the economics of network control

Best Practices in Network Planning and Traffic Engineering

MENOG 5

Wednesday 28 October 2009

Beirut, Lebanon

Thomas Telkamp
telkamp@cariden.com

John Evans
johnevans@cariden.com

Trends:

- Acceptance that simply monitoring per link statistics does not provide the fidelity required for effective and efficient IP/MPLS service delivery
- Shift from expert, guru-led planning to a more systematic approach
- Blurring of the old boundaries between planning, engineering and operations

Best Practices in Network Planning and Traffic Engineering

- The fundamental problem of SLA Assurance is one of ensuring there is sufficient capacity, relative to the actual offered traffic load
- The goal of network planning and traffic engineering (TE) is to ensure there is sufficient capacity to deliver the SLAs required for the transported services
- What tools are available:
  - Capacity planning – *essential*
  - Diffserv – helps with efficient support for multiple services ... *but still need (per class) capacity planning*
    - [Filsfils and Evans 2005]
  - TE – may also help ... *but still need capacity planning*

Best Practices in Network Planning and Traffic Engineering

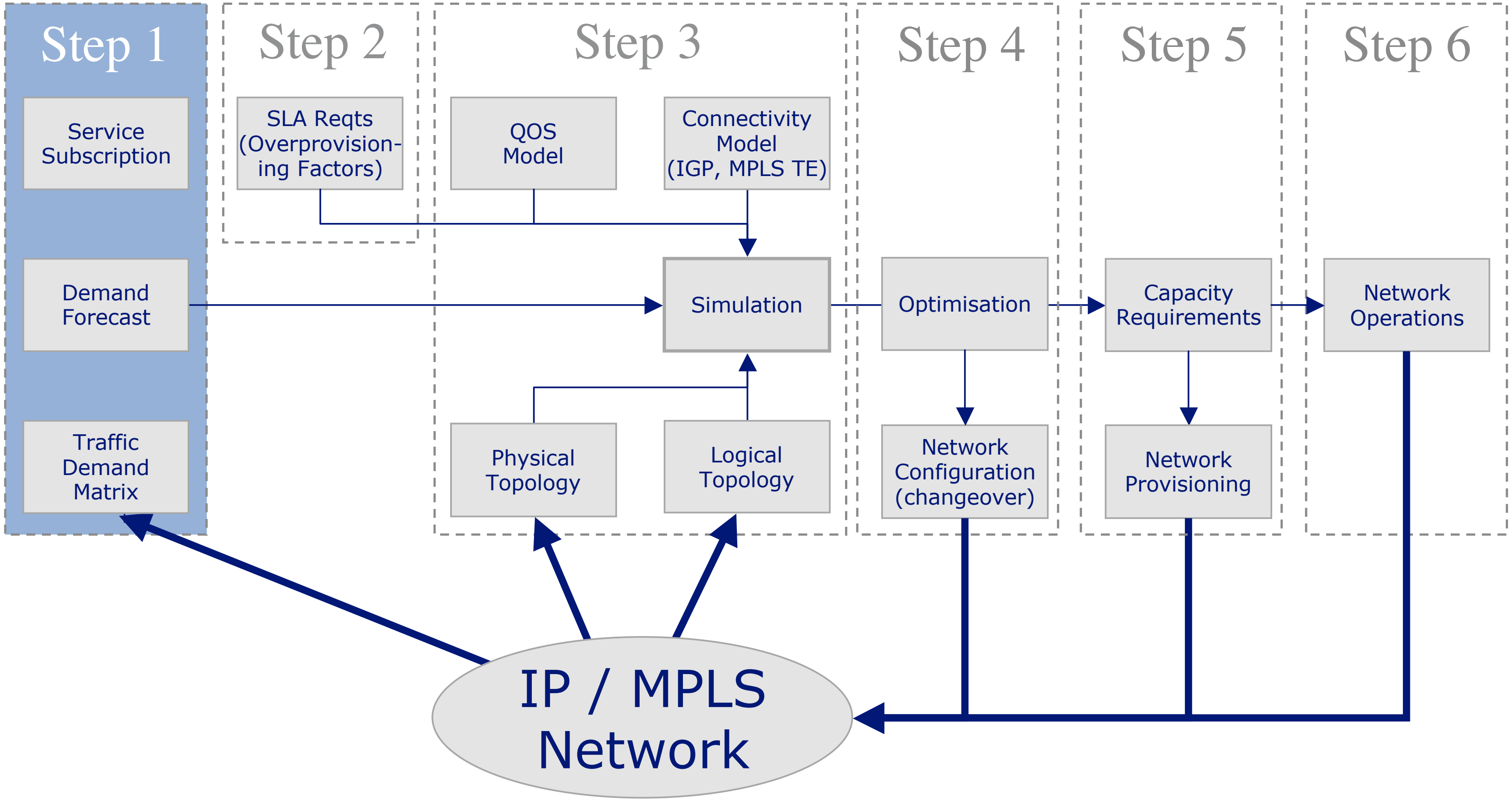
Network Planning and Traffic Engineering are two faces of the same problem.

In simple words:

- Network Planning:
  - building your network capacity where the traffic is
- Traffic Engineering:
  - routing your traffic where the network capacity is
- The better planning, the less TE you need…

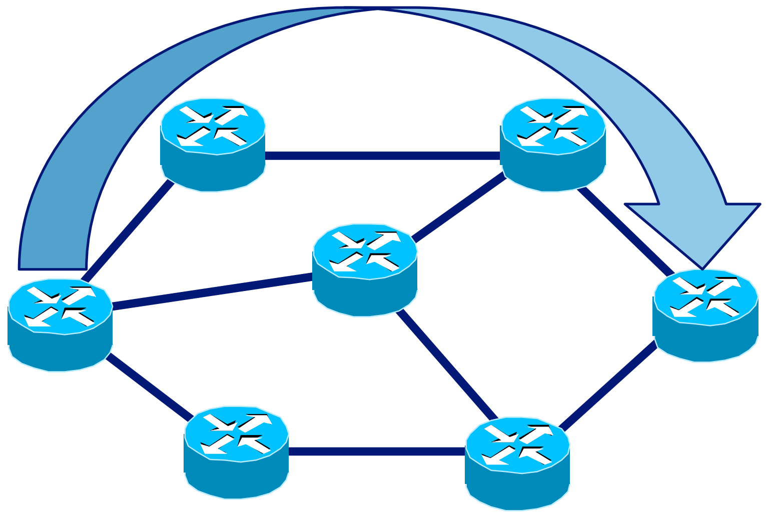
Network Planning Methodology

1. Traffic / Demand matrices ...



Traffic Demand Matrix

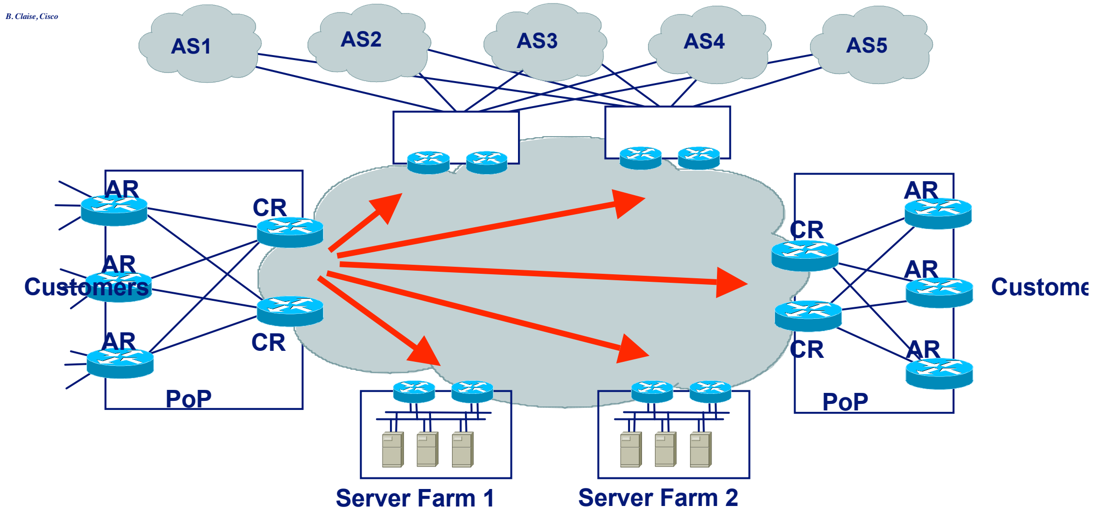
- Traffic demands define the amount of data transmitted between each pair of network nodes
  - Internal vs. external
  - per Class, per application, ...
  - Can represent peak traffic, traffic at a specific time, or percentile
  - Router-level or PoP-level demands
  - May be measured, estimated or deduced

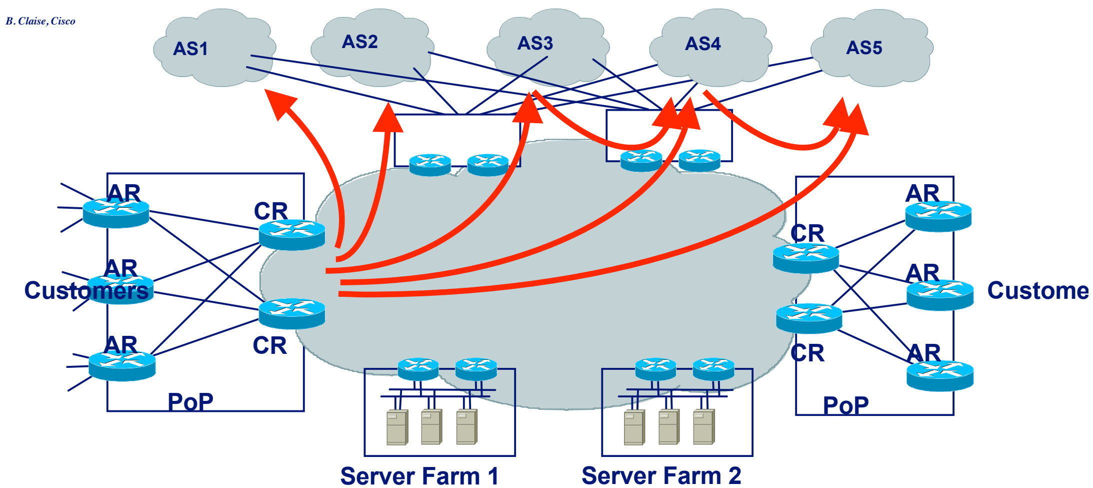


- The matrix of network traffic demands is crucial for analysis and evaluation of other network states than the current:
  - network changes
  - “what-if” scenarios
  - resilience analysis, network under failure conditions
  - optimisation: network engineering and traffic engineering
    - Comparing TE approaches
    - MPLS TE tunnel placement and IP TE

Traffic Matrix

- Internal Traffic Matrix
  - POP to POP, AR-to-AR or CR-to-CR
  - Some PoPs, e.g. regional, may be outside MPLS mesh
- External Traffic Matrix
  - Router (AR or CR) to External AS or External AS to External AS (for transit providers)
  - Useful for analyzing the impact of external failures on the core network
  - Origin-AS or Peer-AS
    - Peer-AS sufficient for capacity planning and resilience analysis
  - See RIPE presentation on peering planning [Telkamp 2006]





IP Traffic Matrix Practices

2001 2003 2007

Direct Measurement

NetFlow, RSVP, LDP, Layer 2, ...

Good when it works (half the time), but*

Measurement Methods

Flows

- NetFlow
  - Routers collect "flow" information
  - Export of raw or aggregated data
- BGP Policy Accounting & Destination Class Usage
  - Routers collect aggregated destination statistics – accounting for traffic according to the route it traverses

MPLS LSPs

- LDP
  - Used for VPNs
  - Measurement of LDP counters
- RSVP-TE
  - Used for MPLS TE
  - Measurement of Tunnel/LSP counters

NetFlow Background

- Router keeps track of (sampled) flows and packet/byte usage per flow
- Different approaches to aggregate flows depending on netflow version:
  - v5 (most common) /v8 (router based aggregation)
    - Enable NetFlow on edge-of-model interfaces
    - Export v5 with IP address or v8 with prefix aggregation (instead of peer-as or destination-as for source and destination)
    - Correlate flows with edge-of-model, e.g. IP to iBGP NextHop
      - V5: BGP passive peer on collector and aggregate flow counts
  - v9
    - Router does Flow-to-BGP Next Hop TOS aggregation – exports traffic matrix (very convenient!)
      - Only for BGP routes; only for IP {IP-to-IP, IP-to-MPLS}
      - configure on ingress interfaces
      - Cisco only
    - MPLS aware netflow - provides flow statistics for MPLS and IP packets
      - FEC implicitly maps to BGP next hop / egress PE
      - Based on the NetFlow version 9 export
      - No router based aggregation

MPLS

- A full mesh of MPLS LSPs (should be able to) provide internal traffic matrix directly
  - LDP: MPLS-LSR-MIB (or equivalent)
    - Mapping FEC to exit point of LDP cloud
    - Counters for packets that enter FEC (ingress)
    - Counters for packets switched per FEC (transit)
  - Full mesh of TE tunnels and Interface MIB
  - $O(N^2)$ measurements required
  - Inconsistencies in vendor implementations [Telkamp 2007]
- Does not provides external traffic matrix

- LSP stats good enough when:
  - Only need internal traffic matrix
  - Have full mesh of LSPs already; but no reason to deploy MPLS just for the TM
  - Not getting bitten by various platform issues
  - Long-term analysis (not quick enough for tactical Ops)

Measuring the Traffic Matrix in Practise

Flows

- NetFlow
  - v5
    - Resource intensive for collection and processing
    - Non-trivial to convert to Traffic Matrix
  - v9
    - BGP NextHop Aggregation scheme provides almost direct measurement of the Traffic Matrix
    - Only supported by newer versions of Cisco IOS
  - Inaccuracies
    - Stats can clip at crucial times
    - NetFlow and SNMP timescale mismatch
- BGP Policy Accounting & Destination Class Usage
  - Limited to 16 / 64 / 126 buckets

MPLS LSPs

- LDP
  - O(N^2) measurements
    - Missing values (expected when making tens of thousands of measurements)
    - Can take many minutes (important for tactical, quick response, TE)
  - Internal matrix only
  - Inconsistencies in vendor implementations
- RSVP-TE
  - Requires a full mesh of TE tunnels
  - Internal matrix only
  - Issues with O(N2): missing values, time, …
  - Inconsistencies in vendor implementations

IP Traffic Matrix Practices

2001 — 2003 — 2007

Direct Measurement	Estimation
NetFlow, RSVP, LDP, Layer 2, ...	Pick one of many solutions that fit link stats (e.g., Tomogravity)
Good when it works (half the time), but*	TM not accurate but good enough for planning

*Measurement issues

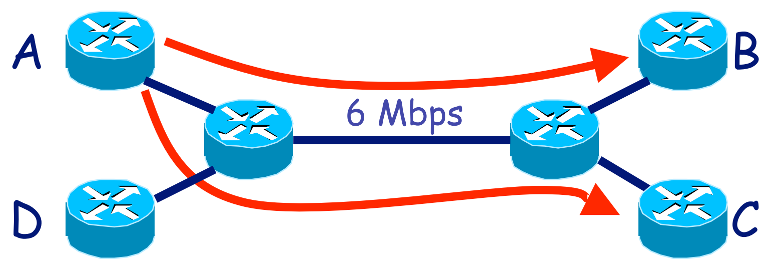
High Overhead (e.g., $O(N^2)$ LSP measurements, NetFlow CPU usage)

End-to-end stats not sufficient:

- Missing data (e.g., LDP ingress counters not implemented)
- Unreliable data (e.g., RSVP counter resets, NetFlow cache overflow)
- Unavailable data (e.g., LSPs not cover traffic to BGP peers)
- Inconsistent data (e.g., timescale differences with link stats)

Demand Estimation

- Goal: Derive Traffic Matrix (TM) from easy to measure variables
- Problem: Estimate point-to-point demands from measured link loads
- Underdetermined system:
  - N nodes in the network
  - O(N) links utilizations (known)
  - $O(N^2)$ demands (unknown)
  - Must add additional assumptions (information)
- Many algorithms exist:
  - Gravity model
  - Iterative Proportional Fitting (Kruithof's Projection)
  - … etc
- Estimation background: network tomography, tomogravity*, etc
  - Similar to: Seismology, MRI scan, etc.
  - [Vardi 1996]
  - * [Zhang et al, 2004]



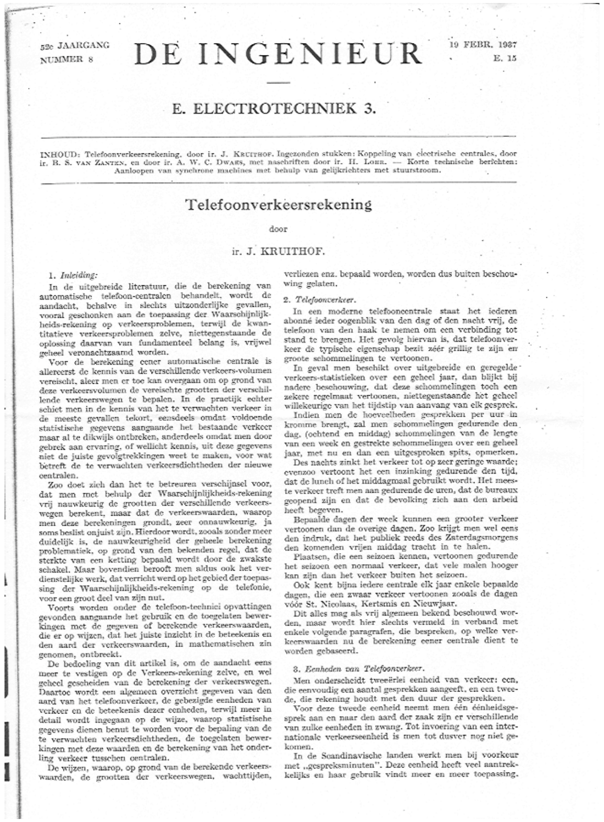
y: link utilizations
A: routing matrix
x: point-to-point demands

Solve: y = Ax -> In this example: 6 = AB + AC

Calculate the **most likely** Traffic Matrix

Is this new?

- Not really…
- ir. J. Kruithof: *Telefoonverkeersrekening*, De Ingenieur, vol. 52, no. 8, feb. 1937 (!)



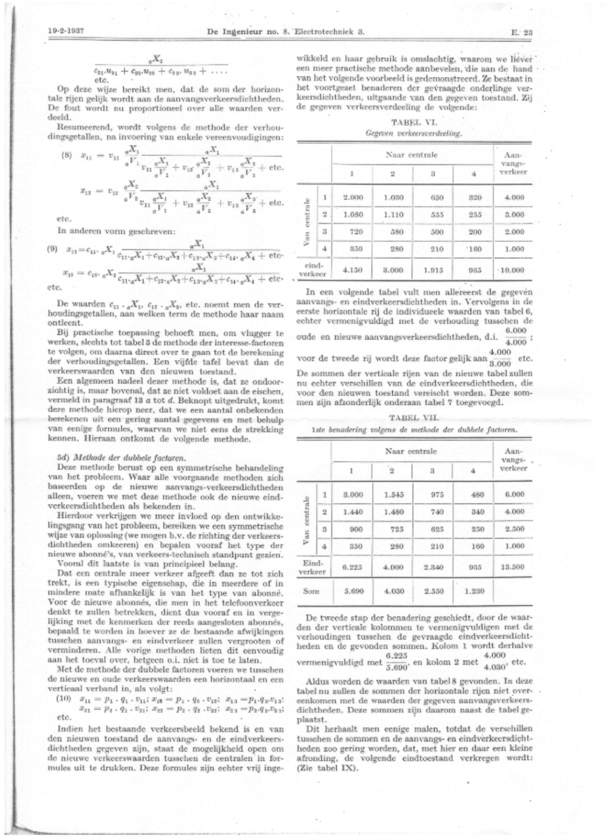
52e JAARGANG NUMMER 8 — DE INGENIEUR — 19 FEBR. 1937 E. 15

E. ELECTROTECHNIEK 3.

Telefoonverkeersrekening

door

ir. J. KRUITHOF.



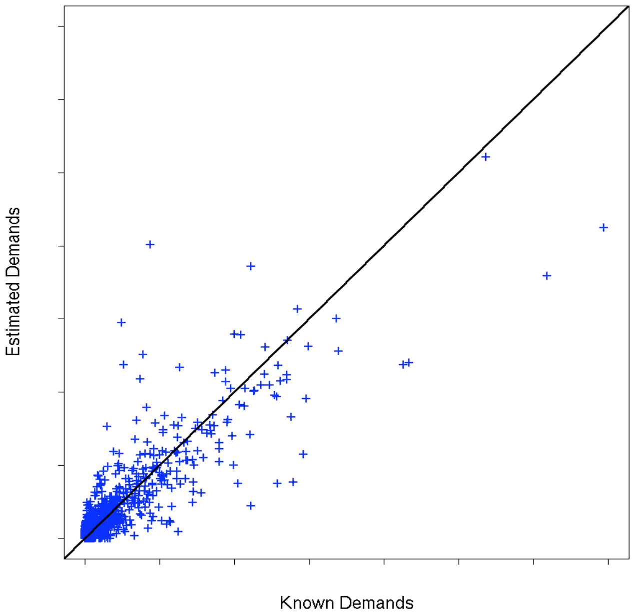
19-2-1937 — De Ingenieur no. 8. Electrotechniek 3. — E. 23

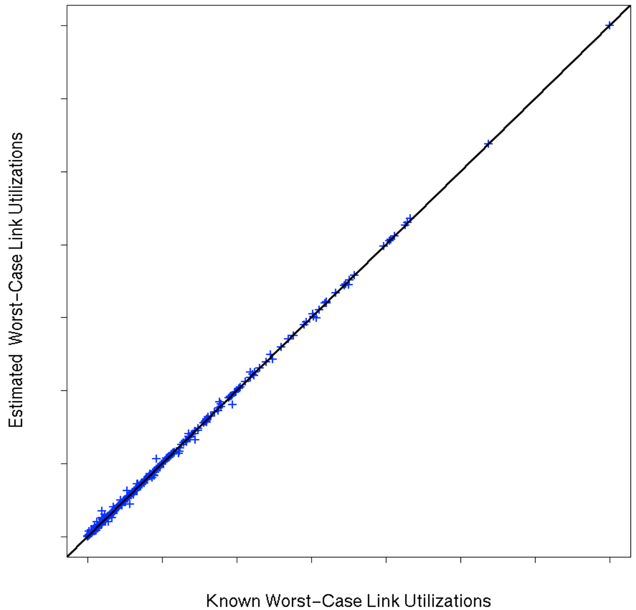
TABEL VI.

TABEL VII.

Demand Estimation Results

- Results from International Tier-1 IP Backbone





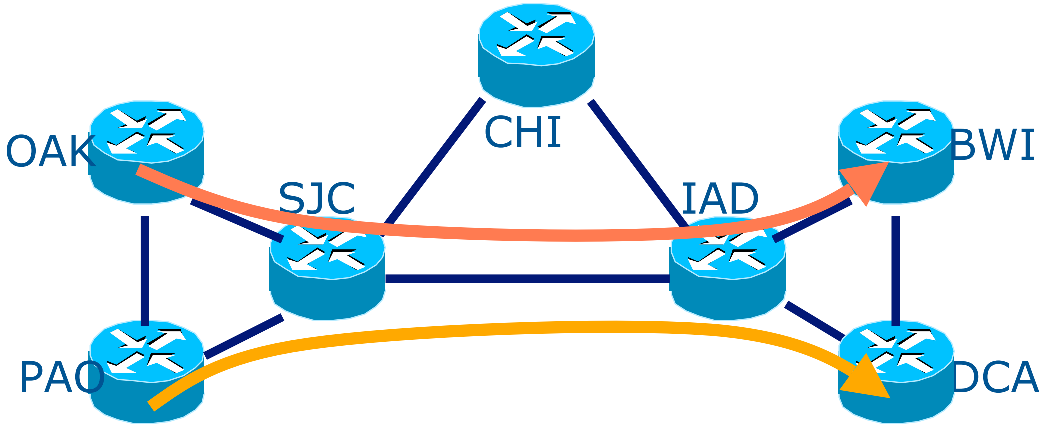
- Individual demand estimates can be inaccurate

- Using demand estimates in failure case analysis is accurate

See also [Zhang et al, 2004]: "How to Compute Accurate Traffic Matrices for Your Network in Seconds"

Results show similar accuracy for AT&T IP backbone (AS 7018)

Estimation Paradox Explained



- Hard to tell apart elements
  - OAK->BWI, OAK->DCA, PAO->BWI, PAO->DCA, similar routings
- Are likely to shift as a group under failure or IP TE
  - e.g., above all shift together to route via CHI under SJC-IAD failure

IP Traffic Matrix Practices

2001	2003	2007
Direct Measurement	**Estimation**	**Regressed Measurement**
NetFlow, RSVP, LDP, Layer 2, ... Good when it works (half the time), but*	Pick one of many solutions that fit link stats (e.g., Tomogravity) TM not accurate but good enough for planning	Use link stats as gold standard (reliable, available) Regression Framework adjusts (corrects/fills in) available NetFlow, MPLS, measurements to match link stats

*Measurement issues

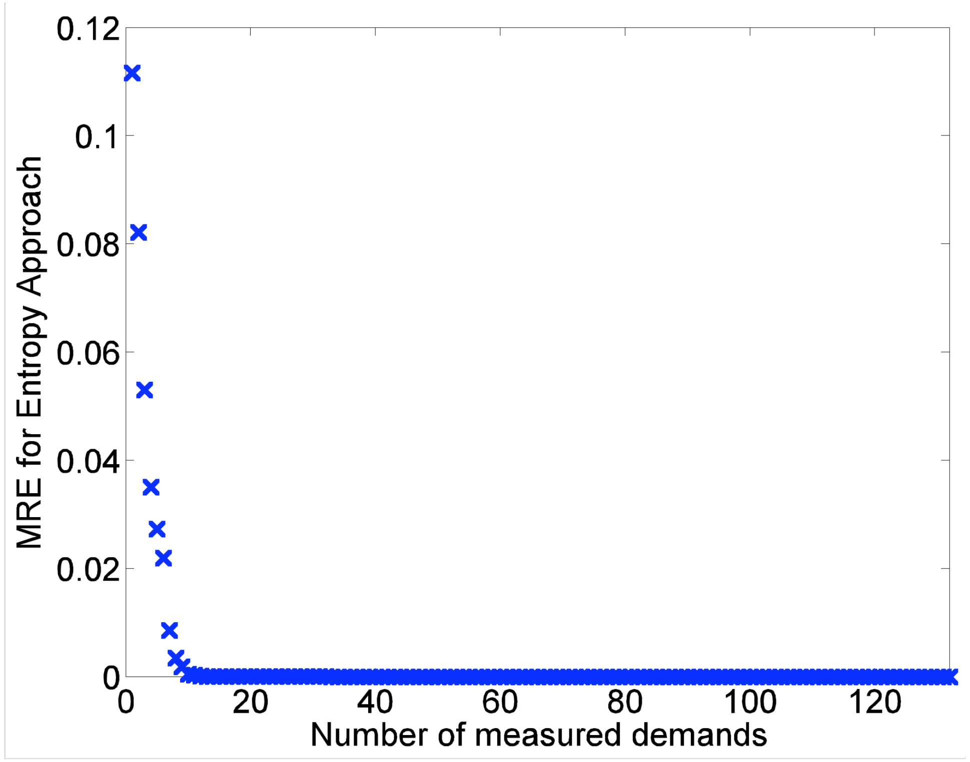
High Overhead (e.g., O(N^2) LSP measurements, NetFlow CPU usage)
End-to-end stats not sufficient:
- Missing data (e.g., LDP ingress counters not implemented)
- Unreliable data (e.g., RSVP counter resets, NetFlow cache overflow)
- Unavailable data (e.g., LSPs not cover traffic to BGP peers)
- Inconsistent data (e.g., timescale differences with link stats)

Regressed Measurements Overview

- Use interface stats as gold standard
  - Traffic management policies, almost always, based on interface stats, e.g.
    - ops alarm if 5-min average utilization goes >90%
    - traffic engineering considered if any link util approach 80%
    - cap planning guideline is to not have link util above 90% under any single failure
- Combine NetFlow, LSP stats, ... to match interface stats

Role of Netflow, LSP Stats,...

- Estimation techniques can be used in combination with demand measurements
  - E.g. NetFlow or partial MPLS mesh
- Can significantly improve TM estimate accuracy with just a few measurements



Regressed Measurements Sample

- Topology discovery done in real-time
- LDP measurements rolling every 30 minutes
- Interface measurement every 2 minutes
- Regression* combines the above information
- Robust TM estimate available every 5 minutes
- (See the DT LDP estimation for another approach for LDP**)

*Cariden's Demand Deduction™ in this case(http://www.cariden.com)
** Schnitter and Horneffer (2004)

Regressed Measurements Summary

- Interface counters remain the most reliable and relevant statistics
- Collect LSP, Netflow, etc. stats as convenient
  - Can afford partial coverage (e.g., one or two big PoPs)
  - more sparse sampling (1:10000 or 1:50000 instead of 1:500 or 1:1000)
  - less frequent measurements (hourly instead of by the minute)
- Use regression (or similar method) to find TM that conforms primarily to interface stats but is guided by NetFlow, LSP stats

Overall Summary

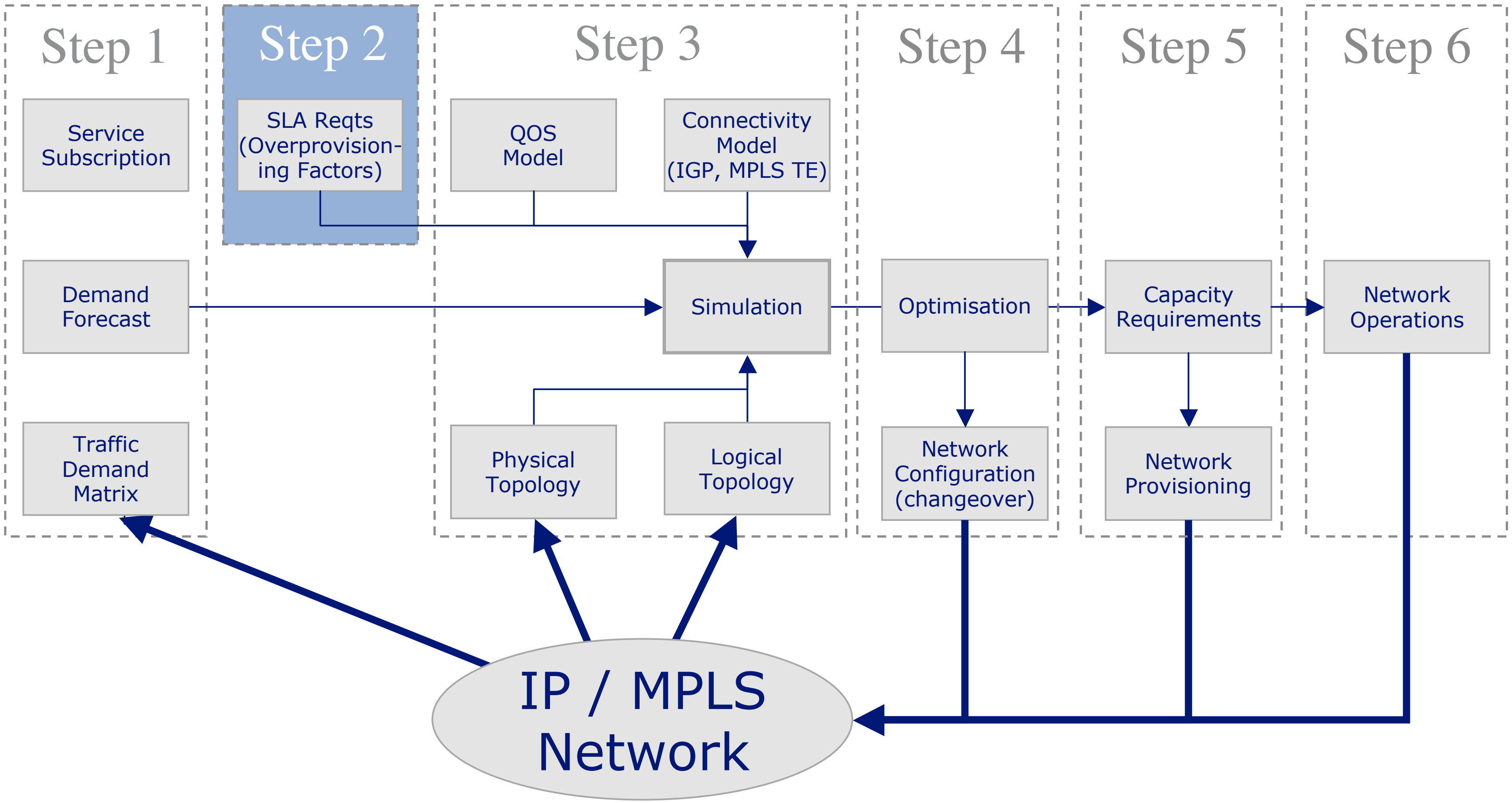
- Direct Measurement works well sometimes
  - Netflow OK on some equipment
  - LSP counters OK on some equipment and if only care for internal traffic matrix
  - Watch out for scaling, speed and measurement mismatch with link stats
- Estimation on link stats works sometimes
  - Has great speed (order of time to measure link stats)
  - Validity for given topology must be verified
- Regression is most flexible
  - Provides a spectrum of solutions between measurement and estimation
- Best practice is to start simple, verify, add complexity only if required
- More details: [Telkamp 2007, Maghbouleh 2007 and Claise 2003]

Best Practice: Start Simple, Verify

- Collect data over a few weeks
  - Link stats plus LSP and NetFlow stats (as available)
  - Make sure data set contains some failures:-)
- LSP or NetFlow stats good enough? (if so stop)
  - Compare sum of LSP, NetFlow against link counters
  - Compare failure utilization prediction against reality
- Link-based estimation good enough? (if so stop)
  - Again, test prediction against reality after failure
- Use Regressed Measurements on available data
  - Test, stop if predictions good enough
  - Otherwise add stats incrementally (e.g., additional NetFlow coverage)
  - Repeat this step until predictions are good

Network Planning Methodology

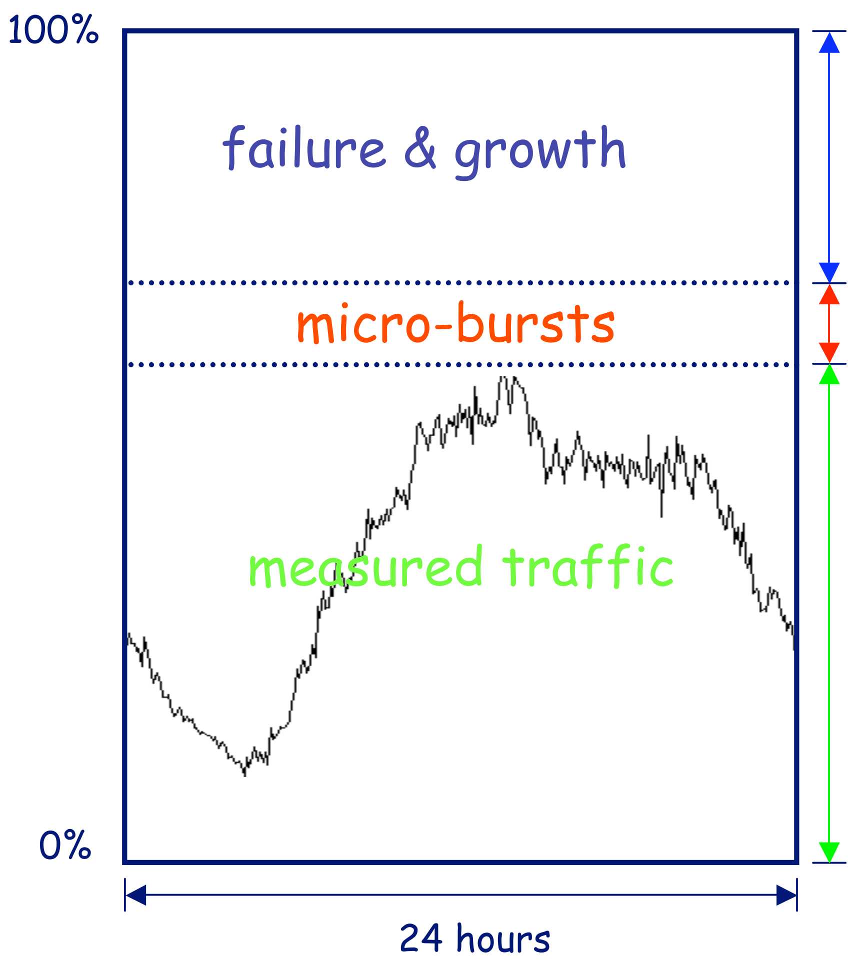
2. The relationship between SLAs and network planning targets ...



IP / MPLS Traffic Characterisation

- Network traffic measurements are normally long term, i.e. in the order of minutes
  - Implicitly the measured rate is an average of the measurement interval
- In the short term, i.e. milliseconds, however, microbursts cause queueing, impacting the delay, jitter and loss

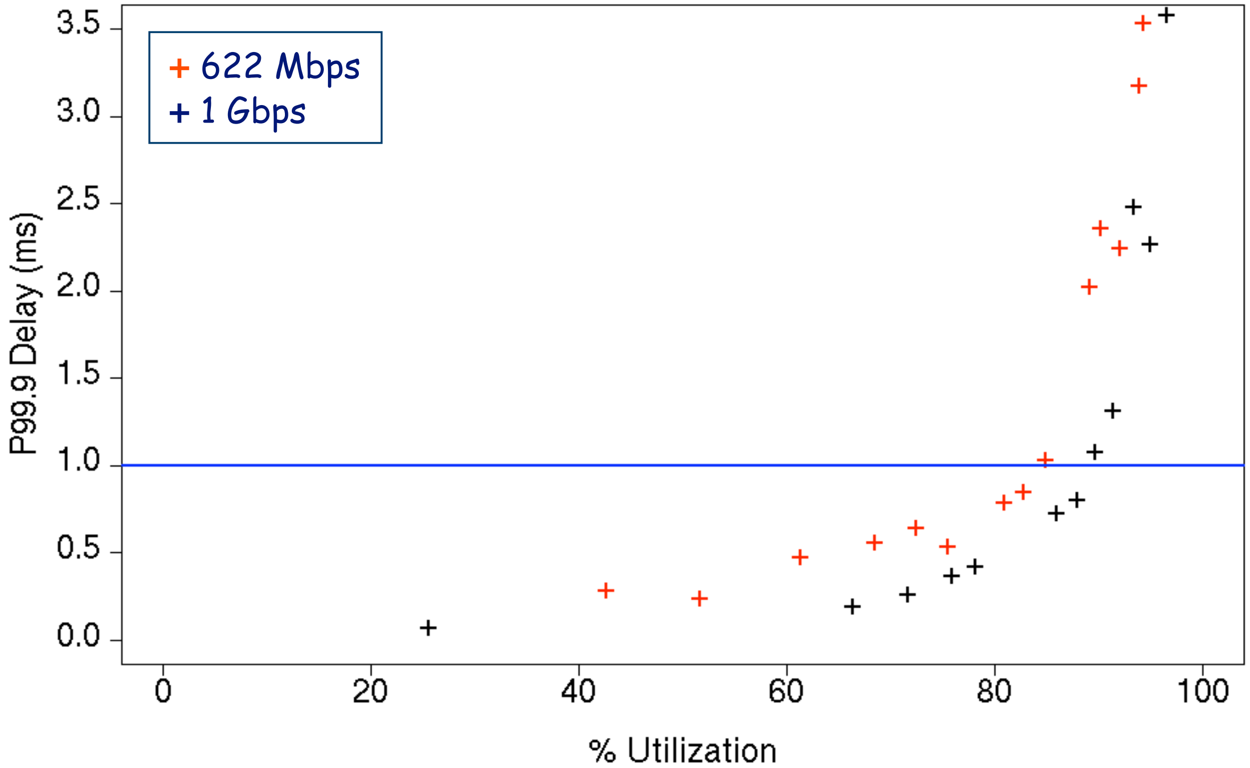
- *What's the relationship between the measured load and the short term microbursts?*
- *How much bandwidth needs to be provisioned, relative to the measured load, to achieve a particular SLA target?*



IP / MPLS Traffic Characterisation

- Opposing theoretical views:
  - M/M/1
    - Markovian, i.e. poisson-process
    - “Circuits can be operated at over 99% utilization, with delay and jitter well below 1ms” [Fraleigh et al. 2003, Cao et al. 2002]
  - Self-Similar
    - Traffic is bursty at many or all timescales
    - “Scale-invariant burstiness (i.e. self-similarity) introduces new complexities into optimization of network performance and makes the task of providing QoS together with achieving high utilization difficult” [Zafer and Sirin 1999]
    - Various reports: 20%, 35%, …
- Results from empirical simulation show characteristics similar to Markovian
  - [Telkamp 2003]

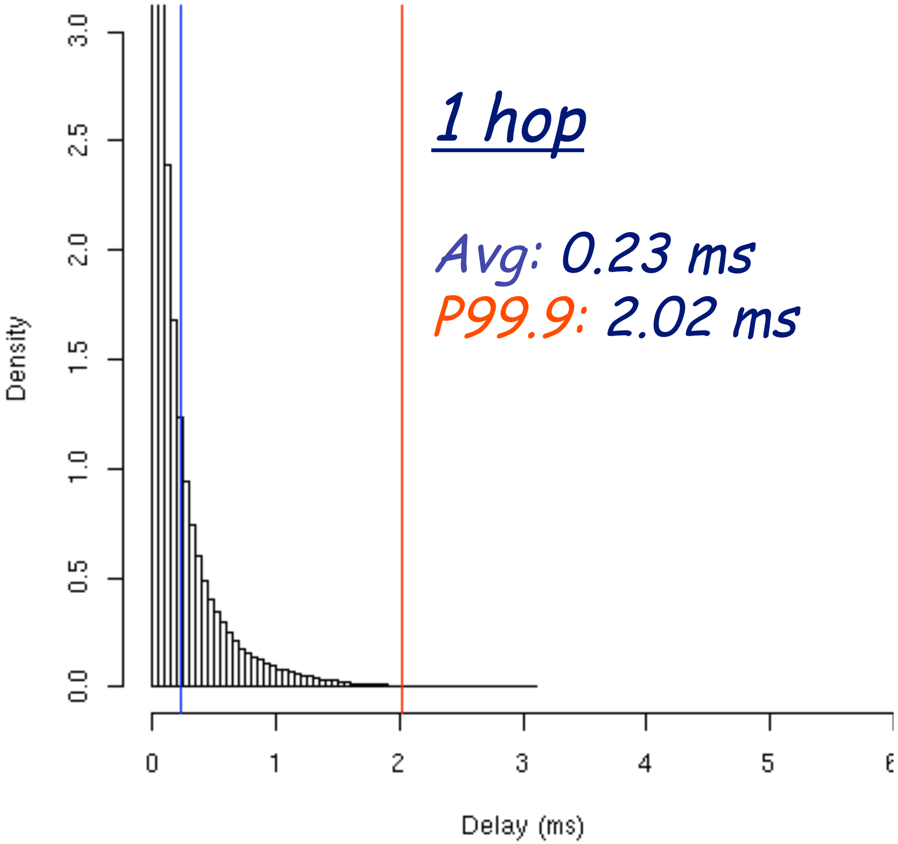
Queueing Simulation Results [Telkamp 2003]

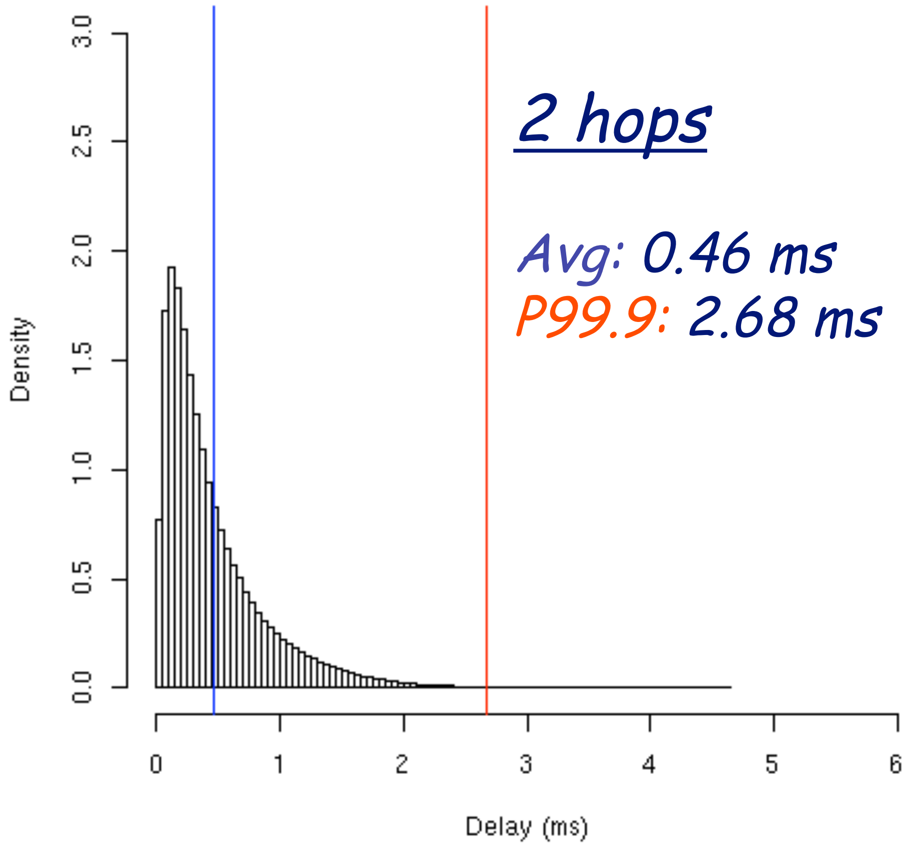


- 622Mbps, 1Gbps links – overprovisioning percentage ~10% is required to bound delay/jitter to 1-2ms
- Lower speeds (≤155Mpbs) – overprovisioning factor is significant
- Higher speeds (2.5G/10G) – overprovisioning factor becomes very small

Multi-hop Queuing [Telkamp 2003]

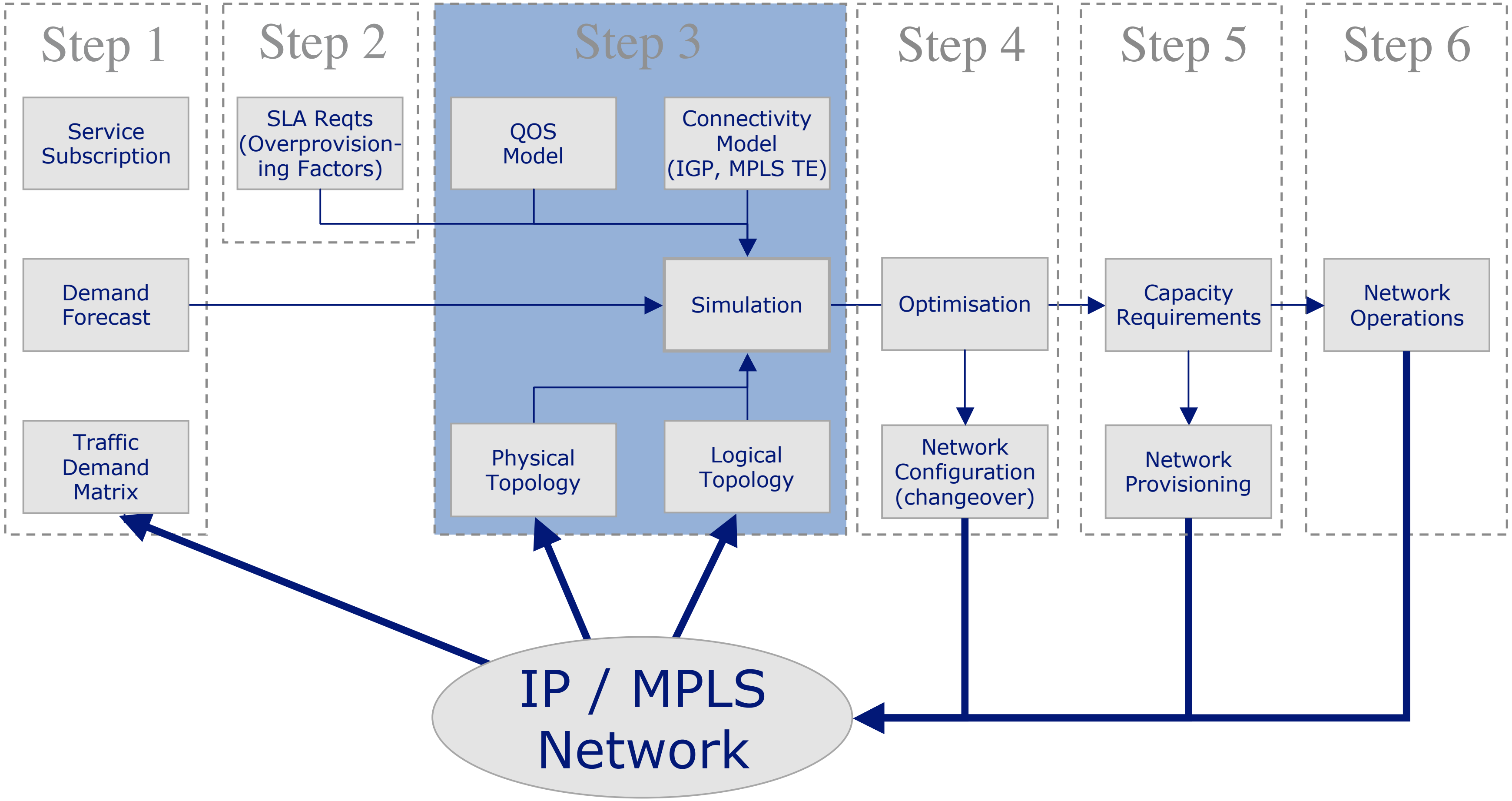
P99.9 multi-hop delay/jitter is not additive



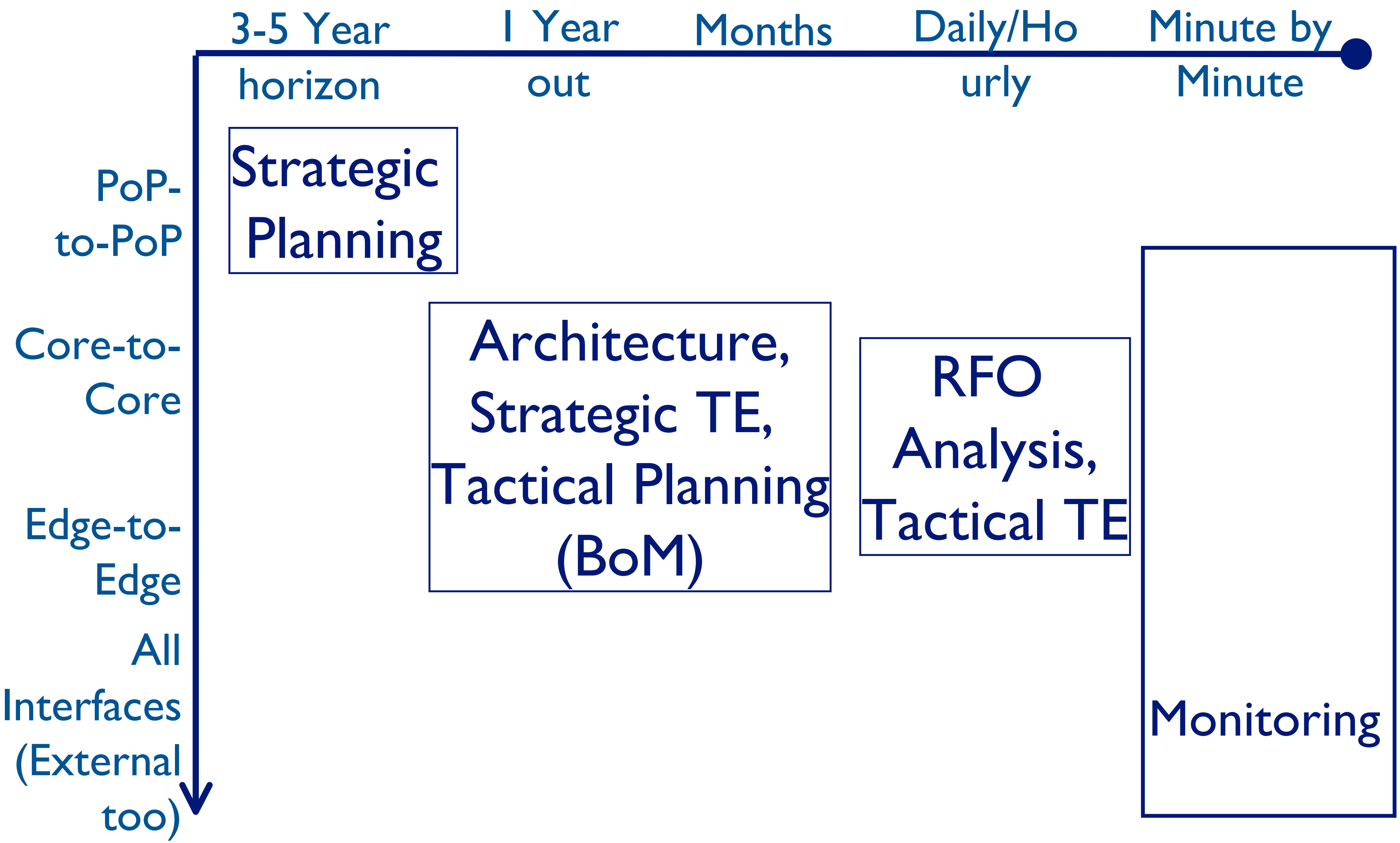


Network Planning Methodology

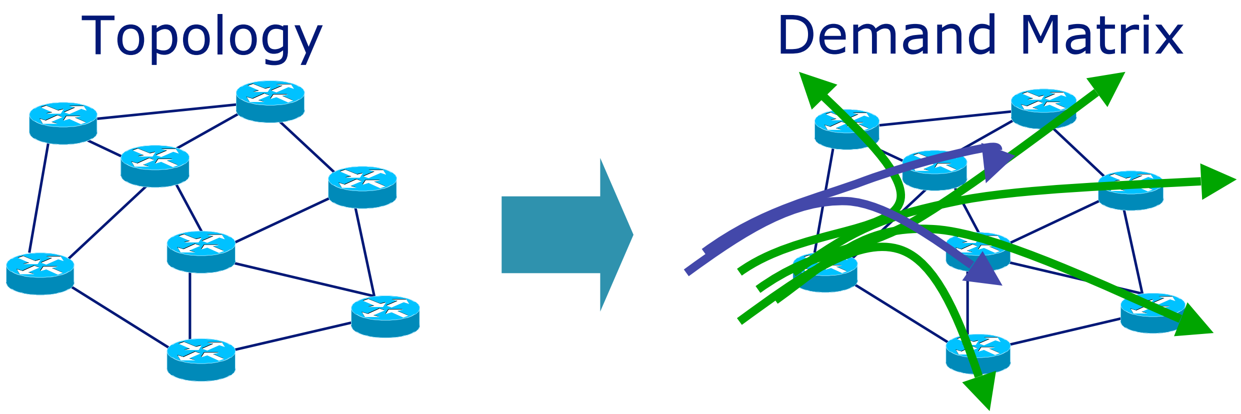
3. Network planning simulation and analysis – working and failure cases, what-if scenarios ...



Traffic Management in Context



Simulation

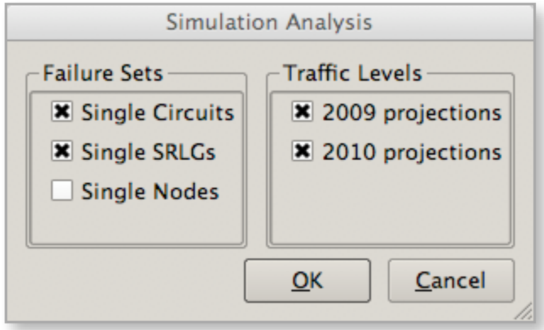


- Map core traffic matrix to topology (logical and physical)
- Simulate for link, node and shared risk (SRLG) failures
  - Can add a traffic growth factor if required
- On a per class basis if Diffserv deployed
- Enables:
  - Forecasting of which links need upgrading when
  - Understand of if topology should be changed
  - Comparison of different TE approaches

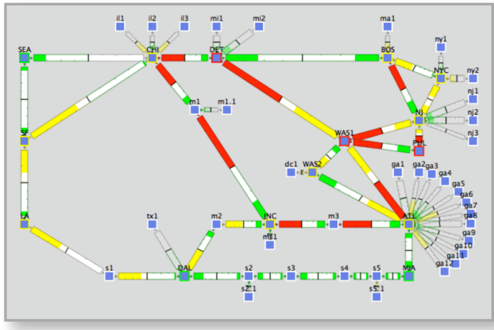
Failure Planning

Scenario: Planning receives traffic projections, wants to determine what buildout is necessary

Simulate using external traffic projections

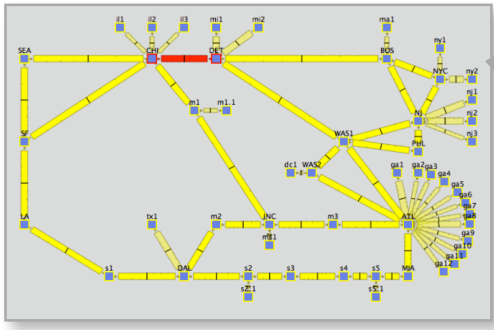


Worst case view



Potential congestion under failure in **RED**

Failure impact view



Failure that can cause congestion in **RED**

Perform topology what-if analysis

Topology What-If Analysis

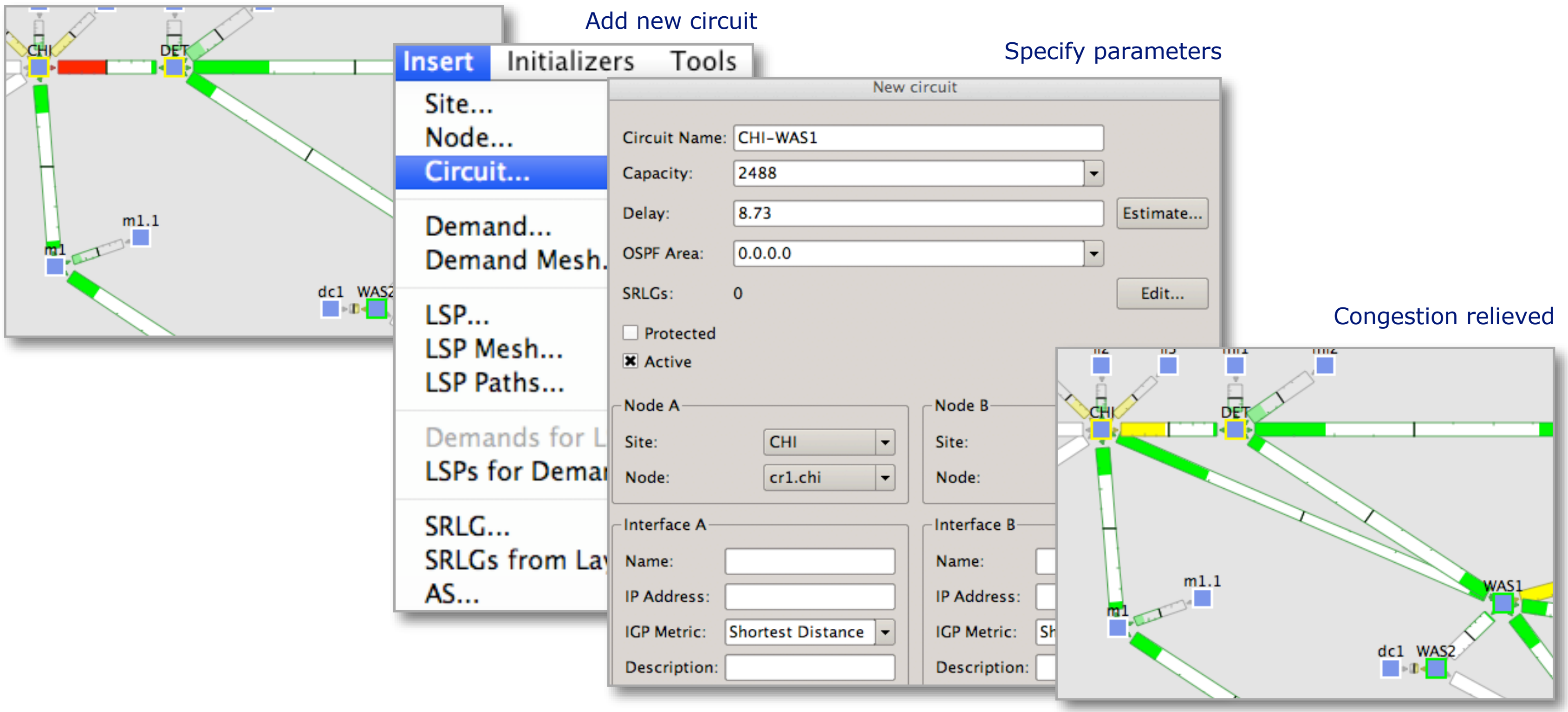
Scenario: Want to know if adding a direct link from CHI to WAS1 would improve network performance

Congestion between CHI and DET

Add new circuit

Specify parameters

Congestion relieved



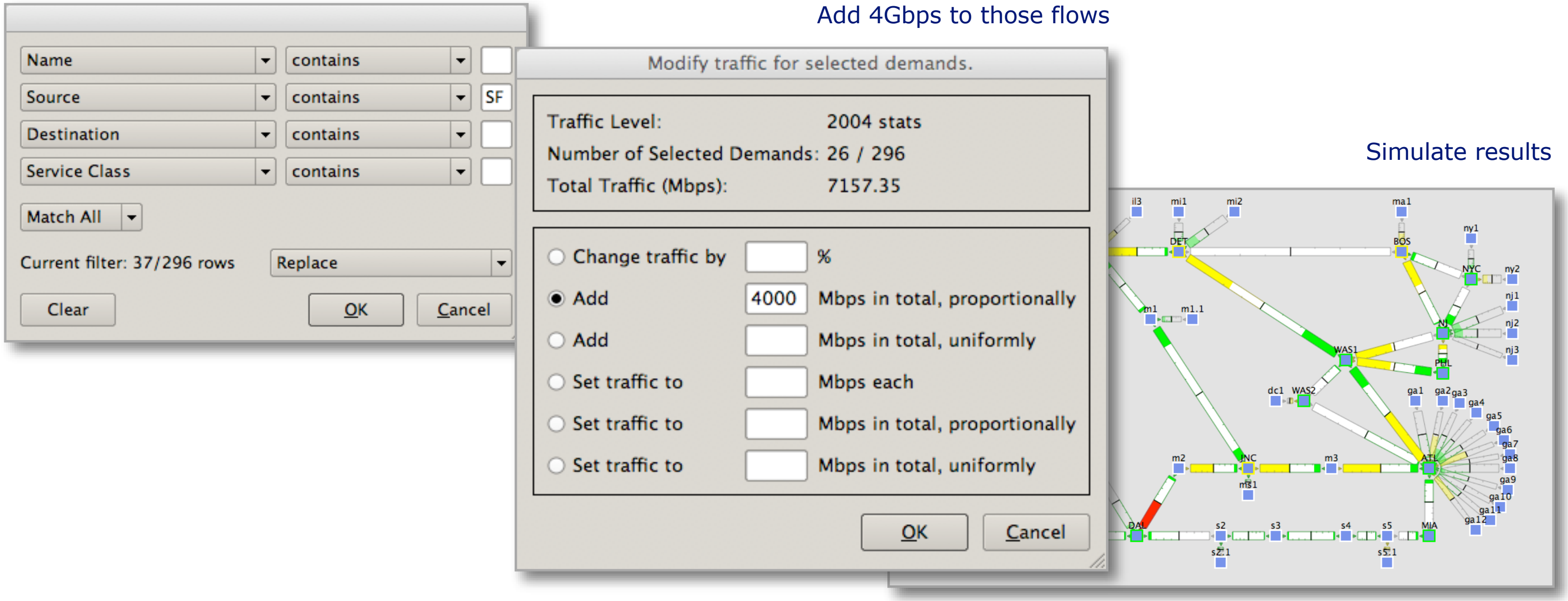
Evaluate New Customer

Scenario: Sales inquires whether network can support a 4 Gbps customer in SF

Identify flows for new customer

Add 4Gbps to those flows

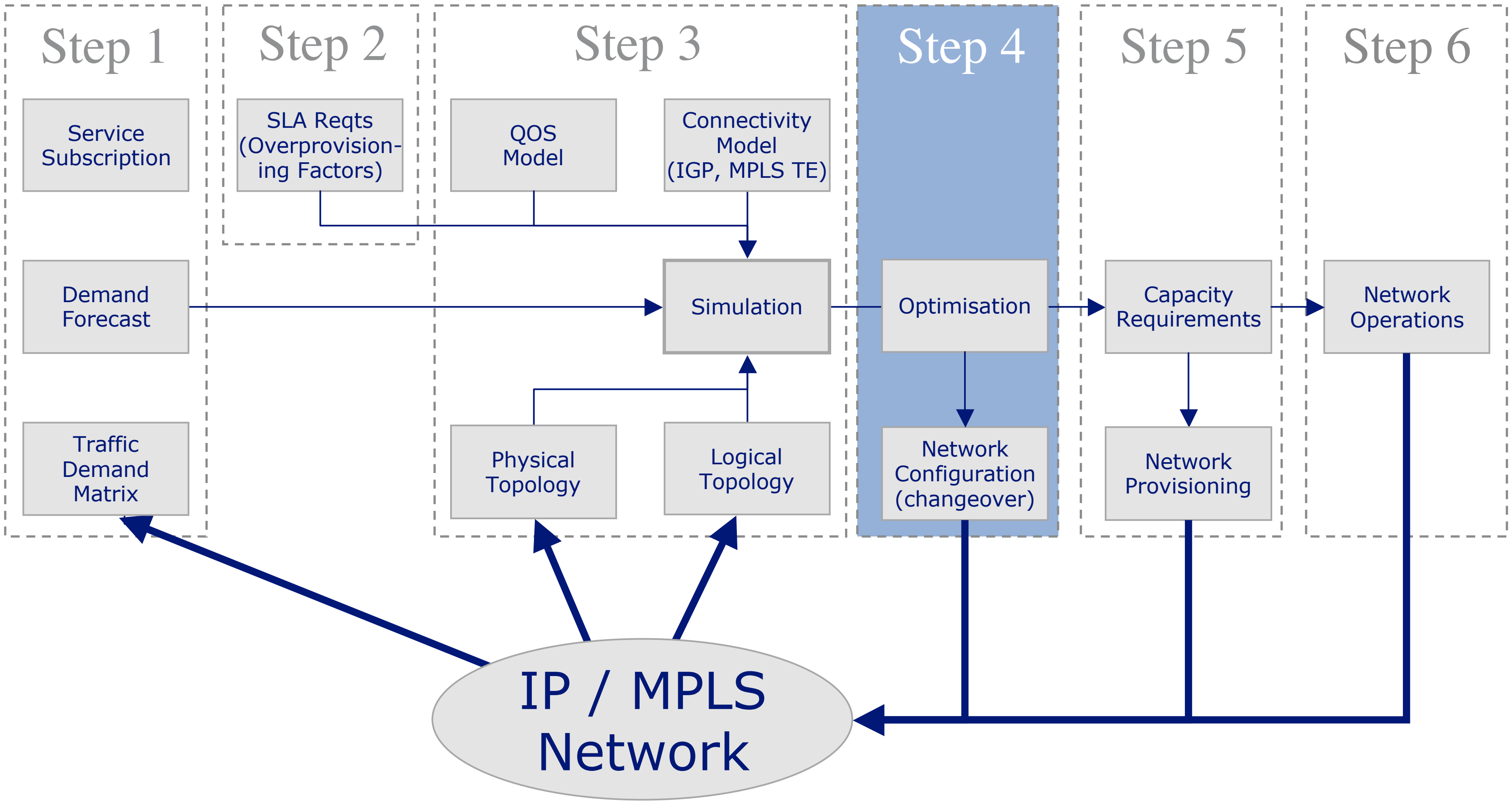
Simulate results



Congested link in **RED**

Network Planning Methodology

4. Traffic Engineering options and approaches: tactical, strategic, MPLS, IGP ...



cariden

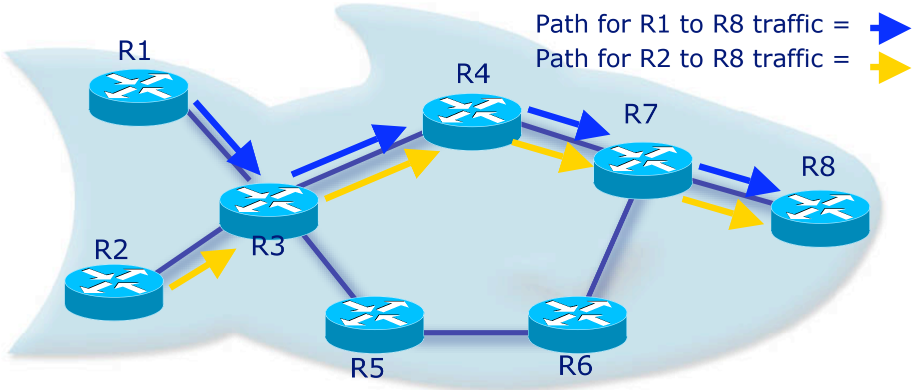
Network Optimisation

- Network Optimisation encompasses network engineering and traffic engineering
  - Network engineering
    - Manipulating your network to suit your traffic
  - Traffic engineering
    - Manipulating your traffic to suit your network

- Whilst network optimisation is an optional step, all of the preceding steps are essential for:
  - Comparing network engineering and TE approaches
  - MPLS TE tunnel placement and IP TE

IP Traffic Engineering: The Problem

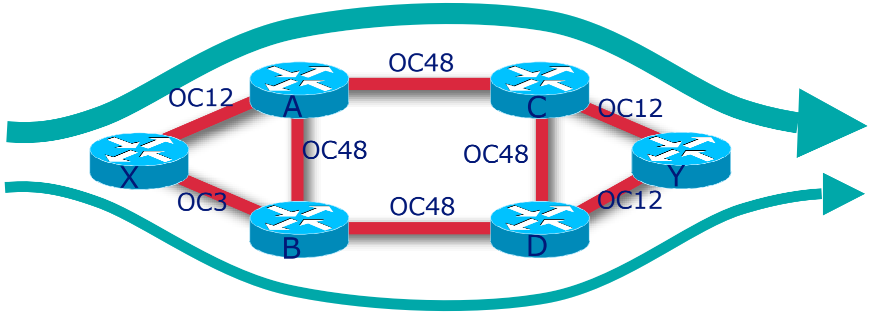
- Conventional IP routing uses pure destination-based forwarding where path computation is based upon a simple additive metric
  - Bandwidth availability is not taken into account



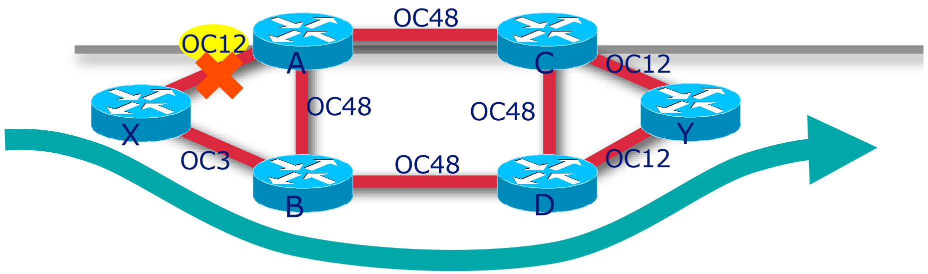
- Some links may be congested while others are underutilized
- The traffic engineering problem can be defined as an optimization problem
  - Definition – "*optimization problem*": A computational problem in which the objective is to find the best of all possible solutions
    - ➔ Given a fixed topology and a fixed source-destination matrix of traffic to be carried, what routing of flows makes most effective use of aggregate or per class (Diffserv) bandwidth?
      - » ➔ *How do we define most effective … ?*

IP Traffic Engineering: The objective

- What is the primary optimization objective?
  - Either …
    - minimizing maximum utilization in normal working (non-failure) case
  - Or …
    - minimizing maximum utilization under single element failure conditions
- Understanding the objective is important in understanding where different traffic engineering options can help and in which cases more bandwidth is required
  - Other optimization objectives possible: e.g. minimize propagation delay, apply routing policy …
- Ultimate measure of success is cost saving

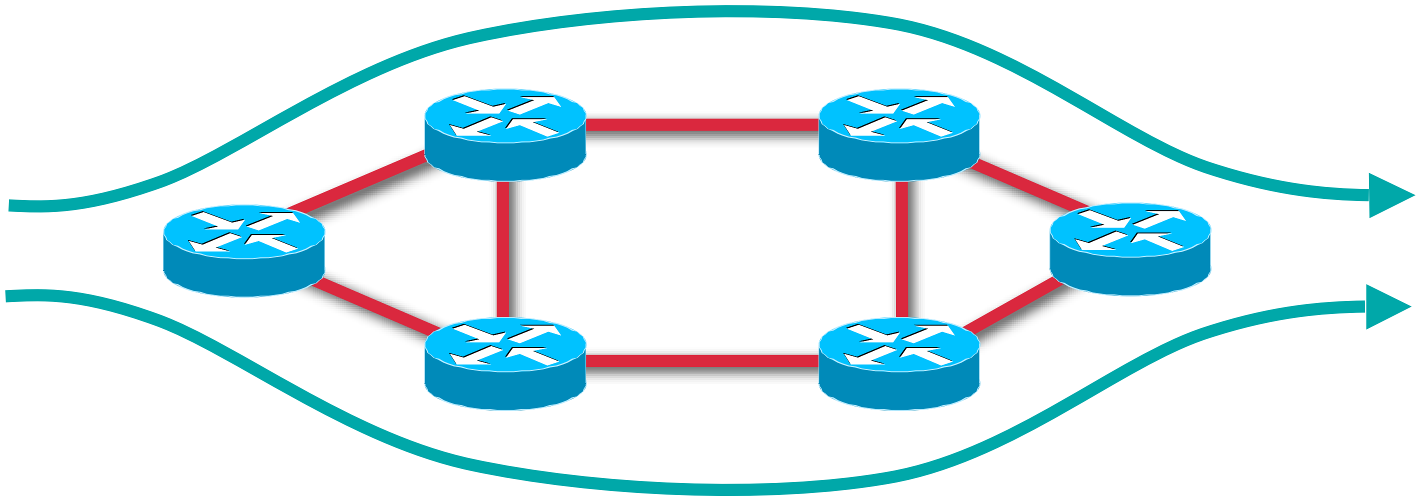


- In this asymmetrical topology, if the demands from X➔Y > OC3, traffic engineering can help to distribute the load when all links are working



- However, in this topology when optimization goal is to minimize bandwidth for single element failure conditions, if the demands from X➔Y > OC3, TE cannot help - must upgrade link X➔B

Traffic Engineering Limitations



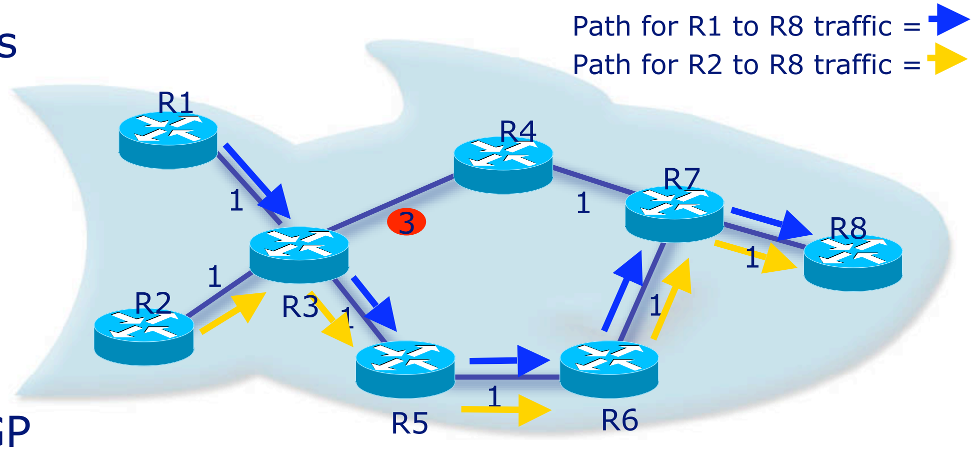
- TE cannot create capacity
  - e.g. "V-O-V" topologies allow no scope strategic TE if optimizing for failure case
    - Only two directions in each "V" or "O" region – no routing choice for minimizing failure utilization
- Other topologies may allow scope for TE in failure case
  - As case study later demonstrates

Traffic Engineering Approaches

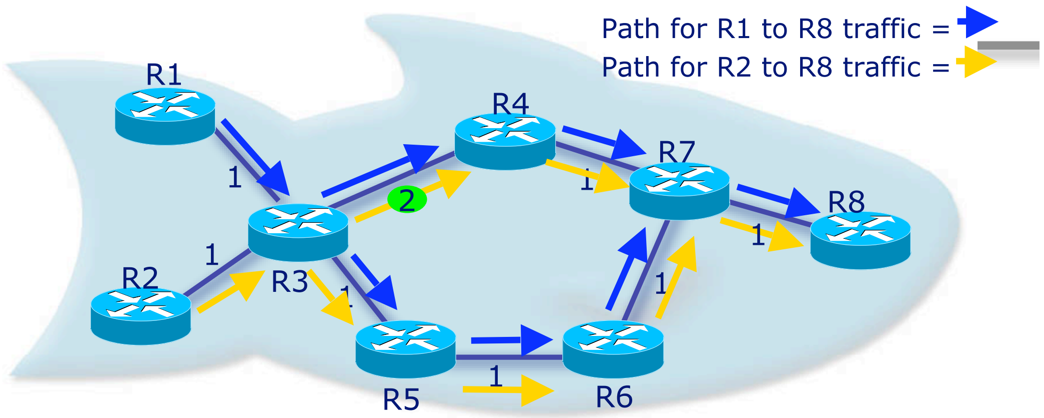
- Technology approaches:
  - MPLS TE
  - IGP Metric based TE
- Deployment models:
  - Tactical TE
    - Ad hoc approach aimed at mitigating specific current congestion spots
    - Short term operational/engineering process
    - Configured in response to failures, traffic changes
  - Strategic TE
    - Systematic approach aimed at cost savings, through traffic engineering the whole network
    - Medium term engineering/planning process
    - Configure in anticipation of failures, traffic changes
      - Resilient metrics, or
      - Primary and secondary disjoint paths, or
      - Dynamic tunnels, or …

IGP metric-based traffic engineering

- … but changing the link metrics will just move the problem around the network?



- …the mantra that tweaking IGP metrics just moves problem around is not generally true in practise
  - Note: IGP metric-based TE can use ECMP

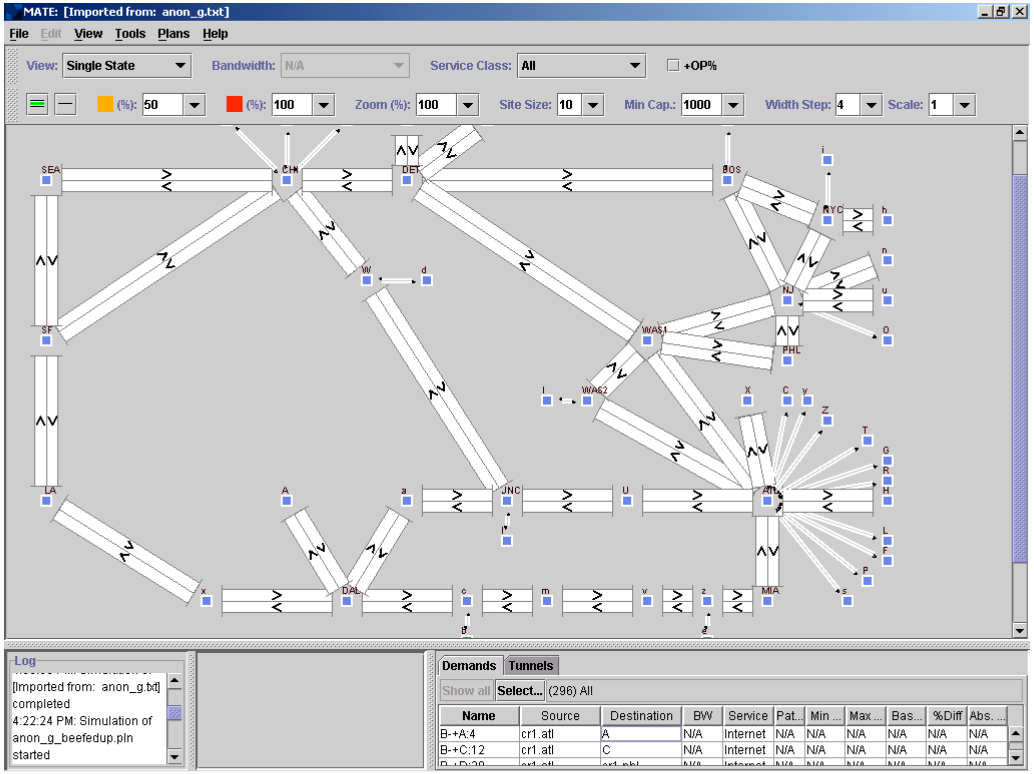


IGP metric-based traffic engineering

- Significant research efforts ...
  - B. Fortz, J. Rexford, and M. Thorup, “Traffic Engineering With Traditional IP Routing Protocols”, IEEE Communications Magazine, October 2002.
  - D. Lorenz, A. Ordi, D. Raz, and Y. Shavitt, “How good can IP routing be?”, DIMACS Technical, Report 2001-17, May 2001.
  - L. S. Buriol, M. G. C. Resende, C. C. Ribeiro, and M. Thorup, “A memetic algorithm for OSPF routing” in Proceedings of the 6th INFORMS Telecom, pp. 187188, 2002.
  - M. Ericsson, M. Resende, and P. Pardalos, “A genetic algorithm for the weight setting problem in OSPF routing” J. Combinatorial Optimization, volume 6, no. 3, pp. 299-333, 2002.
  - W. Ben Ameur, N. Michel, E. Gourdin et B. Liau. Routing strategies for IP networks. Telektronikk, 2/3, pp 145-158, 2001.
  - …

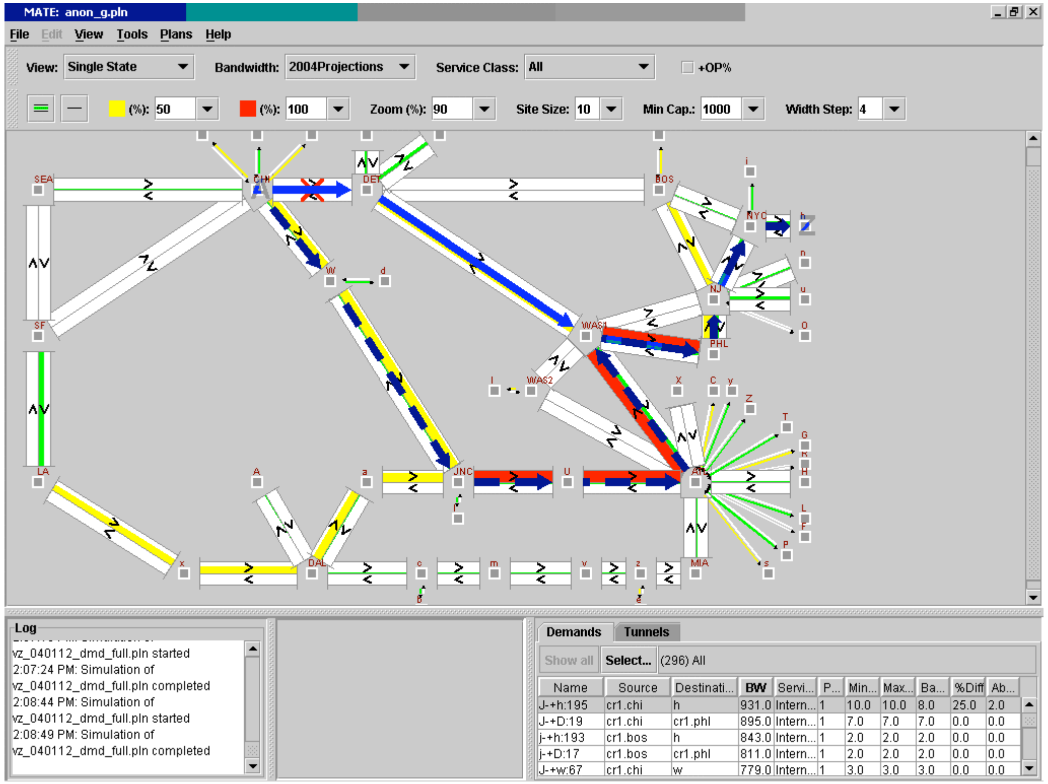
IGP metric-based traffic engineering: Case study

- Proposed OC-192 U.S. Backbone
- Connect Existing Regional Networks
- Anonymized (by permission)



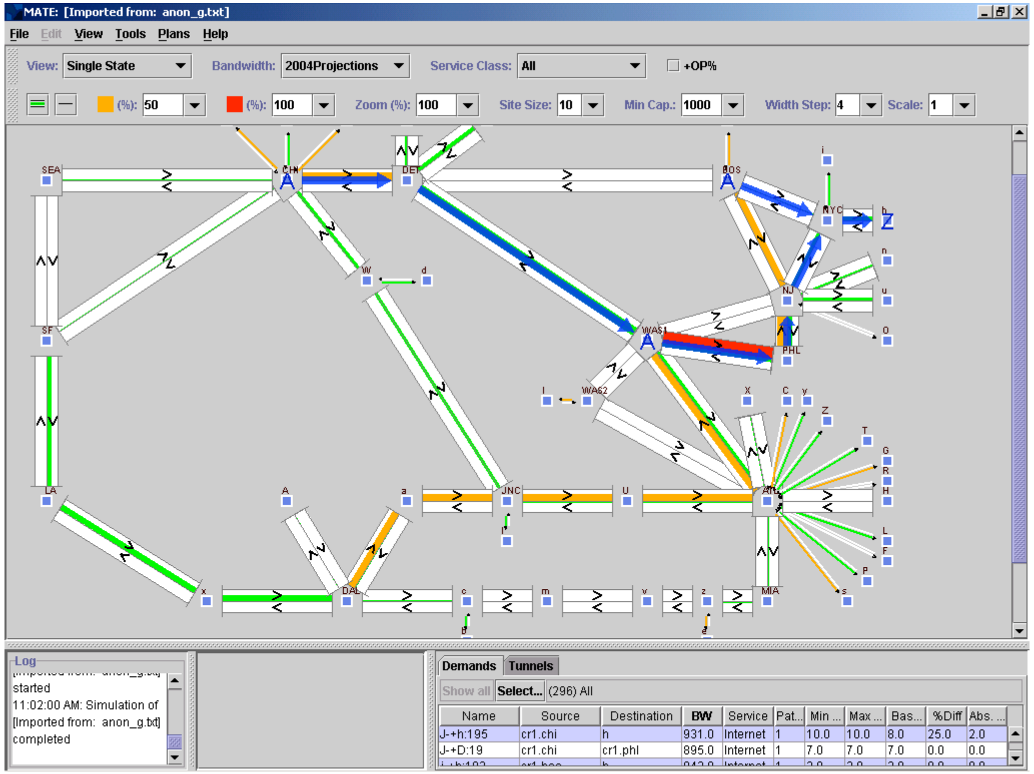
Metric TE Case Study: Plot Legend

- Squares ~ Sites (PoPs)
- Routers in Detail Pane (not shown here)
- Lines ~ Physical Links
  - Thickness ~ Speed
  - Color ~ Utilization
    - Yellow ≥ 50%
    - Red ≥ 100%
- Arrows ~ Routes
  - Solid ~ Normal
  - Dashed ~ Under Failure
- X ~ Failure Location



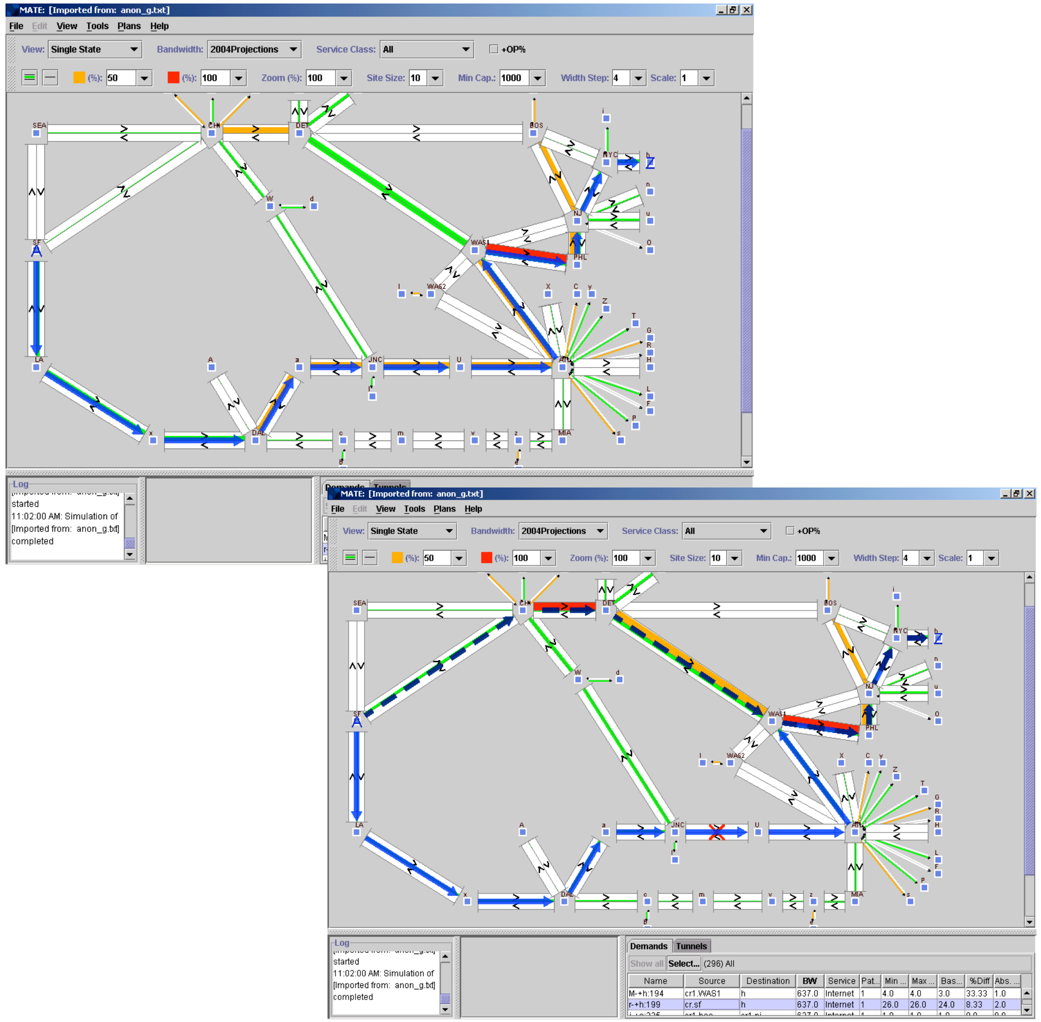
Metric TE Case Study: Traffic Overview

- Major Sinks in the Northeast
- Major Sources in CHI, BOS, WAS, SF
- Congestion Even with No Failure



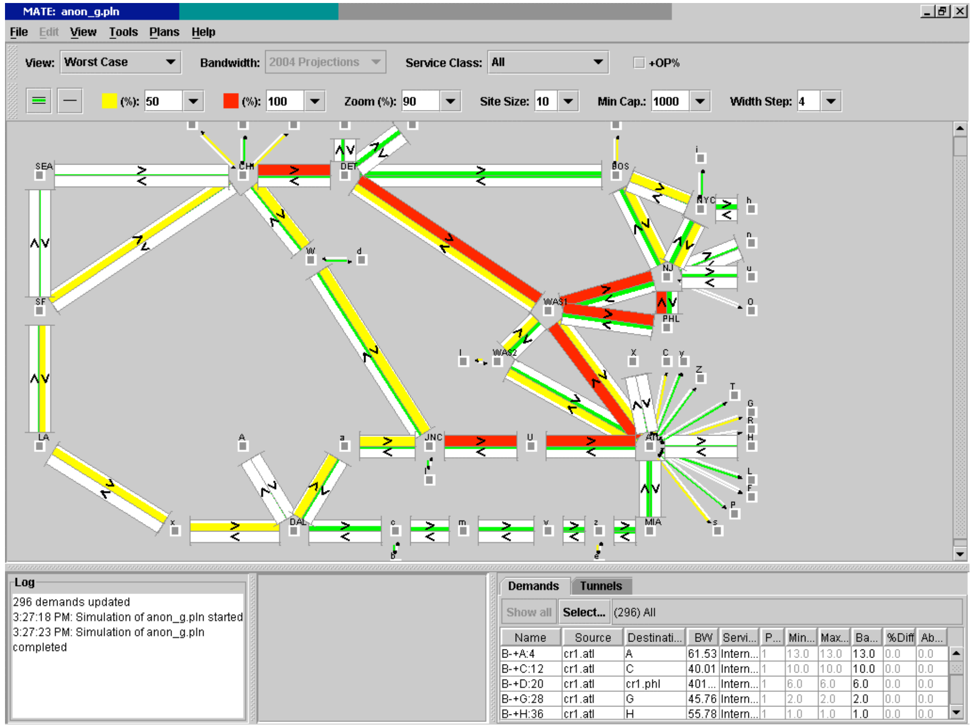
Metric TE Case Study: Manual Attempt at Metric TE

- Shift Traffic from Congested North
- Under Failure traffic shifted back North



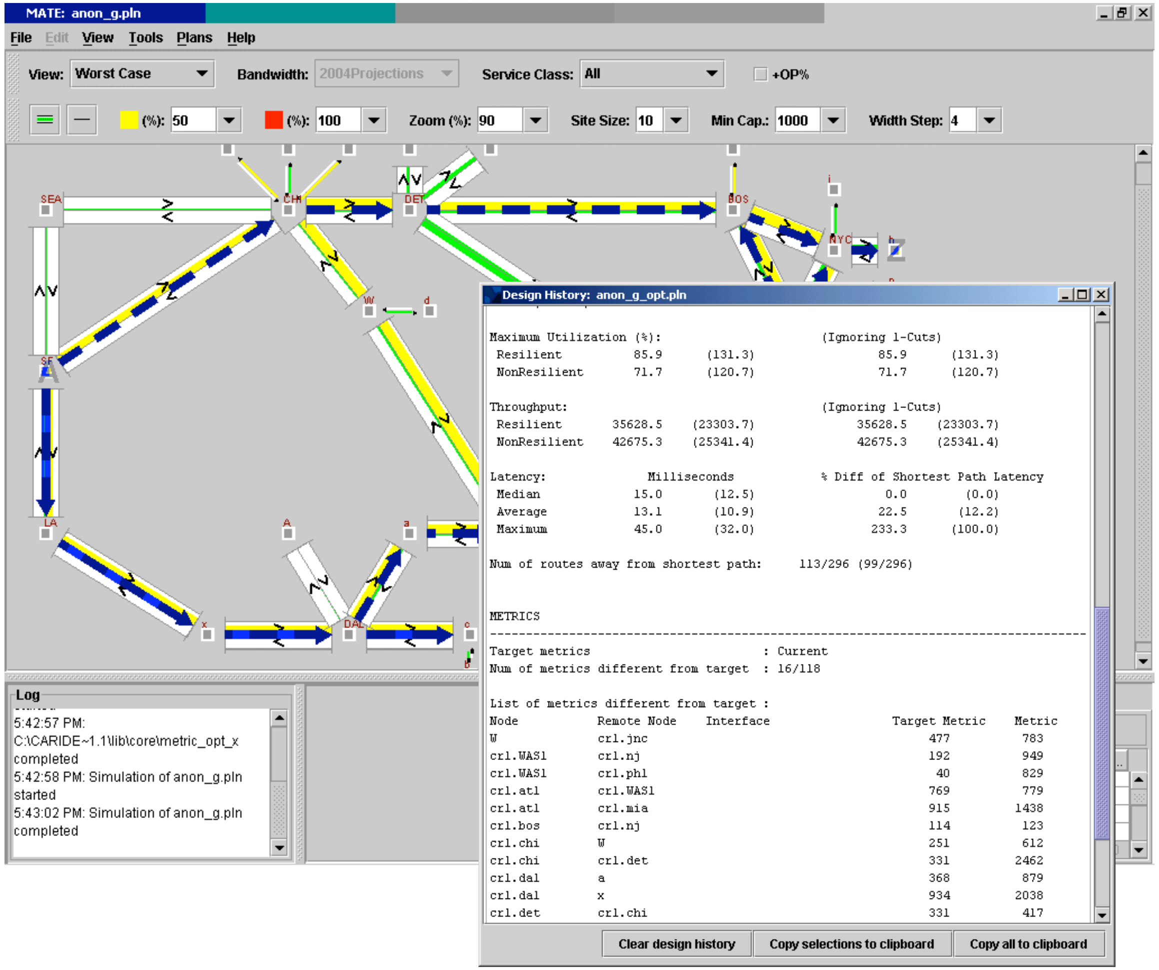
Metric TE Case Study: Worst Case Failure View

- Enumerate Failures
- Display Worst Case Utilization per Link
- Links may be under Different Failure Scenarios
- Central Ring+ Northeast Require Upgrade



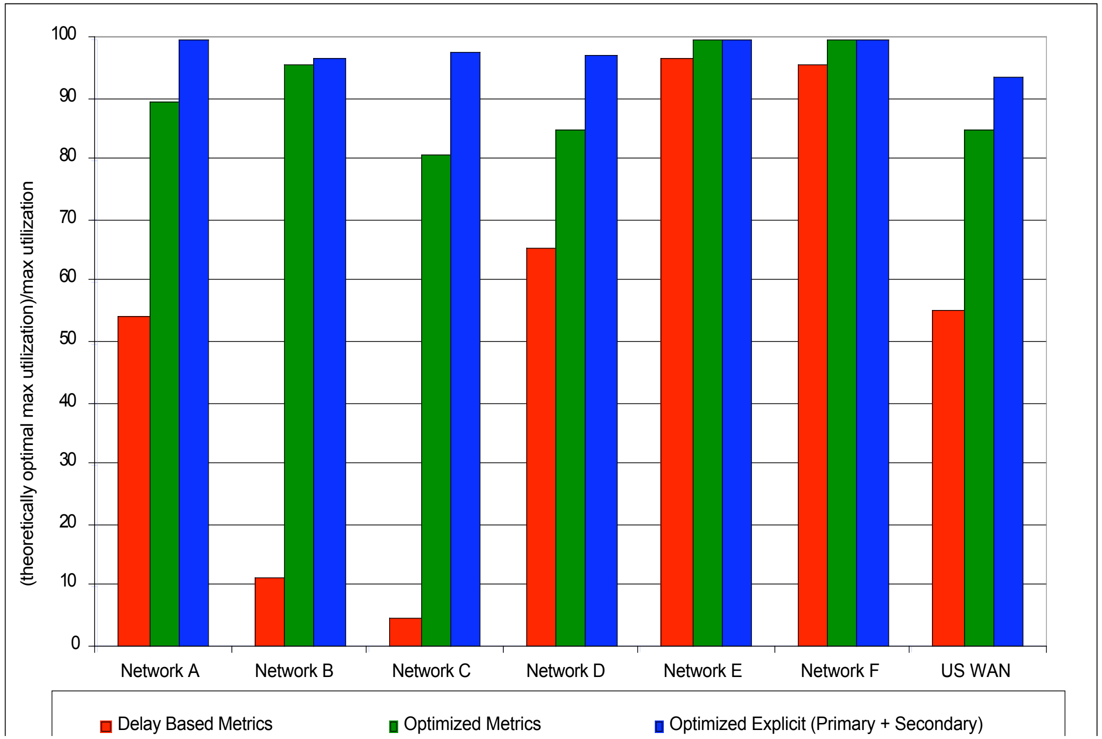
Metric TE Case Study: New Routing Visualisation

- ECMP in congested region
- Shift traffic to outer circuits
- Share backup capacity: outer circuits fail into central ones
- Change 16 metrics
- Remove congestion
  - Normal (121% -> 72%)
  - Worst case link failure (131% -> 86%)



Metric TE Case Study: Performance over Various Networks

- See: [Maghbouleh 2002]
- Study on Real Networks
- Single Set of Metrics Achieve 80-95% of Theoretical Best across Failures



MPLS TE deployment considerations

- Core or edge mesh
- Statically (explicit) or dynamically established tunnels
- Tunnel sizing
- Traffic sloshing

MPLS TE deployment considerations

- Statically (explicit) or dynamically established tunnels
  - Dynamic path option
    - Must specify bandwidths for tunnels
      - Otherwise defaults to IGP shortest path
    - Dynamic tunnels introduce indeterminism and cannot solve “tunnel packing” problem
      - Order of setup can impact tunnel placement
      - Each head-end only has a view of their tunnels
      - Tunnel prioritisation scheme can help – higher priority for larger tunnels
  - Static – explicit path option
    - More deterministic, and able to provide better solution to “tunnel packing” problem
      - Offline system has view of all tunnels from all head-ends
    - If strategic approach then computer-aided tools can ease the task of primary tunnel placement

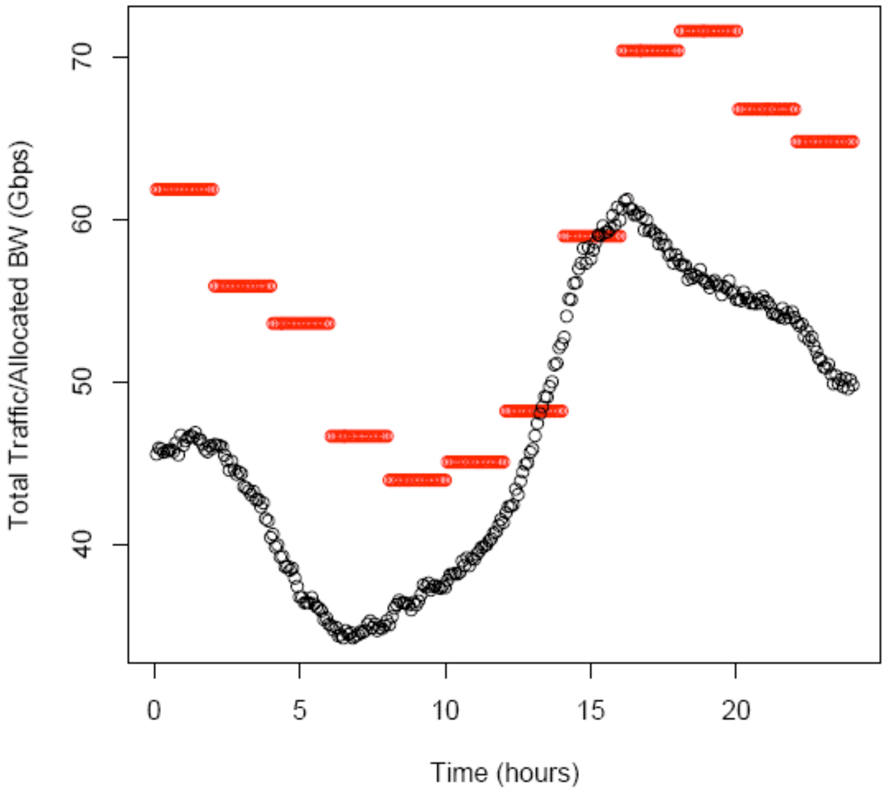
Tunnel Sizing

- Tunnel sizing is key …
  - Needless congestion if actual load >> reserved bandwidth
  - Needless tunnel rejection if reservation >> actual load
    - Enough capacity for actual load but not for the tunnel reservation
- Actual heuristic for tunnel sizing will depend upon dynamism of tunnel sizing
  - Need to set tunnel bandwidths dependent upon tunnel traffic characteristic over optimisation period

Tunnel Sizing

- Online vs. offline sizing:
  - Online sizing: autobandwidth
    - Router automatically adjusts reservation (up or down) based on traffic observed in previous time interval
    - Tunnel bandwidth is not persistent (lost on reload)
    - Can suffer from "bandwidth lag"
  - Offline sizing
    - Statically set reservation to percentile (e.g. P95) of expected max load
    - Periodically readjust – not in real time, e.g. daily, weekly, monthly

"online sizing: bandwidth lag"

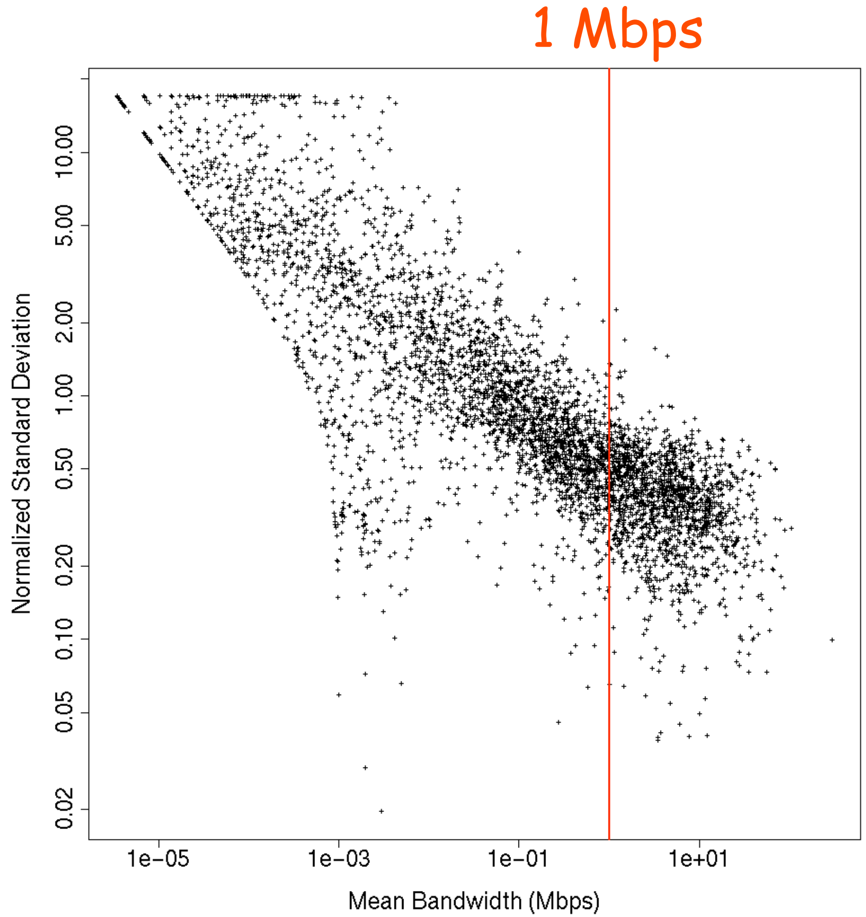


Tunnel Sizing

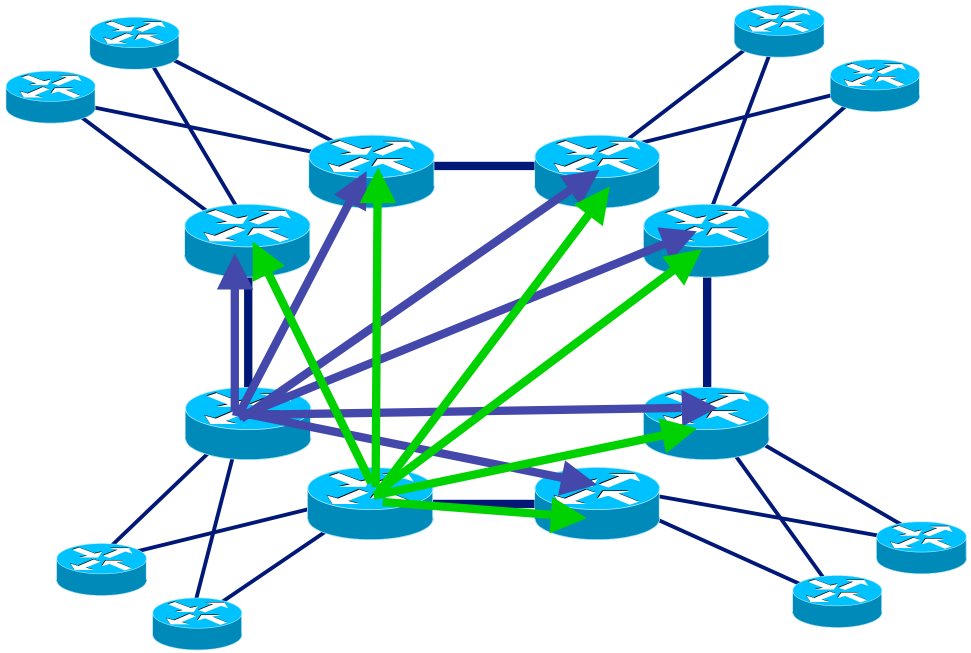
- When to re-optimise?
  - Event driven optimisation, e.g. on link or node failures
    - Won’t re-optimise due to tunnel changes
  - Periodically
    - Tunnel churn if optimisation periodicity high
    - Inefficiencies if periodicity too low
    - Can be online or offline

SP Case Study (Global Crossing) Variance vs. Bandwidth [Telkamp 2003]

- Around 8000 demands between core routers
- Most traffic carried by (relatively) few big demands
  - 97% of traffic is carried by the demands larger than 1 Mbps (20% of the demands!)
- Relative variance decreases with increasing bandwidth
- High-bandwidth demands are well-behaved (predictable) during the course of a day and across days
- Little motivation for dynamically changing routing during the course of a day

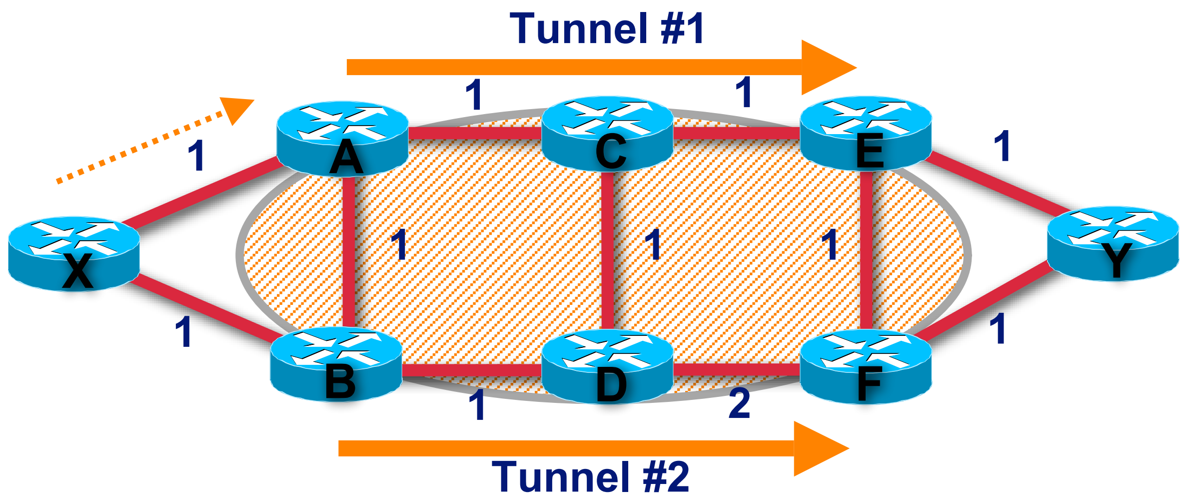


Strategic Deployment: Core Mesh



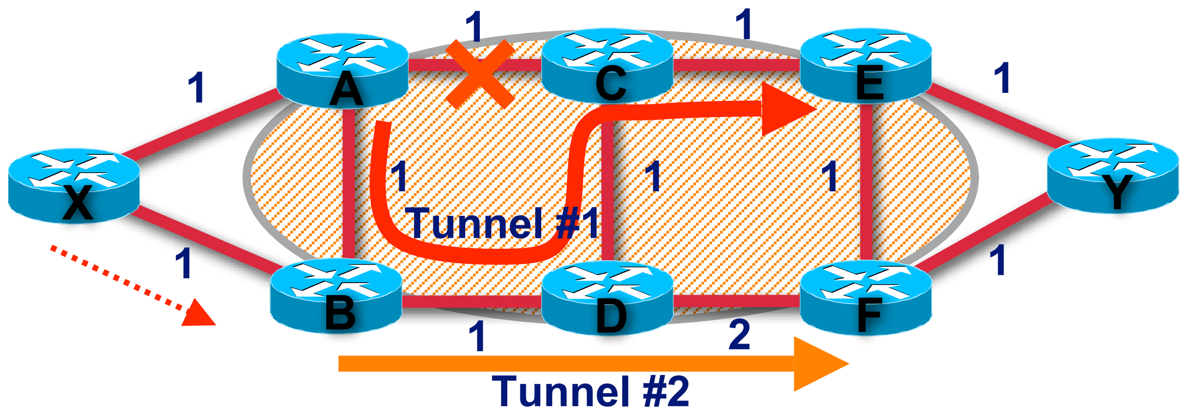
- Reduces number of tunnels required
- Can be susceptible to “traffic-sloshing”

Traffic “sloshing”



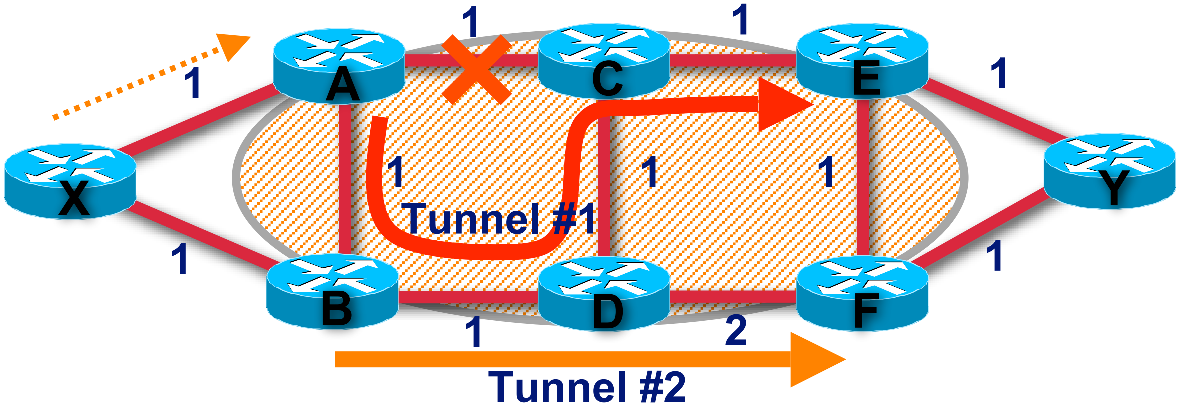
- In normal case:
  - –For traffic from X ➔ Y, router X IGP will see best path via router A
  - –Tunnel #1 will be sized for X ➔ Y demand
  - –If bandwidth is available on all links, Tunnel from A to E will follow path A ➔ C ➔ E

Traffic "sloshing"



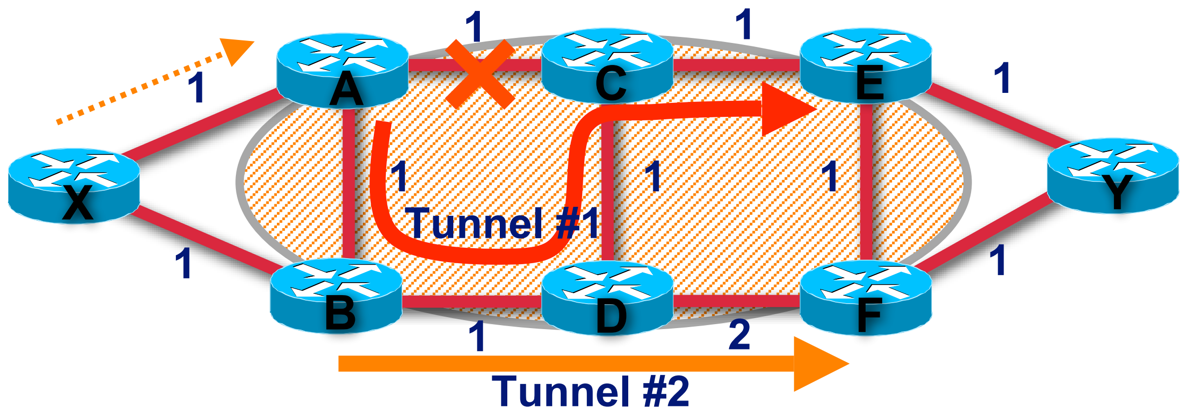
- In failure of link A-C:
  - –For traffic from X ➔ Y, router X IGP will now see best path via router B
  - –However, if bandwidth is available, tunnel from A to E will be re-established over path A ➔ B ➔ D ➔ C ➔ E
  - –Tunnel #2 will not be sized for X ➔ Y demand
  - –Bandwidth may be set aside on link A ➔ B for traffic which is now taking different path

Traffic "sloshing"



- Forwarding adjacency (FA) could be used to overcome traffic sloshing
  - –Normally, a tunnel only influences the FIB of its head-end and other nodes do not see it
  - –With FA the head-end advertises the tunnel in its IGP LSP
    - •Tunnel #1 could always be made preferable over tunnel #2 for traffic from X ➔ Y
- Holistic view of traffic demands (core traffic matrix) and routing (in failures if necessary) is necessary to understand impact of TE

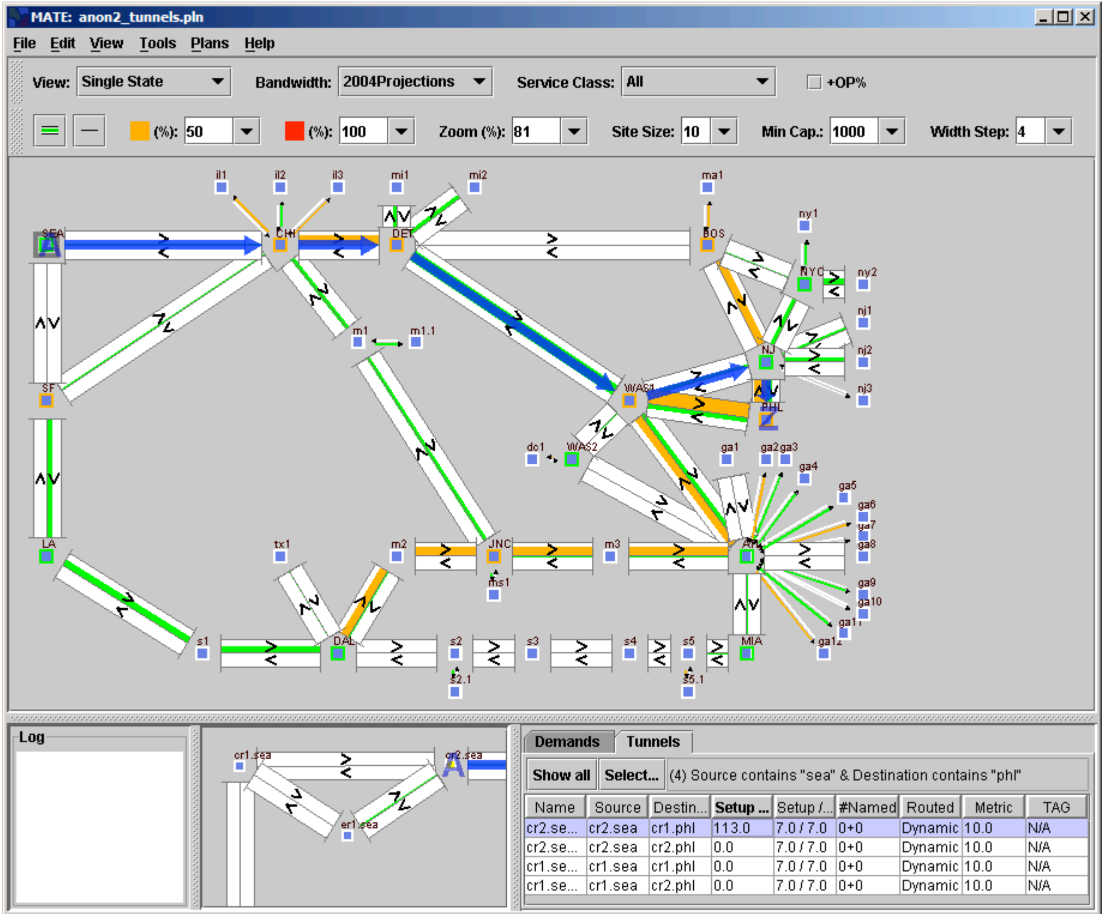
Traffic “sloshing”



- Forwarding adjacency could be used to overcome traffic sloshing
  - –Normally, a tunnel only influences the FIB of its head-end
    - •other nodes do not see it
  - –With Forwarding Adjacency the head-end advertises the tunnel in its IGP LSP
    - •Tunnel #1 could always be made preferable over tunnel #2 for traffic from X ➔ Y

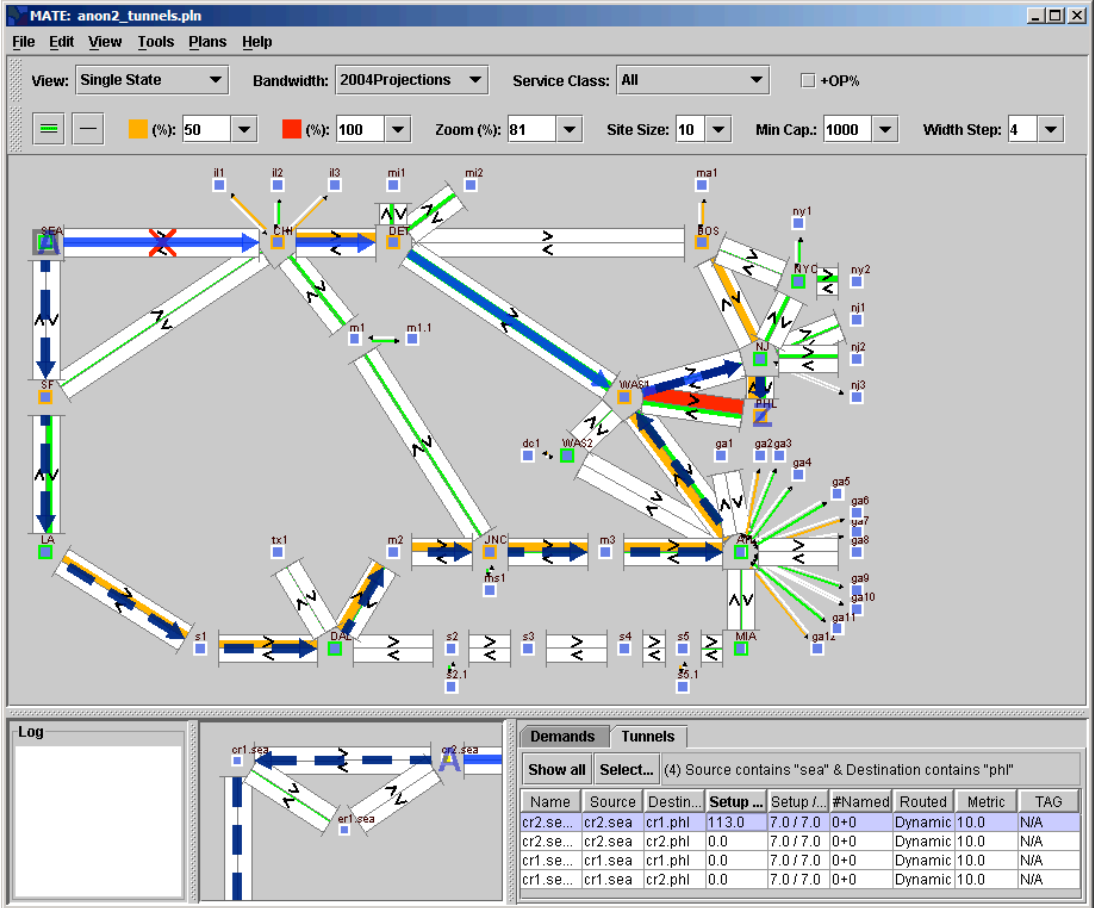
Traffic “sloshing”: A Real Example (I)

- 2 core routers in SEA
- X 2 core routers in PHL
- = 4 tunnels between all pairs
- One of these pairs has the shortest IGP path between them
- So all traffic from SEA-PHL goes on this tunnel



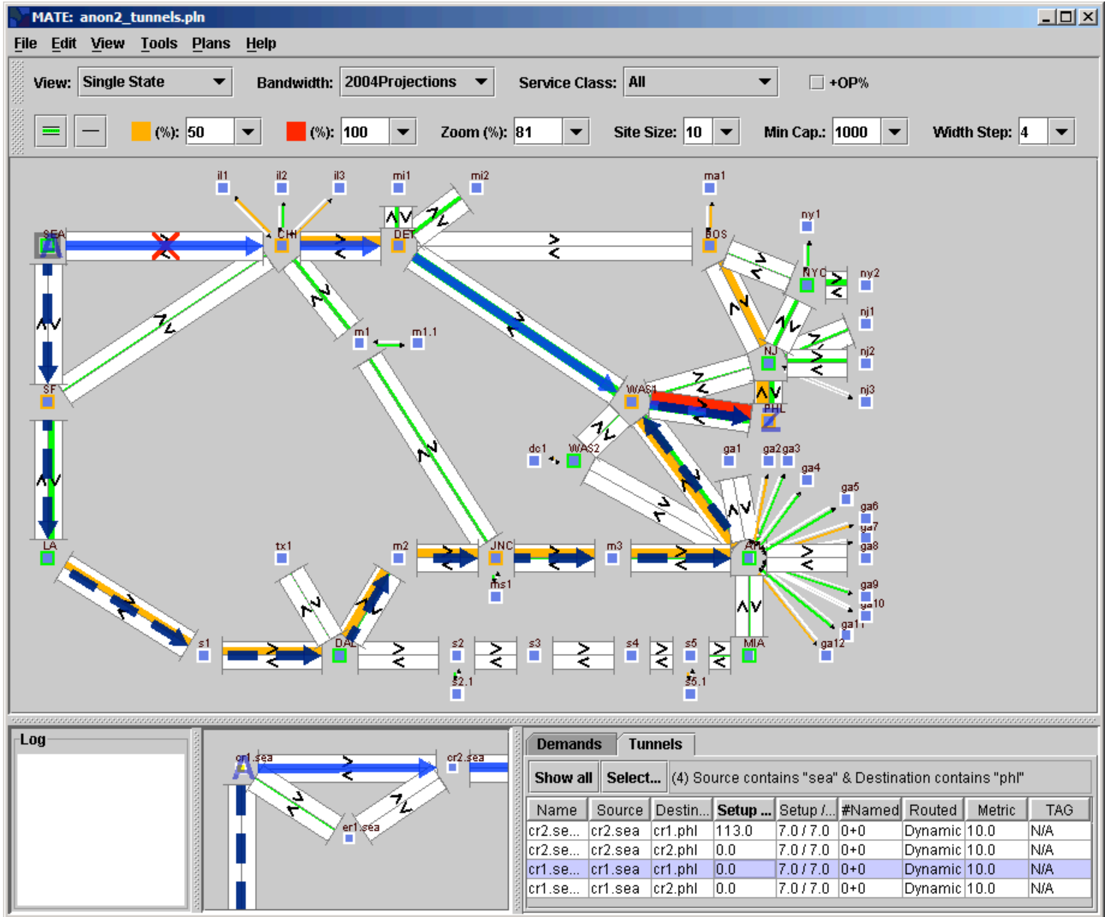
Traffic “sloshing”: A Real Example (II)

- This tunnel reserves enough space for all traffic through it.
- So under failure, finds alternate path avoiding congested links.



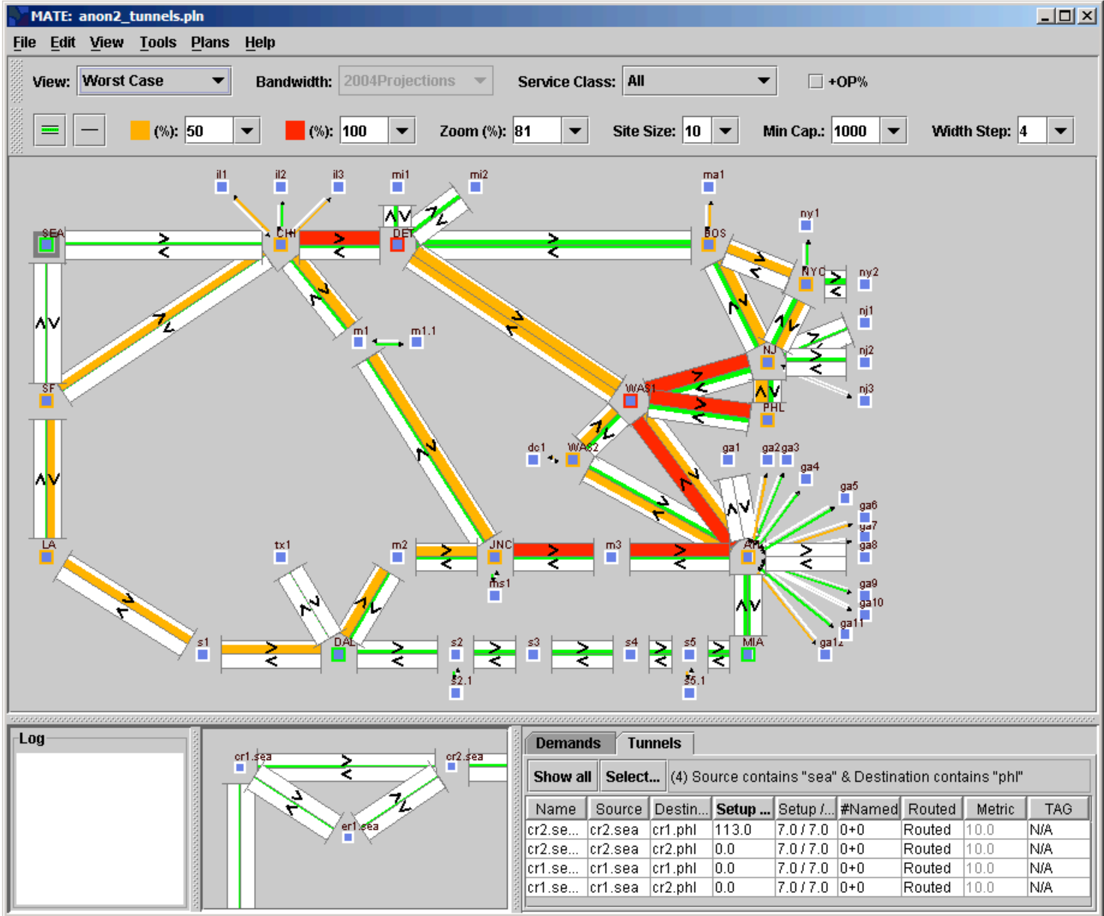
Traffic "sloshing": A Real Example (III)

- BUT, under failure a different pair of core routers is now closest by IGP metric
- So traffic "sloshes" to new tunnel
- New tunnel has zero bandwidth reserved, so has taken congested path.
- Traffic in new tunnel congests network further.



Traffic "sloshing": A Real Example (IV)

- Worst-case view: "sloshing" causes congestion under failure in many circuits.
- cf: Metric-based optimization on same network. Maximum utilization = 86% under any circuit failure.

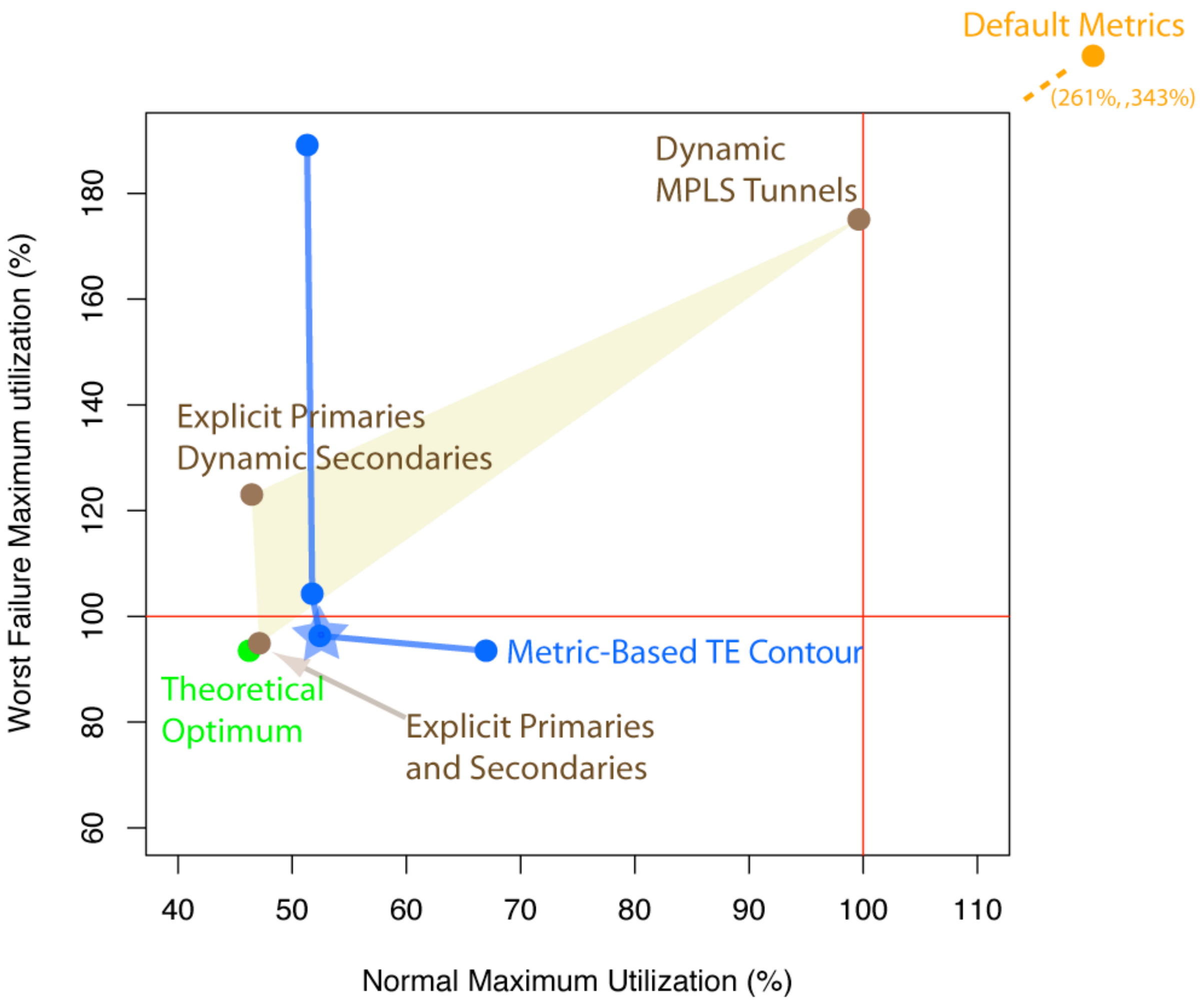


Traffic Engineering Summary

- Need to define whether optimising for working case or failure case
- Need to know traffic matrix to be able to simulate and compare potential approaches
- Deployment choices
  - Tactical vs. strategic
  - IGP metric based TE (works for IP and MPLS LDP)
  - RSVP-TE
    - Choice of core or edge mesh
    - Explicit path options can be more deterministic/optimal, but require offline tool
    - Offline tunnel sizing allows most control
    - Re-optimisation O(days) is generally sufficient
    - Use same tunnel sizing heuristic as is used for capacity planning

TE Case Study 1

Martin Horneffer,
Deutsche Telekom,
Nanog 33

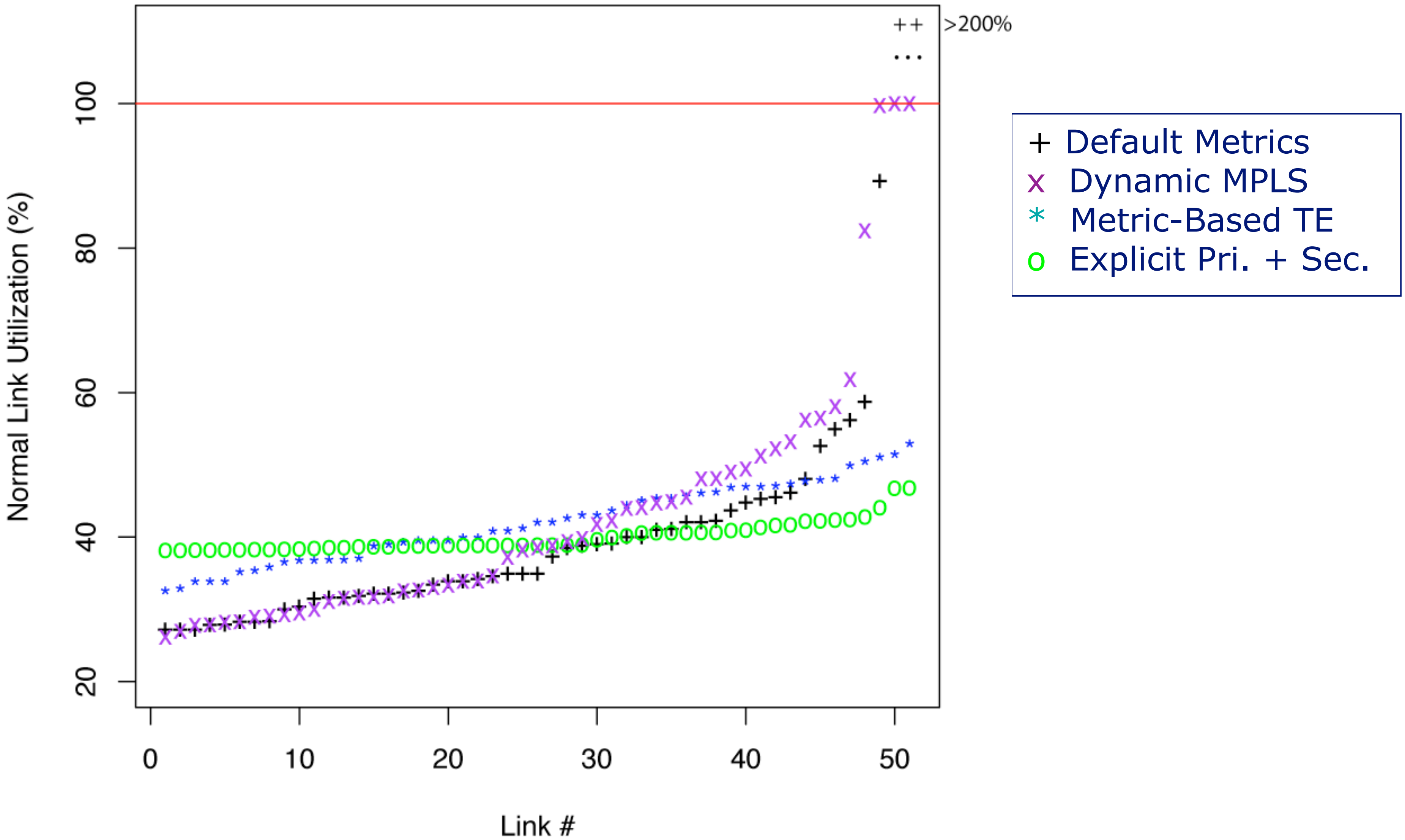


TE Case Study 2

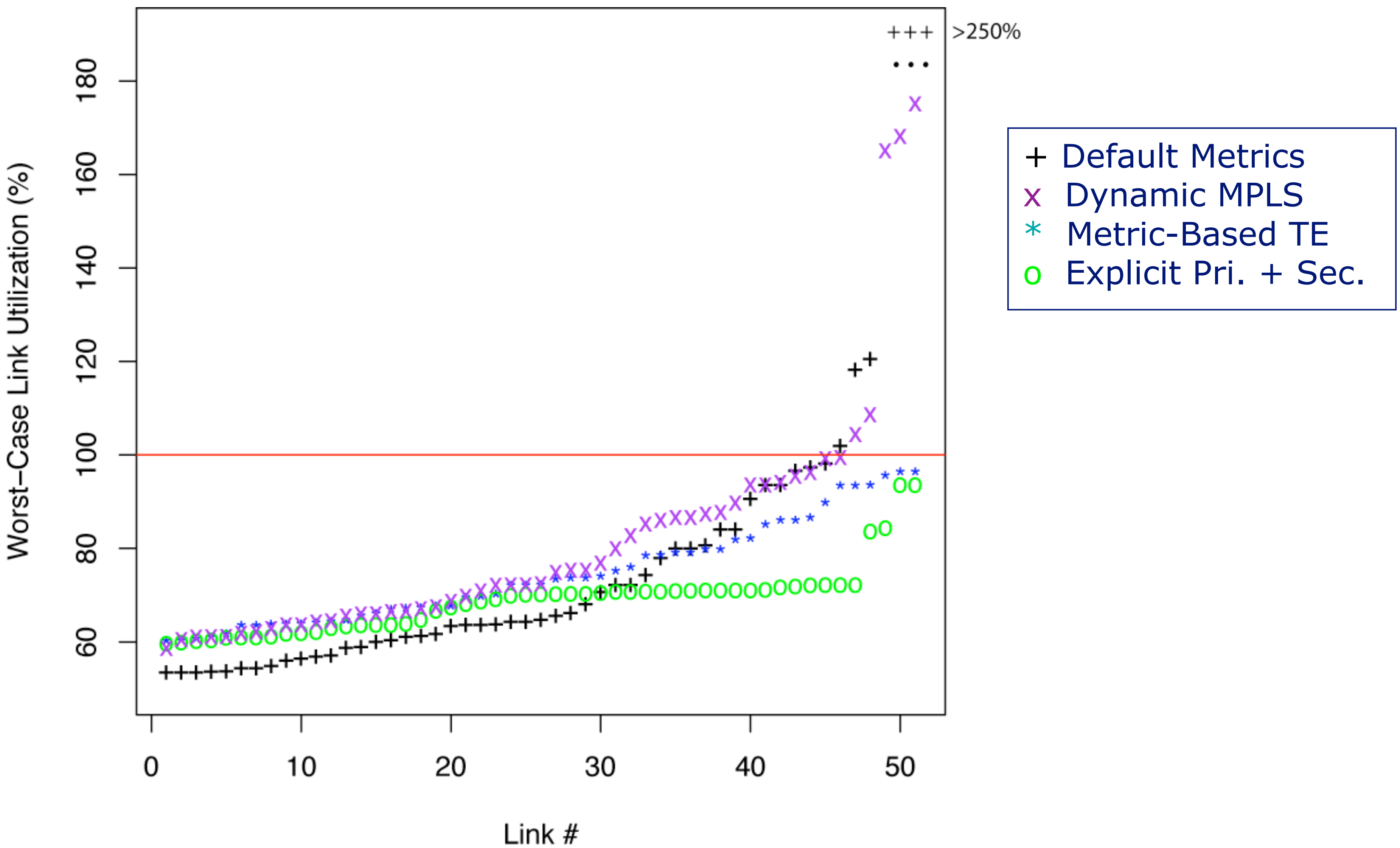
- Anonymous network…
- TE Options:
  - Dynamic MPLS
    - Mesh of CSPF tunnels in the core network
    - "Sloshing" causes congestion under failure scenarios
  - Metric Based TE
  - Explicit Pri. + Sec. LSPs
  - Failures Considered
    - Single-circuit, circuit+SRLG, circuit+SRLG+Node
    - Plot is for single-circuit failures

- Cariden MATE software for simulations and optimizations

Top 50 Utilized Links (normal)

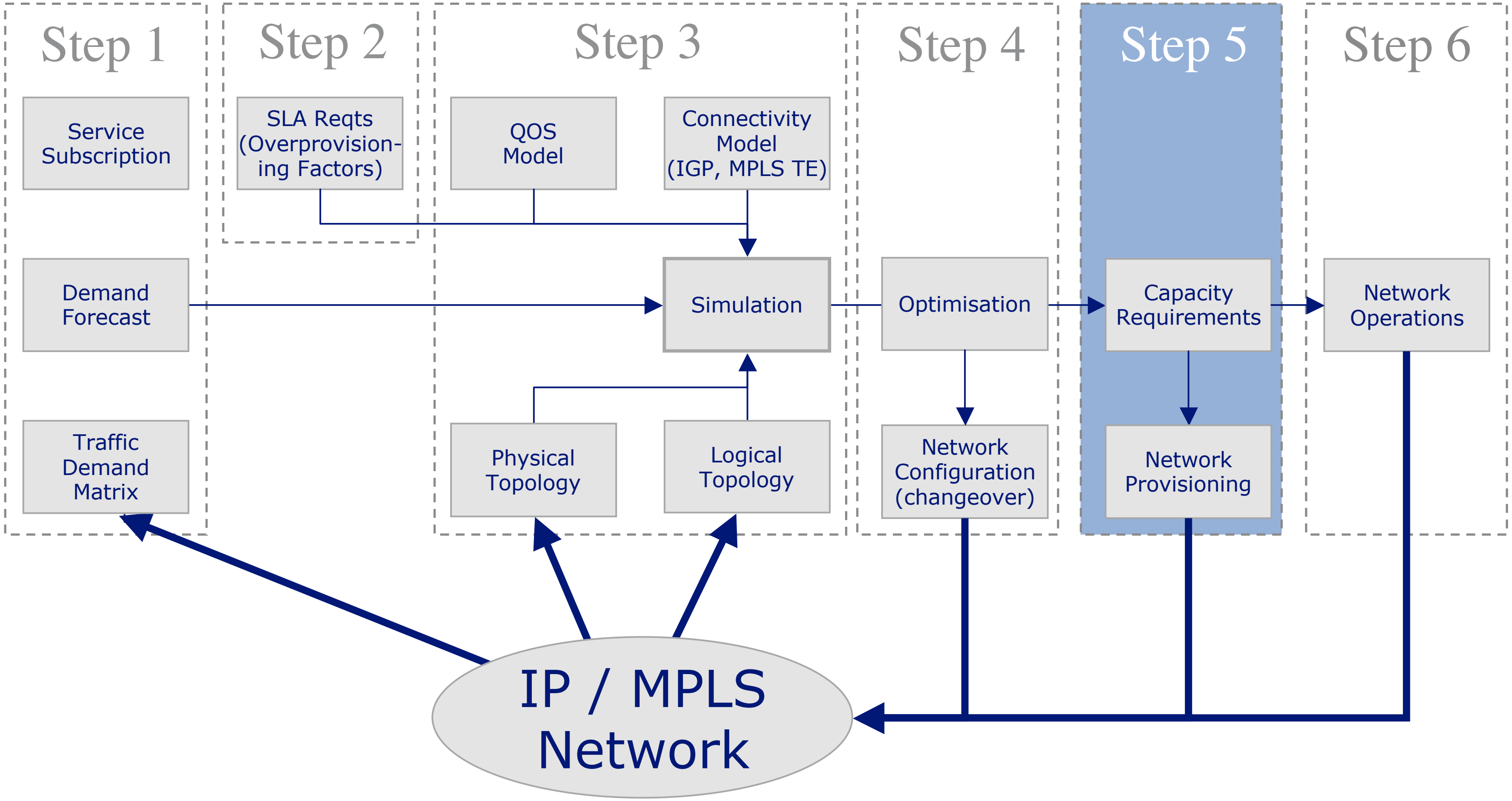


Top 50 Utilized Links (under failure)



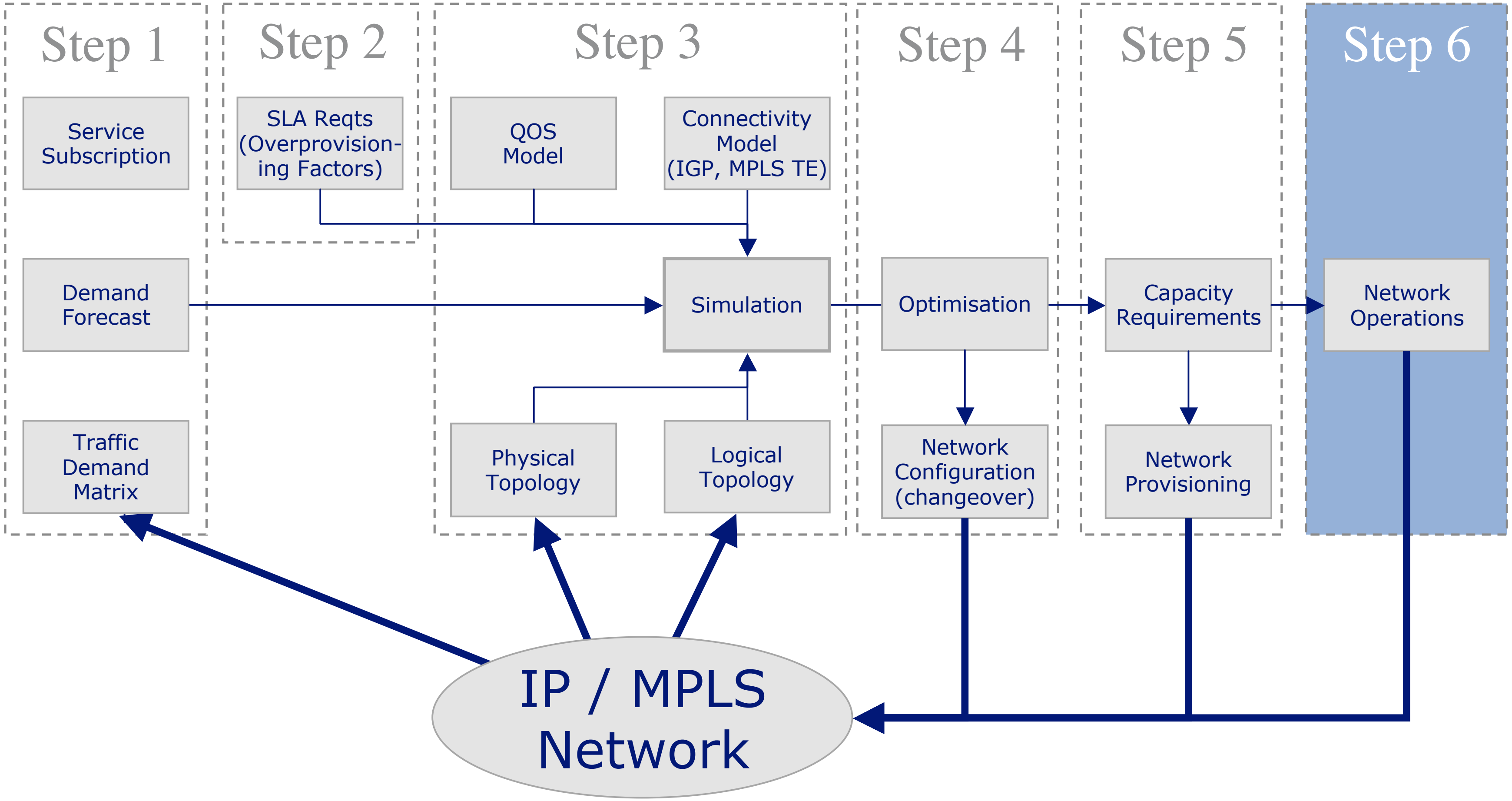
Network Planning Methodology

5. Network capacity provisioning



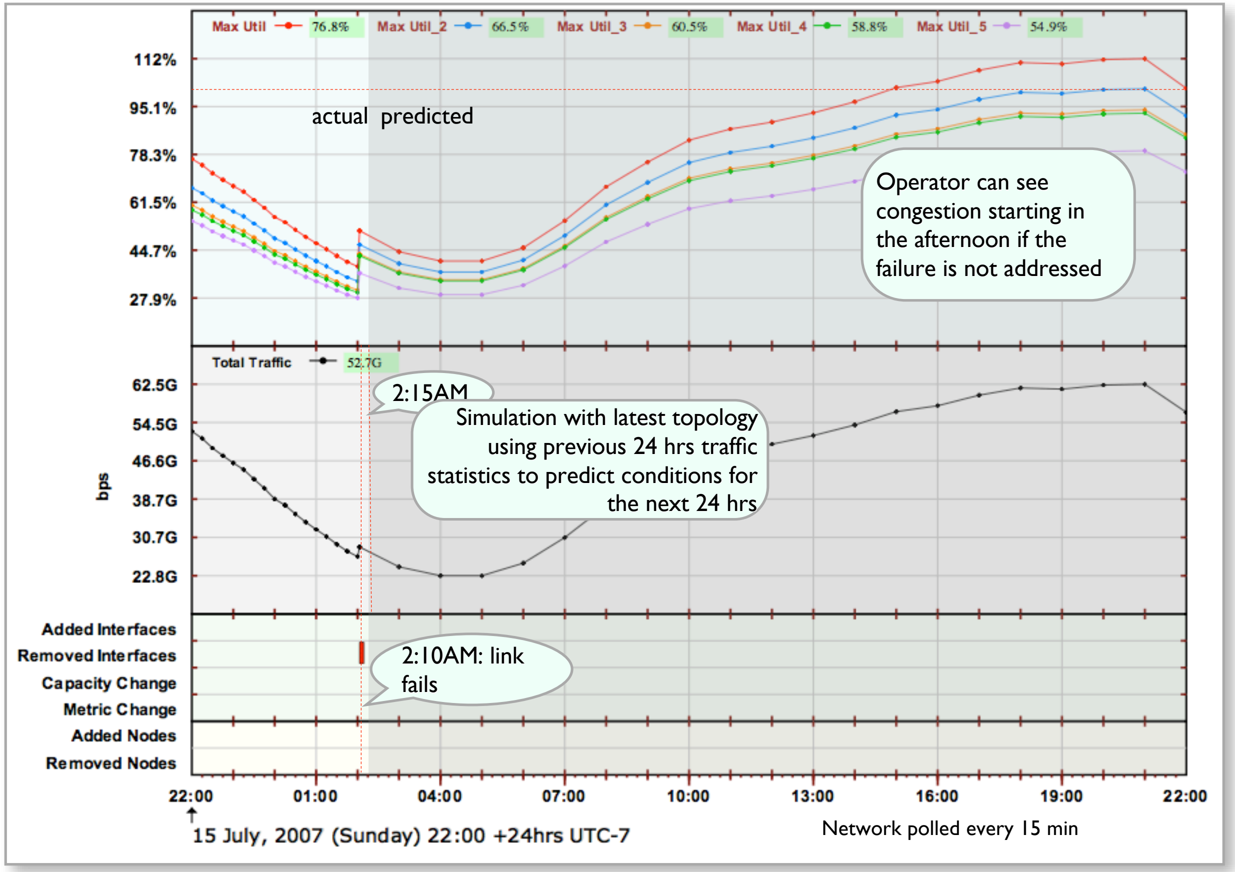
Network Planning Methodology

6. Where planning meets operations



Where planning meets operations

Scenario: Failure at 2:10AM, how severe is the impact?



Same principal could be applied for data from previous week or month, or a combination.

References

- Cao et al. 2002
  - Cao, J., W.S. Cleveland, D. Lin, D.X. Sun, Internet Traffic Tends Towards Poisson and Independent as the Load Increases. In Nonlinear Estimation and Classification, New York, Springer-Verlag, 2002
- Claise 2003
  - Benoit Claise, Traffic Matrix: State of the Art of Cisco Platforms, Intimate 2003 Workshop in Paris, June 2003
  - http://www.employees.org/~bclaise/
- Filsfils and Evans 2005
  - Clarence Filsfils and John Evans, "Deploying Diffserv in IP/MPLS Backbone Networks for Tight SLA Control", IEEE Internet Computing*, vol. 9, no. 1, January 2005, pp. 58-65
  - http://www.employees.org/~jevans/papers.html
- Fraleigh et al. 2003
  - Chuck Fraleigh, Fouad Tobagi, Christophe Diot, Provisioning IP Backbone Networks to Support Latency Sensitive Traffic, Proc. IEEE INFOCOM 2003, April 2003
- Horneffer 2005
  - Martin Horneffer, "IGP Tuning in an MPLS Network", NANOG 33, February 2005, Las Vegas
- Maghbouleh 2002
  - Arman Maghbouleh, "Metric-Based Traffic Engineering: Panacea or Snake Oil? A Real-World Study", NANOG 26, October 2002, Phoenix
  - http://www.cariden.com/technologies/papers.html
- Maghbouleh 2007
  - Arman Maghbouleh, "Traffic Matrices for IP Networks: NetFlow, MPLS, Estimation, Regression", Preparing for the Future of the Internet, Network Information Center, Mexico, November 29, 2007
  - http://www.cariden.com/technologies/papers.html

References

- Schnitter and Horneffer 2004
  - S. Schnitter, T-Systems; M. Horneffer, T-Com. "Traffic Matrices for MPLS Networks with LDP Traffic Statistics." Proc. Networks 2004, VDE-Verlag 2004.
- Telkamp 2003
  - Thomas Telkamp, "Backbone Traffic Management", Asia Pacific IP Experts Conference (Cisco), November 4th, 2003, Shanghai, P.R. China
  - http://www.cariden.com/technologies/papers.html
- Telkamp 2006
  - T. Telkamp, "Peering Planning Cooperation without Revealing Confidential Information." RIPE 52, Istanbul, Turkey, April 2006
  - http://www.cariden.com/technologies/papers.html
- Telkamp 2007
  - Thomas Telkamp, Best Practices for Determining the Traffic Matrix in IP Networks V 3.0, NANOG 39, February 2007, Toronto
  - http://www.cariden.com/technologies/papers.html
- Vardi 1996
  - Y. Vardi. "Network Tomography: Estimating Source-Destination Traffic Intensities from Link Data." J.of the American Statistical Association, pages 365–377, 1996.
- Zafer and Sirin 1999
  - Zafer Sahinoglu and Sirin Tekinay, On Multimedia Networks: "Self-Similar Traffic and Network Performance", IEEE Communications Magazine, January 1999
- Zhang et al. 2004
  - Yin Zhang, Matthew Roughan, Albert Greenberg, David Donoho, Nick Duffield, Carsten Lund, Quynh Nguyen, and David Donoho, "How to Compute Accurate Traffic Matrices for Your Network in Seconds", NANOG29, Chicago, October 2004.
  - See also: http://public.research.att.com/viewProject.cfm?prjID=133/

cariden

the economics of network control

- Web: http://www.cariden.com
- Phone: +1 650 564 9200
- Fax: +1 650 564 9500
- Address: 888 Villa Street, Suite 500
  Mountain View, CA 94041
  USA