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

Limited ACK-NACK

Mirror servers exchange data with each other
Data Exchange for Rendering 3D Videos

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
Parties seek to recover each other’s data by communicating as few bits as possible.
Product Cycle for a Practical Data Compression Scheme

- **Practical Scheme**
  - real data
  - dictionary based models
  - practical encoder/decoder

- **More General Models**
  - Markov source
  - unknown distribution
  - practical encoder/decoder

- **Coding Theorem**
  - iid source
  - known distribution
  - random coding

- **Universal Codes**
  - iid source
  - unknown distribution
  - random coding

- **Optimal Rate**
  - LZ 77, LZ 78, ...
  - Method of Types
  - Zip, gzip, ...


Product Cycle for a Practical Data Compression Scheme

Coding Theorem

Optimal Rate

Universal Codes

More General Models

Markov source
unknown distribution
practical encoder/decoder

LZ 77, LZ 78, ...

Method of Types

real data
dictionary based models
practical encoder/decoder

iid source
known distribution
random coding

Zip, gzip, ...

Practical Scheme

iid source
unknown distribution
random coding

7
Set of parties, $\mathcal{M} = \{1, \ldots, m\}$

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

$\pi$ constitutes an $\epsilon$-omniscience protocol if

$$P\left(\hat{X}_1 = \ldots = \hat{X}_m = X_{\mathcal{M}}^n\right) \geq 1 - \epsilon$$
Minimum Communication for Omniscience

$|\pi| = \text{max. no. of bits communicated during an execution of } \pi$

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

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

$R$ is an $\epsilon$-achievable rate if $\exists$ an $\epsilon$-omniscience protocol $\pi$

with $|\pi| \leq nR$, $\forall n$ suff. large

$R^\epsilon_{\text{co}} (\mathcal{M} | P_{X_M}) = \min\{R : R$ is an $\epsilon$-achievable rate$\}$
Minimum Communication for Omniscience

$|\pi| = \text{max. no. of bits communicated during an execution of } \pi$

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

$R$ is an $\epsilon$-achievable rate if $\exists$ an $\epsilon$-omniscience protocol $\pi$

with $|\pi| \leq nR$, $\forall n$ suff. large

$$R_{C0}^\epsilon (M|P_{X_M}) = \min \{R : R \text{ is an } \epsilon\text{-achievable rate}\}$$

Minimum communication for omniscience:

$$R_{C0} (M|P_{X_M}) = \lim_{\epsilon \to 0} R_{C0}^\epsilon (M|P_{X_M})$$
Minimum Communication for Omniscience

$|\pi| = \text{max. no. of bits communicated during an execution of } \pi$

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

$R$ is an $\epsilon$-achievable rate if $\exists$ an $\epsilon$-omniscience protocol $\pi$

with $|\pi| \leq nR$, $\forall$ $n$ suff. large

$$R^\epsilon_{C0} (\mathcal{M}|P_{X,M}) = \min \{ R : R \text{ is an } \epsilon\text{-achievable rate} \}$$

Minimum communication for omniscience:

$$R_{C0} (\mathcal{M}|P_{X,M}) = \lim_{\epsilon \to 0} R^\epsilon_{C0} (\mathcal{M}|P_{X,M})$$

Minimum average communication for omniscience:

$$R^\text{av}_{C0} (\mathcal{M}|P_{X,M}) \text{ defined similarly with } |\pi|_{av} \text{ in place of } |\pi|$$
Characterization of Min. Comm. for Omniscience

[Csiszár-Naryan 04]

\[ R_{c0} (\mathcal{M}|P_{X_\mathcal{M}}) = R_{c0}^{av} (\mathcal{M}|P_{X_\mathcal{M}}) = \min_{R_1,\ldots,R_m} \sum_{i=1}^{m} R_i, \]

where minimum is over all \((R_1,\ldots,R_m)\) in the set \(\mathcal{R}_{c0}(\mathcal{M})\) given by

\[ \mathcal{R}_{c0}(\mathcal{M}) = \{(R_1,\ldots,R_m) : \sum_{i \in B} R_i \geq H(X_B|X_{B^c}), \ \forall B \subseteq \mathcal{M} \} \]
Characterization of Min. Comm. for Omniscience

[Csiszár-Naryan 04]

\[ R_{c0} (\mathcal{M} | P_{X,\mathcal{M}}) = R_{c0}^{av} (\mathcal{M} | P_{X,\mathcal{M}}) = \min_{R_1,\ldots,R_m} \sum_{i=1}^{m} R_i, \]

where minimum is over all \((R_1,\ldots,R_m)\) in the set \(\mathcal{R}_{c0}(\mathcal{M})\) given by

\[ \mathcal{R}_{c0}(\mathcal{M}) = \{(R_1,\ldots,R_m) : \sum_{i \in B} R_i \geq H(X_B|X_{B^c}), \ \forall \ B \subsetneq \mathcal{M}\} \]

[Chan-Zheng 10]

\[ \min_{(R_1,\ldots,R_m) \in \mathcal{R}_{c0}(\mathcal{M})} \sum_{i=1}^{m} R_i = \max_{\sigma \in \Sigma(\mathcal{M})} \frac{1}{|\sigma| - 1} \mathbb{H}_\sigma, \]

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

- **Coding Theorem**
- **Optimal Rate**
- **Method of Types**
- **Universal Codes**
  - iid source
  - known distribution
  - random coding

- **Practical Scheme**
  - real data
  - dictionary based models
  - practical encoder/decoder

- **More General Models**
  - Markov source
  - unknown distribution
  - practical encoder/decoder

- **LZ 77, LZ 78, ...**

- **Zip, gzip, ...**
- Use $n^\alpha$ symbols to estimate $P_{X,M}$
- This will facilitate estimation within variational distance $O(n^{-\alpha/2})$
- Excess no. of bits communicated over $nR_{c0}(M|P_{X,M})$ is order:

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

- Use $n^\alpha$ symbols to estimate $P_{X_M}$
- This will facilitate estimation within variational distance $\mathcal{O}(n^{-\alpha/2})$
- Excess no. of bits communicated over $nR_{c0}(M|P_{X_M})$ is order:

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

[T-Viswanath-Watanabe 15]

For $m = 2$ when $P_{X_1X_2}$ is known, excess is $\mathcal{O}(n^{1/2})$
Naive Universal Protocol

- Use $n^\alpha$ symbols to estimate $P_{X_\mathcal{M}}$
- This will facilitate estimation within variational distance $O(n^{-\alpha/2})$
- Excess no. of bits communicated over $nR_{C0}(\mathcal{M}|P_{X_\mathcal{M}})$ is order:

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

[T-Viswanath-Watanabe 15]

For $m = 2$ when $P_{X_1X_2}$ is known, excess is $O(n^{1/2})$

Can we obtain a similar excess rate without knowing $P_{X_\mathcal{M}}$?
\( \mathcal{R}_{c_0}(M|P_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\} \)
Protocol for Two Parties

\[ \mathcal{R}_{co}(\mathcal{M}|P_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\} \} \]

**Universal Protocol 1:**

1. Party 1 increases the rate until party 2 can decode
Protocol for Two Parties

\[ R_{c_0}(\mathcal{M}|\mathcal{P}_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\} \]

*Universal Protocol 1:*

1. Party 1 increases the rate until party 2 can decode
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
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
$\mathcal{R}_{c_0}(\mathcal{M} | P_{X_1 X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2 | X_i), i \in \{1, 2\}\}$
Who Starts?

\[ \mathcal{R}_{c0}(\mathcal{M} | P_{X_1 X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2 | X_i), i \in \{1, 2\}\} \]

Observation 1: \( R_1^* - R_2^* = H(X_1 | X_2) - H(X_2 | X_1) = H(X_1) - H(X_2) \)
Who Starts?

\[ \mathcal{R}_{c_0}(\mathcal{M}|P_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\} \]

**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) \( P_{x_i} \) and share
Who Starts?

\[ R_{c_0}(M|P_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\} \]

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 \( H(P_{X_i}) \) initializes communication
Who Starts?

\[ \mathcal{R}_{c_0}(\mathcal{M} | P_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2 | X_i), i \in \{1, 2\}\} \]

![Graph showing the relation between R1 and R2, with H(X2|X1) as a threshold.]

**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 \( H(P_{x_i}) \) initializes communication

2. Party 2 starts communicating when \( R_1 = H(P_{x_1}) - H(P_{x_2}) \)
Who Starts?

\[
\mathcal{R}_{c_0}(\mathcal{M}|P_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\}
\]

\[
R_2
\]

\[
H(X_2|X_1)
\]

\[
H(X_1) - H(X_2) \quad H(X_1|X_2) \quad R_1
\]

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 \( H(P_{x_i}) \) initializes communication
2. Party 2 starts communicating when \( R_1 = H(P_{x_1}) - H(P_{x_2}) \)
3. Parties increase the rates until they recover each other
Who Starts?

\[ \mathcal{R}_{c_0}(M|P_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\} \]

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 \( H(P_{x_i}) \) initializes communication
2. Party 2 starts communicating when \( R_1 = H(P_{x_1}) - H(P_{x_2}) \)
3. Parties increase the rates until they recover each other
Who Starts?

\[ \mathcal{R}_c(\mathcal{M}|P_{X_1X_2}) = \{(R_1, R_2) : R_i \geq H(X_1, X_2|X_i), i \in \{1, 2\}\} \]

**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 \( H(P_{x_i}) \) initializes communication
2. Party 2 starts communicating when \( R_1 = H(P_{x_1}) - H(P_{x_2}) \)
3. Parties increase the rates until they recover each other
- **Continuous rate**: Rate can be increased continuously

- **Ideal decoder**: An ideal decoder with following features is available
  
  1. Returns correct $x_A, A \subset \mathcal{M}$, as soon as $(R_i, i \in A) \in \mathcal{R}_{c0}(A)$
  
  2. If the condition above does not hold for any $A$, returns a NACK
The OMN Subroutine

\( \text{OMN}(\sigma, H, R) \)

**Inputs**

\( H = (H_{\sigma_1}, \ldots, H_{\sigma_k}) \) is a decreasing sequence

\( R = (R_1, \ldots, R_m) \)

**Outputs**

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

\( R^{\text{out}} : \) 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
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 $R$.

If a new subset $A$ attains local omniscience:

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

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

The sum rate $R_A$ is given by

$$R_A = H_{\sigma_f(A)}(A) = \frac{1}{l-1} \sum_{j=1}^{l} H(X_A|X_{\sigma_{i_j}})$$
Protocol under Ideal Assumptions

**Initialization**

\[ R = (0, -1, -1, \ldots, -1) \]

\[ H = (H(P_{x_1}), \ldots, H(P_{x_m})) \]

\[ \sigma = \sigma_f(M) \]

**Execution**

**While** omniscience is not attained

1. **Call** \( \text{OMN}(\sigma, H, R) \); let output be \( O \) and \( R^{\text{out}} \)

2. **Update:**

   \[ R = R^{\text{out}} \]

   \[ \sigma = \text{parts consist of subsets that have attained local omniscience} \]

   \[ H = (H_{\sigma_1}, \ldots, H_{\sigma_k}) \]

3. Go to step 1
Example 1

\(m = 3\)

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

- Finest partition is dominant

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

\( m = 3 \)

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

- 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

\[ m = 3 \]

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

- 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

\[ m = 3 \]

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

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

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

\[ m = 3 \]

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

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

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

\[ m = 3 \]

\[ W_1, W_2 \sim \text{Ber}(1/2), \quad V_1, V_2 \sim \text{Ber}(q), \quad q < 1/2 \]

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

- Partition \( \{12|3\} \) is dominant
Example 2

\( m = 3 \)

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

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

- Partition \( \{12|3\} \) is dominant

1. Parties 1 and 2 start at slope 1
Example 2

$m = 3$

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

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

- Partition \{12|3\} is dominant

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

\( m = 3 \)

\[ W_1, W_2 \sim \text{Ber}(1/2), \quad V_1, V_2 \sim \text{Ber}(q), \quad q < 1/2 \]

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

- Partition \( \{12|3\} \) is dominant

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

\[ m = 3 \]

\[ W_1, W_2 \sim \text{Ber}(1/2), \quad V_1, V_2 \sim \text{Ber}(q), \quad q < 1/2 \]

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

- 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

\( m = 3 \)

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

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

- 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

\[ m = 3 \]

\[ W_1, W_2 \sim \text{Ber}(1/2), \quad V_1, V_2 \sim \text{Ber}(q), \quad q < 1/2 \]

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

- 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
4. Parties 3 starts when \( R_1 + R_2 - R_3 = H(X_1, X_2) - H(X_3) = 1 + h(q) \)
Example 2

\( m = 3 \)

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

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

- 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 \( 1/2 \)
4. Parties 3 starts when \( R_1 + R_2 - R_3 = H(X_1, X_2) - H(X_3) = 1 + h(q) \)
Example 3

\[ m = 4 \]

\[ W_1, W_2, W_3 \sim \text{Ber}(1/2), \quad V_1, V_2 \sim \text{Ber}(q), \quad q < 1/2 \]

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

- Partition \{1234\} is dominant
The facts of the matter
Real World Protocol

- Parties increase rates in steps of $\Delta > 0$

- Use a typical decoder:
  
  \[
  \text{Find the type } P_{\overline{X}_A} \text{ s.t.}
  \]
  
  1. $(R_i, i \in A) \in R_{c0}(A|\overline{X}_A)$, and
  
  2. $\exists$ unique $x_A$ of type $P_{\overline{X}_A}$ consistent with hash values

- Probability of error small, but greater than 0
Theorem

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

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

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

$$nR_{c_0}(\mathcal{M}|P_{x_{\mathcal{M}}}) + nC_2\Delta + C_3 \left( \frac{\log |\mathcal{X}_M|}{\Delta} + m \right) + C_4 \log n.$$
Corollary

For $\Delta = \frac{1}{\sqrt{n}}$ and every distribution $P_{X|M}$, our protocol has a probability of error $\epsilon_n$ vanishing to 0 as $n \to \infty$ and average length $|\pi|_{av}$ less than

$$nR_{C0}(M|P_{X|M}) + O(\sqrt{n \log n}).$$

Furthermore, for a fixed $R > 0$, the fixed-length variant of our protocol has probability of error $\epsilon_n$ vanishing to 0 as $n \to \infty$ for all distributions $P_{X|M}$ that satisfy

$$R > R_{C0}(M|P_{X|M}) + O\left(n^{-1/2}\sqrt{\log n}\right).$$
Product Cycle for Practical Multiparty Data Compression

- Practical Scheme
- More General Models
- Csiszár-Narayan
- Our Universal Protocol
- Optimal Rate
- Method of Types
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
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
   - If No match send
   - Else Accept files as the same

Dare to think beyond rsync!
Paper Cycle for a Practical Data Compression Scheme


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