

The Road towards Optical Networking

www.science.uva.nl/~deLaat

Cees de Laat

EU

SURFnet

University of Amsterdam

SARA
NIKHEF
UCL

What is this buzz about optical networking

(2 of 15)

- **Networks are already optical for ages**
- **Users won't see the light**
- **Almost all current projects are about SONET circuits and Ethernet (old wine in new bags?)**
- **Are we going back to the telecom world, do NRN's want to become telco's**
- **Does it scale**
- **Is it all about speed or is it integrated services**

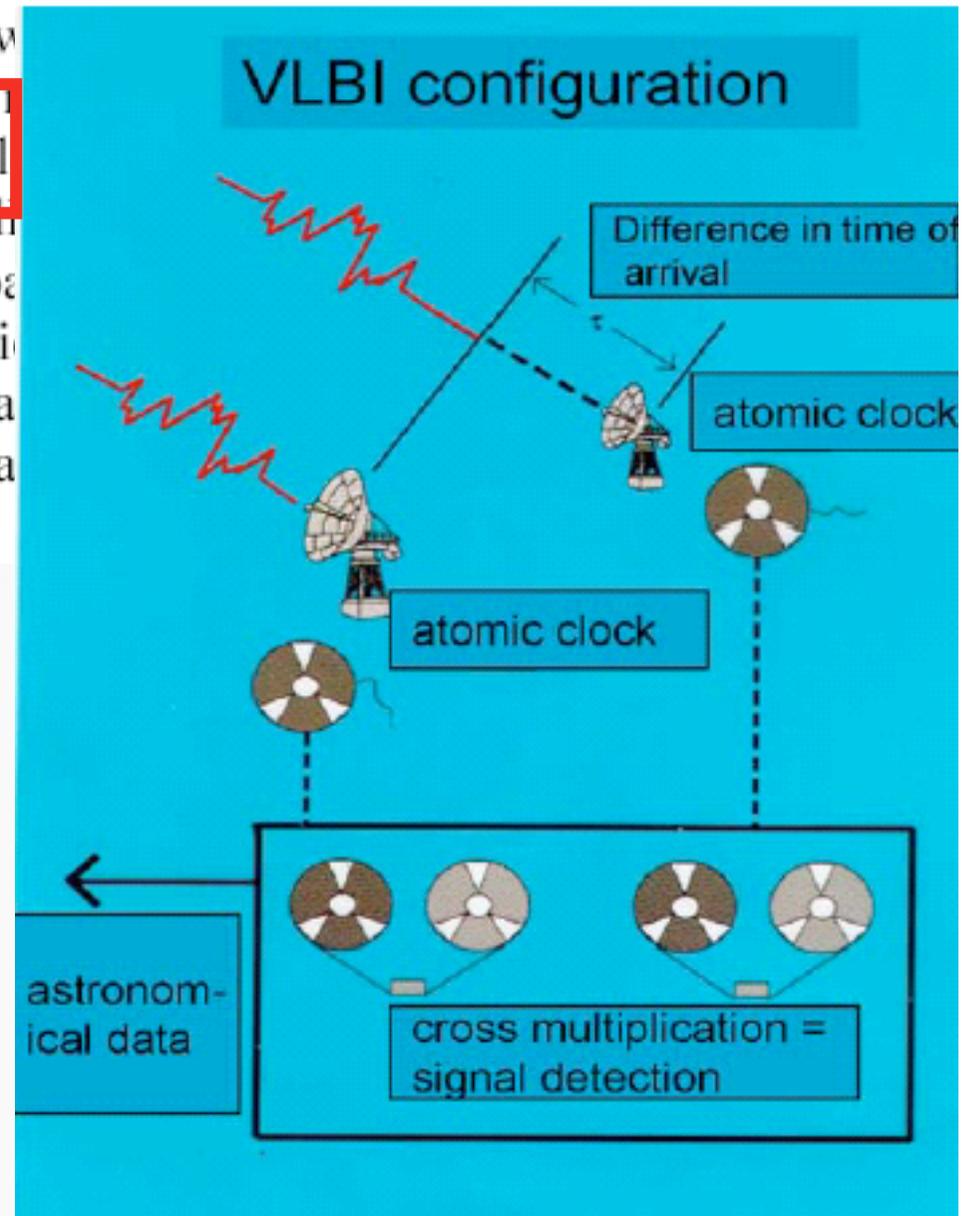
VLBI

(3 of 15)

VLBI is easily capable of generating many Gb of data per second. The sensitivity of the VLBI array scales with the square root of the bandwidth (data-rate) and there is a strong push to increase the data rate. Rates of 8Gb/s or more are entirely feasible with current technology. It is expected that parallel processing will remain the most efficient approach. Distributed processing may have an application in VLBI and multi-gigabit data streams will aggregate into larger data streams and the capacity of the final link to the data center.

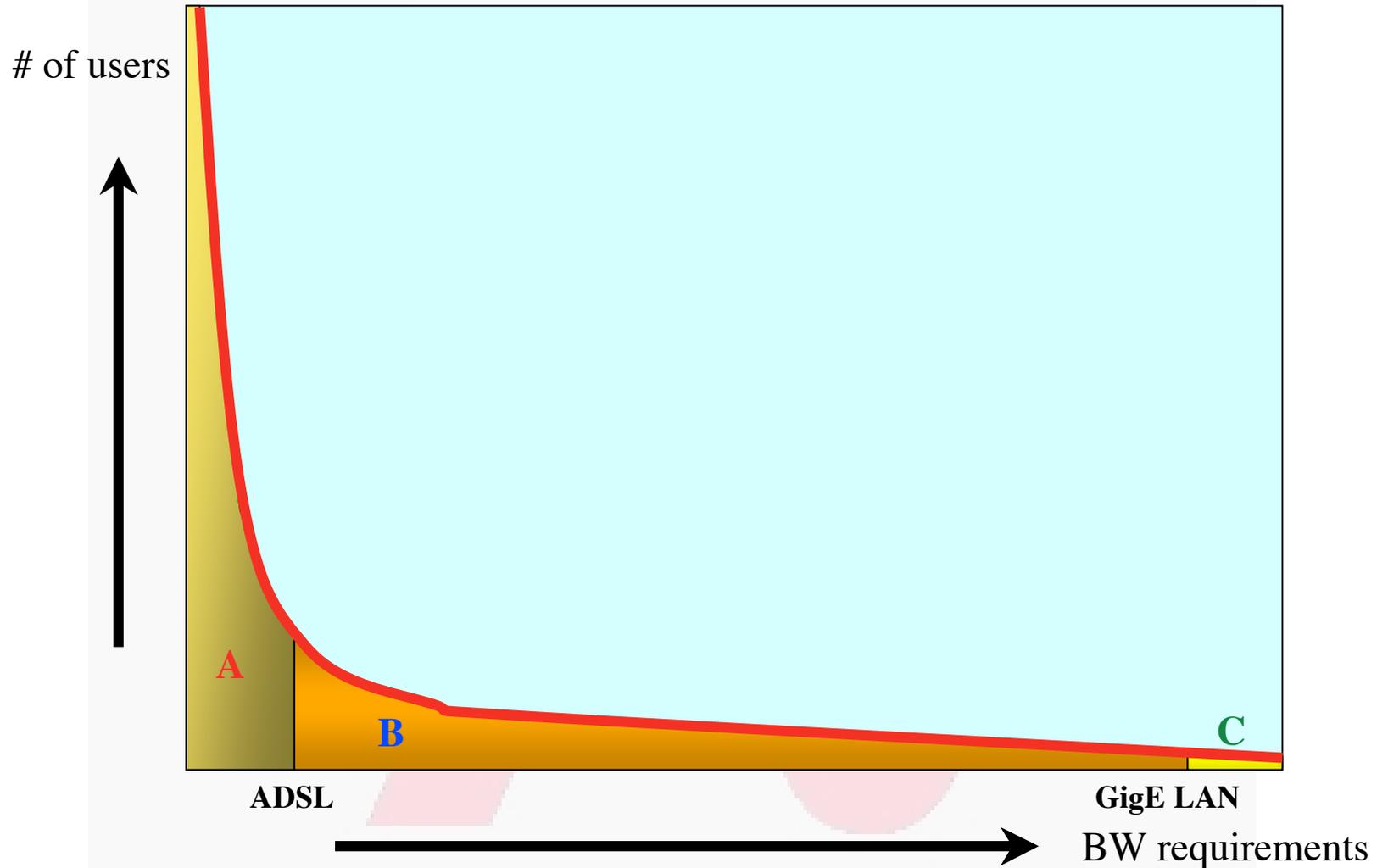


Westerbork Synthesis Radio Telescope - Netherlands



Know the user

(4 of 15)



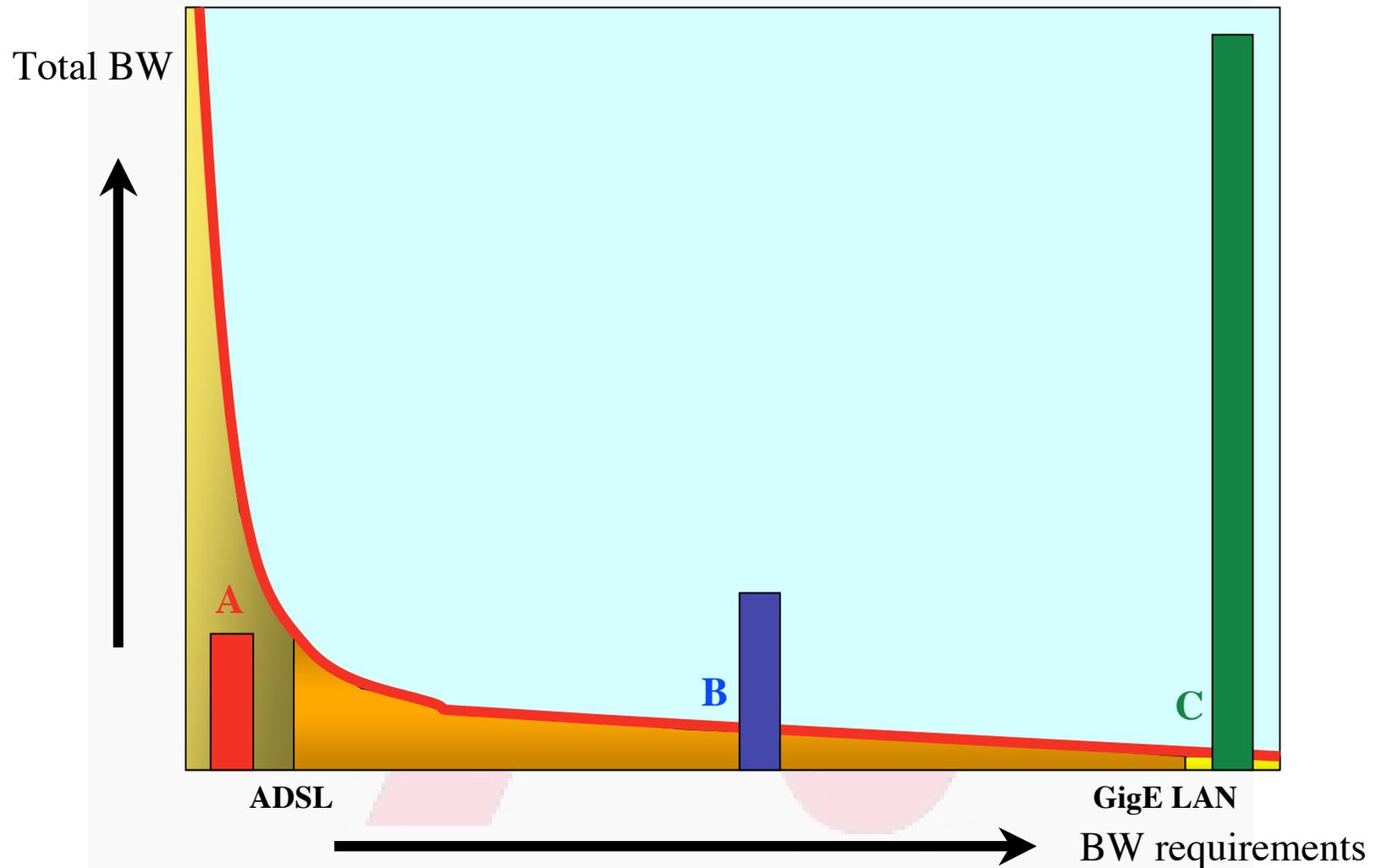
A -> Lightweight users, browsing, mailing, home use

B -> Business applications, multicast, streaming, VPN's, mostly LAN

C -> Special scientific applications, computing, data grids, virtual-presence

What the user

(4a of 15)



A -> Lightweight users, browsing, mailing, home use

B -> Business applications, multicast, streaming

C -> Special scientific applications, computing, data grids, virtual-presence

So, what's up doc

Suppose:

- **Optical components get cheaper and cheaper**
- **Dark (well, dark?) fibers abundant**
- **Number of available λ /user $\rightarrow \infty$**
- **Speeds of 10, 100, 1000 Gbit/s make electrical domain packet handling physically difficult**
 - 150 bytes @ 40 Gbit/s = 30 ns = 15 meter fiber
 - QoS makes no sense at these speeds
- **Cost per packet forwarding lower at L1 / L2**

Then:

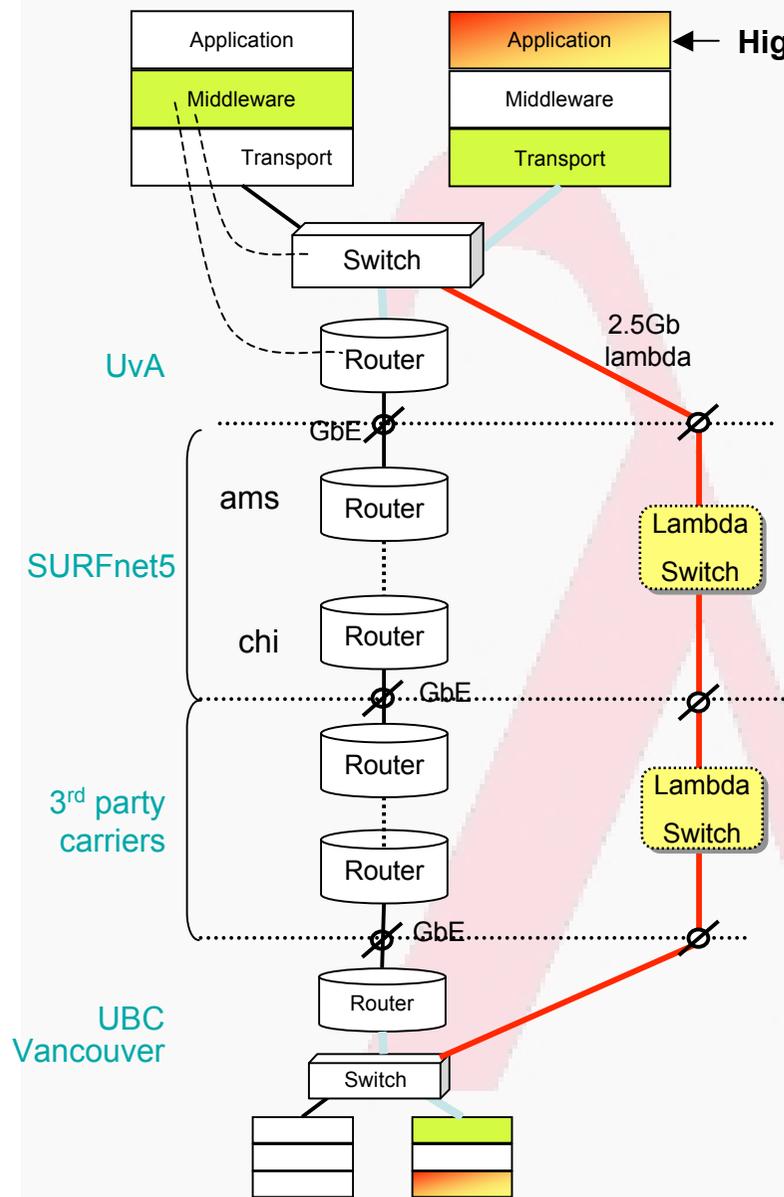
- **Long term view ---> full optical**
- **λ provisioning for grid applications**
- **How low can you go**

Optical networking, 3 scenarios

- **Lambdas for internal ISP bandwidth provisioning**
 - An ISP uses a lambda switching network to make better use of its (suppliers) dark fibers and to provision to the POP's. In this case the optical network is just within one domain and as such is a relatively simple case.
- **Lambda switching as peering point technology**
 - In this use case a layer 1 Internet exchange is build. ISP's peer by instantiating lambdas to each other. Is a $N*(N-1)$ and multi domain management problem.
- **Lambda switching as grid application bandwidth provisioning**
 - This is by far the most difficult since it needs UNI and NNI protocols to provision the optical paths through different domains.

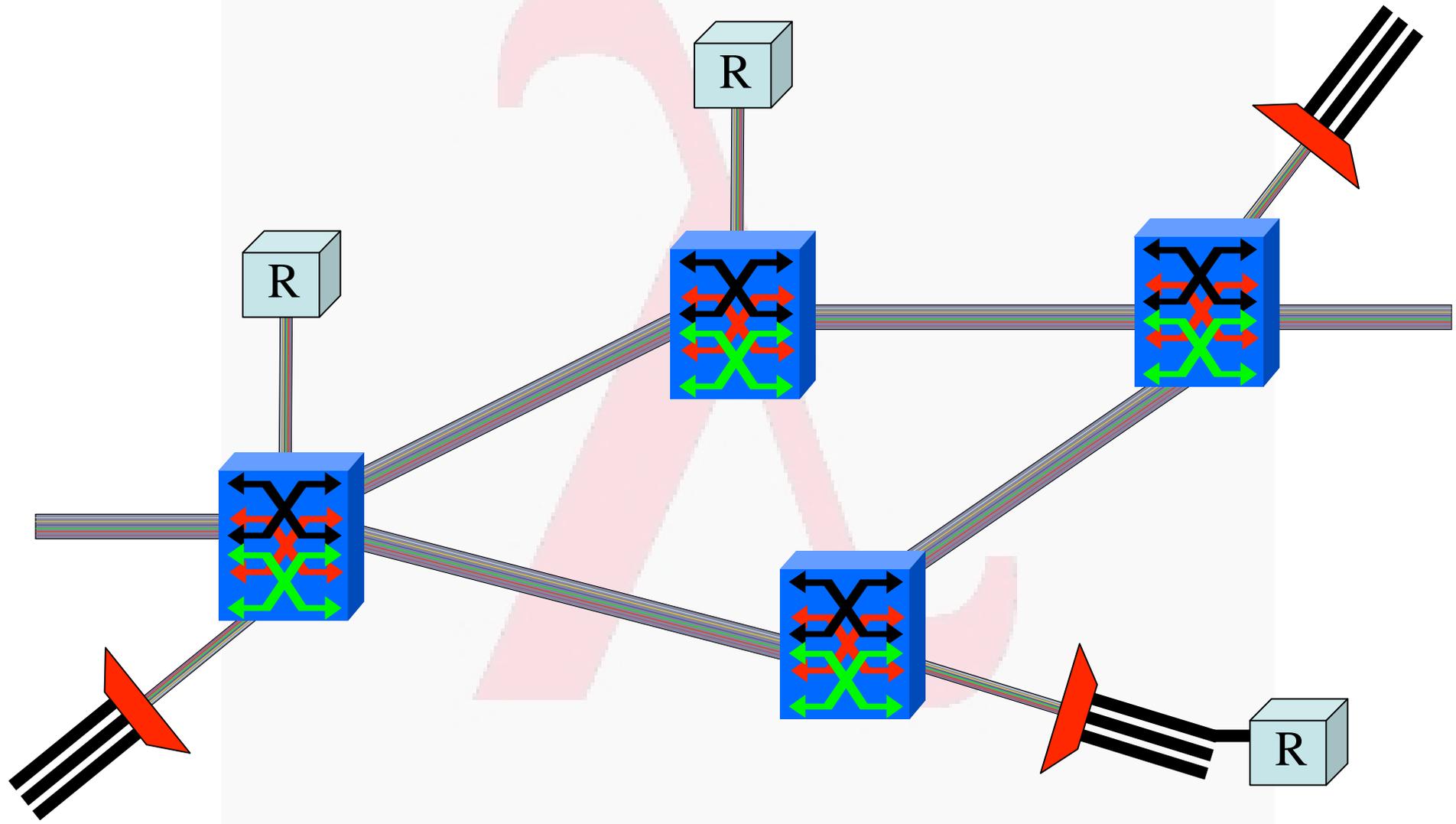
Current technology + (re)definition

- Current (to me) available technology consists of SONET/SDH switches which transport Ethernet
- DWDM+switching coming up
- Starlight and CANET use for the time being VLAN's on Ethernet switches to connect [(not)exactly] two ports
- So redefine a λ as:
 - “a λ is a pipe where you can inspect packets as they enter and when they exit, but principally not when in transit. In transit one only deals with the parameters of the pipe: number, color, bandwidth”



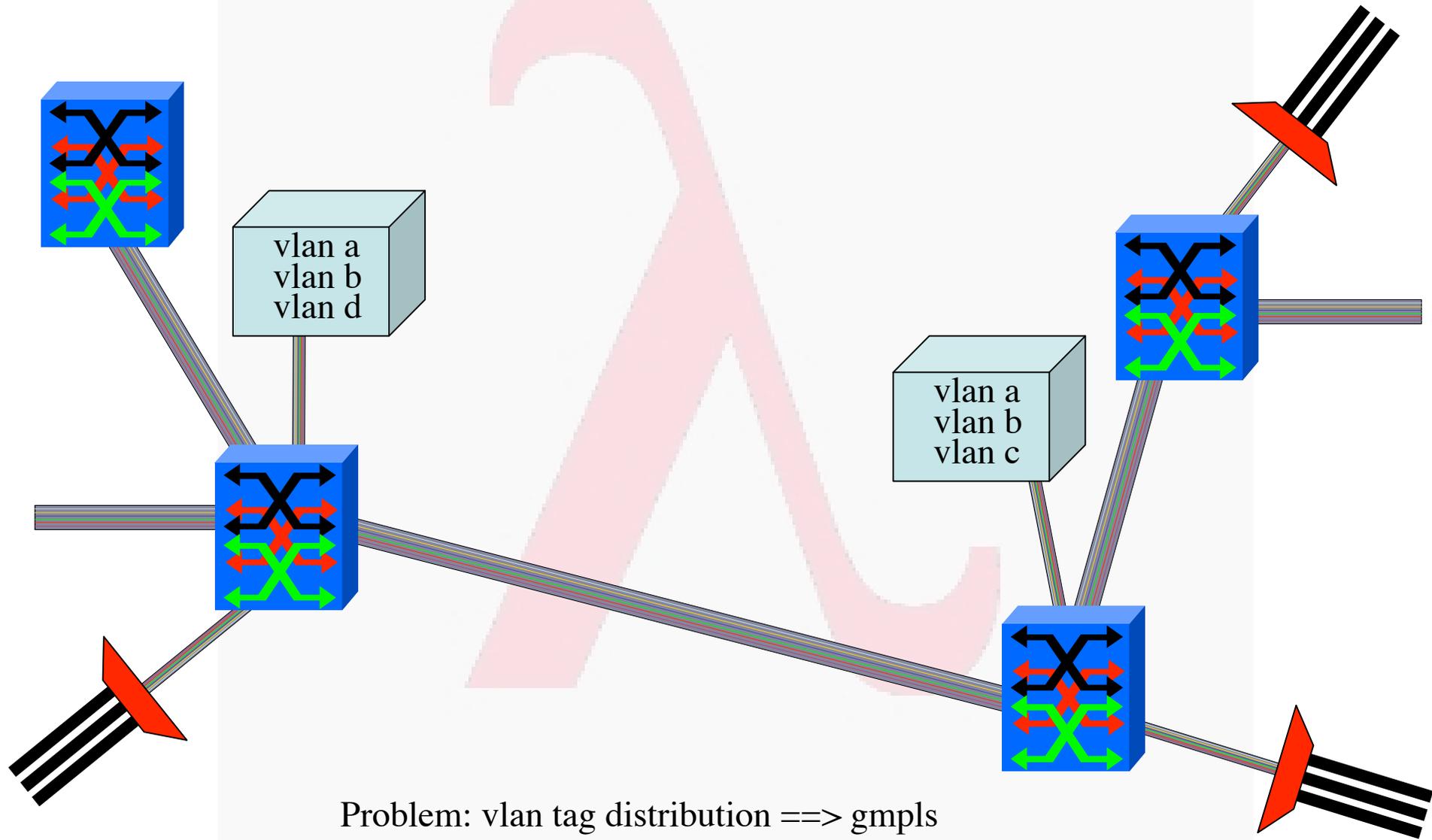
- lambda for high bandwidth applications
 - Bypass of production network
 - Middleware may request (optical) pipe
- RATIONALE:
 - Lower the cost of transport per packet

Architectures - L1 - 3



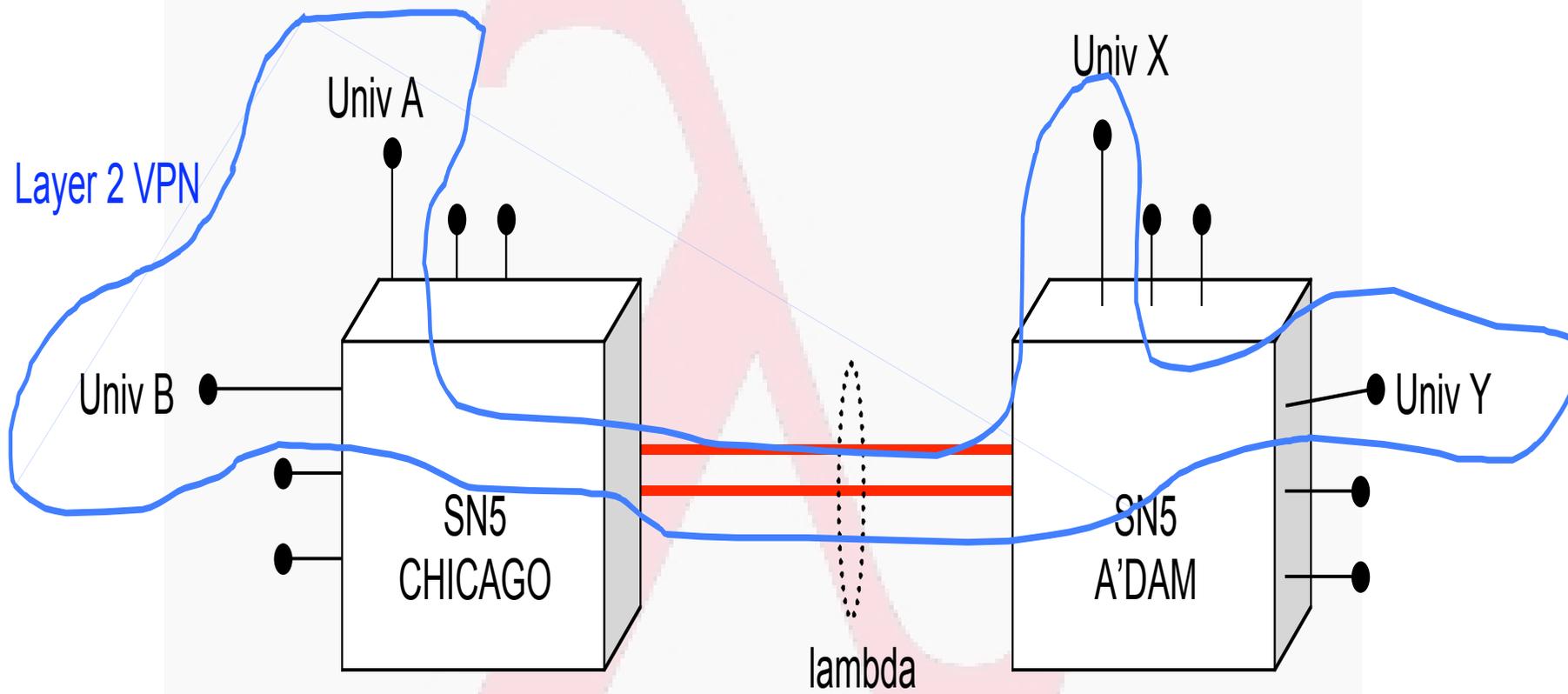
Other architectures - Distributed (10 of 15)

virtual IEX'es

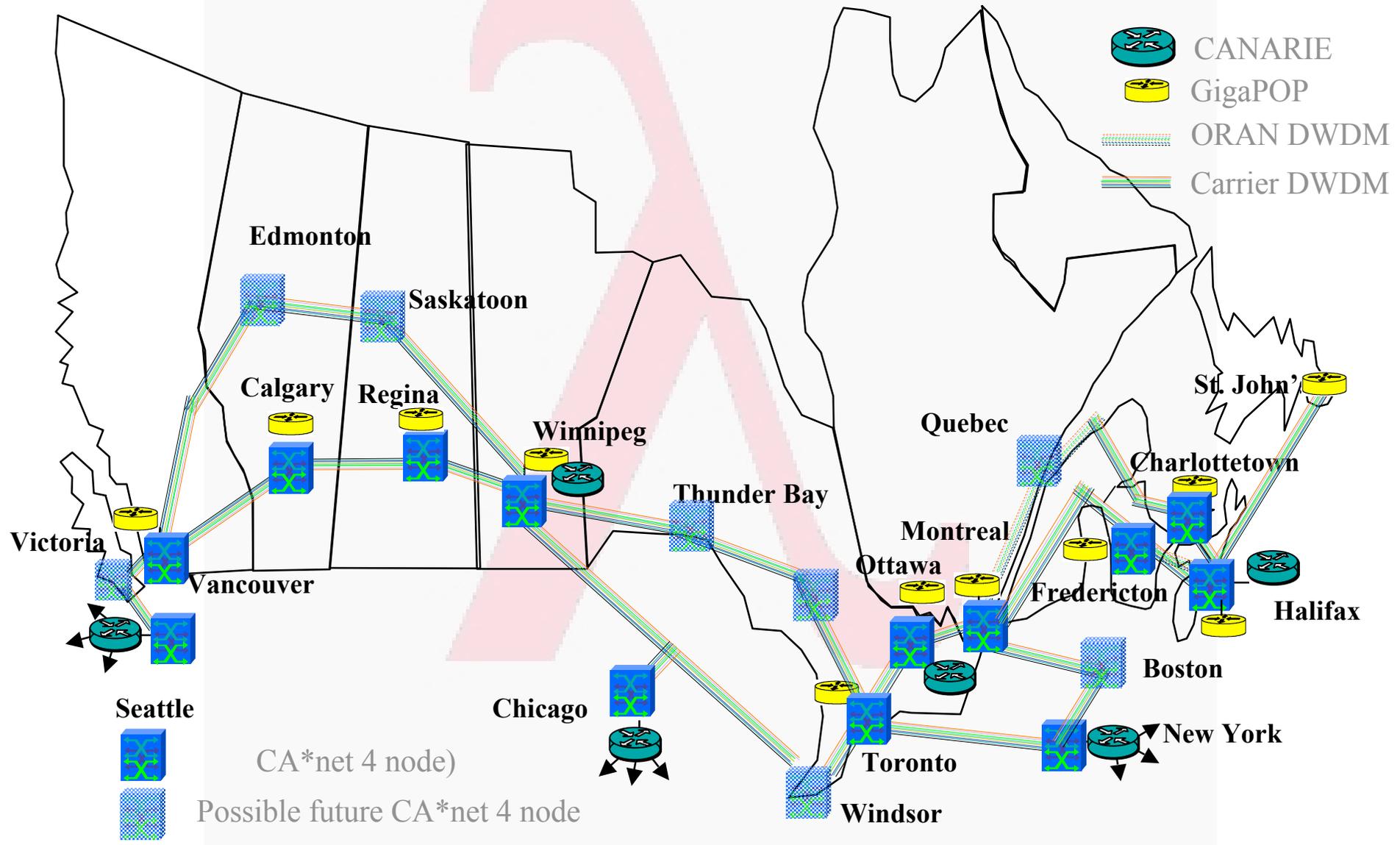


Problem: vlan tag distribution ==> gmpls

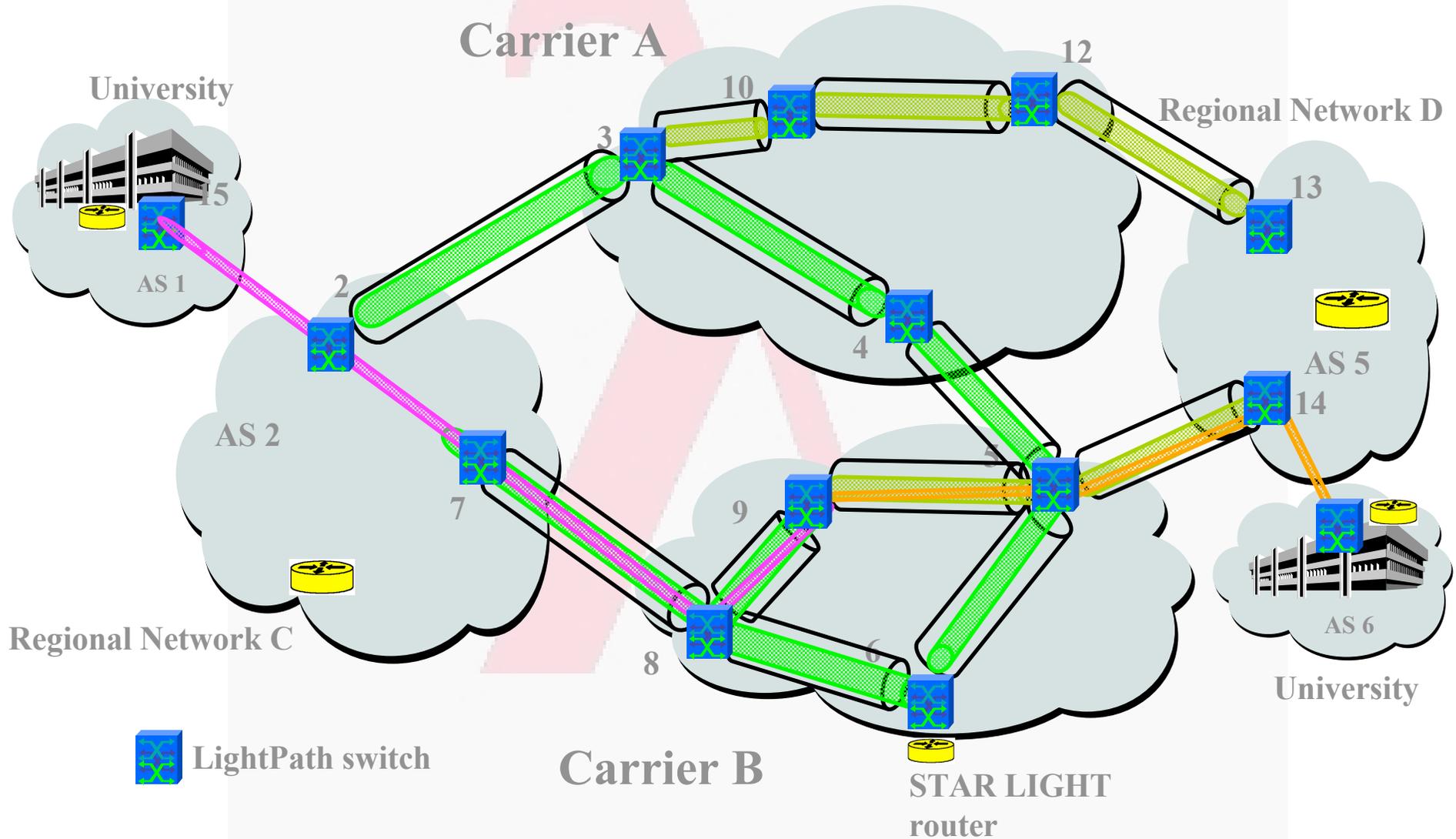
Distributed L2



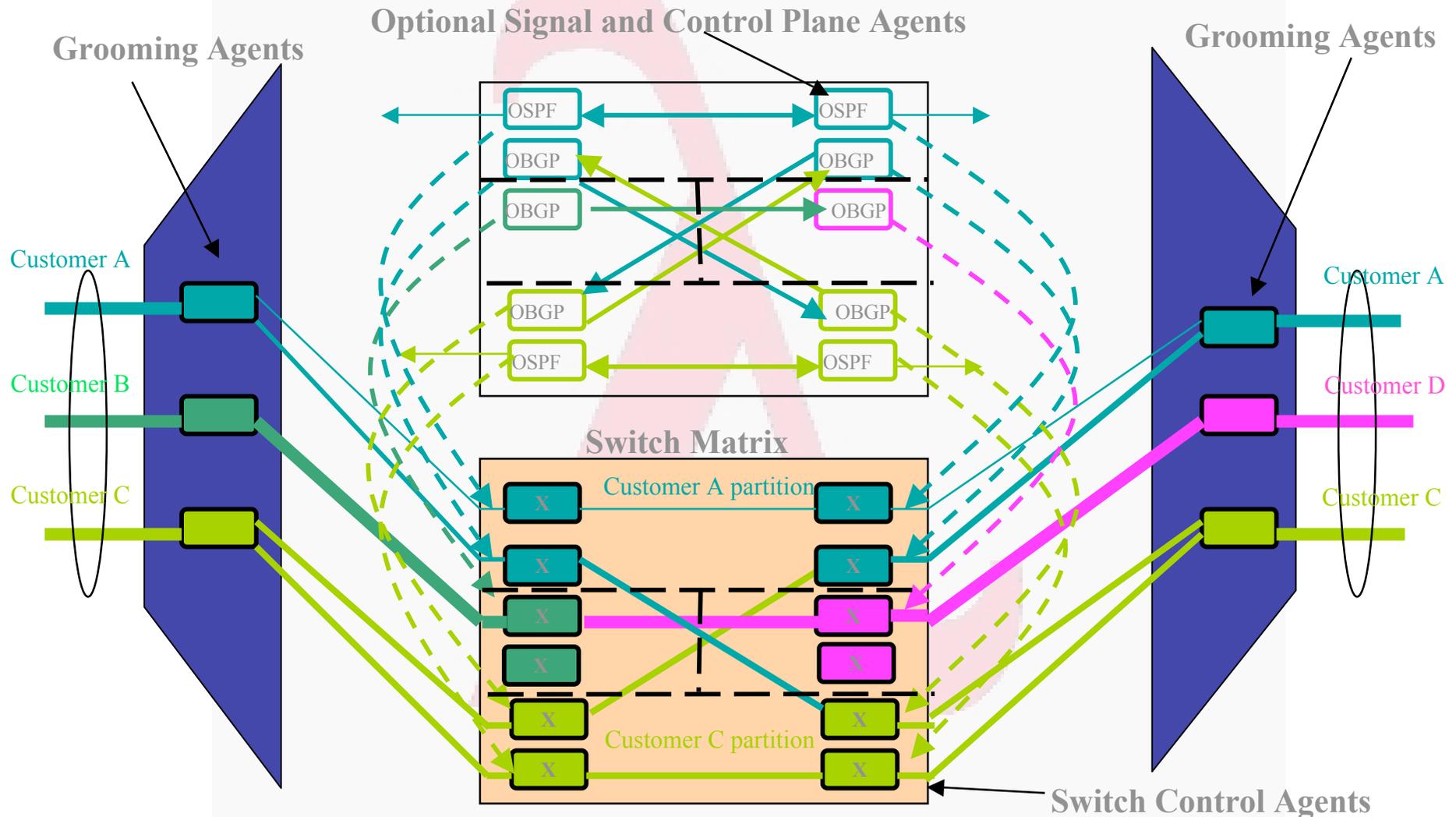
CA*net 4 Architecture



Nested Lightpaths

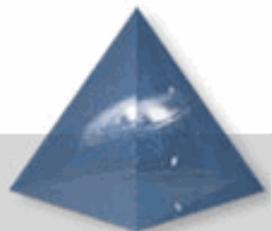
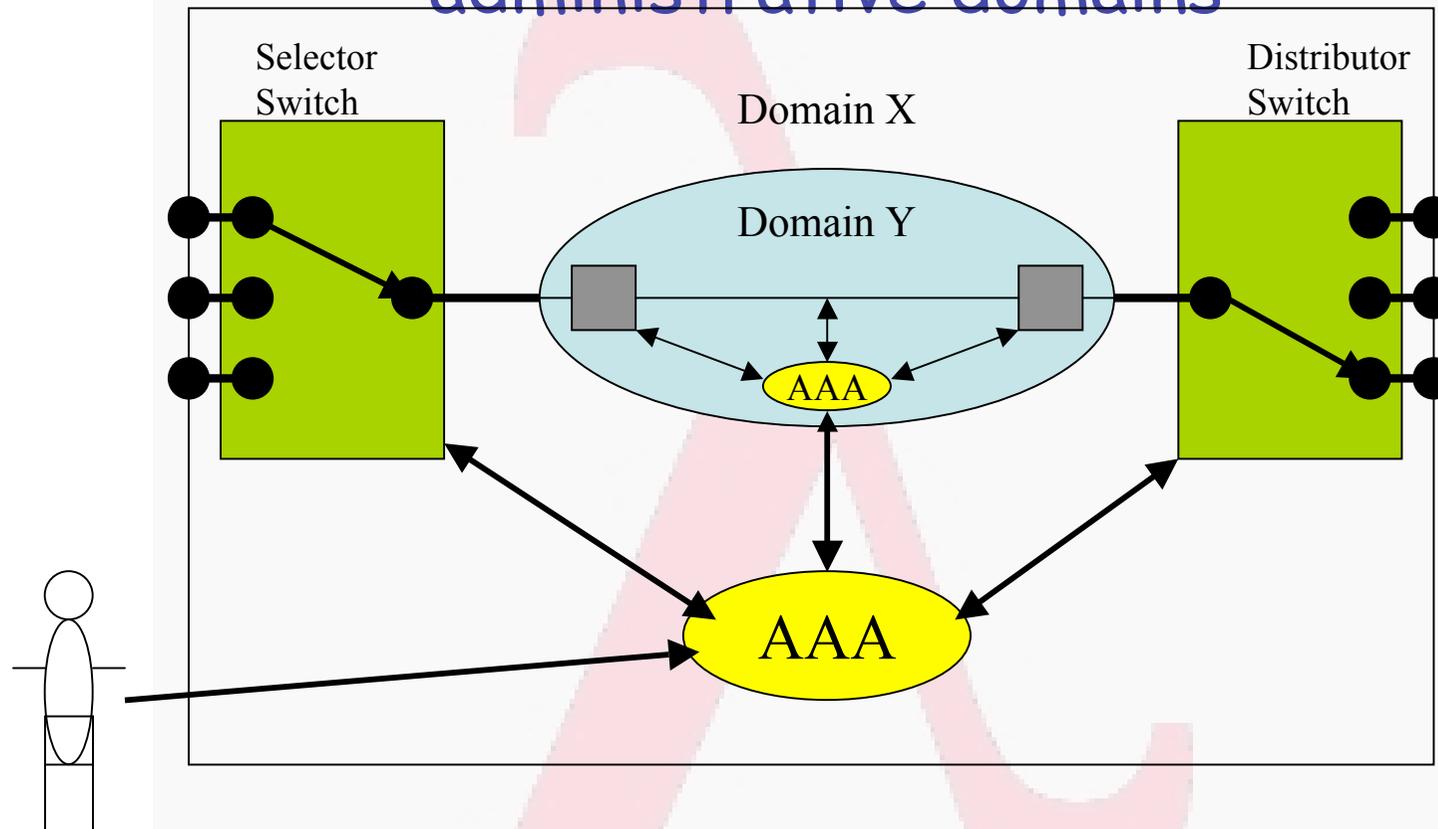


Mini-IX with OON

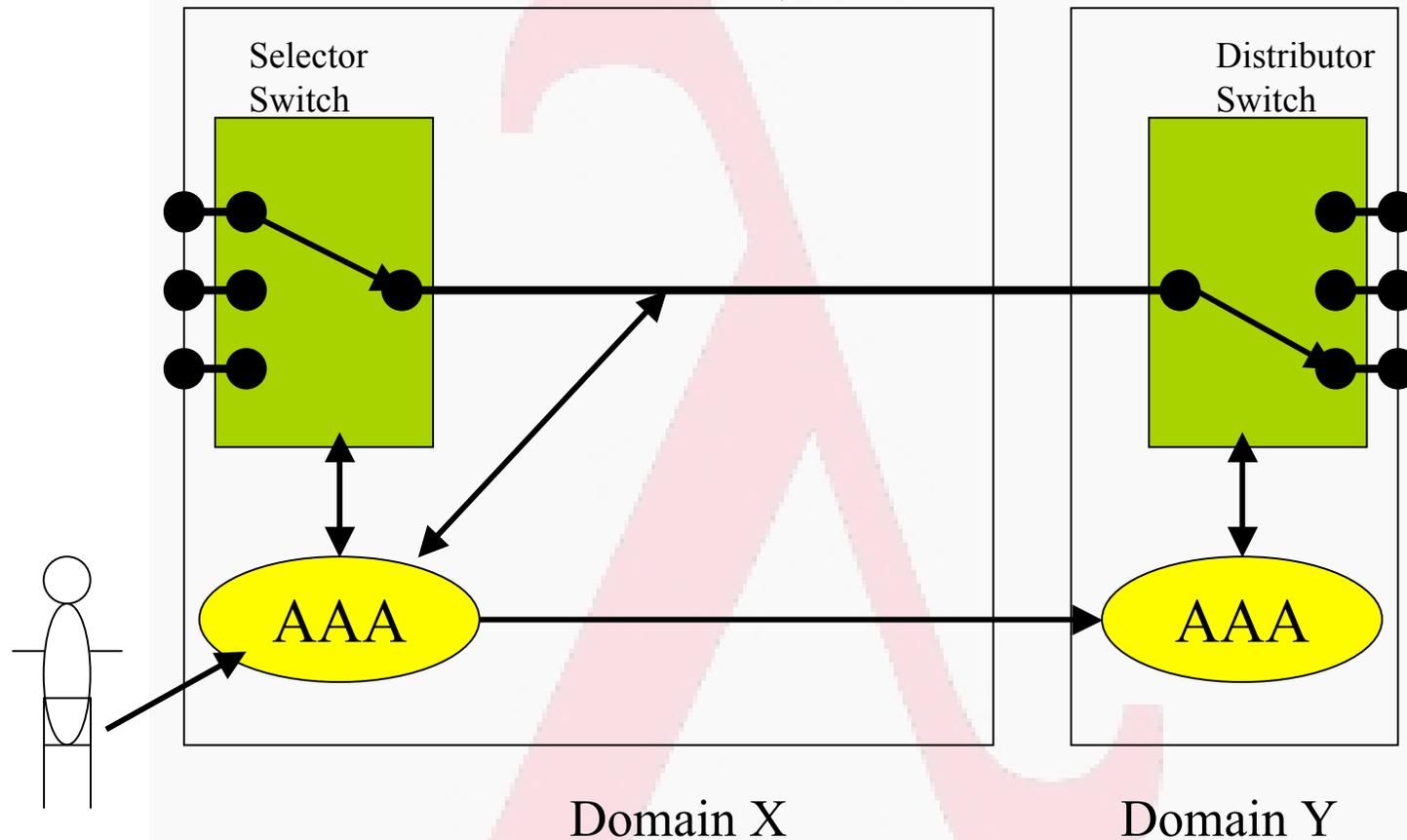


Note: Switch and Signal Planes are partitioned to respective condominium owner. All objects are controlled remotely via SOAP

Recursive Onion control model of administrative domains



Daisy Chain control model of administrative domains



XML based AAA request for Lightpath element

```
-<AAARequest version="0.1" type="Service">  
- <Authentication protocol="simple">  
  <Identity>Arie</Identity>  
  <AuthenticationData  
type="password">abc123</AuthenticationData>  
  </Authentication>  
- <ServiceData>  
  - <SwitchData>  
    <Source>146.50.0.40</Source>  
    <Destination>169.254.93.1</Destination>  
    <Bandwidth>1000</Bandwidth>  
    <StartTime>12:45</StartTime>  
    <Duration>2:31:45</Duration>  
  </SwitchData>  
  </ServiceData>  
</AAARequest>
```



```
if
(
  (
    ASM::RM.CheckConnection(
      Request::ServiceData.SwitchData.Source,
      Request::ServiceData.SwitchData.Destination
    )

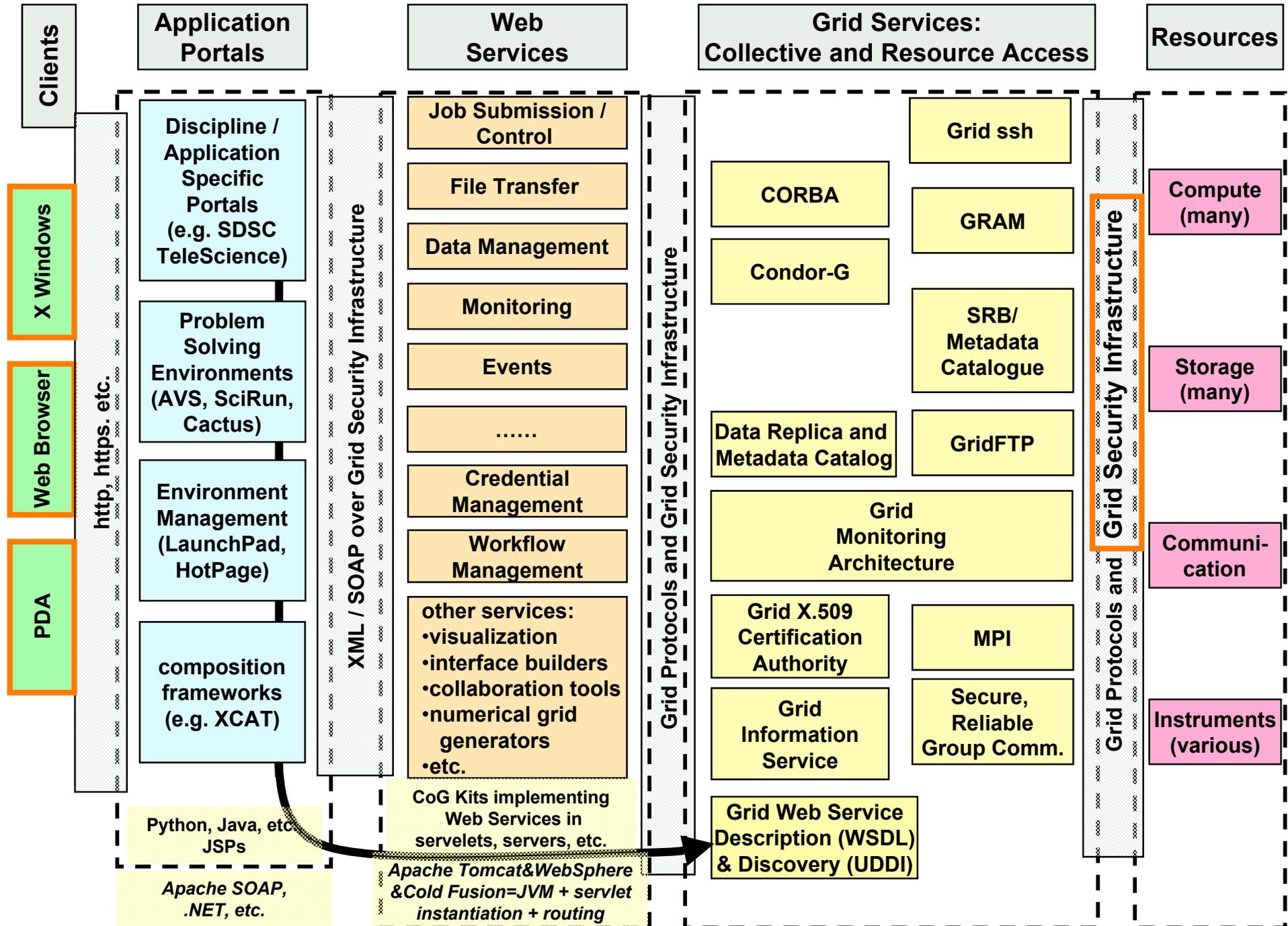
    &&

    ( Request::ServiceData.SwitchData.Bandwidth == 1000 )
  )
)
then
(
  ASM::RM.RequestConnection(
    Request::ServiceData.SwitchData.Source,
    Request::ServiceData.SwitchData.Destination,
    Request::ServiceData.SwitchData.Bandwidth,
    Request::ServiceData.SwitchData.StartTime,
    Request::ServiceData.SwitchData.Duration
  )

  ;
  Reply::Answer.Message = "Request successful"
)
else
(
  Reply::Error.Message = "Request failed"
)
```



GGF-GPA: Combined Grid and Web Services Architecture



iGrid 2002

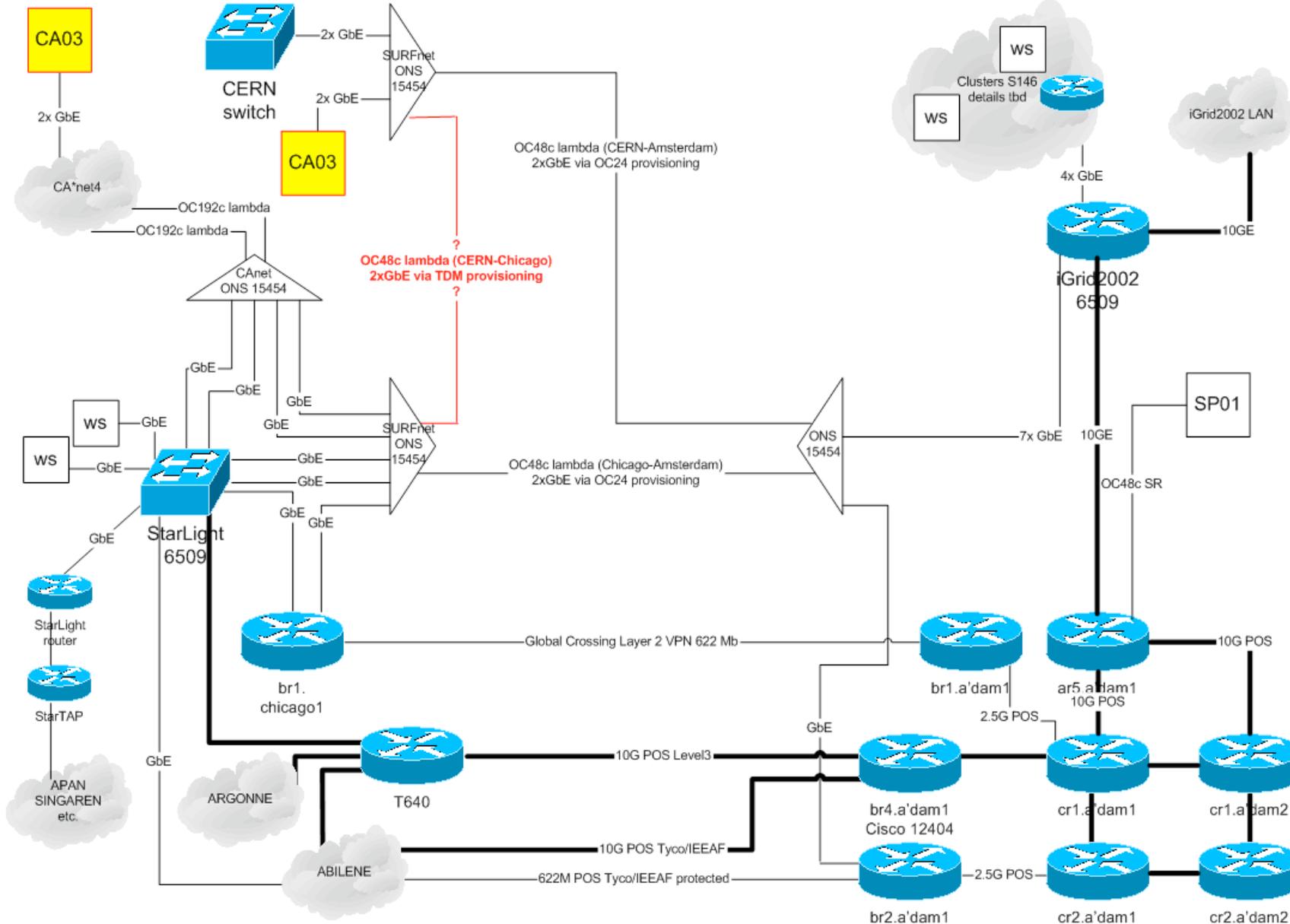
The International Virtual
Laboratory

www.igrd2002.org

24-26 September 2002
Amsterdam Science and Technology Centre (WTCW)
The Netherlands

- A showcase of applications that are “early adopters” of very-high-bandwidth national and international networks
 - What can you do with a 10Gbps network?
 - What applications have insatiable bandwidth appetites?
- Scientists and technologists to optimally utilize 10Gbps experimental networks, with special emphasis on e-Science, Grid and Virtual Laboratory applications
- Registration is open (www.igrd2002.org)
- iGrid is not just a conference/demonstration event, **it is also a testbed!!**
- Contact
 - maxine@startup.net or deLaat@science.uva.nl

iGrid 2002

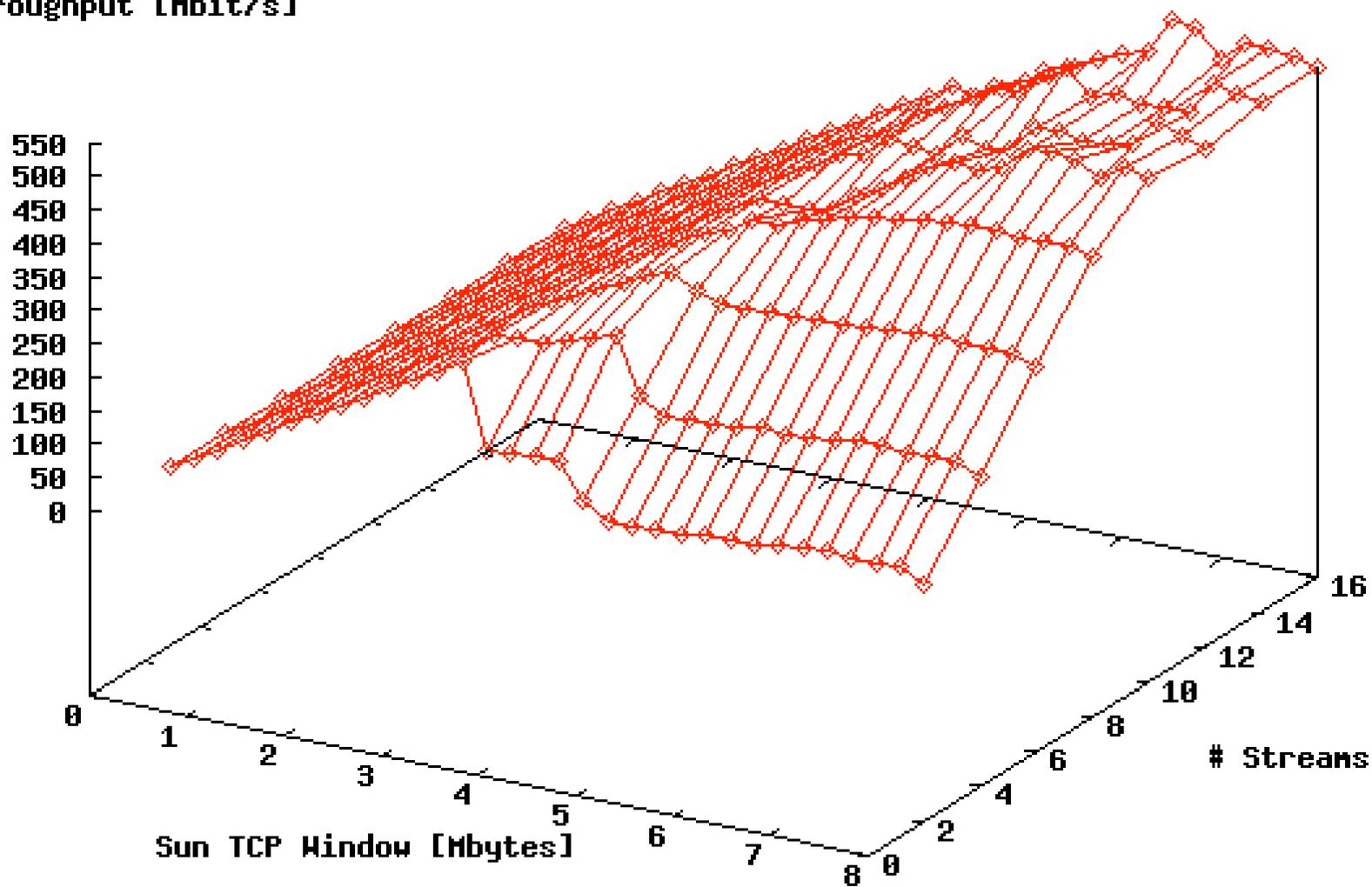


First experiences with SURFnet (14 of 18) pure for research Lambda

- **2.5 Gbit SONET λ delivered dec 2001**
 - Took about 3 months, should be 300 ms
- **First generation equipment delivered nov 2001**
- **Back to back tests => OC12 limit -> 560 Mbit/s**
- **1 unit shipped to Chicago (literally, took 3 weeks)**
- **End to end now 80 Mbit/s**
- **So, what is going on?**
- **Second generation equipment delivered in april 2002**
- **1 unit shipped to Chicago (yes, it took 3 weeks)**

EVL => HCH 

Sun Throughput [Mbit/s]



Layer - 2 requirements from 3/4



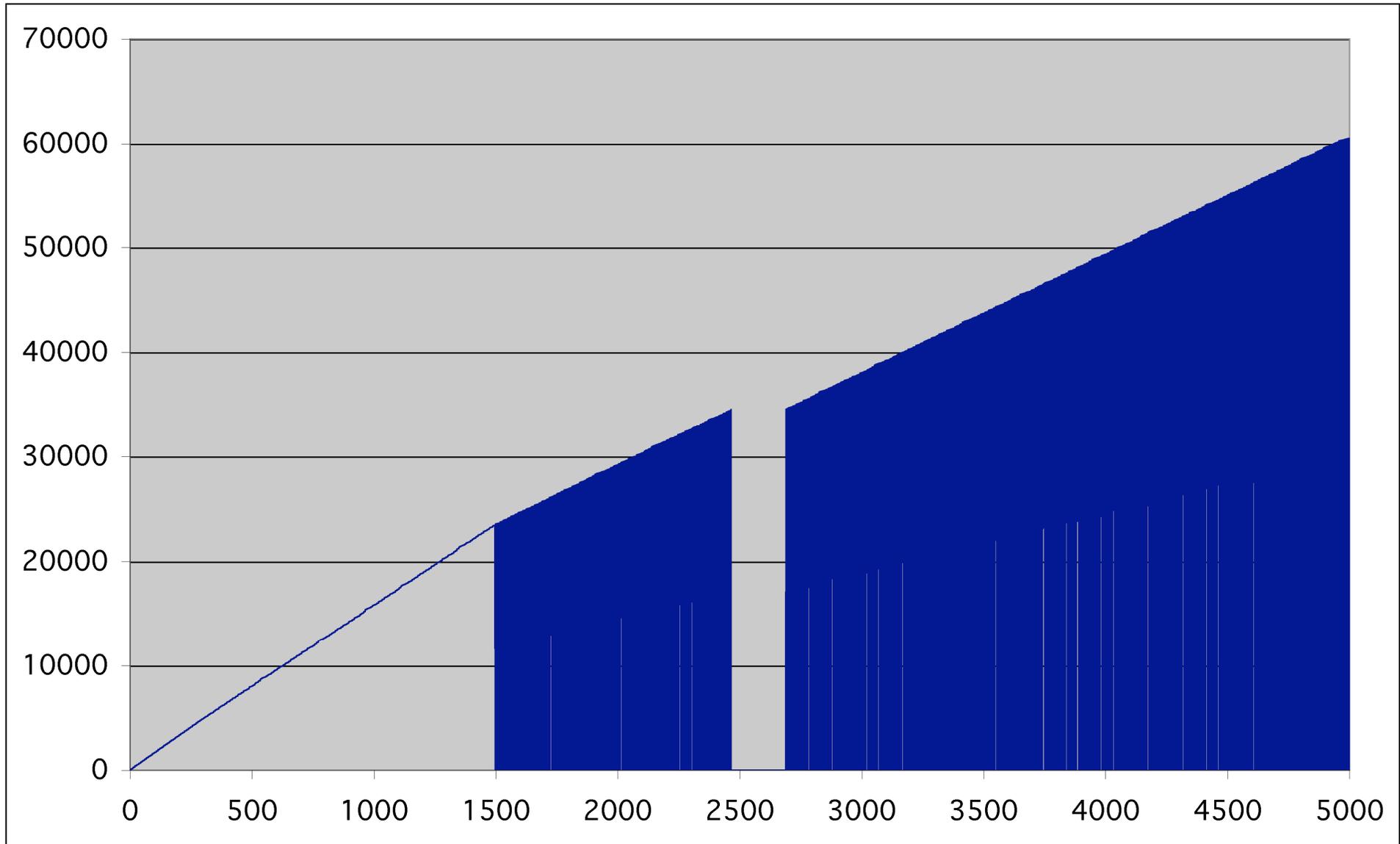
TCP is bursty due to sliding window protocol and slow start algorithm. So pick from menu:

- ◆ *Flow control*
- ◆ *Traffic Shaping*
- ◆ *RED (Random Early Discard)*
- ◆ *Self clocking in TCP*
- ◆ *Deep memory*

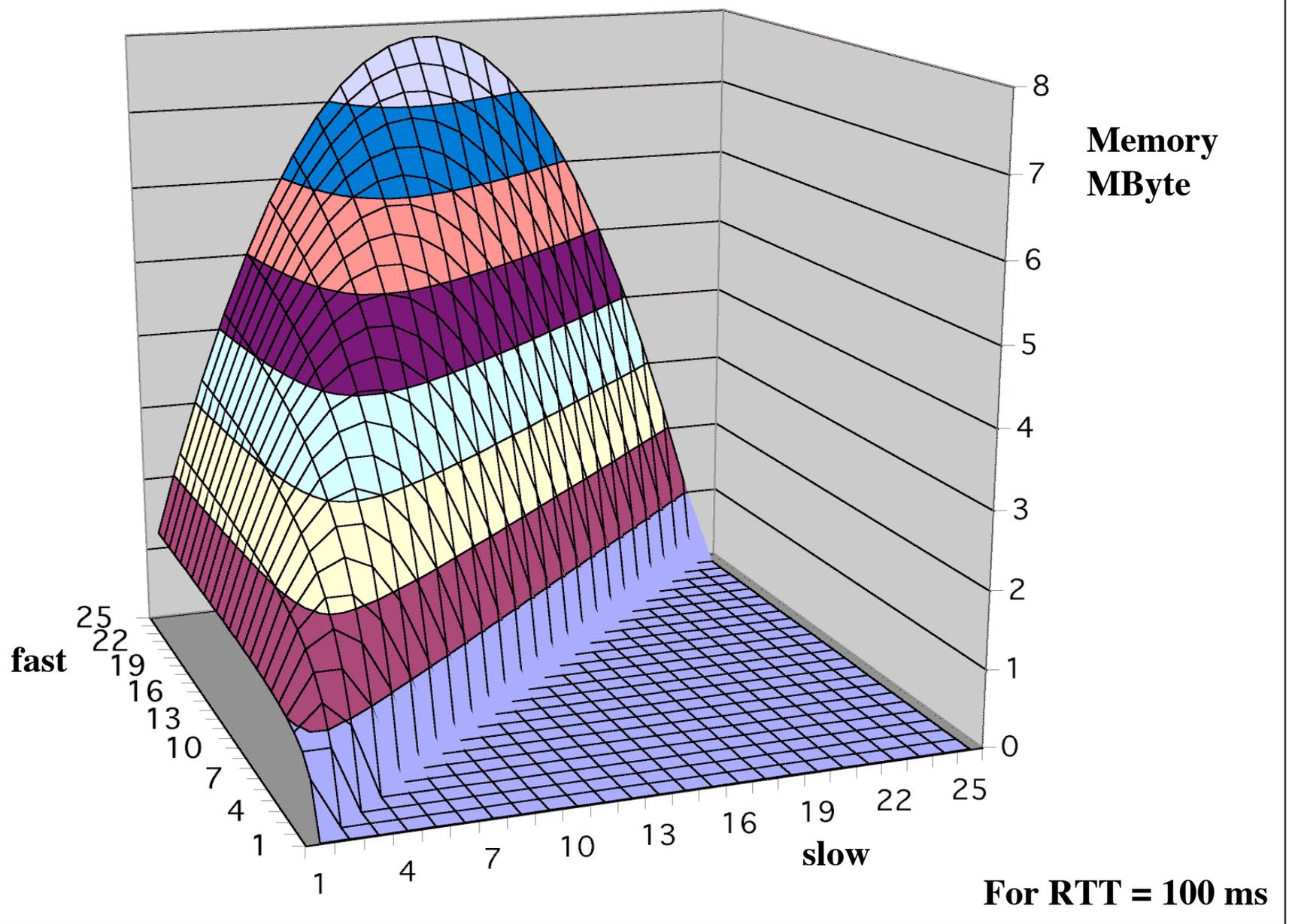
Window = BandWidth * RTT & BW == slow

Memory-at-bottleneck = $\frac{\text{fast} - \text{slow}}{\text{fast}} * \text{slow} * \text{RTT}$

5000 1 kByte UDP packets



$$\text{Memory} = \frac{\text{fast} - \text{slow}}{\text{fast}} * \text{slow} * \text{RTT}$$



Layer - 2 requirements from 3/4



Window = BandWidth * RTT & BW == slow

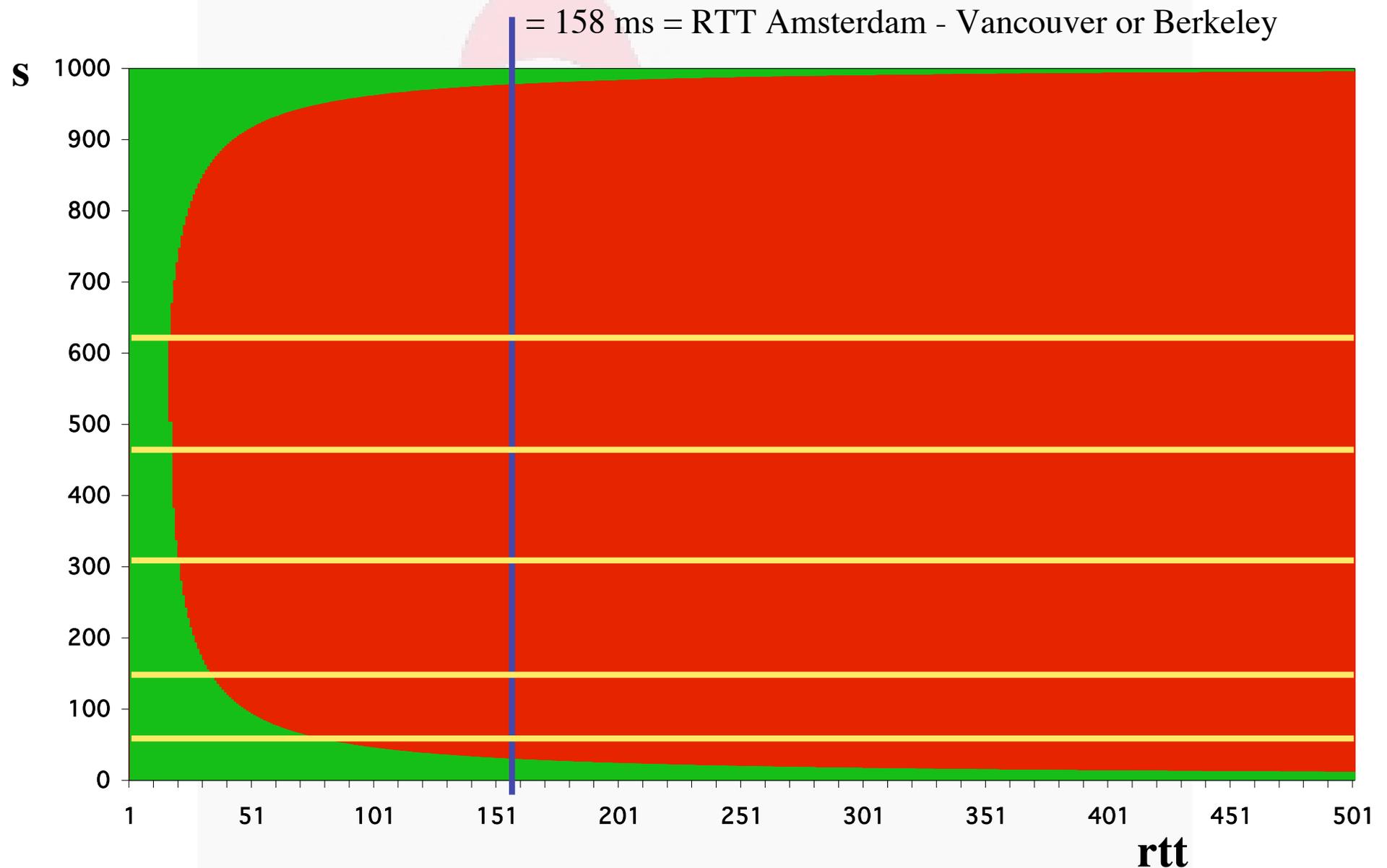
Memory-at-bottleneck = $\frac{\text{fast} - \text{slow}}{\text{fast}} * \text{slow} * \text{RTT}$

Given M and f, solve for slow ==>

$$0 = s^2 - f * s + \frac{f * M}{\text{RTT}}$$

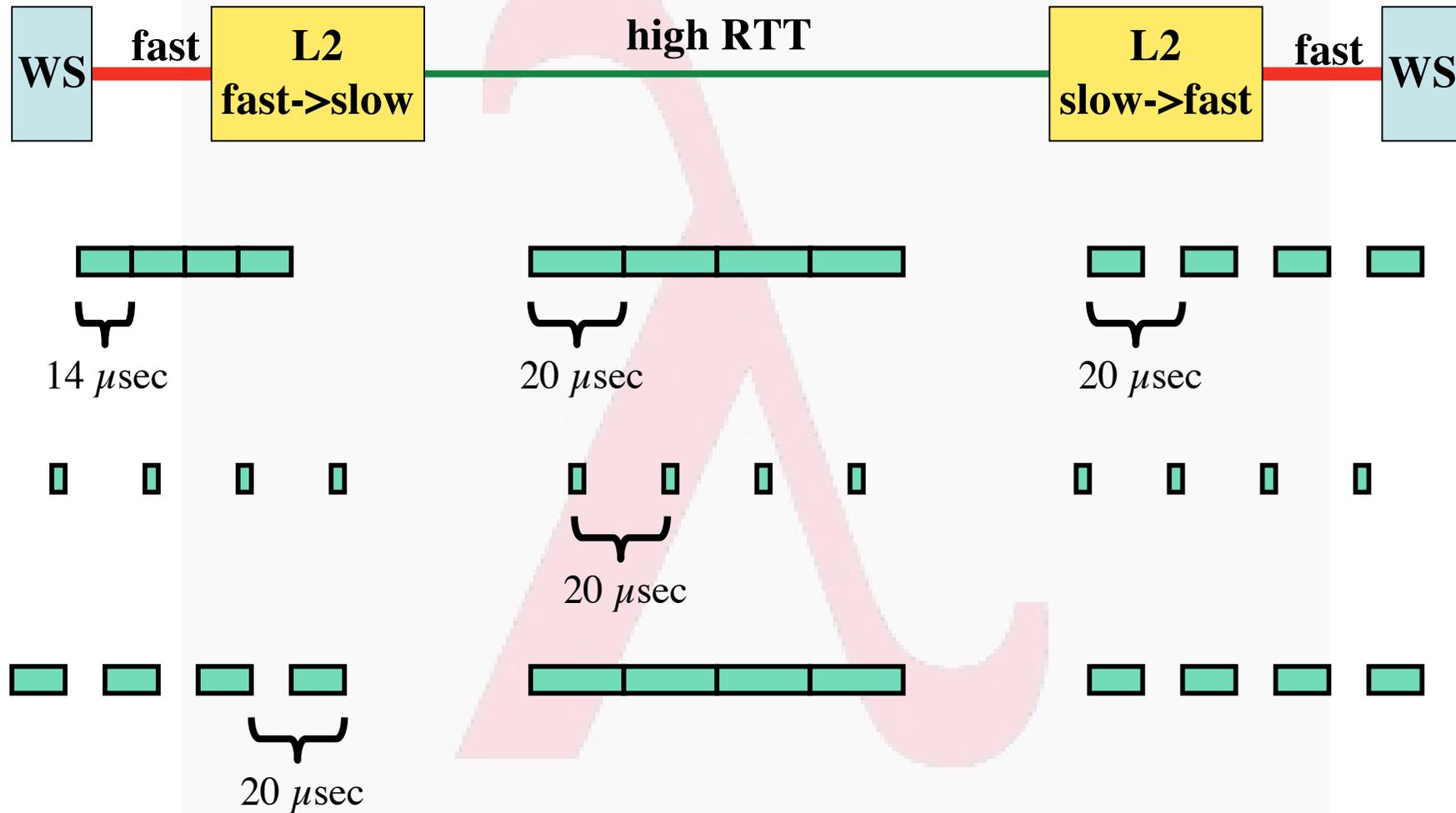
$$s_1, s_2 = \frac{f}{2} \left(1 \pm \sqrt{1 - 4 \frac{M}{f * \text{RTT}}} \right)$$

Forbidden area, solutions for s when $f = 1$ Gb/s, $M = 0.5$ Mbyte AND NOT USING FLOWCONTROL



Self-clocking of TCP

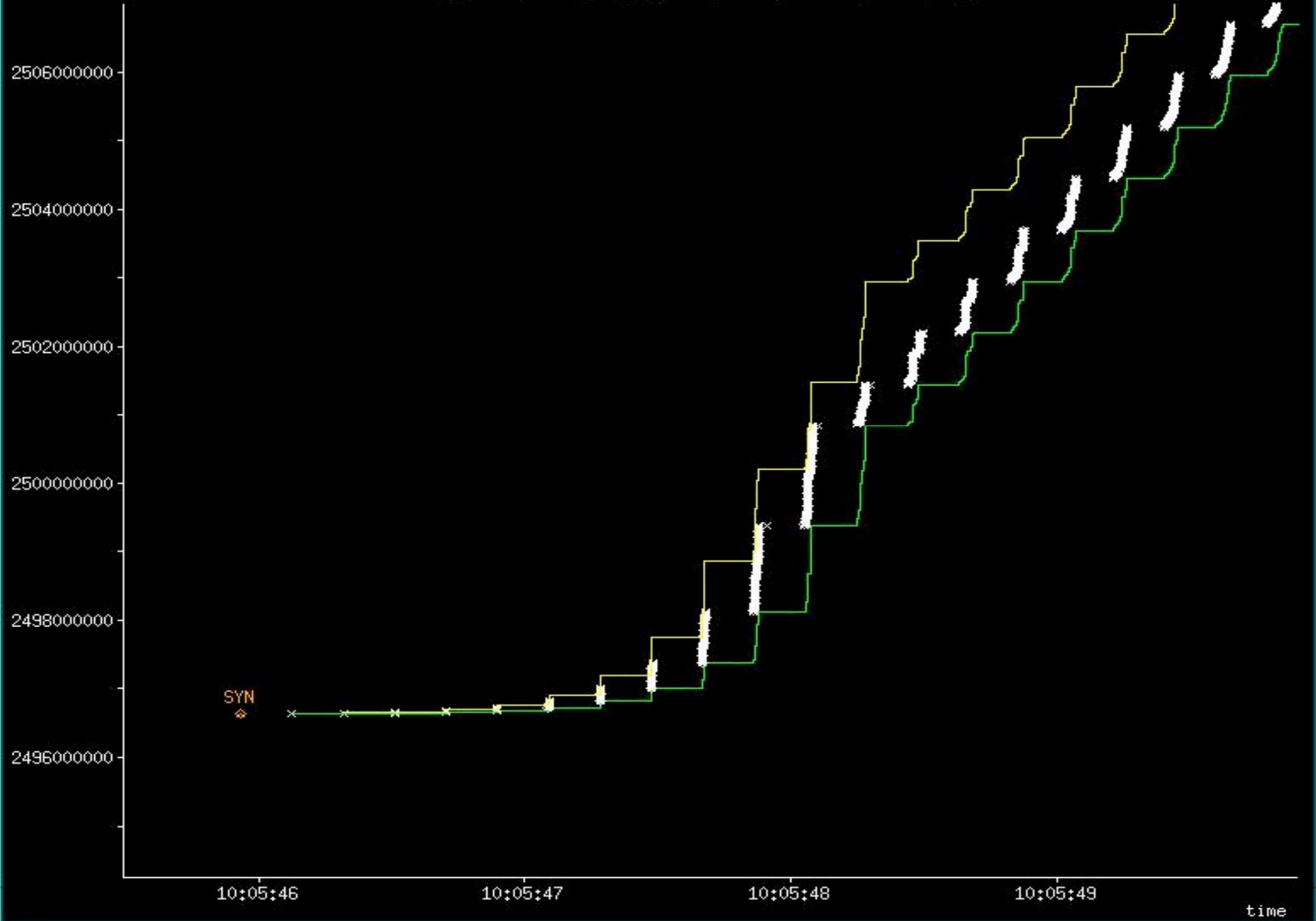
(17 of 18)



xplot

sequence number

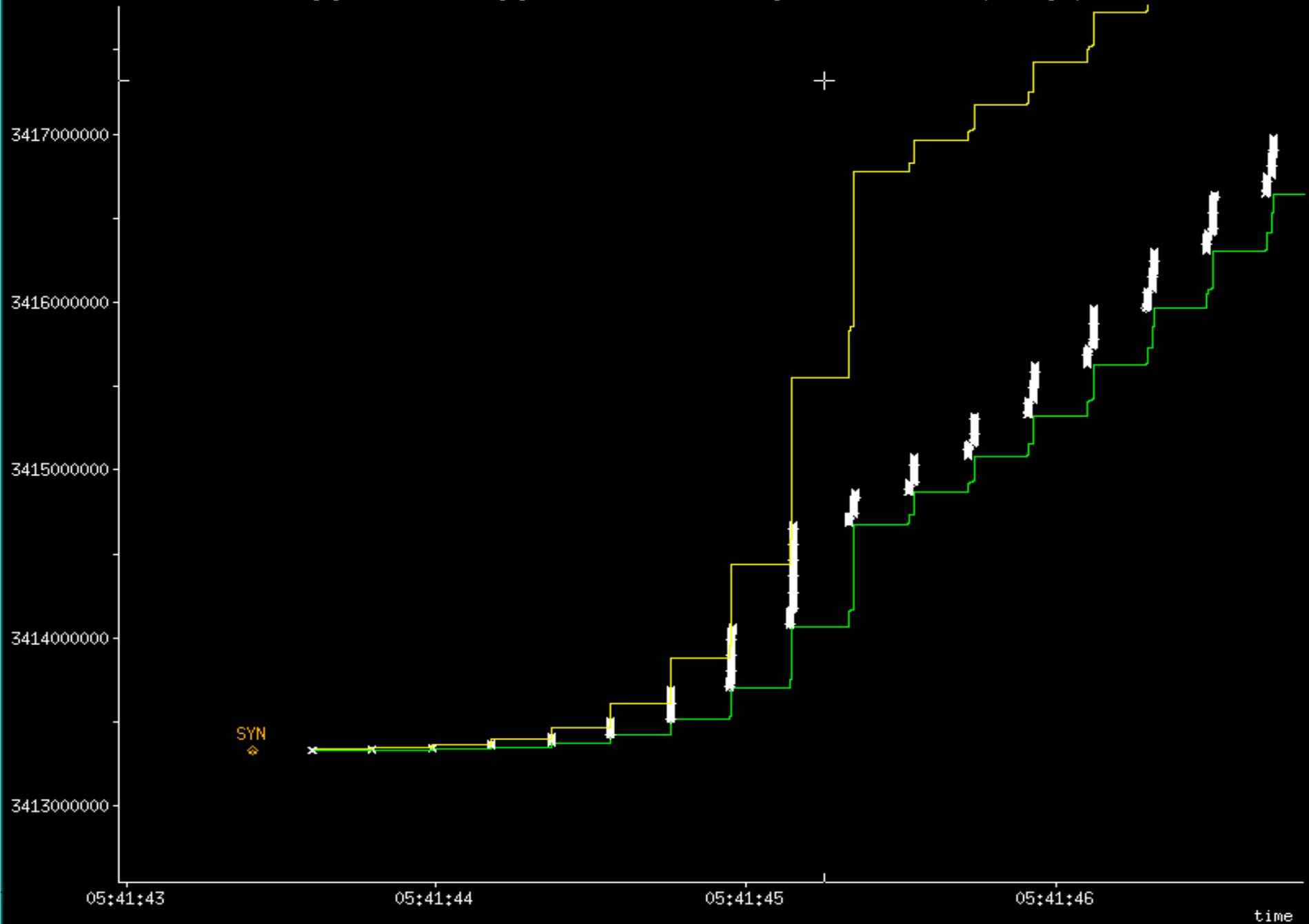
gwgsara3:42141_==>_gwgsara4:5432 (time sequence graph)



xplot

sequence number

gwgsara3:42186_==>_gwgsara5.Amsterdam1.Netherlight.net:5001 (time sequence graph)



Revisiting the truck of tapes

Consider one fiber

- Current technology allows for 320 λ in one of the frequency bands
- Each λ has a bandwidth of 40 Gbit/s
- Transport: $320 * 40 * 10^9 / 8 = 1600$ GByte/sec
- Take a 10 metric ton truck
- One tape contains 50 Gbyte, weights 100 gr
- Truck contains $(10000 / 0.1) * 50$ Gbyte = 5 PByte
- Truck / fiber = $5 \text{ PByte} / 1600 \text{ GByte/sec} = 3125 \text{ s} \approx \text{one hour}$
- For distances further away than a truck drives in one hour (50 km) minus loading and handling 100000 tapes **the fiber wins!!!**

The END

Thanks to

TERENA: David Williams

SURFnet: Kees Neggers

UIC&iCAIR: Tom DeFanti, Joel Mambretti

CANARIE: Bill St. Arnaud

This work is supported by:

SURFnet

EU-IST project DATATAG



SURFnet

