Multi-user information theory and an example: the two-way relay channel

Natasha Devroye  
Assistant Professor  
University of Illinois at Chicago  
http://www.ece.uic.edu/~devroye

Monologues vs. Dialogues

Monologue = one-way = uni-directional

Dialogue = two-way = bi-directional

Two-way communication applications - wired

Video conferencing

Data synchronization

Telesurgery

Two-way communication applications - wireless

1 Dialogue ≠ or = 2 Monologues?

It depends....

we will use information theory to find out.
Overall - much is still unknown

Outline

- Information theory - what, why, when
- Two-way channel - channel coding
- Wireless channels and networks
- Two-way cellular-like networks
- Two-way relay channels - canonical example of wireless network coding

Information theory - what, why, when


Information theory’s claims to fame

Source coding

- Source = random variable
- Ultimate data compression limit is the source’s entropy $H_{X}$

Channel coding

- Channels, mutual information
- Ultimate transmission rate is the channel capacity $C$

Source coding = data compression

Minimum $R$ needed is $H(X) = \sum_x p(x) \log_2(p(x))$

Lossy: $d(\hat{X}^n, X^n) \leq D$

Minimum $R$ needed is described by the Rate-Distortion function $R(D)$

$$R(D) = \min_{p(\hat{x}|x)} I(X; \hat{X})$$

s.t. $d(\hat{X}^n, X^n) \leq D$
What is the capacity of this channel?

Channel capacity

\[ C = \max_{p(x)} I(X; Y) \]

- Information channel capacity:

- Operational channel capacity:
  
  Highest rate (bits/channel use) that can communicate at reliably

- Channel coding theorem says: information capacity = operational capacity

Channel capacity: a cute example

Source \( A, B, C, D \) \( \xrightarrow{\text{Encoder}} \) \( Y^n \)

Source \( A, B, C, D \) \( \xrightarrow{\text{Encoder}} \) \( A = \text{AAA}^? \)

Source \( A, B, C, D \) \( \xrightarrow{\text{Encoder}} \) \( A = \text{AAA}^? \) \( \xrightarrow{\text{Channel}} \) \( \text{AAA} \rightarrow \text{AB}. \)
Channel capacity: a cute example

Source → Encoder → Channel

Use these 9 symbols!

C = \log_2(9)

How to communicate reliably?

AB. → AAA .AZ BBA

Channel capacity: a cute example

Source → Encoder → Channel → Decoder

Mathematical description of capacity

- Can achieve reliable communication for all transmission rates R:

\[ R < C \]

- BUT, probability of decoding error always bounded away from zero if

\[ R > C \]

Capacity in general

- Main idea was to reduce the rate (from a 27-letter input per channel use to a 9-letter input per channel use) so as to produce non-overlapping outputs!

Inputs

Outputs

Capacity: key ideas

- choose input set of codewords so they are “non-confusable” at the output

- number of these that we can choose will determine the channel’s capacity

- number that we can choose will depend on the distribution \( p(y|x) \) which characterizes the channel

One-way channel capacity

\[ C = \max_{p(x)} I(X;Y) \]

bits/channel use

\[ I(X;Y) = \sum_{x,y} p(x,y) \log \left( \frac{p(x,y)}{p(x)p(y)} \right) \]

A few examples
What is capacity / mutual information?

\[ C = \max_{p(x)} I(X; Y) \]

Entropy of a random variable

(A) entropy is the measure of average uncertainty in the random variable

(B) entropy is the average number of bits needed to describe the random variable

(C) entropy is measured in bits?

(D) \( H(X) = -\sum_x p(x) \log_2(p(x)) \)

(E) entropy of a deterministic value is 0
Entropy of a uniform distribution

- Let X be uniformly distributed over 8 outcomes. What is the entropy of X?
  \[ H(X) = \log_2(K) \]

- This is the number of bits needed to describe X!

- By extension, for a discrete random variable taking on K outcomes, the maximal entropy is attained by a uniform distribution and is equal to the number of bits needed to describe K:
  \[ H(X) = \log_2(K) \]

Entropy of a non-uniform distribution

- Suppose X represents the outcome of a horse race with 8 horses, which win with probabilities \( \left( \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64} \right) \)
  \[ H(X) = -\sum_{x} p(x) \log_2(p(x)) = -\left( \frac{1}{2} \log_2 \left( \frac{1}{2} \right) \right) - \left( \frac{1}{4} \log_2 \left( \frac{1}{4} \right) \right) - \left( \frac{1}{8} \log_2 \left( \frac{1}{8} \right) \right) - \left( \frac{1}{16} \log_2 \left( \frac{1}{16} \right) \right) - \left( \frac{1}{32} \log_2 \left( \frac{1}{32} \right) \right) - \left( \frac{1}{64} \log_2 \left( \frac{1}{64} \right) \right) \]
  \[ = 2 \text{ bits (on average!)} \]

Entropy of a continuous random variable

- entropy:
  \[ H(X) = -\sum_{x} p(x) \log_2(p(x)) \]

- differential entropy:
  \[ h(X) = -\int f(x) \log_2(f(x)) \, dx \]

Entropy maximization

- Uniform distribution maximizes entropy for a given # outcomes
  \[ \max_{X:|X|=K} H(X) = \log_2(K) \]

- Gaussian maximizes entropy for a given covariance constraint
  \[ \max_{E[X^2]=K} h(X) = \frac{1}{2} \log \left( \left( 2\pi e \right)^n |K| \right) \]

Entropy of a Gaussian random variable

- differential entropies of Gaussian distributions:
  \[ h(N(0, \sigma^2)) = \frac{1}{2} \log \left( 2\pi e \sigma^2 \right) \]
  \[ h(N(\mu, K)) = \frac{1}{2} \log \left( \left( 2\pi e \right)^n |K| \right) \]

Mutual information between 2 random variables:

\[ I(X;Y) = \sum_{x,y} p(x,y) \log \left( \frac{p(x,y)}{p(x)p(y)} \right) \]
\[ = H(X) - H(X|Y) \]
\[ = H(Y) - H(Y|X) \]
Mutual information between 2 random variables:

\[ I(X;Y) = \sum p(x,y) \log \left( \frac{p(x,y)}{p(x)p(y)} \right) \]
\[ = H(X) - H(X|Y) \]
\[ = H(Y) - H(Y|X) \]

(A) \(I(X;Y)\) is the reduction in the uncertainty about \(X\) due to knowledge of \(Y\)
(B) if \(X, Y\) are independent \(I(X;Y) = 0\)
(C) if \(X=Y\) then \(I(X;Y) = H(X)\)
(D) \(I(X;Y)\) is non-negative

Properties of mutual information

Channel capacity

\[
C = \max_{p(x)} I(X;Y) \text{ bits/channel use}
\]
\[
I(X;Y) = \sum_{x,y} p(x,y) \log \left( \frac{p(x,y)}{p(x)p(y)} \right)
\]

With permission from David J.C. MacKay
Channel capacity: $C = \max I(X; Y)$
1 bit/channel use

- Capacity?
  - Noiseless channel
    - $C = \max p(x)$
      - $I(X; Y)$
      - $max H(X) - H(X|Y)$
      - $max H(X) - 0$
      - $1$

- Noisy typewriter
  - $C = \log_2(9)$
    - $I(X; Y)$
    - $max H(X) - H(X|Y)$
    - $max H(X) - \log_2(3)$
    - $\log_2(27) - \log_2(3) = \log_2(9)$

- Binary erasure channel
  - $C = \max I(X; Y)$
    - $max H(X) - H(X|Y)$
    - $max H(X) - 0$
    - $1$

Discrete memoryless channel capacity

Capacity $C = \max I(X; Y)$ bits/channel use

$I(X; Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$

"mutual information" between X and Y

Continuous alphabet channel capacity

Capacity $C = \max I(X; Y)$ bits/channel use

"mutual information" between X and Y

What if $X$ and $Y$ are not bits, but real numbers?
AWGN channel capacity

\[ C = \max_{p(x|y)} I(X;Y) \]
\[ = \max_{p(x|y)} h(X) - h(X|Y) \]
\[ = \max_{p(x|y)} h(Y) - h(Y|X) \]
\[ = \max_{p(x|y)} h(X+N) - h(X+N|X) \]
\[ = \frac{1}{2} \log (b^2 P + P_N) - \frac{1}{2} \log (2 \pi e P_N) \]
\[ = \frac{1}{2} \log \left( \frac{b^2 P + P_N}{P_N} \right) \] where power is constrained to \( P \)

Source vs. channel coding

Use?

Claude Shannon — Born on the planet Earth (Sol III) in the year 1916 A.D. Generally regarded as the father of the Information Age, he formulated the notion of channel capacity in 1948 A.D. Within several decades, mathematicians and engineers had devised practical ways to communicate reliably at data rates within 1% of the Shannon limit...

- Robert J. McEliece, Allerton 2000

\[ C = \frac{1}{2} \log \left( \frac{b^2 P + P_N}{P_N} \right) \] (bits/channel use)

What about bits/second and bandwidth of the channel?

\[ C = W \log_2 \left( 1 + \frac{P}{W N_0} \right) \] (bits/second)

[Bandwidth \( W \), \( h=1 \), spectral density \( N/2 \)]

• algebraic codes
• convolutional codes
• iterative codes (LDPC, turbo)
Outline

- Information theory - what, why, when
- Two-way channel - channel coding
- Wireless channels and networks
- Two-way cellular-like networks
- Two-way relay channels - canonical example of wireless network coding

Use?

- Benchmark for performance of practical systems
- Guideline in designing systems - what's worth shooting for?
- Theoretical insights can lead to practical insights

One-way channel capacity - notation

\[ X^n \xrightarrow{p(y|x)} Y^n \]

\[ X \xrightarrow{p(y|x)} Y \]

\[ 1 \xrightarrow{p(y|x)} 2 \]

One-way channel capacity

\[ C = \max_{p(x)} I(X; Y) \]

bits/channel use

\[ I(X; Y) = \sum_{x,y} p(x,y) \log \left( \frac{p(x,y)}{p(x)p(y)} \right) \]

symmetric in its arguments!

So now what?

Unsolved

Fundamental

Two-way channel (historical)

\[ x_1 \xrightarrow{y_1} y_2 \]

\[ x_2 \xrightarrow{y_2} y_1 \]

[Shannon '61]
Two-way channel capacity

One-way

\[ x \rightarrow p(y|x) \rightarrow y \]

Two-way

\[ x_1 \rightarrow p(y_1|x_1, x_2) \rightarrow y_1 \]
\[ x_2 \rightarrow y_2 \]

When is

Two-way

\[ W_{12} \rightarrow R_{12} \rightarrow \hat{W}_{12} \]
\[ \hat{W}_{21} \rightarrow R_{21} \rightarrow W_{21} \]

When is\[ W_{12} = W_{21} \] equal to \[ \hat{W}_{12} = \hat{W}_{21} \]

Models for two-way adaptation

One-way: no adaptation possible

\[ x_1^n(w_{12}) \rightarrow W_{12} \rightarrow \hat{W}_{12} \]

Two-way: no adaptation

\[ x_1^n(w_{12}) \rightarrow W_{12} \rightarrow \hat{W}_{12} \]
\[ x_2^n(w_{21}) \rightarrow W_{21} \rightarrow \hat{W}_{21} \]

“Restricted two-way channel”

Two-way: full adaptation

\[ x_1^n(w_{12}, y_1^{n-1}) \rightarrow W_{12} \rightarrow \hat{W}_{12} \]
\[ x_2^n(w_{21}, y_2^{n-1}) \rightarrow W_{21} \rightarrow \hat{W}_{21} \]

Duplex

Two-way: half duplex

Two-way: full duplex

(Aside: \( I(X;Y) \) and \( I(X;Y|Z) \))

\[ I(X;Y) = \sum_{x \in X, y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \]
\[ = E_{p(x,y)} \log \frac{p(X,Y)}{p(X)p(Y)} \]

\[ I(X;Y|Z) = \sum_{x \in X, y \in Y, z \in Z} p(x,y,z) \log \frac{p(x,y|z)}{p(x|z)p(y|z)} \]
\[ = E_{p(x,y,z)} \log \frac{p(X,Y|Z)}{p(X|Z)p(Y|Z)} \]
(Aside: $I(X;Y)$ and $I(X;Y|Z)$)

\[ I(X;Y) = H(Y) - H(Y|X) = \log_2(4)/4 + \log_2(2)/2 + \log_2(4)/4 - \log_2(2) = 0.5 \text{ bits} \]

\[ I(X;Y|Z) = H(Y|Z) - H(Y|X,Z) = \log_2(2) - 0 = 1 \text{ bit} \]

General results

- **Inner bound**
  \[
  R_1 \leq I(X_1;Y_2|X_2) \\
  R_2 \leq I(X_2;Y_1|X_1)
  \]
  where $X_1$ and $X_2$ follow the joint distribution $p(x_1, x_2) = p(x_1)p(x_2)$.

- **Outer bound**
  \[
  R_1 \leq I(X_1;Y_2|X_2) \\
  R_2 \leq I(X_2;Y_1|X_1)
  \]
  where the joint distribution of random variables $X_1$ and $X_2$ is $p(x_1, x_2)$.

When is capacity known

- Parallel two-way channel
- Mod-2 adder
- Two-way restricted channel
- Two-way “push-to-talk” channel
- Two-way Gaussian noise channel (full & half duplex, restricted & unrestricted)

When is capacity unknown

- General unrestricted discrete memoryless channels
- Binary multiplier channel (BMC)

Capacity: binary mod-2 adder channel

\[ y_1 = y_2 = x_1 + x_2 \mod 2. \]

How to achieve capacity region?

Achieving mod 2 adder channel capacity

**Receiver 1:**
\[
\begin{array}{c|c|c|}
 y_1 & 0 & 1 \\
\hline
 x_1 & 0 & 0 & 1 \\
1 & 1 & 0
\end{array}
\]

**Receiver 2:**
\[
\begin{array}{c|c|c|}
 y_2 & 0 & 1 \\
\hline
 x_1 & 0 & 1 & 0 \\
1 & 1 & 0
\end{array}
\]

**Exploit Two-Way!**
### Capacity: restricted channel

- **Capacity region:**
  \[R_1 \leq I(X_1; Y_2|X_2)\]
  \[R_2 \leq I(X_2; Y_1|X_1)\]

  where \(X_1\) and \(X_2\) follow the joint distribution \(p(x_1, x_2) = p(x_1)p(x_2)\).

  \[\text{[Shannon '61]}\]

### Capacity: Gaussian noise channel

- **Power constraint \(P_1\):**
  \[Y_1 = aX_1 + bX_2 + N_1 \sim \mathcal{N}(0, \sigma^2_1)\]

- **Power constraint \(P_2\):**
  \[Y_2 = cX_1 + dX_2 + N_2 \sim \mathcal{N}(0, \sigma^2_2)\]

- **Capacity region:**
  \[R_1 \leq (1/2) \log(1 + c^2 P_1/\sigma^2_2)\]
  \[R_2 \leq (1/2) \log(1 + b^2 P_2/\sigma^2_1)\]

  **No dependence on "a" or "d"** \[\text{[Han '84]}\]

### Capacity: "push-to-talk" channel

- **Two-way: half duplex**

- **Capacity: Gaussian noise channel**

  \[R_1 \leq (1/2) \log(1 + c^2 P_1/\sigma^2_2)\]
  \[R_2 \leq (1/2) \log(1 + b^2 P_2/\sigma^2_1)\]

  **"Feedback" does not help here**

### When is capacity known

- Parallel two-way channel
- Mod-2 adder
- Two-way restricted channel
- Two-way "push-to-talk" channel
- Two-way Gaussian noise channel (full & half duplex, restricted & unrestricted)

### When is capacity unknown

- General unrestricted discrete memoryless channels
- Binary multiplier channel (BMC)

### Capacity unknown: Binary Multiplier Channel

- **Lower/upper bounds**
  - Shannon's lower bound: 0.01895
  - Haghi-Harjager's lower bound: 0.383
  - Schalkwijk's lower bound: 0.63056
  - Shannon's upper bound: 0.89424
  - Zheng's upper bound: 0.8900

- \(y_1 = y_2 = x_1 x_2\), where \(x_1, x_2 \in \{0, 1\}\).
General techniques for two-way channels

Adaptation

Adaptive codewords

Non-adaptive codewords:

\[ X_1 \in \{0, 1\} \]
\[ Y_1 = X_1 X_2 \]
\[ Y_2 = Y_1 \mod 2 \]
\[ X_2 \in \{0, 1, 2\} \]

Code of user 1
\[ u_{11} = 0 \]
\[ u_{11} = 1 \]

Code of user 2
\[ u_{21} = 0 \]
\[ u_{21} = 1 \]

L=N=3 channel uses

Can we take this adaptation into account?

CAUSAL adaptation
From mutual information to directed information

Need to extend the symmetric $I(X^N; Y^N)$ to account for causally adaptive codewords

Marko/Massey’s “Directed Information”

$$I(X^N \to Y^N) := \sum_{n=1}^{N} I(X^n; Y^n | Y^{n-1})$$

Kramer’s “Causally conditioned Directed Information”

$$I(X^N \to Y^N | Z^N) := \sum_{n=1}^{N} I(X^n; Y^n | Y^{n-1} Z^n)$$

Aside - could this be what we need?

Usage: capacity of two-way channels

$$X_1 \in \{0, 1\}, \quad Y_1 = X_1 X_2, \quad Y_2 = Y_1 \mod 2, \quad X_2 \in \{0, 1, 2\}$$

The capacity region of two-way channel is given by the limit as $L \to \infty$ of the regions

$$R_1 = I(A_1^L \rightarrow Y_2^L | X_2^L), \quad R_2 = I(A_2^L \rightarrow Y_1^L | X_1^L)$$

where $A_1^L, A_2^L$ are adaptive independent codewords.

Difficulty: space of codewords hard to compute!

Take away points - AWGN two-way channel

- If have half-duplex constraint and memoryless channels, time-share
- If have full-duplex - obtain two parallel clean channels

For applications - full duplex gains a lot!

Take away points - Discrete memoryless two-way channel

- If have half-duplex constraint (“push-to-talk”), time-share
- If have parallel two-way channels, mod-2 adder
- If have restricted channel
  $$R_1 \leq I(X_1; Y_1(X_2))$$
  $$R_2 \leq I(X_2; Y_2(X_1))$$

where $X_1$ and $X_2$ follow the joint distribution $p(x_1, x_2) = p(x_1)p(x_2)$.

In general may need adaptive codewords

In general OPEN PROBLEM
Two-way source coding = what skipping

- Two-way (lossy) source coding using block-coding and protocols of K rounds
- Set reconciliation framework
- Interactive communication framework / communication complexity

Outline

- Information theory - what, why, when
- Source coding, channel coding, entropy and mutual information, capacity, Gaussian noise channel
- Two-way channel - channel coding
- Wireless channels and networks
- Two-way cellular-like networks
- Two-way relay channels - canonical example of wireless network coding

Point-to-point

- Channel capacity ✔
- How to approach it for memoryless Gaussian noise channels ✔

Is that the end of the story?

NO! Motivation 1: two-way channels

Unsolved
Fundamental

NO! Motivation 2 - networks
Two-way networks!

Capacity and capacity regions

- Point to point capacity
- Multi-user capacity region

Capacity regions

Achievable rate region

• Propose a coding scheme (random codes!

Outer bound

Capacity regions

Achievable region

• Prove that as long as \( R_1 \leq I(X_1; Y|X_2) \) \( R_2 \leq I(X_2; Y|X_1) \) \( R_1 + R_2 \leq I(X_1, X_2; Y) \)

• Limit of communication, NOT how to achieve it in practice necessarily!

• However, benchmark and guidance in practical designs

- Prove that error is bounded away from 0 when \( \uparrow \) not satisfied
- Find a more capable channel whose capacity is known
- Be creative!
The multiple-access channel (MAC) is a channel where multiple transmitters share the same channel bandwidth to send information to a common receiver. The resulting rate pairs are plotted in Figure 8.4.

\[ p(x_1, x_2) = p(x_1)p(x_2) \]

\[ p(y|x_1, x_2) \]

- Introduced by Shannon in 1961
- Capacity known for discrete and Gaussian noise channels
  - Capacity [Ahlswede '71, Liao '72]
  - MIMO [Telatar '99]
  - Fading [Gallager '94, Shamai-Wyner '97, Tse-Hanly '98]

**Three key multi-user channels**

- Broadcast channel
- Relay channel
- Multiple-access channel

**Multiple-access channel (MAC)**

- Capacity region is the closure of the convex hull of all rate pairs \((R_1, R_2)\) satisfying

\[
R_1 \leq I(X_1; Y | X_2) \\
R_2 \leq I(X_2; Y | X_1) \\
R_1 + R_2 \leq I(X_1, X_2; Y)
\]

for some distribution

\[ p(x_1, x_2, y) = p(x_1)p(x_2)p(y|x_1, x_2) \]

**Gaussian MAC**

**Optimal multiple access**

**TDM/FDM multiple access**

\[
R_1 = \frac{1}{2} \log (1 + P_1) \\
R_2 = \frac{1}{2} \log (1 + P_2) \\
R_1 + R_2 < \frac{1}{2} \log_2 (1 + P_1 + P_2)
\]

- GSM
- CDMA
- WiFi

**Multiple access channels in practice**
Capacity region is the closure of the convex hull of all rate pairs \((R_1, R_2)\)

- 3 types of forwarding:
  - Decode+forward (DF), Compress+forward (CF), Amplify+forward (AF)

**Best achievable rate region: Marton’s region**

- Capacity region is the closure of the convex hull of all rate pairs \((R_1, R_2)\) satisfying

\[
0 \leq R_1 \leq I(U_1; Y_1) \\
0 \leq R_2 \leq I(U_2; Y_2) \\
R_1 + R_2 \leq I(U_1; Y_1) + I(U_2; Y_2) - I(U_1, U_2)
\]

for some distribution

\[
p(u_1, u_2, x, y_1, y_2) = p(u_1, u_2) p(x | u_1, u_2) p(y_1, y_2 | x)
\]

**Relay channel**

- Introduced by Van der Meulen in 1968

- Capacity known special cases:
  - Physically degraded relay channels [Cover, El Gamal ’79]

- 3 types of forwarding:
  - Decode+forward (DF), Compress+forward (CF), Amplify+forward (AF)

**Broadcast channel**

- Introduced by Cover in 1972

- Capacity known special cases:
  - Degraded broadcast channels [Bergmans ’73/74, Gallager ’74]
  - General BC with degraded message sets [Korner + Marton ’77]
  - Gaussian MIMO broadcast channel [Weingarten, Steinberg, Shamai ’06]

  - Best achievable rate region [Marton ’79]
MAC with feedback

No feedback

With feedback

Outline

- Information theory - what, why, when
  Source coding, channel coding, entropy and mutual information, capacity, Gaussian noise channel
- Two-way channel - channel coding
  Adaptive codewords, capacity in Gaussian noise = two parallel channels
- Wireless channels and networks
  MAC, relay, BC channels
- Two-way cellular-like networks
- Two-way relay channels - canonical example of wireless network coding

Two-way MAC/BC channels

- account for feedback / data transmission tradeoff in a MAC/BC channel

MAC with feedback results

- Discrete Memoryless MAC
  - capacity region with feedback not known
  - with perfect feedback [Cover+Leung '81], [Bross+Lapidoth '05]
  - with generalized feedback [Carleial '82], [King '78]
  - noisy feedback bounds [Gastpar+Kramer '06]

- Gaussian MAC
  - perfect feedback - capacity is known [Ozarow '84]
  - noisy feedback bounds [Gastpar+Kramer '06]
MAC with feedback visualization

Feedback allows the sources to correlate

Two-way framework

- half-duplex nodes

MAC phase - time $\gamma T$

BC phase - time $(1-\gamma)T$

- what if BC phase used only for feedback?
- power $P_3$ is used to feedback

How large does feedback power need to be?

- $R_1 = R_2 = \frac{P_1}{P_2}$ for $P_1 = P_2$
- What is minimum $P_f$ to have a rate gain?
- Cooperative info $= \log(1+\alpha P) - (1-\eta) \log(1+\beta P)$

Two-way is necessary!

- If have data in both directions, then the joint encoding of up+down links outperforms separate treatment of up+down links!
- combine feedback and new information!
Outline

- Information theory - what, why, when
  Source coding, channel coding, entropy and mutual information, capacity, Gaussian noise channel
- Two-way channel - channel coding
  Adaptive codewords, capacity in Gaussian noise - two parallel channels
- Wireless channels and networks
  MAC, relay, BC channels
- Two-way cellular-like networks
  Tradeoff between forward information and feedback
- Two-way relay channels - canonical example of wireless network coding
  - Single flow, single relay
  - Multiple flows, single relay
  - Single flow, multiple relays

Two-way relay channel

Motivation

Telesurgery
Full-duplex


Half-duplex


Direct link


All through the relay


Early work

Relaying type

Decoding and forward

Compress and forward

Amply and forward

Multiple terminals

A. Nosratinia, A. Stegan, and L. T. Tang. Capture region of the two-terminals multiterminals to

Temporal “phases”: who transmits when

Two-way relay channel: half-duplex

Nodes can either transmit or receive but not both.

Better protocol

Are 4 phases needed? NO!
In particular, if the messages of $a$ and $b$ are $w_a$ and $w_b$ respectively and belong to an algebraic group (such as binary addition), then it is sufficient for the relay node to successfully transmit $w_a \oplus w_b$ simultaneously to $a$ and $b$.

**Key exploits**

- "own message side information" at nodes used to cancel out own message
- "overheard side information" available to nodes when not transmitting
- broadcast nature of wireless channels: relay broadcasts one thing, both nodes hear it.

**Four possible protocols**

**(i) DT: Direct Transmission**

**(ii) MABC: Multiple Access Broadcast Channel**
(iii) TDBC: Time Division Broadcast Channel

(iv) HBC: Hybrid Broadcast

Relaying schemes

- **Amplify and Forward (AF)**
- **Decode and Forward (DF)**
- **Compress and Forward (CF)**
- **Mixed Forward**

Amplify and forward (AF)

- The relay sends a scaled version of the signal it receives.
- Very little computation is needed.

Decode and forward (DF)

- The relay decodes both $w_a$ and $w_b$.
- Much computation, and transmitter codebooks are needed at the relay.
Compress and forward (CF)

- The relay compresses/quantizes the received signal.
- Less computation than DF and transmitter codebooks are not needed at the relay.

Comparison of protocols

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Side information</th>
<th>Phase</th>
<th>Interference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MABC</td>
<td>not present</td>
<td>2</td>
<td>present</td>
</tr>
<tr>
<td>TDBC</td>
<td>present</td>
<td>3</td>
<td>not present</td>
</tr>
<tr>
<td>HBC</td>
<td>present</td>
<td>4</td>
<td>present</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relaying Complexity Noise Relay needs</th>
</tr>
</thead>
<tbody>
<tr>
<td>AF</td>
</tr>
<tr>
<td>DF</td>
</tr>
<tr>
<td>CF</td>
</tr>
<tr>
<td>Mixcl</td>
</tr>
</tbody>
</table>

Mixed Forward (MF)

- The relay decodes \( w_a \) and compresses \( w_b \), combines them into a new message \( w_r \) according to a bijective function, which it encodes and transmits.

Achievable rate regions: one example

- **Theorem 1**: The capacity region of the half-duplex bi-directional relay channel with the MABC protocol is the union of

\[
R_a < \min \left\{ \Delta_1 I(X_a^{(1)}, Y_b^{(1)}|X_b^{(1)}, Q), \Delta_2 I(X_a^{(2)}, Y_b^{(2)}|Q) \right\}
\]

\[
R_b < \min \left\{ \Delta_1 I(X_b^{(1)}, Y_b^{(1)}|X_b^{(1)}, Q), \Delta_2 I(X_b^{(2)}, Y_b^{(2)}|Q) \right\}
\]

\[
R_a + R_b < \Delta_1 I(X_a^{(1)}, X_b^{(1)}, Y_b^{(1)}|Q)
\]

over all joint distributions \( p(q)p^{(1)}(x_a|q)p^{(1)}(x_b|q)p^{(2)}(x_r|q) \) with \( |Q| \leq 5 \).

Achievable rate regions: an example

Outer bounds: cut-set bound

If the rates \( R^{(1)} \) are achievable with a protocol \( P \) and \( R_{\Sigma}(S \rightarrow S') \) denotes the total rate of independent information sent from set \( S \) to set \( S' \) then for all sets \( S \):

\[
R_{\Sigma}(S \rightarrow S') \leq \sum_i \Delta_1 I(X_i^{(i)}(S); Y_i^{(i)}(S)|X_i^{(i)}(S), Q).
\]
Simulations for the Gaussian noise channel

Gaussian simulations

$h_{ar} = h_{br} = 1$, $h_{ab} = 0.2$, $N = 1$, and $P = 50$ dB.

Recent developments

- Capacity is known to within a constant # of bits in Gaussian noise

Gaussian simulations

Relation to network coding?

- bit-level / packet level network coding $\rightarrow$ Decode and Forward (DF)

- excellent systems-level demonstration of 2-way relaying gains (all layers, actual tested)

- physical / analog network coding $\rightarrow$ similar to Amplify and Forward (AF)

- excellent systems-level demonstration of analog network coding (all layers, actual tested)
Key exploits

- "own message side information" at nodes used to cancel out own message
- "overheard side information" available to nodes when not transmitting
- broadcast nature of wireless channels: relay broadcasts one thing, both nodes hear it.

Open questions

- Will it be used beyond academic demonstrations?

Outline

- Information theory - what, why, when
  Source coding, channel coding, entropy and mutual information, capacity, Gaussian noise channel
- Two-way channel - channel coding
  Adaptive codewords, capacity in Gaussian noise = two point channels
- Wireless channels and networks
  MAC, relay, BC channels
- Two-way cellular-like networks
  Tradeoff between forward information and feedback
- Two-way relay channels - canonical example of wireless network coding
  - Single flow, single relay
  - Multiple flows, single relay
  - Single flow, multiple relays

Multiple terminals

Arbitrary (m) number of end users
Half-duplex nodes

Decode + forward relay

Compress + forward end user cooperation

Per-flow network coding of messages at relay

"Protocols" = time "phases"

Protocol 1: FMABC (Full MAC then BC)

Phase 1 = MAC phase

Phase 2 = BC phase

Protocol 2: PMABC (Partial MAC then BC)

Phase 1

Phase 2

Phase 3

Phase 4 = BC phase

Protocol 3: FTDBC (Full Time Division then BC)

Phase 1

Phase 2

Phase 3

Phase 4

Phase 5

To be presented at ISIT 2010.
Protocol 4: PTDBC (Partial Time Division then BC)

Phase 1

Phase 2

Phase 3

Node number

Time

Multiple-access period

Broadcast period

Phase 1 Phase 2 Phase 3

(d) PTDBC protocol

Which protocol is “better”?

1. Extended Marton’s region for broadcasting
2. Per-flow network coding
3. Random-binning to exploit side-information
4. Terminal node cooperation

Marton’s region

Extended Marton’s

Per-flow Network coding (N)

Which protocol is “better”?

1. Extended Marton’s region for broadcasting
2. Per-flow network coding
3. Random-binning to exploit side-information
4. Terminal node cooperation

\[ Y_1 - Y_2 (U_1, U_2) - X \]

\[ R_1 \leq I(U_1; Y_1) \]

\[ R_2 \leq I(U_2; Y_2) \]

\[ R_1 + R_2 \leq I(U_1; Y_1) + I(U_2; Y_2) - I(U_1; U_2) \]

over all joint distributions \( p(u_1, u_2, x) \)

\[ R_1 \leq I(U_1; Y_1) \]

\[ R_2 \leq I(U_2; Y_2) \]

\[ R_1 + R_2 \leq I(U_1; Y_1) + I(U_2; Y_2) - I(U_1; U_2) \]

over all joint distributions \( p(u_1, u_2, x) \)

\[ w_r = w_1 \oplus w_2 \]

\( w_1 \)

\( w_2 \)
Per-flow Network coding (N)

\[ w_r = w_1 \oplus w_2 \]

\[ w_{r,1} = w_{1,0} \oplus w_{0,1} \]

\[ w_{r,2} = w_{2,0} \oplus w_{0,2} \]

\[ x_r(w_{r,1}, w_{r,2}) \]

Example of per-flow Network coding

**FMABC**

\[ R_{0,0,0} \leq \Delta_3 I(U_{2,0}^{(2)}, Y_{2,0}^{(2)}) \]

\[ R_{0,1,0} \leq \Delta_3 I(U_{2,0}^{(2)}, Y_{2,0}^{(2)}) \]

\[ R_{1,0,0} \leq \Delta_3 I(U_{2,0}^{(2)}, Y_{2,0}^{(2)}) \]

**FMABC-N**

\[ R_{0,0,0} \leq \Delta_3 I(U_{2,0}^{(2)}, Y_{2,0}^{(2)}) \]

\[ R_{0,1,0} \leq \Delta_3 I(U_{2,0}^{(2)}, Y_{2,0}^{(2)}) \]

\[ R_{1,0,0} \leq \Delta_3 I(U_{2,0}^{(2)}, Y_{2,0}^{(2)}) \]

\[ R_{0,1,1} \leq \Delta_3 I(U_{2,0}^{(2)}, Y_{2,0}^{(2)}) \]

Which protocol is “better”?

1. Extended Marton’s region for broadcasting
2. Per-flow network coding
3. Random-binning to exploit side-information
4. Terminal node cooperation

Random binning (R) for exploiting overheard information

\[ R_{1,0} \leq \Delta_1 I(X_1^{(1)}; Y_r^{(1)}) \]

\[ R_{1,0} \leq \Delta_1 I(X_1^{(1)}; Y_0^{(1)}) + \Delta_2 I(X_r^{(3)}; Y_0^{(3)}) \]

\[ R_{0,1} \leq \Delta_2 I(X_0^{(2)}; Y_r^{(2)}) \]

\[ R_{0,1} \leq \Delta_2 I(X_0^{(2)}; Y_1^{(2)}) + \Delta_3 I(X_r^{(3)}; Y_1^{(3)}) \]

Random binning and Marton’s binning

\[ I(x_0, y_0, z_0) \]

\[ I(x_1, y_1, z_1) \]
Which protocol is “better”?

1. Extended Marton’s region for broadcasting
2. Per-flow network coding
3. Random-binning to exploit side-information
4. Terminal node cooperation
The PTDBC protocol is outer bounded by the set of all points for all choices of the joint distribution.

### A. Channel model

For each channel use, the corresponding mathematical channel model is:

$$\text{Theorem 20:}$$

These outer bounds will be evaluated in Gaussian noise, as described next.

In this section, we evaluate the achievable rate regions and outer bounds obtained in the previous

- Relay
- Multiple Access
- $0.15$
- $X$
- $0.2$
- $1.5$
- $2.5$
- $0.5$
- $X$
- $X$
- $0.5$
- $0.45$
- $X$
- $2.5$
- $2$
- $4.5$
- $X$
- $0$
- $i$
- $\{\}$

$\{\}$

Outer bounds - half-duplex cut-set

**FMABC**

$$R_{01} + R_{02} \leq \Delta_1(I(X_0^{(1)}; Y_0^{(1)}|X_1^{(1)}, X_2^{(2)})$$

$$R_{20} + R_{21} \leq \Delta_2(I(X_2^{(2)}; Y_0^{(2)})$$

$$R_{10} \leq \Delta_1(I(X_1^{(1)}; Y_1^{(1)}|X_1^{(1)}, X_2^{(2)})$$

$$R_{20} \leq \Delta_2(I(X_2^{(2)}; Y_2^{(2)}|X_1^{(1)}, X_2^{(2)})$$

$$R_{01} + R_{02} \leq \Delta_2(I(X_2^{(2)}; Y_1^{(1)}|X_1^{(1)}, X_2^{(2)})$$

$$R_{02} \leq \Delta_2(I(X_2^{(2)}; Y_2^{(2)})$$

Evaluate and optimize over

- phase durations
- correlation matrices of Marton binning RVs subject to power constraints
- compression parameters

### Table 1

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Multiple Access</th>
<th>Marton’s Broadcast</th>
<th>Network coding</th>
<th>Random binning</th>
<th>User cooperation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>FMABC</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>FMABC-N</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>PMABC-N</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>FMABC-NR</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>PMABC-NR</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>FTDTC-N</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>FTDTC-NR</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>PTDBC</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>PTDBC-NR</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

$N = \text{Network coding}$

$R = \text{Random binning}$

$C = \text{Cooperation between terminals}$

### Simulations in Gaussian noise

$$Y[k] = HX[k] + Z[k]$$

$$H_1 = \begin{bmatrix} 0 & 0.3 & 0.05 & 1 \\ 0.3 & 0 & 1.5 & 1 \\ 0.05 & 1.5 & 0 & 0.2 \\ 1 & 1 & 0.2 & 0 \end{bmatrix}, H_2 = \begin{bmatrix} 0 & 0.9 & 0.4 & 1 \\ 0 & 0 & 0.02 & 1 \\ 0.02 & 0.0 & 0.5 \\ 1 & 1 & 0.5 & 0 \end{bmatrix}$$

### Evaluate and optimize over

- phase durations
- correlation matrices of Marton binning RVs subject to power constraints
- compression parameters

### Simple - MB - MB+NR

### Network coding + random binning

**FMABC**

**PMABC**

**FTDBC**

**FMABC/PMABC**
Multi-flow take-away points

- Most schemes use per-flow network coding

\[
w_r = w_1 \oplus w_2
\]

One flow

\[
w_{r,1} = w_{1,0} \oplus w_{0,1}
\]
\[
w_{r,2} = w_{2,0} \oplus w_{0,2}
\]
\[
x_r(w_{r,1}, w_{r,2})
\]

Multiple flows

Multi-flow take-away points

- Most schemes use per-flow network coding

- Significantly more complex; protocols and opportunities abound. Only starting to understand when to do what.

- Due to practical relevance - crucial to develop insight and thereafter demonstrations (there are none)

Multi-flow take-away points

- Most schemes use per-flow network coding

- Significantly more complex; protocols and opportunities abound. Only starting to understand when to do what.

- Due to practical relevance - crucial to develop insight and thereafter demonstrations (there are none)

- Two-way nature must be explicitly accounted for (side-information, ability to network code and broadcast) in order to see gains.
Outline

- Information theory - what, why, when
  - Source coding, channel coding, entropy and mutual information, capacity, Gaussian noise channel
- Two-way channel - channel coding
  - Adaptive codewords, capacity in Gaussian noise, two parallel channels
- Wireless channels and networks
  - MAC, relay, BC channels
- Two-way cellular-like networks
  - Tradeoff between forward information and feedback
- Two-way relay channels - canonical example of wireless network coding
  - Single flow, single relay
  - Multiple flows, single relay
  - Single flow, multiple relays

Assumptions

- Half-duplex nodes -> protocols in time
- Decode and Forward (DF) and Amplify and Forward (AF) only
- What rates are achievable, what protocols are best?

Protocol 1: (m,2) MABC

Protocol 2: (m,3) TDBC

Protocol 3: (m,m+2) Multi-hop Multi-relay

Multiple relays

Better!

Naive - does not combine messages in 2 directions!
Messages split

\[ w_a \rightarrow w_a(0), w_a(1), \ldots w_a(K) \]

\[ w_b \rightarrow w_b(0), w_b(1), \ldots w_b(K) \]

Initialization

Before communication

After initialization

Exploit overhead information

Performance in Gaussian noise

- In low SNR: (2,4) DF MHMR
- In high SNR: (2,4) DF MHMR or (2,2) AF MABC
Multi-relay take aways

- numerical results indicate that (m, m+2) MHMR protocol with information flowing in 2 directions yields the largest rates

- for low # of hops, or at high SNR, AF relaying may do well, but rapidly degrades as # hops increases or SNR decreases

- fundamentally unsolved, in part due to complexity of space.

Multi-relay take aways

- numerical results indicate that (m, m+2) MHMR protocol with information flowing in 2 directions yields the largest rates

- for low # of hops, or at high SNR, AF relaying may do well, but rapidly degrades as # hops increases or SNR decreases

- fundamentally unsolved, in part due to complexity of space.

- unique two-way characteristics exploited: own message side information to network code (so far only at the bit level rather than the signal level)

Outline

- Information theory - what, why, when
- Two-way channel - channel coding
- Wireless channels and networks
- Two-way cellular-like networks
- Two-way relay channels - canonical example of wireless network coding

Future areas of two-way channels

- one-way information theory "fairly" well understood
- advances in processing power
- never ending desire for bandwidth and limited wireless spectrum

Two-way wireless networks
Questions?

Natasha Devroye
Assistant Professor
University of Illinois at Chicago
http://www.ece.uic.edu/~devroye