Network error correction

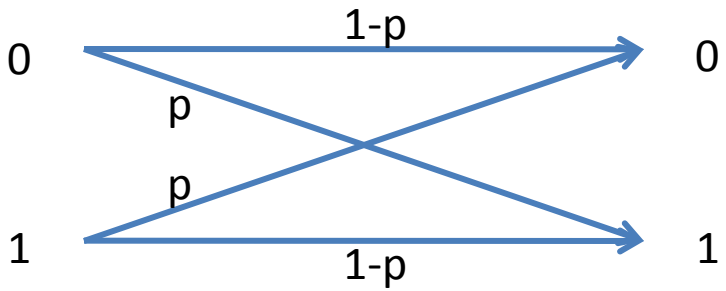
Tracey Ho

Lee Center Workshop

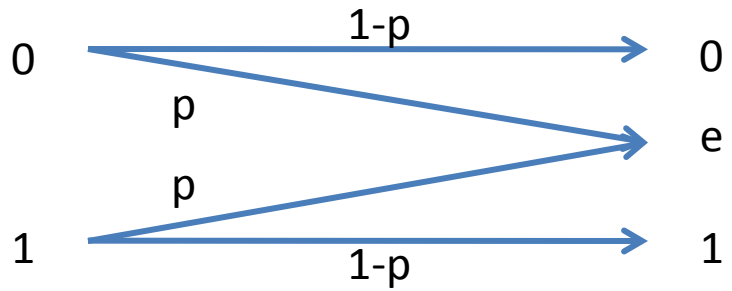
Introduction

- One of the basic problems of information and coding theory is reliable communication over noisy channels by adding redundancy
- Originally studied for point-to-point channels with a single source and sink

e.g. binary error channel



e.g. binary erasure channel



  - E.g. (3,1) repetition code: send three copies of the same bit
    - Corrects one error or two erasures
- Generalization of error correction to networks with multiple sources or sinks introduces new elements into the problem

Error correction in networks

- We consider a network composed of point-to-point channels (links) that may be error-prone
- If errors are not ergodic on each link, it is not sufficient to do error correction on a link by link basis
  - E.g. unpredictable faults or adversarial interference
  - Need network error correction codes that operate across multiple links
- Network error correction is a basic component of information theoretic security for networks

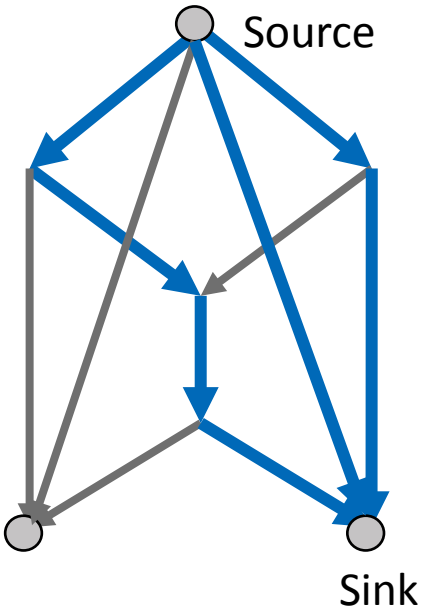
Problem 1: correction of errors in arbitrary locations

- Acyclic network composed of point-to-point unit capacity links
- Arbitrary errors can occur on up to *z* links
- Communication scenarios can be classified into:
  - Single source, single sink
  - Single source multicasting to multiple sinks
  - Multiple sources sending to one sink, or multicasting to multiple sinks
  - Multiple sources sending different information to different subsets of sinks (non-multicast, general case)
- How do the different communication scenarios affect the network error correction problem?

Single source, single sink

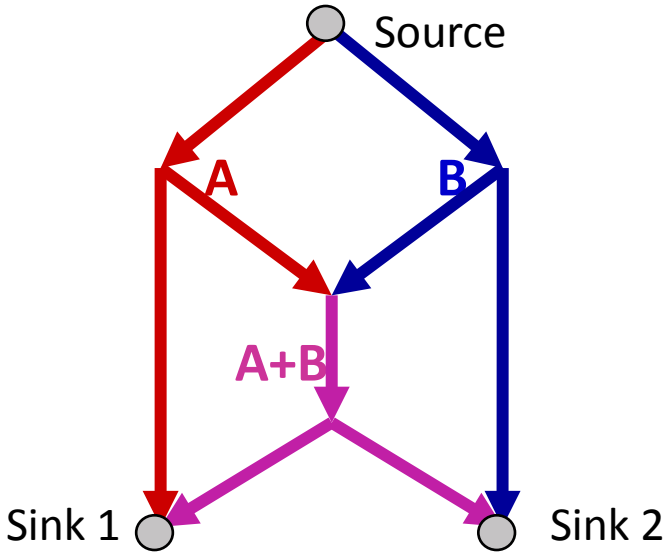
- Simply apply a point-to-point error correction code over multiple disjoint paths from source to sink
- From classical coding results, capacity = minimum cut - *2z*
  - E.g. *z = 1,* capacity = 1

    Applying a (3,1) repetition code corrects any single error



Single source multicast

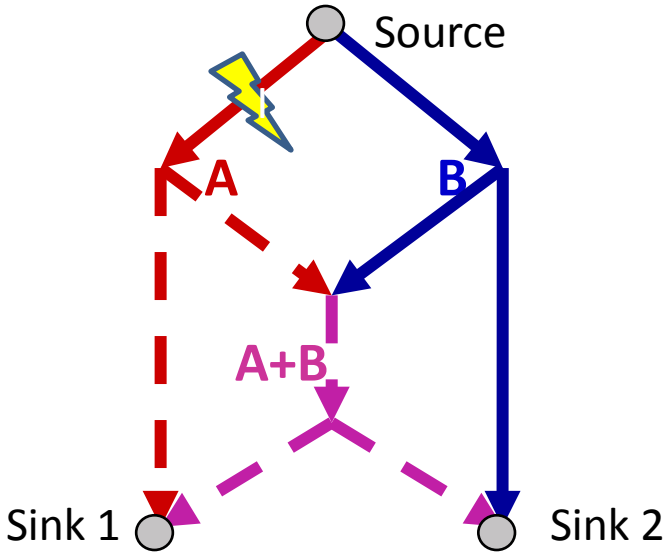
- Even in the error-free case, network coding may be required to achieve capacity, e.g.[1]



[1] R. Ahlswede et al., ``Network information flow,'' *IEEE Trans. Inform. Theory*, IT-46:4 1204-1216, 2000.

Single source multicast

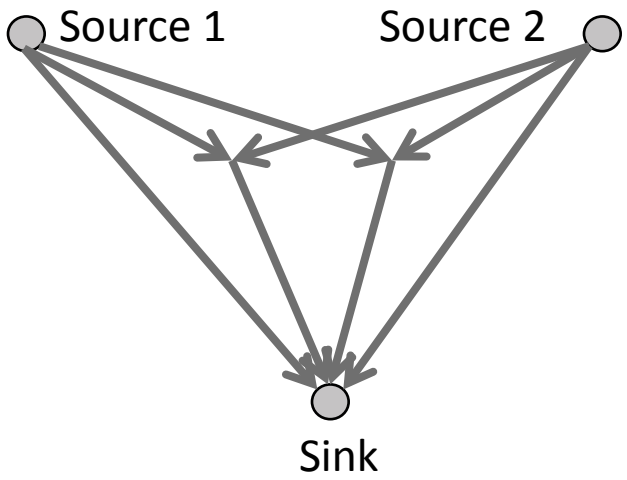
- Even in the error-free case, network coding may be required to achieve capacity, e.g.[1]



- An error on one link may cause errors on many links through coding
  - But the errors are dependent
  - Point-to-point error correction codes don't take this into account

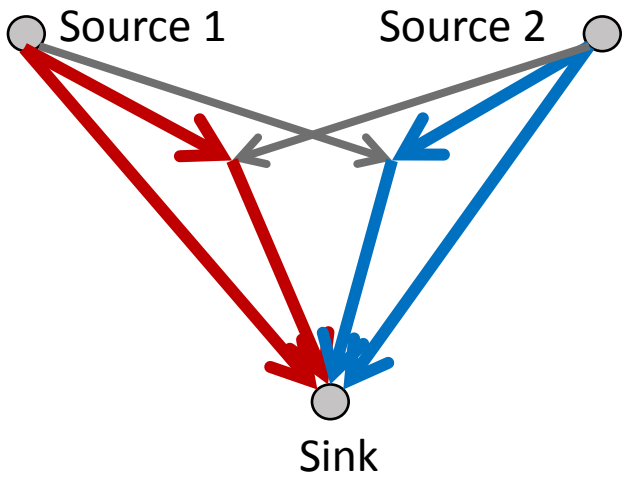
[1] R. Ahlswede et al., ``Network information flow,'' *IEEE Trans. Inform. Theory*, IT-46:4, 1204-1216, 2000.

Multiple sources, single sink



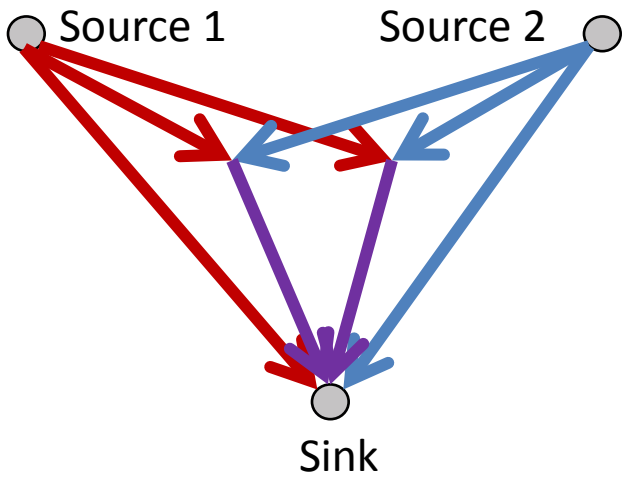
- The sources do not know each other's information
  - Can't send redundant information on behalf of each other

Multiple sources, single sink



- The sources do not know each other's information
  - Can't send redundant information on behalf of each other
- If we allocate part of the network to carry information from source 1 only, and the rest of the network carries information from source 2 only, then each part must be able to deal with *z* errors
  - But this is excessive, since there are only *z* errors altogether

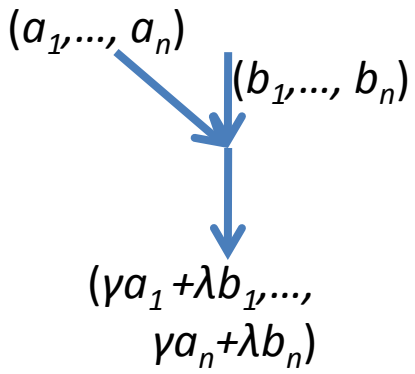
Multiple sources, single sink



- The sources do not know each other's information
  - Can't send redundant information on behalf of each other
- If we allocate part of the network to carry information from source 1 only, and the rest of the network carries information from source 2 only, then each part must be able to deal with *z* errors
  - But this is excessive, since there are only *z* errors altogether
- Coding across the two sources is needed so that they can simultaneously use shared network capacity to send redundant information

Coherent and non-coherent cases

- Coherent case: network topology known, centralized code design
- Non-coherent case: network topology not known, distributed code design
  - Convenient to use random linear network coding[1]
    - each packet is represented as a vector of symbols from a finite field F_q
    - each node sends linear combinations of its received packets with coefficients chosen uniformly at random from F_q
    - Information is transmitted via the subspace spanned by the received vectors[2]

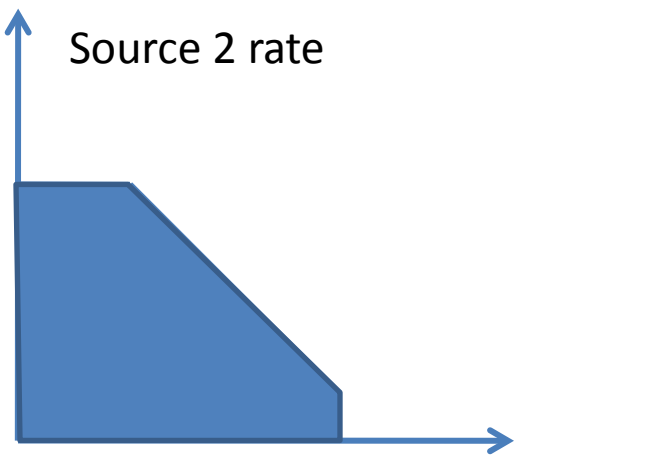


[1] T. Ho et al., ``A random linear coding approach to multicast,'' *IEEE Trans. Inform. Theory*, IT-52:10, 4413-4430, 2006.

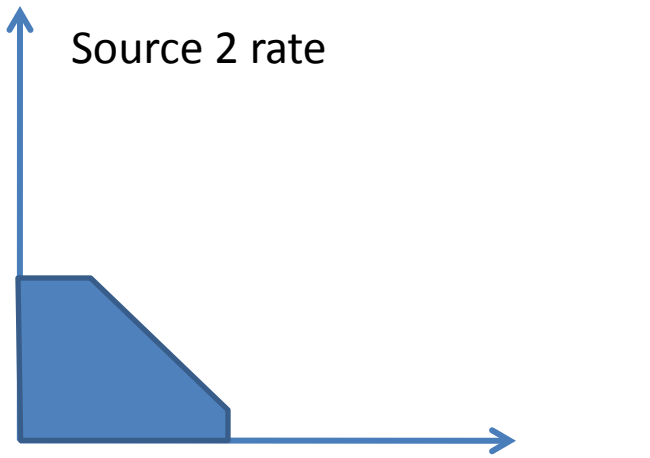
[2] R. Koetter, F. R. Kschischang, ``Coding for errors and erasures in random network coding,'' *IEEE Trans. Inform. Theory*, IT-54:8, 3579-3591, 2008.

Questions

- How to construct network error correction codes for the various cases?
- What are the capacity regions?
  - Except for non-multicast, where the capacity region even in the error free case is an open question



Small *z*

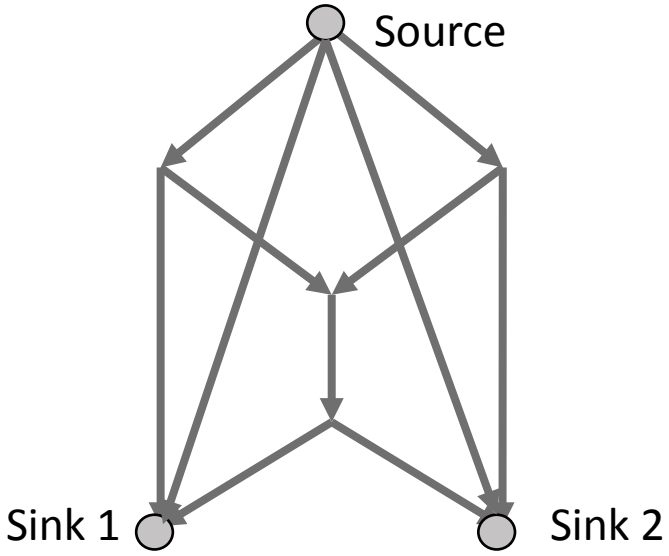


Large *z*

Coherent single-source multicast

- Let *m* be the minimum cut capacity between the source and any sink
- **Theorem**: The capacity under any *z* link errors is *m-2z*
- Example: *z = 1*

  Capacity = 1



R. W. Yeung, N Cai. Network Error Correction, I: Basic Concepts and Upper Bounds. Commun. Inf. Syst., vol. 6, iss 1 (2006), 19-36.

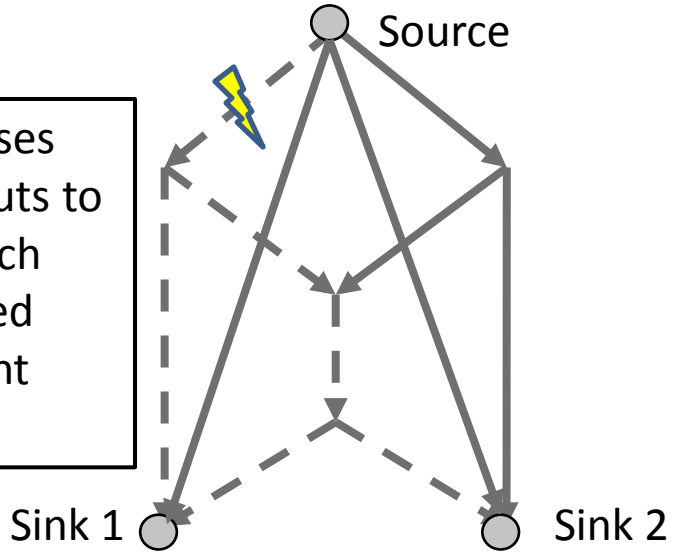
N. Cai, R. W. Yeung. Network Error Correction, II: Lower Bounds. Commun. Inf. Syst., vol. 6, iss 1 (2006), 37-54.

Coherent single-source multicast

- Let *m* be the minimum cut capacity between the source and any sink
- **Theorem**: The capacity under any *z* link errors is *m-2z*
- Example: *z = 1*

Capacity = 1

This one error causes two of sink 1's inputs to be erroneous, which cannot be corrected using point-to-point codes

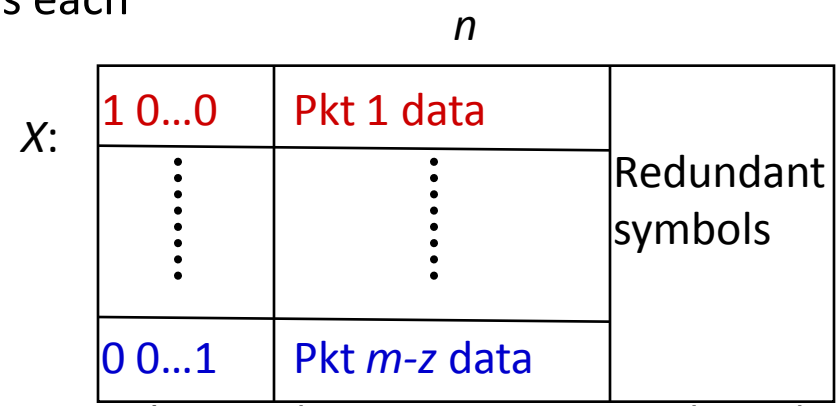


R. W. Yeung, N Cai. Network Error Correction, I: Basic Concepts and Upper Bounds. Commun. Inf. Syst., vol. 6, iss 1 (2006), 19-36.

N. Cai, R. W. Yeung. Network Error Correction, II: Lower Bounds. Commun. Inf. Syst., vol. 6, iss 1 (2006), 37-54.

Non-coherent single-source multicast

- Let *m* be the minimum cut capacity between the source and any sink
- **Theorem**: The capacity under any *z* link errors is *m-2z* and is achievable with high probability with a distributed polynomial complexity algorithm
- Source adds *(z+ε)n* redundant symbols, s.t. the resulting value of the source packets *X* satisfies *(z+ε)n* randomly chosen linear constraints, and forms *m-z* packets of *n* symbols each



- We show that w.h.p. over the random constraints and random code, for all q^{zn} possible values of the set of error packets, the sink can construct and solve a full rank system of linear equations to obtain source data
- Out of the *(m-z)n* symbols, *(z+ε)n* are redundant and $(m-z)^2$ are header information, so rate *m-2z* is achieved asymptotically with packet length *n*

S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, M. Medard, "Resilient Network Coding in the Presence of Byzantine Adversaries," Infocom 2007.

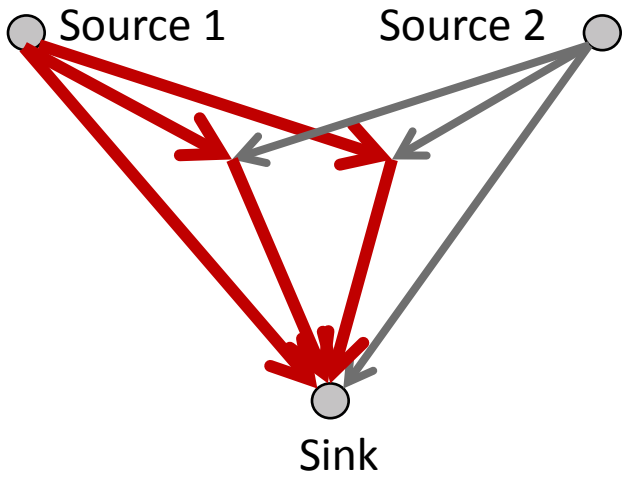
Multiple-source multicast

- Let U be the set of source nodes, and let m_S be the minimum cut capacity between sources in subset S of U and each sink
- Denote by r_i the multicast rate from the i^{th} source to the sinks
- **Theorem**: For both the coherent and noncoherent case, the capacity region under any z link errors is

$$\sum_{i \in S} r_i \leq m_S - 2z, \forall S \subseteq U$$

- Example: $z = 1$

  $r_1 \leq 1$



S. Vyetrenko, T. Ho, M. Effros, J. Kliewer and E. Erez, "Rate regions for coherent and noncoherent multisource network error correction," to appear in IEEE ISIT, Jun. 2009.

Multiple-source multicast

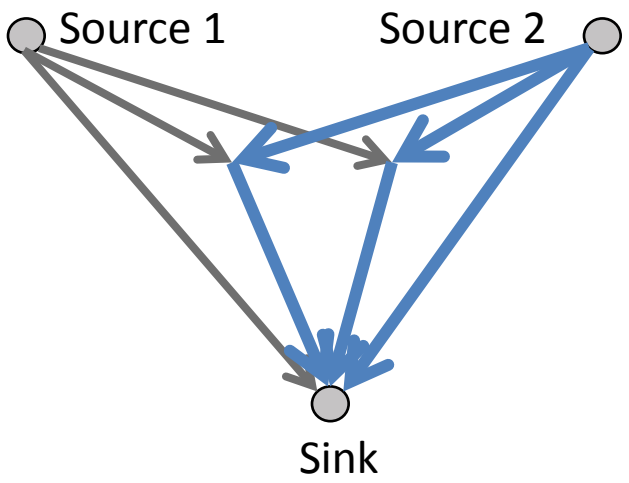
- Let U be the set of source nodes, and let m_S be the minimum cut capacity between sources in subset S of U and each sink
- Denote by r_i the multicast rate from the i^{th} source to the sinks
- **Theorem**: For both the coherent and noncoherent case, the capacity region under any z link errors is

$$\sum_{i \in S} r_i \leq m_S - 2z, \forall S \subseteq U$$

- Example: $z = 1$

  $r_1 \leq 1$

  $r_2 \leq 1$



S. Vyetrenko, T. Ho, M. Effros, J. Kliewer and E. Erez, "Rate regions for coherent and noncoherent multisource network error correction," to appear in IEEE ISIT, Jun. 2009.

Multiple-source multicast

- Let U be the set of source nodes, and let m_S be the minimum cut capacity between sources in subset S of U and each sink
- Denote by r_i the multicast rate from the i^{th} source to the sinks
- **Theorem**: For both the coherent and noncoherent case, the capacity region under any z link errors is

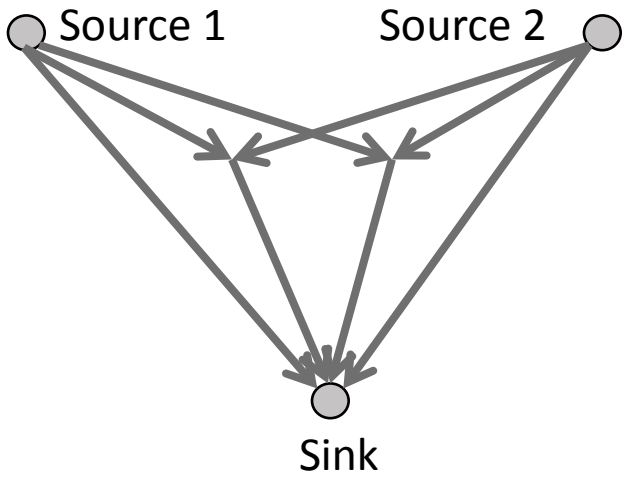
$$\sum_{i \in S} r_i \leq m_S - 2z, \forall S \subseteq U$$

- Example: $z = 1$

  $r_1 \leq 1$

  $r_2 \leq 1$

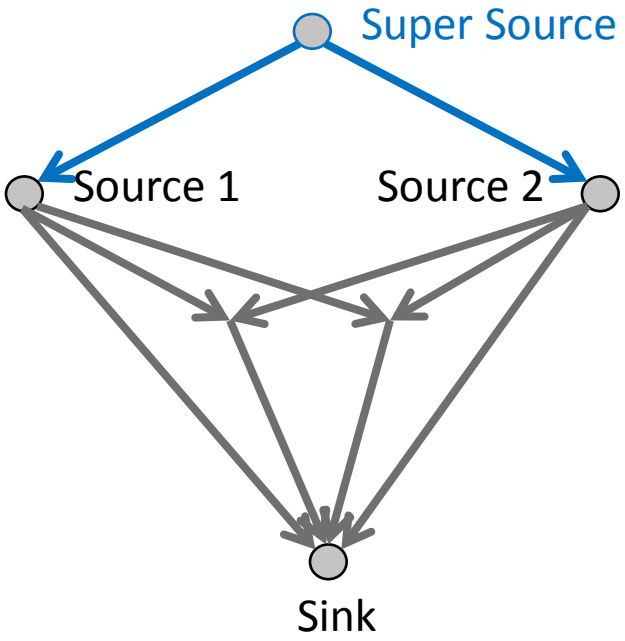
  $(r_1 + r_2 \leq 2)$



S. Vyetrenko, T. Ho, M. Effros, J. Kliewer and E. Erez, "Rate regions for coherent and noncoherent multisource network error correction," to appear in IEEE ISIT, Jun. 2009.

Multiple-source multicast (cont'd)

- Converse:
  - For any subset of sources *S*, add a virtual super source node with k_i links to each source *i* in *S*
  - Apply the Singleton bound to minimum cut



S. Vyetrenko, T. Ho, M. Effros, J. Kliewer and E. Erez, "Rate regions for coherent and noncoherent multisource network error correction," to appear in IEEE ISIT, Jun. 2009.

Multiple-source multicast (cont'd)

- Sketch of achievability (noncoherent case):
  - Code construction
    - Non-source nodes do distributed random linear network coding
    - From the single-source case, for each source we can construct a code C_i where each codeword comprises $k_i > r_i + z$ vectors, that corrects any z additions
    - This implies[1] that for any pair of distinct codewords V_i and V_i' in C_i ,

      $\dim(V_i \cap V_i') < k_i - z$
    - We can make vectors from one source linearly independent of vectors from all other sources by appending a vector of length $\sum k_i$

S. Vyetrenko, T. Ho, M. Effros, J. Kliewer and E. Erez, "Rate regions for coherent and noncoherent multisource network error correction," to appear in IEEE ISIT, Jun. 2009.

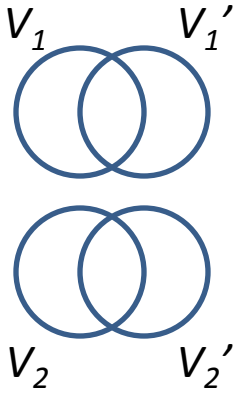
[1] S. Vyetrenko, T. Ho, and E. Erez, "On noncoherent error and erasure correction," to appear in IEEE ISIT, Jun. 2009.

Multiple-source multicast (cont'd)

- Error analysis
  - Consider any subset S of sources, and let V and W be the subspaces spanned by the vectors from sources in S and sources not in S respectively
  - We show that w.h.p., the sink's received subspace could not have been produced by W and a different set V' of codewords from sources in S, under any z errors
  - Consider just the sources in S and the error sources
    - Let P be the sink's received subspace, and let P_V be P excluding the error vectors
    - There exists a routing solution for which we can upper bound the dimension of $P_V \cap V'$ as

      $\sum_{i \epsilon S} \dim(V_i \cap V_i') < \sum_{i \epsilon S} (k_i - z)$

      since by the code design, a vector from V_i is not in any $V_j' \epsilon C_j$



S. Vyetrenko, T. Ho, M. Effros, J. Kliewer and E. Erez, "Rate regions for coherent and noncoherent multisource network error correction," to appear in IEEE ISIT, Jun. 2009.

Multiple-source multicast (cont'd)

- We can also lower bound the dimension of P_V by considering the error-free case
- We can then lower bound the dimension of $P_V \cup V'$, as well as $P \cup V'$ which contains it, by

  $\sum_{i \in S} k_i + z$
- For random linear coding in a sufficiently large field, w.h.p. subspaces are maximally linearly independent, so this bound carries over from the routing case
- We can then show that P contains more than z vectors that are not in V'

- Now we add the contribution from the rest of the sources
  - We can show that the sink's received vector space contains more than z vectors that are not in $V' \cup W$

S. Vyetrenko, T. Ho, M. Effros, J. Kliewer and E. Erez, "Rate regions for coherent and noncoherent multisource network error correction," to appear in IEEE ISIT, Jun. 2009.

Coherent non-multicast

- Finding the capacity region of a general non-multicast network is an open problem
- We give a construction of a network error correction code from a given error-free network code
- **Theorem**: Given any linear network code *C* that achieves rate vector (r_1 , …, r_n), where r_i is the information rate of source *i,* we can obtain a network code *C'* that achieves rate vector ($r_1 - 2z$, …, $r_n - 2z$) under any *z* link errors
- Intuition for proof:
  - Extend the concept of distance to non-multicast network codes:
    - For each sink *t*, let S_t be the subset of sources demanded by *t*
    - The distance between two codewords from S_t is defined as the minimum number of link errors needed to cause one codeword to result in the same received value at *t* as the other, under any choice of codewords from the other sources

S. Vyetrenko, T. Ho, M. Effros, J. Kliewer and E. Erez, "Rate regions for coherent and noncoherent multisource network error correction," to appear in IEEE ISIT, Jun. 2009.

Coherent non-multicast (cont'd)

- Analogously to the point-to-point and multicast cases, it can be shown that a non-multicast code corrects *z* errors, or *2z* erasures, iff it has minimum distance $d > 2z$
- Code *C'* is obtained by applying a random linear pre-code at each source
- We can prove that *C'* can correct any *2z* erasures
- It then follows that *C'* can correct any *z* errors

S. Vyetrenko, T. Ho, M. Effros, J. Kliewer and E. Erez, "Rate regions for coherent and noncoherent multisource network error correction," to appear in IEEE ISIT, Jun. 2009.

Problem 2: Non-worst-case error and erasure locations

- So far, we have considered codes that can correct errors on any *z* arbitrary links (worst-case)
  - Capacity can be achieved with random linear coding at every node in the multicast case
- For other models, e.g. random error and erasure locations,
  - The code constructions and capacity regions obtained for the worst-case model may be pessimistic
  - Random linear coding at every node is not always optimal, even for multicast

S. Vyetrenko, T. Ho, and E. Erez, "On noncoherent error and erasure correction," to appear in IEEE ISIT, Jun. 2009.

Non-worst-case error and erasure locations (cont'd)

- For a given realization of erasure and error locations and network coding/routing scheme, a necessary and sufficient decodability condition can be given in terms of the rank of matrices corresponding to useful and erroneous information received at the sink node
- This can be used to evaluate different coding/routing strategies under different error and erasure models
- The benefit of random network coding versus routing is found to increase with the relative occurrence of erasures versus errors, when they are randomly located

S. Vyetrenko, T. Ho, and E. Erez, "On noncoherent error and erasure correction," to appear in IEEE ISIT, Jun. 2009.

Conclusion

- Summary:
  - The presence of multiple sources or sinks introduces new angles into the error correction problem
  - Capacity results for coherent and noncoherent multi-source multicast
  - Achievability result for non-multicast
  - Analysis for non-worst-case error and erasure locations
- Further work:
  - Improved complexity constructions for multicast
  - Improved achievability results for non-multicast
  - Other error models

Thank you

Detection of adversarial errors in random network coding

Overview:

- High probability detection of adversarial errors in random network coded communication
- Only assumption is that adversary does not know the entire random network code; no limit on adversary's transmission capacity assumed

Approach:

- Augment each source packet with a flexible number of hash symbols
- Let each source packet contain n header/payload symbols $x_1,\ldots, x_n$ and k<n hash symbols $h_1, \ldots, h_k$, where n and k are design parameters which determine overhead
- $h_1 = \phi (x_1,\ldots, x_t) = x_1^2 +\ldots+ x_t^{t+1}$

  …

  $h_k = \phi (x_{t(k-1)+1},\ldots, x_n)$ where $t = \lceil n/k \rceil$

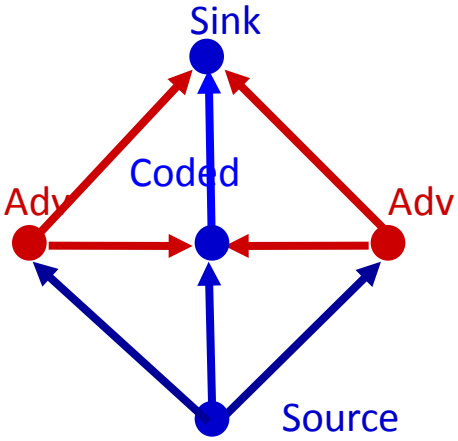
Main result:

- For symbol length log q bits, if the sink receives s linearly independent combinations of source packets whose code coefficients are unknown to the adversary (but which may be coded together with any number of adversarial packets), then if there is an adversarial error,

  a) for at least s decoded packets, the adversary cannot determine which of a set of at least q-1 possible values will be obtained

  b) the detection probability is at least $1-((t+1)/q)^s$

Example where sink can detec error



Example values:

- With 2% overhead (t=50), symbol length=7 bits, s=5, the detection probability is ≥ 98.9%
- With 1% overhead (t=100), symbol length=8 bits, s=5, the detection probability is ≥ 99.0%

Summary of results

- Capacity (achievability and converse):
  - Coherent single-source multicast (Cai & Yeung 06)
  - Noncoherent single-source multicast (Jaggi, Langberg, Katti, Ho, Katabi, Médard 07, Koetter & Kschischang 08)
  - Coherent and noncoherent multi-source multicast (Vyetrenko, Ho, Effros, Kliewer & Erez 09)
- Achievable construction:
  - Coherent non-multicast (Vyetrenko, Ho, Effros, Kliewer & Erez 09)
    - Capacity even in the error-free case remains an open question

Non-worst-case error and erasure locations (cont'd)

- For a given realization of erasure and error locations and network coding/routing scheme, a necessary and sufficient decodability condition can be given in terms of the rank of matrices corresponding to useful and erroneous information received at the sink node
- **Theorem**: For a non-coherent
  - $R - rank(TX + BZ) + 2rank(BZ) < y$