Fault-Tolerant and Secure Data Transmission Using Random Linear Network Coding

Pouya Ostovari and Jie Wu

Computer & Information Sciences
Temple University

Center for Networked Computing
http://www.cnc.temple.edu

Agenda

- Introduction
  - Multi-path network coding
  - Fault tolerance and security
- Fault-tolerant and secure data transmission
  - Problem definition
  - Problem formulation
- Evaluations
- Conclusions

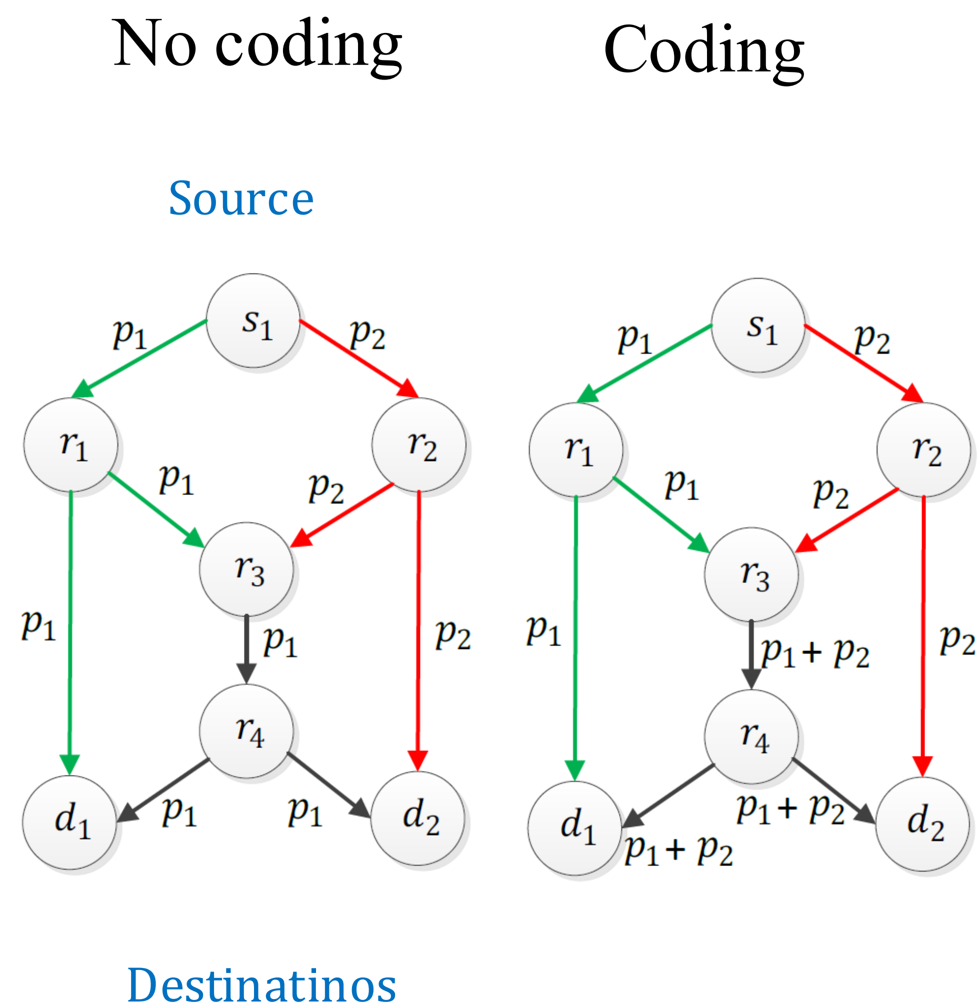
Introduction

- Multi-path transmission
  - Fault tolerance (FT) via redundancy
    - Transmitting data through multiple paths
    - Paths with different reliabilities
    - More redundancy increases FT, but increases the cost as well
  - Security
    - Encryption, public/private keys
    - Overhead of encryption methods

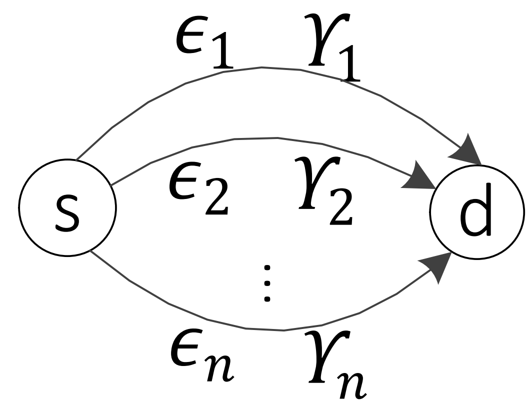
Network Coding

XOR network coding

- Single multicast
- Two packets
- Two destinations d_1 and d_2
- Capacity of each link: one packet



Simple System Setting

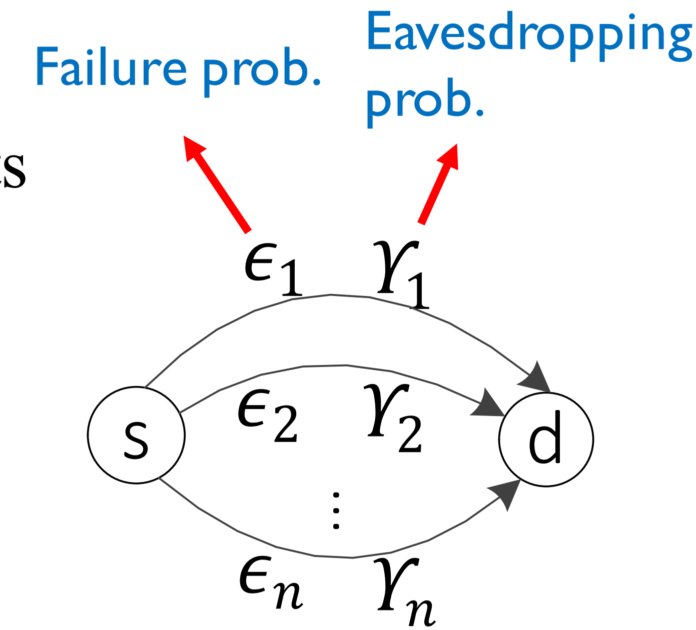


- Transmission a file with *m* packets via *n* disjoint paths
- Path failure model
  - If a path fails, all of the transmitted packets over that path fail
- Eavesdropper probability: fixed
  - e.g. in wireless networks depends on location of the eavesdropper
- Objective
  - Balance fault tolerance and security

Linear Coding

- Random linear network coding
  - Linear combinations of the packets

$$\begin{cases} q_1 = \alpha_{1,1}p_1 + \alpha_{1,2}p_2 + \alpha_{1,3}p_3 \\ q_2 = \alpha_{2,1}p_1 + \alpha_{2,2}p_2 + \alpha_{2,3}p_3 \\ \vdots \\ q_k = \alpha_{k,1}p_1 + \alpha_{k,2}p_2 + \alpha_{k,3}p_3 \end{cases}$$



  - *m*=*3* linearly independent coded packets are sufficient for decoding, using Gaussian elimination

- If we code *m* packets, eavesdropper/destination needs *m* coded packets to retrieve the original packets
- *m* and *n* can be different numbers

Fault Tolerance and Security

- FT
  - *m* linearly independent coded packets are sufficient for retrieving the original data
- Security
  - Eavesdropper cannot decode the coded packets unless it has *m* linearly independent packets
- Challenge

More transmitted coded packets

More robust against failures

More vulnerable against eavesdropping

Problem Formulation

- With n paths, there are 2^n possible failure/eavesdropping cases
  - $\boldsymbol{R_j}$: set of paths that do not fail
  - $\boldsymbol{S_j}$: set of overheared paths by eavesdropper

ith path

Failure prob. of the path d_i

$$p_j = \prod_{d_i \in R_j} (1 - \epsilon_i) \prod_{d_i \notin R_j} \epsilon_i$$

Prob. of paths in set R_j not to fail and the rest fail

$$q_j = \prod_{d_i \in S_j} \gamma_i \prod_{d_i \notin S_j} (1 - \gamma_i)$$

Prob. that an eavesdropper has only access to data transmitted on the set of paths in S_j

Eavesdropping prob. of the ith path

Problem Formulation- Case 1

- Objective function as a function of FT and security.
  - $\boldsymbol{x_i}$: rate of transmitted packets over path d_i
  - Sum of x_i can be greater than 1
  - *R and S:* power set of the paths

$$\max \; U = \overbrace{\sum_{R_j \in R} \alpha \, p_j y_j}^{\text{Reliability}} - \overbrace{\sum_{S_j \in S} (1-\alpha) \, q_j z_j}^{\text{Vulnerability}}$$

Weighted sum

$$y_j \leq \sum_{d_i \in R_j} x_i \quad \forall \; R_j \in R$$

$\boldsymbol{y_j}$: Boolean variable to show if packets transmitted over paths in <u>*Rj*</u> suffice for decoding <u>by destination</u>

$$z_j \leq \sum_{d_i \in S_j} x_i \quad \forall \; S_j \in S$$

$\boldsymbol{z_j}$: Boolean variable to show if packets transmitted over paths in <u>*Sj*</u> suffice for decoding <u>by eavesdropper</u>

$$y_j, z_j \in \{0,1\} \quad \forall \; R_j \in R, S_j \in S$$

Problem Formulation- Case 2

- We set reliability threshold as a constraint.
- We then minimize the eavesdropping probability.

$$\min \; U = \sum_{S_j \in S} q_j z_j$$

Minimizing prob. of successful eavesdropping

$$s.t \quad \sum_{R_j \in R} p_j y_j \geq t$$

Reliability threshold *t*

$$y_j \leq \sum_{d_i \in R_j} x_i \quad \forall \; R_j \in R$$

$$z_j \leq \sum_{d_i \in S_j} x_i \quad \forall \; S_j \in S$$

$$y_j, z_j \in \{0, 1\} \quad \forall R_j \in R, \;\; S_j \in S$$

Problem Formulation- Case 3

- This is the reverse of Case 2.
  - We set eavesdropping prob. threshold as a constraint.
  - We maximize the reliability.

$$\max \ U = \sum_{R_j \in R} p_j y_j$$ Maximizing the reliability

$$s.t \quad \sum_{R_j \in R} q_j z_j \leq t$$ Security threshold *t*

$$y_j \leq \sum_{d_i \in R_j} x_i \quad \forall \ R_j \in R$$

$$z_j \leq \sum_{d_i \in S_j} x_i \quad \forall \ S_j \in S$$

$$y_j, z_j \in \{0, 1\} \quad \forall \ R_j \in R, S_j \in S$$

Relaxation to Linear Programming, Case 1 (LP)

- NP-complete
  - mixed integer and linear programming optimizations
- Modifying the integer variables to real variables

$$\max\ U = \sum_{R_j \in R} \alpha\, p_j y_j - \sum_{S_j \in S} (1-\alpha)\, q_j z_j$$

$$y_j \leq \sum_{d_i \in R_j} x_i \quad \forall\ R_j \in R$$

Relaxing integer variables to real

$$z_j \leq \sum_{d_i \in S_j} x_i \quad \forall\ S_j \in S$$

$$y_j, z_j \in (0, 1) \quad \forall\ R_j \in R, S_j \in S$$

Heuristic Solution- HR

- Complexity of the relaxed linear programming
  - Liner to the number of variables and constraints
  - With n paths, there are 2^n possible failure/eavesdropping cases
- Heuristic
  - Distribution of the transmissions <u>proportional</u> to the <u>failure rate</u> and <u>eavesdropping prob.</u> of the paths

$$\max\ U = \sum_{d_i \in D} \Big[\overbrace{\alpha \underbrace{(1-\epsilon_i)}_{\text{Reliability of the } i\text{th path}} x_i}^{\text{Reward}} - \overbrace{(1-\alpha) \underset{\substack{\downarrow \\ \text{Eavesdropping prob. of the } i\text{th path}}}{\gamma_i} x_i}^{\text{Punishment}} \Big]$$

$$s.t \ \sum_{i=1}^{n} x_i \geq 1$$

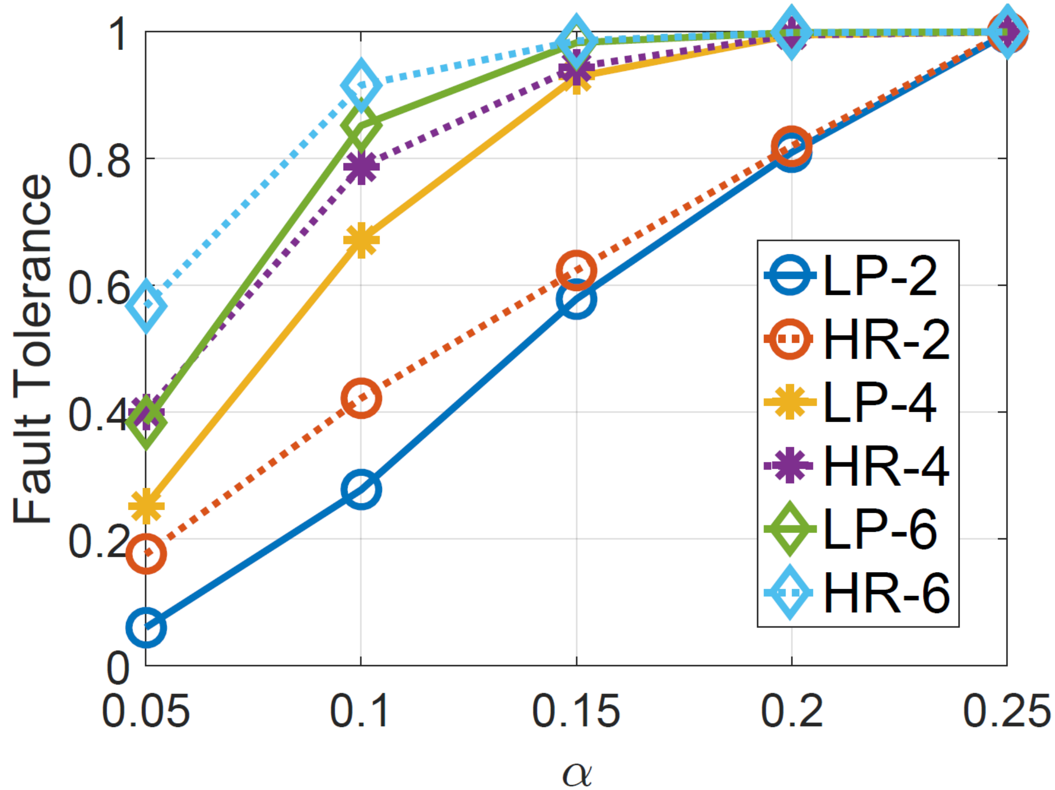
$$x_i \in (0,1) \quad \forall\ d_i \in D$$

Evaluations

- Simulator in Matlab environment
- We use Linprog tool of Matlab to find the solution of the optimizations
- 100 simulation runs
- Two settings
  - LP-*n*: relaxed optimization case 1(linear programming) with *n* paths
  - HR-*n*: heuristic solution with *n* paths

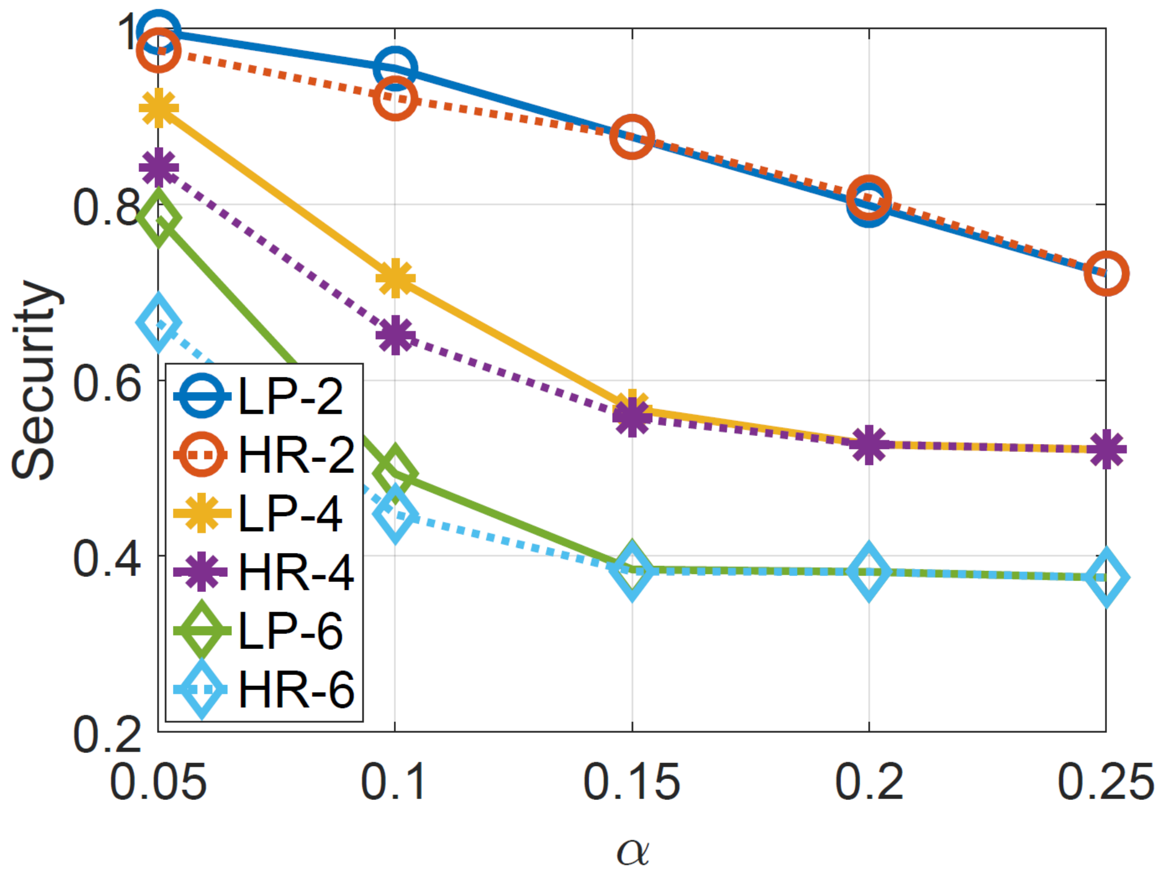
Evaluations- FT

- Path failure prob. of each path: [0,0.1]
- Eavesdropping prob. of each path: [0,0.3]



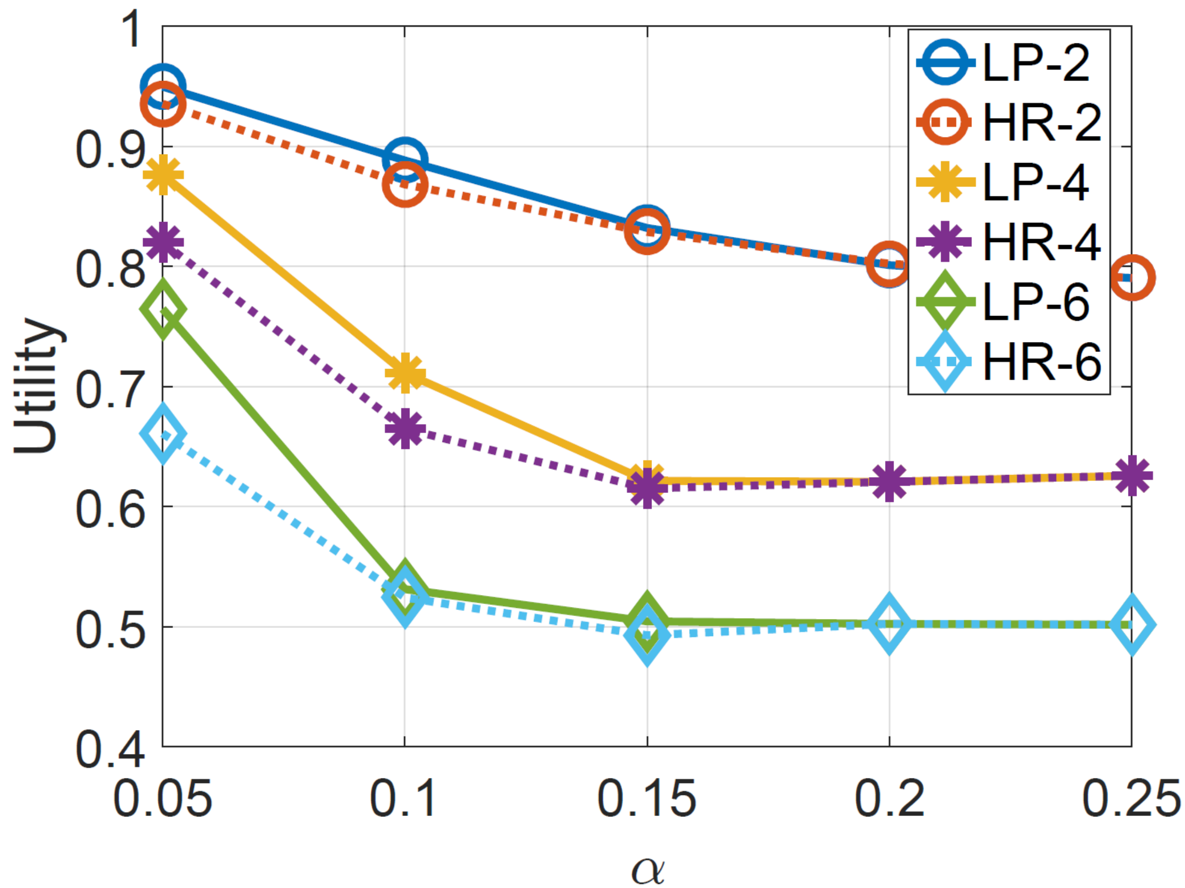
- Reliability of heuristic (HR) is close to LP
- HR over-estimates the reliability
- More paths enhances the reliability

Evaluations- Security



- Security of HR is close to LP
- HR under-estimates the security
- More paths reduces the security

Evaluations- Utility



- The utility of HR and LP is close
- More paths reduces utility (because of the higher eavesdropping prob. selected compared to the path failure prob.)

Future Work

- Using the idea of critical path
  - Finding a critical path in a general graph
- Impact of multi-path on FT and security
  - More realistic and heterogeneous prob. distributions
- Impact of correlation
  - Failure prob. and eavesdropping prob.

Thank you