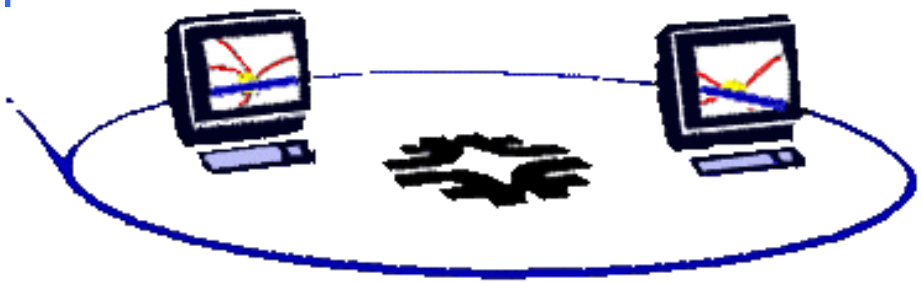
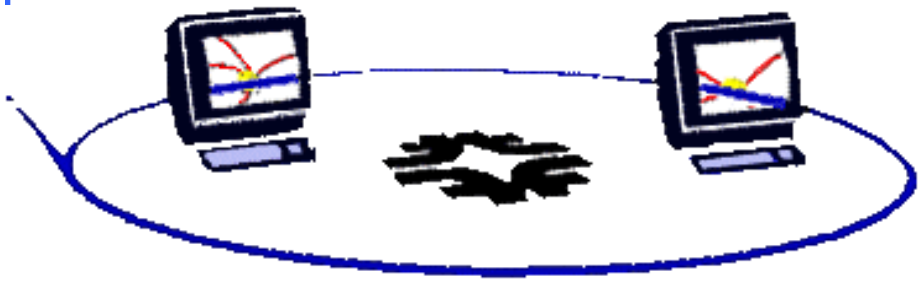


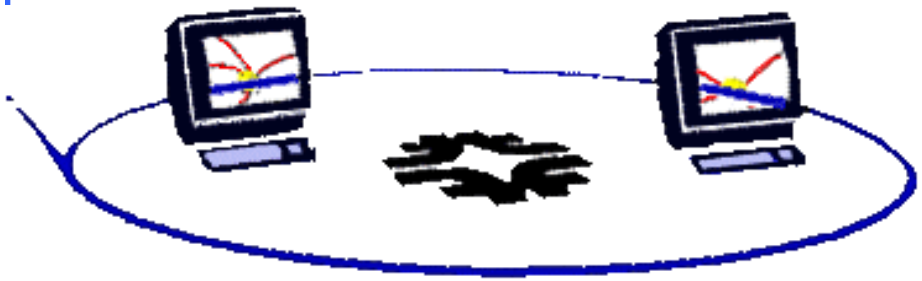
SITE REPORT: FERMI

LISA GIACCHETTI

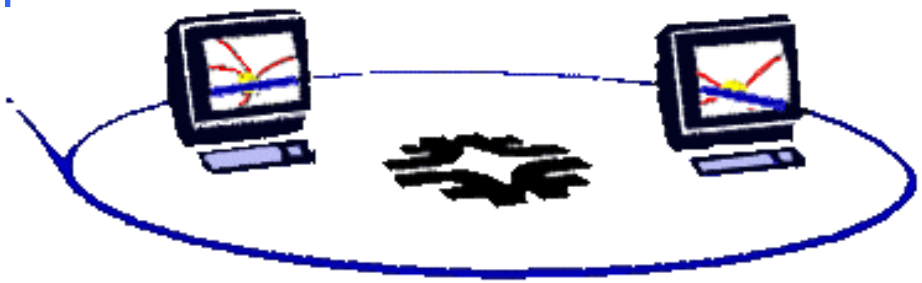


GENERAL Computing

- Finances Tight
  - Budgets flat or < flat
  - In midst of a 5% person staff reduction process
  - Announcement by DOE to cancel Btev a big blow
- 2 new computer rooms deployed in existing experimental halls
  - LCC: Houses LQCD
  - GCC: Houses CDF, D0 and USCMS farms
    - Planning 2 extensions to GCC : robot room and another computer room
    - Construction slated to begin Jun-Jul '05

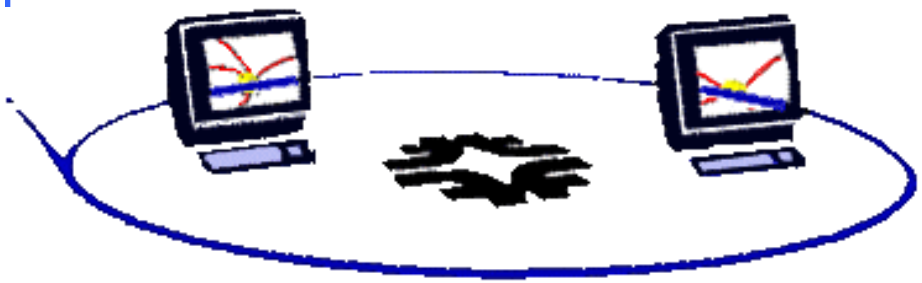


GCC Computer Room



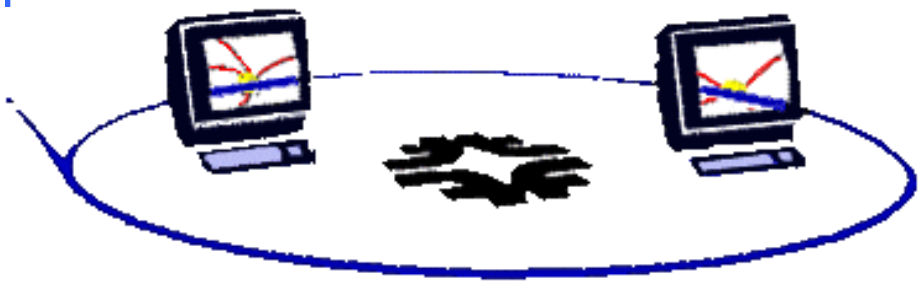
Lattice QCD

- Project now operates the following tightly coupled clusters
  - 128-node dual 2.4Ghz Xeon, coupled with Myrinet
  - 128-node single 2.8GHz Pentium 4 (Prescott), coupled with Myrinet
  - 32-node dual 2.0GHz Xeon, coupled with Infiniband
- Expansions in FY05
  - 260-node single 3.2GHx Pentium 640 coupled with Infiniband
    - Commissioning now, release to production June 1
  - 260-cpu expansion to this cluster, either using Pentium 640's or dual Opterons
    - Commissioning Sept, release to production ~ Oct 1
  - Total capacity should exceed 1.0 Teraflops/sec sustained



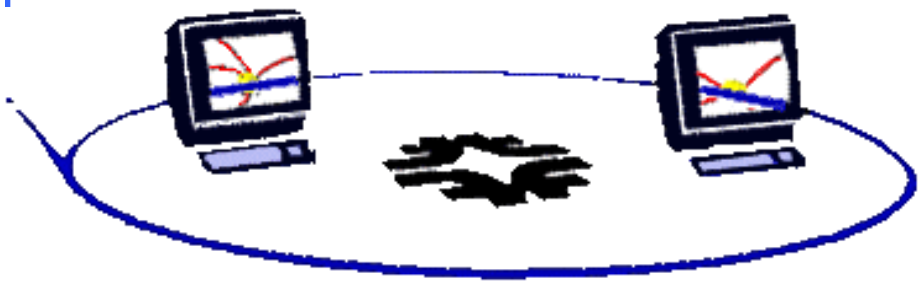
USCMS T1 Facilities

- FNAL_WC1 for USCMS production, LCG, Sam-grid and OSG computing
  - 2 gatekeepers
  - 174 workers (300 more on order); SL(F)3.0.3
  - Condor batch system
  - Publishes information to OSG and LCG
- User Analysis Facility (UAF) for interactive and batch computing
  - 1 fbs headnode (migrating to Condor soon)
  - 1 NFS server; ~10TB Infortrend disk
    - Migrating all of this to IBRIX
  - 53 workers; In process of migrating to SL(F)3.0.3
  - Interactive load balancing currently done with customizations to FBSNG
    - Are ready to implement load balancing via Cisco IOS



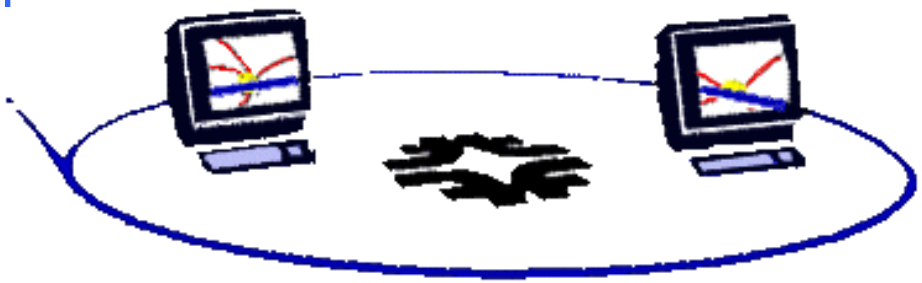
USCMS T1 Facilities

- ~ 10 server/service nodes (web, db, voms, rocks, console)
- Dcache
  - 58TB currently deployed; 38TB more on order; 100TB by October
  - 16 Dual Intel systems serving Infortrend disk via SAN and older 3ware Intel systems
    - New storage will be Nexsan
  - 4 admin nodes
  - Implemented resilient Dcache config using worker node local data drives in February
- All installations being done with ROCKS
  - 2 front ends using mirror of SL(F)3 as our distribution
  - YUM updates from Primary SL(F)3 servers and our own distribution



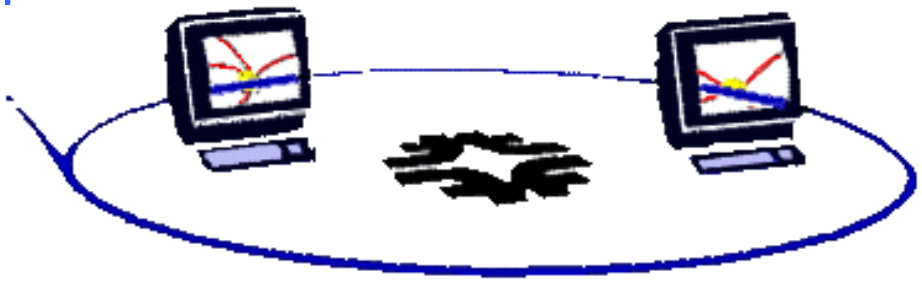
Mass Storage

- Enstore
  - http://hppc.fnal.gov/enstore/
  - ~2.6PB/8PB currently in storage systems
  - 6 STK silos
    - 20 9940A drives, 52 9940B
    - 20,191/34,077 slots used
  - ADIC AML/2
    - 2 arms, 3 towers
    - 9 LTO-1, 14 LTO-2, 8 DLT
    - 6885/13,920 slots used



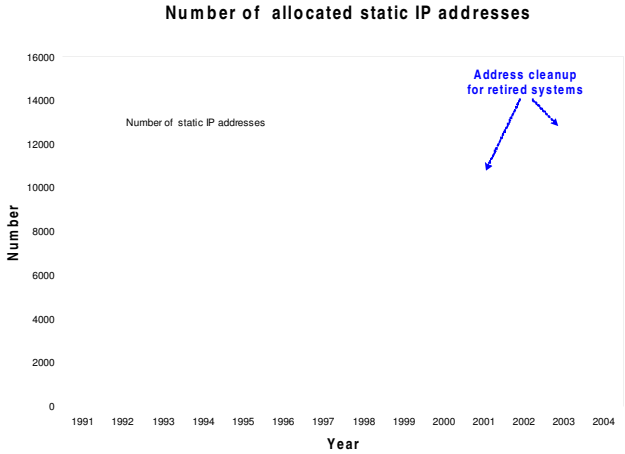
Mass Storage

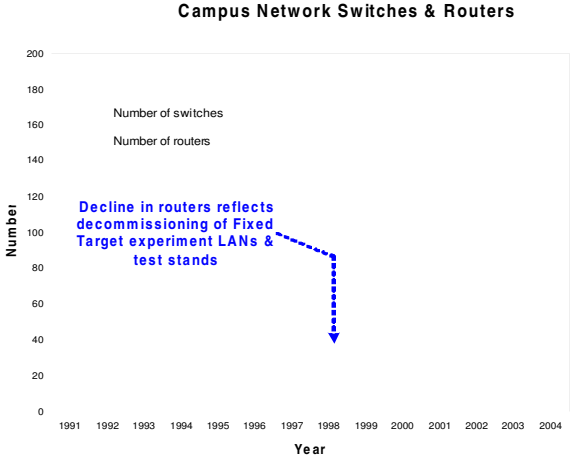
- Dcache
  - 225 TB Dcache front end disk
  - CMS T1 data transfers from CERN to FNAL
    - ~7GB/s to Dcache; 50MB/s to tape
- Administrative Info
  - 4 admins + 6 developers manage
    - 27K tapes, 7 libraries and >150 PC's
    - 24x7

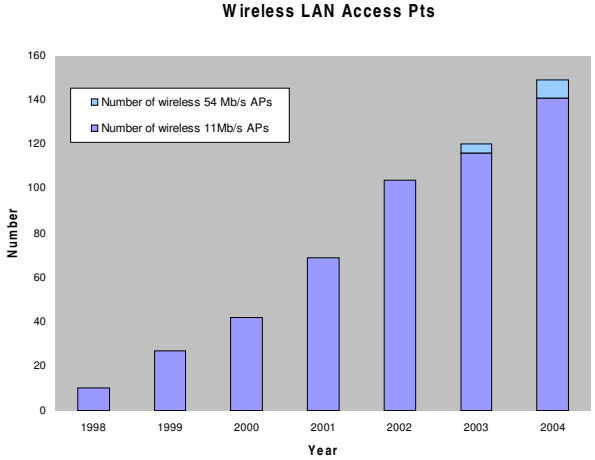


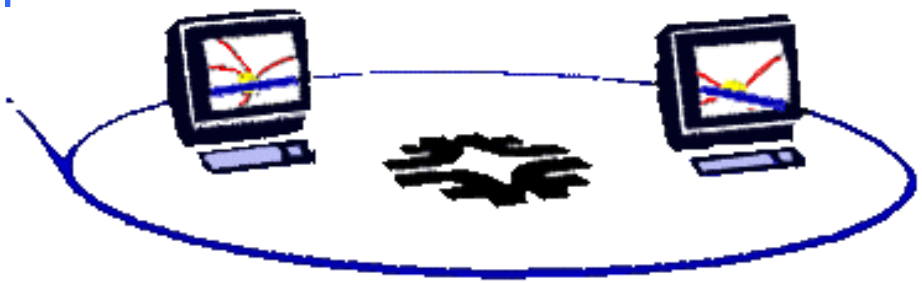
FNAL LAN Growth Trends

- Growth in systems continues at ~1000/year (below left)
  - Necessitates corresponding growth in # of switches (below center)
  - Emerging new technologies (eg., WLAN) adds to growth & complexity
  - System growth rate likely to increase with CMS gearing up
- Upgrades in LAN technologies parallels system growth:
  - Systems now connected at 1000B-T by default
  - New switch uplinks correspondingly deployed at 10 GE



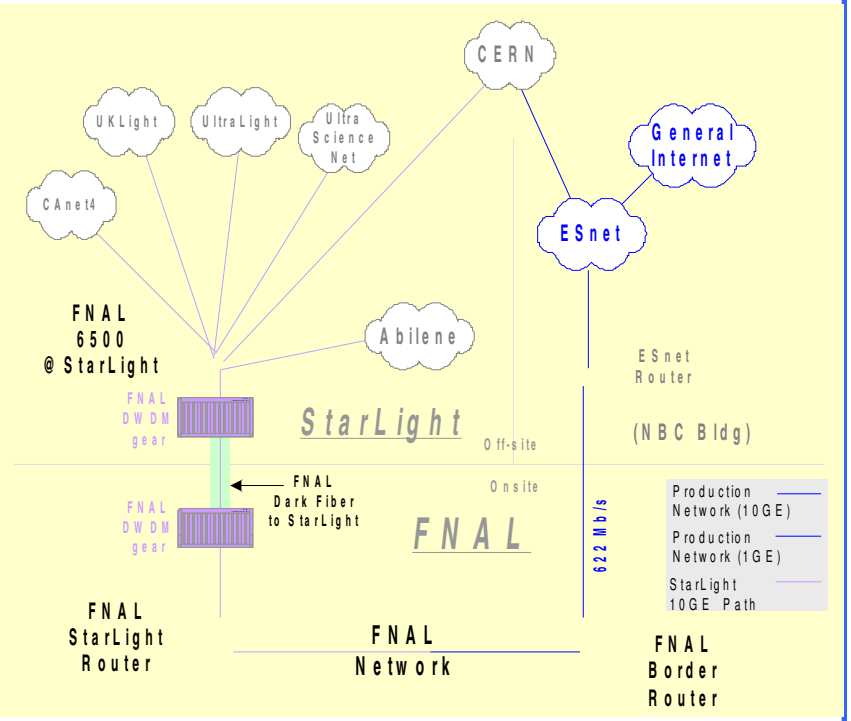


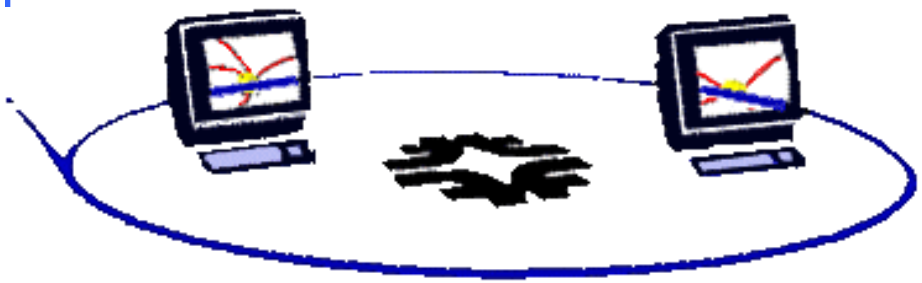




FNAL Wide Area Networking

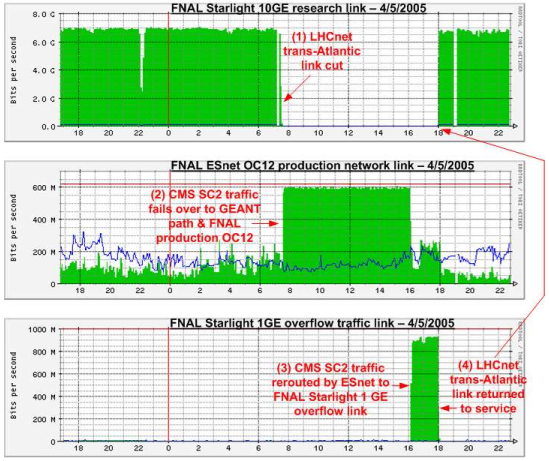
- Production WAN link (ESnet)
  - Only OC12 (622 Mb/s)
  - Compare to CMS challenges:
    - 5,000-10,000 Mb/s
- FNAL-leased dark fiber to StarLight:
  - Initial configuration 12,000 Mb/s
    - Theoretical capacity = 330 Gb/s
  - Intended for programmatic and network R&D, redundancy, & overflow for production traffic
- Soon, FNAL production network rates of 10 Gb/s & higher
  - Good practice = backup link of similar capacity





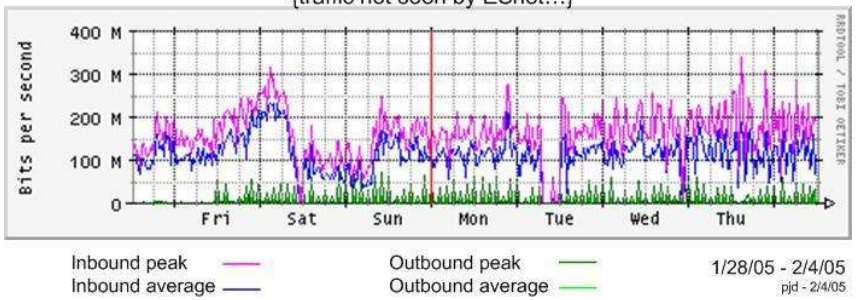
FNAL StarLight Link Usage

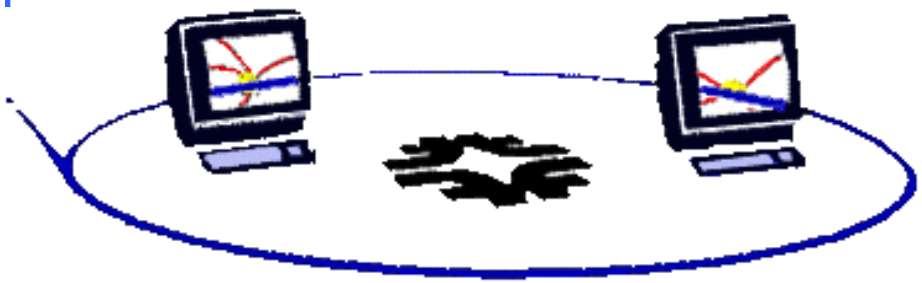
- R & D projects:
  - CMS robust service challenge sustained 7 Gb/s for ~3 days
    - Storage system to storage system
    - Survived trans-Atlantic cable cut…
- Overflow production traffic:
  - CERN CastorGrid traffic
  - Westgrid traffic
  - Working with McGill, UCL, others…
- Redundant off-site link:
  - Automated failover for Esnet link utilized 2-3 times already
  - Reliability still an issue; 2 extended outages last year



OverFlow Production Traffic directed over FNAL R&D StarLight Link

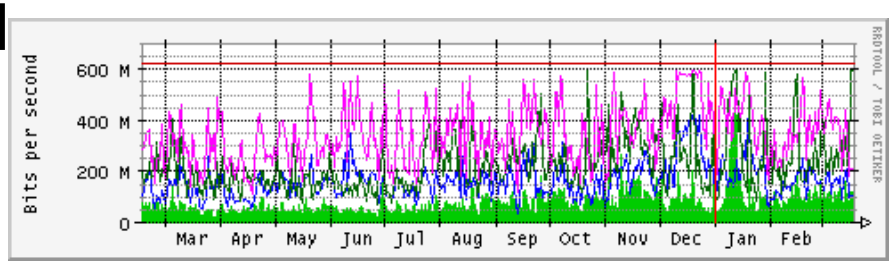
{traffic not seen by ESnet…}





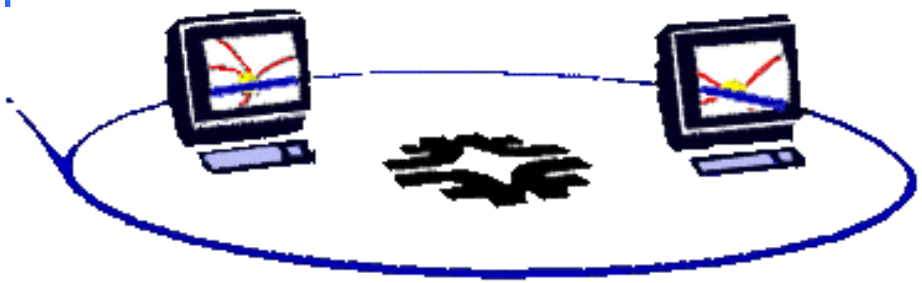
General WAN Growth Trends

- Historical trend – FNAL WAN traffic doubles every year:
  - More than held last year
  - Graph doesn't include overflow link traffic (Jan & Feb, '05…)

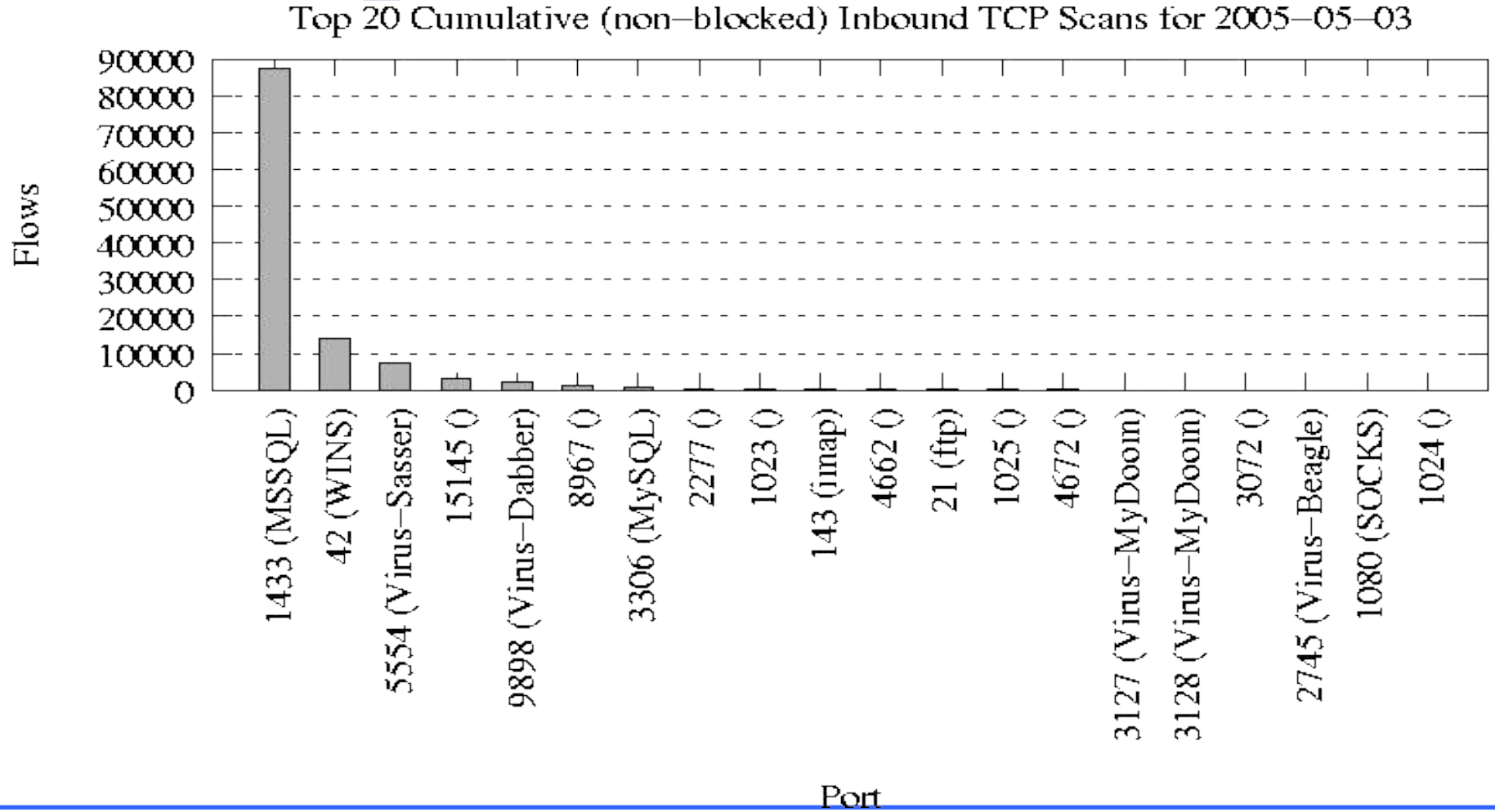


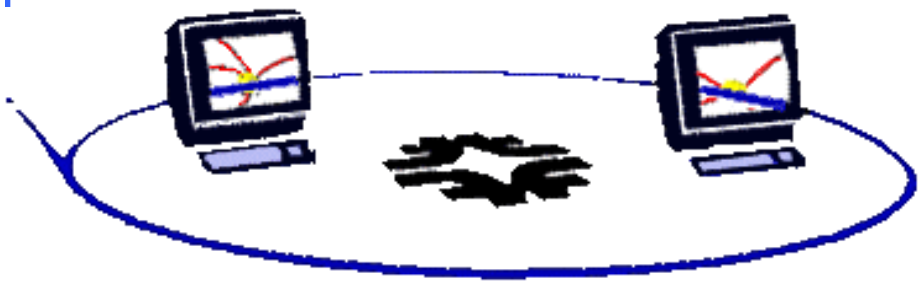
- StarLight link bandwidth is enabling technology:
  - Sites that have requested high bandwidth paths to FNAL →
  - Nearly all projects are intended to migrate quickly to production network use

Project	Collaboration	Remote Site	Status	Max B.W.	Type
UKlight - CDF	CDF	UCL	(semi)Active	4 Gb/s	periodic
CMS/CERN	CMS	CERN	Active	1-10 Gb/s	sustained
LambdaStation	CMS	CalTech/UCSD	Start 3/2005	1-10 Gb/s	sustained
Toronto HEP	CDF	U. Toronto	Pending	1-10 Gb/s	periodic
Vanderbilt	OSG(?)	Vanderbilt	Pending	1-10Gb/s	periodic
WestGrid (Ca)	D0	Simon Fraser	Active	1 Gb/s	sustained
LHC Tier 1	CMS	BNL	Active	< 622 Mb/s	periodic
UltraLight	CMS	Cal Tech, UF	Pending	1-10 Gb/s	sustained
UKlight – B.W.	CD	Manchester	Pending	1-10 Gb/s	periodic
UKlight – D0	D0	Lancaster	Pending	4 Gb/s	sustained
ASnet	CDF	Sinica, Taiwan	Active	2-4 Gb/s	periodic
CMS Tier 2	CMS	Florida	Pending	< 622 Mb/s	sustained
CMS Tier-2	CMS	Wisconsin	Pending	1-10 Gb/s	sustained
McGill	CDF & D0	McGill	Active	1 Gb/s	sustained

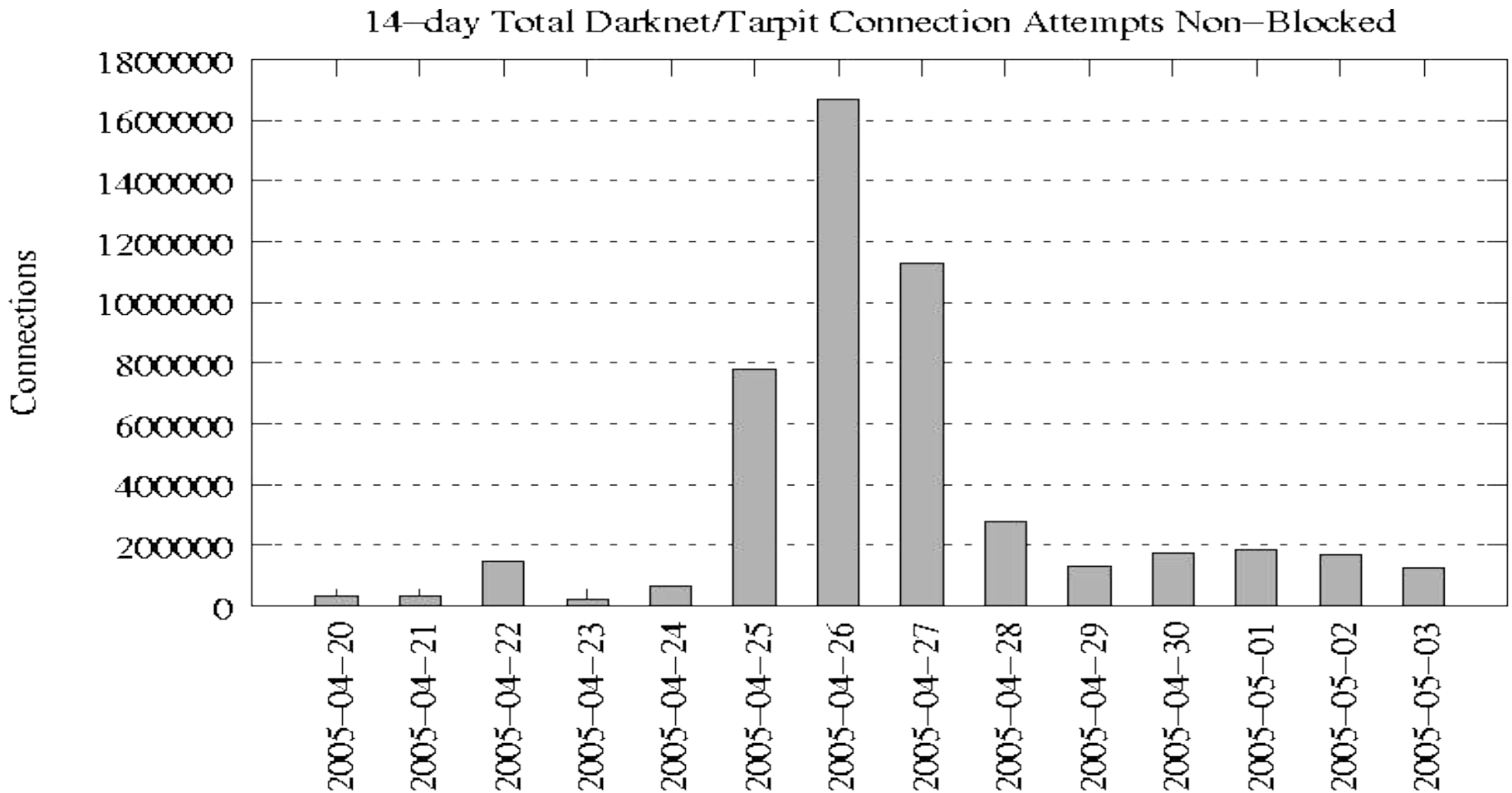


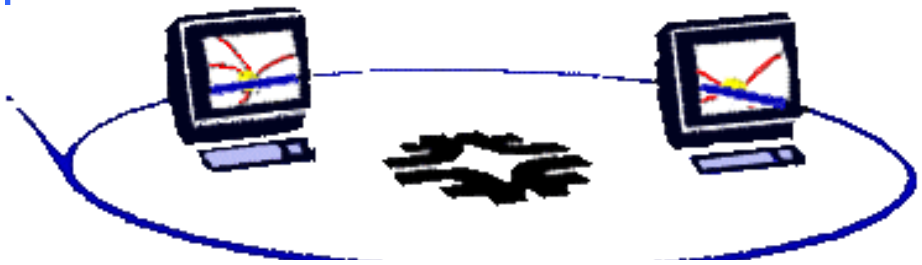
Security: Darknets and Tarpits
Port scan traffic flows



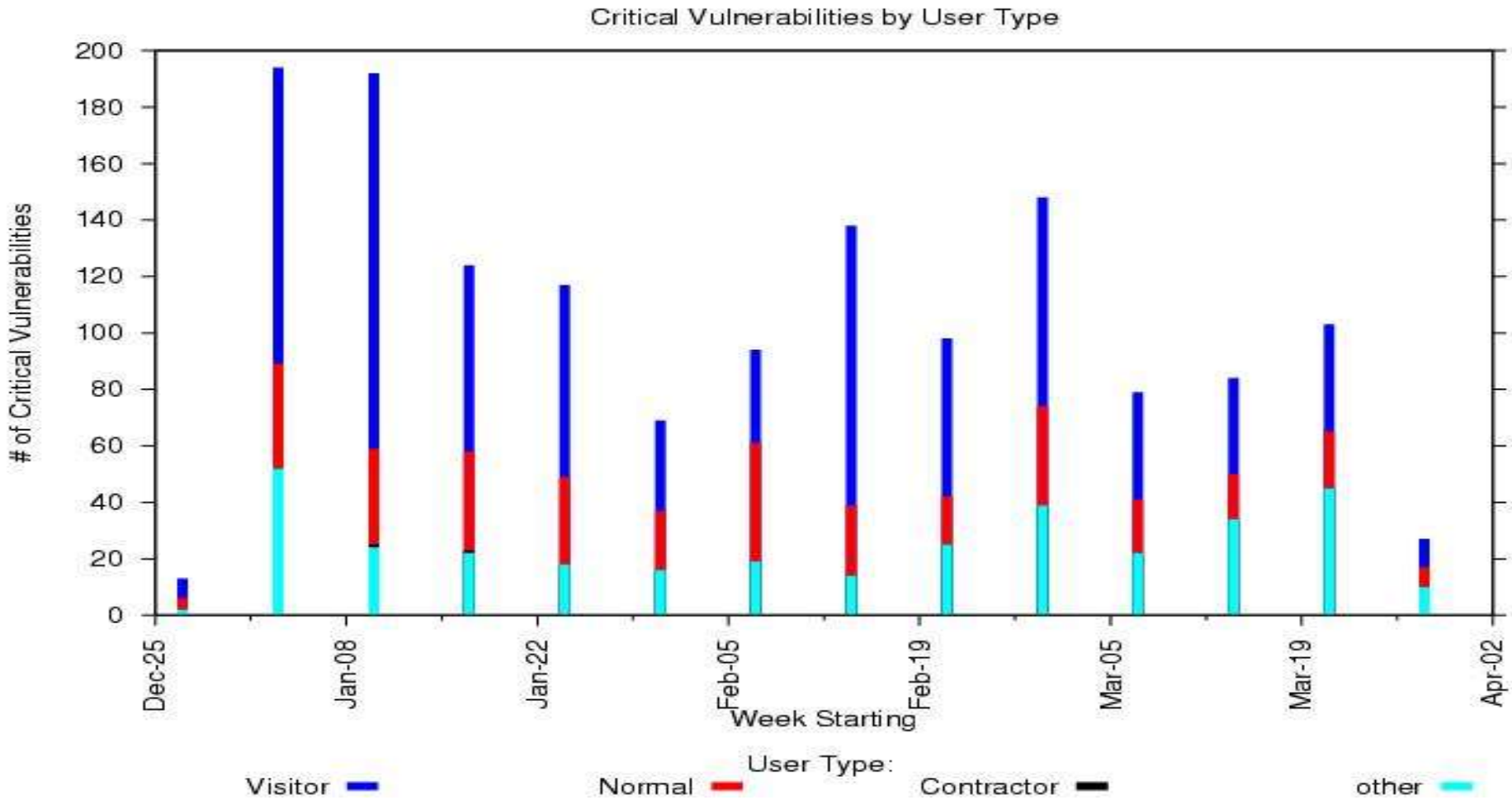


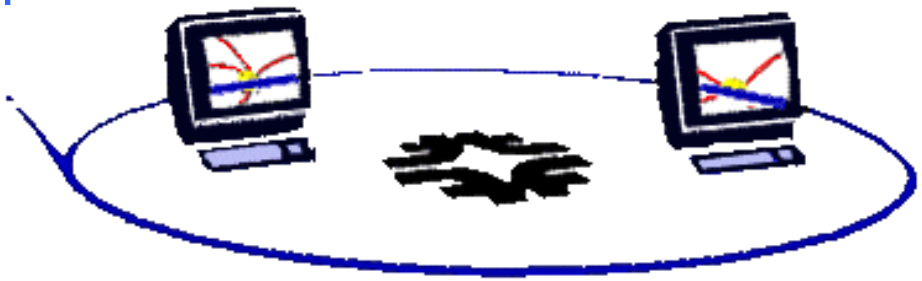
Security: Darknets and Tarpits Connection attempts





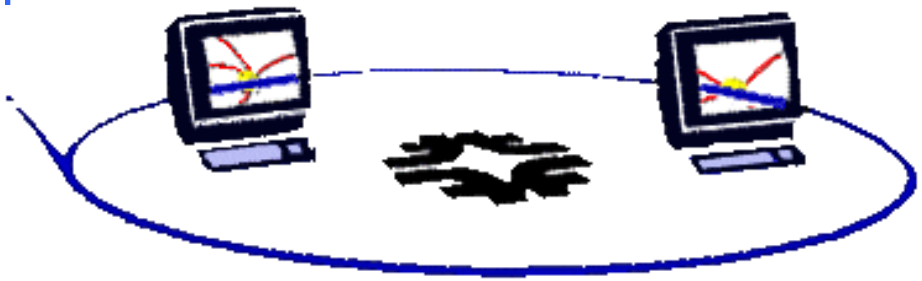
Security: Vulnerabilities Found





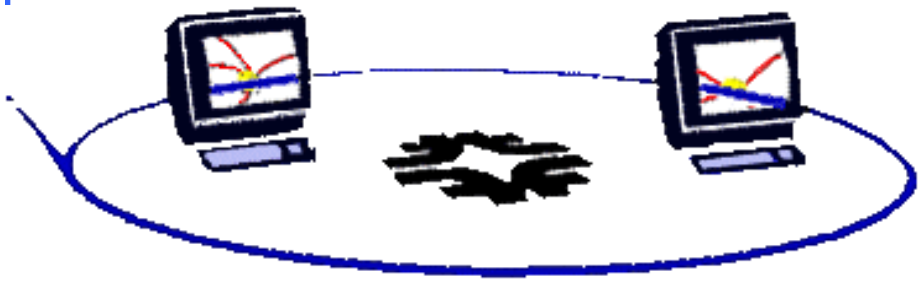
Scientific Computing

- General Purpose Farms
  - Migrating to Condor batch system,
  - New Linux head node,
  - Added grid interface
  - 102 production + 46 test nodes.
  - (See K. Schumacher talk in batch system WS)
- Run II Farms
  - 168 nodes for CDF, 444 for D0 reconstruction farms
- Linux cluster installs utilizing ROCKS
  - One master frontend serves several smaller clusters
  - Configured only dhcp server on smaller cluster which talks to DB on frontend



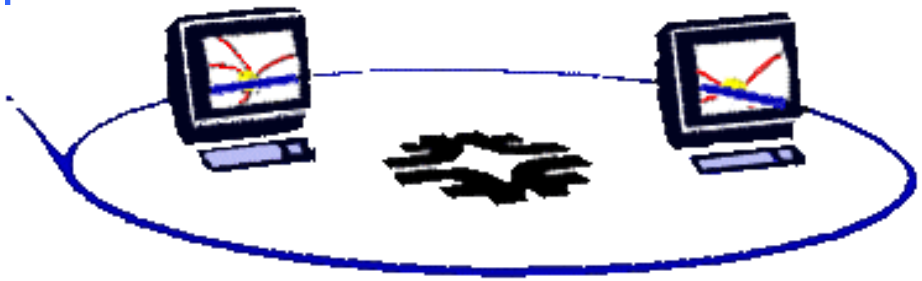
Scientific Computing

- Evaluation of Opteron CPU's,
  - (see K. Schumacher talk)
- Using Tibs for all SCS backups
  - ~300 systems, 3TB
- Managing and deploying gateway for FERMI Grid, site grid service
- New interactive login cluster for MINOS



Core Servers & Infrastructure

- Rollout of SMS began Marc '05
  - Using SMS for patch management, software deployment, license compliancy and inventory
  - Currently 355/2750 machines have client installed
- Upgraded to v8.5 of Meetingmaker calendaring system
- Windows 2003 server migration well underway.
  - Alpha test domain running Windows 2003 Active Directory
- Migrating windows backups to TIBS
- Upgrading Windows desktops to XP and use of XP firewall
- Increasing use of GPO (group policies) in W2K domain
  - Help to enhance and control desktop/server policies and software deployment



Core Servers & Infrastructure

- Anti-spam being refined
  - Deployed bayesian filtering and reporting tool
- Continued Plone development
  - See Plone talk being given by Marc Mengel
- Production Jabber instant messenger server rolled out
- Open source software continues to be deployed via The Written Word
- LSF – currently at v5.1
  - Usage dropping – 158 active licenses
- NGOP- 55310 monitored elements on 2362 hosts
  - New server deployed to alleviate performance issues
- WEB
  - Urchin web reporting in production
  - Preparing to add 5th host to web farm