Results for Two Wireless Network Models

Babak Hassibi
California Institute of Technology

Joint work with R. Gowaikar, A.F. Dana, R. Palanki, M. Effros, B. Hochwald

Mathematics of Relaying and Cooperation in Communications Networks
Mathematical Sciences Research Institute

Berkeley, CA, April 11, 2006

Outline

- Introduction
- Wireless Erasure Networks[a]
  - model and problem statement
  - main result
- Scaling Laws for Random Wireless Networks[b]
  - model and problem statement
  - main result
  - two-scale and three-scale networks
- Conclusion

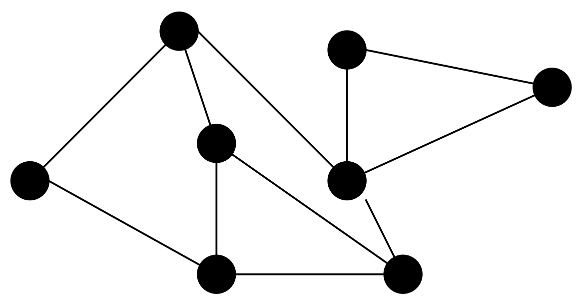
---

[a] Joint work with A.F. Dana, R.Gowaikar, R. Palanki, M. Effros

[b] Joint work with R.Gowaikar and B. Hochwald

Introduction

A wireless network is more than a collection of nodes and edges

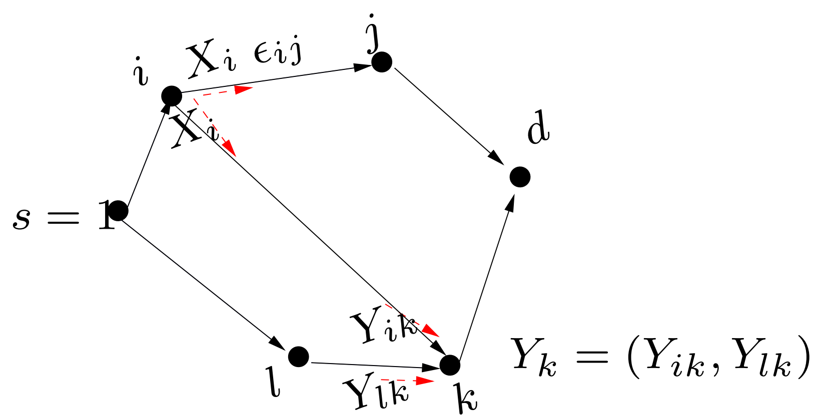


since the medium is *shared*, and we have *interference, broadcast, path-loss, and fading*.

- Given all the vagaries of wireless networks, obtaining “exact” capacity results appears beyond reach
- One must therefore resort to either
  1. simplified models that are mathematically tractable, yet somewhat faithful to the true physical model
  2. asymptotic results (large networks, high SNR, low SNR, large bandwidth, etc.)

Wireless Erasure Networks

- Network is an *acyclic directed graph* $\mathcal{G} = (\mathcal{V} = \{1, 2, \ldots, |\mathcal{V}|\}, \mathcal{E})$
- Each edge $(i, j) \in \mathcal{E}$ is an *memoryless packet erasure channel* from i to j with probability of erasure ϵ_{ij}
- *Broadcast:* Each node i sends the same signal through all its outgoing edges, X_i
- Each node receives all the signals from its incoming edges without interference, $Y_j = (Y_{ij}, \quad (i, j) \in \mathcal{E})$.



Wireless Erasure Networks–Continued

- This is an appropriate model for systems where transmissions are packet-based and where some form of error-correction is used to detect erasures
- Interference is not *directly* modeled in the network (primarily because it is not clear what interference means for erasure channels)
- However, it can be *indirectly* modeled by allowing for dependencies on the erasures coming into any node

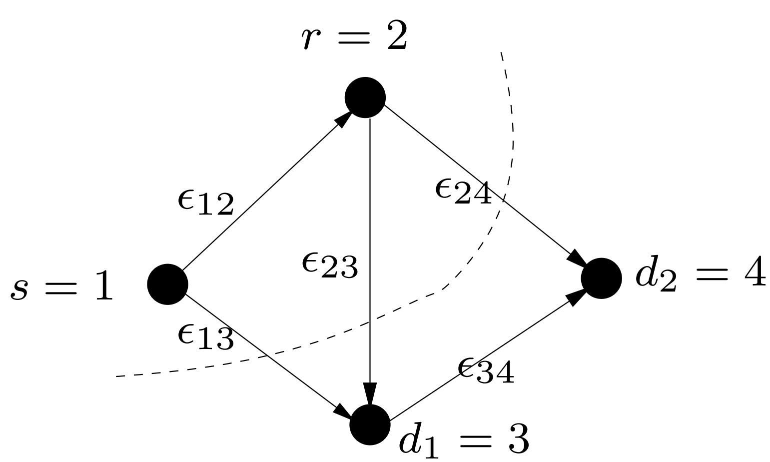
We wil the consider single source, multiple destination multicast problem. $s = 1 \in \mathcal{V}$ is the source and $\mathcal{D} = \{d_1, \ldots, d_{|\mathcal{D}|}\}$ is the set of destinations all requiring the same information.

Cuts and Cut-Sets

For graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

- An s-d cut is characterized by a set of nodes that contains s but not d
- For an s-d cut $\mathcal{V}_s$ we define the cut-set as

$$E(\mathcal{V}_s) = \{(i,j) | (i,j) \in \mathcal{E}, i \in \mathcal{V}_s, j \in \mathcal{V}_s^c\}$$



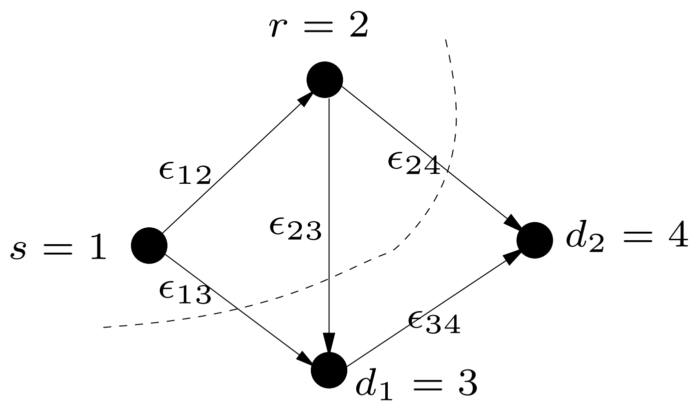
$s - d_2\, \text{cut} \mathcal{V}_s = \{1, 2\}$

$E(\mathcal{V}_s) = \{(2,4), (1,3), (2,3)\}$

Cut-Capacity

Definition 1 *Consider an erasure wireless network represented by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Suppose $\mathcal{V}_s$ is a subset of nodes containing source s but not d destination. Then the cut-capacity corresponding to this source set is*

$$C(\mathcal{V}_s) = \sum_{i:(i,j)\in E(\mathcal{V}_s)} (1 - \prod_{j:(i,j)\in E(\mathcal{V}_s)} \epsilon_{ij})$$



$s - d_2 \,\text{cut}\, \mathcal{V}_s = \{1, 2\}$

$E(\mathcal{V}_s) = \{(2,4), (1,3), (2,3)\}$

$C(\mathcal{V}_s) = 1 - \epsilon_{13} + 1 - \epsilon_{24}\epsilon_{23}$

Main Result

Key side information: the destination knows all the erasure occurrences on the links of the network

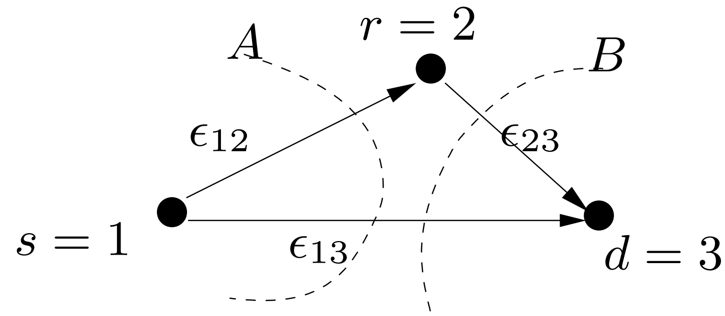
Theorem 1 *Capacity of a single source/destination erasure broadcast network, C, is given by the value of the min-cut, i.e.,*

$$C = \min_{\mathcal{V}_s \textit{ an s-d cut}} C(\mathcal{V}_s)$$

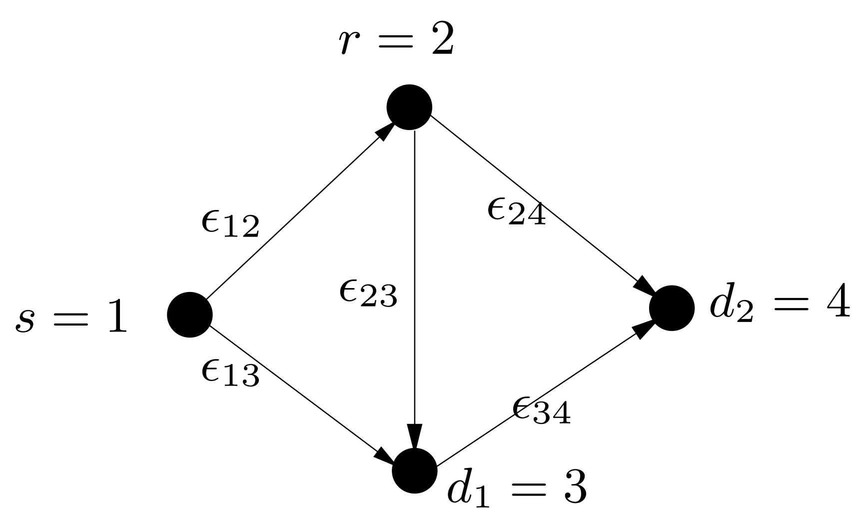
Theorem 2 *Capacity of a single source/multiple destination erasure broadcast network, C, is*

$$C = \min_{d_i \in \mathcal{D}} \min_{\mathcal{V}_s \textit{ an s-}d_i \textit{ cut}} C(\mathcal{V}_s)$$

Some Examples



$$C = \min\{1 - \epsilon_{13}\epsilon_{12}, 1 - \epsilon_{13} + 1 - \epsilon_{23}\}$$



$$C = \min\{1 - \epsilon_{13}\epsilon_{12}, 1 - \epsilon_{13} + 1 - \epsilon_{23}, 1 - \epsilon_{12} + 1 - \epsilon_{34}, 1 - \epsilon_{24} + 1 - \epsilon_{34}\}$$

Wireless Erasure Networks: Formal Definitions

A $(2^{nR}, n)$ code consists of the following components:

- **Message set:** $\mathcal{W} = \{1, 2, \ldots, 2^{nR}\}$. Messages are equally likely
- **Encoding function for source node:** $f_1 : \mathcal{W} \to \{0,1\}^n$. In order to transmit $k \in \mathcal{W}$, source node transmits $f_1(k)$
- **Encoding function for other nodes:** for node $i \in \mathcal{V}$:

$$f_i : \{0, 1, e\}^{nd_I(i)} \to \{0,1\}^n$$

  Upon receiving $Y_i = y_i$, node i sends out $x_i = f_i(y_i)$

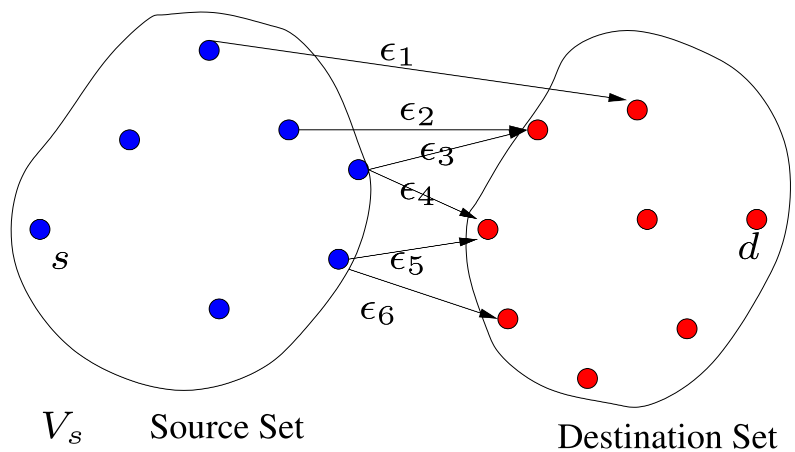
- **Decoding function at a destination:** For destination $d_i \in \mathcal{D}$,

$$g_{d_i} : \{0,1,e\}^{nd_I(d_i)} \times \{0,1\}^{n|\mathcal{E}|} \to \mathcal{W} = \{1, \ldots, 2^{nR}\}$$

  Destination d_i outputs $\hat{w}_{d_i} = g_{d_i}(Y_{d_i} = y_{d_i}, S = s)$ as the transmitted message. S is the vector of erasure occurrences across the network.
  $d_I(v)$ is the in degree of node v.

Upper Bound

- The upper bound is readily found from a min-cut argument, assuming cooperation among the nodes in the source set and the nodes in the destination set
- The cut-capacity defined earlier is essentially the resulting point-to-point capacity assuming full cooperation among these two sets



$$C(V_s) = 1 - \epsilon_1 + 1 - \epsilon_2 + 1 - \epsilon_3\epsilon_4 + 1 - \epsilon_5\epsilon_6$$

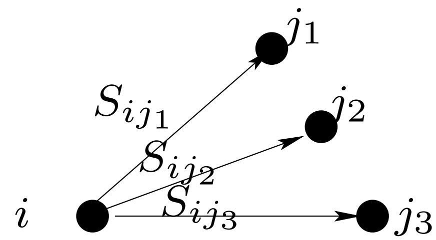
Achievability

- Achievability is established by a random coding argument
- **Encoding:** For each y in the domain of f_i we generate each element of $f_i(y)$ i.i.d. from $\{0, 1\}$ and according to a Bernoulli distribution with parameter $\frac{1}{2}$
- **Decoding:** Suppose the transmitted message is w and the destination d has received $y_d(w)$ at all its incoming edges.

  The destination "simulates" the network for all the messages in $\mathcal{W}$ and outputs the message(s) that result in $y_d(w)$. Knowing the error locations S, the destination node can unambiguously determine the message set of its incoming edges for any transmitted input. Therefore, it can compute $Y_d(k)$ for all the messages.
- **Error Event:** If there is a unique message $w' \in \mathcal{W}$ s.t. $Y_d(w') = y_d(w)$ then w' is the declared as the transmitted message. Otherwise an error is declared.

Achievability: Prob. of Error Analysis

- **Joint strong typicality and Union bound:**
  - We look at "strongly typical" instances of the network. According to the strong law of large numbers for large block lengths, the network is "strongly typical" with probability approaching one, i.e. $\Pr(A_\delta^{(n)}) \to 1$.



$(S_{ij_1}, S_{ij_2}, S_{ij_3})$ is δ-strongly typical.

  - Define: $E_w = \{Y_d(w') = Y_d(w)\}, \quad w \neq w' \in \mathcal{W}$. Averaging over all the functions f_i:

$$\begin{aligned} \mathrm{E}P_e^{(n)} &= \mathrm{E}\lambda(w) = \Pr(\bigcup_{w' \neq w} E_{w'}) \\ &\leq \Pr(A_\delta^{(n)}) \sum_{w' \neq w} \Pr(E_{w'} | A_\delta^{(n)}) + \Pr({A_\delta^{(n)}}^c) \end{aligned}$$

Achievability: Prob. of Error Analysis

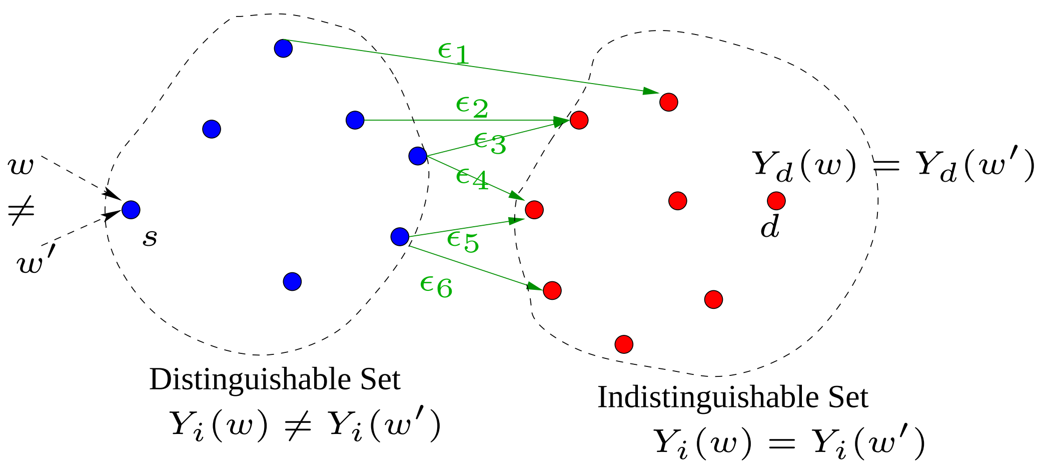
- **Distinguishable sets:** Define for any s-d cut V_s:

$$B[V_s] = \left[\bigcap_{i \in V_s^c} \{Y_i(w) = Y_i(w')\}\right] \bigcap \left[\bigcap_{i \in V_s} \{Y_i(w) \neq Y_i(w')\}\right]$$

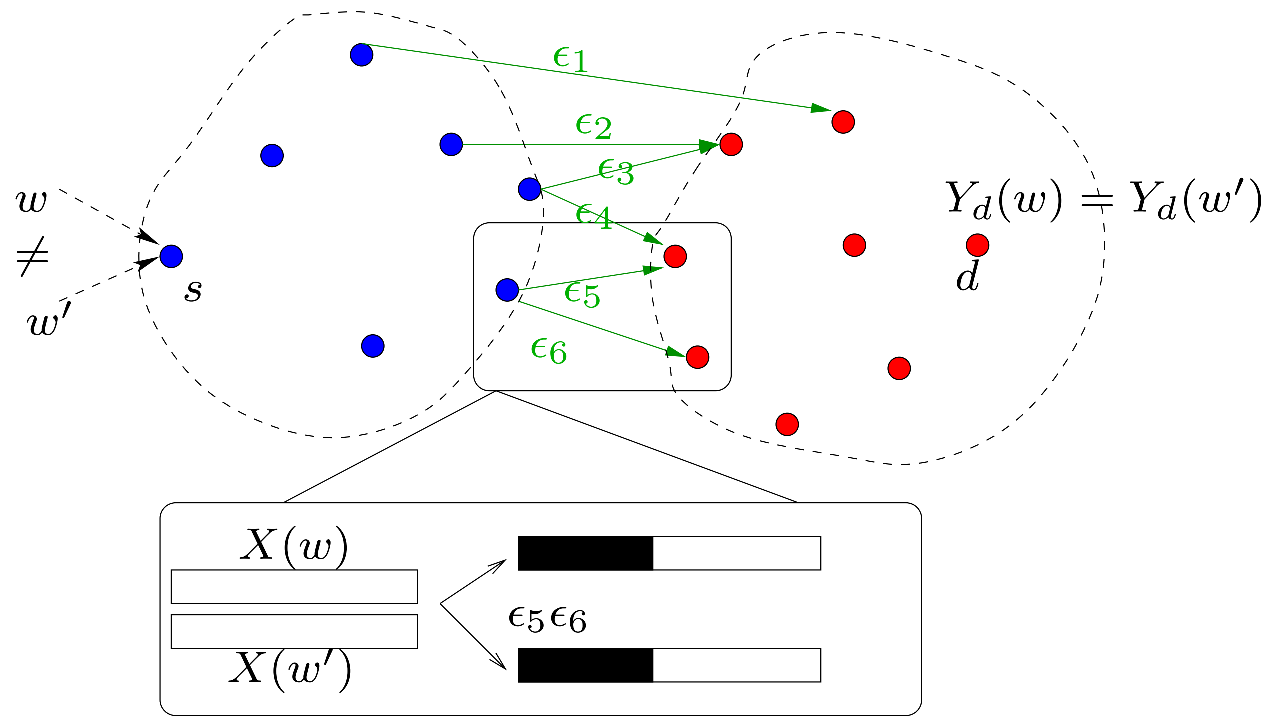
then

$$E_{w'} = \bigcup_{V_s:\,\text{s-d cut}} B[V_s].$$

$$\Pr(E'_w | A_\delta^{(n)}) = \Pr(\cup_{V_s:\,\text{s-d cut}} B[V_s] | A_\delta^{(n)}) \leq \sum_{V_s:\,\text{s-d cut}} \Pr(B[V_s] | A_\delta^{(n)})$$



Achievability: Prob. of Error Analysis



$$\Pr\left(B[V_s] | A_\delta^{(n)}\right) \le 2^{n|\mathcal{V}|\delta} 2^{-n(1-\epsilon_1)} 2^{-n(1-\epsilon_2)} 2^{-n(1-\epsilon_3\epsilon_4)} 2^{-n(1-\epsilon_5\epsilon_6)}$$
$$\le 2^{-nC(V_s)} 2^{n|\mathcal{V}|\delta}$$

$$\Rightarrow \mathrm{E}\, P_e^{(n)} \le (1-\mu(\delta)) \sum_{V_s \text{ :s-d cut}} 2^{n(R-C(V_s))} + \mu(\delta)$$

$\Rightarrow R \le \min_{V_s \text{ :s-d cut}} C(V_s)$ is achievable

Linear Block Codes

- It can be shown that we can achieve capacity with random linear coding
- With linear encoding, the decoding complexity becomes polynomial time (we need only to solve a system of linear equations)
- In this case each f_i is a linear mapping from the outputs of the incoming edges to i to $\{0,1\}^n$ and it can be presented by a matrix multiplication.

$$Y_i(w) = \begin{bmatrix} d \\ e \\ e \\ d \\ d \\ e \end{bmatrix} \longrightarrow Y_i'(w) = \begin{bmatrix} d \\ 0 \\ 0 \\ d \\ d \\ 0 \end{bmatrix} \longrightarrow X_i(w) = f_i \cdot Y_i'(w)$$

Discussion

- Note that we do not have to perform separate channel and network coding to achieve capacity, In fact, making each link error-free is demonstrably sub-optimal.
- The assumption on the side information of the erasure locations is key
  - the cut-set bounds are independent of the side information assumed
  - what we have shown is that this particular side information (erasure locations) makes the cut-set bound achievable
  - can a similar approach be applied to other classes of networks?

- In practice, the side information can be provided by adding headers to the packets
- For large packet lengths (or small size networks), the size of the side-information is negligible to the data length. In fact the effective capacity is

$$C_{eff} = \frac{nCL}{nL + n|\mathcal{E}|} = \frac{C}{1 + \frac{|\mathcal{E}|}{L}}$$

where L is the packet length, and so as $L \to \infty$, we have $C_{eff} \to C$

Further and Future Work

- Practical encoding/decoding schemes (ala network coding)
- How best to model a physical wireless network as a wireless erasure network?
- What if the side-information is not available at the destination?
- Non-multicast network problems, especially ones with more than one source. (We have recently solved the broadcast problem for such networks).

Scaling Laws in Ad Hoc Networks

There has been a great deal of interest in the capacity of ad hoc wireless networks

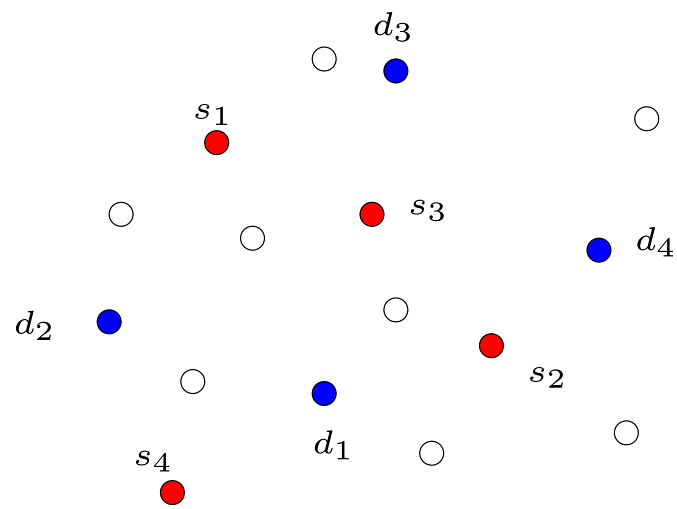
- Seminal work of Kumar-Gupta: scaling law of $O(\sqrt{n})$
- Many variations (Xie and Kumar, Baccelli et al, Dousse and Thiran, Leveque and Telatar, Franceschetti et al, Tonguz and Ferrari, Toumpis and Goldsmith, Weber et al)
- Improvement obtained from mobility, $O(n)$, (Grossglauser and Tse, Diggavi et al)

All the above results are based on geometric considerations

- Recent studies suggest that in some situations (e.g., shadow fading) the network may behave more like a random graph (Hekmat and van Mieghem, Ballister et al, Franceschetti et al)

We will explore the implications in this talk

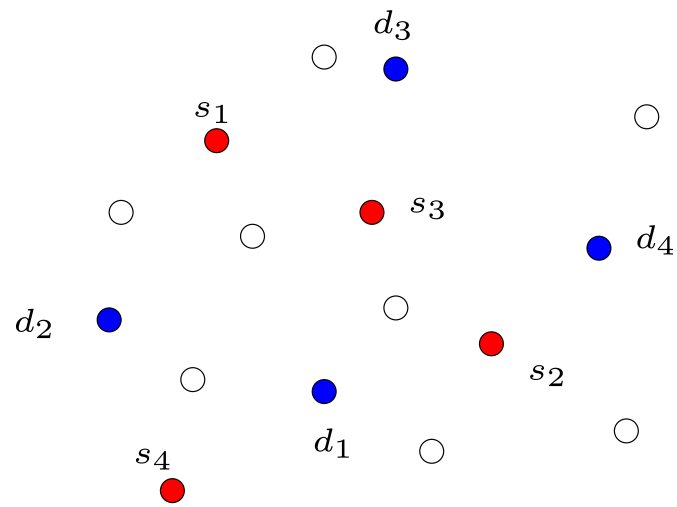
Model and Problem Statement



- n nodes, k source-destination pairs
- channel h_{ij} between nodes i and j, ($\gamma_{ij} = |h_{ij}|^2$)
- we have broadcast, interference and additive noise (AWGN)

We are interested in determining the total throughput of the network

Existing Results – Distance Based Models



Kumar and Gupta basically used a decay law:

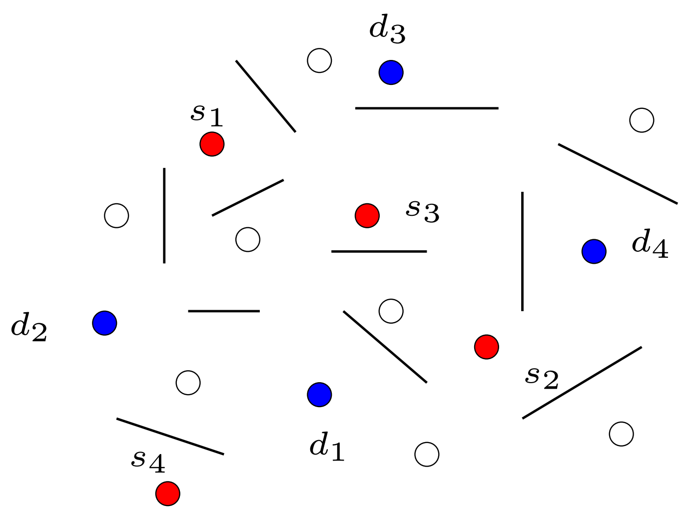
$$\gamma_{ij} \propto \frac{1}{(\text{distance}_{i,j})^m}, \quad m > 2$$

Rough idea: $O(n)$ nodes can simultaneously communicate to their close neighbors. However, getting to the desired destination requires $O(\sqrt{n})$ hops.

- Pretty much all variants to this model give $O(\sqrt{n})$ throughput

New Model – Random Links

- Strength of connection does not depend on distance
- All connection strengths are identically and independently distributed according to $f_n(\gamma)$



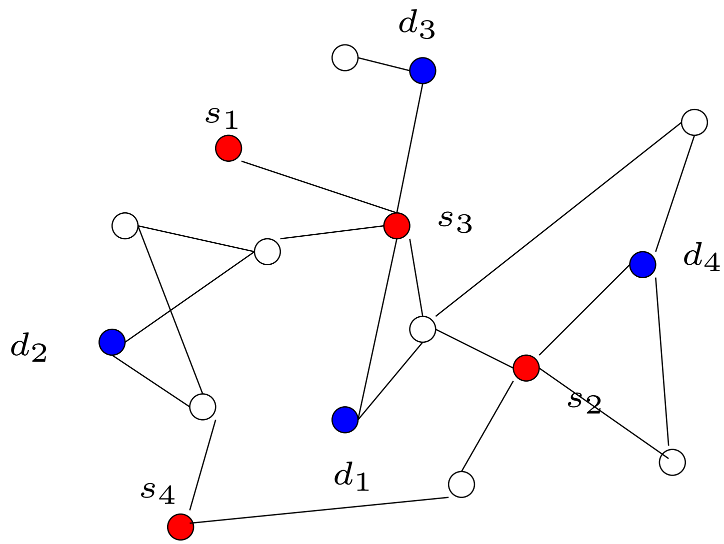
- Model is justified over small area with many obstructions
- Line-of-sight is absent, multipath dominates

How to Determine the Throughput?

- We would still like to use the idea of multi-hop routing (messages get decoded at relay nodes and passed along routes)
- However, we can no longer use geometric intuition (transmission to nearest neighbors to reduce interference)

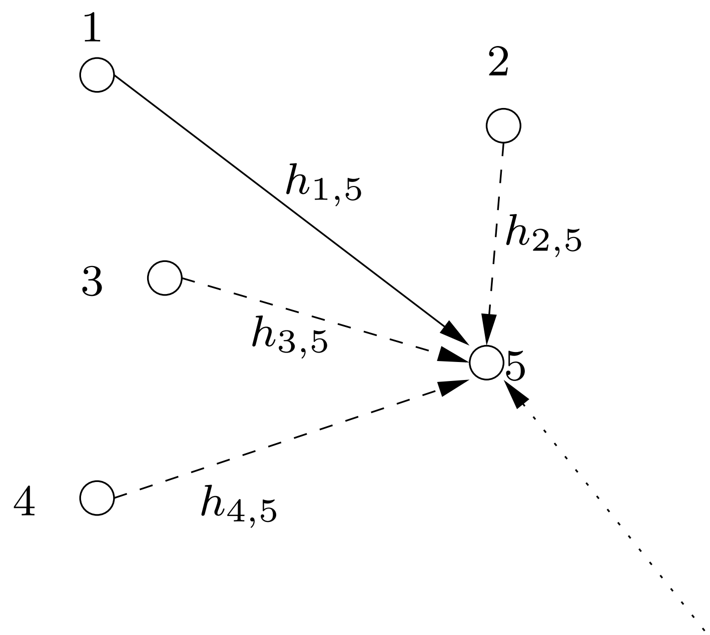
Main idea: Introduce the conecpt of "good" edges

- Introduce a parameter β_n. Let $\mathrm{P}(\gamma \geq \beta_n) = Q_n(\beta_n) = p_n$.
- Edges with $\gamma \geq \beta_n$ are called "good."



- Keeping only good edges and eliminating the rest gives us an instance of the random graph model $G(n, p)$
  - $G(n, p)$ is a well-studied random graph model
  - for example, if p behaves like $\frac{\log n}{n}$, the diameter of the graph behaves like $\frac{\log n}{\log \log n}$
  - thus, there *may* be *some* hope in obtaining throughput $O(\frac{n}{\log n})$
  - this depends on finding enough source-destimation paths that do not collide and whose interference is kept under control

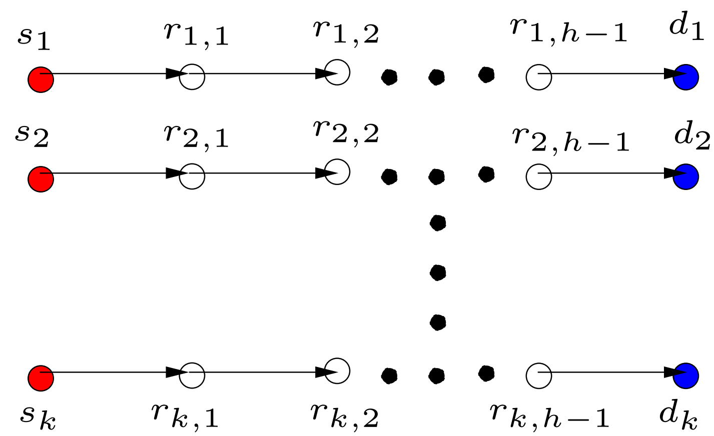
Successful Communication



A node can decode only if the signal-to-interference-plus-noise-ratio (SINR) exceeds some threshold ρ_0. Node 5 can decode only if

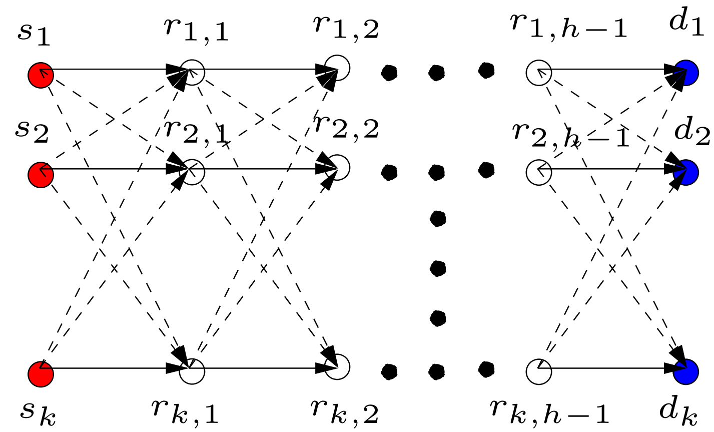
$$\frac{P|h_{1,5}|^2}{\sigma^2 + P(|h_{2,5}|^2| + |h_{3,5}|^2 + |h_{4,5}|^2)} = \frac{P\gamma_{1,5}}{\sigma^2 + P(\gamma_{2,5} + \gamma_{3,5} + \gamma_{4,5})} \geq \rho_0.$$

Scheduling Multiple Hops



- Relay nodes decode and retransmit message
- Nodes cannot transmit and receive simultaneously
  - therefore non-colliding schedule required
- h hops are used for each communication

Error Free Decoding



- SINR condition must be fulfilled at every relay node.
- Want to maximize throughput

$$T = (1-\epsilon)\frac{\text{Number of source-destination pairs}}{\text{Avg number of hops}}\log(1+\rho_0)$$

$$T = (1-\epsilon)\frac{k}{h}\log(1+\rho_0).$$

Vertex Disjoint Paths

Theorem 1 *(Broder et al. 1996) Suppose that* $G = G(n,p)$ *and* $p \geq \frac{\log n + \omega_n}{n}$*, where* $\omega_n \to \infty$*. Then there exists a positive constant* α *such that, with probability going to 1, there are vertex-disjoint paths connecting* s_i *to* d_i *for any set of disjoint, randomly chosen source-destination pairs*

$$F = \{(s_i, d_i) | s_i, d_i \in \{1, \ldots, n\}, i = 1, \ldots, k\}$$

provided k *is not greater than* $\alpha n \frac{\log np}{\log n}$*.*

- We can thus establish upto $k = \alpha n \frac{\log np}{\log n}$ non-colliding (in fact, vertex-disjoint) paths. Theorem is tight upto constant.
- Can show that every message makes around $h = \frac{\log n}{\alpha \log np}$ hops
- We still need to ensure that the probability of error can be made to go to zero (i.e., that the interference is under control)

Probability of Error

$$\begin{aligned} \text{P(Message } i \text{ fails)} &= \text{P}(\bigcup_{j=1}^{h} \text{SINR at hop } j \leq \rho_0) \\ &\leq \#\text{ hops } \times \text{P(SINR at hop } 1 \leq \rho_0) \end{aligned}$$

$$\epsilon_n = \text{P(Message } i \text{ fails)} \leq \frac{\log n}{\alpha \log np} \frac{\sigma_\gamma^2/(k-1)}{(\frac{P\beta_n - \rho_0 \sigma^2}{(k-1)P\rho_0} - \mu_\gamma)^2}.$$

Can use Chebyshev bound if $\rho_0 \leq \dfrac{P\beta_n}{\sigma^2 + P(k-1)\mu_\gamma}$.

Put Scheduling and Probability of Error results together to get Main Result.

Main Result

Theorem 2 *Let $Q_n(x) = \mathrm{P}(\gamma \geq x)$. Choose any β_n such that $Q_n(\beta_n) = \frac{\log n + \omega_n}{n}$, where $\omega_n \to \infty$. Let $\mu_\gamma = E\gamma$, $\sigma_\gamma^2 = E(\gamma - \mu_\gamma)^2$. Then there exists a positive constant α such that a throughput of*

$$T = (1 - \epsilon_n)\ k_n(\beta_n)\alpha \frac{\log(nQ_n(\beta_n))}{\log n}\ \log\left(1 + \frac{a_n \beta_n}{\frac{\sigma^2}{P} + (k_n(\beta_n) - 1)\mu_\gamma}\right)$$

is achievable for any positive a_n such that $a_n \leq 1$ and any $k_n(\beta_n)$ that satisfy the conditions:

1.

$$k_n(\beta_n) \leq \alpha n \frac{\log(nQ_n(\beta_n))}{\log n}$$

2.

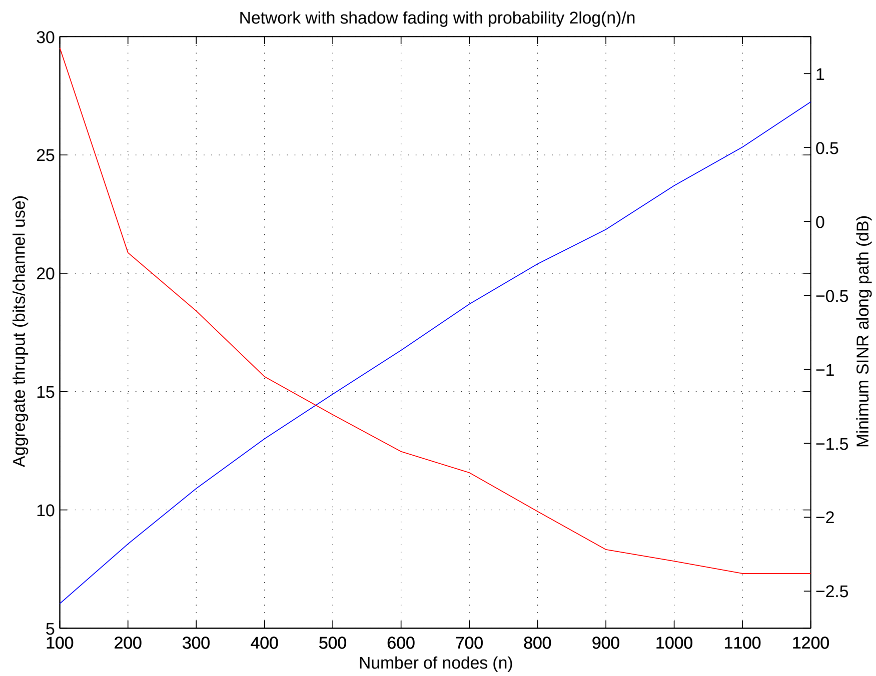
$$\epsilon_n \leq \frac{a_n^2}{\alpha(1 - a_n)^2} \frac{(k_n(\beta_n) - 1)\sigma_\gamma^2}{(\frac{\sigma^2}{P} + (k_n(\beta_n) - 1)\mu_\gamma)^2} \frac{\log n}{\log(nQ_n(\beta_n))} \to 0$$

Heavy Dependence on $f_n(\gamma)$

Distribution	$f_n(\gamma)$	Throughput
Shadow	$(1-p_n)\delta(\gamma)+p_n\delta(\gamma-1)$	$\frac{1}{w_n}\frac{\log^2(\log n)}{\log^3 n}\ n$
Exponential	$e^{-\gamma}$	$\log n$
Decay	$\frac{4\pi\Delta}{nm}\frac{1}{\gamma^{1+\frac{2}{m}}}, m>2$	$\frac{1}{w_n}\frac{\log^2(\log n)}{(\log n)^{2+m/2}}\ n$
Heavy Tail	$\frac{c}{1+\gamma^4}$	$\frac{\log\log n}{\log^{4/3} n}\ n^{1/3}$

Shadow Fading Distribution – Scaling Law

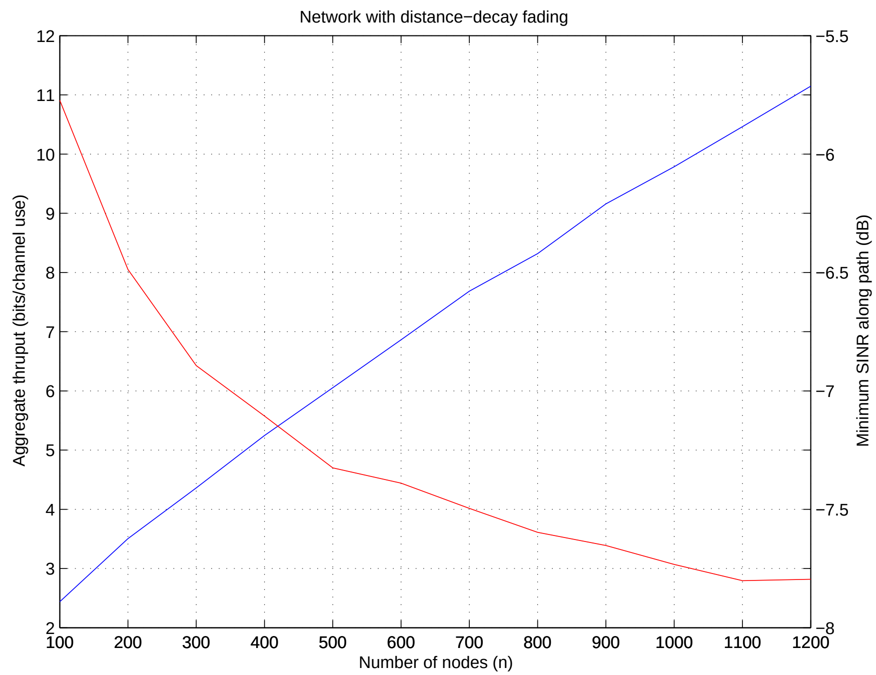
- For $p_n = \frac{\log n + \omega_n}{n}$, $h = \frac{\log n}{\log(\log n + \omega_n)}$, $T = \frac{1}{w_n} \frac{\log^2(\log n)}{\log^3 n}\ n$



$p = \frac{2 \log n}{n}$, 100 – 1200 nodes.

Density obtained from Decay Law

- $f_n(\gamma) = \frac{4\pi\Delta}{nm} \frac{1}{\gamma^{1+\frac{2}{m}}}$. Optimum choice of β_n is $\frac{2\pi\Delta}{(\log n + \omega_n)^{m/2}}$
- $p_n = \frac{\log n + \omega_n}{n}$, $h = \frac{\log n}{\log(\log n + \omega_n)}$, $T = \frac{1}{\omega_n} \frac{\log^2(\log n + \omega_n)}{\log^2 n (\log n + \omega_n)^{3/2}}\, n$



$m = 3$, $d = 1$, $\Delta = 1$. Throughput increases almost linearly.

Criticisms of the Model

- While all this is fine and encouraging, the model only applies to networks of small physical size?
- It is not a good model for very large (again, physically large) networks

But does this have any implications for networks that span a large physical area?

Two-Scale Networks

To capture both the "random" effects that dominates for small networks and the "geometric" decay that dominates for large networks, we propose the following two-scale model

- The network consists of N nodes randomly placed on a sphere of radius R
- *local scale:* connections that are of a distance less than r, are chosen iid from a distribution $f(\gamma)$
- *global scale:* connections that are of a distance greater than r are chosen from Rayleigh distribution whose power satisfies a decay law
- more precisely the connections are drawn from

$$p_x(\gamma) = \begin{cases} f(\gamma) & \text{if } x \leq r \\ \frac{\mu_\gamma r^m}{x^m} \exp(-\gamma \frac{x^m}{\mu_\gamma r^m}) & \text{if } x > r \end{cases}$$

where x represents the distance, $m > 2$, the decay law and $\mu_\gamma = E_f \gamma$

Analysis of Two-Scale Networks

We can, in fact, analyze such networks using a combination of the tools developed for random networks and the tools used by Kumar-Gupta

- We start with a tesselation of the sphere so that each cell is at the local scale
- We obtain a *super-schedule* for communication between the tesselations
- Communication at the level of tesselations is that of a random network and so is determined by a *sub-schedule*
- the local and global interference are each dealt with

Main Result

Theorem 3 *Consider the two-scale network just decribed.Let $F(\gamma)$ denote the cumulative distribution function of $f(\gamma)$ and $Q(\gamma) = 1 - F(\gamma)$. Let $n = c_2 N \sin^2 \frac{r}{24R}$ where c_2 is a known constant. Choose any β such that $p = Q(\beta) = \frac{\log n + \omega_n}{n}$, where $\omega_n \to \infty$ as $n \to \infty$. Then a throughput of*

$$T = (1-\epsilon)\frac{\alpha K r \log np \cdot \log\left(1 + \frac{P\beta}{\sigma^2 + a'(PK \sin^2 \frac{r}{4R}\mu_\gamma + PK\frac{r^2 \mu_\gamma}{2R^2})}\right)}{\log n \cdot c_1 R \cdot \log N}$$

is achievable where α and c_1 are constants and K and $a' \geq 1$ are chosen such that the following conditions are satisfied.

1. $K \leq N/(8\cos^2 \frac{r}{24R})$.
2. $K \leq \alpha N \log np/(\log n \cdot \cos^2 \frac{r}{24R})$.
3. $\epsilon \leq \frac{\log n}{\alpha \log np} \cdot \frac{R}{r} \cdot \frac{1}{a'} \to 0$

An Example

Again, the result is very dependent on $f(\gamma)$, the local scale distribution

- Here is an example:
  - Suppose
    $$f(\gamma) = \frac{K}{(1+\gamma)^t}, \quad t > 2$$
    then a throughput of
    $$\frac{N^{\frac{1}{t-1}}}{\log^2 N}$$
    For t almost 2, this is almost linear in N. For $t > 3$ it is worse than $\sqrt{N}$.

Some Further Models

- *Three-Scale Models*

$$p_x(\gamma) = \begin{cases} f(\gamma) & \text{if } x \le r_1 \\ \frac{r_2 - x}{r_2 - r_1} f(\gamma) + \frac{x - r_1}{r_2 - r_1} \frac{\mu_\gamma r^m}{x^m} \exp(-\gamma \frac{x^m}{\mu_\gamma r^m}) & \text{if } r_1 < x \le r_2 \\ \frac{\mu_\gamma r^m}{x^m} \exp(-\gamma \frac{x^m}{\mu_\gamma r^m}) & \text{if } x > r_2 \end{cases}$$

The analysis and results are almost identical to two-scale networks.

- *Mixture Models*

$$p_x(\gamma) = \frac{R - x}{R} f(\gamma) + \frac{x}{R} \frac{\mu_\gamma r^m}{x^m} \exp(-\gamma \frac{x^m}{\mu_\gamma r^m})$$

This one seems much more difficult to analyze.

Concluding Remarks

- Random models give significantly different scaling laws than distance-based models
  - They are heavily dependent on the pdf $f_n(\gamma)$.
- We can also give upperbounds on the achievable throughput using multihop protocols
  - Bound of $O(n)$ is expected in general
  - Bound of $O(\log n)$ can be shown for exponential distribution. Same as achieved.
- Future work
  - Analysis of mixture models
  - Information-theoretic upperbounds
  - Decentralized implementation schemes