Network Tomography using Network Coding

Athina Markopoulou
EECS, UC Irvine

Joint work with
Christina Fragouli, Suhas Diggavi, Ramya Srinivasan
at EPFL, Lausanne

Problem Context

- Network Monitoring and Diagnosis
  - Network Tomography
- Network Coding
- Goal of this work:
  - How to do tomography in networks with network coding already implemented?

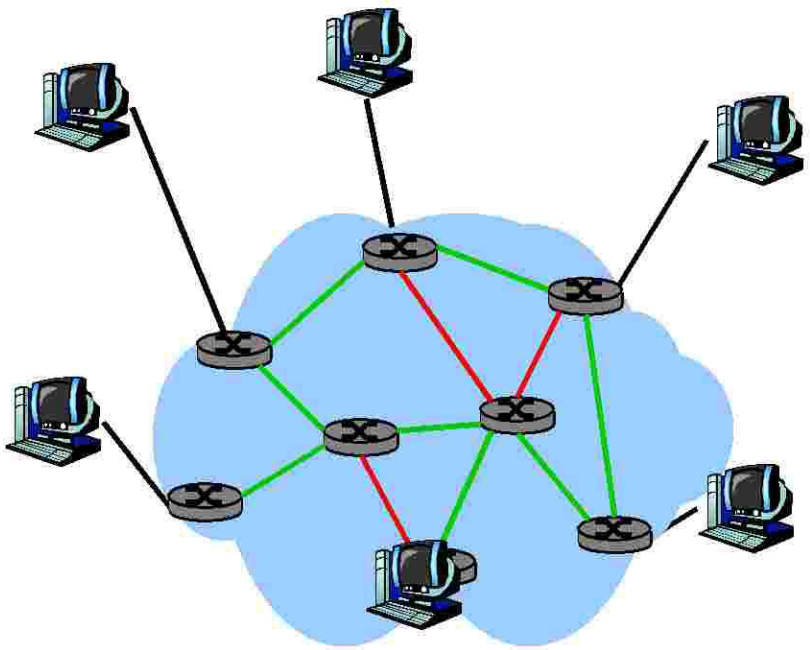
Outline

- Background
  - Network tomography
  - Network coding
- Topology Inference using Network Coding
- Link Loss Inference using Network Coding
- Conclusions

What is Network Tomography

Goal: obtain detailed picture of a network/internet from end-to-end views

- infrastructure
- infer link-level
  - loss
  - delay
  - utilization

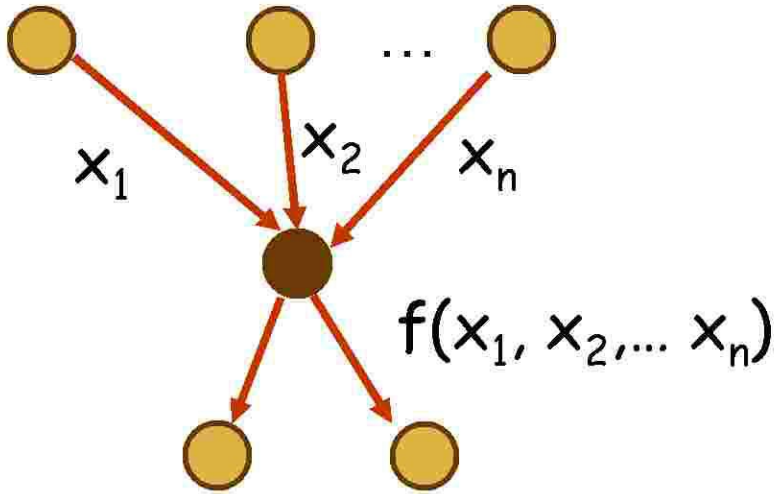


Why end-to-end measurements?

- no participation by network needed
- no administrative access needed
- inference across multiple domains
  - no cooperation required
  - e.g. to monitor service level agreements

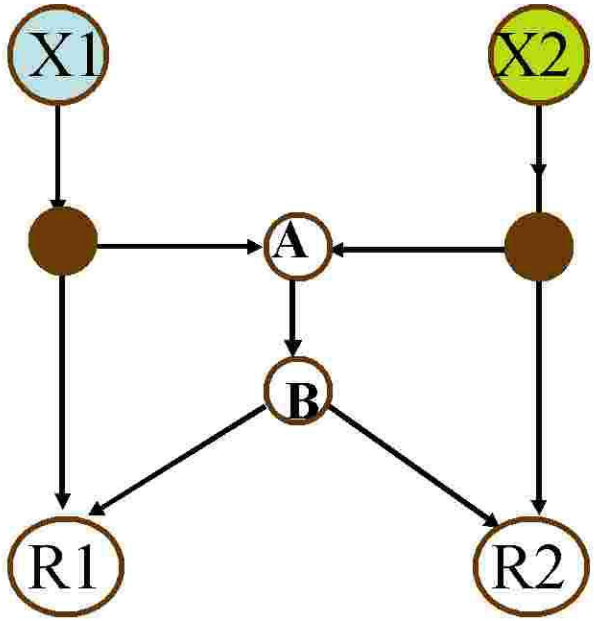
What is Network Coding?

- Allow intermediate nodes to perform operations on incoming packets before forwarding them



Example (1)

Ahlswede,Cai,Li,Yeung 2000

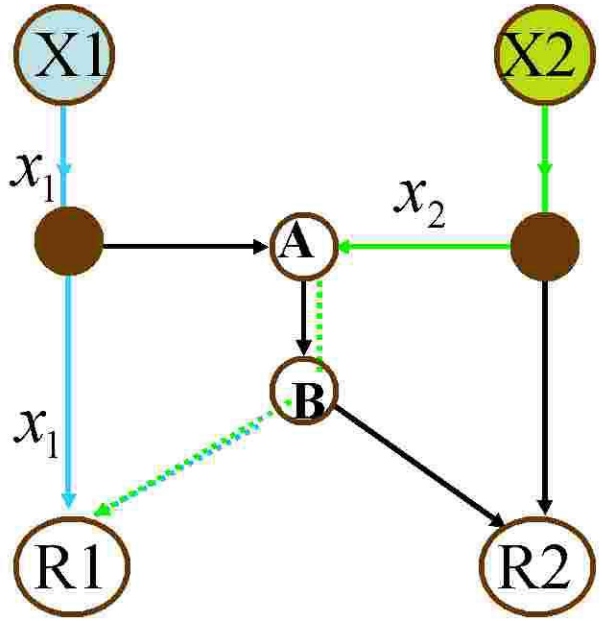


Receiver 1

Receiver 2

Example (2)

Ahlswede,Cai,Li,Yeung 2000

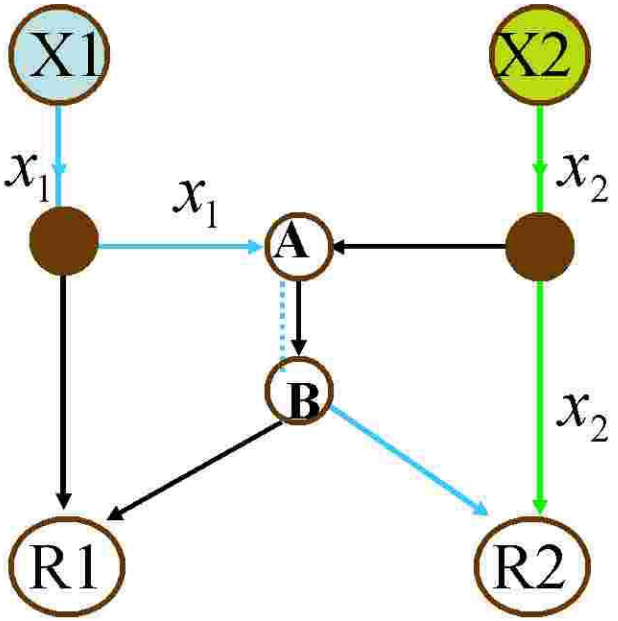


Receiver 1

Receiver 2

Example (3)

Ahlswede,Cai,Li,Yeung 2000

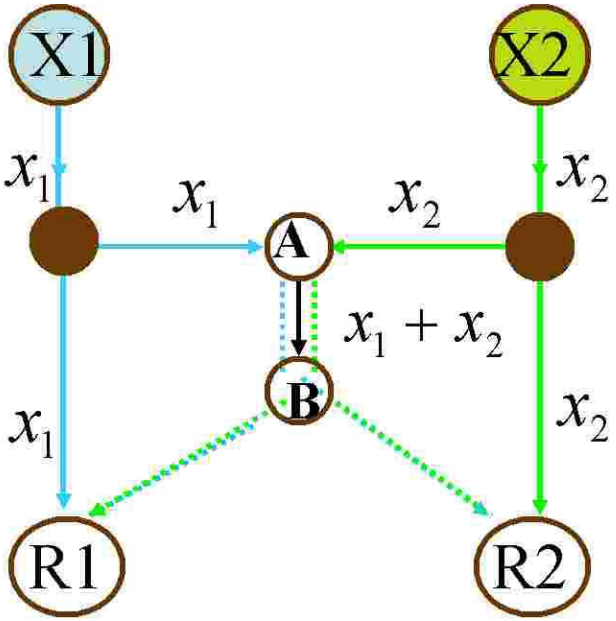


Receiver 1

Receiver 2

Example (4)

Ahlswede,Cai,Li,Yeung 2000

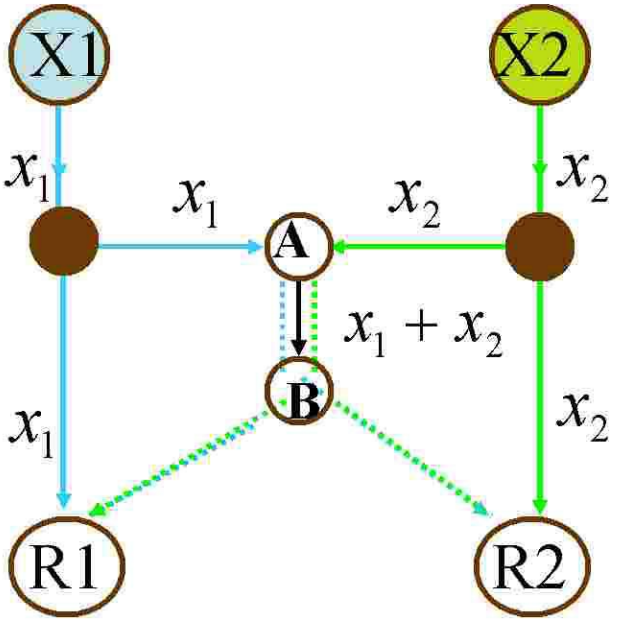


Receiver 1

Receiver 2

Example (5)

Ahlswede,Cai,Li,Yeung 2000



$\{x_1, x_1 + x_2\}$ $\{x_2, x_1 + x_2\}$

Receiver 1 *Receiver 2*

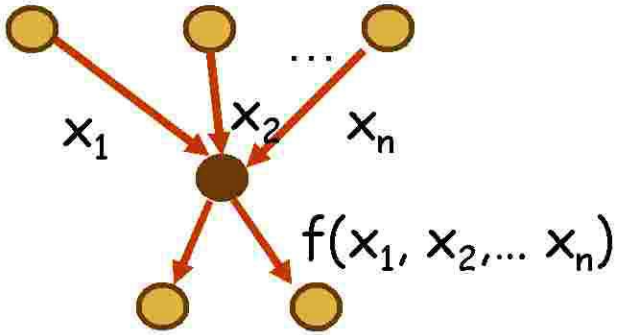
When is Network Coding useful?

- Practical Applications today:
  1. Wireless Multi-hop Networks
     - Throughput benefits [Katabi, Sigcomm 06]
  2. Content Distribution in P2P networks
     - Coupon collector problem [Avalanche '05-'06]

- Potential benefits at the cost of processing at intermediate nodes

Problem Statement

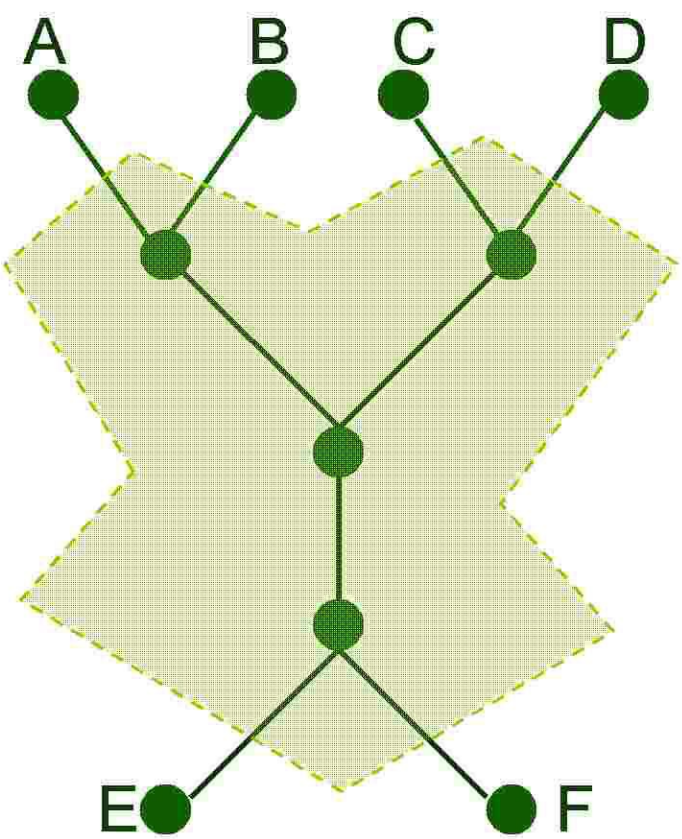
- Hypothesis: there will be networks in the near future that deploy Network Coding (NC).
- Question: Can we exploit NC to improve other operations? E.g. tomography?
- Answer: Tomography turns out to be easier in networks with Network Coding.
- Insight: NC introduces topology-dependent correlation, which can be exploited for inference.



Outline

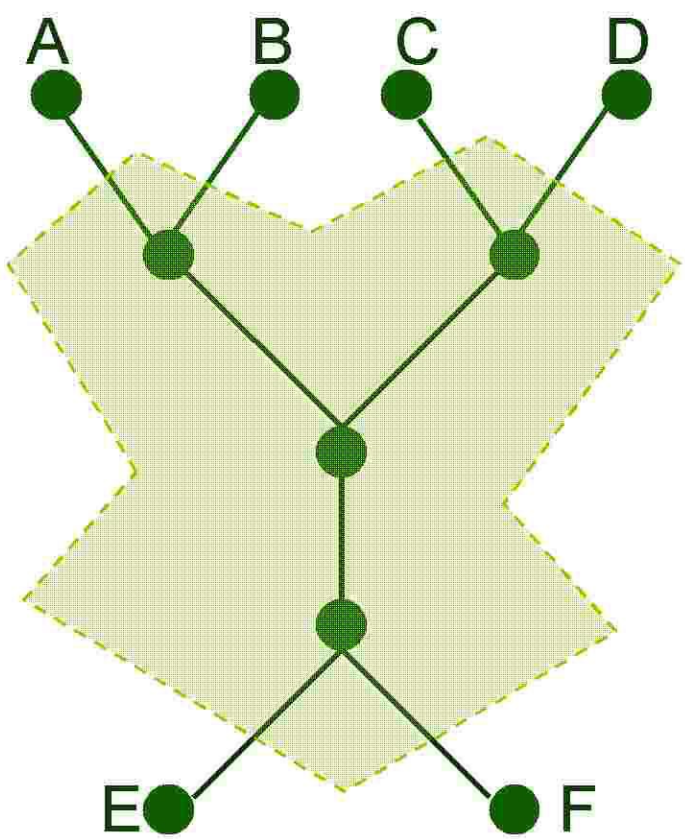
- Background
  - Network tomography
  - Network coding
- Topology Inference using Network Coding
- Link Loss Inference using Network Coding
- Conclusions

Topology Inference
Problem statement



Infer tree topology from measurements at end nodes.

Topology Inference
Problem statement

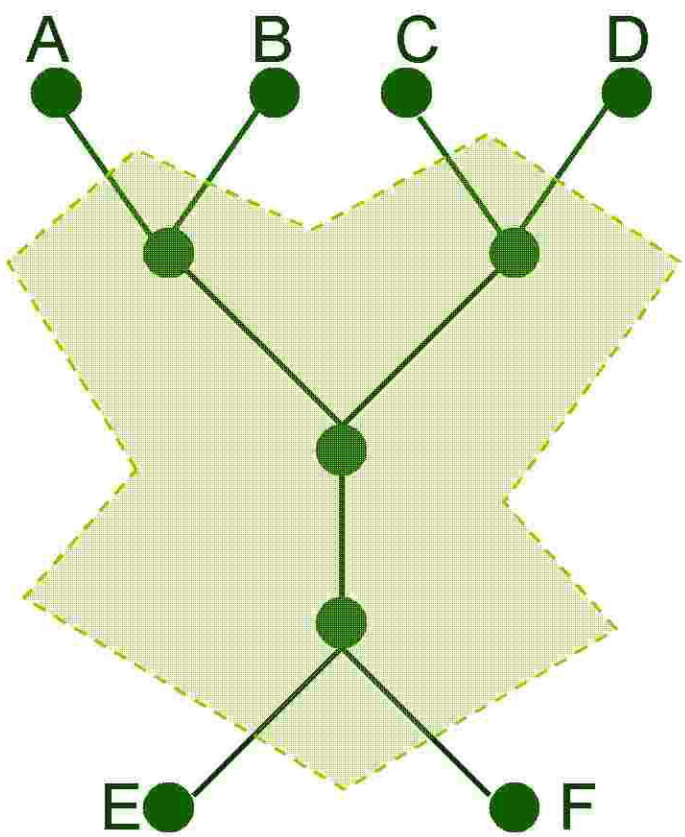


- Infer tree topology from measurements at end-nodes.

- *S. Ratnasamy and S. McCanne, 1999*
- *N.G. Duffield, J. Horowitz, F. Lo Presti, and D. Towsley, 2002*
- *Coates et al. 2000*
- *Byers et al. 2002*
- *Castro et al., 2004*

....

Traditional Approach (1)
Hierarchical Clustering

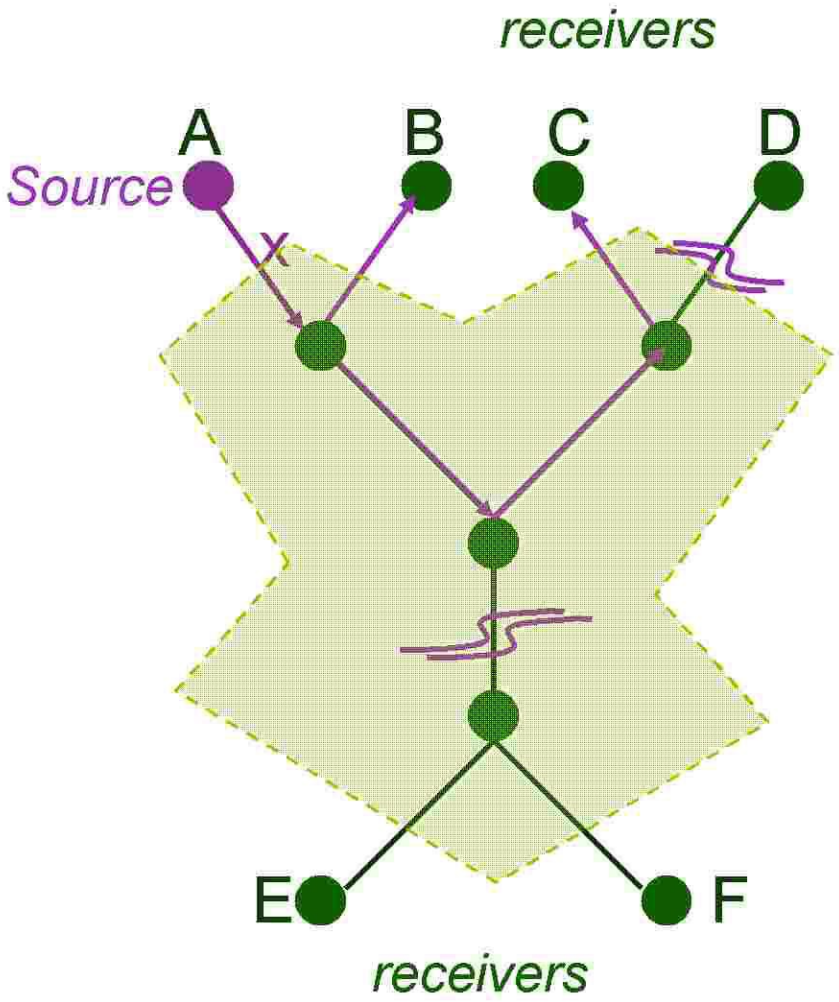


- Take advantage of a "monotonic" property, such as link loss or delay, to cluster together end-nodes

- *S. Ratnasamy and S. McCanne, 1999*
- *N.G. Duffield, J. Horowitz, F. Lo Presti, and D. Towsley, 2002*
- *Coates et al. 2000*
- *Byers et al. 2002*
- *Castro et al., 2004*

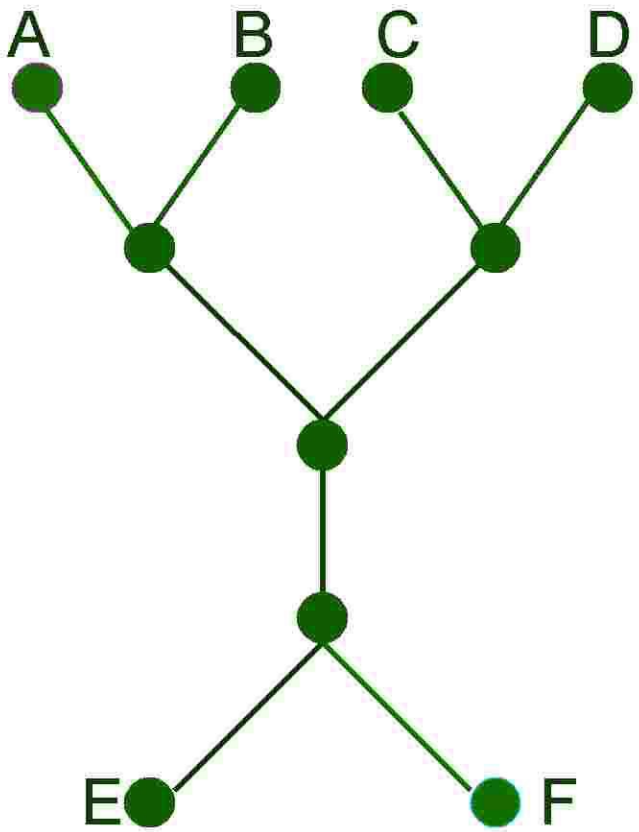
....

Traditional Approach (2)
Hierarchical Clustering



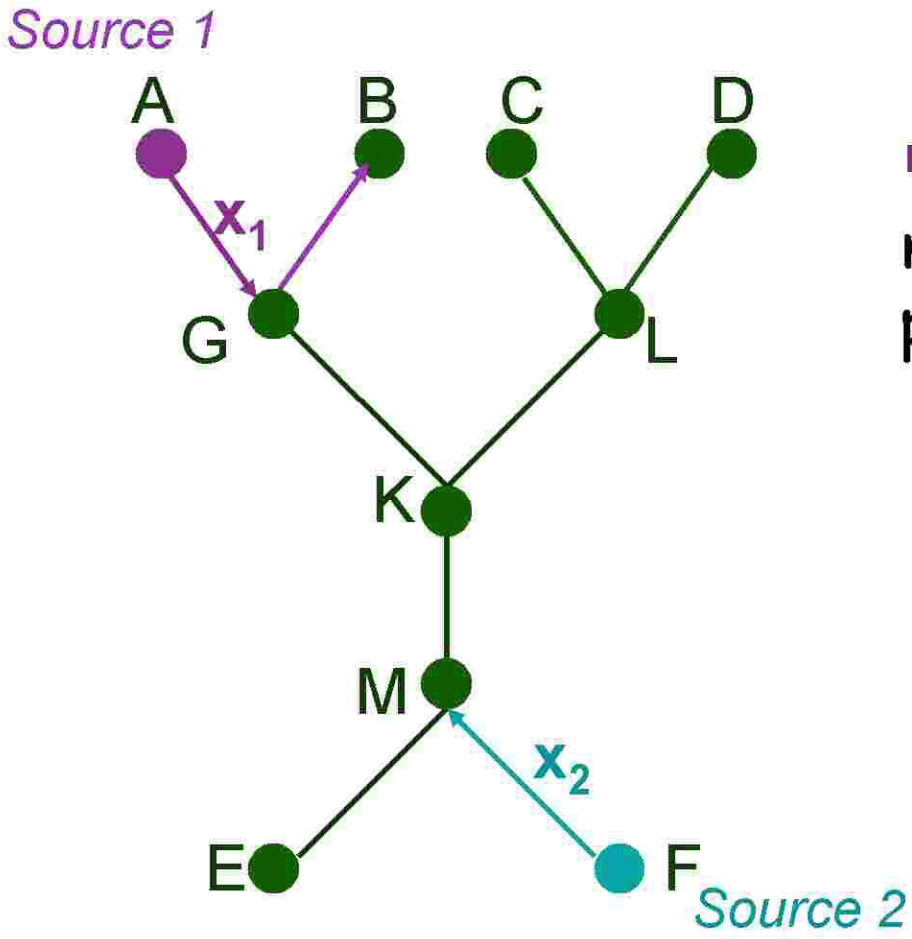
- Send n probes
- Observe correlation at receivers, in terms of a "monotonic" property, e.g. link loss or delay variation
- Cluster together receivers that see correlated patterns

Network Coding Approach



- Consider
  - a binary tree w/o loss
- Leaves
  - can act as sources or receivers of probes
- Intermediate node:
  - Within a window w
  - If it receives one packet x1, it forwards it
  - If it receives two packets (x1,x2), it forwards x1+x2

Network Coding Approach



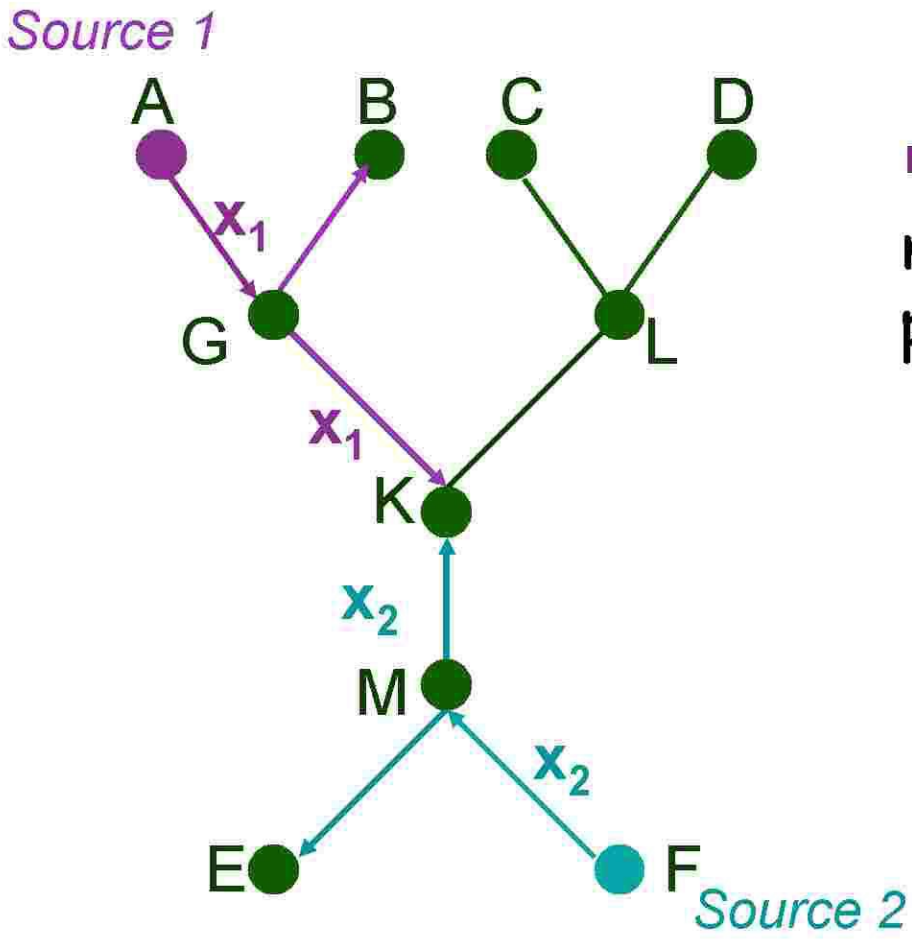
Our Algorithm

- Randomly pick two leaf nodes to act as sources of probe packets:

x_1=(1 0)

x_2=(0 1)

Network Coding Approach



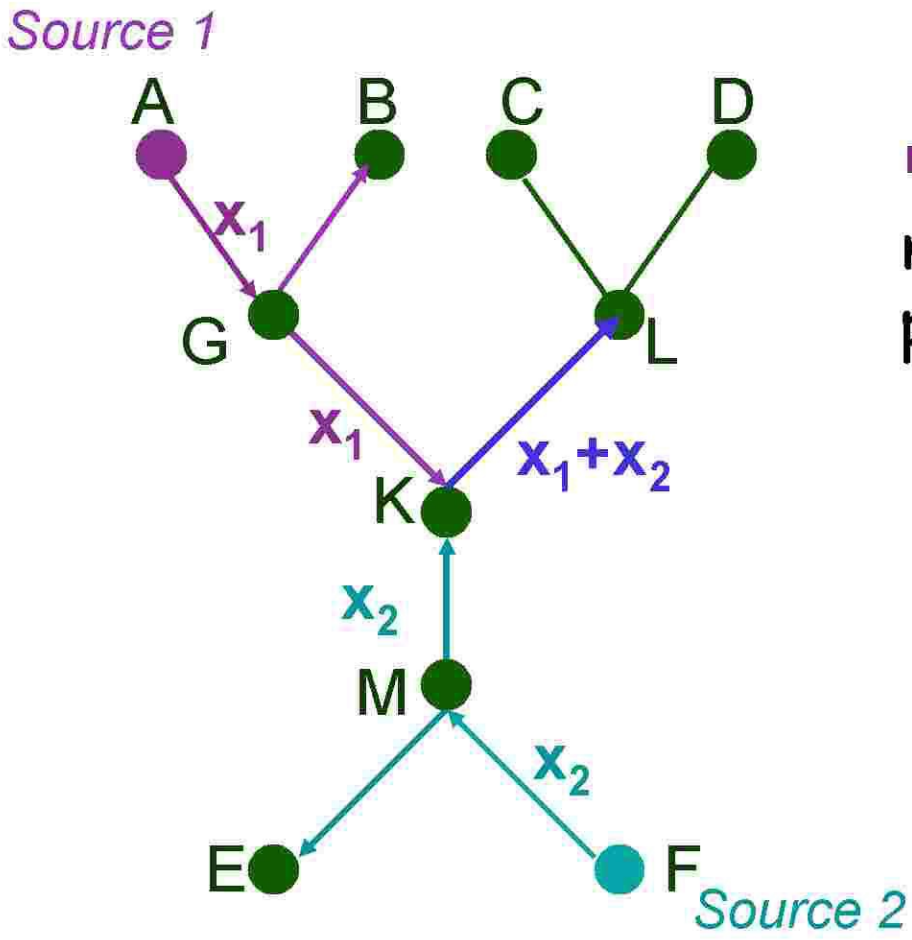
Algorithm

- Randomly pick two leaf nodes to act as sources of probe packets:

x_1=(1 0)

x_2=(0 1)

Network Coding Approach



Algorithm

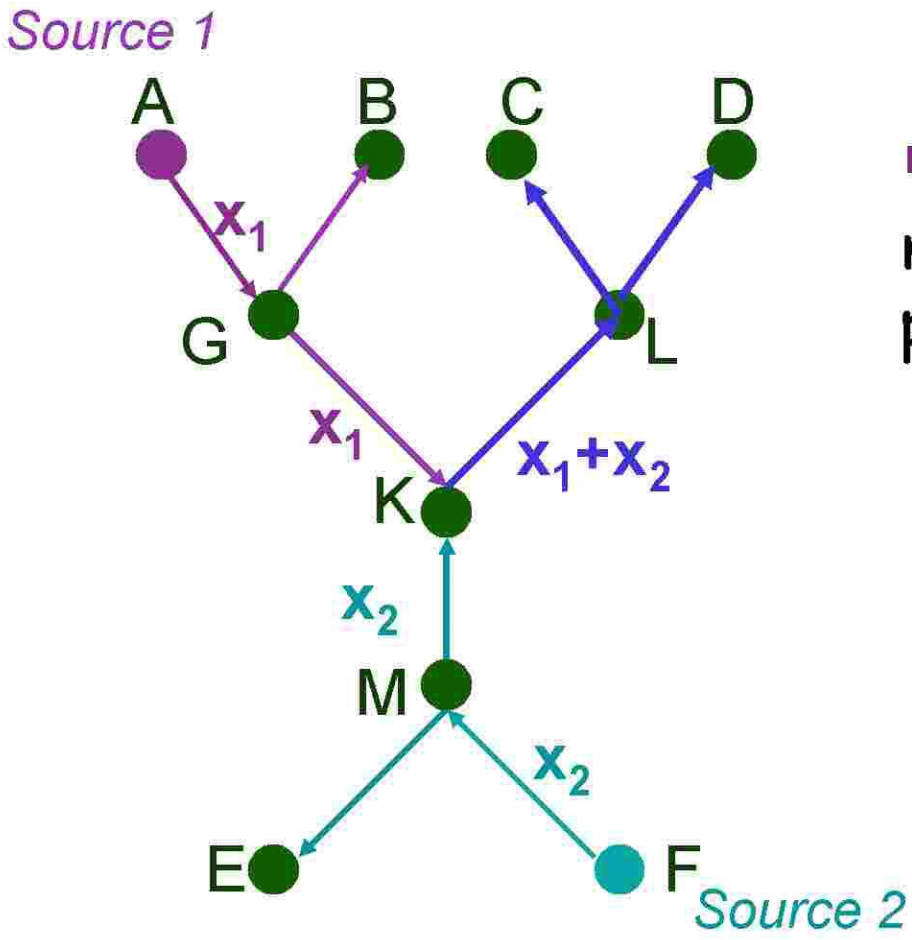
- Randomly pick two leaf nodes to act as sources of probe packets:

$x_1=(1\ 0)$

$x_2=(0\ 1)$

$x_3=x_1+x_2=(1\ 1)$

Network Coding Approach



Algorithm

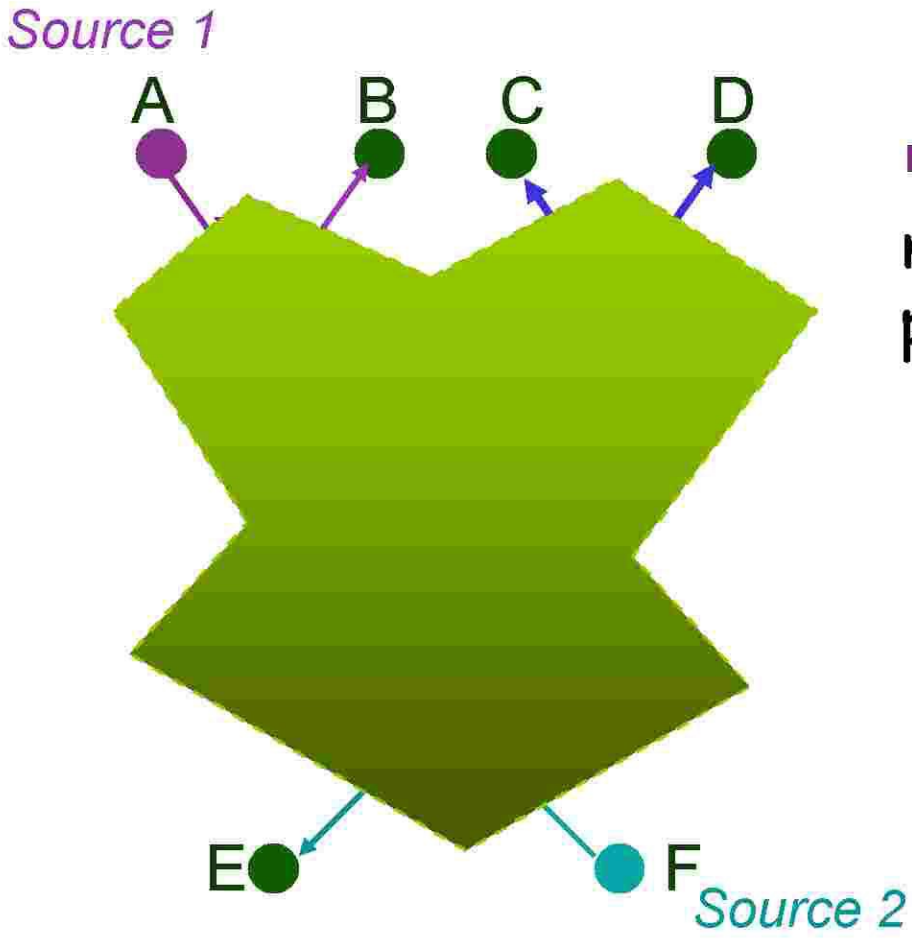
- Randomly pick two leaf nodes to act as sources of probe packets:

$x_1=(1\ 0)$

$x_2=(0\ 1)$

$x_3=x_1+x_2=(1\ 1)$

Network Coding Approach



Algorithm

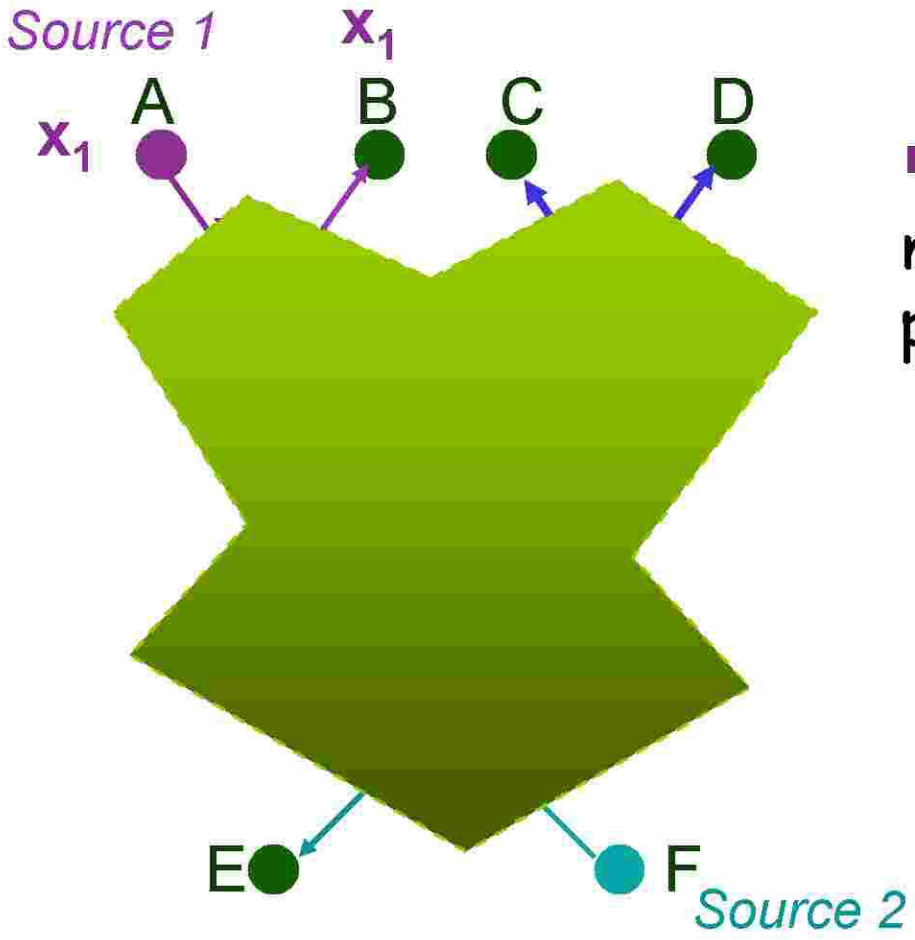
- Randomly pick two leaf nodes to act as sources of probe packets:

$x_1 = (1\ 0)$

$x_2 = (0\ 1)$

$x_3 = x_1 + x_2 = (1\ 1)$

Network Coding Approach



Algorithm

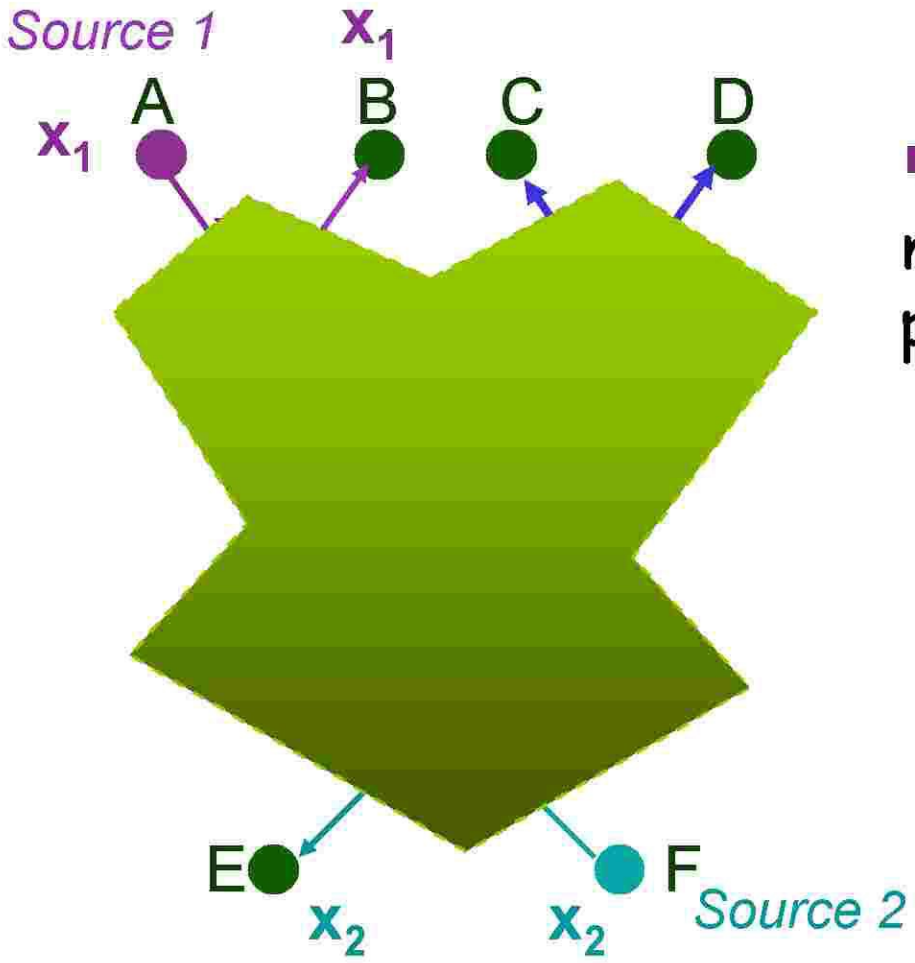
- Randomly pick two leaf nodes to act as sources of probe packets:

$x_1=(1\ 0)$

$x_2=(0\ 1)$

$x_3=x_1+x_2=(1\ 1)$

Network Coding Approach



Algorithm

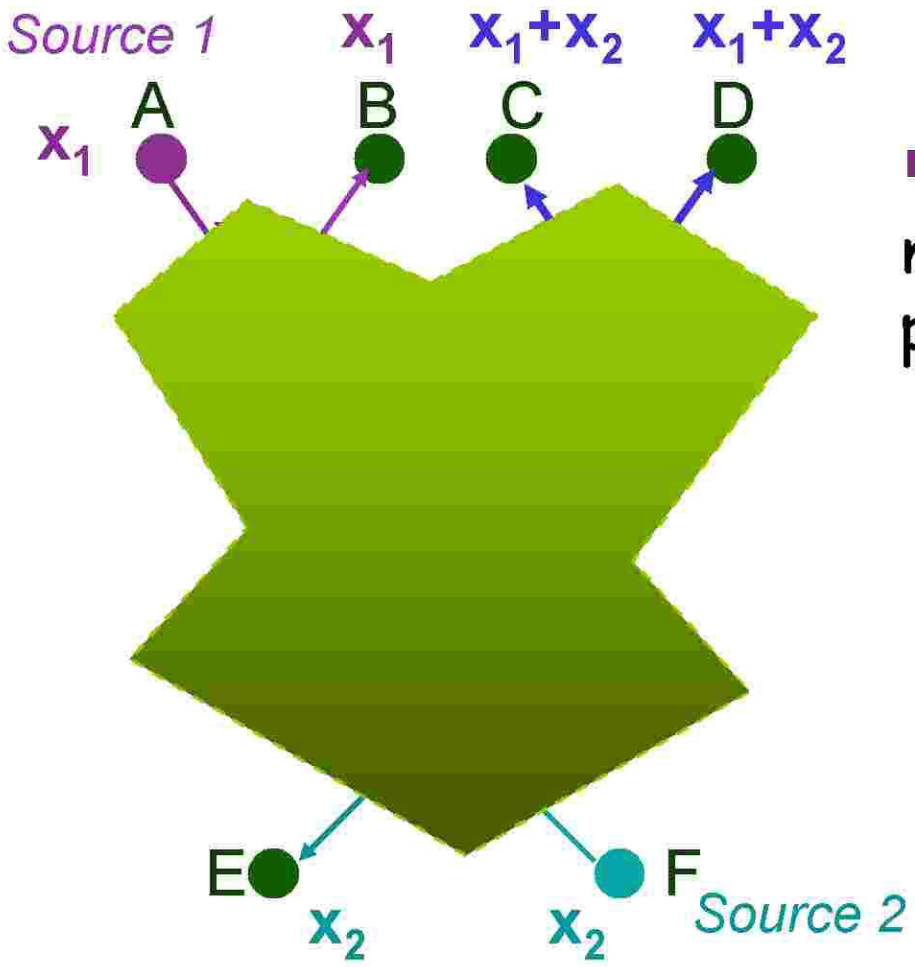
- Randomly pick two leaf nodes to act as sources of probe packets:

$x_1=(1\ 0)$

$x_2=(0\ 1)$

$x_3=x_1+x_2=(1\ 1)$

Network Coding Approach



Algorithm

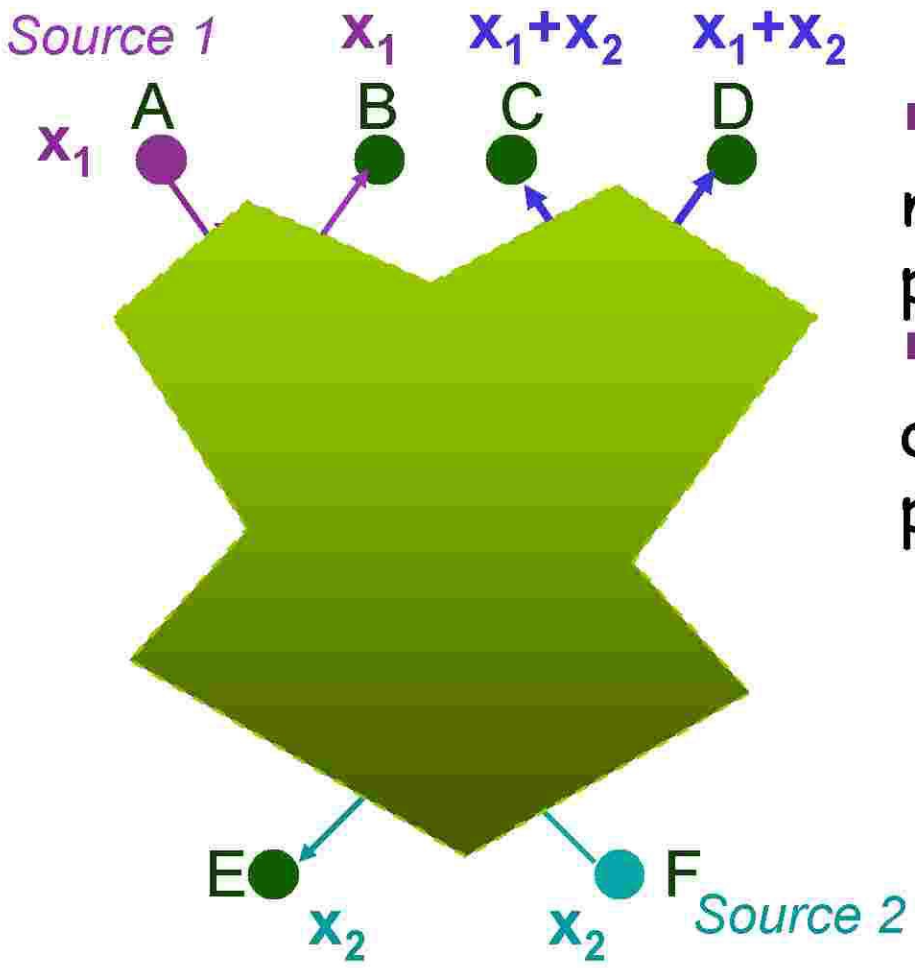
- Randomly pick two leaf nodes to act as sources of probe packets:

$x_1=(1\ 0)$

$x_2=(0\ 1)$

$x_3=x_1+x_2=(1\ 1)$

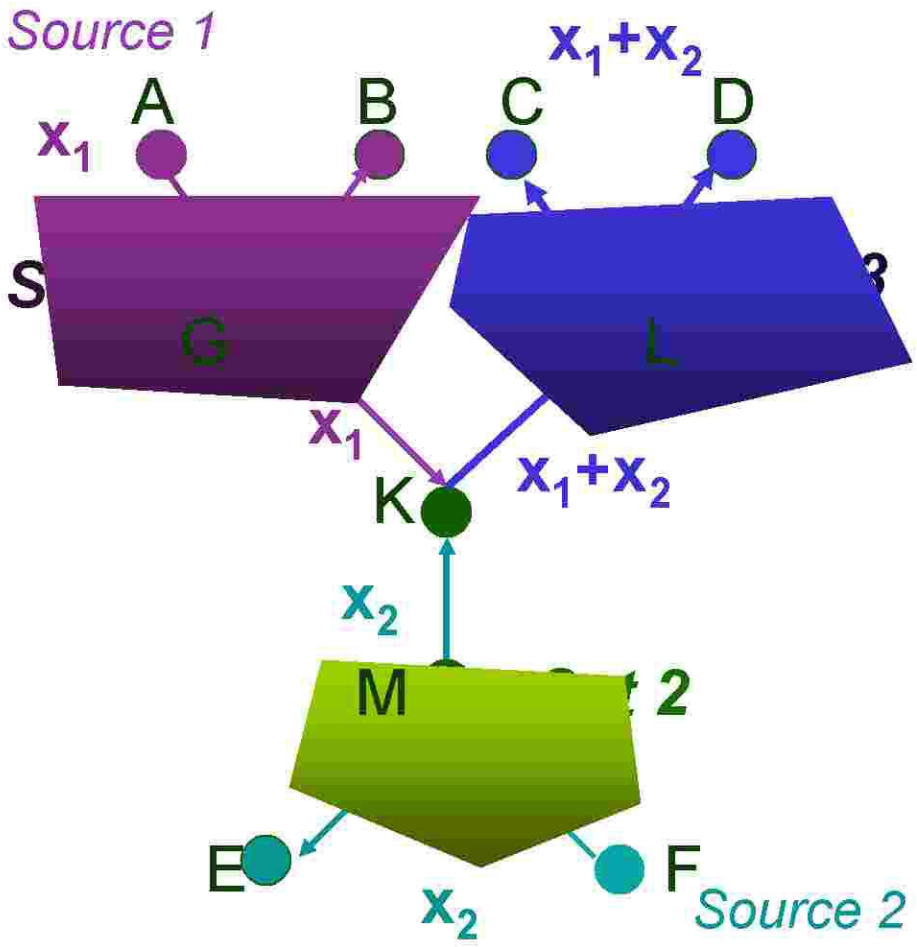
Network Coding Approach



Algorithm

- Randomly pick two leaf nodes to act as sources of probe packets
- Group receivers in 3 sets, depending on the received packet: x1, x2 or x1+x2

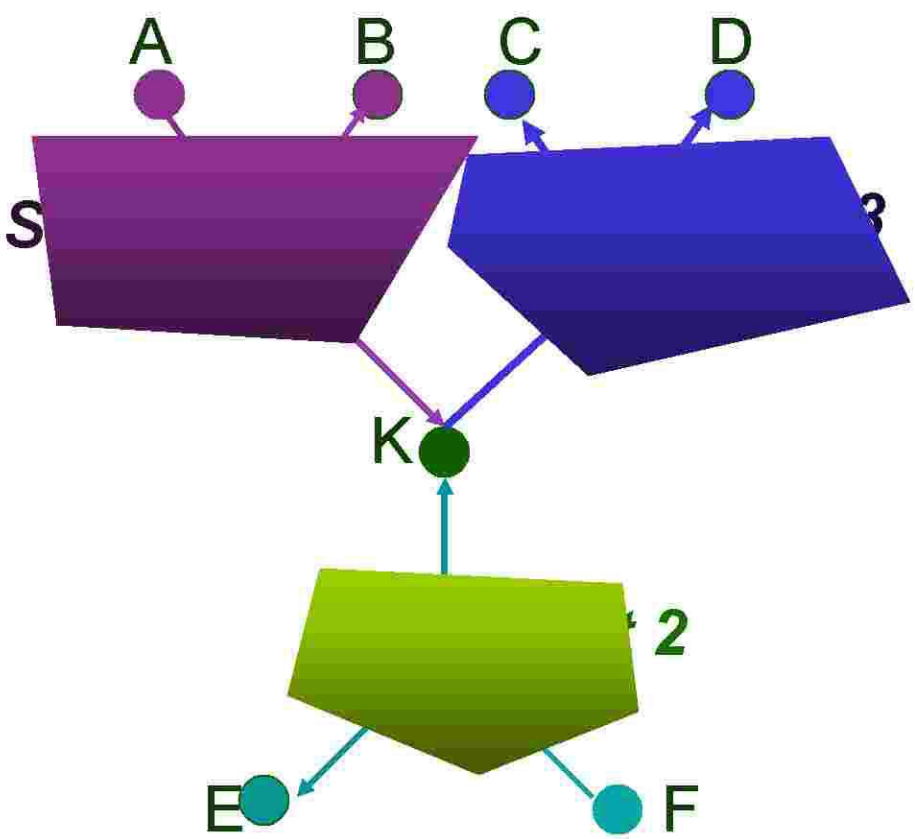
Network Coding Approach



Algorithm

- Randomly pick two leaf nodes to act as sources of probe packets:
- Group receivers in 3 sets, depending on what packet they received: {x_1, x_2, x_1+x_2}
- Reveal three inner edges: KG, KL, KM.

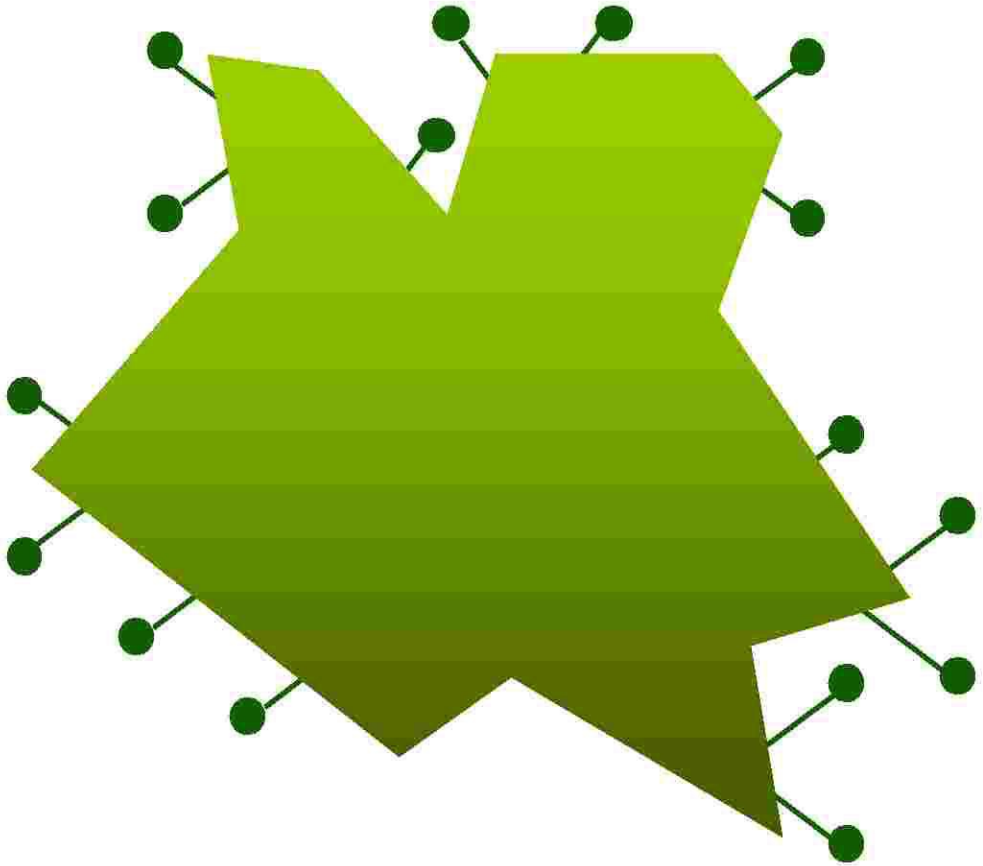
Network Coding Approach



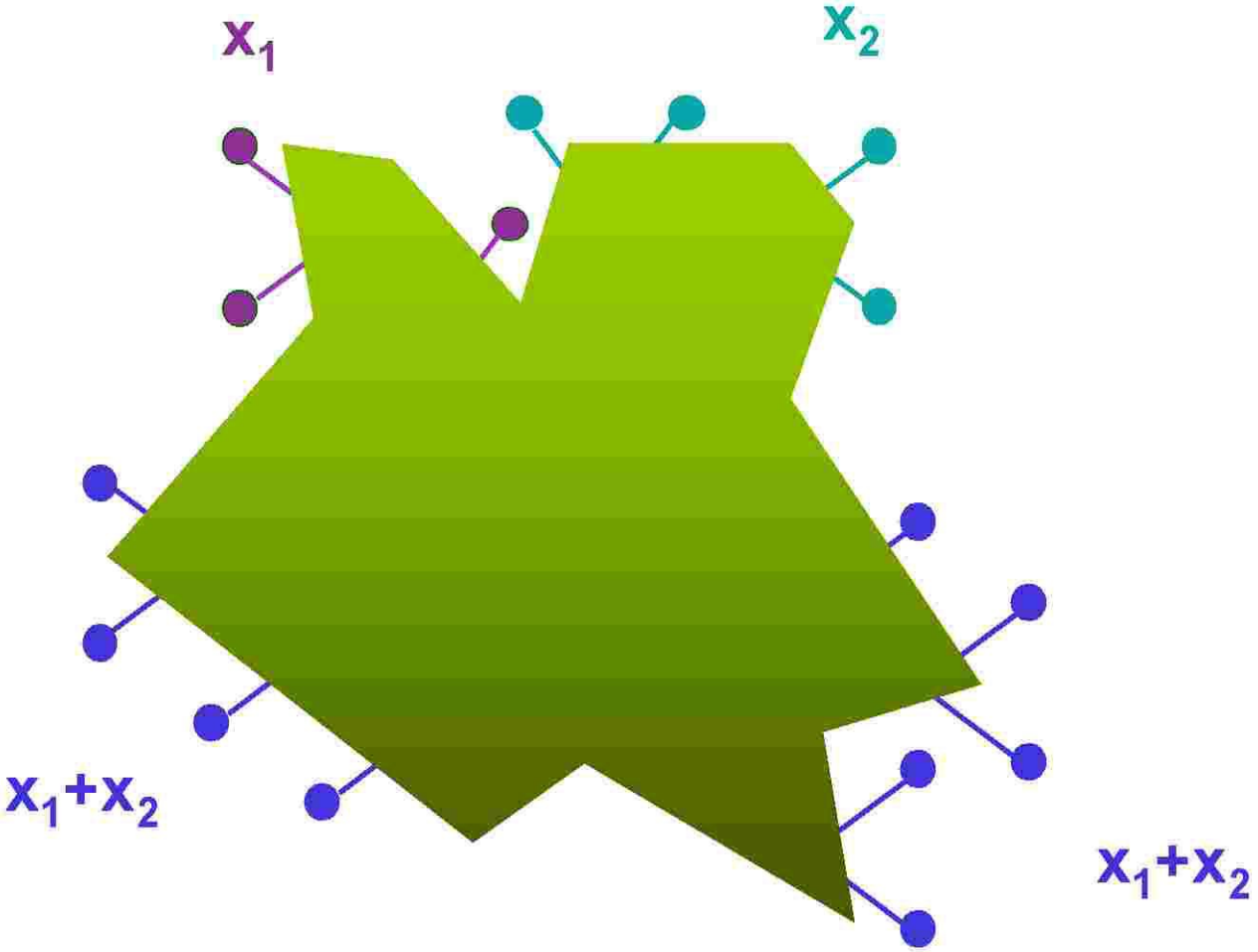
Algorithm

- …
- .continue recursively
  - (revealed nodes act as aggregate receivers)
- until all edges revealed
  - (<=2 leaves in each set)
- number of steps
  - at most number of edges

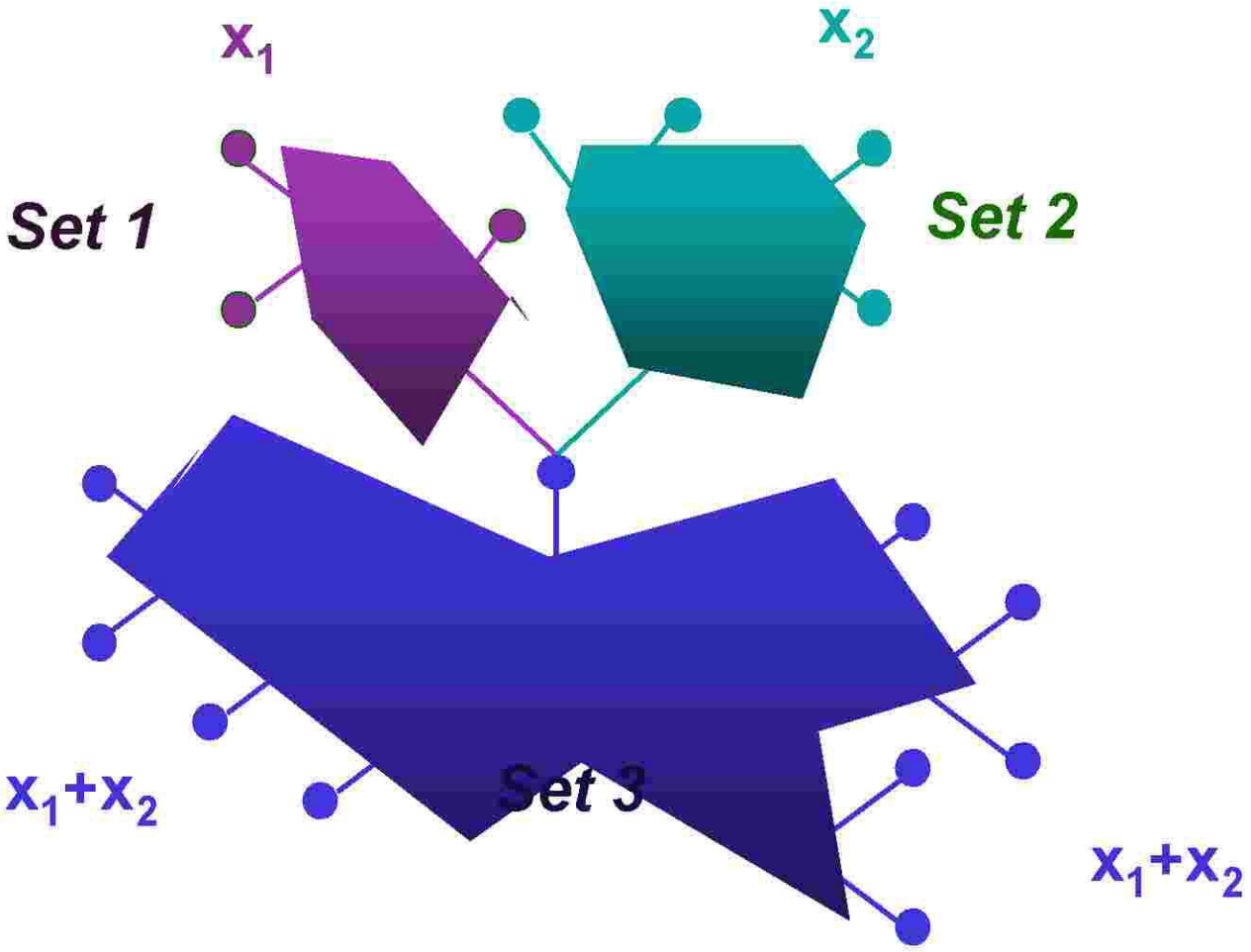
Larger example



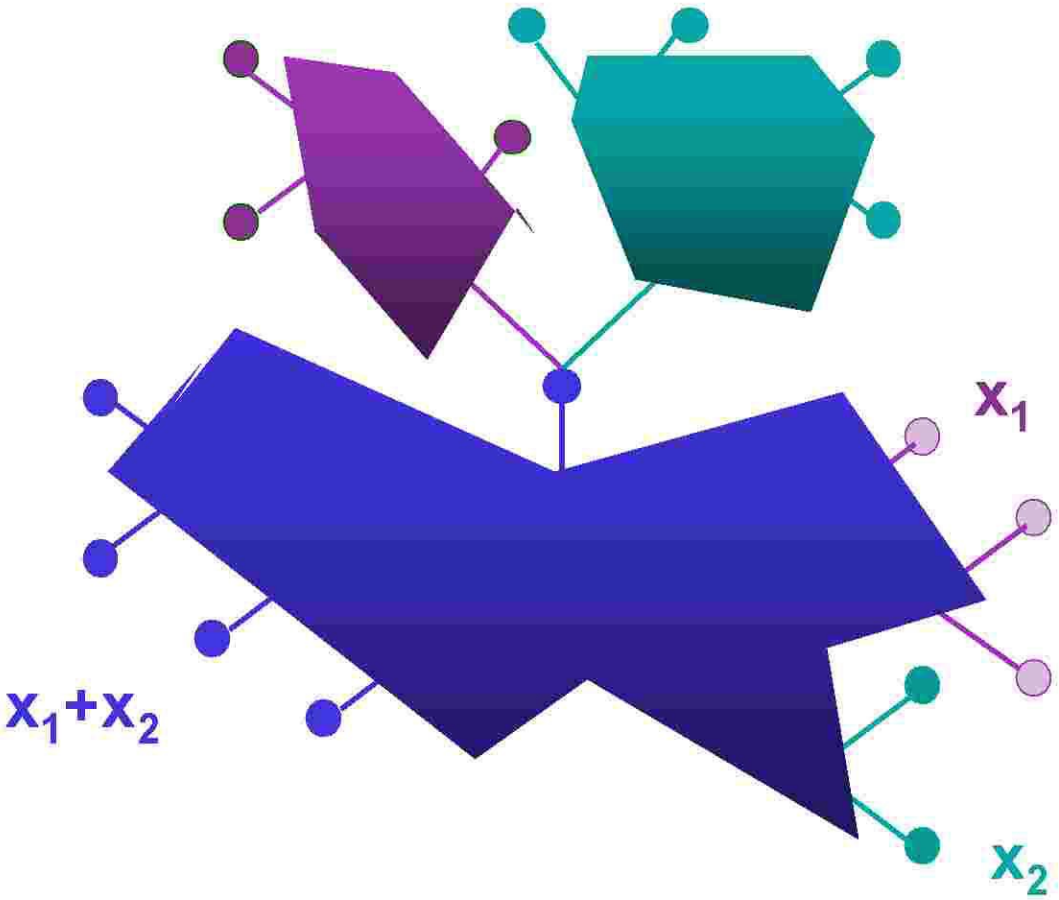
Larger example



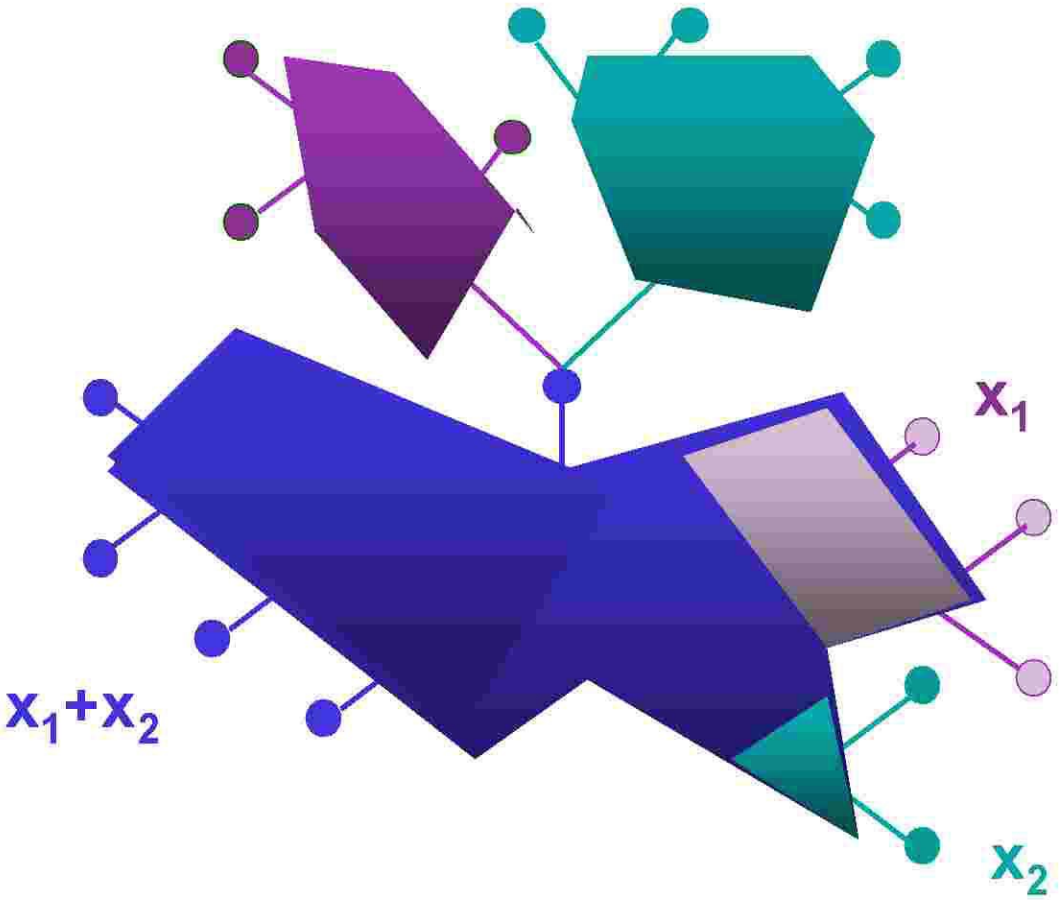
First iteration



Second iteration

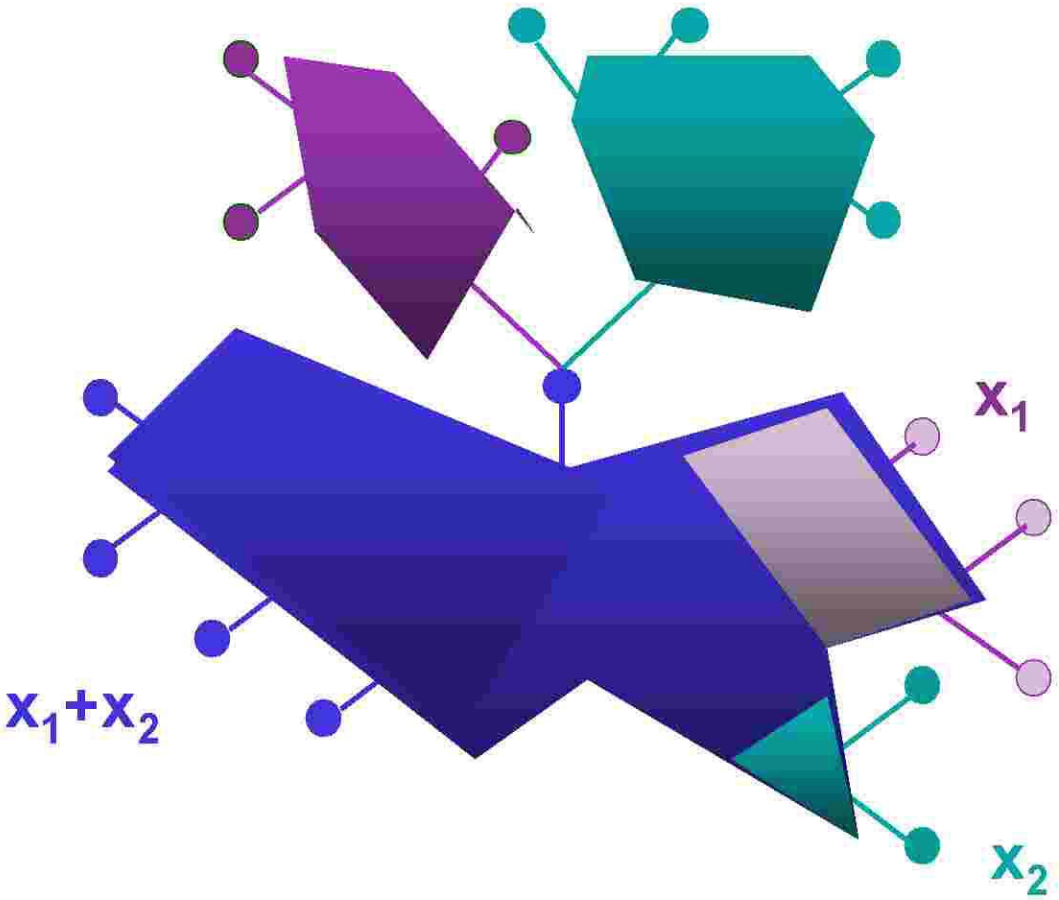


Second iteration



Deterministic inference in <n steps

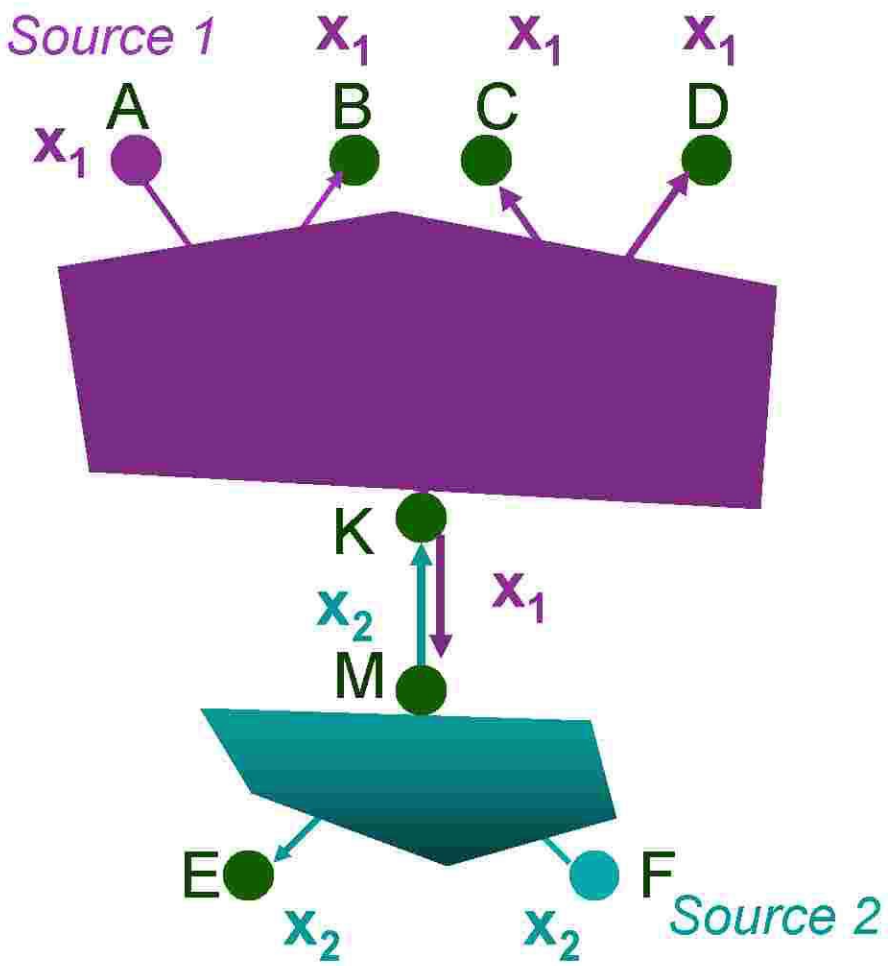
(in trees without loss)



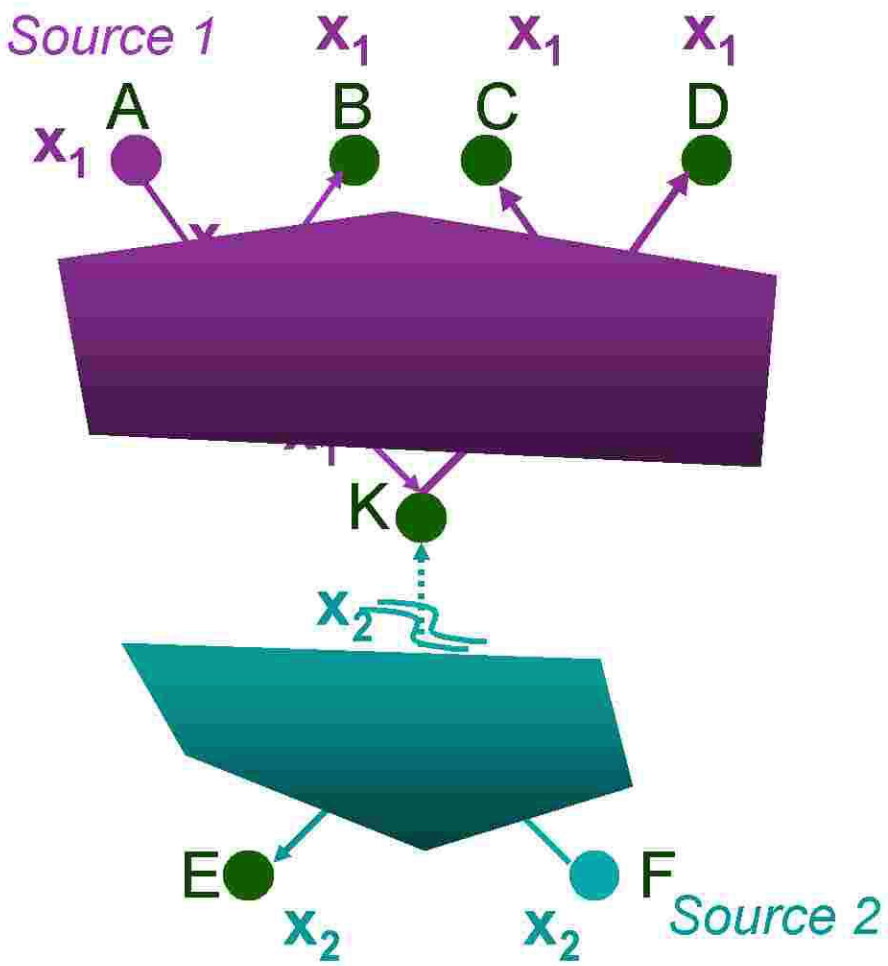
Some technicalities…

- Link delays and selection of sources
- Time window
- Trees with larger degree

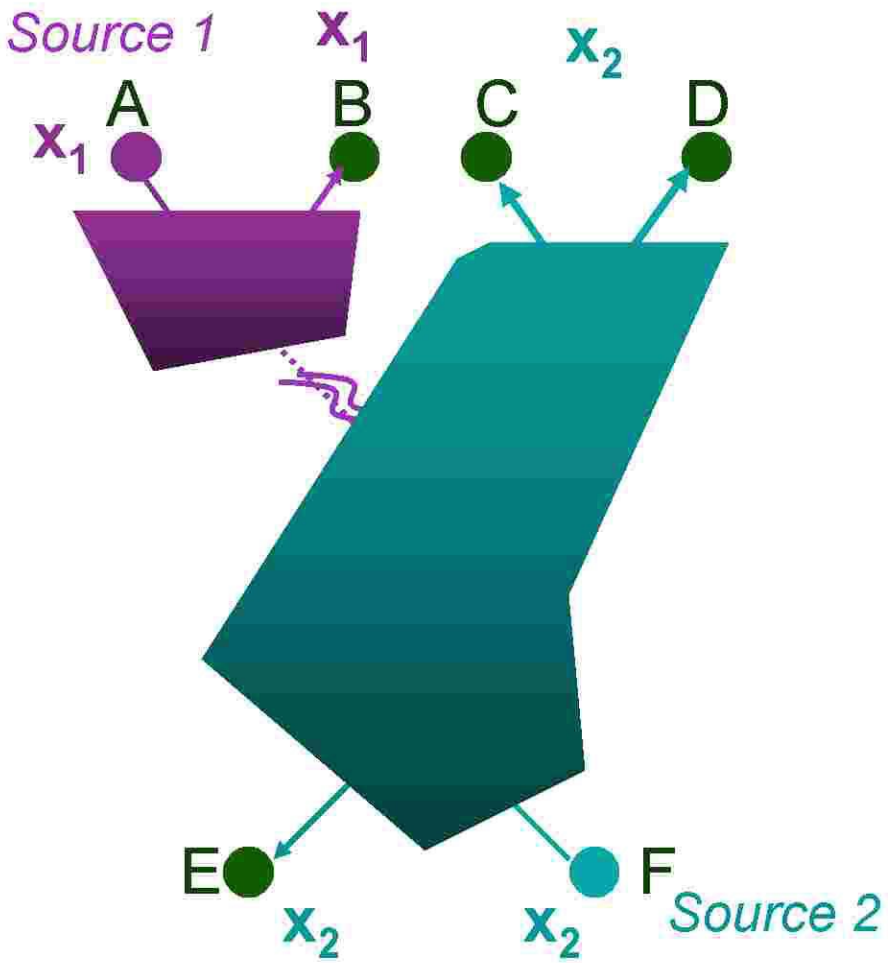
What if two probes cross each other?



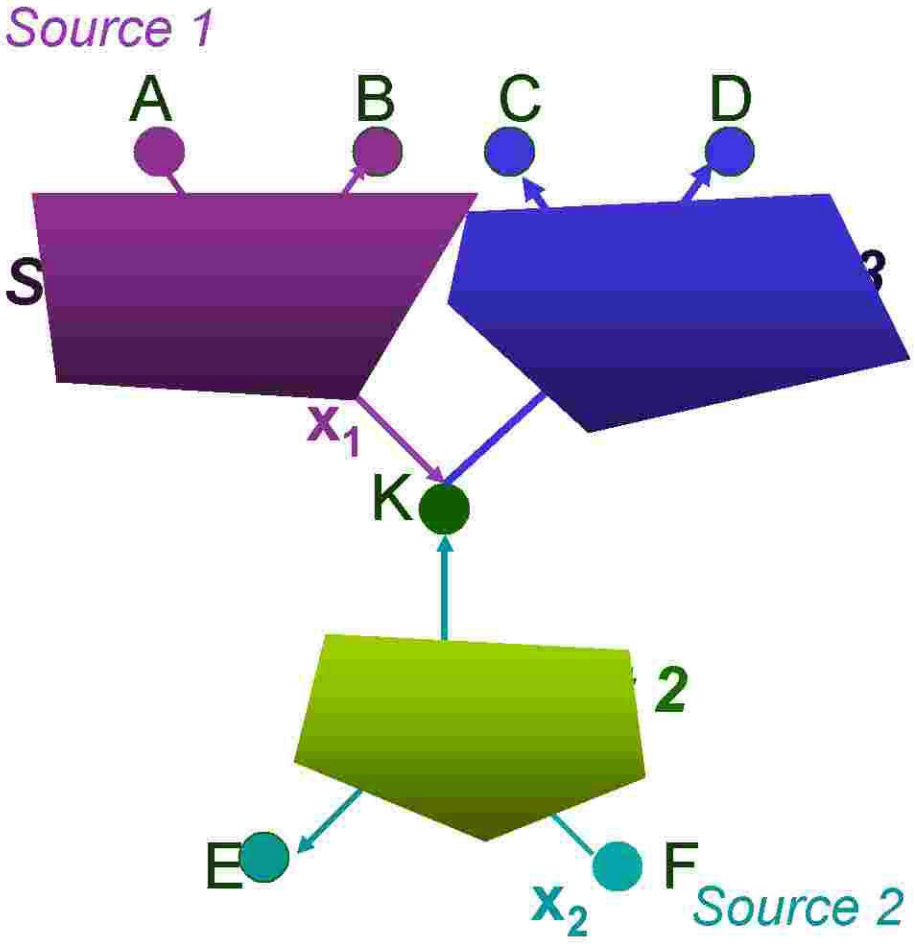
Packet losses cause ambiguity



Packet losses cause ambiguity

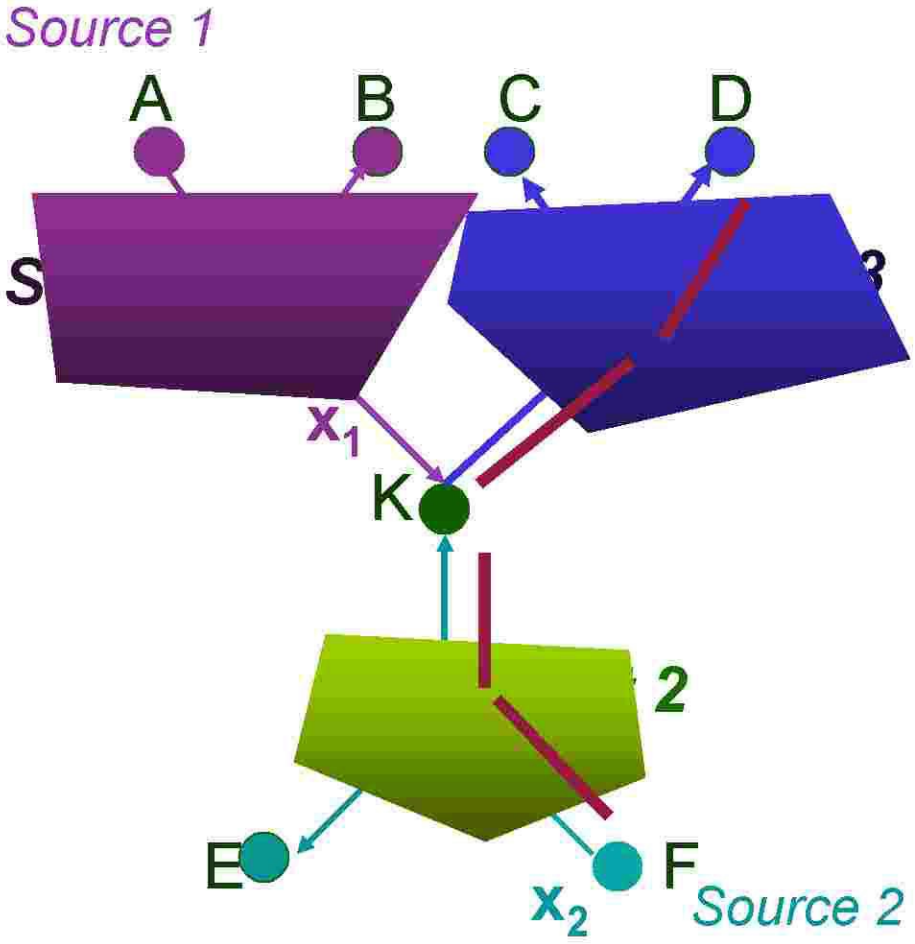


Resilience to packet loss



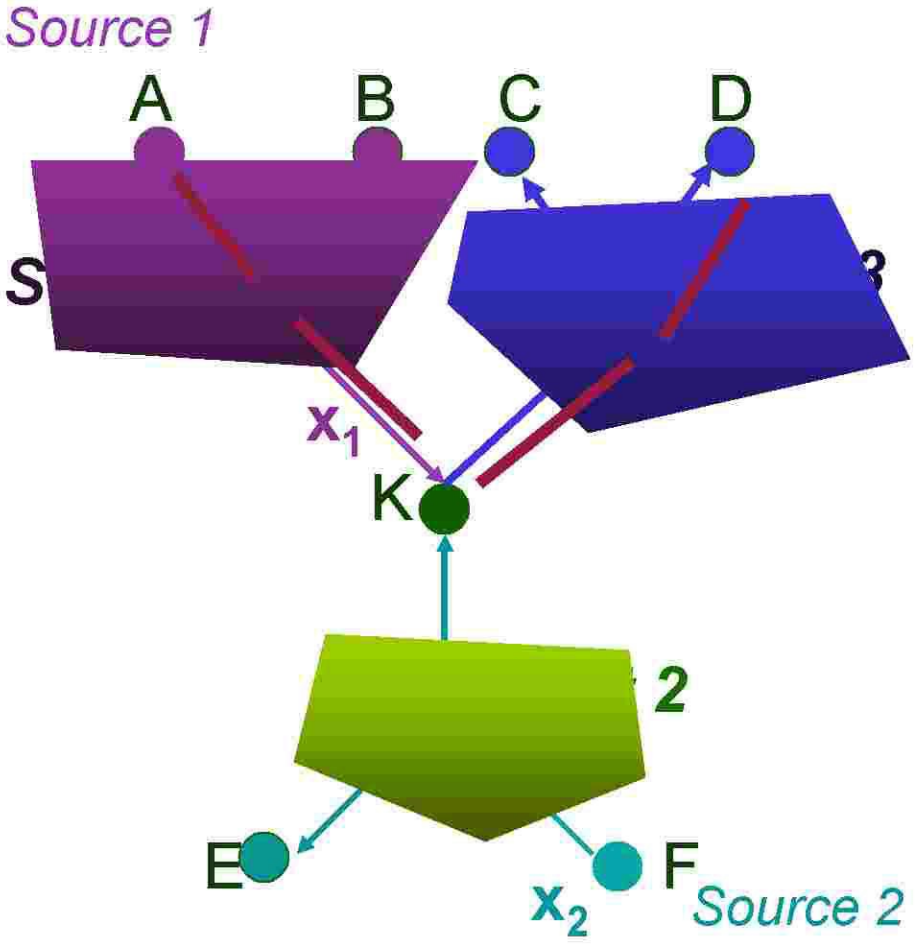
- Send multiple probe packets per iteration
- Classification of a receiver R
  - only x1's received → Set 1
  - only x2's received → Set 2
  - {x1+x2} , {x1's & x2's} →Set 3

Resilience to packet loss



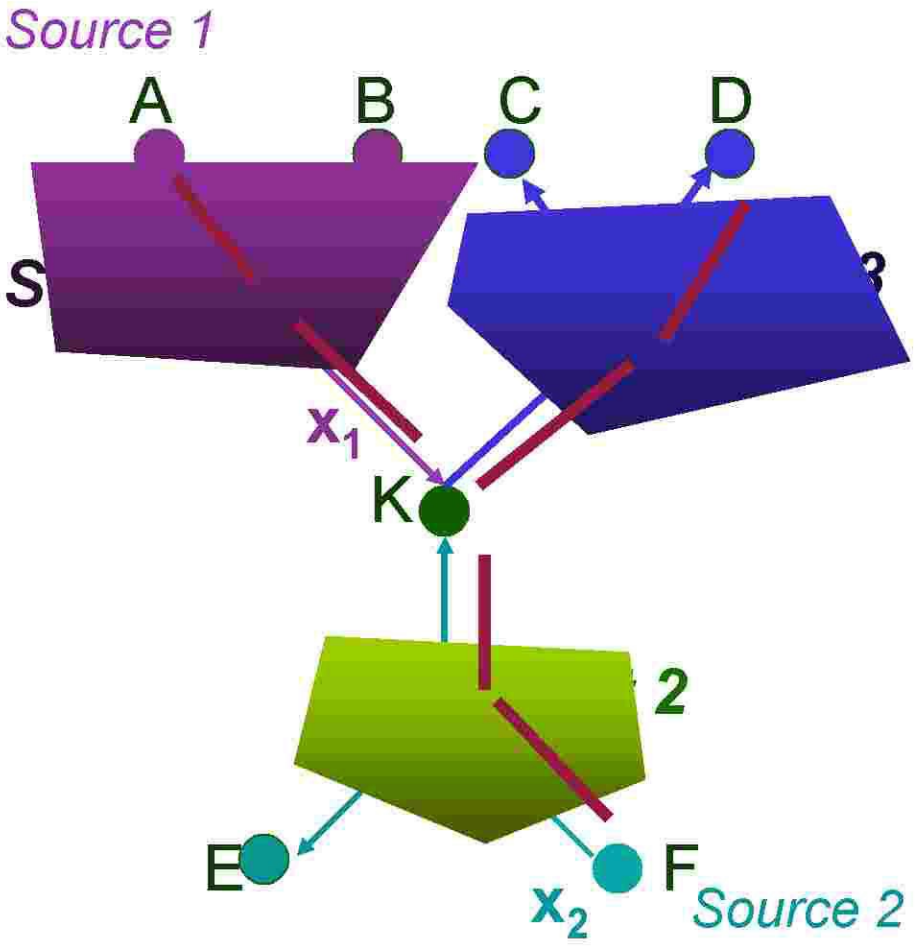
- Send multiple probe packets per iteration
- Classification of a receiver R
  - only x1's received → Set 1
  - only x2's received → Set 2
  - {x1+x2} , {x1's & x2's} →Set 3

Resilience to packet loss



- Send multiple probe packets per iteration
- Classification of a receiver R
  - only x1's received → Set 1
  - only x2's received → Set 2
  - {x1+x2} , {x1's & x2's} →Set 3

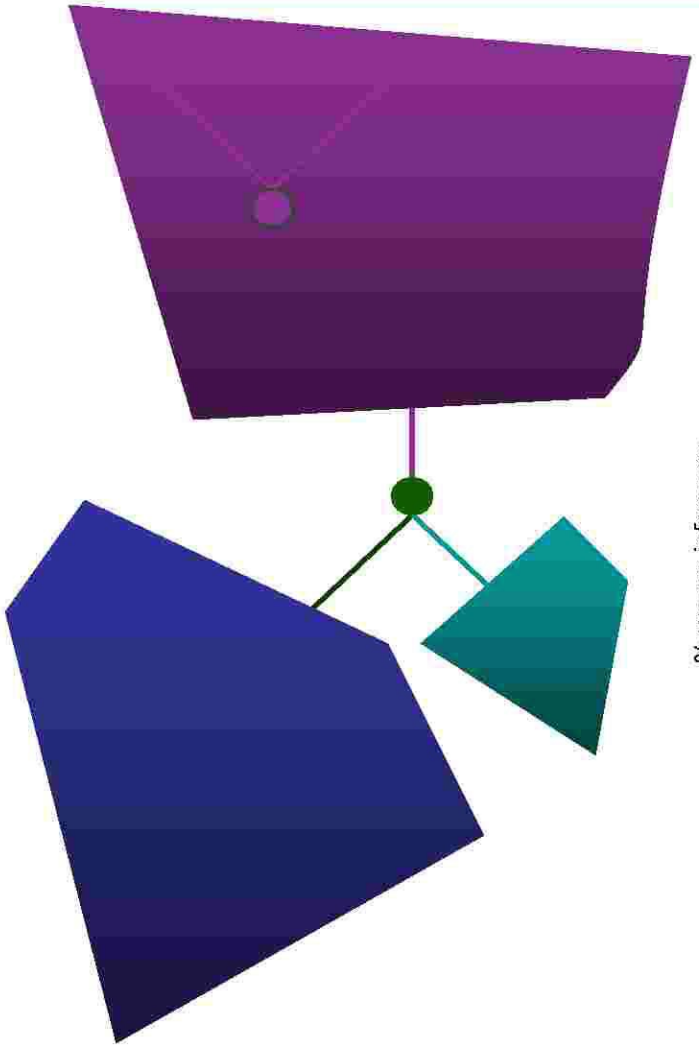
Resilience to packet loss

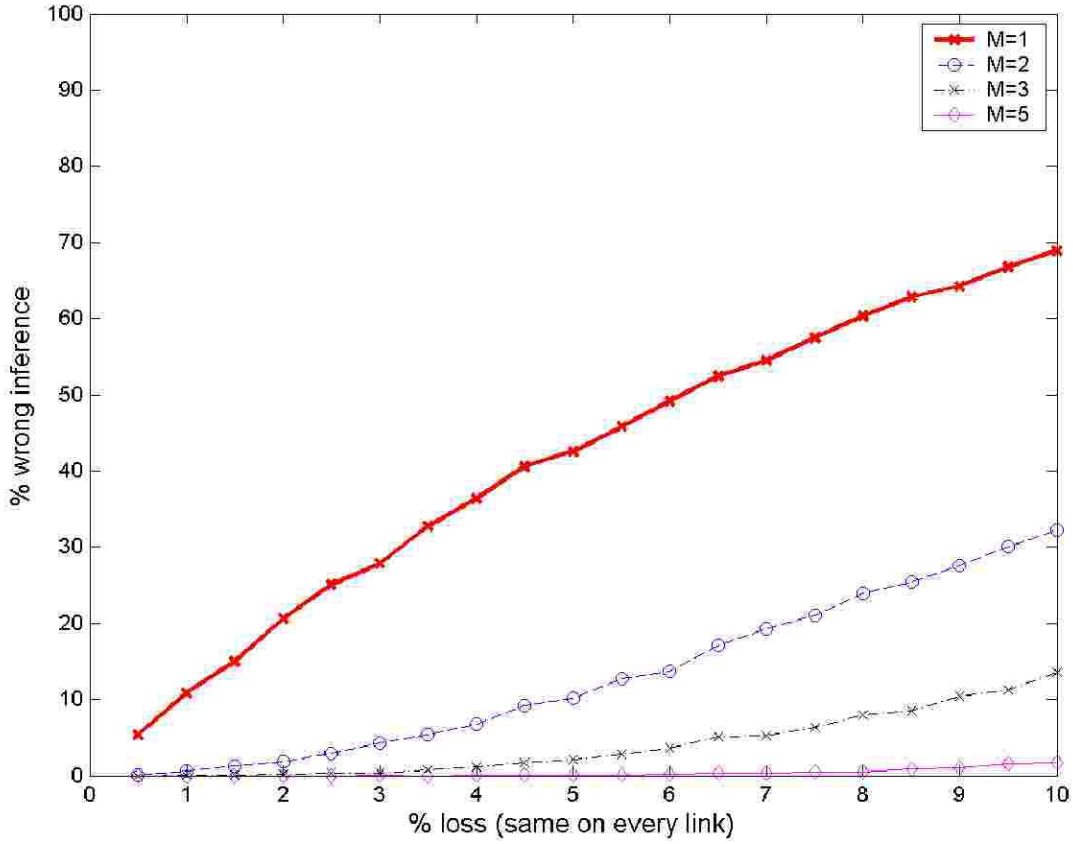


- Send multiple probe packets per iteration
- Classification of a receiver R
  - only x1's received → Set 1
  - only x2's received → Set 2
  - {x1+x2} , {x1's & x2's} →Set 3
- Sufficient for each source-receiver path to succeed **once**
  - faster than collecting statistics

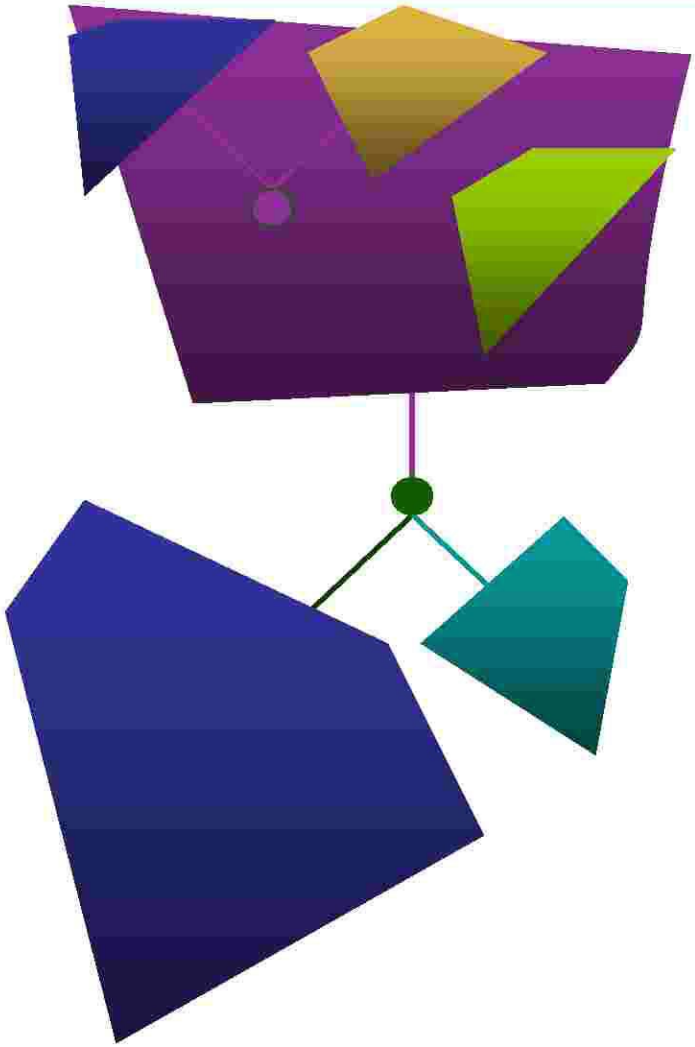
Preliminary Simulation Results

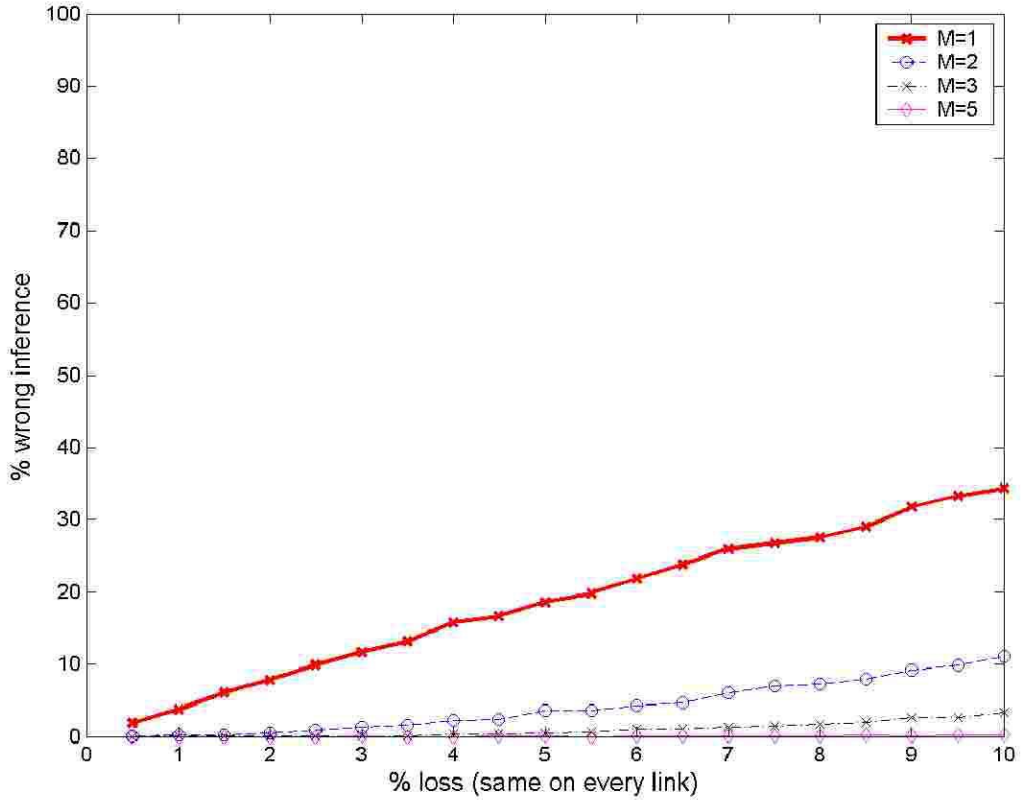
Iteration 1



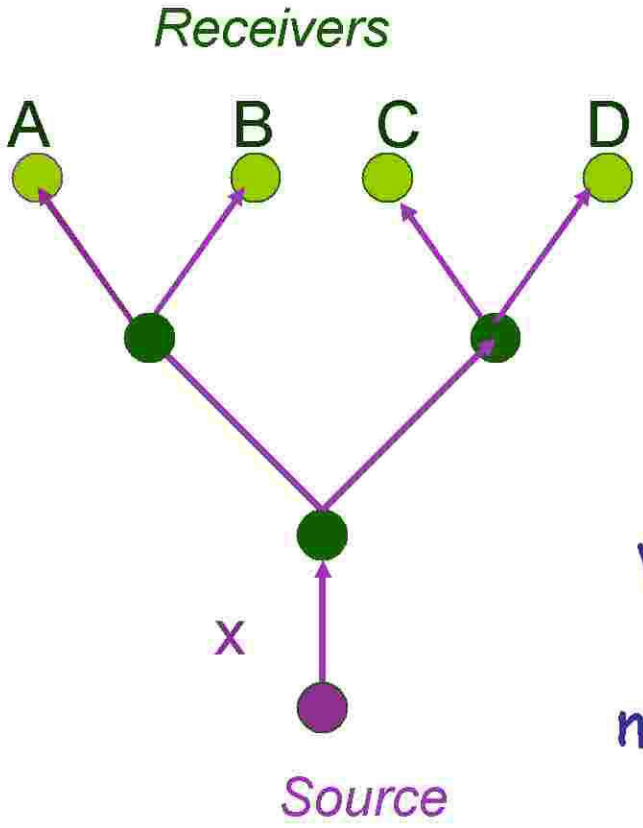


Preliminary Simulation Results
Iteration 2





The case for locality of operations

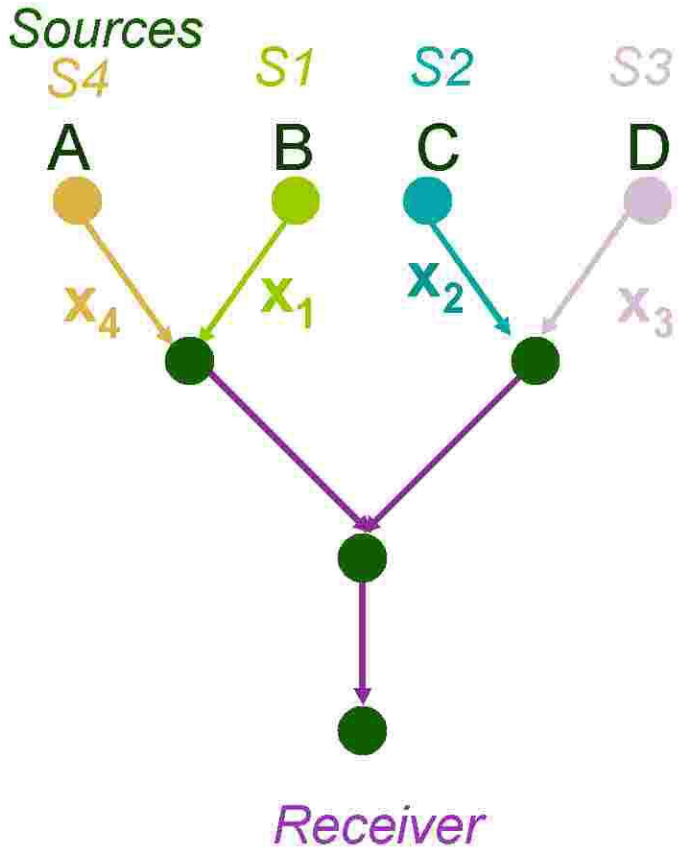


N. Duffield, J. Horowitz, F. Lo Presti, and D. Towsley, 2002:

"…collecting measurements from remote parts of the networks is a difficult task"

When there are many receivers, we still need to collect the measurements from end-nodes to a processing center

A special case of network coding: inverse multicast tree

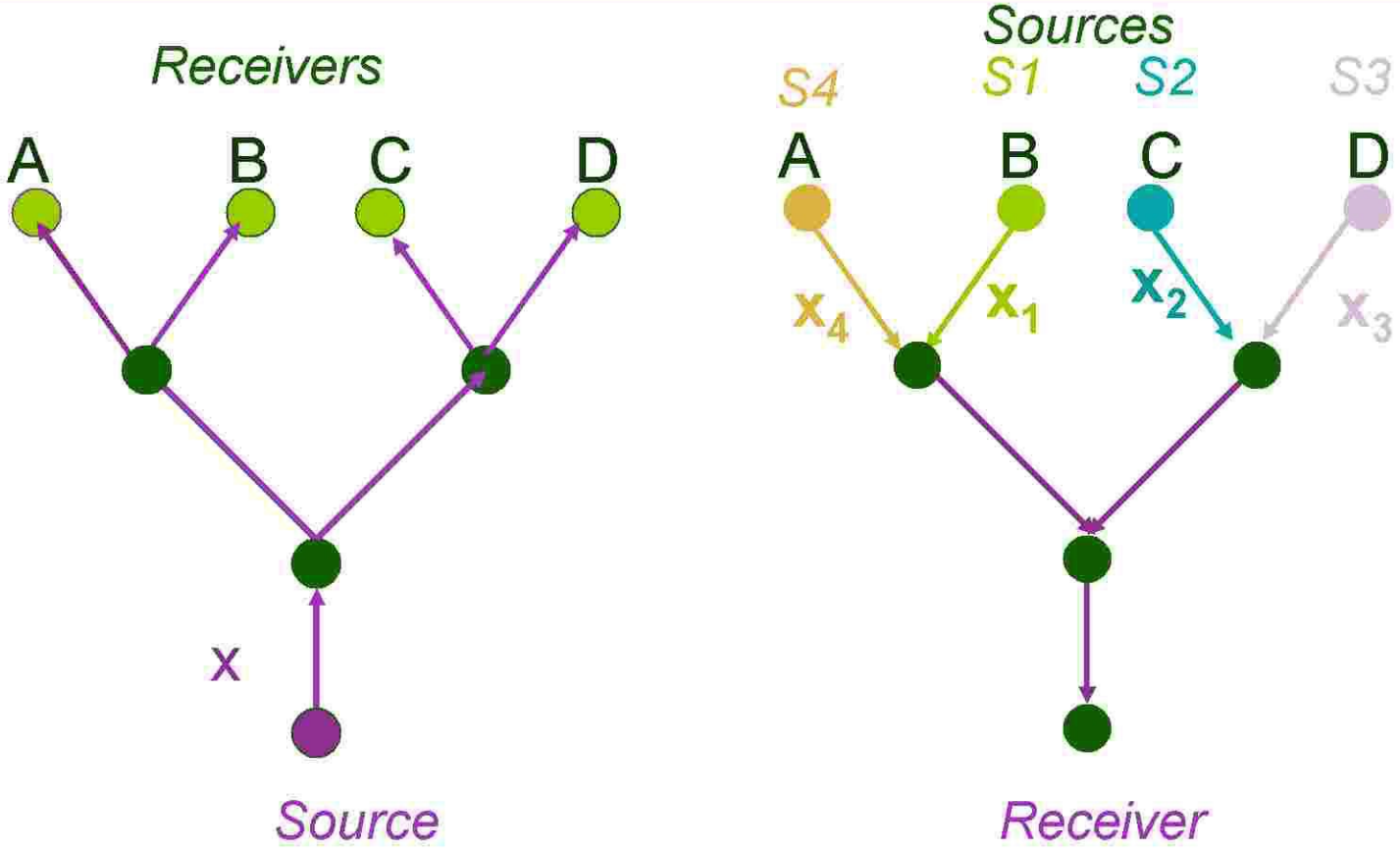


Sources send:

x1=(1 0 0 0)
x2=(0 1 0 0)
x3=(0 0 1 0)
x4=(0 0 0 1)

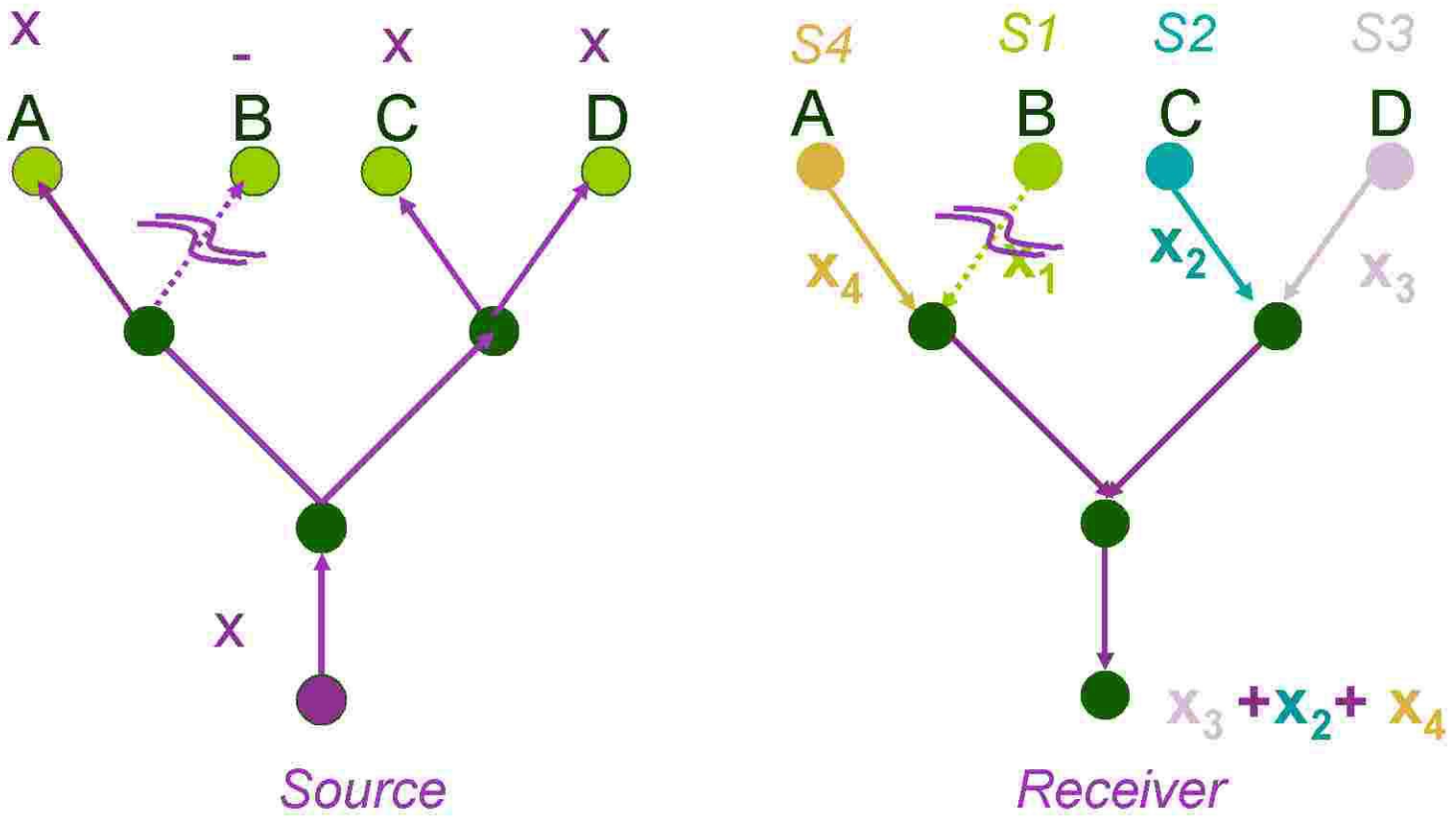
Receiver observes the XOR of a subset of the source packets

Multicast and Inverse Multicast Tree: Equivalent



Theorem: the MLE for a multicast tree and its inverse coincide.

Multicast and Inverse Multicast Tree: Equivalent



Intuition: there is a 1-1 correspondence between observable outcomes (and with the same probability) in the 2 trees.

Topology Inference
Conclusions and Ongoing Work

- Proposed algorithms for topology inference using network coding capabilities [Allerton 06]
  - Lossless Trees: deterministic inference in O(n)
  - Lossy Trees: rapid inference
  - Equivalence of multicast and inverse multicast tree
- Intuition:
  - when internal nodes combine incoming flows using NC → they reveal information about topological structure
- Ongoing work
  - Exploit correlation introduced by both link losses and NC
  - Extend from trees to arbitrary topologies
  - Passive topology inference

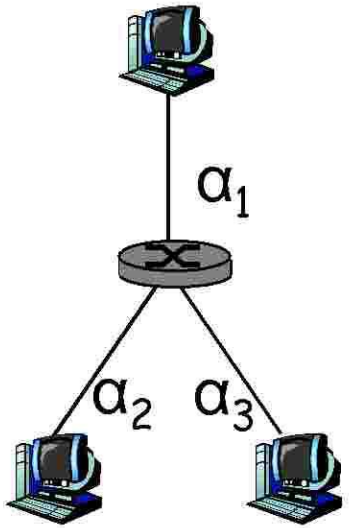
Outline

- Background
  - Network tomography
  - Network coding
- Topology Inference using Network Coding
- Link Loss Inference using Network Coding
- Conclusion & Ongoing Work

Traditional Loss Inference

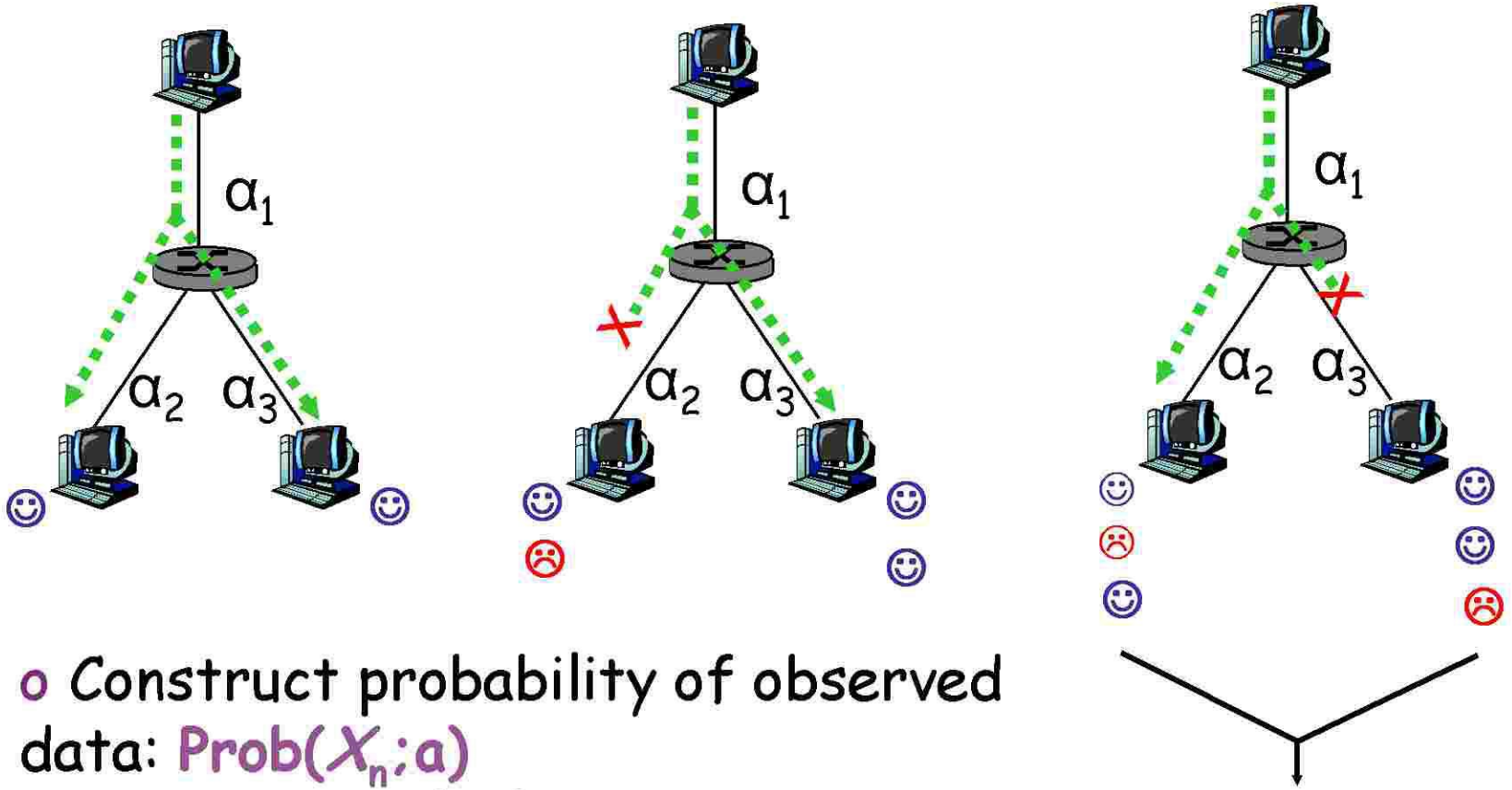
[Caceres, Duffield, Horowitz, Towsley, @IT'99]

- Multicast probes
  - receivers observe correlated performance
  - *exploit* correlation to infer link loss behavior
- Loss model:
  - Bernoulli losses a_k, $k \in L$
  - independent between links
- Data:
  - n probes, X_n : record of probes
- Goal:
  - estimate link probabilities $\alpha = \{\alpha_k : k \in L\}$ from X



Traditional Loss Inference

[Caceres, Duffield, Horowitz, Towsley, @IT'99]



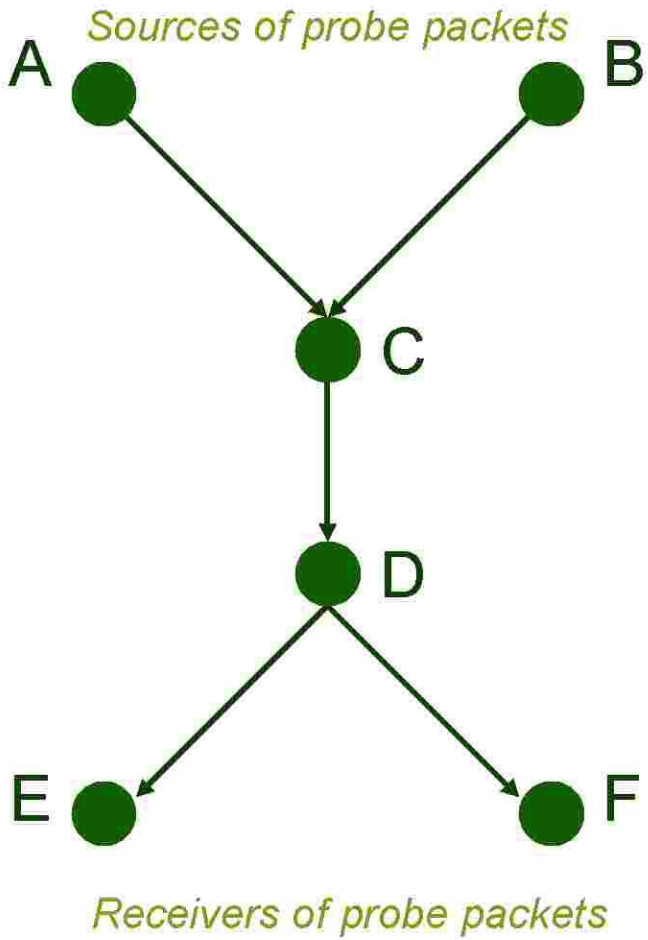
- Construct probability of observed data: $Prob(X_n;a)$
- Estimate a = {ak}, to maximize likelihood: $a(n)=\arg\max_a Prob(X_n;a)$

Prior Tomography Work

- **Single multicast tree (MINC)**
  - R. Caceres, N. Duffield, J. Horowitz and D. Towsley, "Multicast-based inference of network-internal loss characteristics", *Transactions on Inf. Theory, 1999*
  - Link Metrics: loss, delay
- **Extensions:**
  - Computationally efficient, suboptimal algorithms vs. MLE
  - General topologies: covering the network with several trees [Lo Presti & Duffield]
  - Unicast probes: back-to-back probes [Nowak]
- **Most relevant to us:**
  - [Rabbat, Nowak '04]
    - Multiple sources, unicast probes share fate
    - Joint topology and link loss inference
  - [Yan et al., Sigcomm 04-06]
    - Overlay network measurements

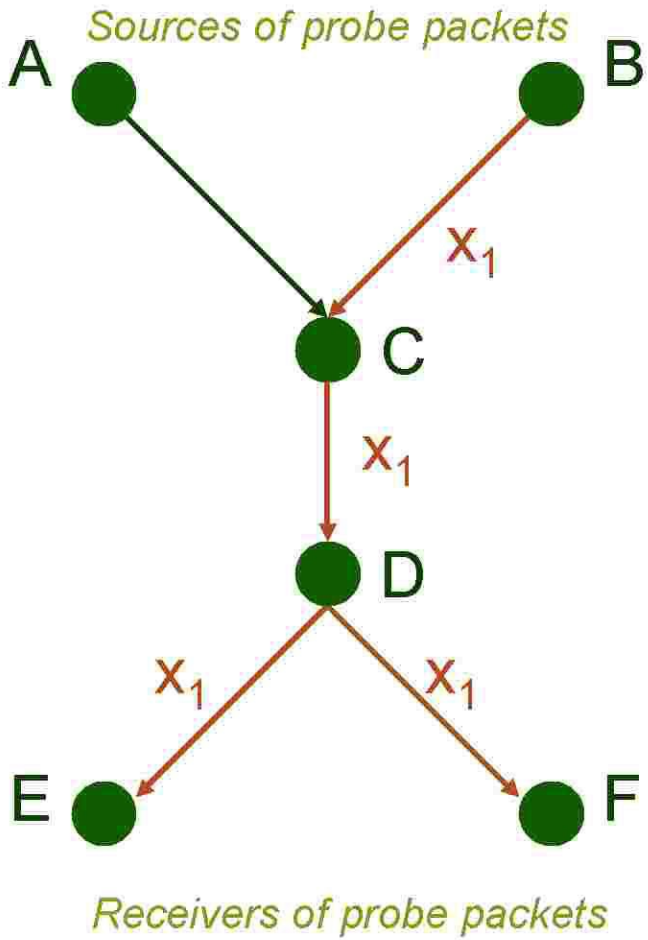
Loss Inference w. Network Coding

Basic Example



- We want to infer the link loss rates a_k on all links $k \in \{AB, AC, CD, DE, DF\}$
- using end-to-end probes from {A,B}, to {E,F}

Traditional Loss Inference



- Use multicast trees

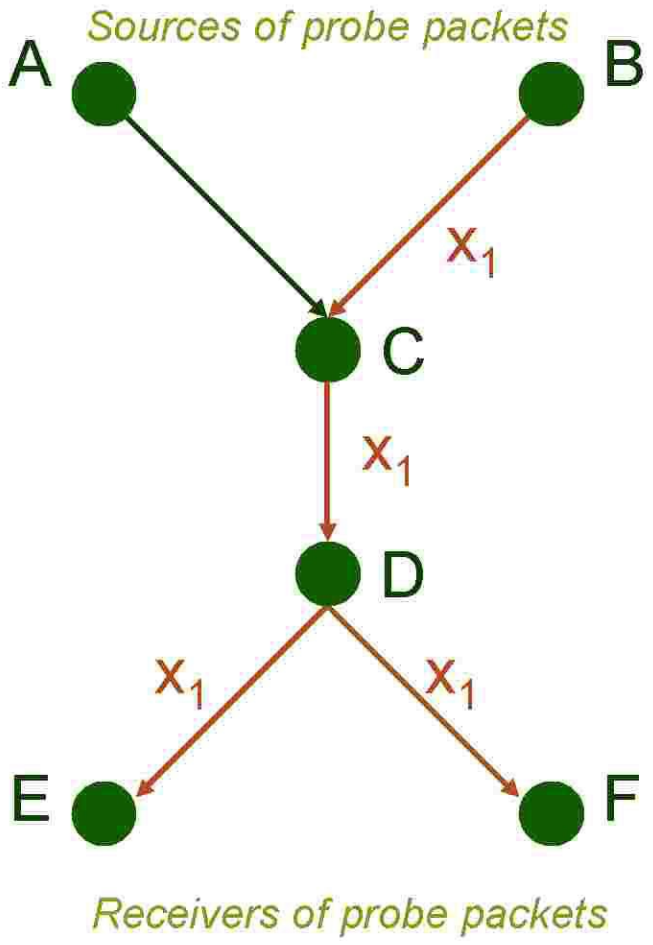
- *R. Caceres, N. Duffield, J. Horowtz and D. Towsley, "Multicast-based inference of network-internal loss characteristics", Trans. Inf. Theory, 1999*

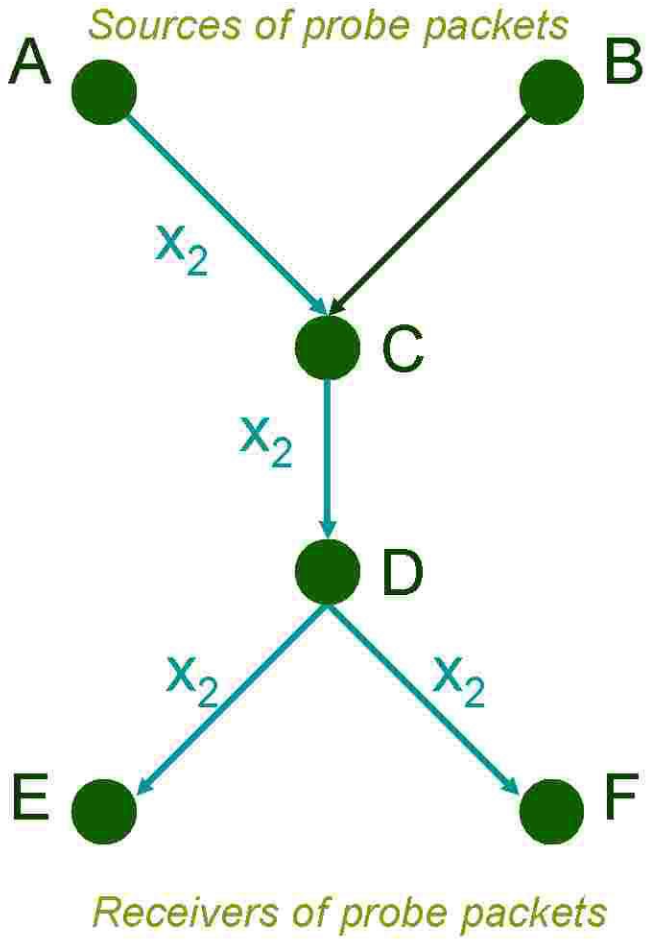
- *M. Rabat, R. Nowak and M. Coates, "Multiple source, multiple destination network tomography", Infocom 2004.*

- *M. Adler, T. Bu, R. Sitaraman and D. Towsley, "Tree layout for internal network characterizations in multicast networks", ACM NGC 2001.*

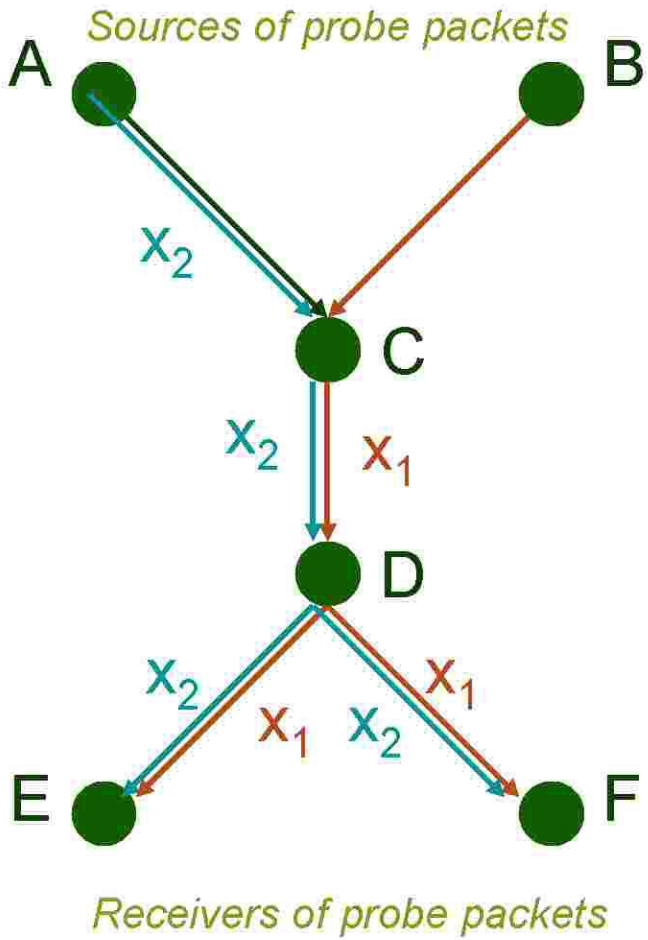
- *.....*

Covering the graph with trees

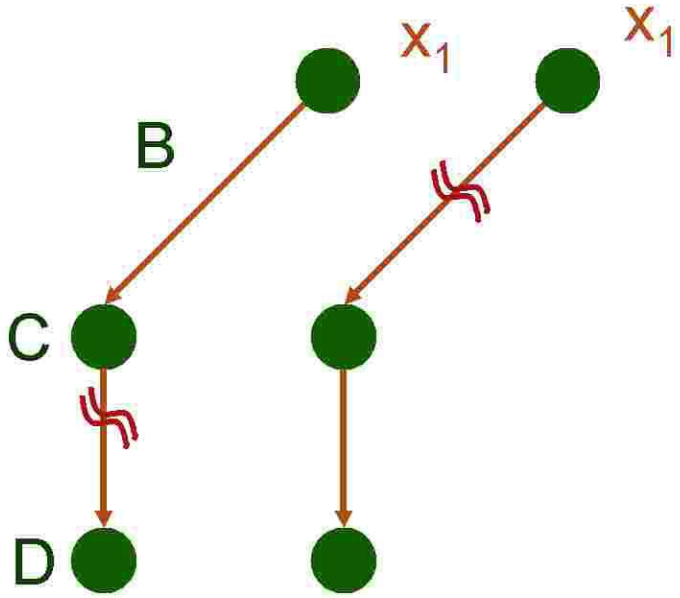




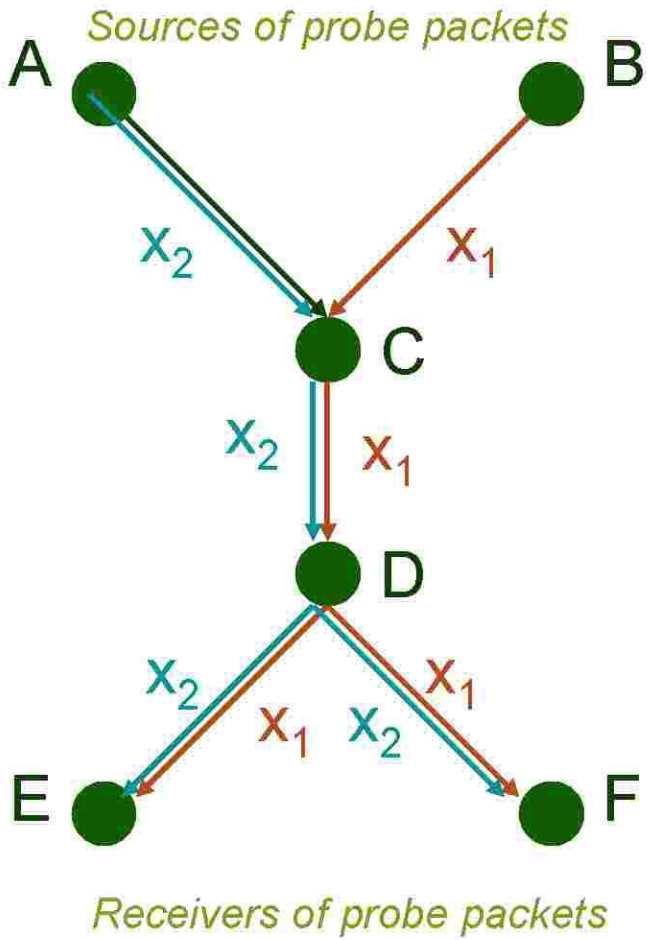
Drawbacks



1. We cannot infer the loss rate for edge CD



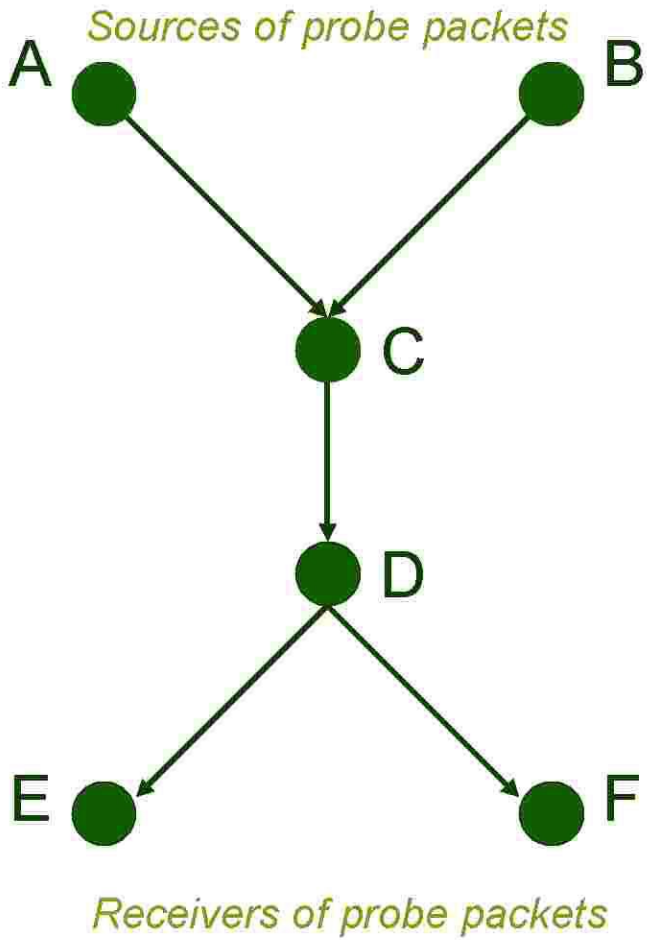
Drawbacks



1. We cannot infer the loss rate for edge CD
2. Minimum cost covering with multicast trees is NP-hard
3. Paths overlap from C and downstream
4. Combining observations from 2 tress leads to suboptimal estimation

Network coding approach

[C.Fragouli, A. Markopoulou Allerton 05]

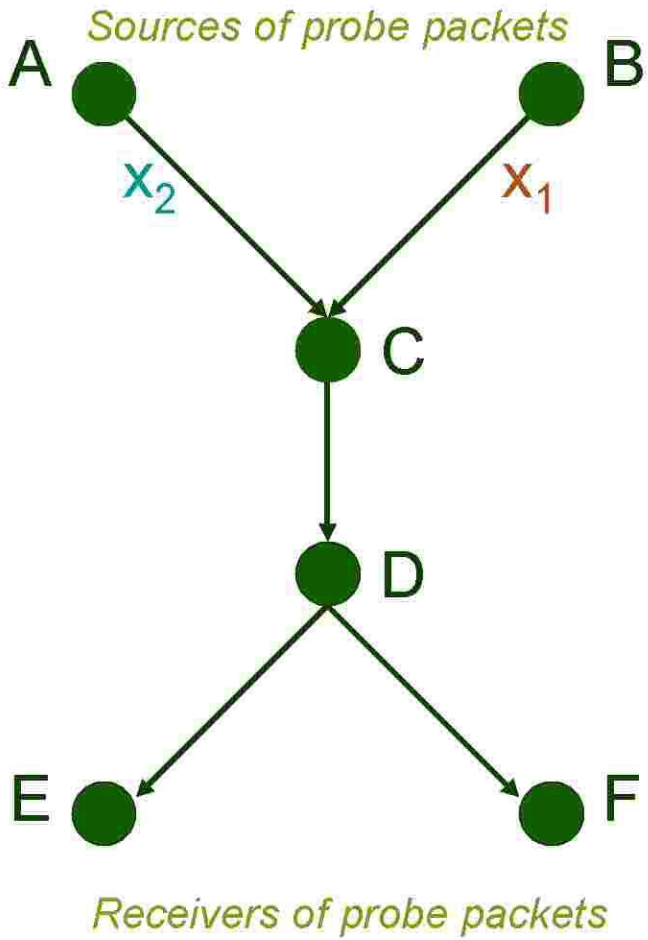


Intermediate node (C):

Within a time window

- if received 2 incoming packets,
  - XOR them and forward
- if received 1 incoming packet
  - just forward

Network coding approach

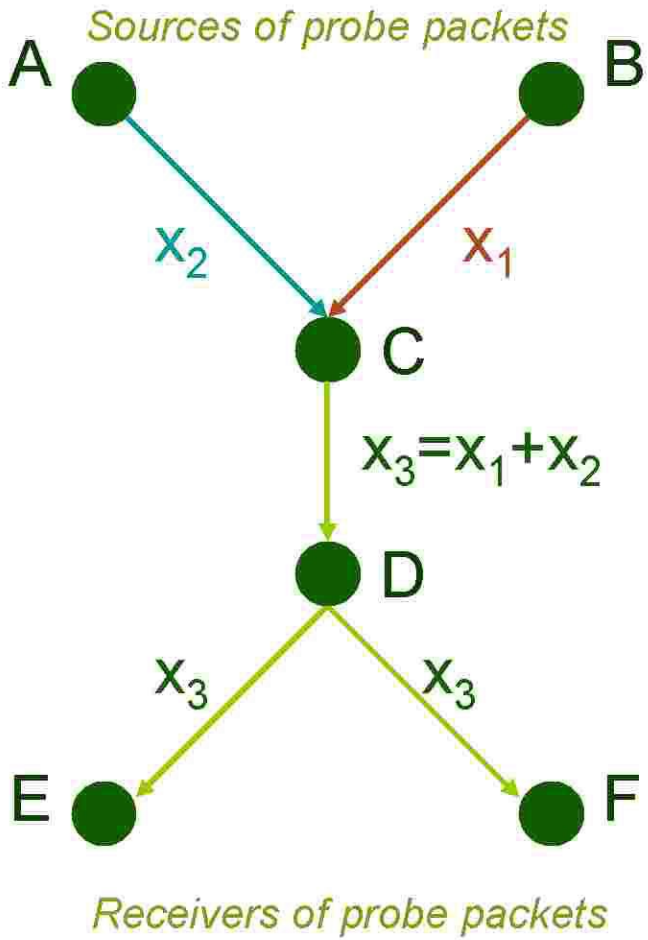


Example:

Nodes A and B send packets

$x_1 = [\ 1\ 0\ ]$, $x_2 = [\ 0\ 1\ ]$

Network coding approach



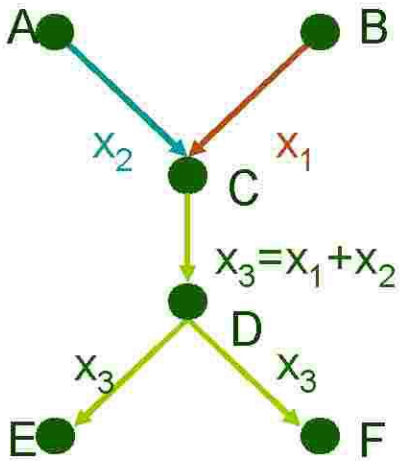
Example:

Nodes A and B send packets

$x_1 = [\ 1\ 0\ ],\ x_2 = [\ 0\ 1\ ]$

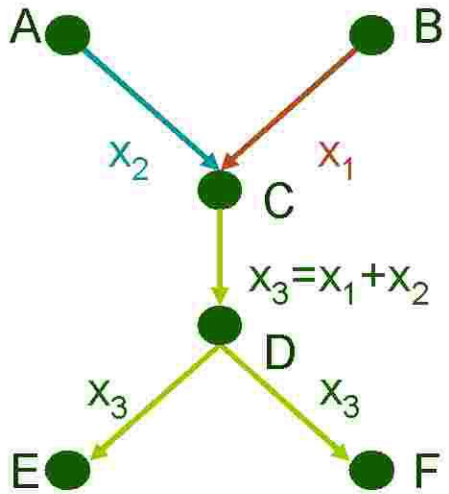
Observations → Events

E	F	AB	BC	CD	DE	DF
-	-	all events not listed below				
x1	-	1	0	1	1	0
x2	-	0	1	1	1	0
x3	-	1	1	1	1	0
-	x1	1	0	1	0	1
x1	x1	1	0	1	1	1
-	x2	0	1	1	0	1
x2	x2	0	1	1	1	1
-	x3	1	1	1	0	1
x3	x3	1	1	1	1	1



Why NC does better?

- Each observed probe conveys information for paths from two (instead of one) sources
  - more information per probe!
- NC combines packets on (otherwise ovelapping paths) into exactly one probe per link
  - we can have more sources for no additional bandwidth!



Problem Decomposition

Link loss inference involves the following steps:

1. Identifiability
2. Select probe paths, sources/receivers
3. Packet Design
4. Estimation

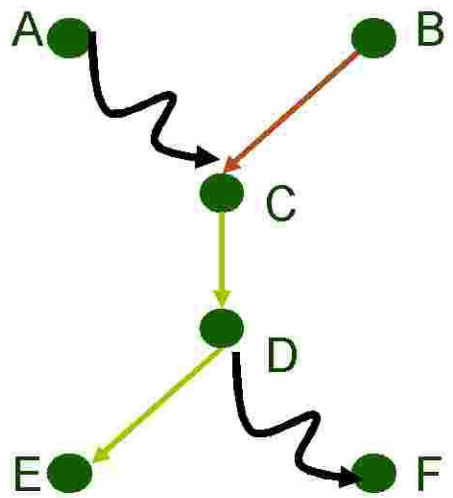
Summary of Results

With NC, we get the following benefits:

1. Identifiability
   - We can identify more links
2. Select sources/receivers, probe paths
   - Covering the graph becomes easier
   - More sources help and "for free"
3. Packet Design
4. Estimation
   - Better, even with suboptimal algorithms

Benefit #1:
We can identify more links

- Theorem [ITA'07]:
  - Link CD is identifiable if and only if
    {C is a source or a coding point} and
    {D is a receiver or a branching point}



Benefit #2:
Easier to select the probe paths

- **Minimum-cost covering of the graph**
  - with multicast trees is NP-hard
  - With NC, estimating a set of links is LP, [Allerton 05]
  - A useful special case:
    - to estimate all identifiable edges, no need to even solve the LP; just have sources emit their probes

- **The "points of view" matters**
  - Equivalence between multicast and inverse multicast
  - Guidelines for selecting sources and receivers [ITA '07]
  - More sources always help. With NC, they also come for free.

Issue #3
Packet Design

- If the graph is a tree:
  - Source i sends xi=[0 …1 …0], with 1 in the i^{th} position
    - n sources, vector of length n
  - Intermediate nodes just XOR
  - Receiver can distinguish among different subsets

- Graphs with cycles?
  - $(x_1+x_2)+x_1= x_2$(in F_2) or $2x_{1+}x_2$(in F_4)

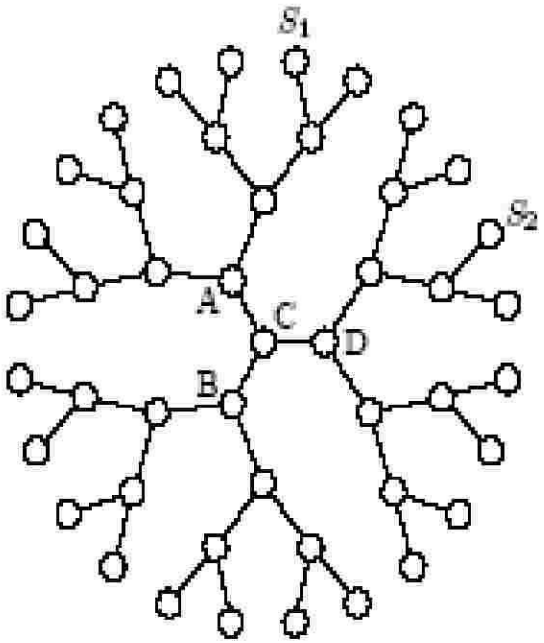
Benefit #4:
Estimation improves

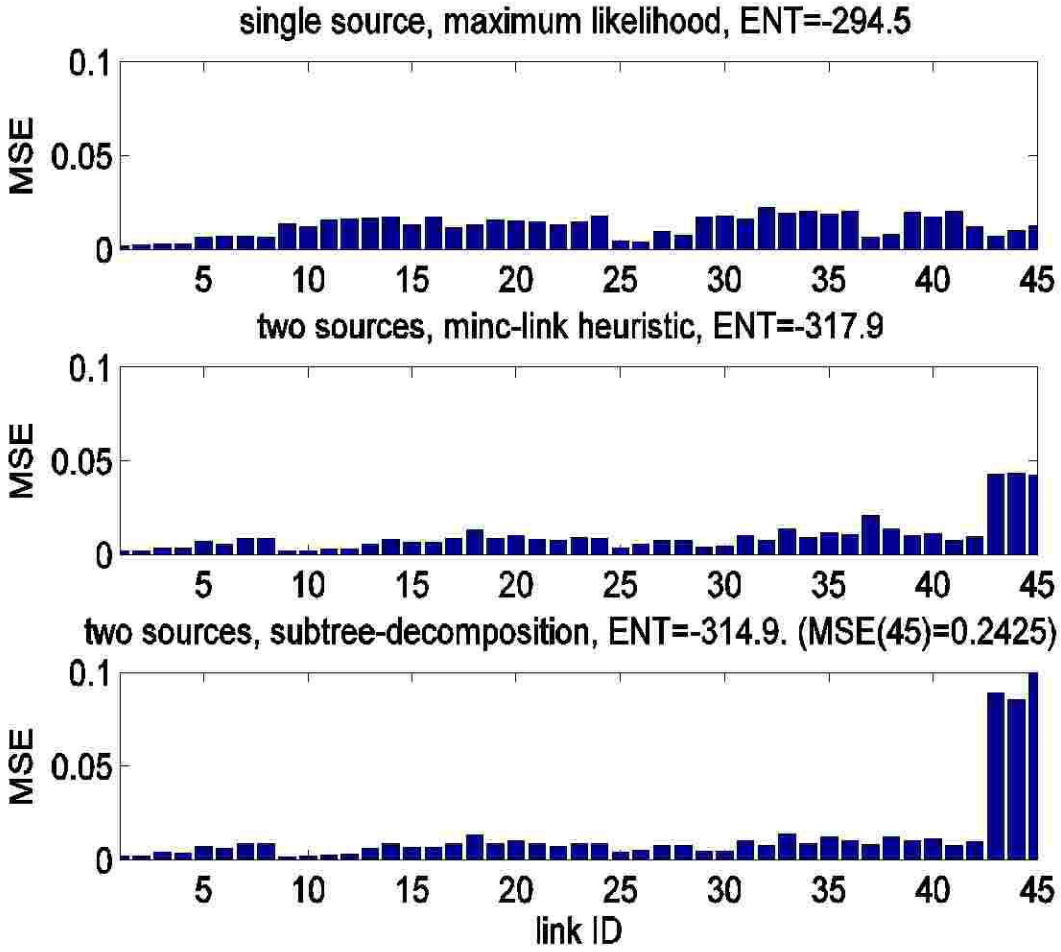
- MLE is prohibitively complex
- We develop suboptimal algorithms [ITA '07]
  - Subtree decomposition
  - Belief propagation

- More sources improve estimation
  - Multiple sources+NC (even with suboptimal estimation) better than single source+multicast (with MLE) .
- Intuition:
  - Every probe has more information
  - multiple sources use no additional bandwidth

Preliminary simulations

Two sources (with suboptimal estimation) do better than one (with MLE)





Link Loss Inference using NC
Summary and Ongoing Work

- Preliminary results show that links loss inference improves with NC
- Intuition:
  - more sources help and NC removed the bandwidth duplication

- Ongoing work [ITA'07]
  - Estimation depends on the "points of view":
    - how to select the number and placement of sources.
  - From trees to graphs with cycles
  - Suboptimal algorithms

Outline

- Background
  - Network tomography
  - Network coding
- Topology Inference using Network Coding
- Link Loss Inference using Network Coding
- Conclusions

Tomography vs. Network Coding:
contradicting concepts?

- Not in networks that already employ NC
- Internal nodes are still "simple"
  - NC not more complex than forwarding or multicasting
  - other processing delegated to "special" nodes
- No need to reveal internal nodes' identity
- NC allows for rapid inference

Thank you!

- More information:
  - Fragouli, Markopoulou, "Network Monitoring using Network Coding Techniques", *Allerton '05*
  - Fragouli, Markopoulou, Diggavi, "Topology Inference using Network Coding", *Allerton '06*
  - Fragouli, Markopoulou, Srinivasan, Diggavi, "Network Monitoring: it depends on your point of view", *ITA '07*

- Contact
  - athina@uci.edu, newport.eecs.uci.edu/~athina