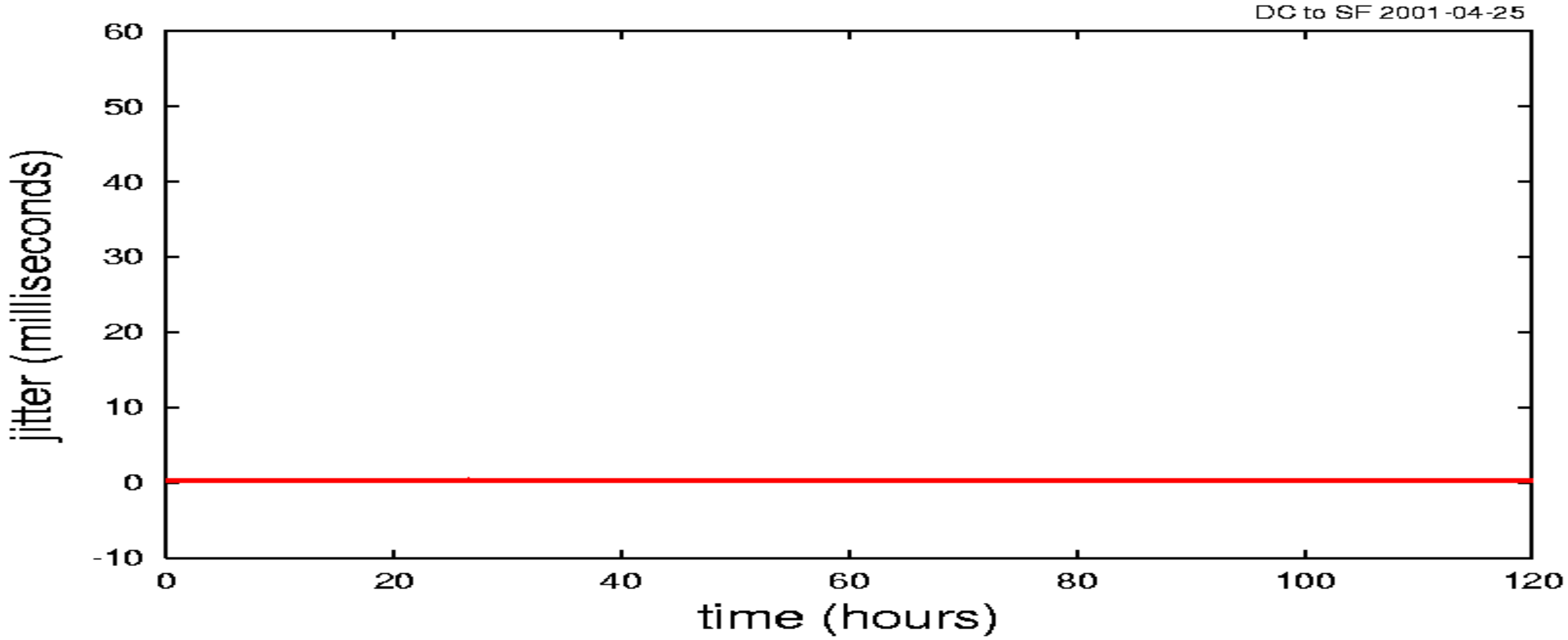
Diagnosing Performance Degradation with Route Analytics

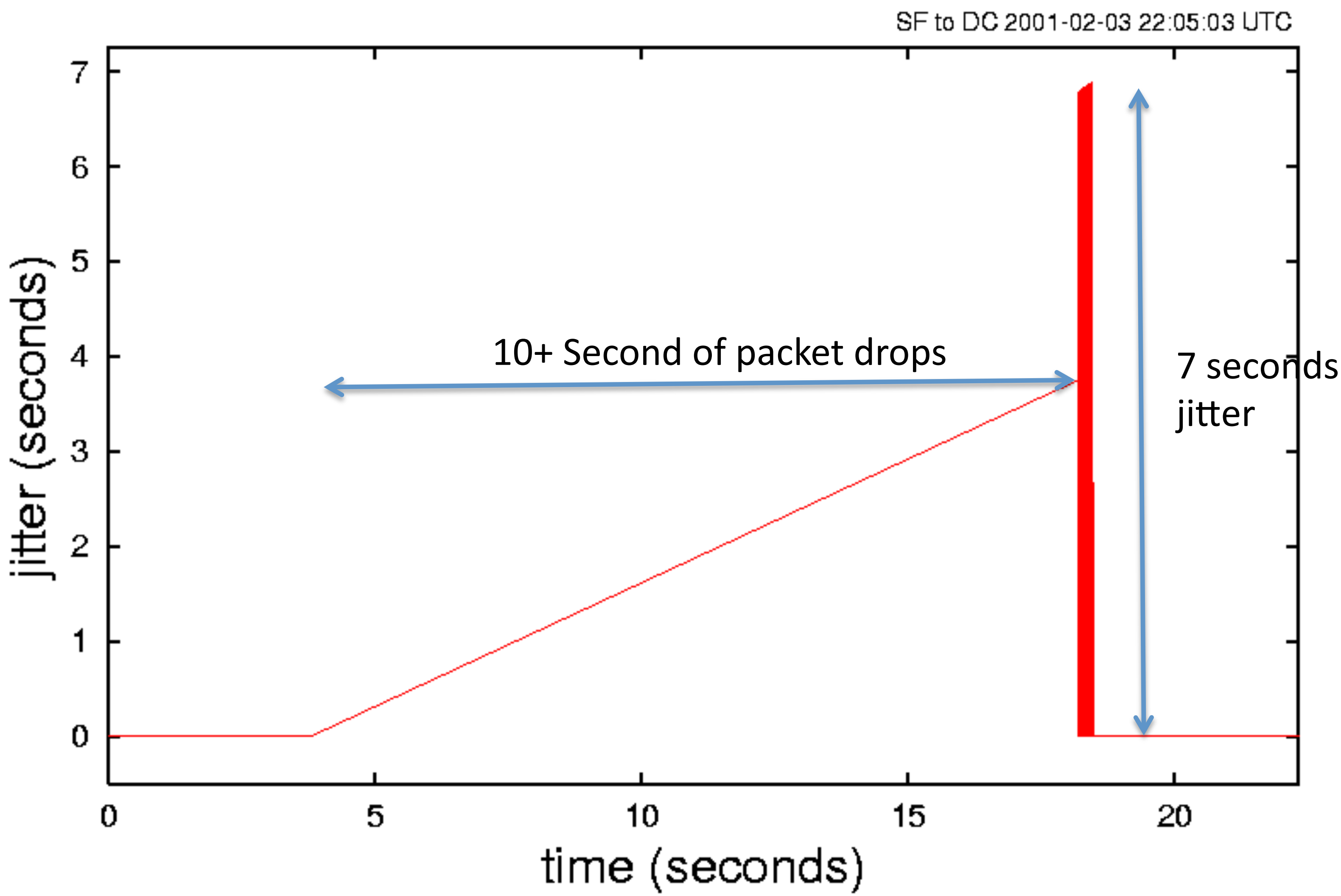
Cengiz Alaettinoglu

It all started with a Jitter Study (2000)

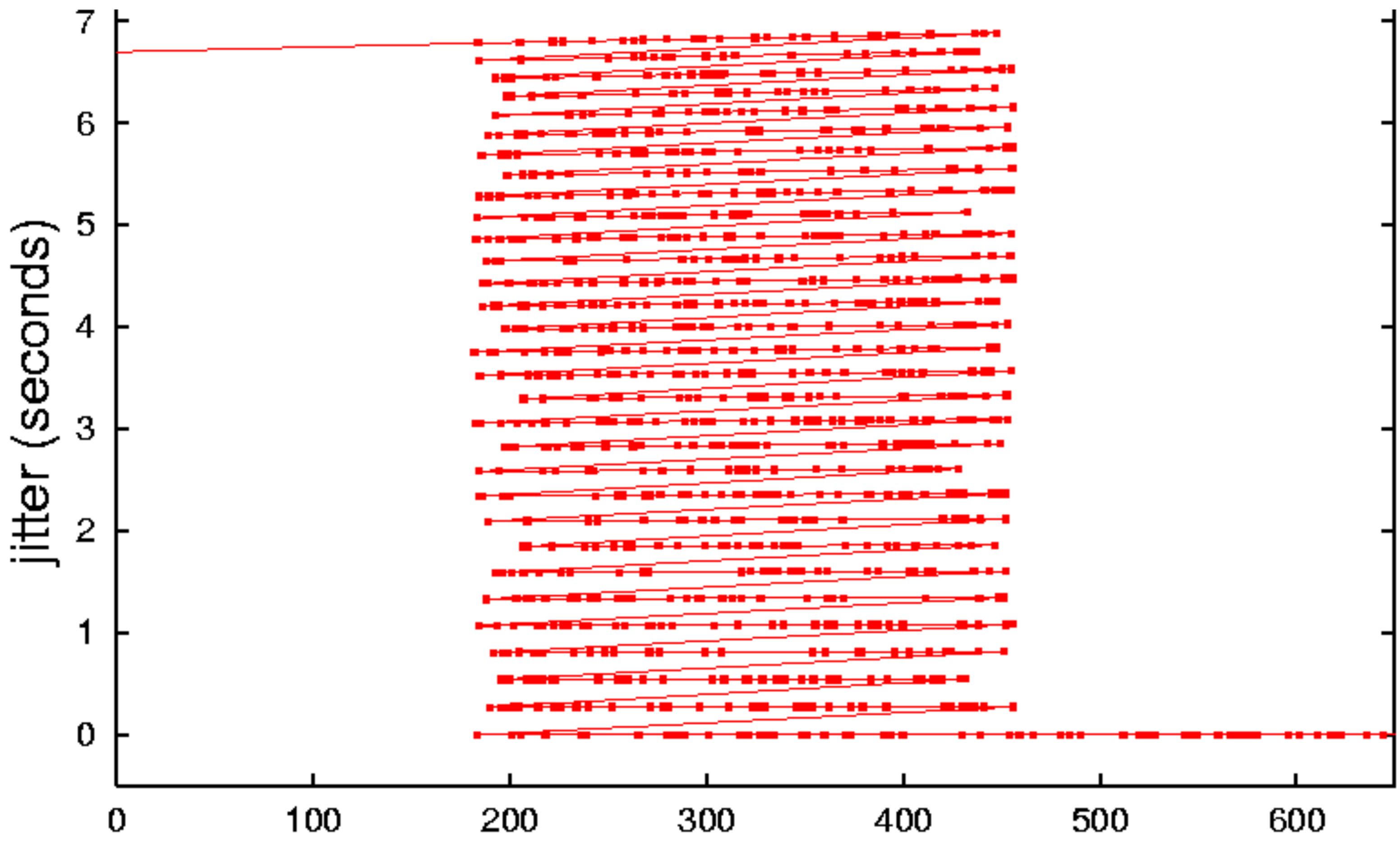
- Studied jitter on 2 US, and 1 European backbones for several weeks
- For 99.99% packets, measured jitter < 1ms



However, 0.01% of Jitter was Severe

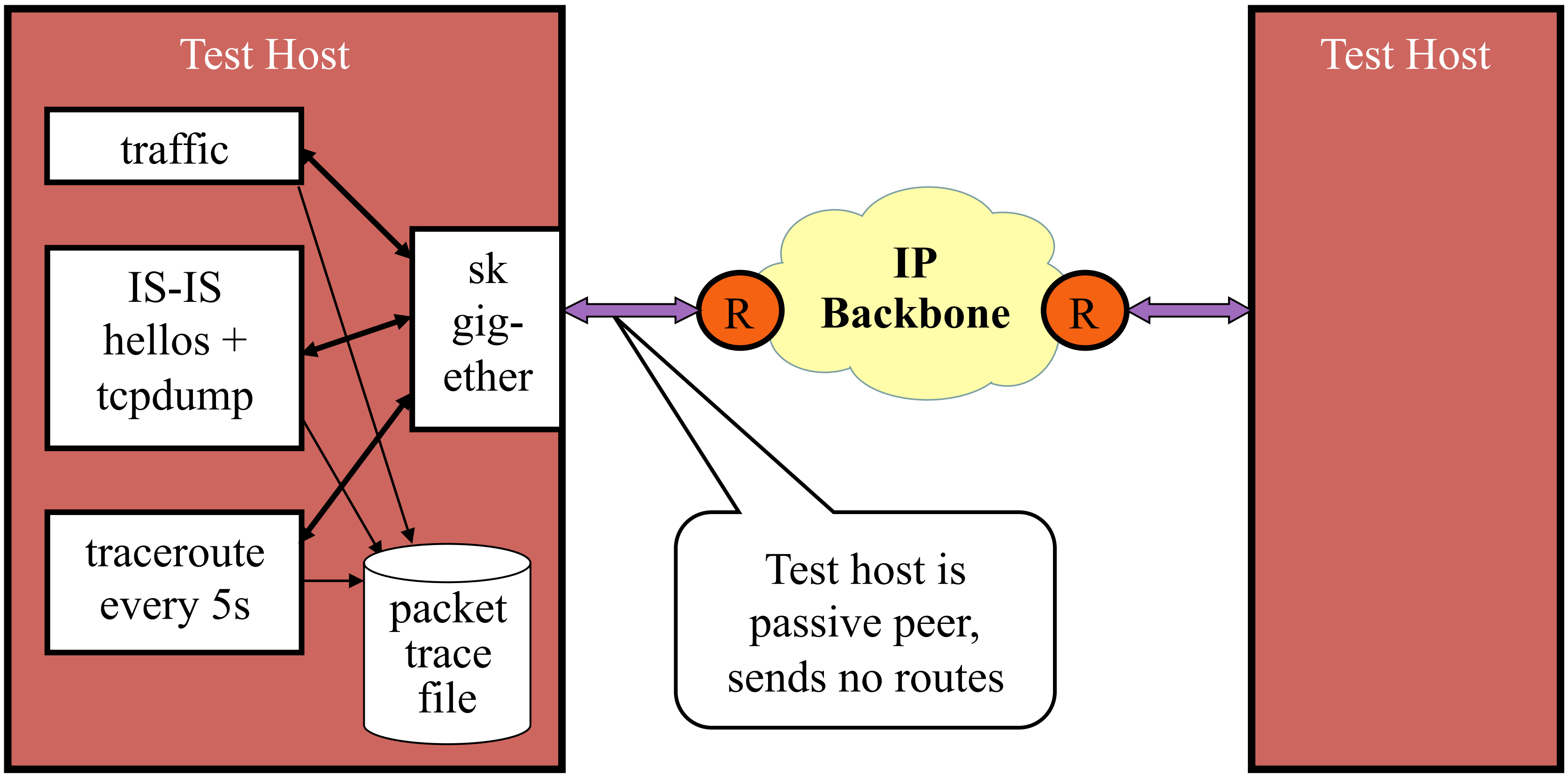


Severe Packet Reordering

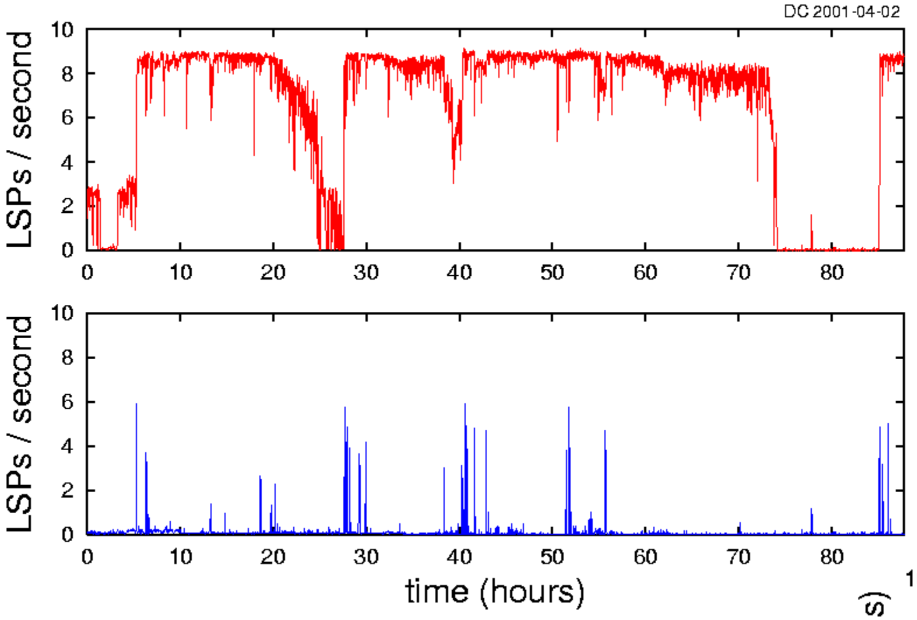


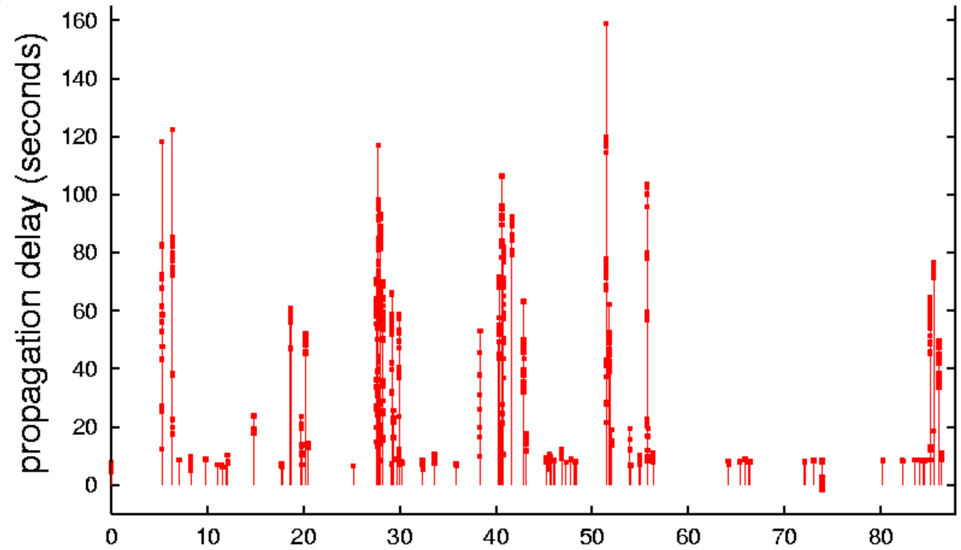
Theory: Packets being spewed out from an unwinding routing loop…

- Did we really have long routing loops in the network?
- Did ISIS really take 10+ seconds to convergence?
- So, we ***analyzed routing*** along with jitter



Excessive ISIS Churn caused excessive LSP Propagation Delay

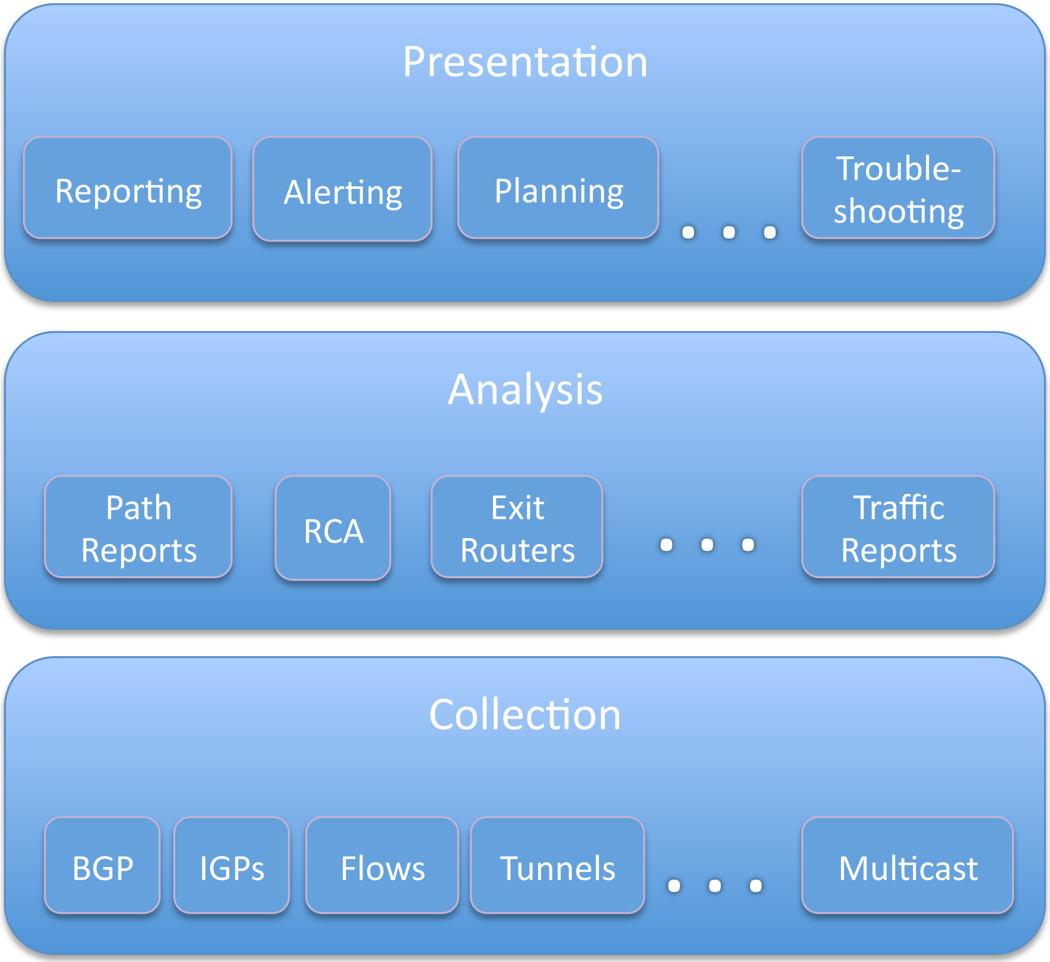




Explanation

- Link state databases were not in sync:
  - Very large LSP databases
  - High churn rate $\Rightarrow$ many LSPs to flood
  - LSP rate-control slowed down flooding
  - SPF updates may also have been delayed by rate limits
  - Any topology change could result in a loop under these conditions
- We realized being able to look at routing was key for powerful network performance analysis
- Today, we see very high churn in very large TE databases with auto-bandwidth with large number of tunnels

Route Analytics Today



Route Analytics Applications

- Troubleshooting and visualization
  - Ability to look at network state at any given time
  - Inspecting and playing events learned both from the routers and routing protocols
  - Comparing routing state and paths when a service/ application is performing well and when it is not
- Service/application monitoring and alerting
  - Monitor paths for changes in hops, metric, delay, and bandwidth
  - Monitor excessive protocol behavior

Route Analytics Applications (cont.)

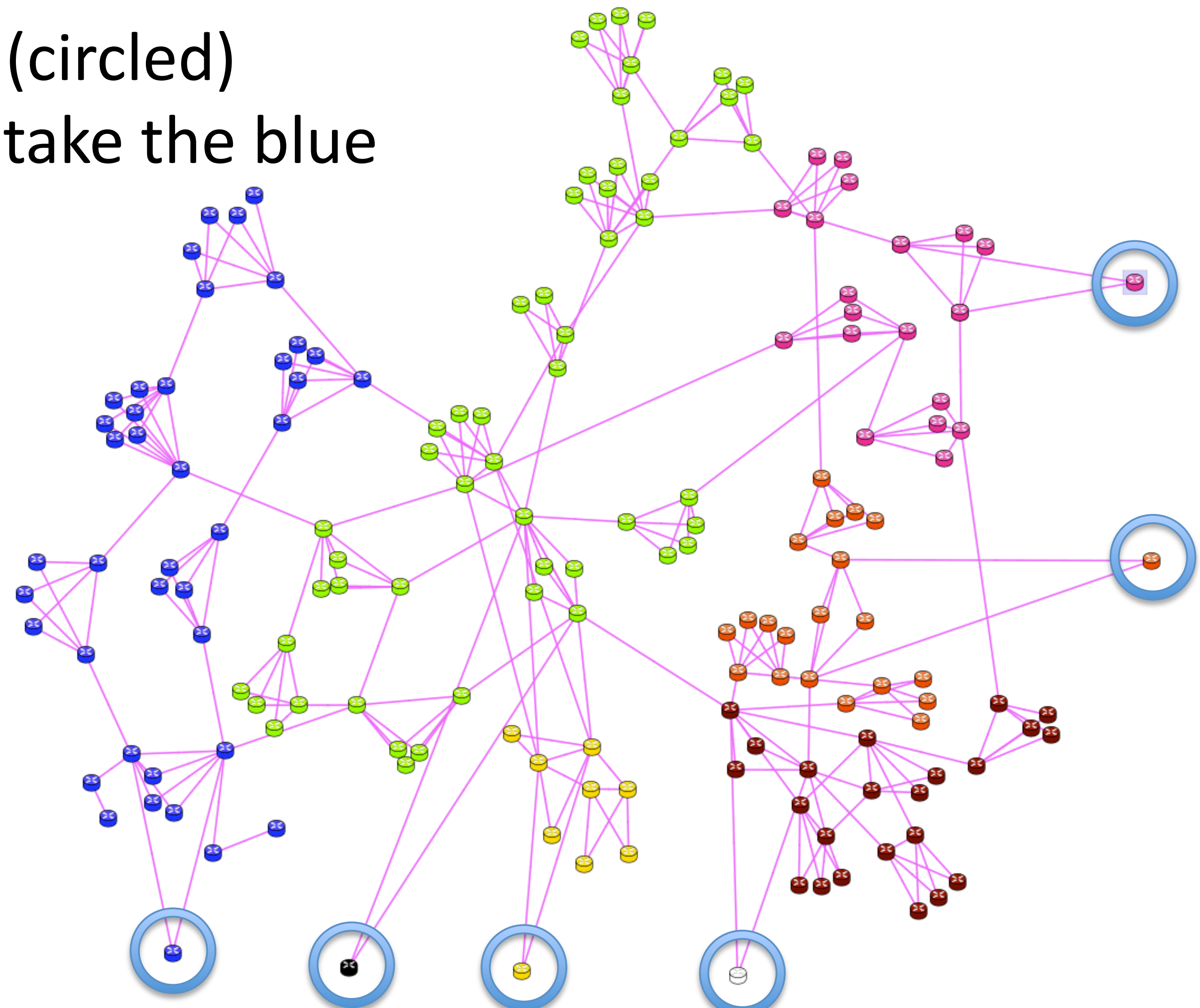
- Network health assessment
  - Capacity planning
  - Anomaly detection
  - Path bottleneck analysis (bandwidth, delay, metric)
  - Failure analysis
- Topology-aware traffic analysis
  - Where is the traffic coming from, going to, its path and why?
  - *Feed this back into path computation as a traffic matrix*
- Proactive change modeling
  - Add/drop routers/switches/links/prefixes/peerings
  - Add/drop applications/services
  - Analyze the impact on paths and traffic levels
  - *Feed this into provisioning*

Use Case: Diagnosing Black Holing

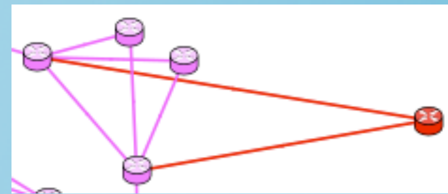
- A peering router to a major service provider crashed
  - Hot swappable card was not quite so…
- Traffic to the SP was black-holed network-wide
  - Traffic that exited in all 6 locations was all black-holed
- About 3 minutes of routing outage
  - 3 minutes was too short to diagnose the issue at human speed
  - Had a 45 minute ad-revenue impact on the services
    - Users who can not use the service do something else

Expected Exit-Points Before Incident

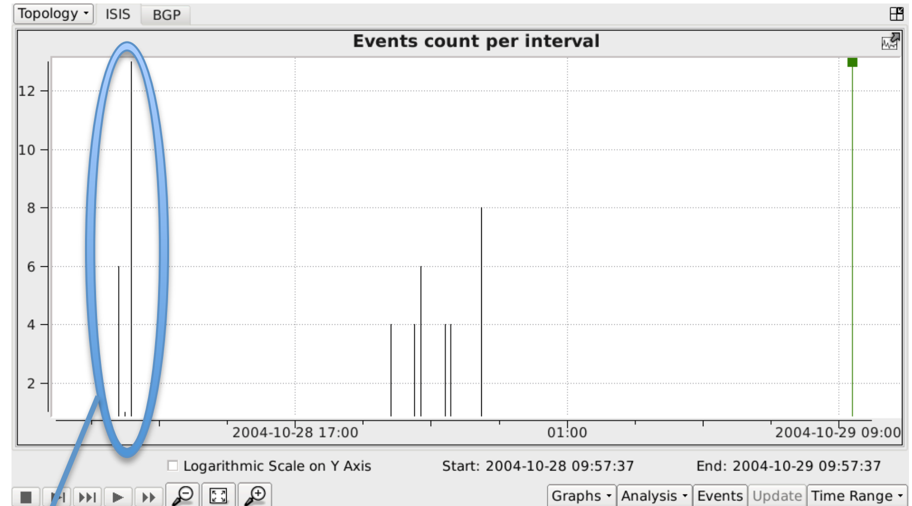
- 6 Exit-Points (circled)
- Blue routers take the blue exit router…



The Incident



- ISIS activity during incident



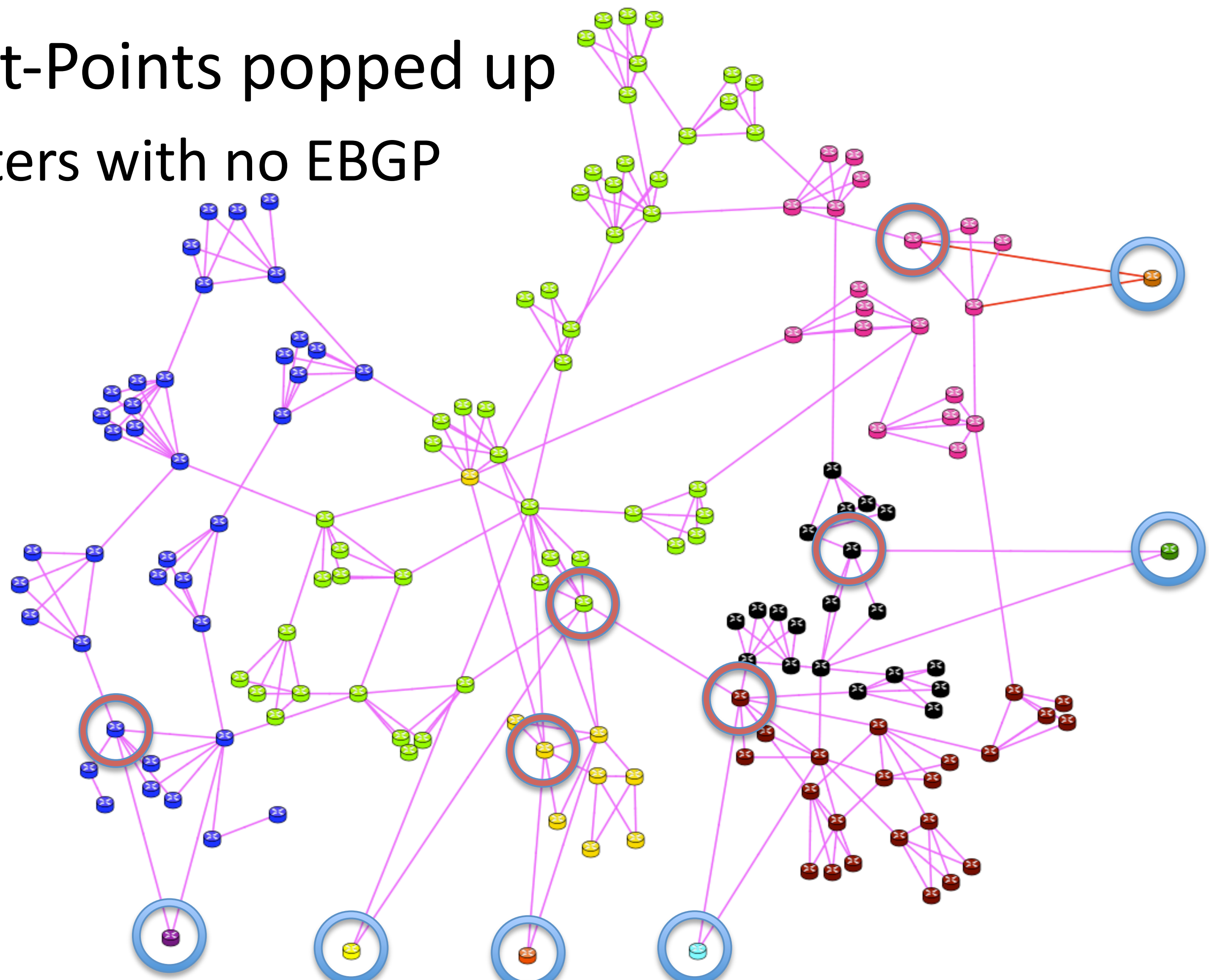
edge-ord-02

Time	Router	Operation	Operand	Attributes
2004-10-28 07:36:11.974206	core-ord-01	Drop Neighbor	edge-ord-02	Metric: Down (TE)
2004-10-28 07:36:12.374093	core-ord-02	Drop Neighbor	edge-ord-02	Metric: Down (TE)
2004-10-28 07:38:09.063564	core-ord-01	Add Neighbor	edge-ord-02	Metric: 503 (TE)
2004-10-28 07:38:36.071999	core-ord-02	Add Neighbor	edge-ord-02	Metric: 503 (TE)

4 entries 2004-10-28 03:32:47 - 10:58:08

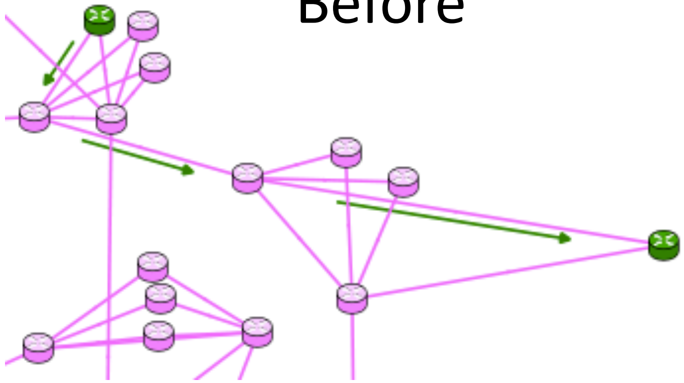
Exit-Points During Incident

- 6 more Exit-Points popped up
  - Core routers with no EBGP

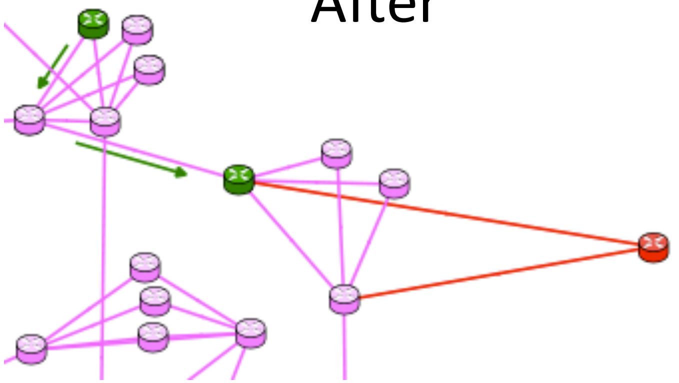


A Path Before and After the Incident

Before



After



BGP Next hop resolution: before 128.9.129.1/32 in ISIS vs after 128.9.128.0/19 in BGP

Path	Source Node	Destination Node	Protocol	Resolved by Prefix
edge-dfw-03→ 199.221.80.0/24				
Hop 1	edge-dfw-03	core-dfw-01	BGP	199.221.80.0/24
Hop 2	core-dfw-01	core-aus-01	BGP	199.221.80.0/24
Hop 3	core-aus-01	edge-aus-01	BGP	199.221.80.0/24
Lookup 1			ISIS	128.9.129.1/32

Route Recursion

Path	Source Node	Destination Node	Protocol	Resolved by Prefix
edge-dfw-03→ 199.221.80.0/24				
Hop 1	edge-dfw-03	core-dfw-01	BGP	199.221.80.0/24
Hop 2	core-dfw-01	core-aus-01	BGP	199.221.80.0/24
Self Hop	core-aus-01	core-aus-01	BGP	199.221.80.0/24
Lookup 1			BGP	128.9.128.0/19

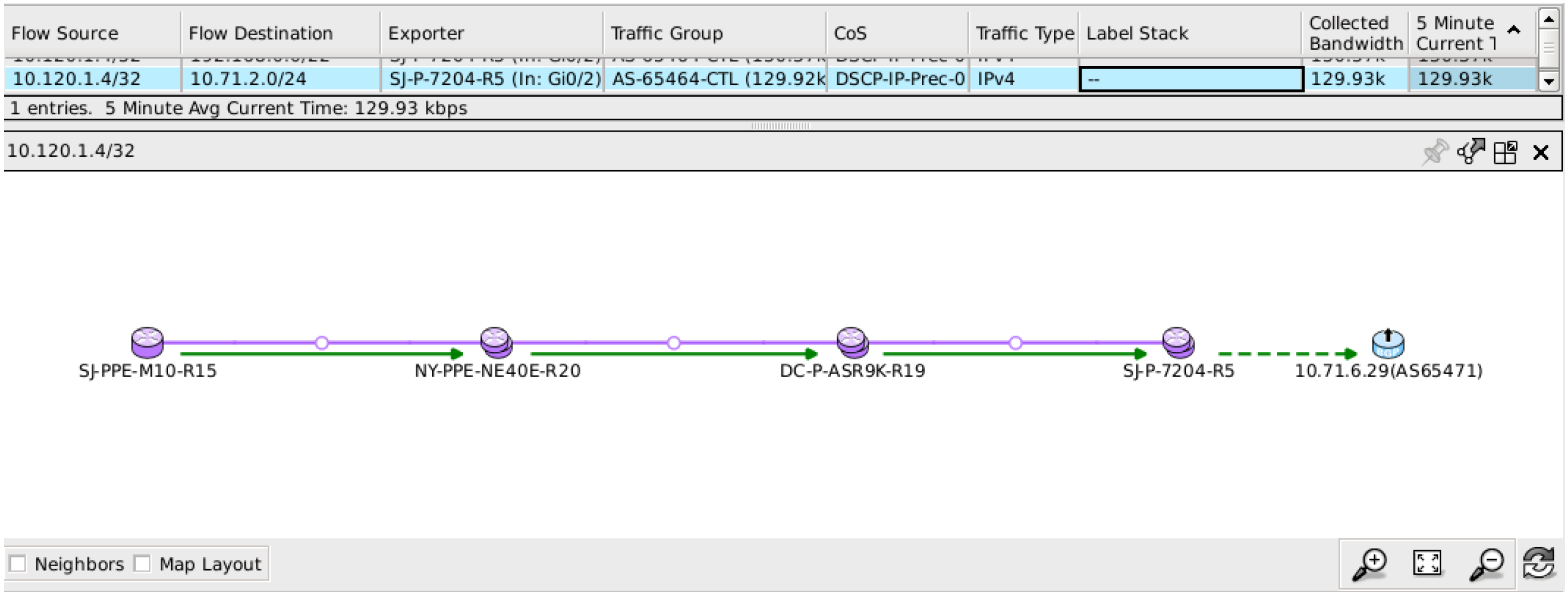
Cause of Black Holing

- When the peering router crashed
  - IGP routes were withdrawn in seconds
  - BGP routes were not withdrawn
    - 3 KEEPALIVEs of 60 seconds each – router rebooted before this
  - The BGP routes were now resolved by the /19 prefix in BGP
    - The /19 BGP announces internal address space – not meant for this
    - Injected by 6 core routers – cost from any router to a core router is very low
- Remedy:
  - Insert a really expensive static route for the /19 to ISIS
    - Cost more than longest possible path in IGP
  - Now, when a peering router crashes, the traffic will choose a true exit
  - See http://www.nanog.org/meetings/nanog34/presentations/gill.pdf
- Do not:
  - Make IBGP session to converge faster (like running BFD)
    - You will lose the IBGP session each time the IGP path of the session changes

Use Case: BGP Peering Traffic Analysis

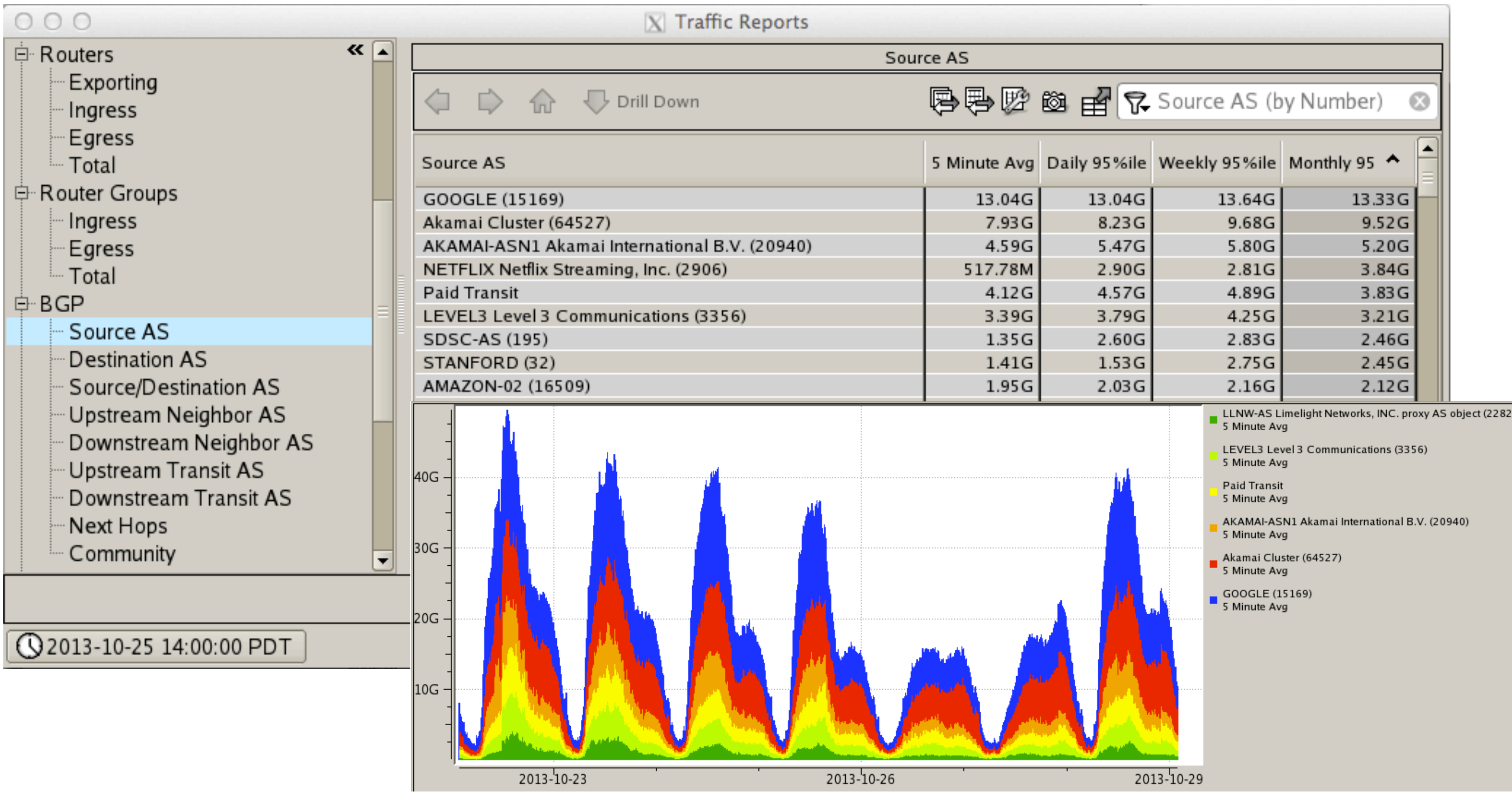
- For most regional networks, incoming traffic from the Internet is higher than outgoing traffic
  - What ASs (sources) is the traffic coming from
  - What neighbor AS did it come from
  - How is it distributed across the edge routers
- Need answers for:
  - Should I upgrade external links or get new links?
    - To whom? - The same neighbor ASs or new neighbor ASs?
  - Who can I peer with to cut transit cost?

Routing-Aware Traffic Flow Analysis



- Route-Flow Fusion determines each flow's ingress/egress and path across the network and onto the Internet
  - Delivers traffic matrices for planning and path computation
  - Helps make peering decisions to save transit cost

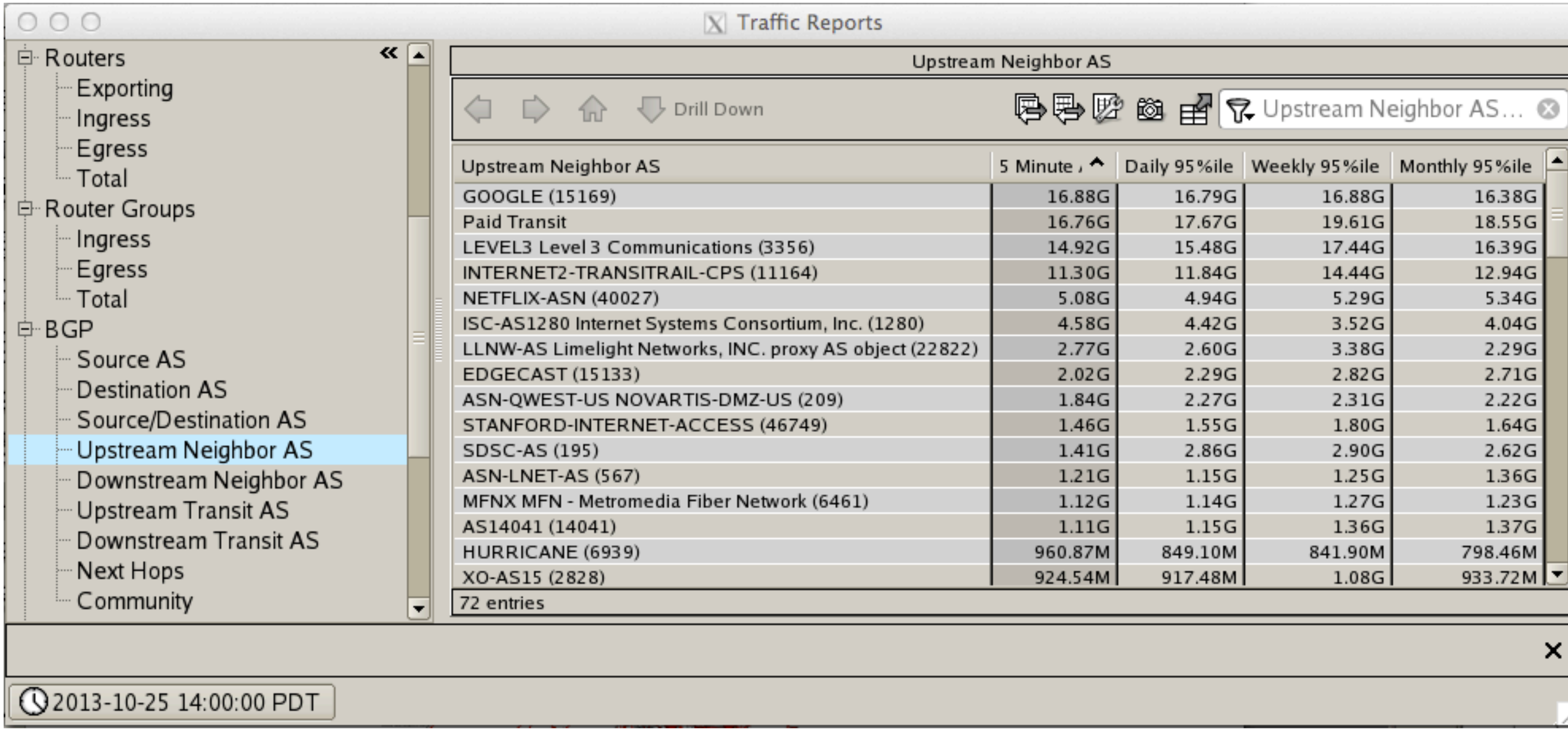
Traffic by Source AS



Source AS	5 Minute Avg	Daily 95%ile	Weekly 95%ile	Monthly 95
GOOGLE (15169)	13.04G	13.04G	13.64G	13.33G
Akamai Cluster (64527)	7.93G	8.23G	9.68G	9.52G
AKAMAI-ASN1 Akamai International B.V. (20940)	4.59G	5.47G	5.80G	5.20G
NETFLIX Netflix Streaming, Inc. (2906)	517.78M	2.90G	2.81G	3.84G
Paid Transit	4.12G	4.57G	4.89G	3.83G
LEVEL3 Level 3 Communications (3356)	3.39G	3.79G	4.25G	3.21G
SDSC-AS (195)	1.35G	2.60G	2.83G	2.46G
STANFORD (32)	1.41G	1.53G	2.75G	2.45G
AMAZON-02 (16509)	1.95G	2.03G	2.16G	2.12G

- Peering with Google can save the most cost
- Not all sources may be feasible to peer with

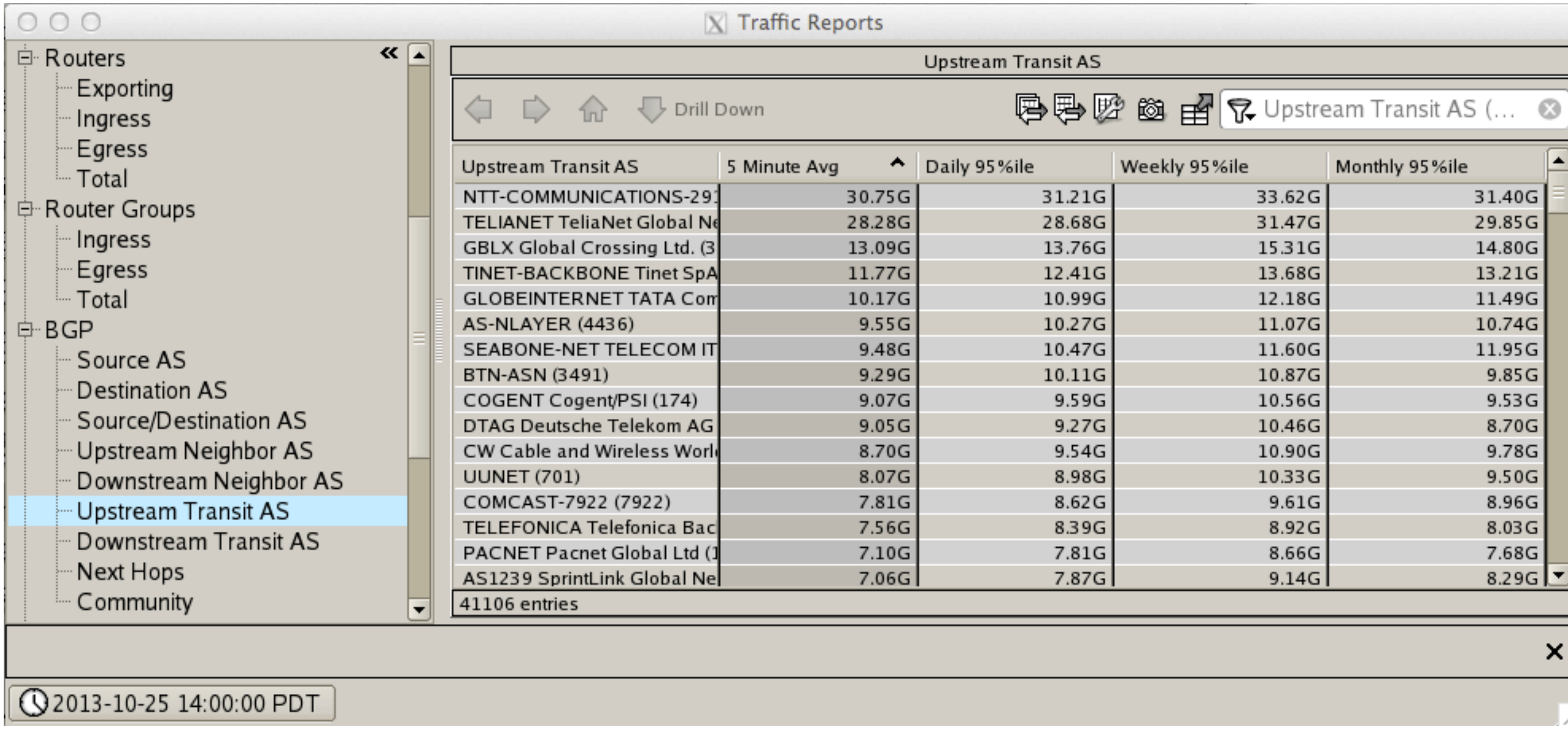
Neighbor ASs



Upstream Neighbor AS	5 Minute	Daily 95%ile	Weekly 95%ile	Monthly 95%ile
GOOGLE (15169)	16.88G	16.79G	16.88G	16.38G
Paid Transit	16.76G	17.67G	19.61G	18.55G
LEVEL3 Level 3 Communications (3356)	14.92G	15.48G	17.44G	16.39G
INTERNET2-TRANSITRAIL-CPS (11164)	11.30G	11.84G	14.44G	12.94G
NETFLIX-ASN (40027)	5.08G	4.94G	5.29G	5.34G
ISC-AS1280 Internet Systems Consortium, Inc. (1280)	4.58G	4.42G	3.52G	4.04G
LLNW-AS Limelight Networks, INC. proxy AS object (22822)	2.77G	2.60G	3.38G	2.29G
EDGECAST (15133)	2.02G	2.29G	2.82G	2.71G
ASN-QWEST-US NOVARTIS-DMZ-US (209)	1.84G	2.27G	2.31G	2.22G
STANFORD-INTERNET-ACCESS (46749)	1.46G	1.55G	1.80G	1.64G
SDSC-AS (195)	1.41G	2.86G	2.90G	2.62G
ASN-LNET-AS (567)	1.21G	1.15G	1.25G	1.36G
MFNX MFN - Metromedia Fiber Network (6461)	1.12G	1.14G	1.27G	1.23G
AS14041 (14041)	1.11G	1.15G	1.36G	1.37G
HURRICANE (6939)	960.87M	849.10M	841.90M	798.46M
XO-AS15 (2828)	924.54M	917.48M	1.08G	933.72M

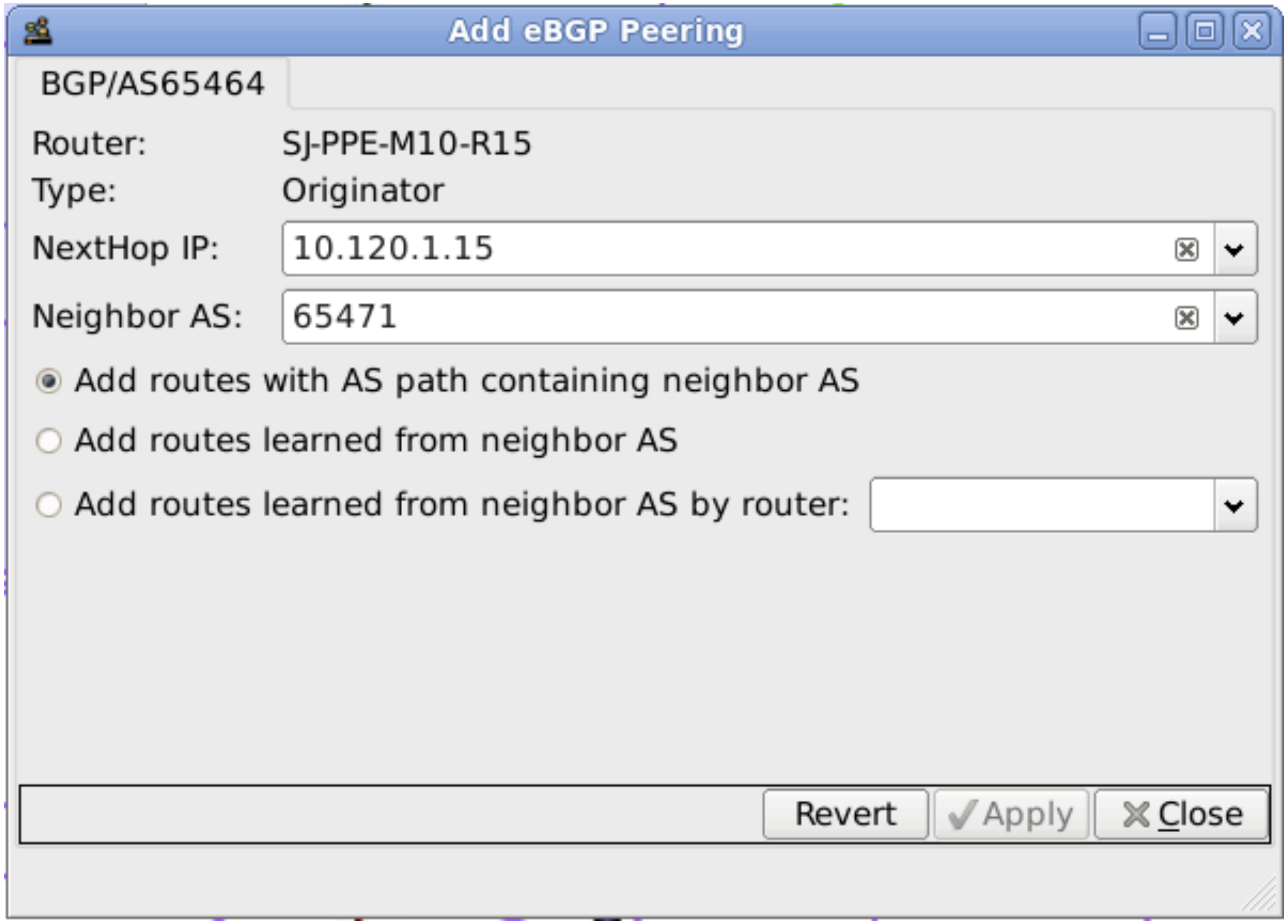
- Reveals where the traffic enters your network
- Is the traffic balanced the way you would like?
- Capacity planning with this data tells when it is time to upgrade or add new peering

Upstream Transit ASs

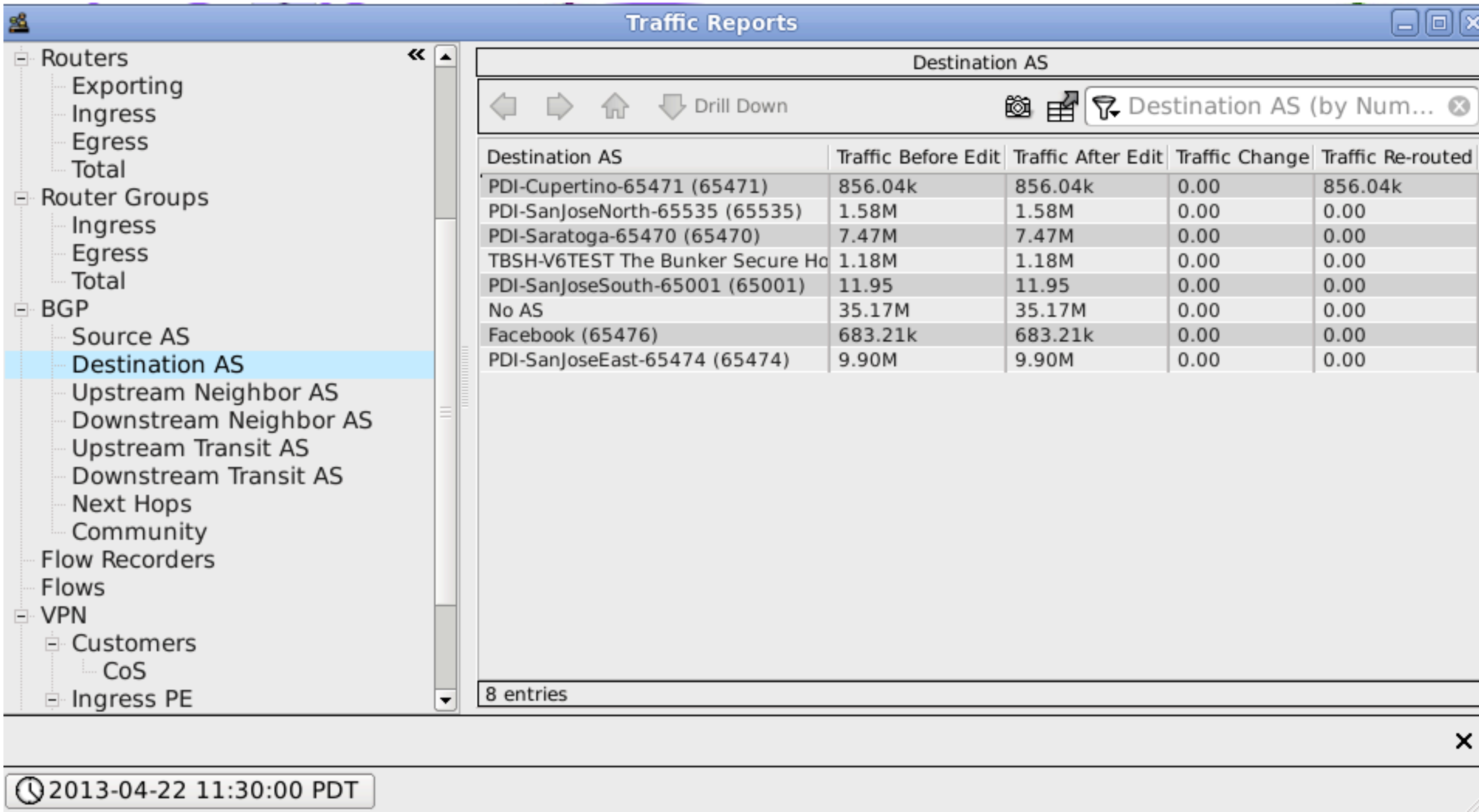


- Offers the good choices for peering – very hard to compute
  - This is not in your control, rather in the source AS's control
  - A heuristics-based solution:
    - Alternate routes for a given source are exposed during BGP convergence
    - Use these and graph search to compute these values

Modeling a New BGP Peer



Before-and-After Comparison Reveals Transit Traffic Shifts

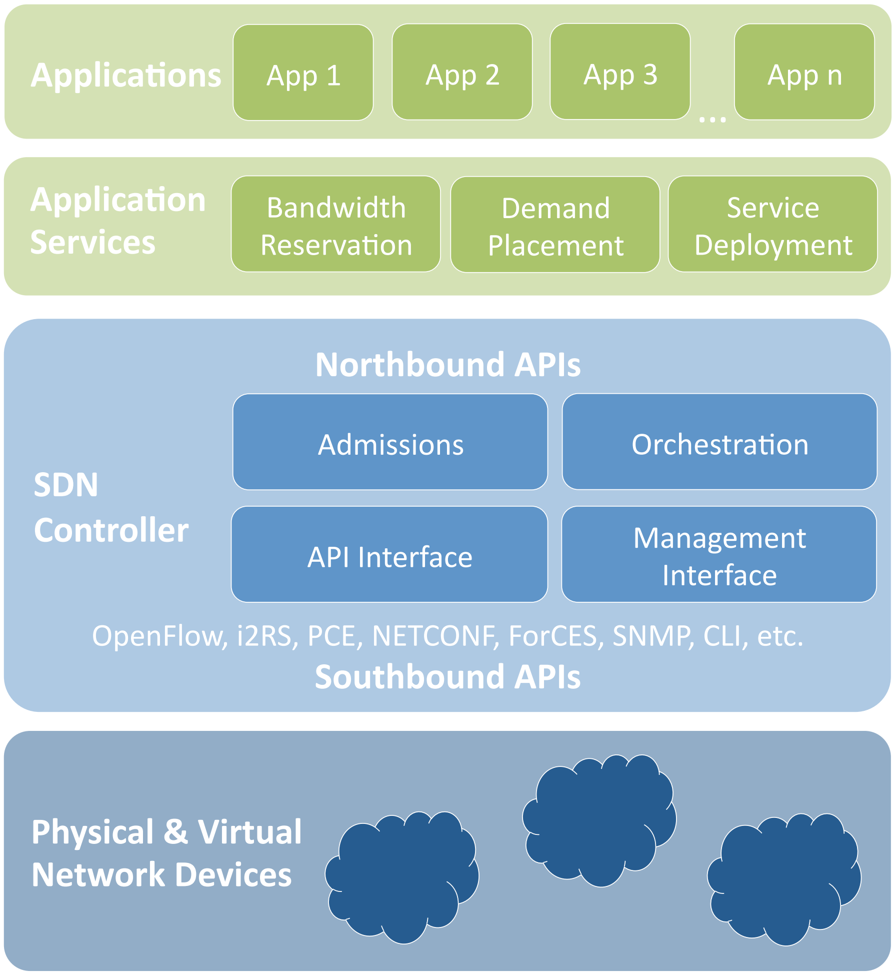


Destination AS	Traffic Before Edit	Traffic After Edit	Traffic Change	Traffic Re-routed
PDI-Cupertino-65471 (65471)	856.04k	856.04k	0.00	856.04k
PDI-SanJoseNorth-65535 (65535)	1.58M	1.58M	0.00	0.00
PDI-Saratoga-65470 (65470)	7.47M	7.47M	0.00	0.00
TBSH-V6TEST The Bunker Secure Ho	1.18M	1.18M	0.00	0.00
PDI-SanJoseSouth-65001 (65001)	11.95	11.95	0.00	0.00
No AS	35.17M	35.17M	0.00	0.00
Facebook (65476)	683.21k	683.21k	0.00	0.00
PDI-SanJoseEast-65474 (65474)	9.90M	9.90M	0.00	0.00

Route Analytics Simplifies Planning for BGP Peering Analysis

- Traffic volumes are obtained by fusing flows with routes
- Create a bunch of “mocked-up” BGP routes
- Route-Flow Fusion will now follow these routes
  - Simple and elegant
  - No simulation, emulation or inaccurate modeling used
- Impact analysis
  - Route analytics knows what your network will do

Challenges in Software Defined Networking



SDN makes networks programmable for

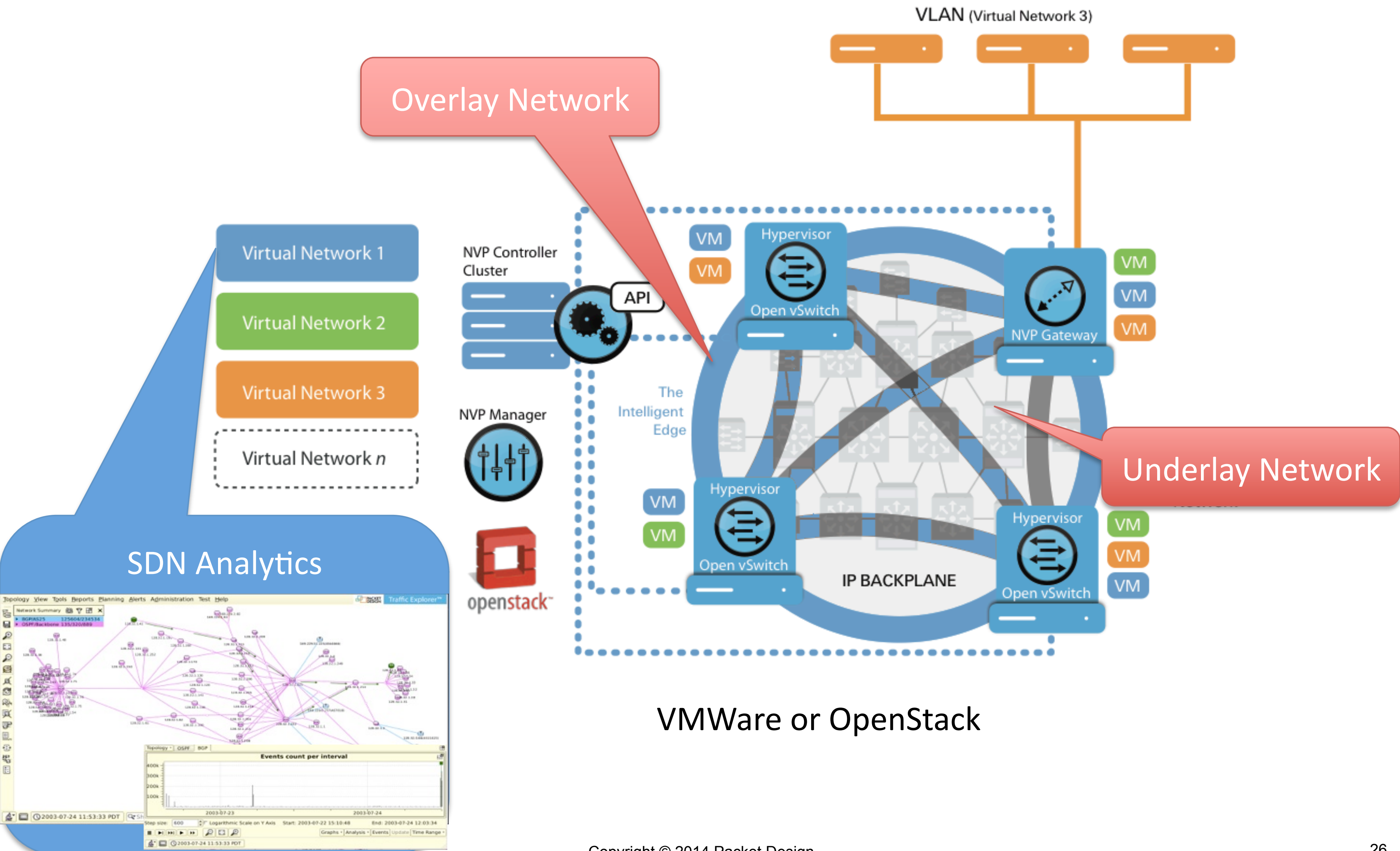
- Network overlays
- Bandwidth reservation
- Demand placement
- Service deployment
- Etc.

What governs whether or not these programmatic changes should be made?

What will be their impact?

Network Virtualization SDN Analytics

Tying Overlay and Underlay Networks



VMWare or OpenStack

Concluding Remarks

- Routing impacts network performance
  - Availability and reachability
  - Sub-optimal paths with longer delays, jitter
- Route analytics proves to be very effective
  - Troubleshooting, monitoring, alerting
  - Reporting and network health assessment
  - Routing-aware traffic analysis
    - BGP peering analysis
    - Traffic matrices
- Route analytics provides foundation for SDN Analytics