Multi-layer Lambda Grid Properly Using Lambdas and Sub-lambdas

Yukio Tsukishima(1), Yasunori Sameshima(1), Akira Hirano(1), Masahiko Jinno(1),
Tomohiro Kudoh(2), Fumihiro Okazaki(2)
1: NTT Network Innovation Laboratories, NTT Corporation, 1-1 Hikari-no-oka, Yokosuka, 239-0847 Japan, Email:
tsukishima.yukio@lab.ntt.co.jp, 2: Information Technology Research Institute, AIST, E-mail: t.kudoh@aist.go.jp

Abstract

This paper proposes a multi-layer lambda Grid that properly provides the exact requested bandwidth to each client traffic flow among dynamically allocated computers according to each client request.

Introduction

To realize efficient and high performance utility computing, high-end applications will require resource management schemes that dynamically co-allocate computing and network resources with a guaranteed level of performance according to client requests. In this regard, [1-4] report experimental results on technology to co-allocate computers and lambdas. However, some of the applications require far less bandwidth than that provided by the lambdas. For various applications, not only lambdas but also sub-lambdas using packet switching should be provided in terms of cost performance. However, little has been done to propose a scheme that provides the exact requested bandwidth to each client traffic flow properly using lambdas and sub-lambdas.

To use properly a sub-lambda or lambda on the same network according to a client request, network edge routers (NERs) must be deployed in photonic networks, and the NERs must activate a packet filter with the abilities to filter incoming and outgoing packets between allocated computers, to control the throughput of a sub-lambda by policing, and to route only filtered traffic flows to the corresponding sub-lambda. The packet filter controls packets based on their content as indicated by their header. However, the packet content changes depending on the allocated computers, and therefore it is difficult to activate dynamically a correct packet filter according to each client request.

In this paper, we propose a multi-layer lambda grid that provides a lambda or sub-lambda with sufficient bandwidth to the allocated computers according to a client request. To achieve this, we propose a control scheme to configure the correct packet filter for the NERs. Finally, we present the results of the first demonstration of the lambda plus sub-lambda grid.

Multi-layer lambda grid

Figure 1 shows the architecture of the multi-layer lambda grid. This system comprises optical cross connects (OXCs), a grid resource scheduler (GRS), a network resource manager (NRM), computer resource managers (CRMs), NERs, and computers. Users can book suitable computer and network resources via grid middleware using request messages that contain the number of CPUs, bandwidth between computer sites, and duration time. Based on the request from the grid middleware, the GRS reserves a sub-lambda or lambda between allocated computer sites with the NRM via grid network service-web service interface version 2 [2], and reserves computers with the CRMs. The NRM and the CRMs manage reservation timetables of the managed resources and activate the resources at each reserved time.

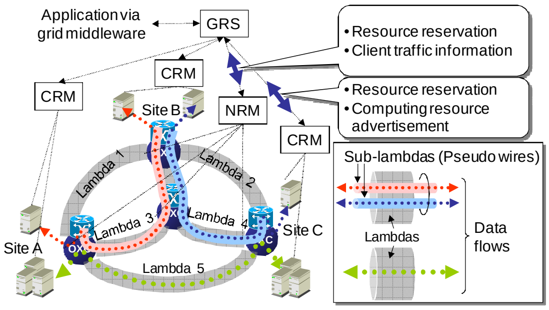


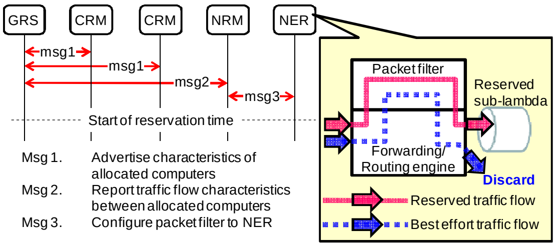
Figure 1: Multi-layer lambda grid architecture

Before activation, the NRM collaborates with the GRS and CRMs to configure dynamically packet filters to the NERs as indicated in Fig. 2. The GRS, NRM, and CRMs do not know the information pertaining to the packets flowing between allocated computers before the collaboration starts. However, the CRMs know the computers allocated in their own domain, and the GRS also knows the allocated computer sites. The CRMs advertise the characteristics of the allocated computers in their own domains to the GRS using message (msg) 1s, which generally contain the MAC and IP address of the allocated computer interface. The GRS computes and specifies the characteristics of the packets flowing between the allocated computers based on msg 1s, and creates msg 2. Msg 2 contains some of the source and destination IP addresses, MAC addresses, and other packet information. The GRS sends msg 2 to the NRM, and the NRM creates msg 3s based on the combination of msg 2 and the request from the grid middleware.

978-1-4244-2229-6/08/$25.00 ©2008 IEEE

Moreover, the NRM configures the correct packet filters for the NERs via msg 3s.

The NERs filter the incoming and outgoing packets in the traffic flow between the allocated computers using the packet filter, and control the throughput of the allocated sub-lambda via the packet filter. In addition, the NERs route the filtered packets to the allocated sub-lambda based on the packet filter and routing and forwarding configuration. The NERs can also discard the best effort traffic to limit its throughput.



Msg 1. Advertise characteristics of allocated computers
Msg 2. Report traffic flow characteristics between allocated computers
Msg 3. Configure packet filter to NER

Figure 2: Collaboration for multi-layer lambda grid

Experimental configuration over a testbed

Figure 3 shows the experimental setup to evaluate the feasibility of the multi-layer lambda grid. Domain 1 and domain 2 are interconnected via JGN2 [2] and GEMnet2 [3]. AIST-NRM-managed domain 1 comprises a router and two OXCs, and NTT-NRM-managed domain 2 comprises two routers and two OXCs. The routers are connected through Giga-bit Ethernet links via lambdas. NTT-NRM controls OXCs 1-2 via GMPLS, and configures routers 1-2 via telnet to set up lambdas and sub-lambdas. AIST-NRM controls OXCs 3-4 and router 3 via SSH to control lambdas and sub-lambdas. As for applications, the Message Passing Interface (MPI) [4], Super High Definition video (SHD) [5] and FTP are used, and they request sub-lambdas via AIST-GRS. The best effort traffic continually traverses over the network.

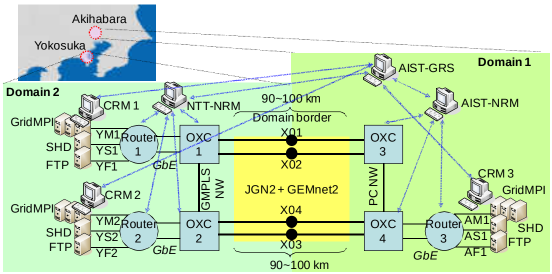
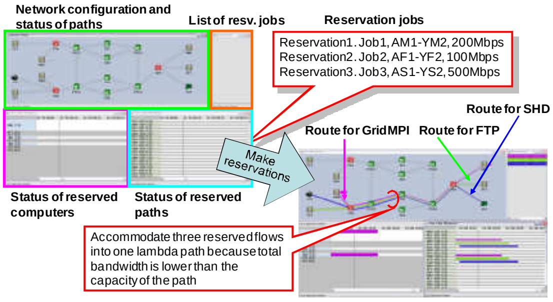


Figure 3: Experimental network configuration

Experimental results

Figure 4 shows screen shots of experimental results extracted from the NRM and GRS execution monitor. MPI, FTP, and SHD requested sub-lambdas of 200, 100, and 500 Mbps, respectively, between routers 2-3 in an overlapped term. In this case, both NRMs calculated the total traffic volume of the three application flows, planned to use a 1-Gbps lambda for these flows, and allocated sub-lambdas with sufficient bandwidth to those flows. As shown in Fig. 4, the three traffic flows are properly reserved on the lambda between routers 2-3 via X03 (X03-lambda). When the scheduled time to set up the reserved sub-lambda for the MPI came prior to FTP and SHD, both NRMs established X03-lambda and the allocated sub-lambda for MPI. Subsequently, the allocated sub-lambdas for FTP and SHD were established and activated. Each flow was absolutely filtered via the correct packet filters based on its own source and destination IP addresses, controlled to guarantee the required throughput, and routed to the correct sub-lambda based on packet filtering and static routing. As anticipated, we successfully confirmed sufficient quality of experience for each application and the guaranteed bandwidth against interfering traffic with the help of the novel filtering scheme for the first time.



Reservation1. Job1, AM1-YM2, 200Mbps
Reservation2. Job2, AF1-YF2, 100Mbps
Reservation3. Job3, AS1-YS2, 500Mbps

Accommodate three reserved flows into one lambda path because total bandwidth is lower than the capacity of the path

Figure 4: Network resource reservation for multiple-applications based on requested bandwidth

Conclusions

We showed for the first time that according to each client request, the multi-layer lambda grid successfully allocated computers and a path using a sub-lambda and a lambda properly, routed a client traffic flow between the allocated computers to the path, and controlled the bandwidth of the path in an actual field environment.

Acknowledge

AIST is partially funded by the MEXT Science and Technology Promotion Program "Optical Paths Network Provisioning based on Grid Technologies."

References

1. A. Takefusa *et al.*, FGCS, Vol. 22 (2006), pp. 868-875.
2. M. Hayashi *et al.*, ECOC2006, We4.1.1.
3. Y. Tsukishima *et al.*, OECC/IOOC2007, 12A2-5.
4. S. R. Thorpe *et al.*, GridNets2007.
5. G. Zervas *et al.*, OFC2008, PDP34.
6. JGN 2, http://www.jgn.nict.go.jp/e/index.html
7. H. Uose *et al.*, Tridentcom 2005, pp. 232-241.
8. GridMPI, http://www.gridmpi.org/
9. D. Shirai *et al.*, APCCAS 2006, pp. 1855-1858.