ASCC Site Report

Simon C. Lin

Computing Centre, Academia Sinica

Service Challenge Workshop

Apr. 26, 2005

Outline

- Networking Status
- ASCC Service Challenge Status
- Tentative Resource Plan for Tier1
- Other Related Activities
- Site Visit

Taipei Status

- ASNet* runs one STM16 IPLC from Taipei to Amsterdam
- One GE Local loop from Amsterdam to Geneva (via NetherLight LightPath Service), second GE is waiting to turn on (in process)
- will double the IPLC on July 1st (contract will be signed on May 27th)
- ASCC LCG facilities now have multiple GE uplinks to ASNet's core router; will be replaced by one or two 10GE during this summer.
- ASNet connects (10GE*2 + STM64*n) to domestic backbone, a.k.a. TANet/TWAREN joint backbone, to reach T2's in Taiwan.
- Every T2 in Taiwan has it's own 10GE link to the domestic backbone

* ASNet (Academic Services Network, as#9264) is the network division and is also the network name that registered in APNIC.

Asia Pacific Link Status

- How many T2s in Asia Pacific region?
- Taipei – Tokyo
  - STM-4 IPLC
  - 2*GE to APAN-JP (for JP Universities)
  - Will have GE to SiNet/SuperSiNet (for Tokyo Univ. & KEK)
  - IPLC should be increased before LCG production run.
- Taipei – Singapore
  - STM-1 IPLC to SingAREN (National Grid Office of Singapore)

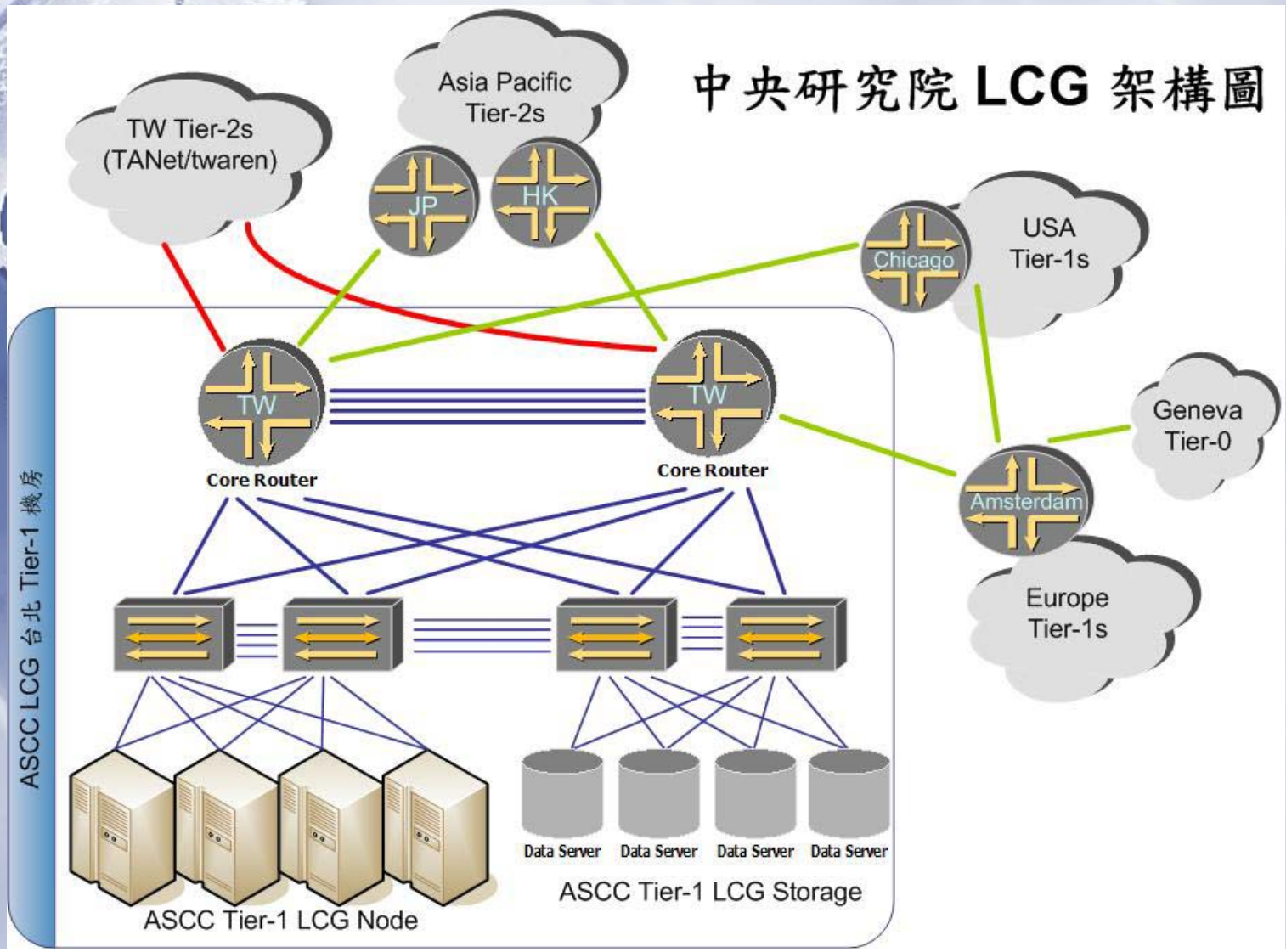
Asia Pacific Link Status

- Taipei – Hong Kong
  - STM-4 IPLC, will upgrade to STM-16 on Feb 1st, 2006 (contract already been signed on March 30th)
  - GE link to CERnet, GE link to CSTnet
    - CERnet : China Education and Research network
    - CSTnet : China Science and Technology network
- Taipei – Korea
  - Maybe have GE link with KOREN (still work with KOREN engineer)

Technical Issues

- LightPath
  - How urgent is it for the lightpath between T1 and T2 ?
  - Today, there is no lambda can be brought in AP region (there is only SDH circuits) because of the design of cable systems
  - Can MPLS L2 VPN and/or premium IP services provide same (or almost the same) quality compares to LightPath in the LCG environment ?
  - Even though, ASNet is still surveying the WDM equipments

Networking Topology for LCG (New)

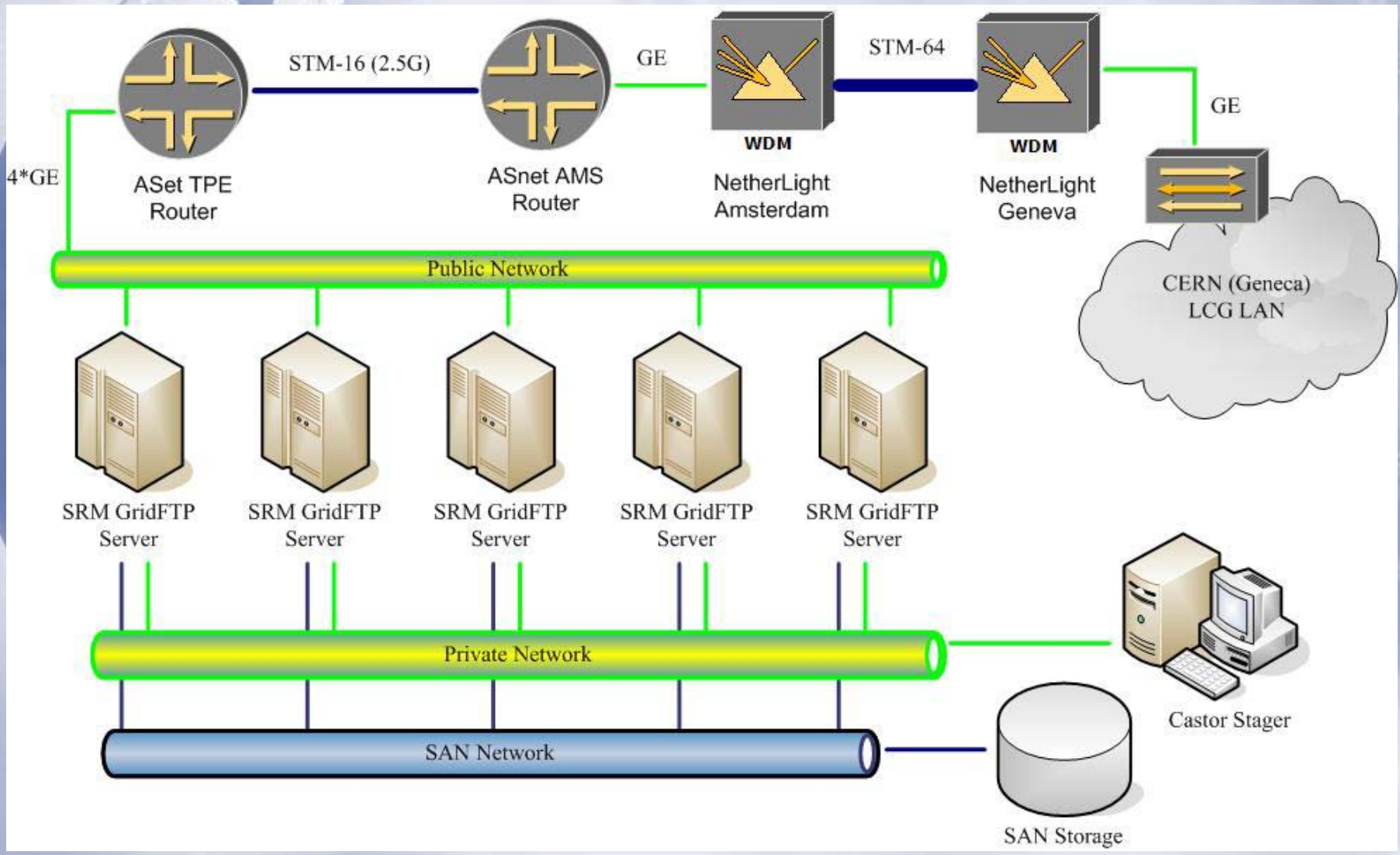



Service Challenge

ASCC-CERN Service Challenge

- Objectives
  - Ensure the baseline services, framework for problem resolution, and operational/emergency procedures
  - Reliable data transfer, job flooding/exerciser, incident response, interoperability
- Should reach ~ 200MB under current internetworking status
- Still working with CERN to get better performance

SC test bed @ ASCC



Testbed

- 4 nodes
  - 2 GByte Memory
  - ~200GB HD
  - Scientific Linux 3.0.4
  - ClassicSE configuration / GridFTP
  - Each has 1Gb connectivity (Intel Pro1000/MT NIC)

LAN Performance Test

without system tuning

```
[root@lcg00115 sc]# iperf -c lcg00116.grid.sinica.edu.tw -i 5 -w 64M -t 36000 | tee iperf-0426.txt
------------------------------------------------------------
Client connecting to lcg00116.grid.sinica.edu.tw, TCP port 5001
TCP window size:   128 MByte (WARNING: requested 64.0 MByte)
------------------------------------------------------------
[  3] local 140.109.248.4 port 32819 connected with 140.109.248.5 port 5001
[  3]  0.0- 5.0 sec    562 MBytes    843 Mbits/sec
[  3]  5.0-10.0 sec    486 MBytes    816 Mbits/sec
[  3] 10.0-15.0 sec    486 MBytes    815 Mbits/sec
[  3] 15.0-20.0 sec    474 MBytes    794 Mbits/sec
[  3] 20.0-25.0 sec    467 MBytes    784 Mbits/sec
[  3] 25.0-30.0 sec    478 MBytes    802 Mbits/sec
[  3] 30.0-35.0 sec    499 MBytes    837 Mbits/sec
[  3] 35.0-40.0 sec    493 MBytes    827 Mbits/sec
[  3] 40.0-45.0 sec    465 MBytes    780 Mbits/sec
[  3] 45.0-50.0 sec    467 MBytes    784 Mbits/sec
[  3] 50.0-55.0 sec    486 MBytes    815 Mbits/sec
[  3] 55.0-60.0 sec    485 MBytes    813 Mbits/sec
[  3] 60.0-65.0 sec    487 MBytes    817 Mbits/sec
```

Up to 800 Mbits/sec with single stream test

WAN Test
Taipei-to-CERN
without system tuning

```
[root@lcg00115 sc]# iperf -c 128.142.224.200 -i 5 -w 4M -t 36000
------------------------------------------------------------
Client connecting to 128.142.224.200, TCP port 5001
TCP window size: 8.00 MByte (WARNING: requested 4.00 MByte)
------------------------------------------------------------
[  3] local 140.109.248.4 port 32774 connected with 128.142.224.200 port 5001
[  3]  0.0- 5.0 sec  18.3 MBytes  30.7 Mbits/sec
[  3]  5.0-10.0 sec  28.3 MBytes  47.4 Mbits/sec
[  3] 10.0-15.0 sec  20.9 MBytes  35.1 Mbits/sec
[  3] 15.0-20.0 sec  2.00 MBytes  3.36 Mbits/sec
[  3] 20.0-25.0 sec  14.7 MBytes  24.7 Mbits/sec
[  3] 25.0-30.0 sec  14.6 MBytes  24.5 Mbits/sec
[  3] 30.0-35.0 sec  14.0 MBytes  23.4 Mbits/sec
[  3] 35.0-40.0 sec  16.0 MBytes  26.9 Mbits/sec
[  3] 40.0-45.0 sec  14.1 MBytes  23.7 Mbits/sec
[  3] 45.0-50.0 sec  16.3 MBytes  27.3 Mbits/sec
[  3] 50.0-55.0 sec  14.2 MBytes  23.8 Mbits/sec
[  3] 55.0-60.0 sec  16.4 MBytes  27.6 Mbits/sec
[  3] 60.0-65.0 sec  14.4 MBytes  24.2 Mbits/sec
[  3] 65.0-70.0 sec  14.5 MBytes  24.3 Mbits/sec
```

Up to 40Mbits/s in single stream test
with small window size & system buffer & NIC buffer


Performance Tuning
on
High Bandwidth/Latency Production
Network

Enabling High Performance Data Transfers

- High Latency Network
  - RTT is above 300ms
- Tuning Parameter
  - TCP Window Size
  - System Buffer Size
  - NIC Queue length (txqueuelen)

WAN Test
Taipei-to-CERN

```
------------------------------------------------------------
Client connecting to 128.142.224.200, TCP port 5001
TCP window size:   128 MByte (default)
------------------------------------------------------------
[  3] local 140.109.248.4 port 33812 connected with 128.142.224.200 port 5001
[  3]  0.0- 5.0 sec  95.4 MBytes    160 Mbits/sec
[  3]  5.0-10.0 sec    485 MBytes    814 Mbits/sec
[  3] 10.0-15.0 sec    528 MBytes    886 Mbits/sec
[  3] 15.0-20.0 sec    450 MBytes    755 Mbits/sec
[  3] 20.0-25.0 sec    500 MBytes    839 Mbits/sec
[  3] 25.0-30.0 sec    502 MBytes    842 Mbits/sec
[  3] 30.0-35.0 sec    424 MBytes    711 Mbits/sec
[  3] 35.0-40.0 sec    500 MBytes    840 Mbits/sec
[  3] 40.0-45.0 sec    502 MBytes    842 Mbits/sec
[  3] 45.0-50.0 sec    427 MBytes    716 Mbits/sec
[  3] 50.0-55.0 sec    503 MBytes    843 Mbits/sec
[  3] 55.0-60.0 sec    500 MBytes    840 Mbits/sec
```

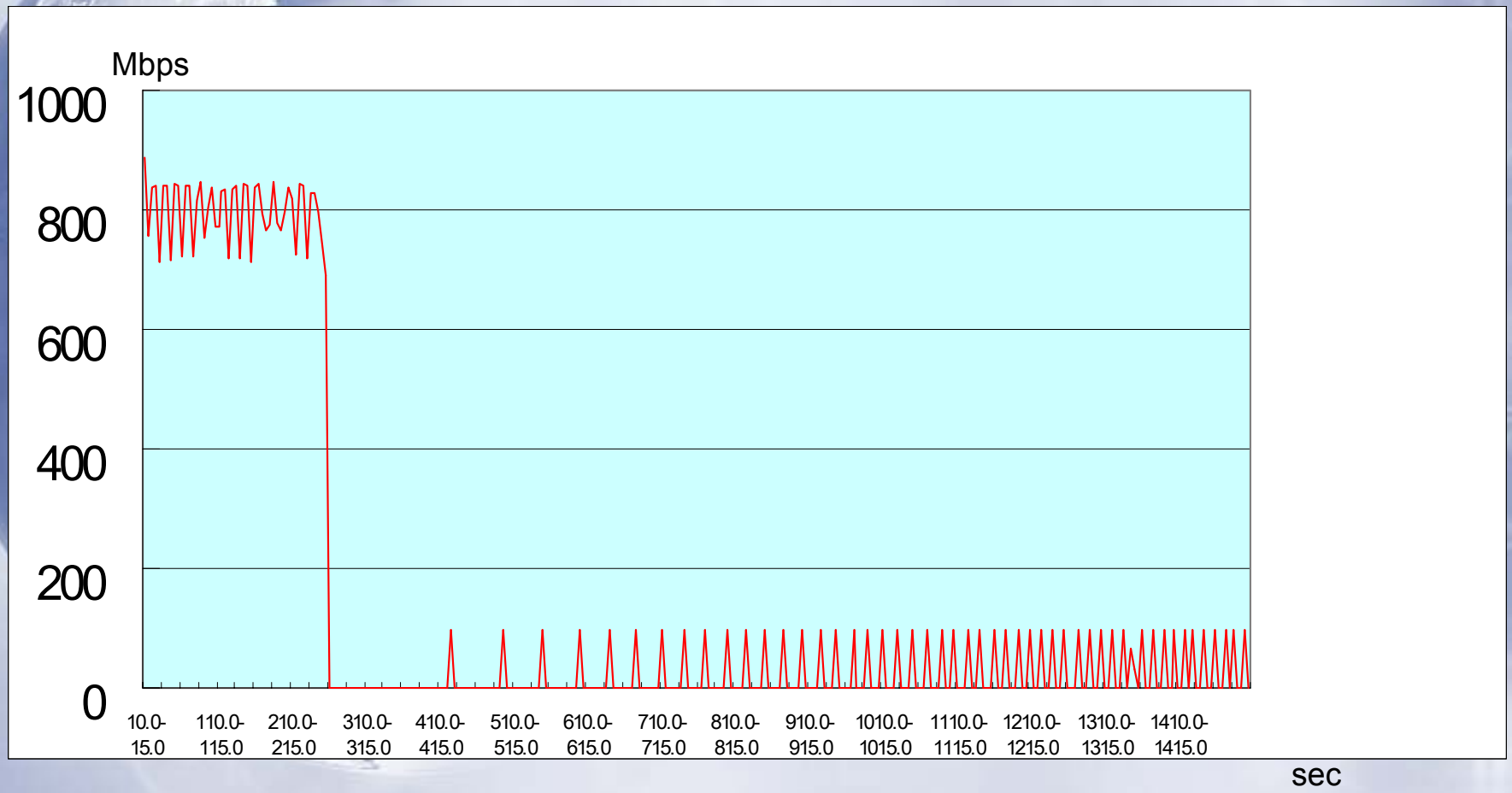
TCP test reaches to 800 Mbits/s, ~77% of max. PCI speed.

128M window size

256M buffersize

txqueuelen 1000000

WAN Test
Taipei-to-CERN (Cont.)



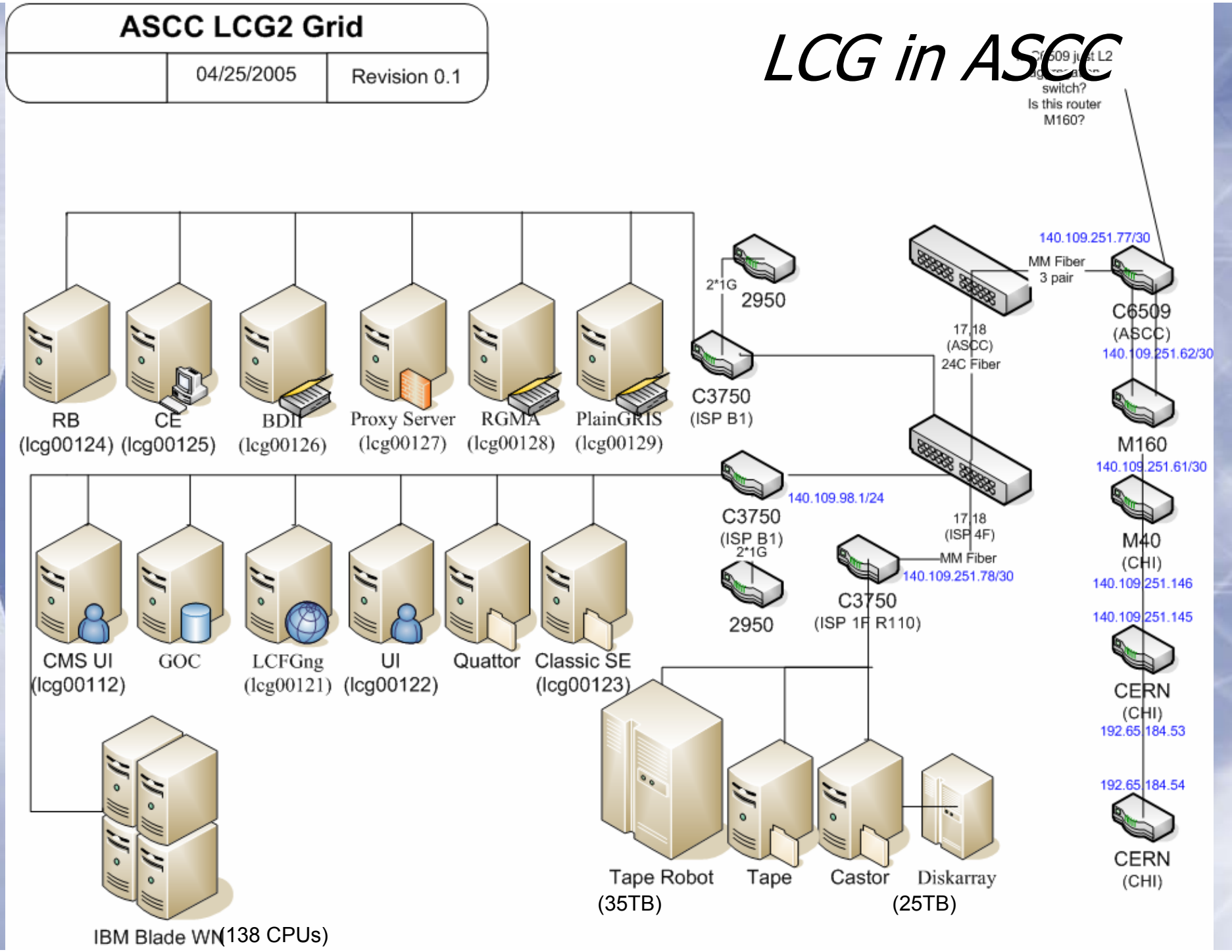
Not stable enough! Need further investigation.

To Do

- Work with CERN
  - for the performance data from CERN to ASCC
  - Tuning for stable state
  - Testing router by router
- Node Architecture and CPU effects
  - SMP nodes
  - Itanium CPUs
- Middleware Testing
  - GridFTP
  - SRM


Resource Plan for T1

LCG in ASCC


ASCC LCG2 Grid
04/25/2005
Revision 0.1
switch?
Is this router
M160?
140.109.251.77/30
MM Fiber
3 pair
C6509
(ASCC)
140.109.251.62/30
2*1G
2950
17,18
(ASCC)
24C Fiber
C3750
(ISP B1)
M160
140.109.251.61/30
RB
(lcg00124)
CE
(lcg00125)
BDII
(lcg00126)
Proxy Server
(lcg00127)
RGMA
(lcg00128)
PlainGRIS
(lcg00129)
140.109.98.1/24
C3750
(ISP B1)
2*1G
2950
17,18
(ISP 4F)
MM Fiber
140.109.251.78/30
C3750
(ISP 1F R110)
M40
(CHI)
140.109.251.146
140.109.251.145
CERN
(CHI)
192.65.184.53
192.65.184.54
CERN
(CHI)
CMS UI
(lcg00112)
GOC
LCFGng
(lcg00121)
UI
(lcg00122)
Quattor
Classic SE
(lcg00123)
IBM Blade WN
(138 CPUs)
Tape Robot
(35TB)
Tape
Castor
Diskarray
(25TB)

Castor Deployment

- Installed at Apr. 2004
  - 20 TB SATA Disk Array
  - 35 TB IBM LTO Tape Library
- Why Castor
  - Support and sharable with TSM, IBM 3494, IBM 3584
  - SRM interface is ready, but is not good enough in both performance and stability
- Current Status
  - Version
  - Management of disk cache and tape library

Tentative Resource Plan

Taipei Tier1	2004	2005	2006	2007	2008	2009	2010
CPU (kSI2K)	166	500	950	1770	3400	3600	
Disk (Tbytes)	24	120	400	900	1500	2373	
Tape (Tbytes)	20	140	500	800	1300	2000	
Nominal WAN (Mbits/sec)	2400	2400	4800	10000	10000	10000	

Split 2008/9	ALICE	ATLAS	CMS	LHCb	SUM 2008/9
Offered		1870	1530		3400
% of Total		7%	10%		8%
Offered		825	675		1500
% of Total		5%	9%		6%
Offered		715	585		1300
% of Total		7%	5%		6%

Extending Grid Applications

- Atmospheric Science
  - NTU, NCU and NTNU
  - SRB-based Data Grid, will integrate analysis applications in the next phase
- Data Grid for Digital Archives
- Bioinformatics: mpiBLAST-g2
- Grid for Medical Imagery (planning)
- Island-wide Grid Testbed
- Geospatial Information Infrastructure
- Biodiversity Informatics

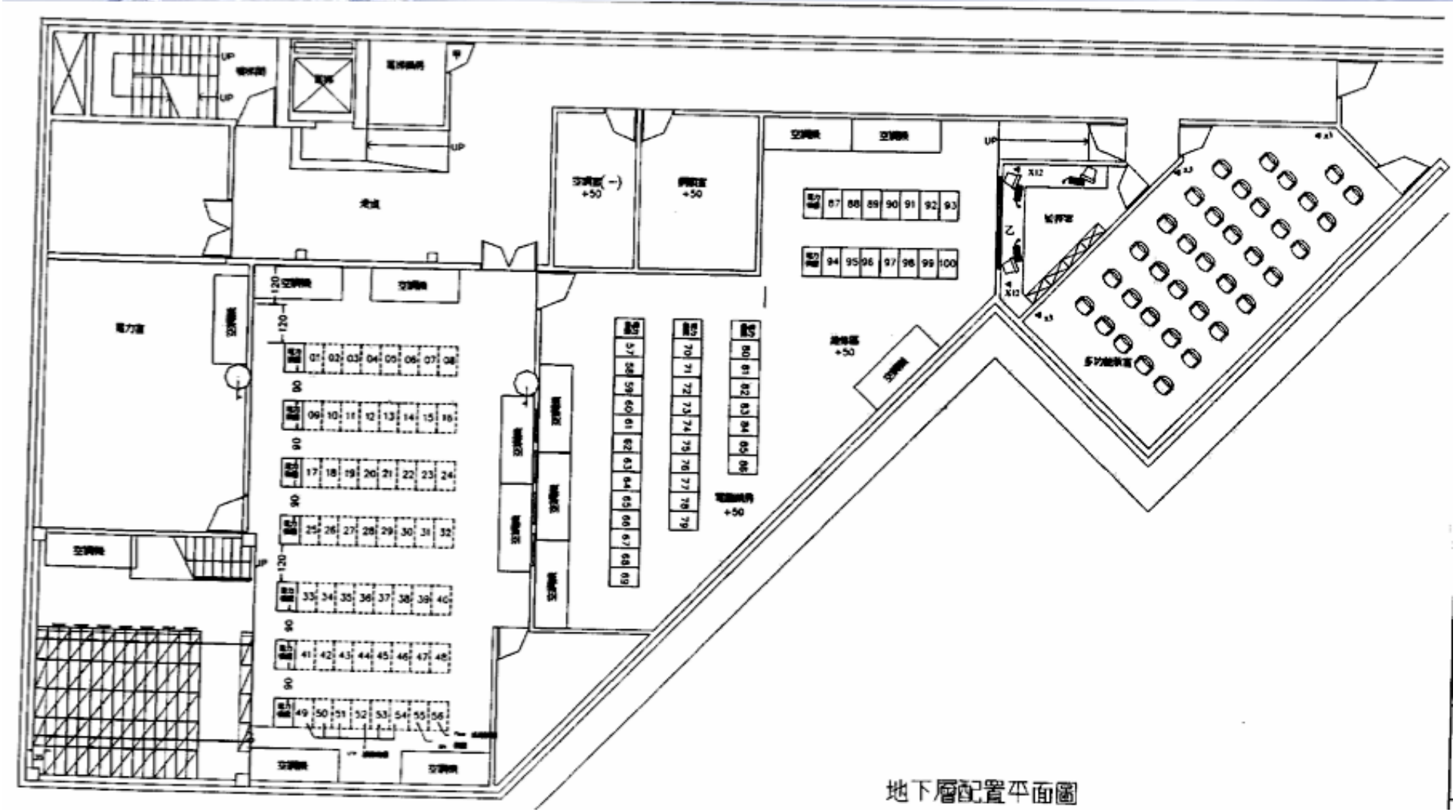
Education and Training

Event	Date	Attendant	Remarks
China Grid LCG Training	May 16-18,'04	40	Beijing, China
ISGC'04	Jul.26,'04	50	AS, Taipei
Grid Workshop	Aug.16-18,'04	50	Shang-Dong, CN
NTHU	Dec.22-23,'04	110	Shin-Chu
NCKU	Mar.9-10,'05	80	Tainan
ISGC'05	Apr.25,'05	~80	AS, Taipei
Tung-Hai U	June, 2005		Tai-chung
Hua-Lian/ Tai-Tung	Sep., 2005		Planning

IPAS Computer Room

- Space Plan
  - 60% CPU & Networking
  - 30% Disk
  - 10% Tape
- Resource Plan
- Movement Plan
  - Apr. 14 ~ Apr.20
- Layout

Layout of Computer Room in IPAS



LCG/EGEE Technology Involvement

ASCC in ARDA (1)

- gLite Testing
  - gLite WMS monitoring
    - Program has been started to detect WMS efficiency and stability through the heart-beat job operations since Feb. 2005
    - Realtime report is available at here on ARDA webpage
  - gLite IO monitoring
    - The same philosophy of WMS monitoring
    - The program has been started to collect data since Mar. 2005 (web report is under construction)
  - gLite certification test suite
    - Help on porting 3 testing items from LCG to gLite

ASCC in ARDA (2)

- ATLAS Distributed Analysis
  - Evaluating and testing the frameworks for ATLAS distributed analysis on LCG and gLite Grid environments
    - Still on the early phase of enabling the frameworks on LCG and gLite
- ARDA Group in Taiwan
  - Middleware (gLite) Testing & Integration
  - ALTAS
  - CMS

3D -- Distributed Deployment of Databases

- Objectives
  - Define distributed databases services and application
  - Help to avoid the costly parallel development of data distributions, backup and HA features
  - Minimize number of LCG database administration personnel
- ASCC Involvement
  - is one of the 3D T1 site.
  - Service Definition and Implementation Group
  - intend to support 3D software development and DB design
  - Incorporation of CA with 3D

CIC and ROC Services

- Objectives
  - CIC Responsibilities
    - Centralized service operations (RB, BDII, VOMS, RLS)
    - Monitoring and Operations Support
  - ROC Responsibilities
    - Monitoring and Operations Support:
    - Deployment support
    - Testbed Operations
    - User support
  - General
    - Management
    - Tools development
- VOMS:
  - System Establishment
  - AUP reviewing