Available online at www.sciencedirect.com
SCIENCE DIRECT®

Future Generation Computer Systems 22 (2006) 949–954

FGCS
FUTURE GENERATION COMPUTER SYSTEMS
www.elsevier.com/locate/fgcs

AMROEBA: Computational astrophysics modeling enabled by dynamic lambda switching

Joe Mambretti*, Rachel Gold, Fei Yeh, Jim Chen

International Center for Advanced Internet Research, Northwestern University, Chicago, IL 60611, United States

Available online 2 May 2006

Abstract

Many data and compute intensive Grid applications, such as computational astrophysics, may be able to benefit from networking supported by dynamically provisioned lightpaths. To date, the majority of high performance distributed environments have been based on traditional routed packet networks, provisioned as external services rather than as integrated components within those environments. Because this approach often cannot provide high performance capabilities required by these applications, an alternative distributed infrastructure architecture is being designed based on dynamic lightpaths, supported by optical networks. These designs implement communication services and infrastructure as integral components of distributed infrastructure. The resultant environments resemble large scale specialized instruments. Presented here is one such architecture, implemented on a wide-area, optical Grid test bed, featuring a closely integrated dedicated lightpath mesh. The test bed was used to conduct a series of experiments to explore its potential for supporting adaptive mesh refinement (AMR) astrophysics simulations. While preliminary, the results of these experiments indicate that this architecture may provide the deterministic capabilities required by a wide range of high performance distributed services and applications, especially for computational science.

Keywords: Grid; Grid networking; Optical networking; Lambda Grid; Lightpath; Distributed infrastructure; Computational science; Wavelength switching

1. Introduction

AMROEBA (Adaptive Mesh Refinement Optical Enzo Backplane Architecture) is an experimental architecture designed to investigate the behavior of resource intensive applications within a distributed Grid environment, based on dynamic lightpath provisioning. This architecture was implemented on a wide area optical Grid test bed, closely integrated with a dedicated lightpath mesh. A series of experiments were conducted on this test bed to explore the potential of this approach for supporting high performance computational science, including adaptive mesh refinement (AMR) astrophysics simulations.

Currently, almost all Grids [1] are supported by traditional data networks, which provide a general, best effort service. Although this model has been highly scalable and successful for meeting multiple general data communication needs, it can be a restrictive barrier to the deployment of specialized capabilities. Such capabilities include highly differentiated and deterministic services and support for large scale sustained data flows, especially those used for live interactive access to data that is continuously updated [2,3]. Deterministic services are those that can be specified and predicted in advance, with a high degree of certainty that they will be provided by the network. In contrast, best effort communications provide no firm service quality guarantees. To address these issues, recent research has been investigating the potential for supporting large scale data intensive distributed processes by closely integrating Grids with high performance, dynamic lightpaths, supported by optical networks [4–9]. Such integration transforms external data networking from an external service to a "first class" Grid resource, one that can be fully controlled by Grid processes.

2. Computation clusters

The computational clusters used for the experiments were comprised of different mixtures of high performance data and

* Corresponding address: Northwestern University, International Center for Advanced Internet Research, 750 North Lake Shore Drive, Suite 600, Chicago, IL 60611, United States. Tel.: +1 312 503 0735; fax: +1 312 503 0745.
E-mail address: j-mambretti@northwestern.edu (J. Mambretti).

0167-739X/$ - see front matter
doi:10.1016/j.future.2006.03.008

computational dual processor nodes (2 GHz Xeon processors), with at least 1.5 Gb of RAM. Each node was provisioned with RedHat 7.3 running Linux 2.4.20-20.7 smp and had/home NSF mounted across the nodes. Job submission was controlled by OpenPBS (2.3.16). MPICH (1.2.5) and Lam-MPI (7.0.4) libraries are used for messaging. Each node was implemented with GCC 2.96 or GCC 3.0.x. The non-optimized internal data rate among any two random nodes within these clusters ranged from 920 Mbps to 940 Mbps. Each cluster node was connected through a Gigabit Ethernet (GE) interface with a high performance Layer 2 (L2) switch supporting a meshed point-to-point connections and reconfigurations among the clusters at each site. The L2 switches provided direct channels to the wide area optical test-bed fabric through optical switches. The core optical fabric was extended directly into the research labs with dedicated fiber at each site.

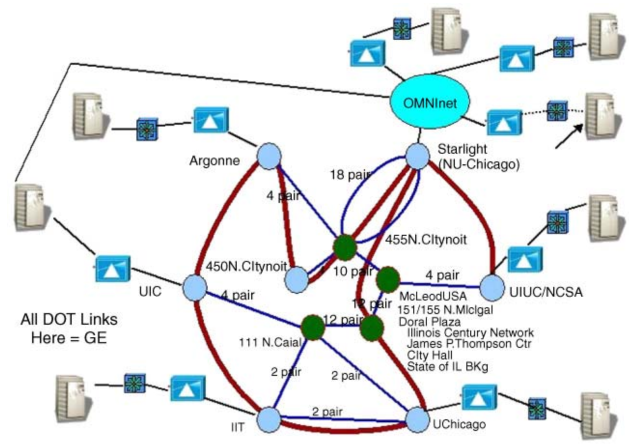

Fig. 1. Experimental optical testbeds.

3. Optical backplane components

The AMROEBA experiments were conducted on the large scale Distributed Optical Testbed (DOT), designed to investigate methods for providing applications with direct access to a controllable network environment. Data communications on this test bed are completely non-routed—transport is supported exclusively by Layer 2 (L2) and Layer 1 (L1) paths. DOT was established by interconnecting, with a mesh topology, distributed computational and data clusters on five separate, but interrelated, optical network domains. OMNInet [10], a metro area optical test bed, I-WIRE, a state-wide optical network (ref. Fig. 1), StarLight, an international optical network exchange facility [11], NetherLight, a trans-Atlantic research lightpath [12], and StarPlane, an optical test bed in the Netherlands [13]. Each domain transports L2 framed traffic through L1 bi-directional lightpaths, based on Dense Wave Division Multiplexing technology (DWDM). The distributed backplane used for these experiments was based on multiple persistent, high performance deterministic paths—dedicated local and remote optical transport channels. These dedicated channels can support high performance data flows (e.g., over 980 Mbps) with little or no packet loss for many hours.

The OMNInet optical test bed allows individual applications to define their own networks, to dynamically change the topology of the network by directly signaling to individually addressable lightpaths. Grid applications can directly create and reconfigure the specific topologies and levels of performance that they require, enabling them to take advantage of lightpaths that are flexible, adaptable, and deterministic. OMNInet incorporates a signaling technique that is part of a service layer, comprised of an implementation of the Optical Dynamic Intelligent Network (ODIN) service layer architecture [9]. (ref. Fig. 2.) This experimental service layer provides capabilities for virtualizing ad hoc networks for applications, and for provisioning those networks discovering, using, and discarding individually addressable lightpaths. The service layer architecture consists of components that can accept and fulfill requests from clients for network

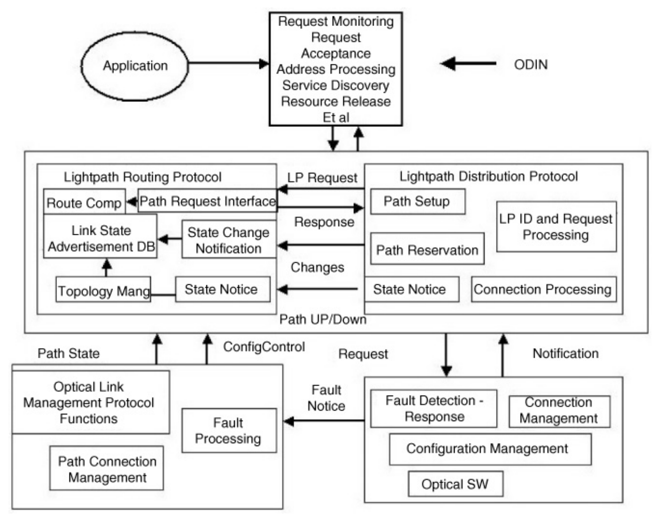

Fig. 2. ODIN architecture.

resources. The service layer creates all mechanisms required to transport data over the specified path, and then signals to edge processes and resources to enable the configuration required to use the virtual network that has been provisioned. A key component is an API, based on the Simple Path Control Protocol (SPC), that enables applications to directly signal for network resources [14]. The Simple Path Control Protocol was developed to enable external processes, including applications, to communicate messages that allow for such paths to be created, deleted, reconfigured, and monitored. The service layer uses Generalized Multiprotocol Label Switching (GMPLS) as a provisioning interface to configure photonic switches to implement the lightpaths (i.e., for resource discovery, link provisioning and deletion) [15]. GMPLS is an IETF standard that allows for IP based control of devices based on time-division multiplexing, wavelengths, and spatial switches. Forwarding decisions can be determined by time slots, wavelengths or ports.