Verification of High Data Rate Bandwidth-on-Demand networks: User Based Test Equipment

Jimmy Cullen

Research Associate, University of Manchester

MANCHESTER 1824

The University of Manchester

NEXPReS

Contents

- NEXPReS
- Very Long Baseline Interferometry
- Test equipment
- BoD tests
- Conclusions

What is NEXPReS?

- NEXPReS = Novel EXplorations Pushing Robust e-VLBI Services
- Three years (start 1 July 2010)
- Funded through the European Community's Seventh Framework Programme (FP7/2007-2013), Contract n°: RI-261525
- Budget: 5,745,000 € (EC contribution: 3,500,000 €)
- Objective: further improve the astronomy technique of electronic Very Long Baseline Interferometry (e-VLBI) with the objective of incorporating it into every experiment conducted by the European VLBI Network (EVN)
- Means: develop data caching and implement dynamically provisioned network resources to offer the best of both worlds: the data archiving and re-processing afforded by traditional disk-based VLBI and the speed and flexibility of e-VLBI

NEXPReS Partners

Coordinator

- Joint Institute for VLBI in Europe (JIVE), EU (The Netherlands)

National Astronomy Institutes

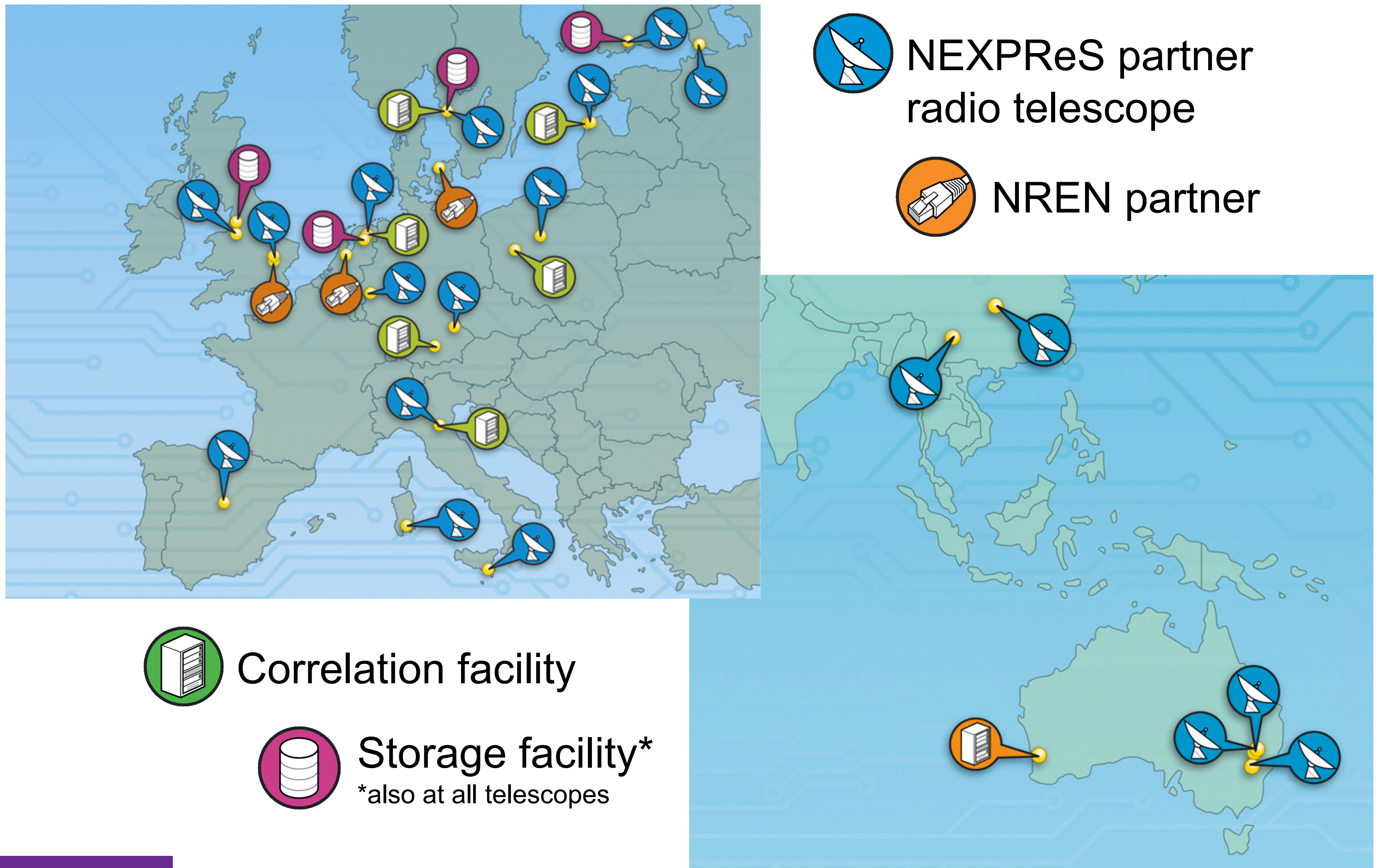
- The Netherlands Institute for Radio Astronomy (ASTRON), The Netherlands
- Istituto Nazionale di Astrofisica (INAF), Italy
- Max Planck Gesellschaft zur Foerderung der Wissenschaften E.V. (MPG), Germany
- The University of Manchester (UMAN), United Kingdom
- Chalmers Tekniska Hoegskola AB (OSO), Sweden
- Ventspils Augstskola (VENT), Latvia
- Fundación General de la Universidad de Alcalá, together with Instituto Geográfico Nacional (FG-IGN), Spain
- Aalto University Metsähovi Radio Observatory (AALTO), Finland
- Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia

NREN Providers and Advanced Computing Facilities

- NORDUnet A/S (NORDUnet), Denmark
- SURFnet bv (SURFnet), The Netherlands
- Poznan Supercomputing and Networking Center (PSNC), Poland
- Delivery of Advanced Network Technology to Europe Limited (DANTE), EU (United Kingdom)
- Technische Universität München (TUM), Germany

MANCHESTER 1824
The University of Manchester

NEXPReS Infrastructure



Activities

#	Description	Leader
WP1	Management of the Consortium	T. Charles Yun, JIVE
WP2	EVN-NREN	Richard Hughes-Jones, DANTE
WP3	eVSAG	Francisco Colomer, FG
WP4	Communication	Kristine Yun, JIVE
WP5	Cloud Correlation	Arpad Szomoru, JIVE
WP6	High Bandwidth on Demand	Paul Boven, JIVE
WP7	Computing in a Shared Infrastructure	Mark Kettenis, JIVE
WP8	Provisioning High-Bandwidth, High-Capacity Networked Storage on Demand	Ari Mujunen, Aalto

MANCHESTER 1824
The University of Manchester

Interferometry

- Angular resolution ≈ wavelength / collecting area diameter
- For fixed wavelengths (radio), must increase collecting area (dish size) to gain higher resolution images
- Lovell Telescope (JBO) D = 76m

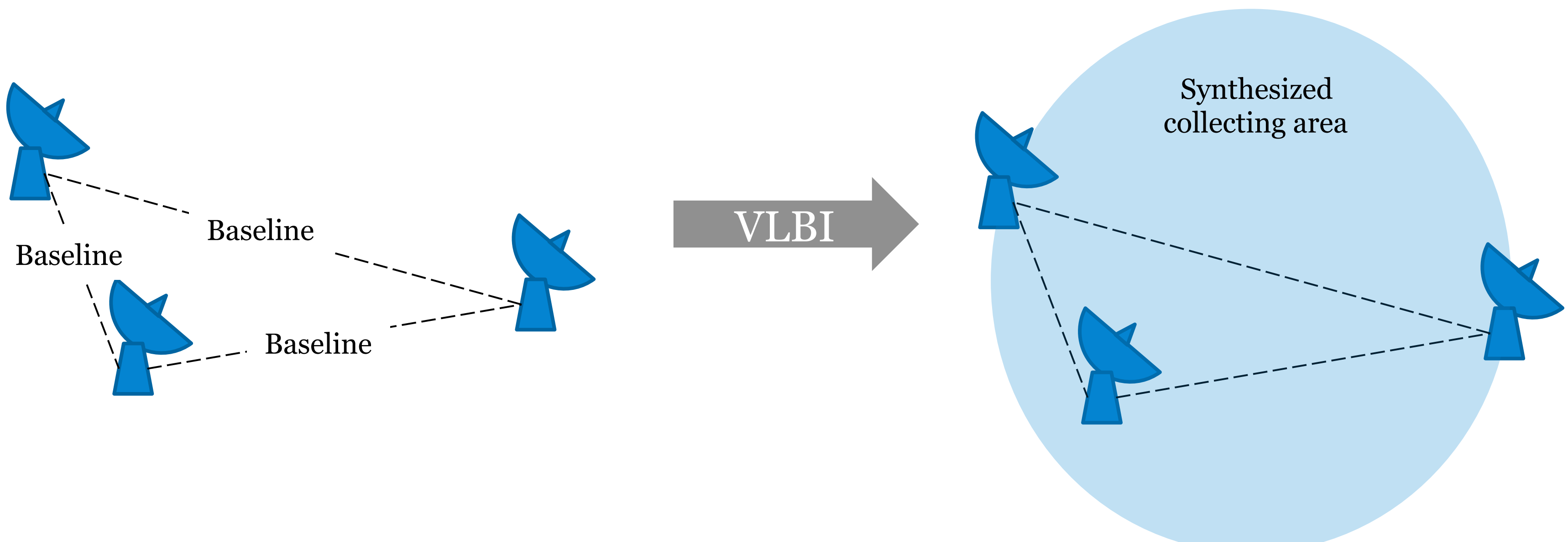
$$\sin\theta \approx 1.220\frac{\lambda}{D}$$

- Interferometry combines signals from two or more telescopes to generate images
- Resolution proportional to the largest separation of dishes, not individual dish size

MANCHESTER 1824
The University of Manchester

Very Long Baseline Interferometry (VLBI)

- Distance between individual telescopes described as baselines
- Baselines are often hundreds or thousands of km - transglobal



VLBI and eVLBI

- Large volumes of data recorded at each telescope, typically 1 Gbps or higher for several hours (8 – 12 hours)
- Data are stored on hard disks, then shipped to a central correlator for processing
- Observation results can take weeks to be processed

- Electronic VLBI (eVLBI) uses dedicated light paths from observatories to the correlator
  - Allows faster processing of data (⇒ observation results)
  - Realtime monitoring of telescopes and error detection and correction
  - Enabled new science – Target of Opportunity (ToS) observations

- VLBI observations infrequent however, therefore inefficient use of the light paths (~ once a month)

Bandwidth on Demand and eVLBI

- BoD offers an ideal solution for eVLBI:
  - Resources needed only for limited periods of time
  - Dedicated links required since using UDP

- NEXPReS, the NRENs and GÉANT aim to provide a BoD service for eVLBI in Europe

- Essential to verify connectivity, bandwidth and packet loss on BoD links prior to observation

Test equipment

- Tests to evaluate different hardware approaches to characterise networks
- Evaluated two solutions for characterising networks
  - FPGAs
  - PCs

MANCHESTER 1824
The University of Manchester

Test Network

- A dedicated 10 Gbps light path between Jodrell Bank Observatory and main university campus (Schuster Building, physics)
- ~30km direct distance, but fibre is ~80km
- No other users, uncontended bandwidth

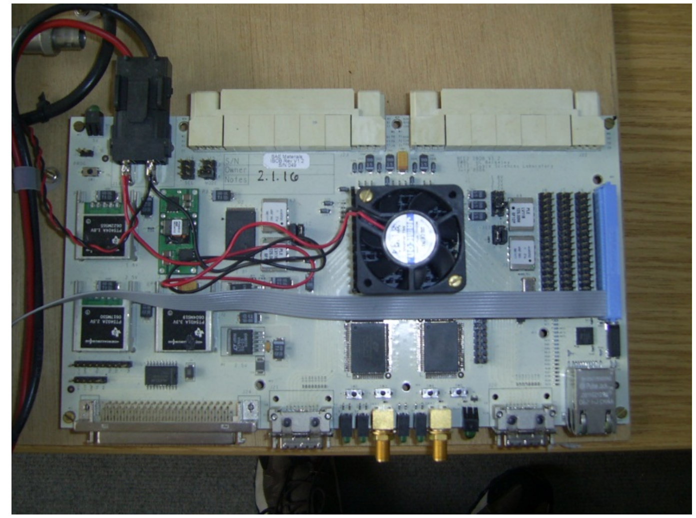
Schuster

JBO

MANCHESTER 1824
The University of Manchester

Test Hardware - FPGA

- Interconnect Break-out Board (IBOB, from CASPER, Berkley)
- Xilinx Virtex-II Pro 2VP50 Field Programmable Gate Array
- 200MHz PowerPC
- 2 x CX4 10 Gbps
- 1 x RJ45 100 Mbps interfaces



Test Hardware - PCs

- Asus Crosshair IV Formula motherboard
- AMD Phenom(tm) II X6 1090T Processor
- 4 x Hynix 4GB DDR3 PC3-10600 (1333)
- Chelsio N310E-CXA 10GbE NIC and Myricom 10G-PCIE-8A-C NIC.

MANCHESTER 1824
The University of Manchester

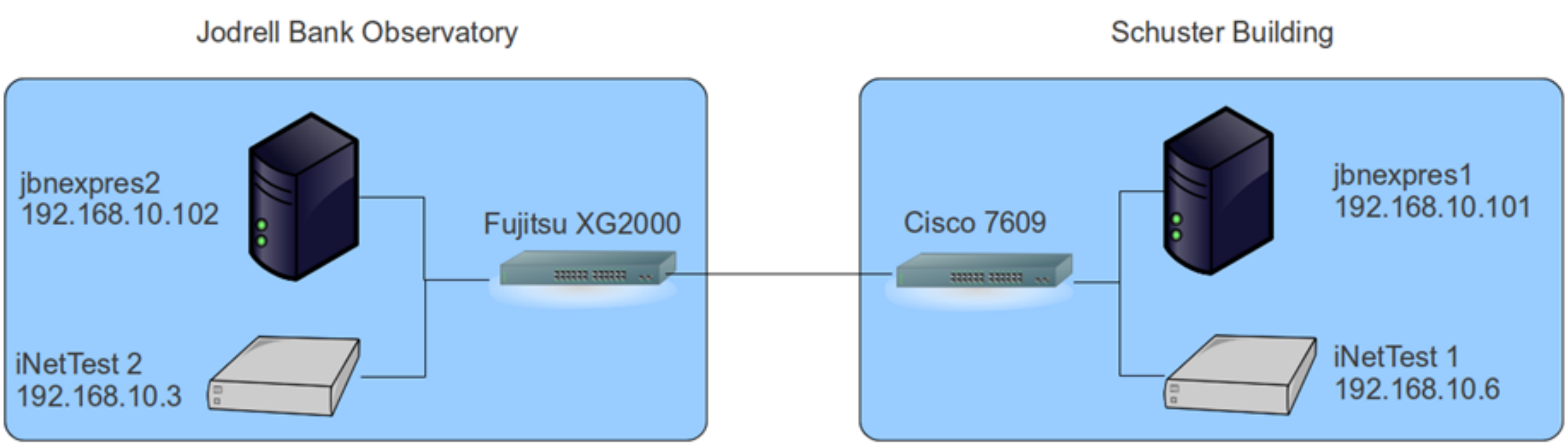
Test Software

- UDPmon* is a network diagnostic program which uses UDP datagrams to test endhost and network performance
- Client/server model written in C
- Measures many aspects of network communication including packet loss, packet reordering and variation in interpacket arrival times (jitter)
- iNetTest is a control framework designed to control the “gateware” in the IBOB in a manor compatible with UDPmon

*http://www.hep.man.ac.uk/u/rich/net/index.html

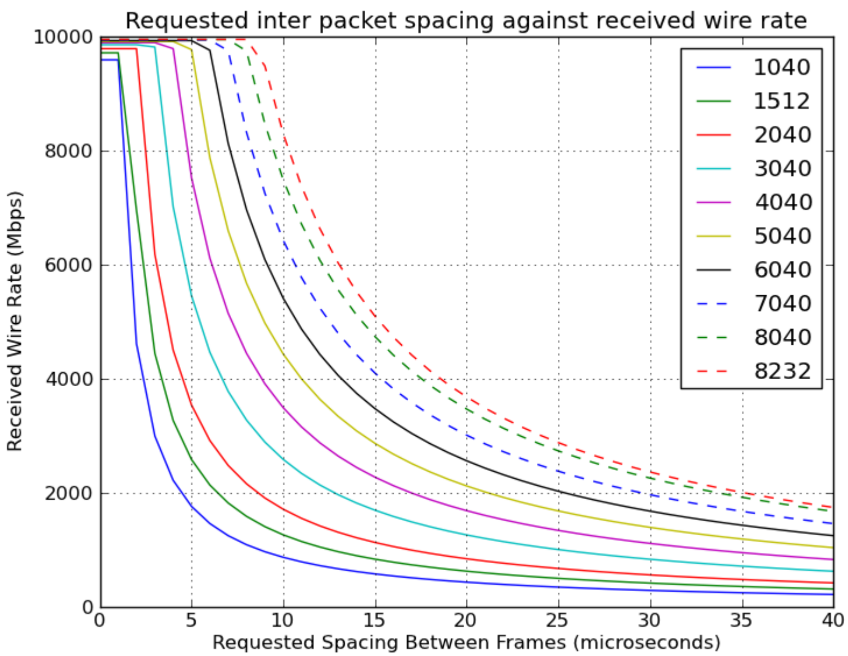
Test setup

- Two PCs and two IBOBs used
  - Manchester – jbnexpres1 & iNetTest1
  - JBO – jbnexpres2 & iNetTest2

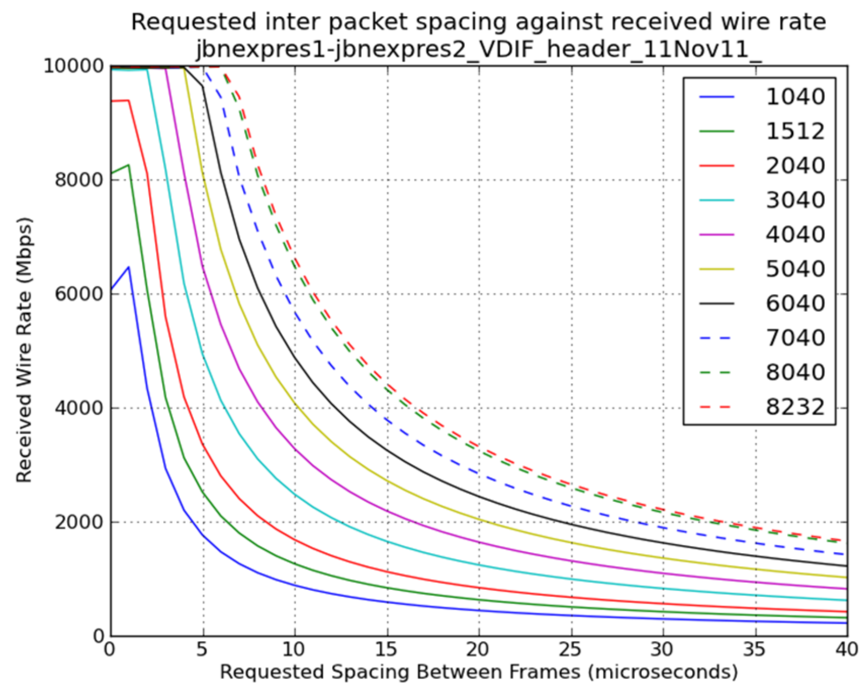


Received Wire Rate

- Received wire rates measured for 10 different packet sizes
- At small requested inter-packet separation, steady data rate
- As separation increases, data rate decreases (inverse power law)



IBOB



PC

MANCHESTER 1824
The University of Manchester

Received Wire Rate (2)

- Both hardware achieved > 9.9 Gbps data rates
- At smaller packet sizes, the hardware must process more packets per second to achieve high data rates
- IBOBs were able to achieve higher data rates than PCs for small packet sizes
- PCs achieved higher data rates than IBOBs
  - IBOB has a 200 MHz processor ⇒ 5 ns temporal resolution
  - PC has a 3.2 GHZ processor ⇒ ~0.3 ns temporal resolution

Received Wire Rate (3)

- IBOB to IBOB

Packet Size (bytes)	Number of Packets	Time between sending packets (nanoseconds)	Requested inter-packet delay (microseconds)	Mean Time Between receiving successive packets (nanoseconds)	Received Wire Rate (Mbps)
8232	100000	6643.05	0	6643.13	9944.7
8232	100000	6643.05	1	6643.13	9944.7
8232	100000	6643.05	2	6643.13	9944.7
8232	100000	6643.05	3	6643.13	9944.7
8232	100000	6643.06	4	6643.13	9944.7
8232	100000	6643.06	5	6643.13	9944.71
8232	100000	6643.06	6	6643.13	9944.7
8232	100000	6643.06	7	6643.13	9944.7
8232	100000	6643.06	8	6643.13	9944.7
8232	100000	6969.93	9	6969.95	9478.41
8232	100000	7969.92	10	7969.94	8289.15

Received Wire Rate (4)

- PC to PC

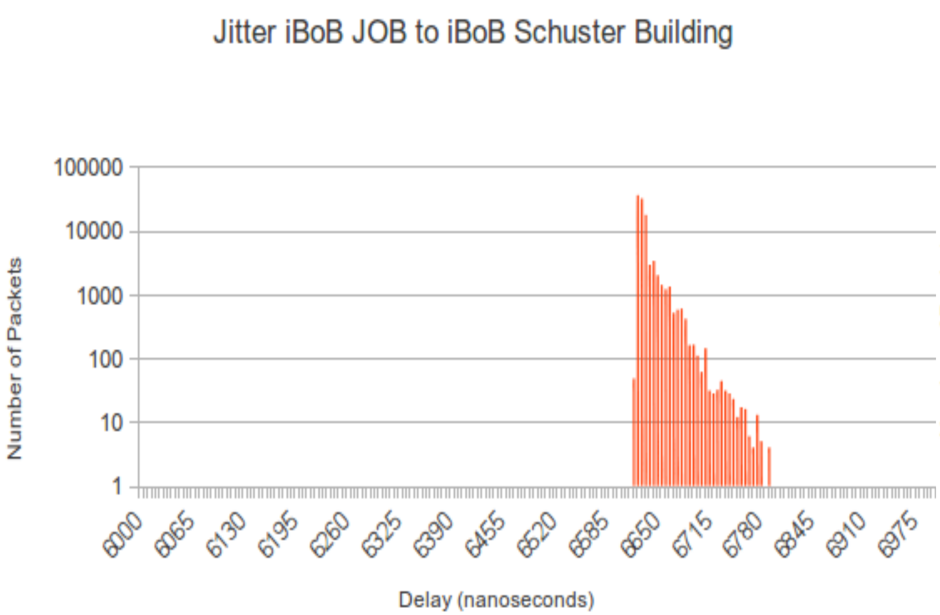
Packet Size (bytes)	Number of Packets	Time between sending packets (nanoseconds)	Requested inter-packet delay (microseconds)	Mean Time Between receiving successive packets (nanoseconds)	Received Wire Rate (Mbps)
8232	100000	6001.6	0	6649.52	9983.28
8232	100000	6288.5	1	6655.61	9974.14
8232	100000	6314.3	2	6654.99	9975.07
8232	100000	6084.0	3	6652.61	9978.64
8232	100000	6255.2	4	6656.12	9973.38
8232	100000	6180.4	5	6655.21	9974.74
8232	100000	6117.9	6	6656.16	9973.32
8232	100000	7024.4	7	7027.83	9445.87
8232	100000	8035.6	8	8036.74	8260.07
8232	100000	9019.4	9	9020.85	7358.95
8232	100000	10015.2	10	10010.6	6631.37

Jitter

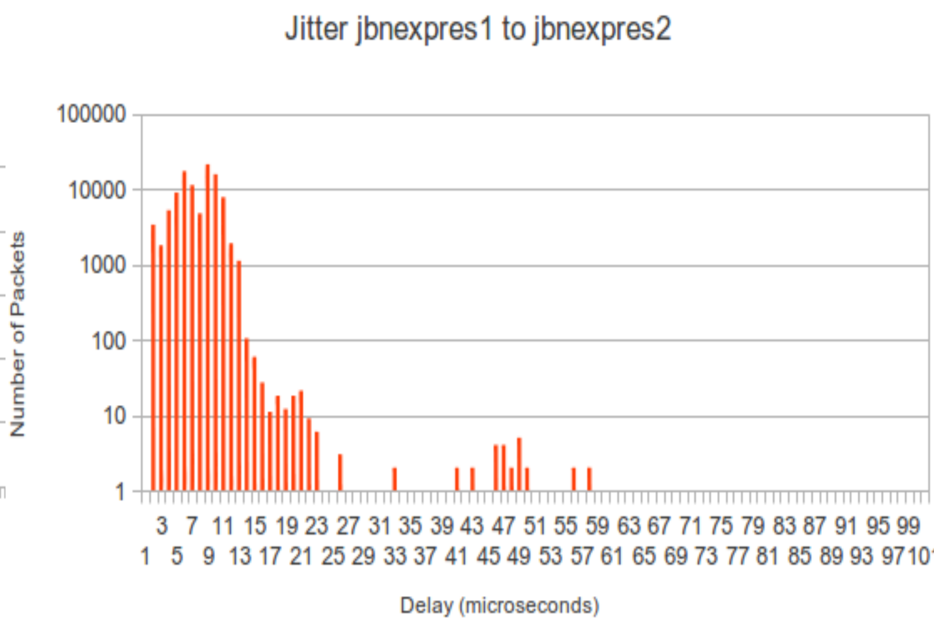
- Defined here as the variability in inter-packet reception times
- Measured jitter for requested inter-packet delays of zero and 50 μs
- Histogram bin widths:
  - IBOBs – 5 nanoseconds
  - PCs – 1 microsecond

Jitter (1)

- Zero inter-packet delay requested



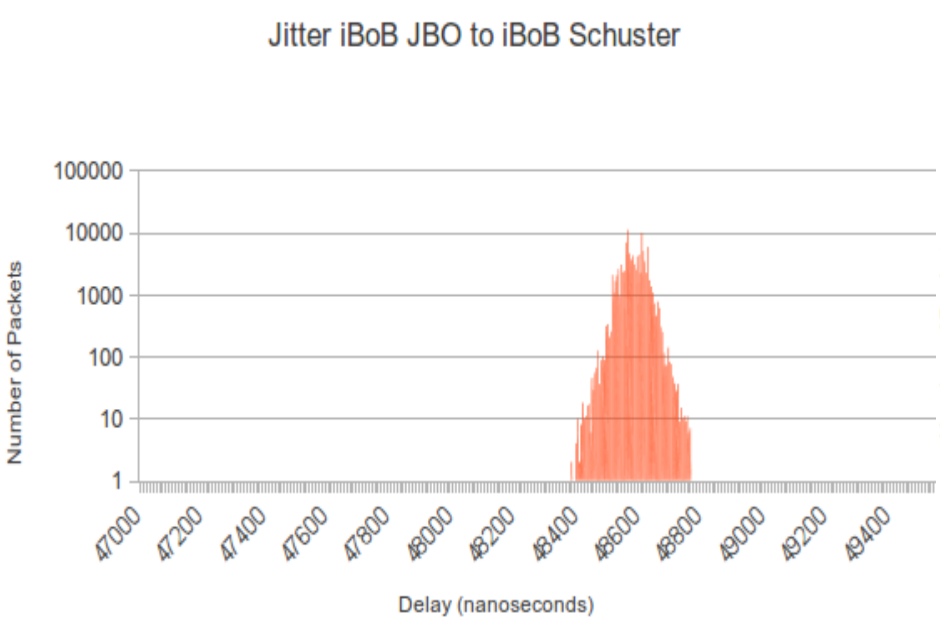
Range ~ 150 nanoseconds
FWHM ~ 15 nanoseconds



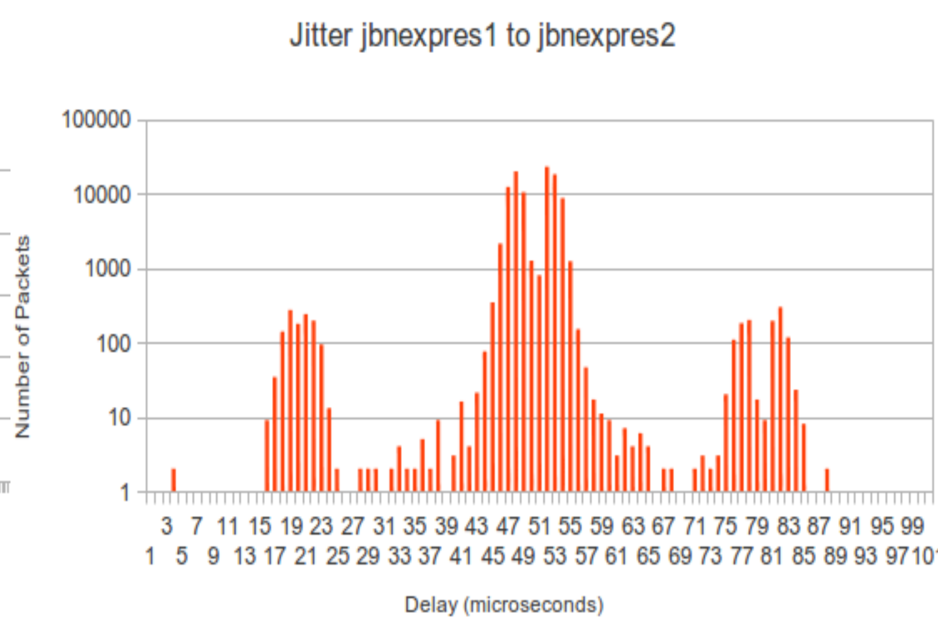
Range ~ 60 microseconds
FWHM ~ 8 microseconds

Jitter (2)

- 50 μs inter-packet delay requested



Range ~ 400 nanoseconds
FWHM ~ 80 nanoseconds



Range ~ 70 microseconds
FWHM ~ 8 microseconds

MANCHESTER 1824
The University of Manchester

Jitter (3)

Client	Server	Histogram bin width (µs)	Requested inter-packet delay (µs)	Achieved inter-packet delay				
				Min (µs)	Max (µs)	Mode (µs)	Mean (µs)	Standard deviation (µs)
PC - PC								
jbnexpres1	jbnexpres2	1	0	1	83	11	6.060	4.104
jbnexpres2	jbnexpres1	1	0	1	50	9	6.061	3.615
jbnexpres1	jbnexpres2	1	50	3	89	52	49.438	2.573
jbnexpres2	jbnexpres1	1	50	2	83	47	49.453	2.565
IBOB - IBOB								
JBO iBoB	Schu iBoB	0.005	0	6.620	6.805	6.625	6.633	0.076
Schu iBoB	JBO iBoB	0.005	0	6.620	6.815	6.625	6.633	0.077
JBO iBoB	Schu iBoB	0.005	50	48.370	48.775	48.565	48.585	0.077
Schu iBoB	JBO iBoB	0.005	50	48.365	48.805	48.565	48.585	0.080

Local Testing Conclusions

- Both IBOBs and PCs can transmit and receive UDP data at 10 Gbps without packet loss and packet reordering
  - PCs can send data marginally faster than IBOBs due to faster CPUs
  - IBOB data rates are more reproducible than PCs

- IBOBs have a very small jitter in inter-packet reception time
  - Deterministic behaviour
- PCs have larger jitter
  - General purpose machine that has other processes to control besides network communications
  - Data caching in socket buffers and on the NIC

- Standard deviation of PC jitter at 10 Gbps < length of time for packet delivery

- In conclusion, IBOBs are shown to be more accurate and reproducible in network testing, however they are expensive and hard to programme when compared to modern PCs. PCs, whilst less accurate, are of sufficient precision for our needs.

International BoD Links

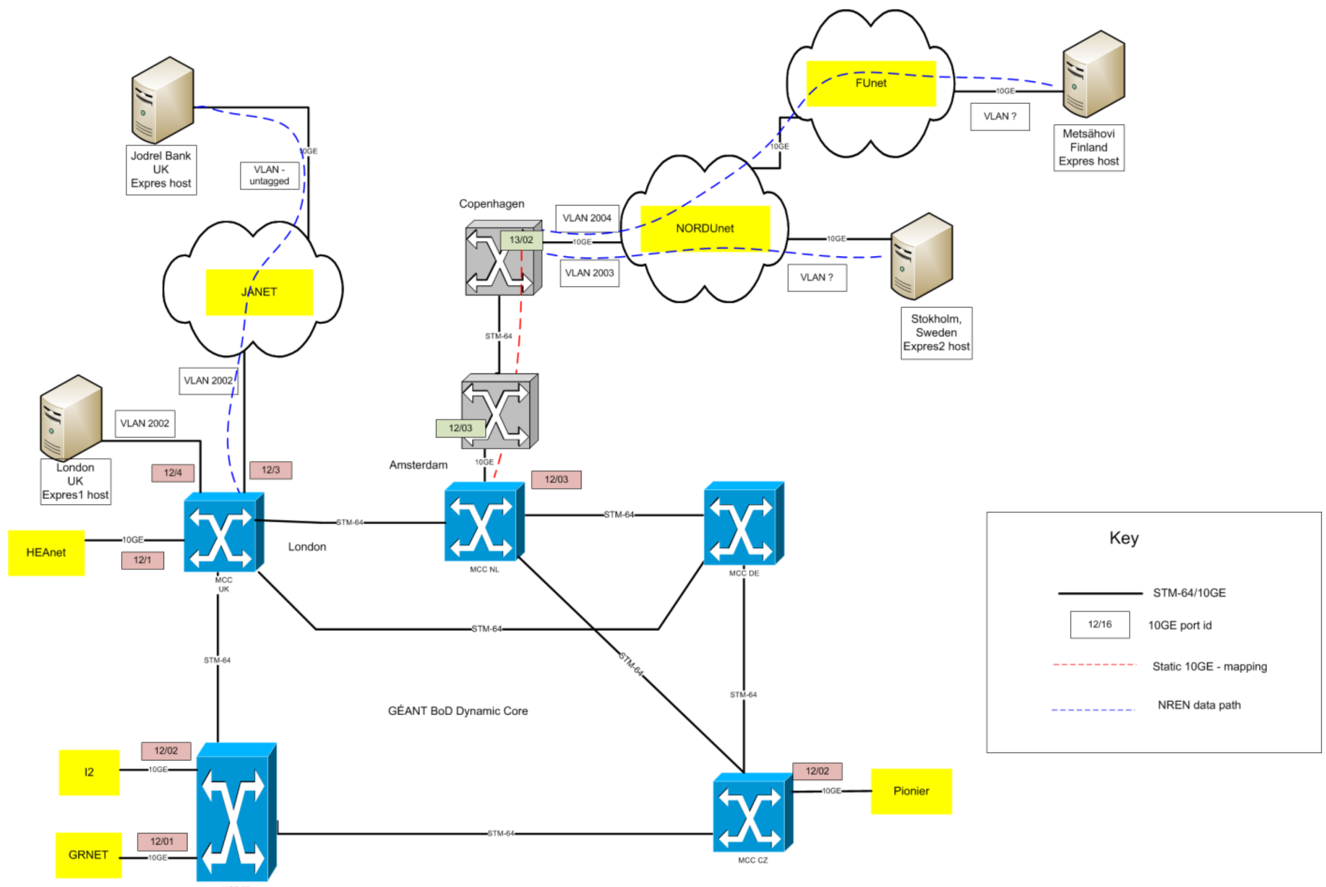
- BoD service still experimental, not production
- GÉANT pilot BoD system uses Inter-Domain Controller (IDC) protocol
- Networks used in test a combination of BoD and static paths to connect endhosts
- Data rates of 4 Gbps used

International BoD Links

- Using the GÉANT core dynamic network, a BoD circuit was set up with endhosts in:
  - JBO, UK
  - London, UK
  - Stockholm, Sweden
  - Metsähovi, Finland
- Static light paths were provided by:
  - JANET – JBO/Manchester to London
  - GÉANT – Amsterdam to Copenhagen
  - NORDUnet – Copenhagen to Stockholm
  - NORDUnet & Funet – Copenhagen to Metsähovi

International BoD Links

NEXPReS BoD connectivity

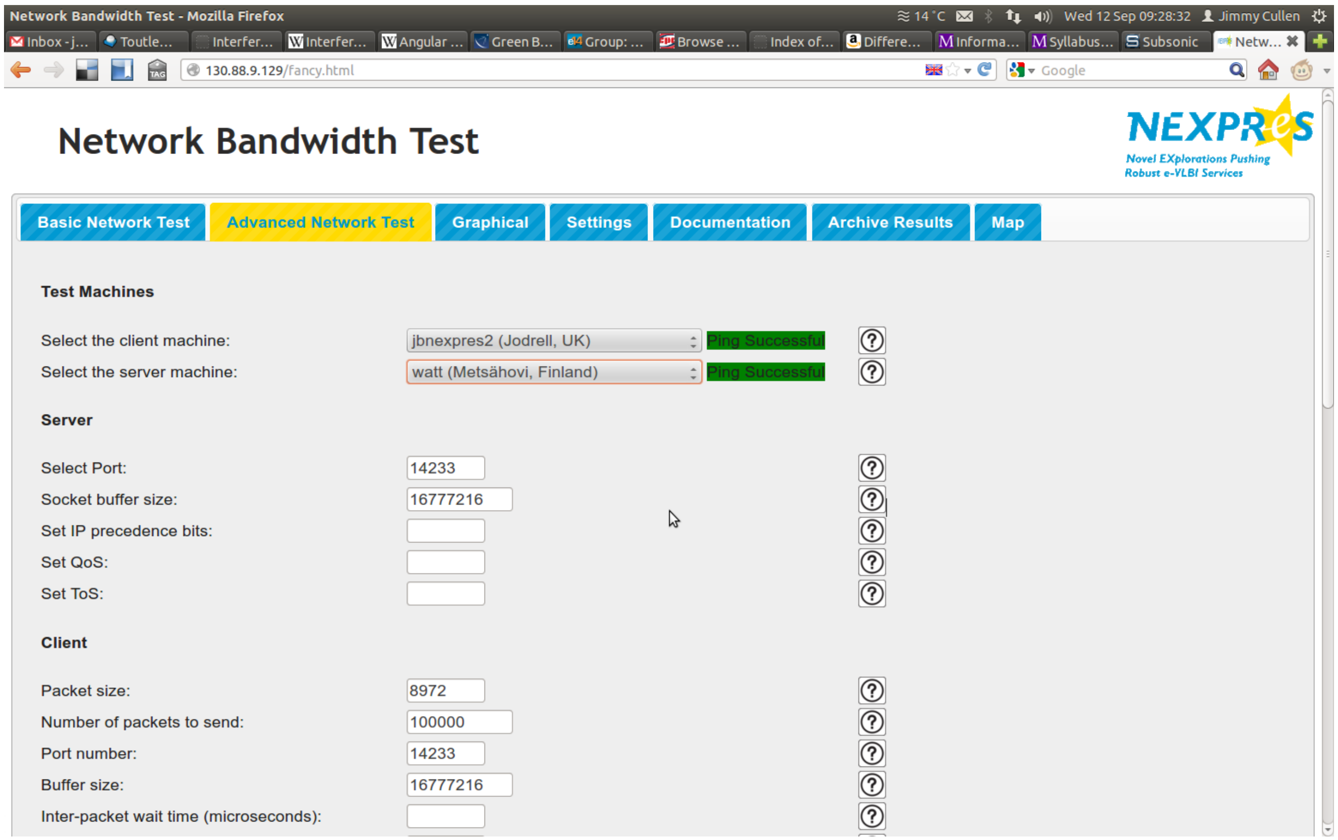


NORDUnet conference - 18th September 2012 - Oslo, Norway

MANCHESTER 1824
The University of Manchester

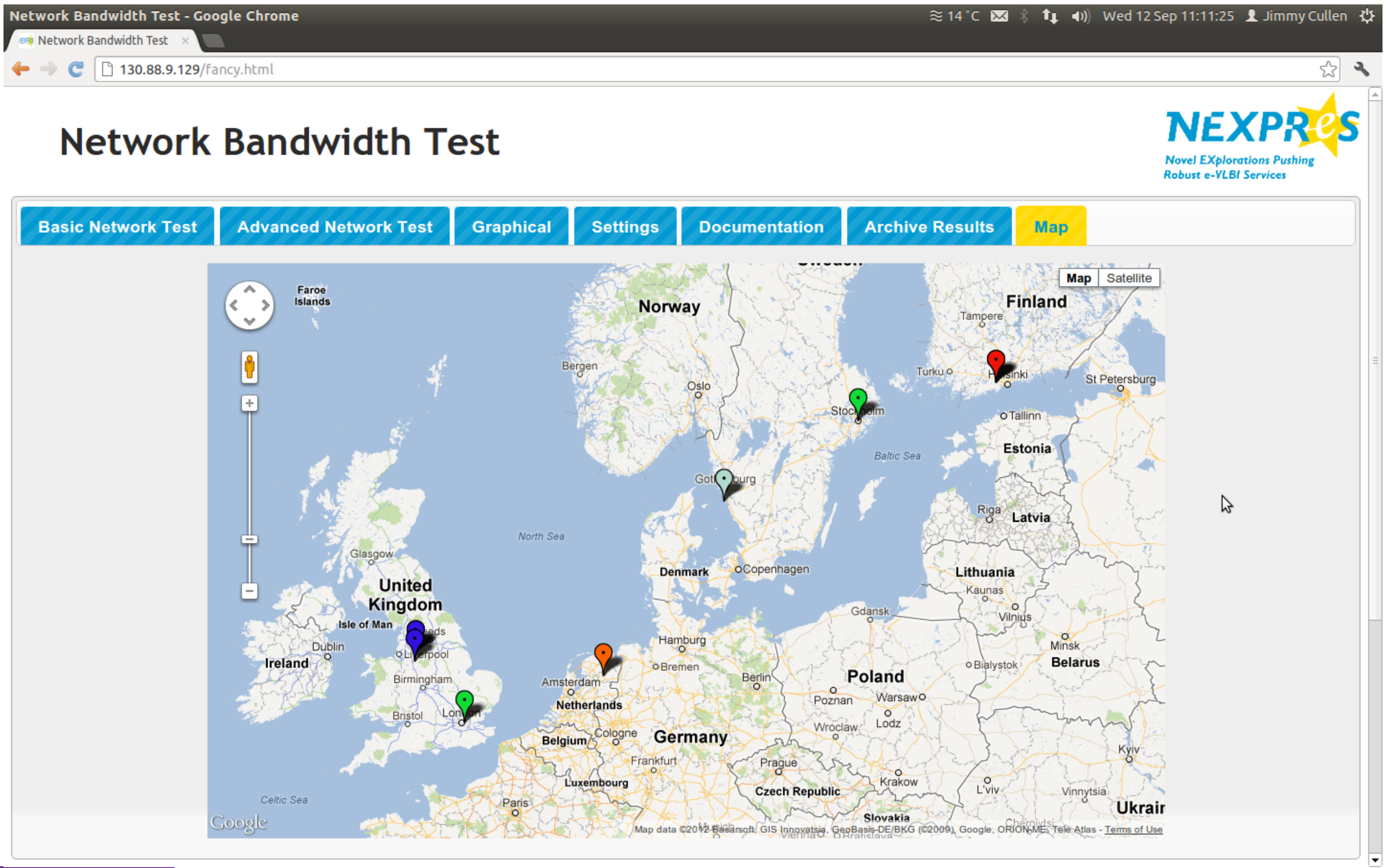
Network Test Website

- Graphical interface to run UDPmon tests between preconfigured endhosts



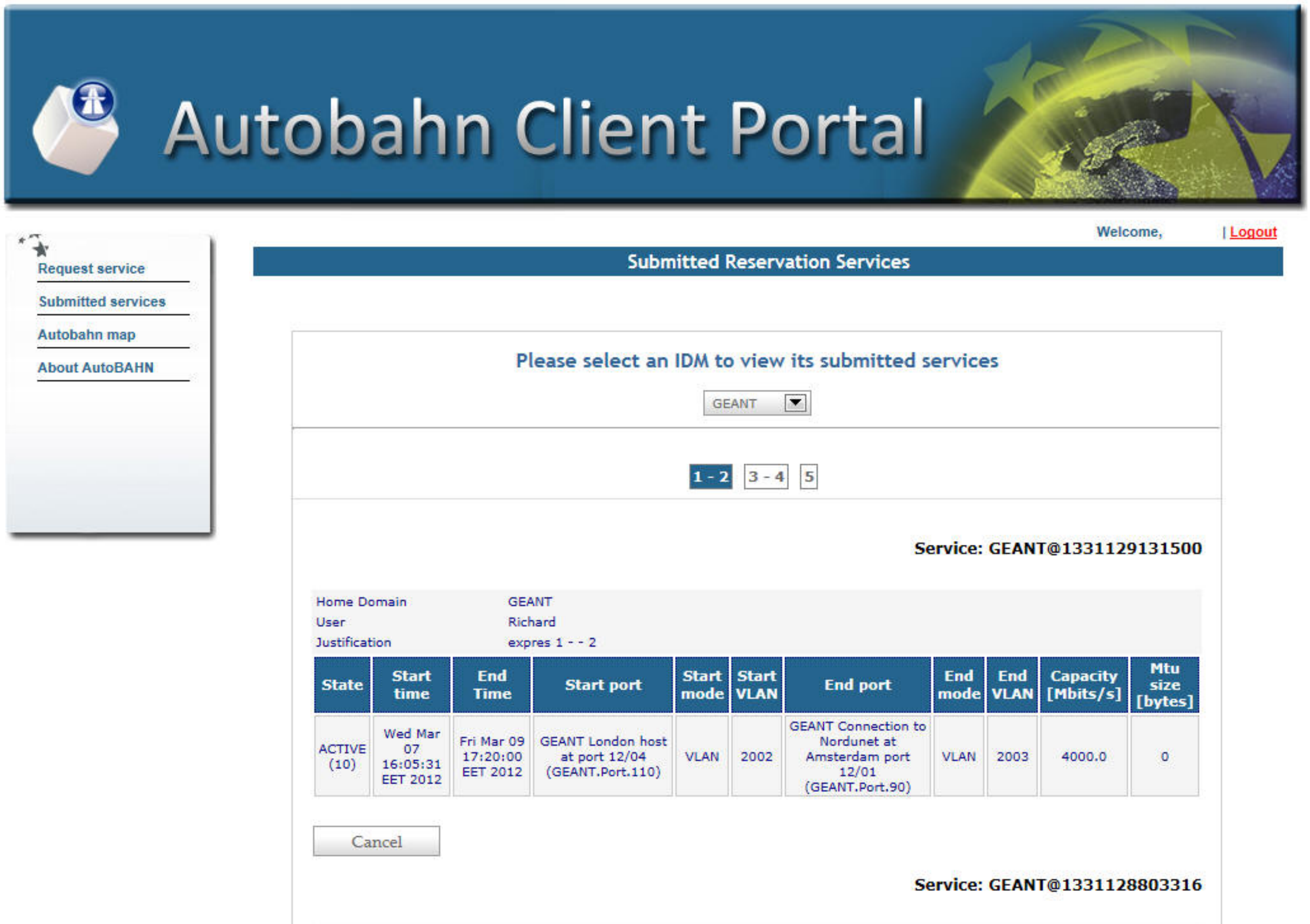
MANCHESTER 1824
The University of Manchester

Network Test Website



MANCHESTER 1824
The University of Manchester

AutoBAHN Reservation Interface



Autobahn Client Portal

Welcome, | Logout

Request service
Submitted services
Autobahn map
About AutoBAHN

Submitted Reservation Services

Please select an IDM to view its submitted services

GEANT

1 - 2 | 3 - 4 | 5

Service: GEANT@1331129131500

Home Domain GEANT
User Richard
Justification expres 1 - - 2

State	Start time	End Time	Start port	Start mode	Start VLAN	End port	End mode	End VLAN	Capacity [Mbits/s]	Mtu size [bytes]
ACTIVE (10)	Wed Mar 07 16:05:31 EET 2012	Fri Mar 09 17:20:00 EET 2012	GEANT London host at port 12/04 (GEANT.Port.110)	VLAN	2002	GEANT Connection to Nordunet at Amsterdam port 12/01 (GEANT.Port.90)	VLAN	2003	4000.0	0

Cancel

Service: GEANT@1331128803316

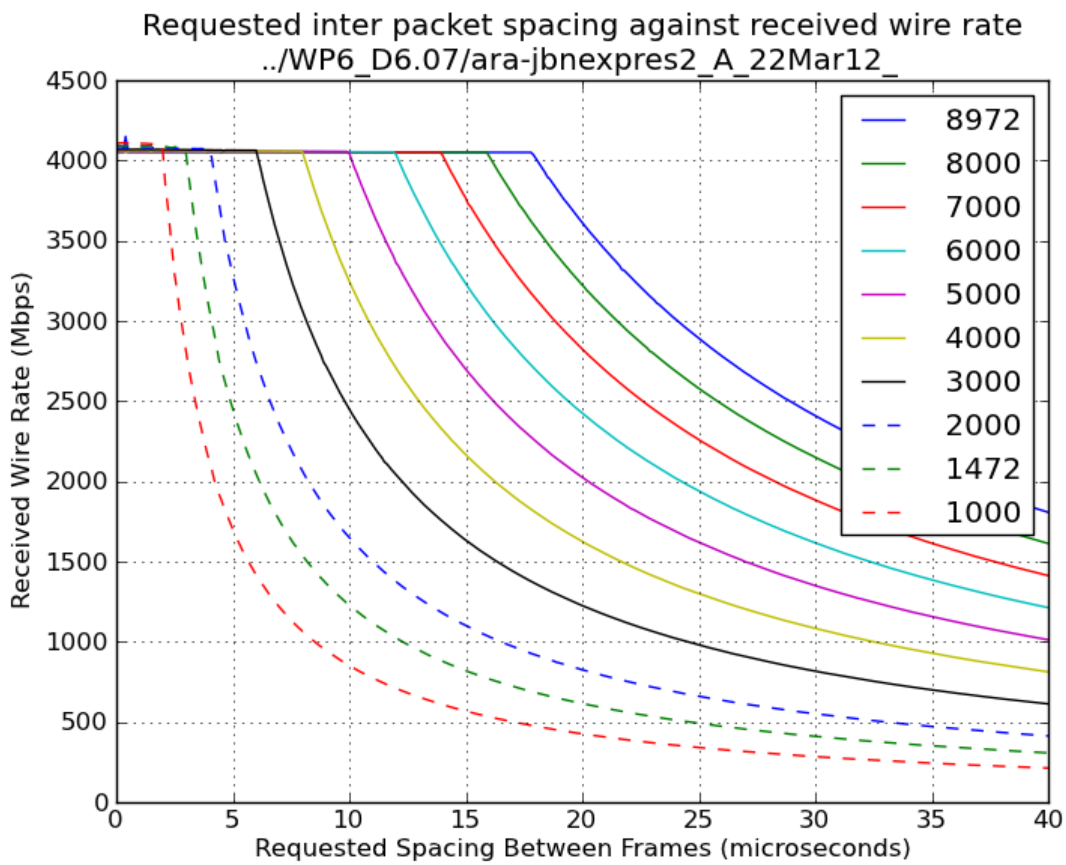
Start time
End time
Source
Destination
Bandwidth
Packet size

MANCHESTER 1824
The University of Manchester

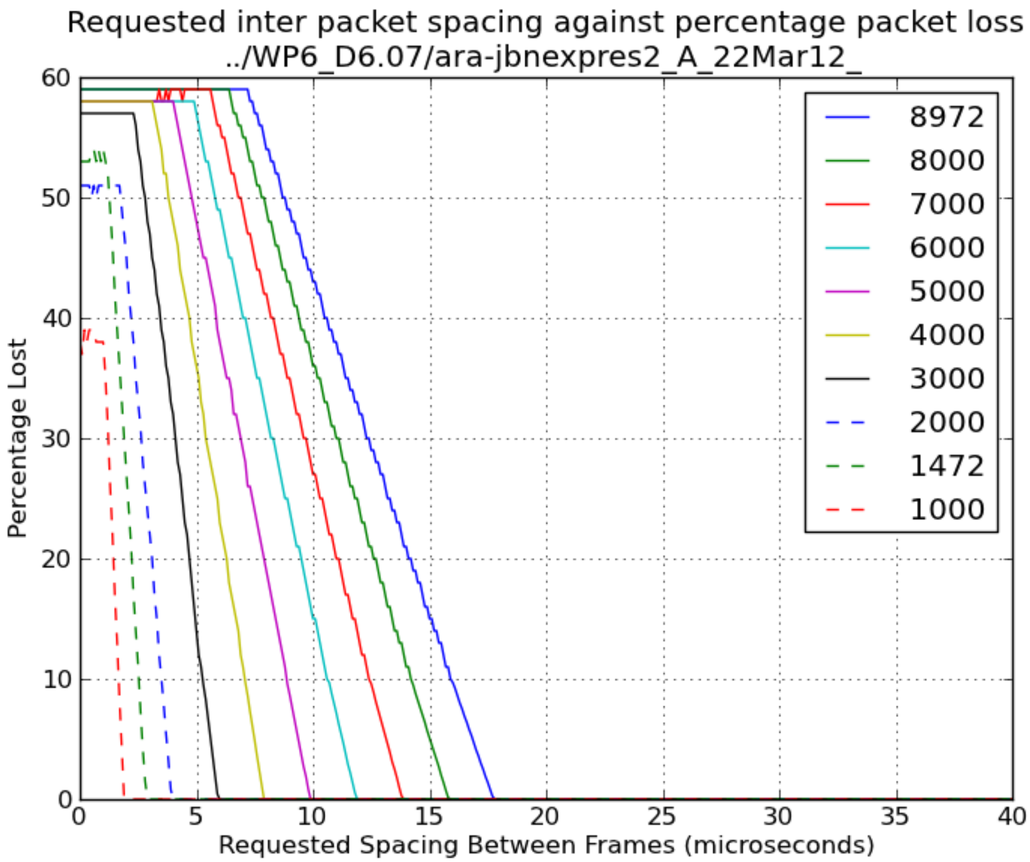
Tests Performed

- Network test website used to verify connectivity of the BoD links, and measurement of available bandwidth and any packet loss
- Command line used to run longer running, more detailed tests

Received Wire Rate



Packet Loss



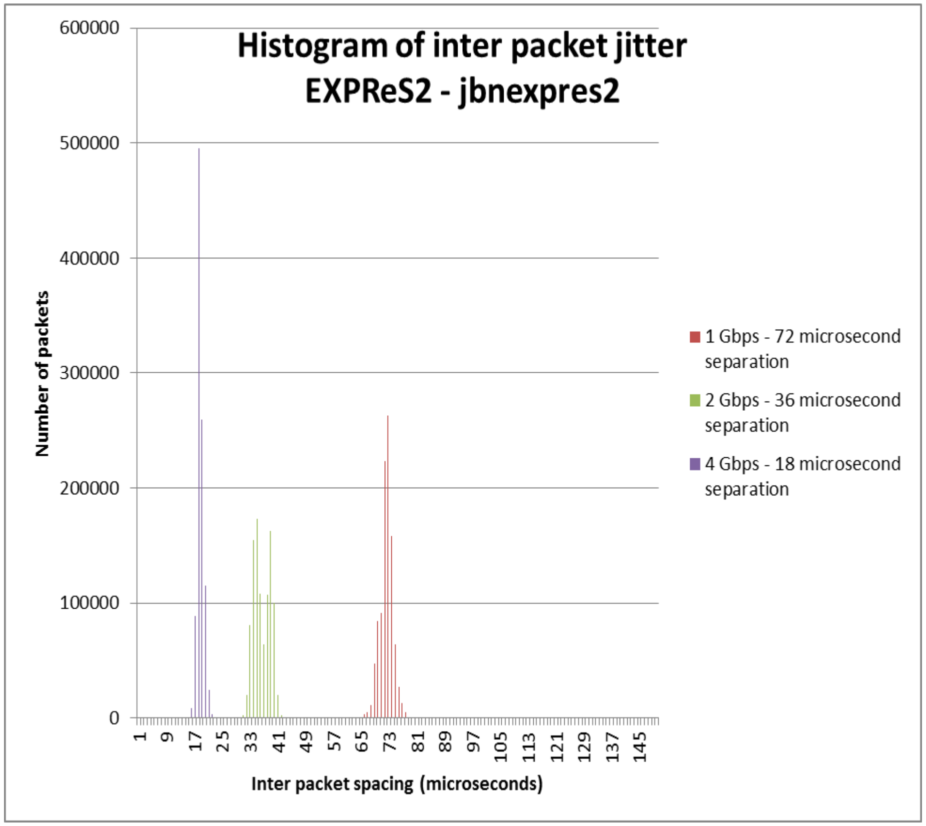
Requested inter packet spacing against percentage packet loss
../WP6_D6.07/ara-jbnexpres2_A_22Mar12_
Percentage Lost
Requested Spacing Between Frames (microseconds)
0
10
20
30
40
50
60
0
5
10
15
20
25
30
35
40
8972
8000
7000
6000
5000
4000
3000
2000
1472
1000

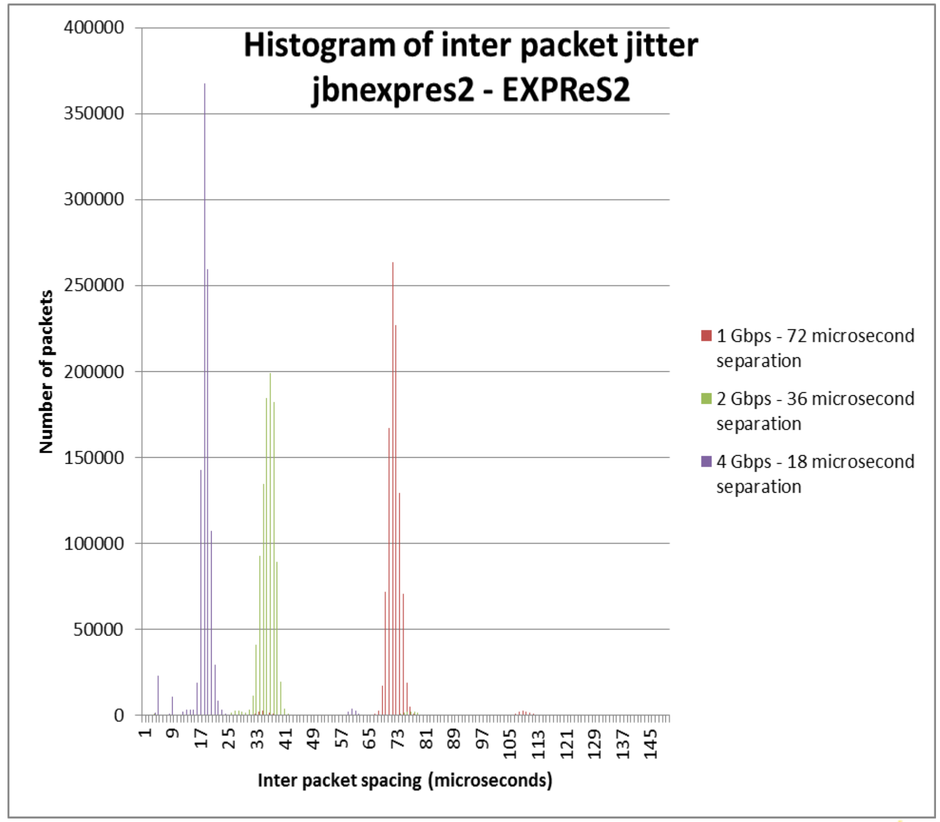
Jitter

- One million packets sent between hosts and jitter in inter-packet reception times recorded
- Three inter-packet delays used to generate different data rates
  - 72 microseconds – 1 Gbps
  - 36 microseconds – 2 Gbps
  - 18 microseconds – 4 Gbps

Hosts	Requested inter packet delay (µs)	Mean inter packet delay (µs)	Mode inter packet delay (µs)	Standard deviation of inter packet delay (µs)
Stockholm – JBO	72	71.4832	72	2.14274
	36	35.4732	34	2.60401
	18	17.4783	17	1.29231
JBO – Stockholm	72	71.4342	71	5.91037
	36	35.4516	36	5.93888
	18	17.4664	17	5.35474

Jitter





MANCHESTER 1824
The University of Manchester

24 Hour Tests

- UDPmon tests at 4 Gbps were run in each direction between JBO and Stockholm

Client	Server	Number of packets Received	Number of packets lost	Number of packets Reordered	Bytes received	Mean received wire rate (Mbps)
Stockholm	JBO	4794303918	0	18288	4.30145E+13	4012.810546
JBO	Stockholm	4794875087	2102	18292	4.30196E+13	4013.286133

- Stockholm → JBO bit error rate of less than 1 in 3.45×10^{14}

MANCHESTER 1824
The University of Manchester

Conclusions

- BoD matches very closely the requirements of eVLBI observations
- PCs are suitable for network parameter verification up to 10 Gbps
- BoD links are:
  - Stable
  - Reliable
  - Suitable for eVLBI operations
- However, important to verify links before experiments

Acknowledgements

- NetNorthWest – Anthony Ryan
- Metsähovi – Ari Mujunen, Tomi Salminen
- JANET – Dave Tinkler, David Salmon
- NORDUnet – Fredrik Pettai
- Funet - Jani Myyry
- DANTE/GÉANT

Questions/Comments

- Contact information

Jimmy Cullen
Research Associate
The University of Manchester
jcullen@jb.man.ac.uk

- Additional Information at http://nexpres.eu

- NEXPReS is an Integrated Infrastructure Initiative (I3), funded under the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no RI-261525.

eVLBI

- Data has a constant bit rate
- Transport protocols
  - TCP
    - \+ Reliably transfer all data
    - \- Large TCP buffer (window) required for long links
    - \- Lost packets cause reduction window size and hence data rate
    - \- Retransmission of lost packets and lower data rates cause large delay in data delivery
  - UDP
    - \+ No buffering
    - \+ Timely delivery of data
    - \- Unreliable transfer
- UDP most suitable format for VLBI data since timely arrival is more important than small amounts of data loss
  - $SNR \propto \sqrt{(1 - fractional\ loss)}$
- VLBI observations infrequent however, therefore inefficient use of the light paths (~ once a month)

MANCHESTER 1824
The University of Manchester

GÉANT AutoBAHN

- Uses VLANs across the network
- Data streams are tagged with the appropriate VLAN tag
- VLAN tags dynamically altered based upon BoD path requests.
  - e.g. BoD Metsähovi to JBO:
  - Data are tagged with VLAN 2004 over Funet and NORDUnet, then altered to VLAN 2002 at GÉANT, which identifies the VLAN to JBO

MANCHESTER 1824
The University of Manchester