

Internet Traffic Engineering by Optimizing OSPF/IS-IS Weights

Mikkel Thorup
AT&T Labs—Research

Joint work with

Bernard Fortz

Université Catholique de Louvain, Belgium

Shortst paths routing (OSPF/IS-IS)

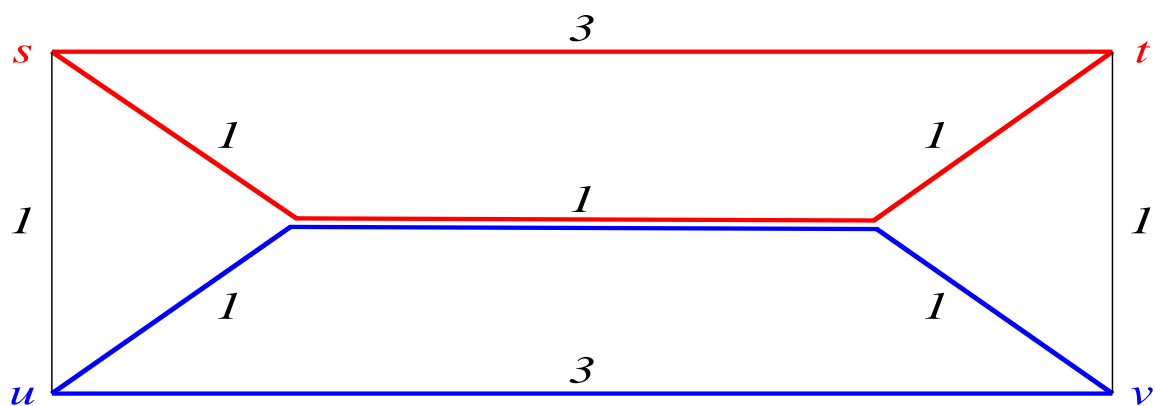
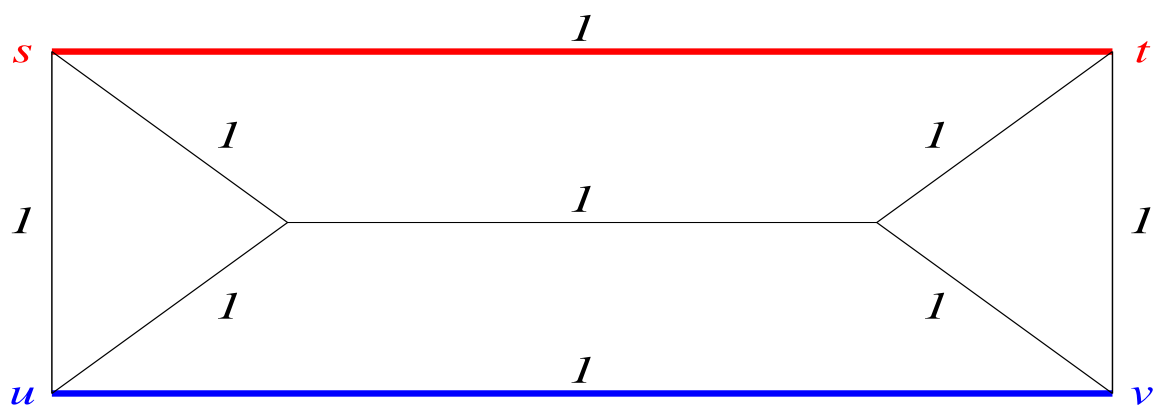
- Each link has an operator-specified weight.
- Each router knows all the links and weights.
- When a packet arrives at a router, it is sent out on the first link of a shortest (i.e., minimum weight) path to its destination (precomputed by the router).
- If several outgoing links on shortest paths to a destination, split traffic evenly.
- Weights are often set by simple rules of thumb. Frequently cited schemes:
 1. All equal
 2. Proportional to link length or air miles
 3. Inversely proportional to link capacity (Cisco default)
 4. *Ad hoc* tweaks to the above made in hopes of better balancing the traffic

OSPF Competitors

- **Multicommodity Flow Solution**
 - Complete freedom in distributing traffic on paths from sources to destinations
 - Requires knowledge of traffic matrix and ability to route packets fractionally
 - Useful lower bound on what's possible
- **MPLS**
 - If used for all traffic could approximate the multicommodity flow solution (assuming we know the traffic matrix)
 - Not likely to be used for all traffic
 - Not yet widely implemented/used

Demand of 1 for (s, t) and (u, v)

Shortest paths (OSPF/IS-IS): max load 1



General routing (MPLS): max load $2/3$



Basic Research Question

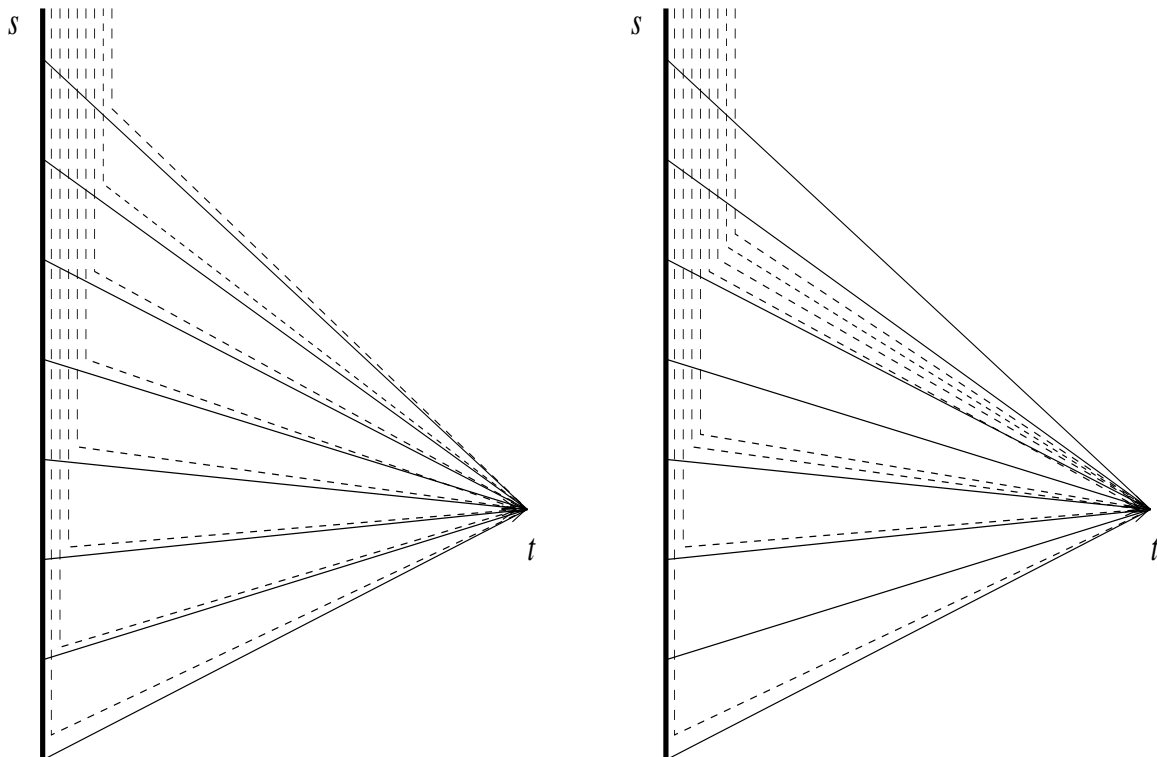
If one is allowed to tailor the weights to the traffic matrix, how well can OSPF do compared to

- multicommodity flow lower bound
- standard OSPF weight settings

in theory and practice?

Theory

Horrible gaps between MPLS and OSPF:



Moreover, it is NP-hard to find even near-optimal weight settings.

Practice:

Surprisingly good ...

Approach:

Use sophisticated local search heuristic **HeurOSPF** to determine good weights, given network and traffic matrix.

Quality Metrics

Performance guarantee

$\max \{utilization(e)\}$ over all links e

where $utilization(e) = \frac{load(e)}{capacity(e)}$

Best effort

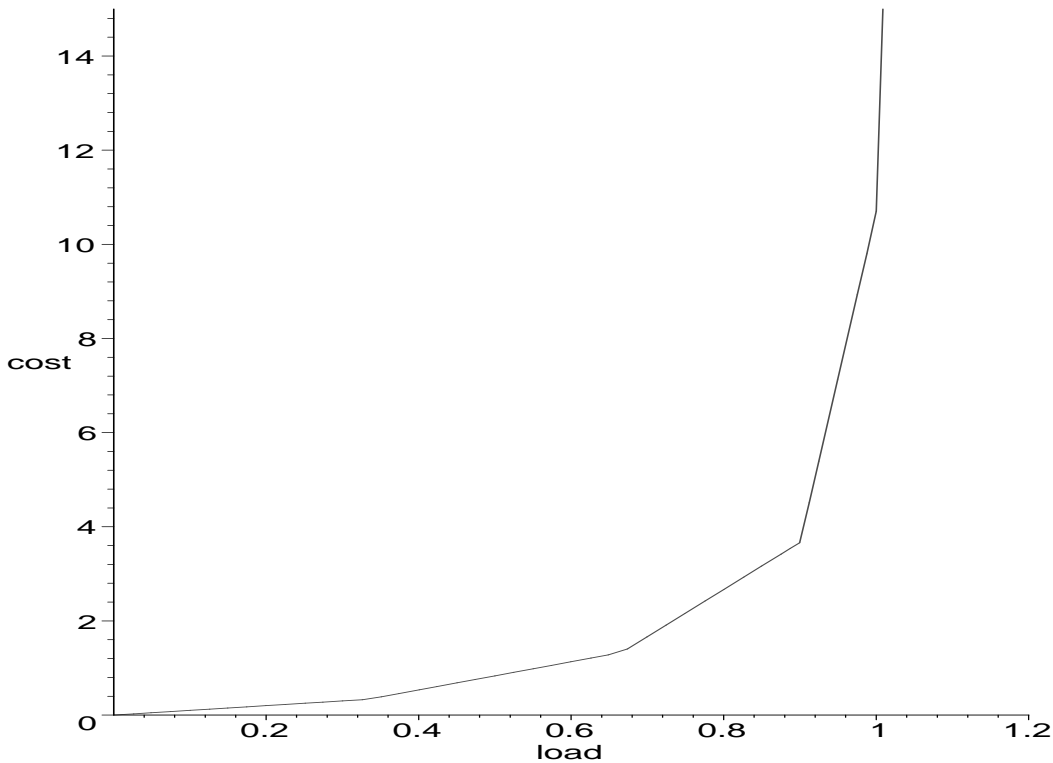
$$\sum_e \Phi_{capacity(e)}(load(e))$$

for a suitable penalty function Φ

Penalty function

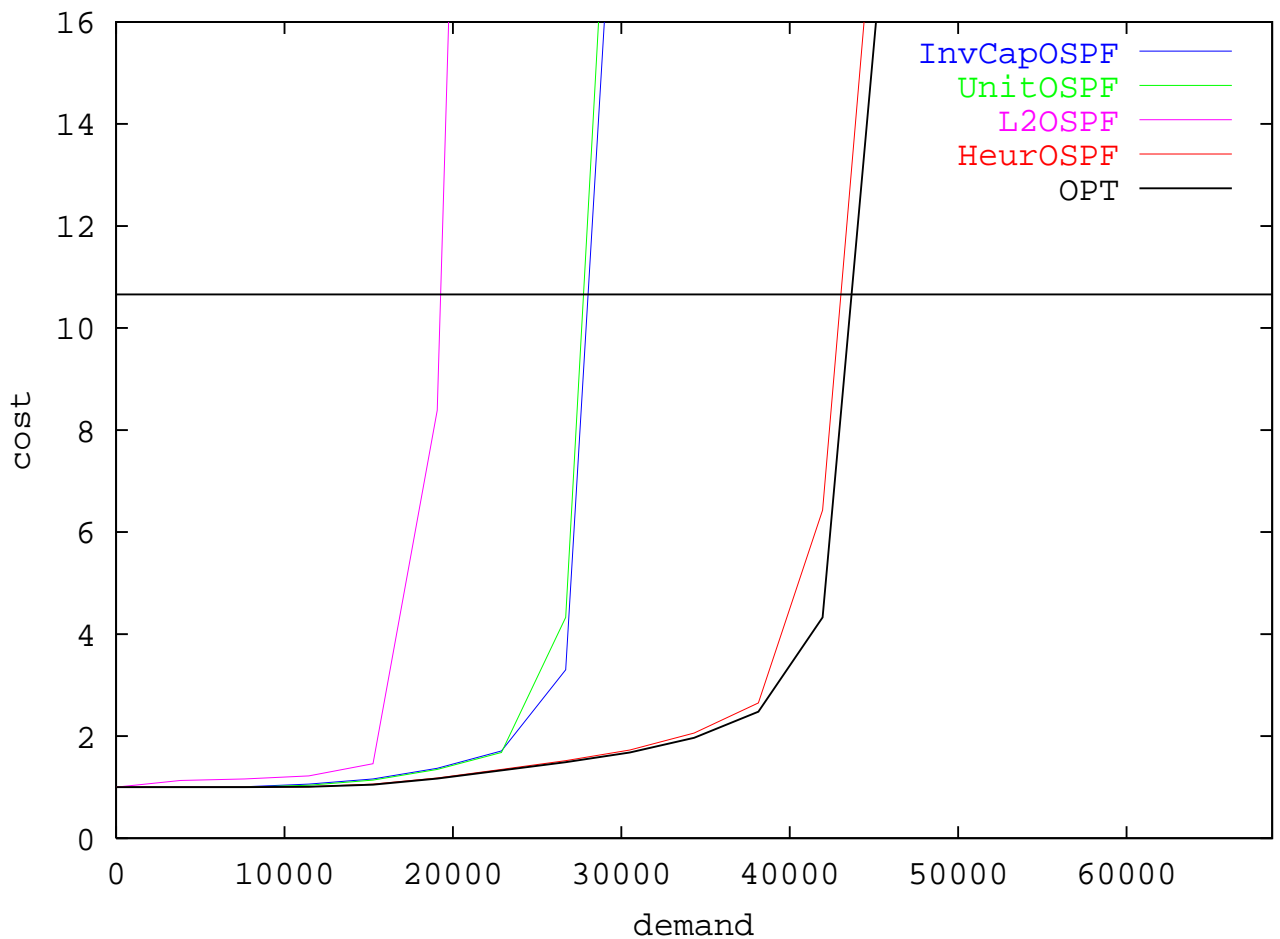
$$\Phi_c(0) = 0$$

$$\Phi'_c(\ell) = \begin{cases} 1 & \text{for } \ell/c \in [0, 1/3) \\ 3 & \text{for } \ell/c \in [1/3, 2/3) \\ 10 & \text{for } \ell/c \in [2/3, 9/10) \\ 70 & \text{for } \ell/c \in [9/10, 1) \\ 500 & \text{for } \ell/c \in [1, 11/10) \\ 5000 & \text{for } \ell/c \in [11/10, \infty) \end{cases}$$



AT&T WorldNet

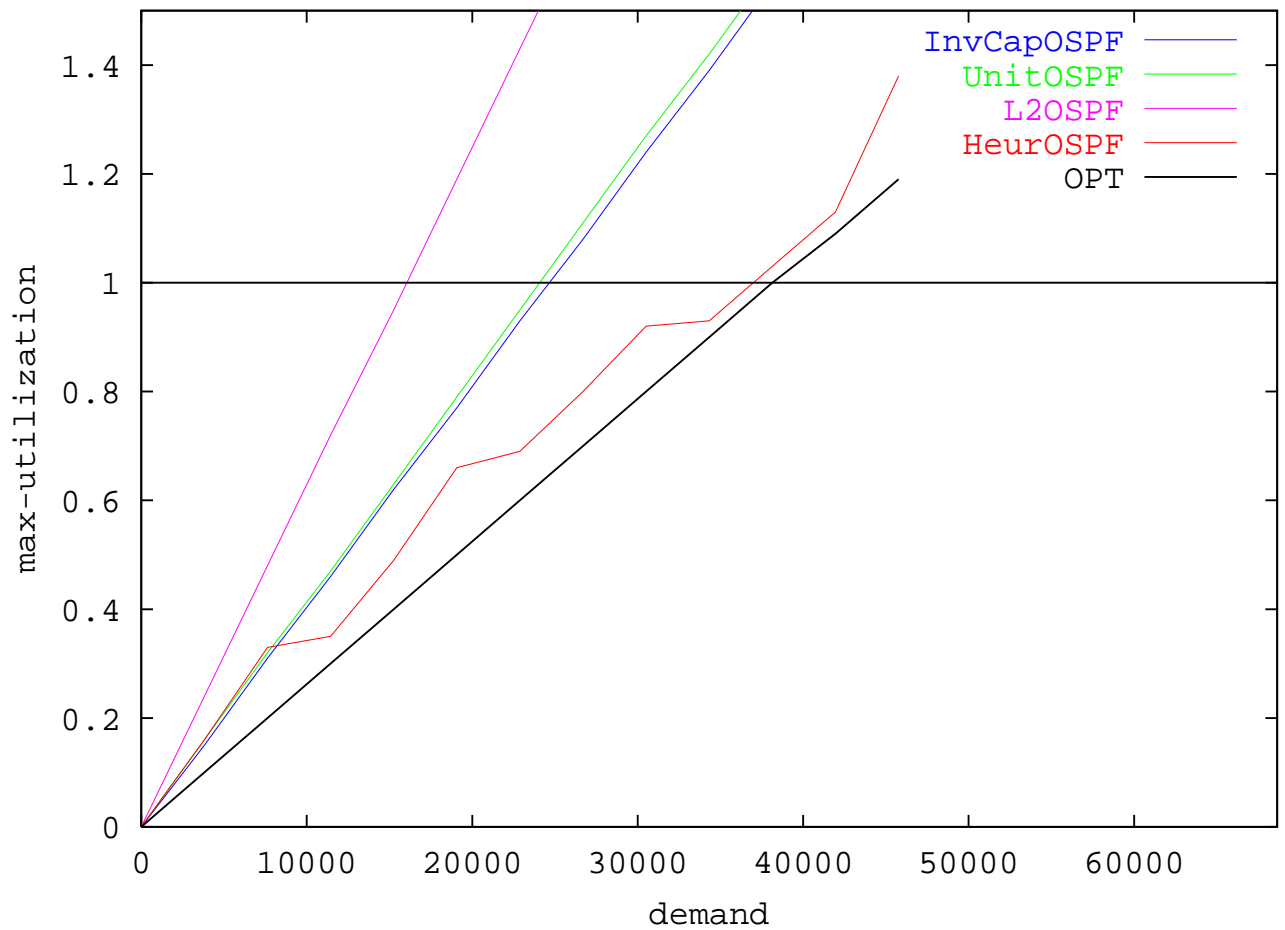
Proposed OC48 network with 90 nodes, 274 links, and projected demands



HeurOSPF uses weights $\{1, \dots, 20\}$ but $\{1, 2\}$ nearly as good

WorldNet, Continued

Max-utilization with same weight setting

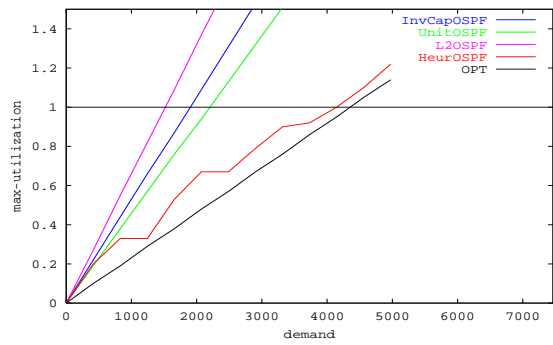
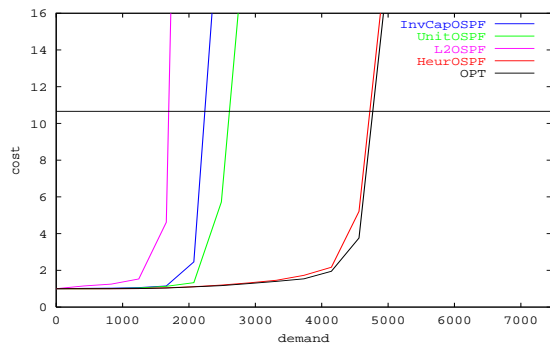


Synthetic Networks

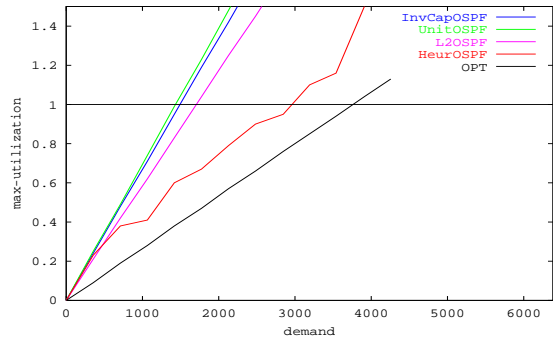
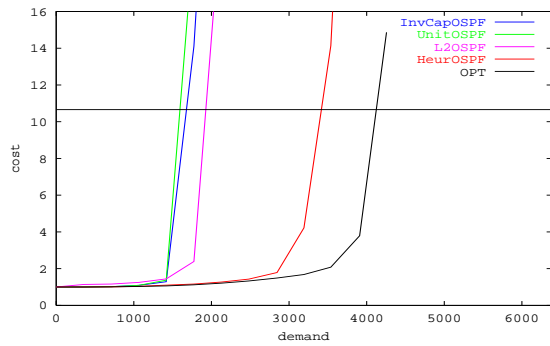
Zegura et al.'s 2-level graphs (INFOCOM'96, *IEEE Communications Magazine*, 1997)

- long distance arcs with capacity 1000
- local access arcs with capacity 200
- $Demand[i, j] =$

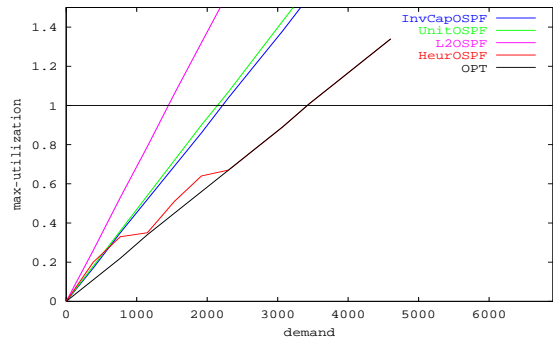
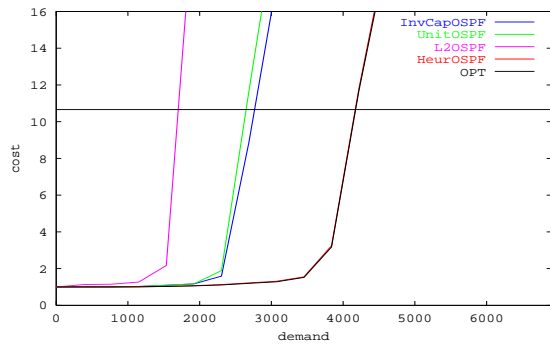
$$\alpha \times O_i \times D_j \times C_{(i,j)} \times e^{\delta(i,j)/2\Delta}$$



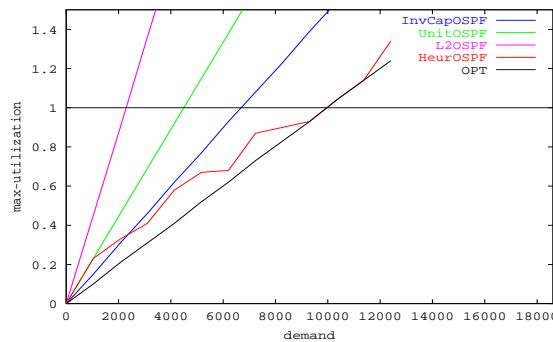
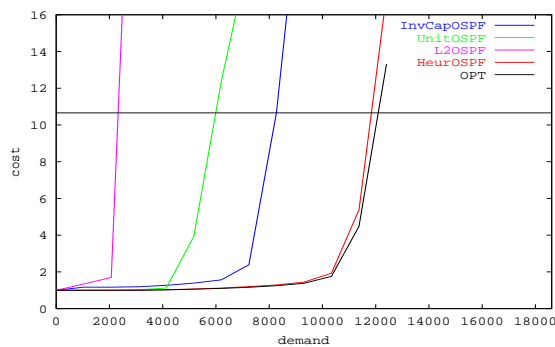
2-level graph with 50 nodes and 148 arcs.



2-level graph with 50 nodes and 212 arcs.



2-level graph with 100 nodes and 280 arcs.



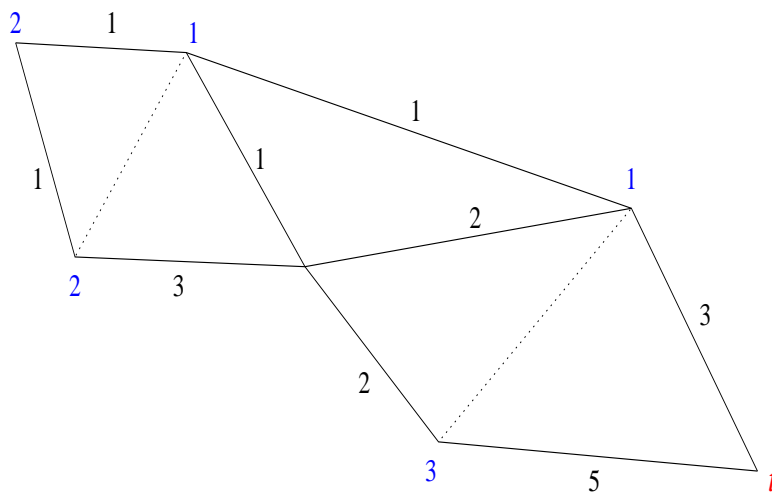
2-level graph with 100 nodes and 360 arcs.

Algorithmic Details

OSPF evaluation

Weight setting \mapsto routing \mapsto cost

- Compute flow to each destination t :
 - compute shortest path graph to t
 - accumulate flow starting furthest away



- For each arc e , add up the flows through e to all destination and compute the penalty function
- Add up the penalties for all the arcs e

Running time $O(nm \log n)$

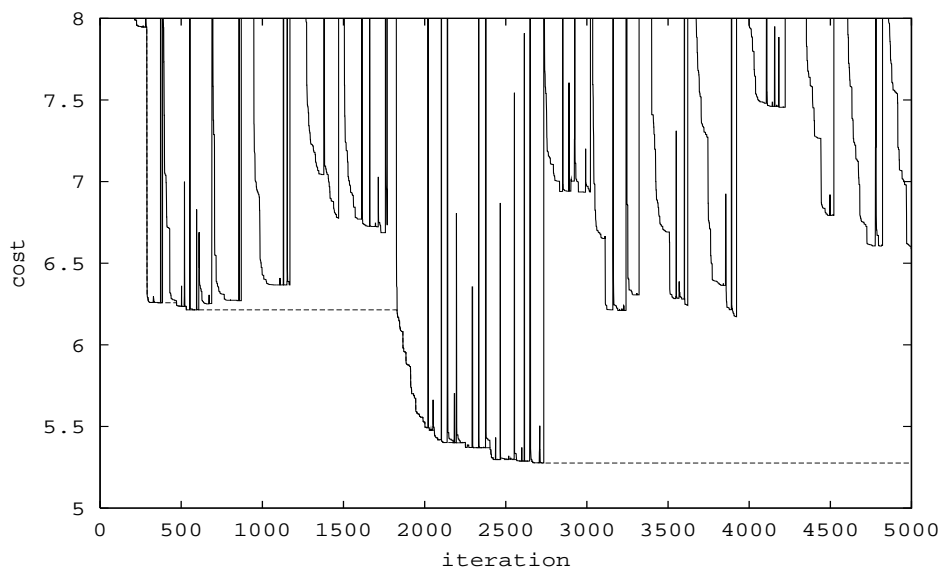
Local search

Pick random weight setting

Repeat 5000 times

Go to the “best” neighbor

Output best solution seen



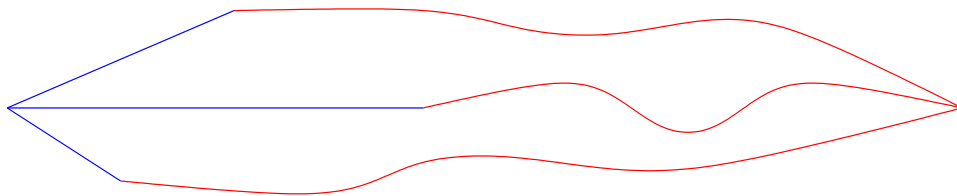
Neighbors

The obvious one:

change one weight

The less obvious one:

for one node, change weights of outgoing arcs so as to get even splitting of flow towards one destination



Picking neighbors

We pick best from random subset of neighborhood.

subset size decreases with non-improving moves
... making best more random
... thus getting out of valleys
... quickly

vector hashing used to

- avoid cycling
- avoid evaluating redundant neighbors

Speed

A typical run can evaluate $\approx 10,000,000$ weight settings.

Each evaluation includes an all pairs shortest path computation.

Nevertheless, we typically spend only 1 hour on a network graph with 100 nodes and 360 arcs

- ... thanks to lazy evaluation of neighbor weights
- ... reusing as much information as possible
- ... which gains a factor of 15 in speed

OSPF Basic Conclusions

New OSPF heuristic

- Typically close to optimal
- Allows for 50%-110% increase in traffic that can be carried without congestion relative to standard weight settings
- Works well for different objectives
- Robust for random perturbation of demands
 - multiply $D(s, t)$ by random $r(x, y) \in [0, 2]$
- But is this realistic?

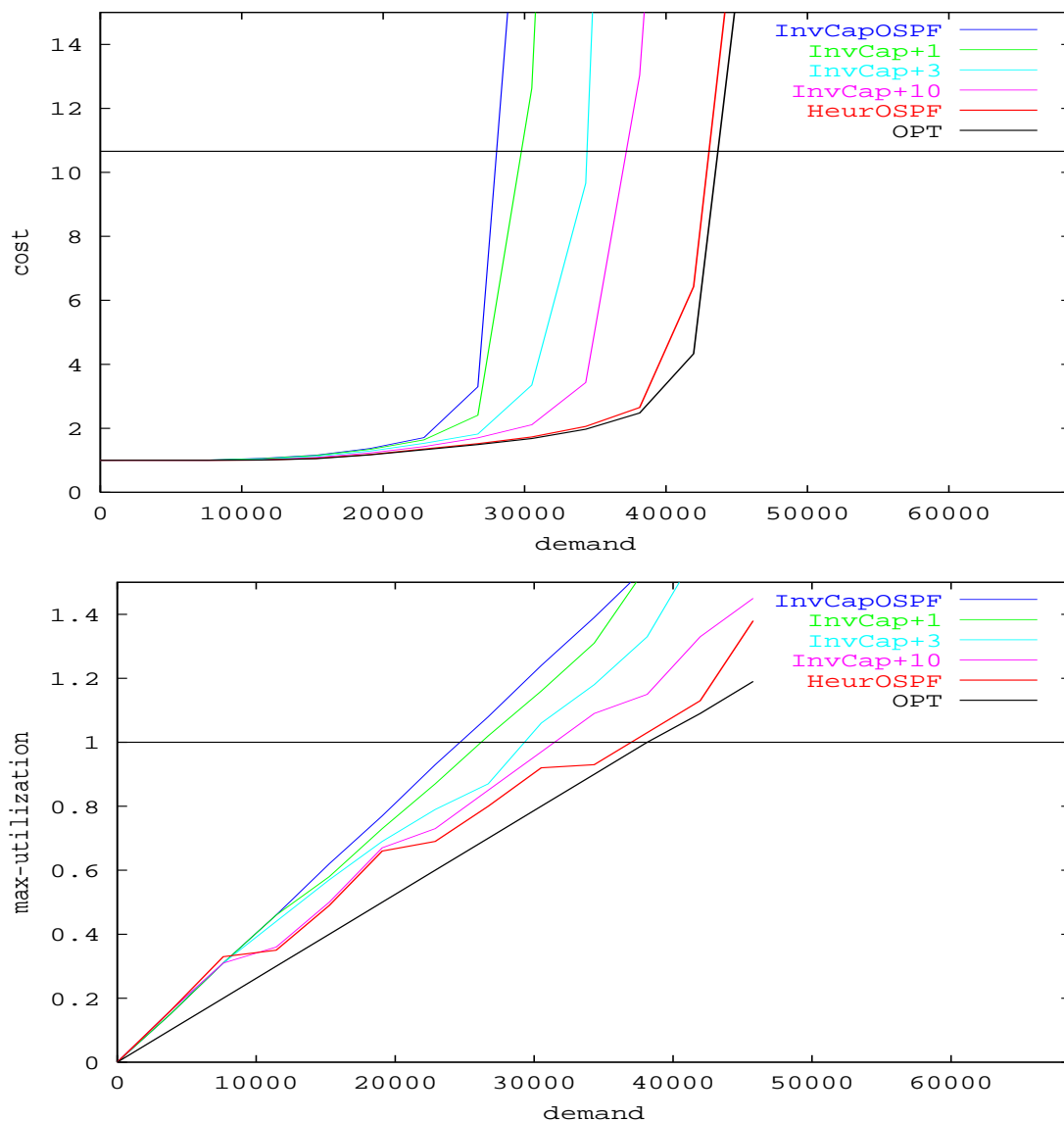
Traffic Management

- Adapting weight settings with few weight changes, e.g., in connection with link-failures.
- Weight settings for predicted periodic changes in demands.

Few weight changes

On real AT&T IP backbone with current weights, single weight change improved max-utilization by 8%.

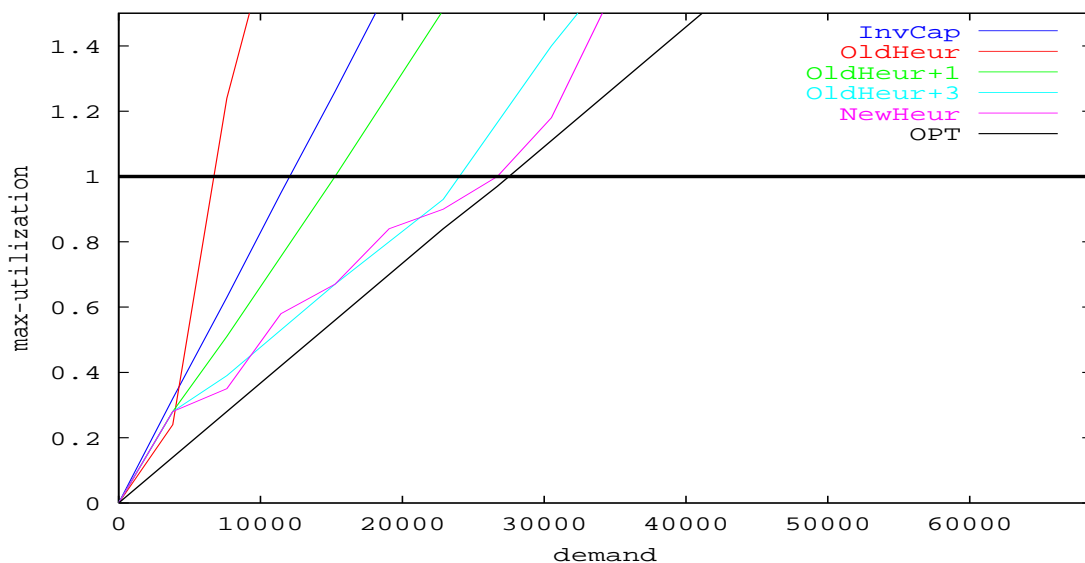
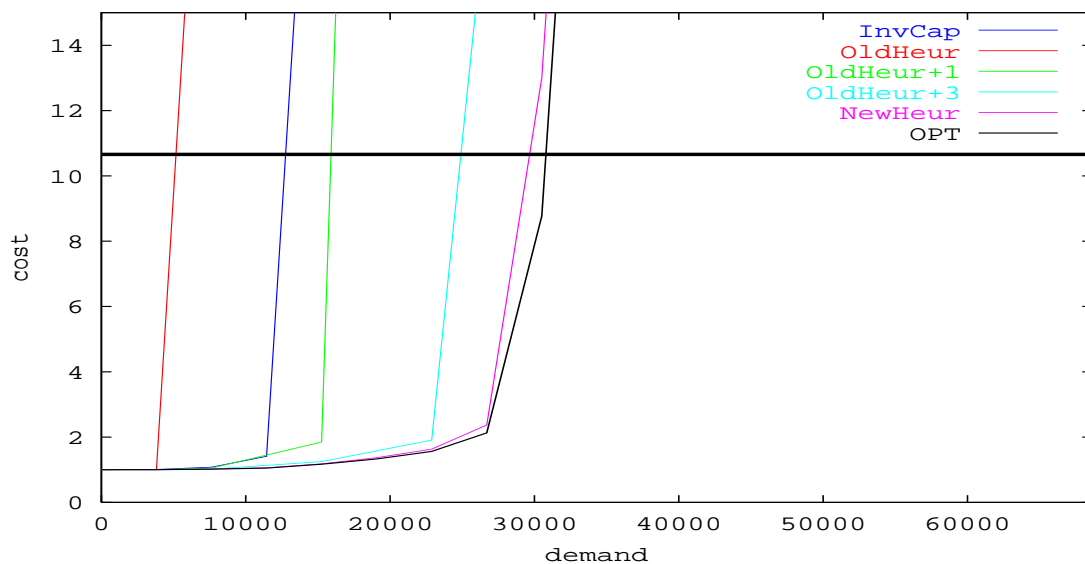
On proposed AT&T OC48 network



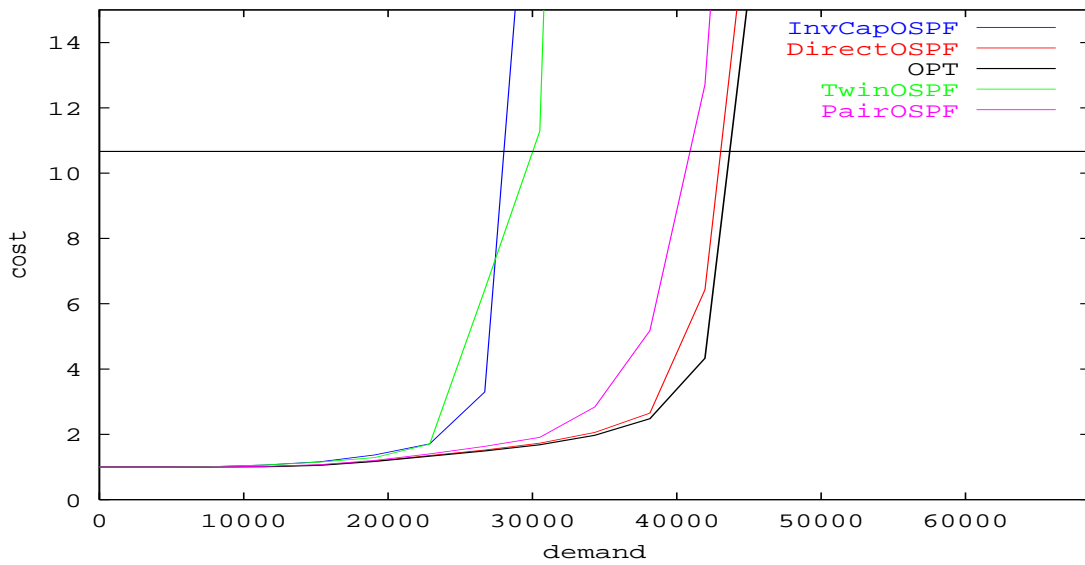
Link-failure

Our heuristic weight setting remained within 10% of optimum for almost all link failures.

For a critical link-failure, we can precompute weight changes.



Same weight setting day



and night

