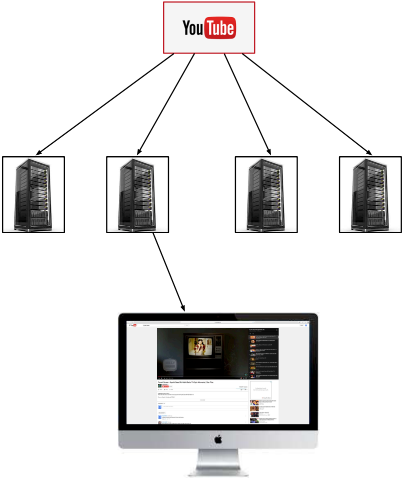
Universal Multiparty Data Exchange

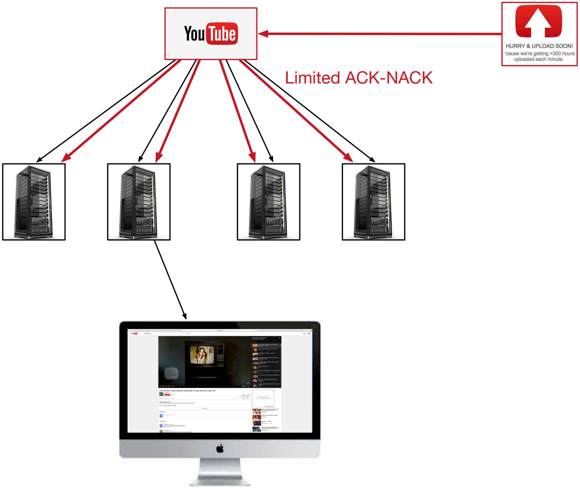
Himanshu Tyagi
Department of Electrical Communication Engineering

Joint work with Shun Watanabe

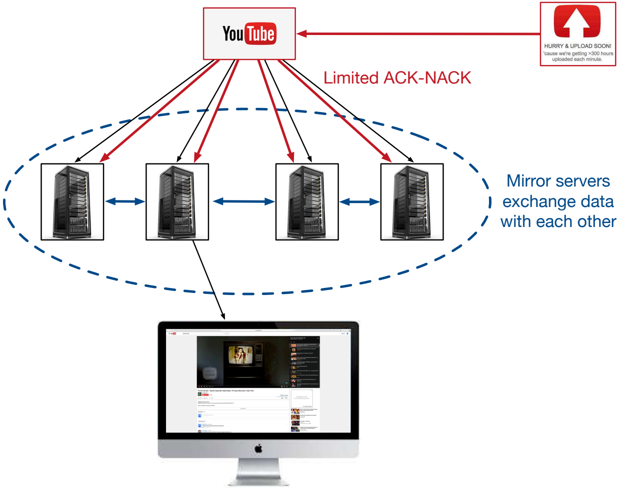
Data Exchange for Maintaining Mirror Servers



Data Exchange for Maintaining Mirror Servers



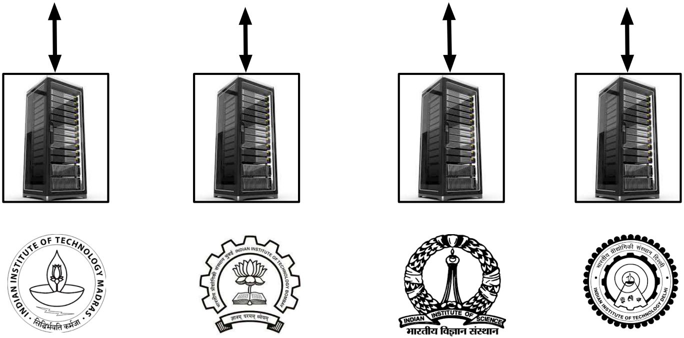
Data Exchange for Maintaining Mirror Servers



Data Exchange for Rendering 3D Videos[1]

[1]Suggested by Parimal Parag

Data Exchange for Project Management Server



Outline

1. The Multiparty Data Exchange Problem

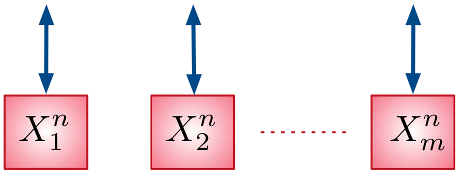
2. Description of Protocol

3. Examples

4. Results and Discussion

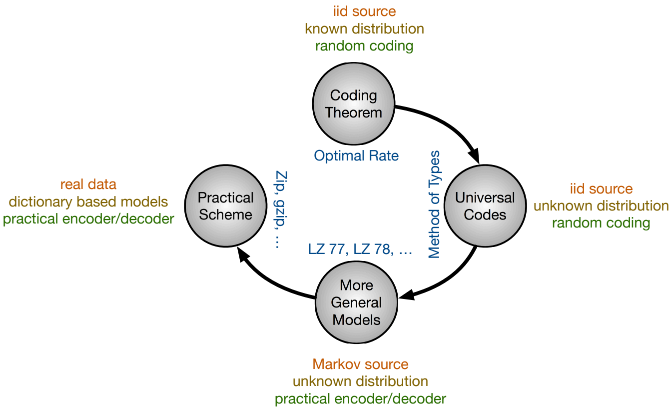
The Multiparty Data Exchange Problem

Multiparty Data Exchange

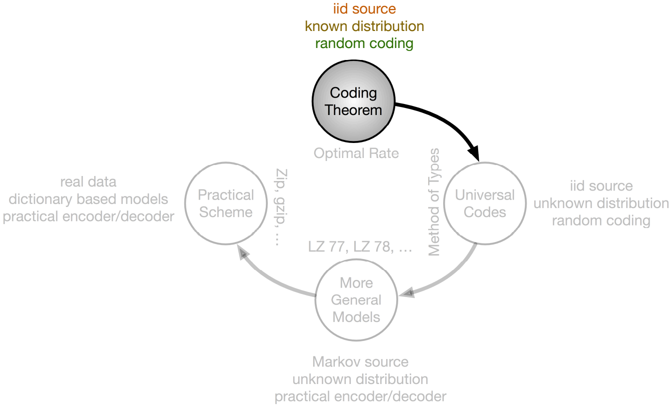


Parties seek to recover each other's data by communicating as few bits as possible

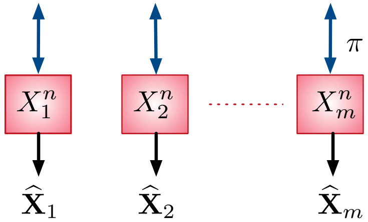
Product Cycle for a Practical Data Compression Scheme



Product Cycle for a Practical Data Compression Scheme



Source Model for Data Exchange



Set of parties, $\mathcal{M} = \{1, ..., m\}$

Observations $X_{\mathcal{M}}^n = \{X_{\mathcal{M}t}\}_{t=1}^n$ are iid with common pmf $\mathrm{P}_{X_{\mathcal{M}}}$

π constitutes an ϵ-omniscience protocol if

$$\mathrm{P}\left(\widehat{\mathbf{X}}_1 = ... = \widehat{\mathbf{X}}_m = X_{\mathcal{M}}^n\right) \geq 1 - \epsilon$$

Minimum Communication for Omniscience

$|\pi|$ = max. no. of bits communicated during an execution of π

$|\pi|_{\mathtt{av}}$ = avg. no. of bits communicated during an execution of π

Minimum Communication for Omniscience

$|\pi|$ = max. no. of bits communicated during an execution of π

$|\pi|_{\mathtt{av}}$ = avg. no. of bits communicated during an execution of π

R is an ϵ-achievable rate if $\exists$ an ϵ-omniscience protocol π with $|\pi| \leq nR$, $\forall\, n$ suff. large

$$R^{\epsilon}_{\mathsf{CO}}\left(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}\right) = \min\{R : R \text{ is an } \epsilon\text{-achievable rate}\}$$

Minimum Communication for Omniscience

$|\pi| =$ max. no. of bits communicated during an execution of π

$|\pi|_{\mathtt{av}} =$ avg. no. of bits communicated during an execution of π

R is an ϵ-achievable rate if $\exists$ an ϵ-omniscience protocol π with $|\pi| \leq nR$, $\forall\, n$ suff. large

$$R^{\epsilon}_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}\right) = \min\{R : R \text{ is an } \epsilon\text{-achievable rate}\}$$

Minimum communication for omniscience:

$$R_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}\right) = \lim_{\epsilon \to 0} R^{\epsilon}_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}\right)$$

Minimum Communication for Omniscience

$|\pi|$ = max. no. of bits communicated during an execution of π

$|\pi|_{\mathtt{av}}$ = avg. no. of bits communicated during an execution of π

R is an ϵ-achievable rate if $\exists$ an ϵ-omniscience protocol π with $|\pi| \leq nR$, $\forall\, n$ suff. large

$$R_{\mathtt{CO}}^{\epsilon}(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}) = \min\{R : R \text{ is an } \epsilon\text{-achievable rate}\}$$

Minimum communication for omniscience:

$$R_{\mathtt{CO}}(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}) = \lim_{\epsilon \to 0} R_{\mathtt{CO}}^{\epsilon}(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}})$$

Minimum average communication for omniscience:

$R_{\mathtt{CO}}^{\mathtt{av}}(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}})$ defined similarly with $|\pi|_{\mathtt{av}}$ in place of $|\pi|$

Characterization of Min. Comm. for Omniscience

[Csiszár-Naryan 04]

$$R_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}\right) = R^{\mathtt{av}}_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}\right) = \min_{R_1,\dots,R_m} \sum_{i=1}^{m} R_i,$$

where minimum is over all $(R_1, ..., R_m)$ in the set $\mathcal{R}_{\mathtt{CO}}(\mathcal{M})$ given by

$$\mathcal{R}_{\mathtt{CO}}(\mathcal{M}) = \{(R_1, ..., R_m) : \sum_{i \in B} R_i \geq H(X_B|X_{B^c}), \quad \forall\, B \subsetneq \mathcal{M}\}$$

Characterization of Min. Comm. for Omniscience

[Csiszár-Naryan 04]

$$R_{\mathtt{CO}}\left(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}\right) = R_{\mathtt{CO}}^{\mathtt{av}}\left(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}\right) = \min_{R_1,\ldots,R_m} \sum_{i=1}^{m} R_i,$$

where minimum is over all $(R_1, ..., R_m)$ in the set $\mathcal{R}_{\mathtt{CO}}(\mathcal{M})$ given by

$$\mathcal{R}_{\mathtt{CO}}(\mathcal{M}) = \{(R_1, ..., R_m) : \sum_{i \in B} R_i \geq H(X_B|X_{B^c}), \quad \forall\, B \subsetneq \mathcal{M}\}$$

[Chan-Zheng 10]

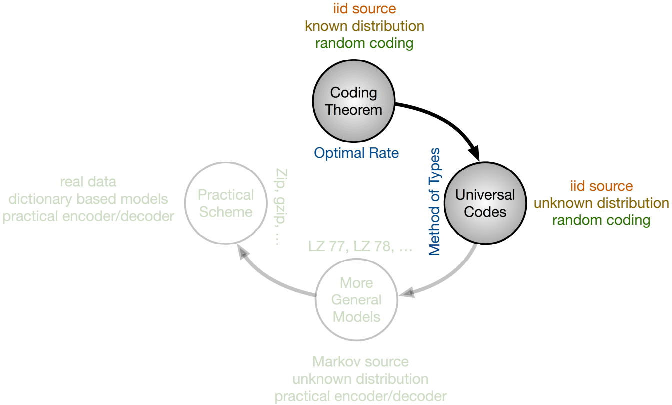
$$\min_{(R_1,\ldots,R_m)\in\mathcal{R}_{\mathtt{CO}}(\mathcal{M})} \sum_{i=1}^{m} R_i = \max_{\sigma\in\Sigma(\mathcal{M})} \frac{1}{|\sigma|-1}\mathbb{H}_\sigma,$$

where

$$\mathbb{H}_\sigma = \sum_{i=1}^{|\sigma|} H(X_{\mathcal{M}}|X_{\sigma_i})$$

The Protocol

Product Cycle for a Practical Data Compression Scheme



Naive Universal Protocol

- Use n^{α} symbols to estimate $\mathrm{P}_{X_{\mathcal{M}}}$
- This will facilitate estimation within variational distance $\mathcal{O}(n^{-\alpha/2})$
- Excess no. of bits communicated over $nR_{\mathsf{CO}}(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}})$ is order:

$$\min_{\alpha \in (0,1)} n^{\alpha} + n^{1-\alpha/2} = n^{2/3}$$

Naive Universal Protocol

- Use n^α symbols to estimate $\mathrm{P}_{X_\mathcal{M}}$
- This will facilitate estimation within variational distance $\mathcal{O}(n^{-\alpha/2})$
- Excess no. of bits communicated over $nR_{\mathsf{CO}}(\mathcal{M}|\mathrm{P}_{X_\mathcal{M}})$ is order:

$$\min_{\alpha\in(0,1)} n^\alpha + n^{1-\alpha/2} = n^{2/3}$$

[T-Viswanath-Watanabe 15]

For $m = 2$ when $\mathrm{P}_{X_1X_2}$ is known, excess is $\mathcal{O}(n^{1/2})$

Naive Universal Protocol

- Use n^α symbols to estimate $\mathrm{P}_{X_\mathcal{M}}$
- This will facilitate estimation within variational distance $\mathcal{O}(n^{-\alpha/2})$
- Excess no. of bits communicated over $nR_{\mathsf{CO}}(\mathcal{M}|\mathrm{P}_{X_\mathcal{M}})$ is order:

$$\min_{\alpha \in (0,1)} n^\alpha + n^{1-\alpha/2} = n^{2/3}$$

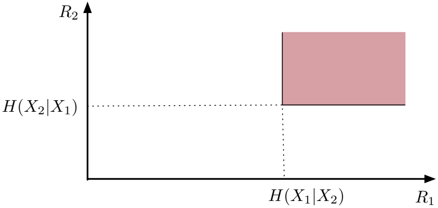
[T-Viswanath-Watanabe 15]

For $m = 2$ when $\mathrm{P}_{X_1 X_2}$ is known, excess is $\mathcal{O}(n^{1/2})$

Can we obtain a similar excess rate without knowing $\mathrm{P}_{X_\mathcal{M}}$?

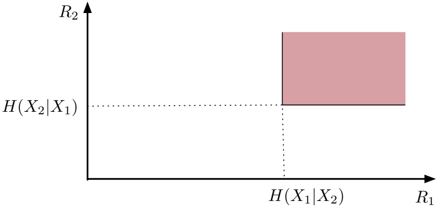
Protocol for Two Parties

$$\mathcal{R}_{\mathsf{CO}}(\mathcal{M}|\mathrm{P}_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\}$$



Protocol for Two Parties

$$\mathcal{R}_{\mathsf{CO}}(\mathcal{M}|\mathrm{P}_{X_1 X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\}$$

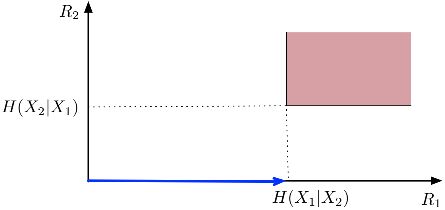


Universal Protocol 1:

1. Party 1 increases the rate until party 2 can decode

Protocol for Two Parties

$$\mathcal{R}_{\mathtt{CO}}(\mathcal{M}|\mathrm{P}_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\}$$

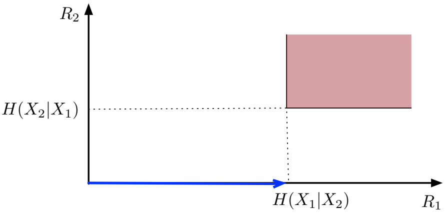


Universal Protocol 1:

1. Party 1 increases the rate until party 2 can decode

Protocol for Two Parties

$$\mathcal{R}_{\mathsf{CO}}(\mathcal{M}|\mathrm{P}_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\}$$

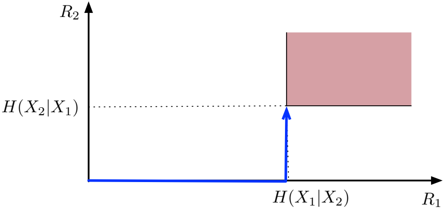


Universal Protocol 1:

1. Party 1 increases the rate until party 2 can decode
2. Party 2 increases the rate until Party 1 can decode

Protocol for Two Parties

$$\mathcal{R}_{\mathtt{CO}}(\mathcal{M}|\mathrm{P}_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\}$$

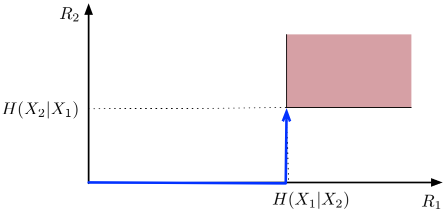


Universal Protocol 1:

1. Party 1 increases the rate until party 2 can decode
2. Party 2 increases the rate until Party 1 can decode

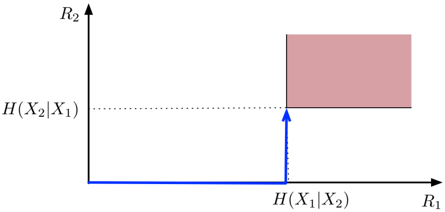
Who Starts?

$$\mathcal{R}_{\mathsf{CO}}(\mathcal{M}|\mathrm{P}_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\}$$



Who Starts?

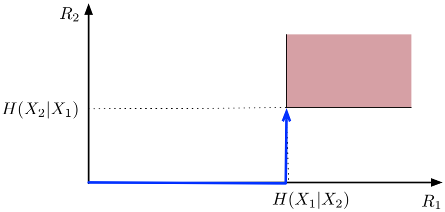
$$\mathcal{R}_{\mathsf{CO}}(\mathcal{M}|\mathrm{P}_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\}$$



Observation 1: $R_1^* - R_2^* = H(X_1|X_2) - H(X_2|X_1) = H(X_1) - H(X_2)$

Who Starts?

$$\mathcal{R}_{\mathsf{CO}}(\mathcal{M}|\mathrm{P}_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\}$$



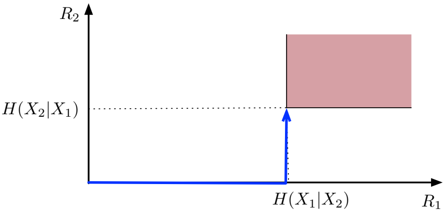
Observation 1: $R_1^* - R_2^* = H(X_1|X_2) - H(X_2|X_1) = H(X_1) - H(X_2)$

Universal Protocol 2:

1. Parties compute their types (empirical distributions) $\mathrm{P}_{\mathbf{x}_i}$ and share

Who Starts?

$\mathcal{R}_{\mathsf{CO}}(\mathcal{M}|\mathrm{P}_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\}$



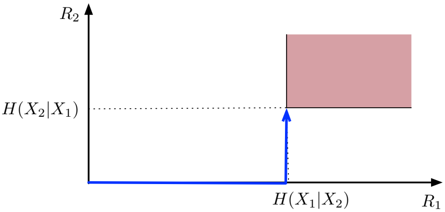
Observation 1: $R_1^* - R_2^* = H(X_1|X_2) - H(X_2|X_1) = H(X_1) - H(X_2)$

Universal Protocol 2:

1. Party with higher value of $H(\mathrm{P}_{\mathbf{x}_i})$ initializes communication

Who Starts?

$$\mathcal{R}_{\text{CO}}(\mathcal{M}|\mathrm{P}_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\}$$



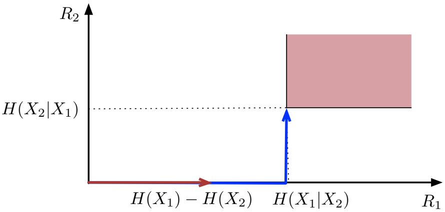
Observation 1: $R_1^* - R_2^* = H(X_1|X_2) - H(X_2|X_1) = H(X_1) - H(X_2)$

Universal Protocol 2:

1. Party with higher value of $H(\mathrm{P}_{\mathbf{x}_i})$ initializes communication
2. Party 2 starts communicating when $R_1 = H(\mathrm{P}_{\mathbf{x}_1}) - H(\mathrm{P}_{\mathbf{x}_2})$

Who Starts?

$\mathcal{R}_{\text{CO}}(\mathcal{M}|\mathrm{P}_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\}$



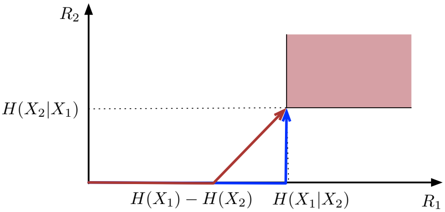
Observation 1: $R_1^* - R_2^* = H(X_1|X_2) - H(X_2|X_1) = H(X_1) - H(X_2)$

Universal Protocol 2:

1. Party with higher value of $H(\mathrm{P}_{\mathbf{x}_i})$ initializes communication
2. Party 2 starts communicating when $R_1 = H(\mathrm{P}_{\mathbf{x}_1}) - H(\mathrm{P}_{\mathbf{x}_2})$
3. Parties increase the rates until they recover each other

Who Starts?

$$\mathcal{R}_{\text{CO}}(\mathcal{M}|\mathrm{P}_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\}$$



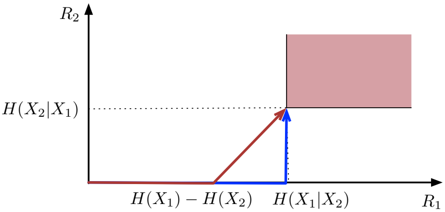
Observation 1: $R_1^* - R_2^* = H(X_1|X_2) - H(X_2|X_1) = H(X_1) - H(X_2)$

Universal Protocol 2:

1. Party with higher value of $H(\mathrm{P}_{\mathbf{x}_i})$ initializes communication
2. Party 2 starts communicating when $R_1 = H(\mathrm{P}_{\mathbf{x}_1}) - H(\mathrm{P}_{\mathbf{x}_2})$
3. Parties increase the rates until they recover each other

Who Starts?

$$\mathcal{R}_{\mathsf{CO}}(\mathcal{M}|\mathrm{P}_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1,2\}\}$$



Observation 1: $R_1^* - R_2^* = H(X_1|X_2) - H(X_2|X_1) = H(X_1) - H(X_2)$

Observation 2: Both parties will simultaneously decode each other

Universal Protocol 2:

1. Party with higher value of $H(\mathrm{P}_{\mathbf{x}_i})$ initializes communication
2. Party 2 starts communicating when $R_1 = H(\mathrm{P}_{\mathbf{x}_1}) - H(\mathrm{P}_{\mathbf{x}_2})$
3. Parties increase the rates until they recover each other

Ideal Assumptions: Oracle model

- *Continuous rate:* Rate can be increased continuously
- *Ideal decoder:* An ideal decoder with following features is available
  1. Returns correct $\mathbf{x}_A$, $A \subset \mathcal{M}$, as soon as $(R_i, i \in A) \in \mathcal{R}_{\texttt{CO}}(A)$
  2. If the condition above does not hold for any A, returns a NACK

The OMN Subroutine

$\texttt{OMN}(\sigma, \mathbf{H}, \mathbf{R})$

Inputs

$\mathbf{H} = (H_{\sigma_1}, ..., H_{\sigma_k})$ is a decreasing sequence

$\mathbf{R} = (R_1, ..., R_m)$

Outputs

$\mathcal{O}$: the set of subsets that attain omniscience

$\mathbf{R}^{\text{out}}$: rates of communication when OMN terminates

Execution

While all decoders output NACK

1. All parties with $R_i > 0$, $i \in \sigma_l$, increase their rates at "slope" $1/|\sigma_l|$
2. A new party $j \equiv \sigma_j$ starts communicating if

$$R_{\sigma_1} - R_{\sigma_j} = H_{\sigma_1} - H_{\sigma_j}$$

3. Each party is running the ideal decoder

Main Observation: The Recursive Structure of OMN

If `OMN` is called with a valid rate vector $\mathbf{R}$

If a new subset A attains local omniscience:

(i) A is of the form $\{\sigma_{i_1}, ..., \sigma_{i_l}\}$;

(ii) $\mathbf{R}^{\text{out}}$ is as if the parties in A were together from the start

Main Observation: The Recursive Structure of OMN

If `OMN` is called with a valid rate vector $\mathbf{R}$

If a new subset A attains local omniscience:

(i) A is of the form $\{\sigma_{i_1}, ..., \sigma_{i_l}\}$;

(ii) $\mathbf{R}^{\text{out}}$ is as if the parties in A were together from the start

The sum rate R_A is given by

$$R_A = \mathbb{H}_{\sigma_{f(A)}}(A) = \frac{1}{l-1} \sum_{j=1}^{l} H(X_A | X_{\sigma_{i_j}})$$

Protocol under Ideal Assumptions

Initialization

$\mathbf{R} = (0, -1, -1, ..., -1)$

$\mathbf{H} = (H(\mathrm{P}_{\mathbf{x}_1}), ..., H(\mathrm{P}_{\mathbf{x}_m}))$

$\sigma = \sigma_f(\mathcal{M})$

Execution

While omniscience is not attained

1. Call $\mathtt{OMN}(\sigma, \mathbf{H}, \mathbf{R})$; let output be $\mathcal{O}$ and $\mathbf{R}^{\mathtt{out}}$
2. **Update:**

   $\mathbf{R} = \mathbf{R}^{\mathtt{out}}$

   $\sigma =$ parts consist of subsets that have attained local omniscience

   $\mathbf{H} = (H_{\sigma_1}, ..., H_{\sigma_k})$
3. Go to step 1

Examples

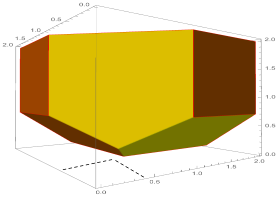
Example 1

$\underline{m = 3}$

$X_1 \sim \texttt{Ber}(1/2), \quad X_3 \sim \texttt{Ber}(q), \quad X_2 = X_1 \oplus X_3, \qquad h(q) > 1/2$

- Finest partition is dominant

- The unique optimal rate assignment is given by $\mathbf{R}^* = (1/2, 1/2, h(q) - 1/2)$



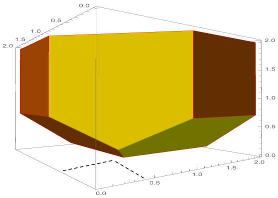
Example 1

$\underline{m = 3}$

$X_1 \sim \texttt{Ber}(1/2), \quad X_3 \sim \texttt{Ber}(q), \quad X_2 = X_1 \oplus X_3, \quad h(q) > 1/2$

- Finest partition is dominant

- The unique optimal rate assignment is given by $\mathbf{R}^* = (1/2, 1/2, h(q) - 1/2)$



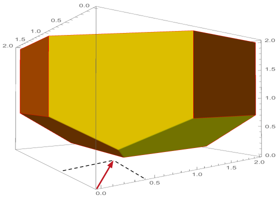
1. $H(X_1) = H(X_2) = 1 > H(X_3) \Rightarrow$ Parties 1 and 2 start at slope 1

Example 1

$\underline{m = 3}$

$X_1 \sim \texttt{Ber}(1/2), \quad X_3 \sim \texttt{Ber}(q), \quad X_2 = X_1 \oplus X_3, \quad h(q) > 1/2$

- Finest partition is dominant
- The unique optimal rate assignment is given by $\mathbf{R}^* = (1/2, 1/2, h(q) - 1/2)$



1. $H(X_1) = H(X_2) = 1 > H(X_3) \Rightarrow$ Parties 1 and 2 start at slope 1

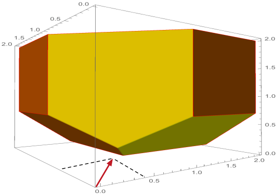
Example 1

$\underline{m = 3}$

$X_1 \sim \texttt{Ber}(1/2), \quad X_3 \sim \texttt{Ber}(q), \quad X_2 = X_1 \oplus X_3, \quad h(q) > 1/2$

- Finest partition is dominant

- The unique optimal rate assignment is given by $\mathbf{R}^* = (1/2, 1/2, h(q) - 1/2)$



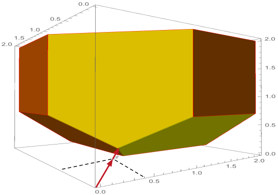
1. $H(X_1) = H(X_2) = 1 > H(X_3) \Rightarrow$ Parties 1 and 2 start at slope 1
2. Party 3 starts when $R_1 = R_2 = H(X_1) - H(X_3) = 1 - h(q)$

Example 1

$\underline{m = 3}$

$X_1 \sim \texttt{Ber}(1/2), \quad X_3 \sim \texttt{Ber}(q), \quad X_2 = X_1 \oplus X_3, \qquad h(q) > 1/2$

- Finest partition is dominant
- The unique optimal rate assignment is given by $\mathbf{R}^* = (1/2, 1/2, h(q) - 1/2)$



1. $H(X_1) = H(X_2) = 1 > H(X_3) \Rightarrow$ Parties 1 and 2 start at slope 1
2. Party 3 starts when $R_1 = R_2 = H(X_1) - H(X_3) = 1 - h(q)$

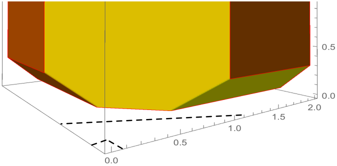
Example 2

$\underline{m = 3}$

$W_1, W_2 \sim \texttt{Ber}(1/2), \quad V_1, V_2 \sim \texttt{Ber}(q), \quad q < 1/2$

$X_1 = (W_1, W_2), \quad X_2 = (W_1 \oplus V_1, W_2), \quad X_3 = W_2 \oplus V_2$

- Partition $\{12|3\}$ is dominant



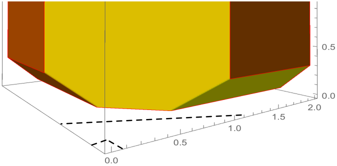
Example 2

$\underline{m=3}$

$W_1, W_2 \sim \texttt{Ber}(1/2), \quad V_1, V_2 \sim \texttt{Ber}(q), \qquad q < 1/2$

$X_1 = (W_1, W_2), \quad X_2 = (W_1 \oplus V_1, W_2), \quad X_3 = W_2 \oplus V_2$

- Partition $\{12|3\}$ is dominant



1. Parties 1 and 2 start at slope 1

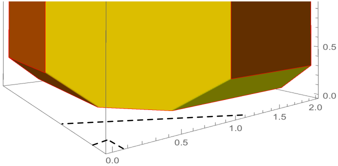
Example 2

$\underline{m = 3}$

$W_1, W_2 \sim \texttt{Ber}(1/2), \quad V_1, V_2 \sim \texttt{Ber}(q), \qquad q < 1/2$

$X_1 = (W_1, W_2), \quad X_2 = (W_1 \oplus V_1, W_2), \quad X_3 = W_2 \oplus V_2$

- Partition $\{12|3\}$ is dominant



1. Parties 1 and 2 start at slope 1
2. They attain local omniscience when $R_1 = h(q) = R_2$

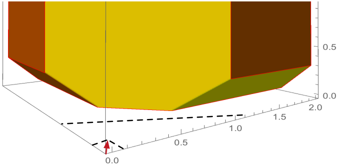
Example 2

$\underline{m = 3}$

$W_1, W_2 \sim \texttt{Ber}(1/2), \quad V_1, V_2 \sim \texttt{Ber}(q), \qquad q < 1/2$

$X_1 = (W_1, W_2), \quad X_2 = (W_1 \oplus V_1, W_2), \quad X_3 = W_2 \oplus V_2$

- Partition $\{12|3\}$ is dominant



1. Parties 1 and 2 start at slope 1
2. They attain local omniscience when $R_1 = h(q) = R_2$

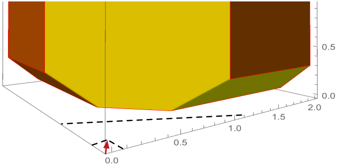
Example 2

$\underline{m = 3}$

$W_1, W_2 \sim \texttt{Ber}(1/2), \quad V_1, V_2 \sim \texttt{Ber}(q), \qquad q < 1/2$

$X_1 = (W_1, W_2), \quad X_2 = (W_1 \oplus V_1, W_2), \quad X_3 = W_2 \oplus V_2$

- Partition $\{12|3\}$ is dominant



1. Parties 1 and 2 start at slope 1
2. They attain local omniscience when $R_1 = h(q) = R_2$
3. Parties 1 and 2 increase rates at slope $1/2$

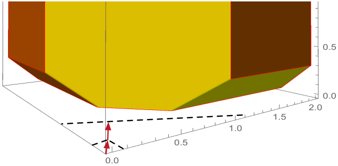
Example 2

$\underline{m = 3}$

$W_1, W_2 \sim \texttt{Ber}(1/2), \quad V_1, V_2 \sim \texttt{Ber}(q), \quad q < 1/2$

$X_1 = (W_1, W_2), \quad X_2 = (W_1 \oplus V_1, W_2), \quad X_3 = W_2 \oplus V_2$

- Partition $\{12|3\}$ is dominant



1. Parties 1 and 2 start at slope 1
2. They attain local omniscience when $R_1 = h(q) = R_2$
3. Parties 1 and 2 increase rates at slope $1/2$

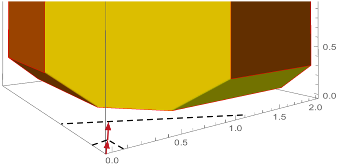
Example 2

$\underline{m = 3}$

$W_1, W_2 \sim \texttt{Ber}(1/2), \quad V_1, V_2 \sim \texttt{Ber}(q), \qquad q < 1/2$

$X_1 = (W_1, W_2), \quad X_2 = (W_1 \oplus V_1, W_2), \quad X_3 = W_2 \oplus V_2$

- Partition $\{12|3\}$ is dominant



1. Parties 1 and 2 start at slope 1
2. They attain local omniscience when $R_1 = h(q) = R_2$
3. Parties 1 and 2 increase rates at slope $1/2$
4. Parties 3 starts when $R_1 + R_2 - R_3 = H(X_1, X_2) - H(X_3) = 1 + h(q)$

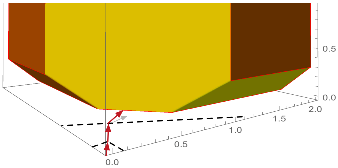
Example 2

$\underline{m = 3}$

$W_1, W_2 \sim \texttt{Ber}(1/2), \quad V_1, V_2 \sim \texttt{Ber}(q), \qquad q < 1/2$

$X_1 = (W_1, W_2), \quad X_2 = (W_1 \oplus V_1, W_2), \quad X_3 = W_2 \oplus V_2$

- Partition $\{12|3\}$ is dominant



1. Parties 1 and 2 start at slope 1
2. They attain local omniscience when $R_1 = h(q) = R_2$
3. Parties 1 and 2 increase rates at slope $1/2$
4. Parties 3 starts when $R_1 + R_2 - R_3 = H(X_1, X_2) - H(X_3) = 1 + h(q)$

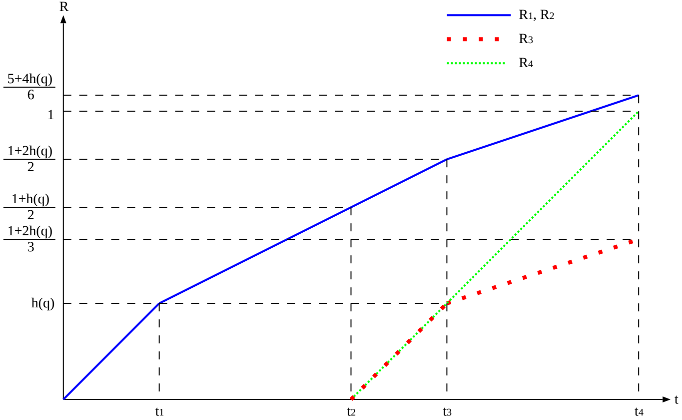
Example 3

$\underline{m = 4}$

$W_1, W_2, W_3 \sim \texttt{Ber}(1/2), \quad V_1, V_2 \sim \texttt{Ber}(q), \qquad q < 1/2$

$X_1 = (W_1, W_2), \quad X_2 = (W_1 \oplus V_1, W_2), \quad X_3 = W_2 \oplus V_2, \quad X_4 = W_3$

- Partition $\{123|4\}$ is dominant



The facts of the matter

Real World Protocol

- Parties increase rates in steps of $\Delta > 0$
- Use a typical decoder:

  Find the type $\mathrm{P}_{\overline{X}_A}$ s.t.

  1. $(R_i, i \in A) \in \mathcal{R}_{\mathtt{CO}}(A|\overline{X}_A)$, and
  2. $\exists$ unique $\mathbf{x}_A$ of type $\mathrm{P}_{\overline{X}_A}$ consistent with hash values

- Probability of error small, but greater than 0

Individual Sequence Performance

Theorem

For every $\Delta > 0$ and every sequence $\mathbf{x}_{\mathcal{M}}$, the probability of error for our protocol is bounded above by

$$C_1 \left(\frac{\log |\mathcal{X}_{\mathcal{M}}|}{\Delta} + m \right) p(n) 2^{-n\Delta}.$$

Furthermore, if an error does not occur, the number of bits communicated by the protocol for input $\mathbf{x}_{\mathcal{M}}$ is bounded above by

$$nR_{\mathsf{CO}}(\mathcal{M}|\mathrm{P}_{\mathbf{x}_{\mathcal{M}}}) + nC_2\Delta + C_3 \left(\frac{\log |\mathcal{X}_{\mathcal{M}}|}{\Delta} + m \right) + C_4 \log n.$$

Universal Performance

Corollary

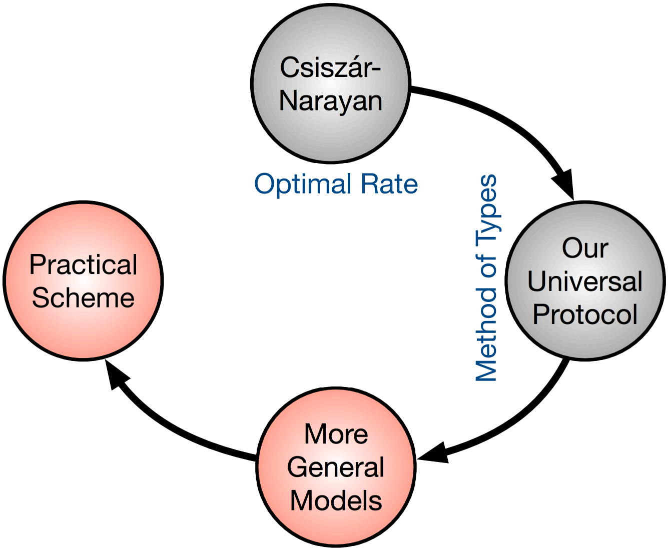
For $\Delta = \frac{1}{\sqrt{n}}$ *and every distribution* $\mathrm{P}_{X_{\mathcal{M}}}$*, our protocol has a probability of error* ϵ_n *vanishing to* 0 *as* $n \to \infty$ *and average length* $|\pi|_{\mathtt{av}}$ *less than*

$$nR_{\mathsf{CO}}(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}) + \mathcal{O}(\sqrt{n \log n}).$$

Furthermore, for a fixed $R > 0$*, the fixed-length variant of our protocol has probability of error* ϵ_n *vanishing to* 0 *as* $n \to \infty$ *for all distributions* $\mathrm{P}_{X_{\mathcal{M}}}$ *that satisfy*

$$R > R_{\mathsf{CO}}\left(\mathcal{M}|\mathrm{P}_{X_{\mathcal{M}}}\right) + \mathcal{O}\left(n^{-1/2}\sqrt{\log n}\right).$$

Product Cycle for Practical Multiparty Data Compression



rsync

How rsync works:

1. File 1 sends an easy hash (rolling checksum)
2. File 2 compares with its own hash
3. **If** no match, send the file
4. **Else** Send better hash (MD5)
   - **If** No match send the file
   - **Else** Accept files as the same

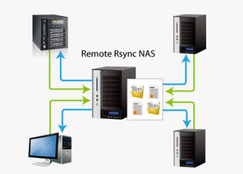


*image from RGS

rsync

How rsync works:

1. File 1 sends an easy hash (rolling checksum)
2. File 2 compares with its own hash
3. **If** no match, send the
4. **Else** Send better hash
   - **If** No match send
   - **Else** Accept files a

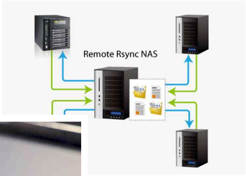
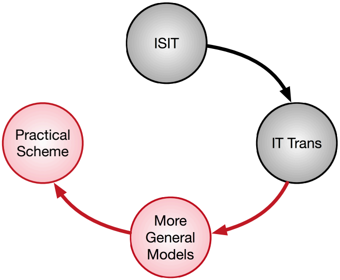


*image from RGS

Dare to think beyond rsync!

Paper Cycle for a Practical Data Compression Scheme

H. Tyagi and S. Watanabe, "*Universal Multiparty Data Exchange,*" accepted ISIT 2016



H. Tyagi and S. Watanabe, "*Universal Multiparty Data Exchange and Secret Key Agreement,*" submitted. (available on arXiv)

INTELLECTUAL PROPERTY CELL

Data Compression Conference

ACM SIGMETRICS