Traffic-Redundancy Aware Network Design

Siddharth Barman

Joint work with Shuchi Chawla

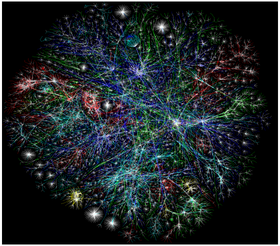
University of Wisconsin–Madison

January 18, 2012

Internet is "a series of tubes"

Commodity Network

versus

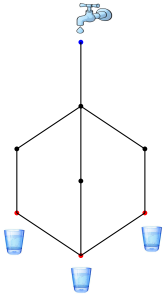


Internet

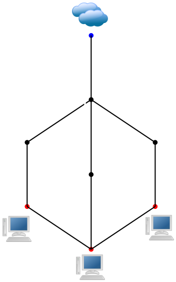
- In contrast to physical commodity networks, information networks carry data, which can be easily duplicated, compressed, and combined.
- Classical network-design models represent physical flow.

Routing Physical Flow vs. Routing Data

Objective: Route flow from the source to all the sinks at a low cost.



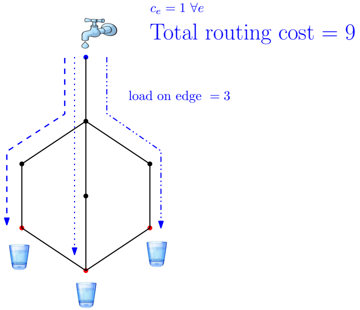
Objective: *Multicast* data from the source to all the sinks at a low cost.



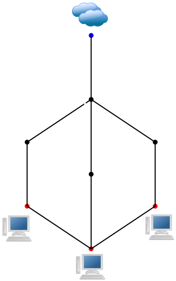
Routing cost $= \sum_{\text{edges}}$ Cost of the edge $\times$ Load on the edge

Routing Physical Flow vs. Routing Data

Objective: Route flow from the source to all the sinks at a low cost.



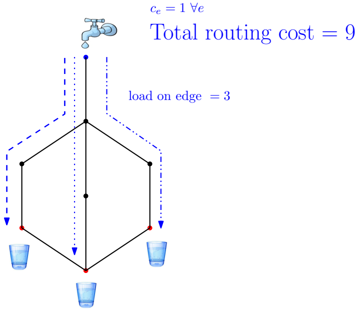
Objective: *Multicast* data from the source to all the sinks at a low cost.



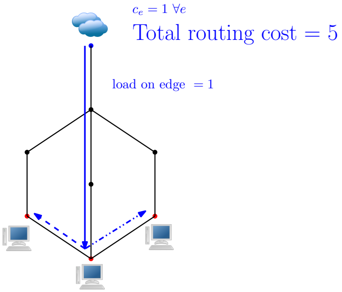
Routing cost $= \sum_{\text{edges}}$ Cost of the edge $\times$ Load on the edge

Routing Physical Flow vs. Routing Data

Objective: Route flow from the source to all the sinks at a low cost.



Objective: *Multicast* data from the source to all the sinks at a low cost.



Routing cost $= \sum_{\text{edges}}$ Cost of the edge $\times$ Load on the edge

Overarching Goals

- Develop expressive and tractable combinatorial models to represent information flow.
- In particular, focus on a cost structure that captures savings obtained by eliminating redundancy in data.
- Design approximation algorithms within this framework.

Outline

- Related Work
- Redundancy Aware Network Design
- Redundancy Aware Facility Location

Related Work

- Shmoys-Swamy-Levi '04: $O(1)$ approximation algorithm for facility location with service installation costs.
- Svitkina-Tardos '06: $O(1)$ approximation algorithm for facility location with hierarchical costs.
- Hayrapetyan-Swamy-Tardos '05 considered submodular costs on edges.
  - $O(\log |V|)$ approximation for single-source network design.

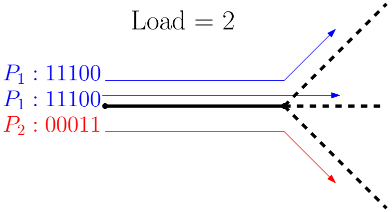
Outline

- Related Work
- Redundancy Aware Network Design
- Redundancy Aware Facility Location

Redundancy Aware Network Design

- We focus on redundant-data elimination.
- Apply the following cost structure:

Load on an edge $l_e = \#$ **distinct** data packets on it.

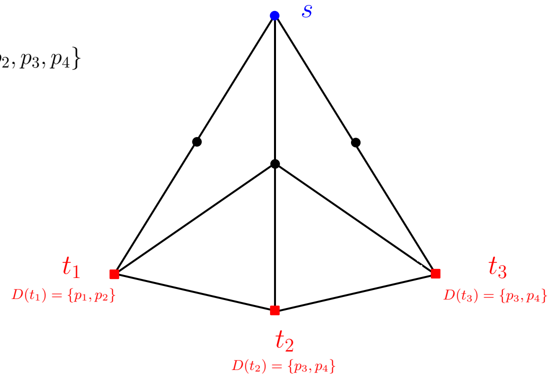


Problem Definition

Input:

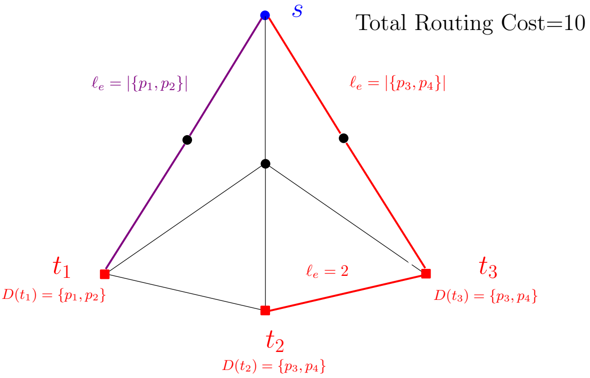
- Graph with cost on edges, a universal source s and a set of terminals (sinks) $\{t\}$.
- Each terminal t has a demand set $D(t)$ of packets.
- Global set of **distinct** packets $\Pi = \cup_t D(t)$.

$\Pi = \{p_1, p_2, p_3, p_4\}$



Problem Definition

- **Goal:** Connect each t to s and route packets in $D(t)$ on the connecting path.
- **Objective:** Minimize total routing cost: $\sum_{\text{edges}} \text{Cost of edge} \times \text{Load on edge}$



Problem Definition

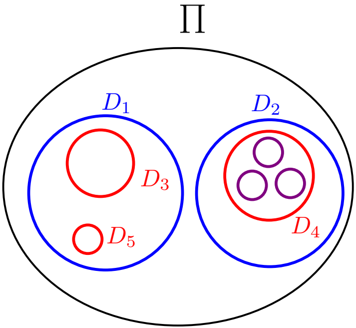
- **Input:**
  - Graph with cost on edges and a universal source s.
  - A global set of packets Π.
  - Each terminal t has a demand set $D(t) \subset \Pi$.
- **Goal:** Connect each t to s and route packets in $D(t)$ on the connecting path.
- **Objective:** Minimize total routing cost: $\sum_{\text{edges}} \text{Cost of edge} \times \text{Load on edge}$

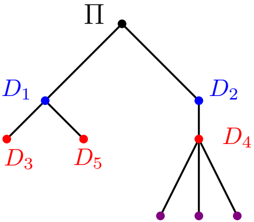
Result

We develop an algorithm that achieves an $O(\log |\Pi|)$ approximation when the set of demands, $\{D(t)\}_t$, is laminar.

Laminar Set of Demands

The demand family $\{D(t)\}_t$ is said to be **laminar** if $\forall\ i, j$ one of the following holds: $D_i \subset D_j$ or $D_j \subset D_i$ or $D_i \cap D_j = \phi$.

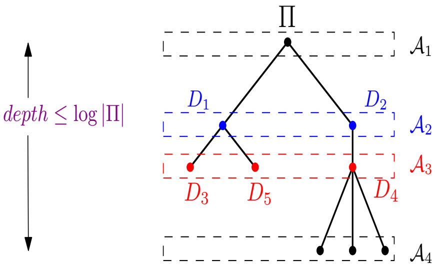




$O(\log|\Pi|)$ Approximation

Algorithm when the packets have uniform weight:

1. Depth Reduction: Transform demand tree by merging demand sets that are within a factor of two ($|D(child)| \leq \frac{1}{2}|D(parent)|$).



$\mathcal{A}_2 = \text{Steiner}(D_1) + \text{Steiner}(D_2)$

$\mathcal{A}_3 = \text{Steiner}(D_3) + \text{Steiner}(D_4) + \text{Steiner}(D_5)$

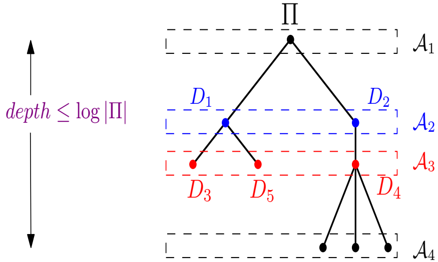
$\mathcal{A}_i \leq O(1)\ \text{OPT}\ \forall i$

$\sum_i \mathcal{A}_i \leq O(depth) \times \text{OPT}$

$O(\log|\Pi|)$ Approximation

Algorithm when the packets have uniform weight:

2. Composition: Solve for each level of the demand tree independently by constructing low cost Steiner trees. Take the union of the solutions.



$\mathcal{A}_2 = \text{Steiner}(D_1) + \text{Steiner}(D_2)$

$\mathcal{A}_3 = \text{Steiner}(D_3) + \text{Steiner}(D_4) + \text{Steiner}(D_5)$

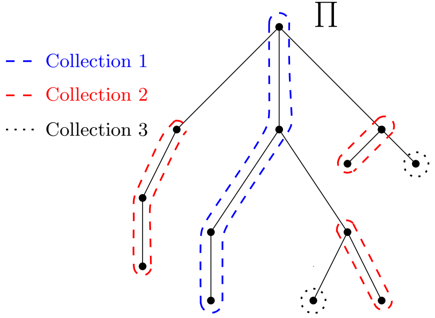
$\mathcal{A}_i \leq O(1)\ \text{OPT}\ \forall i$

$\sum_i \mathcal{A}_i \leq O(depth) \times \text{OPT}$

Extension: Packets with Weights

- Load on an edge = Sum of the weights of distinct packets through it.
- Transform demand tree s.t. weight(child) $\leq \frac{1}{2}$ weight(parent).

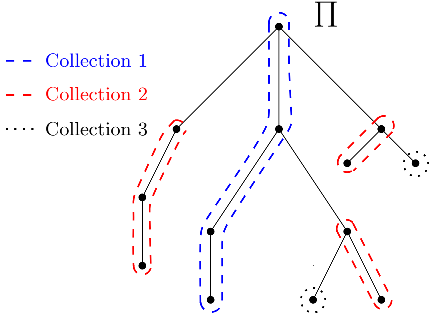
Partition the demand tree into $O(\log |\Pi|)$ collections each containing disjoint descending paths.



Extension: Packets with Weights

- Load on an edge $=$ Sum of the weights of distinct packets through it.
- Transform demand tree s.t. weight(child) $\leq \frac{1}{2}$ weight(parent).

Partition the demand tree into $O(\log|\Pi|)$ collections each containing disjoint descending paths.



Result

We achieve an $O(\log|\Pi|)$-approximation ratio for the weighted case as well.

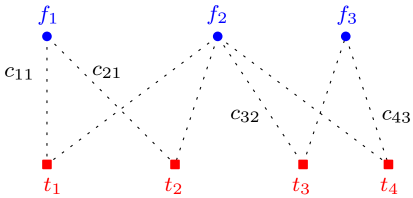
Outline

- Related Work
- Redundancy Aware Network Design
- Redundancy Aware Facility Location

Redundancy Aware Facility Location

- **Input:**
  - A set of facilities $\{f_i\}$ (with production costs λ_i) and a set of terminals $\{t\}$ with connection costs between them.
  - Each terminal has a demand set: $D(t)$ for terminal t.

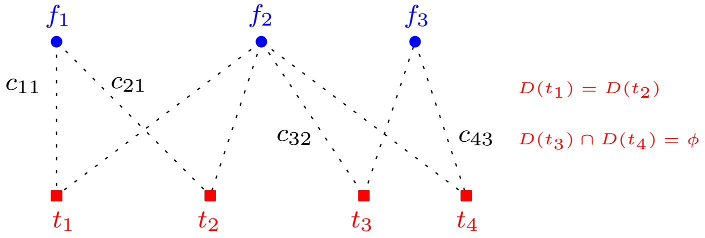
Facility opening cost at $f_i = \lambda_i \times$ Number of distinct packets produced at f_i



Redundancy Aware Facility Location

- **Goal:** Connect each terminal to a facility and produce its demand set there.
- **Objective:** Minimize the total facility opening and routing cost.

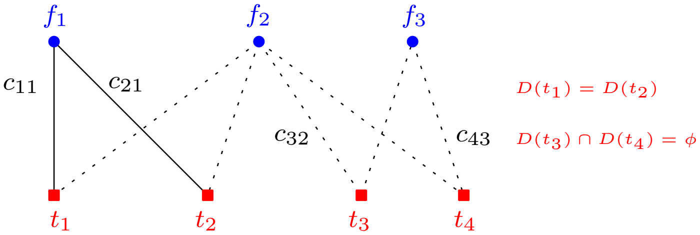
Redundancy Elimination: At any facility it suffices to produce one copy of each desired packet.



Redundancy Aware Facility Location

- **Goal:** Connect each terminal to a facility and produce its demand set there.
- **Objective:** Minimize the total facility opening and routing cost.

Redundancy Elimination: At any facility it suffices to produce one copy of each desired packet.



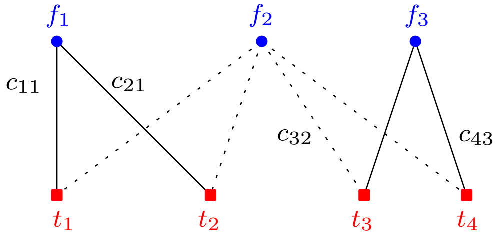
Facility Opening Cost: $\lambda_1 \, |D(t_1)|$

Routing Cost: $c_{11}|D(t_1)| + c_{21}|D(t_2)|$

Redundancy Aware Facility Location

- **Goal:** Connect each terminal to a facility and produce its demand set there.
- **Objective:** Minimize the total facility opening and routing cost.

Redundancy Elimination: At any facility it suffices to produce one copy of each desired packet.



$D(t_1) = D(t_2)$

$D(t_3) \cap D(t_4) = \phi$

Facility Opening Cost: $\lambda_1\,|D(t_1)| + \lambda_3\,|D(t_3)| + \lambda_3\,|D(t_4)|$

Routing Cost: $c_{11}|D(t_1)| + c_{21}|D(t_2)| + c_{33}|D(t_3)| + c_{43}|D(t_4)|$

Integer Programming Formulation

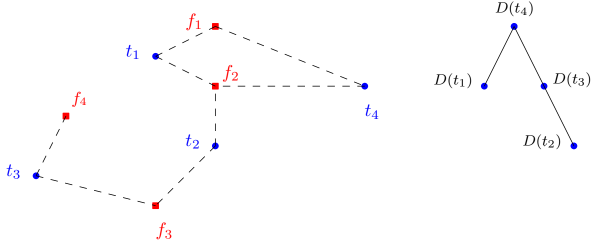
$$
\begin{aligned}
\text{minimize} \quad & \sum_f \sum_p \lambda_f \, y_{f,p} + \sum_t \left(\sum_f |D(t)| x_{t,f} \; c(t,f) \right) \\
\text{subject to} \quad & \sum_f x_{t,f} \geq 1 \quad \forall t \in T \\
& y_{f,p} \geq x_{t,f} \quad \forall t, f, p \in D(t) \\
& x_{t,f}, y_{f,p} \in \{0,1\}
\end{aligned}
$$

$x_{t,f} = 1$ iff terminal t connected to facility f
$y_{f,p} = 1$ iff packet p is produced at facility f

Constant-Factor Approximation for Laminar Demands

Multi-phase LP rounding algorithm for uniform facility opening costs.

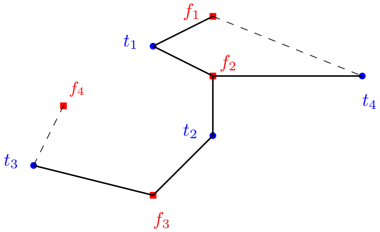
1. Filtering: Each terminal is fractionally connected to facilities that are at a distance at most twice the fractional optimal.

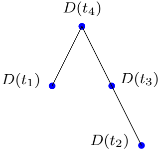


Constant-Factor Approximation for Laminar Demands

Multi-phase LP rounding algorithm for uniform facility opening costs.

2. Temporary Assignment: Consider terminals in increasing order of connection cost and temporarily assign facilities to them.

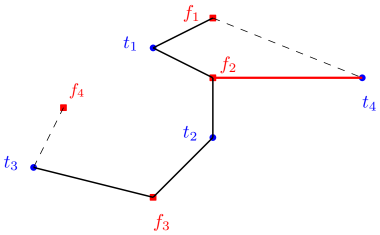


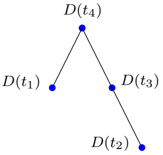


Constant-Factor Approximation for Laminar Demands

Multi-phase LP rounding algorithm for uniform facility opening costs.

3. Permanent Assignment: Consider terminals in decreasing order of $|D(t)|$ and assign permanent facilities.

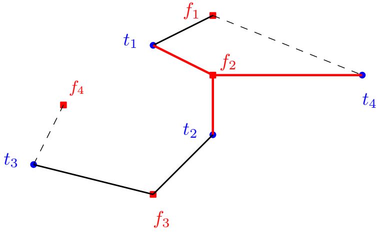


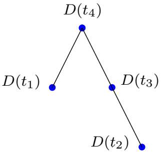


Constant-Factor Approximation for Laminar Demands

Multi-phase LP rounding algorithm for uniform facility opening costs.

3. Permanent Assignment: Consider terminals in decreasing order of $|D(t)|$ and assign permanent facilities.

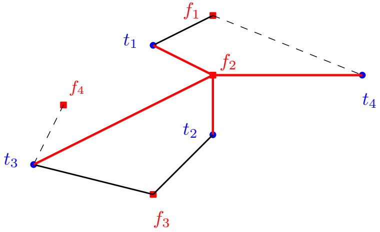


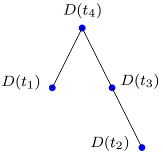


Constant-Factor Approximation for Laminar Demands

Multi-phase LP rounding algorithm for uniform facility opening costs.

3. Permanent Assignment: Consider terminals in decreasing order of $|D(t)|$ and assign permanent facilities.





Constant-Factor Approximation for Laminar Demands

We can extending the multi-phase rounding algorithm to account for production cost λ_i at facilities.

Constant-Factor Approximation for Laminar Demands

We can extending the multi-phase rounding algorithm to account for production cost λ_i at facilities.

Result

We obtain an approximation ratio of 27 for Redundancy Aware Facility Location.

Summary

- A combinatorial model which captures cost savings obtained by eliminating redundancy in data.
- $O(\log |\Pi|)$ approximation for Redundancy Aware Network Design
- 27 approximation for Redundancy Aware Facility Location

Questions?

Thanks!