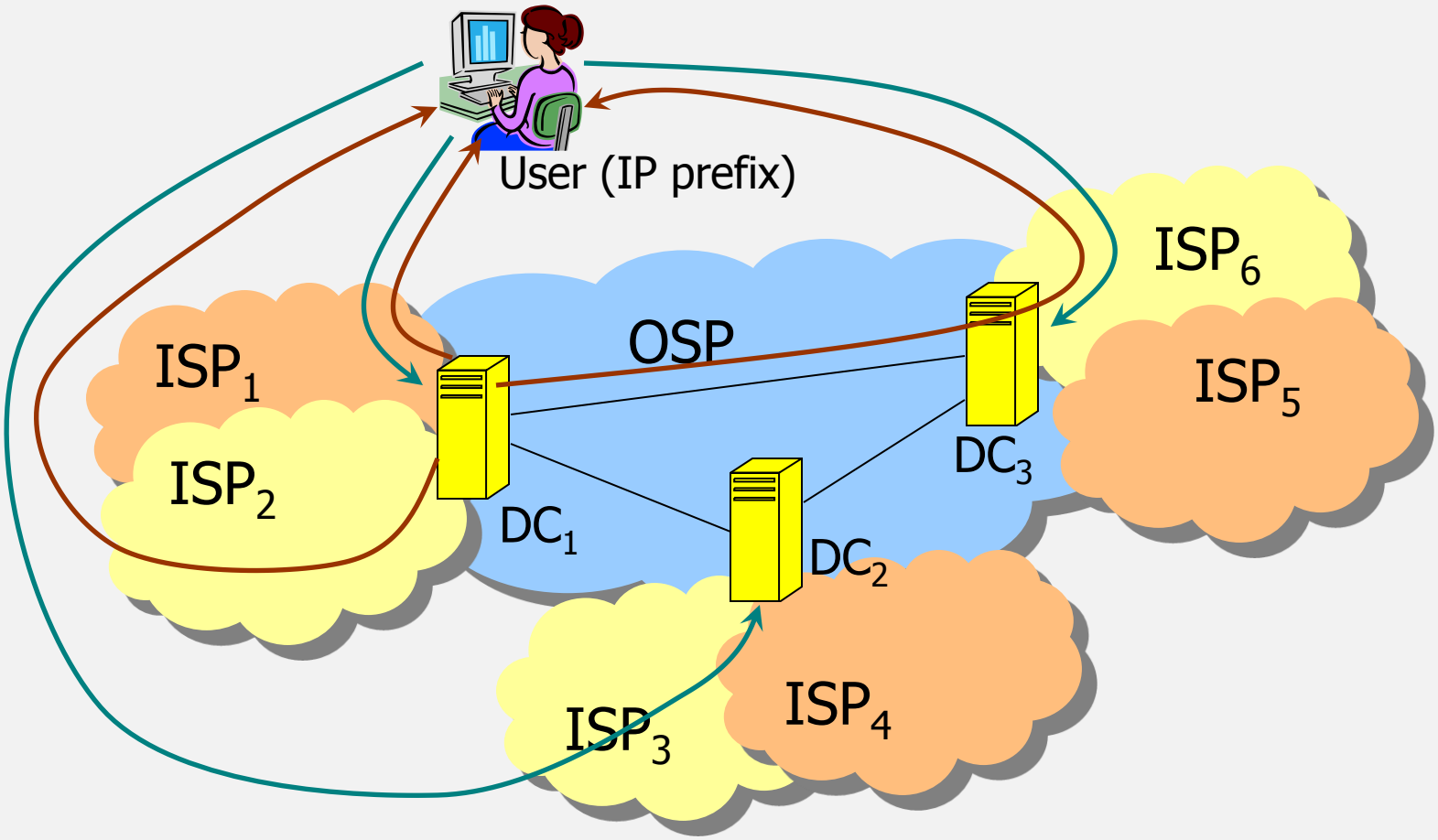
Optimizing Cost and Performance in Online Service Provider Networks

Ming Zhang

Microsoft Research

Joint work with Zheng Zhang (Purdue), Albert Greenberg (MSR), Y. Charlie Hu (Purdue), Ratul Mahajan (MSR), and Blaine Christian (Microsoft)

OSP network



Google Microsoft YAHOO! amazon.com

Key factors in OSP traffic engineering

- Cost
  - Google Search: 5B queries/month
  - MSN Messenger: 330M users/month
  - Traffic volume exceeding a PB/day
- Performance
  - Directly impacts user experience and revenue
    - Purchases, search queries, ad click-through rates

Current TE solution is limited

- Current practice is mostly manual
  - Incoming: DNS redirection, nearby DC
  - Outgoing: BGP, manually configured

- Complex TE strategy space
  - (~300K prefixes) x (~10 DC) x (~10 routes/prefix)
  - Link capacity creates dependencies among prefixes

Prior work on TE

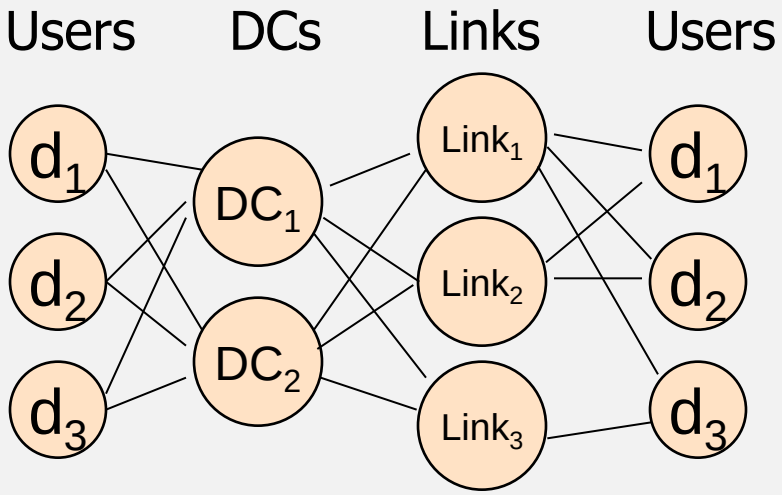
- Intra-domain TE for transit ISPs
  - Balancing load across internal paths
  - Not considering end-to-end performance

- Route selection for multi-homed stub networks
  - Single site
  - Small number of ISPs

Our contributions

- Formulation of OSP TE problem
- Design & implementation of Entact
  - A route-injection-based measurement technique
  - An online TE optimization framework
- Extensive evaluations in MSN
  - 40% cost reduction
  - Low operational overheads

Problem formulation

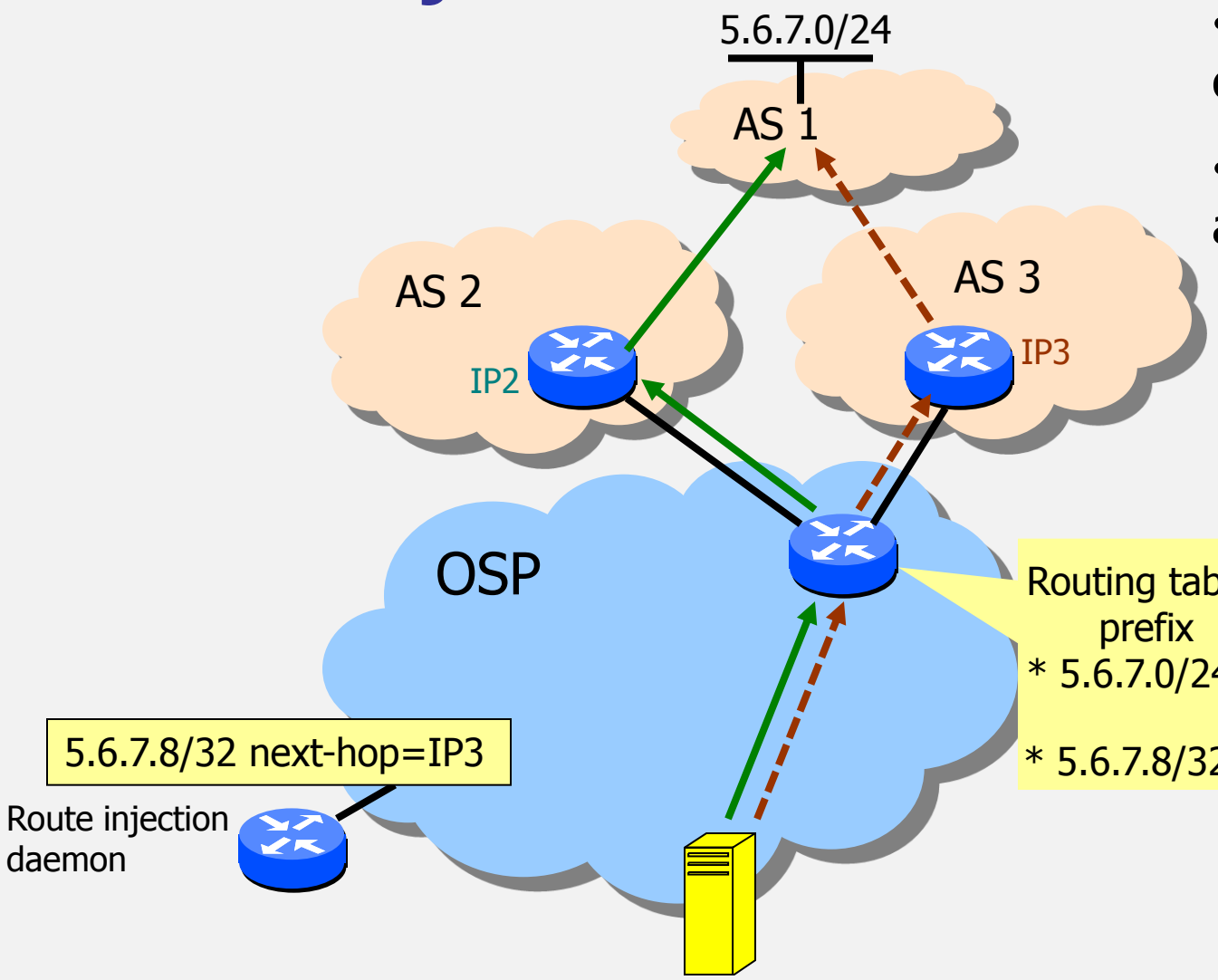


- INPUT: user prefixes, DCs, & external links
- OUTPUT: TE strategy, user prefix → (DC, external link)
- CONSTRAINTS: link capacity, route availability

Cost & performance measures

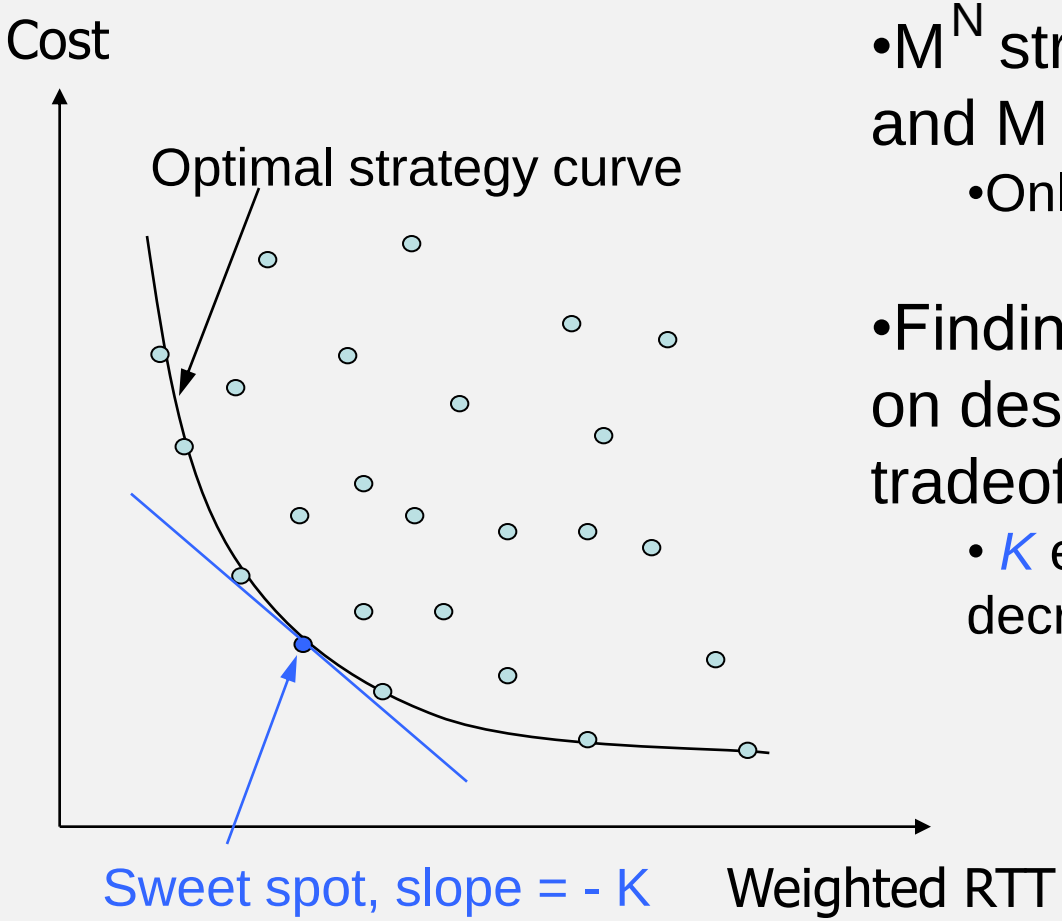
- Use RTT as the performance measure
  - Many latency-sensitive apps: search, email, maps
  - Apps are chatty: N x RTT quickly gets to 100+ms
- Transit cost: $F(v) = price \times v$
  - Ignore internal traffic cost

Measuring alternative paths with route injection



- Minimal impact on current traffic
- Existing approaches are inapplicable

Selecting desirable strategy



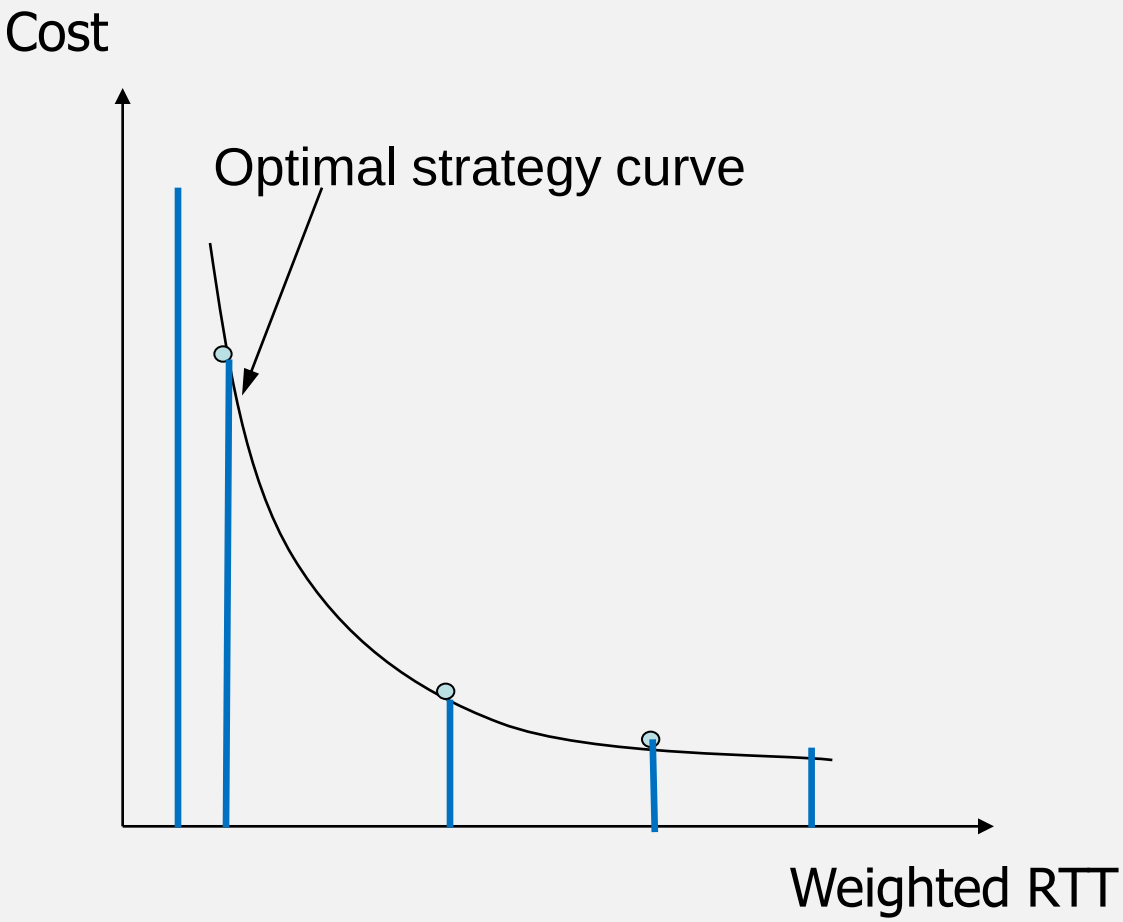
- M^N strategies for N prefixes and M alternative paths/prefix
  - Only consider optimal strategies
- Finding "sweet spot" based on desirable cost-performance tradeoff
  - *K* extra cost for unit latency decrease

Computing optimal strategy

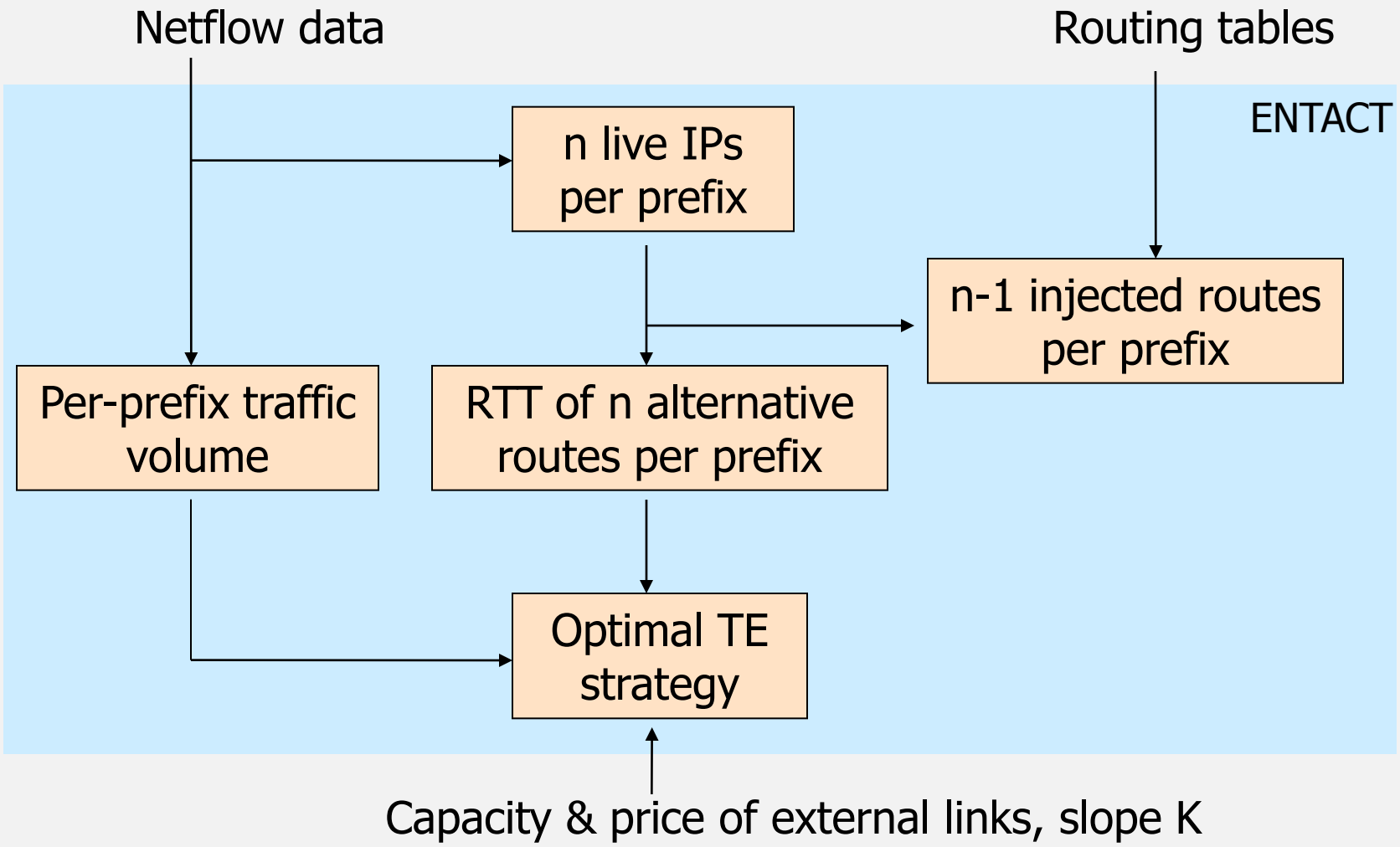
- P95 cost optimization is complex
  - Optimize short-term cost online
  - Evaluate using P95 cost

- Reduced an ILP problem
  - Find a fractional solution
  - Convert to an integer solution

Finding optimal strategy curve



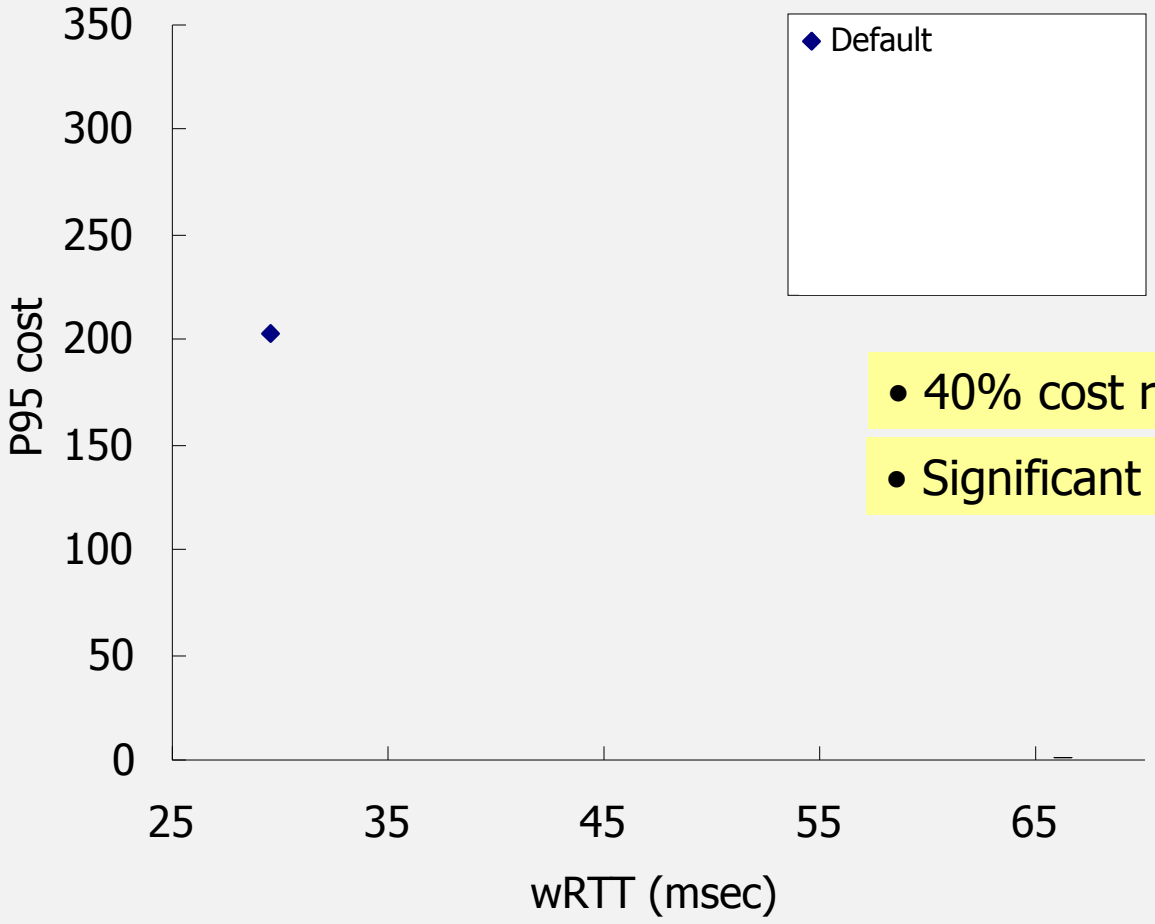
Entact architecture



Experimental setup

- MSN: one of the largest OSP networks
  - 11 DCs, 1,000+ external links
- Assumptions in evaluation
  - Traffic and performance do not change with TE strategies
- 6K destination prefixes from 2,791 ASes
  - High-volume, single-location, representative

Benefits of Entact



- 40% cost reduction
- Significant cost/perf tradeoff

Where does cost reduction come from?

path chosen by Entact	prefixes (%)	wRTT difference (msec)	short-term cost difference
same	88.2	0	0
cheaper & shorter	1.7	-8	-309
cheaper & longer	5.5	+12	-560
pricier & shorter	4.6	-15	+42
pricier & longer	0.1	0	0

- Entact makes "intelligent" performance-cost tradeoff
- Automation is crucial for handling complexity & dynamics

Conclusions

- TE automation is crucial for large OSP network
  - Multiple DCs
  - Many external links
  - Dependencies between prefixes

- Entact -- first online TE scheme for OSP network
  - 40% cost reduction w/o performance degradation
  - Low operational overhead