the economics of network control

Best Practices in Network Planning

PLNOG 3 – 11 Sep 2009

John Evans

johnevans@cariden.com

Trends:

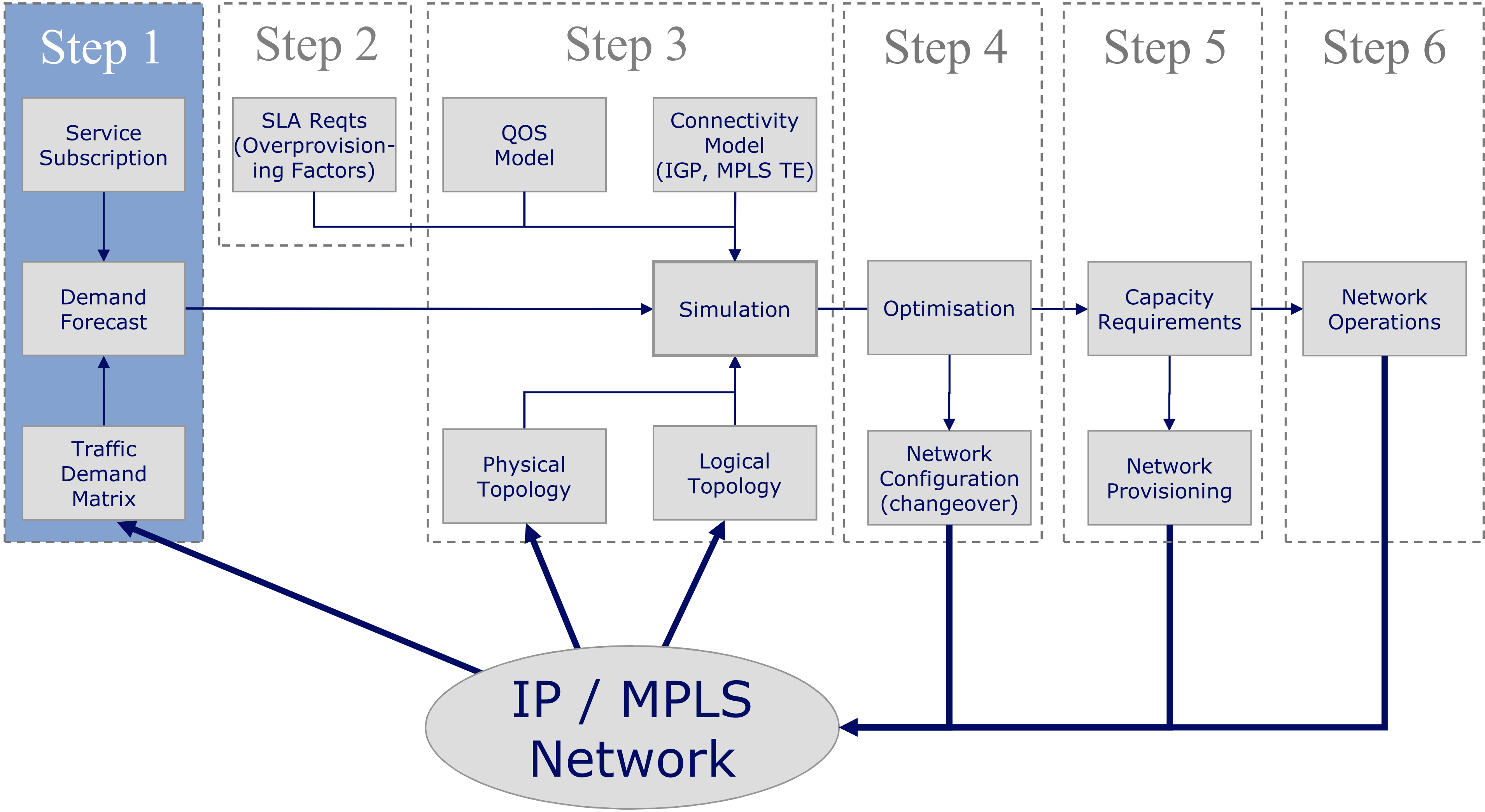
- Acceptance that simply monitoring per link statistics does not provide the fidelity required for effective and efficient IP / MPLS service delivery
- Shift from expert, guru-led planning to a more systematic approach
- Blurring of the old boundaries between planning, engineering and operations

Why does this matter?

- 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 is to ensure there is sufficient capacity to deliver the SLAs required for the transported services [without gross overprovisioning]
- 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*

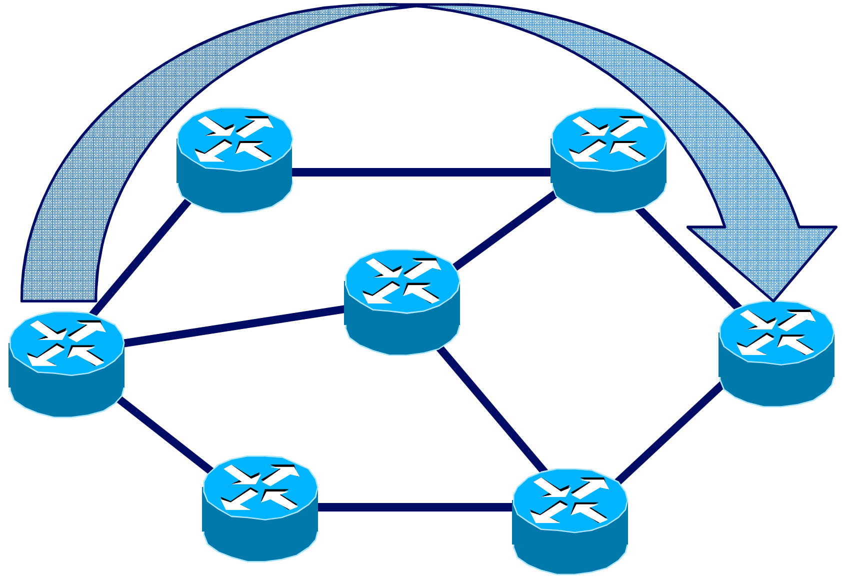
Network Planning Methodology

1. Traffic / demand matrices …



Traffic Demand Matrix

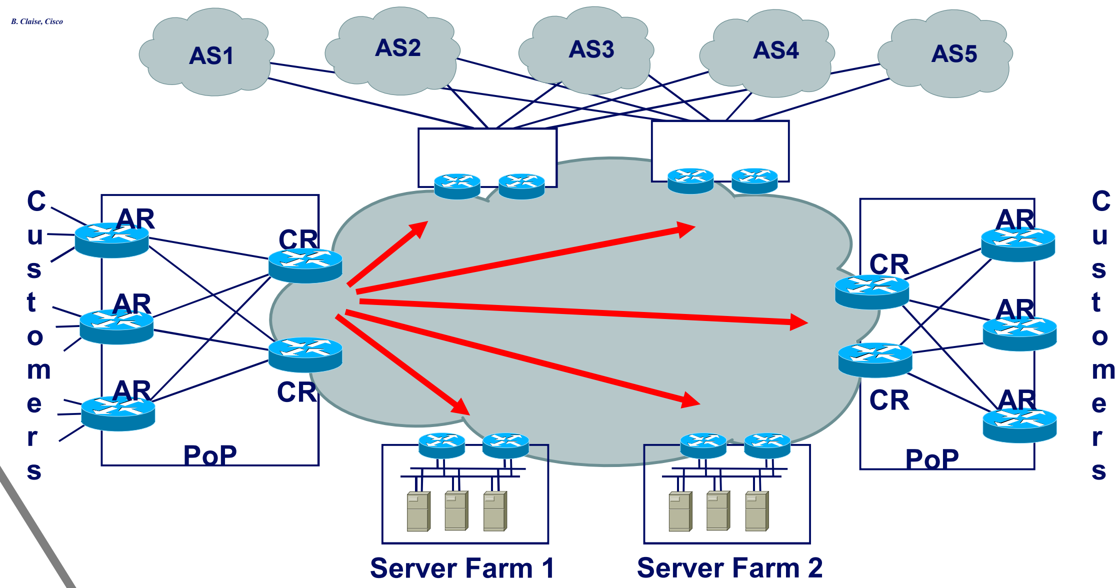
- Traffic demands define the amount of data transmitted between each pair of network nodes
  - Internal vs. external
  - per Class, per application, ...
  - Can represent peak traffic, traffic at a specific time, or percentile
  - Router-level or PoP-level demands
  - May be measured, estimated or deduced



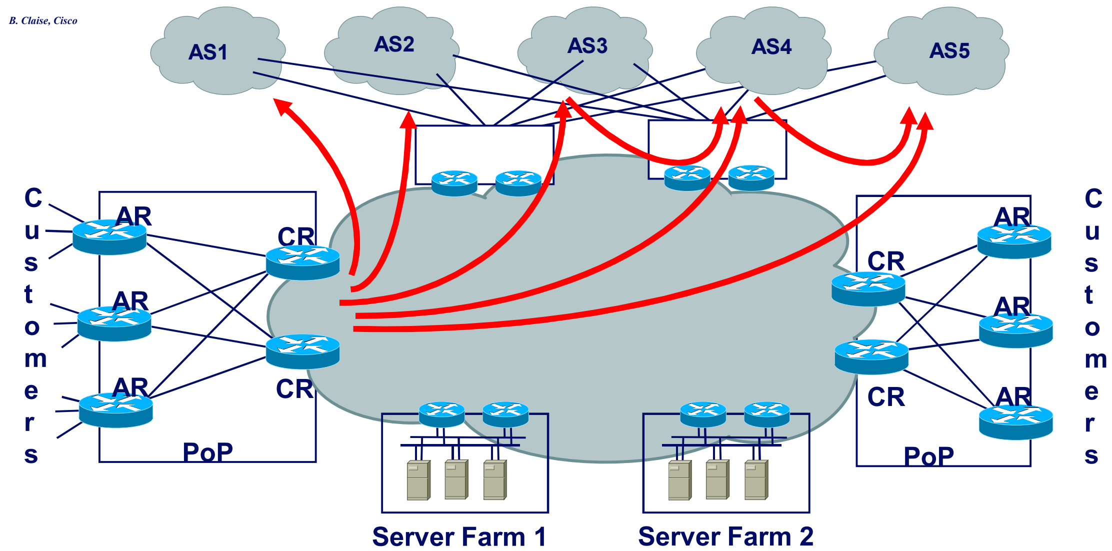
- The matrix of network traffic demands is crucial for analysis and evaluation of other network states than the current:
  - network changes
  - "what-if" scenarios
  - resilience analysis, network under failure conditions
  - optimisation: network engineering and traffic engineering
    - Comparing 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	**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)

Measuring the Traffic Matrix in Practise

Flows

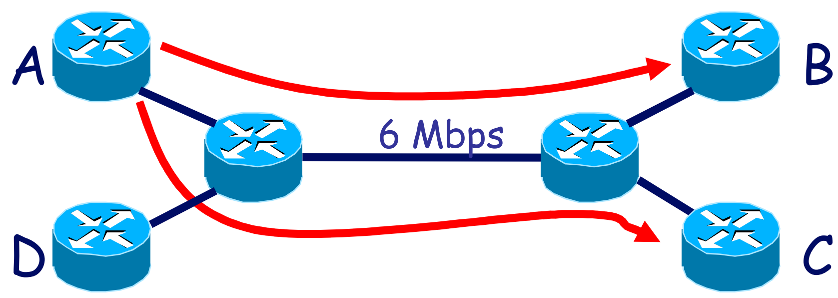
- NetFlow
  - v5
    - Resource intensive for collection and processing
    - Non-trivial to convert to Traffic Matrix
  - v9
    - BGP NextHop Aggregation scheme provides almost direct measurement of the Traffic Matrix
    - Only supported by newer versions of Cisco IOS
  - Inaccuracies
    - Stats 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 $O(N^2)$: missing values, time, ...
  - Inconsistencies in vendor implementations

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(N2) 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]



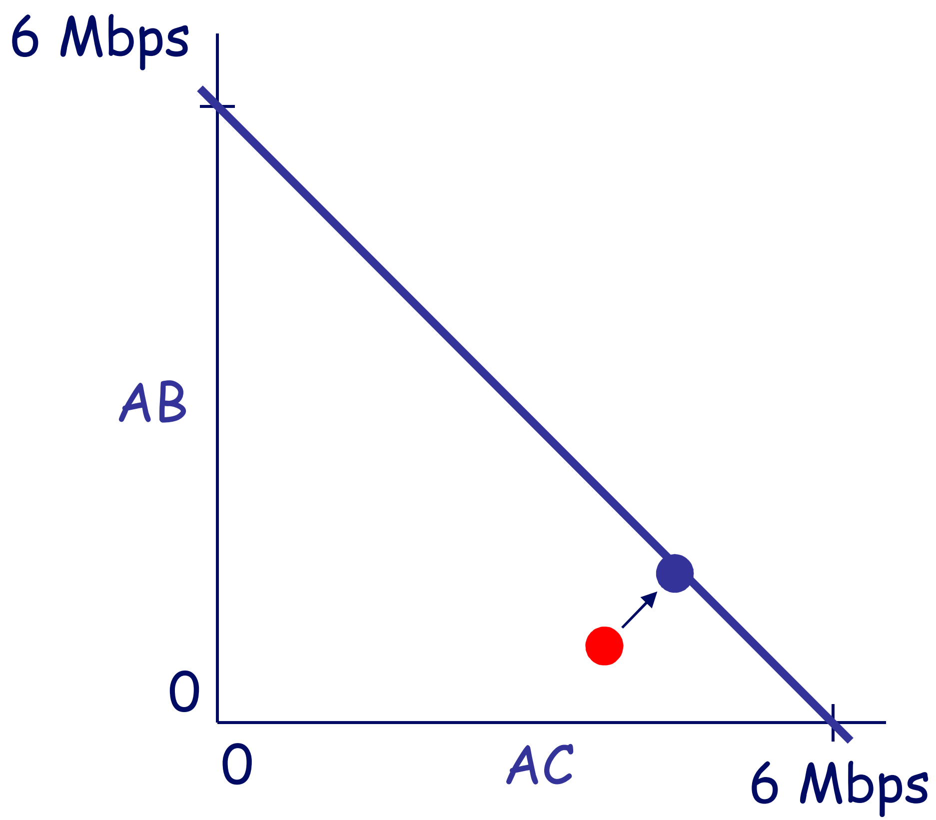
y: link utilizations
A: routing matrix
x: point-to-point demands

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

Calculate the **most likely** Traffic Matrix

Demand Estimation: Example

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



Additional information

E.g. Gravity Model (every source sends the same percentage as all other sources of it's total traffic to a certain destination)

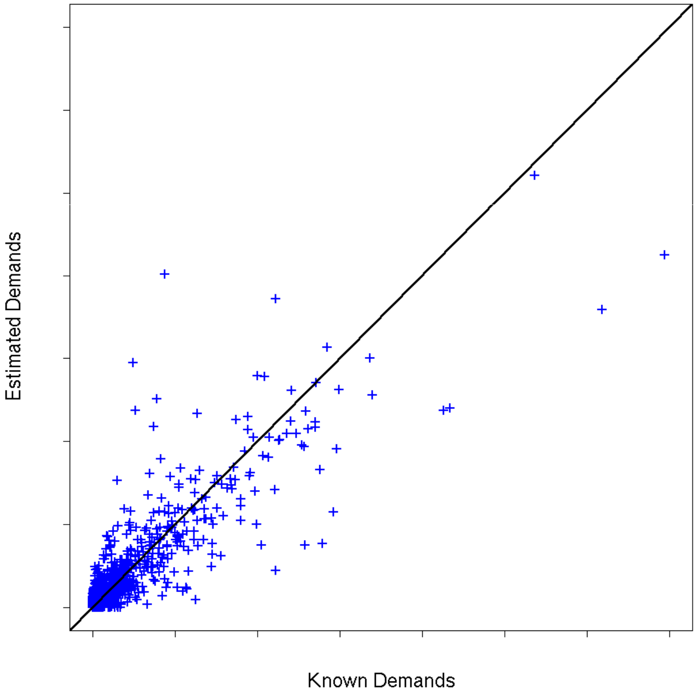
Example: Total traffic sourced at Site A is *50Mbps*.
Site B sinks *2%* of total network traffic, C sinks *8%*.

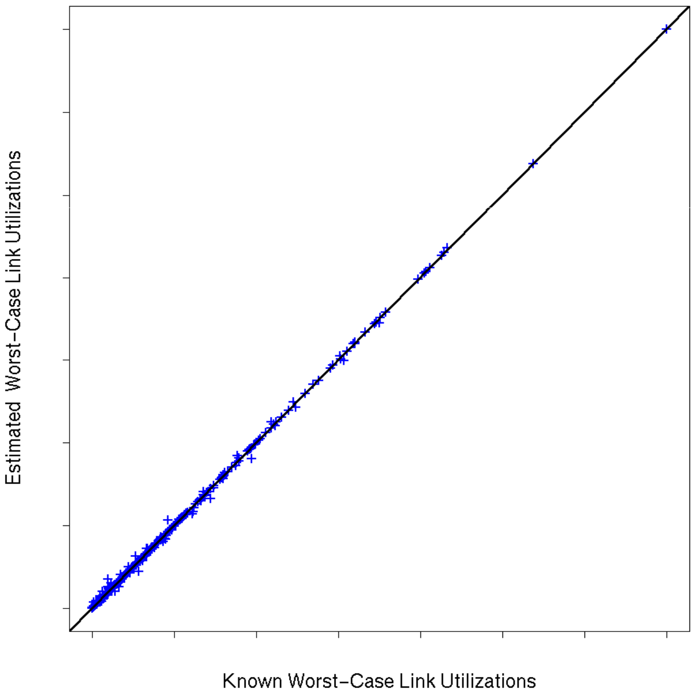
AB = 1 Mbps and AC = 4 Mbps

Final Estimate: AB = 1.5 Mbps and AC = 4.5 Mbps

Demand Estimation Results

- [Gunner et al]
  Results from International
  Tier-1 IP Backbone





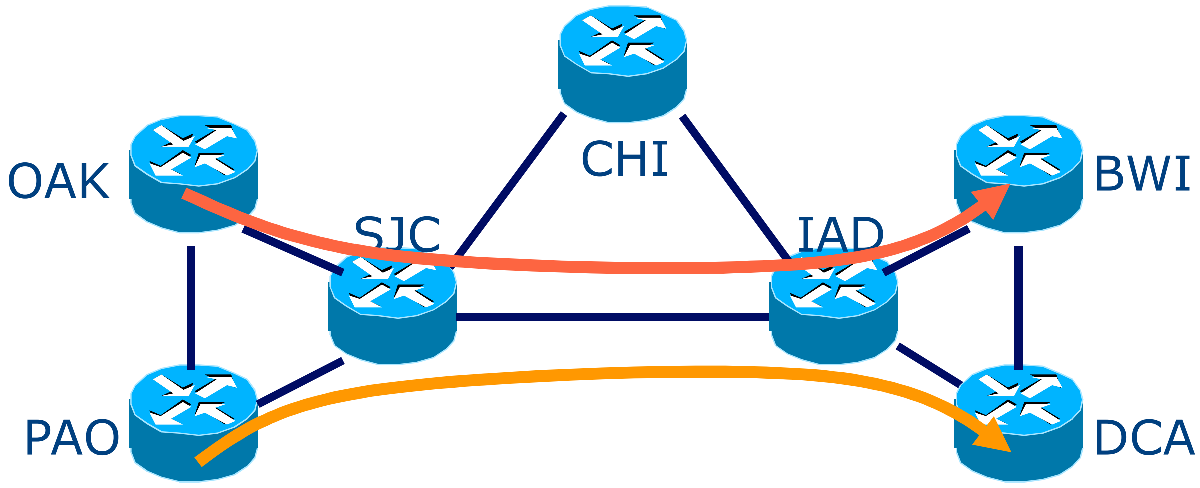
- Individual demand estimates can be inaccurate

- 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

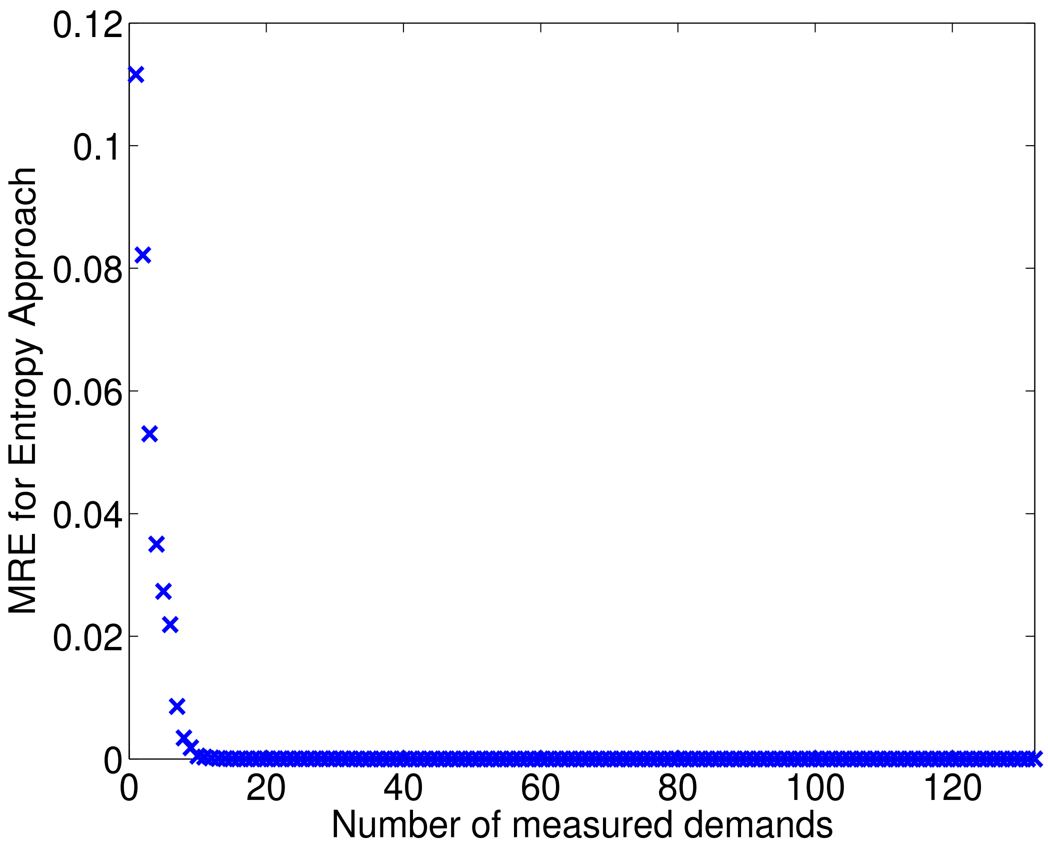


- Hard to tell apart elements
  - OAK->BWI, OAK->DCA, PAO->BWI, PAO->DCA, similar routings
- Are likely to shift as a group under failure or IP TE
  - e.g., above all shift together to route via CHI under SJC-IAD failure

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 [Gunner et al]



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

Regressed Measurements Example

- 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)

Overall Summary

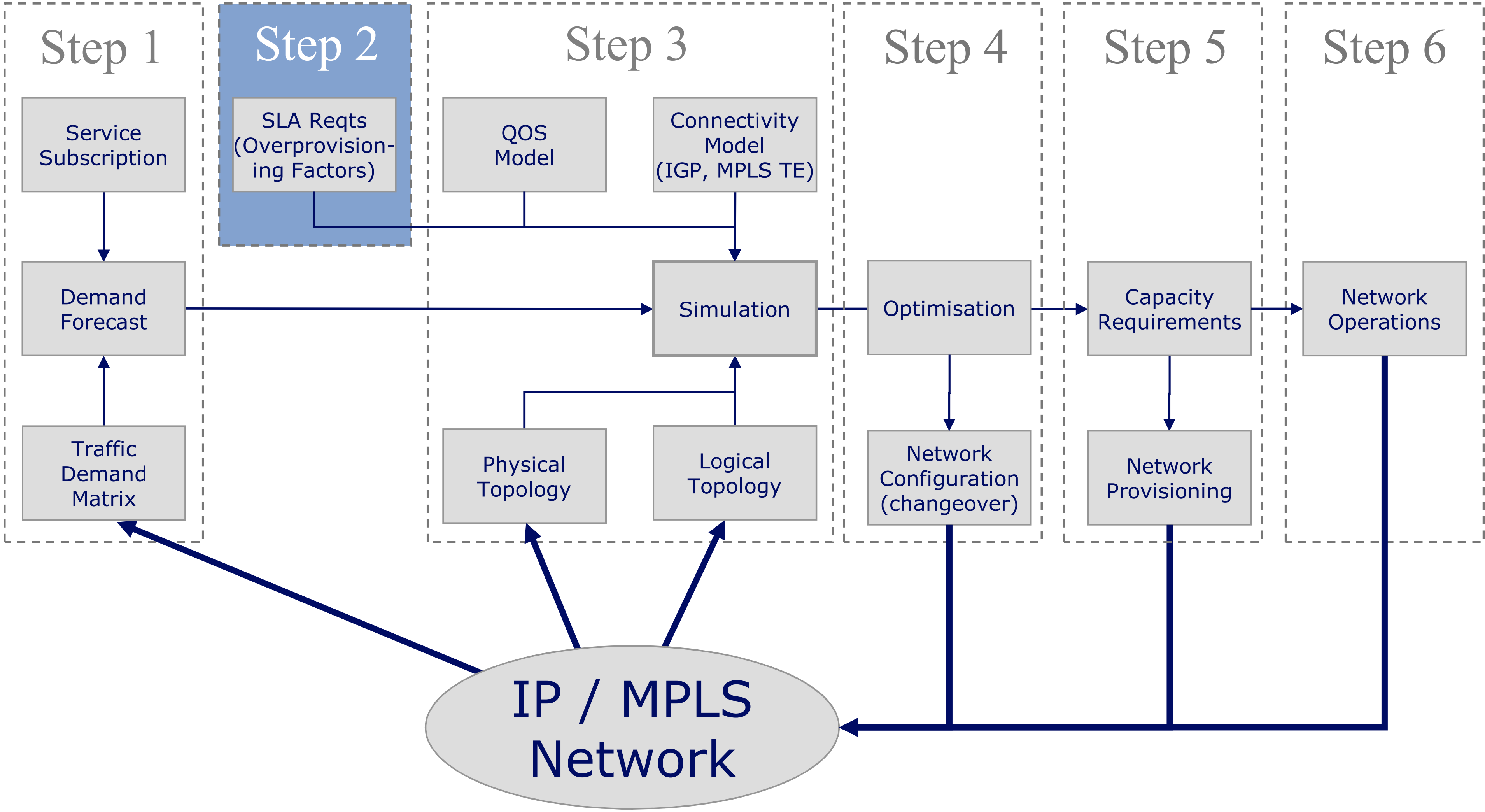
- Direct Measurement works well sometimes
  - Netflow OK on some equipment
  - LSP counters OK on some equipment and if only care for internal traffic matrix
  - Watch out for scaling, speed and measurement mismatch with link stats
- Estimation on link stats works sometimes
  - Has great speed (order 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

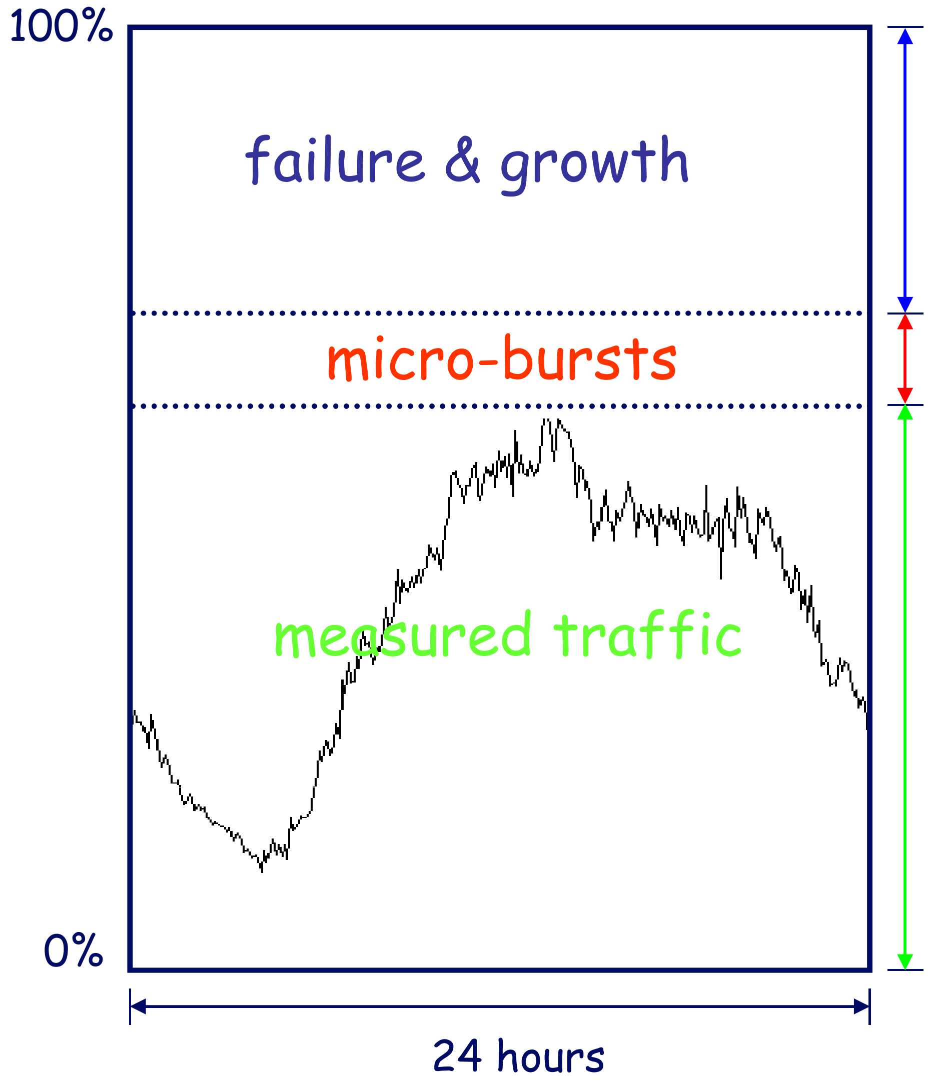
Network Planning Methodology

2. The relationship between SLAs and network planning targets …



IP / MPLS Traffic Characterisation

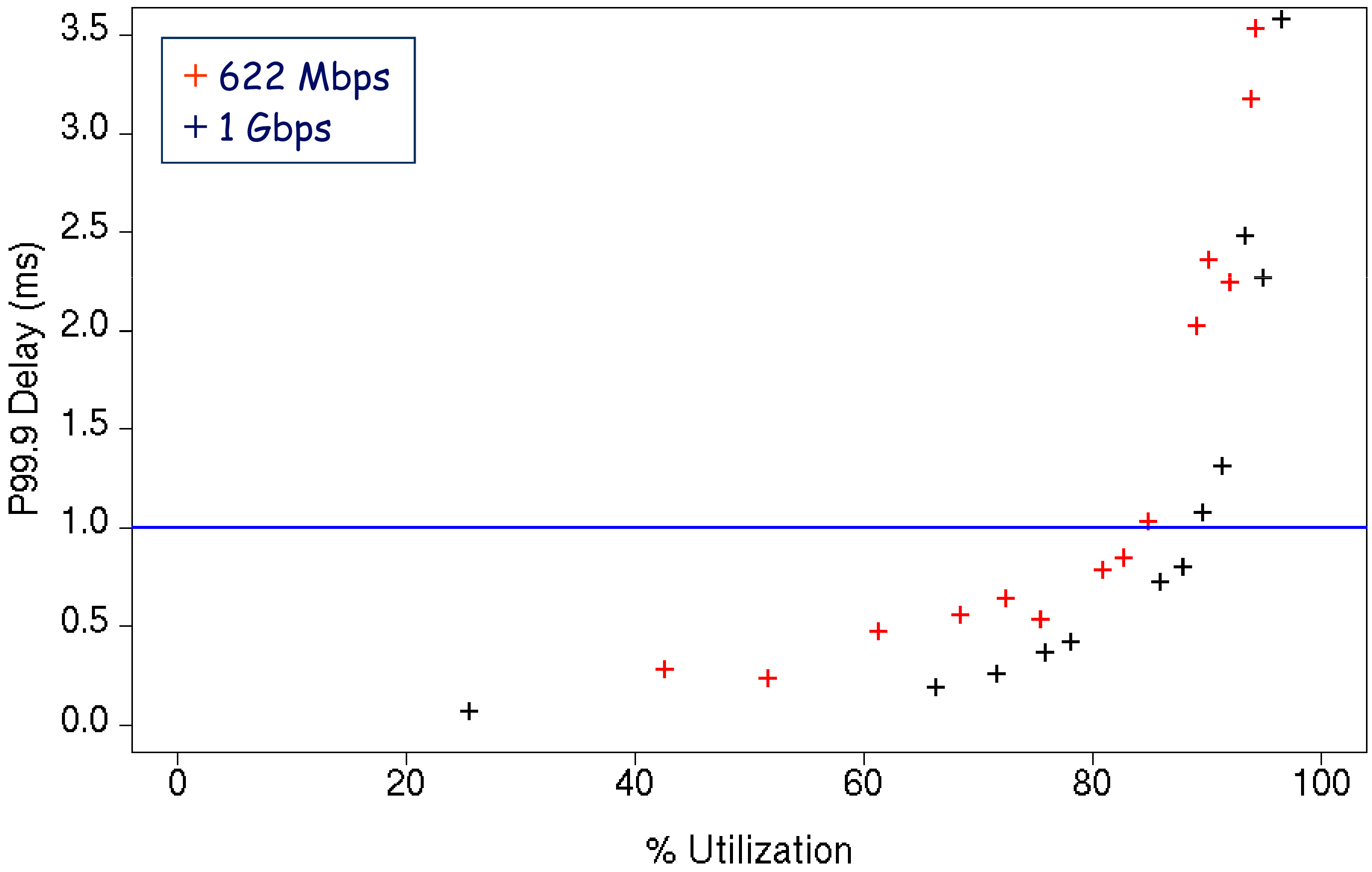
- Network traffic measurements are normally long term, i.e. in the order of minutes
  - Implicitly the measured rate is an average of the measurement interval
- In the short term, i.e. milliseconds, however, microbursts cause queueing, impacting the delay, jitter and loss
- *What's the relationship between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

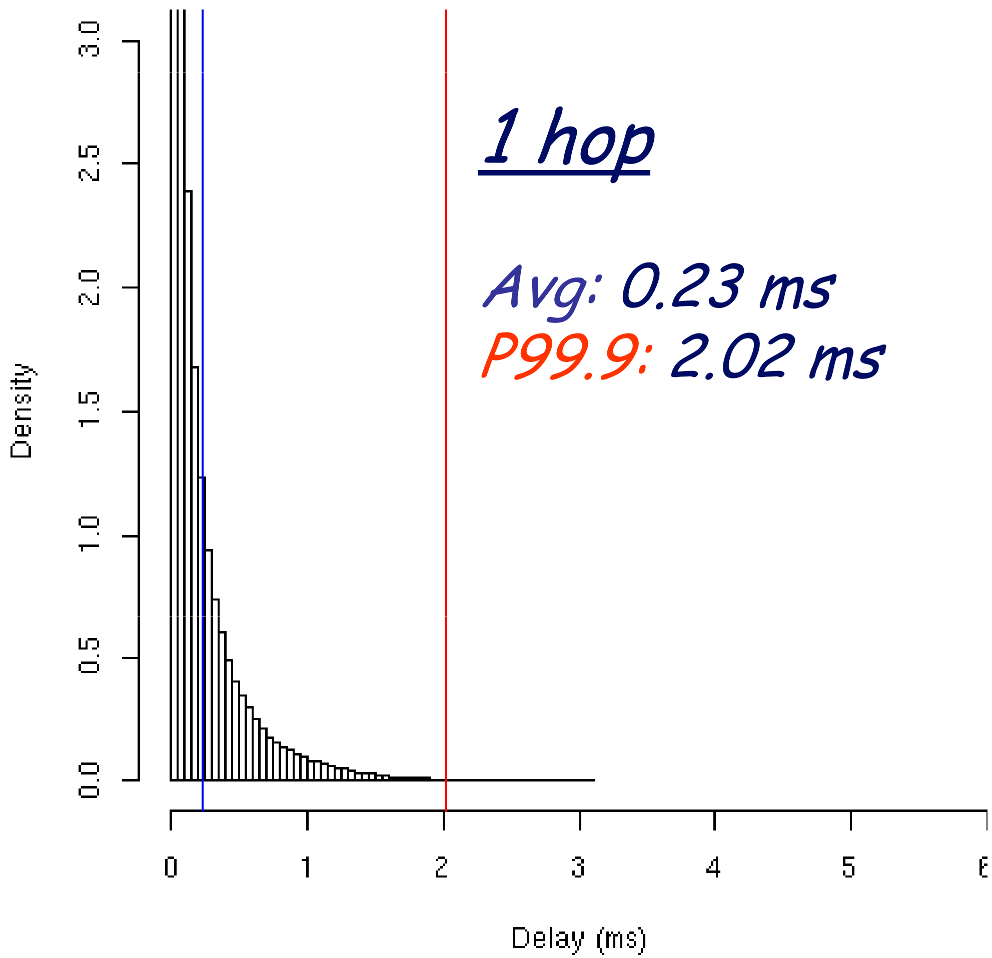
Queueing Simulation Results
[Telkamp 2003]

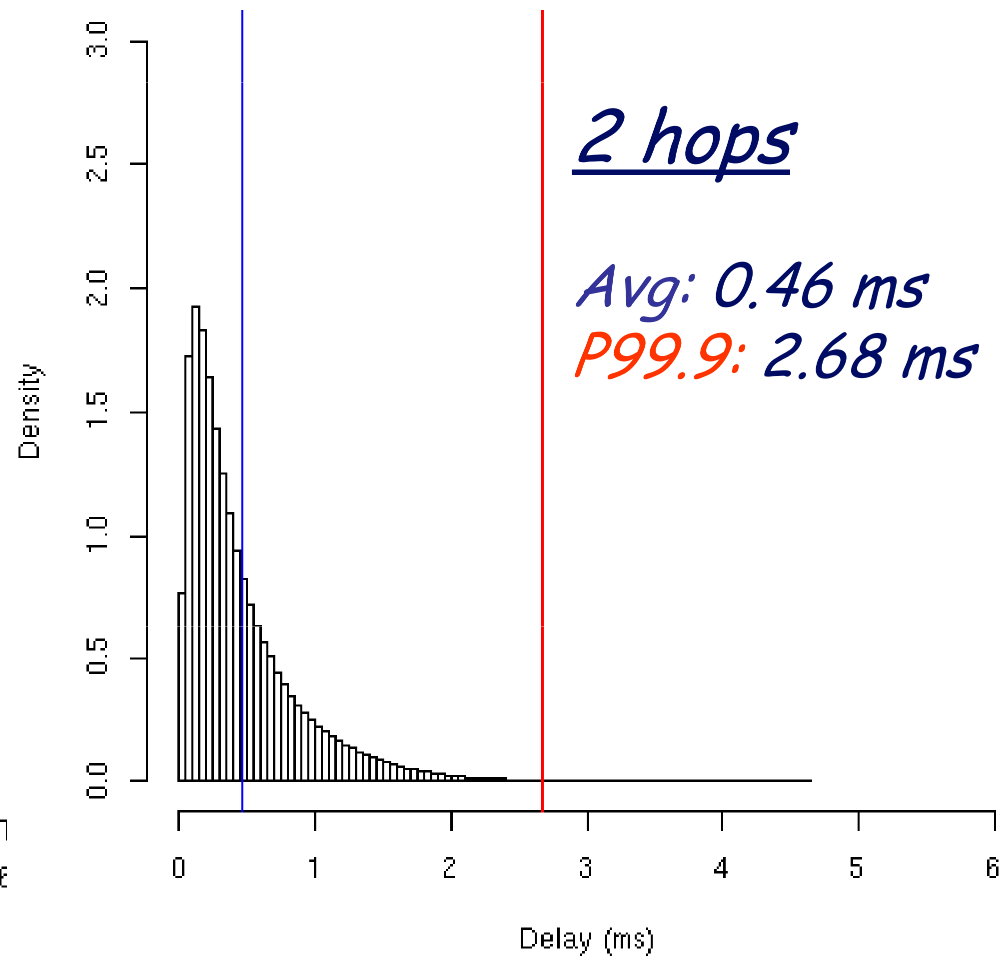


- 622Mbps, 1Gbps links – overprovisioning percentage ~10% is required to bound delay/jitter to 1-2ms
- Lower speeds (≤155Mpbs) – overprovisioning factor is significant
- Higher speeds (2.5G/10G) – overprovisioning factor becomes very small

Multi-hop Queuing [Telkamp 2003]

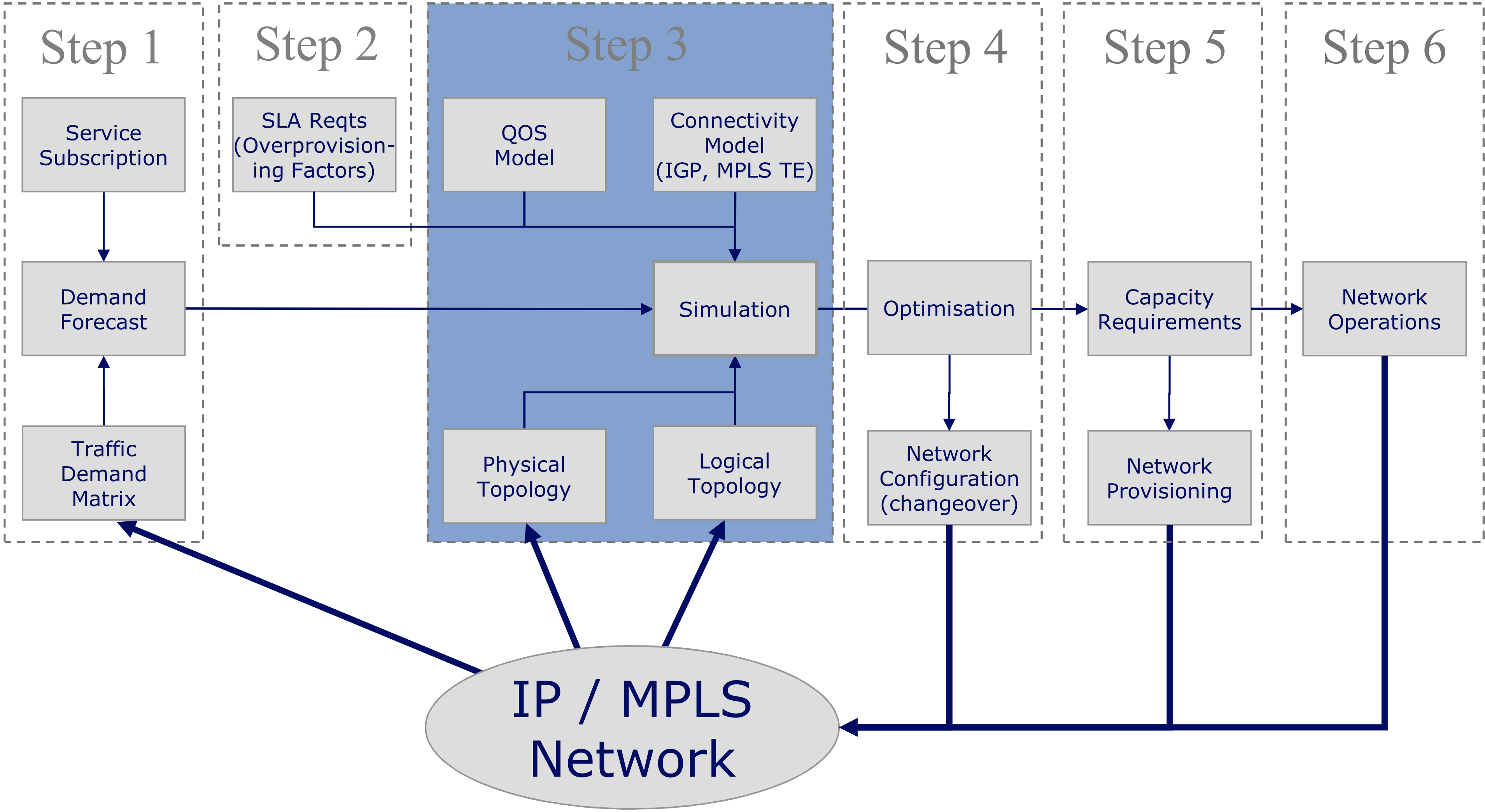
P99.9 multi-hop delay/jitter is not additive



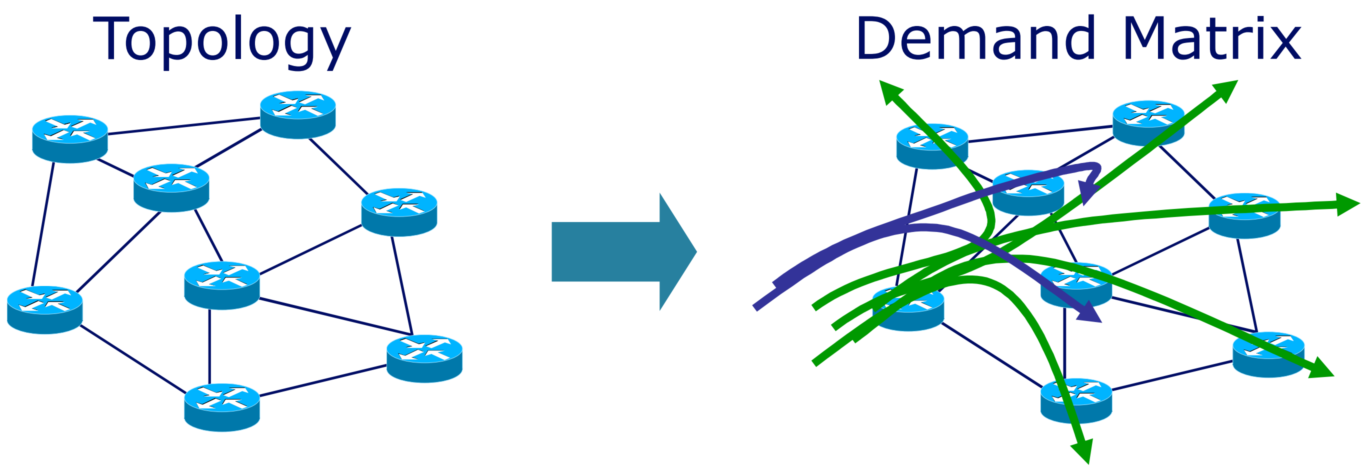


Network Planning Methodology

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



Simulation

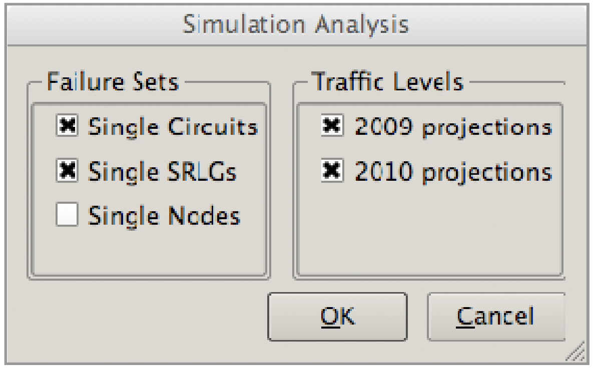


- Map core traffic matrix to topology (logical and physical)
- Simulate for link, node and shared risk (SRLG) failures
  - Can add a traffic growth factor if required
- On a per class basis if Diffserv deployed
- Enables:
  - Forecasting of which links need upgrading when
  - Understand of if topology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

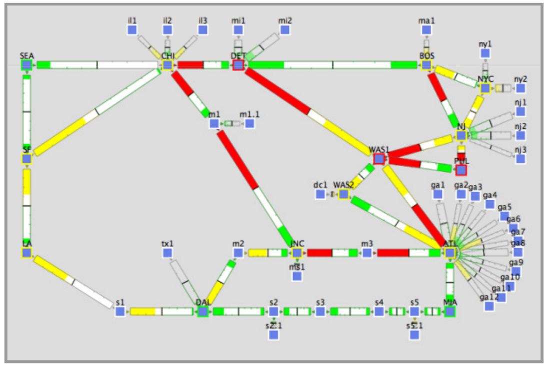
Failure Planning

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

Simulate using external traffic projections

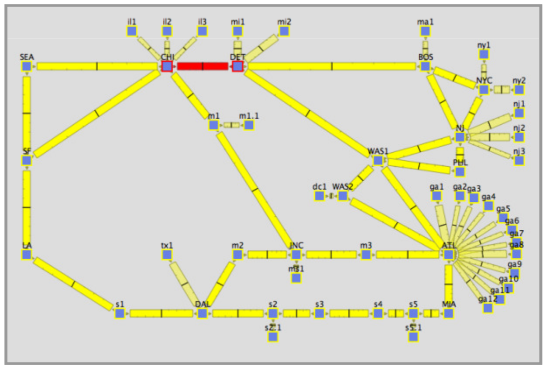


Worst case view



Potential congestion under failure in **RED**

Failure impact view



Failure that can cause congestion in **RED**

Perform topology what-if analysis

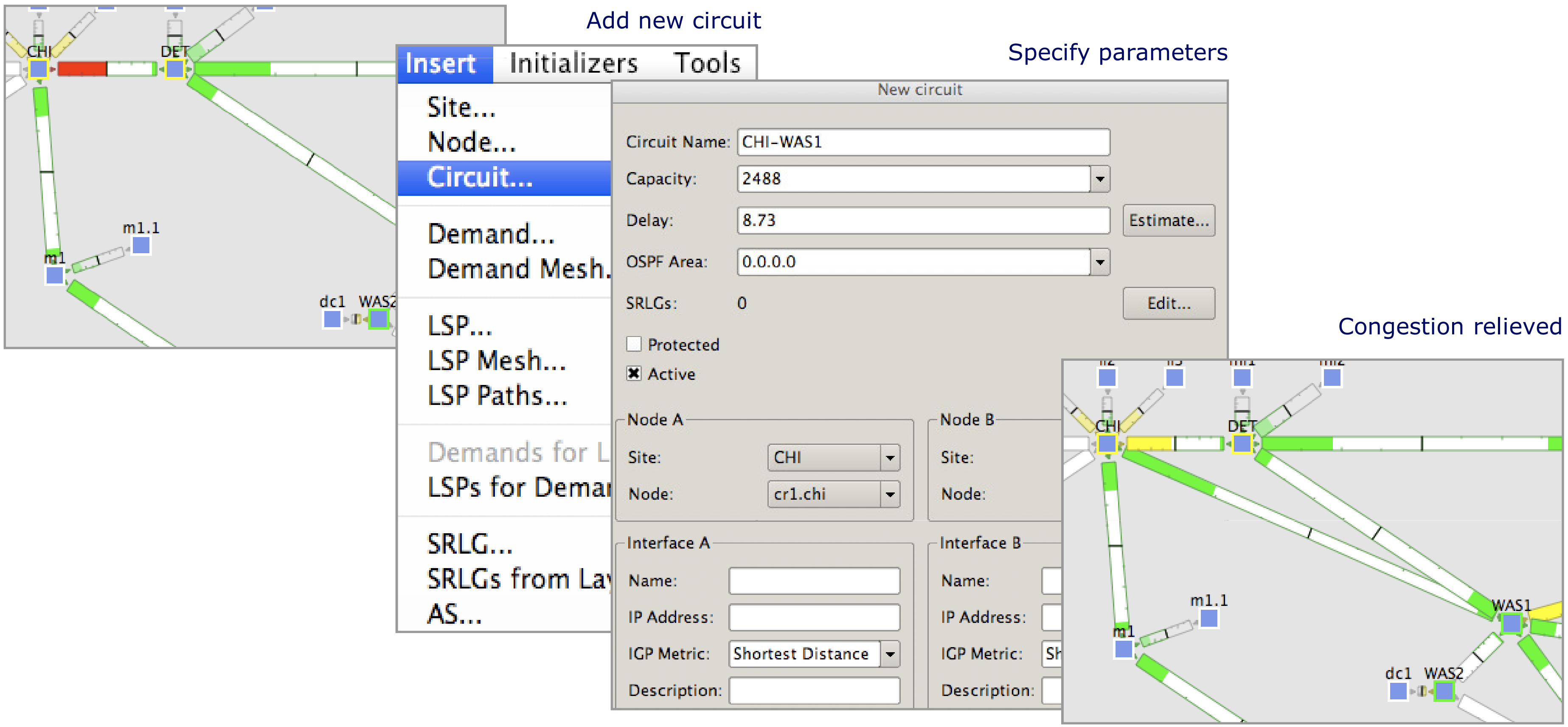
Topology What-If Analysis

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

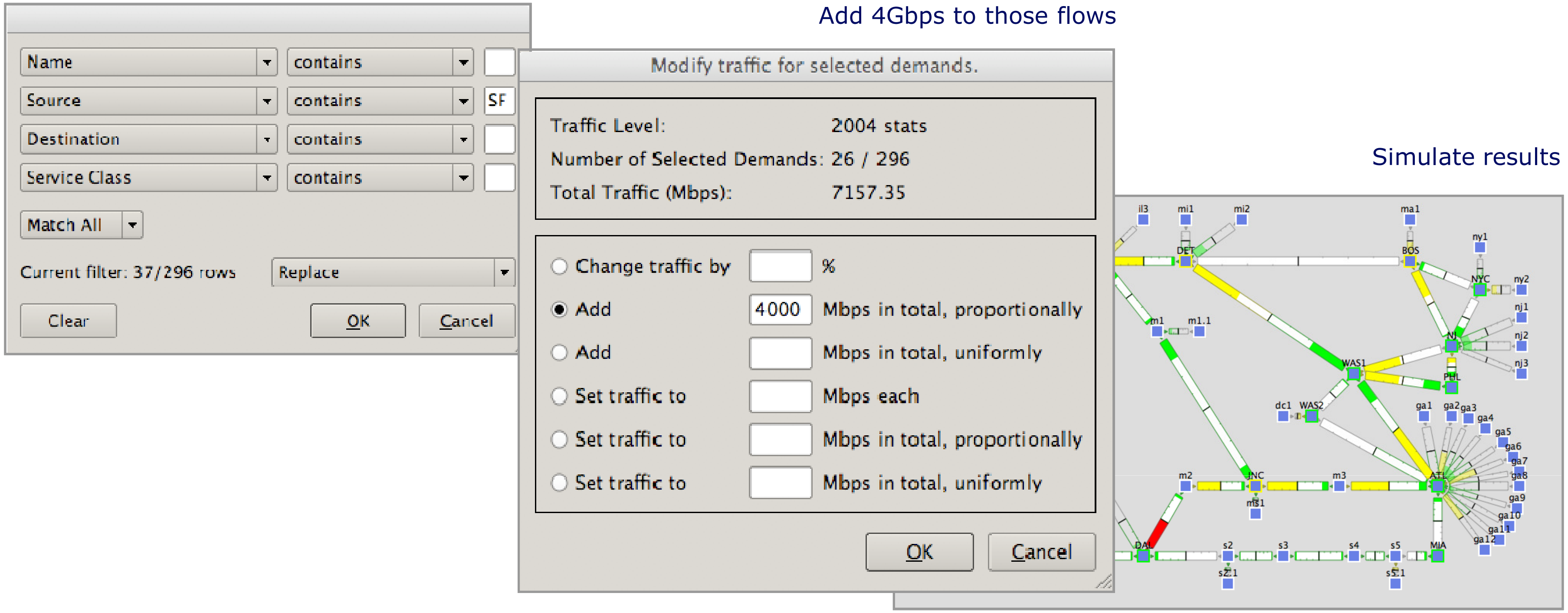
Evaluate New Customer

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

Identify flows for new customer

Add 4Gbps to those flows

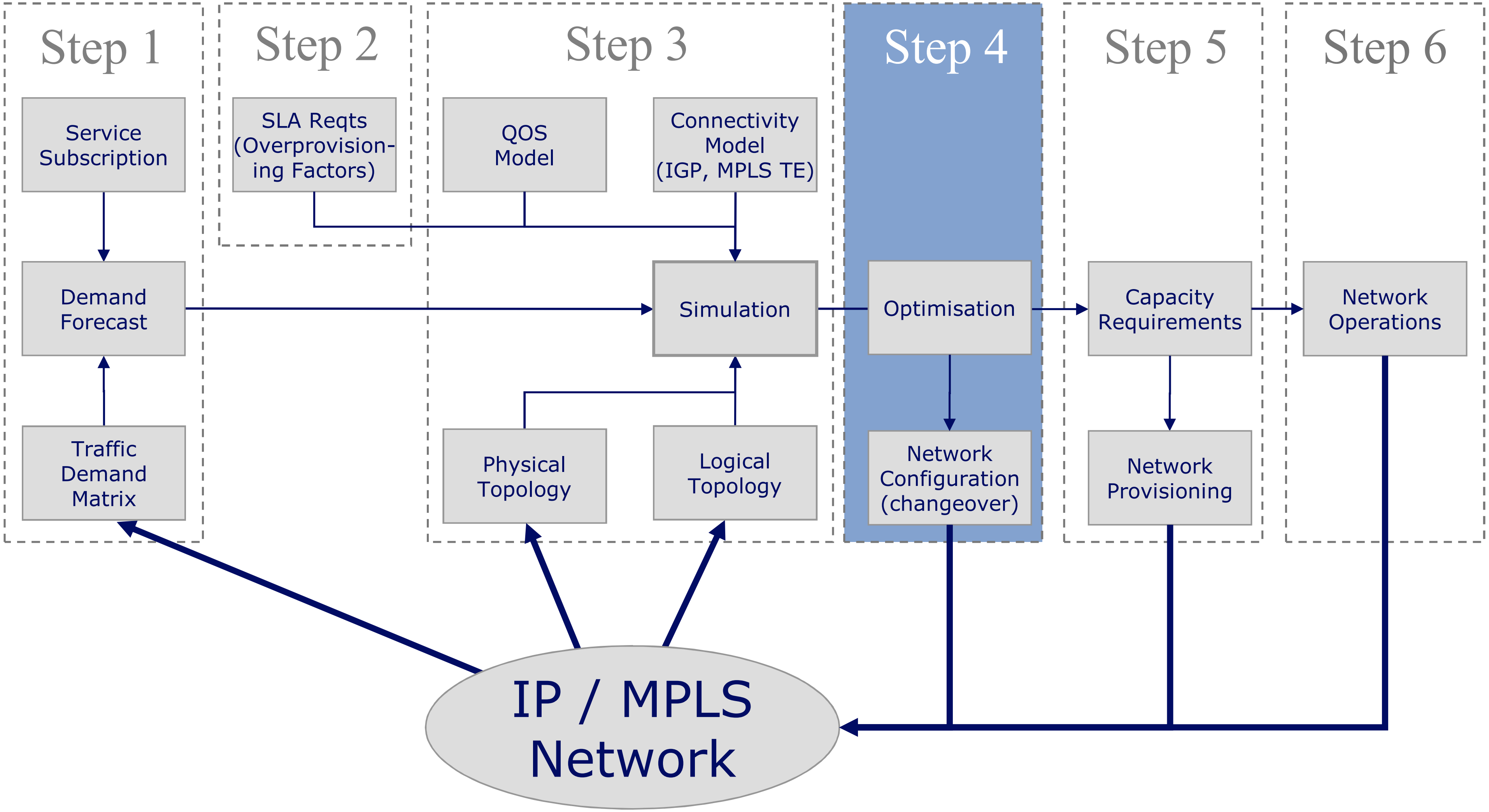
Simulate results



Congested link in **RED**

Network Planning Methodology

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



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

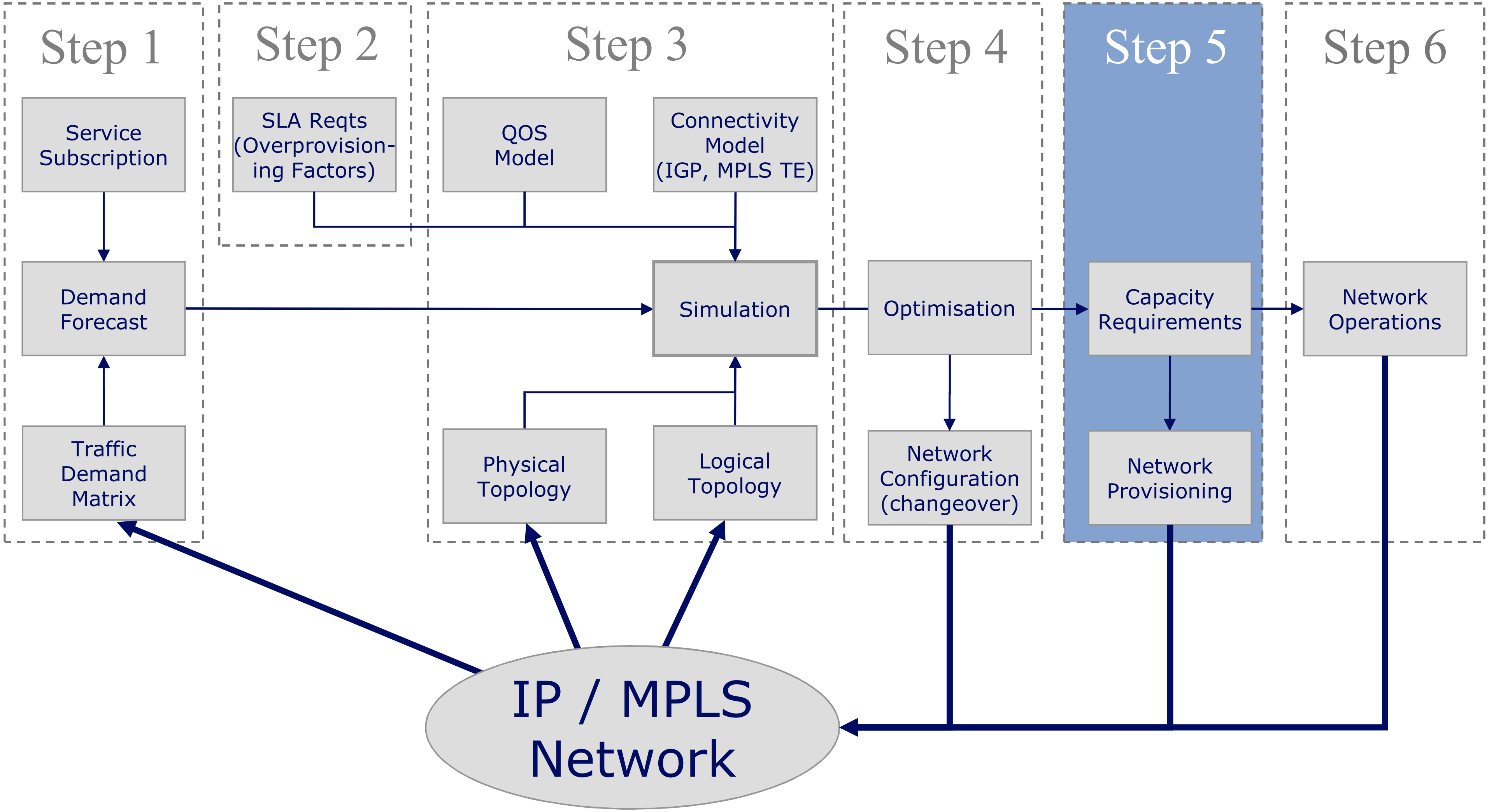
- Whilst network optimisation is an optional step, all of the preceding steps are essential for:
  - Comparing network engineering and TE approaches
  - MPLS TE tunnel placement and IP TE

Network Optimisation: Questions

- What optimisation objective?
- Which approach?
  - IGP TE or MPLS TE
- Strategic or tactical?
- How often to re-optimise?
- If strategic MPLS TE chosen:
  - Core or edge mesh
  - Statically (explicit) or dynamically established tunnels
  - Tunnel sizing
  - Online or offline optimisation
  - Traffic sloshing

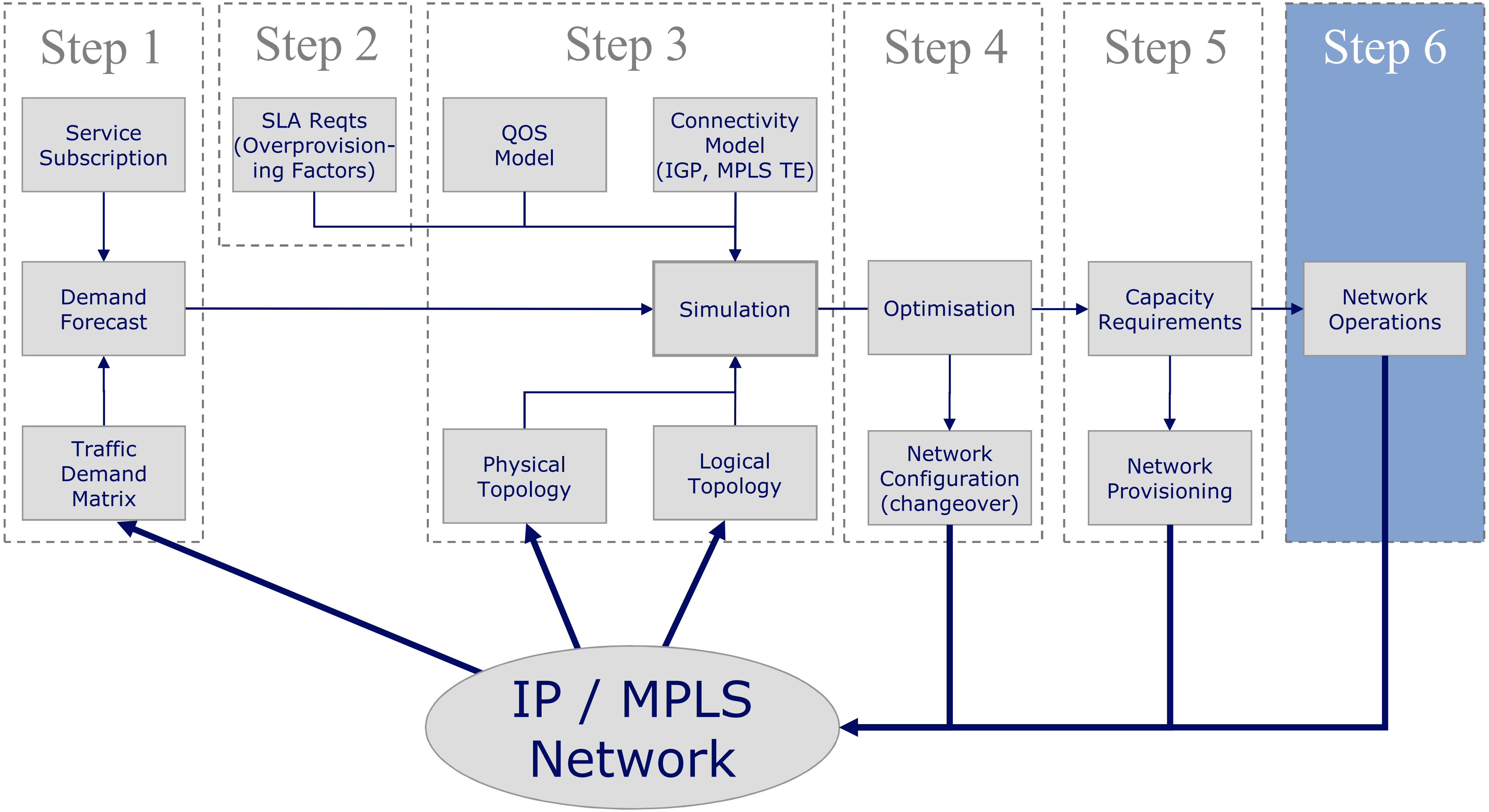
- Answers left for a future session ...

5. Network capacity provisioning



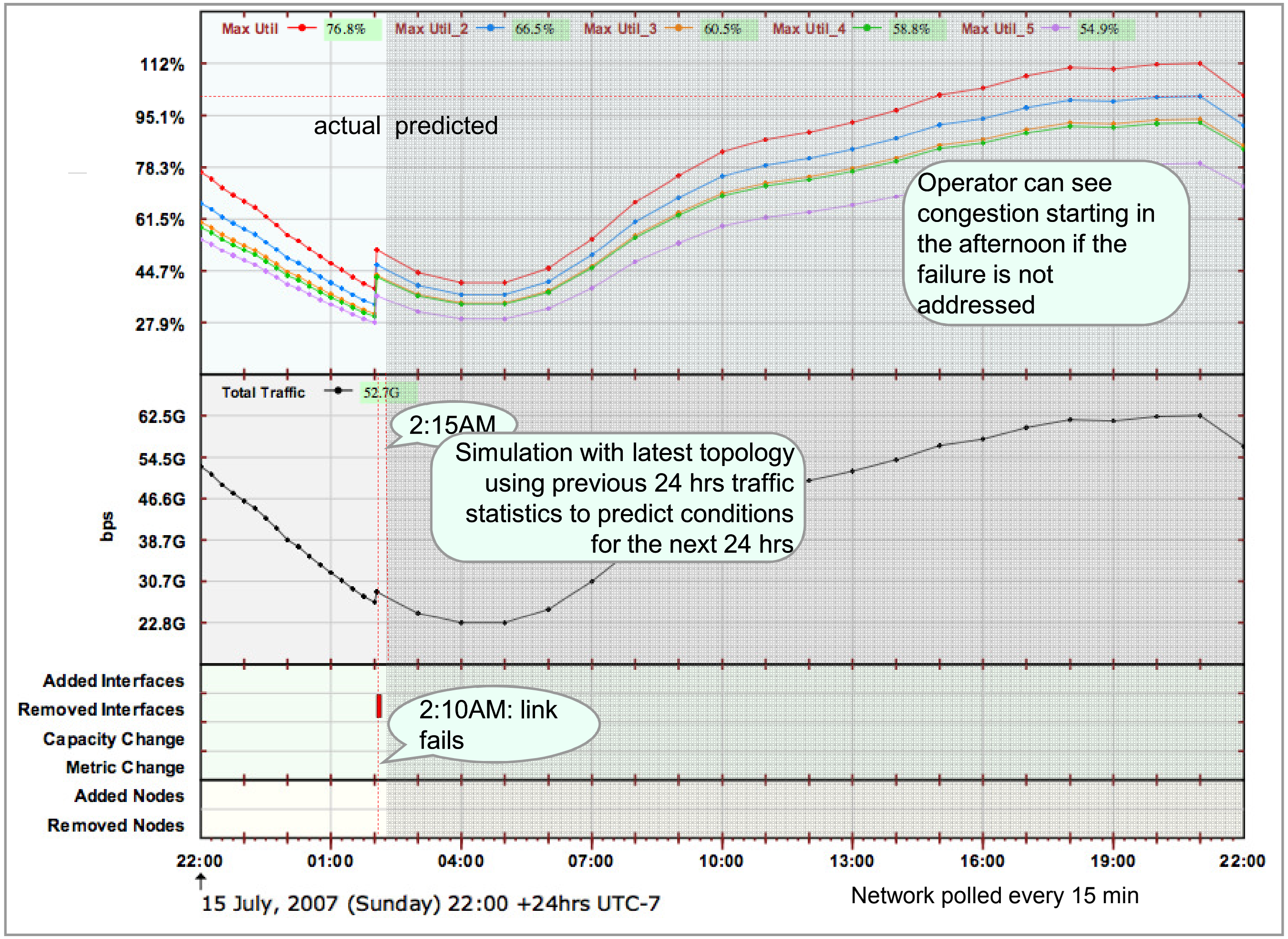
Network Planning Methodology

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/
- Gunner et al
  - Anders Gunnar, Mikael Johansson, Thomas Telkamp, "Traffic Matrix Estimation on a Large IP Backbone – A Comparison on Real Data", Internet Measurement Conference, October 2004, Sicily
  - http://www.cariden.com/technologies/papers.html#tm-imc
- 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/

- 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