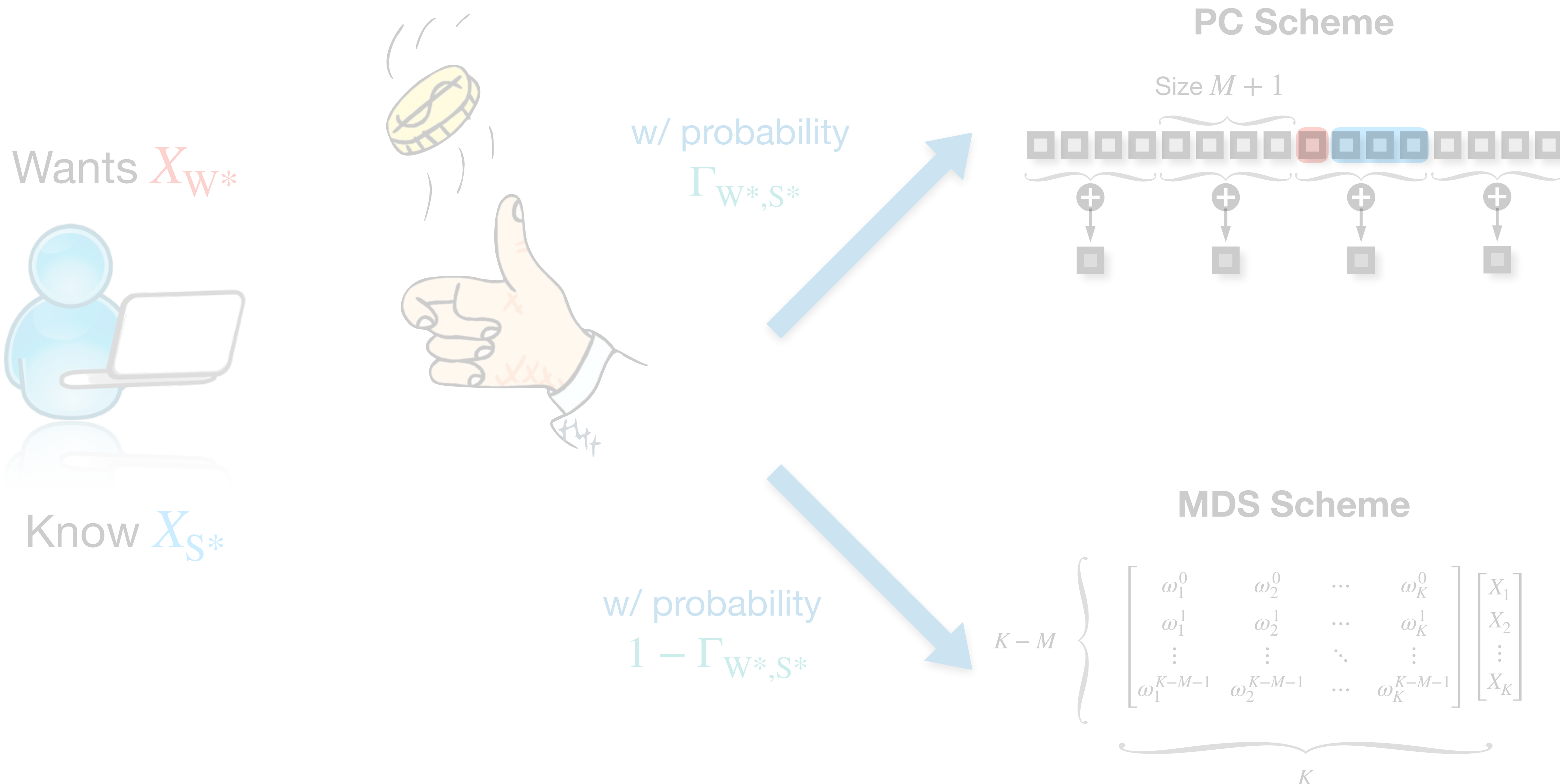
Achievability Scheme

Random Code Selection (RCS) Scheme

- Given realization $(\mathrm{W}*, \mathrm{S}*)$

$$\Gamma_{\mathrm{W}*,\mathrm{S}*} \triangleq \Gamma_{\{1\},[2:M+1]} \frac{p_{\mathbf{W},\mathbf{S}}(\{1\},[2:M+1])p_{\mathbf{W}}(\mathrm{W}*)}{p_{\mathbf{W},\mathbf{S}}(\mathrm{W}*,\mathrm{S}*)p_{\mathbf{W}}(\{1\})},$$

Where does this choice come from?



The $M = 1$ Case

- Consider parameters $\Gamma_{i,j}$ for each pair $(i, j) \in \mathscr{K} \times \mathscr{K}, i \neq j$
- Given (i, j), follow Partition-and-Code Scheme w.p. $\Gamma_{i,j}$ or MDS Scheme w.p $1 - \Gamma_{i,j}$
- Want to maximize RCS rate …

$$\overbrace{\left(\sum_{i,j \in \mathscr{K} \times \mathscr{K}, i \neq j} p_{\mathbf{w},\mathbf{s}}(i,j) \times \left[\underbrace{\Gamma_{i,j} \left(\frac{K}{M+1} \right)}_{\text{1/rate of Partition-and-Code Scheme}} + \underbrace{(1 - \Gamma_{i,j})(K - M)}_{\text{1/rate of MDS}} \right] \right)^{-1}}^{\mathbb{E}_{\sim(\mathbf{w},\mathbf{s})}[\,\cdot\,]}$$

… subject to privacy condition.

The $M = 1$ Case

Maximize

$$\overbrace{\left(\sum_{i,j \in \mathcal{K} \times \mathcal{K}, i \neq j} p_{\mathbf{W},\mathbf{S}}(i,j) \times \left[\underbrace{\Gamma_{i,j} \left(\frac{K}{M+1} \right)}_{\text{1/rate of Partition-and-Code Scheme}} + \underbrace{(1 - \Gamma_{i,j})(K - M)}_{\text{1/rate of MDS}} \right] \right)^{-1}}^{\mathbb{E}_{\sim(\mathbf{W},\mathbf{S})}[\,\cdot\,]}$$

s.t. $\mathbb{P}(\mathbf{W} = i \,|\, \mathbf{Q} = \mathrm{Q}) = \mathbb{P}(\mathbf{W} = i) \quad \forall i \in \mathcal{K}$.

The $M = 1$ Case

Maximize

$$\overbrace{\left(\sum_{i,j \in \mathcal{K} \times \mathcal{K}, i \neq j} p_{\mathbf{W},\mathbf{S}}(i,j) \times \left[\underbrace{\Gamma_{i,j} \left(\frac{K}{M+1} \right)} + \underbrace{(1 - \Gamma_{i,j})(K - M)} \right] \right)^{-1}}^{\mathbb{E}_{\sim(\mathbf{W},\mathbf{S})}[\,\cdot\,]}$$

1/rate of Partition-and-Code Scheme

1/rate of MDS

s.t. $\mathbb{P}(\mathbf{W} = i \,|\, \mathbf{Q} = \mathrm{Q}) = \mathbb{P}(\mathbf{W} = i) \quad \forall i \in \mathcal{K}$.

Is this optimization over $K^2 - K$ variables?

The $M = 1$ Case

Maximize

$$\overbrace{\left(\sum_{i,j \in \mathcal{K} \times \mathcal{K}, i \neq j} p_{\mathbf{W},\mathbf{S}}(i,j) \times \left[\underbrace{\Gamma_{i,j} \left(\frac{K}{M+1} \right)}_{\text{1/rate of Partition-and-Code Scheme}} + \underbrace{(1 - \Gamma_{i,j})(K - M)}_{\text{1/rate of MDS}} \right] \right)^{-1}}^{\mathbb{E}_{\sim(\mathbf{W},\mathbf{S})}[\,\cdot\,]}$$

s.t. $\mathbb{P}(\mathbf{W} = i \,|\, \mathbf{Q} = \mathrm{Q}) = \mathbb{P}(\mathbf{W} = i) \quad \forall i \in \mathcal{K}\,.$

Is this optimization over $K^2 - K$ variables?

No; the privacy condition allows us to reduce the problem to a single variable.

The $M = 1$, $K = 6$ Case, Enumerating Queries

- Recall MDS query is the same for all (W, S)
- Only Partition-and-Code Scheme queries have dependence on (W, S)
- We can represent each Partition-and-Code Scheme query as a partition.
- For example,

{1,2}
{3,4}
{5,6}

describes any query requesting

$$X_1 + X_2, \quad X_3 + X_4, \quad X_5 + X_6$$

The $M = 1$, $K = 6$ Case, Enumerating Queries

All 15 possible queries

Q_1: {1,2} {3,4} {5,6}	Q_2: {1,2} {3,6} {4,5}	Q_3: {1,2} {3,5} {4,6}
Q_4: {1,3} {2,4} {5,6}	Q_5: {1,3} {2,6} {4,5}	Q_6: {1,3} {2,5} {4,6}
Q_7: {1,4} {2,3} {5,6}	Q_8: {1,4} {2,6} {3,5}	Q_9: {1,4} {2,5} {3,6}
Q_{10}: {1,5} {2,3} {4,6}	Q_{11}: {1,5} {2,6} {3,4}	Q_{12}: {1,5} {2,4} {3,6}
Q_{13}: {1,6} {2,3} {4,5}	Q_{14}: {1,6} {2,5} {3,4}	Q_{15}: {1,6} {2,4} {5,3}

The $M = 1$, $K = 6$ Case, Enumerating Queries

All 15 possible queries

Q_1: {1,2} {3,4} {5,6}	Q_2: {1,2} {3,6} {4,5}	Q_3: {1,2} {3,5} {4,6}
Q_4: {1,3} {2,4} {5,6}	Q_5: {1,3} {2,6} {4,5}	Q_6: {1,3} {2,5} {4,6}
Q_7: {1,4} {2,3} {5,6}	Q_8: {1,4} {2,6} {3,5}	Q_9: {1,4} {2,5} {3,6}
Q_{10}: {1,5} {2,3} {4,6}	Q_{11}: {1,5} {2,6} {3,4}	Q_{12}: {1,5} {2,4} {3,6}
Q_{13}: {1,6} {2,3} {4,5}	Q_{14}: {1,6} {2,5} {3,4}	Q_{15}: {1,6} {2,4} {5,3}

Privacy condition yields some useful identities.

The $M = 1, K = 6$ Case, Enumerating Queries

All 15 possible queries

Q_1 {1,2} {3,4} {5,6}	Q_2 {1,2} {3,6} {4,5}	Q_3 {1,2} {3,5} {4,6}
Q_4 {1,3} {2,4} {5,6}	Q_5 {1,3} {2,6} {4,5}	Q_6 {1,3} {2,5} {4,6}
Q_7 {1,4} {2,3} {5,6}	Q_8 {1,4} {2,6} {3,5}	Q_9 {1,4} {2,5} {3,6}
Q_{10} {1,5} {2,3} {4,6}	Q_{11} {1,5} {2,6} {3,4}	Q_{12} {1,5} {2,4} {3,6}
Q_{13} {1,6} {2,3} {4,5}	Q_{14} {1,6} {2,5} {3,4}	Q_{15} {1,6} {2,4} {5,3}

Privacy condition yields some useful identities.

$$\mathbb{P}(\mathrm{W} = 1 \mid Q_1)$$

The $M = 1, K = 6$ Case, Enumerating Queries

All 15 possible queries

Q_1: {1,2} {3,4} {5,6}	Q_2: {1,2} {3,6} {4,5}	Q_3: {1,2} {3,5} {4,6}
Q_4: {1,3} {2,4} {5,6}	Q_5: {1,3} {2,6} {4,5}	Q_6: {1,3} {2,5} {4,6}
Q_7: {1,4} {2,3} {5,6}	Q_8: {1,4} {2,6} {3,5}	Q_9: {1,4} {2,5} {3,6}
Q_{10}: {1,5} {2,3} {4,6}	Q_{11}: {1,5} {2,6} {3,4}	Q_{12}: {1,5} {2,4} {3,6}
Q_{13}: {1,6} {2,3} {4,5}	Q_{14}: {1,6} {2,5} {3,4}	Q_{15}: {1,6} {2,4} {5,3}

Privacy condition yields some useful identities.

$$\mathbb{P}(\mathrm{W} = 1 \mid Q_1)$$

$$= \frac{\mathbb{P}(Q_1 \mid \mathrm{W} = 1, \mathrm{S} = 2) \times \mathbb{P}(\mathrm{W} = 1, \mathrm{S} = 2)}{\mathbb{P}(Q_1)}$$

The $M = 1$, $K = 6$ Case, Enumerating Queries

All 15 possible queries

Q_1: {1,2} {3,4} {5,6}	Q_2: {1,2} {3,6} {4,5}	Q_3: {1,2} {3,5} {4,6}
Q_4: {1,3} {2,4} {5,6}	Q_5: {1,3} {2,6} {4,5}	Q_6: {1,3} {2,5} {4,6}
Q_7: {1,4} {2,3} {5,6}	Q_8: {1,4} {2,6} {3,5}	Q_9: {1,4} {2,5} {3,6}
Q_{10}: {1,5} {2,3} {4,6}	Q_{11}: {1,5} {2,6} {3,4}	Q_{12}: {1,5} {2,4} {3,6}
Q_{13}: {1,6} {2,3} {4,5}	Q_{14}: {1,6} {2,5} {3,4}	Q_{15}: {1,6} {2,4} {5,3}

Privacy condition yields some useful identities.

$$\mathbb{P}(\mathrm{W} = 1 \mid Q_1)$$

$$= \frac{\mathbb{P}(Q_1 \mid \mathrm{W} = 1, \mathrm{S} = 2) \times \mathbb{P}(\mathrm{W} = 1, \mathrm{S} = 2)}{\mathbb{P}(Q_1)}$$

$$= \frac{\Gamma_{1,2} \times \frac{1}{L} \times \mathbb{P}(\mathrm{W} = 1, \mathrm{S} = 2)}{\mathbb{P}(Q_1)}$$

The $M = 1$, $K = 6$ Case, Enumerating Queries

All 15 possible queries $L = 3$

Q_1: {1,2} {3,4} {5,6}	Q_2: {1,2} {3,6} {4,5}	Q_3: {1,2} {3,5} {4,6}
Q_4: {1,3} {2,4} {5,6}	Q_5: {1,3} {2,6} {4,5}	Q_6: {1,3} {2,5} {4,6}
Q_7: {1,4} {2,3} {5,6}	Q_8: {1,4} {2,6} {3,5}	Q_9: {1,4} {2,5} {3,6}
Q_{10}: {1,5} {2,3} {4,6}	Q_{11}: {1,5} {2,6} {3,4}	Q_{12}: {1,5} {2,4} {3,6}
Q_{13}: {1,6} {2,3} {4,5}	Q_{14}: {1,6} {2,5} {3,4}	Q_{15}: {1,6} {2,4} {5,3}

Privacy condition yields some useful identities.

$$\mathbb{P}(\mathrm{W} = 1 \mid Q_1)$$

$$= \frac{\mathbb{P}(Q_1 \mid \mathrm{W} = 1, \mathrm{S} = 2) \times \mathbb{P}(\mathrm{W} = 1, \mathrm{S} = 2)}{\mathbb{P}(Q_1)}$$

$$= \frac{\Gamma_{1,2} \times \frac{1}{L} \times \mathbb{P}(\mathrm{W} = 1, \mathrm{S} = 2)}{\mathbb{P}(Q_1)}$$

The $M = 1$, $K = 6$ Case, Enumerating Queries

All 15 possible queries

Q_1: {1,2} {3,4} {5,6}	Q_2: {1,2} {3,6} {4,5}	Q_3: {1,2} {3,5} {4,6}
Q_4: {1,3} {2,4} {5,6}	Q_5: {1,3} {2,6} {4,5}	Q_6: {1,3} {2,5} {4,6}
Q_7: {1,4} {2,3} {5,6}	Q_8: {1,4} {2,6} {3,5}	Q_9: {1,4} {2,5} {3,6}
Q_{10}: {1,5} {2,3} {4,6}	Q_{11}: {1,5} {2,6} {3,4}	Q_{12}: {1,5} {2,4} {3,6}
Q_{13}: {1,6} {2,3} {4,5}	Q_{14}: {1,6} {2,5} {3,4}	Q_{15}: {1,6} {2,4} {5,3}

Privacy condition yields some useful identities.

$$\mathbb{P}(\mathrm{W} = 1 \mid Q_1)$$

$$= \frac{\mathbb{P}(Q_1 \mid \mathrm{W} = 1, \mathrm{S} = 2) \times \mathbb{P}(\mathrm{W} = 1, \mathrm{S} = 2)}{\mathbb{P}(Q_1)}$$

$$= \frac{\Gamma_{1,2} \times \frac{1}{L} \times \mathbb{P}(\mathrm{W} = 1, \mathrm{S} = 2)}{\mathbb{P}(Q_1)}$$

$$= \mathbb{P}(\mathrm{W} = 1)$$

The $M = 1$, $K = 6$ Case, Enumerating Queries

All 15 possible queries

Q_1: {1,2} {3,4} {5,6}	Q_2: {1,2} {3,6} {4,5}	Q_3: {1,2} {3,5} {4,6}
Q_4: {1,3} {2,4} {5,6}	Q_5: {1,3} {2,6} {4,5}	Q_6: {1,3} {2,5} {4,6}
Q_7: {1,4} {2,3} {5,6}	Q_8: {1,4} {2,6} {3,5}	Q_9: {1,4} {2,5} {3,6}
Q_{10}: {1,5} {2,3} {4,6}	Q_{11}: {1,5} {2,6} {3,4}	Q_{12}: {1,5} {2,4} {3,6}
Q_{13}: {1,6} {2,3} {4,5}	Q_{14}: {1,6} {2,5} {3,4}	Q_{15}: {1,6} {2,4} {5,3}

Privacy condition yields some useful identities.

$$\mathbb{P}(\mathrm{W} = 1 \mid Q_1)$$

$$= \frac{\mathbb{P}(Q_1 \mid \mathrm{W} = 1, \mathrm{S} = 2) \times \mathbb{P}(\mathrm{W} = 1, \mathrm{S} = 2)}{\mathbb{P}(Q_1)}$$

$$= \frac{\Gamma_{1,2} \times \frac{1}{L} \times \mathbb{P}(\mathrm{W} = 1, \mathrm{S} = 2)}{\mathbb{P}(Q_1)}$$

$$= \mathbb{P}(\mathrm{W} = 1)$$

$$\Longrightarrow \mathbb{P}(Q_1) = \frac{\Gamma_{1,2} \times \frac{1}{L} \times \mathbb{P}(\mathrm{W} = 1, \mathrm{S} = 2)}{\mathbb{P}(\mathrm{W} = 1)}$$

The $M = 1, K = 6$ Case, Enumerating Queries

All 15 possible queries

Q_1 {1,2} {3,4} {5,6}	Q_2 {1,2} {3,6} {4,5}	Q_3 {1,2} {3,5} {4,6}
Q_4 {1,3} {2,4} {5,6}	Q_5 {1,3} {2,6} {4,5}	Q_6 {1,3} {2,5} {4,6}
Q_7 {1,4} {2,3} {5,6}	Q_8 {1,4} {2,6} {3,5}	Q_9 {1,4} {2,5} {3,6}
Q_{10} {1,5} {2,3} {4,6}	Q_{11} {1,5} {2,6} {3,4}	Q_{12} {1,5} {2,4} {3,6}
Q_{13} {1,6} {2,3} {4,5}	Q_{14} {1,6} {2,5} {3,4}	Q_{15} {1,6} {2,4} {5,3}

Privacy condition yields some useful identities.

$$\mathbb{P}(\mathrm{W} = 2 \mid Q_1)$$

The $M = 1$, $K = 6$ Case, Enumerating Queries

All 15 possible queries

Q_1: {1,2} {3,4} {5,6}	Q_2: {1,2} {3,6} {4,5}	Q_3: {1,2} {3,5} {4,6}
Q_4: {1,3} {2,4} {5,6}	Q_5: {1,3} {2,6} {4,5}	Q_6: {1,3} {2,5} {4,6}
Q_7: {1,4} {2,3} {5,6}	Q_8: {1,4} {2,6} {3,5}	Q_9: {1,4} {2,5} {3,6}
Q_{10}: {1,5} {2,3} {4,6}	Q_{11}: {1,5} {2,6} {3,4}	Q_{12}: {1,5} {2,4} {3,6}
Q_{13}: {1,6} {2,3} {4,5}	Q_{14}: {1,6} {2,5} {3,4}	Q_{15}: {1,6} {2,4} {5,3}

Privacy condition yields some useful identities.

$$\mathbb{P}(\mathrm{W} = 2 \mid Q_1)$$

$$= \frac{\mathbb{P}(Q_1 \mid \mathrm{W} = 2, \mathrm{S} = 1) \times \mathbb{P}(\mathrm{W} = 2, \mathrm{S} = 1)}{\mathbb{P}(Q_1)}$$

The $M = 1$, $K = 6$ Case, Enumerating Queries

All 15 possible queries

Q_1: {1,2} {3,4} {5,6}	Q_2: {1,2} {3,6} {4,5}	Q_3: {1,2} {3,5} {4,6}
Q_4: {1,3} {2,4} {5,6}	Q_5: {1,3} {2,6} {4,5}	Q_6: {1,3} {2,5} {4,6}
Q_7: {1,4} {2,3} {5,6}	Q_8: {1,4} {2,6} {3,5}	Q_9: {1,4} {2,5} {3,6}
Q_{10}: {1,5} {2,3} {4,6}	Q_{11}: {1,5} {2,6} {3,4}	Q_{12}: {1,5} {2,4} {3,6}
Q_{13}: {1,6} {2,3} {4,5}	Q_{14}: {1,6} {2,5} {3,4}	Q_{15}: {1,6} {2,4} {5,3}

Privacy condition yields some useful identities.

$$\mathbb{P}(\mathrm{W} = 2 \mid Q_1)$$

$$= \frac{\mathbb{P}(Q_1 \mid \mathrm{W} = 2, \mathrm{S} = 1) \times \mathbb{P}(\mathrm{W} = 2, \mathrm{S} = 1)}{\mathbb{P}(Q_1)}$$

$$= \frac{\Gamma_{2,1} \times \frac{1}{L} \times \mathbb{P}(\mathrm{W} = 2, \mathrm{S} = 1)}{\mathbb{P}(Q_1)}$$

The $M = 1, K = 6$ Case, Enumerating Queries

All 15 possible queries

Q_1: {1,2} {3,4} {5,6}	Q_2: {1,2} {3,6} {4,5}	Q_3: {1,2} {3,5} {4,6}
Q_4: {1,3} {2,4} {5,6}	Q_5: {1,3} {2,6} {4,5}	Q_6: {1,3} {2,5} {4,6}
Q_7: {1,4} {2,3} {5,6}	Q_8: {1,4} {2,6} {3,5}	Q_9: {1,4} {2,5} {3,6}
Q_{10}: {1,5} {2,3} {4,6}	Q_{11}: {1,5} {2,6} {3,4}	Q_{12}: {1,5} {2,4} {3,6}
Q_{13}: {1,6} {2,3} {4,5}	Q_{14}: {1,6} {2,5} {3,4}	Q_{15}: {1,6} {2,4} {5,3}

Privacy condition yields some useful identities.

$$\mathbb{P}(\mathrm{W} = 2 \mid Q_1)$$

$$= \frac{\mathbb{P}(Q_1 \mid \mathrm{W} = 2, \mathrm{S} = 1) \times \mathbb{P}(\mathrm{W} = 2, \mathrm{S} = 1)}{\mathbb{P}(Q_1)}$$

$$= \frac{\Gamma_{2,1} \times \frac{1}{L} \times \mathbb{P}(\mathrm{W} = 2, \mathrm{S} = 1)}{\mathbb{P}(Q_1)}$$

The $M = 1$, $K = 6$ Case, Enumerating Queries

All 15 possible queries

Q_1: {1,2} {3,4} {5,6}	Q_2: {1,2} {3,6} {4,5}	Q_3: {1,2} {3,5} {4,6}
Q_4: {1,3} {2,4} {5,6}	Q_5: {1,3} {2,6} {4,5}	Q_6: {1,3} {2,5} {4,6}
Q_7: {1,4} {2,3} {5,6}	Q_8: {1,4} {2,6} {3,5}	Q_9: {1,4} {2,5} {3,6}
Q_{10}: {1,5} {2,3} {4,6}	Q_{11}: {1,5} {2,6} {3,4}	Q_{12}: {1,5} {2,4} {3,6}
Q_{13}: {1,6} {2,3} {4,5}	Q_{14}: {1,6} {2,5} {3,4}	Q_{15}: {1,6} {2,4} {5,3}

Privacy condition yields some useful identities.

$$\mathbb{P}(\mathrm{W} = 2 \mid Q_1)$$

$$= \frac{\mathbb{P}(Q_1 \mid \mathrm{W} = 2, \mathrm{S} = 1) \times \mathbb{P}(\mathrm{W} = 2, \mathrm{S} = 1)}{\mathbb{P}(Q_1)}$$

$$= \frac{\Gamma_{2,1} \times \frac{1}{L} \times \mathbb{P}(\mathrm{W} = 2, \mathrm{S} = 1)}{\mathbb{P}(Q_1)}$$

$$= \mathbb{P}(\mathrm{W} = 2)$$

The $M = 1$, $K = 6$ Case, Enumerating Queries

All 15 possible queries

Q_1: {1,2} {3,4} {5,6}	Q_2: {1,2} {3,6} {4,5}	Q_3: {1,2} {3,5} {4,6}
Q_4: {1,3} {2,4} {5,6}	Q_5: {1,3} {2,6} {4,5}	Q_6: {1,3} {2,5} {4,6}
Q_7: {1,4} {2,3} {5,6}	Q_8: {1,4} {2,6} {3,5}	Q_9: {1,4} {2,5} {3,6}
Q_{10}: {1,5} {2,3} {4,6}	Q_{11}: {1,5} {2,6} {3,4}	Q_{12}: {1,5} {2,4} {3,6}
Q_{13}: {1,6} {2,3} {4,5}	Q_{14}: {1,6} {2,5} {3,4}	Q_{15}: {1,6} {2,4} {5,3}

Privacy condition yields some useful identities.

$$\mathbb{P}(\mathrm{W} = 2 \mid Q_1)$$

$$= \frac{\mathbb{P}(Q_1 \mid \mathrm{W} = 2, \mathrm{S} = 1) \times \mathbb{P}(\mathrm{W} = 2, \mathrm{S} = 1)}{\mathbb{P}(Q_1)}$$

$$= \frac{\Gamma_{2,1} \times \frac{1}{L} \times \mathbb{P}(\mathrm{W} = 2, \mathrm{S} = 1)}{\mathbb{P}(Q_1)}$$

$$= \mathbb{P}(\mathrm{W} = 2)$$

$$\Longrightarrow \mathbb{P}(Q_1) = \frac{\Gamma_{2,1} \times \frac{1}{L} \times \mathbb{P}(\mathrm{W} = 2, \mathrm{S} = 1)}{\mathbb{P}(\mathrm{W} = 2)}$$

The $M = 1$, $K = 6$ Case, Enumerating Queries

All 15 possible queries

Q_1: {1,2} {3,4} {5,6}	Q_2: {1,2} {3,6} {4,5}	Q_3: {1,2} {3,5} {4,6}
Q_4: {1,3} {2,4} {5,6}	Q_5: {1,3} {2,6} {4,5}	Q_6: {1,3} {2,5} {4,6}
Q_7: {1,4} {2,3} {5,6}	Q_8: {1,4} {2,6} {3,5}	Q_9: {1,4} {2,5} {3,6}
Q_{10}: {1,5} {2,3} {4,6}	Q_{11}: {1,5} {2,6} {3,4}	Q_{12}: {1,5} {2,4} {3,6}
Q_{13}: {1,6} {2,3} {4,5}	Q_{14}: {1,6} {2,5} {3,4}	Q_{15}: {1,6} {2,4} {5,3}

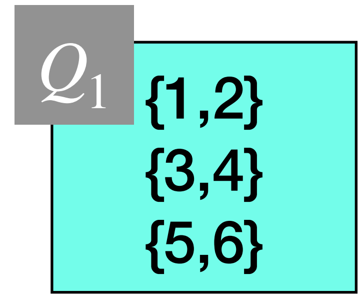
Privacy condition yields some useful identities.

$$\mathbb{P}(Q_1)$$

$$= \frac{\Gamma_{1,2} \times \cancel{\frac{1}{L}} \times \mathbb{P}(\mathrm{W} = 1, \mathrm{S} = 2)}{\mathbb{P}(\mathrm{W} = 1)}$$

$$= \frac{\Gamma_{2,1} \times \cancel{\frac{1}{L}} \times \mathbb{P}(\mathrm{W} = 2, \mathrm{S} = 1)}{\mathbb{P}(\mathrm{W} = 2)}$$

The $M = 1$, $K = 6$ Case, Identities



$$\mathbb{P}(Q_1) = \frac{\Gamma_{1,2} \times p_{\mathbf{W},\mathbf{S}}(1,2)}{p_{\mathbf{W}}(1)} = \frac{\Gamma_{2,1} \times p_{\mathbf{W},\mathbf{S}}(2,1)}{p_{\mathbf{W}}(2)} = \frac{\Gamma_{3,4} \times p_{\mathbf{W},\mathbf{S}}(3,4)}{p_{\mathbf{W}}(3)} = \frac{\Gamma_{4,3} \times p_{\mathbf{W},\mathbf{S}}(4,3)}{p_{\mathbf{W}}(4)} = \frac{\Gamma_{5,6} \times p_{\mathbf{W},\mathbf{S}}(5,6)}{p_{\mathbf{W}}(5)} = \frac{\Gamma_{6,5} \times p_{\mathbf{W},\mathbf{S}}(6,5)}{p_{\mathbf{W}}(6)}$$

The $M = 1, K = 6$ Case, Identities

Q_1

{1,2}
{3,4}
{5,6}

$$\mathbb{P}(Q_1) = \frac{\Gamma_{1,2} \times p_{\mathbf{w},\mathbf{s}}(1,2)}{p_{\mathbf{w}}(1)} = \frac{\Gamma_{2,1} \times p_{\mathbf{w},\mathbf{s}}(2,1)}{p_{\mathbf{w}}(2)} = \frac{\Gamma_{3,4} \times p_{\mathbf{w},\mathbf{s}}(3,4)}{p_{\mathbf{w}}(3)} = \frac{\Gamma_{4,3} \times p_{\mathbf{w},\mathbf{s}}(4,3)}{p_{\mathbf{w}}(4)} = \frac{\Gamma_{5,6} \times p_{\mathbf{w},\mathbf{s}}(5,6)}{p_{\mathbf{w}}(5)} = \frac{\Gamma_{6,5} \times p_{\mathbf{w},\mathbf{s}}(6,5)}{p_{\mathbf{w}}(6)}$$

Q_2

{1,2}
{3,6}
{4,5}

$$\mathbb{P}(Q_2) = \frac{\Gamma_{1,2} \times p_{\mathbf{w},\mathbf{s}}(1,2)}{p_{\mathbf{w}}(1)} = \frac{\Gamma_{2,1} \times p_{\mathbf{w},\mathbf{s}}(2,1)}{p_{\mathbf{w}}(2)} = \frac{\Gamma_{3,6} \times p_{\mathbf{w},\mathbf{s}}(3,6)}{p_{\mathbf{w}}(3)} = \frac{\Gamma_{6,3} \times p_{\mathbf{w},\mathbf{s}}(6,3)}{p_{\mathbf{w}}(6)} = \frac{\Gamma_{5,4} \times p_{\mathbf{w},\mathbf{s}}(5,4)}{p_{\mathbf{w}}(5)} = \frac{\Gamma_{4,5} \times p_{\mathbf{w},\mathbf{s}}(4,5)}{p_{\mathbf{w}}(4)}$$

The $M = 1$, $K = 6$ Case, Identities

$$\mathbb{P}(Q_1) = \frac{\Gamma_{1,2} \times p_{\mathbf{w},\mathbf{s}}(1,2)}{p_{\mathbf{w}}(1)} = \frac{\Gamma_{2,1} \times p_{\mathbf{w},\mathbf{s}}(2,1)}{p_{\mathbf{w}}(2)} = \frac{\Gamma_{3,4} \times p_{\mathbf{w},\mathbf{s}}(3,4)}{p_{\mathbf{w}}(3)} = \frac{\Gamma_{4,3} \times p_{\mathbf{w},\mathbf{s}}(4,3)}{p_{\mathbf{w}}(4)} = \frac{\Gamma_{5,6} \times p_{\mathbf{w},\mathbf{s}}(5,6)}{p_{\mathbf{w}}(5)} = \frac{\Gamma_{6,5} \times p_{\mathbf{w},\mathbf{s}}(6,5)}{p_{\mathbf{w}}(6)}$$

$$\mathbb{P}(Q_2) = \frac{\Gamma_{1,2} \times p_{\mathbf{w},\mathbf{s}}(1,2)}{p_{\mathbf{w}}(1)} = \frac{\Gamma_{2,1} \times p_{\mathbf{w},\mathbf{s}}(2,1)}{p_{\mathbf{w}}(2)} = \frac{\Gamma_{3,6} \times p_{\mathbf{w},\mathbf{s}}(3,6)}{p_{\mathbf{w}}(3)} = \frac{\Gamma_{6,3} \times p_{\mathbf{w},\mathbf{s}}(6,3)}{p_{\mathbf{w}}(6)} = \frac{\Gamma_{5,4} \times p_{\mathbf{w},\mathbf{s}}(5,4)}{p_{\mathbf{w}}(5)} = \frac{\Gamma_{4,5} \times p_{\mathbf{w},\mathbf{s}}(4,5)}{p_{\mathbf{w}}(4)}$$

$$\mathbb{P}(Q_3) = \frac{\Gamma_{1,2} \times p_{\mathbf{w},\mathbf{s}}(1,2)}{p_{\mathbf{w}}(1)} = \frac{\Gamma_{2,1} \times p_{\mathbf{w},\mathbf{s}}(2,1)}{p_{\mathbf{w}}(2)} = \frac{\Gamma_{3,5} \times p_{\mathbf{w},\mathbf{s}}(3,5)}{p_{\mathbf{w}}(3)} = \frac{\Gamma_{5,3} \times p_{\mathbf{w},\mathbf{s}}(5,3)}{p_{\mathbf{w}}(5)} = \frac{\Gamma_{6,4} \times p_{\mathbf{w},\mathbf{s}}(6,4)}{p_{\mathbf{w}}(6)} = \frac{\Gamma_{4,6} \times p_{\mathbf{w},\mathbf{s}}(4,6)}{p_{\mathbf{w}}(4)}$$

$$\vdots$$

Q_{15}

{1,6}
{2,4}
{5,3}

$$\mathbb{P}(Q_{15}) = \frac{\Gamma_{1,6} \times p_{\mathbf{w},\mathbf{s}}(1,6)}{p_{\mathbf{w}}(1)} = \frac{\Gamma_{6,1} \times p_{\mathbf{w},\mathbf{s}}(6,1)}{p_{\mathbf{w}}(6)} = \frac{\Gamma_{2,4} \times p_{\mathbf{w},\mathbf{s}}(2,4)}{p_{\mathbf{w}}(2)} = \frac{\Gamma_{4,2} \times p_{\mathbf{w},\mathbf{s}}(4,2)}{p_{\mathbf{w}}(4)} = \frac{\Gamma_{5,3} \times p_{\mathbf{w},\mathbf{s}}(5,3)}{p_{\mathbf{w}}(5)} = \frac{\Gamma_{3,5} \times p_{\mathbf{w},\mathbf{s}}(3,5)}{p_{\mathbf{w}}(3)}$$

The $M = 1, K = 6$ Case, Identities

$$\mathbb{P}(Q_1) = \frac{\Gamma_{1,2} \times p_{\mathbf{w},\mathbf{s}}(1,2)}{p_{\mathbf{w}}(1)} = \frac{\Gamma_{2,1} \times p_{\mathbf{w},\mathbf{s}}(2,1)}{p_{\mathbf{w}}(2)} = \frac{\Gamma_{3,4} \times p_{\mathbf{w},\mathbf{s}}(3,4)}{p_{\mathbf{w}}(3)} = \frac{\Gamma_{4,3} \times p_{\mathbf{w},\mathbf{s}}(4,3)}{p_{\mathbf{w}}(4)} = \frac{\Gamma_{5,6} \times p_{\mathbf{w},\mathbf{s}}(5,6)}{p_{\mathbf{w}}(5)} = \frac{\Gamma_{6,5} \times p_{\mathbf{w},\mathbf{s}}(6,5)}{p_{\mathbf{w}}(6)}$$

$$\mathbb{P}(Q_2) = \frac{\Gamma_{1,2} \times p_{\mathbf{w},\mathbf{s}}(1,2)}{p_{\mathbf{w}}(1)} = \frac{\Gamma_{2,1} \times p_{\mathbf{w},\mathbf{s}}(2,1)}{p_{\mathbf{w}}(2)} = \frac{\Gamma_{3,6} \times p_{\mathbf{w},\mathbf{s}}(3,6)}{p_{\mathbf{w}}(3)} = \frac{\Gamma_{6,3} \times p_{\mathbf{w},\mathbf{s}}(6,3)}{p_{\mathbf{w}}(6)} = \frac{\Gamma_{5,4} \times p_{\mathbf{w},\mathbf{s}}(5,4)}{p_{\mathbf{w}}(5)} = \frac{\Gamma_{4,5} \times p_{\mathbf{w},\mathbf{s}}(4,5)}{p_{\mathbf{w}}(4)}$$

$\mathbb{P}(Q_3) = \frac{\Gamma_{1,2} \times p_{\mathbf{w},\mathbf{s}}(1,2)}{p_{\mathbf{w}}(1)} = \frac{\Gamma_{2,1} \times p \text{[illegible]}}{\text{[illegible]}} = \frac{\Gamma_{3,5} \times p_{\mathbf{w},\mathbf{s}}(3,5)}{\text{[illegible]}} = \frac{\Gamma_{5,3} \times p_{\mathbf{w},\mathbf{s}}(5,3)}{p_{\mathbf{w}}(5)} = \frac{\Gamma_{6,4} \times p_{\mathbf{w},\mathbf{s}}(6,4)}{p_{\mathbf{w}}(6)} = \frac{\Gamma_{4,6} \times p_{\mathbf{w},\mathbf{s}}(4,6)}{p_{\mathbf{w}}(4)}$

⋮

Q_{15}

{1,6}
{2,4}
{5,3}

$\mathbb{P}(Q_{15}) = \frac{\Gamma_{1,6} \times p_{\mathbf{w},\mathbf{s}}(1,6)}{p_{\mathbf{w}}(1)} = \frac{\Gamma_{6,1} \times p \text{[illegible]}}{p_{\mathbf{w}}(6)} = \text{[illegible]} = \frac{\Gamma_{4,2} \times p_{\mathbf{w},\mathbf{s}}(4,2)}{p_{\mathbf{w}}(4)} = \frac{\Gamma_{5,3} \times p_{\mathbf{w},\mathbf{s}}(5,3)}{p_{\mathbf{w}}(5)} = \frac{\Gamma_{3,5} \times p_{\mathbf{w},\mathbf{s}}(3,5)}{p_{\mathbf{w}}(3)}$

The $M = 1$ Case

Maximize

$$\overbrace{\left(\sum_{i,j \in \mathscr{K} \times \mathscr{K}, i \neq j} p_{\mathbf{W},\mathbf{S}}(i,j) \times \left[\underbrace{\Gamma_{i,j} \left(\frac{K}{M+1} \right)}_{\text{1/rate of Partition-and-Code Scheme}} + \underbrace{(1 - \Gamma_{i,j})(K - M)}_{\text{1/rate of MDS}} \right] \right)^{-1}}^{\mathbb{E}_{\sim(\mathbf{W},\mathbf{S})}[\,\cdot\,]}$$

s.t.

$$\mathbb{P}(\mathbf{W} = i \,|\, \mathbf{Q} = \mathrm{Q}) = \mathbb{P}(\mathbf{W} = i) \quad \forall i \in \mathscr{K}\,.$$

The $M = 1$ Case

$$\left(K - M - \left(K - M - \frac{K}{M+1} \right) \times \Gamma_{1,2} \frac{p_{\mathbf{w},\mathbf{s}}(\{1\}, \{2\}])}{p_{\mathbf{w}}(\{1\})} \binom{K-1}{M} \right)^{-1}$$

$$\Gamma_{1,2} = \min_{i \in [K-1:K]} \left\{ 1, \frac{p_{\mathbf{w},\mathbf{s}}(\{i\}, [K-1:K] \backslash \{i\}) p_{\mathbf{w}}(\{1\})}{p_{\mathbf{w},\mathbf{s}}(\{1\}, \{2\}) p_{\mathbf{w}}(\{i\})} \right\}$$

The $M = 1$ Case

$$\left(K - M - \left(K - M - \frac{K}{M+1}\right) \times \Gamma_{1,2} \frac{p_{\mathbf{W},\mathbf{S}}(\{1\},\{2\}])}{p_{\mathbf{W}}(\{1\})} \binom{K-1}{M}\right)^{-1}$$

$$\Gamma_{1,2} = \min_{i \in [K-1:K]} \left\{ 1, \frac{p_{\mathbf{W},\mathbf{S}}(\{i\},[K-1:K] \backslash \{i\}) p_{\mathbf{W}}(\{1\})}{p_{\mathbf{W},\mathbf{S}}(\{1\},\{2\}) p_{\mathbf{W}}(\{i\})} \right\}$$

We generalize this technique to M, K satisfying the divisibility condition, thus obtaining the lower bound on the capacity.

$$R_{\text{LB}} = \left(K - M - \left(K - M - \frac{K}{M+1}\right) \times \Gamma_{\{1\},[2:M+1]} \frac{p_{\mathbf{W},\mathbf{S}}(\{1\},[2:M+1])}{p_{\mathbf{W}}(\{1\})} \binom{K-1}{M}\right)^{-1}$$

$$\Gamma_{\{1\},[2:M+1]} = \min_{i \in [K-M:K]} \left\{ 1, \frac{p_{\mathbf{W},\mathbf{S}}(\{i\},[K-M:K] \backslash \{i\}) p_{\mathbf{W}}(\{1\})}{p_{\mathbf{W},\mathbf{S}}(\{1\},[2:M+1]) p_{\mathbf{W}}(\{i\})} \right\}$$

Capacity of PA-PIR-SI

Theorem 1. *For PA-PIR-SI with K messages and M side info. messages such that $M+1$ is a divisor of K and strictly less than $\sqrt{K}$, under any popularity profile Λ, the capacity is upper bounded by*

$$R_{\text{UB}} = \frac{M+1}{K}$$

and is lower bounded by

$$R_{\text{LB}} = \left(K - M - \left(K - M - \frac{K}{M+1} \right) \times \Gamma_{\{1\},[2:M+1]} \frac{p_{\mathbf{W},\mathbf{S}}(\{1\},[2:M+1])}{p_{\mathbf{W}}(\{1\})} \binom{K-1}{M} \right)^{-1}$$

where

$$\Gamma_{\{1\},[2:M+1]} = \min_{i \in [K-M:K]} \left\{ 1, \frac{p_{\mathbf{W},\mathbf{S}}(\{i\},[K-M:K] \backslash \{i\}) p_{\mathbf{W}}(\{1\})}{p_{\mathbf{W},\mathbf{S}}(\{1\},[2:M+1]) p_{\mathbf{W}}(\{i\})} \right\}$$

The Upper Bound

- Proof relies on a simple, but effective consequence of the decodability/privacy conditions.

The Upper Bound

- Proof relies on a simple, but effective consequence of the decodability/privacy conditions.
  - For a query generated by any PA-PIR-SI protocol, consider any message $X_{\mathrm{W}*}$.

The Upper Bound

- Proof relies on a simple, but effective consequence of the decodability/privacy conditions.
    - For a query generated by any PA-PIR-SI protocol, consider any message $X_{\mathrm{W}*}$.
    - There must exist M messages $X_{\mathrm{S}*}$ such that $X_{\mathrm{W}*}$ can be recovered given $X_{\mathrm{S}*}$.

The Upper Bound

- Proof relies on a simple, but effective consequence of the decodability/privacy conditions.
  - For a query generated by any PA-PIR-SI protocol, consider any message $X_{\mathrm{W}*}$.
  - There must exist M messages $X_{\mathrm{S}*}$ such that $X_{\mathrm{W}*}$ can be recovered given $X_{\mathrm{S}*}$.
  - Otherwise, from the server's perspective, the user wants $X_{\mathrm{W}*}$ with 0 probability, contradicting the assumption that $\lambda_{\mathrm{W}*} > 0$.

The Upper Bound

- Proof relies on a simple, but effective consequence of the decodability/privacy conditions.
  - For a query generated by any PA-PIR-SI protocol, consider any message $X_{\mathrm{W}*}$.
  - There must exist M messages $X_{\mathrm{S}*}$ such that $X_{\mathrm{W}*}$ can be recovered given $X_{\mathrm{S}*}$.
  - Otherwise, from the server's perspective, the user wants $X_{\mathrm{W}*}$ with 0 probability, contradicting the assumption that $\lambda_{\mathrm{W}*} > 0$.

$\Longrightarrow$ Server's answer must have at least $\dfrac{K}{M+1}$ linear combinations, each of length B bits.

The Upper Bound

- Proof relies on a simple, but effective consequence of the decodability/privacy conditions.
    - For a query generated by any PA-PIR-SI protocol, consider any message $X_{\mathrm{W}*}$.
    - There must exist M messages $X_{\mathrm{S}*}$ such that $X_{\mathrm{W}*}$ can be recovered given $X_{\mathrm{S}*}$.
    - Otherwise, from the server's perspective, the user wants $X_{\mathrm{W}*}$ with 0 probability, contradicting the assumption that $\lambda_{\mathrm{W}*} > 0$.

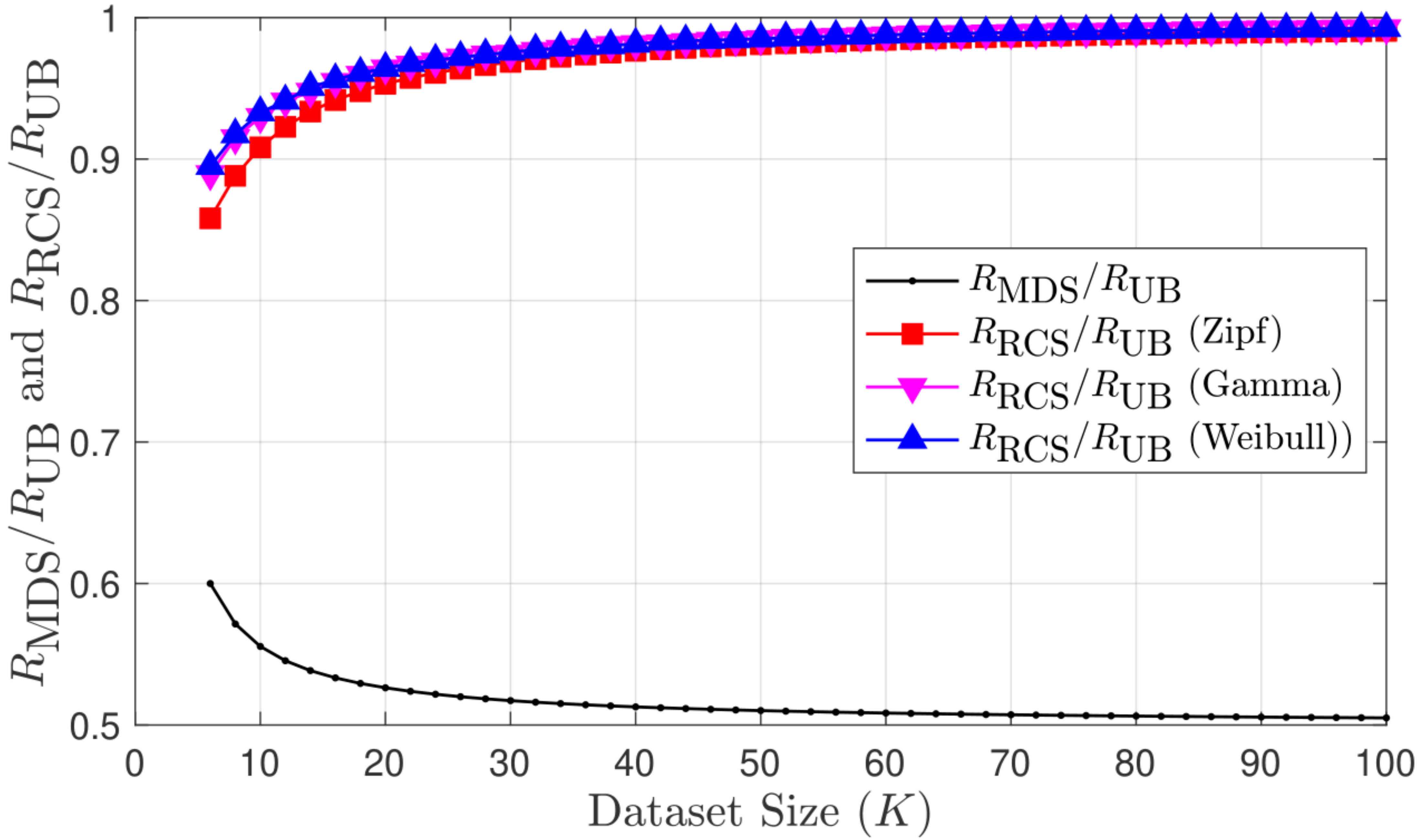
$\Longrightarrow$ Server's answer must have at least $\dfrac{K}{M+1}$ linear combinations, each of length B bits.

$\Longrightarrow$ Rate upper bound is $\dfrac{B}{[K/(M+1)]B} = \dfrac{M+1}{K}$

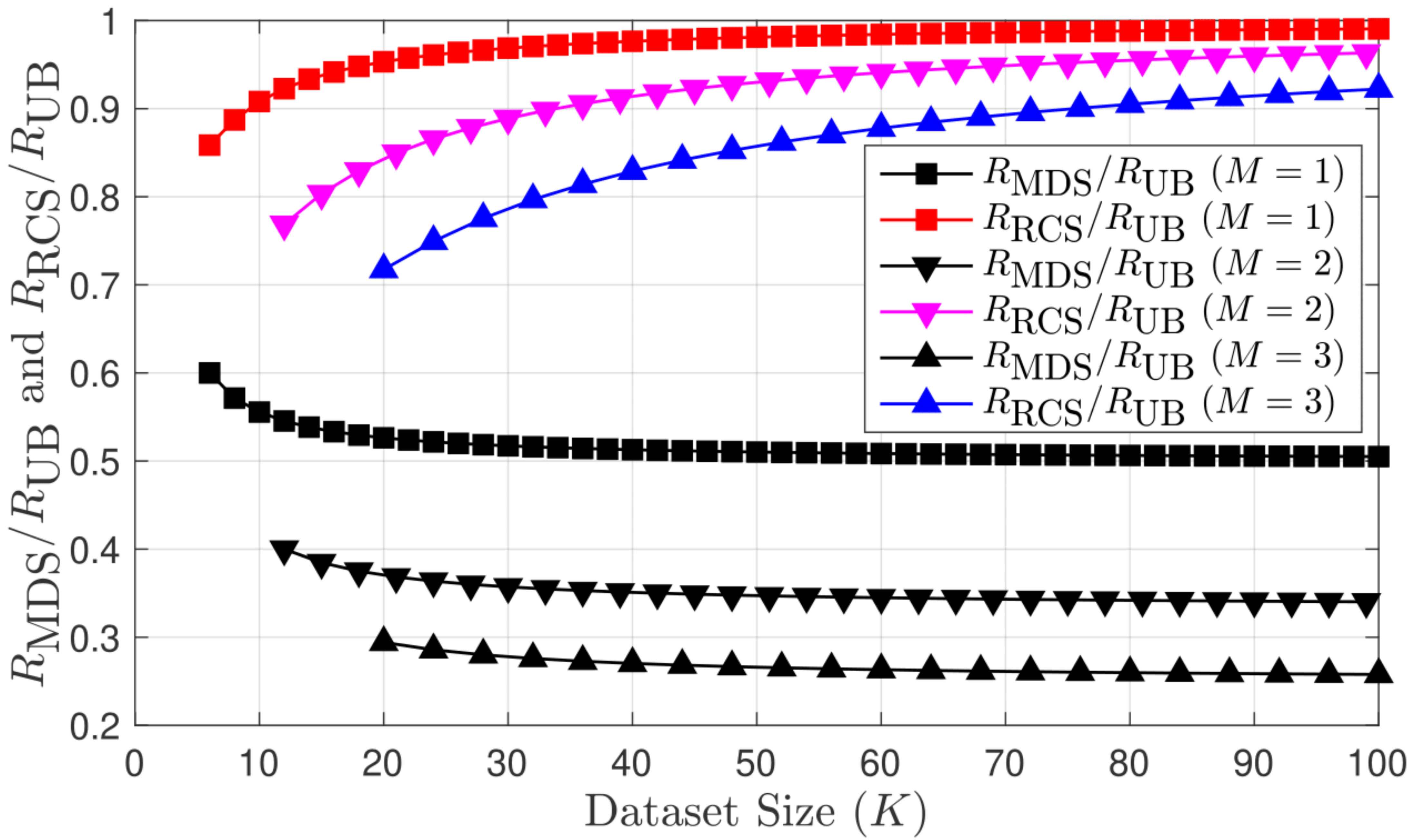
Outline

- Model + Assumptions
- A Motivating Example
- Main Results
- Simulations
- Summary and Open Problems

$\frac{R_{\mathrm{RCS}}}{R_{UB}}$ and $\frac{R_{\mathrm{MDS}}}{R_{UB}}$ vs. K, for $M = 1$ and different models for the popularity profile.



$\frac{R_{\text{RCS}}}{R_{UB}}$ and $\frac{R_{\text{MDS}}}{R_{UB}}$ vs. K, for different M and the Zipf model for the popularity profile.



Outline

- Model + Assumptions
- A Motivating Example
- Main Results
- Simulations
- Summary and Open Problems

In This Work:

- Introduced the PA-PIR-SI problem—a popularity-aware generalization of PIR-SI
- Studied the pitfalls of the existing PIR-SI schemes in the case of non-uniform popularities
- Derived bounds on the capacity of the PA-PIR-SI problem
  - Upper Bound: Using information-theoretic arguments.
  - Lower Bound: New achievability scheme (Randomized Code Selection).

Open Problems

- Capacity of multi-server PA-PIR-SI?
- Capacity of multi-message PA-PIR-SI?
- Capacity of multi-user PA-PIR-SI?
- Information leakage due to inaccurate statistics (e.g. noisy popularity profile)?