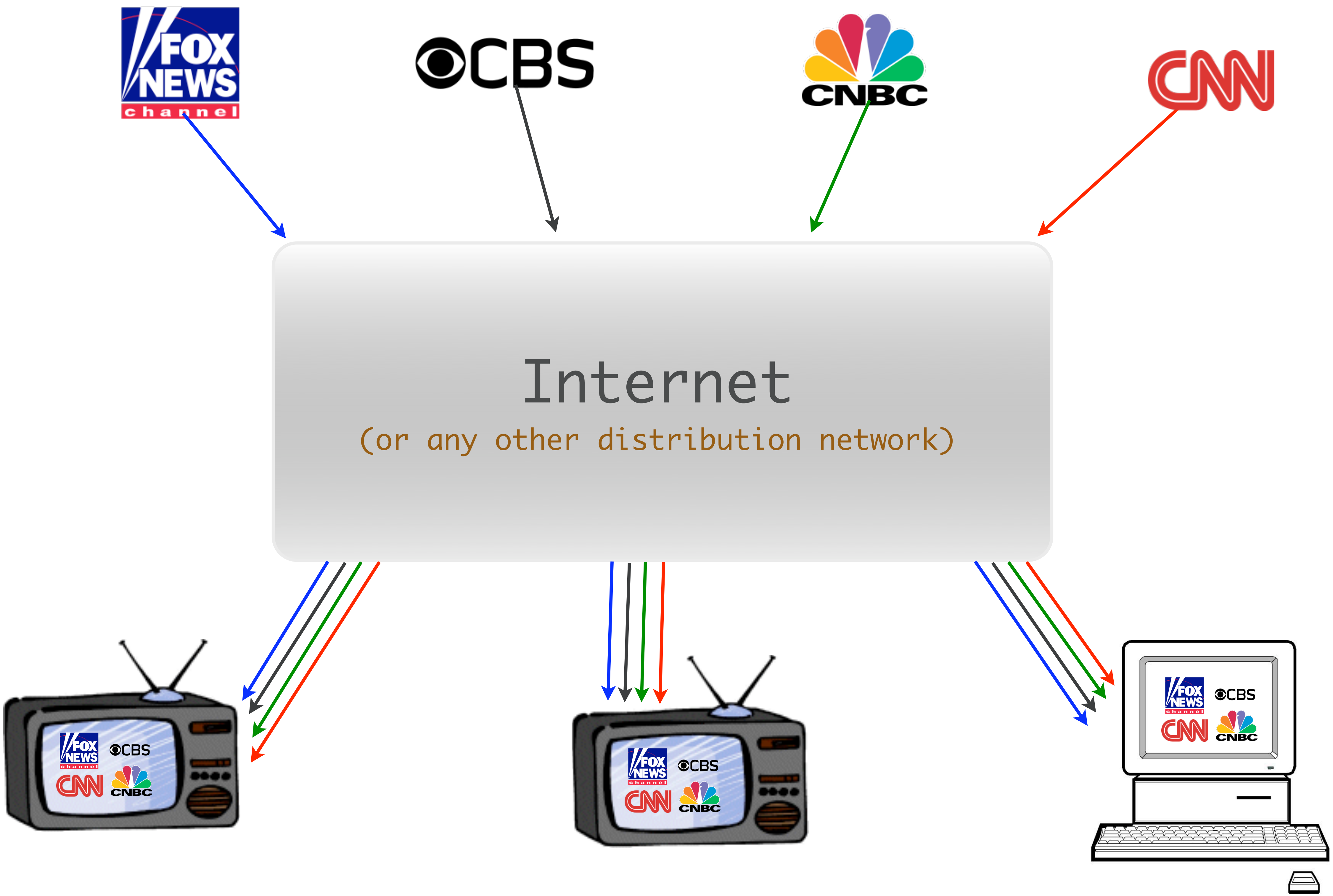
On network capacity and the power of side information

Mayank Bakshi, Michelle Effros
Department of Electrical Engineering,
California Institute of Technology

Motivation



An example where all users want everything

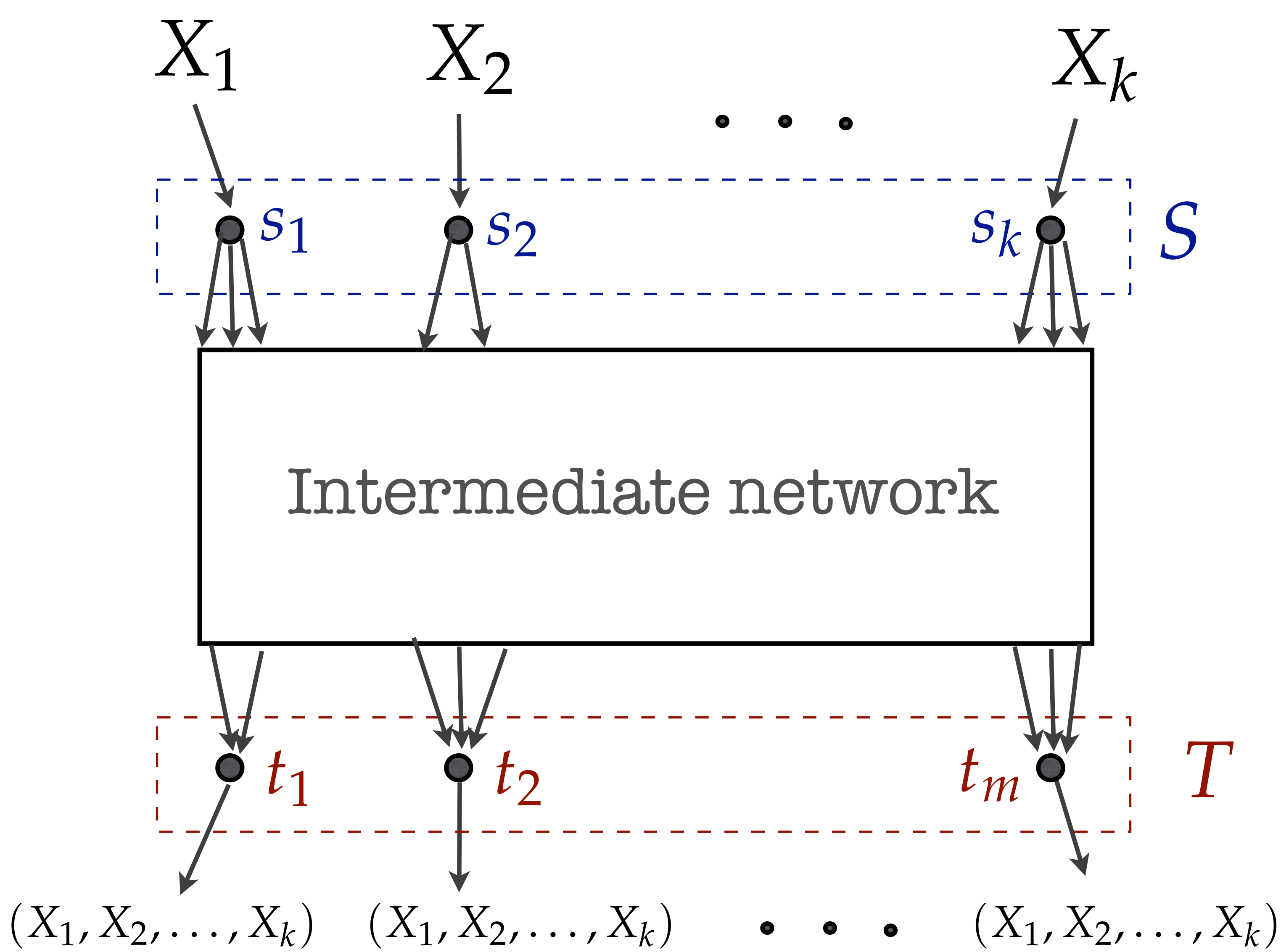
The multicast model for networks

- Directed Graph : $\mathcal{N} = (V, E)$

Vertices $V = \{1, 2, \ldots, n\}$

Edges $E = \{(1,2), (2,3), \ldots, (n-1, n)\}$

- Source nodes : $S = (s_1, s_2, \ldots, s_k)$
- Sink nodes : $T = (t_1, t_2, \ldots, t_m)$
- *Error-free* links, *capacities* $(C_e : e \in E)$



Basic question: Given $(C_e : e \in E)$, what demands are *feasible* ?

[Ho et al, 2006] :

Demands are feasible with capacities $(C_e : e \in E)$ iff

$$\sum_{e \in \Gamma_o(C)} C_e \geq H(X_{S \cap C} | X_{S \setminus C})$$

for all $C \subseteq V$.

sources already present outside C

sum of rates on the outgoing links from a subset C of V

sources present in C

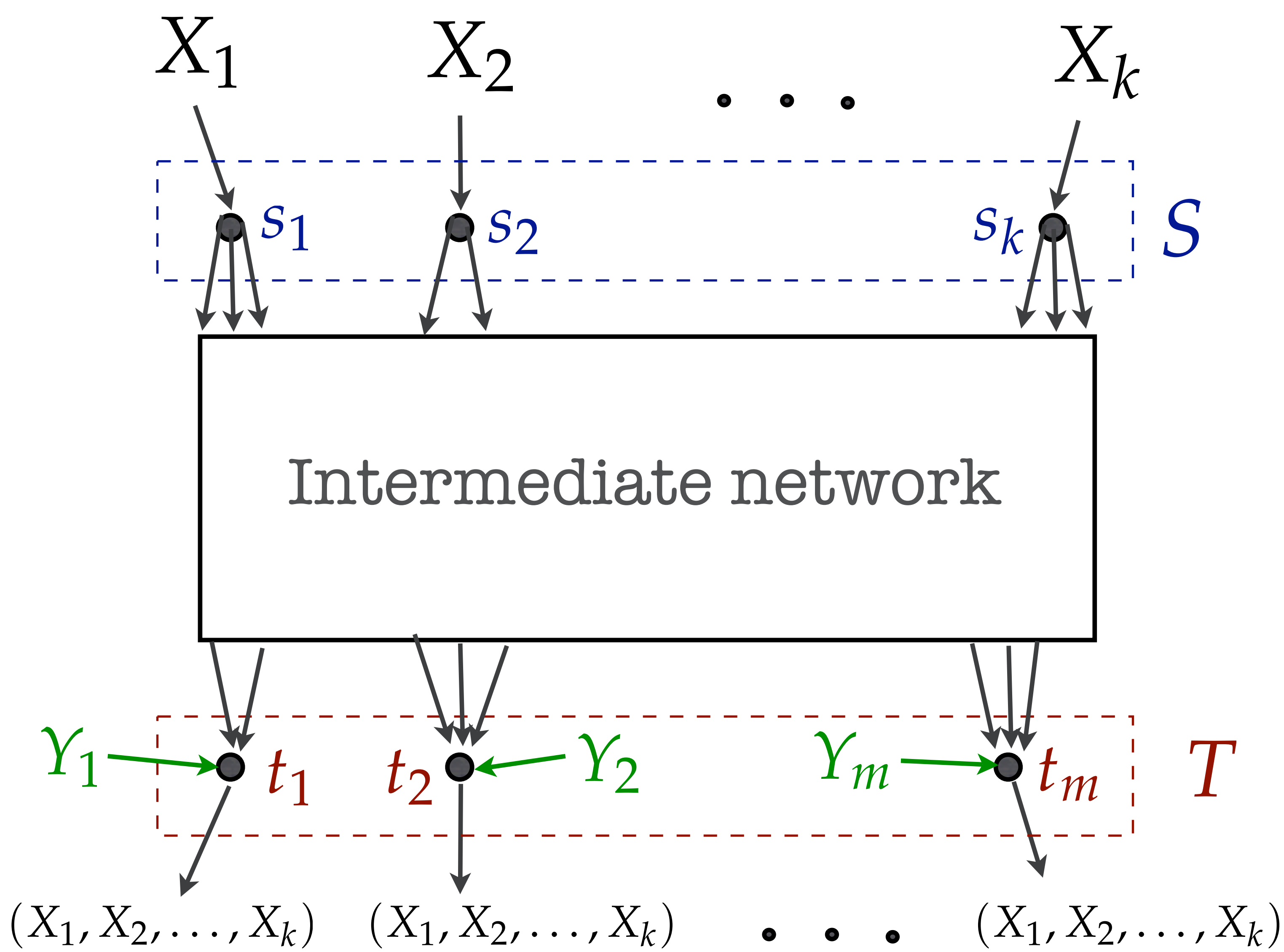
- Linear codes suffice for these rates

Problem with this model: Often, the sinks have some extra information already available, or may not want all the sources present in the network.

This work:

Generalizations of the above result to multicast with side information

Side information at sink nodes



- Source nodes : $S = (s_1, s_2, \ldots, s_k)$
- Sink nodes : $T = (t_1, t_2, \ldots, t_m)$
- Side information at sink nodes : $Y_1, Y_2, \ldots, Y_m$

Result 1 :

Demands are feasible with capacities $(C_e : e \in E)$ *iff*

$$\underbrace{\sum_{e\in\Gamma_o(C)} C_e}_{\text{sum of rates on the outgoing links from a subset } C \text{ of } V} \geq H(X_{S\cap C}|X_{S\setminus C}, Y_{T\setminus C})$$

for all $C \subseteq V$.

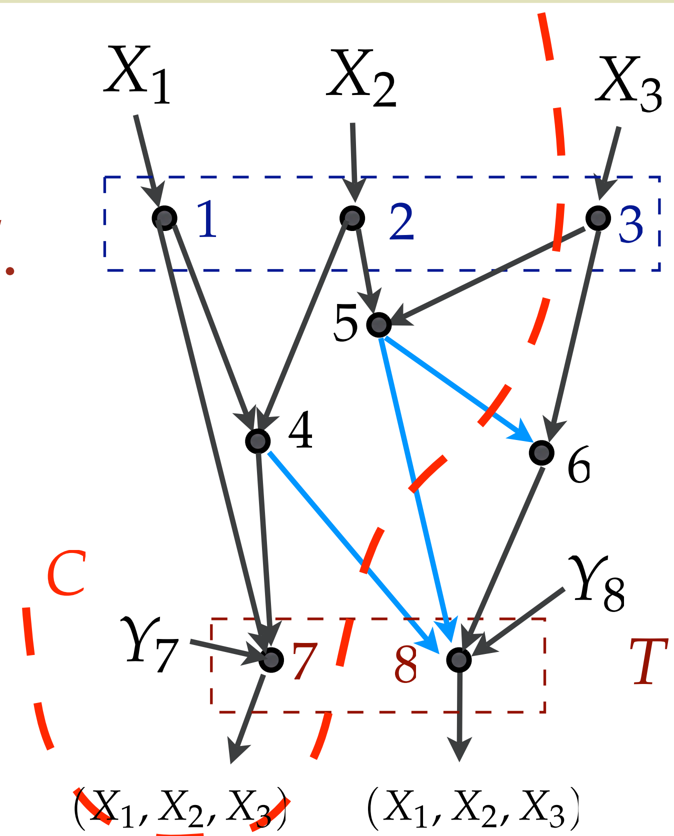
side information present outside C

sources already present outside C

sources present in C

- Linear codes suffice for achieving these rates.

- Expression similar to the result for Multicast without side information given in *[Ho et al, 2006]*.



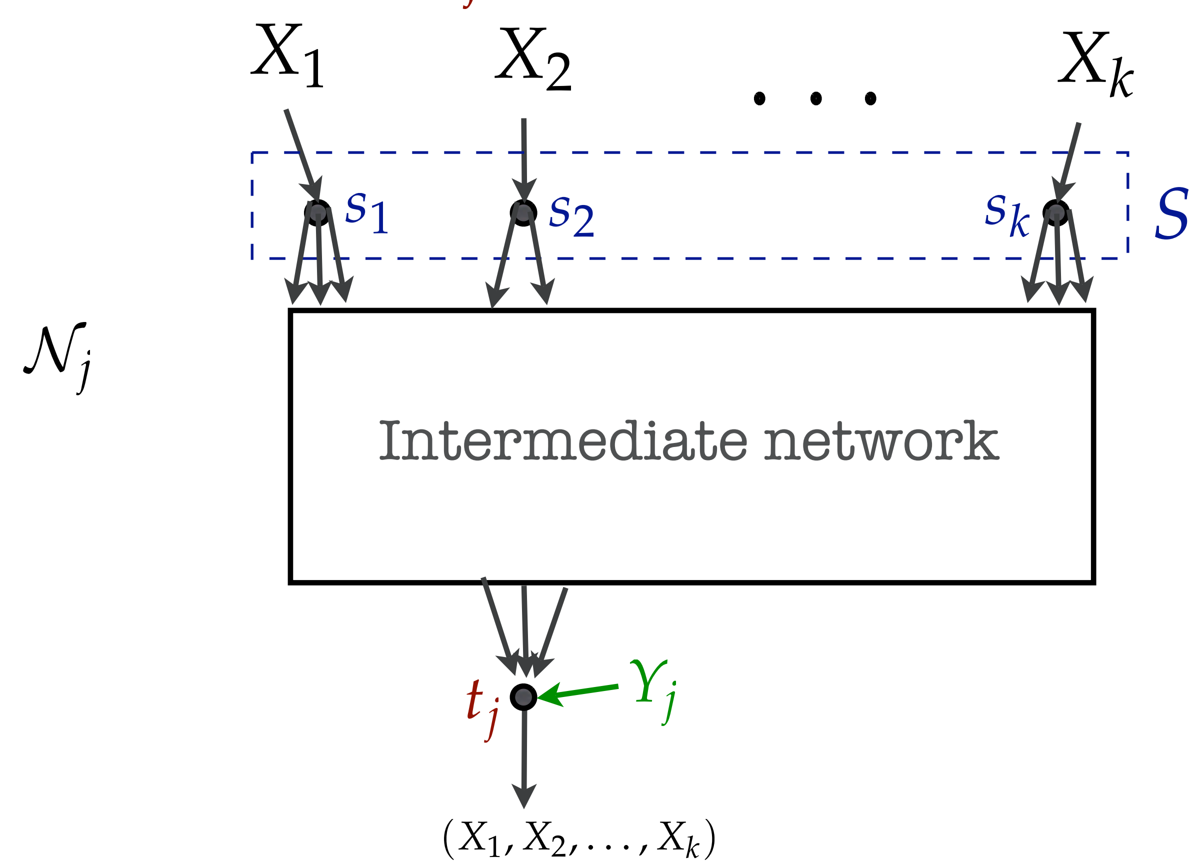
<u>Converse:</u>

- Cutset bound

  (Cover and Thomas, 1991)

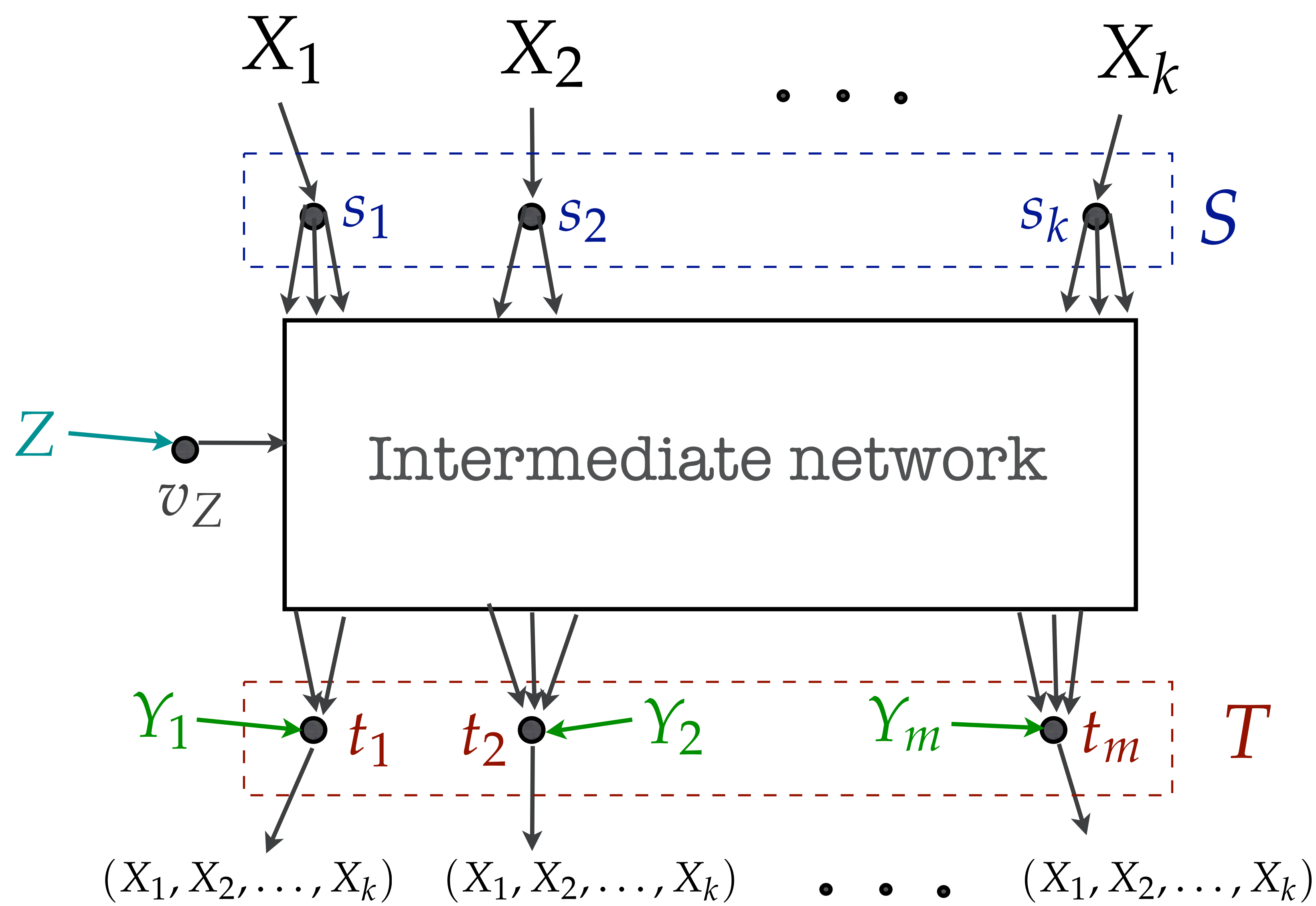
Achievability:

- network $\mathcal{N}_j$ - delete all sinks except t_j



- pure multicast network with sources $(X_1, X_2, \ldots, X_k, Y_j)$
- random linear codes for $\mathcal{N}_j$ as in [Ho et al, 2006]
  - need only input and output rates at each node
  - good for each $\mathcal{N}_j$ => good for $\mathcal{N}$
- can be done for all rates in the claimed region

Side information at a non-sink nodes

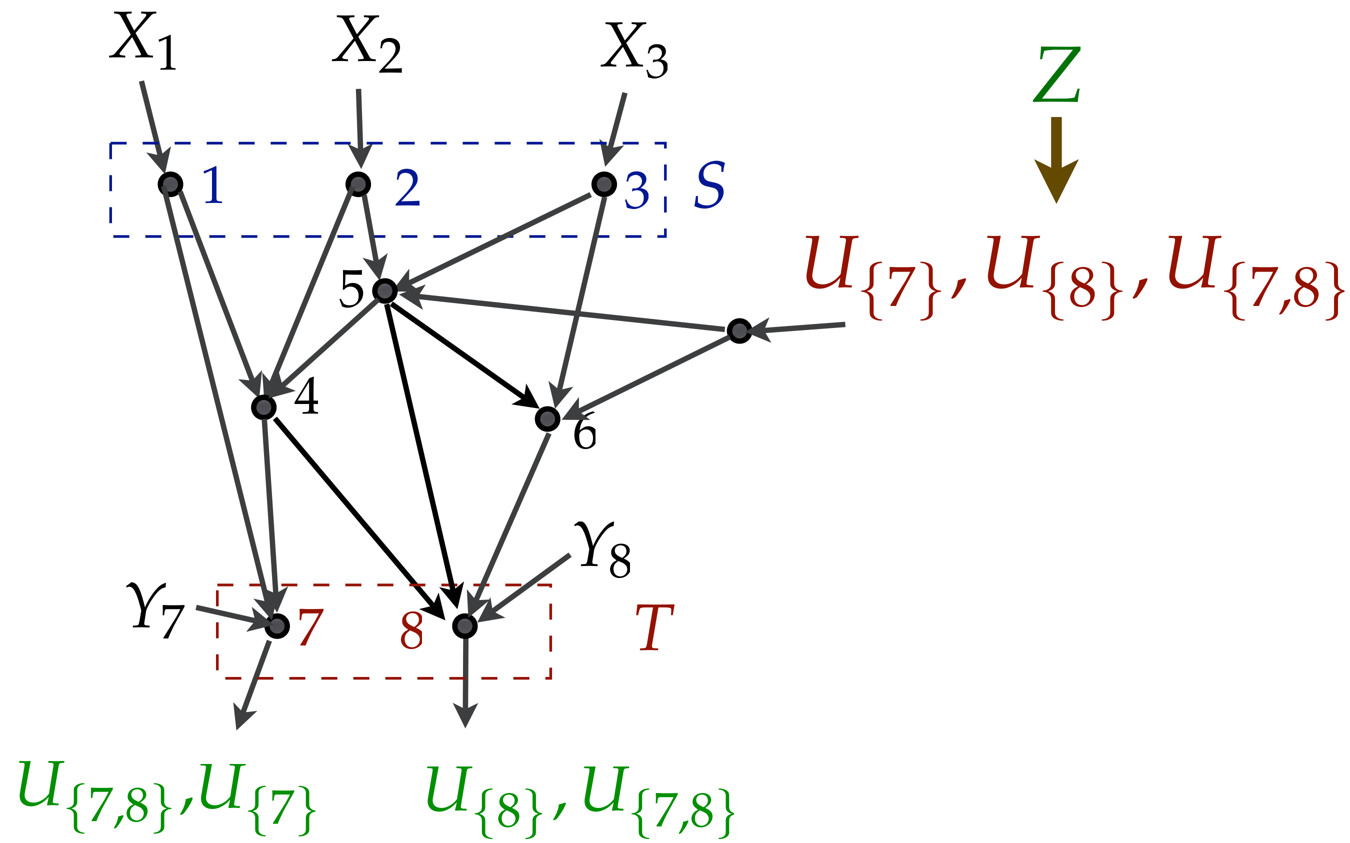


- Source nodes : $S = (s_1, s_2, \ldots, s_k)$
- Sink nodes : $T = (t_1, t_2, \ldots, t_m)$
- Side information at sink nodes : $Y_1, Y_2, \ldots, Y_m$
- Side information Z at a non-sink node v_Z

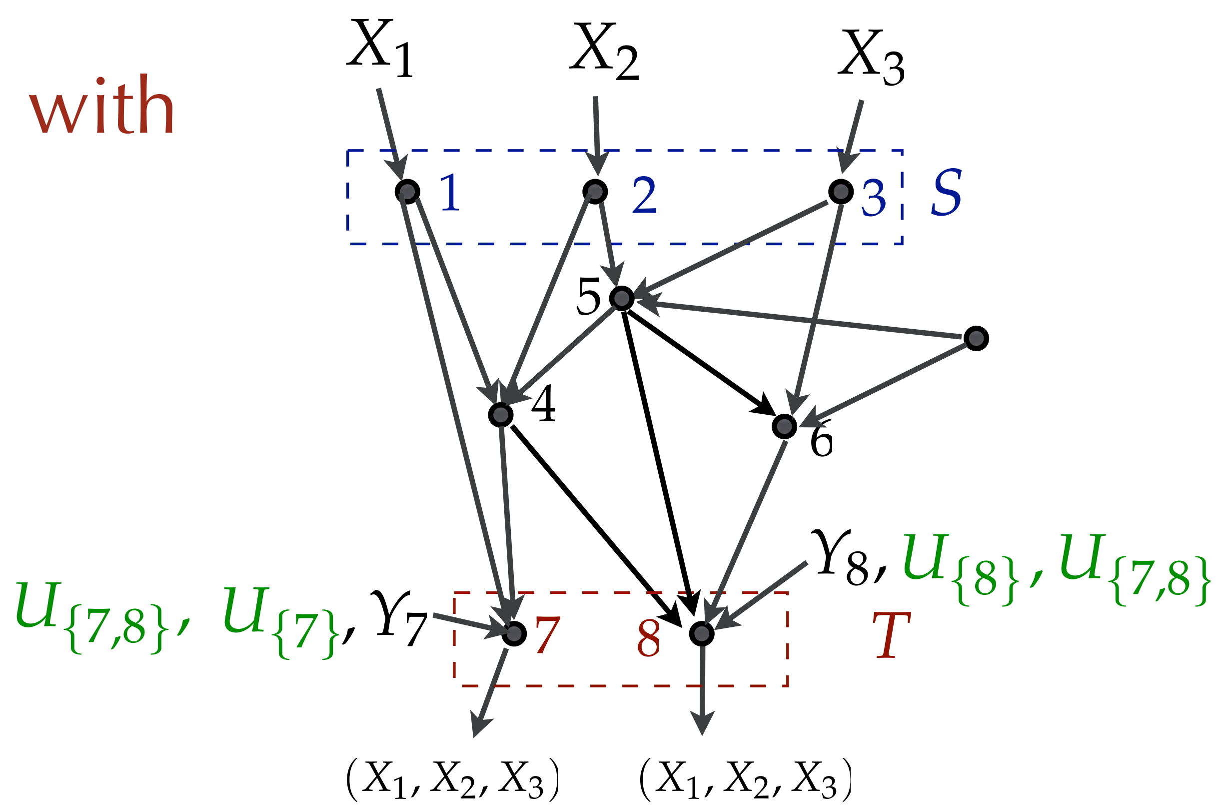
Coding scheme

- Separate codewords for each subset of the sink nodes.

- Sequence of multicast sessions to transmit U_τ's

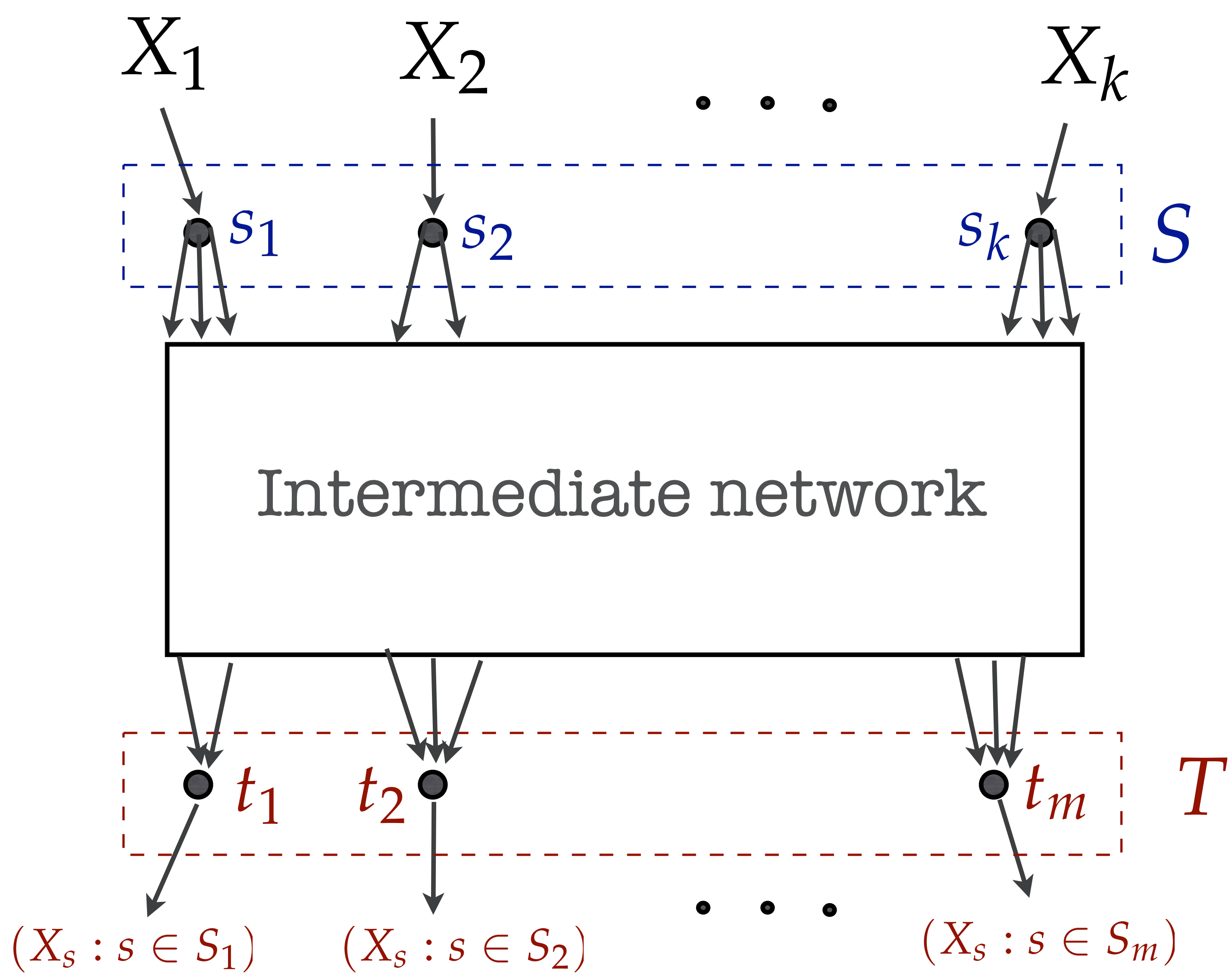


- Multicast $(X_1, X_2, \ldots, X_k)$ with U_τ's as side information



Result 2 :
An sufficient set of feasibility conditions

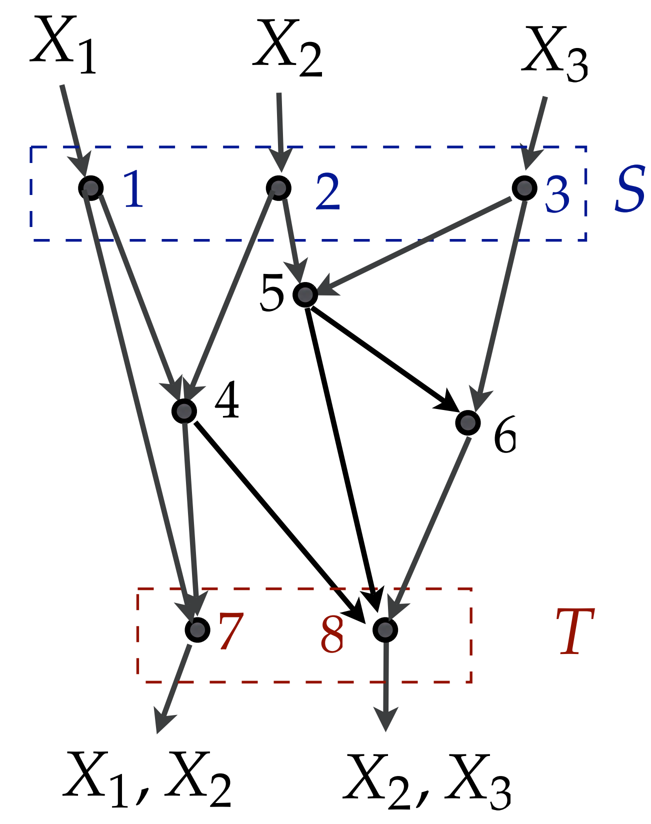
General source-demand structures



- Source nodes : $S = (s_1, s_2, \ldots, s_k)$

- Sink nodes : $T = (t_1, t_2, \ldots, t_m)$

- Demand $(X_s : s \in S_j)$ at the sink node t_j



- Fix permutation $\sigma(\cdot)$ of $(1, 2, \ldots, k)$.
- Multicast X_i's in order $\sigma(\cdot)$

Result 3 :

Demands are feasible with capacities $(C_e : e \in E)$ if

$$\sum_{e \in \Gamma_o(C)} C_e \geq \sum_{i=1}^{k} \max_{j : t_j \in T \setminus C} H(X_{\{\sigma(i)\} \cap S_j} | X_{\{\sigma(1), \sigma(2), \ldots, \sigma(i-1)\} \cap S_j})$$

for all $C \subseteq V$.

Summary

Key ideas

- *Use existing result for multicast to design coding strategies for cases where side information is present only at the sink nodes.*
- *Use the result obtained for case with side information at the sinks to design coding strategies for other cases.*

Results

- *Characterization of the entire set of feasible rates for multicast when side information is present only at the sinks*
  - Linear codes suffice
- *Sufficient condition for feasibility for*
  - Multicast when side information is present at a non-sink node
  - Networks with general source-demand structures

Future directions

- *Better coding strategies for the cases where feasibility conditions are not tight*
  - Optimal ordering of multicast sessions ?
  - Other transmission strategies ?
- *Lower bound on the rates when the feasibility conditions are not tight*
  - Bound on the difference between the lower bound and the capacity region?
- *Complexity issues*
  - When do linear codes suffice?