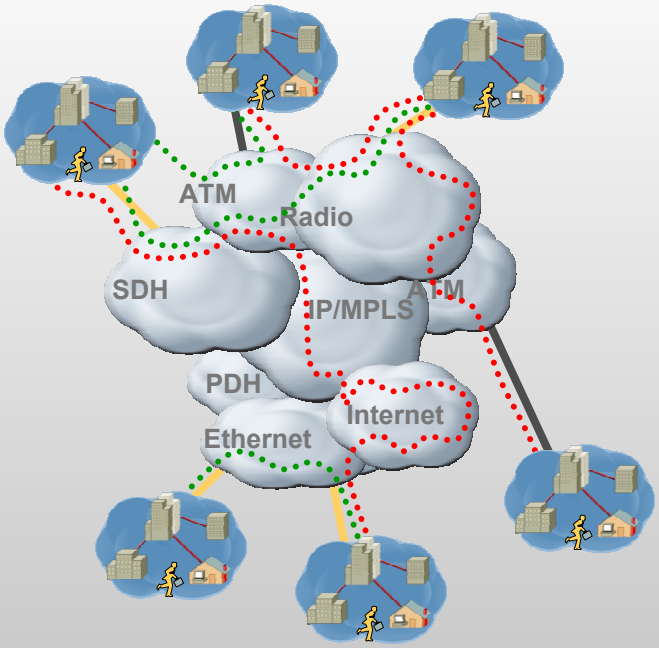
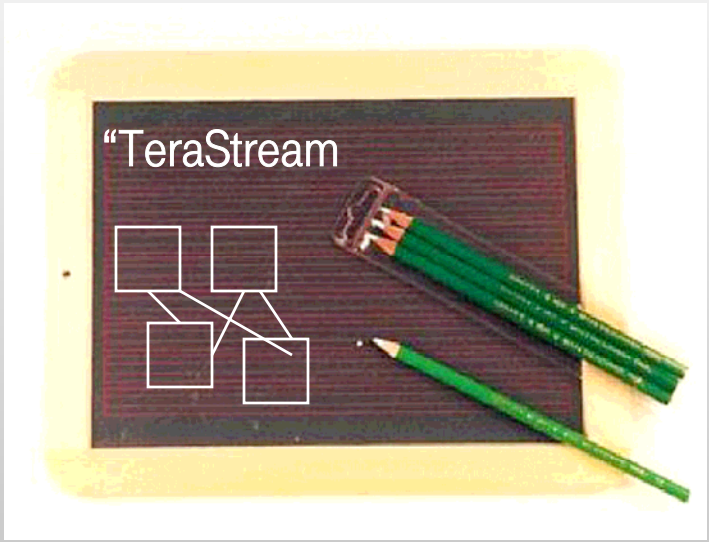




TeraStream –

A Simplified ~~IP Network~~ Service Delivery Model

Peter Lothberg, CSN, Deutsche Telekom AG

Life is for sharing.

I HAVE BEEN DOWN THIS ROUTE BEFORE…. …..

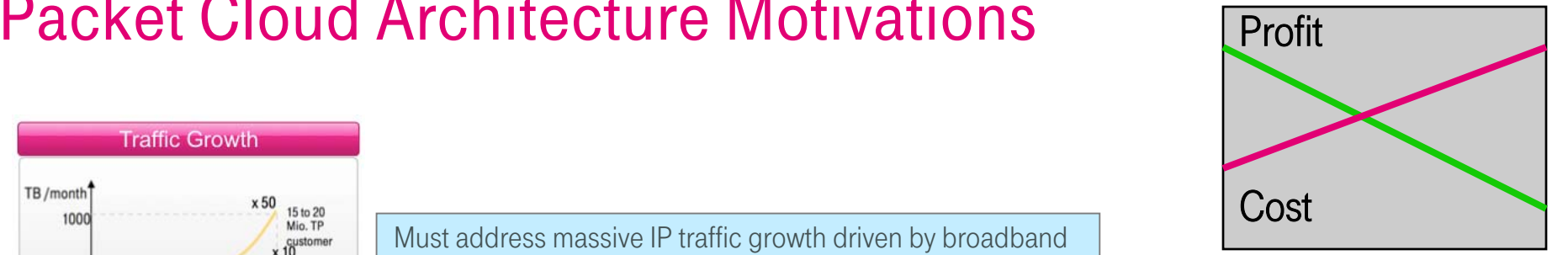
- 20 Years ago I was talking at Ripe meetings about how to combine existing network islands in to what became “Ebone”
- Introducing BGP3, trying to not use IGRP, HELLO, default route.....
- Moving to CIDR and BGP4 it became a reference model for a basic IP-transit operator configuration, EBS -> GW -> PE
- Many packets under the bridge, gray hair and I turned DNS this year
- Remember, “Keyed IPv6 Tunnel”

TODAY'S TALK ....

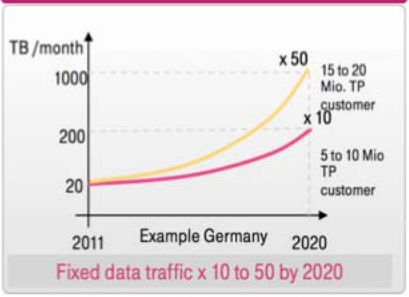
- Some years ago I was talked in to making a suggestion on how to build a future customer access system, target 2020 Starting with a empty white A3 paper (stolen from the office copie
- The inner packet in my head was saying;
  - As few boxes as possible (but do carry all packets)
  - As few interfaces as possible
  - No special HW (eg pingmaster2000)
  - No “services” in the network elements
  - IPv6 only, use only L3 tools, no layer violations / Carrier_Ethernet_SERVICE
  - IPv4 is a service, L3 VPN is a service, L2 is a service
  - If a technology is missing today, make it happen ASAP.
  - Fully automated operations
  - NNI (bidirectional with visibility)
  - Driven by data models (Netconf/Yang)
  - Use multi homing as the tool to pass policy to the home network
  - A IPv6 /56 per provider to every home. (that’s 2^72 hosts in your house)

TERASTREAM

Packet Cloud Architecture Motivations

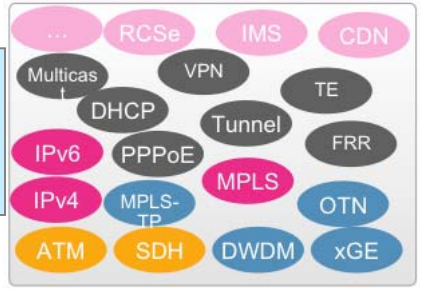


Traffic Growth



Must address massive IP traffic growth driven by broadband access and new Internet services and Internet business models

Cost & Complexity



Many networks and technologies, complex systems – long service lead-times, high-cost evolution to converged network architecture

Competitive Pressures

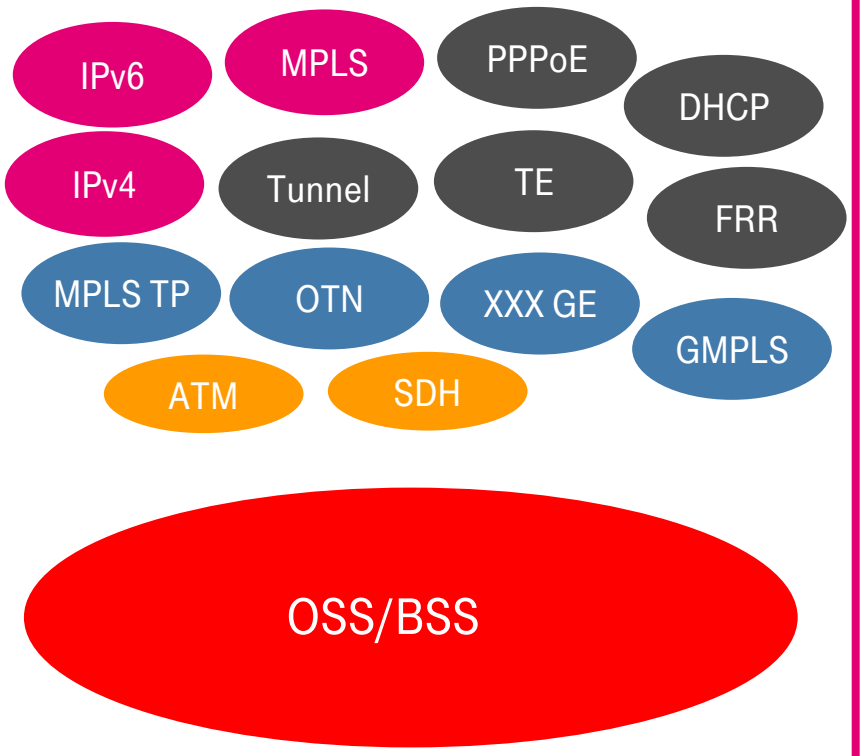
Competitors offer better performance, more service flexibility and more features, faster provisioning, lower price

Long Innovation Cycles

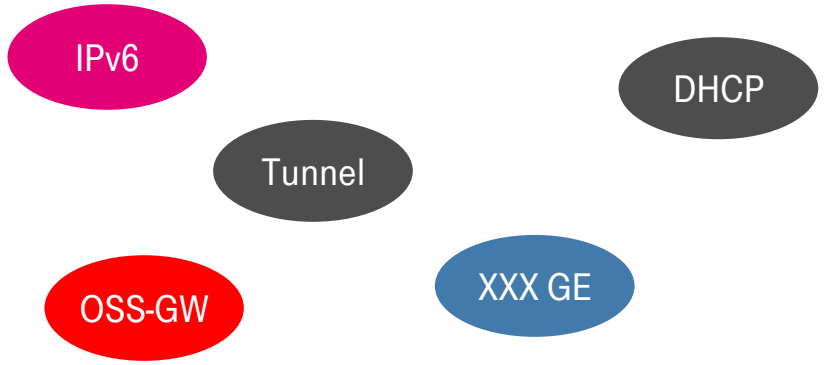
Multi-layer system complexity results in slow or lack of service innovation, low customer satisfaction, impacting revenue

KAIKAKU FOR IP NETWORKS
INDUSTRY LEADERSHIP

From

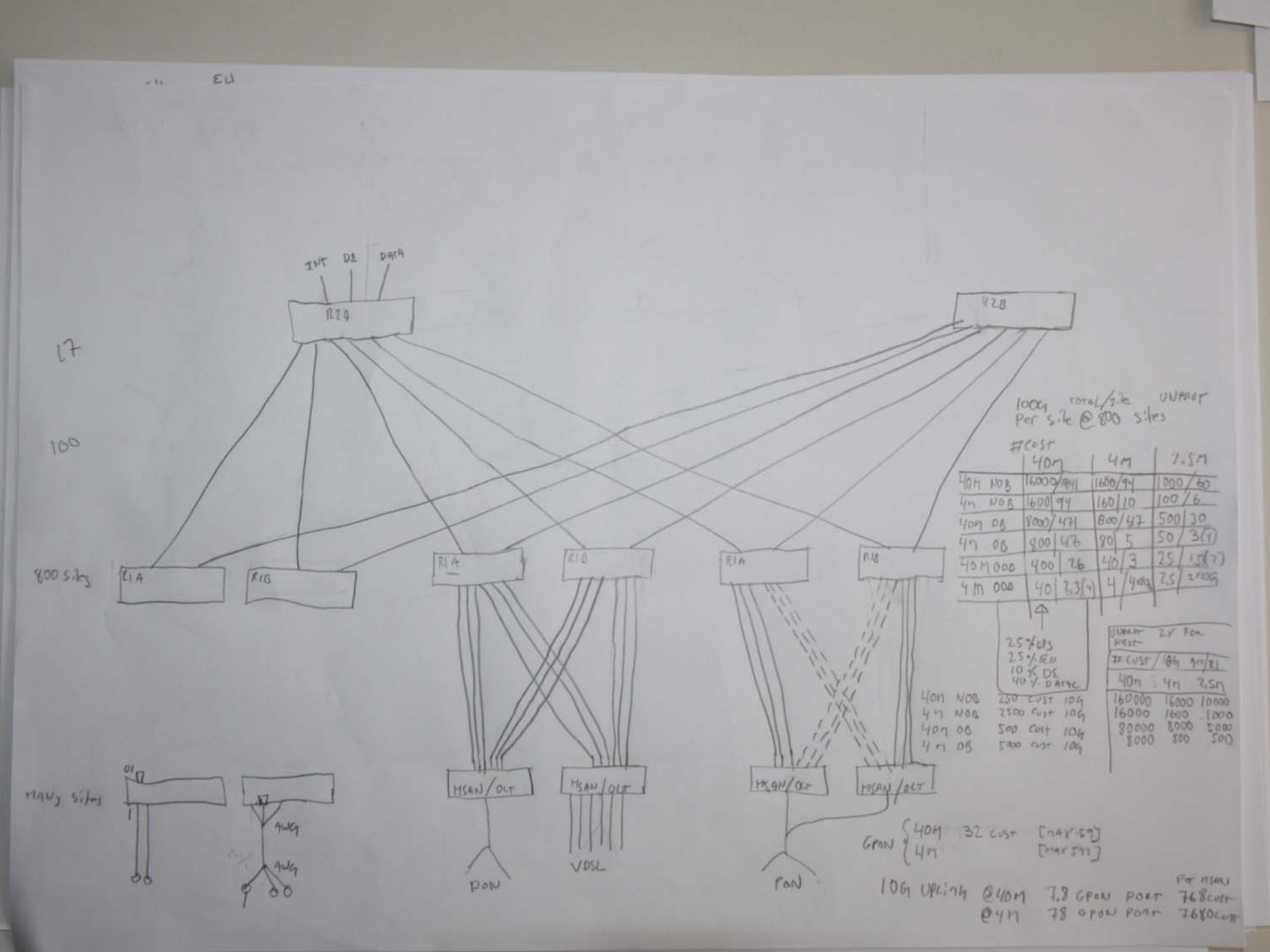


To

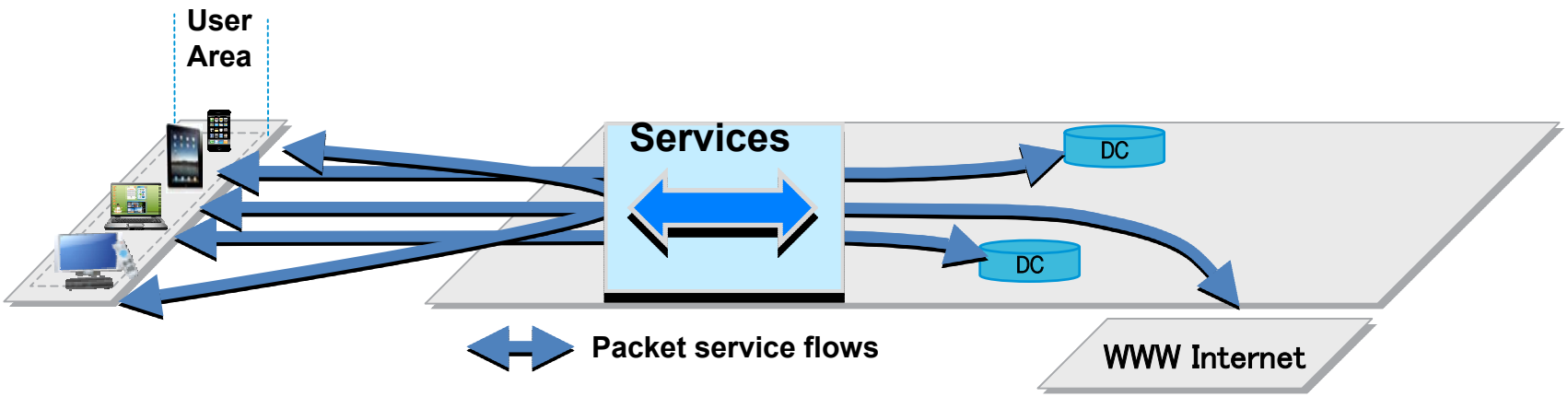


TeraStream

- Drastic simplification of IP networks
- IP&Optical integration
- Infrastructure Cloud model


INT
DE
DATA
R2A
R2B
17
100
800 sites
E1A
R1B
R1A
E1B
R1A
R1B
MSAN / OLT
MSAN / OLT
MSAN / OLT
MSAN / OLT
PON
VDSL
PON
MANY sites
01
AWG
AWG
1000 total/site UNPROT
Per site @ 800 sites
cost
40M
4M
2.5M
40M NOB 16000/944 1600/94 1000/60
4M NOB 1600/94 160/10 100/6
40M OB 8000/471 800/47 500/30
4M OB 800/46 80/5 50/3(?)
40M 000 400/26 40/3 25/1.5(?)
4M 000 40/2.3(?) 4/4xlg 2.5/2x10G
25% LTS
25% FTU
10% DS
40% DATA
UNPROT 2x PON PORT
CUST / 10G 4M/81
40M 4M 2.5M
40M NOB 250 CUST 10G 160000 16000 10000
4M NOB 2500 CUST 10G 16000 1600 1000
40M OB 500 CUST 10G 80000 8000 5000
4M OB 5000 CUST 10G 8000 800 500
GPON { 40M 32 CUST [MAX 64]
4M [MAX 512]
10G UPLINK @40M 7.8 GPON PORT 768 CUST
@4M 78 GPON PORT 7680 CUST
PER MSAN

CONVERGING TO PACKET CENTRIC NETWORK, WHY?

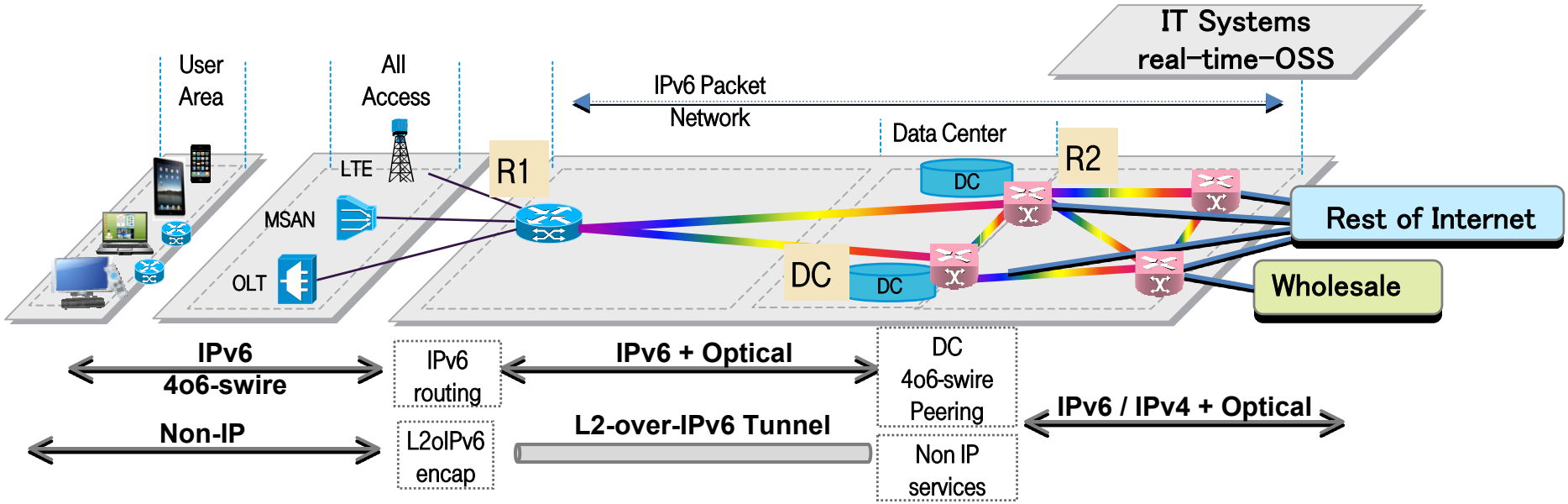


- Improve user experience, Internet services to more users
- Use just enough complexity to do the job and no more
- Get the revenue and cost balance right

TERASTREAM DESIGN PRINCIPLES

Principle	Applied to TeraStream design
Reduce the amount of technologies used	Use IP and optical transmission only No OTN, L2, MPLS switching
Use IPv6 for all internal functions and services	No native IPv4 support in the network IPv4 is a service IPv6 based “carrier Ethernet service”
Avoid internal interfaces	Minimize non-customer, non-peering facing interfaces Distribute Internet peerings, offload external traffic ASAP
Size the network to handle all IP traffic without IP packets losses	Dimension the network for peak hour IP traffic, no oversubscription, packet loss is extreme exception
Integrate optical networks and IP networks as much as possible	Integrate IP and optical layers into routers to simplify the network, avoid redundant mechanisms e.g. failure handling, reduce total cost
Use one network for all services – Internet, IP TV, business, …	Single converged packet network Note: Dominant traffic drives the design!
Deterministic and short routing path for all on-net traffic	Network distance between R1 access routers is at most two R2 backbone routers away and R1 is multi homed to two R2
Service policy for packets are outside the payload	Encode service type, traffic class, direction etc in the IPv6 address
Data Centers are directly connected to backbone routers	DCs connect directly to R2s to avoid building internal IP interfaces for very large amount of traffic

TERASTREAM – DESIGN IN A NUTSHELL



TeraStream key functional elements

R1

- Terminate access interfaces
- Runs IPv6 routing only, integrates optical
- Access services
  - IPv6 - dealt with natively
  - IPv4 – IPv4 over IPv6 softwire between HGW / CPE and DC, R1 not involved
  - non-IP - L2-over-IPv6 encapsulation
- User configuration
  - using Netconf / Yang
  - Driven by real-time OSS i.e. self-service portal

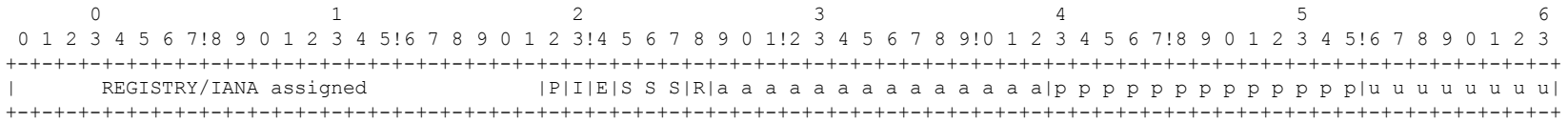
R2

- Connects R1s, Data Centers and Internet peerings
- Runs IPv6 and IPv4 routing, integrates optical
- Closely integrated with Data Centers
  - Optimized handling of locally sourced services
- High scale IP bandwidth

Data Center / Services

- Distributed design
  - fully virtualized x86 compute and storage environment
- Network support functions - DNS, DHCP, NMS
- Real-time OSS incl. user self-service portal
- Cloud DC applications, XaaS services
- Complex network services e.g. high-touch subscriber handling

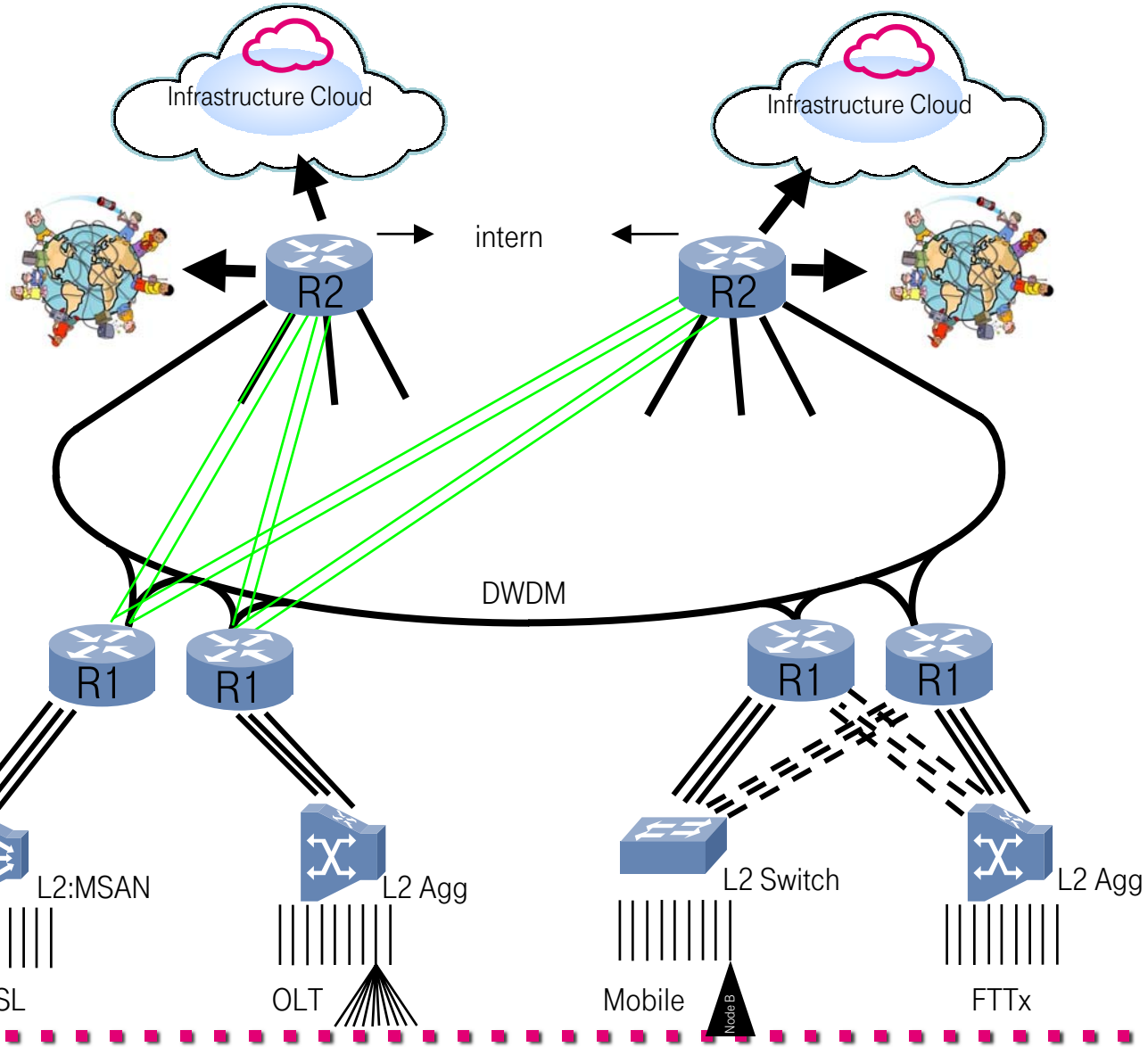
IPV6 ADDRESSING FORMAT, USERS



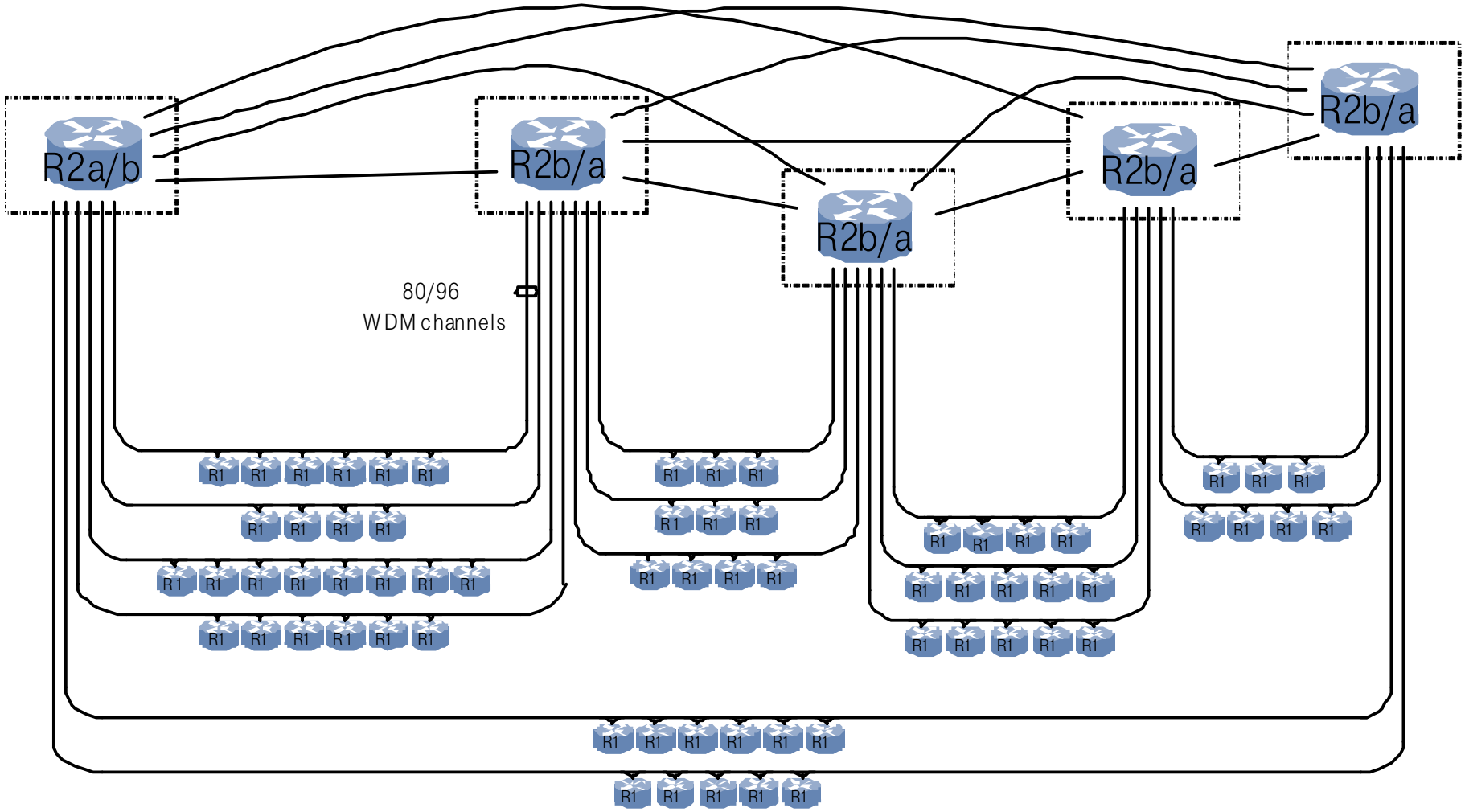
P Public	0=traffic internal to local SP
I Infrastructure	0=user traffic
E Endpoint/Service	0=network endpoint, 1=service
S Logical Network (Internal ISP#)	0=res, 1=res, 2=internet, 3=res, 4=video, 5=L2 service, 6=voice, 7=management
R Reserved	
a R1 Area 14 bit	Indicates what R1 that the address is delegated from, max 16,384 R1
p R1 User 13 bit	User identifier, max 8192 users
u User subnet	Delegated to user

Examples:	Source PIESSS	Destination PIESSS
User -> Voice	000110	011110
Voice -> User	011110	000110
User -> User (best effort)	X00001	X00001
User -> Internet (best effort)	100001	XXXXXX
Internet -> User (best effort)	XXXXXX	100001
Lan-Lan service	010101	010101

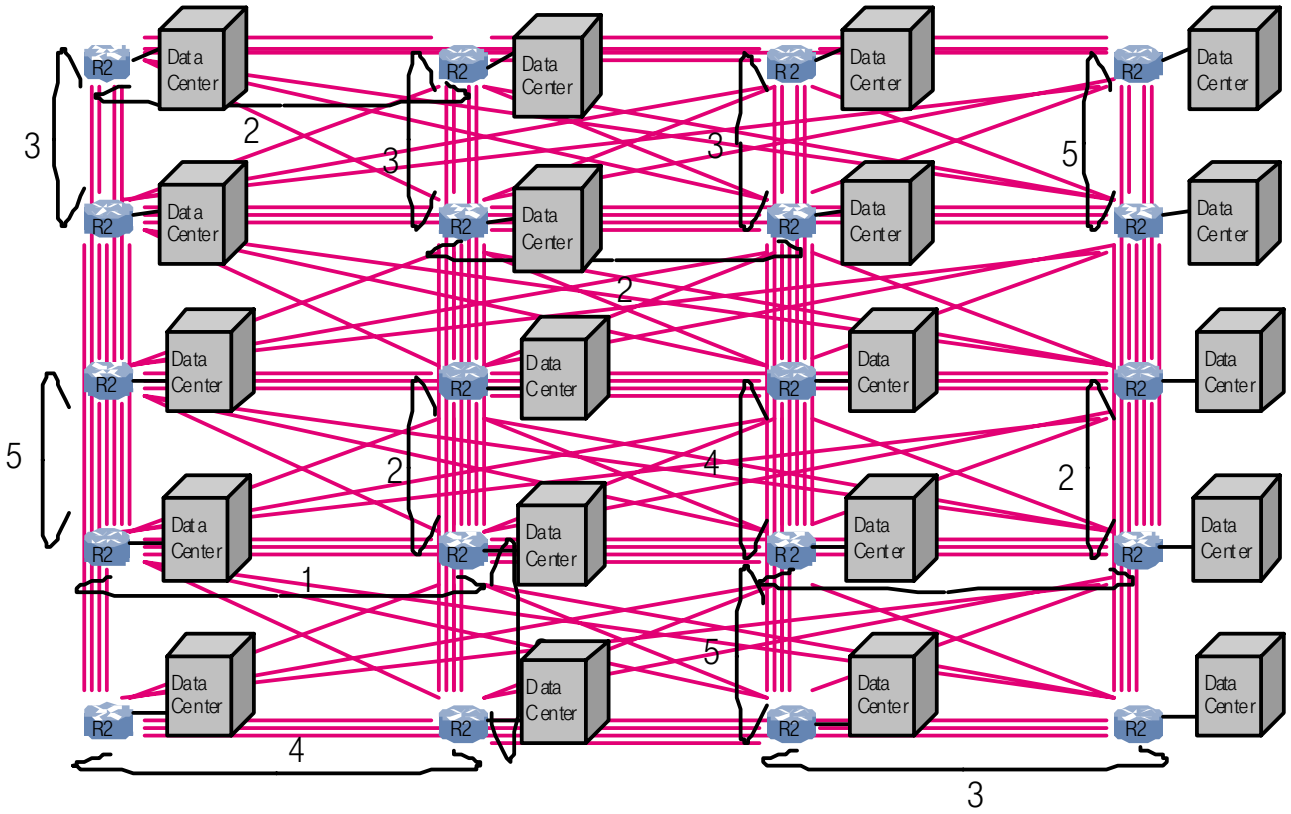
THE TERASTREAM ARCHITECTURE



R1 <-> R2 OPTICAL FIBER LINKS



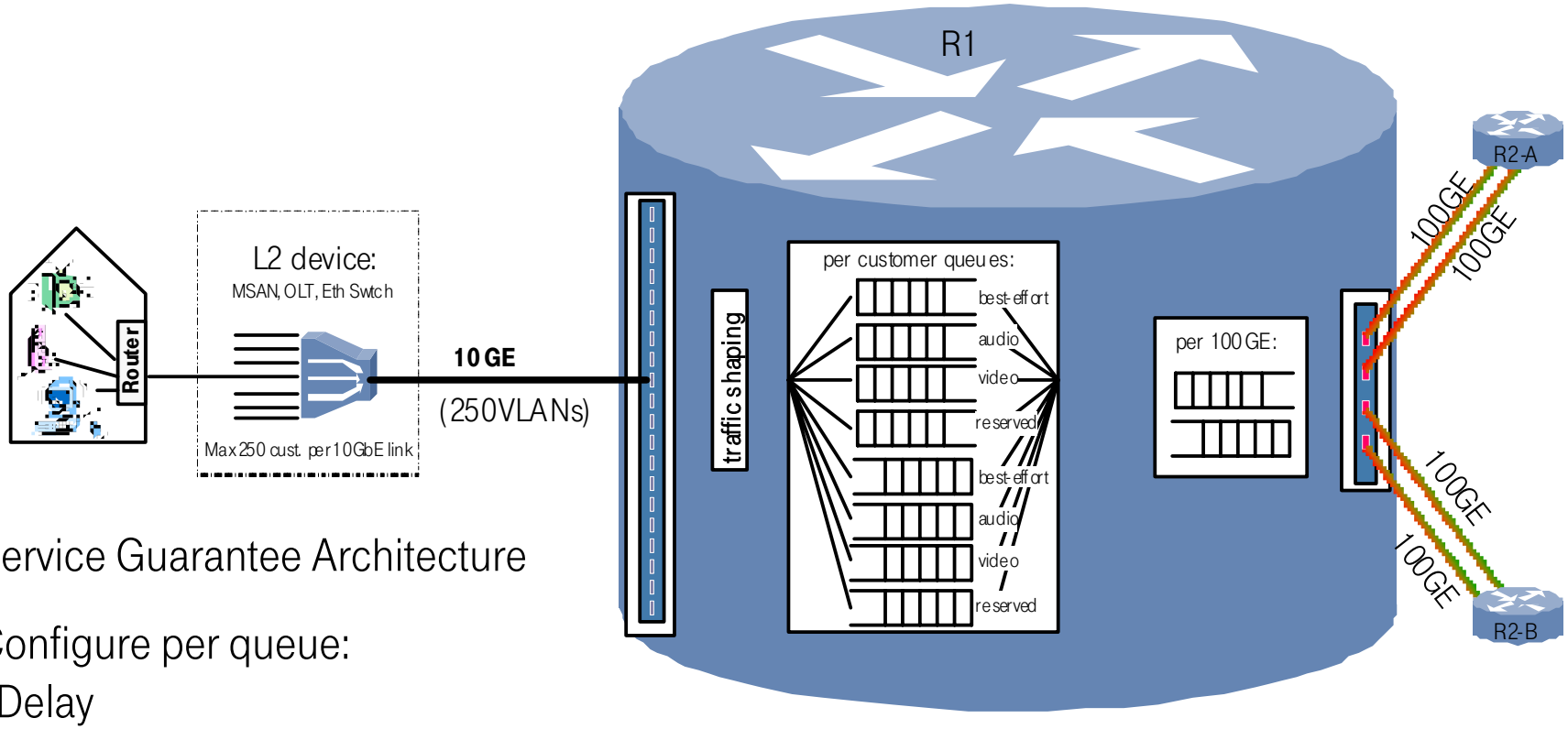
IP “R2 GRAPH”



R2 deployment characteristics (examples):

- The black lines are the horseshoes
- The numbers indicate the amount of horseshoes between R2 pairs
- The red lines indicate extra links to implement a fully meshed model (not all red links are shown, approx 400 links)
- Each R2 has its own data center. Data centers are deployed as a “cloud” so that services can be accessed on any of them.

TERASTREAM USER FACING ROUTER R1



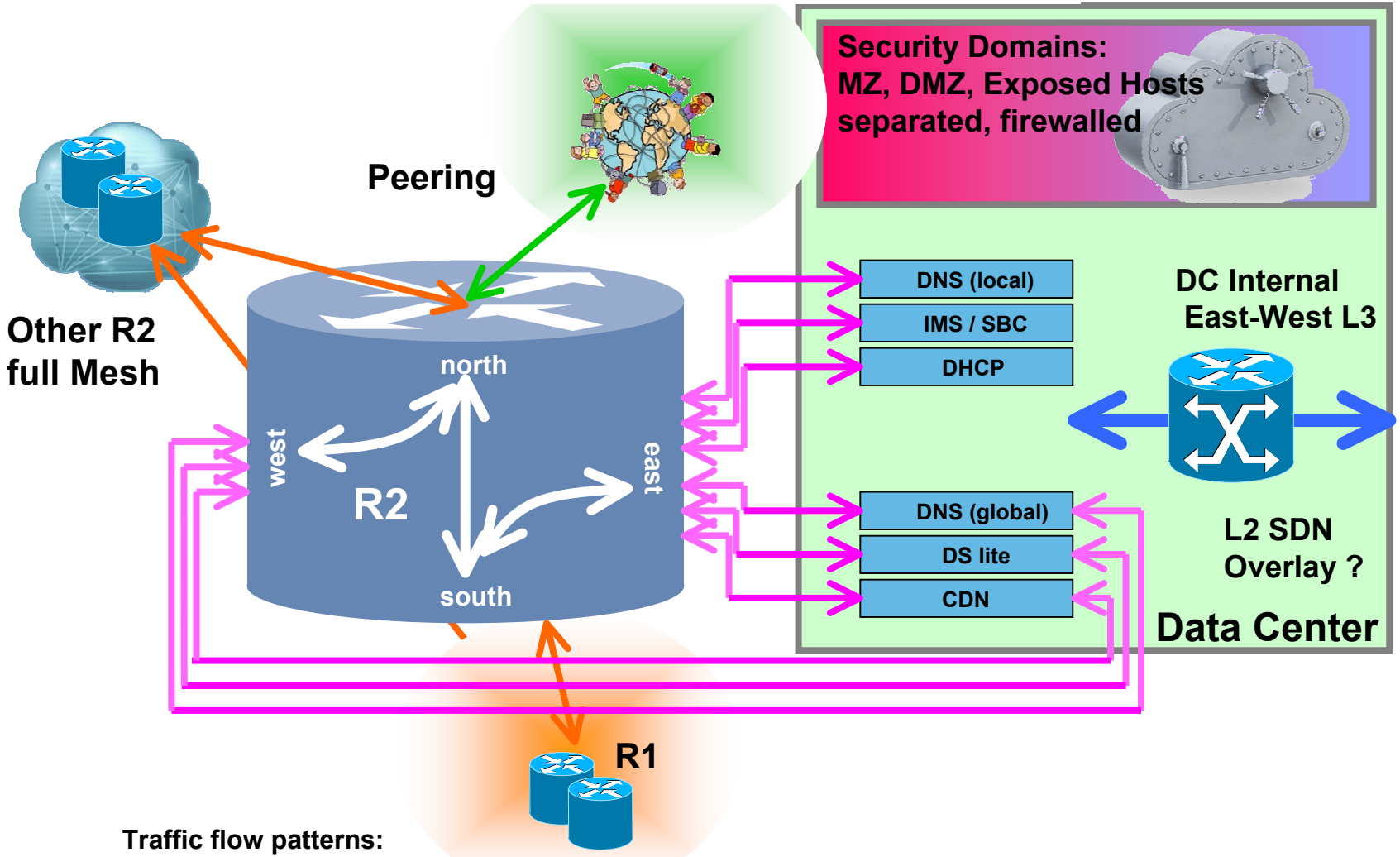
Service Guarantee Architecture

Configure per queue:

- Delay
- Drop
- Bandwidth
- Reorder
- Etc...

- IP traffic shaped to capabilities of L2 device
- 5000 customers connections per R1
- 20 * 10GE port for L2 device
- 4 * 100GE for R2 link

R2 ROUTER AND TRAFFIC PATTERNS



Traffic flow patterns:

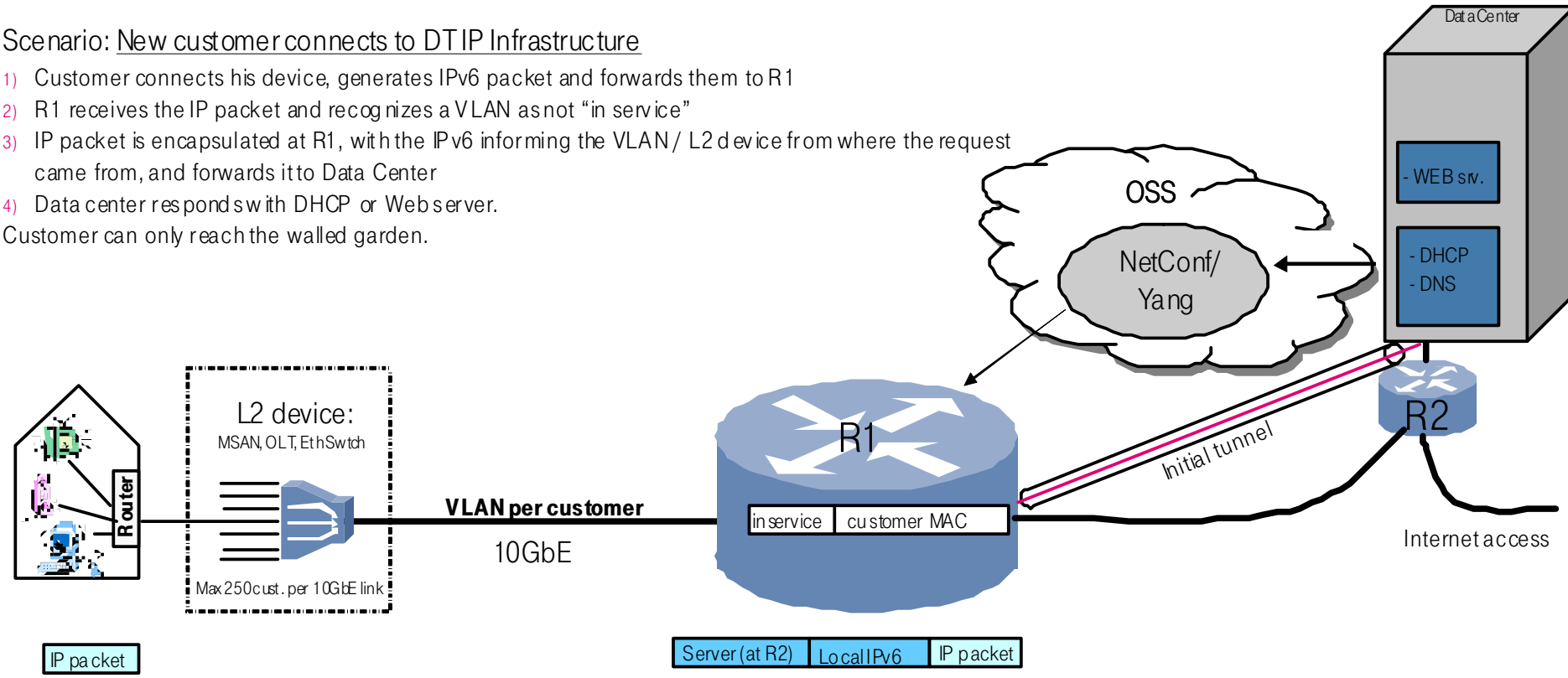
- **R1 ⇔ Peers and Other R2 going north ⇔ south (example: IPv6 Internet traffic)**
- **R1 ⇔ Data Center services going south ⇔ east (example: DHCP)**
- **R1 ⇔ Data Center ⇔ Peers going south ⇔ east ⇔ west ⇔ north (example: IPv4 Internet traffic)**

IF NOT IPV6, USE THE NETWORK AS A PTP ETHERNET

Scenario: New customer connects to DT IP Infrastructure

1) Customer connects his device, generates IPv6 packet and forwards them to R1
2) R1 receives the IP packet and recognizes a VLAN as not "in service"
3) IP packet is encapsulated at R1, with the IPv6 informing the VLAN / L2 device from where the request came from, and forwards it to Data Center
4) Data center responds with DHCP or Web server.

Customer can only reach the walled garden.



Scenario: customer registers

1) Web server at Data Center generates a request to OSS to configure a new customer via NetConf / Yang at router R1, Line ID.
2) The OSS via NetConf configures the R1 as "in service" for a customer located at a specific interface (IPv6 address).
3) From now on, the customer is outside the walled garden and can reach other Internet addresses.

IPV4 DECOMMISSIONING STRATEGY

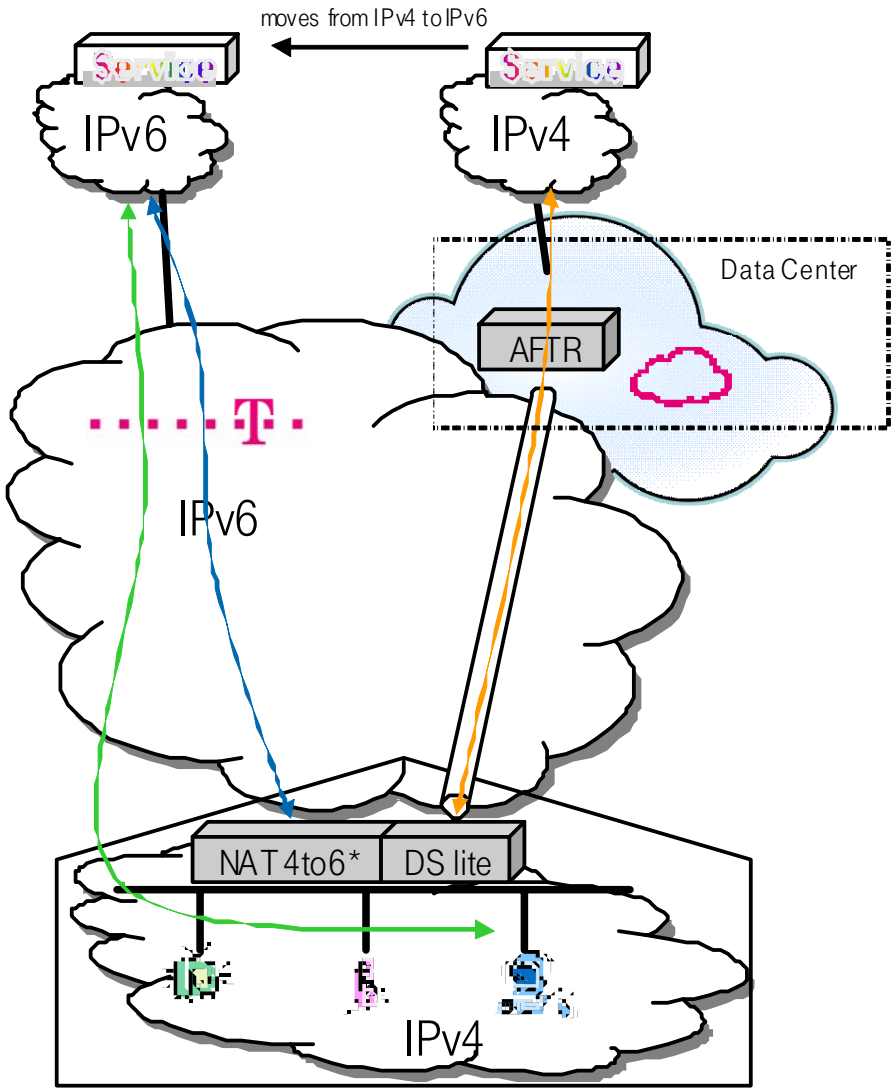
The Internal IP network of DT is IPv6. All IPv4 traffic to and from the customer will be translated to IPv6 at the borders of the network. 2 alternatives are seen as viable:
1) Customer IPv4 traffic is encapsulated on IPv6 via DS-lite to a AFTR element located at the Data Center. RFC 6333.

2) Customer IPv4 traffic is translated to IPv6 at the customer's device (NAT 4to6). (Standard not defined)

In the long term, the expectation is that most customers will be IPv6 capable and that the services will move to IPv6.

In the transition time DS lite should provide the mechanism to connect IPv4 devices to other networks.

There is no standard describing NAT 4to6, i.e. translating IPv4 packets to IPv6. This standard remains for further work.



* NAT 4to6 – Standard not defined

HIERARCHY OF CLOUD BASED SERVICES

Backend DC
"classic" large scale/business Cloud Services

Decoupled - based on any commercial cloud technology

Focus on commercial, highly sensitive, low BW applications

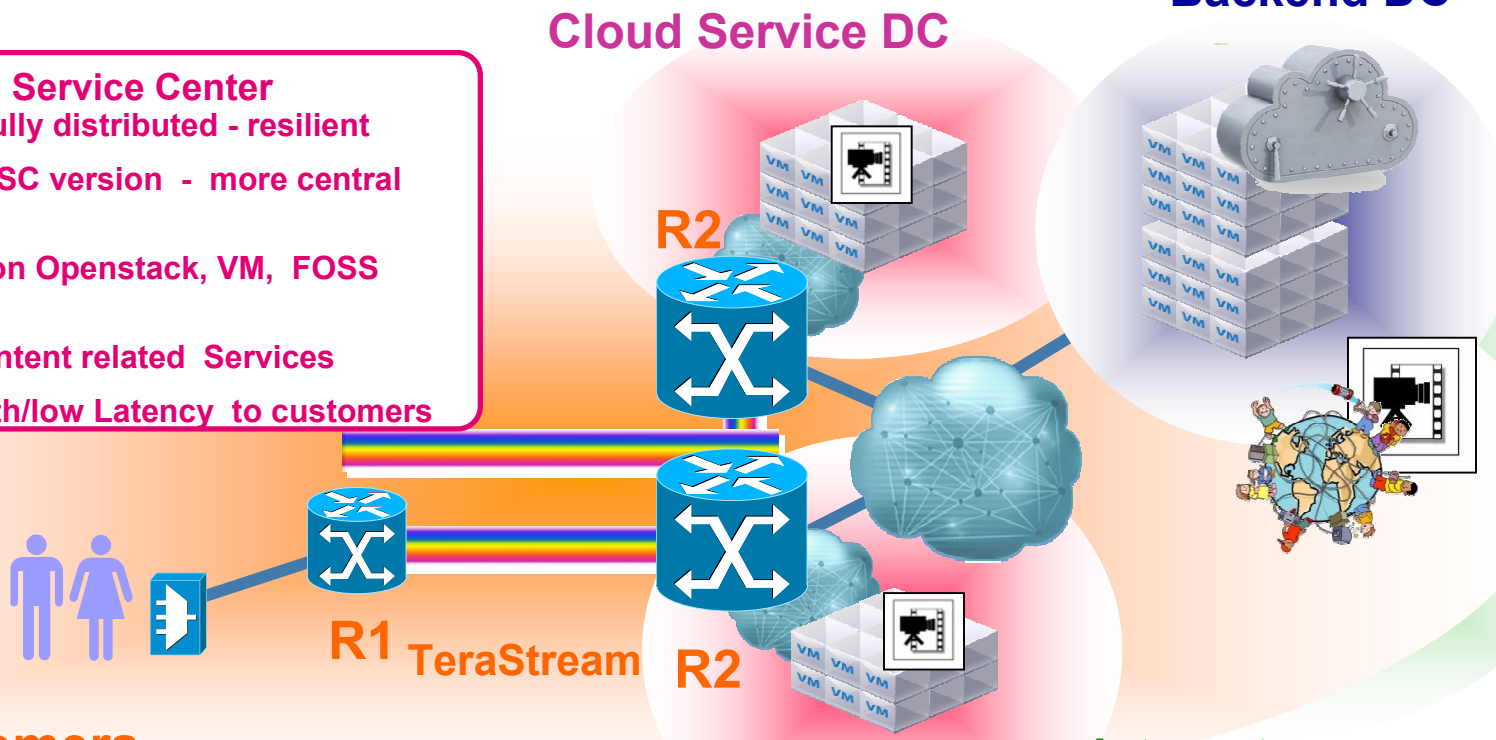
Frontend - Cloud Service Center
adjacent to R2 - > fully distributed - resilient

(Optional Master CSC version - more central functions)

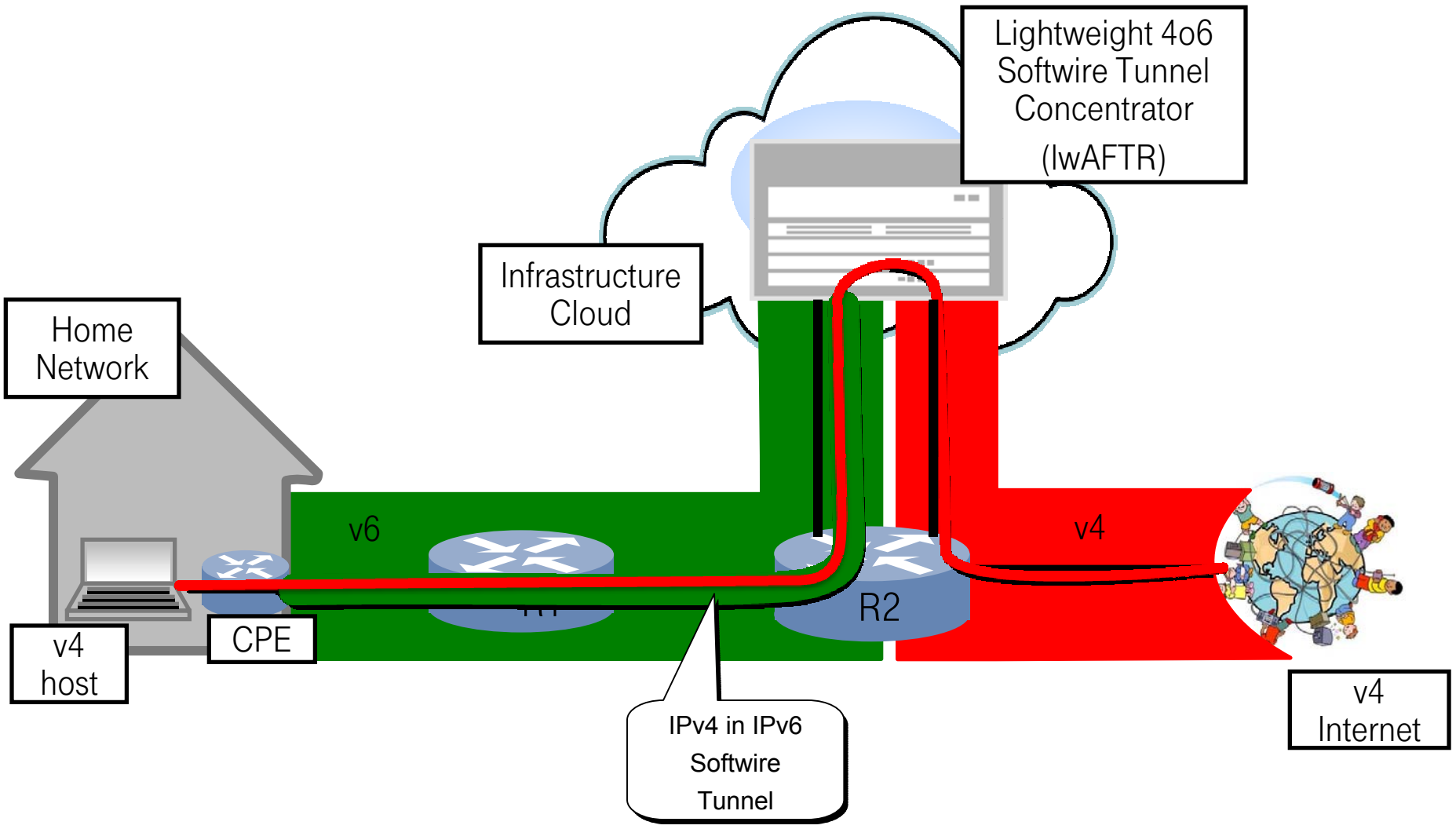
Integrated - based on Openstack, VM, FOSS

Focus on:
Infrastructure & Content related Services

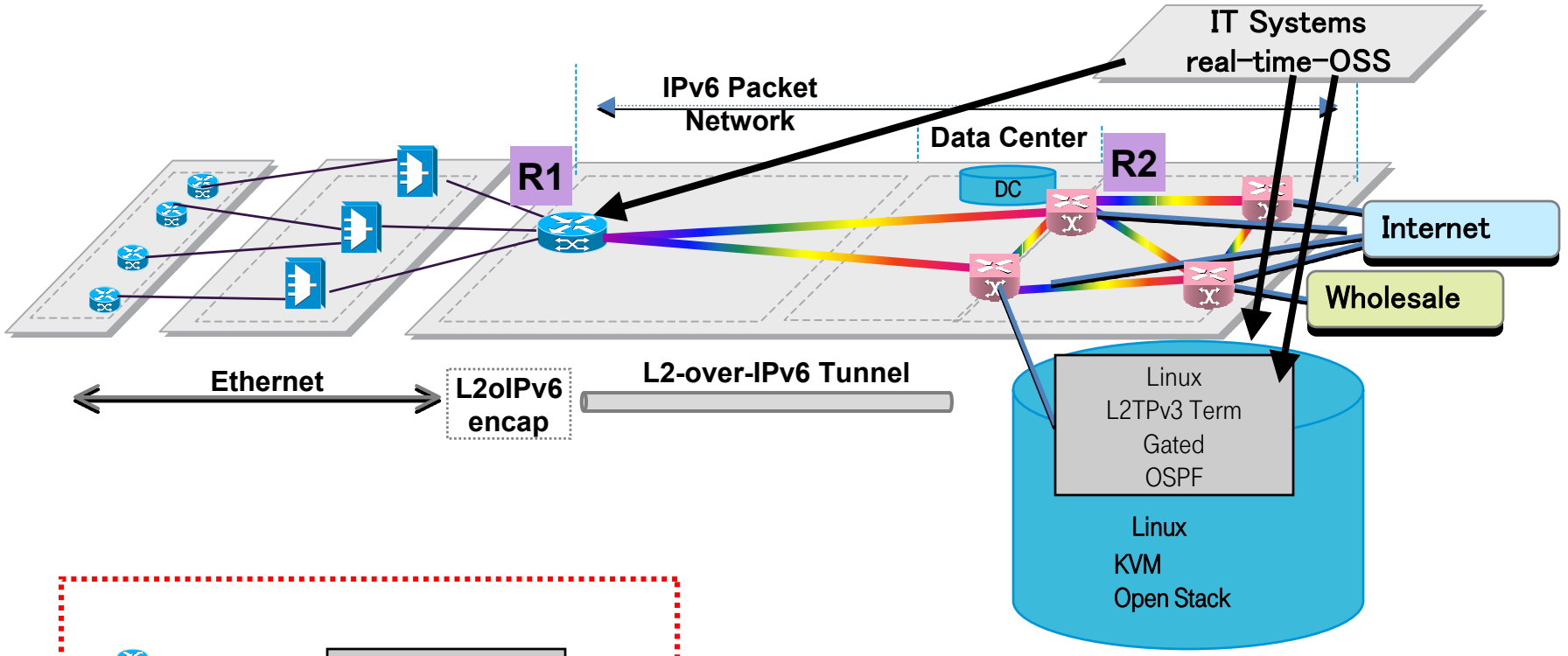
Very high Bandwidth/low Latency to customers

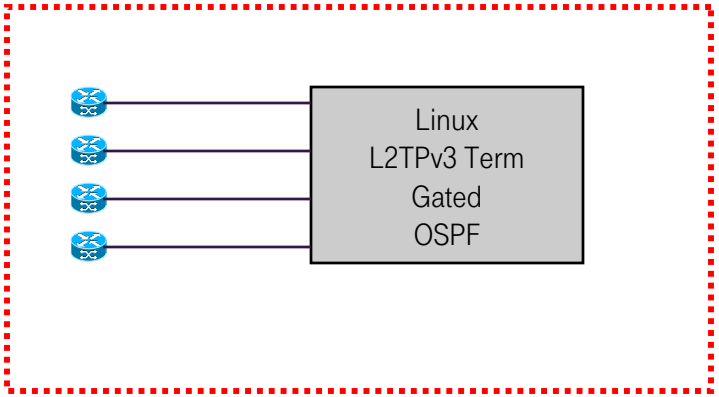


IPV4 AS A SERVICE – LIGHTWEIGHT 4O6 SOFTWIRES

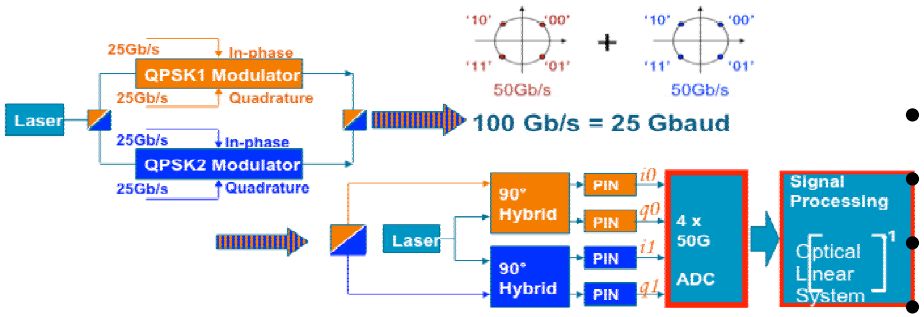


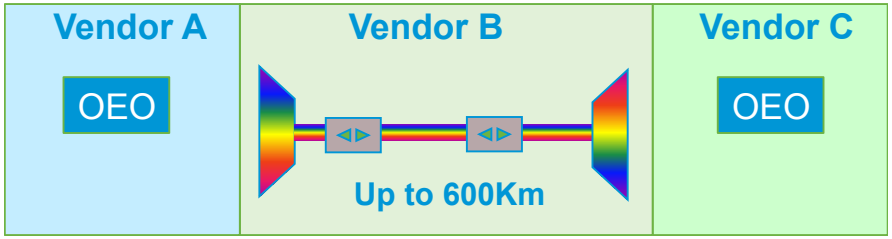
VIRTUALISATION OF SERVICES – L3 VPN

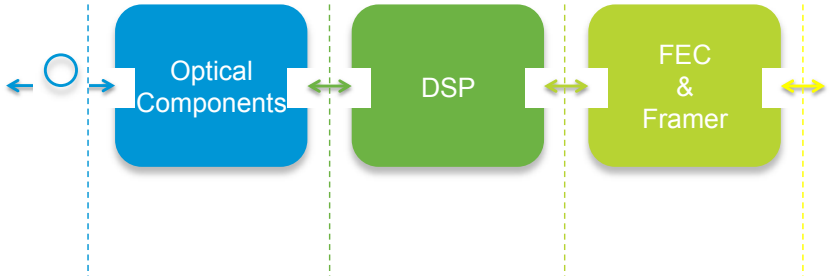




100G COHERENT DWDM INTEROP
and Pluggable Technology





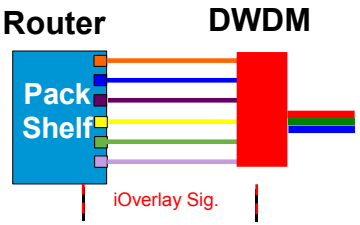


- Agree on a common set of parameters for the 100G line side
- Enable innovation by many players in the silicon optics arena
- Work driven by DT, Cisco, ALU, Cortina
- Hard staircase FEC, typ 800km
- If price is right, use in data center
- Coding
- Carrier Recovery
- Acquisition (blind)
- Reach
- Framing (works with both OTU4.4 and OTU4.10)
- Forward Error Correction (Hard FEC Staircase)

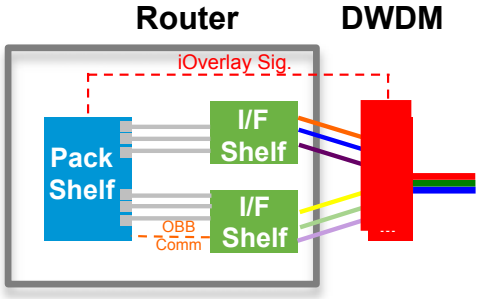
PACKET OPTICAL INTEGRATION

- Packet Optical Integration takes physical layer OAM&P close to the service
- Provides upper layer awareness of physical layer performance
  - Ability to provide a proactive network rather then reactive network
- Simplifies circuit turn up removing redundant layers
- Two forms of implementation
  - Physical Integration
  - Logical Integration

Physical Integration

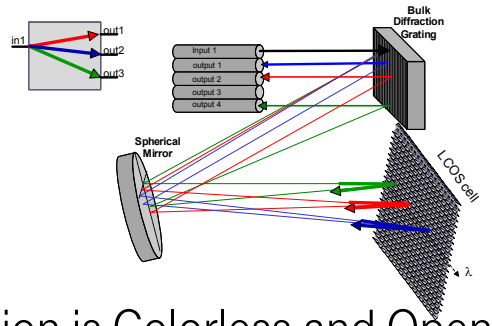


Logical Integration



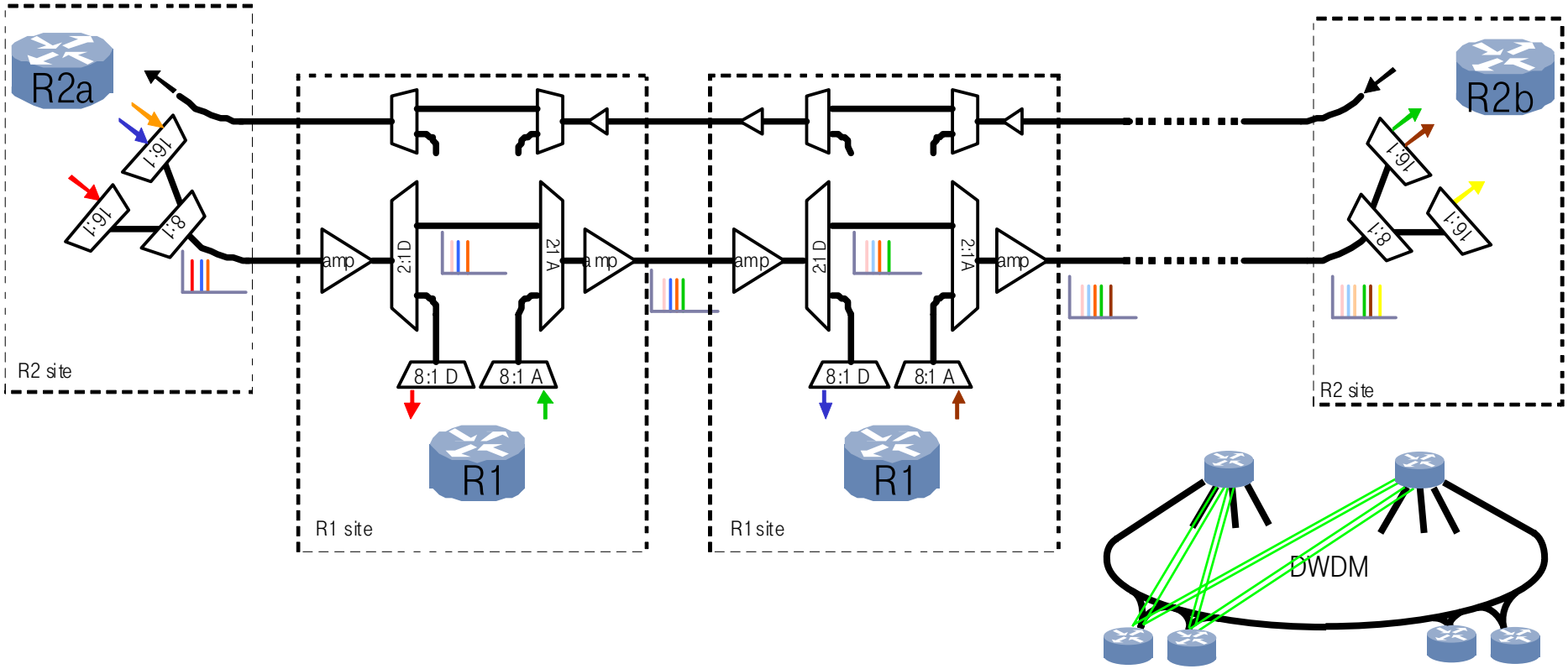
FLEXIBLE SPECTRUM

- Most common approach is to leverage LCoS technology in ROADMs
  - LCoS – Liquid Crystal on Silicon – common fabrication as in consumer electronics industry
  - Provides 12.5GHz of granularity providing Shaping and Filtering



- An Optical Splitter, by definition is Colorless and Open Spectrum
  - Low Cost, flexible port ratios
  - Fused Glass – no moving parts
  -

TERASTREAM OPTICAL “DROP & WASTE” MODEL



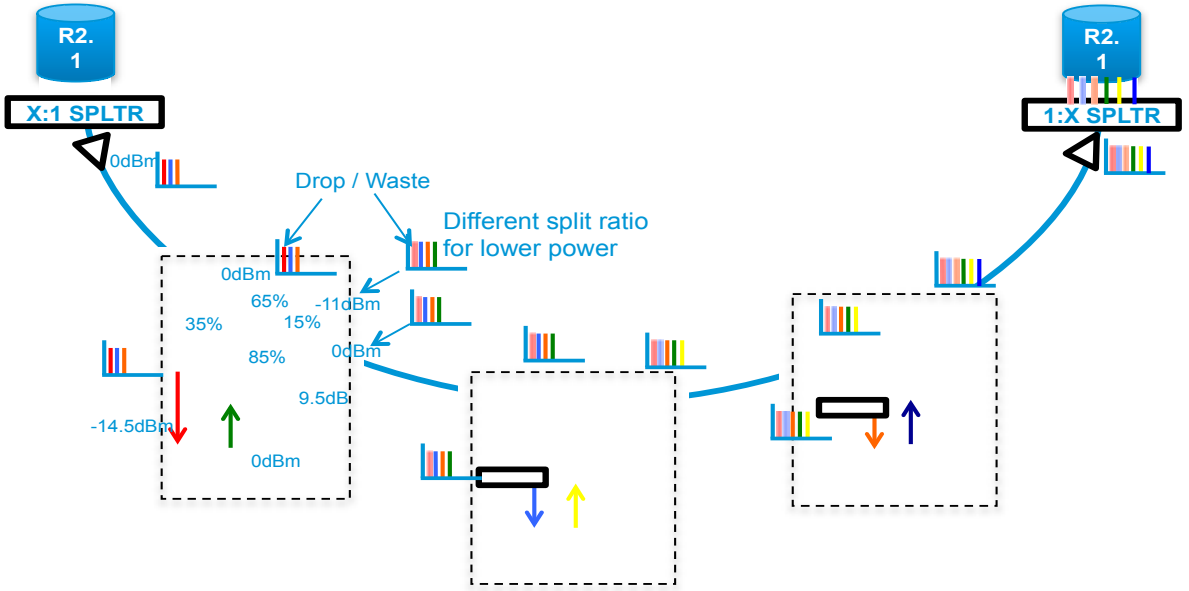
- The diagram shows only the lambdas in the direction from R2a to R2b. The direction R2b to R2a works in a similar form.
- The diagram also only shows one link from R2a to R1s, while TeraStream uses two links.

PROS AND CONS

- TeraStream Provides
  - Streamlined Circuit Provisioning by taking provisioning closer to the Service
  - Leverages Low Cost, Simple, Colorless and Open Spectrum Splitters
  - L3 protection and ultra fast Restoration if needed
  - Enhanced availability by reducing Components from the network
- This comes at the cost of
  - Wasting wavelengths based on Drop and Waste nature of Splitter architecture
  - Policing of wavelengths at Ingress is not available
  - Channel equalization provided only at the TX port
  - Coherent only solution due to channel selection in splitter network

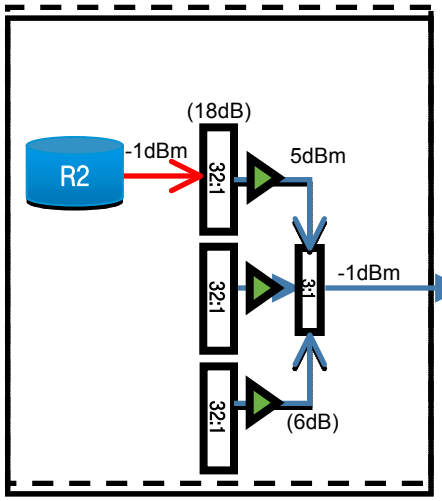
TERASTREAM, OPTICAL REALITY

- Drop and Waste Architecture is utilized
  - Based on Splitters hence all channels express including dropped channels
  - Channel Selection based on Coherent RX (6.25GHz Center Freq.)
- All power balancing will take place at the TX port of the DWDM interface with 10dB of freedom – address peak to peak variation
- Properly selected Splitters are used to ensure proper channel combining



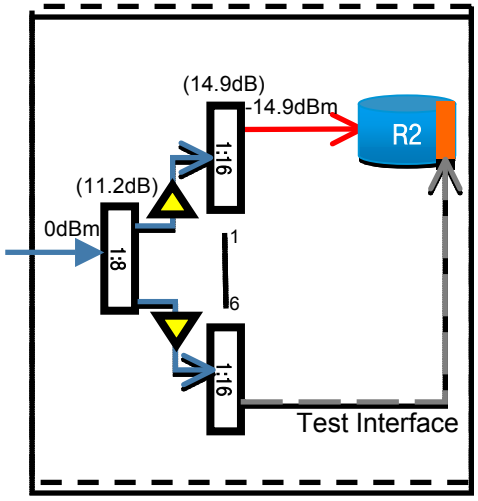
SPLITTER COMBINATIONS

R2 Add

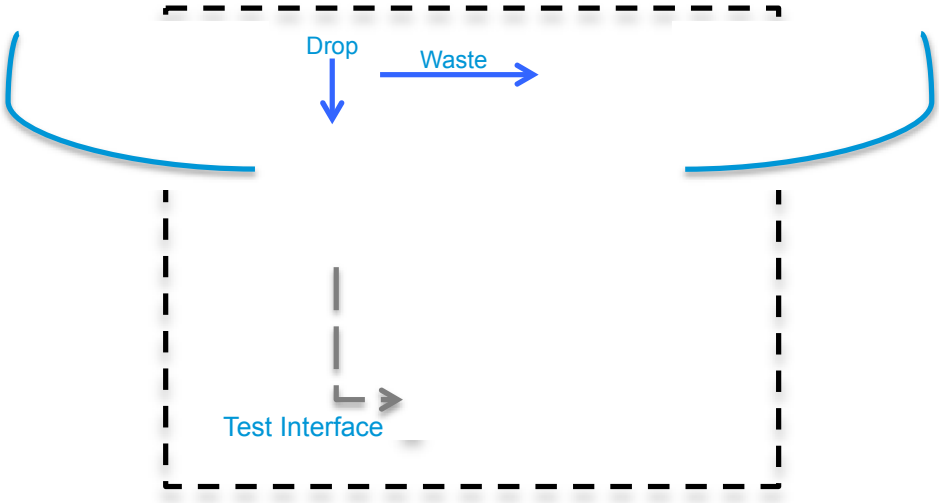


EDFA Selection is based on:
IF input to EDFA is >=-15.5dBm/ch
THEN EDFA-17
ELSEIF Pin <-15.5dBm/ch
THEN EDFA-24
Per Channel Launch Power shall not exceed 0dBm into the fiber

R2 Drop



R1 Add / Drop Sites



All Power levels are per channel
EDFAs are set in Constant Gain Mode
Power set via VOA on Interface shelf

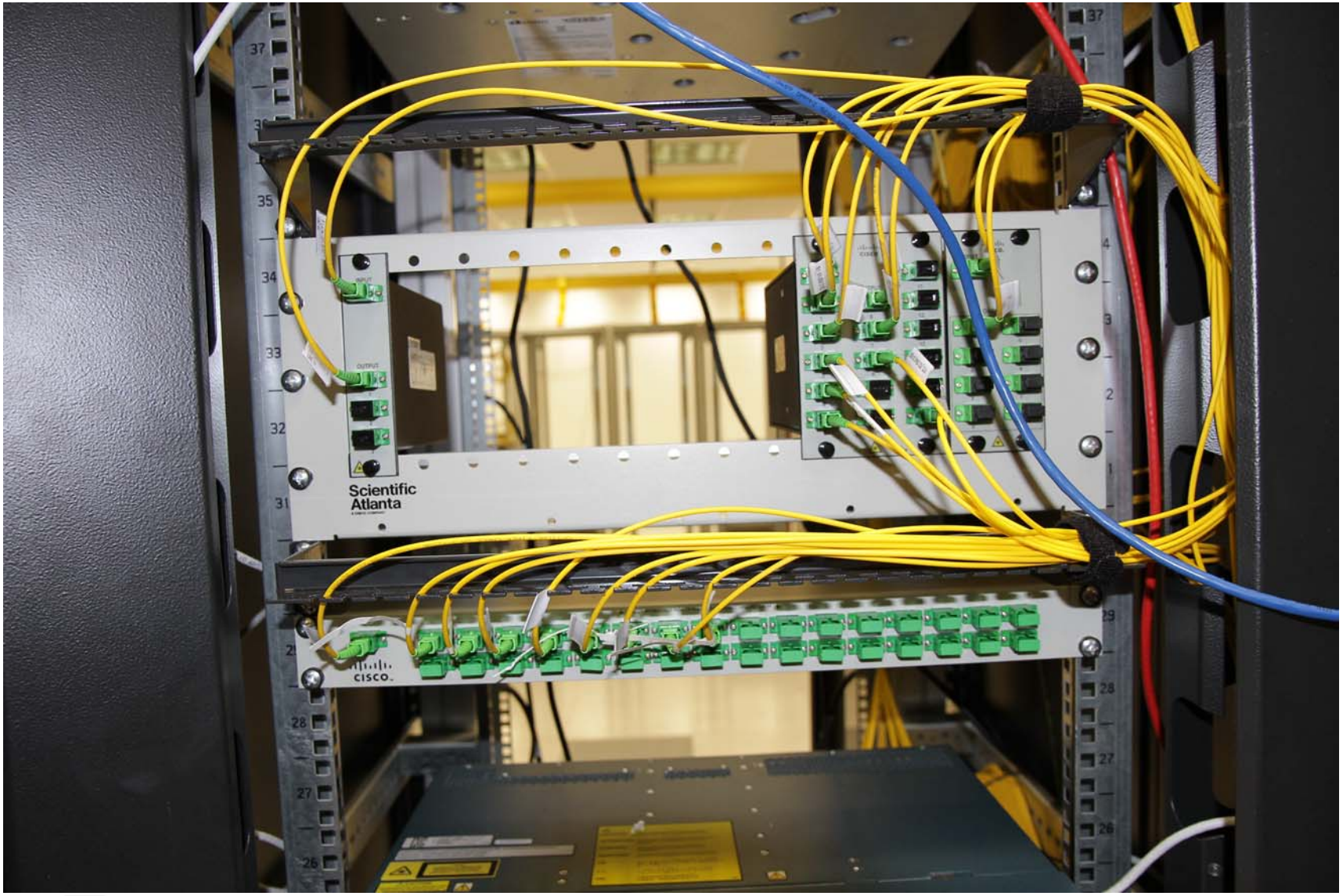
65/35 Splitter

80/20 Splitter

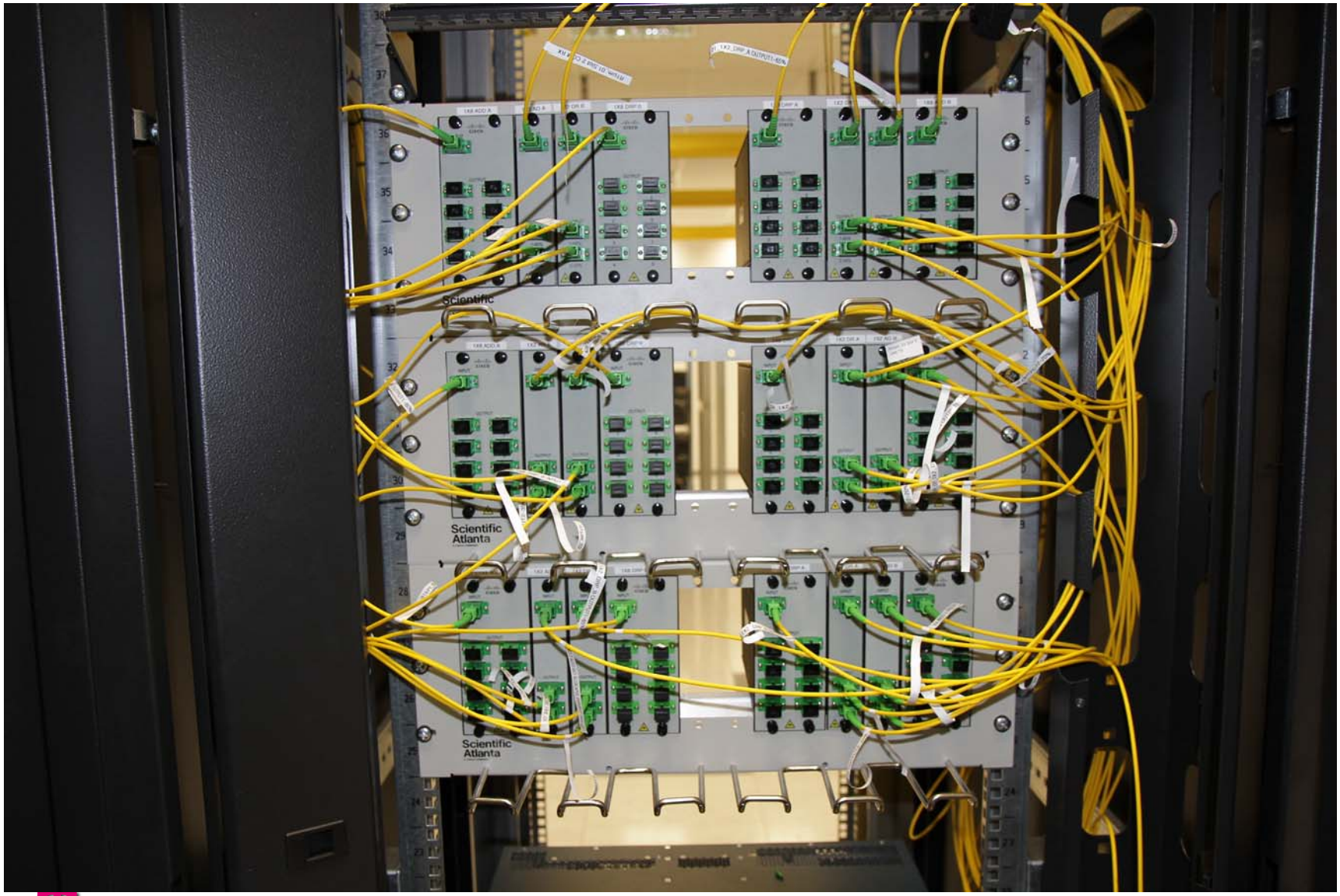
EDFA 17

EDFA 24

R2 SITE 32 DROPS



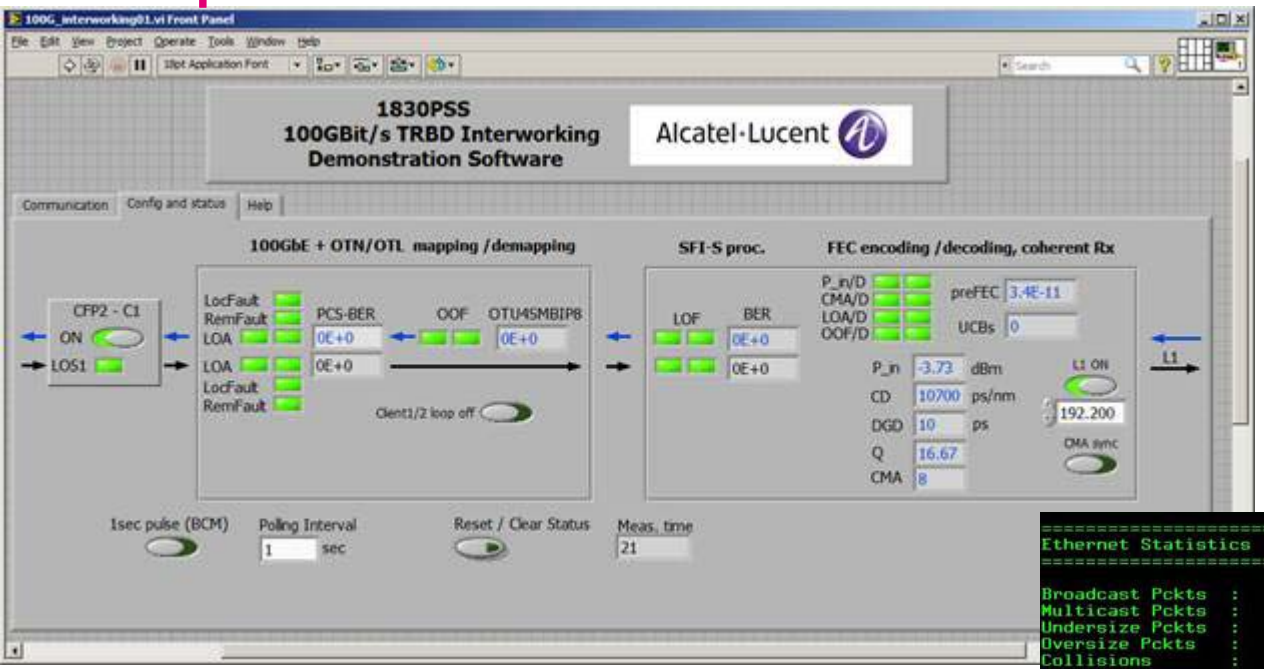
6 ADD DROP NODES (TO SIMULATE MORE SITES)




AUSTRIA
Graz
Klagenfurt
HUNGARY
Cakovec
Varazdin
Koprivnica
Ljubljana
Bjelovar
Virovitica
Zagreb
SERBIA
Osijek
Sisak
Karlovac
Vukovar
Slavonski Brod
Umag
Porec
Rovinj
Pula
Cres
Krk
Plitvice Lakes National Park
Losinj
BOSNIA & HERCEGOVINA
Zadar
Sarajevo
Ancona
Split
Mostar
Makarska
Brac
Medjugorje
Vis
Hvar
Neum Corridor
MONTENEGRO
Korcula
Dubrovnik
Mljet
Cavtat
ITALY
Pescara

FIRST 100G ETHERNET LINE SIDE INTEROPERABILITY

27 Sep 2013



ALU

Cisco

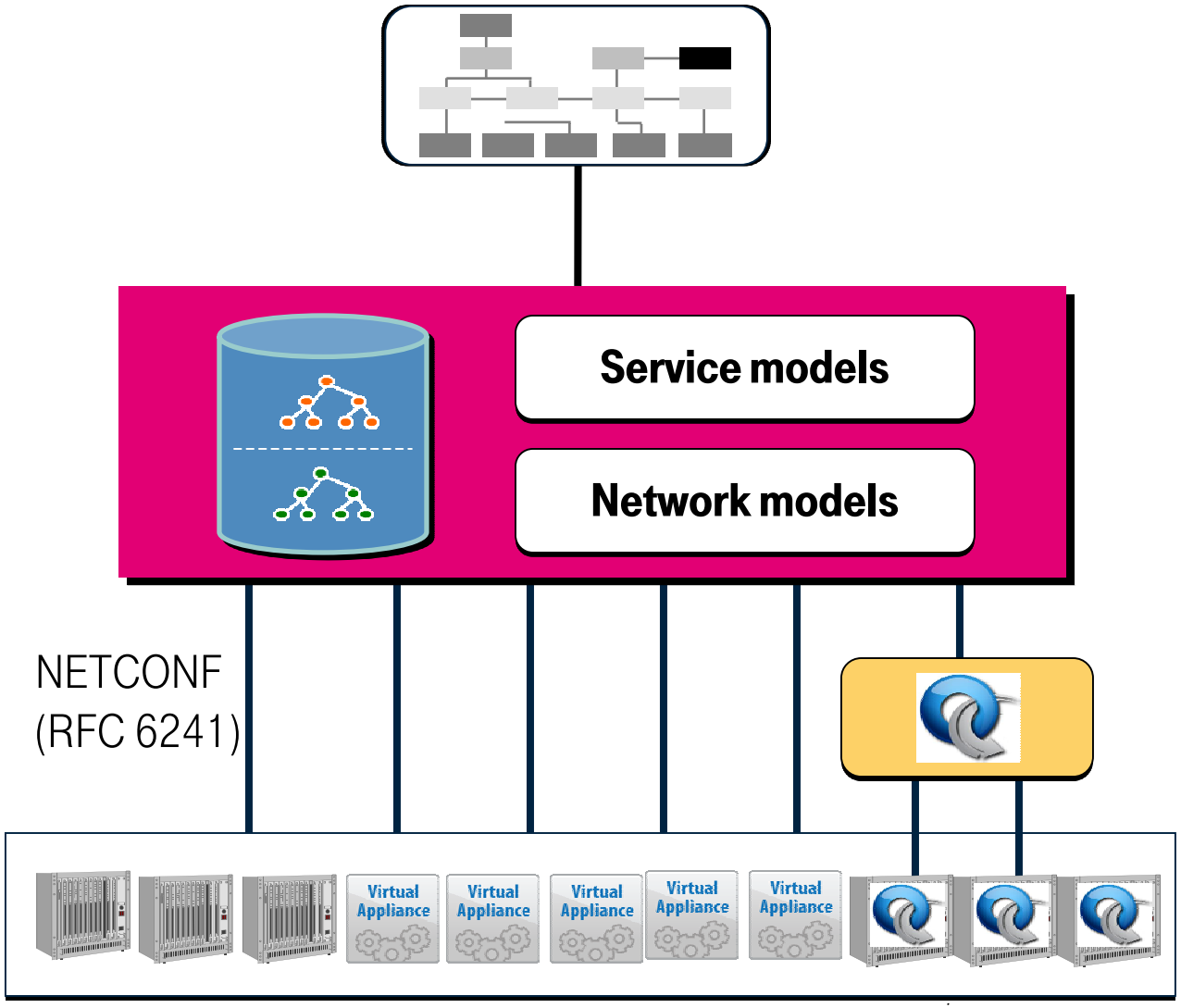
```
Ethernet Statistics
Broadcast Pckts  :          12   Drop Events       :          0
Multicast Pckts  :       19433   CRC/Align Errors  :          0
Undersize Pckts  :           0   Fragments         :          0
Oversize Pckts   :           0   Jabbers           :          0
Collisions       :           0

Octets                          :    1184145624
Packets                         :      10991227
Packets of 64 Octets            :         10146
Packets of 65 to 127 Octets     :      10581242
Packets of 128 to 255 Octets    :        272821
Packets of 256 to 511 Octets    :         39805
Packets of 512 to 1023 Octets   :         17826
Packets of 1024 to 1518 Octets  :         56233
Packets of 1519 or more Octets  :         13154

Port Statistics
                                        Input              Output
Unicast Packets                       4248405             6723377
Multicast Packets                        7291               12142
Broadcast Packets                           2                  10
Discards                                    0                   0
Unknown Proto Discards                      0

Ethernet-like Medium Statistics
Alignment Errors  :           0   Sngl Collisions   :          0
FCS Errors        :           0   Mult Collisions   :          0
SQE Test Errors   :           0   Late Collisions   :          0
CSE               :           0   Excess Collisns   :          0
Too long Frames   :           0   Int MAC Tx Errs   :          0
Symbol Errors     :           0   Int MAC Rx Errs   :          0
In Pause Frames   :           0   Out Pause Frames  :          0
```

http://www.stupi.se/Standards/100G-long-haul4.pdf

EVOLVING THE DWDM LAYER

- Packet, more and more is driving a wavelength
- Provide Physical Layer awareness to upper Layer
  - Packet Optical Integration does just that
- Colorless and Flex Spectrum become key
  - Ability to turn on any wave at anytime
  - Pushing Shannon's Limit in 50GHz window
- Integrated Channel Selection
  - Coherent Rx provides 6.25GHz granularity of Channel Tuning
- Increase Bit/Hz efficiency and decrease cost
  - Leverage Advanced Multi Layer Modulation
- Next step 400G/Multirate, 1.6T/Multirate

TERASTREAM IS DT'S FIRST SDN



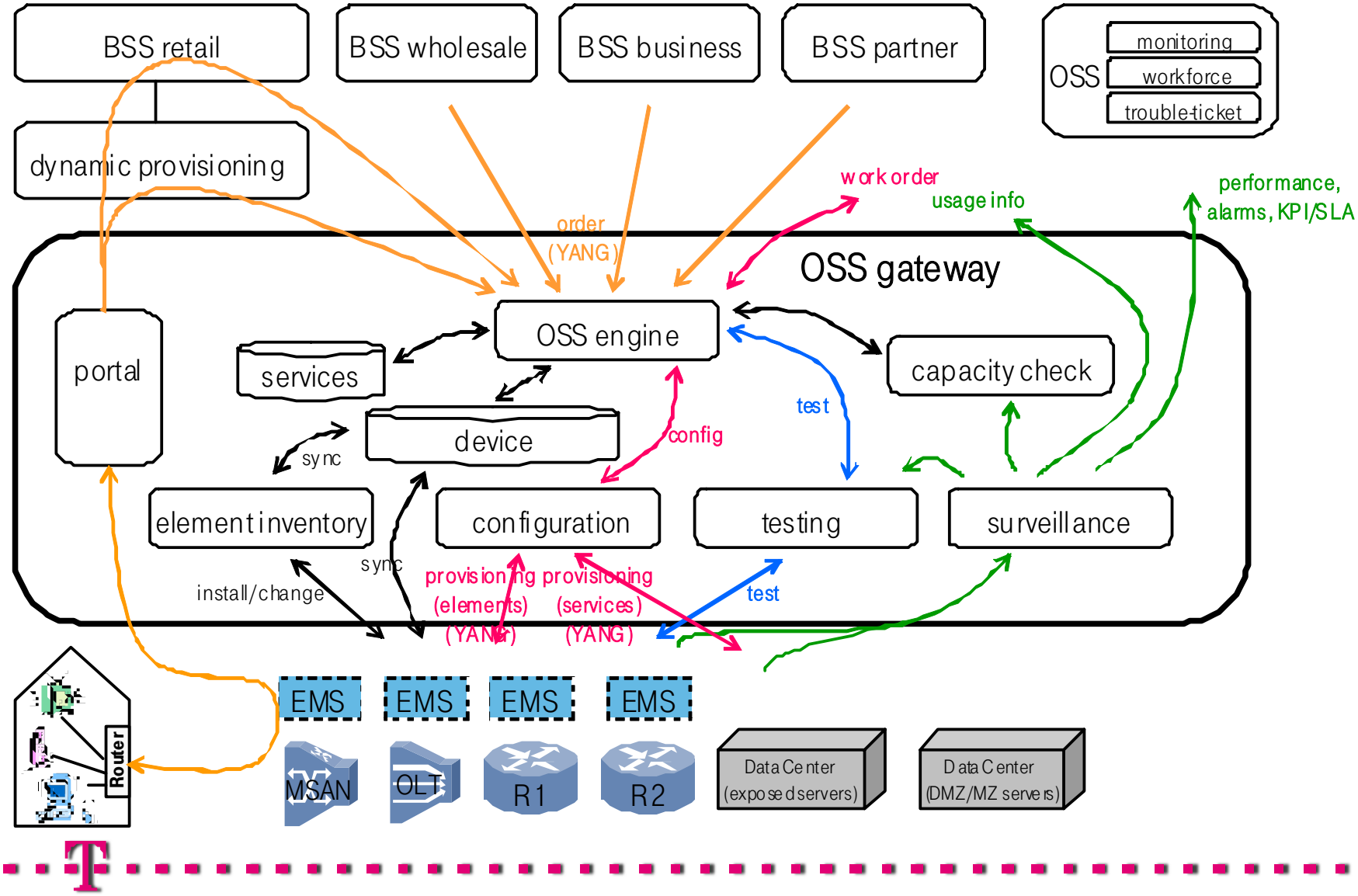
Management Applications

Real-time OSS

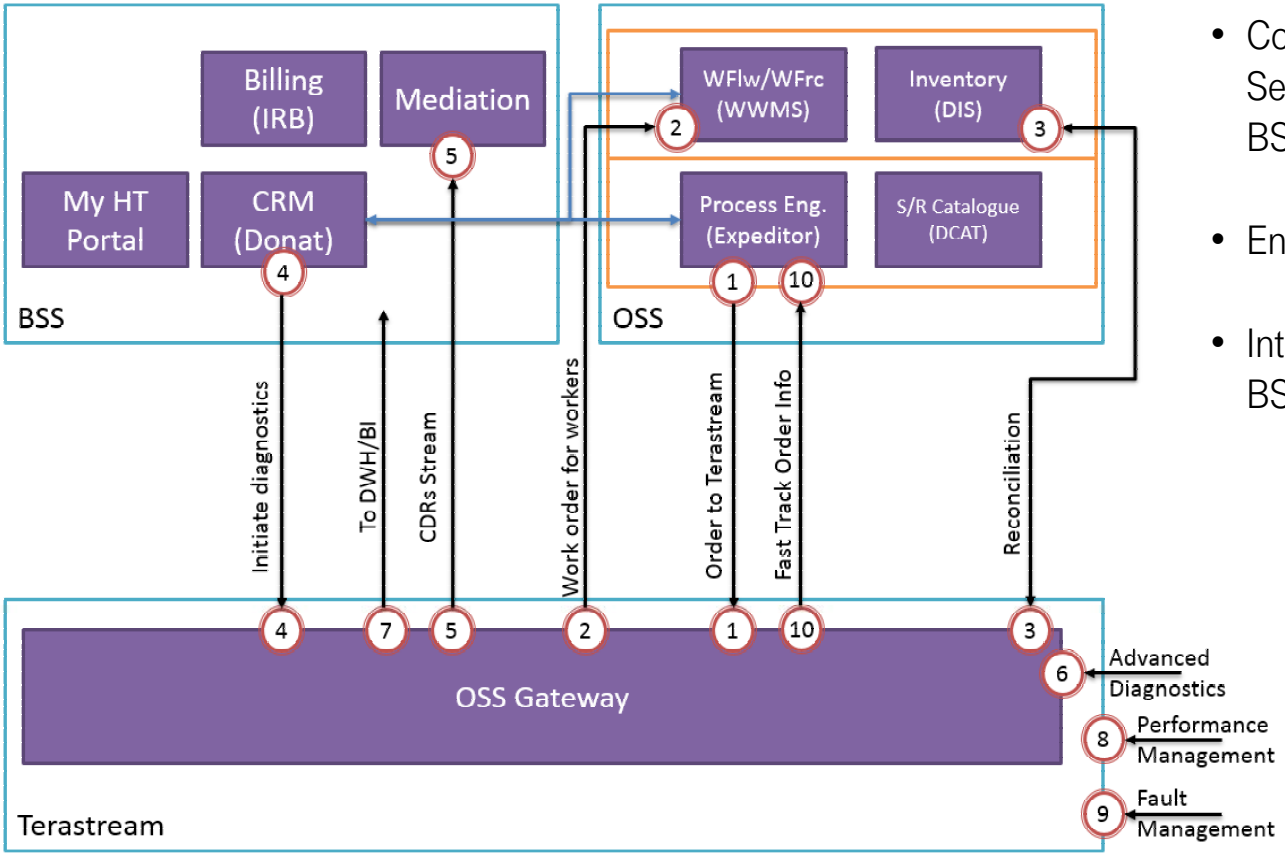
YANG (RFC 6020)

Multi-Vendor Multi-Technology

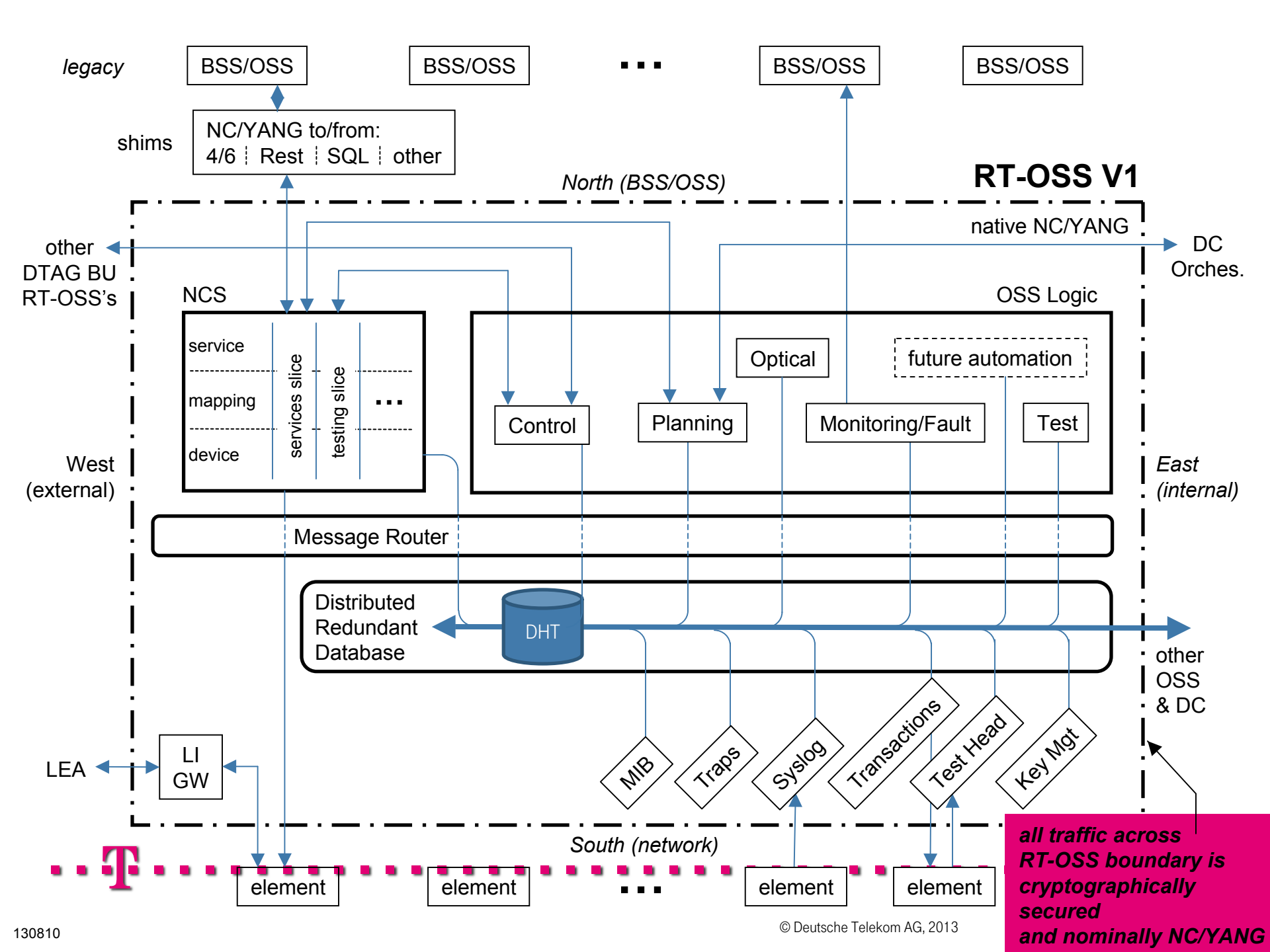
TERASTREAM OSS “GATEWAY”



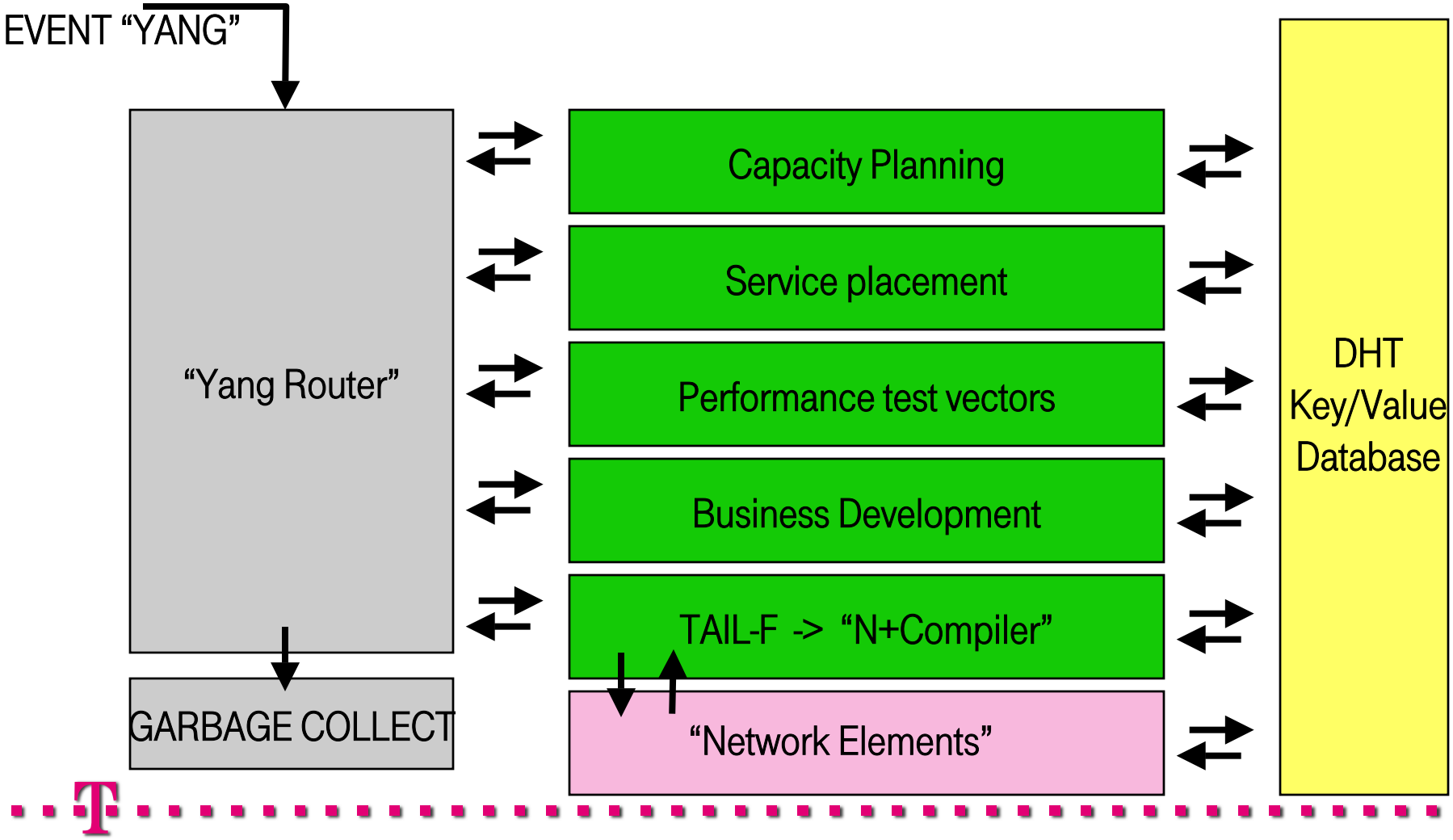
CONNECTING TO LEGACY BSS/OSS



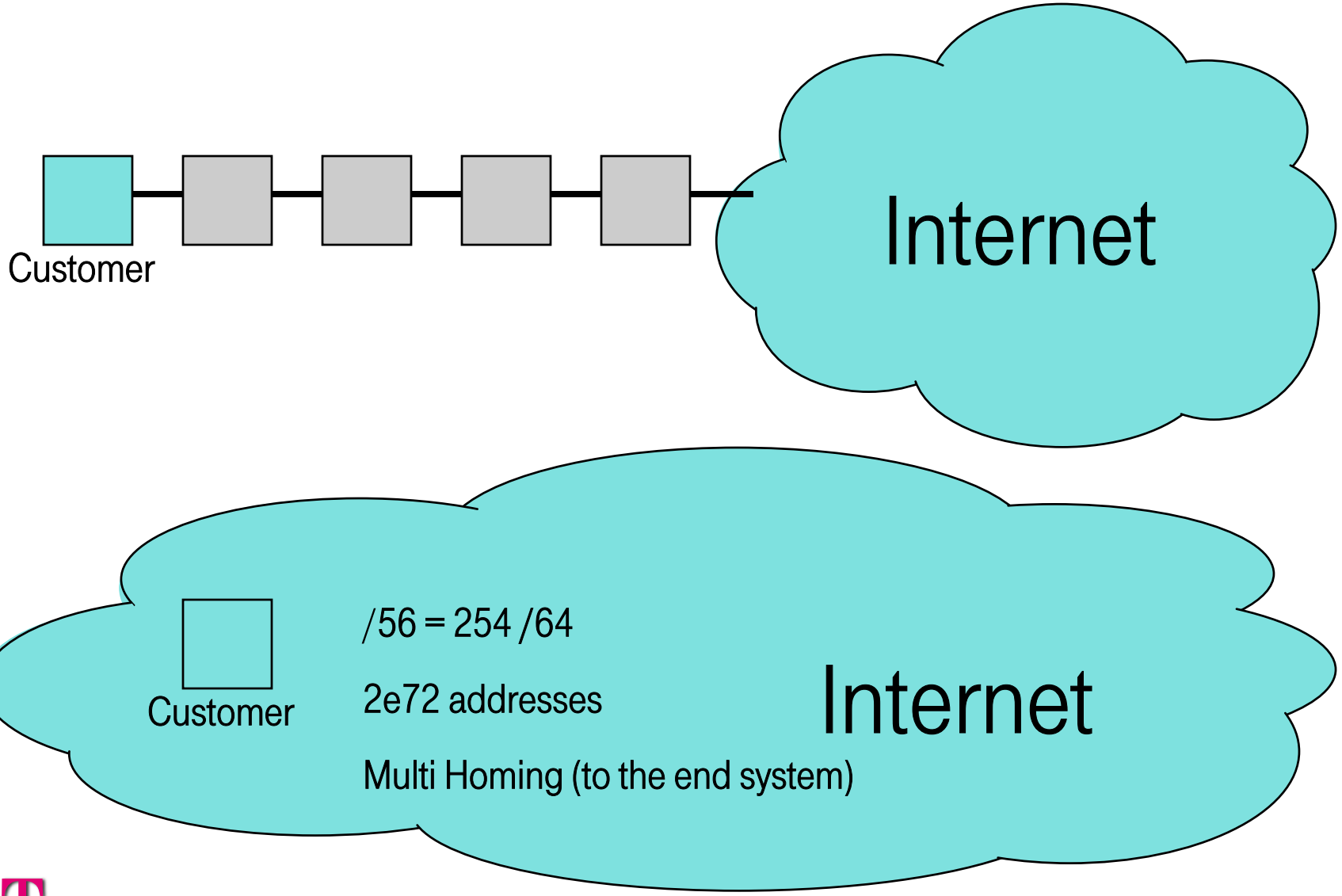
- Connecting Terastream as a new Service Delivery Platform to existing BSS/OSS systems
- Enable existing processes
- Interfaces btw Terastream – BSS/OSS:
  1. Order management
  2. Work orders
  3. Inventory reconciliation
  4. Initiate diagnostics
  5. CDRs Stream
  6. Advanced diagnostics
  7. DWH/BI
  8. Performance Management
  9. Fault Management
  10. Fast Track Order Info



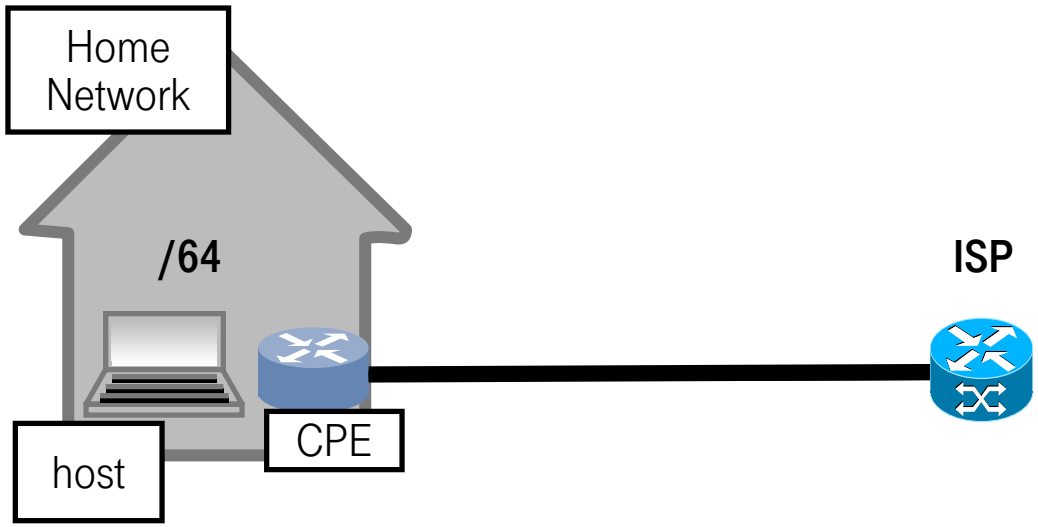
Wait for response from lowest priority required



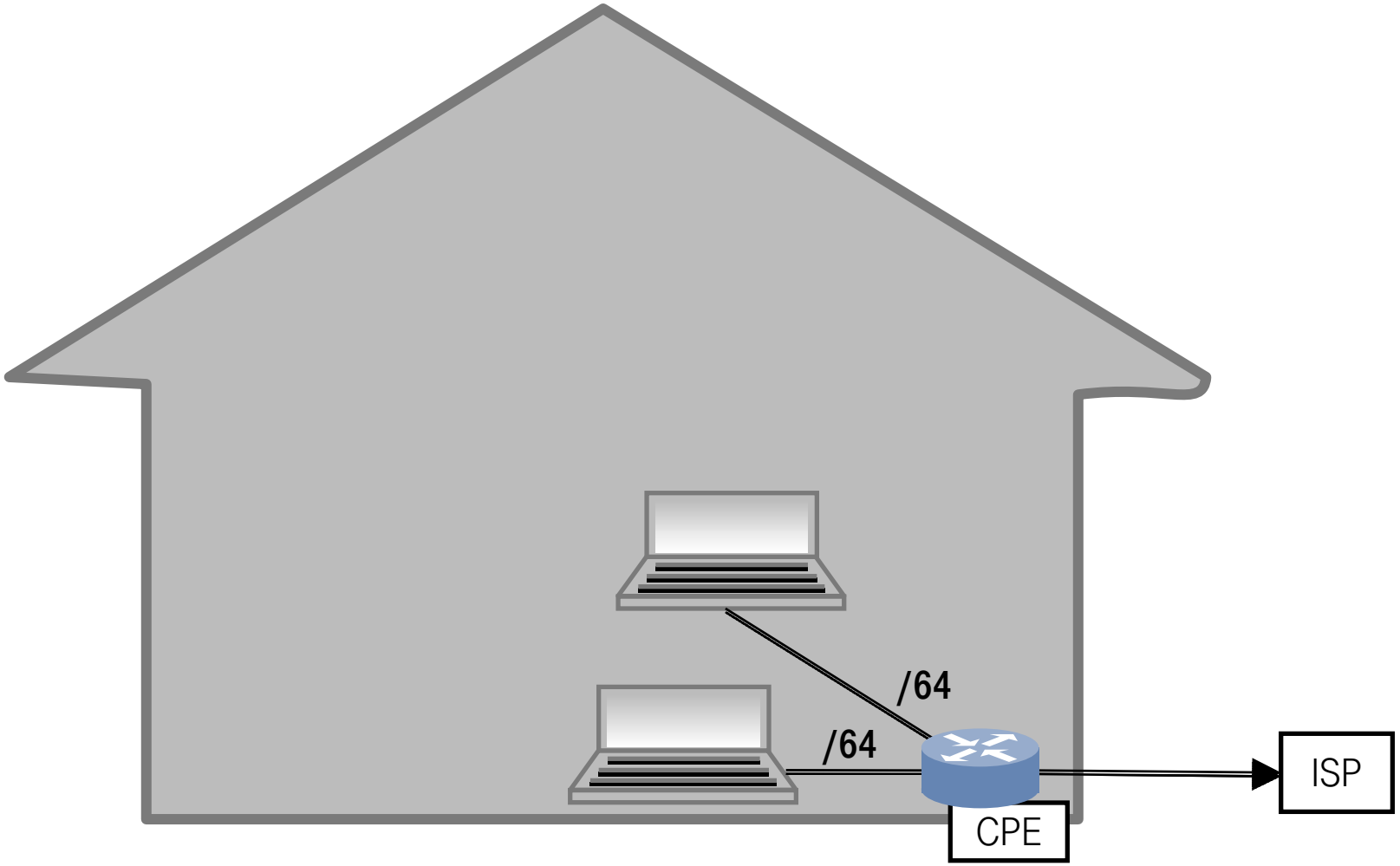
CHANGING THE BROADBAND PARADIGM



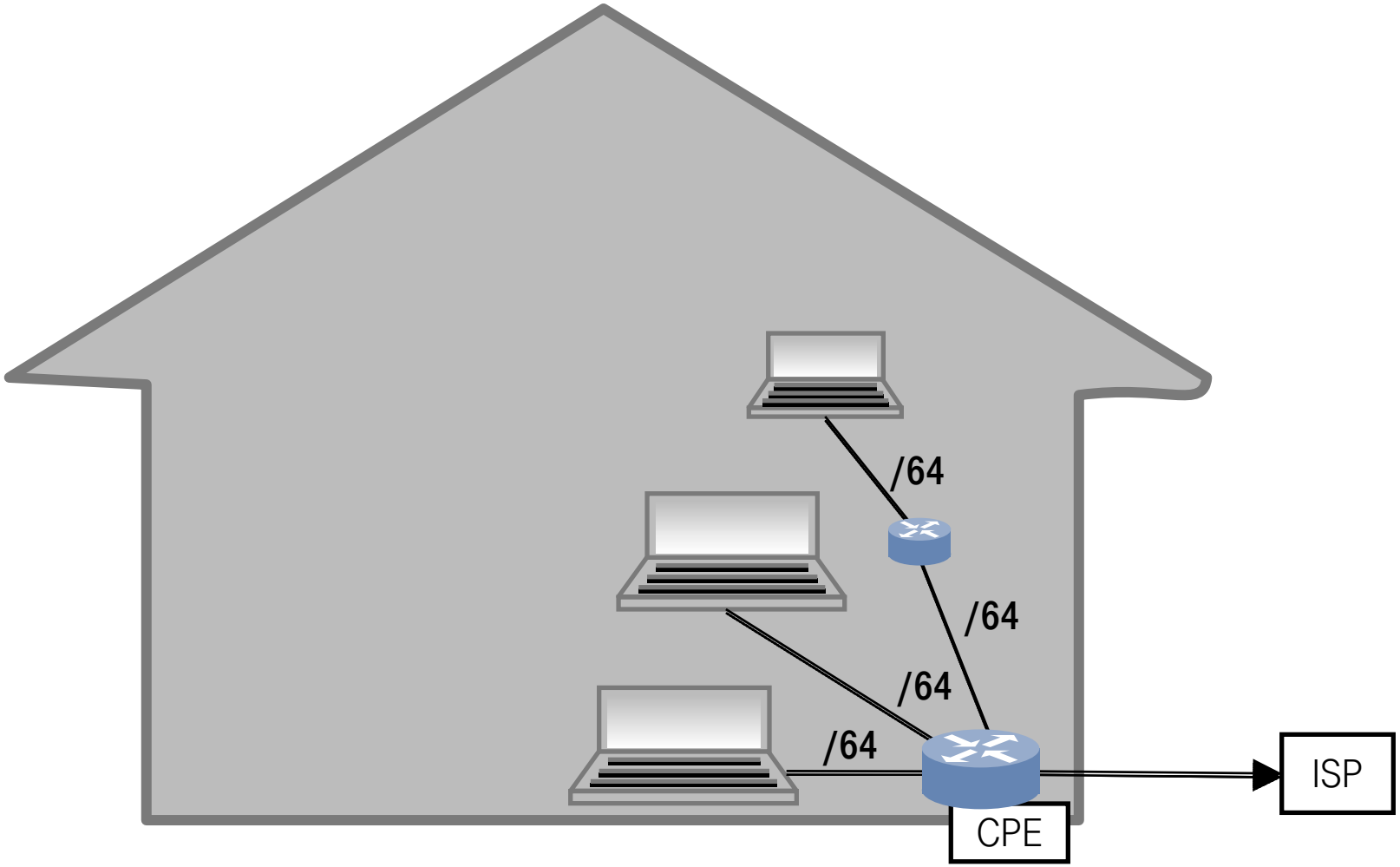
USUAL HOME NETWORK DRAWINGS



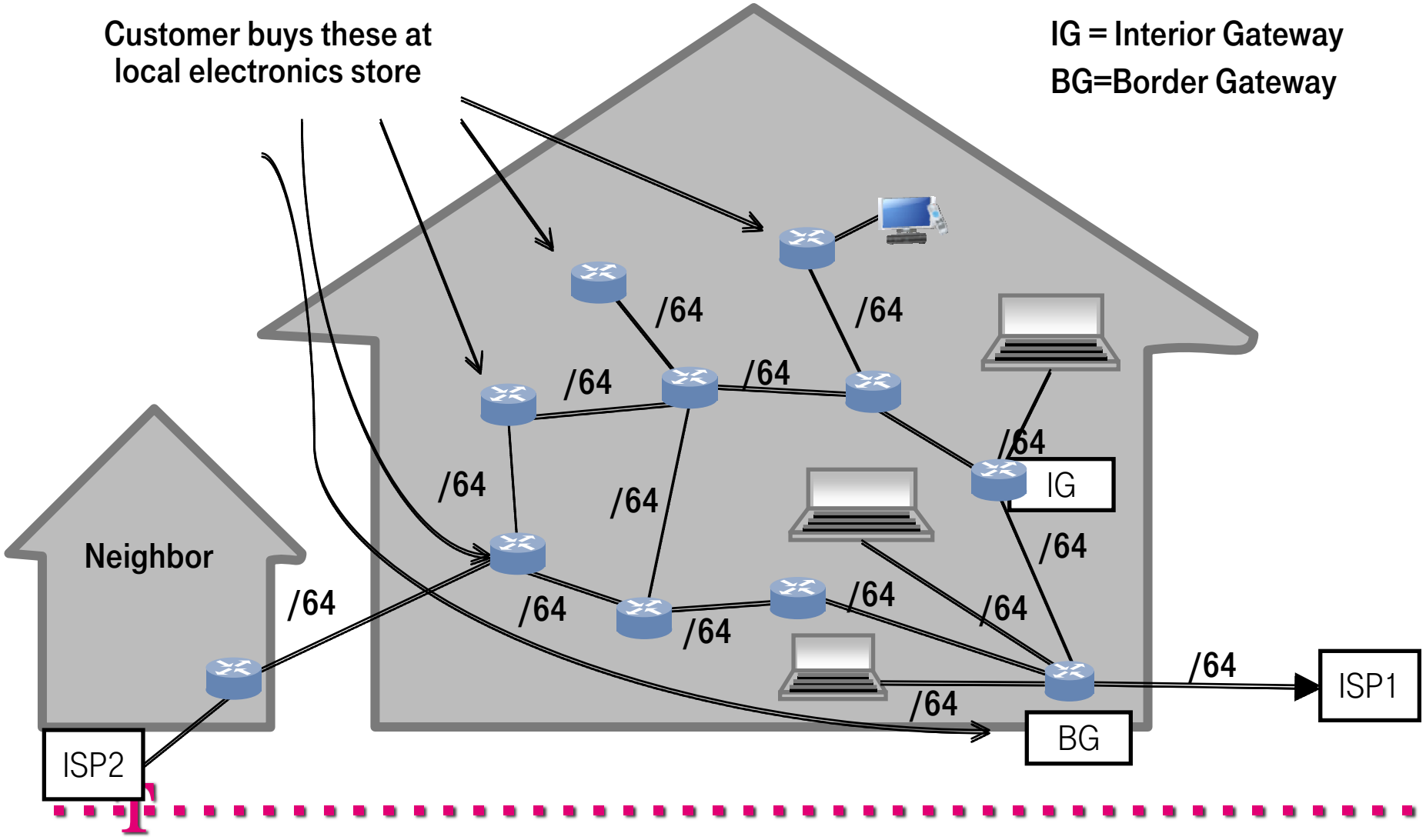
SOMETIMES LIKE THIS



PERHAPS EVEN PREFIX DELEGATION!



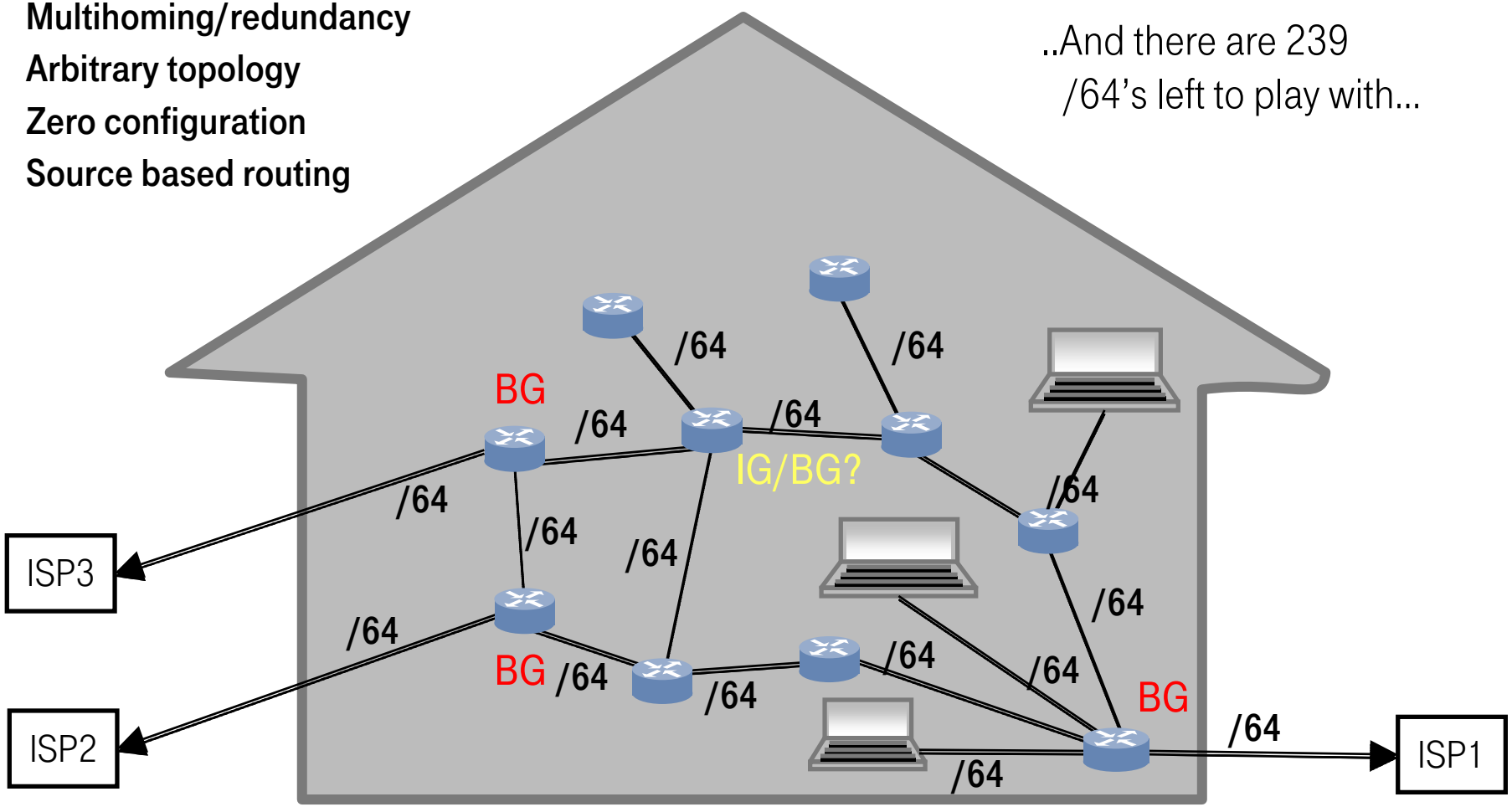
WE'RE AIMING TO HANDLE THIS



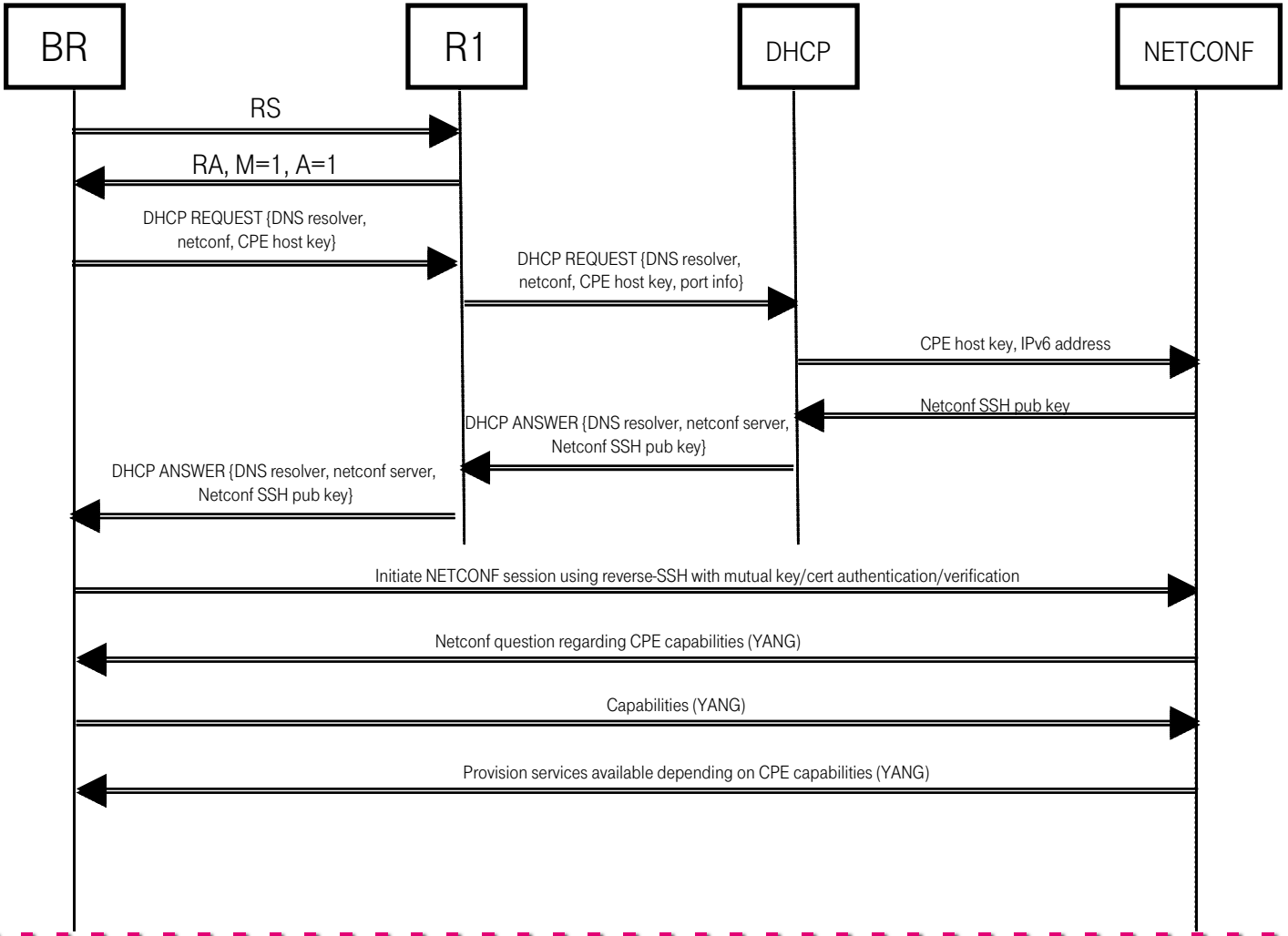
… AND THIS

Multihoming/redundancy
Arbitrary topology
Zero configuration
Source based routing

..And there are 239
/64's left to play with…



BOOTSTRAP PROCESS BR



QUESTIONS?

Now you can bring out your tar and feathers and start throwing things at me..

THANKS!