INFORMATION FOR WHAT? FROM X TO F(X)

Muriel Médard

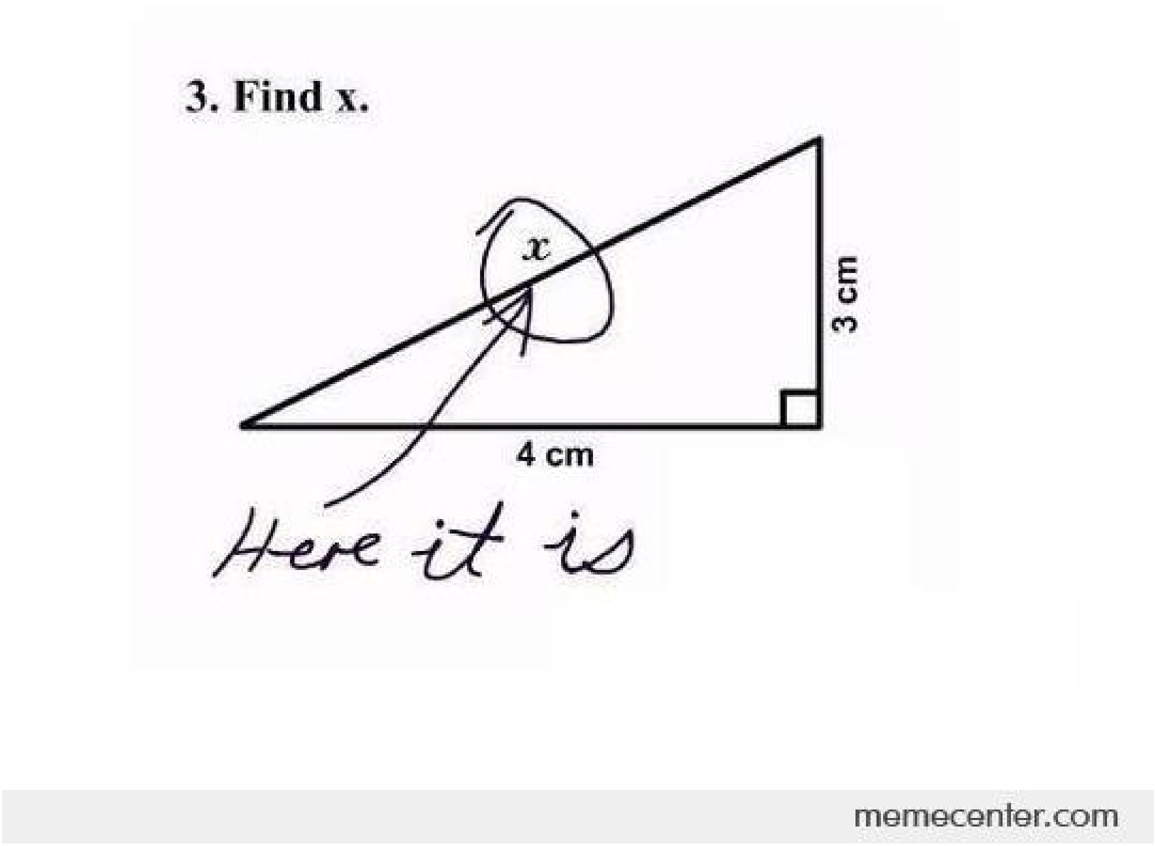
RLE

MIT

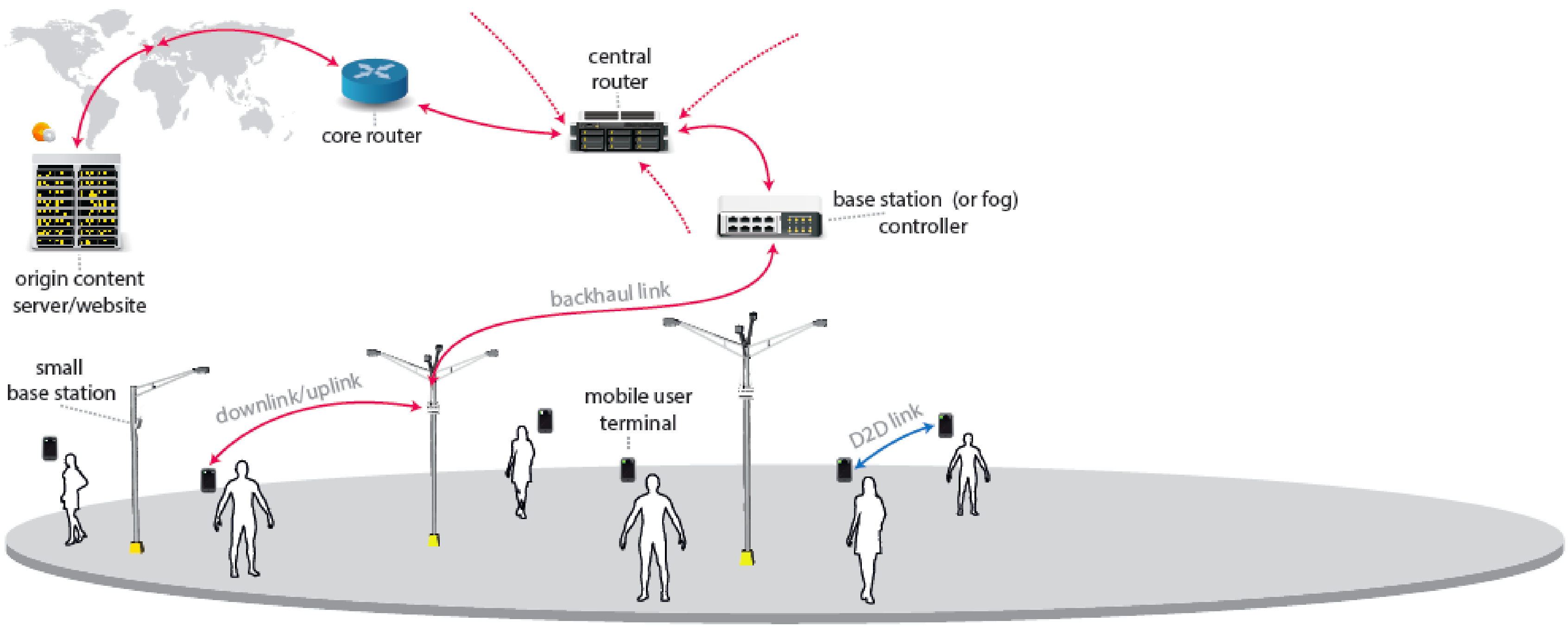
Collaborators

- MIT: Alejandro Cohen, Rafael D'Oliveira, Vishal Doshi (Nike), Soheil Feizi (University of Maryland), Litian Liu, Derya Malak (RPI), Devavrat Shah, Salman Salamatian, Amit Solomon
- Caltech: Michelle Effros
- Rutgers: Salim El Rouayheb.

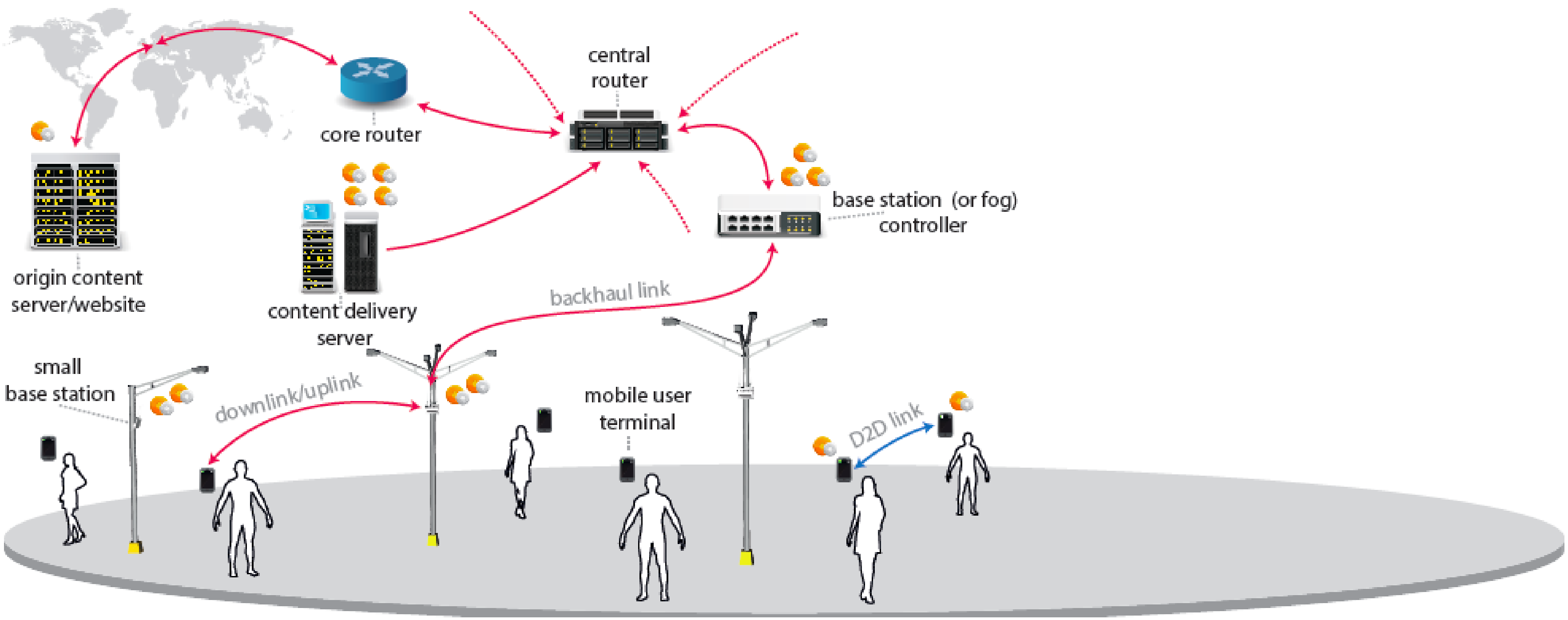
What is x?



The vision - data

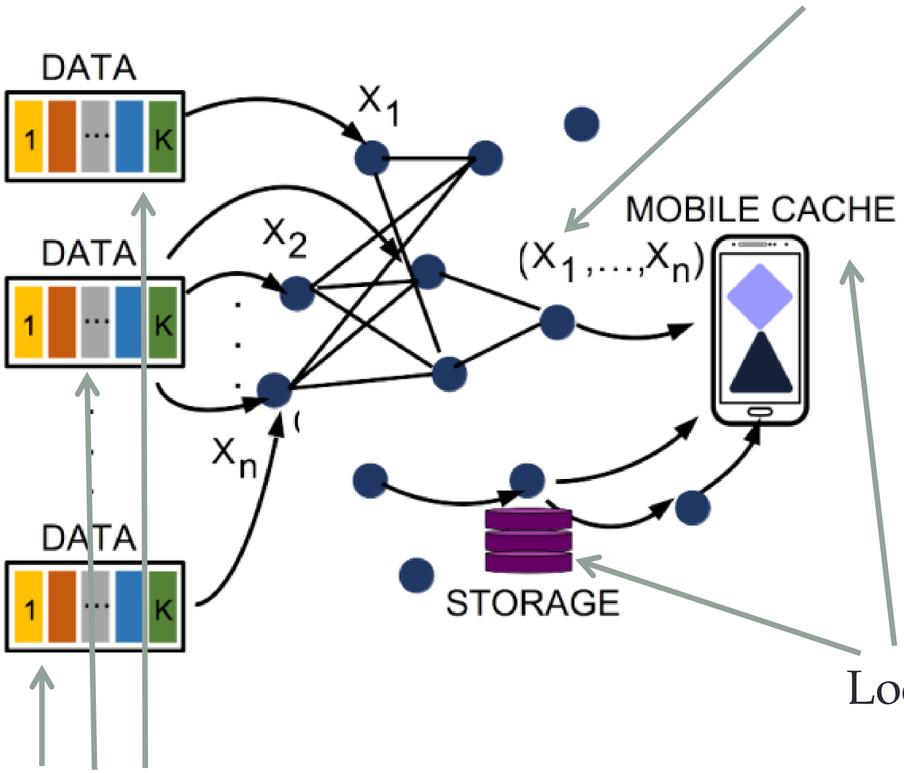


The vision - data



What does it mean?

Your work requires you to retrieve data

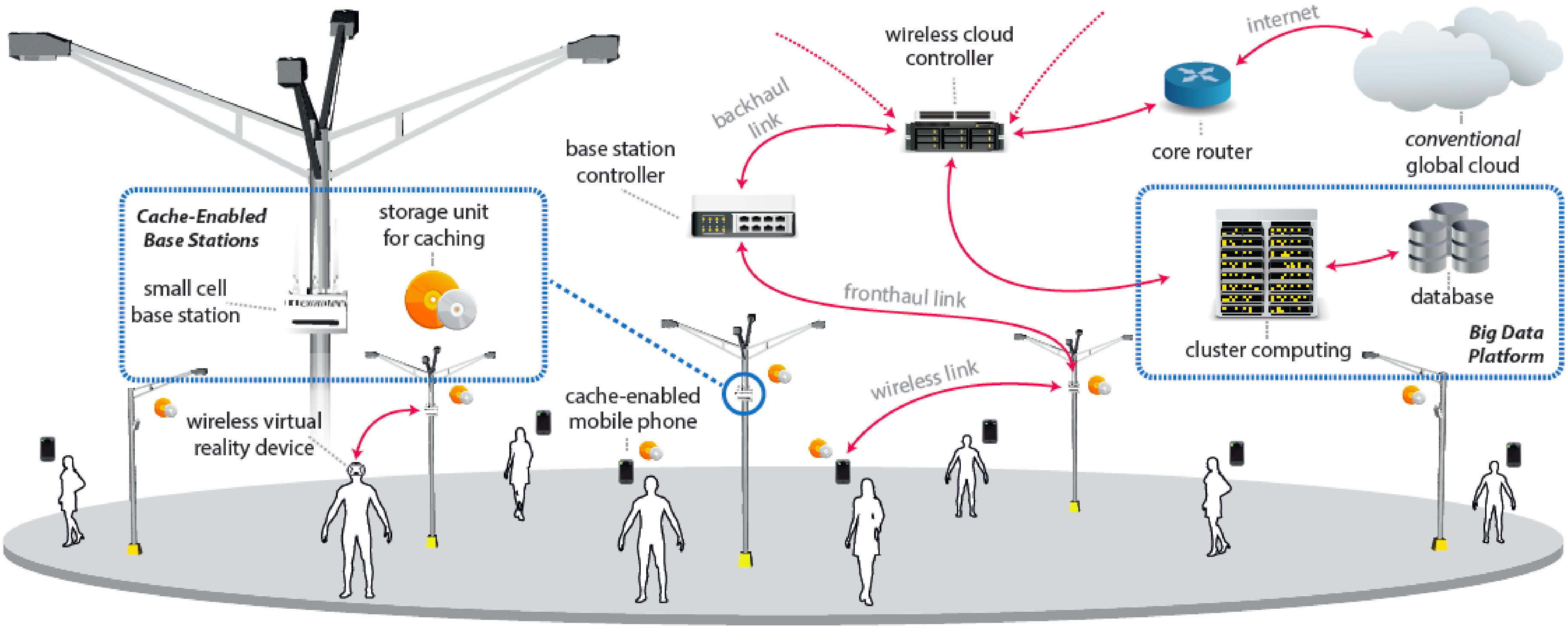


Storage, communications are inseparable

Local data may be necessary

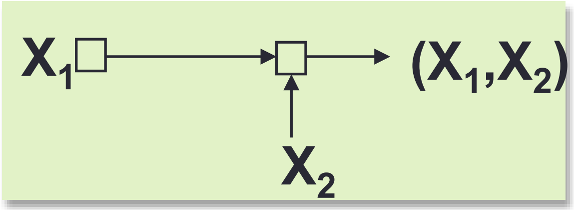
The data resides across different domains and locations

The vision - computation

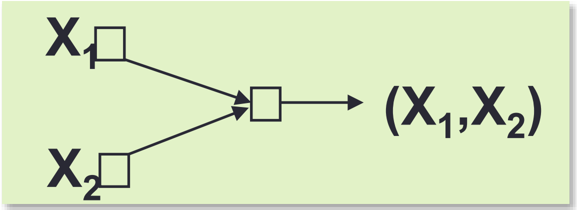


Data

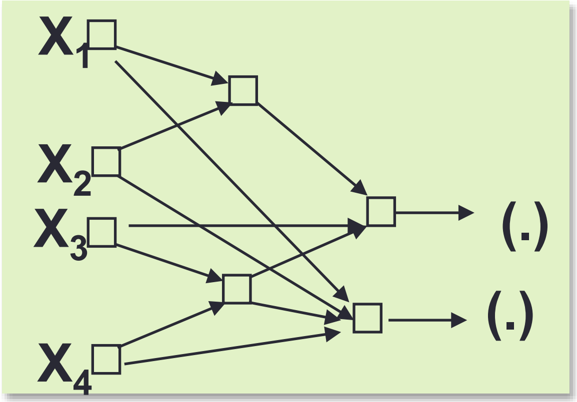
- Side information problem (Wyner-Ziv)



- Depth one trees (Slepian-Wolf)
  - Can be generalized to trees

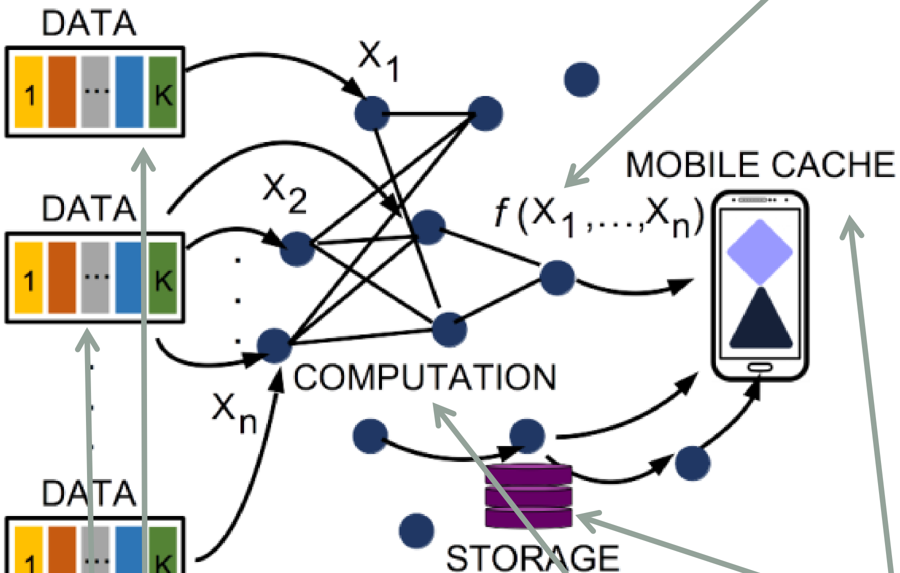


- General Networks (Ho et al)
  - Multicast
  - Random Linear Network Coding
  - Error exponents via method of types generalize Csiszar error exponents



What does it mean?

Your work requires you to exploit data



Computation storage, communications are inseparable

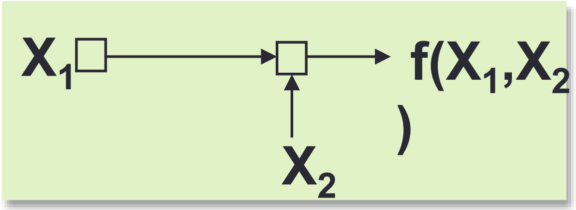
Local data may be necessary

The data resides across different domains and locations

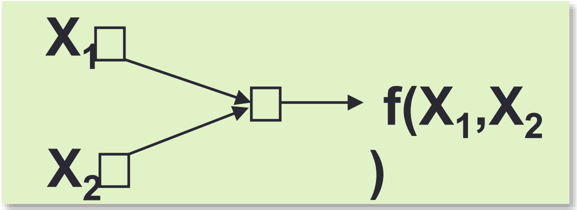
For reasons of dimensioning, scaling, security Computation needs to occur in network

Functions

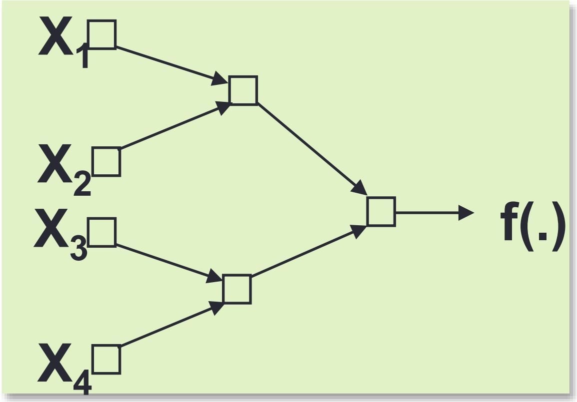
- Side Information Problem
  (Orlitsky and Roche, Ortlitsky and Alon)



- Depth One Trees (Doshi et al.)
  - Characterizing the rate region
  - A necessary and sufficient condition for achievability



- General Trees (Feizi and Médard)
  - Rate lower bounds for a general case
  - For independent sources: optimal coding schemes
  - Polynomial time (practical) coding schemes for some conditions
  - Feedback can improve rate bounds



V. Doshi, Shah, D., Médard, M., and Effros, M., "Functional Compression Through Graph Coloring", *IEEE Transactions on Information Theory*, Vol. 56, No. 8, Aug. 2010
S. Feizi, and Médard, M., "On Network Functional Compression", *Transactions on Information Theory,* vol. 60, no. 9, Sept.2014.

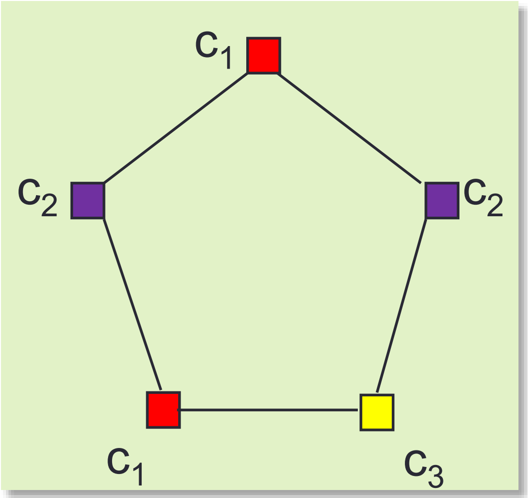
Characteristic Graph

- **Example:** X_1 is a RV with uniform distribution over {0,1,2,3,4}. X_2 is a RV such that we have the following graph G_{X_1} for .

  $c_{G_{X_1}} = \{c_1, c_2, c_3\}.$

  $P(c_1) = P(c_2) = 2/5 \qquad P(c_3) = 1/5$

  $H(c_{G_{X_1}}) \approx 1.52$



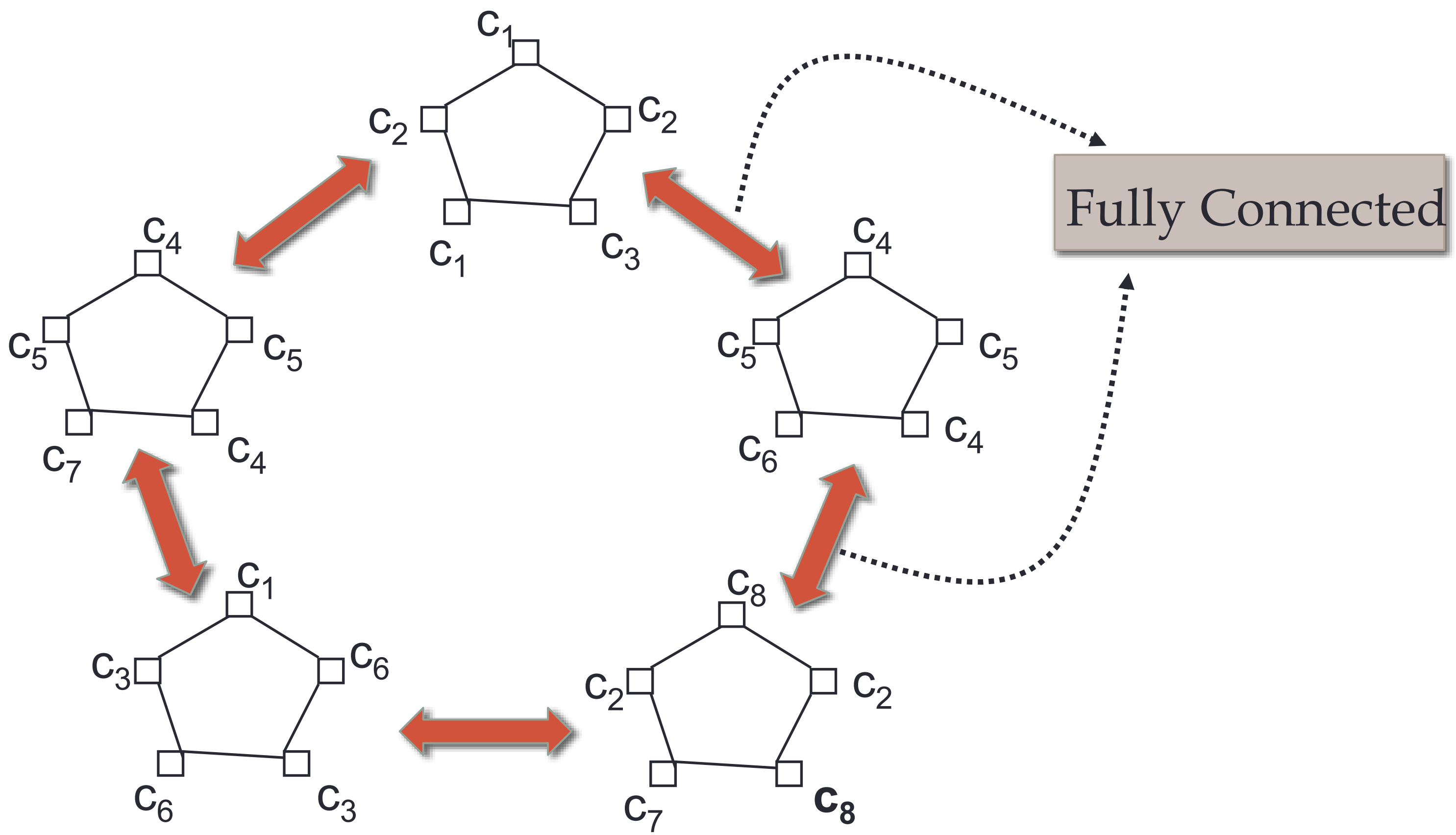
- X_1^2 can take 25 values {00,01,…,44}.
- To make characteristic graph of X_1^2, we connect two vertices if at least one of coordinates are connected in G_{X_1}.

Characteristic graph:

– Vertices are different sample values

– Two vertices are connected if they should be distinguished

Power Graph Example

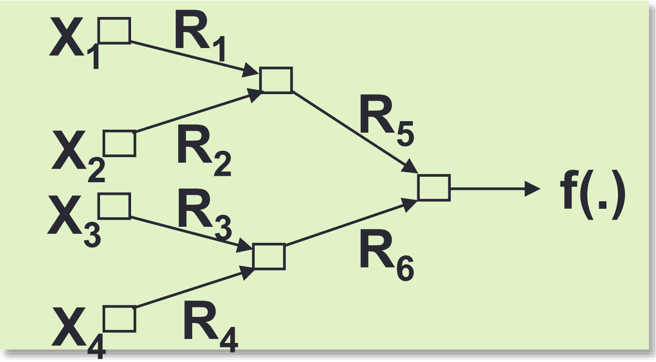


- One can color this graph by using **8 colors**.

$$\frac{1}{2}H(c_{G^2_{X_1}}) \approx 1.48 < H(c_{G_{X_1}}) \approx 1.52$$

General Trees

- Intermediate nodes are allowed to compute some functions.
- Rate lower bounds by using cut-set bounds on graph entropies:



$$R_1 + R_2 + R_3 + R_4 \geq H_{G_{X_1}, G_{X_2}, G_{X_3}, G_{X_4}}(X_1, X_2, X_3, X_4)$$

$$R_1 + R_2 \geq H_{G_{X_1}, G_{X_2}}(X_1, X_2 | X_3, X_4)$$

$$R_5 + R_6 \geq H_{G_{X_1,X_2}, G_{X_3,X_4}}(X_1, X_2, X_3, X_4), \ldots$$

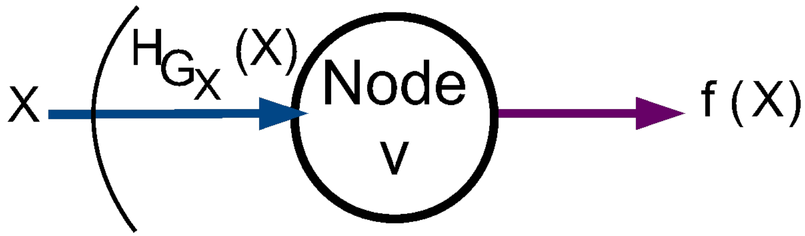
- **Theorem**

  When sources are independent, these bounds are tight.

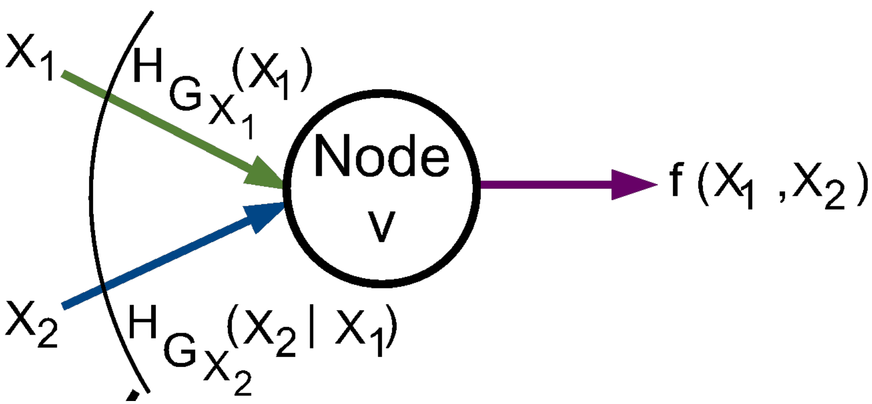
- For independent sources, functions to be computed at intermediate nodes are **coloring functions**.
- Unlike regular entropy, chain rule **does not hold** for graph entropies in general:

$$H_{G_{X_1}, G_{X_2}}(X_1, X_2) \neq H_{G_{X_1,X_2}}(X_1, X_2)$$

Rate region for distributed functional compression



$$R_X > H_{G_X}(X)$$

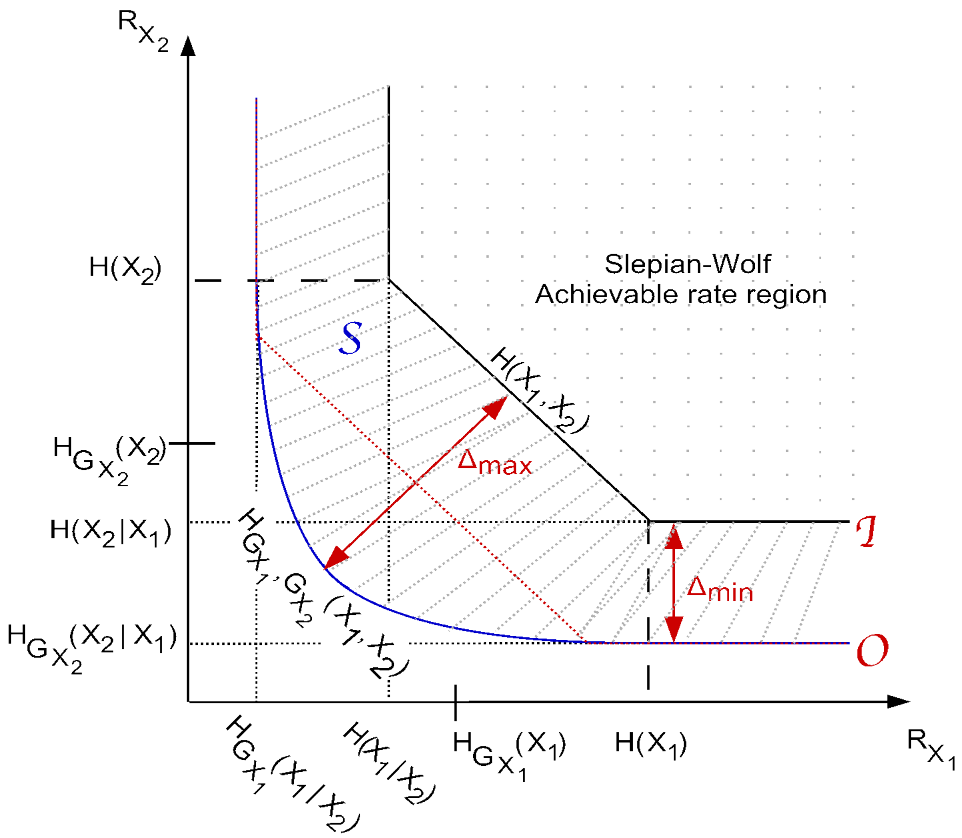


$$R_{X_1} + R_{X_2} > H_{G_{X_1}}(X_1) + H_{G_{X_2}}(X_2 | X_1)$$

$$> H_{G_{X_1} G_{X_2}}(X_1, X_2)$$

Exploit Körner's graph entropy to compute the true rate region for distributed functional compression.

Functional Compression versus Slepian-Wolf



Entropic Surjectivity

Defn. Entropic surjectivity of a function is how well can be compressed wrt the compression rate of its domain :

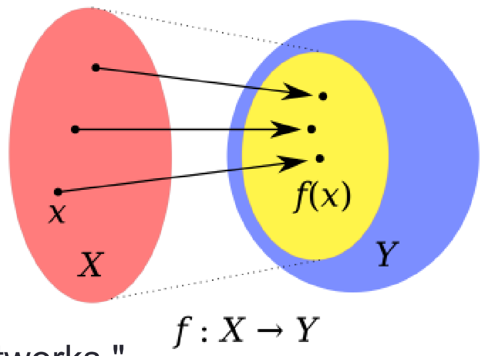
$$\Gamma_c(f) = \Gamma_c = \frac{\mathbf{H}(f_c(X))}{H(X)}$$

A minimal representation of function *f*, e.g. if coloring is used [DSME, 10], [FM, 14]:

$$\mathbf{H}(f(X)) = H_{G_X}(X)$$

A non-surjective function has less redundancy vs surjective function:

$$\frac{\mathbf{H}(f_c(X))}{H(X)} \approx 0$$



D. Malak, A. Cohen and M. Médard, "How to Distribute Computation in Networks," *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*,

Compression and Communication

Distributed source compression [Slepian and Wolf, 73], [Pradhan and Ramchandran, 13], [Coleman et al, 06], [Wyner and Ziv, 76]

Rate region and graph entropy [Körner, 73], [Alon and Orlitsky, 96], [Orlitsky and Roche, 01], [Doshi et al, 10], [Feizi and Médard, 14], [Feng et al, 04], [Gallager, 88], [Kamath and D. Manjunath, 08], [Shah et al, 13]

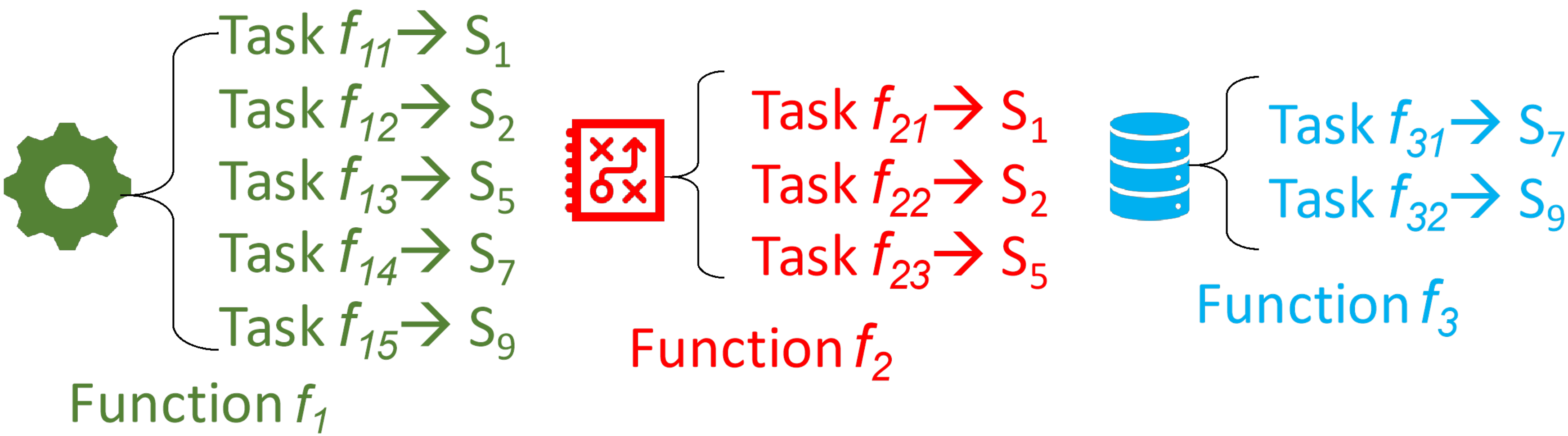
Network coding and linear functions [Ho et al, 06], [Kowshik and Kumar, 10, 12], [Appuswamy and M. Franceschetti, 14], [Koetter et al, 04], [Koetter and Médard, 03], [Huang et al, 18], [Li et al, 03]

Coding for computation/communications [Li et al, 18], [Kamran et al, 19], [Yu, Maddah-Ali, and Avestimehr, 18]

Functions with special structures [Shen et al, 18], [Giridhar and Kumar, 05], [Gorodilova, 19]

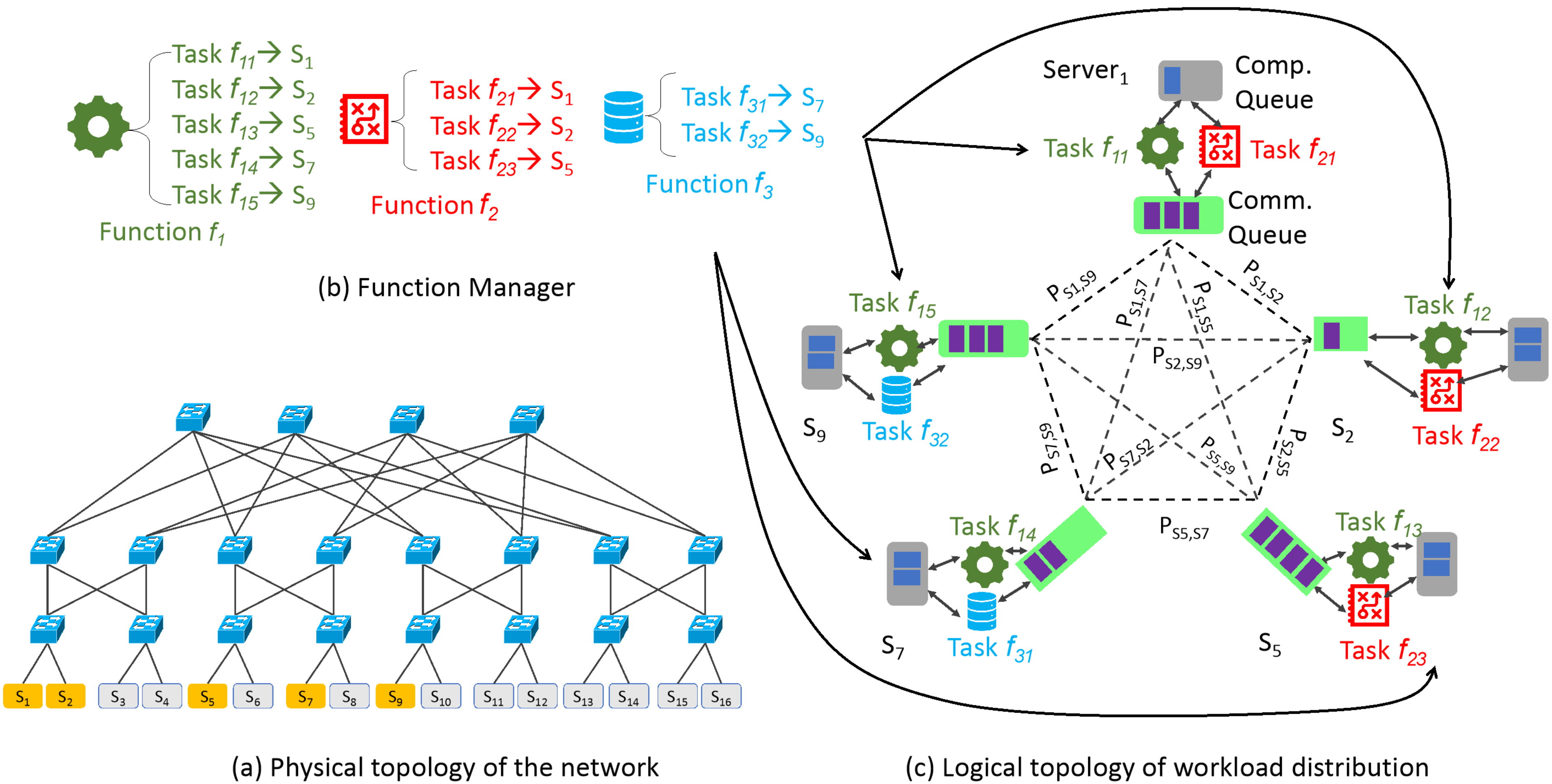
OUR GOAL: Use underlying redundancy both in data and functions, and recover a sparse representation, or labeling, at the destination.

How to Manage Functions

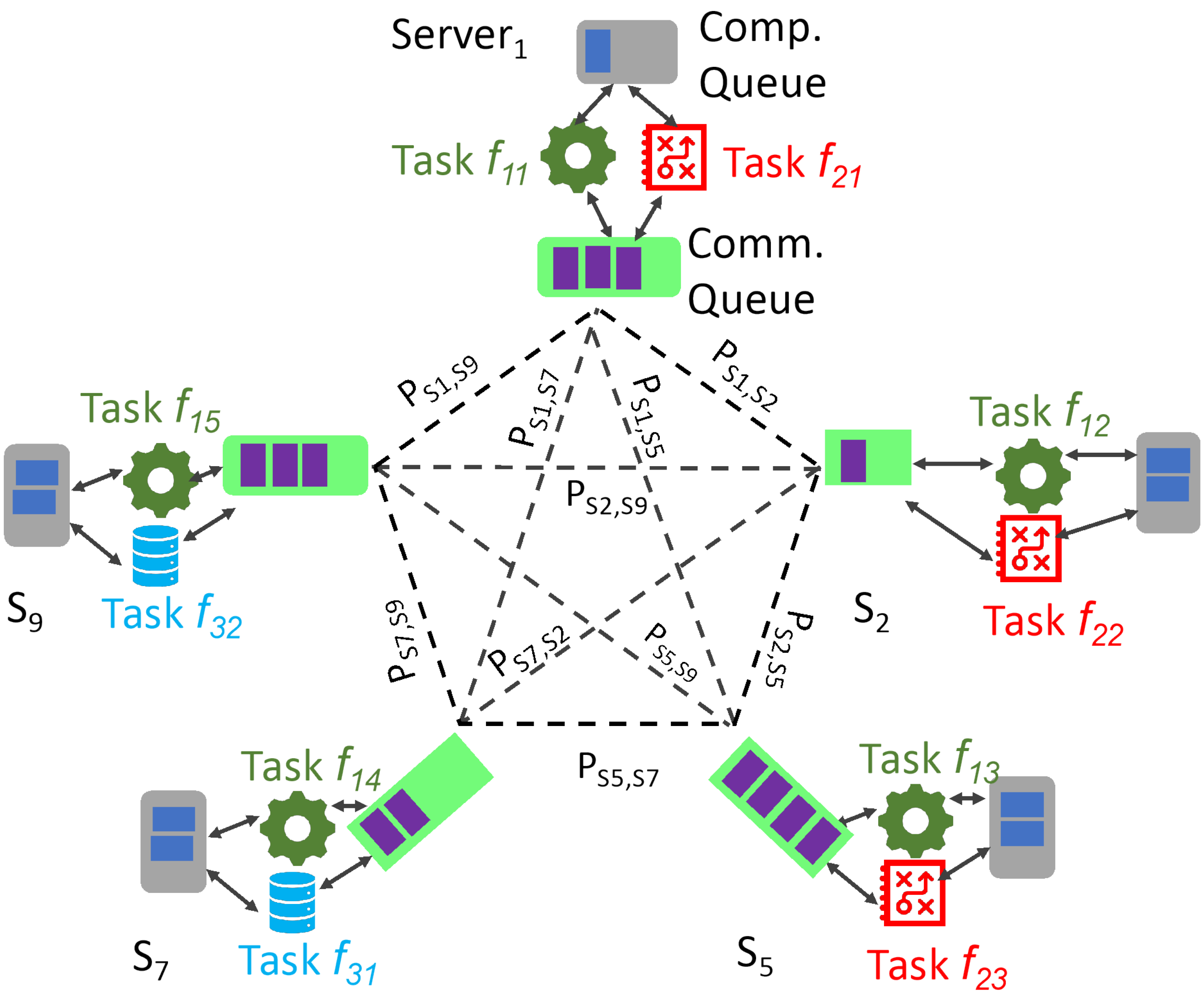


Task manager decides how to distribute the task/computation accordingly (by looking at the routing information).

Architecture

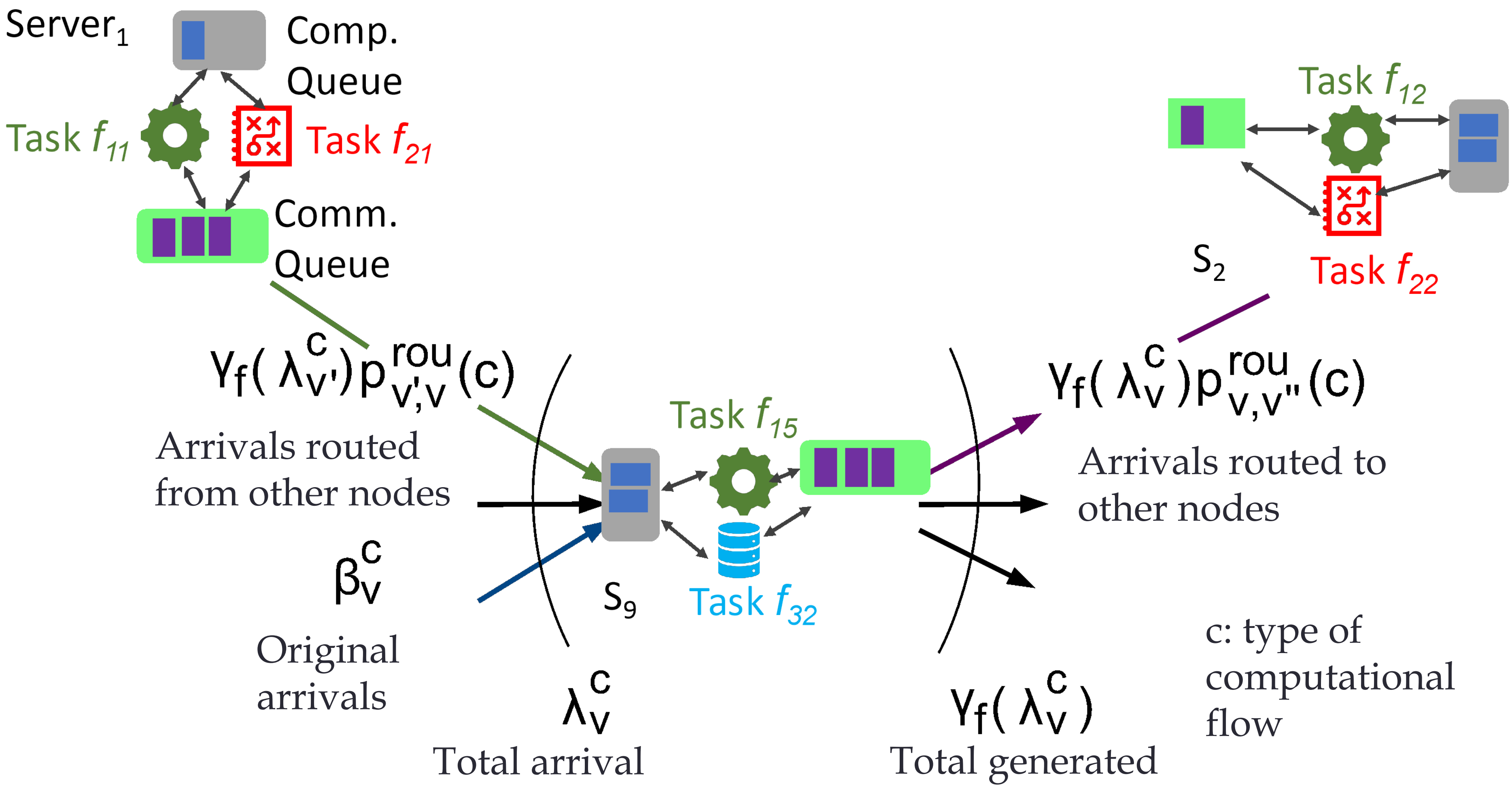


Routing for Computing



For tractability, consider each node in isolation, i.e quasi-reversible or product form as in a Jackson network [Walrand

Routing for Computing



Network is in product form. Hence, nodes (servers) can be considered in isolation.

Average Delay (per node)

The total delay of computation and communications for processing functions of type $c \in \mathcal{C}$ at node v

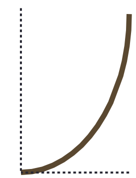
$$W_v^c = C_{v,comp}^c + C_{v,comm}^c$$

time complexity of computing

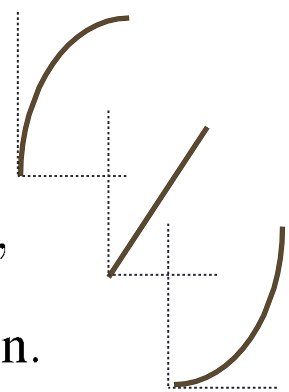
$$C_{v,comp}^c = \frac{1}{\lambda_v^c} d_f(M_v^c)$$

time complexity of communicating

$$C_{v,comm}^c = \frac{1}{\mu_v^c - \gamma_f(\lambda_v^c)}$$



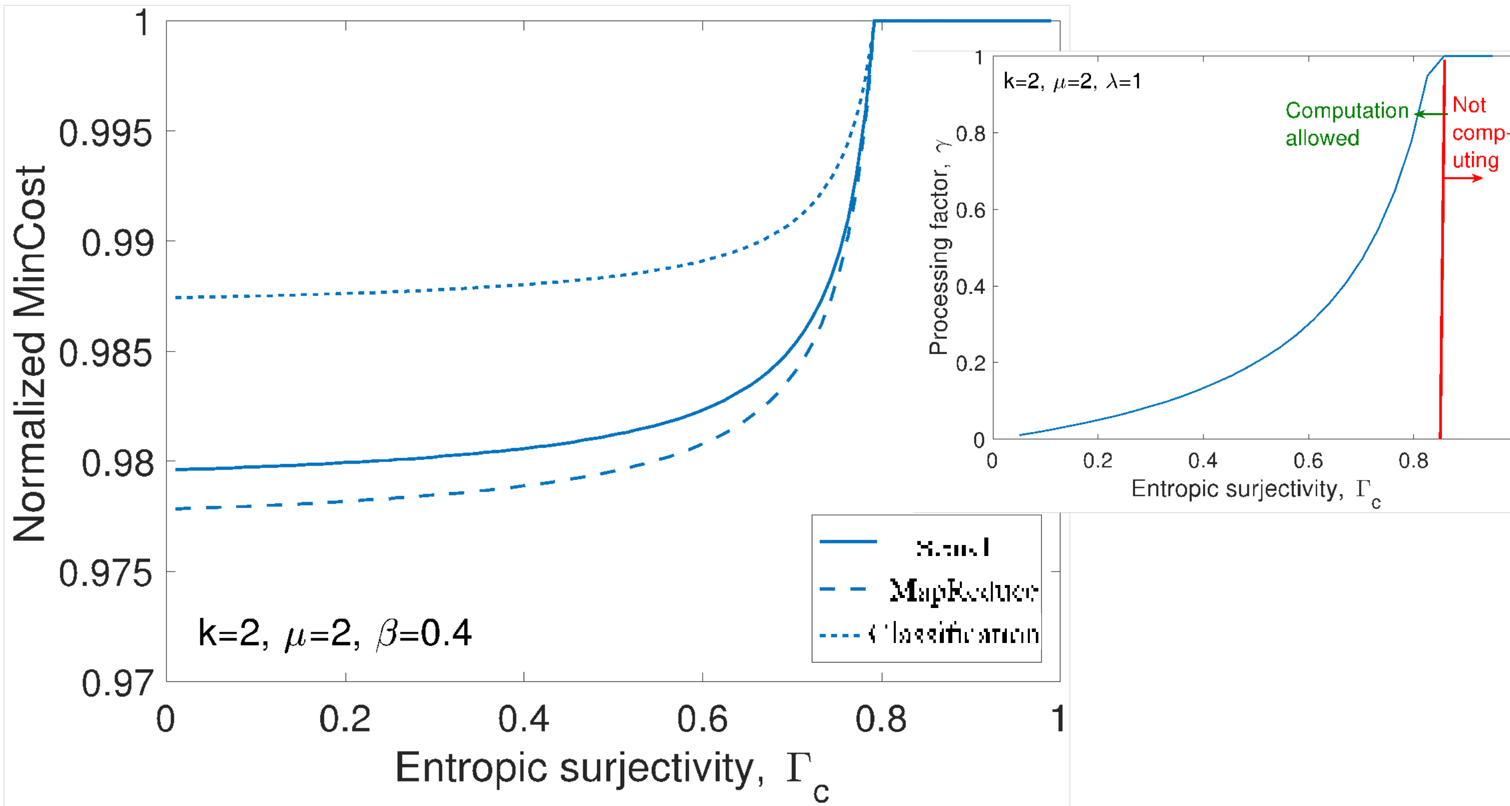
$$d_f(M_v^c) = \begin{cases} O(\log(M_v^c)), & \text{Search}, \\ O(M_v^c), & \text{MapReduce}, \\ O(\exp(M_v^c)), & \text{Classification}. \end{cases}$$



$$\rho_v^c = \lambda_v^c / \mu_v^c \in [0, 1)$$

M_v^c is the long-term average number of packets waiting for service.

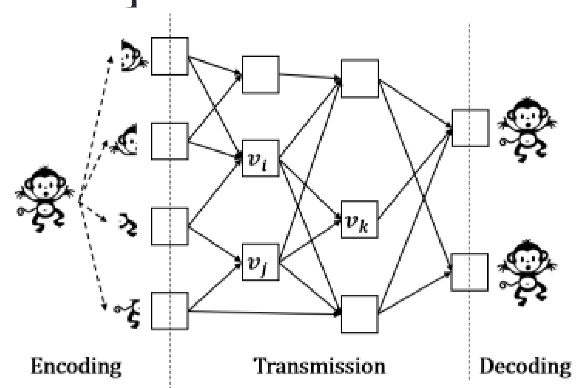
Cost change with surjectivity



The normalized MinCost versus Γ_c .

What about neural networks?

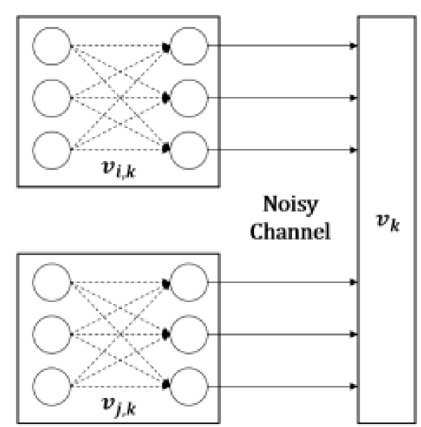
- Point-to-point NN-based joint source-channel coding
  - Images [Bourtsoulatze, Kurka, and Gündüz 2019]
  - Text [Farsad, Rao, and Goldsmith 2018]



(a)

Neural network coding:

- No assumption on sources.
- Joint source-network coding scheme.
- Practical decoders.
- Applicable to arbitrary network topology
- Applicable to arbitrary power constraints.



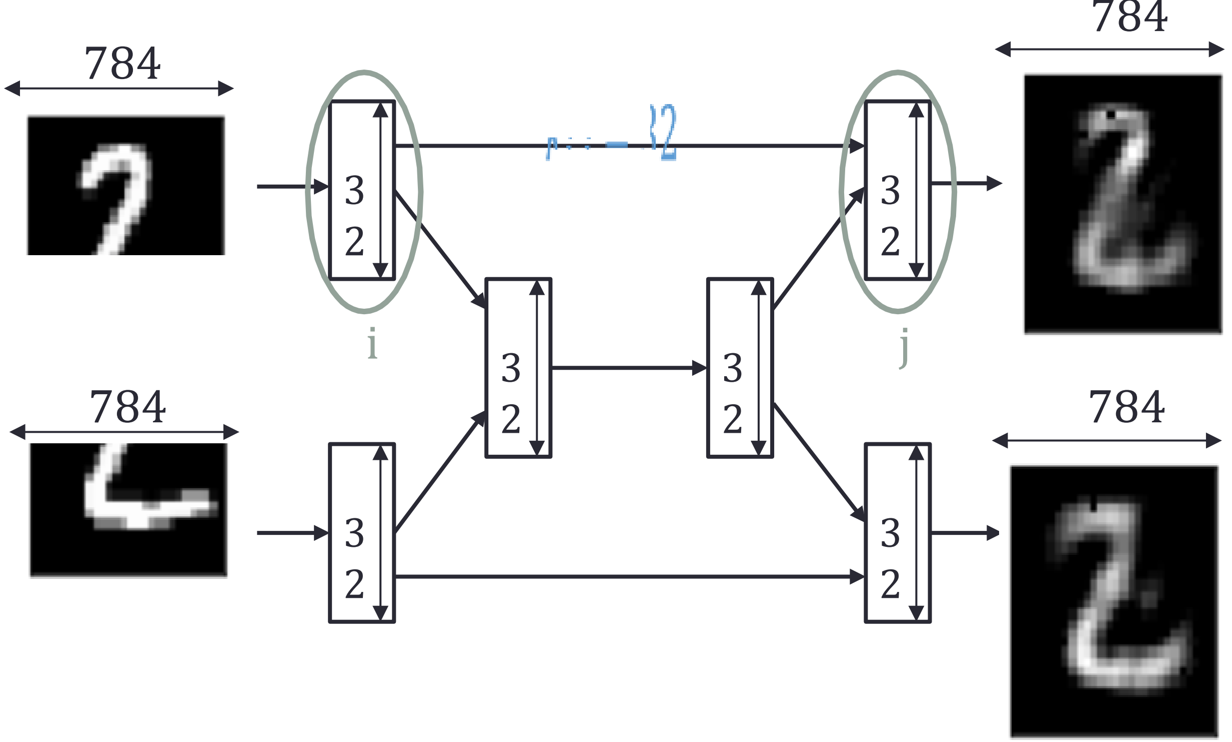
(b)

L. Liu, A. Solomon, S. Salamatian and M. Médard, "Neural Network Coding," *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*

Performance Evaluation

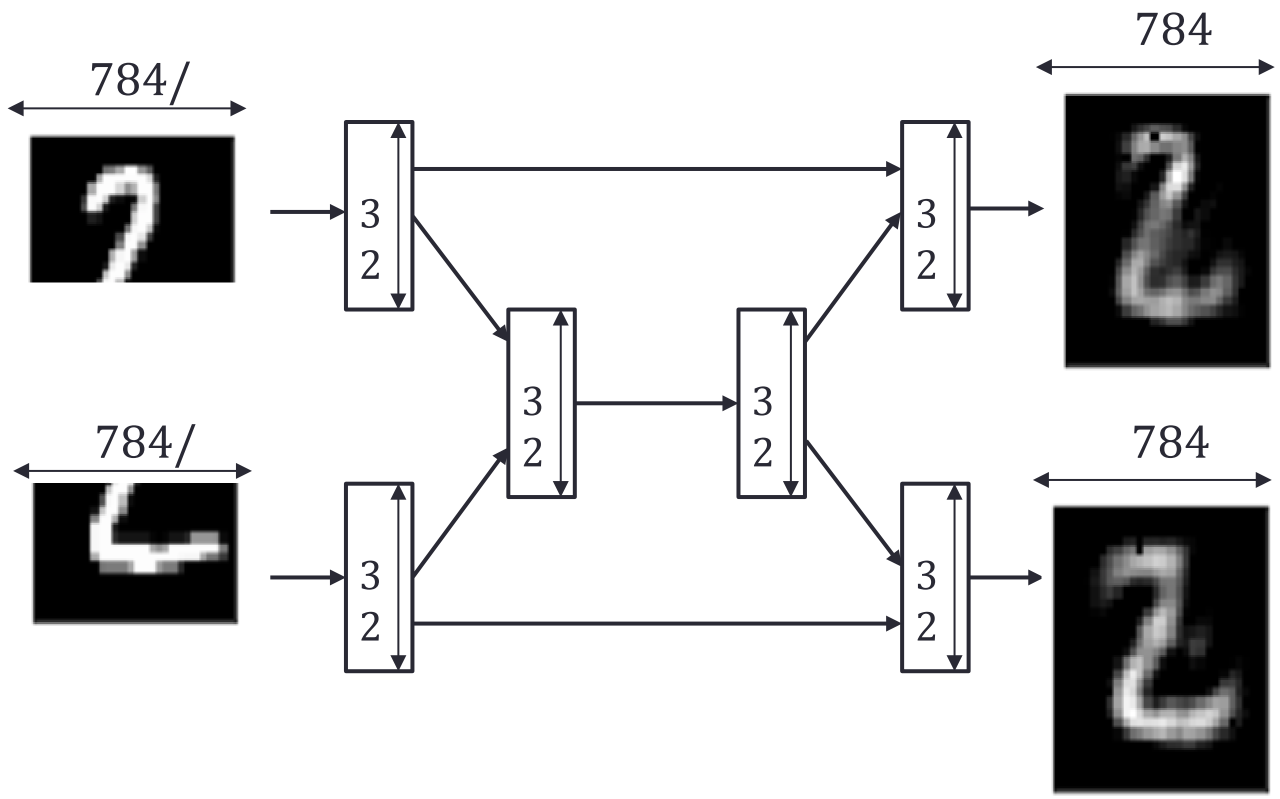
- Reconstruction metric

$$pSNR = 10\log_{10}\left(\frac{\max\{X_i\}^2}{\mathbb{E}[(\vec{X} - \hat{\vec{X}})^2]}\right)$$



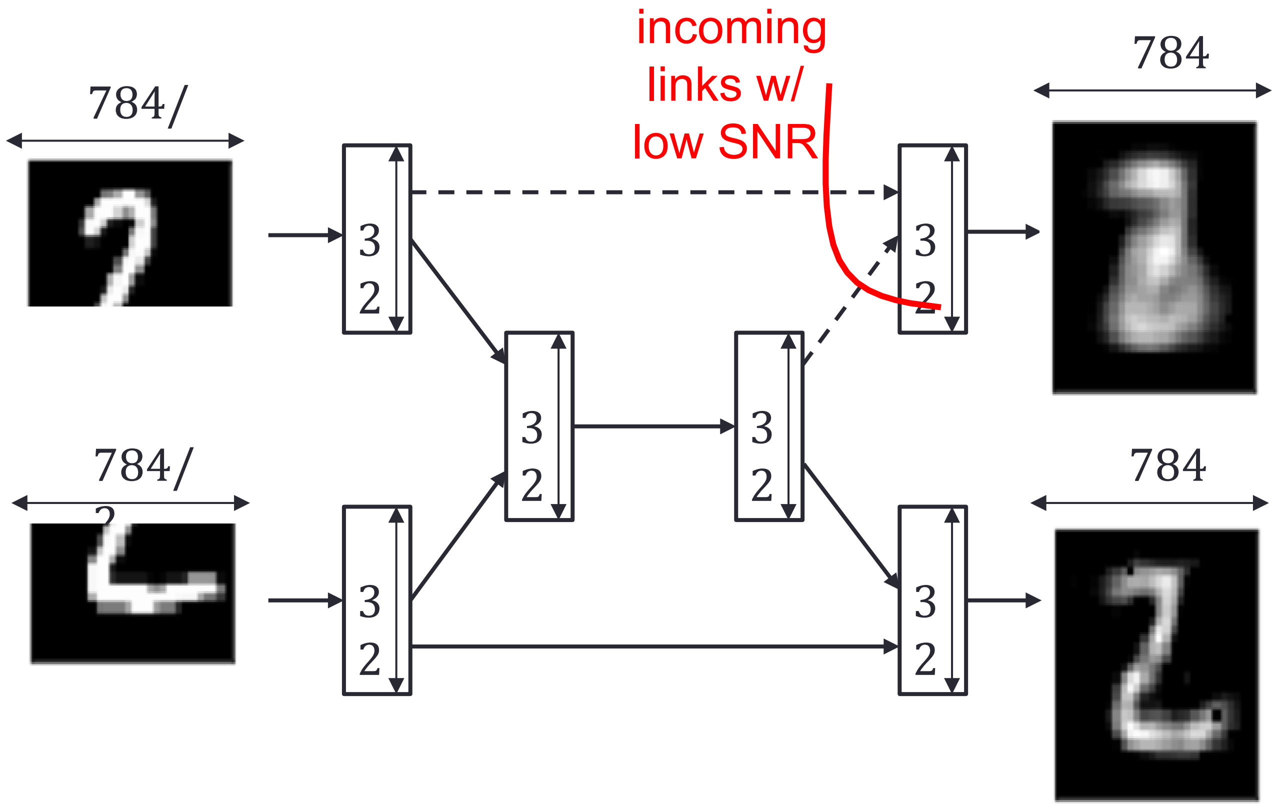
Performance Evaluation

High SNR on all links



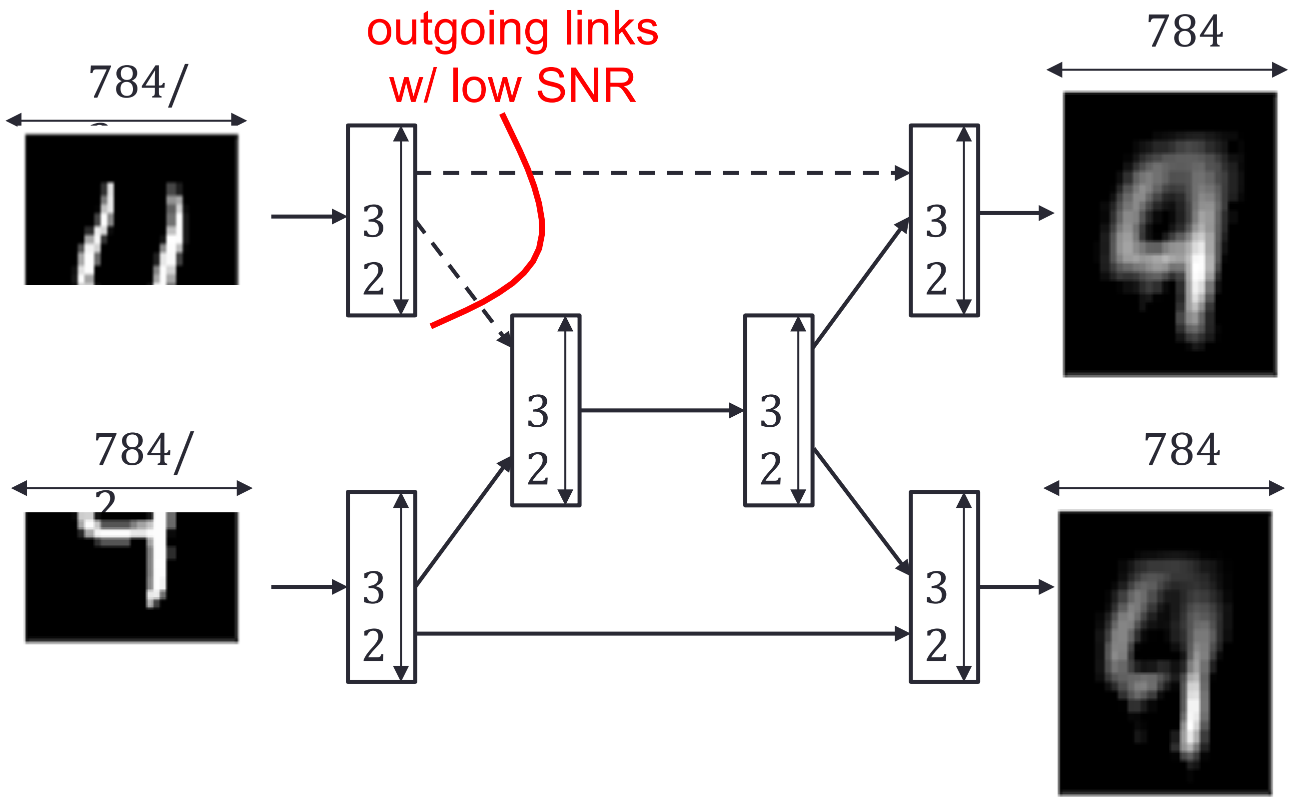
Performance Evaluation

One destination node with weak receiver



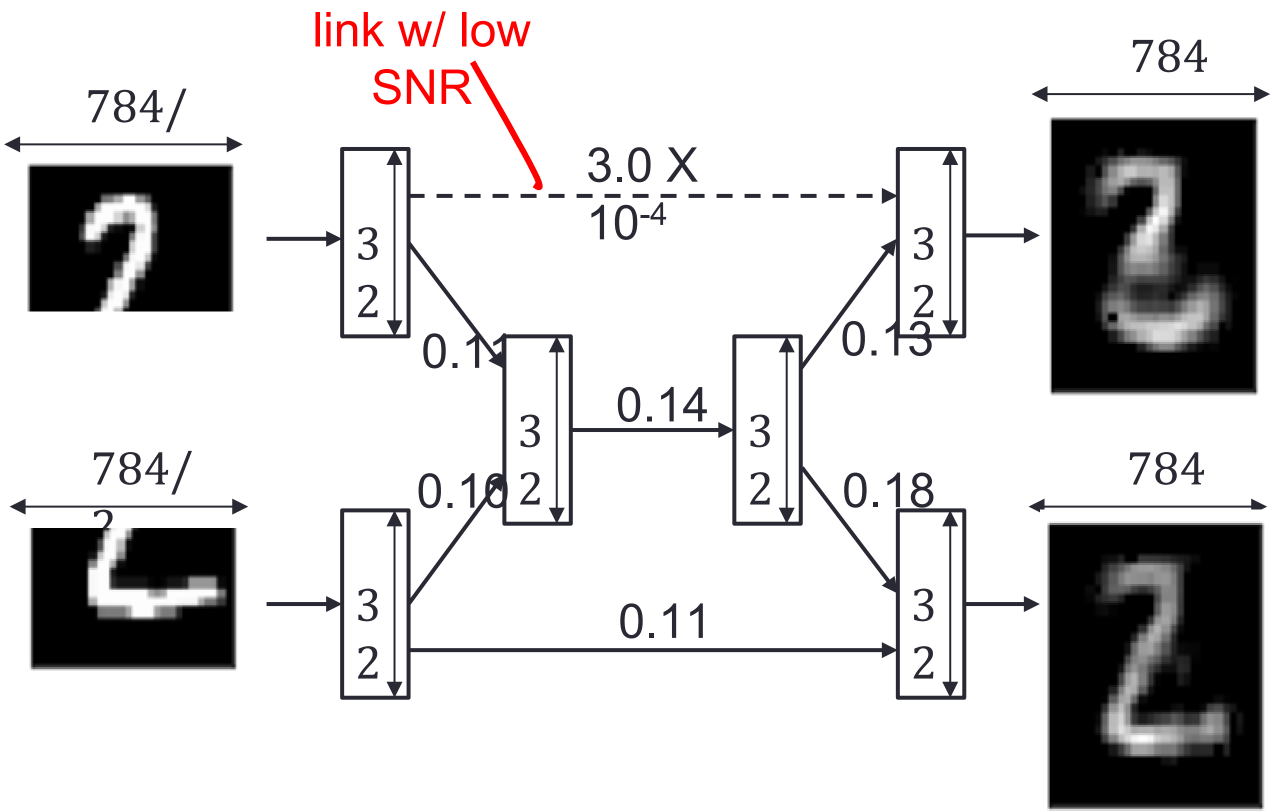
Performance Evaluation

One source node with weak sender



Performance Evaluation

One weak link

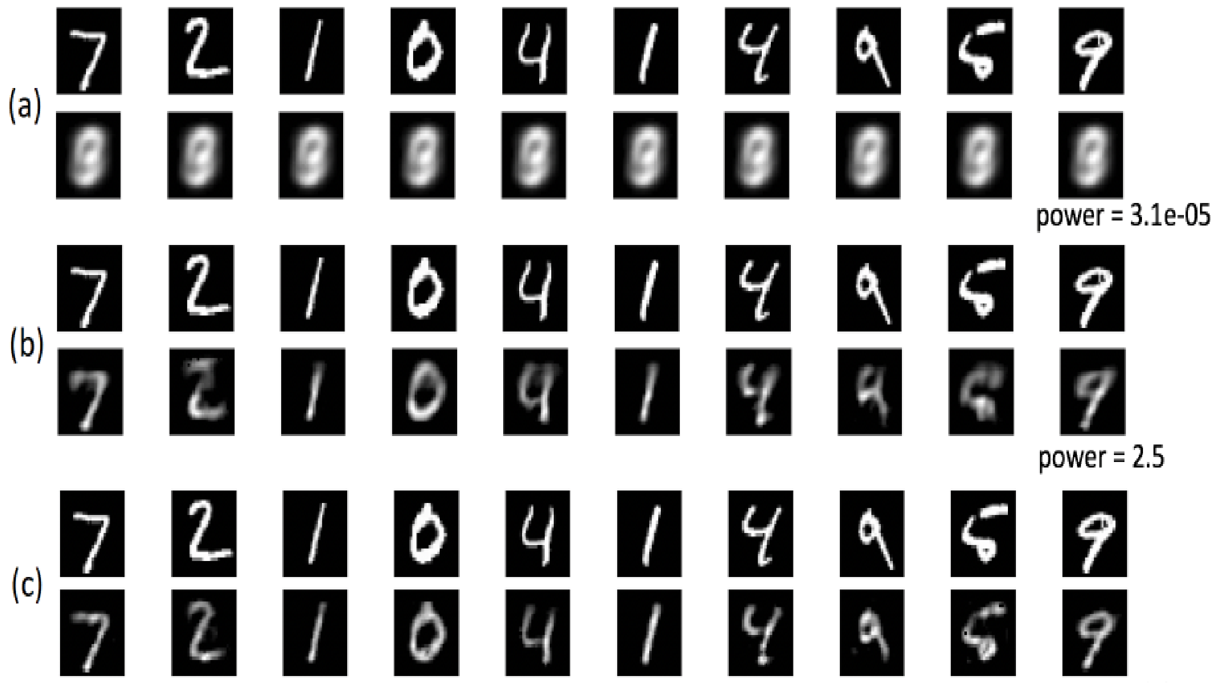


Performance Evaluation

All links equally strong

Power-Distortion Tradeoff

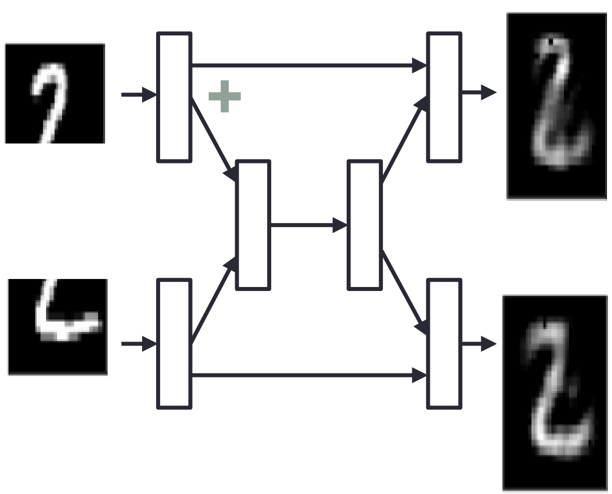
Water-filling

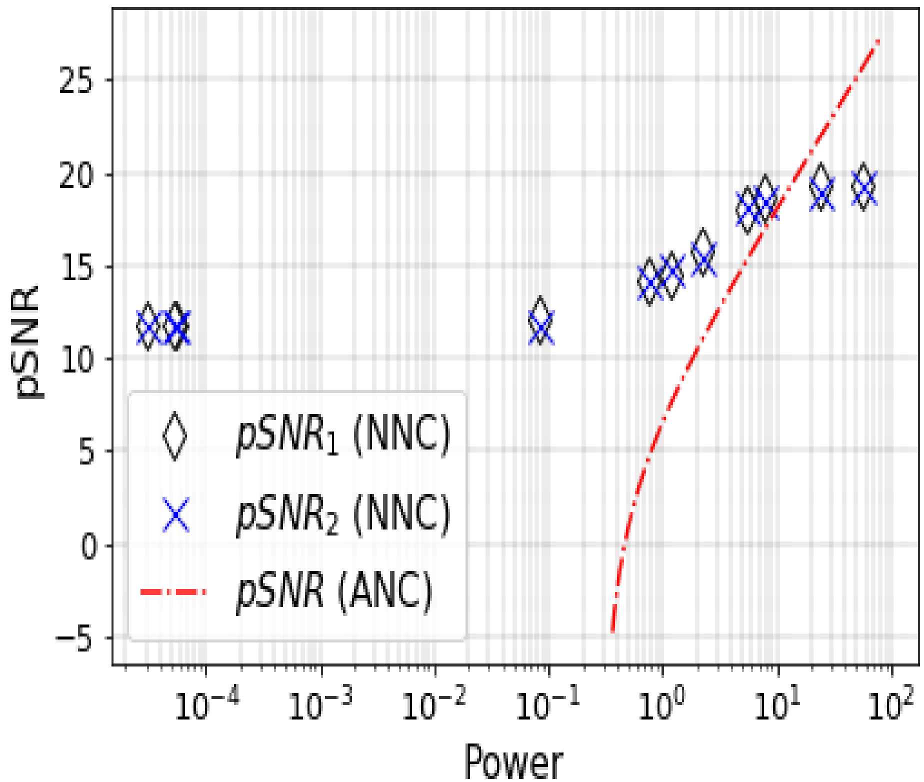


Performance Evaluation

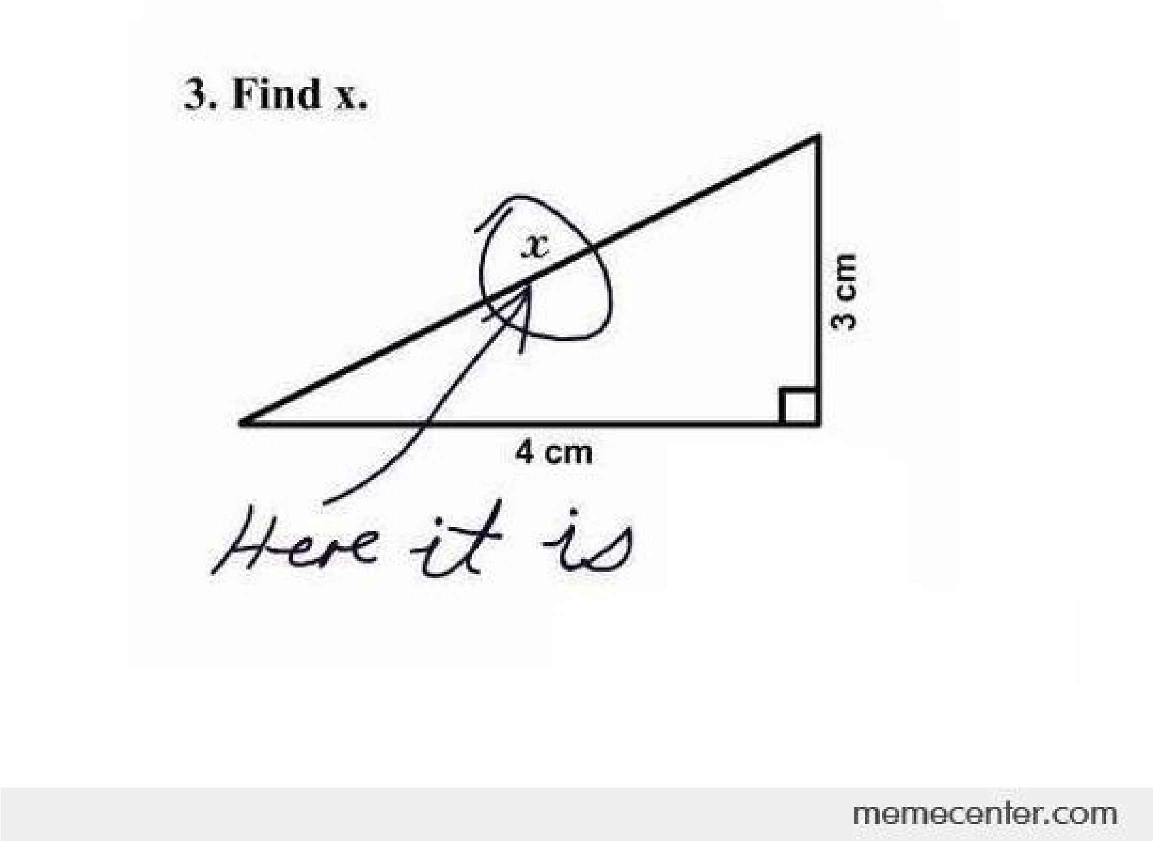
Comparison with Analog Network Coding

- No source compression
- All distortion from channel noise

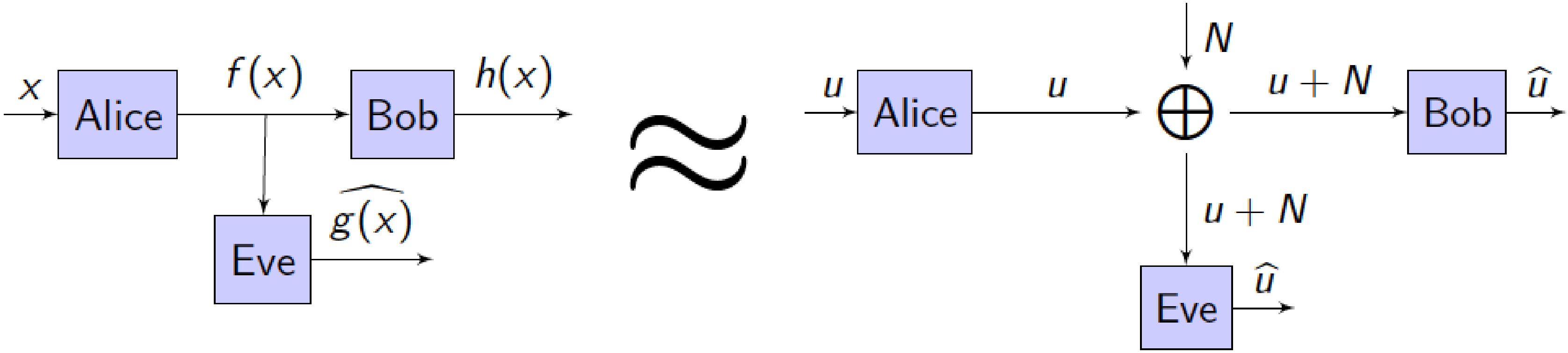




What is x?



Low influence functions



- Real valued function $f, g : \{-1, 1\}^n \rightarrow \mathbb{R}$
- Low influence functions: No coordinate has too much
- control on the function.
- Well behaved random variable: Conditions on x.

Rafael G. L. D'Oliveira, Salim El Rouayheb, Muriel Médard:, "The Computational Wiretap Channel". Allerton 2018.

Where to?

- Computation, networking and communication are increasingly united
- Information theory has tools to study an exploit this unification
- Further work:
  - Implementation: ongoing work with Alejandro Cohen, Manya Ghobadi, Benoît Pit-Claudel, Ganesh Ananthanarayanan (Microsoft), Derya Malak (RPI)
  - Characterize low influence functions
  - Multiterminal computational wiretap.