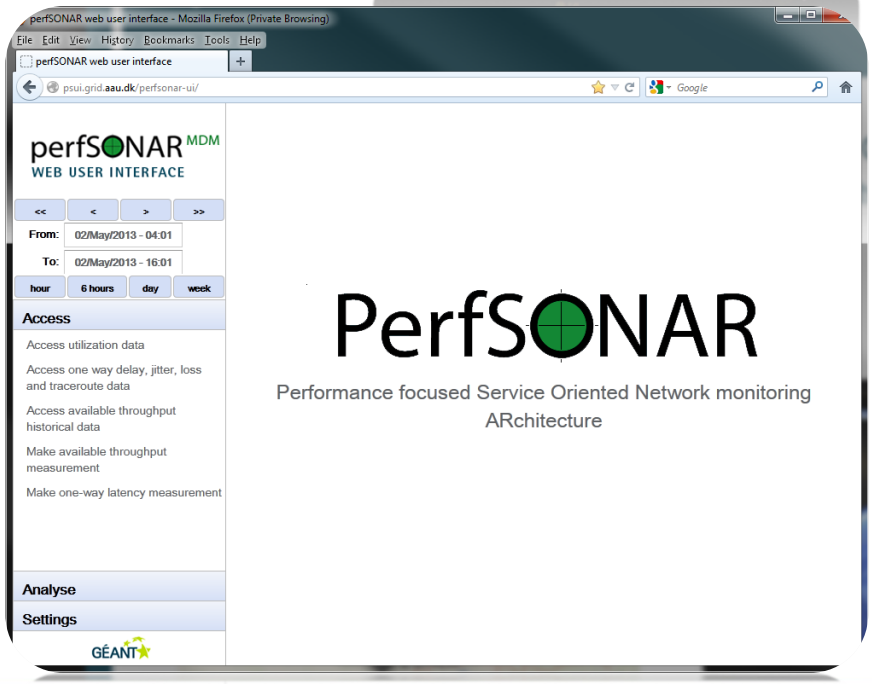
perfSONAR MDM and TCP buffers

Domenico Vicinanza
DANTE, Cambridge, UK
domenico.vicinanza@dante.net

EGI Technical Forum 2013, Madrid, Spain

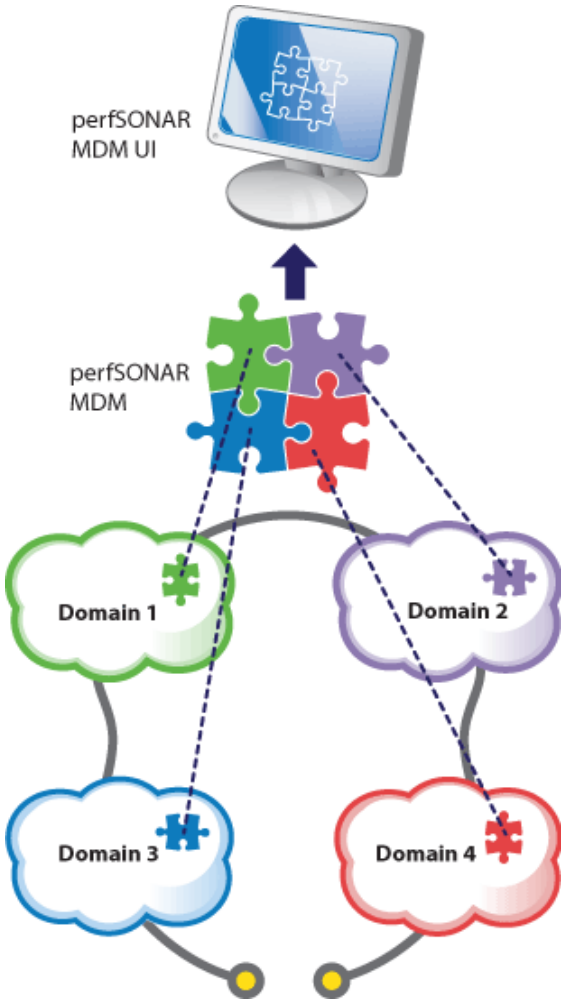
perfSONAR MDM

- Multi-domain monitoring service
- Based on monitoring probes installed in the network
- Based on a standard (perfSONAR) protocol
- Hundreds of deployments around the world
- Web interface!

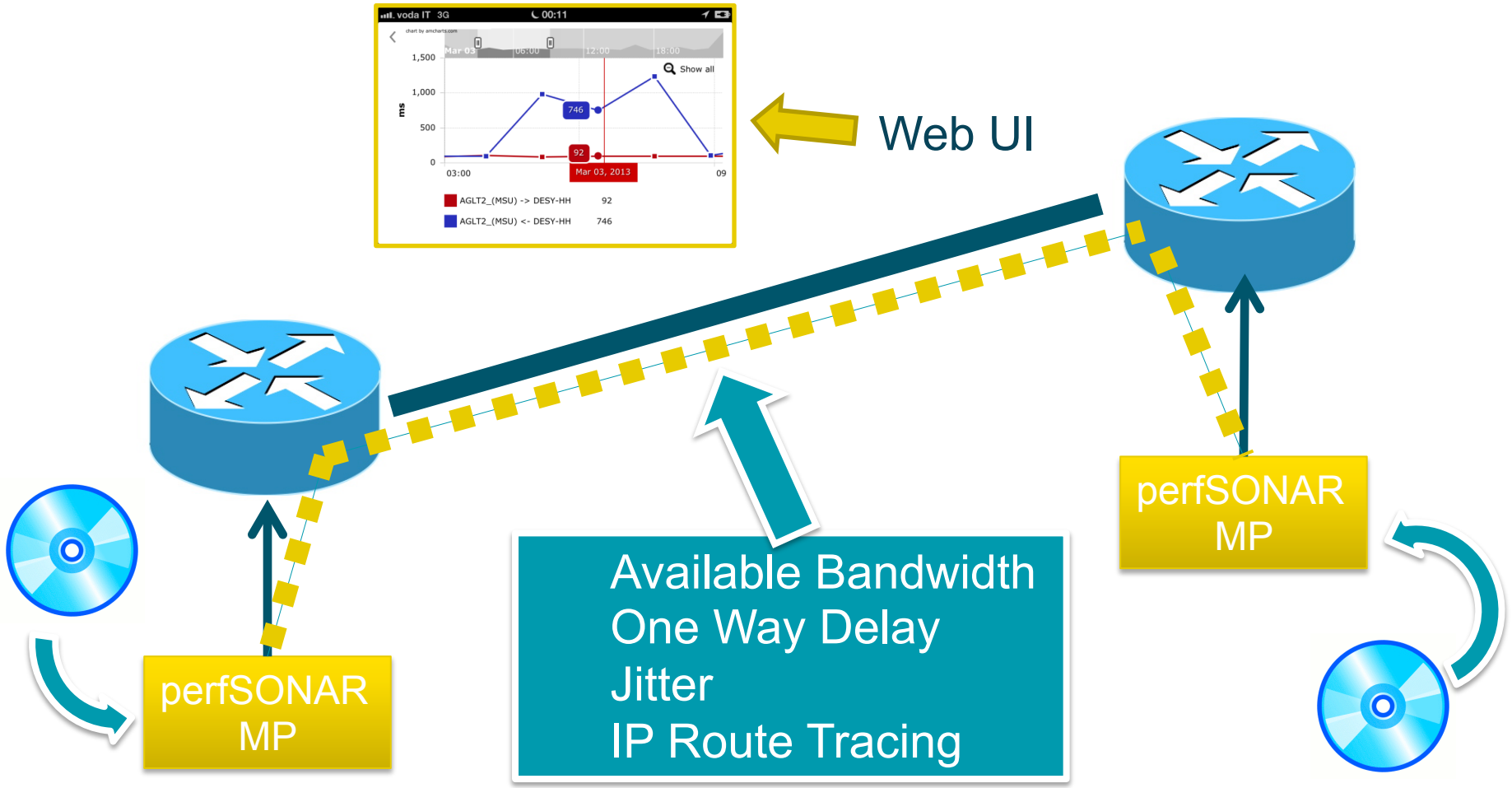


perfSONAR in a nutshell

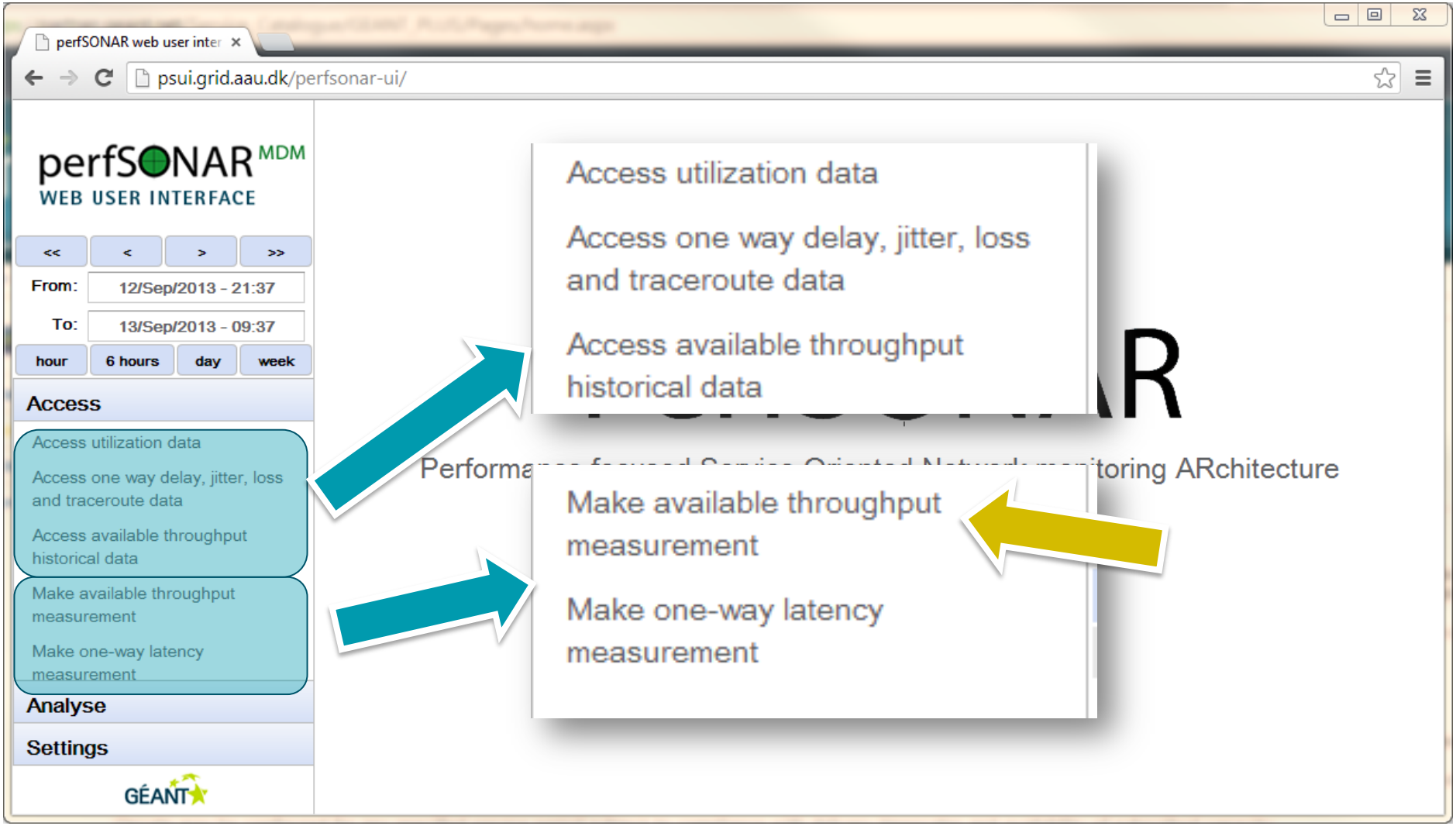
- **Perf**ormance focused **S**ervice **O**riented **N**etwork monitoring **Ar**chitecture
- International collaboration for network monitoring
- Contributors are GÉANT, Internet2, ESnet, and RNP
- Two main implementations committed to interoperate:
  - perfSONAR MDM within GÉANT: http://perfsonar.geant.net
  - perfSONAR PS within I2/ESnet: http://psps.perfsonar.net/
- Open OGF protocol to exchange data
- Web-service based
- Design goals: flexibility, extensibility, openness, and decentralization.



How does perfSONAR work?



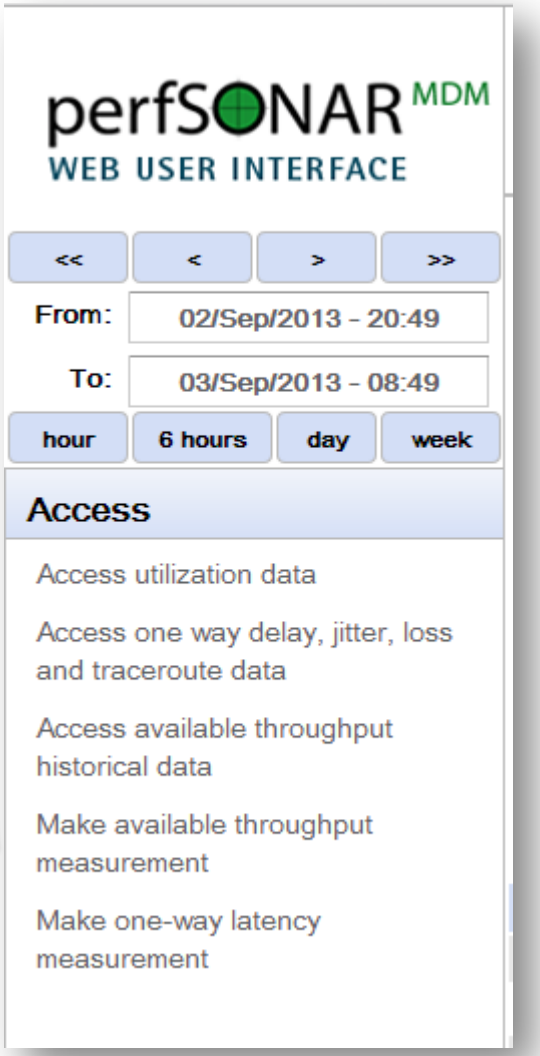
perfSONAR web user interface



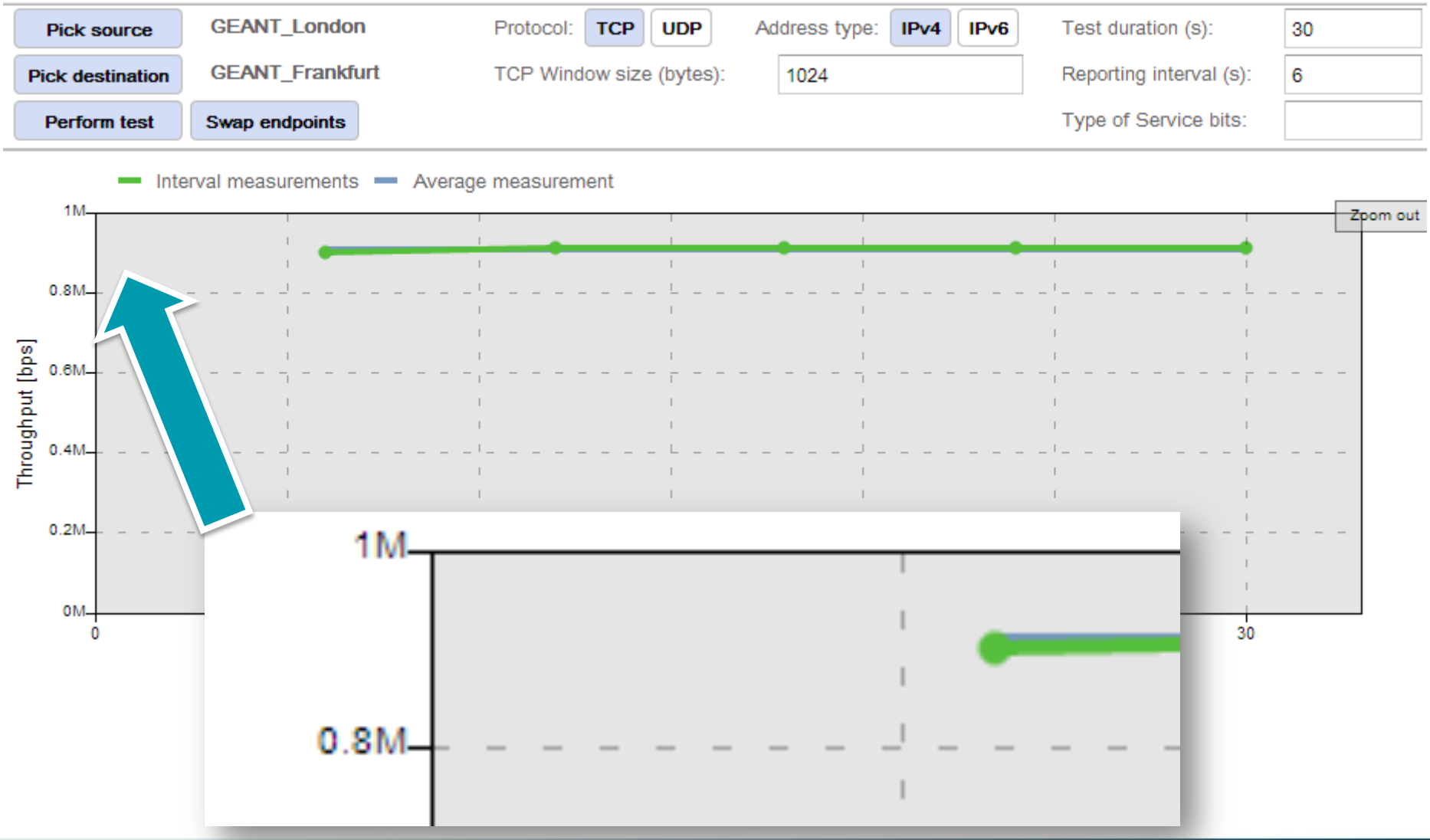
Using perfSONAR UI to run TCP measurements

- Left Panel on perfSONAR UI
- Click on “Make available throughput measurement”

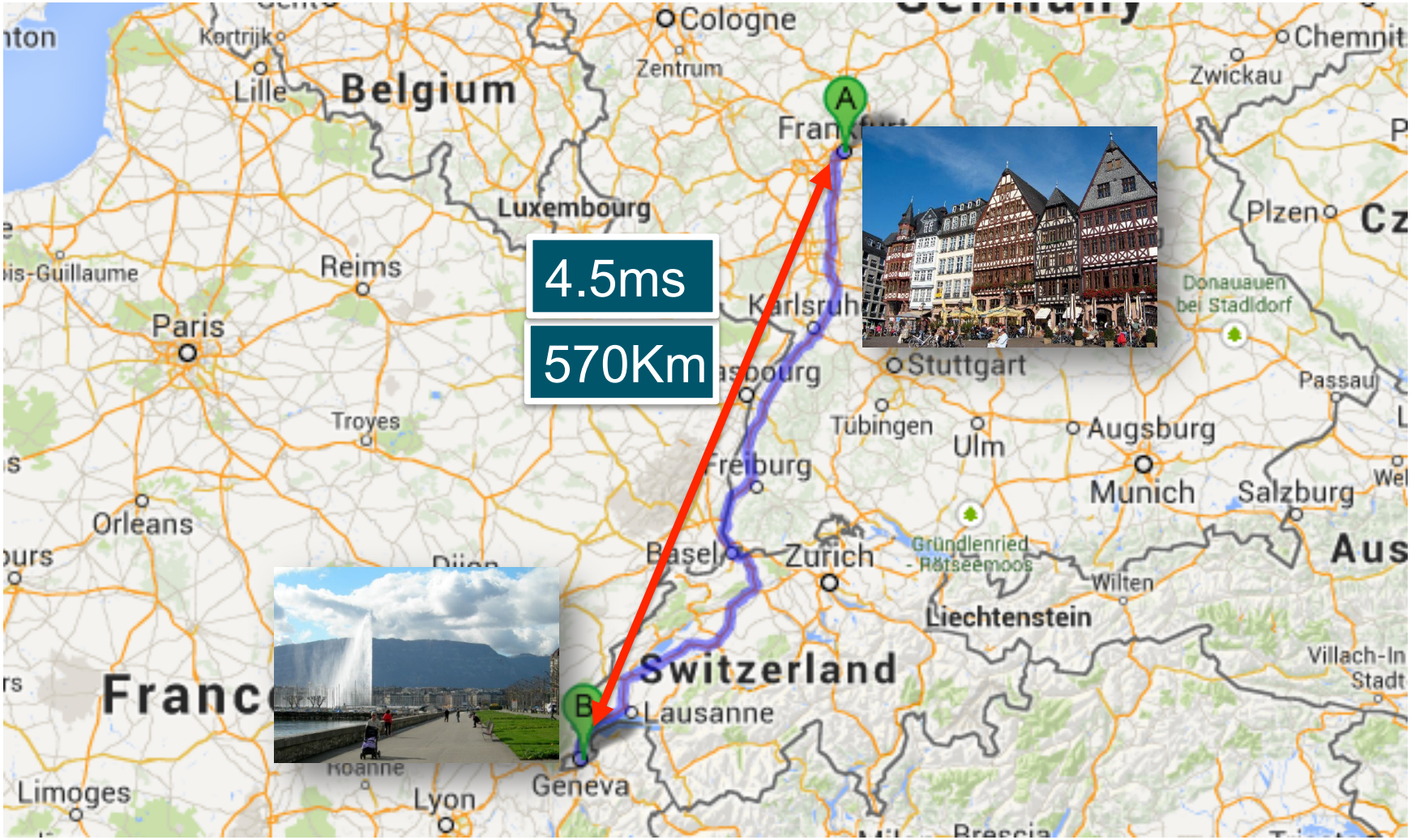
Make available throughput measurement



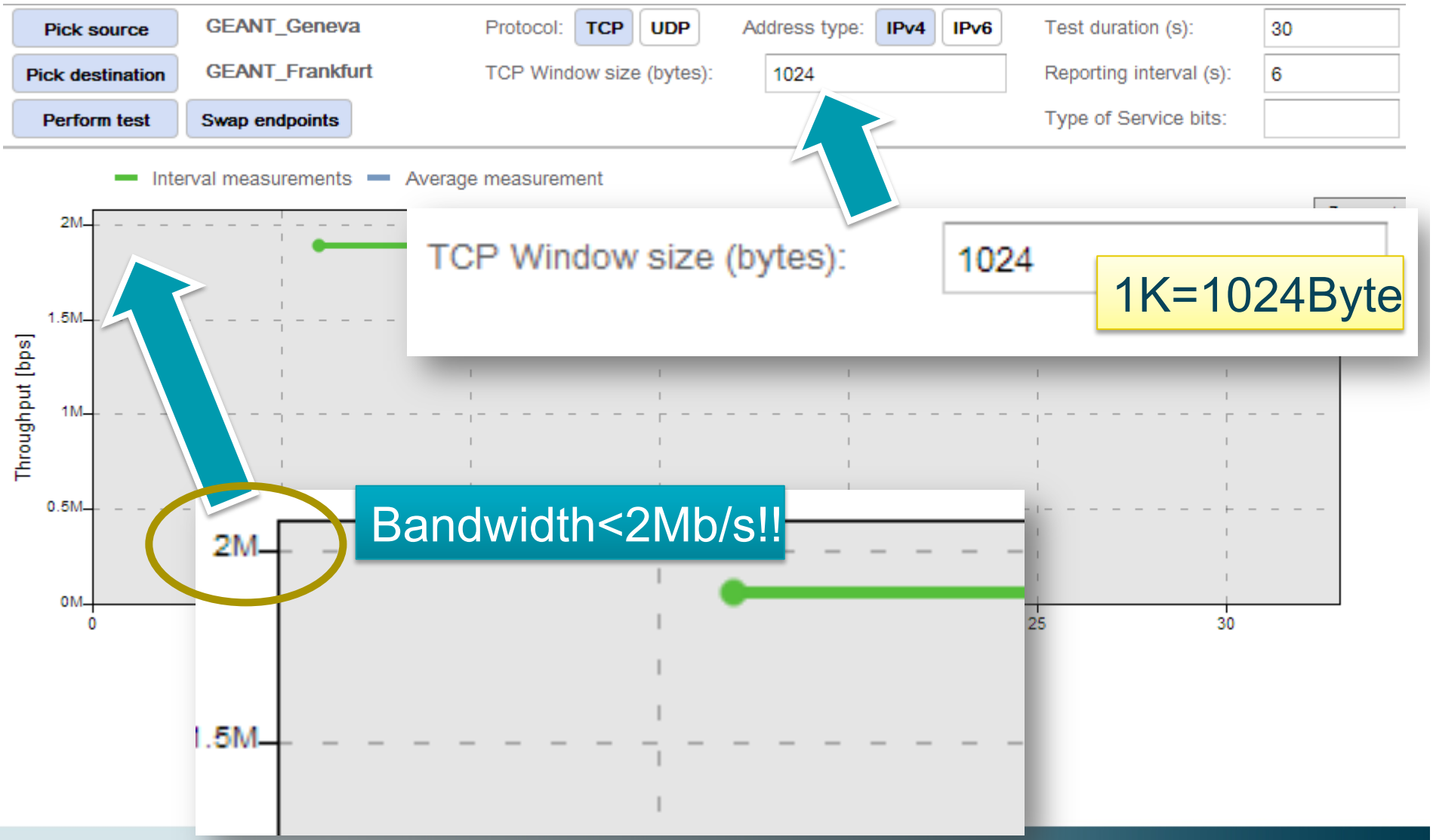
London-Frankfurt
1Kbyte buffer



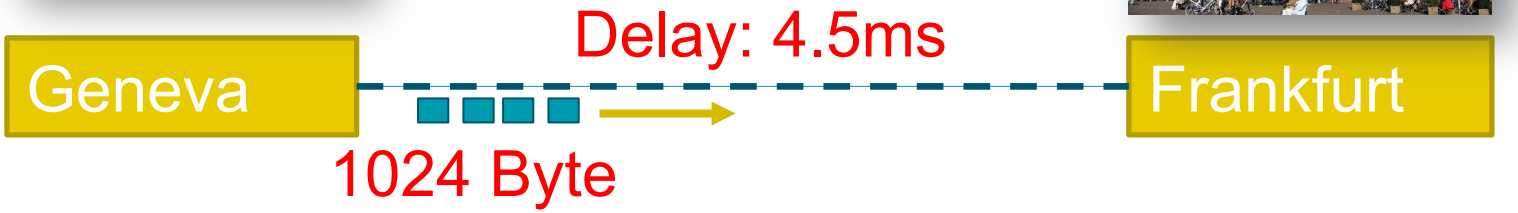
First measurement: Geneva-Frankfurt



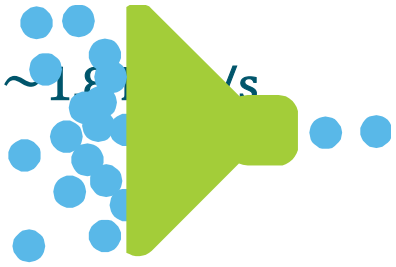
Geneva-Frankfurt
1Kbyte buffer



What happens with TCP buffer=1KB?



- 1024Bytes sent every 4.5ms
- Max Throughput≤(TCP buffer)/Delay
- $Max\ Throughput \leq 1024 \cdot 8 / 4.5 \cdot 10\uparrow -3$ ~1
- **Bandwidth limited by the buffer!**



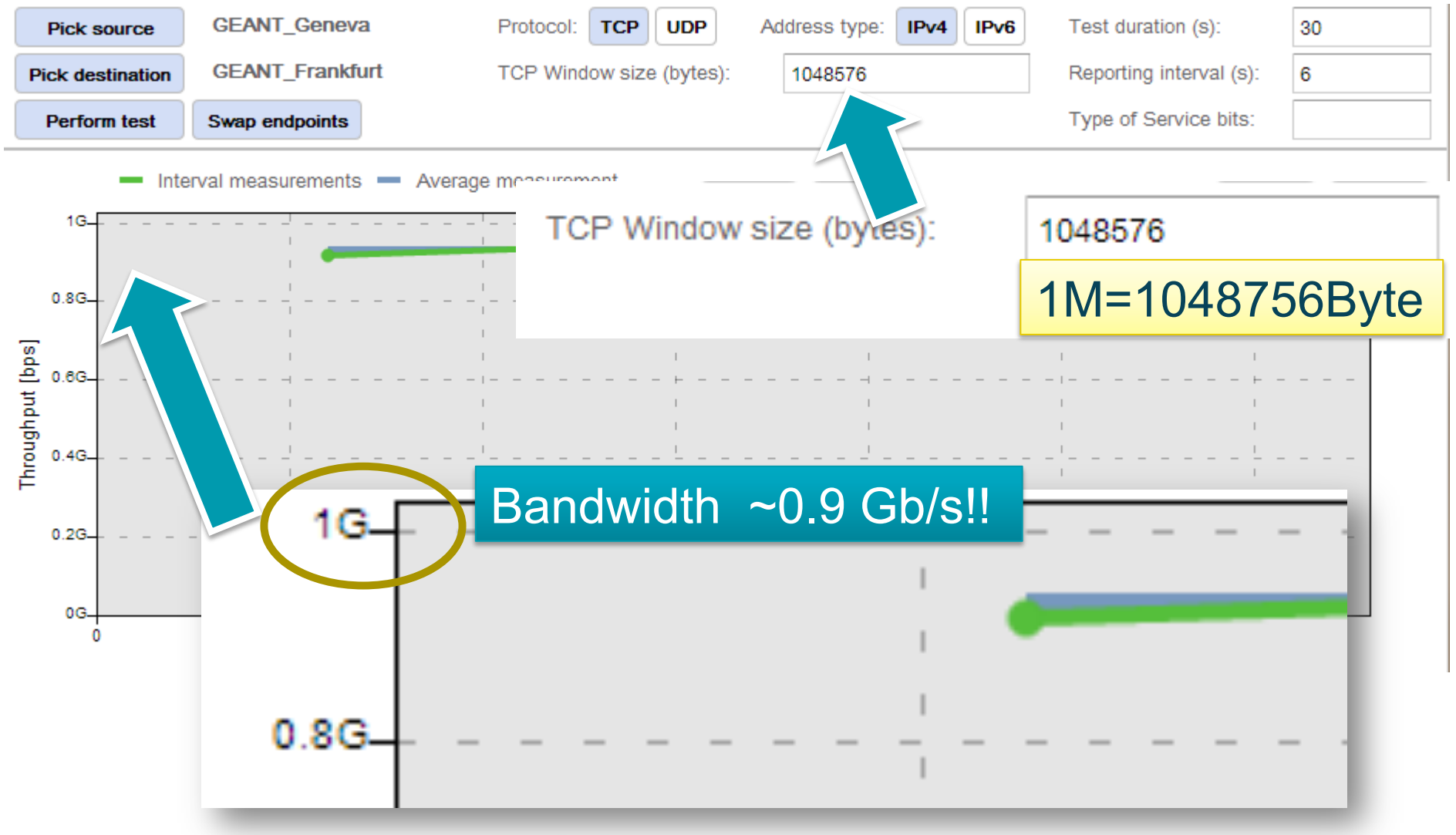
Optimal TCP Window size: Geneva-Frankfurt

- RTT: 9ms
- 1Gb/s network interface

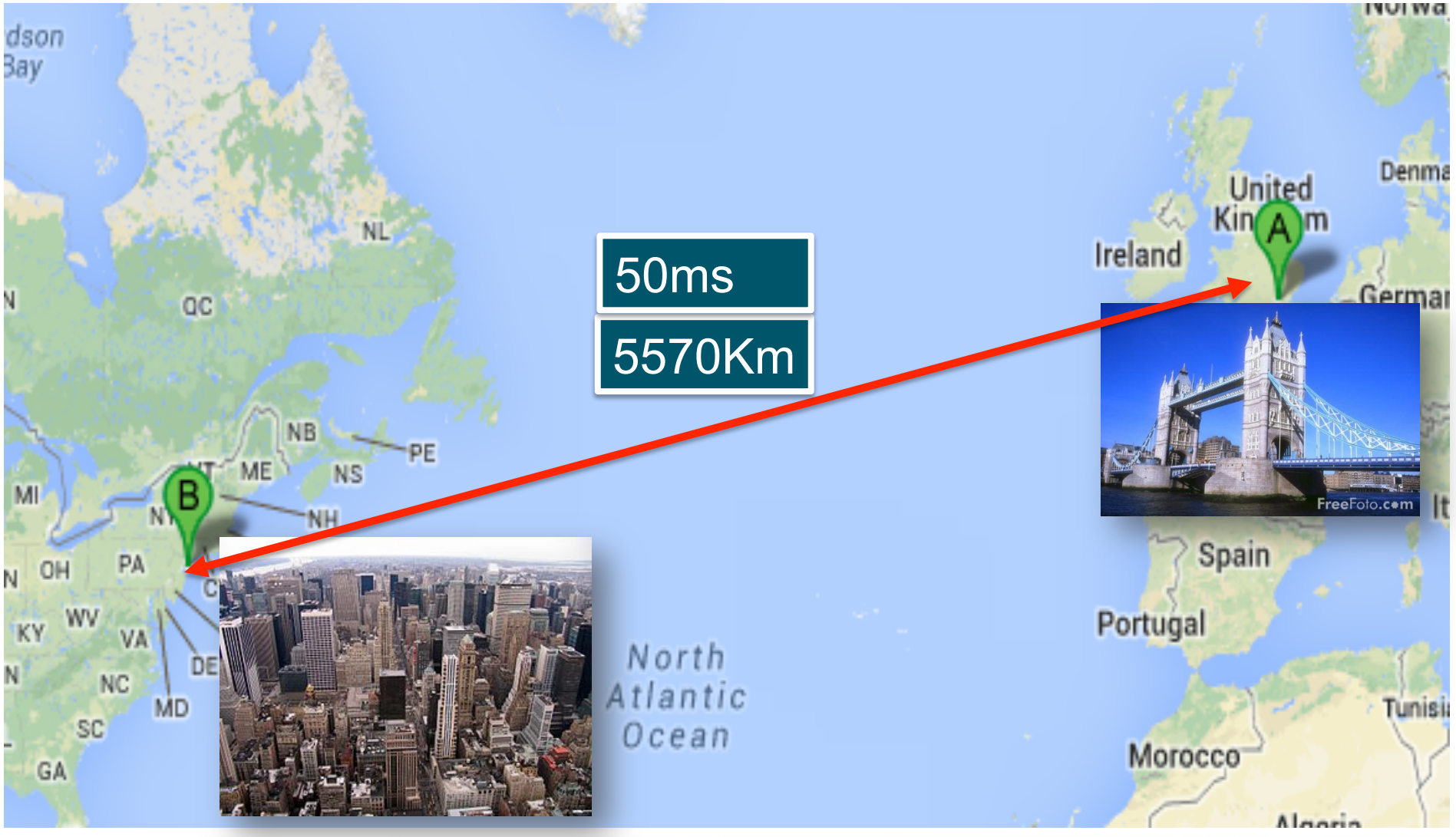
Calculating Bandwidth Delay Product

- BDP=$9 \cdot 10^{-3} \cdot 1 \cdot 10^{9} = 9 \cdot 10^{6} = 9 Mbit = 1.07 MByte$
- Optimal TCP window = BDP
  - TCP window = **1MByte for Geneva-Frankfurt**

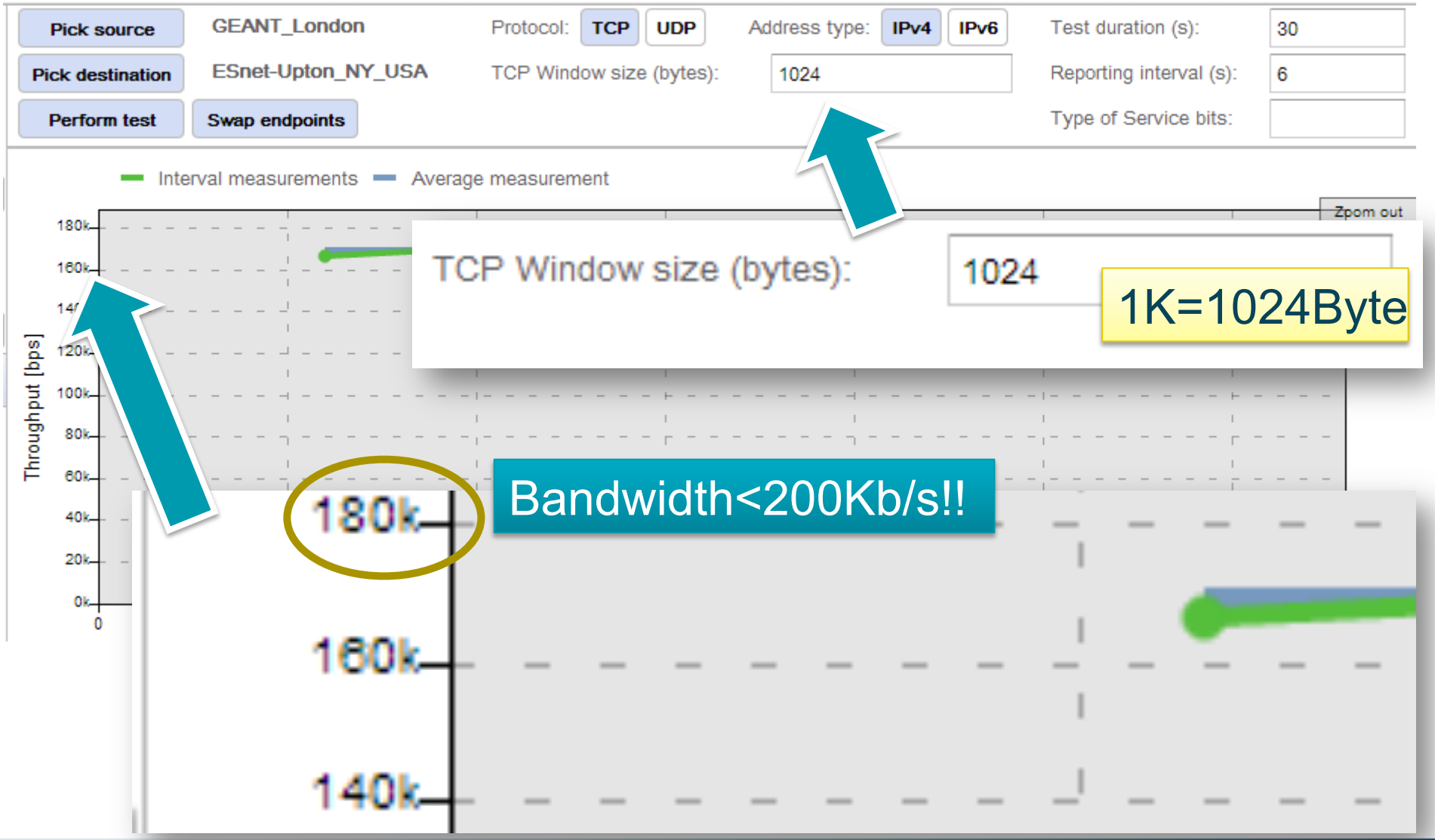
Geneva-Frankfurt
1Mbyte TCP Buffer



Second measurement: London – New York



London-New York
1Kbyte Buffer

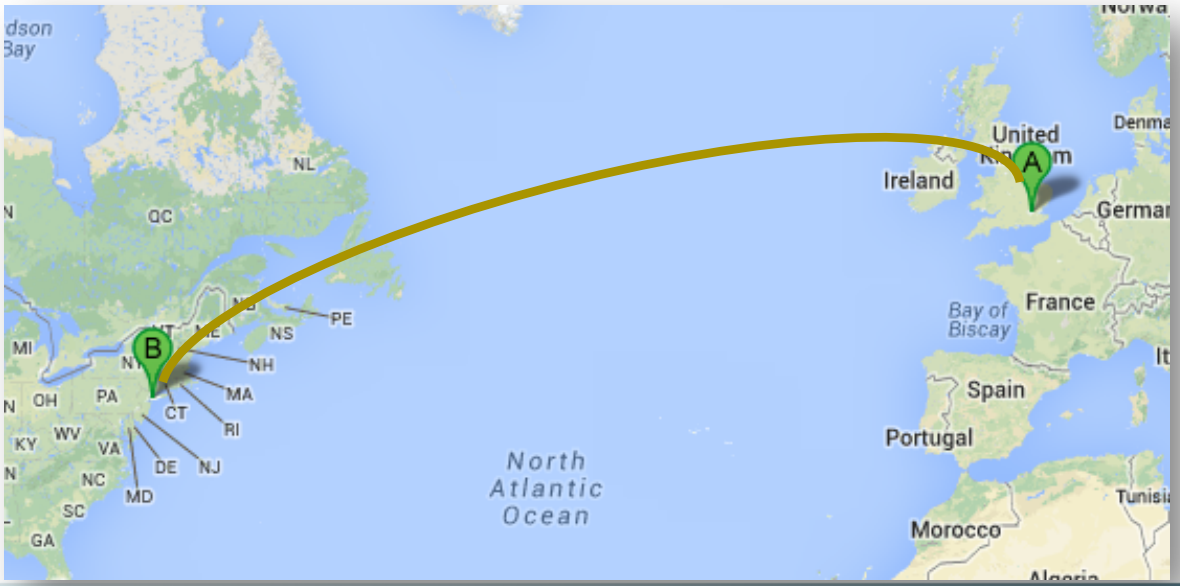


Optimal TCP Window size: London-New York

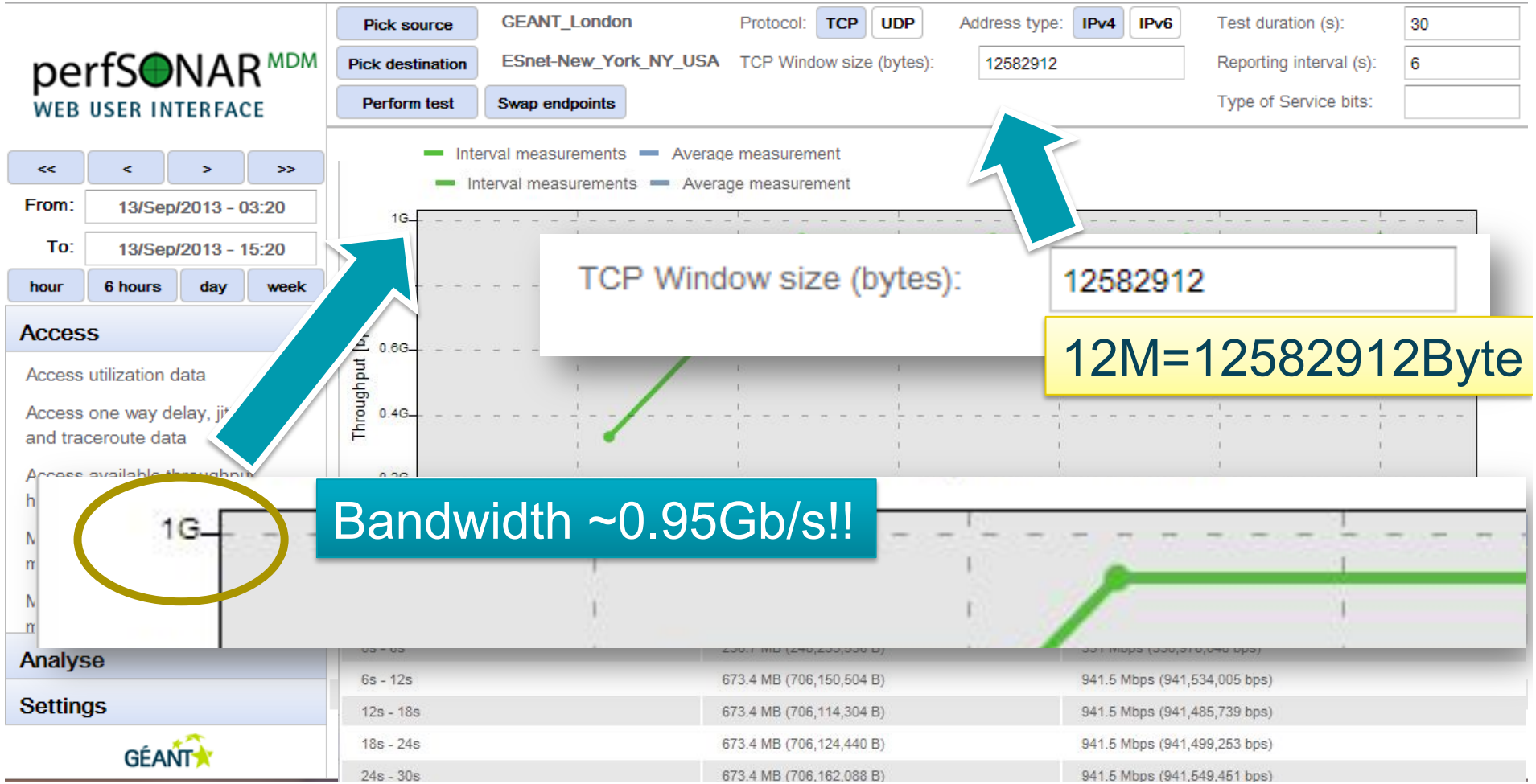
- RTT: 100ms
- 1Gb/s network interface

Calculating Bandwidth Delay Product

- BDP=$100 \cdot 10^{-3} \cdot 1 \cdot 10^{9} = 100 \cdot 10^{6} = 12 MByte$
- Optimal TCP window = BDP
  - TCP window = **12MByte for London-New York**



London-New York
12Mbyte Buffer



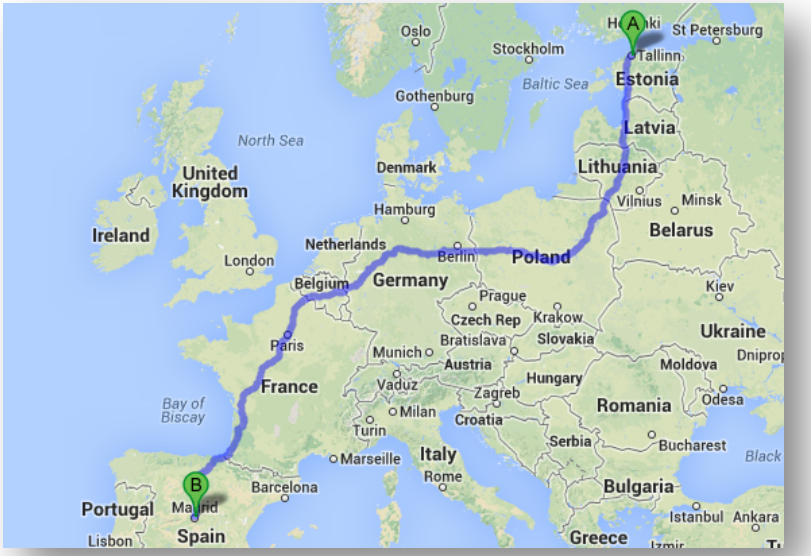
Live demo: Madrid-Tallin

Optimal TCP Window size: Madrid-Tallinn

- RTT: 54ms
- 1Gb/s network interface

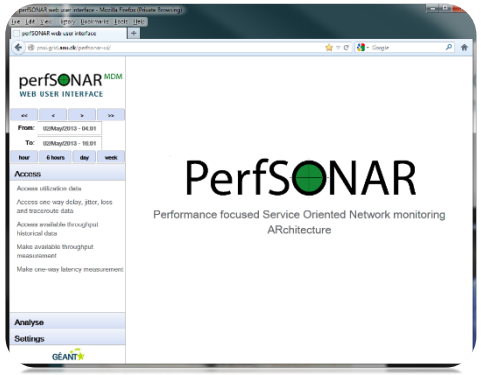
Calculating Bandwidth Delay Product

- BDP=$54 \cdot 10^{-3} \cdot 1 \cdot 10^{9} = 54 \cdot 10^{6} = 54 Mbit = 6.43 MByte$
- Optimal TCP window = BDP
  - TCP window = **6.43MByte for Madrid-Tallin**

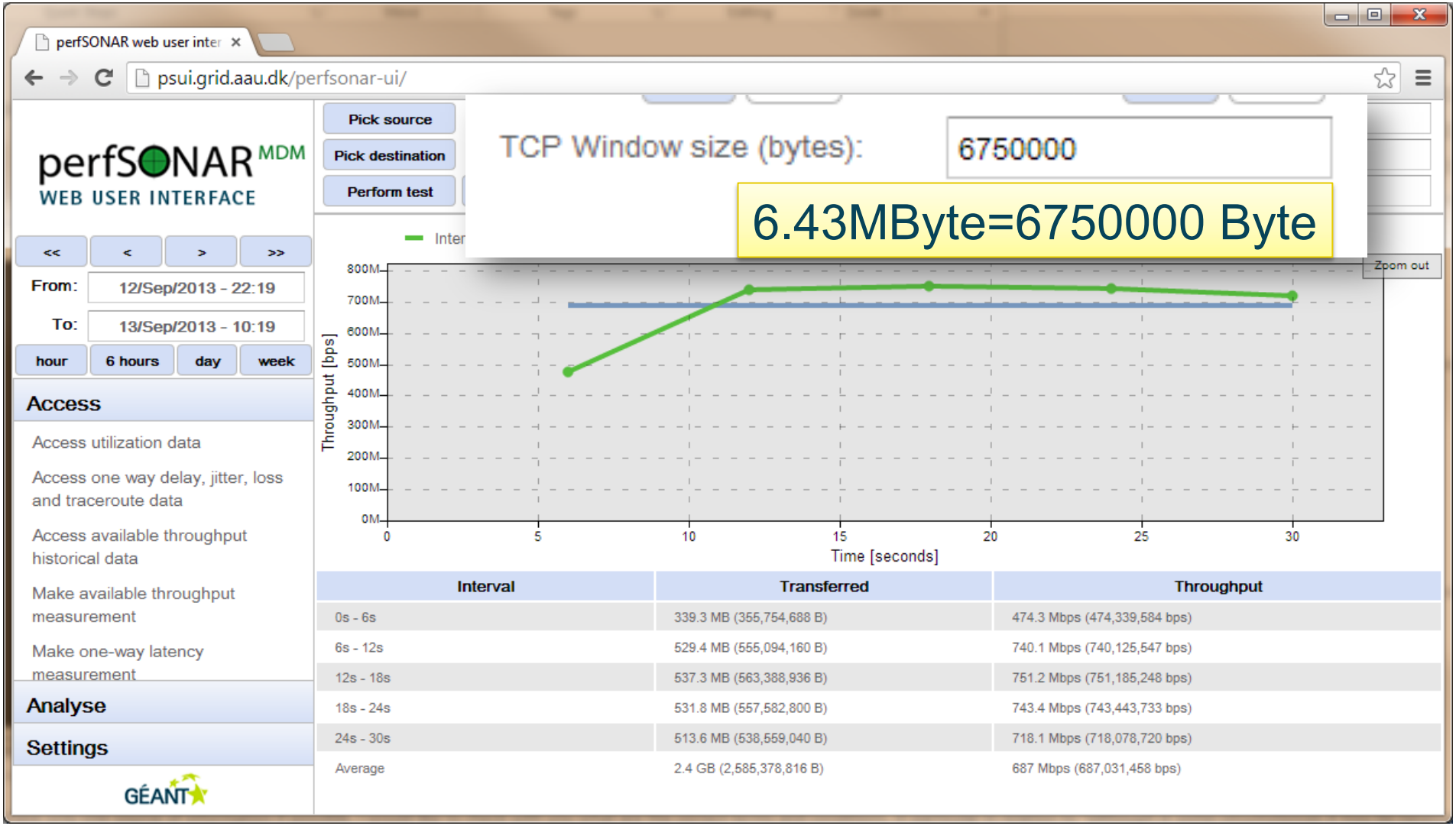


Let's try on the perfSONAR UI!

- We'll set:
  - Source: GEANT Madrid
  - Destination: GEANT Tallinn
  - TCP Window size to 6.43MByte=6750000 Byte



Madrid-Tallinn
Test with optimal TCP window



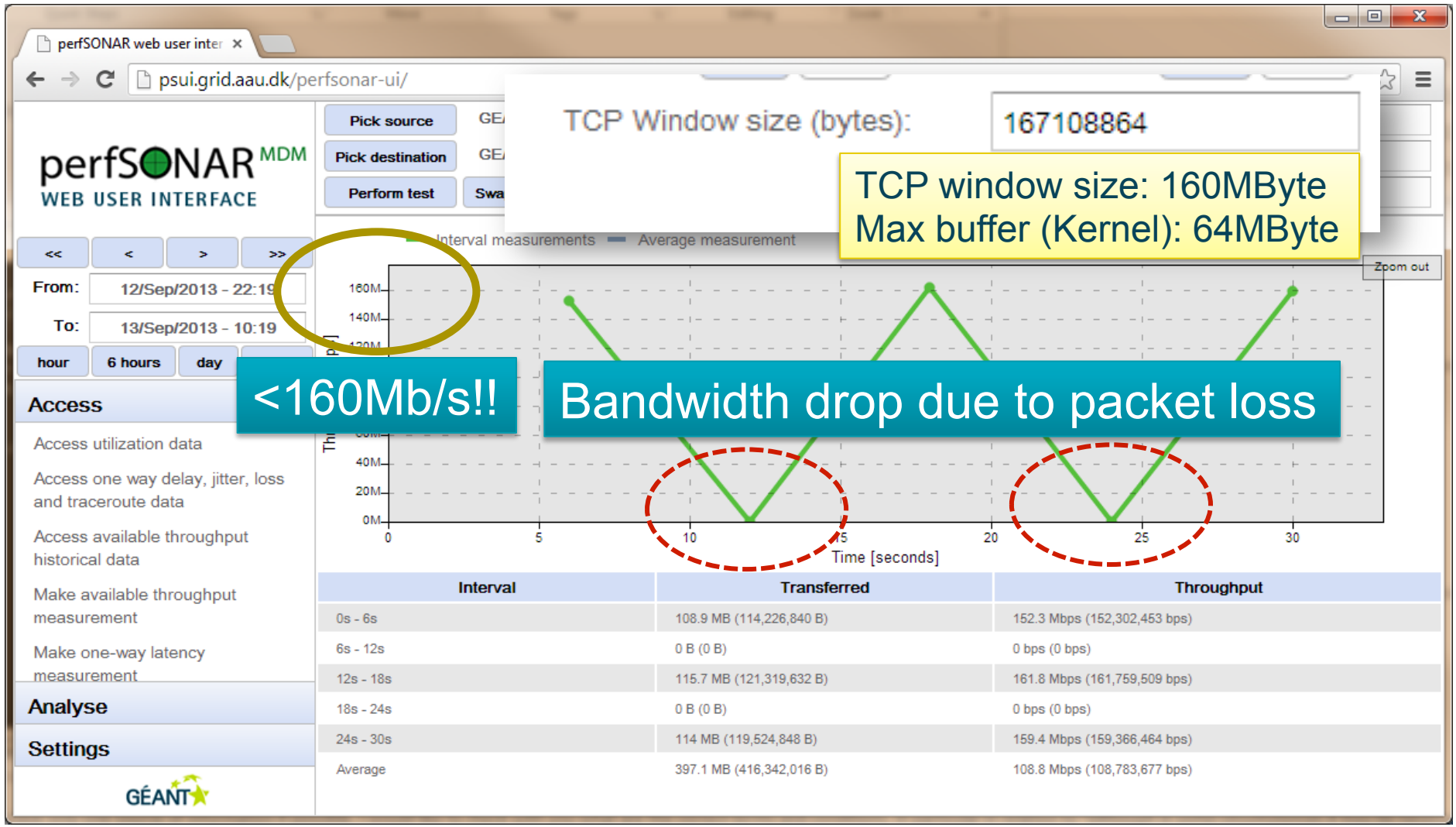
Madrid-Tallinn

- Sometimes we have the possibility to check the buffer set in kernel
- Max receiver (Tallinn) buffer: 67108864 bytes = 64MBytes

```
[root@mp1 ~]# cat /proc/sys/net/core/rmem_max
67108864
[root@mp1 ~]# cat /proc/sys/net/core/wmem_max
67108864
```

- If we set TCP buffers> 67108864 bytes
  - there will be packet loss!
- Let's try!

TCP Window too large → Packet drops



Conclusions

- Buffers are critical in TCP communication
- Not too small
  - But not too big!
- Particular care with long-haul, high-bandwidth network
  - Long Fat Networks, LFN
- Buffers can be the limiting factor
- perfSONAR MDM
  - Add interactive, on-demand testing capability
  - Particularly useful for TCP debugging
- Measurement strategy
  - Calculate the BDP
  - Set the TCP Window size=BDP

Thank you!

Connect | Communicate | Collaborate

www.geant.net

www.twitter.com/GEANTnews | www.facebook.com/GEANTnetwork | www.youtube.com/GEANTtv