

Designing Overlay Multicast Networks for Commercial Streaming

Konstantin Andreev

Bruce Maggs

Adam Meyerson

Ramesh Sitaraman

Differences for Streaming



- Player rather than browser
- Streaming server rather than web server
- Feed from live event must be distributed to server
- Sufficient bandwidth must be consistently available

Stream Quality: What to Measure?



- Metrics should capture important elements of user experience.
- Focused, simple, universal, and objective.

Our Streaming Quality Metrics

Start Up


Play

Freeze

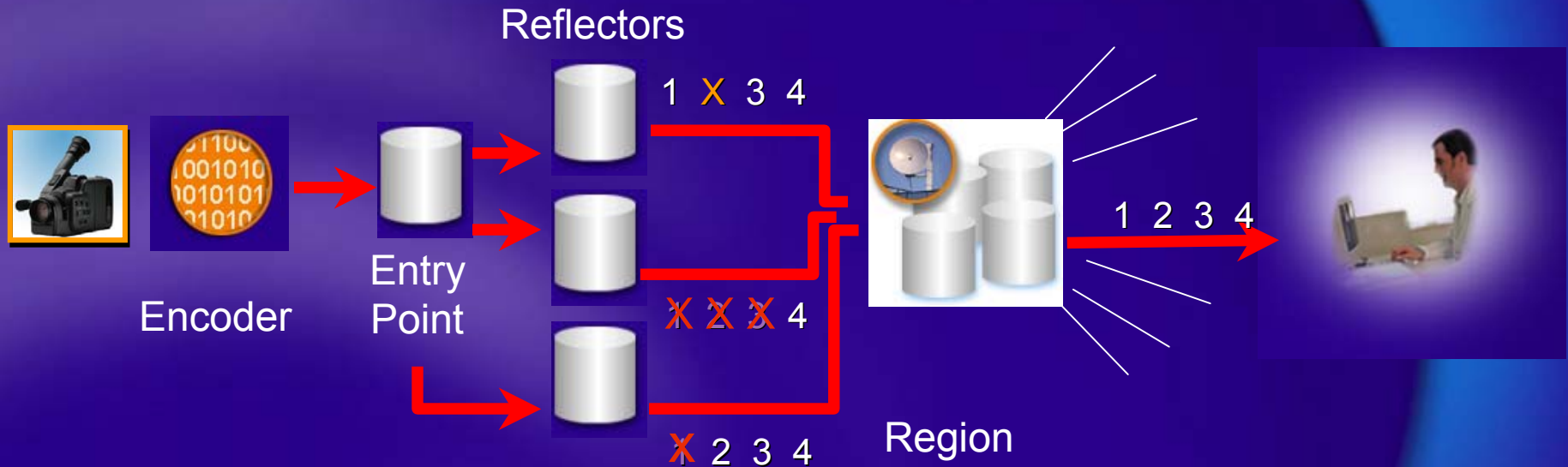
Play

Freeze

Play

- 
- A horizontal timeline diagram with a white arrow pointing to the right. The timeline is divided into segments by vertical lines. The segments are labeled 'Start Up', 'Play', 'Freeze', 'Play', 'Freeze', and 'Play' from left to right. The 'Play' segments are highlighted in red, and the 'Freeze' segments are highlighted in green.
- I. **Failure Rate:** Rate at which streams fail to play.
 - II. **Startup Time:** Time to start playing after user hits “play”.
 - III. **Thinning and Loss Rate:** Reduction in “playback bandwidth”.
 - Playback Bwth (PB) = Useful Bits/Stream Time, where Useless = Thinned, or Unrecoverably lost, or Arrives late.
 - Thinning & Loss Rate = (Ideal PB – Actual PB)/Ideal PB.
 - IV. **Interruptions:** Rebuffers per min, Rebuffer Time per min.

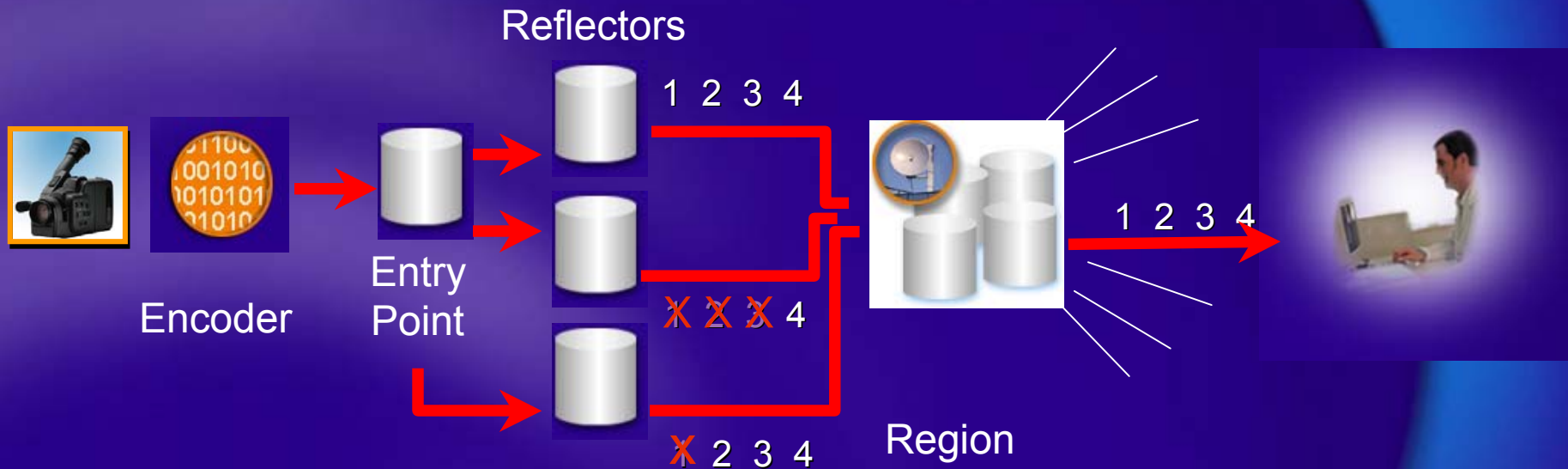
Live Streaming Architecture



End-to-End Metrics: Minimize loss, “lateness”, cost.

1. **Multi-Path**: Detect loss, pull more copies, reconstruct clean copy.
2. **Information dispersal** across multiple paths to reduce overhead.
 - Example, Odd-Even-XOR, Reed-Solomon, etc.

Live Streaming Architecture



3. **Fast failover** to alternate path without disrupting stream to the end-user.
4. **Link-level recovery**: Retransmits and Forward Error Correction.

Modeling Server Load and Capacity

Simple models such raw machine BW do not work!

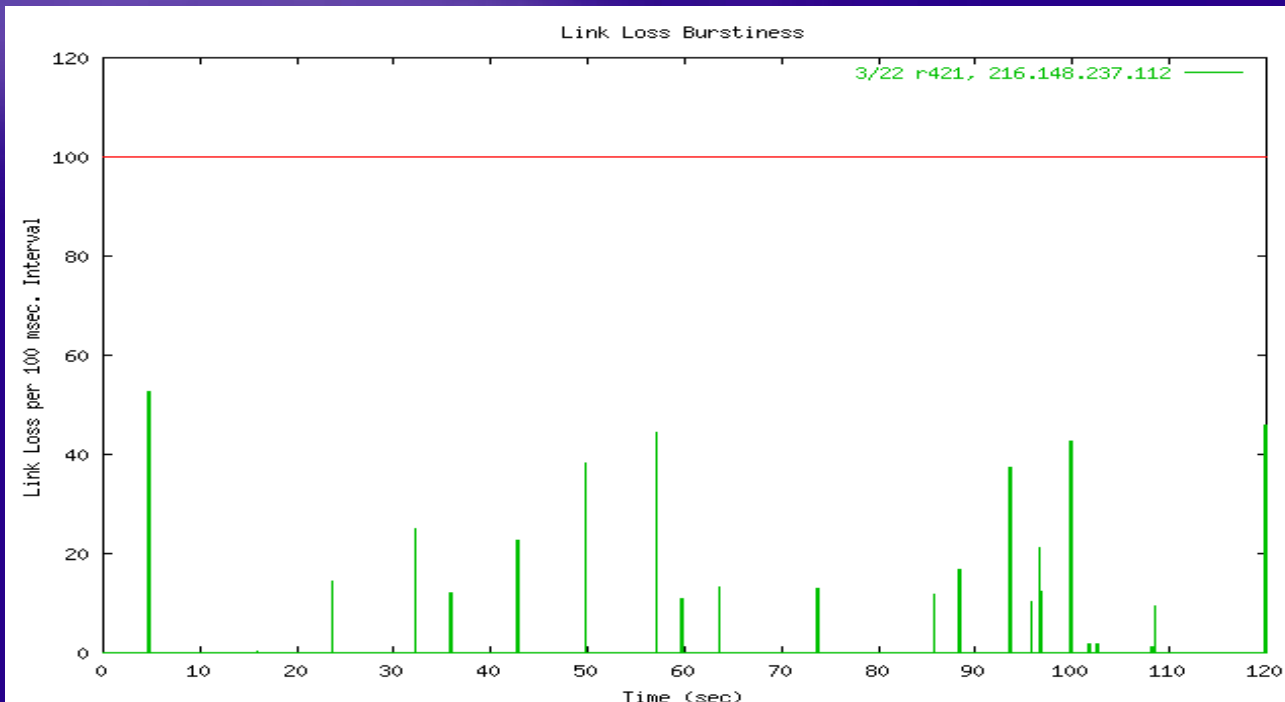
Stream	BW Limit (one distinct)	BW Limit (all distinct)	Factor
300K	62 Mbps	14 Mbps	4.5
100K	60 Mbps	8 Mbps	7.5
36K	54 Mbps	3.9 Mbps	14.0

- Capacity = multi-dimensional polytope.
- Dimensions include bandwidth, #conns, #distinct streams



Link-Level Recovery: FEC

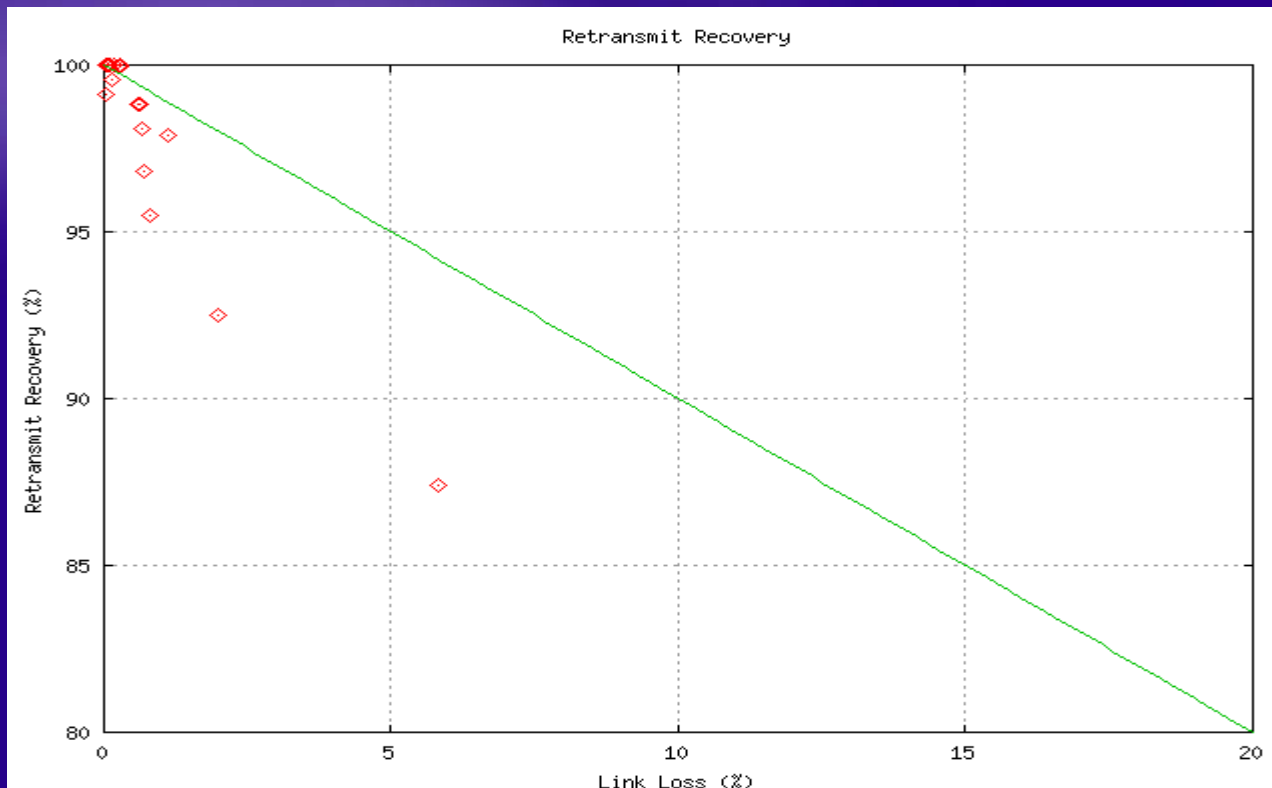
Loss is bursty on the Internet, so simple parity performs poorly. Less than 50% loss recovery even if cost overhead is 33%! More sophisticated FEC schemes work better, e.g., interleaved parity.



Link loss =
0.6%

Link-Level Recovery: Retransmits

Good bang for the buck, esp. on low latency links. Low bwth overhead – it is proportional to link loss rate. High loss recovery. Acceptable time to recovery.



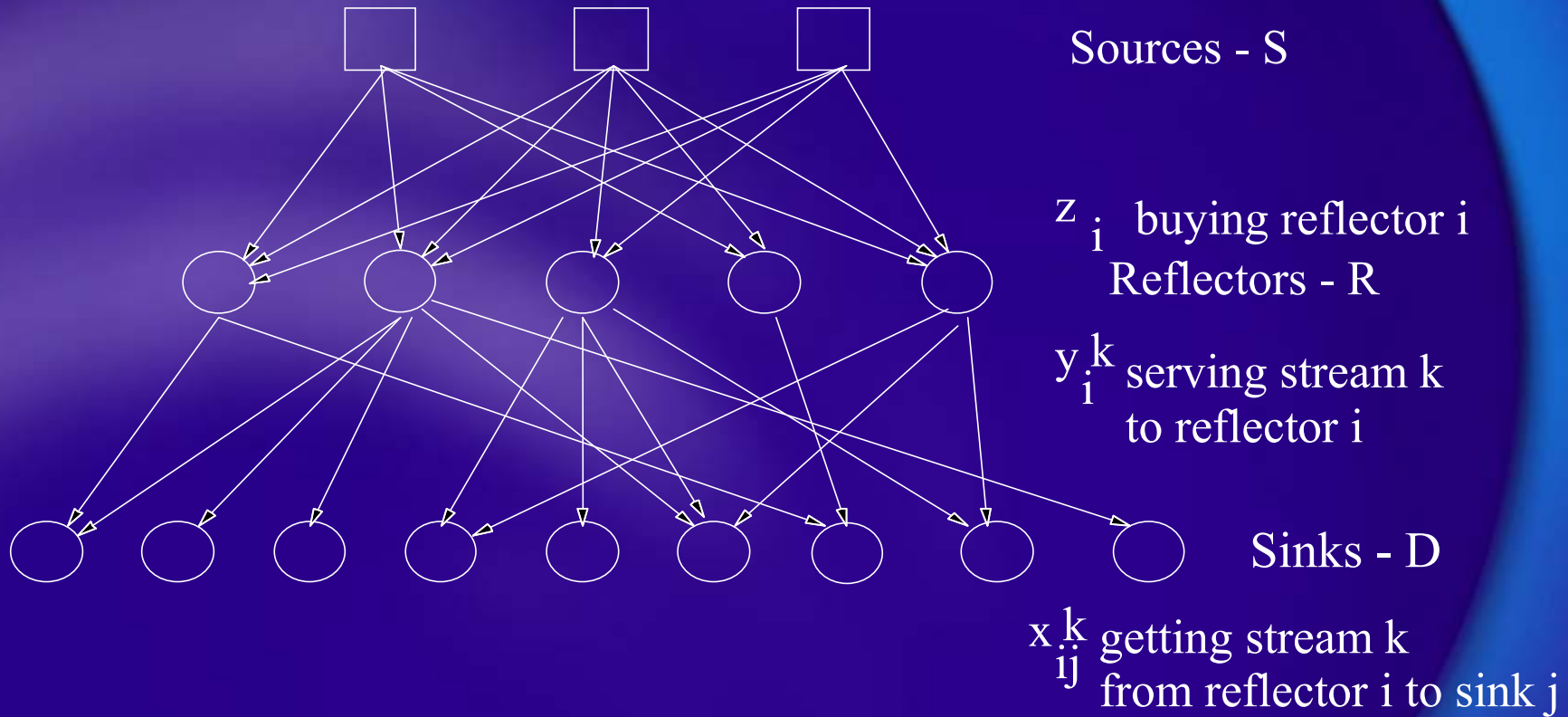
Goals when building the overlay multicast network

- Satisfy reliability requirements
- Minimize cost
- Obey capacity constraints

Packet loss model

- The algorithm receives as an input the probability of failure on each link and every packet on the link can be lost with that probability on average
- We assume loss of packets on different links are independent (in the extensions we consider a model in which some link losses are related)

Problem Setup



Formal definition

3-level network reliability min-cost multicommodity flow problem

➤ Tripartite digraph $V = S \cup R \cup D$

➤ Costs on the edges
(depending on the commodity) $c^u: E^u \rightarrow \mathbb{R}^u$

➤ Cost of using a reflector $r: R \rightarrow \mathbb{R}$

➤ Fanout constraint on reflectors $F: R \rightarrow \mathbb{R}$

➤ Probability of failure on the edges
 $p: E \rightarrow [0, 1]$

➤ Demand thresholds for each destination and commodity pair
 $\Phi^u: D^u \rightarrow [0, 1]^u$

Here u is the number of commodities



Problem transformation

- Converting probabilities into weights*
- Edge weights

$$w_{ij}^k = -\log(p_{ki} + p_{ij} - p_{ki}p_{ij})$$

- Demand weights

$$W_j^k = -\log(1 - \Phi_j^k)$$

- *Disclaimer: WLOG we can assume that each source serves only one stream and each sink demands only one commodity.

Previous related work

- General framework of non-metric facility location
- We can encode Set Cover, thus no approximation better than $\log n$ on the cost
- We want to cover sinks with a commodity from multiple reflectors.
- Similar to facility location with redundancy (but non-metric) and Weighted Set Cover (but coverage depends on the set, element pair)

More previous work

- We are also constructing a fault tolerant network
- For general networks, the Network Reliability Problem is #P-complete.
- There is an FPRAS that approximates it with in $1+\epsilon$
- In 3-level networks one can compute the exact reliability in polynomial time.

IP formulation/LP relaxation

$$\min \sum_{i \in R} r_i z_i + \sum_{i \in R} \sum_{k \in S} c_i^k y_i^k + \sum_{i \in R} \sum_{k \in S} \sum_{j \in D} c_{ij}^k x_{ij}^k$$

s.t.

1. $y_i^k \leq z_i \quad \forall i \in R, k \in S$
2. $x_{ij}^k \leq y_i^k \quad \forall i \in R, j \in D, k \in S$
3. $\sum_{k \in S, j \in D} x_{ij}^k \leq F_i z_i \quad \forall i \in R$
4. $\sum_{j \in D} x_{ij}^k \leq F_i y_i^k \quad \forall i \in R, k \in S$
5. $\sum_{i \in R} x_{ij}^k w_{ij}^k \geq W_j^k \quad \forall j \in D, k \in S$
6. $x_{ij}^k \in \{0, 1\}, y_i^k \in \{0, 1\}, z_i \in \{0, 1\}$

Phases of the solution

- Relaxation / Randomized rounding
- Modified *GAP* conversion
- In the extensions: Srinivasan and Teo technique

Randomized rounding analysis

- The expected cost after the rounding is at most $(c \log n) \times C^{\text{OPT}}$
- Weight constraint violation

$$\Pr \left(\sum_{i \in R} w_{ij}^k \bar{x}_{ij}^k \leq (1 - \delta) W_j^k \right) \leq \frac{1}{n^{\delta^2 \cdot c/2}}$$

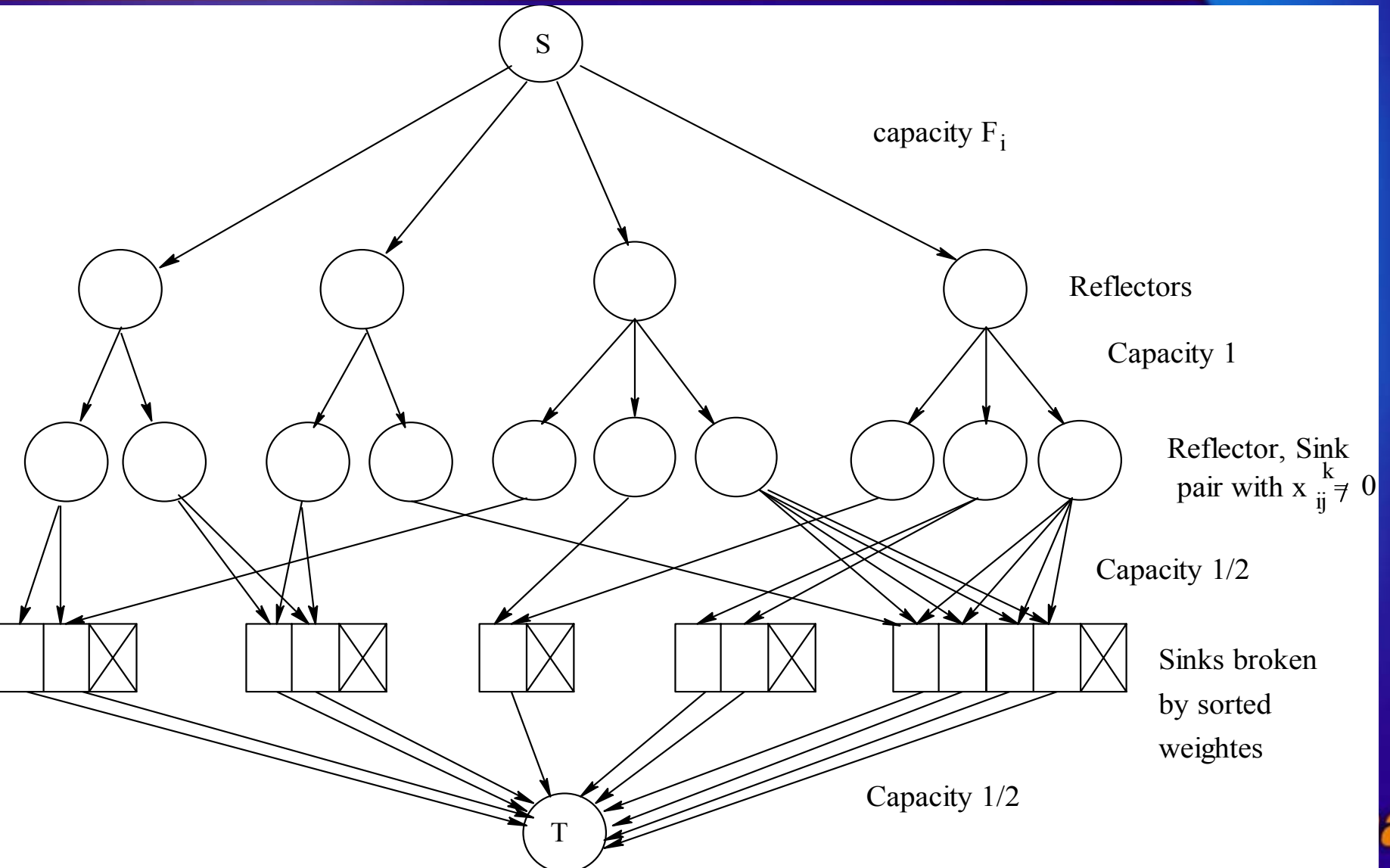
- Fanout constraint violation if $c \leq 24$

$$\Pr \left(\sum_{k \in S} \sum_{j \in D} \bar{x}_{ij}^k \geq 2F_i \right) \leq \frac{1}{2n^2}$$

Second phase

- The only fractional variables left after the randomized rounding are \bar{x}_{ij}^k
- To round those we apply modified Generalized Assignment Problem conversion similar to Shmoys and Tardos
- It doubles cost and violates weight and fanout constraints by a factor of two

Modified GAP Approximation



GAP approximation analysis

- There exists an optimal flow with values only 0, $\frac{1}{2}$ and 1 (Shmoys and Tardos).
- We double this solution
- We violate the fanout and capacity constraints by at most a factor of 2

Running time

- Let $|S| = \#$ of streams
- $|D| = \#$ of (stream, sink) pairs
- The initial LP has $O(|S| \cdot |R| \cdot |D|)$ variables and constraints
- The randomized rounding procedure takes as many iterations as the number of LP variables
- The modified GAP network has $O(|R| \cdot |D|)$ nodes and edges
- The running time of solving the GAP flow problem is absorbed by the LP solver running time.

Extensions

- Constraints added to the IP
- Capacities between reflectors and sinks

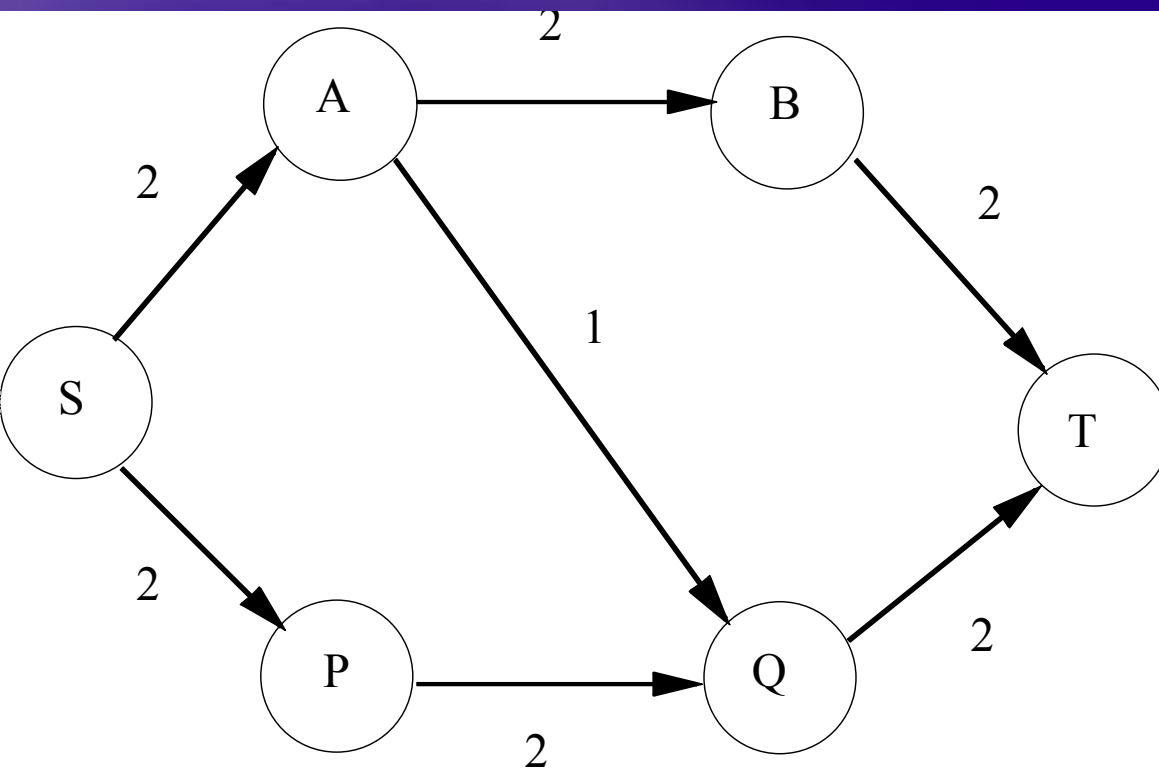
$$\sum_{k \in S} x_{ij}^k \leq u_{ij} \quad \forall i \in R, j \in D$$

- Color constraints

$$\sum_{i \in R_\ell} x_{ij}^k \leq 1 \quad \forall j \in D, k \in S, \ell \in [m]$$

where $R = R_1 \cup R_2 \cup \dots \cup R_m$

Flow problem with additional set constraints



- Additional set capacity constraint
- E.g. $\{ab, pq\}$ has capacity 3
- There is an LP/IP gap
- E.g. Fractional flow is 3.5, best integral flow is 3

Flow problem setup

- We find an integral solution within a constant factor of optimal that violates the constraints by at most a constant
- In order to do that we reformulate the flow problem in terms of paths

Flow problem solution

- Theorem: Let A be a real valued $r \times s$ matrix and y be an s -vector. Assume that in every column of A
 1. the sum of all positive entries is at most t
 2. the sum of all negative entries is at least $-t$Then we can round the solution to $Ay=b$ component wise up or down (say y' is the rounded integral vector) in such a way that $Ay'=b'$ where $b'_i - b_i < t$
- In the path reformulation the sum of the coefficient in front of y_p is at most 7
- The running time is at most $O(|R|^3 \cdot |D|^3)$