Something Ancient and Something Recent

Raymond W. Yeung
Institute of Network Coding, CUHK
Something Ancient
ITW '93
1993 IEEE Information Theory Workshop
June 4 (FRI) - June 8 (TUE), 1993

Fuji Institute of Education and Training
Suzuwari, Shizuoka, Japan

Sponsored by:
IEEE Information Theory Society

In cooperation with:
IEICE Information Society Group
The Society of Information Theory
and Its Applications

Co-sponsored by:
International Communications Foundation
Inoue Foundation for Science
The Telecommunications Advancement Foundation
ADVANCE COMPANY, LTD.
Canon Inc.
IBM Japan, Ltd.
Mitsubishi Electric Corporation
NEC Corporation
Nippon Telegraph and Telephone Corporation
NTT Mobile Communications Network Inc.
OKI Electric Industry Company Ltd.
Toshiba Corporation
Diversified Coding with One Distortion Criterion

Raymond W. Yeung
Department of Information Engineering
The Chinese University of Hong Kong
email: whyeung@ie.cuhk.hk

1 Introduction

In a Diversified Coding System (DCS), an information source is encoded by a number of encoders. There are a number of decoders, each of which can access a certain subset of the encoders. Each decoder is to reconstruct the source either perfectly or subject to a distortion criterion. The problem is to determine the coding rate region for a particular configuration of a DCS subject to certain distortion criteria.

Diversified coding has wide application in distributed information storage (e.g. [3]), fault-tolerant communication network (e.g. [4]), and secret sharing (e.g. [5]). Most of these works are application of the pioneering work of Singleton [1] on maximum distance error-correcting codes.

Diversified coding from the rate-distortion point of view is discussed in the work of El Gamal and Cover [2] on the multiple descriptions problem. In their work, each decoder makes it best effort to reconstruct the source with no reference to the reconstructions by other decoders. By contrast, in our problem, the decoders are divided into classes, and it is required that the reconstructions of the source by decoders within the same class are identical. This is a natural requirement for many applications. For example, if the users of decoders within the same class are to discuss the information they receive subsequently, it would be critical that the information they receive are identical.
for all \( j \in B \). Now each DC decoder can reconstruct the output of the source encoder with probability of error less than \( \epsilon'' \). Thus the source decoder can reconstruct the source with average distortion of less than
\[
(1 - \epsilon'')(D + \epsilon') + \epsilon''d_{\text{max}}
\]
where \( d_{\text{max}} \) is the largest possible value of the distortion function (we assume that \( d_{\text{max}} \) is finite). For any \( \epsilon > 0 \), by taking \( \epsilon', \epsilon'' > 0 \) satisfying
\[
\epsilon' + \epsilon'' < \epsilon
\]
and
\[
(1 - \epsilon'')(D + \epsilon') + \epsilon''d_{\text{max}} < D + \epsilon,
\]
we have shown the existence of a diversity coding scheme with coding rates satisfying
\[
\sum_{i \in G_j} R_i < R(D) + \epsilon
\]
for all \( j \in B \), where each decoder can reconstruct the source with average distortion of less than \( D + \epsilon \). This proves that the inequalities in (2) are necessary and sufficient conditions for the coding rates in the usual Shannon sense. This observation appears to be new.

### B. Subtlety of Multilevel Diversity Coding

The class of MDCS’s we treat in this paper, which will be described in the next section, have two levels of decoders. In this subsection, we first illustrate the subtlety of such problems by means of a simple example. Consider the i.i.d. source \( \{X_k, Y_k\} \) where \( \{X_k\} \) and \( \{Y_k\} \) are independent bit streams with rates 1 bit per unit time. The MDCS we consider is shown in Fig. 3, where it is required that Decoder 1 can reconstruct \( \{X_k\} \) with zero error, and Decoder \( i, i = 2, 3, 4 \) can reconstruct \( \{(X_k, Y_k)\} \) with zero error; \( \{X_k\} \) can be regarded as a degraded version of \( \{(X_k, Y_k)\} \). Now \( \{X_k\} \) has to be delivered to all decoders, while \( \{Y_k\} \) has to be delivered to Decoders 2, 3, and 4. Since \( \{X_k\} \) and \( \{Y_k\} \) are independent bit streams, one may expect that optimality can be achieved by coding \( \{X_k\} \) and \( \{Y_k\} \) separately, as for the case of point-to-point communication. This will be referred to as the principle of superposition. The argument is that as \( \{X_k\} \) and \( \{Y_k\} \) are independent bit streams, the coding rates contributing to the coding of \( \{X_k\} \) do not contribute to the coding of \( \{Y_k\} \). At this point, however, it is not clear whether the principle of superposition actually applies.

We now define the coding rate region corresponding to coding \( \{X_k\} \) and \( \{Y_k\} \) separately, \( R_{\text{sup}} \), as the set containing \( (R_1, R_2, R_3) \) such that for \( i = 1, 2, 3 \)
\[
1 R_i = r_i^x + r_i^y
\]
where \( r_i^x, r_i^y \geq 0 \), and
\[
\begin{align*}
1 r_1^x &\geq 1 \\
r_1^x + r_2^y &\geq 1 \\
r_2^x + r_3^y &\geq 1 \\
r_3^x + r_3^y &\geq 1
\end{align*}
\]
and
\[
\begin{align*}
r_1^y &\geq 1 \\
r_1^y + r_2^y &\geq 1 \\
r_2^y + r_3^y &\geq 1 \\
r_1^y + r_3^y &\geq 1
\end{align*}
\]
(1, 1, 1) is achievable

![Diagram of MDCS](image-url)
for all \( j \in B \). Now each DC decoder can reconstruct the output of the source encoder with probability of error less than \( \epsilon'' \). Thus the source decoder can reconstruct the source with average distortion of less than

\[
(1 - \epsilon'') (D + \epsilon') + \epsilon'' d_{\text{max}}
\]

where \( d_{\text{max}} \) is the largest possible value of the distortion function (we assume that \( d_{\text{max}} \) is finite). For any \( \epsilon > 0 \), by taking \( \epsilon', \epsilon'' > 0 \) satisfying

\[
\epsilon' + \epsilon'' < \epsilon
\]

and

\[
(1 - \epsilon'') (D + \epsilon') + \epsilon'' d_{\text{max}} < D + \epsilon,
\]

we have shown the existence of a diversity coding scheme with coding rates satisfying

\[
\sum_{i \in G_j} R_i < R(D) + \epsilon
\]

for all \( j \in B \), where each decoder can reconstruct the source with average distortion of less than \( D + \epsilon \). This proves that the inequalities in (2) are necessary and sufficient conditions for the coding rates in the usual Shannon sense. This observation appears to be new.

**B. Subtlety of Multilevel Diversity Coding**

The class of MDCS’s we treat in this paper, which will be described in the next section, have two levels of decoders. In this subsection, we first illustrate the subtlety of such problems by means of a simple example. Consider the i.i.d. source \( \{(X_k, Y_k)\} \) where \( \{X_k\} \) and \( \{Y_k\} \) are independent bit streams with rates 1 bit per unit time. The MDCS we consider is shown in Fig. 3, where it is required that Decoder 1 can reconstruct \( \{X_k\} \) with zero error, and Decoder \( i, i = 2, 3, 4 \) can reconstruct \( \{(X_k, Y_k)\} \) with zero error; \( \{X_k\} \) can be regarded as a degraded version of \( \{(X_k, Y_k)\} \). Now \( \{X_k\} \) has to be delivered to all decoders, while \( \{Y_k\} \) has to be delivered to Decoders 2, 3, and 4. Since \( \{X_k\} \) and \( \{Y_k\} \) are independent bit streams, one may expect that optimality can be achieved by coding \( \{X_k\} \) and \( \{Y_k\} \) separately, as for the case of point-to-point communication. This will be referred to as the principle of superposition. The argument is that as \( \{X_k\} \) and \( \{Y_k\} \) are independent bit streams, the coding rates contributing to the coding of \( \{X_k\} \) do not contribute to the coding of \( \{Y_k\} \). At this point, however, it is not clear whether the principle of superposition actually applies.

We now define the coding rate region corresponding to coding \( \{X_k\} \) and \( \{Y_k\} \) separately, \( \mathcal{R}_{\text{sup}} \), as the set containing \((R_1, R_2, R_3)\) such that for \( i = 1, 2, 3 \)

\[
R_i = r_i^x + r_i^y
\]

where \( r_i^x, r_i^y \geq 0 \), and

\[
r_i^x \geq 1
\]

\[
r_1^x + r_2^x \geq 1
\]

\[
r_2^x + r_3^x \geq 1
\]

\[
r_1^x + r_3^x \geq 1
\]

and

\[
r_i^y \geq 1
\]

\[
r_1^y + r_2^y \geq 1
\]

\[
r_2^y + r_3^y \geq 1
\]

\[
r_1^y + r_3^y \geq 1
\]

\((1, 1, 1)\) is achievable.
Why is this interesting?

- When $X$ and $Y$ are independent, $H(X) + H(Y) = H(X,Y)$.
- From classical information theory, we know that there is no difference between compressing $X$ and $Y$ separately or together.
- Therefore, in classical information theory, there is no distinction between single-source or multi-source data compression.
- The last example shows that the behavior of information in a network deviates from what we would expect from classical information theory.
- Transmission of well-compressed information sources in a network is not a commodity flow.
- A gold mine ahead $\rightarrow\rightarrow\rightarrow$ Network coding
Other Works Leading to Network Coding


Network Coding and Entropy Function

This problem is either trivial or great. (1995)

LYC03, KM03, JSCEEJT04, ... DFZ05, SYC06, DFZ07, CG08, YYZ12, ...
Something about Network Coding

A very unique class of multi-user information theory problems:

- Non-trivial even with very simplistic assumptions
  - independent sources
  - individual sources well compressed
  - no distortion consideration

- Exists a unifying implicit single-letter characterisation of the capacity region in terms of the entropy function region $\Gamma^*$
Something Recent
BATched Sparse (BATS) Code


Shenghao Yang
CUHK (Shenzhen)
Transmission through Packet Networks (Erasure Networks)

One 20MB file \( \approx \) 20,000 packets

- A practical solution
- Low computational and storage costs
- High transmission rate
- Small protocol overhead

\[
\begin{array}{c}
    b_1 \quad b_2 \quad \cdots \quad b_K \\
    s \\
    t_1 \\
    t_2
\end{array}
\]
Transmission through Packet Networks (Erasure Networks)

One 20MB file \(\approx 20,000\) packets

A practical solution
- low computational and storage costs
- high transmission rate
- small protocol overhead

\[
b_1 \ b_2 \ \ldots \ b_K
\]
Routing Networks

Retransmission
- Example: TCP
- Not scalable for multicast
- Cost of feedback

\[ s \quad \rightarrow \quad u \quad \rightarrow \quad t \]

(re)transmission \quad forwarding \quad feedback
Routing Networks

Retransmission
- Example: TCP
- Not scalable for multicast
- Cost of feedback

Forward error correction
- Example: fountain codes
- Scalable for multicast
- Neglectable feedback cost

s → u → t

encoding  forwarding  decoding
Complexity of Fountain Codes with Routing

- $K$ packets, $T$ symbols in a packet.
- Encoding: $O(T)$ per packet.
- Decoding: $O(T)$ per packet.
- Routing: $O(1)$ per packet and fixed buffer size.

Both links have a packet loss rate 0.2.
The capacity of this network is 0.8.

<table>
<thead>
<tr>
<th>Intermediate</th>
<th>End-to-End</th>
<th>Maximum Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>forwarding</td>
<td>retransmission</td>
<td>0.64</td>
</tr>
<tr>
<td>forwarding</td>
<td>fountain codes</td>
<td>0.64</td>
</tr>
<tr>
<td>network coding</td>
<td>random linear codes</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Achievable Rates: $n$ hops

All links have a packet loss rate 0.2.

<table>
<thead>
<tr>
<th>Intermediate Operation</th>
<th>Maximum Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>forwarding</td>
<td>$0.8^n \rightarrow 0$, $n \rightarrow \infty$</td>
</tr>
<tr>
<td>network coding</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Multicast capacity of erasure networks

Theorem

Random linear network codes achieve the capacity of a large range of multicast erasure networks.

References:


Complexity of Linear Network Coding

- Encoding: $O(TK)$ per packet.
- Decoding: $O(K^2 + TK)$ per packet.
- Network coding: $O(TK)$ per packet. Buffer $K$ packets.
Batched Sparse (BATS) Codes

outer code
(matrix fountain code)

inner code
(network code)


Apply a “matrix fountain code” at the source node:

1. Obtain a degree $d$ by sampling a degree distribution $\Psi$.
2. Pick $d$ distinct input packets randomly.
3. Generate a batch of $M$ coded packets using the $d$ packets.

Transmit the batches sequentially.

\[
X_i = \begin{bmatrix} b_{i1} & b_{i2} & \cdots & b_{id_i} \end{bmatrix} \quad G_i = B_i G_i.
\]
The batches traverse the network.

Encoding at the intermediate nodes forms the inner code.

Linear network coding is applied in a causal manner within a batch.

\[ Y_i = X_i H_i, \quad i = 1, 2, \ldots \]
Belief Propagation Decoding

1. Find a check node $i$ with degree $i = \text{rank}(G_i H_i)$.
2. Decode the $i$th batch.
3. Update the decoding graph. Repeat 1).

The linear equation associated with a check node: $Y_i = B_i G_i H_i$. 
Precoding

- Precoding by a fixed-rate erasure correction code.
- The BATS code recovers $(1 - \eta)$ of its input packets.

### Complexity of Sequential Scheduling

<table>
<thead>
<tr>
<th>Node Type</th>
<th>Encoding/Decoding Time Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source node encoding</td>
<td>$\mathcal{O}(TM)$ per packet</td>
</tr>
<tr>
<td>Destination node decoding</td>
<td>$\mathcal{O}(M^2 + TM)$ per packet</td>
</tr>
<tr>
<td>Intermediate Node buffer</td>
<td>$\mathcal{O}(TM)$ per packet</td>
</tr>
<tr>
<td>Intermediate Node network coding</td>
<td>$\mathcal{O}(TM)$ per packet</td>
</tr>
</tbody>
</table>

- **$T$**: length of a packet
- **$K$**: number of packets
- **$M$**: batch size
Achievable Rates for Line Networks

![Graph showing achievable rates for line networks with different values of M (64, 32, 16, 8, 4, 2, 1) against network length.]
Potential applications

- 5G mobile network
- Wireless mesh network
- Vehicular ad-hoc network
- Mobile ad-hoc network
- Satellite network
- Content delivery network (CDN)
- Internet of Things (IoT)
An all-software prototype running BATS code was recently built.
Source node, relay nodes, and receiving nodes are all notebook computers.
A notebook with Intel i7 CPU was employed for decoding.
A transmission rate > 500 Mb/s was achieved.
Will collaborate with P2MT to implement BATS code in mesh network products (802.11).
BATS codes provide a digital fountain solution with linear network coding:

- Outer code at the source node is a matrix fountain code.
- Linear network coding at the intermediate nodes forms the inner code.
- Prevents BOTH packet loss and delay from accumulating along the way.

The more hops between the source node and the sink node, the larger the benefit.

Future work:

- Finite-length analysis
- Proof of (nearly) capacity achieving
- Design of intermediate operations to maximize the throughput and minimize the buffer size