I2AEX BOF, 83rd IETF @ Paris
Large Bandwidth Use Cases

Greg Bernstein, Grotto Networking

Young Lee, Huawei

Diego Lopez Garcia, Telefonica

Outline

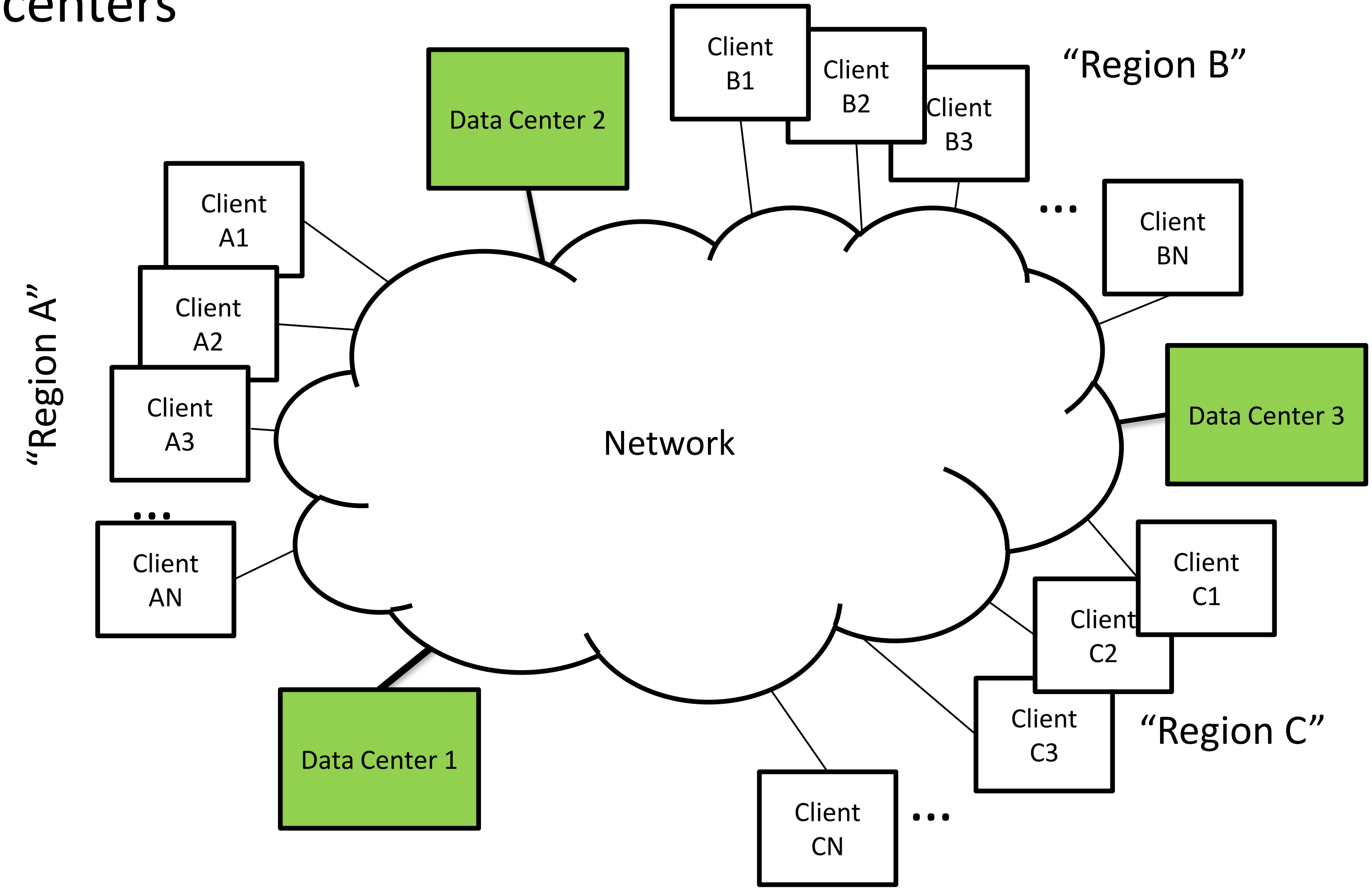
- High Bandwidth Technologies and Techniques
- Use Cases:
  - End System Aggregation (VoD)
  - Express Lanes with Assured Service Quality (ASQ)
  - Data Center to Data Center Communications
    - Application Overlays, Reliability and Recovery
- Infrastructure to Application Exchanges
  - Capacity, Latency, Bottlenecks, and graphs
  - Network to Application Notifications
  - Network Resource Reservations

Technologies & Techniques

- Data Plane Technologies
  - Wavelength Division Multiplexing (WDM) systems typically feature 40, 80, 120, or 160 wavelengths on a fiber, 10Gbps, 40Gbps, or 100Gbps per wavelength.
  - OTN (G.709) light weight TDM multiplexing with FEC and error monitoring
  - SONET/SDH, MPLS, Carrier Ethernet
- Technique: Traffic Engineering via “Express Lanes”
  - AKA “Optical bypass”, “optical grooming” at WDM, OTN, SDH layers. Enhanced by GMPLS control plane.
  - AKA “MPLS tunnels”, “MPLS-TE”. Enhanced by MPLS control plane.
  - Other: Provider Based Ethernet (PBE), Open Flow, etc…

End System Aggregation

- Many clients using services offered at two or more “data centers”



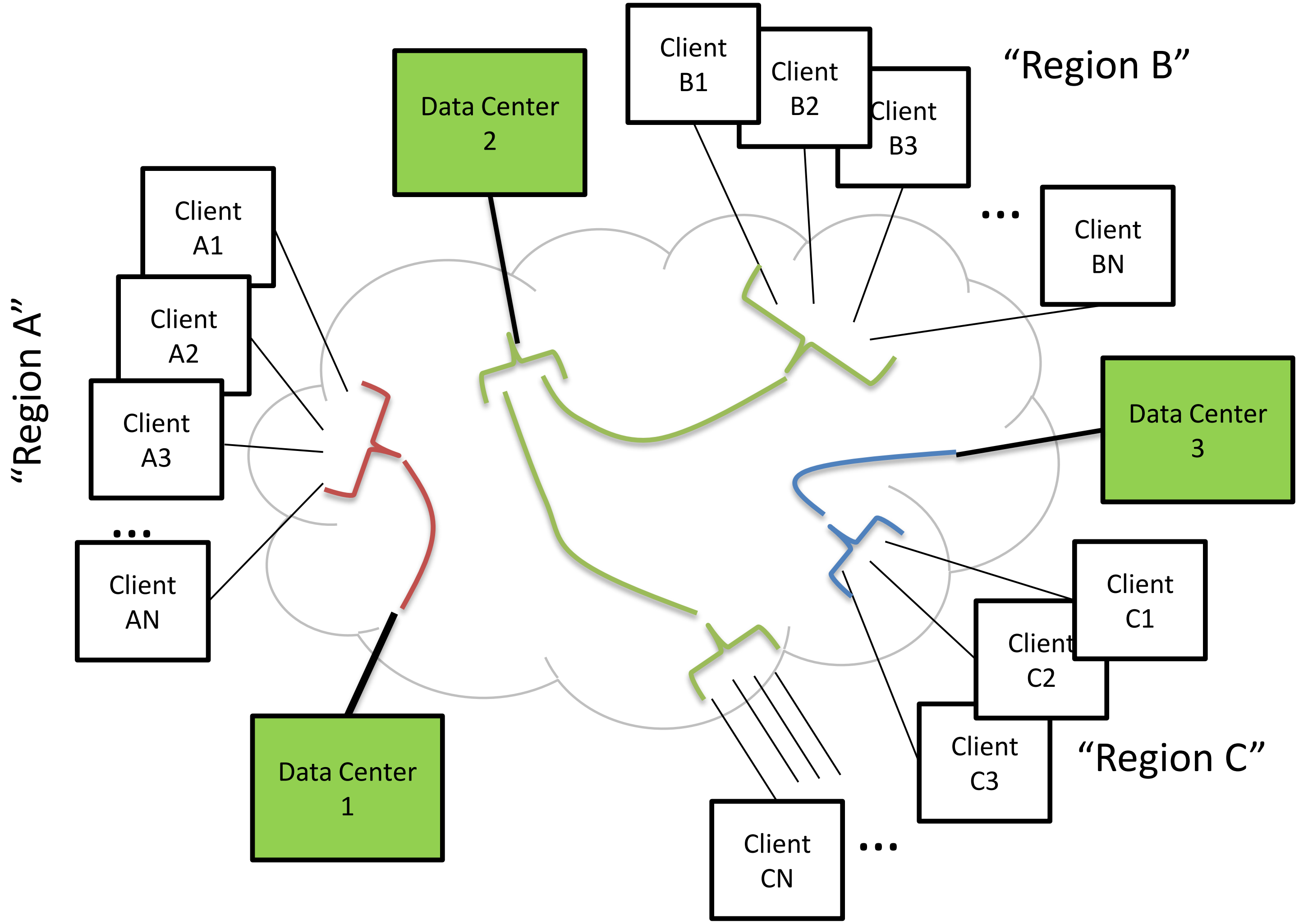
For our purposes here we consider a *data center* any computation facility with ***significant*** access bandwidth to the network (this does not include relatively low bandwidth internet clients)

End System Aggregation Common Example: VoD

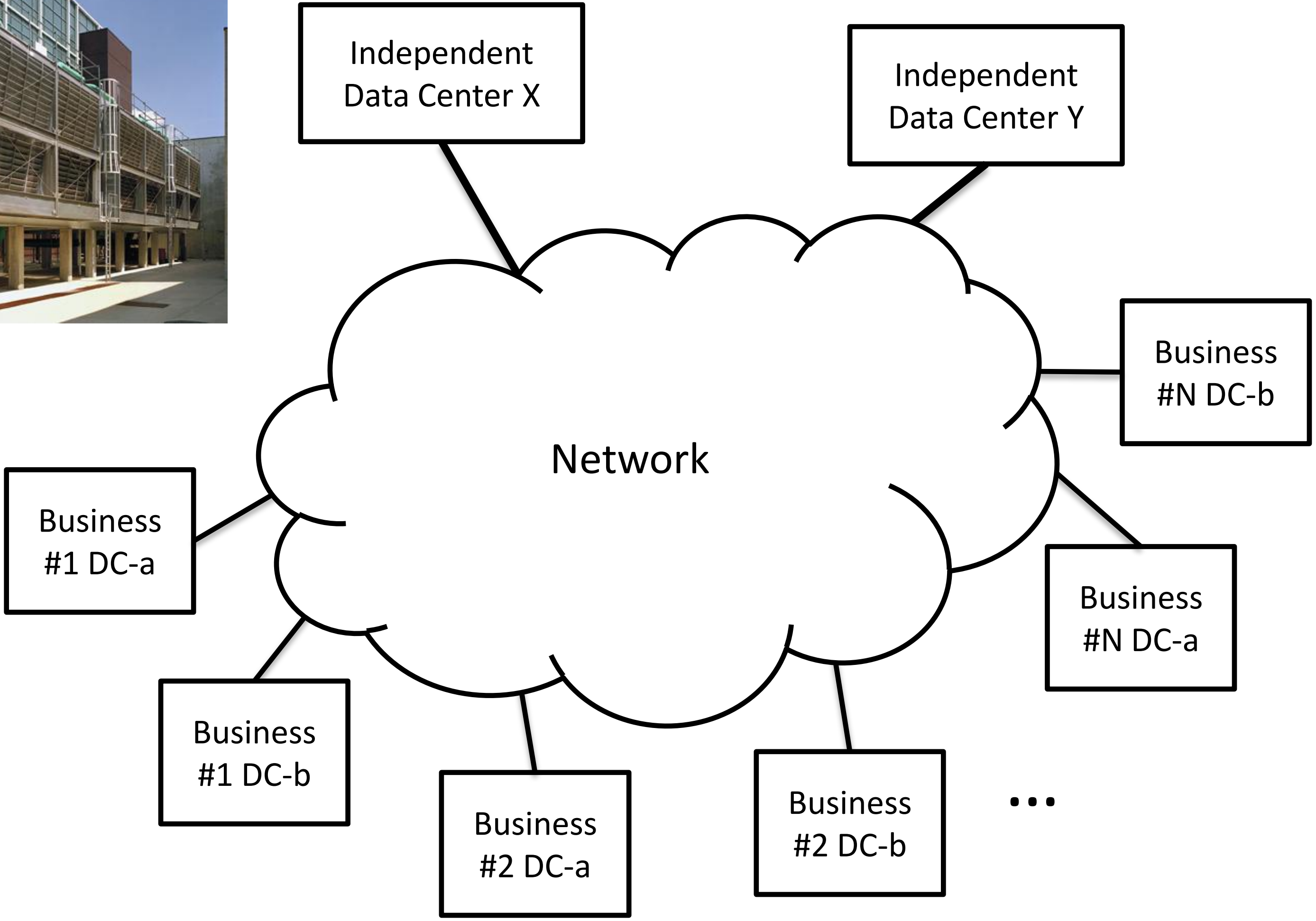
- Clients
  - Millions of customers for a variety of application providers
- Bandwidth
  - Standard definition (STD) quality ~ 1.5mbps, HDTV quality ~ 10mbps per stream client stream
  - Only 6,666 STD or 1,000 High Definition (HDF) streams needed to fill a 10Gbps WDM wavelength (assuming no multicast or peer assist)
- ***Dynamic demand(!)***
  - Time of day, day of week, time of year
  - Scheduled events: new releases, sporting events, concerts, elections, etc…

High Bandwidth Express Lanes

– Traffic engineered "express lanes" between data centers and end user regions



Data Center to Data Center Communications: Application Overlays, Recovery



Reliability and Recovery: Examples

- Application Data Backup
  - Recovery Point Objective (RPO) – how much data may be lost ➔ *Requires periodic bulk data transfer*
  - Recovery Time Objective (RTO) – how quickly can one bring the system back up.
  - Flexibility in scheduling
- Critical Application Resilience
  - SAN replication
  - Geographical database replication
  - Both require dedicated bandwidth
- Live Virtual Machine Migration
  - Data Migration GB to TB to …
  - Machine Size 1-100 GB
- Network Fault Assistance
  - Can application resilience during times of network fault by cooperatively migrating to alternative data centers not affected by network fault.

Line Rate Gbps	Data Transfer per hour
0.1	45 GB
1	450 GB
2.5	1.125 TB
10	4.5 TB
40	18 TB
100	45 TB

Enhanced Information Exchange

- Costs
  - As in existing ALTO model and protocol
  - ***Latency***
  - Others: reliability – usually turned into a cost via probability of failure information (link or path).
- ***Capacity*** (Bandwidth)
  - Could show in a cost map with maximum bandwidth available to each pair (exclusive of other pairs) given
  - However bottleneck links can reduce the usefulness of such a simple representation when it is desired to understand capacity trade offs between multiple source-destination pairs.

Simplified Example from Draft

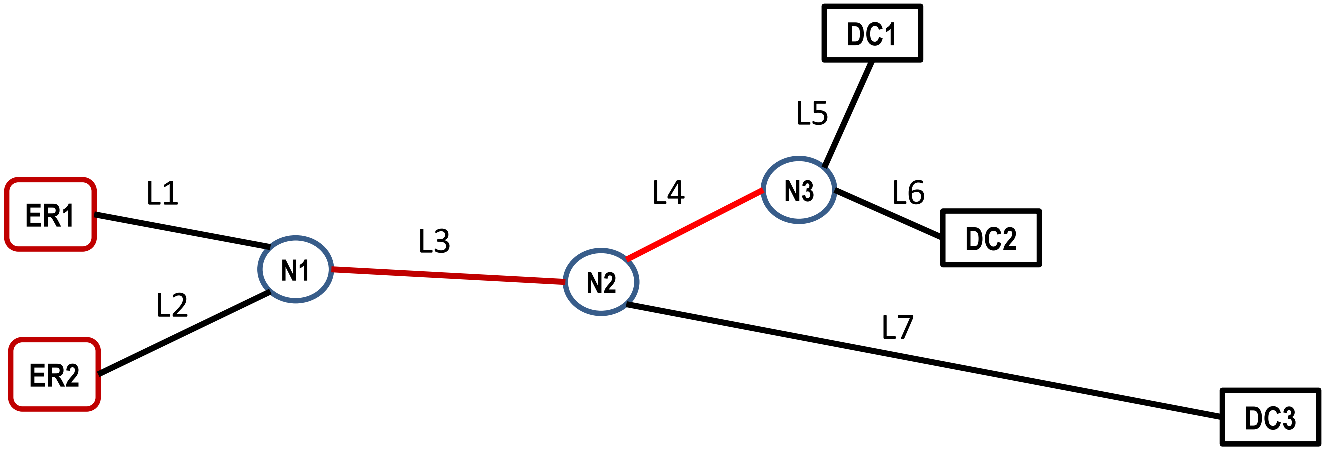


Table 1. ALTO Network Cost

	DC1	DC2	DC3
ER1	5	5	8
ER2	6	6	9

Table 2. Maximum Capacity (as a cost)

	DC1	DC2	DC3
ER1	5	5	5
ER2	5	5	5

Maximum bandwidth exclusive – Doesn't show bottleneck at L3 and L4, this is why we prefer an approximate graph for optimization purposes.

Table 3. Graph Representation

Link	***Capacity***	**Cost**
L1 (ER1, N1)	***5***	1
L2 (ER2, N1)	***5***	2
L3 (N1, N2)	***8***	1
L4 (N2, N3)	***6***	2
L5 (N3, DC1)	***5***	1
L6 (N3, DC2)	***5***	1
L7 (N2, DC3)	***10***	6

Rough Cut Graph Encoding

JSON Object for cost/constraint graph:

```
object {
  LinkEntry [LinkName]<0..*>;
} CostConstraintGraphData;
```

Where a link name is formatted like a PIDName (but names a link), and PID names are used for both provider defined location and provider defined internal model node identification.

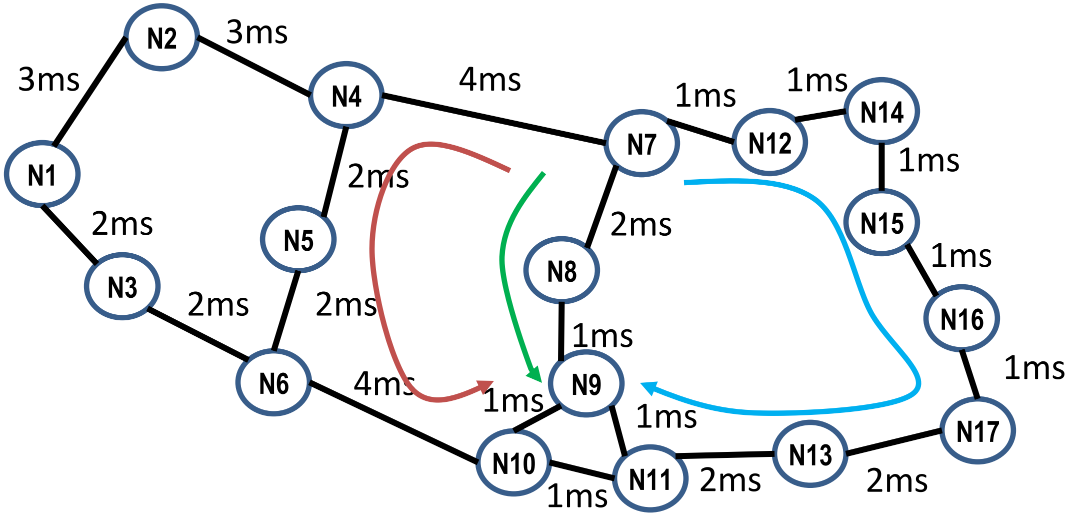
```
object {
  PIDName:    a-end; // Node name at one side of the link
  PIDName:    z-end; // Node name at the other side of the link
  Weight:     wt;
  JSONNumber: latency;
  Capacity:   r-cap; // Reservable capacity
} LinkEntry;
```

Applied to example in high bandwidth draft:

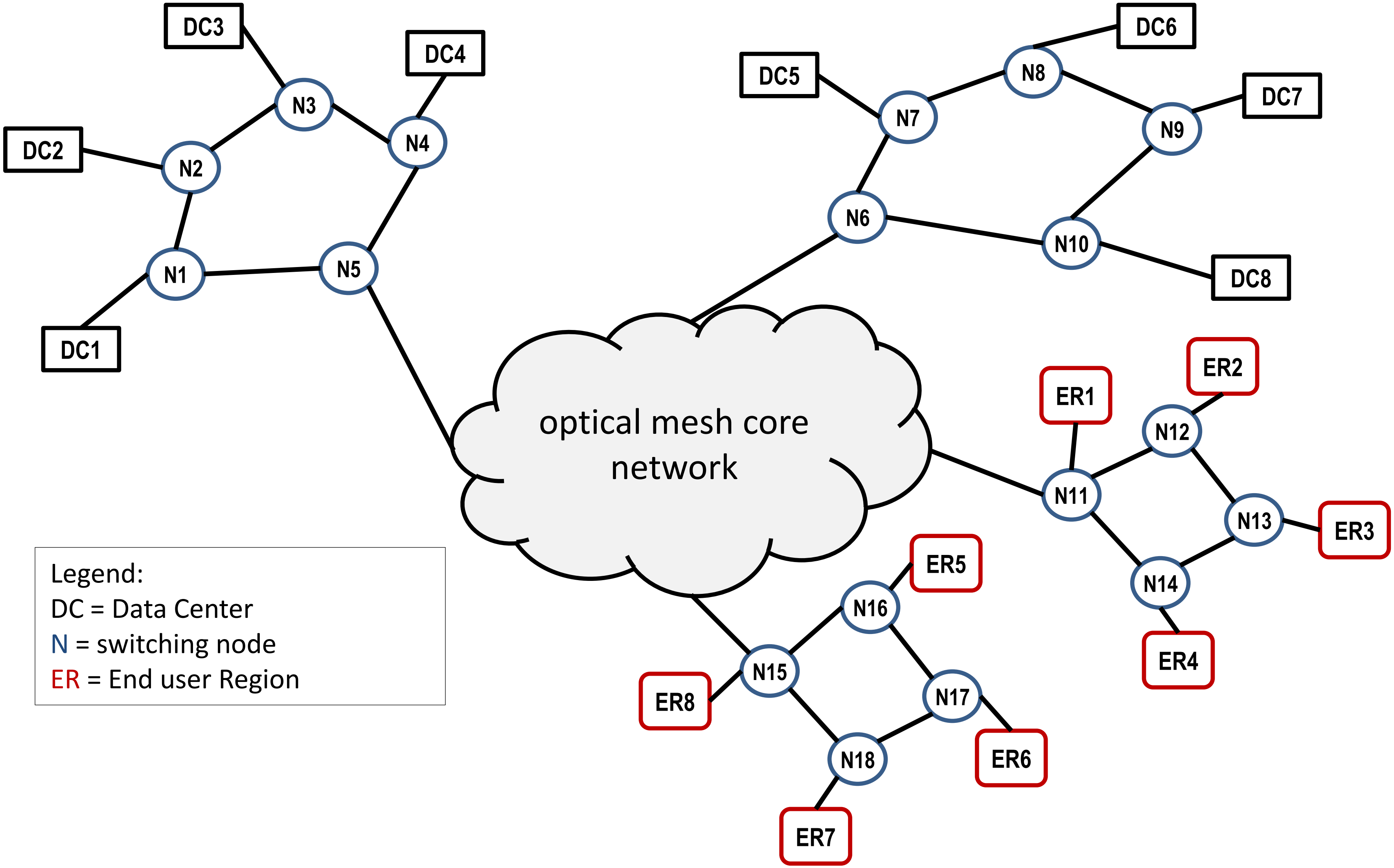
```
 {
 "meta" : {},
 "data" : {
   "graph": {
     "L1": {"a-end":"ER1", "z-end":"N1", "wt":1, “latency”: 1, "r-cap":5},
     "L2": {"a-end":"ER2", "z-end":"N1", "wt":2, “latency”: 1, "r-cap":5},
     "L3": {"a-end":"N1", "z-end":"N2", "wt":1, “latency”: 2, "r-cap":8},
     "L4": {"a-end":"N2", "z-end":"N3", "wt":2, “latency”: 1, "r-cap":6},
     "L5": {"a-end":"N3", "z-end":"DC1", "wt":1, “latency”: 1, "r-cap":5},
     "L6": {"a-end":"N3", "z-end":"DC2", "wt":1, “latency”: 1, "r-cap":5},
     "L7": {"a-end":"N2", "z-end":"DC3", "wt":6, “latency”: 5, "r-cap":10}
} } }
```

Latency Example

- Mesh network N7 and N9 communications:
  - Lowest latency path N7-N8-N9 → ***3ms***
  - Other way around low hop count ring N7-N4-N5-N6-N10-N9 gives ***13ms***
  - Other way around high hop count ring N7-N12-N14-N15-N16-N17-N13-N11-N9 gives ***10ms*** but uses much more link bandwidth



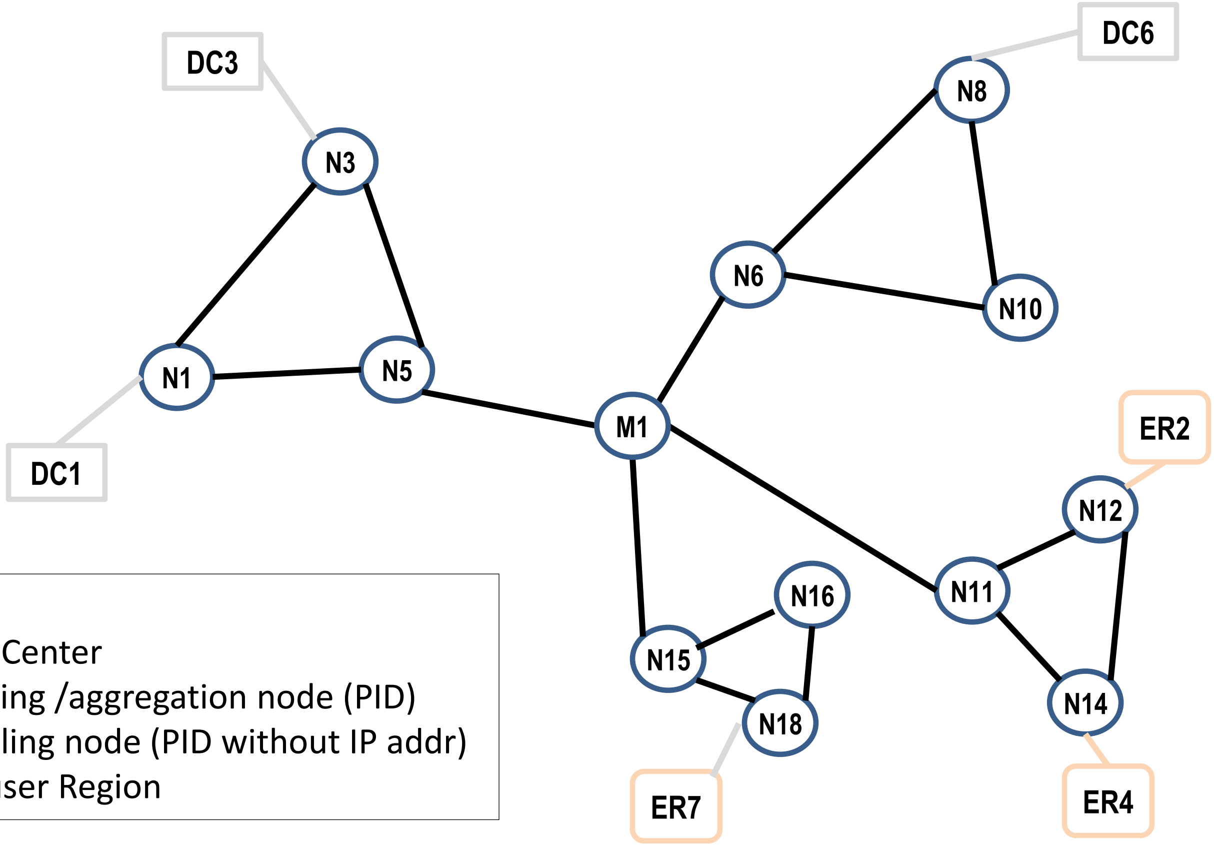
Example Regional Network



Legend:
DC = Data Center
N = switching node
ER = End user Region

Example Approximate Graph Representation

- Cost/Constraint query [DC1, DC3, DC6] to[ER2,ER4, ER7]
- Graph returned specific to query, must include PIDs for end systems of interest
- Graph not necessarily unique, method for deriving graph does not need to be standardized, though we can make recommendations…



Legend:
DC = Data Center
N = switching /aggregation node (PID)
M = Modeling node (PID without IP addr)
ER = End user Region

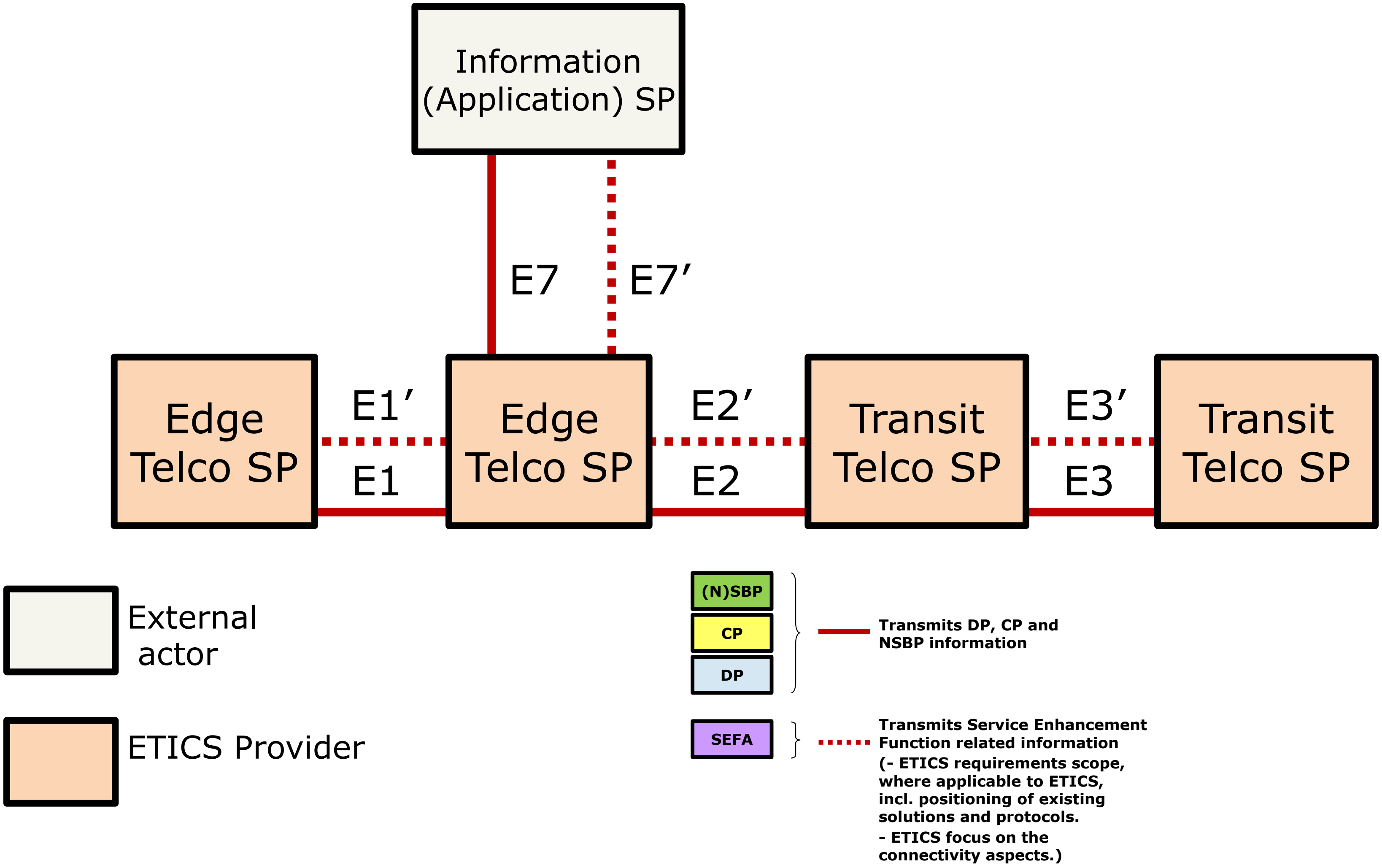
High Bandwidth Interfaces (I)

- General Notions
  - Smaller, closed user community, not the entire internet. Application controllers ↔ Network interface, not individual end users.
  - Both current and future network information and network resource reservations are useful. (Planned, On Demand, etc…)
- Network Info Sharing
  - Make use of all ALTO concepts
  - Would like ***more costs***: reservable bandwidth, latency
  - Would like ***graphs*** to approximate networks; Approximate graphs for CSO use similar but much simpler than previously studied topology aggregation problem.

High Bandwidth Interfaces (II)

- Network Notification Interface
  - Changing conditions in the network such as costs or capacity may need to be relayed to the application layer in suitable form and in a time frame relative to their importance to ***service QoS***, ***service delivery***, or ***cross layer optimization***.
- Network Resource Reservation Interface
  - A way for the application controller to indicate to the network that it should create “express lanes” for particular IP or Ethernet flows. The application would not be involved with network technology specific layers as is done in UNIs, PCE, or GMPLS.
  - Negotiation for the QoS/QoE requirements/SLA between APP and NET

Multi-domain possibility
ETICS ASQ



Information (Application) SP

E7

E7’

Edge Telco SP

E1’

E1

Edge Telco SP

E2’

E2

Transit Telco SP

E3’

E3

Transit Telco SP

External actor

ETICS Provider

(N)SBP

CP

DP

Transmits DP, CP and NSBP information

SEFA

Transmits Service Enhancement Function related information (- ETICS requirements scope, where applicable to ETICS, incl. positioning of existing solutions and protocols.
- ETICS focus on the connectivity aspects.)