the economics of network control

TUTORIAL

Best Practices for Determining the Traffic Matrix in IP Networks V 3.0

NANOG 39
Toronto

February 5th 2007, 2:00pm-3:30pm

Thomas Telkamp, *Cariden Technologies, Inc.*

Agenda

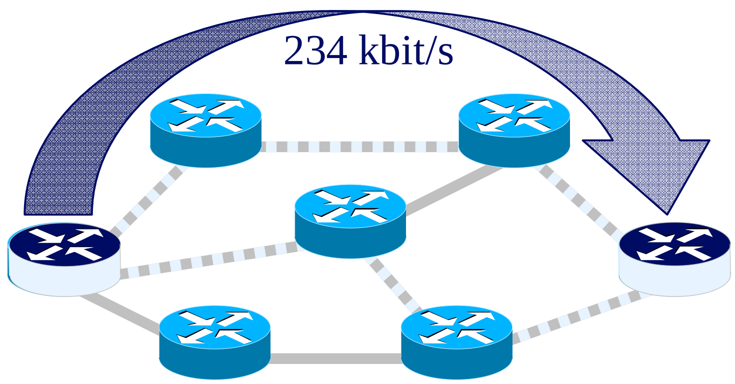
- Introduction
  - Traffic Matrix Properties
- Measurement in IP networks
  - NetFlow
  - BGP Policy Accounting
  - DCU (Juniper)
- MPLS Networks
  - RSVP based TE
  - LDP
    - Data Collection
    - LDP deployment in Deutsche Telekom
- Estimation Techniques
  - Theory
  - Example Data
  - Case-Study
    - Simulation
    - Metric Optimization
    - Adding p-t-p Measurements
- Traffic Matrices in Partial Topologies
- Multicast
- Summary

Contributors

- Stefan Schnitter, *T-Systems*
  - MPLS/LDP, Partial Topologies
- Benoit Claise, *Cisco Systems, Inc.*
  - Cisco NetFlow
- Tarun Dewan, *Juniper Networks, Inc.*
  - Juniper DCU
- Mikael Johansson, *KTH*
  - Traffic Matrix Properties
- Simon Leinen, *SWITCH*
  - BGP Policy Accounting, SNMP

Traffic Matrix

- Traffic matrix: the amount of data transmitted between every pair of network nodes
  - Demands
  - "end-to-end" in the core network
- Traffic Matrix can represent peak traffic, or traffic at a specific time
- Router-level or PoP-level matrices



Determining the Traffic Matrix

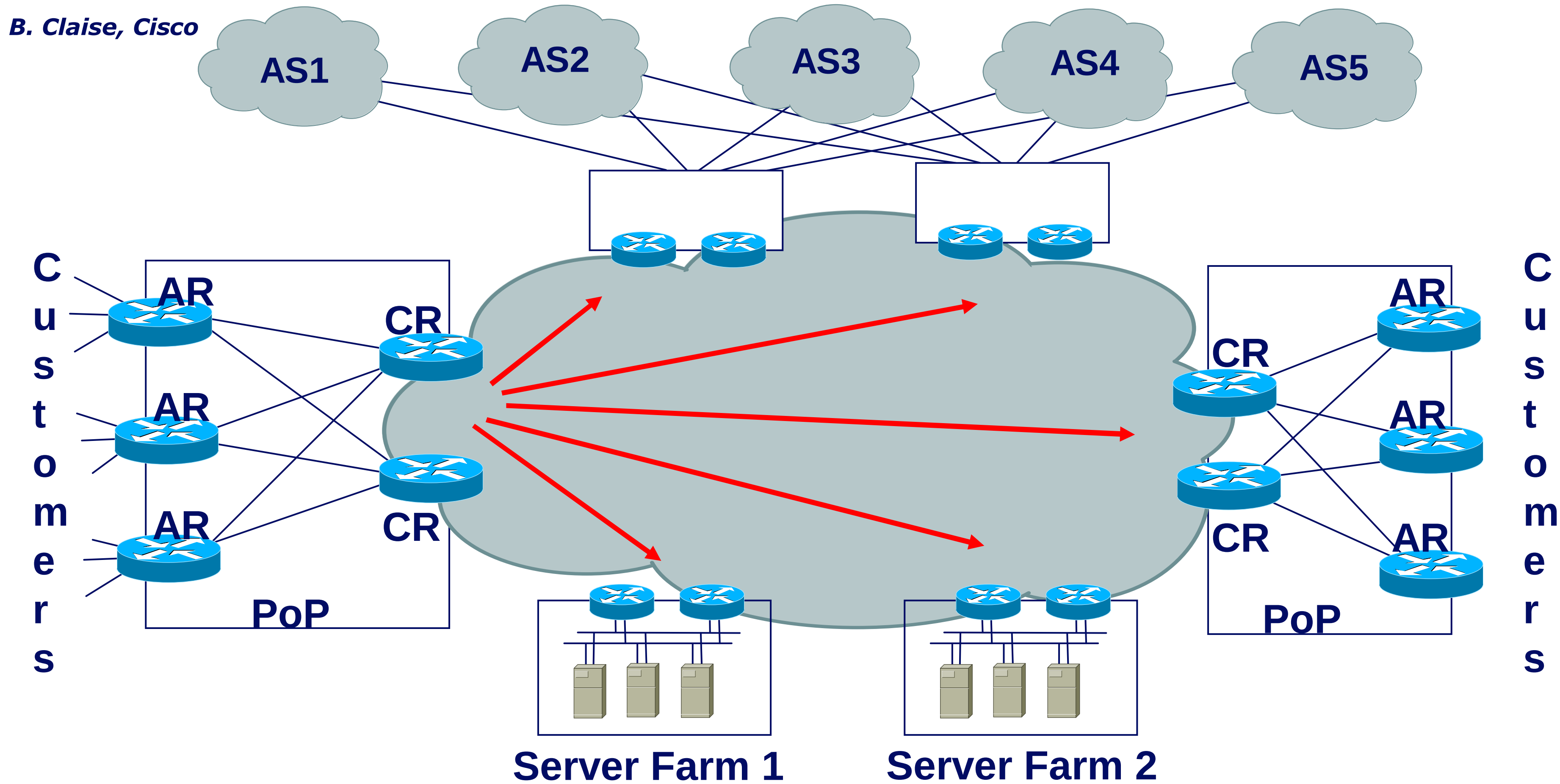
- For what purpose?
  - Analysis and Evaluation of other network states than the current:
  - Capacity Planning
    - network changes
    - “what-if” scenarios
    - Could be per class
  - Resilience Analysis
    - network under failure conditions
  - Optimization
    - Topology
      - Find bottlenecks
    - OSPF/IS-IS metric optimization/TE
    - MPLS TE tunnel placement

Types of Traffic Matrices

- Internal Traffic Matrix
  - PoP to PoP matrix
    - Can be from core (CR) or access (AR) routers
  - Class based

- External Traffic Matrix
  - PoP to External AS
    - BGP
    - Origin-AS or Peer-AS
      - Peer-AS sufficient for Capacity Planning and Resilience Analysis
    - Useful for analyzing the impact of external failures on the core network (capacity/resilience)
    - See RIPE presentation on peering planning [8]

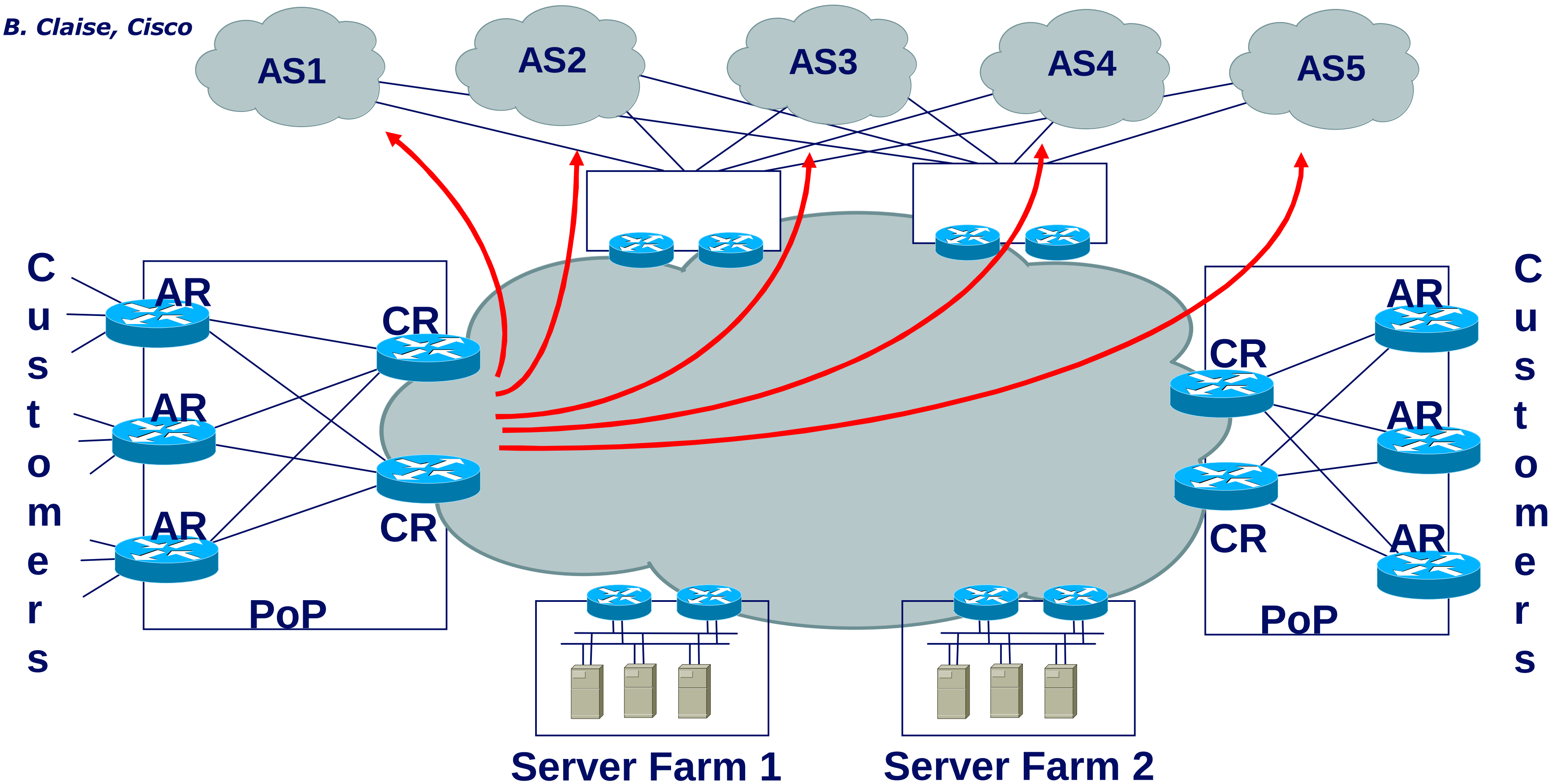
Internal Traffic Matrix



"PoP to PoP", the PoP being the **AR** or **CR**

External Traffic Matrix

B. Claise, Cisco



From "PoP to BGP AS", the PoP being the **AR** or **CR**

The external traffic matrix can influence the internal one

Traffic Matrix & Routing

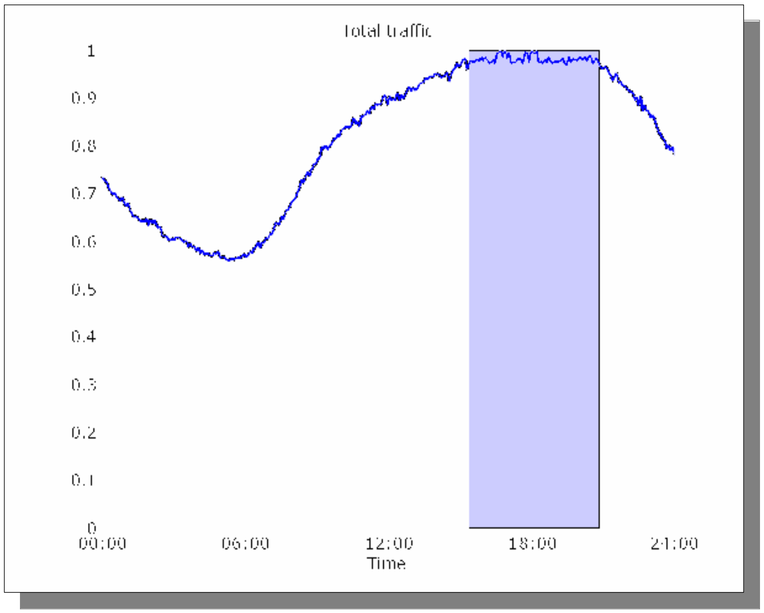
- Presented at NANOG35 ([7]; Jacobson, Yu, Mah):
  - Observation: traffic doesn’t just go everywhere, most of it goes to a few places
  - Routing data:
    - BGP neighbor AS
      - Customers, transit providers, peers, etc.
    - BGP IGP next-hop
      - Locations where they attach
  - Combine IGP/BGP routing data with traffic data
  - Provides more info than just the matrix itself

Traffic Matrix Properties

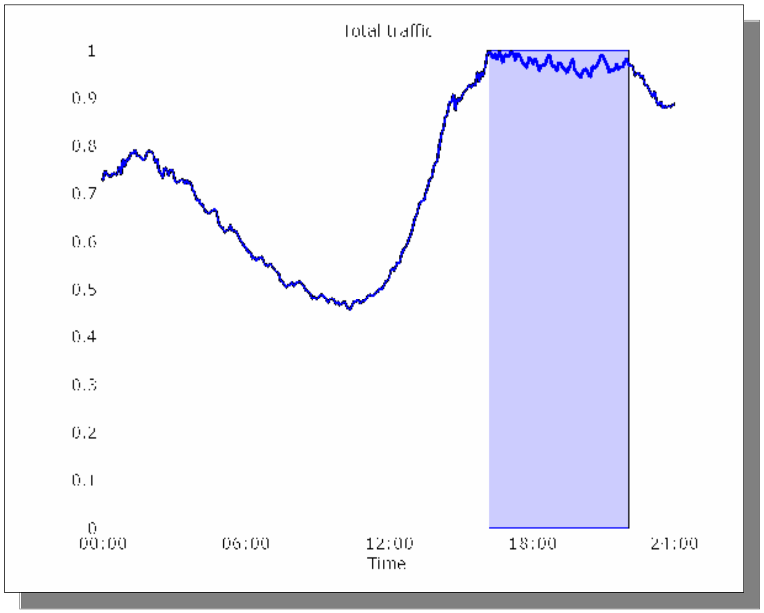
- Example Data from Tier-1 IP Backbone
  - Measured Traffic Matrix (MPLS TE based)
  - European and American subnetworks
  - 24h data
  - See [1]
- Properties
  - Temporal Distribution
    - How does the traffic vary over time
  - Spatial Distribution
    - How is traffic distributed in the network?
  - Relative Traffic Distribution
    - “Fanout”

Total traffic and busy periods

European subnetwork



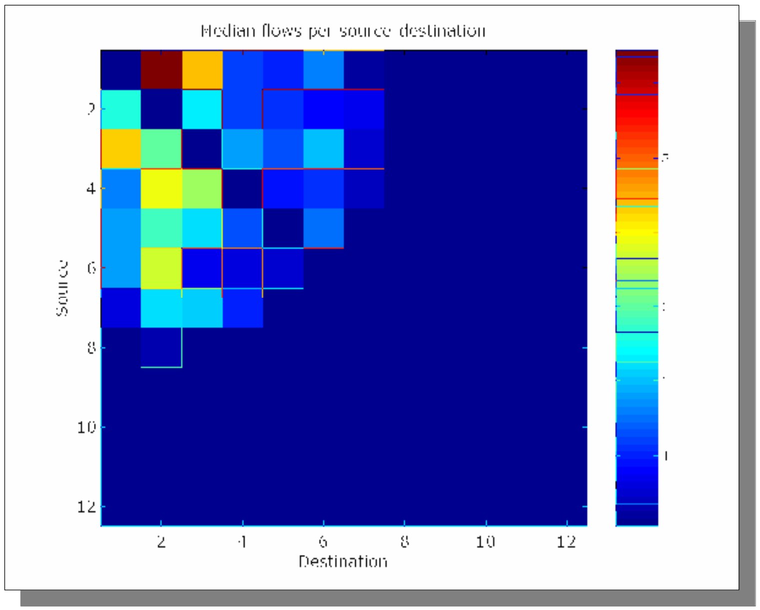
American subnetwork



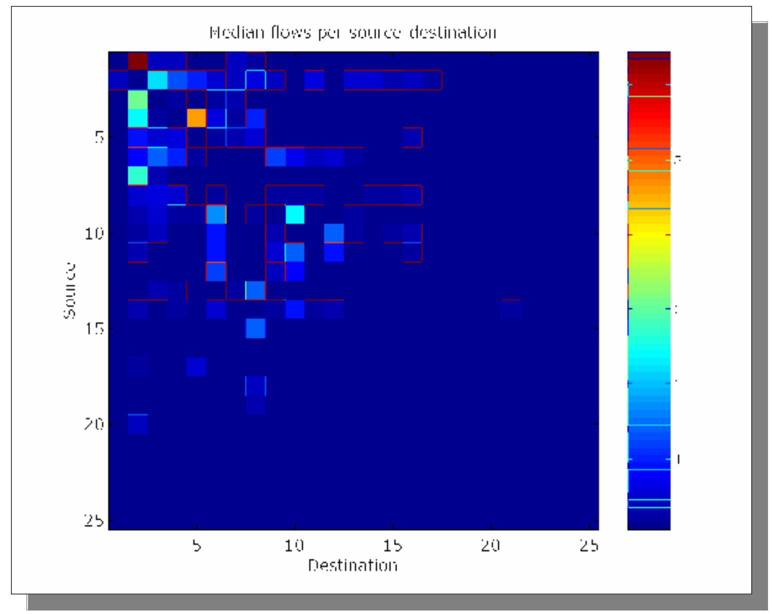
Total traffic very stable over 3-hour busy period

Spatial demand distributions

European subnetwork



American subnetwork

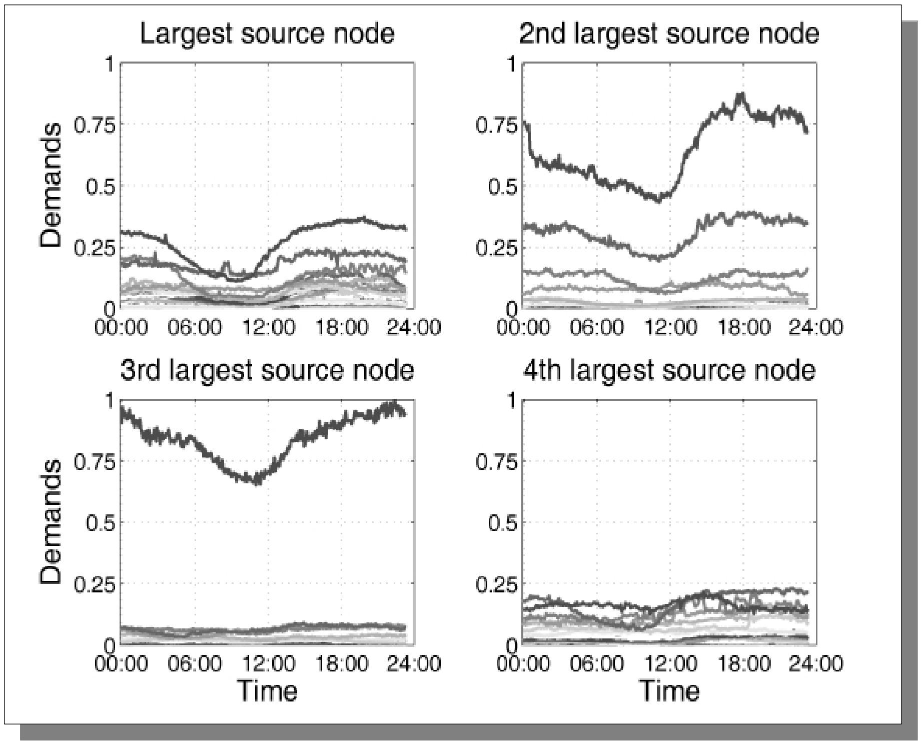


Few large nodes contribute to total traffic (20% demands – 80% of total traffic)

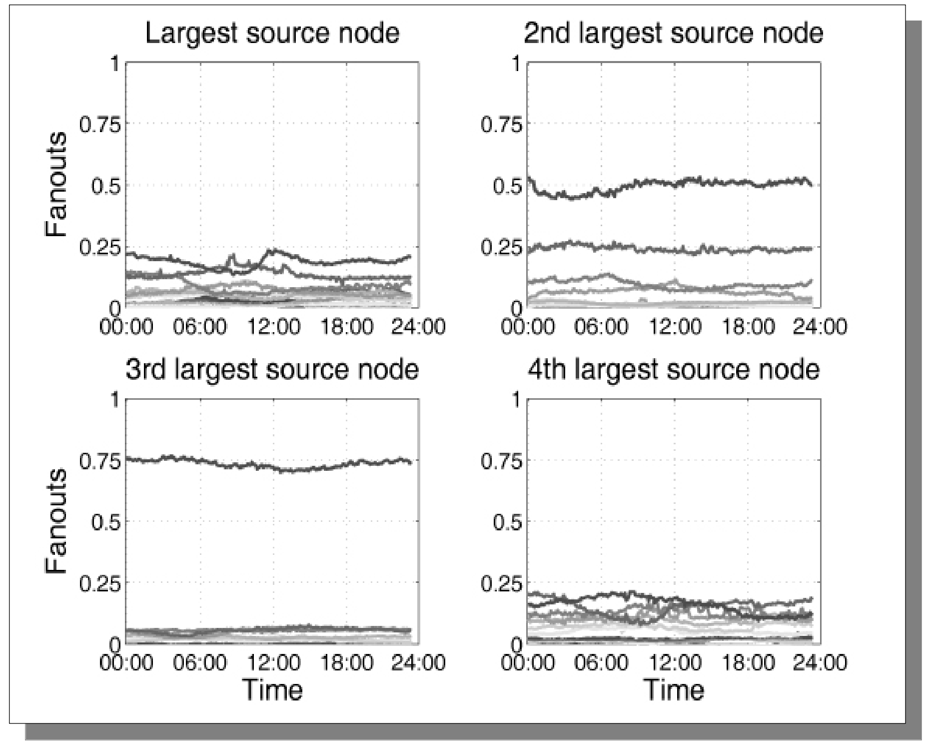
Fanout factors

Fanout: relative amount of traffic (as percentage of total)

Demands for 4 largest nodes, USA



Corresponding fanout factors



Fanout factors much more stable than demands themselves!

Traffic Matrix Collection

- Data is collected at fixed intervals
  - E.g. every 5 or 15 minutes
- Measurement of *Byte Counters*
  - Need to convert to rates
  - Based on measurement interval
  - Counter roll-over issues
- Create Traffic Matrix
  - Peak Hour Matrix
    - 5 or 15 min. average at the peak hour
  - Peak Matrix
    - Calculate the peak for every demand
    - Real peak or 95-percentile

Collection Methods

- NetFlow
  - Routers collect “flow” information
  - Export of raw or aggregated data
- BGP Policy Accounting/DCU
  - Routers collect aggregated destination statistics
- MPLS
  - RSVP
    - Measurement of Tunnel/LSP counters
  - LDP
    - Measurement of LDP counters
- Estimation
  - Estimate Traffic Matrix based on Link Utilizations

NetFlow based Methods

NetFlow

- A “Flow” is defined by
  - Source address
  - Destination address
  - Source port
  - Destination port
  - Layer 3 Protocol Type
  - TOS byte
  - Input Logical Interface (ifIndex)
- Router keeps track of Flows and usage per flow
  - Packet count
  - Byte count

NetFlow Versions

- □Version 5
  - the most complete version
- Version 7
  - on the switches
- Version 8
  - the Router Based Aggregation
- Version 9
  - the new flexible and extensible version

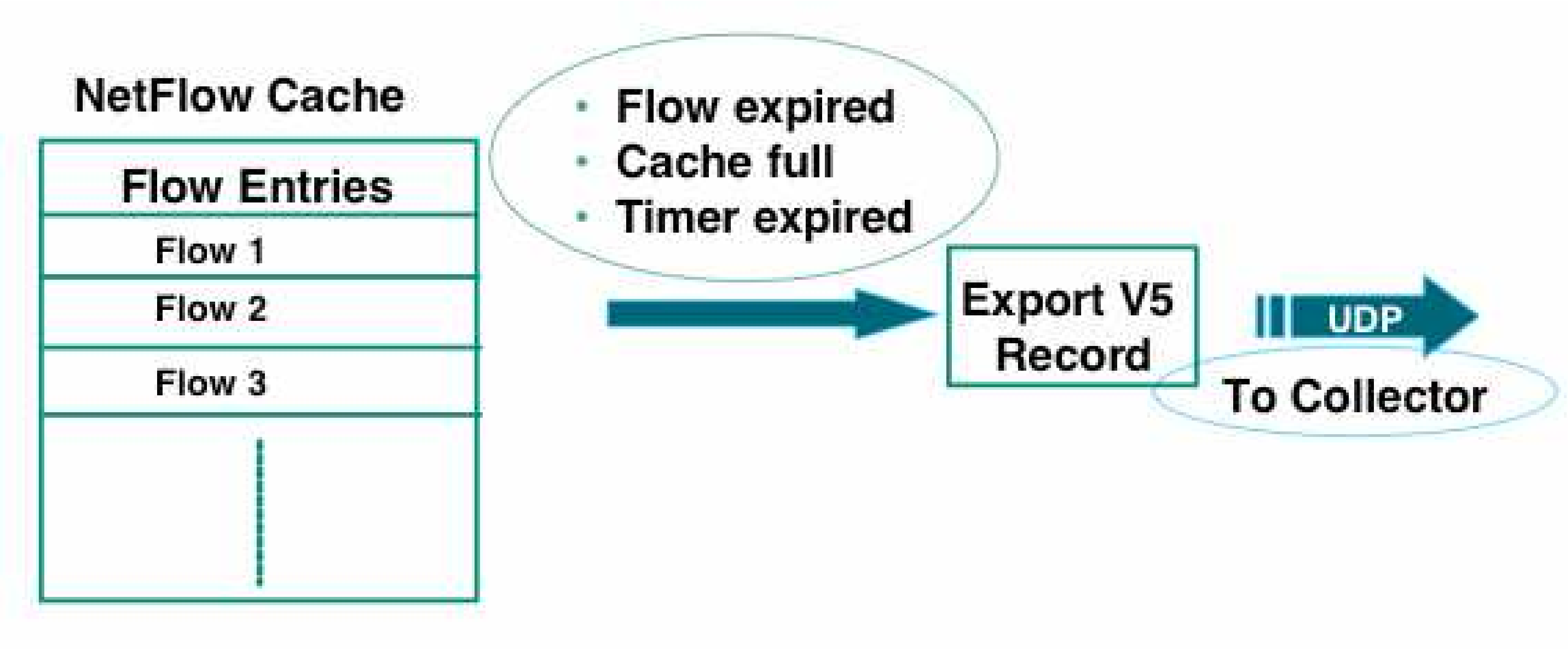
- Supported by multiple vendors
  - Cisco
  - Juniper
  - others

NetFlow Export

- A Flow is exported when
  - Flow expires
  - Cache full
  - Timer expired
- Expired Flows are grouped together into "NetFlow Export" UDP datagrams for export to a collector
  - Including timestamps
- UDP is used for speed and simplicity
- Exported data can include extra information
  - E.g. Source/Destination AS

NetFlow Export

B. Claise, Cisco



NetFlow Deployment

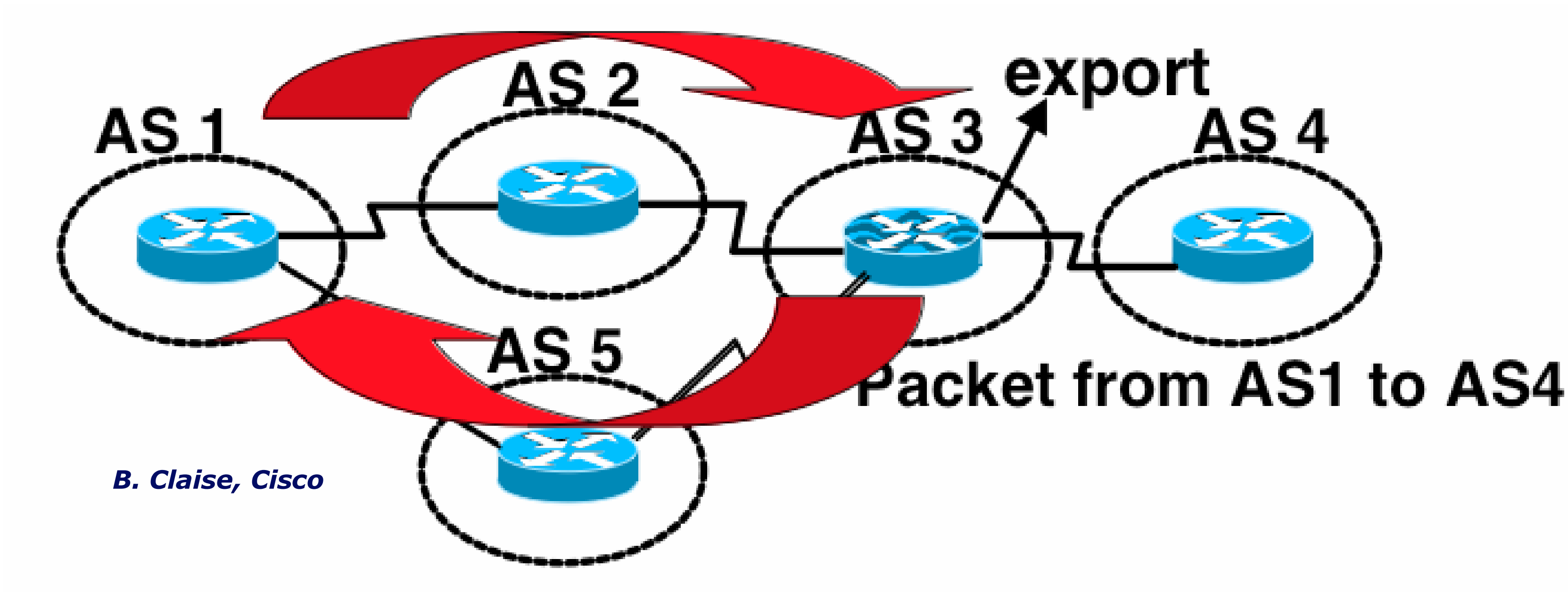
- How to build a Traffic Matrix from NetFlow data?
  - Enable NetFlow on all interfaces that source/sink traffic into the (sub)network
    - E.g. Access to Core Router links (AR->CR)
  - Export data to central collector(s)
  - Calculate Traffic Matrix from Source/Destination information
    - Static (e.g. list of address space)
    - BGP AS based
      - Easy for peering traffic
      - Could use "live" BGP feed on the collector
    - Inject IGP routes into BGP with community tag

BGP Passive Peer on the Collector

- Instead of exporting the peer-as or destination-as for the source and destination IP addresses for the external traffic matrix:
  - Don't export any BGP AS's
  - Export version 5 with IP addresses or version 8 with an prefix aggregation
- A BGP passive peer on the NetFlow collector machines can return all the BGP attributes:
  - source/destination AS, second AS, AS Path, BGP communities, BGP next hop, etc…
- Advantages:
  - Better router performance – less lookups
  - Consume less memory on the router
  - Full BGP attributes flexibility
  - See "Route-Flow Fusion" [7] again

NetFlow: Asymmetric BGP traffic

- Origin-as
  - Source AS1, Destination AS4
- Peer-as
  - Source AS5, Destination AS4 WRONG!
- Because of the source IP address lookup in BGP



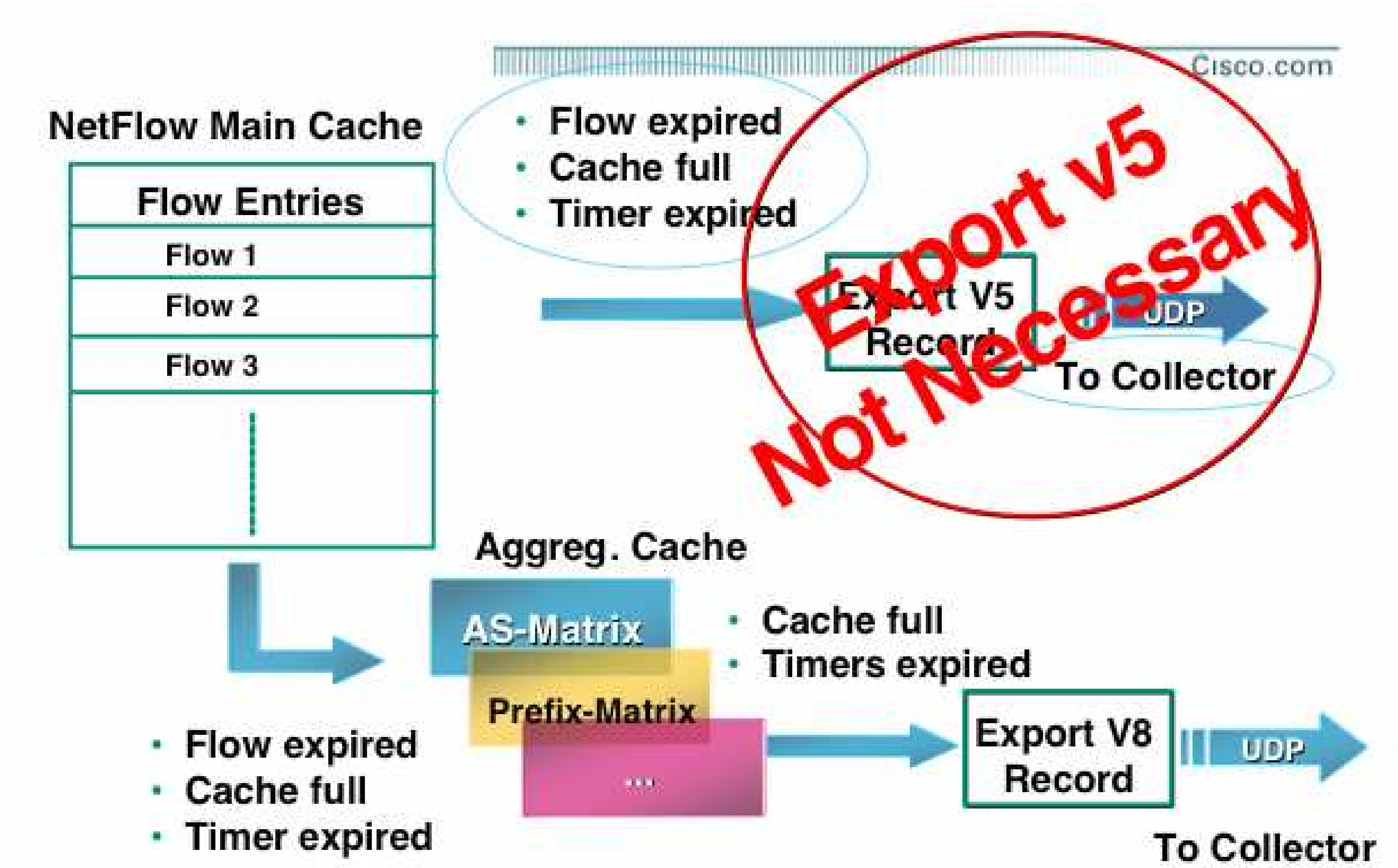
NetFlow Version 8

- Router Based Aggregation
- Enables router to summarize NetFlow Data
- Reduces NetFlow export data volume
  - Decreases NetFlow export bandwidth requirements
  - Makes collection easier

- Still needs the main (version 5) cache
- When a flow expires, it is added to the aggregation cache
  - Several aggregations can be enabled at the same time
- Aggregations:
  - Protocol/port, AS, Source/Destination Prefix, etc.

NetFlow: Version 8 Export

B. Claise, Cisco

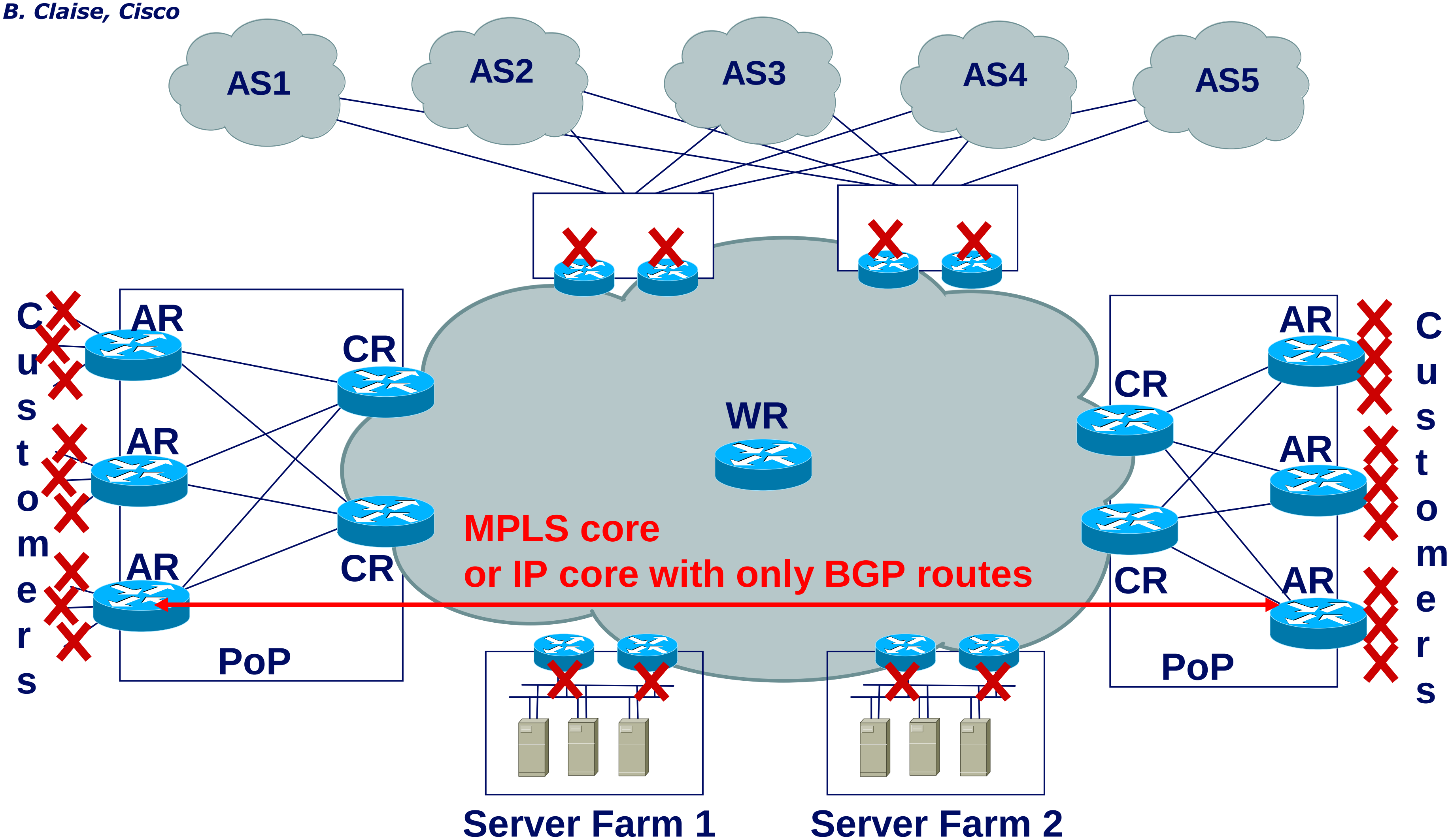


BGP NextHop TOS Aggregation

- New Aggregation scheme
  - Only for BGP routes
    - Non-BGP routes will have next-hop 0.0.0.0
- Configure on Ingress Interface
- Requires the new Version 9 export format
- Only for IP packets
  - IP to IP, or IP to MPLS

BGP NextHop TOS Aggregation

B. Claise, Cisco

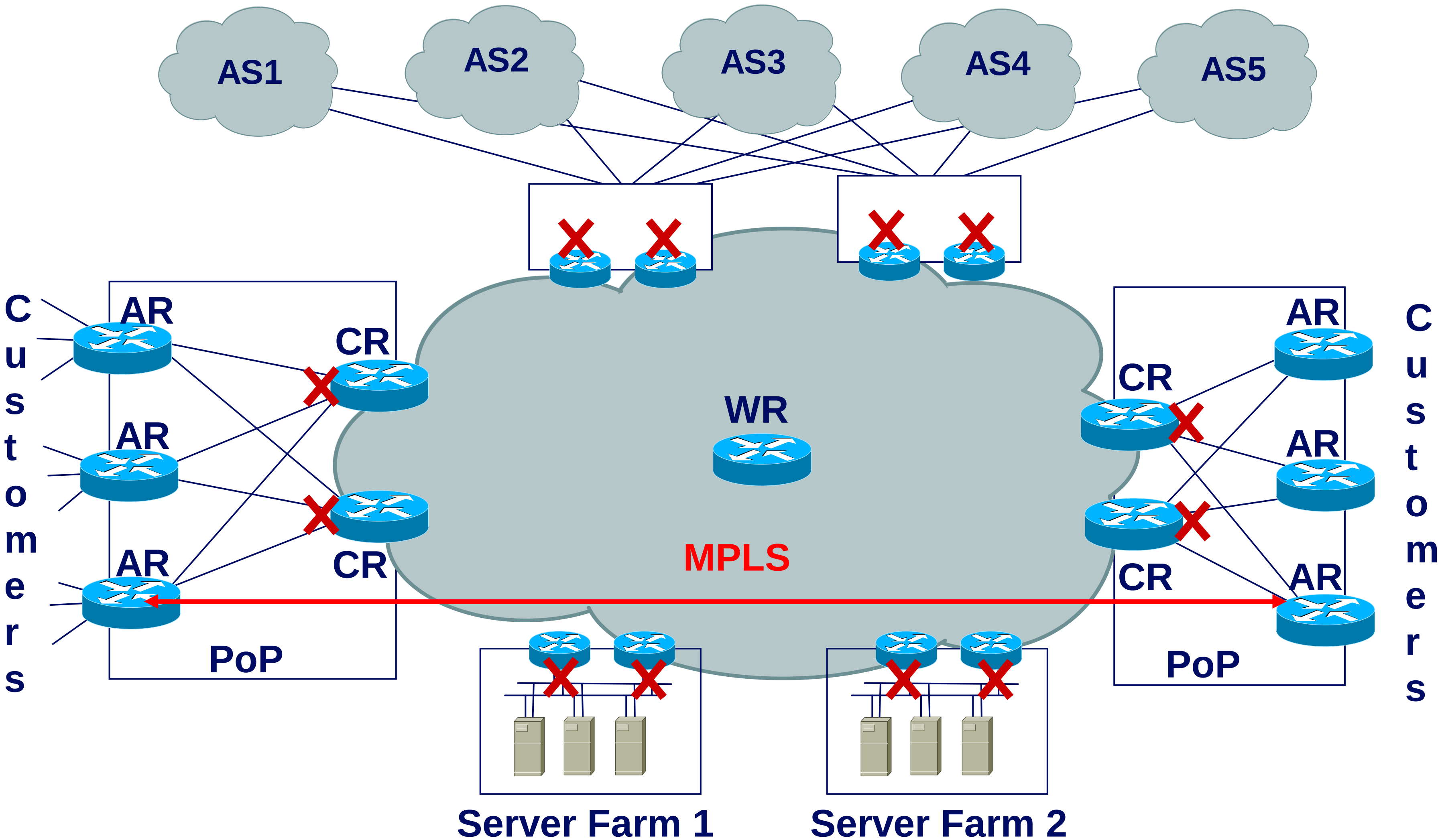


MPLS aware NetFlow

- Provides flow statistics per MPLS and IP packets
  - MPLS packets:
    - Labels information
    - And the V5 fields of the underlying IP packet
  - IP packets:
    - Regular IP NetFlow records
- Based on the NetFlow version 9 export
  No more aggregations on the router (version 8)
- Configure on ingress interface
- Supported on sampled/non sampled NetFlow

MPLS aware NetFlow: Example

B. Claise, Cisco



NetFlow Summary

- Building a Traffic Matrix from NetFlow data is not trivial
  - Need to correlate Source/Destination information with routers or PoPs
- "origin-as" vs "peer-as"
  - Asymmetric BGP traffic problem
- BGP NextHop aggregation comes close to directly measuring the Traffic Matrix
  - NextHops can be easily linked to a Router/PoP
  - BGP only
- NetFlow processing is CPU intensive on routers
  - Use Sampling
    - E.g. only use every 1 out of 100 packets
    - Accuracy of sampled data

NetFlow Summary

- Various other features are available
- Ask vendors (Cisco, Juniper, etc.) for details on version support and platforms
- Commercial collector systems are available:
  - Arbor
    - Also for other purposes, like DDoS
  - Adlex
  - Etc.
- For Cisco, see Benoit Claise's webpage:
  - http://www.employees.org/~bclaise/

BGP Policy Accounting

BGP Policy Accounting

- Accounting traffic according to the route it traverses
- Account for IP traffic by assigning counters based on:
  - BGP community-list
  - AS number
  - AS-path
  - destination IP address
- 64 buckets

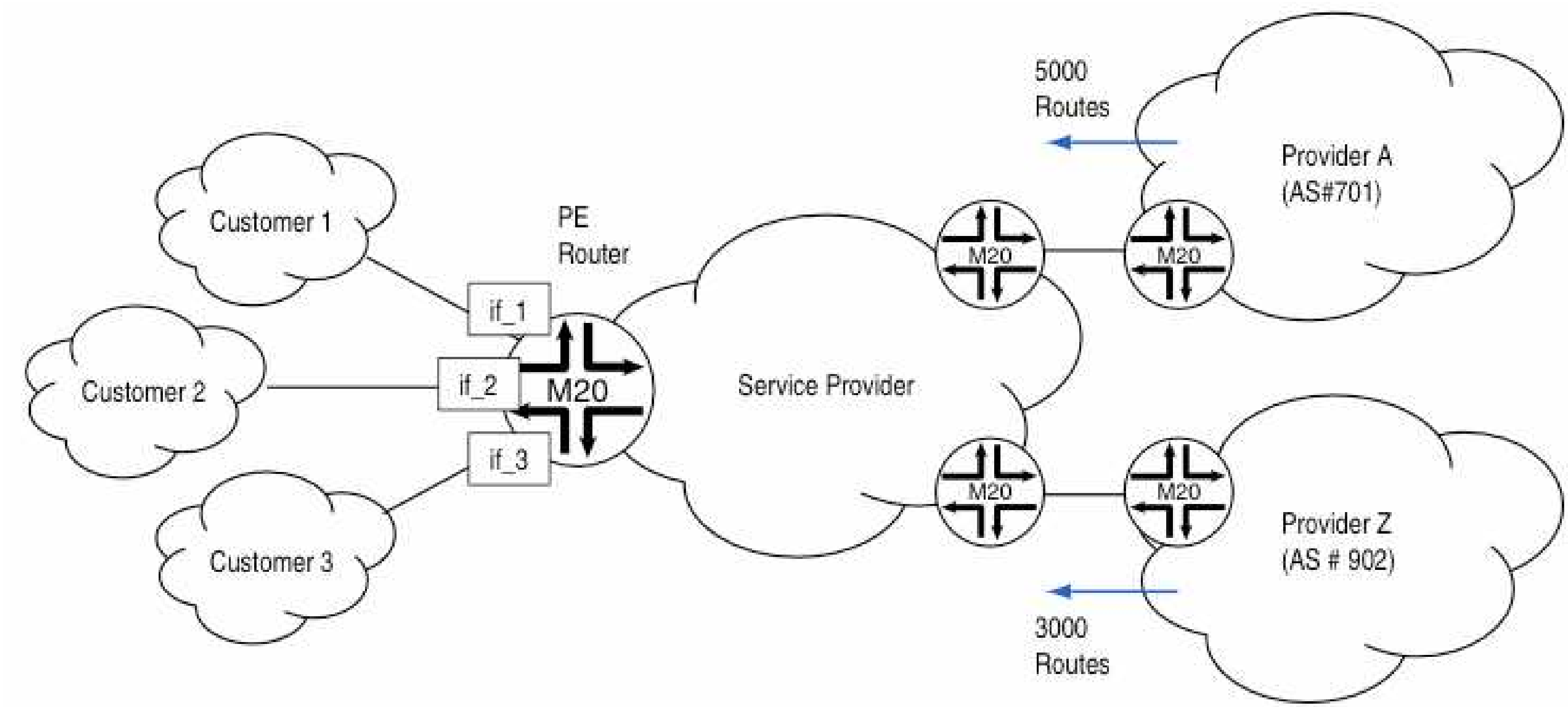
Destination Class Usage (DCU)

Destination Class Usage (DCU)

- Juniper specific!
- Similar to Cisco BGP Policy Accounting
  - Policy based accounting mechanism
  - For example based on BGP communities
- Supports up to 16 different traffic destination classes (126 in more recent releases)
- Maintains per interface packet and byte counters to keep track of traffic per class
- Data is stored in a file on the router, and can be pushed to a collector
- *16 destination classes is in most cases too limited to build a useful full Traffic Matrix, 126 is more realistic though.*

DCU Example

- Routing policy
  - associate routes from provider A with DCU class 1
  - associate routes from provider B with DCU class 2
- Perform accounting on PE



BGP Policy Accounting & DCU

- Easy to configure on the routers
- Results available via SNMP
  - MIBs:
    - CISCO-BGP-POLICY-ACCOUNTING-MIB
    - JUNIPER-DCU-MIB
  - See Simon Leinen's page on this subject [6]:

    http://www.switch.ch/misc/leinen/snmp/monitoring/bucket-accounting.html

MPLS Based Methods

MPLS Based Methods

- Two methods to determine traffic matrices:
  - Using RSVP-TE tunnels
  - Using LDP statistics

- Some comments on Deutsche Telekom's practical implementation

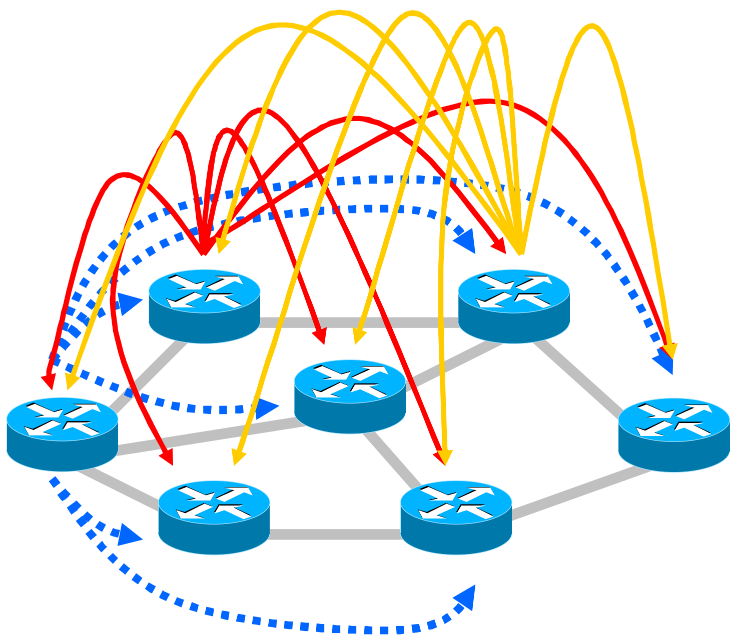
RSVP-TE in MPLS Networks

- RSVP-TE (RFC 3209) can be used to establish LSPs
- Example (IOS):

```
interface Tunnel99
 description RouterA => RouterB
 tag-switching ip
 tunnel destination 3.3.3.3
 tunnel mode mpls traffic-eng
 tunnel mpls traffic-eng priority 5 5
 tunnel mpls traffic-eng bandwidth  1
 tunnel mpls traffic-eng path-option 3 explicit identifier 17
 tunnel mpls traffic-eng path-option 5 dynamic
!
ip explicit-path identifier 17 enable
 next-address 1.1.1.1
 next-address 2.2.2.2
 next-address 3.3.3.3
!
```

RSVP-TE in MPLS Networks

- Explicitly routed Label Switched Paths (TE-LSP) have associated byte counters
- A full mesh of TE-LSPs enables to measure the traffic matrix in MPLS networks directly



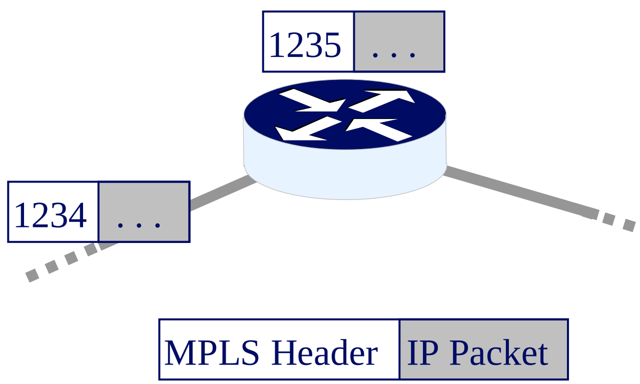
RSVP-TE in MPLS Networks

Pro’s and Con’s

- Advantages:
  - Method that comes closest a traffic matrix measurement
  - Easy to collect data
- Disadvantages:
  - A full mesh of TE-LSPs introduces an additional routing layer with significant operational costs;
  - Emulating ECMP load sharing with TE-LSPs is difficult and complex:
    - Define load-sharing LSPs explicitly;
    - End-to-end vs. local load-sharing;
  - Only provides Internal Traffic Matrix, no Router/PoP to peer traffic

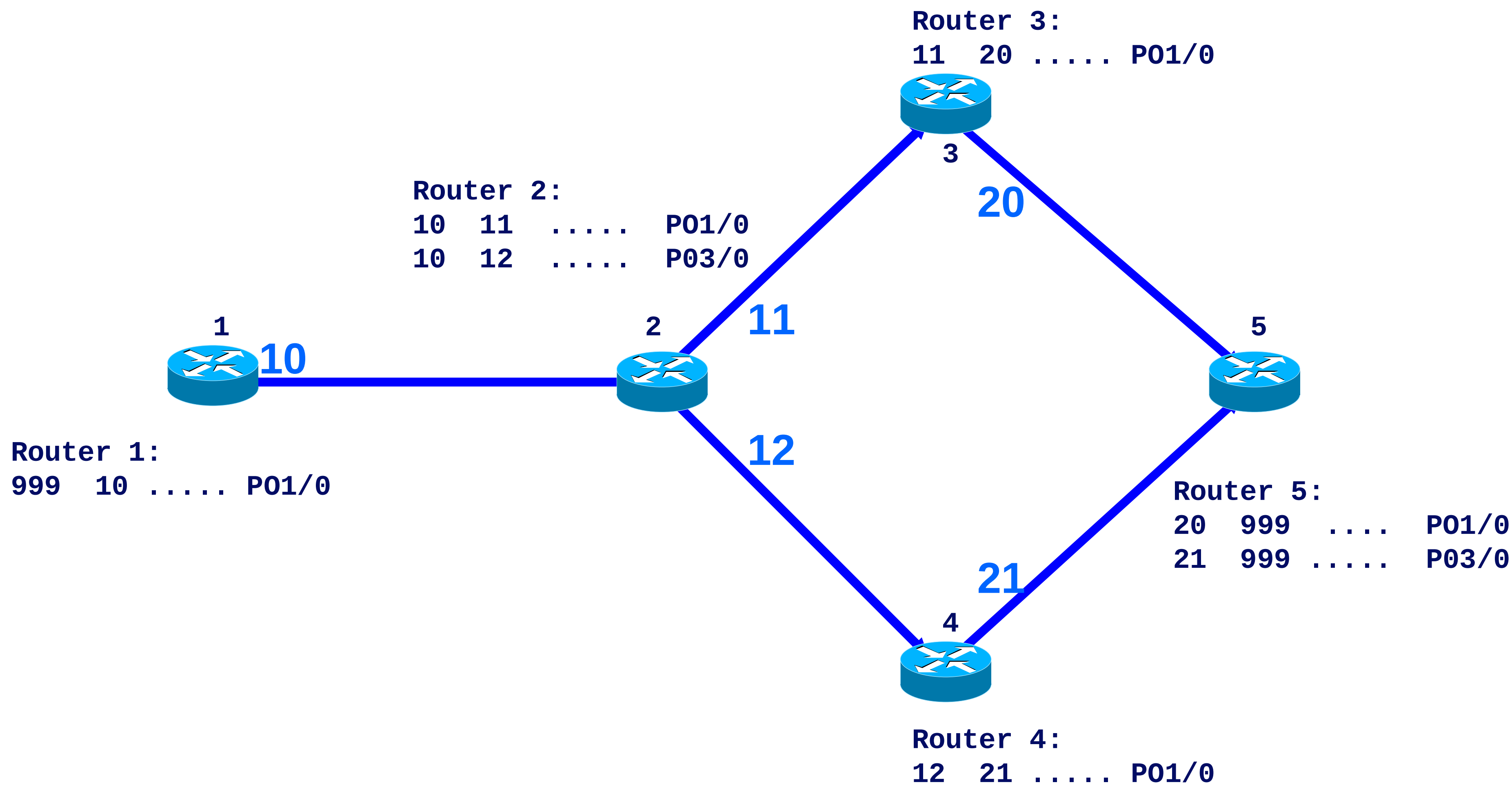
Traffic matrices with LDP statistics

- In a MPLS network, LDP can be used to distribute label information
- Label-switching can be used without changing the routing scheme (e.g. IGP metrics)
- Many router operating systems provide statistical data about bytes switched in each *forwarding equivalence class* (FEC):



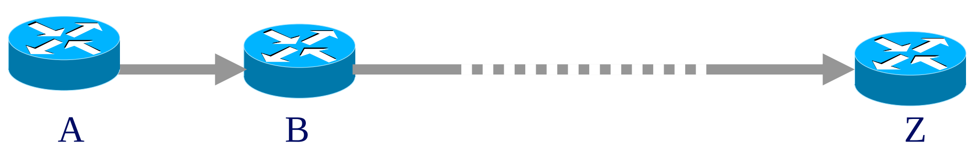
InLabel	OutLabel	Bytes	FEC	OutInt
1234	1235	**4124**	10.10.10.1/32	PO1/2
…	…	…		…

Traffic matrices with LDP statistics
Use of ECMP load-sharing



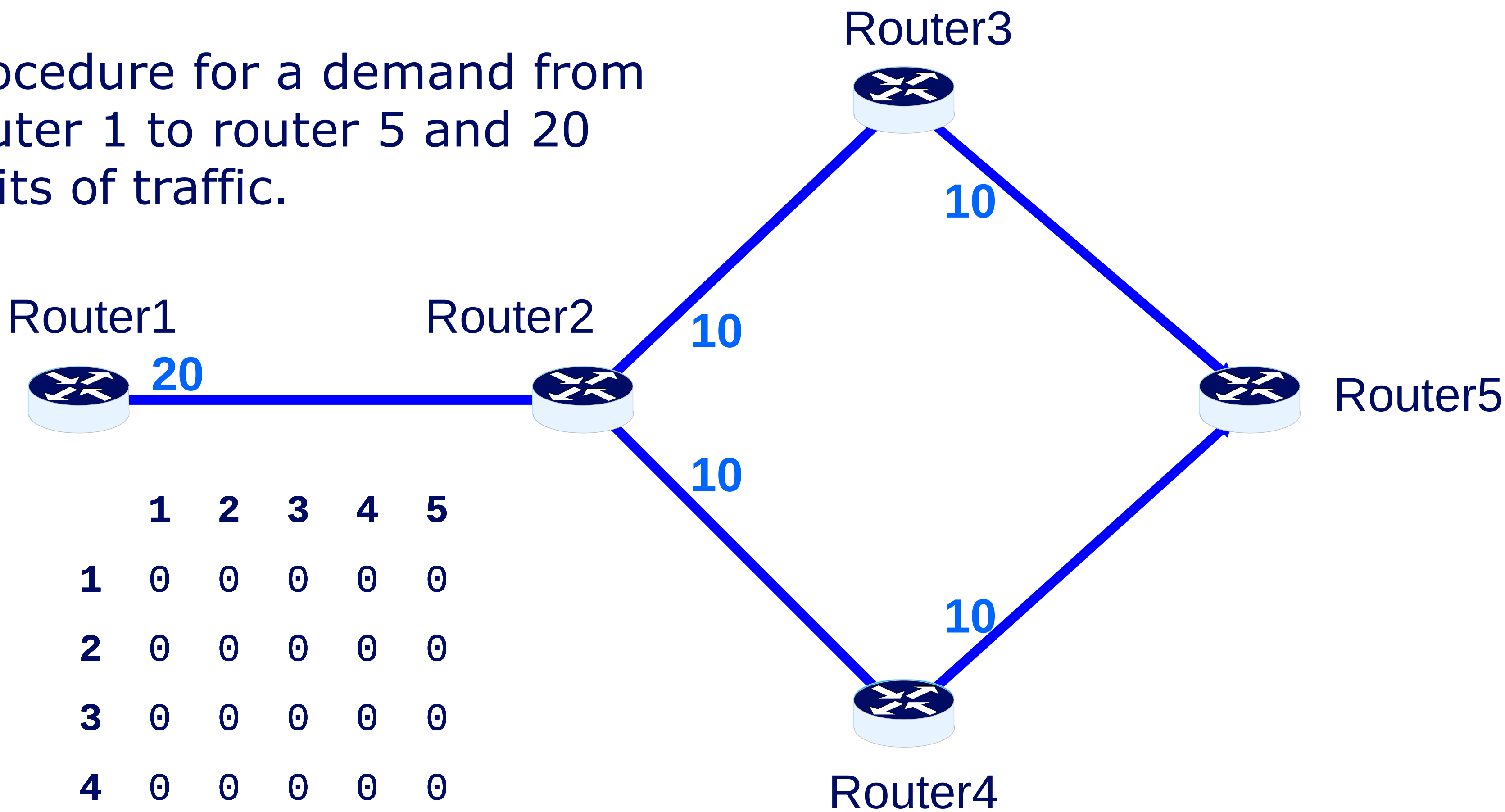
Traffic matrices with LDP statistics

- The given information allows for a forward chaining
- For each router and FEC a set of residual paths can be calculated (given the topology and LDP information)
- From the LDP statistics we gather the bytes switched on each residual path
- Problem: It is difficult to decide whether the router under consideration is the beginning or transit for a certain FEC
- Idea: For the traffic matrix *TM*, add the paths traffic to *TM(A,Z)* and subtract from *TM(B,Z)*. [4]



Traffic matrices with LDP statistics
Example

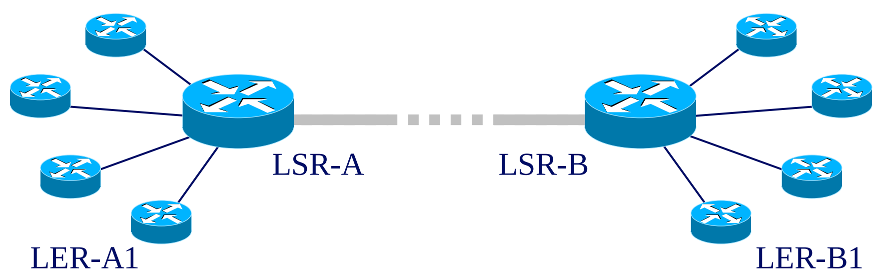
Procedure for a demand from router 1 to router 5 and 20 units of traffic.



	1	2	3	4	5
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	20	-20	10	10	0

Practical Implementation
Cisco IOS

- LDP statistical data available through "show mpls forwarding" command
- Problem: Statistic contains no ingress traffic (only transit)
- If separate routers exist for LER- and LSR-functionality, a traffic matrix on the LSR level can be calculated
- A scaling process can be established to compensate a moderate number of combined LERs/LSRs.



Practical Implementation
Cisco IOS

LDP statistics on IOS:

```
router# show mpls forwarding-table
Local   Outgoing    Prefix            Bytes tag  Outgoing   Next Hop
tag     tag or VC   or Tunnel Id      switched   interface
26      26          62.225.24.184/30  0          PO10/0     point2point
        46          62.225.24.184/30  0          PO13/1     point2point
27      Pop tag     62.225.17.203/32  56529738   PO2/0      point2point
[..]
```

Martin Horneffer, NANOG33

Practical Implementation
Juniper JUNOS

- LDP statistical data available through “show ldp traffic-statistics” command
- Problem: Statistic is given only per FECs and not per outgoing interface
- As a result one cannot observe the branching ratios for a FEC that is split due to load-sharing (ECMP);
- Assume that traffic is split equally
- Especially for backbone networks with highly aggregated traffic this assumption is met quite accurately

Practical Implementation
Juniper JUNOS

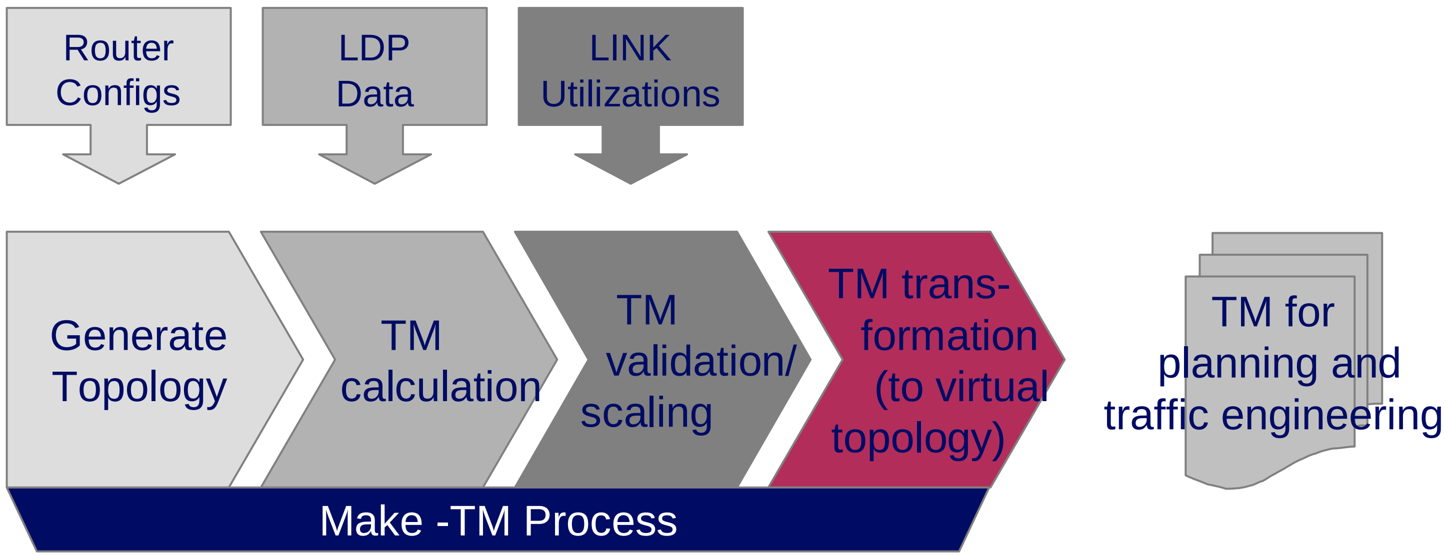
```
user@router> show ldp traffic-statistics
FEC                    Type      Packets        Bytes  Shared
 62.225.16.134/32      Transit   1236933    103984630  No
                       Ingress         0            0  No
[..]
user@router> show route table inet.3
[..]
62.225.16.134/32     *[LDP/9] 06:08:27, metric 1
                      > via so-0/0/0.0, Push 39
[..]
```

Martin Horneffer, NANOG33

Practical Implementation
Results

- The method has been successfully implemented in Deutsche Telekom's global MPLS Backbone
- A continuous calculation of traffic matrices (15min averages) is accomplished in real-time for a network of 180 routers
- The computation requires only one commodity PC
- No performance degradation through LDP queries
- Calculated traffic matrices are used in traffic engineering and network planning

Practical Implementation
Deployment Process



Conclusions for LDP method

- This method can be implemented in a multi-vendor network
- It does not require the definition of explicitly routed LSPs
- It allows for a continuous calculation
- There are some restrictions concerning
  - vendor equipment
  - network topology
- See Ref. [4]

Estimation Techniques

What do we have?

- NetFlow
  - v5 deployment is complex
  - newer versions (aggregation, v8) not complete
- DCU (Juniper)
  - only 16/126 classes
- BGP Policy Accounting
  - only 64 classes, BGP only
- MPLS
  - TE tunnels
    - requires full TE mesh
  - LDP counters
    - nice solution, only minor issues (see [4])

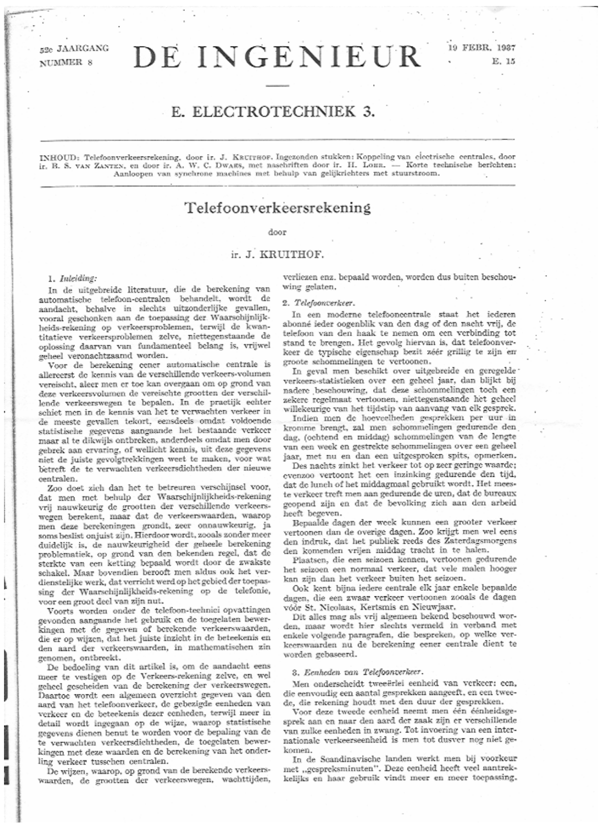
What do we want?

- Derive Traffic Matrix (TM) from easy to measure variables
  - No complex features to enable
- Link Utilization measurements
  - SNMP
  - easy to collect, e.g. MRTG
- Problem:

  Estimate point-to-point demands from measured link loads
- Network Tomography
  - Y. Vardi, 1996
  - Similar to: Seismology, MRI scan, etc.

Is this new?

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



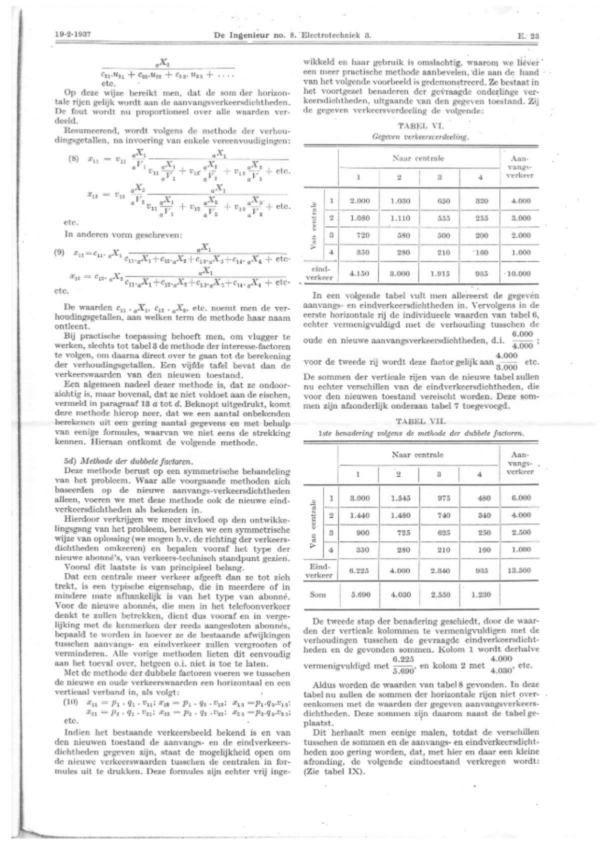
52e JAARGANG NUMMER 8 DE INGENIEUR 19 FEBR. 1937 E. 15

E. ELECTROTECHNIEK 3.

Telefoonverkeersrekening

door

ir. J. KRUITHOF.



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

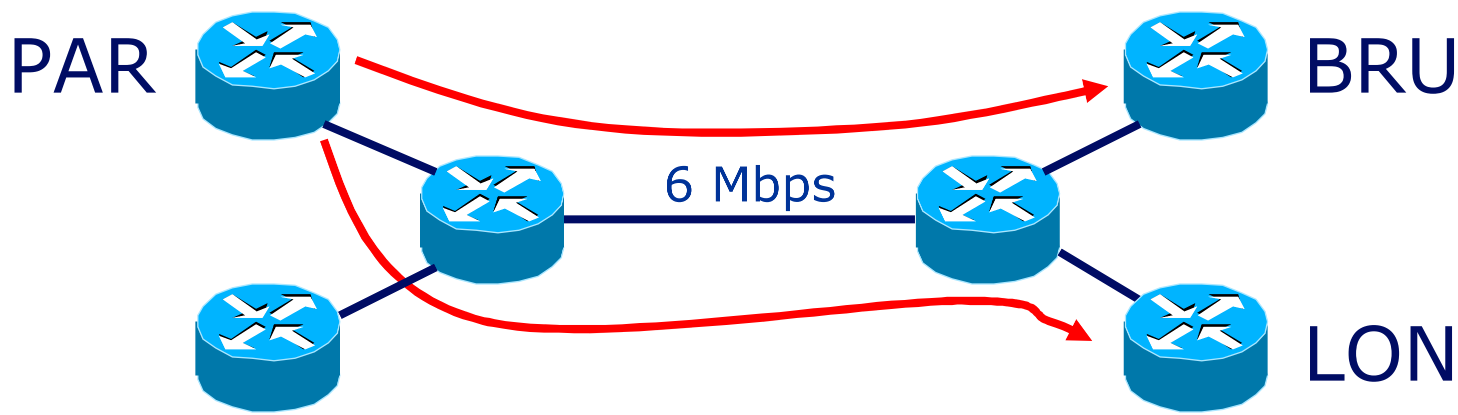
TABEL VI.

TABEL VII.

Demand Estimation

- 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)*
  - *Maximum Likelihood Estimation*
  - *Entropy maximization*
  - *Bayesian statistics (model prior knowledge)*
  - *Etc...!*
- Calculate the **most likely** Traffic Matrix

Example

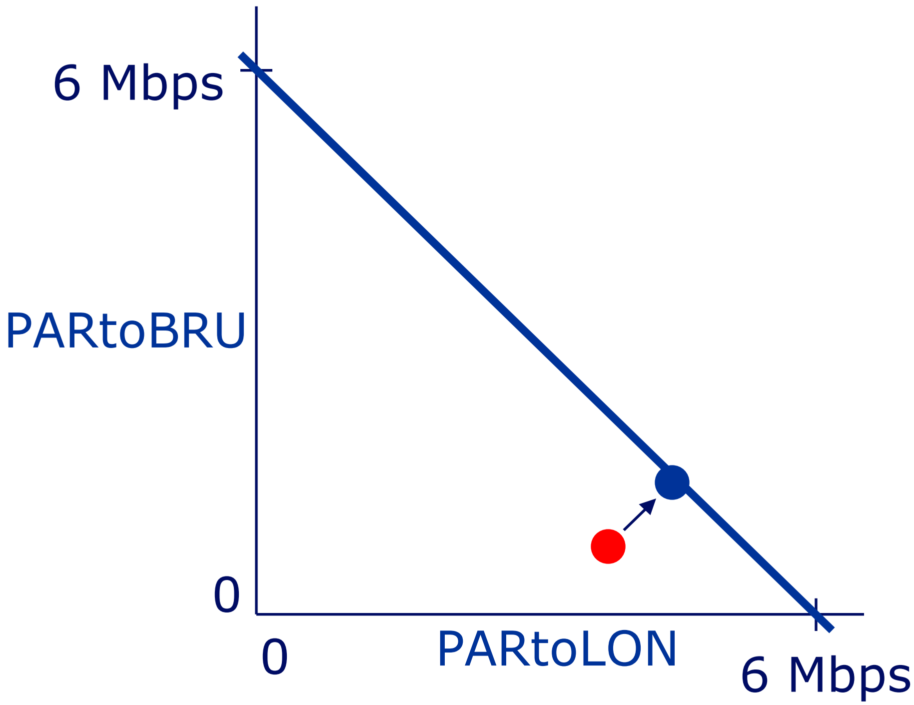


y: link utilizations
A: routing matrix
x: point-to-point demands

Solve: <u>y = Ax</u>
In this example: <u>6 = PARtoBRU + PARtoLON</u>

Example

Solve: y = Ax -> 6 = PARtoBRU + PARtoLON



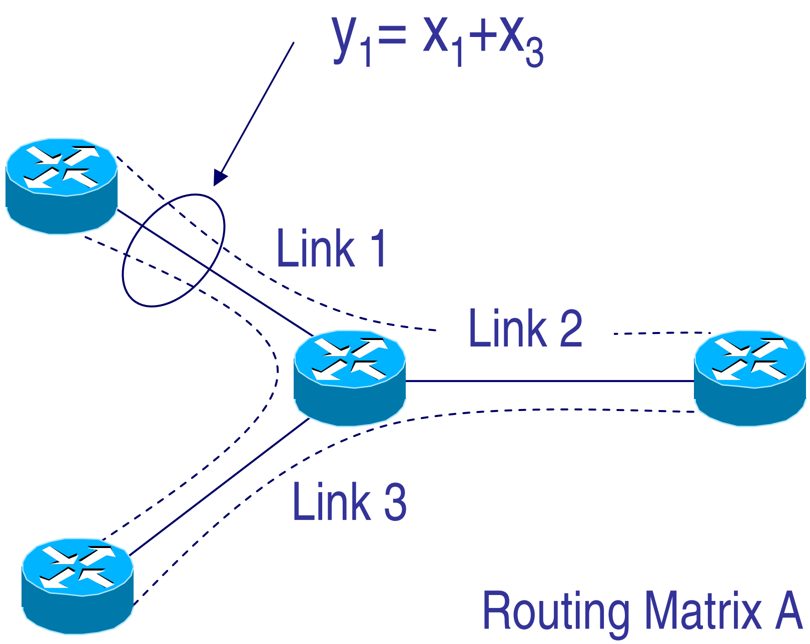
Additional information

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

Example: Total traffic sourced at PAR is *50Mbps.* BRU sinks *2%* of total traffic, LON sinks *8%:*
PARtoBRU =1 Mbps and
PARtoLON =4 Mbps

Final Estimate: PARtoBRU = 1.5 Mbps and PARtoLON = 4.5 Mbps

General Formulation



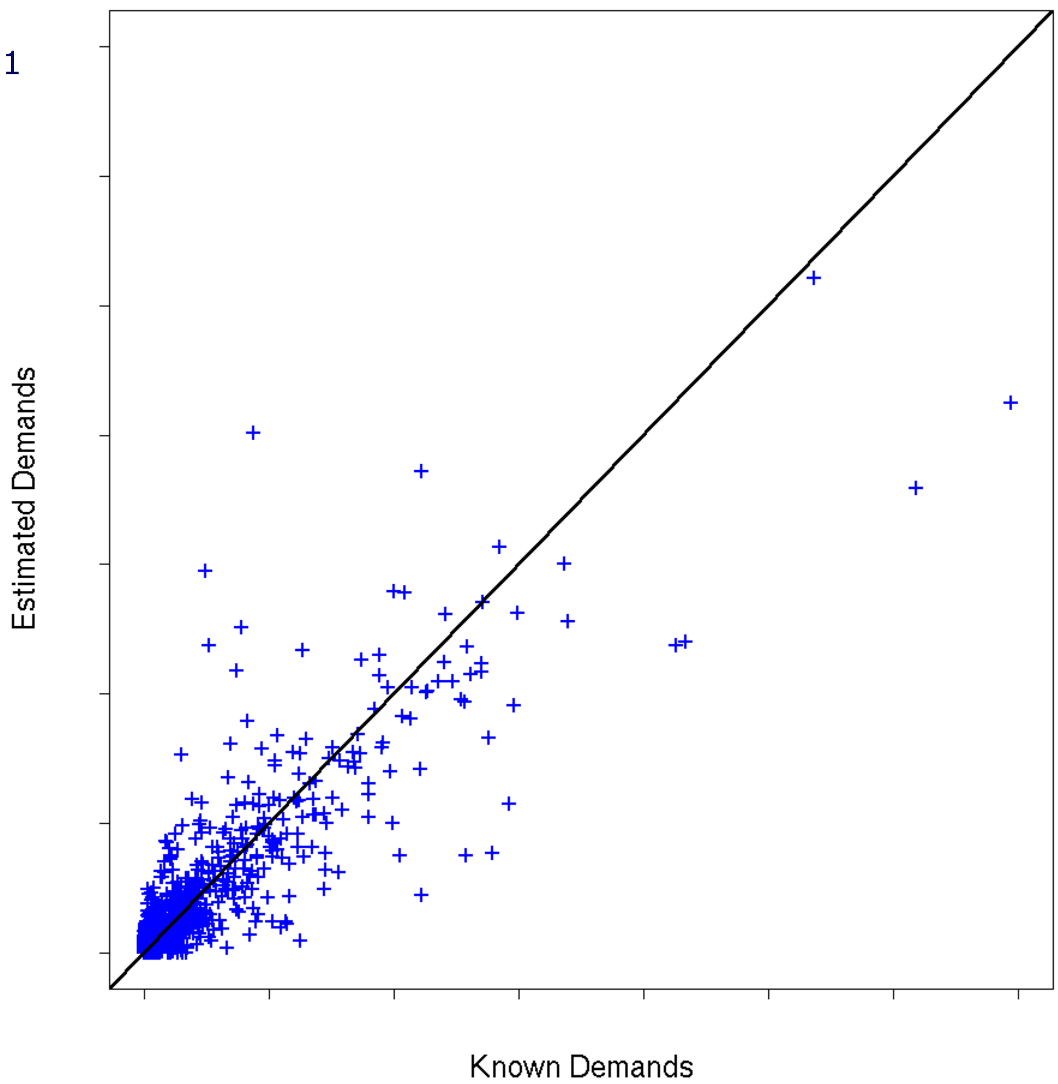
- The total traffic on each link is the sum of all the source destination flows that route over that link
- Given Y and the routing matrix A solve for X

Routing Matrix A

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

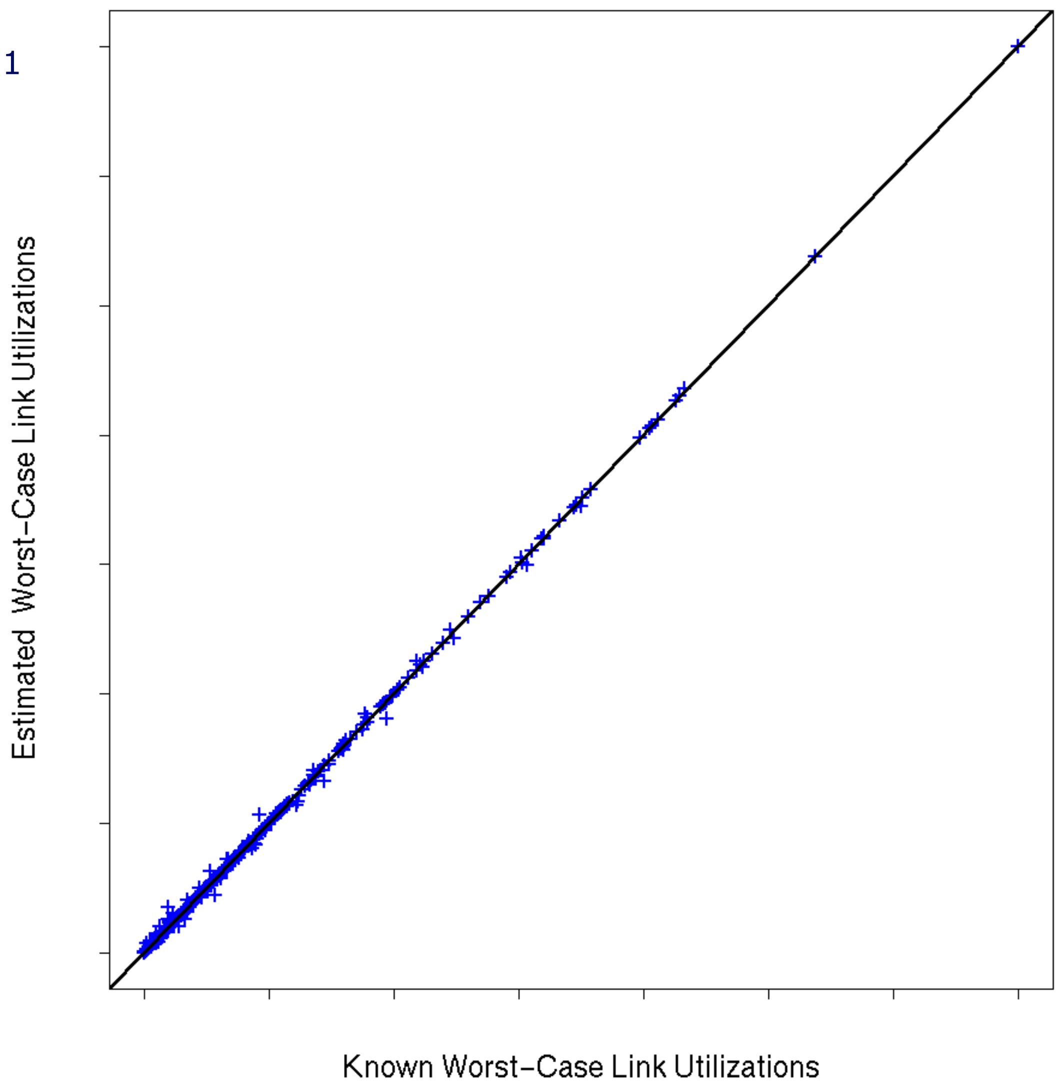
Network Results: Estimated Demands

International Tier-1
IP Backbone



Estimated Link Utilizations!

International Tier-1
IP Backbone



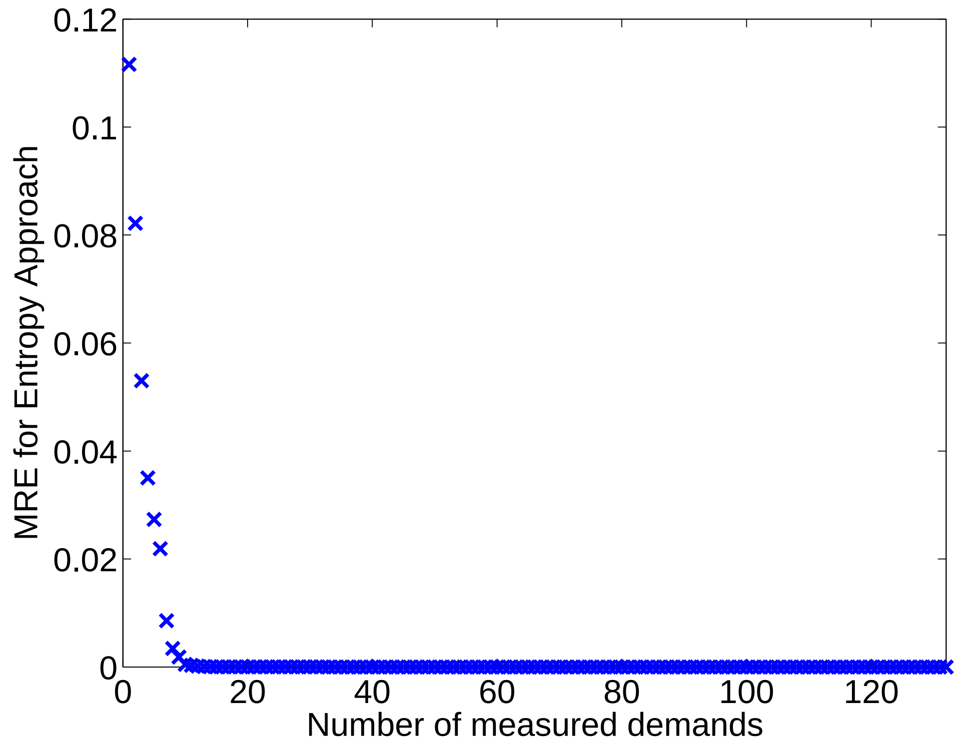
Demand Estimation Results

- Individual demands
  - Inaccurate estimates…
- Estimated worst-case link utilizations
  - Accurate!
- Explanation
  - Multiple demands on the same path indistinguishable, but their sum is known
  - If these demands fail-over to the same alternative path, the resulting link utilizations will be correct

Estimation with Measurements

- Estimation techniques can be used in combination with demand meadsurements
  - E.g. NetFlow or partial MPLS mesh

- This example: Greedy search to find demands which decreases MRE (Mean Relative Error) most.
  - A small number of measured demands account for a large drop in MRE



Data from [1]

Traffic Matrix Estimation
Case-Study

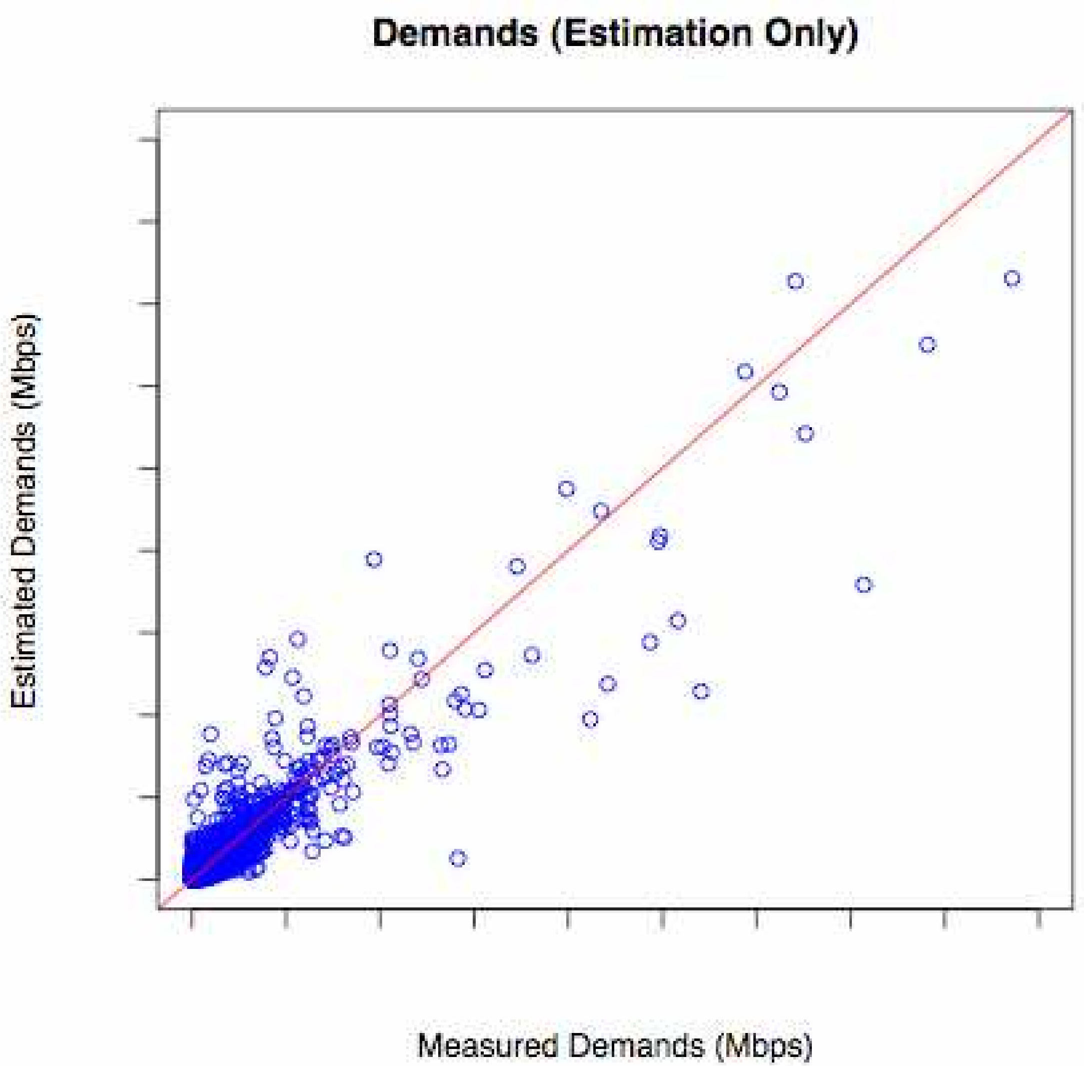
TM Estimation Deployment Case

- Large ISP network
  - about 80 Routers and 200 Circuits
  - 2550 TM entries
    - not all routers source/sink traffic (e.g. core)
- Known Traffic Matrix
  - Direct MPLS measurement
- Case-study will evaluate:
  - How does estimated TM compare to known TM?
  - How well does the estimated TM predict worst-case link utilizations?
  - How well do tools that require a TM work when given the estimated TM?
  - How much can the estimated TM be improved by adding point-to-point measurements?

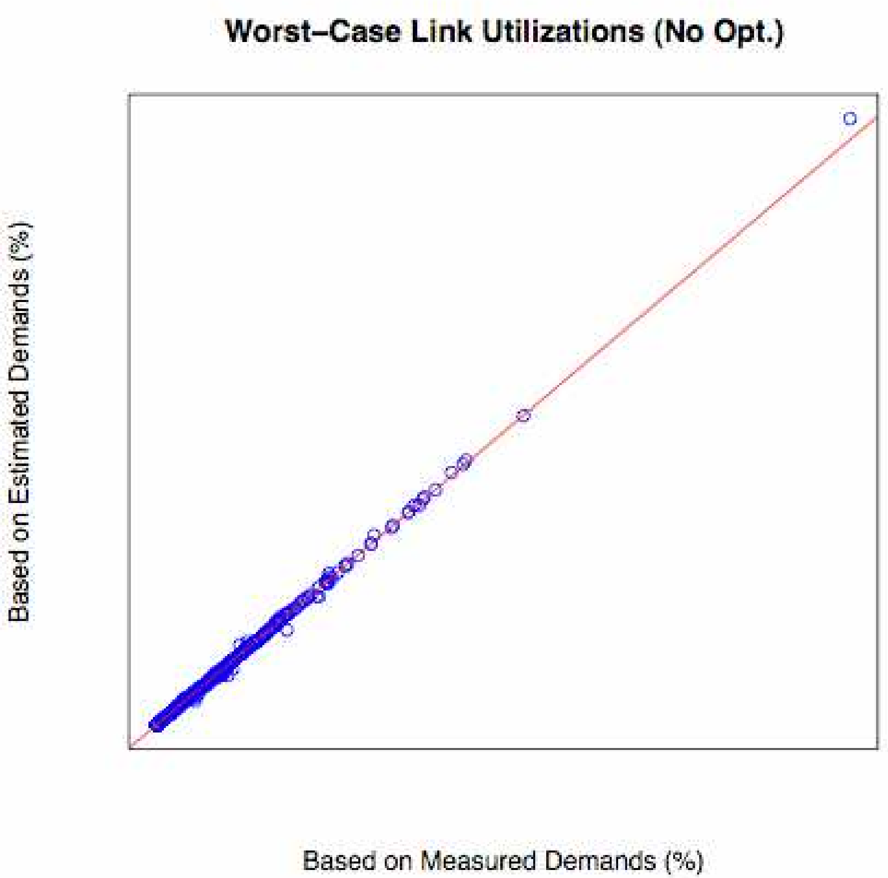
Procedure

- TM estimation using Cariden MATE Software
  - Demand Deduction tool
- Start with current network and known TM
  - save as "PlanA" (with TM "Known")
- IGP Simulation for non-failure
- Save Link Utilizations and Node In/Out traffic
- Estimate Traffic Matrix
  - New TM: "Estimated", Save as "PlanB"
- Do an IGP *Metric Optimization* on both networks
  - Using known TM in planA, estimated TM in PlanB
- Simulate IGP routing on both optimized networks
  - using *known* Traffic matrix for both
- Compare Results!

Estimated Demands

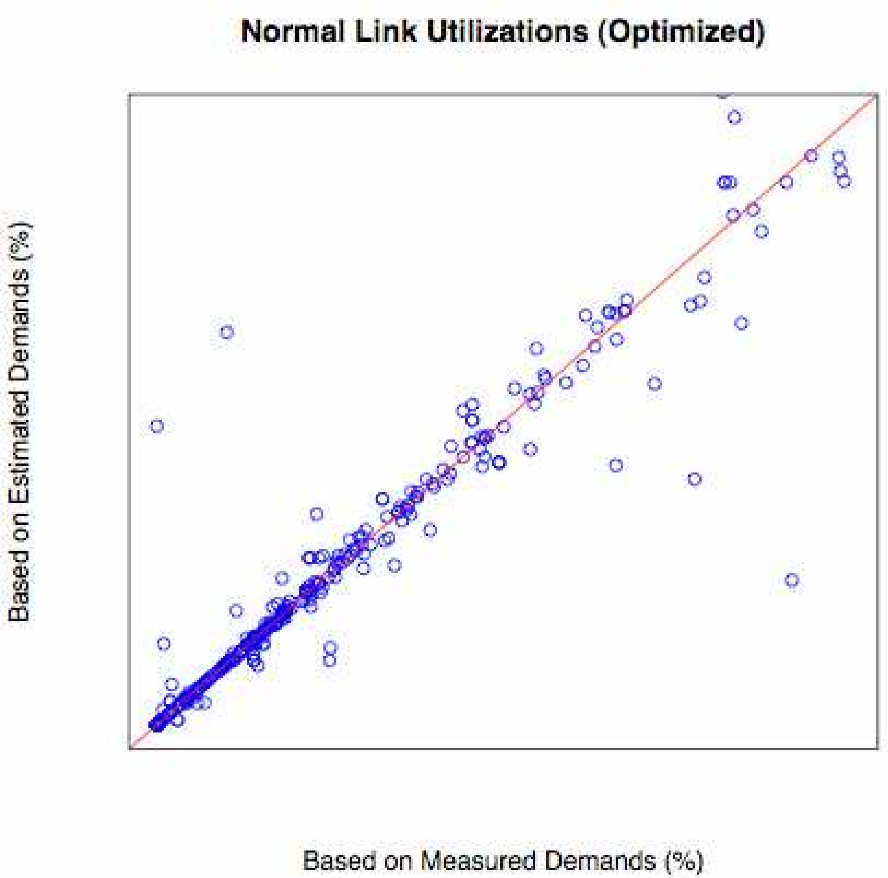


Worst-Case Link Util. (No. Opt)



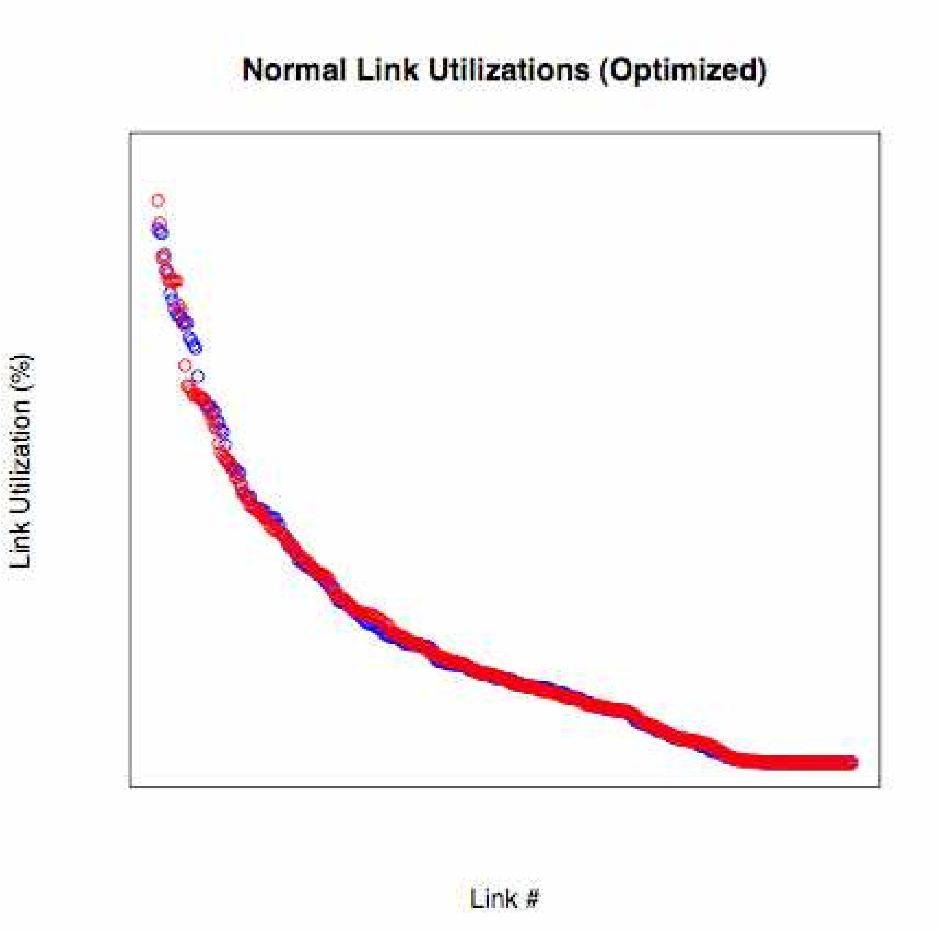
- No Metric Optimization
- PlanA Traffic Matrix:
  - Known
- PlanB Traffic Matrix:
  - Estimated
- IGP Simulation
  - Circuit + SRLG failures
- Compare Worst-Case Link Utilizations (in %)

Normal Link Utilizations (Opt.)



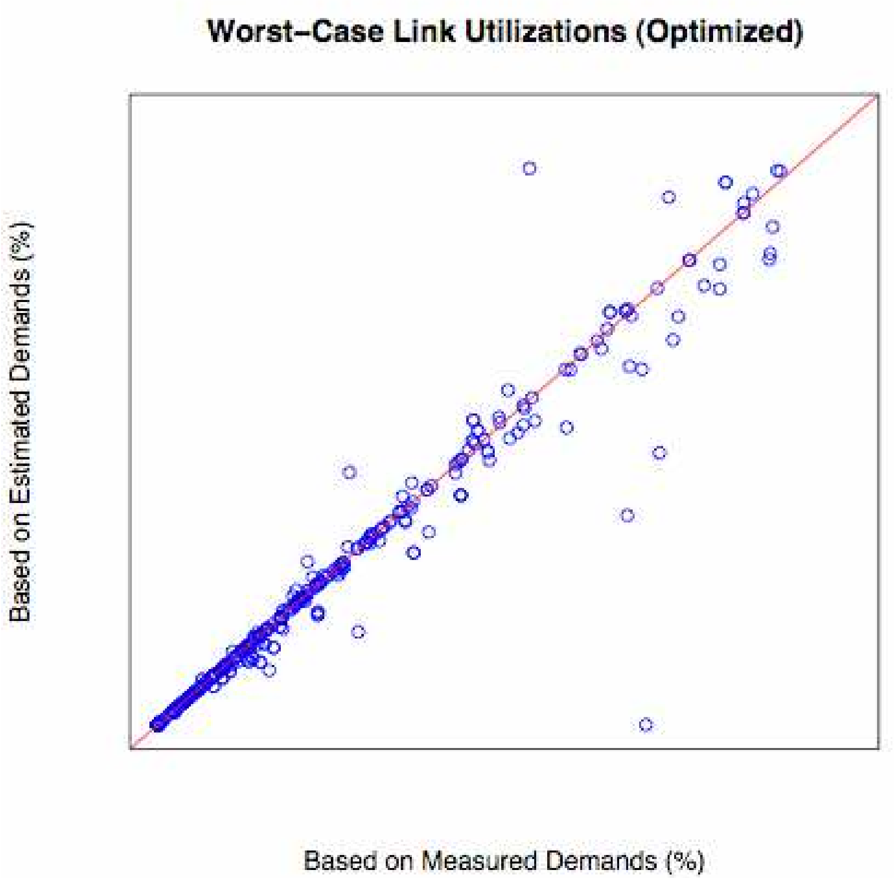
- IGP Metric Optimization
  - PlanA Traffic Matrix:
    - Known
  - PlanB bandwidth level:
    - Estimated
- IGP Simulation
  - PlanA Traffic Matrix:
    - Known
  - PlanB bandwidth level:
    - Original
- Compare Base Link Utilizations (in %)
  - non-failure

Normal Link Utilizations (Opt.)



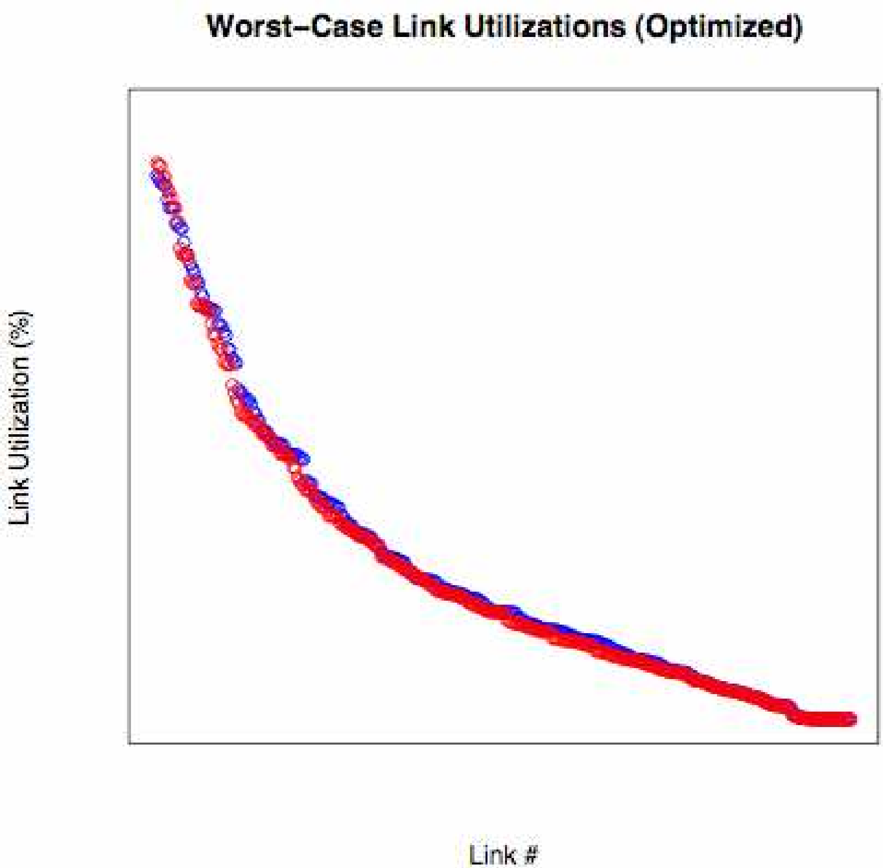
- Scenario: same as previous slide
- Compare *Sorted* Link Utilizations
  - non-failure
- Colors:
  - based on measured demands: BLUE
  - based on estimated demands: RED

Worst-Case Link Utilizations (Opt)



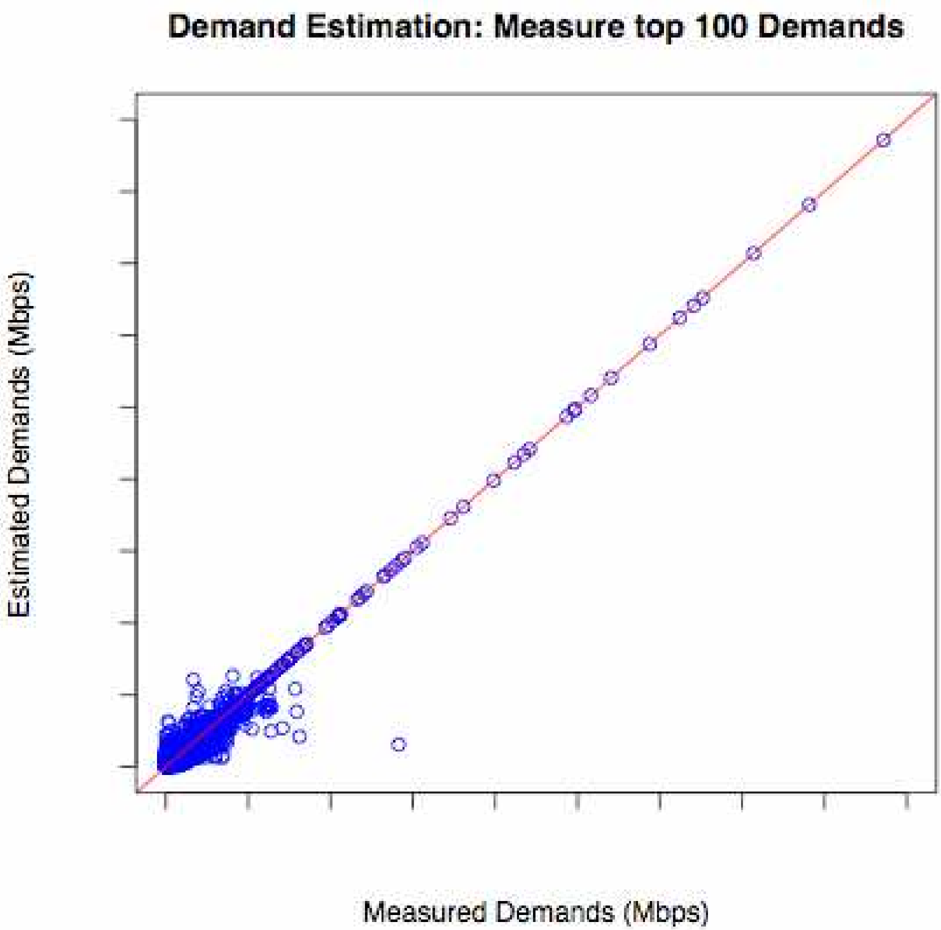
- Scenario: same
- Compare Worst-Case Link Utilizations (in %)
  - Circuits + SRLG failures

Worst-Case Link Utilizations (Opt)



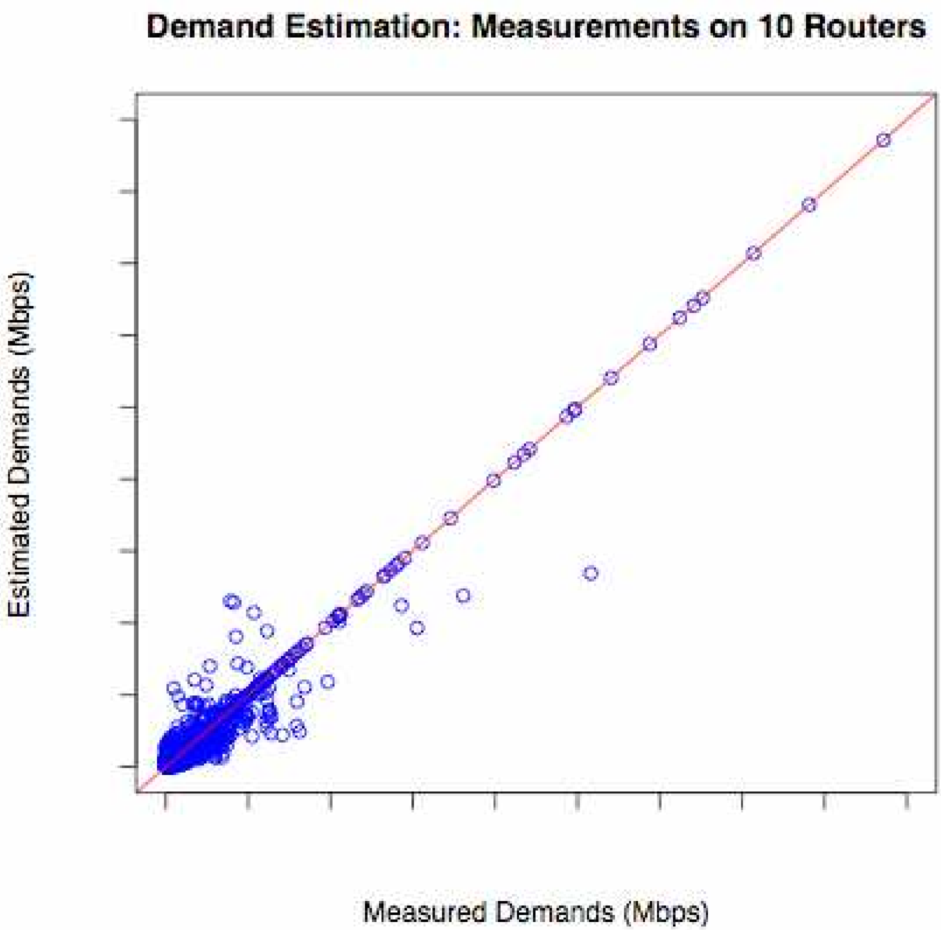
- Scenario: same
- Compare *Sorted* Worst-Case Link Utilizations (in %)
  - Circuits + SRLG failures
- Colors:
  - based on measured demands: BLUE
  - based on estimated demands: RED

Add Measurements (1)



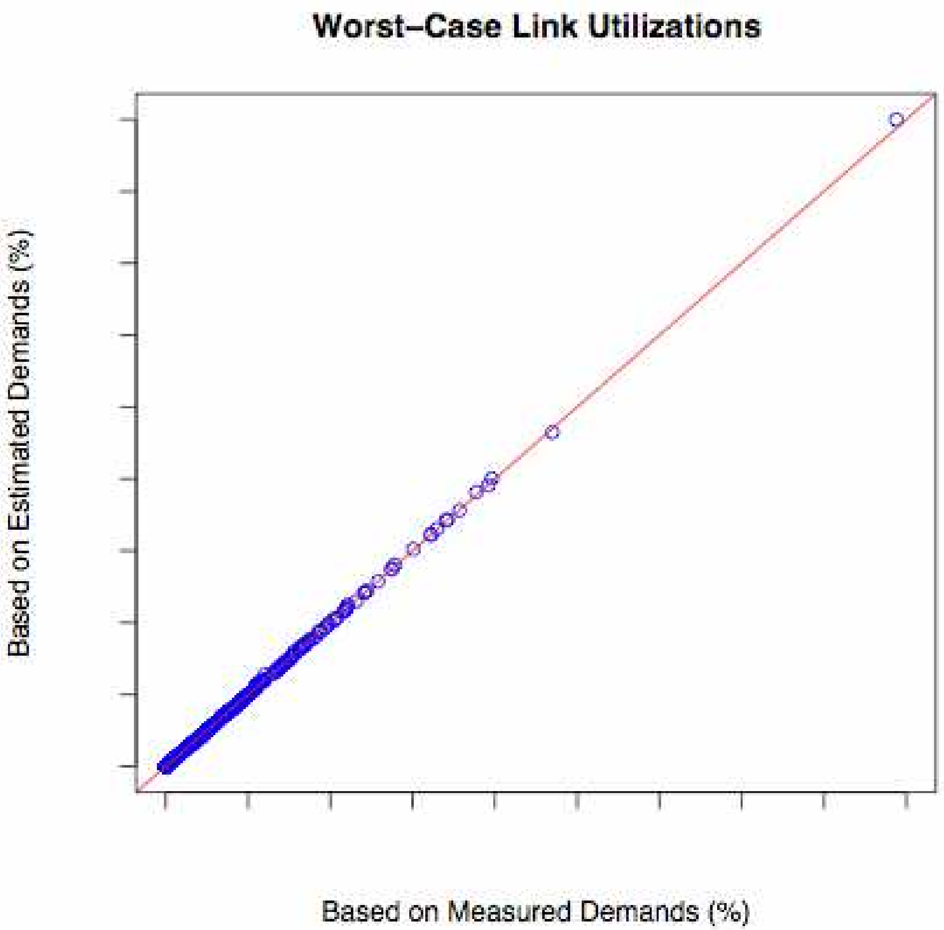
- Select the top 100 demands from the estimated traffic matrix
- Setup Juniper DCU (or equivalent) to measure these demands
  - on 23 routers
  - 38% of total traffic
- Add measured data to the TM estimation process
- (FYI: 990 demands represent 90% of total traffic)

Add Measurements (2)



- Select the top 10 routers from the measured in/out traffic
- Select the top 16 demands on each of these routers (estimated)
- Setup Juniper DCU to measure these demands
  - 160 demands
  - 10 routers
  - 41% of total traffic
- Add measured data to the TM estimation process

Worst-Case Link Utilization (2)



- Revisit the worst-case link utilizations, now with more accurate TM
- Similar results as before
- Hardly any room for improvement!

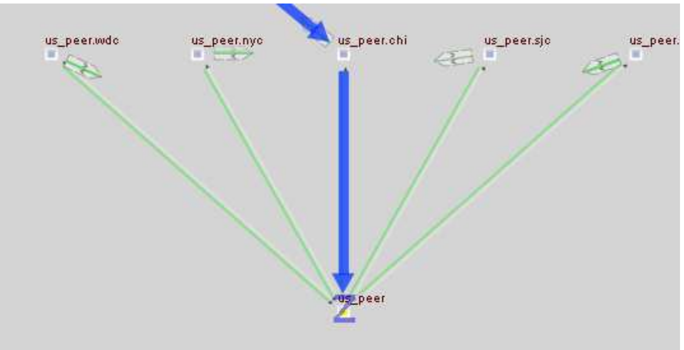
TM Estimation Case-Study

- Works very well on this ISP topology/traffic!
  - Also on AT&T, and all other networks we tried
- Even more accurate if used in combination with demand measurements
  - E.g. from NetFlow, DCU or MPLS

External/Inter-AS TM

- Traffic Matrix on a network will change due to core failures (closest-exit), or peering link failures

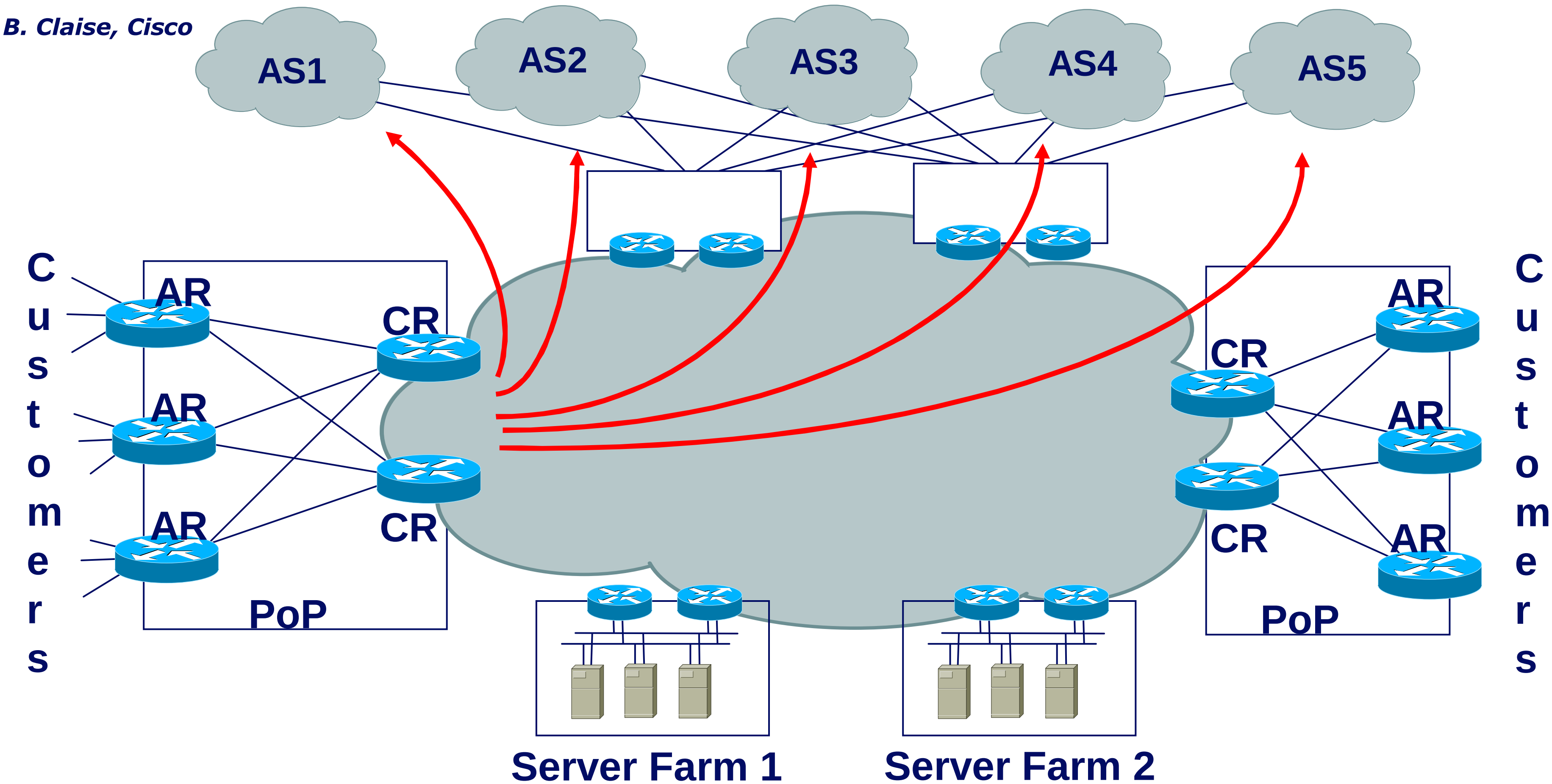
- Create *router-to-peer* TM
- Estimation procedure is similar
- Routing is different
  - policy restrictions:
    e.g. no traffic from peer to peer
  - Virtual model of remote AS



- Estimation can make use of a known core TM

External Traffic Matrix

B. Claise, Cisco



From "Router to BGP AS", the router being the **AR** or **CR**

The external traffic matrix can influence the internal one

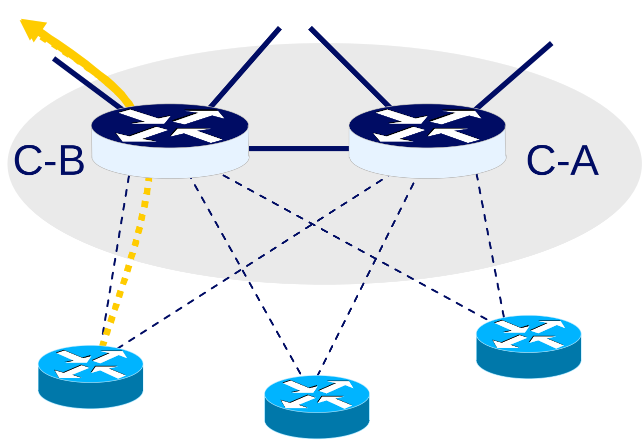
TM Estimation Summary

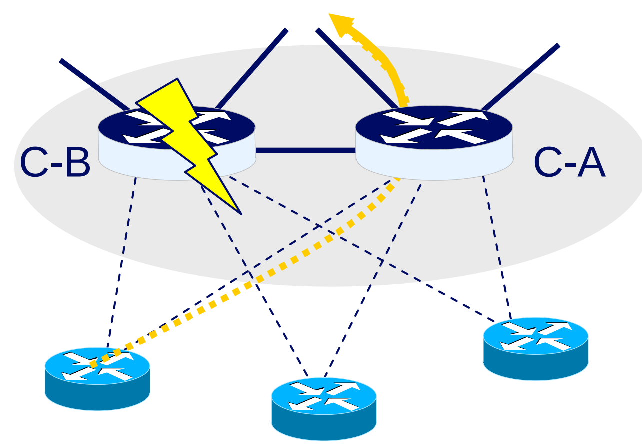
- Algorithms have been published
  - Implement yourself (e.g. IPF procedure)
  - Commercial tools are available
- Can be used in multiple scenarios:
  - Fully estimate Traffic Matrix
  - Combine with NetFlow/DCU/etc.
    - Measure large demands, estimate small ones
  - Estimate unknown demands in a network with partial MPLS mesh (LDP or RSVP)
  - Estimate Peering traffic when Core Traffic Matrix is known
- Also see AT&T work
  - E.g. Nanog29: *How to Compute Accurate Traffic Matrices for Your Network in Seconds* [2]

Traffic Matrices in Partial Topologies

Traffic Matrices in Partial Topologies

- In larger networks, it is often important to have a TM for a partial topology (not based on every router)
- Example: TM for core network (planning and TE)
- Problem: TM changes in failure simulations
- Demand moves to another router since actual demand starts outside the considered topology (red):



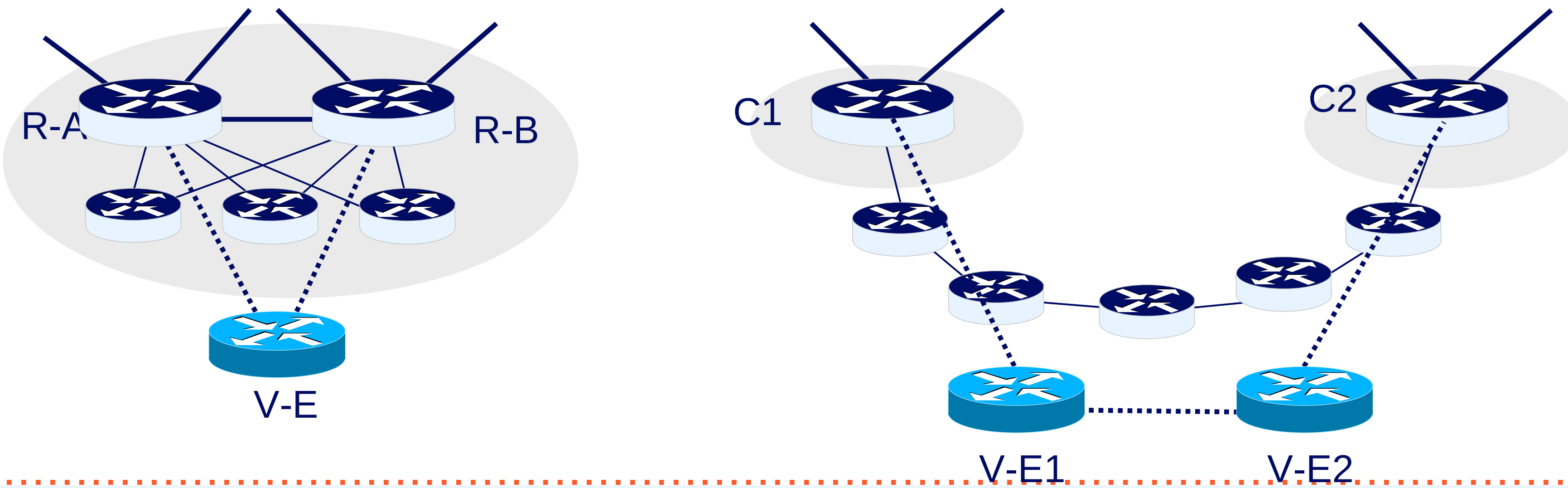


Traffic Matrices in Partial Topologies

- The same problem arises with link failures
- Results in inaccurate failure simulations on the reduced topology
- Metric changes can introduce demand shifts in partial topologies, too.
- But accurate (failure) simulations are essential for planning and traffic engineering tasks

Traffic Matrices in Partial Topologies

- Introduce virtual edge devices as new start-/endpoints for demands
- Map real demands to virtual edge devices
- Model depends on real topology
- Tradeoff between simulation accuracy and problem size.



Multicast

Multicast deployment

- Recent deployment of “IPTV-like” services
- Traffic becomes significant on backbone links
- Using SSM model
  - ‘streams’ are identified by multicast groups
- Unfortunately, no byte counters in the IF-MIB
  - only packets…
- IPMROUTE MIB provides (on every router):
  - Traffic per multicast group
  - Next-hops (outgoing interfaces)
- PIM MIB could be used to build topology
  - Neighbor discovery
- See GRNET Multicast Weathermap [9]

Summary & Conclusions

Overview

- “Traditional” NetFlow (Version 5)
  - Requires a lot of resources for collection and processing
  - Not trivial to convert to Traffic Matrix
- BGP NextHop Aggregation NetFlow provides almost direct measurement of the Traffic Matrix
  - Verion 9 export format
  - Only supported by Cisco in newer IOS versions
- BGP Policy Accounting and Juniper DCU are limited (only 64 or 16/126 classes) to build a full Traffic Matrix
  - But could be used as adjunct to TM Estimation

Overview

- MPLS networks provide easy access to the Traffic Matrix
  - Directly measure in RSVP TE networks
  - Derive from switching counters in LDP network
- Very convenient if you already have an MPLS network, but no reason to deploy MPLS just for the TM
- Estimation techniques can provide reliable Traffic Matrix data
  - Very useful in combination with partially know Traffic Matrix (e.g. NetFlow, DCU or MPLS)

Contact

Thomas Telkamp

Cariden Technologies, Inc.

telkamp@cariden.com

References

1. A. Gunnar, M. Johansson, and T. Telkamp, "Traffic Matrix Estimation on a Large IP Backbone - A Comparison on Real Data", *Internet Measurement Conference 2004.* Taormina, Italy, October 2004.
2. 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.
3. AT&T Tomogravity page:
   http://public.research.att.com/viewProject.cfm?prjID=133/
4. S. Schnitter, T-Systems; M. Horneffer, T-Com. "Traffic Matrices for MPLS Networks with LDP Traffic Statistics." Proc. Networks 2004, VDE-Verlag 2004.
5. Y. Vardi. "Network Tomography: Estimating Source-Destination Traffic Intensities from Link Data." J.of the American Statistical Association, pages 365–377, 1996.
6. "Simon's SNMP Hacks":
   http://www.switch.ch/misc/leinen/snmp/
7. Van Jacobson, Haobo Yu, Bruce Mah, "Route-Flow Fusion: Making traffic measurement useful." NANOG 35, October 2005.
8. T. Telkamp, "Peering Planning Cooperation without Revealing Confidential Information." RIPE 52, Istanbul, Turkey
9. GRNET Multicast Weathermap: http://netmon.grnet.gr/multicast-map.shtml