

Title:	TeraPaths: A QoS Enabled Collaborative Data Sharing Infrastructure for Peta-scale Computing Research ¹
PIs:	Bruce G. Gibbard and Dantong Yu
PI Institution:	Brookhaven National Lab
Senior Personnel:	Shawn Mckee
	University of Michigan

Summary of report:

The primary goal of this project is to investigate the use of LAN QoS and MPLS technology in the ATLAS data intensive distributed computing environment. A number of corollary objectives will also be achieved.

- 1) Expertise in MPLS based QoS technology will be developed which will be important to ATLAS and the LHC community more generally.
- 2) At Brookhaven National Laboratory (BNL) the ability to dedicate an equitable fraction of the available WAN bandwidth via MPLS to ATLAS Tier 1 data movement will assure adequate throughput and limit its disruptive impact on the RHIC program and other more general laboratory network needs.
- 3) The project will enhance technical contact between the ATLAS Tier 1 at BNL and its network partners including the Tier 0 center at CERN, ATLAS Tier 2's and other members of the Open Science Grid community of which it is a part.

Detailed Activities and Achievements:

Terapaths deployed and prototyped the use of differentiated networking services based on a range of new transfer protocols in support of global data movement. The managed network capability being enabled by this project will be integrated, and scheduled as part of Grid computing systems, along with the managed CPU and storage resources, to enhance the overall performance and efficiency of DOE computing facilities.

The Terapath project is enabling data transfers with guarantees of speed and reliability that are crucial to applications with deadlines, expectations and critical decision-making requirements. Brookhaven National Lab needs to do RHIC (Relativistic Heavy Ion Collider) production data transfers and LHC (Large Hardon Collider) Monte Carlo challenges between BNL and the remote collaborators. The aggregate of their peak network requirements is beyond BNL capacity. Our project can modulate LHC data transfers to opportunistically utilize available bandwidth, ensuring that the RHIC production data transfer is not impacted.

Since last project report, we experimentally proved that LAN QoS did not impact the overall network utilization. We did not see performance deterioration when we enabled the QoS policy. We acquired capability to configure network equipment to dedicate fractions of available bandwidth via QoS to various BNL data movement/replications to assure adequate throughput and limit their disruptive impact upon each other.

Software development proceeds smoothly. The system architecture can easily be integrated with other network management tools to provide a complete end-to-end QoS. We also created for software development and testing a [full-featured testbed](http://www.atlasgrid.bnl.gov/terapaths) with the same cisco switches as used in

¹ Project Website: <http://www.atlasgrid.bnl.gov/terapaths>

the production network. We are collaborating with OSCAR, Lambda Station and DWMl projects to define common interface and communication protocol to facility system integration.

- Designed several LAN QoS mechanisms and evaluated their performance:
 - Weighted sharing method: for each type of class of service (COS), a weight is assigned to this class of service. When there is no contention, all types of traffic can utilize unused bandwidth. During network congestion, the forwarding is based on the ratios of the assigned weights, thus allowing us to define the relative allocated bandwidths for the classes of service.
 - Marking down: the traffic is classified into several classes of services. Whenever the high level class of service uses up its allocation, the packet will be marked down to a lower class of service. Alternately we can enforce the allocation by dropping excess packets. The implications of this policy choice (mark-down vs. drop) will be explored.
- Setup, use, manage and remove Lan QoS paths from our end hosts to BNL boarder routers with different priority and bandwidth allocations for different applications from multiple organizations.
- Collaborate with OSCAR to make end-to-end network paths with allocated bandwidth. We studied the LAN QoS/MPLS path and verified that the traffic with specified source, destination ip addresses was correctly labeled and tunneled. The traffic was provided with pre-arranged Bandwidth. There was no performance impact to network when QoS is enabled in BNL production network. The high speed network (10 Gbps and up) will be tested to ensure that these special QoS/MPLS configurations do not interfere with the high data transfer rate.
- We have been developing a network manage systems based on Web service architecture: Web Services portal receives user requests in the format of HTML request or programming API; several backend processes handle user requests, construct routes, and schedules network to satisfy these requests; A database stores all information on current reservations, history, network utilization; Several modules are responsible to negotiate with remote peers, push the QoS into local network routers, and generate reports for accounting and auditing.
- We have been collaborating with the SLAC team to analyze network monitoring needs and requirements. SLAC DWMl network monitoring tools were deployed at BNL. IEPM_BW software was installed on a gigabit-connected computer node on the QoS enabled network path. The SLAC team provides help with the system installations.

The Impact to Specific DOE Science Application:

Our project is a central component of a responsive, managed infrastructure for eScience. Such an infrastructure can ensure that physics event data can be delivered to collaborators a few hours after its generation, regardless of the collaborators' physical distances to data source, local or overseas, providing equal opportunities for all participants.

Synergy Developed with DOE Application Developers to facilitate Technology Transfer:

Project participants are involved in management for LHC network and ATLAS Tier 1 facility. We are directly involved with Open Science Grid. Shawn Mckee is the Co-chair OSG networking group to ensure the technology transfer to this newly formed Grid community. We are collaborating with OSCAR, DWMl, LambdaStation, UltraLight and their home institutions, University of Michigan, SLAC, Fermi Lab and Esnet, to jointly set up end-to end QoS paths. Our BNL colleague designed and developed the OSG user management systems to handle Grid authentication and Authorization. It will be used to enhance the security and usefulness of Terapaths.