Variations on the Index Coding Problem: Caching and Pliable Index Coding

Daniela Tuninetti (UIC),
in collaboration with T. Liu (UIC),
K. Wan (Supelec), and P. Piantanida (Supelec).

Supported by NSF 1527059 (collaborative with C. Fragouli).

University of Illinois at Chicago, danielat AT uic.edu

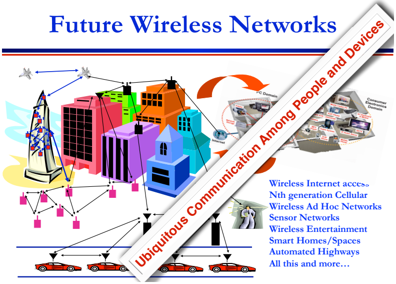
IPAM Emerging Wireless Networks, UCLA, February 2017

Outline of our Work

1. Index Coding:
   - Strict improvement over the composite coding inner bound of [Arbabjolfaei et al. (Jul. 2013)]. To be presented at [(Wan et al., Feb 2017, ITA 2017)].
2. Caching, under the constraint of uncoded cache placement:
   - Outer bounds by using "IC acyclic outer bound" of [Arbabjolfaei et al. (Jul. 2013)].
   - Optimality of [Maddah-Ali and Niesen (May 2014),Maddah-Ali and Niesen (August 2015)] for the case of less users than files. Appeared [(Wan et al., Jul. 2016, ISIT 2016),(Wan et al., Sep. 2016, ITW 2016)].
   - Use inner bound in point 1 as alternative to the one in [Yu et al. (Sep. 2016)] for the case of more users than files so as to matches our outer bounds. To be presented at [(Wan et al., Feb 2017, ITA 2017)].
   - Novel delivery in the finite file size regime [(Wan et al., May 2017, ICC 2017)].
3. Pliable Index Coding:
   - Refined analysis of the fraction of users satisfied by a single transmission from [Brahma and Fragouli (Nov 2015)]. Appeared [(Liu and Tuninetti, Sep. 2016, ITW 2016)].
   - Optimality of the linear schemes in [Brahma and Fragouli (Nov 2015)] when users have side information sets of equal cardinality. Again use IC acyclic outer bound [work in progress].

Motivation, Network Coding (NC), and Index Coding

- Communications in complex wireless networks.
- Challenges: noise, multi-hop (relaying), multi-cast (compound setting), generalized feedback (overheard information interference or "side information"?), etc.
- Simplification: noiseless links without "multiaccess" but still with "broadcast".

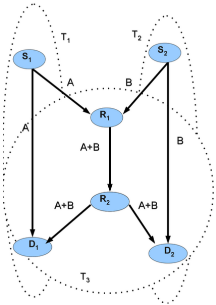


My introduction slide for ECE 437 Wireless Communications.

From S. Goldmsmith's slides.

Network Coding (NC)

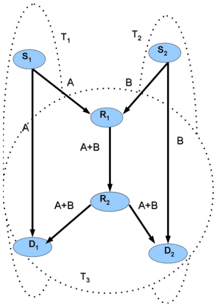
- A network is a directed graph with capacitated edges connecting nodes.
- Source nodes have independent messages destined to subset of destination nodes; wlog multiple unicast.
- Links are noiseless, with broadcast but no multiaccess interference.
- A node can implement *any* function of its received messages.



Example of a NC (Ahlswede, Cai, Li, Yeung [TIT2000]). If relay sends first A and then B, two transmission satisfies both receivers. If instead it sends A+B, one transmission satisfies both receivers, which results in 50% bandwidth savings!

Network Coding (NC)

- NC is **hard** problem.
- Non-linear strategies needed in general; related to entropic region / non-Shannon inequalities.
- NB: an *interference* channel with *relays*, i.e., a combination of two problems believed to be difficult in isolation. We hope to make progress because of its deterministic nature.

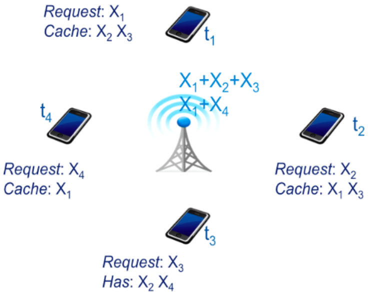


Example of a NC (Ahlswede, Cai, Li, Yeung [TIT2000]). If relay sends first A and then B, two transmission satisfies both receivers. If instead it sends A+B, one transmission satisfies both receivers, which results in 50% bandwidth savings!

NETWORKS INFORMATION COMMUNICATIONS ENGINEERING SYS- TEMS LABORATORY

ELECTRICAL AND COMPUTER ENGINEERING COLLEGE OF ENGINEERING UIC

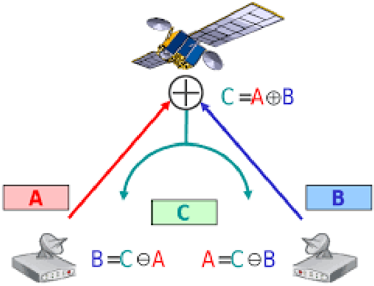
- Special case of NC: one transmitter, no relays, and multiple unicast.
- The TX / server / base station has msgs $\mathcal{X} := \{x_1, ..., x_N\}$.
- RX_j / client_j / user_j requires x_j and knows $\mathcal{S}_j \subseteq \mathcal{X} \backslash \{j\}$.
- Goal: satisfy all clients with the minimum number of transmissions from TX.



Example of a IC. If the TX serves the requested messages one by one, 4 transmissions are necessary. If instead it uses coded transmission, 2 transmissions suffice. 50% bandwidth savings!

Index Coding (IC)

- IC recently attracted LOTS of attention due to its theoretical significance and applications in wireless ad-hoc networks.
- IC is **equivalent** to NC but simpler to describe.
- IC solved for no more than five messages, where linear codes suffice; some 'structured' classes too.
- General inner (graph theoretic quantities or random coding) and outer (cut-set, non-Shannon inequalities) bounds have exponential complexity.



Satellite transmission (Birk and Kol [INFOCOM1998]).

Index Coding

Index Coding: Model

- N (not necessarily unicast) independent messages.
- K (non cooperative) receivers.
- Message $W_j \in [1 : 2^{nR_j}]$, uniform and independent.
- Codeword $X^n(W_1, ..., W_N) \in [1 : 2^n]$, received by all receivers.
- RX_j wants $W_{\mathcal{D}_j} := (W_i : i \in \mathcal{D}_j)$ and has $W_{\mathcal{S}_j} := (W_i : i \in \mathcal{S}_j)$.
- RX_j estimates $\widehat{W_{\mathcal{D}_j}}(X^n, W_{\mathcal{S}_j})$. Assume $\mathcal{S}_j \neq [1 : N], \mathcal{D}_j \neq \emptyset, \ \mathcal{D}_j \cap \mathcal{S}_j = \emptyset$ for all $j \in [1 : K]$.
- Find largest rate region for $(R_1, \ldots, R_N)$ such that probability of error vanishes for each receiver as $n \to \infty$.
- Open problem in general.
- Often one is interested in smallest $1/R_{\text{sym}}$: $(R_1, \ldots, R_N) = (R_{\text{sym}}, \ldots, R_{\text{sym}})$ is achievable.

Index Coding: Example

- Multiple unicast index coding problem with $K = 6$ users and

$$\begin{aligned} \mathcal{D}_1 &= \{A\}, & \mathcal{S}_1 &= \{B, C\}, \\ \mathcal{D}_2 &= \{B\}, & \mathcal{S}_2 &= \{D, E\}, \\ \mathcal{D}_3 &= \{C\}, & \mathcal{S}_3 &= \{A, B\}, \\ \mathcal{D}_4 &= \{D\}, & \mathcal{S}_4 &= \{B, E\}, \\ \mathcal{D}_5 &= \{E\}, & \mathcal{S}_5 &= \{A, F\}, \\ \mathcal{D}_6 &= \{F\}, & \mathcal{S}_6 &= \{C, D\}, \end{aligned}$$

here, for better readability, messages are indicated with the letters A through F.

- Messages are iid binary vectors of length b bits. TX sends

$$\begin{aligned} X^{3b} = (&A \oplus B \oplus C, \\ &B \oplus D \oplus E, \\ &C \oplus D \oplus F). \end{aligned}$$

- RX_5 sums the components in X^{3b} to get $A \oplus E \oplus F$; then, he recovers E.
- Other RX's can directly 'read' the desired message from one component of X^{3b}.
- Rate $R_{\text{sym}} = 1/3$ is achievable; or, $1/R_{\text{sym}} = 3$ transmissions suffice to satisfy all users.

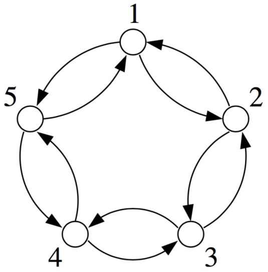
Index Coding: Outer Bound 1

- RX_j desires $\mathcal{D}_j$ and has $\mathcal{S}_j$.
- Theorem: Outer bound

$$\sum_{\ell \in \mathcal{D}_j} R_\ell \le T_{\mathcal{S}_j} - T_{\mathcal{D}_j \cup \mathcal{S}_j}$$

for some monotonic submodular function $T_{\{\cdot\}}$ such that $T_\emptyset \le 1$. Proof: (Fano)

$$\sum_{\ell \in \mathcal{D}_j} R_\ell \le \frac{1}{n} I\left(W_{\mathcal{D}_j}; X^n | W_{\mathcal{S}_j}\right)$$
$$= \frac{1}{n} H\left(X^n | W_{\mathcal{S}_j}\right)$$
$$- \frac{1}{n} H\left(X^n | W_{\mathcal{D}_j \cup \mathcal{S}_j}\right)$$



Index Coding Problem Example.

$R_1 \le T_{2,5} - T_{2,5,1}$,

$R_2 \le T_{1,3} - T_{1,3,2}$,

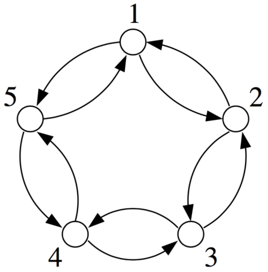
etc.

The optimize over all possible submodular function $T_{\{\cdot\}}$.

Index Coding: Outer Bound 2

- Unicast case: RX_j desires $\mathcal{D}_j = \{j\}$ and has $\mathcal{S}_j$.
- Directed graph representation of side information sets.
- Corollary [**Acyclic Outer Bound**]: (looser bound but easier to evaluate)

$$\sum_{\substack{\ell \in \mathcal{J}: \\ \mathcal{J} \text{ is subgraph} \\ \mathcal{J} \text{ not a d-cyclic}}} R_\ell \le \max H(X) = 1.$$



Index Coding Problem Example.

Corollary: $R_1 + R_3 \le 1$, etc. $(2 \le 1/R_{\text{sym}})$.

Theorem: $R_1 + R_2 + R_3 + R_4 + R_5 \le 2$. $(2.5 \le 1/R_{\text{sym}})$.

Index Coding: Example

- Multiple unicast index coding problem with $K = 6$ users and

$$\mathcal{D}_1 = \{A\}, \quad \mathcal{S}_1 = \{B, C\},$$
$$\mathcal{D}_2 = \{B\}, \quad \mathcal{S}_2 = \{D, E\},$$
$$\mathcal{D}_3 = \{C\}, \quad \mathcal{S}_3 = \{A, B\},$$
$$\mathcal{D}_4 = \{D\}, \quad \mathcal{S}_4 = \{B, E\},$$
$$\mathcal{D}_5 = \{E\}, \quad \mathcal{S}_5 = \{A, F\},$$
$$\mathcal{D}_6 = \{F\}, \quad \mathcal{S}_6 = \{C, D\},$$

  here, for better readability, messages are indicated with the letters A through F.

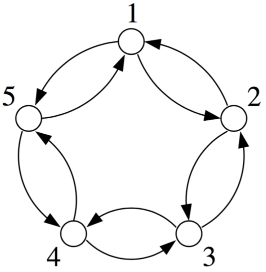
- Messages A, D, F form an acyclic subgraph. [I could not draw a non-messy figure!]
- Also: let users 1 and 4 cooperate == super-user with $\mathcal{D} = \{A, D\}$ and $\mathcal{S} = \{B, C, E\}$.
  After the super-user recovers D, it can mimic user 6, and thus recover F. Thus

$$R_1 + R_4 + R_6 \leq \frac{1}{n} I(A, D, F; X^n | B, C, E) \leq 1,$$

  or $R_{\text{sym}} \leq 1/3$.

Index Coding: Composite Coding

- Two-step encoding:
  1) Generate a ‘composite index’ $X^n_{\mathcal{P}}(W_{\mathcal{P}}) \in [1 : 2^{nS_{\mathcal{P}}}]$, $\forall \mathcal{P} \subseteq [1 : N]$.
  2) Bin all composite indices in $X^n \in [1 : 2^n]$.
- Two-step decoding:
  1) RX_j recovers all composite indices in X^n given the message side information $W_{\mathcal{S}_j}$.
  2) RX_j recovers the messages in $\mathcal{D}_j \subseteq \mathcal{K}_j \subseteq [1 : N]\backslash\mathcal{S}_j$, only those in $\mathcal{D}_j$ uniquely.



Index Coding Problem Example.

E1: (W_1, W_2) to index $X_{1,2}$ at rate $S_{1,2}$;

(W_2, W_3) to index $X_{2,3}$ at rate $S_{2,3}$; etc.

E2: $X^n(X_{1,2}, X_{2,3}, X_{3,4}, X_{4,5}, X_{5,1})$ at rate $\max H(X) = 1$.

D1: $S_{1,2} + S_{2,3} + S_{3,4} + S_{4,5} + S_{5,1} \le 1$. Same for every RX.

D2: $R_1 \le S_{1,2} + S_{1,5}$; $R_2 \le S_{1,2} + S_{2,3}$; etc.

FourierMotzkin elimination of the composite index rates ...

Index Coding: Composite Coding

- Achievable region: Composite Coding.
  A non-negative rate tuple $\mathbf{R} := (R_1, \ldots, R_N)$ is achievable if

$$\mathbf{R} \in \bigcap_{j \in [1:K]} \bigcup_{\mathcal{K}_j : \mathcal{D}_j \subseteq \mathcal{K}_j \subseteq [1:N] \setminus \mathcal{S}_j} \mathscr{R}(\mathcal{K}_j | \mathcal{S}_j, \mathcal{D}_j),$$

$$\mathscr{R}(\mathcal{K} | \mathcal{S}, \mathcal{D}) := \bigcap_{\mathcal{J} : \mathcal{J} \subseteq \mathcal{K}, \mathcal{D} \cap \mathcal{J} \neq \emptyset} \left\{ \sum_{i \in \mathcal{J}} R_i < v_{\mathcal{J}} \right\},$$

$$v_{\mathcal{J}} := \sum_{\mathcal{P} : \mathcal{P} \subseteq \mathcal{S} \cup \mathcal{K}, \mathcal{P} \cap \mathcal{J} \neq \emptyset} S_{\mathcal{P}},$$

$$\sum_{\mathcal{J} : \mathcal{J} \in [1:N], \mathcal{J} \not\subseteq \mathcal{S}_j} S_{\mathcal{J}} \leq 1, \quad \forall j \in [1:K].$$

Index Coding: Example

- Multiple unicast index coding problem with $K = 6$ users and

$$\begin{aligned} \mathcal{D}_1 &= \{A\}, & \mathcal{S}_1 &= \{B, C\}, \\ \mathcal{D}_2 &= \{B\}, & \mathcal{S}_2 &= \{D, E\}, \\ \mathcal{D}_3 &= \{C\}, & \mathcal{S}_3 &= \{A, B\}, \\ \mathcal{D}_4 &= \{D\}, & \mathcal{S}_4 &= \{B, E\}, \\ \mathcal{D}_5 &= \{E\}, & \mathcal{S}_5 &= \{A, F\}, \\ \mathcal{D}_6 &= \{F\}, & \mathcal{S}_6 &= \{C, D\}, \end{aligned}$$

here, for better readability, messages are indicated with the letters A through F.

- Composite coding allows for $R_{\text{sym}} \leq 1/4$.
- Proof: no matter the choice of decoding sets $(\mathcal{K}_j, j \in [1:6]) : |\mathcal{K}_j| \leq 4$ the symmetric rate is cannot exceed $R_{\text{sym}} \leq 1/4$.
- Strictly suboptimal compared to the scheme introduced previously.

NETWORKS INFORMATION COMMUNICATIONS ENGINEERING SYS-TEMS LABORATORY

ELECTRICAL AND COMPUTER ENGINEERING COLLEGE OF ENGINEERING UIC

- A non-negative rate tuple $\mathbf{R} := (R_1, \ldots, R_N)$ is achievable if

$$\mathbf{R} \in \bigcap_{j\in[1:K]} \bigcup_{\mathcal{K}_j:\mathcal{D}_j\subseteq\mathcal{K}_j\subseteq[1:N]\setminus\mathcal{S}_j} \mathscr{R}(\mathcal{K}_j|\mathcal{S}_j, \mathcal{D}_j),$$

$$\mathscr{R}(\mathcal{K}|\mathcal{S}, \mathcal{D}) := \bigcap_{\mathcal{J}:\mathcal{J}\subseteq\mathcal{K}, \mathcal{D}\cap\mathcal{J}\neq\emptyset} \left\{ \sum_{i\in\mathcal{J}} R_i < v_{\mathcal{J}} \right\},$$

$$v_{\mathcal{J}} := I\Big((U_i : i \in \mathcal{J}) \ ; \ (X_{\mathcal{P}} : \mathcal{P} \subseteq [1:N]) \Big| (U_i : i \in \mathcal{S}_j \cup \mathcal{K}_j \setminus \mathcal{J}) \Big),$$

$$H\Big((X_{\mathcal{P}} : \mathcal{P} \subseteq [1:N]) \Big| (U_i : i \in \mathcal{S}_j) \Big) \leq 1, \quad \forall j \in [1:K].$$

for some $P_{U_1,\ldots,U_N} = \prod_{i\in[1:N]} P_{U_i}$ and functions $X_{\mathcal{P}}\big((U_i : i \in \mathcal{P})\big)$.

Index Coding: Example

- Multiple unicast index coding problem with $K = 6$ users and

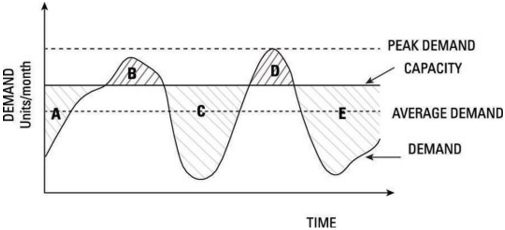
$$\begin{aligned}
\mathcal{D}_1 &= \{A\}, & \mathcal{S}_1 &= \{B, C\}, \\
\mathcal{D}_2 &= \{B\}, & \mathcal{S}_2 &= \{D, E\}, \\
\mathcal{D}_3 &= \{C\}, & \mathcal{S}_3 &= \{A, B\}, \\
\mathcal{D}_4 &= \{D\}, & \mathcal{S}_4 &= \{B, E\}, \\
\mathcal{D}_5 &= \{E\}, & \mathcal{S}_5 &= \{A, F\}, \\
\mathcal{D}_6 &= \{F\}, & \mathcal{S}_6 &= \{C, D\},
\end{aligned}$$

here, for better readability, messages are indicated with the letters A through F.

- New inner bound $R_{\text{sym}} \leq 1/3$.
- Proof: first achieveble scheme.
- Strict improvement over composite coding.

Caching

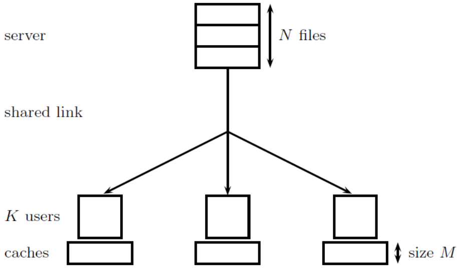
Caching: Motivation



- High temporal variability of network traffic leads to high traffic in peak-traffic times and low traffic in peak-off times.
- Idea: Caching (prefetching) can help to smooth the traffic in peak-traffic times by storing part of the contents in users' local memories during peak-off periods.

Caching: Problem Formulation

- Placement: Cache content creation, without knowledge of later demands. Limit M.
- Delivery: Broadcast packets based on demands and cache contents. Limit R.
- Goal: Minimize the broadcast rate R for the worst case demands.
- Connection to IC: can choose the best side side information for the worst case multicast demands.



Caching system. [Maddah-Ali and Niesen 2016 ITSoc Paper Award].

NETWORKS INFORMATION COMMUNICATIONS ENGINEERING SYS-TEMS LABORATORY

ELECTRICAL AND COMPUTER ENGINEERING COLLEGE OF ENGINEERING UIC

Centralized vs Decentralized placement

- Centralized: Users in placement and delivery phases are the same.
- Decentralized: Otherwise.

Uncoded vs Coded placement

- Uncoded placement: Each user directly stores M out of N files.
- Coded placement: Otherwise.

Maddah-Ali and Niesen (MAN) Coded Caching Schemes

- Centralized: Combinatorial uncoded cache placement, coded delivery. Cut-set outer bound. Optimality within factor ca 4.
- Decentralized: i.i.d. bit storage, coded delivery. Centralized outer bound.

Caching

Centralized vs Decentralized placement

- Centralized: Users in placement and delivery phases are the same.
- Decentralized: Otherwise.

Uncoded vs Coded placement

- Uncoded placement: Each user directly stores M out of N files.
- Coded placement: Otherwise.

Maddah-Ali and Niesen (MAN) Coded Caching Schemes

- Centralized: Combinatorial uncoded cache placement, coded delivery. Cut-set outer bound. Optimality within factor ca 4.
- Decentralized: i.i.d. bit storage, coded delivery. Centralized outer bound.

NETWORKS INFORMATION COMMUNICATIONS ENGINEERING SYS-TEMS LABORATORY

ELECTRICAL AND COMPUTER ENGINEERING COLLEGE OF ENGINEERING UIC

Centralized vs Decentralized placement

- Centralized: Users in placement and delivery phases are the same.
- Decentralized: Otherwise.

Uncoded vs Coded placement

- Uncoded placement: Each user directly stores M out of N files.
- Coded placement: Otherwise.

Maddah-Ali and Niesen (MAN) Coded Caching Schemes

- Centralized: Combinatorial uncoded cache placement, coded delivery. Cut-set outer bound. Optimality within factor ca 4.
- Decentralized: i.i.d. bit storage, coded delivery. Centralized outer bound.

Caching: cMAN and dMAN

- Let $M = t\frac{N}{K}$, $t \in [0:K]$. Files $F_1, \ldots, F_N$ of size B bits.
- Placement phase: each file is split into $\binom{K}{t}$ non-overlapping sub-files of equal size $\frac{B}{\binom{K}{t}}$ bits. The sub-files of F_i are denoted by $F_{i,\mathcal{W}}$ for $\mathcal{W} \subseteq [1:K]$ where $|\mathcal{W}| = t$.
- User k fills his cache as

$$Z_k = \Big(F_{i,\mathcal{W}} : k \in \mathcal{W} \subseteq [1:K], |\mathcal{W}| = t,\ i \in [1:N] \Big).$$

- Delivery phase: the server transmits

$$X_{\mathbf{Z},\mathbf{d}} = \Big(\oplus_{s \in \mathcal{S}} F_{d_s, \mathcal{S} \setminus \{s\}} : \mathcal{S} \subseteq [1:K], |\mathcal{S}| = t+1 \Big),$$

which requires broadcasting $B\binom{K}{t+1}/\binom{K}{t}$ bits.

Caching: cMAN and dMAN

- cMAN [Maddah-Ali and Niesen (May 2014)]

$$R_{\text{cMAN}}[t] := \binom{K}{t+1} / \binom{K}{t}.$$

Removing the redundant transmissions [Yu et al. (Sep. 2016)]

$$R_{\text{c,uncoded placement}}[t] := \left(\binom{K}{t+1} - \binom{K-\min(K,N)}{t+1}\right) / \binom{K}{t}.$$

- dMAN [Maddah-Ali and Niesen (August 2015)]

$$R_{\text{dMAN}}(M) := \left(\frac{N}{M} - 1\right)\left[1 - \left(1 - \frac{M}{N}\right)^{K}\right].$$

Removing the redundant transmissions [Yu et al. (Sep. 2016)]

$$R_{\text{d,uncoded placement}}(M) := \left(\frac{N}{M} - 1\right)\left[1 - \left(1 - \frac{M}{N}\right)^{\min(K,N)}\right].$$

Caching: Our Contributions

- Leveraged the connection between IC and caching (for uncoded placement).
- Used "IC acyclic outer bound" and showed optimality of MAN for uncoded placement for $N \geq K$, or for $N = 2$. Note: Result holds when the total cache size is fixed or/and the total file size is fixed.
- Used new IC inner bound to show optimality of MAN for uncoded placement for $N < K$, as alternate proof to [Yu et al. (Sep. 2016)].
- Consequence: Improvement over MAN can be obtained only by **coded cache placement**.
- Novel delivery scheme for dMAN placement in the finite file size regime [ICC2017].

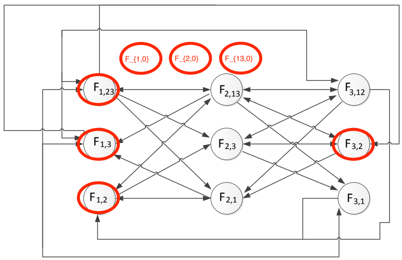
Caching: Outer Bound

- After the users' demands are revealed in a caching scheme with uncoded cache placement, the delivery phase is equivalent to a general multicast index coding problem.
- We used "IC acyclic outer bound" and leveraged the intrinsic symmetries of the caching problem
- Let $\mathcal{N}(\mathbf{d})$ be the set of demanded files. For each $i \in \mathcal{N}(\mathbf{d})$ and for each $\mathcal{W} \subseteq [1:K]$, the sub-file $F_{i,\mathcal{W}}$ (containing the bits of file F_i within the cache of the users indexed by $\mathcal{W}$) is an independent message in the index coding problem with user set $[1:K]$.
- IC with $M = tM/N, t \in [0:K]$, $N = |\mathcal{N}(\mathbf{d})|\binom{K}{t}$ for centralized caching or $N = |\mathcal{N}(\mathbf{d})|(2^t - 1)$ for decentralized caching, and

$$\mathcal{D}_k = \left(F_{d_k,\mathcal{W}} : \mathcal{W} \subseteq [1:K], k \notin \mathcal{W}\right),$$
$$\mathcal{S}_k = \left(F_{i,\mathcal{W}} : \mathcal{W} \subseteq [1:K], i \in \mathcal{N}(\mathbf{d}), k \in \mathcal{W}\right).$$

Caching: Example $N = K = 3$

- Each file is divided into $2^K = 8$ disjoint parts, denoted as $F_{i,\mathcal{W}}$.
- For each demand vector $\mathbf{d}$ we generate an index coding problem with $K2^{K-1} = 12$ independent messages.
- In the constructed graph we want to find sets of nodes not forming a cycle.
- $F_{1,\emptyset}, F_{2,\emptyset}, F_{3,\emptyset}$ are always in such sets $\mathcal{J}$ the acyclic outer bound.



Directed graph representing the equivalent IC. Demand $\mathbf{d} = (1, 3, 2)$; set not containing a cycle is $(F_{1,\emptyset}, F_{1,2}, F_{1,3}, F_{1,23}, F_{3,\emptyset}, F_{3,2}, F_{2,\emptyset})$.

Caching: General Case

- In general, we can find a bound for all possible pairs $(\mathbf{d}, \mathbf{u}) \in$ all permutations of the integers $[1 : K]$.

$$R_{\mathrm{c,u}} \geq \sum_{i \in [0:K]} \frac{\binom{K}{i+1} - \binom{K-\min(K,N)}{i+1}}{\binom{K}{i}} \frac{x_i}{NB}, \quad (1)$$

$$x_t := \sum_{j \in [1:N]} \sum_{\mathcal{W} \subseteq [1:K]:|\mathcal{W}|=t} |F_{j,\mathcal{W}}|, \quad t \in [0:K], \quad (2)$$

where x_t in (2) represents the total number of bits known by t users, and where the constraints on the file size and on the cache size imply

$$x_0 + x_1 + \cdots + x_K = NB, \quad (3)$$

$$0x_0 + 1x_1 + 2x_2 + \cdots + ix_i + \cdots + Kx_K \leq \frac{KM}{N} NB. \quad (4)$$

- Fourier-Motzkin elimination $(x_0, \ldots, x_K)$.

Caching: Main Result: Converse

NETWORKS INFORMATION COMMUNICATIONS ENGINEERING SYS- TEMS LABORATORY

ELECTRICAL AND COMPUTER ENGINEERING COLLEGE OF ENGINEERING UIC

Theorem 1

- *Outer bound for centralized caching systems: lower convex envelope*

$$(M, R) = \left(t\frac{N}{K}, \frac{\binom{K}{t+1} - \binom{K-\min(K,N)}{t+1}}{\binom{K}{t}} \right), \quad t \in [0 : K].$$

- *Outer bound for decentralized caching systems:*

$$R = \frac{1 - \frac{M}{N}}{\frac{M}{N}} \left[1 - \left(1 - \frac{M}{N} \right)^{\min(K,N)} \right].$$

- *Further results: optimality for* $N = 2$.

Theorem 2

- *Inner bound for centralized caching systems: our new IC inner bound matched applied to centralized caching systems attains the outer bound.*
- *Inner bound for decentralized caching systems: our new IC inner bound matched applied to decentralized caching systems attains the outer bound.*

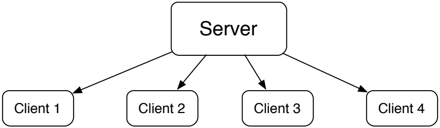
Pliable Index Coding (PICOD)

PICOD: Motivation

- Today's communication networks increasingly deliver *type of content* rather than specific messages. Example: content distribution networks, advertisement networks, social networks, etc.
- PICOD clients are flexible/pliable and happy to receive any one message they do not already have (for instance, any search results they have not already received, etc.).
- Model: Same as IC except that client$_j$ is satisfied by decoding successfully *any* message not in its side information.
- Connection to IC: *choose optimal demands; multi-casting allowed.* Still side information sets fixed.

IC vs PICOD

$$\mathcal{M} = \{b_1, b_2, b_3, b_4\}$$

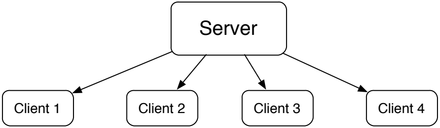


$\mathcal{I}_1 = \{b_2, b_4\}$ $\mathcal{I}_2 = \{b_1, b_3\}$ $\mathcal{I}_3 = \{b_2, b_4\}$ $\mathcal{I}_4 = \{b_1, b_3\}$

$\mathcal{R}_1 = \{b_1\}$ $\mathcal{R}_2 = \{b_2\}$ $\mathcal{R}_3 = \{b_3\}$ $\mathcal{R}_4 = \{b_4\}$

IC optimal transmission: $t_1 = b_1 + b_2, \quad t_2 = b_3 + b_4$.

$$\mathcal{M} = \{b_1, b_2, b_3, b_4\}$$



$\mathcal{I}_1 = \{b_2, b_4\}$ $\mathcal{I}_2 = \{b_1, b_3\}$ $\mathcal{I}_3 = \{b_2, b_4\}$ $\mathcal{I}_4 = \{b_1, b_3\}$

$\mathcal{R}_1 = \mathcal{I}_1^c$ $\mathcal{R}_2 = \mathcal{I}_2^c$ $\mathcal{R}_3 = \mathcal{I}_3^c$ $\mathcal{R}_4 = \mathcal{I}_4^c$

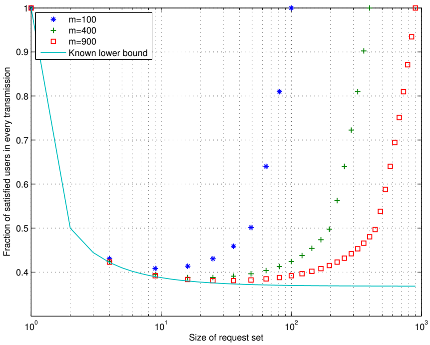
PICOD optimal transmission: $t_1 = b_1 + b_2$. 50% bandwidth saving! ($\mathcal{I}$ side information set, $\mathcal{R}$ possible request set.)

PICOD: Past Work

- m files, n users, $\mathcal{R}$ possible request set, $\mathcal{S}$ side information set.
- PICOD is NP-hard, as general IC.
- Sufficiency of $O(\log n)$ transmissions for $|\mathcal{S}_j| = \text{constant}, \forall j$.
- Sufficiency of $O(\min\{(\log m)(1 + \log^+(\frac{n}{\log m})), m, n\})$ transmissions.
- There exists PICOD's requiring $\Omega(\log n)$ transmissions.
- Achievability: Greedy algorithm with binary linear codes to satisfy the largest fraction of yet-to-be-satisfied clients.

PICOD: Our Work

- Refined analysis of the fraction of clients that can be satisfied by one transmission.
- Deterministic (not probabilistic) approach.
- Strictly better than the known lower bound.
- Capture effect of size of side information set (not seen in IC).
- No converse bounds.



Fraction of satisfied users with a single transmission vs. size of side information set [ITW2016].

PICOD: Main Result

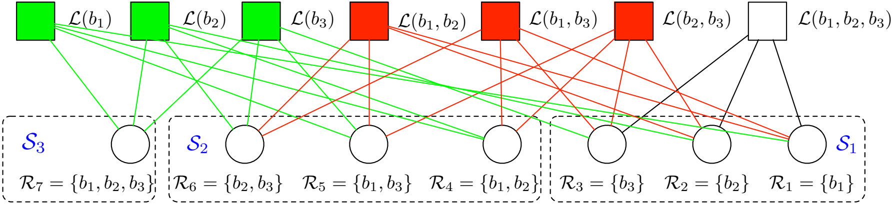
Theorem 3

For a PICOD instance with m messages and where the request set of all n clients has cardinality d, the fraction of unsatisfied clients that can be satisfied by one transmission is lower bounded by
$\max\left\{\left(1-\frac{d}{m}\right)^{\frac{m}{d}-1} e\left(1-\frac{1}{d}\right)^{d-1}, \frac{d}{m}\right\}$*, which is at least* $1/e$.

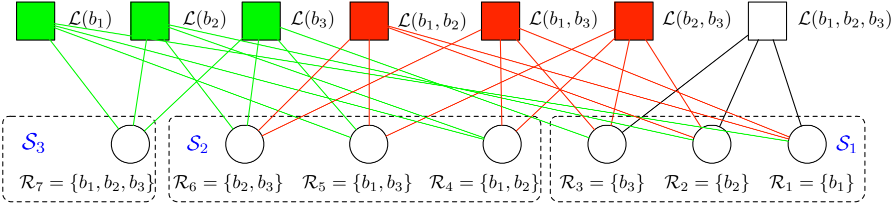
- Refined analysis of the fraction of clients that can be satisfied by one transmission.
- Deterministic approach.

Maximum Number of Satisfied Clients by one Transmission

- m is # of messages; $k \in [1:m]$ is # of messages in linear combinations; $j \in [1:\binom{m}{k}]$ for each k; $D_{kj}, k \in [1:m], j \in [1:\binom{m}{k}]$ number of satisfied clients with one transmission; Group S_d contains all clients with $|\mathcal{D}_i| = d$. Let $C_i = |\mathcal{S}_i|$.



Main Result

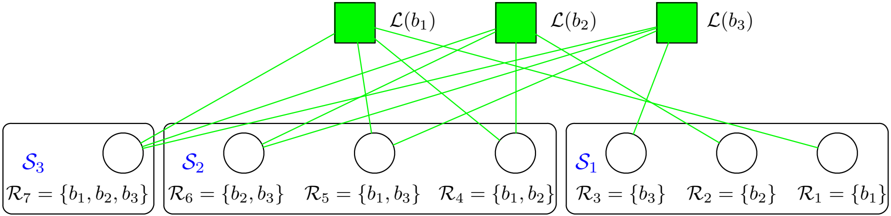


$$\max_{kj} D_{kj} \geq \max_k \frac{D_{k1} + \cdots + D_{k\binom{m}{k}}}{\binom{m}{k}}$$

$$= \max_k \frac{\sum_{i=1}^{m} \binom{i}{1}\binom{m-i}{k-1} C_i}{\binom{m}{k}}$$

$$:= \max_k \sum_{i=1}^{m} R_{i,m,k} C_i$$

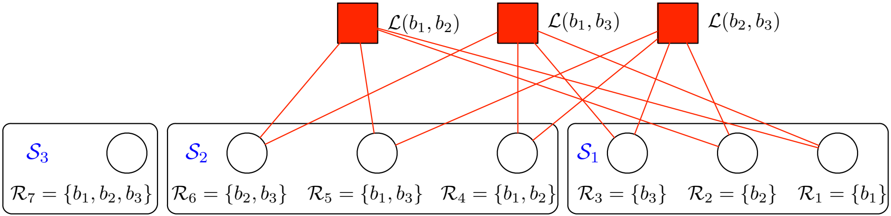
Number of Satisfied Clients by $k = 1$



- Maximum number of client satisfied by one transmission

$$\frac{\sum_{i=1}^{3} \binom{i}{1}\binom{3-i}{1-1} C_i}{\binom{3}{1}} = \frac{C_1 + 2C_2 + 3C_3}{3} = \frac{12}{3} = 4.$$

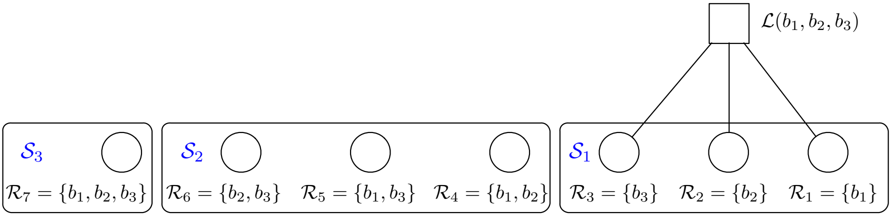
Number of Satisfied Clients by $k = 2$



- Maximum number of client satisfied by one transmission

$$\frac{\sum_{i=1}^{3} \binom{i}{1}\binom{3-i}{2-1} C_i}{\binom{3}{2}} = \frac{2C_1 + 2C_2}{3} = \frac{12}{3} = 4.$$

Number of Satisfied Clients by $k = 3$



- Maximum number of client satisfied by one transmission

$$\frac{\sum_{i=1}^{3} \binom{i}{1}\binom{3-i}{3-1} C_i}{\binom{3}{3}} = \frac{C_1}{1} = 3.$$

We now consider the case where all clients have the same size d for the request set, i.e., $C_d = n, C_i = 0, \forall i \neq d$.

- To derive the maximum fraction of clients that can be satisfied by one transmission, we solve

$$\max_k R_{d,m,k} = \max_k \frac{\binom{d}{1}\binom{m-d}{k-1}}{\binom{m}{k}}.$$

- Note that $R_{d,m,1} = d/m$, i.e., d/m is the fraction of clients that can be satisfied with one transmission.
- For $k \geq 2$, use Stirling's approximation:

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}}$$

and some algebra.

- Finally we can have

$$\max_k R_{d,m,k} > \left(1 - \frac{d}{m}\right)^{\frac{m}{d}-1} e \left(1 - \frac{1}{d}\right)^{d-1}.$$

- Combined with the result for $k = 1$ we have

$$\begin{aligned}\max_k R_{d,m,k} &\geq \max\left\{\left(1 - \frac{d}{m}\right)^{\frac{m}{d}-1} e \left(1 - \frac{1}{d}\right)^{d-1}, \frac{d}{m}\right\} \\ &\geq \left(1 - \frac{1}{\sqrt{m}}\right)^{2\sqrt{m}-2} e \\ &> 1/e.\end{aligned}$$

This proves the main result.

Application: PICOD instance with m messages and n clients, whose side information sets are generated i.i.d. according to a Bern(p) distribution.

- Not exactly the same request set size. However, by the law of large number, for sufficiently large m, almost all clients will be in the "typical client set" defined as

$$\mathcal{T}_{\epsilon}^{(m)} = \left\{ \text{client}_i, \left| \frac{|\mathcal{R}_i|}{m} - (1-p) \right| \leq \varepsilon(1-p) \right\}.$$

- Fraction of clients satisfied by single transmission at least

$$r_{m,p} = \max \left\{ 1-p, p^{\frac{p}{1-p}} \left(1 - \frac{1}{m(1-p)} \right)^{m(1-p)-1} e \right\} > 1/e$$

with high probability.

Computation Complexity

- In PICOD $1 \leq n \leq 2^m - 1$.
- Known algorithms have running time $O\left(mn^2\right)$ [1] and $O(m^2 n \log(n))$ [2].
- Our achievability has running time $O(\frac{2^m}{\sqrt{m}})$, which is independent of n.
- The two known algorithms are faster when n is small, but slower when n is exponential in m.

[1] S. Brahma and C. Fragouli, "Pliable index coding," IEEE Trans. Information Theory, Nov 2015.

[2] L. Song and C. Fragouli, "A deterministic algorithm for pliable index coding," arXiv:1601.05516, Jan. 2016.

Conclusions

- IC is a fundamental open problem in network information theory.
- It is deceivingly simple. It is representative of the general NC.
- It is the building block in several data delivery & management system problems.
- For IC and its variations, new directions needed for:
  - Inner bounds (right now mainly based on linear codes, or too complex / combinatorial).
  - Outer bounds (right now very few techniques but essentially all cut-set-based, and in general too complex / combinatorial).
- New ideas ...
- THANK YOU

References

Fatemeh Arbabjolfaei, Bernd Bandemer, Young-Han Kim, Eren Sasoglu, and Lele Wang. On the capacity region for index coding. *in IEEE Int. Symp. Inf. Theory*, Jul. 2013.

Siddhartha Brahma and Christina Fragouli. Pliable index coding. *IEEE Trans. Information Theory*, 61(11):6192–6203, Nov Nov 2015. ISIT2012.

Tang Liu and Daniela Tuninetti. Pliable index coding: Novel lower bound on the fraction of satisfied clients with a single transmission and its application. *in IEEE Infor. Theory Workshop*, Sep. 2016.

M. A. Maddah-Ali and U. Niesen. Decentralized coded caching attains order-optimal memory-rate tradeoff. *IEEE/ACM Transactions on Networking*, 23(?):1029–1040, August 2015. Available at arXiv:1301.5848.

M. A. Maddah-Ali and U. Niesen. Fundamental limits of caching. *IEEE Trans. Infor. Theory*, 60(5):2856–2867, May 2014. Available at arXiv:1209.5807.

K. Wan, D. Tuninetti, and P. Piantanida. A novel index coding scheme and its application to coded caching. *submitted to IEEE ITA*, Feb 2017.

K. Wan, D. Tuninetti, and P. Piantanida. On caching with more users than files. *in IEEE Int. Symp. Inf. Theory*, pages 135–139, Jul. 2016. Available at arXiv:1601.06383.

K. Wan, D. Tuninetti, and P. Piantanida. Novel delivery schemes for decentralized coded caching in the finite file size regime. *submitted to IEEE ICC*, May 2017. Available at arXiv:1612.01133.

K. Wan, D. Tuninetti, and P. Piantanida. On the optimality of uncoded cache placement. *in IEEE Infor. Theory Workshop*, Sep. 2016. Available at arXiv:1511.02256.

Q. Yu, M. A. Maddah-Ali, and A. Salman. The exact rate-memory tradeoff for caching with uncoded prefetching. *Available at arXiv:1609.07817*, Sep. 2016.