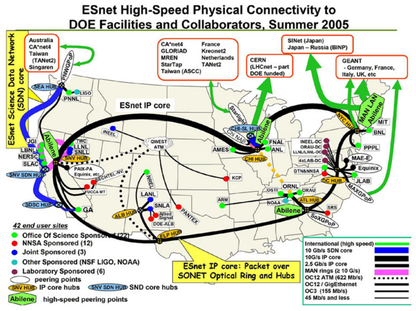
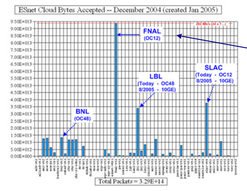
Production WAN Service / ESnet

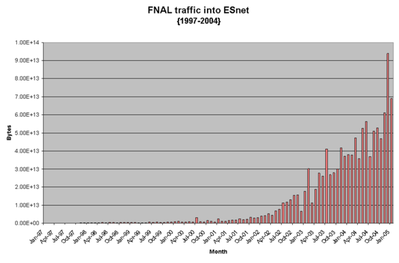


- ESnet is a high-speed network serving thousands of Department of Energy scientists & collaborators
  - Funded by the DOE Office of Science in support of the agency's research missions
  - The network service provider for the National Labs, including Fermilab



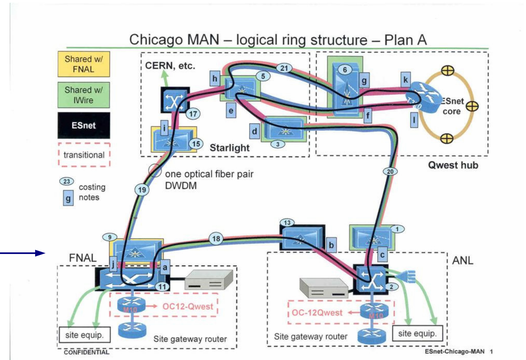
Fermilab & ESnet:

- Fermilab: ESnet's #1 traffic generator
- Fermilab's traffic levels grow at an exponential rate
  - US-CMS traffic will accelerate this traffic growth rate

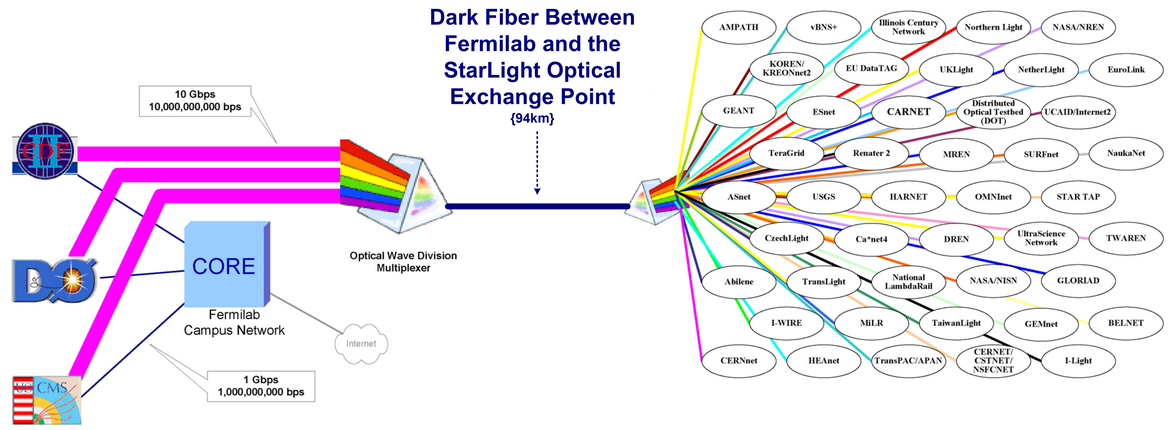


DOE Network Plans for the Future:

- A 3-tiered network architecture:
  - Production IP general network service (ESnet-provided)
  - Scientific Data Network (SDN) for high impact data movement, such as HEP experiment data (ESnet-provided)
  - Research network for development of advanced optical network technologies & applications (UltraScience Net)
- ESnet Chicago Metropolitan Area Network (MAN)
  - Optical infrastructure to support the DOE 3-tiered architecture
  - Fermilab's StarLight dark fiber will be a cornerstone of the MAN

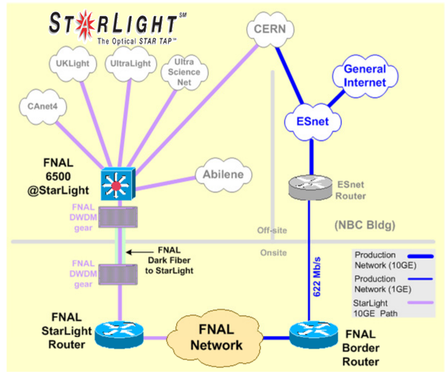


Fermilab's StarLight Dark Fiber Infrastructure

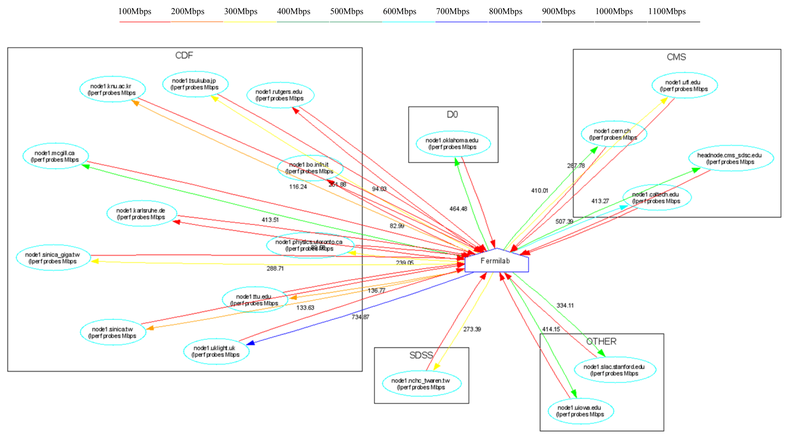


In 2004, Fermilab Leased an optical fiber pair to Starlight (downtown Chicago):

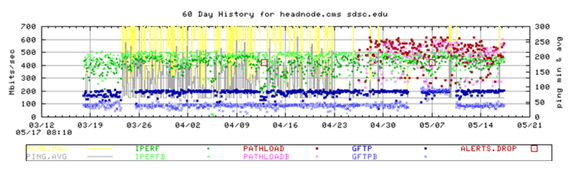
- StarLight = international optical network exchange point
- Link optics by Ciena DWDM gear
  - Capable of supporting 33 λs
  - Each λ can carry 1-10 Gb/s today, and 40 Gb/s in the future
  - Initial configuration: one 10Gb/s channel & two 1 Gb/s channels
- Intended uses for the StarLight channels:
  - WAN advanced research & development projects
  - Overflow link for our production WAN traffic
  - Redundancy for ESnet link
- Today, our research network traffic and US-CMS pre-production traffic (purple) use the very high bandwidth StarLight paths
  - Normal production traffic (blue) still supported on the ESnet link



Network Performance Monitoring

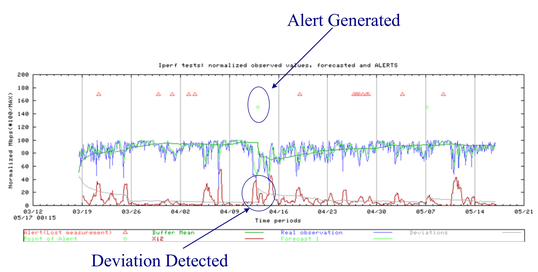


- IEPM-BW collects network path performance data
  - Network path capacity (Pathload)
  - Optimal memory-to-memory data transfer rate (Iperf)
  - Disk-to-disk data transfer rates (GFTP)
  - Round trip response time (Ping) & packet loss (pingMax)
- Monitoring conducted in both directions



- Large, distributed HEP collaborations, such as D0, CDF and US-CMS, depend on stable, robust network paths between major world research centers
- Critical need for WAN monitoring with:
  - Bi-directional active network probes
  - Path characterization
- Fermilab has an active WAN monitoring infrastructure, based on Internet End-to-End Performance Monitoring (IEPM-BW) Tool

- Anomalous event detection used to indicate possible change in network path performance
- Detection based on tracking degree of deviation in observed throughput versus forecast throughput:
  - Forecasting technique: tri-exponential smoothing



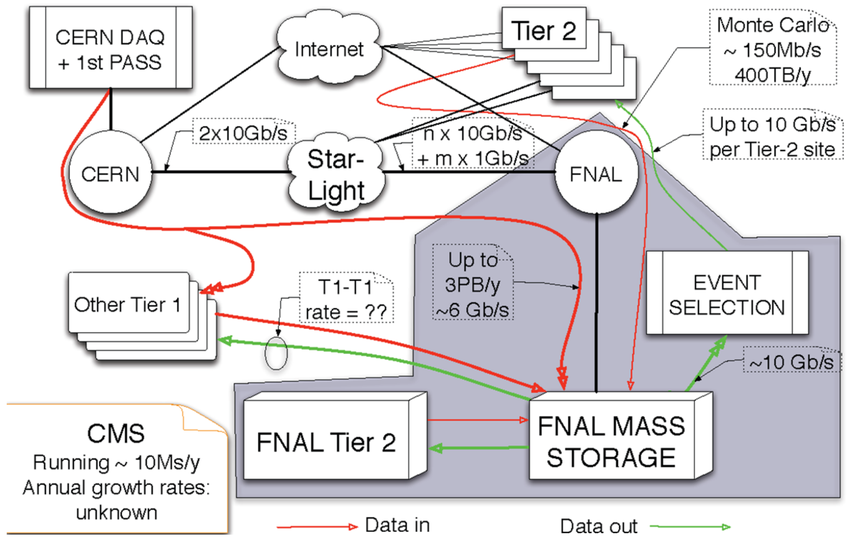
Large Scale Data Movement Projects

- StarLight link bandwidth is a technology "enabler"
  - Offsite data analysis becomes feasible with high bandwidth
- A significant number of sites have requested high volume data movement projects to/from FNAL:
  - Fermilab Wide-Area Working Group (WAWG) coordinates & facilitates these projects (wawg@fnal.gov)
- Nearly all projects are intended to migrate to production use quickly

Table of StarLight Link Projects

Project	Collaboration	Remote Site	Status	Max B.W.	Type
UKlight - CDF	CDF	UCL	Active	1 Gb/s	periodic
CMS/CERN	CMS	CERN	Active	1-10 Gb/s	sustained
LambdaStation	CMS	CalTech	Active	1-10 Gb/s	sustained
WestGrid (Ca)	D0	Simon Fraser	Active	1 Gb/s	sustained
LHC Tier 1	CMS	BNL	Active	< 622 Mb/s	periodic
UltraLight	CMS	Cal Tech, UFL	Pending	1-10 Gb/s	sustained
UKlight – D0	D0	Lancaster	Pending	1 Gb/s	sustained
ASnet	CDF	Sinica, Taiwan	Pending	2-4 Gb/s	periodic
CMS Tier 2	CMS	UFL, UCSD, Wisconsin	Active	< 622 Mb/s	sustained
McGill	CDF & D0	McGill, Ca	Active	1 Gb/s	sustained
Twaren	SDSS	NCHC, Taiwan	Active	1 Gb/s	periodic
Prague, Cz	D0	IoP	Active	1 Gb/s	sustained
Toronto HEP	CDF	U. Toronto	Pending	1-10 Gb/s	periodic

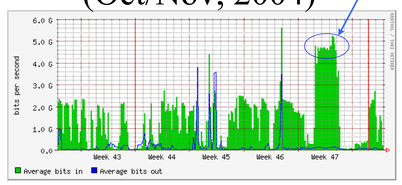
Planning for LHC Traffic



CMS Robust Service Challenges: Moving Pre-Production Data in Preparation for LHC

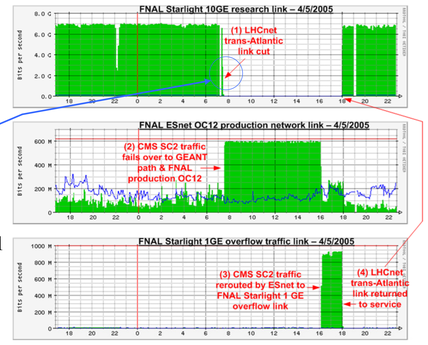
- Sustained ~2.5Gb/s for approximately 5 weeks
- Data transfers were storage system to storage system
  - GridFTP -
  - Standard 1500 Byte MTUs
- 3-day peak load @ 4.75 Gb/s

Service Challenge I (Oct/Nov, 2004)



- Sustained ~7 Gb/s for 3-4 days
- Again, storage system to storage system transfers
- CERN<-> StarLight link failed during test
  - Traffic failed over to alternate path, saturating 622Mb/s production network link
  - Network path manually rerouted onto FNAL StarLight 1 Gb/s channel
  - Restoration of CERN <-> StarLight link resulted in resumption of 7 Gb/s throughput

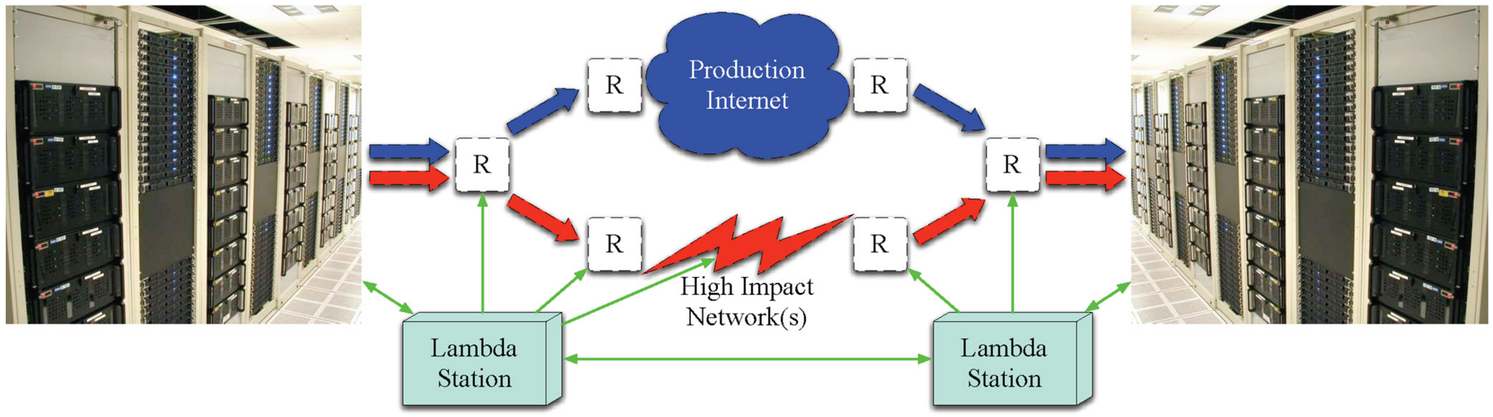
Service Challenge II (April, 2005)



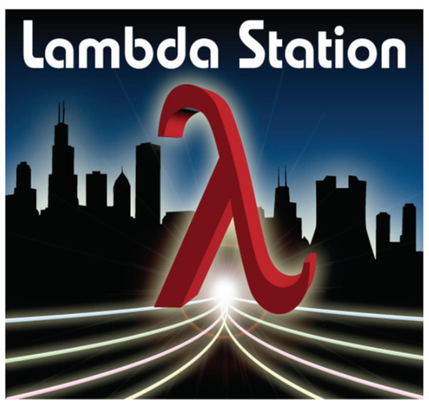
Lambda Station Project

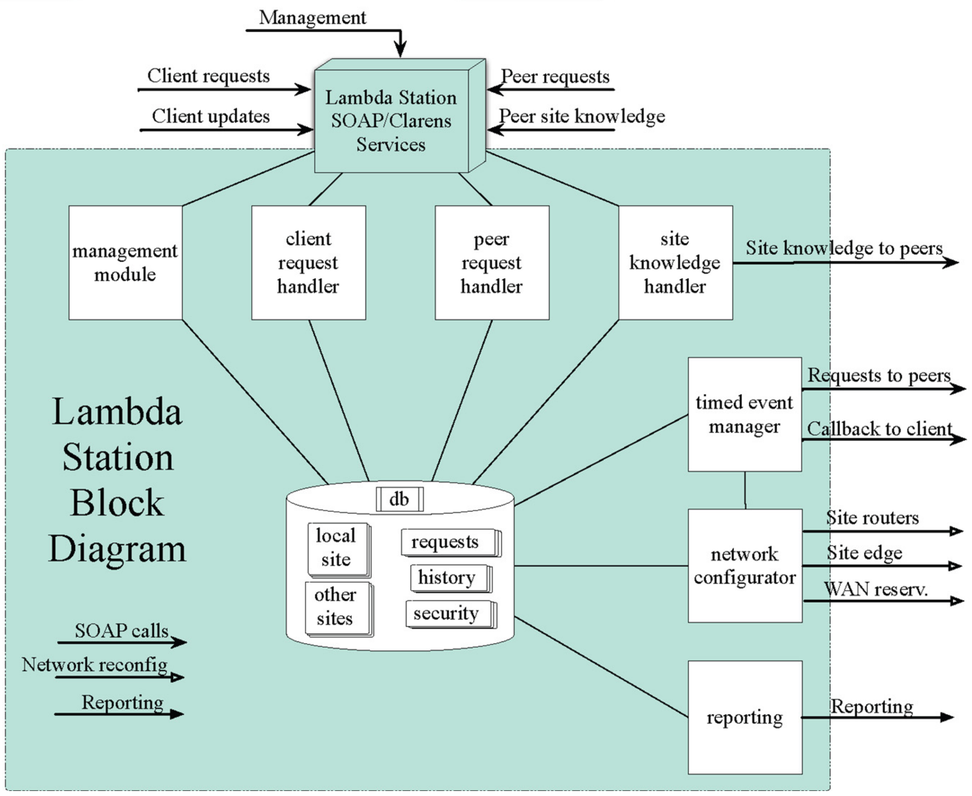
Problem space:
- "High impact" data movements benefit from special WAN links
- Other traffic should not be disturbed
- Hosts should not have to know network details
- Network administrators must retain control

Lambda Station's function:
- Allocate network paths for high-impact applications
- Adjust internal site routing and site-edge routing
- Perform any required WAN reservation & setup actions
- Steer selected flows on the most granular possible basis



Lambda Station has a service-oriented architecture with grid security methods. It is built on the Clarens web services (SOAP) middleware.





Effect of Lambda Station path changes on established TCP data flows

The figure below is a record of traffic sent from a single computer at Fermilab to a computer at Caltech. The traffic consisted of 18 parallel TCP streams. The source machine had dual 750MHz CPUs and a 1 gigabit/second network interface.

The network configuration is similar to the diagram to the left, with the blue network being ESNET and the red network UltraLight.

One of the requirements on Lambda Station is that data connections must survive a change in the network path. In this test, the wide-area path was switched every 15 minutes between the production network (ESNET) and a dedicated network (UltraLight). Corresponding changes in internal site routing were also made.