

perfSONAR MDM and TCP buffers

Domenico Vicinanza

DANTE, Cambridge, UK

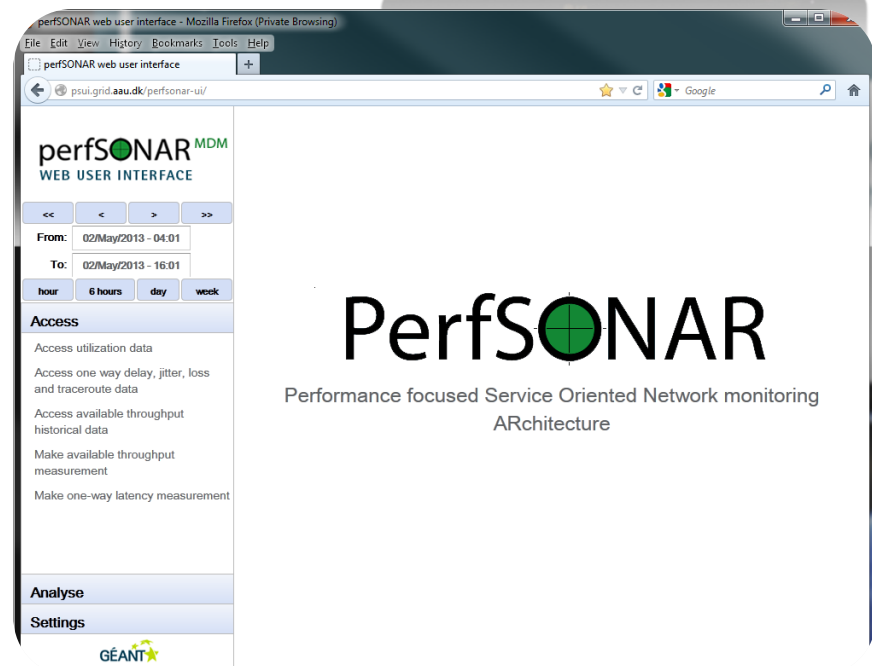
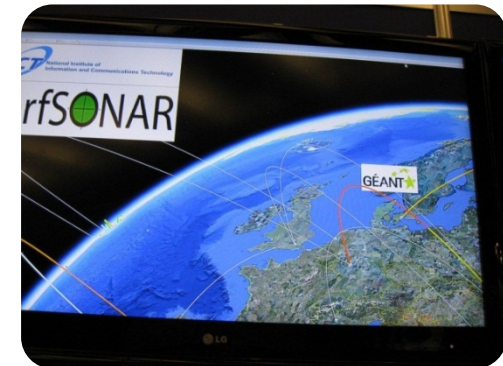
domenico.vicinanza@dante.net

EGI Technical Forum 2013, Madrid, Spain

perfSONAR MDM



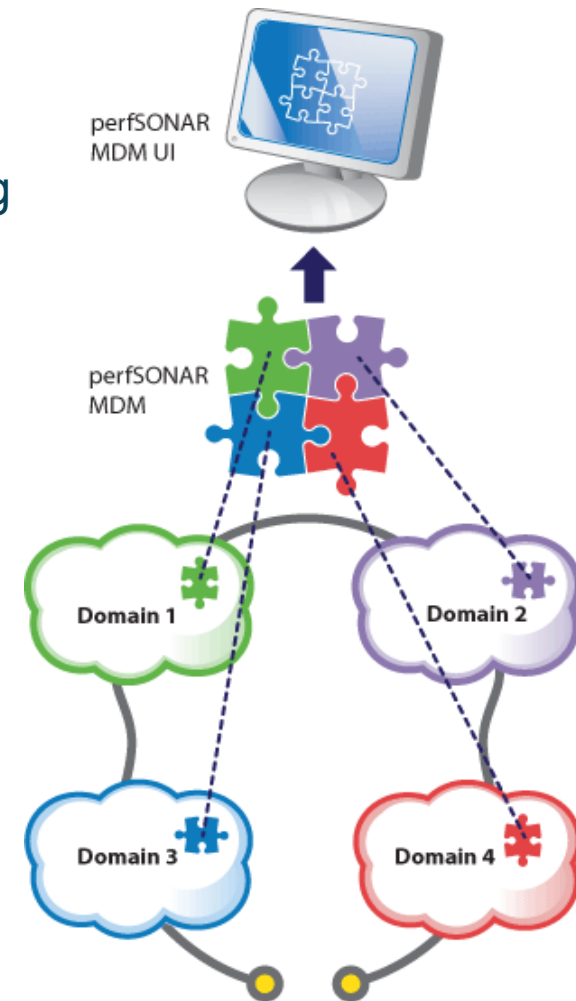
- Multi-domain monitoring service
- Based on monitoring probes installed in the network
- Based on a standard (perfSONAR) protocol
- Hundreds of deployments around the world
- Web interface!



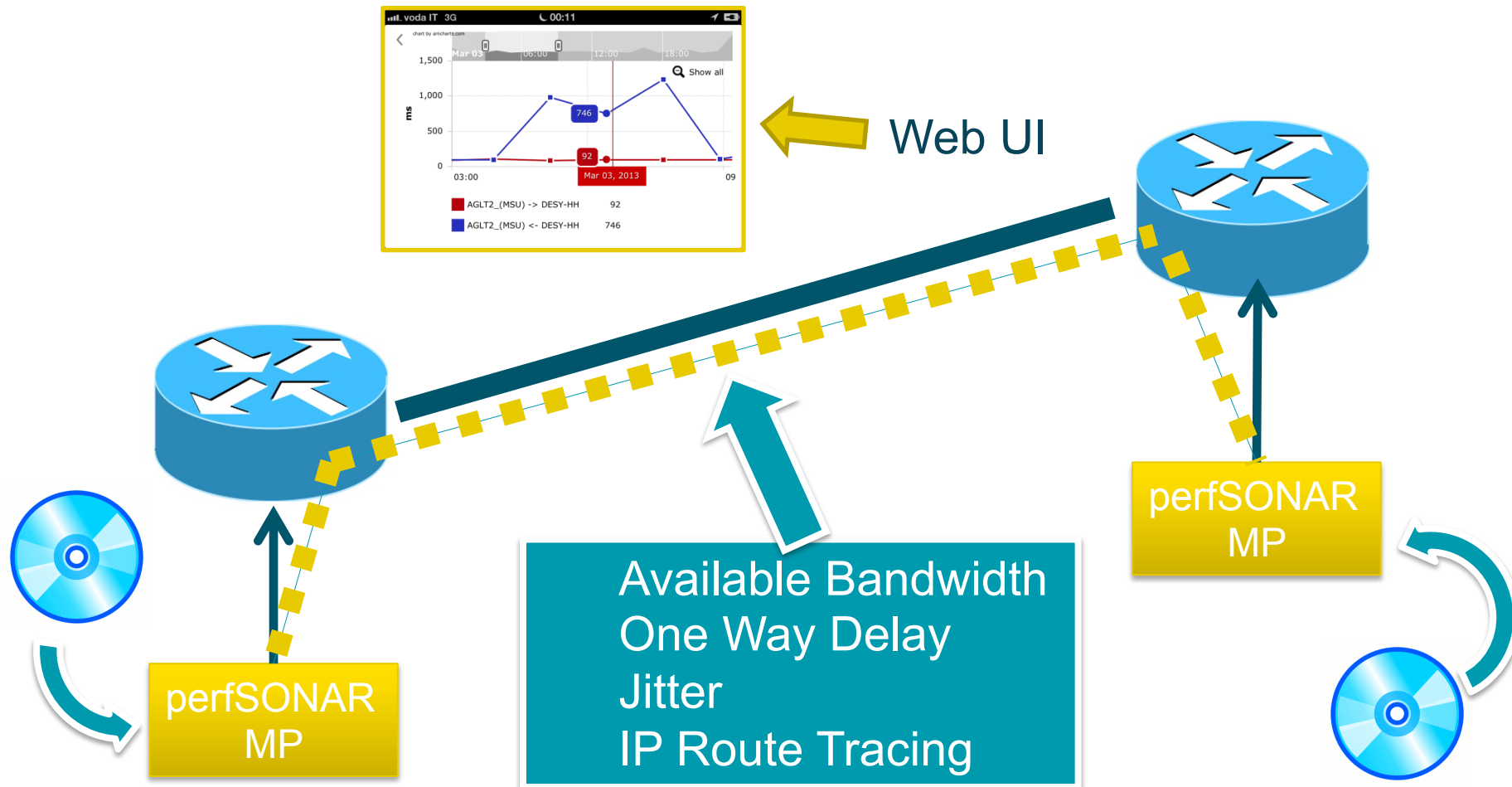
perfSONAR in a nutshell



- **P**erformance focused **S**ervice **O**riented **N**etwork monitoring **A**rchitecture
- 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



A screenshot of the perfSONAR web user interface in a browser window. The browser's address bar shows the URL "psui.grid.aau.dk/perfsonar-ui/". The page header includes the "perfSONAR MDM WEB USER INTERFACE" logo. Below the header, there are navigation buttons: "<<", "<", ">", and ">>". A date range selector shows "From: 12/Sep/2013 - 21:37" and "To: 13/Sep/2013 - 09:37". A time interval selector has buttons for "hour", "6 hours", "day", and "week". The main content area is divided into three sections: "Access", "Analyse", and "Settings". The "Access" section contains three items: "Access utilization data", "Access one way delay, jitter, loss and traceroute data", and "Access available throughput historical data". The "Analyse" section contains two items: "Make available throughput measurement" and "Make one-way latency measurement". The "Settings" section is currently empty. Three callout boxes are overlaid on the interface. The first callout box, with a teal arrow pointing to the "Access" section, lists: "Access utilization data", "Access one way delay, jitter, loss and traceroute data", and "Access available throughput historical data". The second callout box, with a teal arrow pointing to the "Analyse" section, lists: "Make available throughput measurement" and "Make one-way latency measurement". The third callout box, with a yellow arrow pointing to the "Analyse" section, lists: "Make available throughput measurement" and "Make one-way latency measurement".

Using perfSONAR UI to run TCP measurements



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

Make available throughput measurement

A screenshot of the perfSONAR MDM Web User Interface. The interface has a header with the "perfSONAR MDM" logo and "WEB USER INTERFACE" text. Below the header are navigation buttons: "<<", "<", ">", and ">>". There are two date/time input fields: "From: 02/Sep/2013 - 20:49" and "To: 03/Sep/2013 - 08:49". Below these are four buttons: "hour", "6 hours", "day", and "week". A section titled "Access" contains a list of measurement options: "Access utilization data", "Access one way delay, jitter, loss and traceroute data", "Access available throughput historical data", "Make available throughput measurement", and "Make one-way latency measurement".

perfSONAR^{MDM}
WEB USER INTERFACE

<< < > >>

From: 02/Sep/2013 - 20:49

To: 03/Sep/2013 - 08:49

hour 6 hours day week

Access

- Access utilization data
- Access one way delay, jitter, loss and traceroute data
- Access available throughput historical data
- Make available throughput measurement
- Make one-way latency measurement

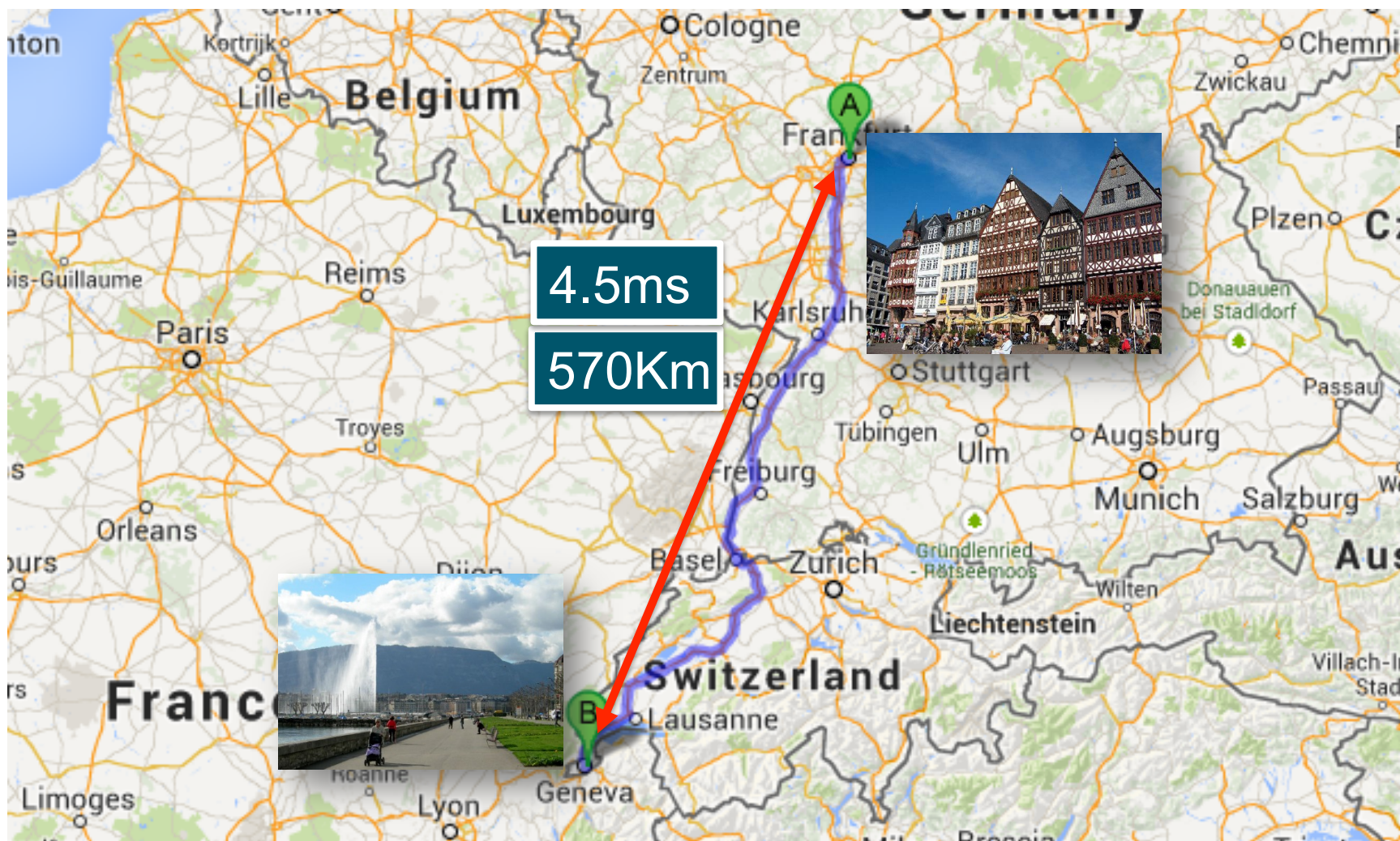
London-Frankfurt 1Kbyte buffer



Pick source	GEANT_London	Protocol:	TCP	UDP	Address type:	IPv4	IPv6	Test duration (s):	30
Pick destination	GEANT_Frankfurt	TCP Window size (bytes):		1024		Reporting interval (s):		6	
Perform test	Swap endpoints	Type of Service bits:							



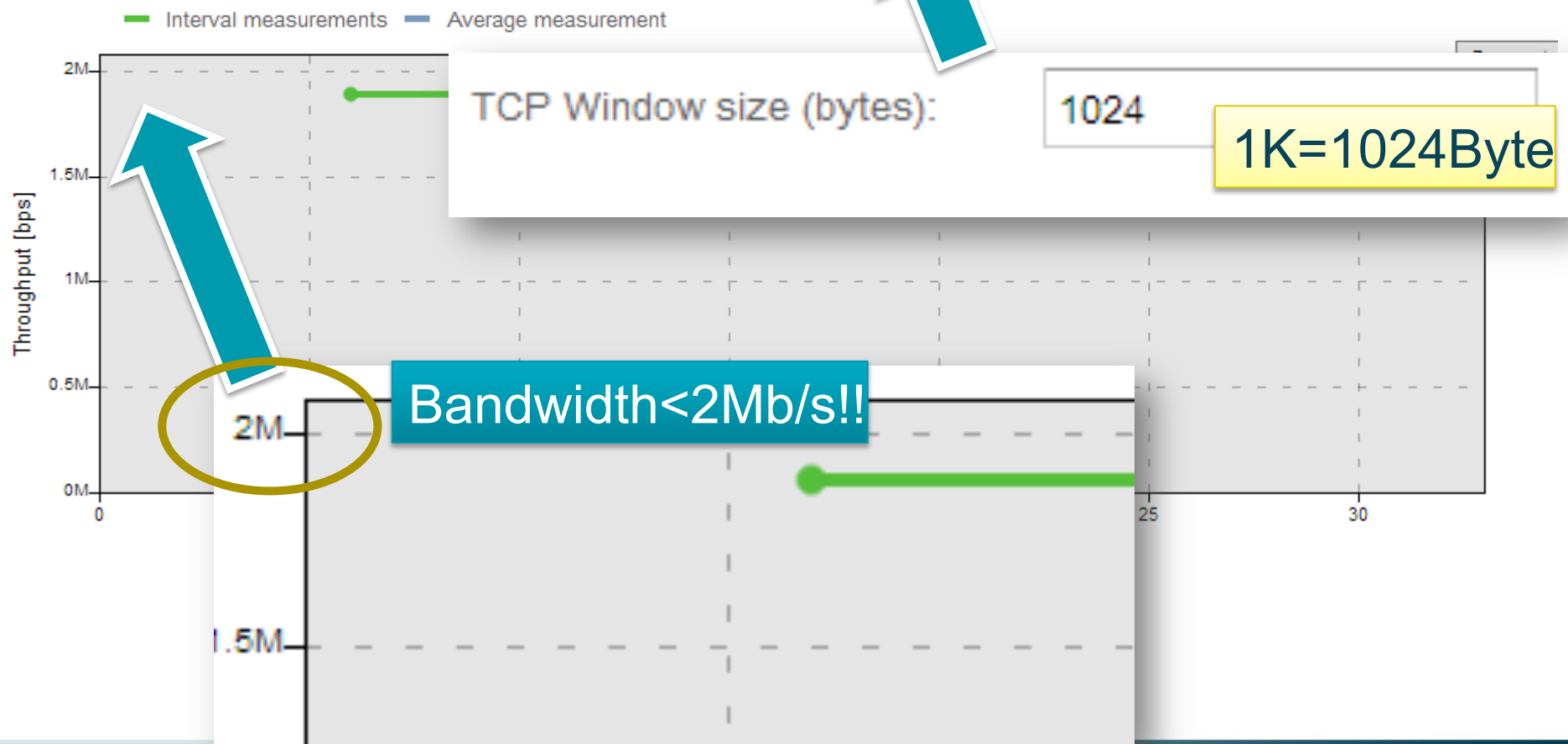
First measurement: Geneva-Frankfurt



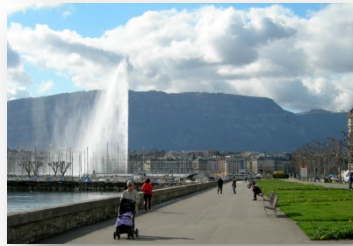
Geneva-Frankfurt 1Kbyte buffer



Pick source	GEANT_Geneva	Protocol:	☒ TCP ☐ UDP	Address type:	☒ IPv4 ☐ IPv6	Test duration (s):	30
Pick destination	GEANT_Frankfurt	TCP Window size (bytes):	1024		Reporting interval (s):	6	
Perform test	Swap endpoints	Type of Service bits:					



What happens with TCP buffer=1KB?



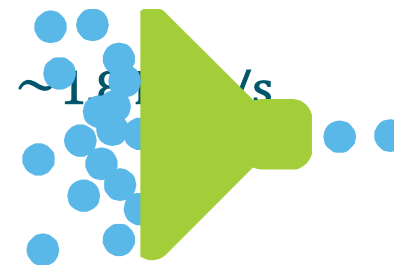
Geneva

Delay: 4.5ms
1024 Byte



Frankfurt

- 1024Bytes sent every 4.5ms
- $\text{Max Throughput} \leq (\text{TCP buffer}) / \text{Delay}$
- $\text{Max Throughput} \leq 1024 \cdot 8 / 4.5 \cdot 10^{-3} \sim 1.8 \text{ MB/s}$
- **Bandwidth limited by the buffer!**



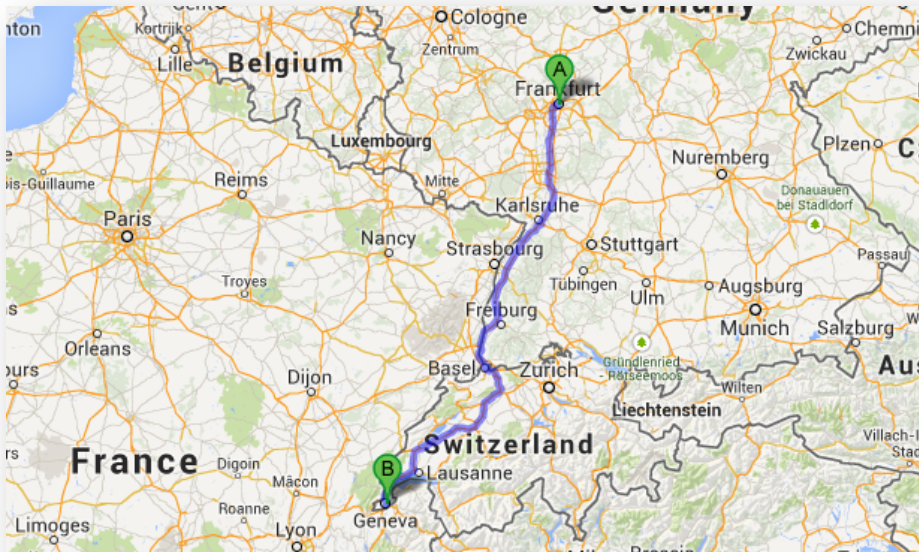
Optimal TCP Window size: Geneva-Frankfurt



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

Calculating Bandwidth Delay Product

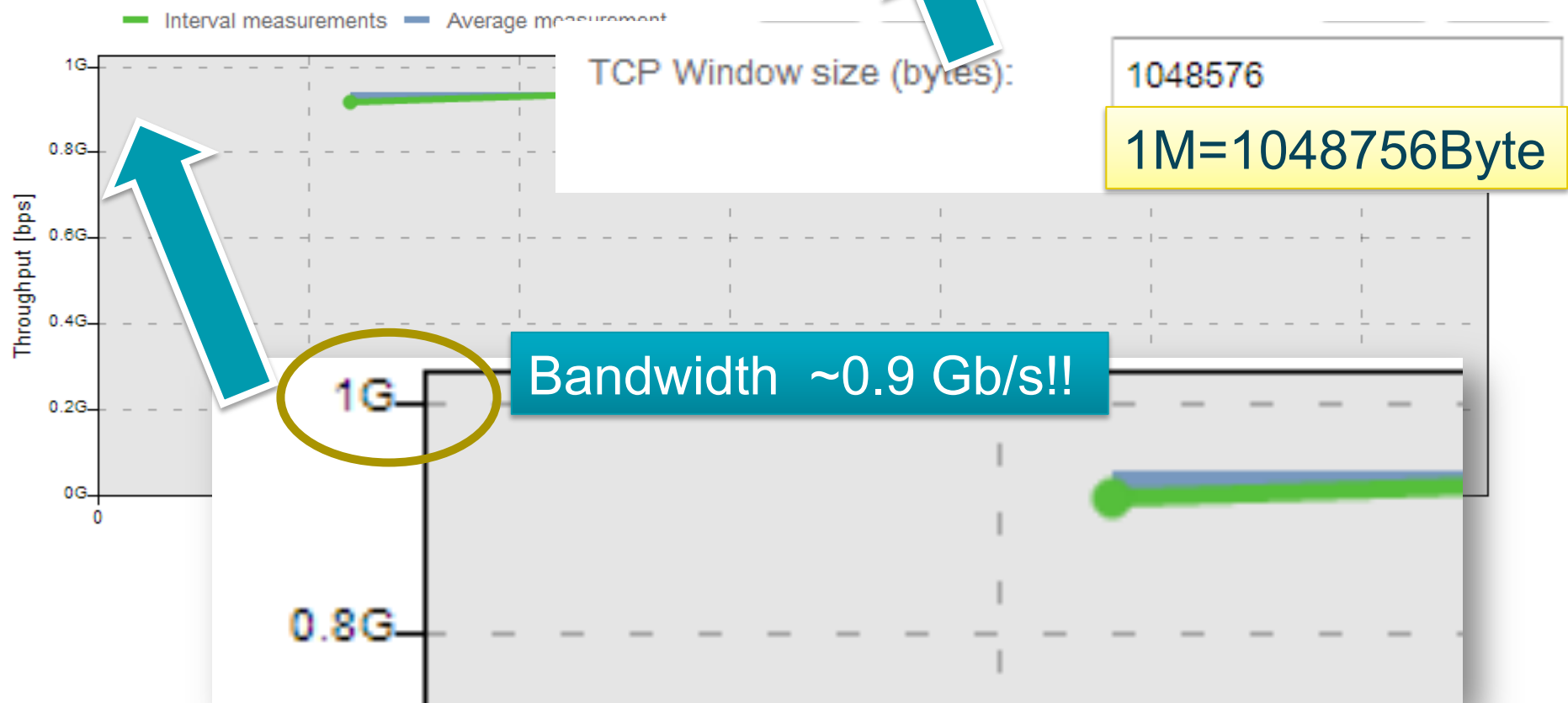
- $BDP = 9 \cdot 10^1 - 3 \cdot 1 \cdot 10^{19} = 9 \cdot 10^{16} = 9 \text{ Mbit} = 1.07 \text{ MByte}$
- Optimal TCP window = BDP
 - TCP window = **1MByte for Geneva-Frankfurt**



Geneva-Frankfurt 1Mbyte TCP Buffer



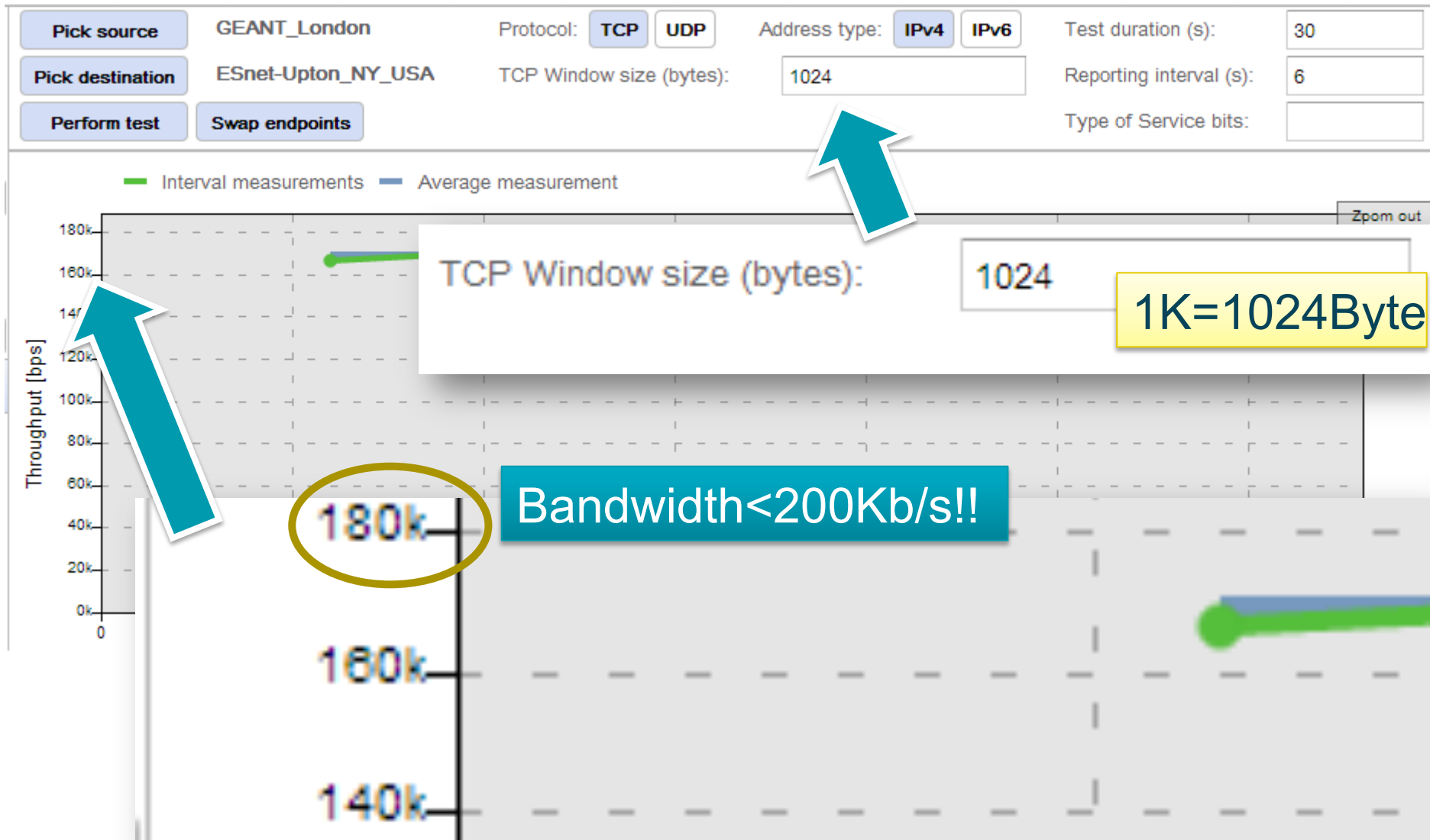
Pick source	GEANT_Geneva	Protocol:	<input checked="" type="button" value="TCP"/> UDP	Address type:	<input checked="" type="button" value="IPv4"/> IPv6	Test duration (s):	30
Pick destination	GEANT_Frankfurt	TCP Window size (bytes):	1048576		Reporting interval (s):	6	
Perform test	Swap endpoints	Type of Service bits:					



GÉANT



London-New York 1Kbyte Buffer



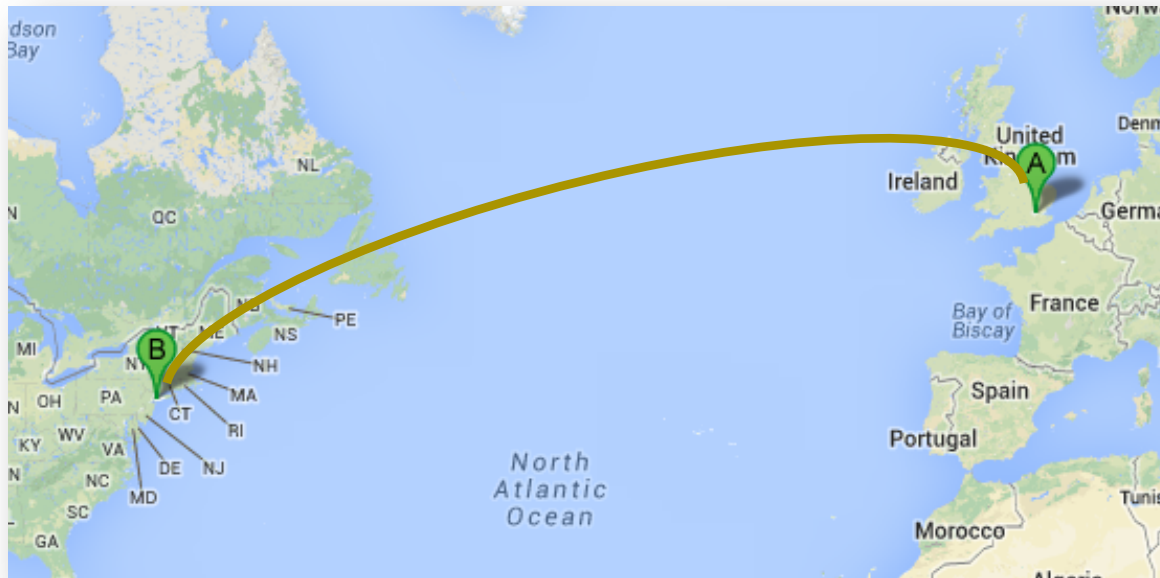
Optimal TCP Window size: London-New York



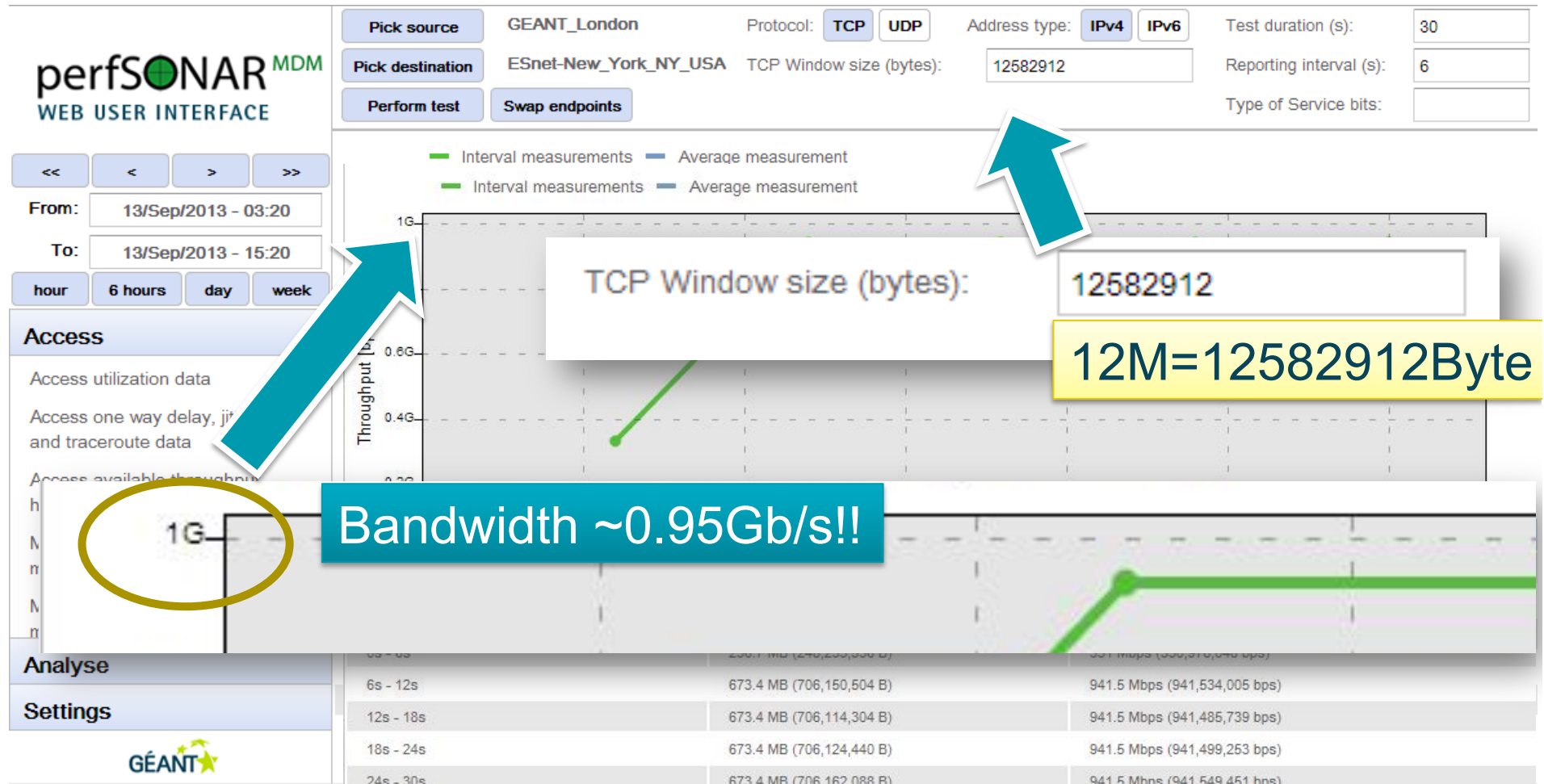
- RTT: 100ms
- 1Gb/s network interface

Calculating Bandwidth Delay Product

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



London-New York 12Mbyte Buffer



Live demo: Madrid-Tallin



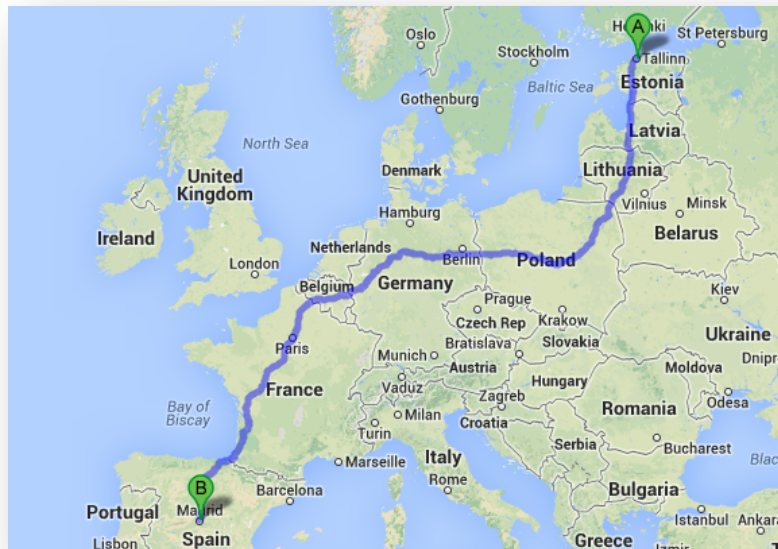
Optimal TCP Window size: Madrid-Tallinn



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

Calculating Bandwidth Delay Product

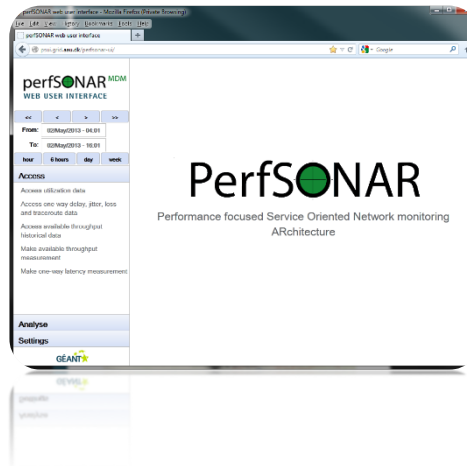
- $BDP = 54 \cdot 10^{-3} \cdot 1 \cdot 10^9 = 54 \cdot 10^6 = 54 \text{ Mbit} = 6.43 \text{ MByte}$
- Optimal TCP window = BDP
 - TCP window = **6.43MByte for Madrid-Tallin**



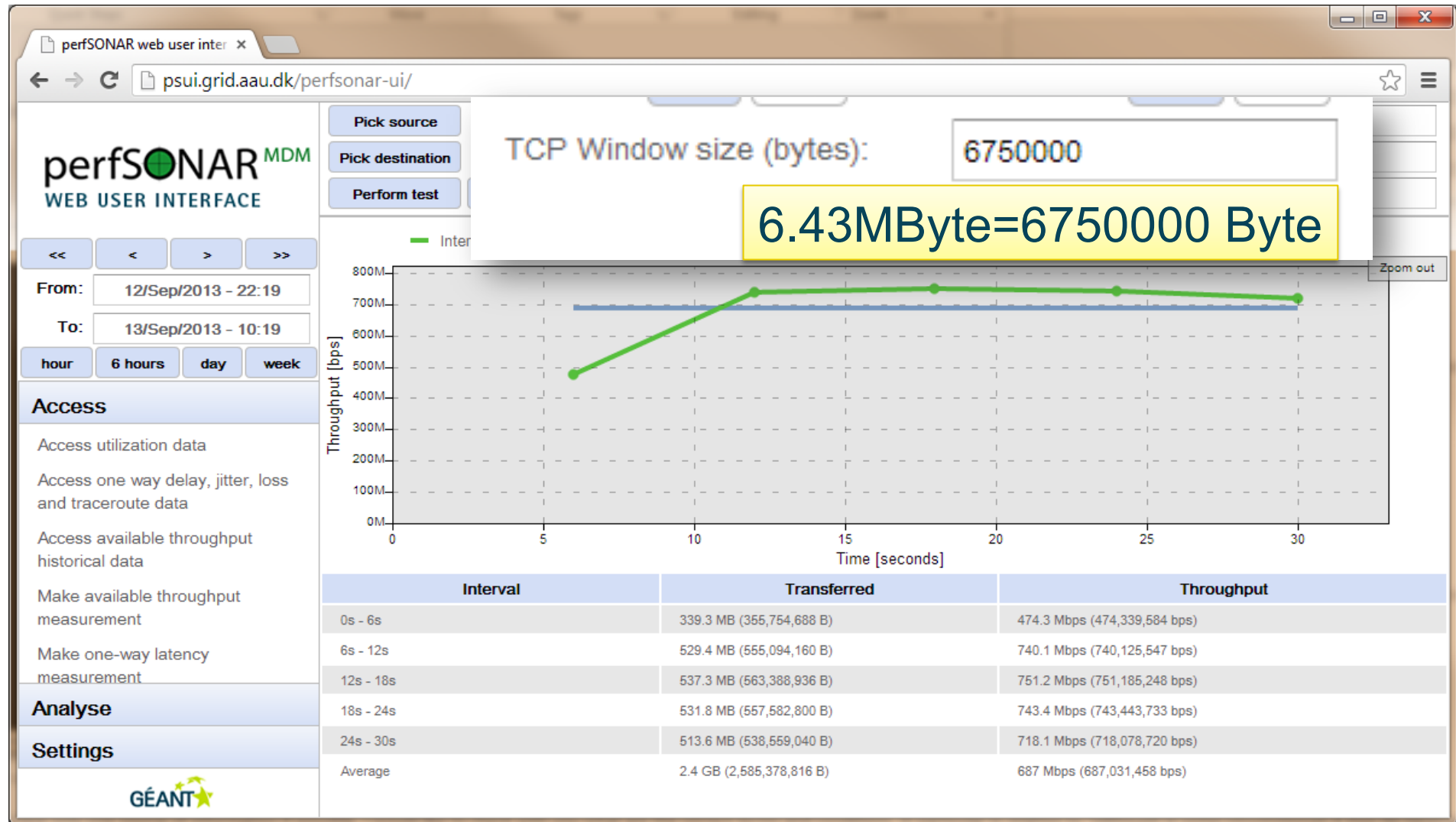
Let's try on the perfSONAR UI!



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



Madrid-Tallinn Test with optimal TCP window

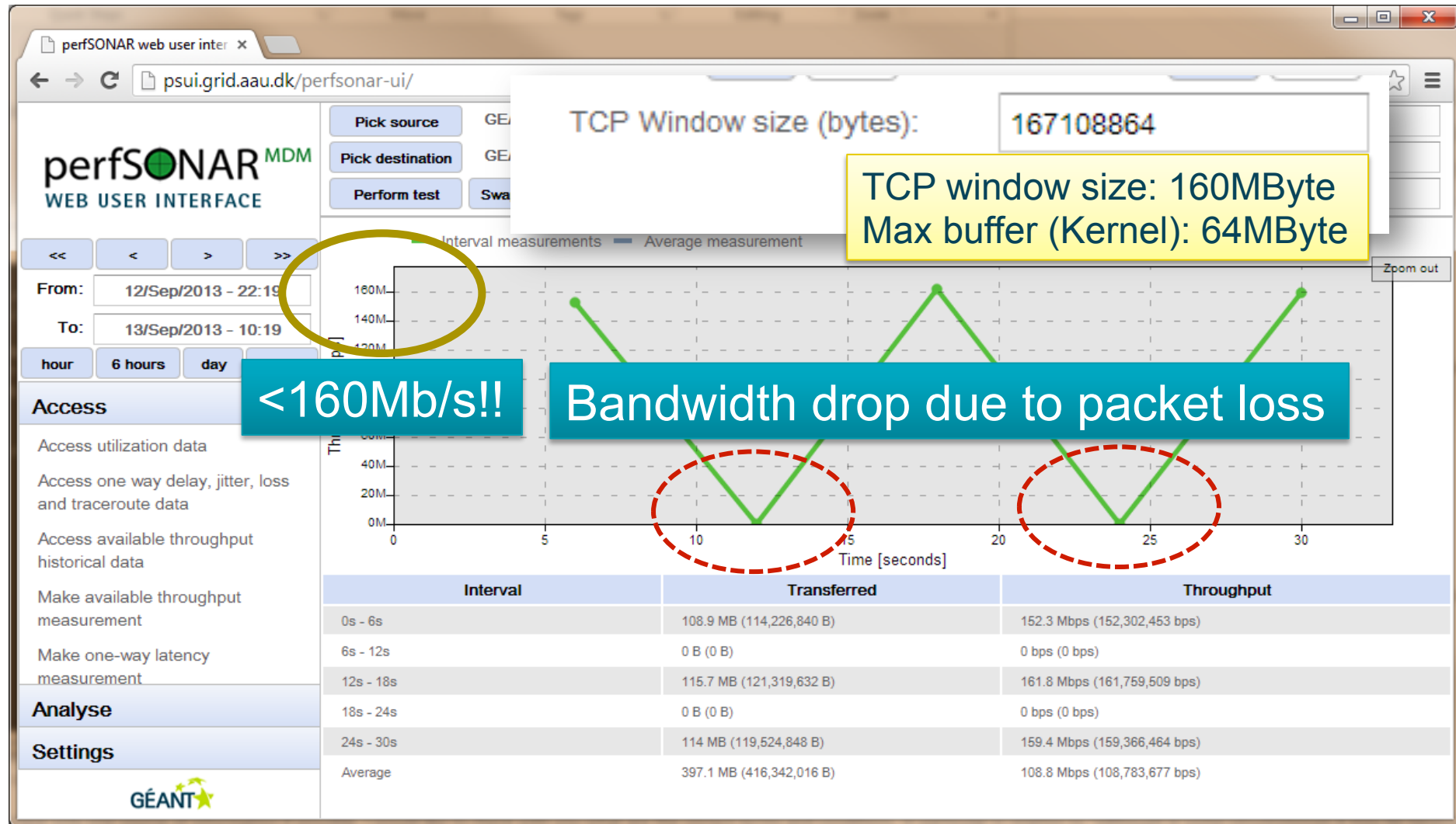


- Sometimes we have the possibility to check the buffer set in kernel
- Max receiver (Tallinn) buffer: 67108864 bytes = 64MBytes

```
[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

