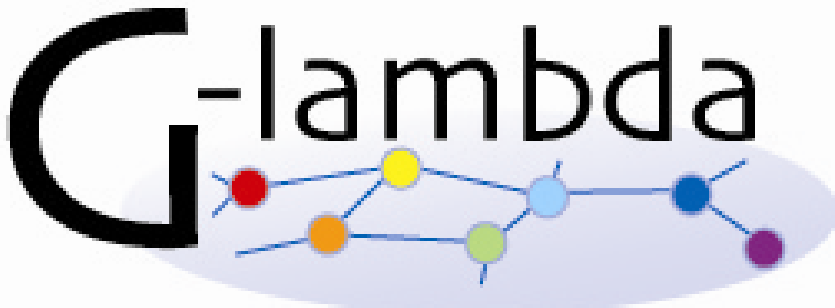
G-lambda and Enlightened Middleware and Control Plane interactions

Tomohiro Kudoh
Grid Technology Research Center
National Institute of Advanced Industrial Science and Technology (AIST)
and
Jon MacLaren
Center for Computation & Technology
Louisiana State University



LIVE DEMO

Inter-domain advance reservation of coordinated network and computing resources over the Pacific

A G-lambda & Enlightened collaboration

At 11th floor of THIS building
(AIST meeting room)

- Sep.11
  - 1:00PM-2:00PM
  - 6:00PM-
- Sep.12
  - 1:00PM-2:00PM
- Sep.13
  - 12:30PM-1:30PM

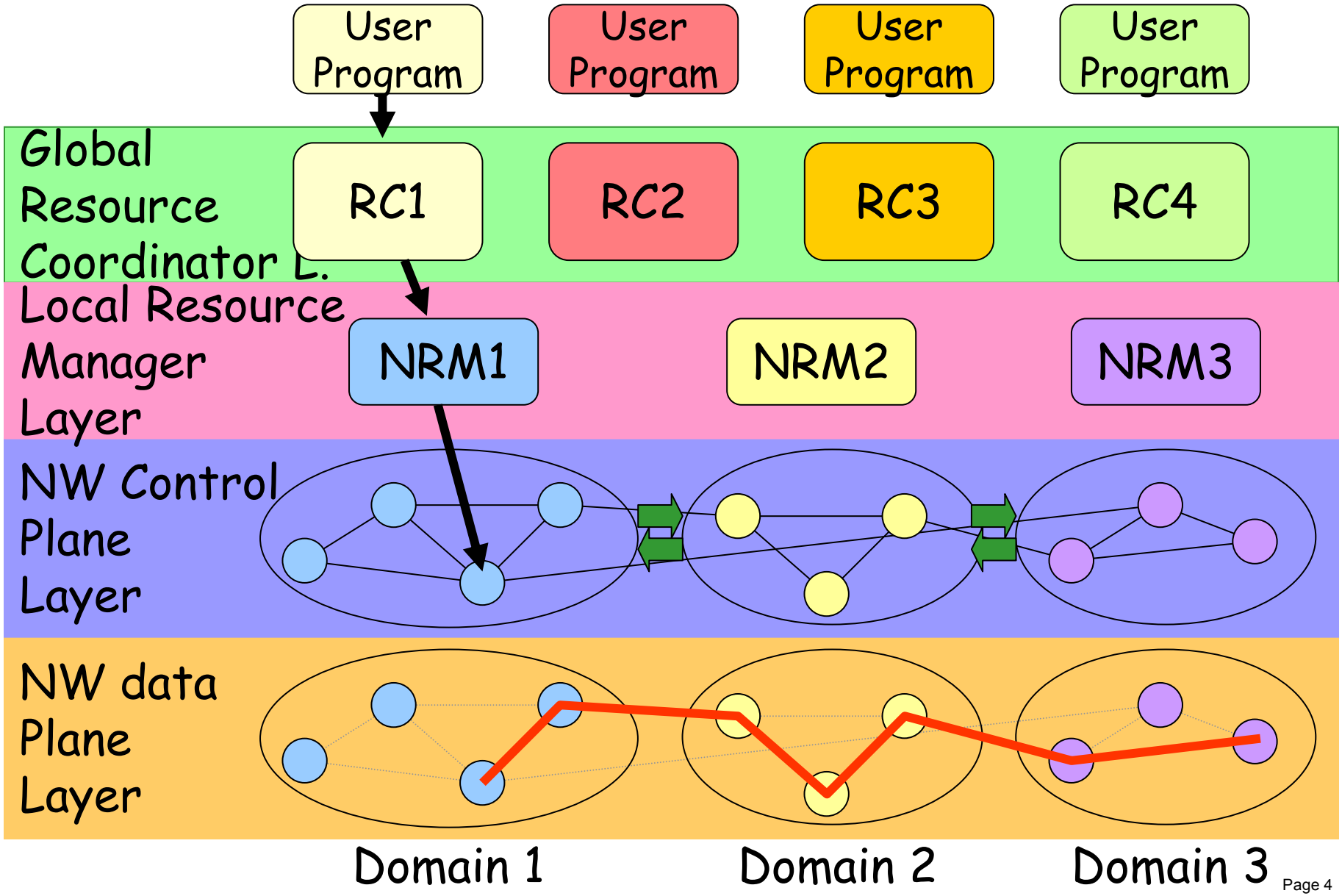
enλIGHTened computing

G-lambda

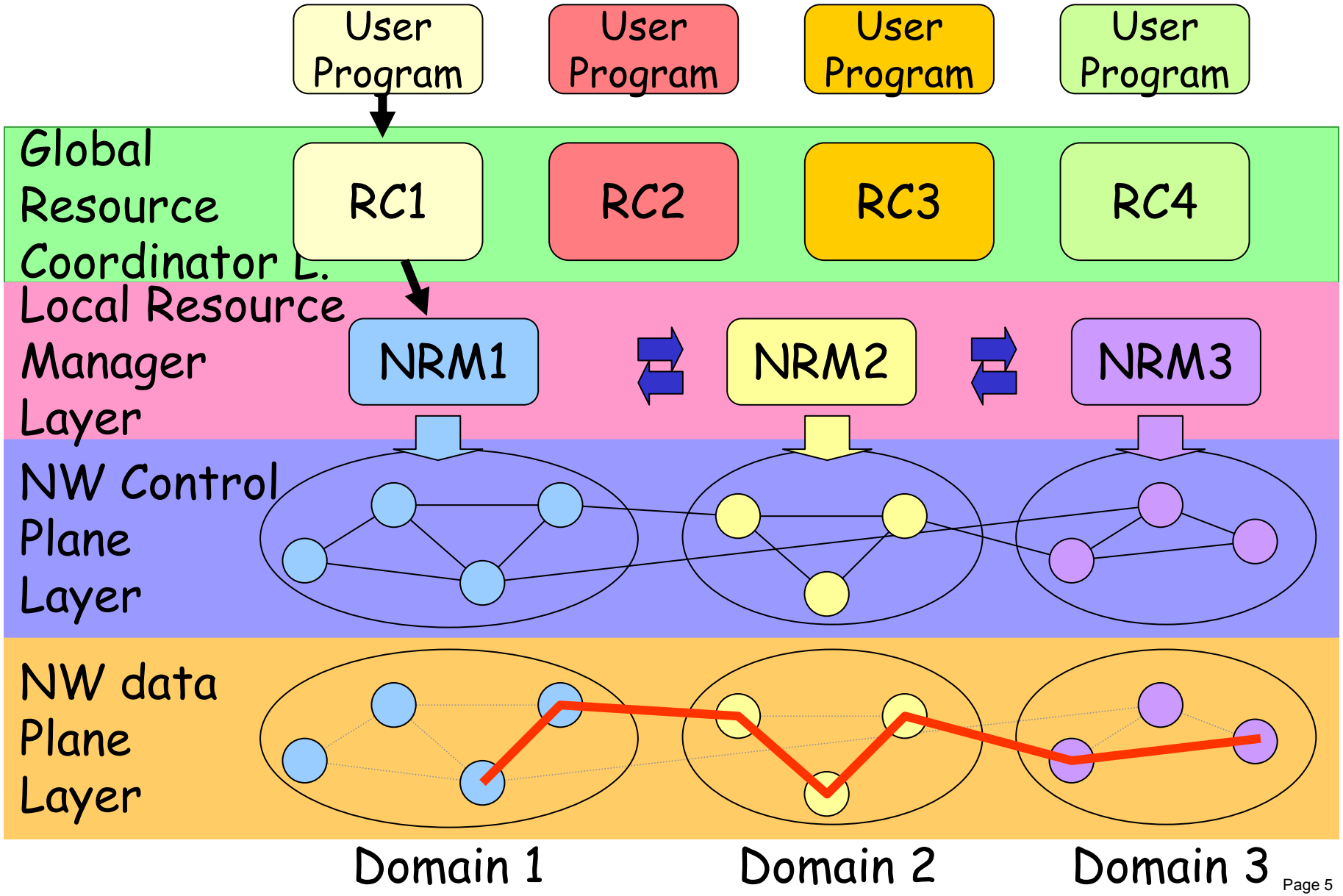
What we have done

- “Automated” interoperability between network and computing resources in two countries’ grid computing research testbeds is shown
  - the first such experiment of this scale between two countries
- Integrated computing and communication technology
  - Automated simultaneous in-advance reservation of network bandwidth between the US and Japan, and computing resources in the US and Japan
  - World’s first inter-domain coordination of resource mangers for in-advance reservation
    - Resource managers have different I/F and are independently developed

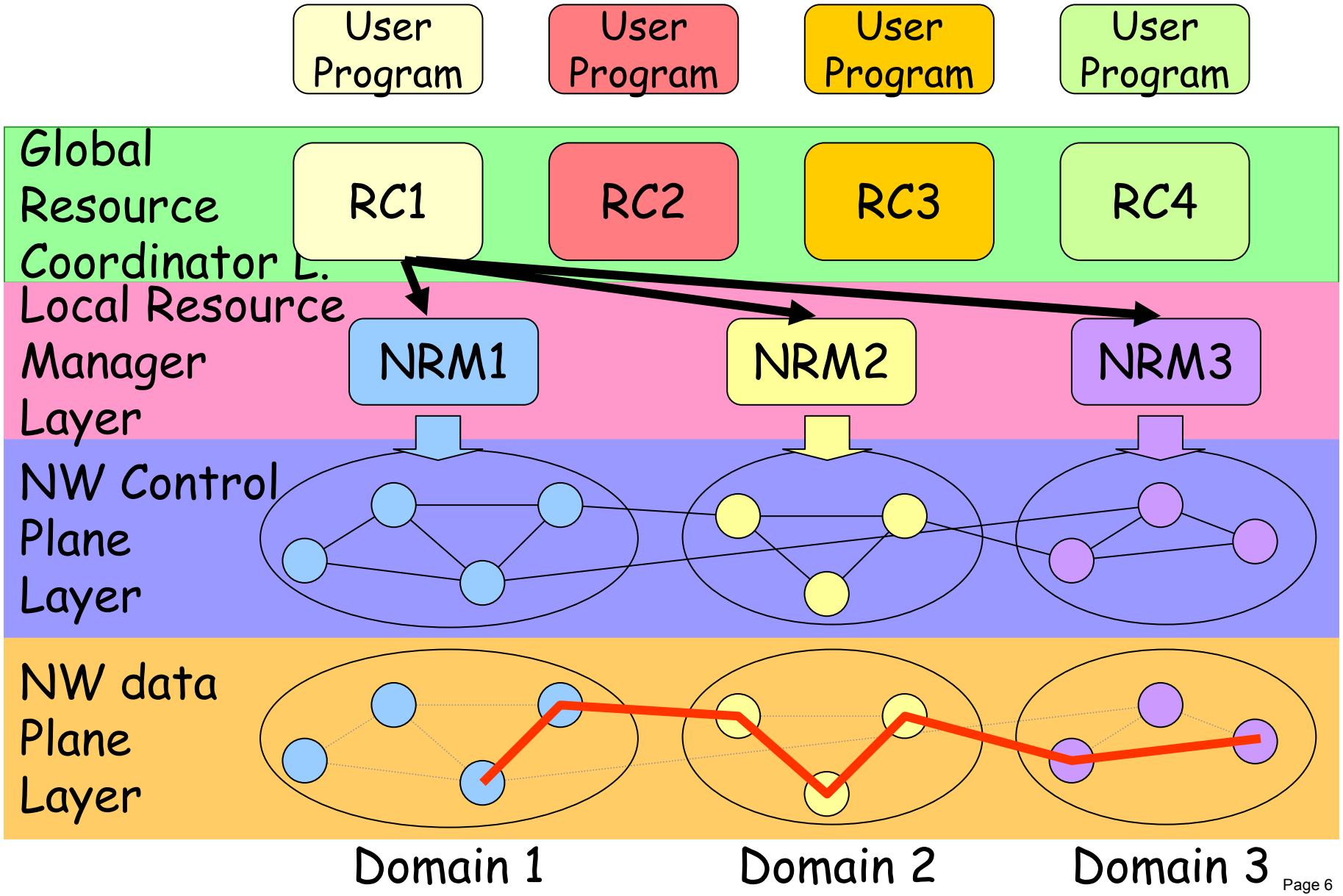
Three models of inter-domain coordination
(1) NW Control Plane Layer inter-working (ex. GMPLS E-NNI)



Three models of inter-domain coordination
(2) Local Resource Manager Layer inter-working



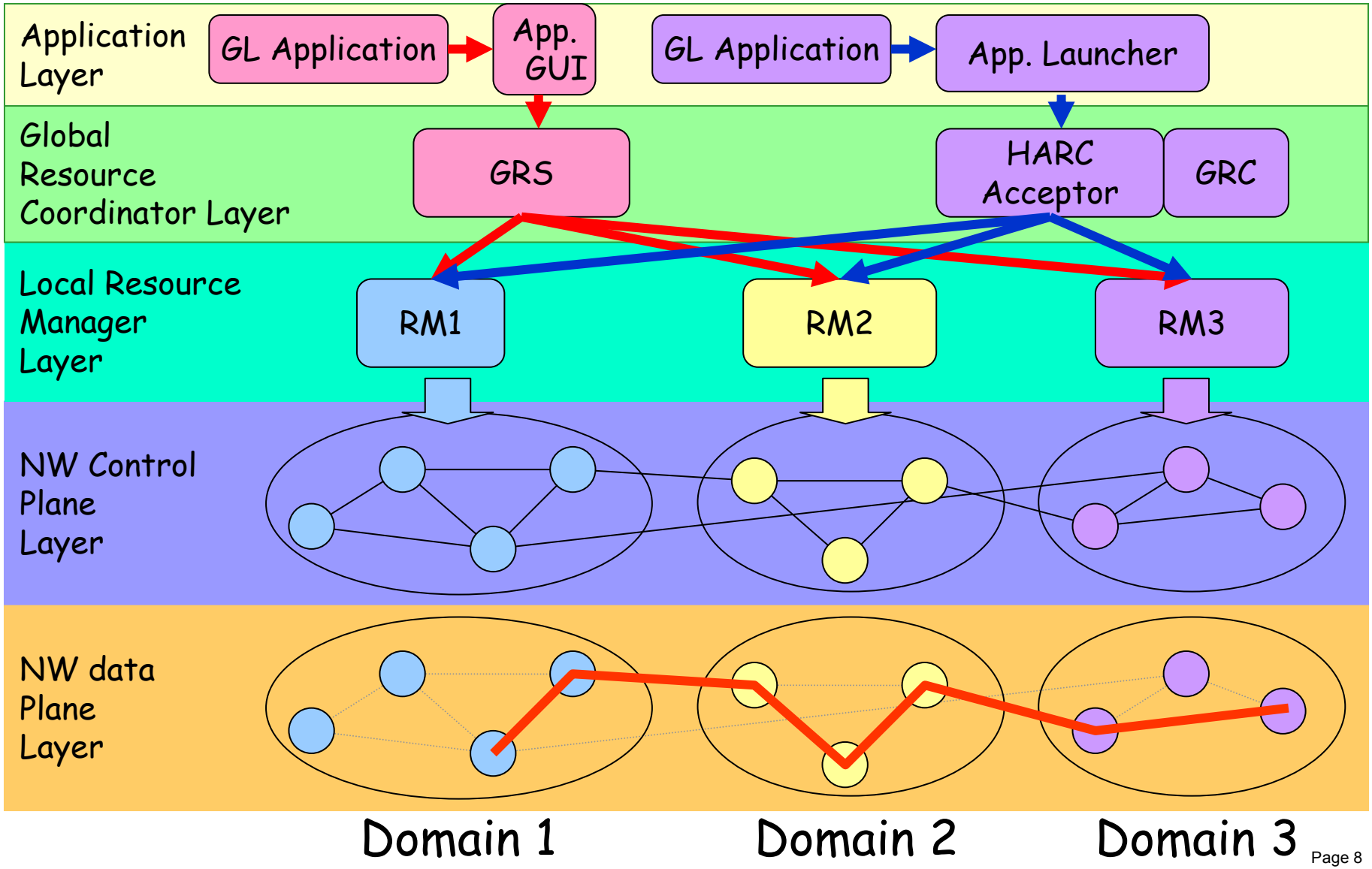
Three models of inter-domain coordination
(3) Global Resource Coordinator Layer inter-working

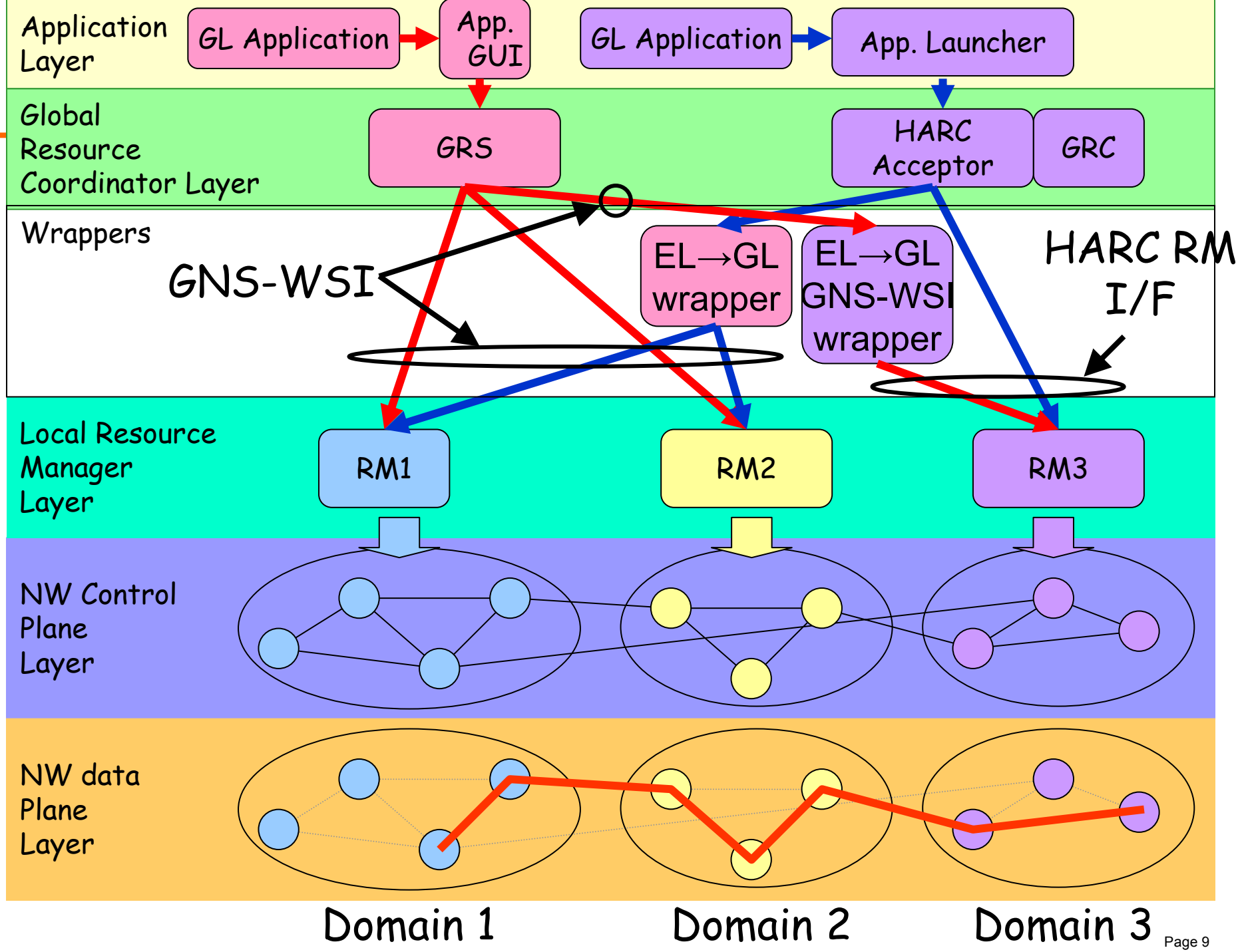


Pros and Cons of the three models

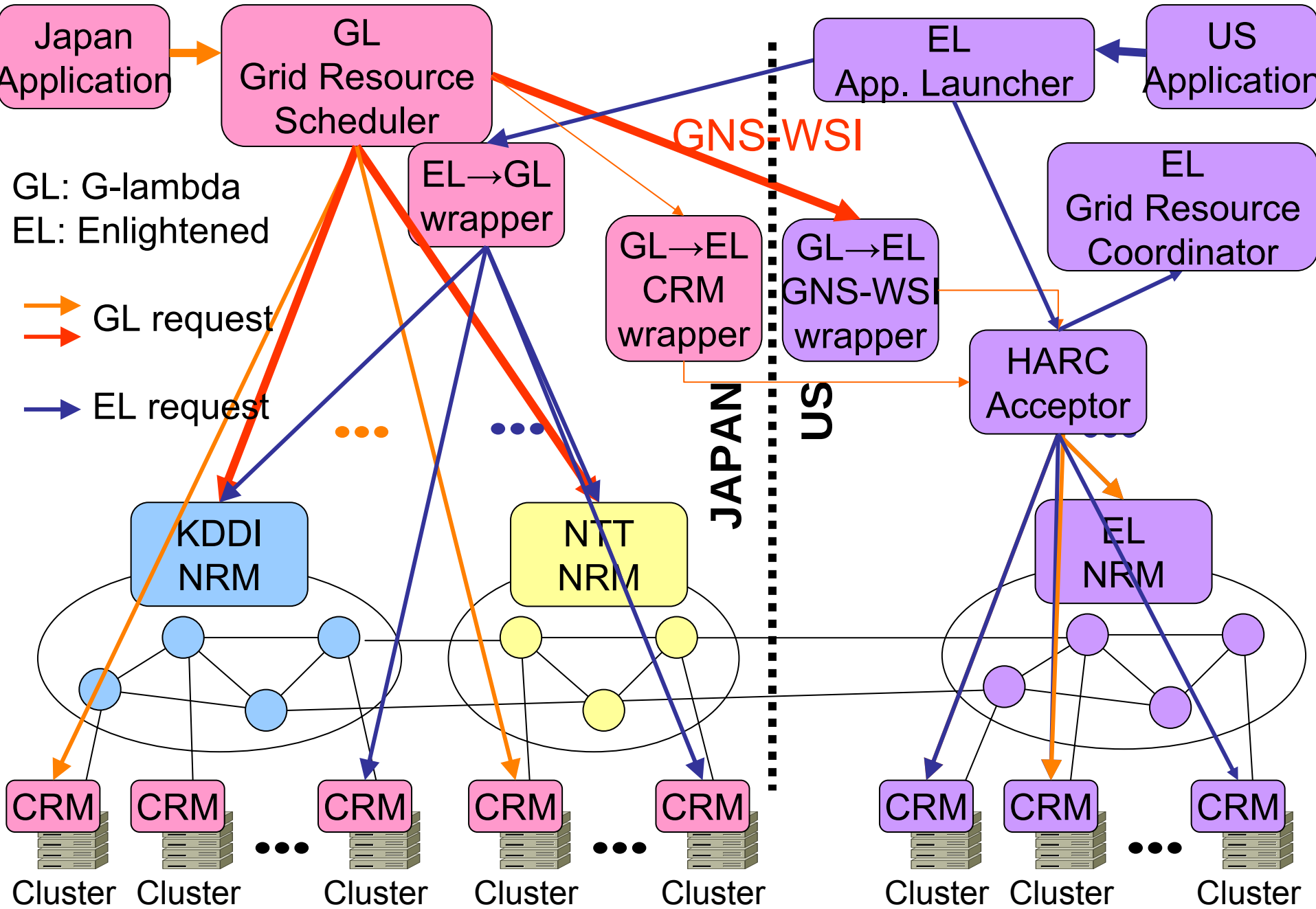
1. NW Control Plane Layer inter-working (ex. GMPLS E-NNI)
   - Pros: User do not have to care about “multiple domains”
   - Cons: GMPLS is an on-demand protocol and can not support advance reservation
   - Cons: Very close relationship between domains is required. May not be always possible for commercial service.
2. Resource Manager Layer inter-working
   - Pros: User do not have to care about “multiple domains”.
   - Cons: Requested NRM may make a reservation which is advantageous for the domain
3. Global Resource Coordinator Layer inter-working
   - Pros: User can control combination of domains
   - Pros: No under-layer interaction is required
   - Cons: User must have knowledge of inter-domain connection

WE EMPLOYED THIS MODEL FOR INTER-DOMAIN CONNECTION

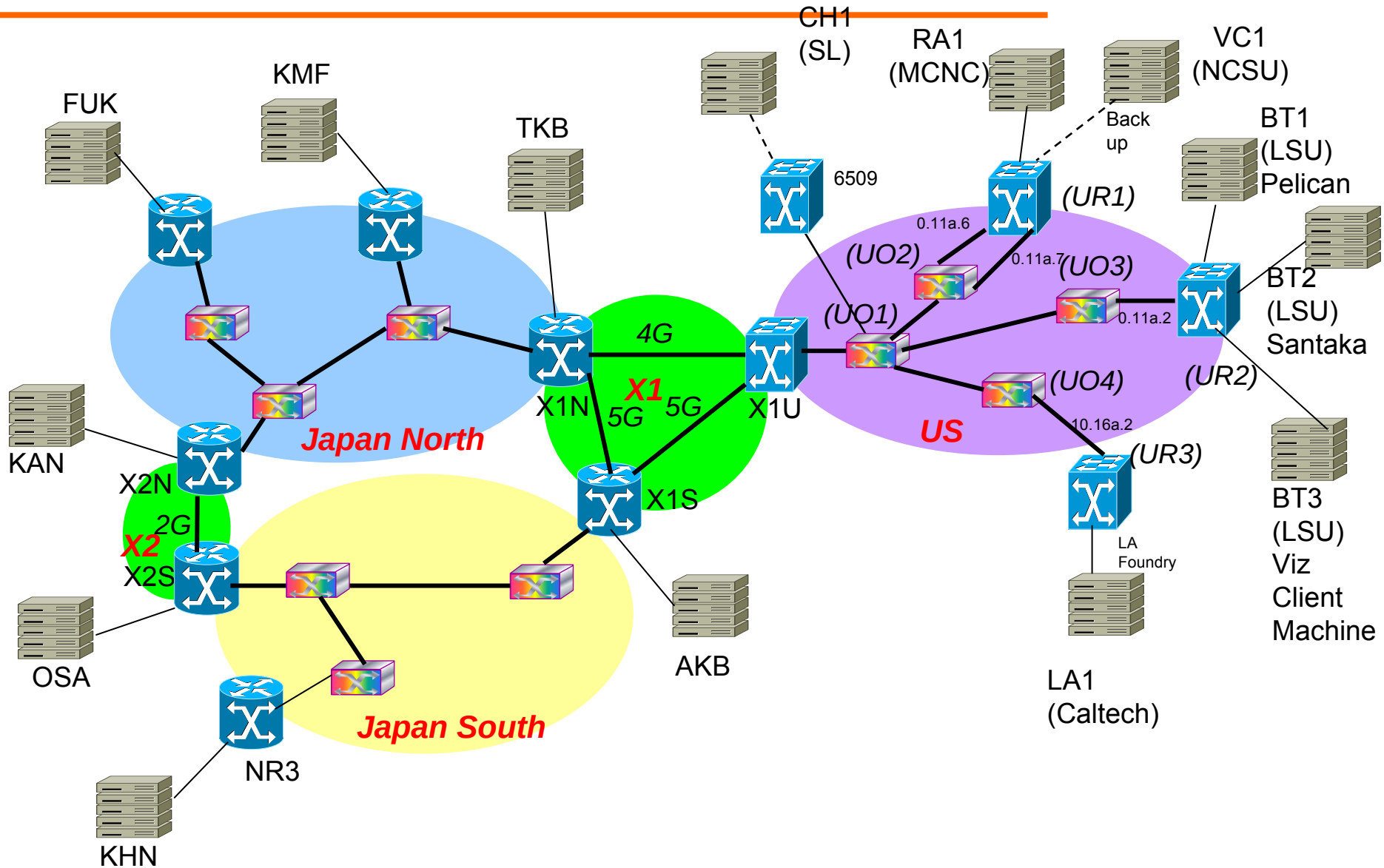

Application Layer
GL Application
App. GUI
GL Application
App. Launcher
Global Resource Coordinator Layer
GRS
HARC Acceptor
GRC
Local Resource Manager Layer
RM1
RM2
RM3
NW Control Plane Layer
NW data Plane Layer
Domain 1
Domain 2
Domain 3


Application Layer
GL Application
App. GUI
GL Application
App. Launcher
Global Resource Coordinator Layer
GRS
HARC Acceptor
GRC
Wrappers
GNS-WSI
EL→GL wrapper
EL→GL GNS-WSI wrapper
HARC RM I/F
Local Resource Manager Layer
RM1
RM2
RM3
NW Control Plane Layer
NW data Plane Layer
Domain 1
Domain 2
Domain 3

G-lambda/Enlightened middleware coordination diagram



Resource map of the demo



G- lambda project overview

- Joint project of KDDI R&D labs., NTT, NICT and AIST.
- G-lambda project has been started in December 2004.
- The goal of this project is to establish a standard web services interface (GNS-WSI) between Grid resource manager and network resource manager provided by network operators.

National Institute of Advanced Industrial Science and Technology
AIST

NTT

NiCT National Institute of Information and Communications Technology

JGNII

KDDI
KDDI R&D LABS

GNS-WSI (Grid Network Service / Web Services Interface)

- Web services interface to reserve bandwidth in advance
  - Network Resource Manager provides this service
- Polling-based operations
  - Advance reservation of a path between end points
  - Modification of reservation (i.e. reservation time or duration)
  - Query of reservation status
  - Cancellation of reservation
- GNS-WSI2
  - WSRF(Web Services Resource Framework) based interface
    - GT4 (Globus Toolkit 4) Java WS Core http://www.globus.org/toolkit/
  - 2-phase commit

GNS-WSI (Grid Network Service / Web Services Interface)

- Grid Network Service-Web Services Interface
- Interface to realize advance reservation of bandwidth
- Based on the Web Services interface technology
- Can be used for inter-domain coordination
- Polling-based operations
  - Advance reservation of a path between end points
  - Modification of reservation (i.e. reservation time or duration)
  - Query of reservation status
  - Cancellation of reservation
- GNS-WSI2
  - WSRF(Web Services Resource Framework) based interface
    - GT4 (Globus Toolkit 4) Java WS Core http://www.globus.org/toolkit/
  - 2-phase commit

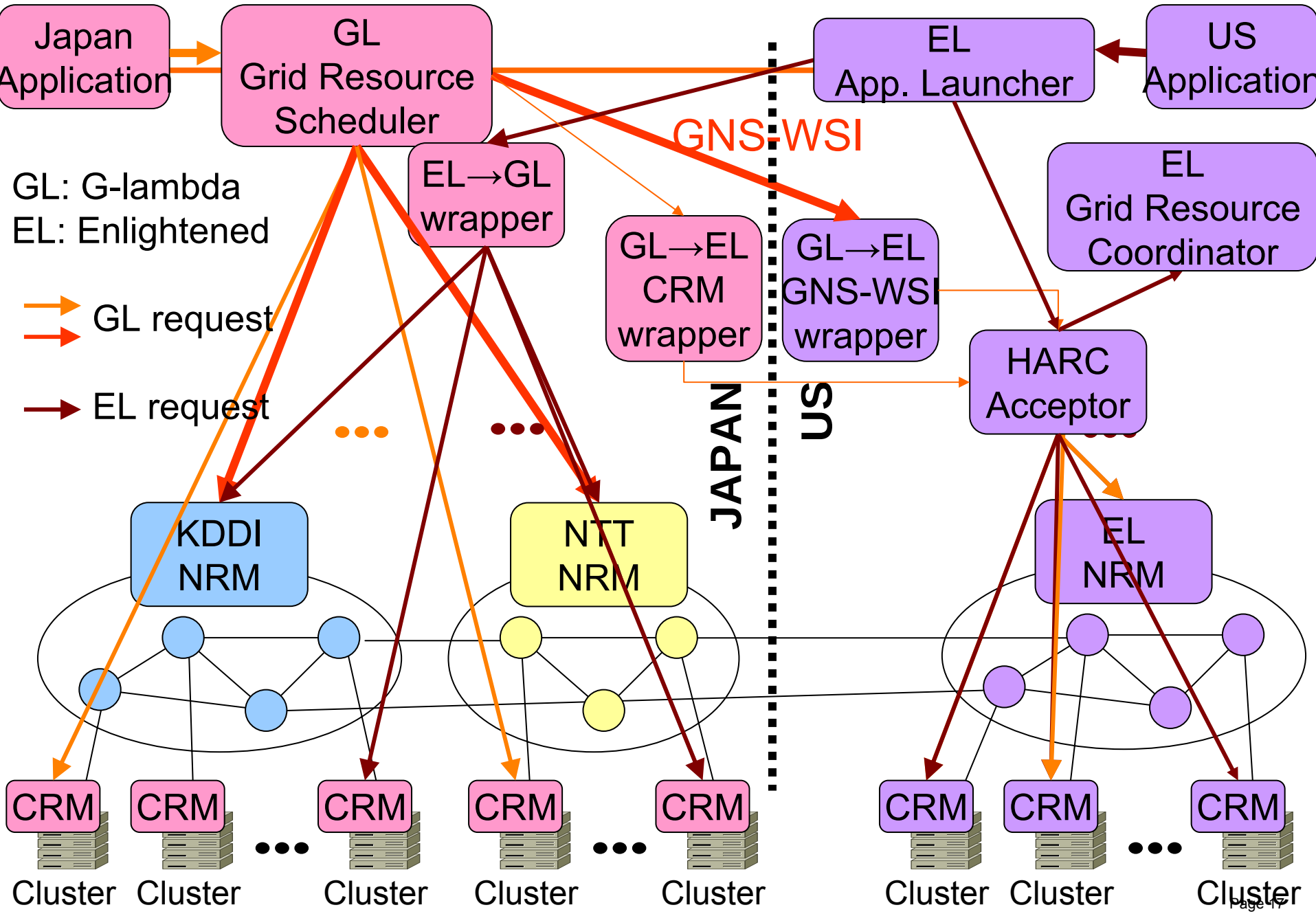
Service Parameters

Parameter	Usage	Value	Remarks
Site ID (APoint, ZPoint)	ID to specify A and Z points	String	Name or ID of sites
bandwidth	Bandwidth of the resource	Positive integer (kbit/s)	
latency	Latency between end points	Positive integer (msec)	
availability	Network protection of network resource	Integer (-2^{32} □ $2^{32}-1$)	0 = Un-protected 1 = Protected
Reservation time (startTime, endTime)	Start time and end time of the reservation	xsd:dateTime	YYYY-MM-DDTHH:MM:SSZ
localUsername	user name of certificate	String	GT4 GSI
reservationStatus	status of reservation	String	p. 15
commandStatus	status of each command	String	p. 16
resourceStatus	status of network resource	String	*Available* / *NotAvailable*

An example XML exchanged through GNS-WSI

```
<requirements>
    <network
      aPoint=“AKB"
      zPoint=“RA1"
      startTime="2006-09-07T04:15:00Z"
      endTime="2006-09-07T06:15:00Z“
      bandwidth="1000000"
      latency="1000"/>
</requirements>
```

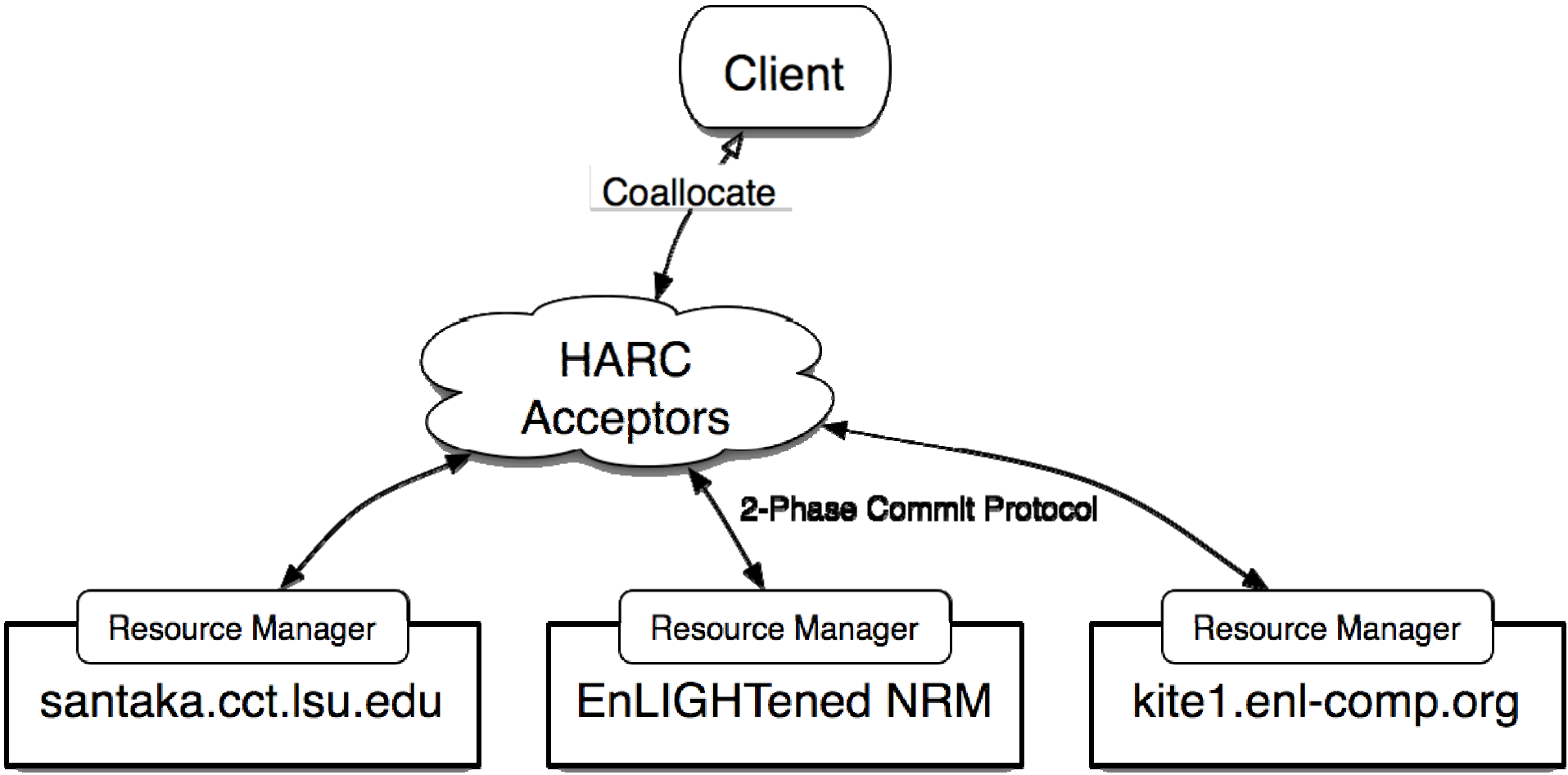
G-lambda/Enlightened middleware coordination diagram



QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

Coallocation of Compute and Network Resources using HARC

HARC – Highly Available Robust Coallocator



Robustness/Redundancy

- Based on **Paxos Commit** (Lamport/Gray)
- The Classic 2PC Transaction Manager functionality is replicated in multiple **acceptors**
  - Algorithm makes progress provided a majority of acceptors are working
  - So the RMs don't get stuck in "Prepared" state
  - Messages can be lost, repeated, arrive in an arbitrary order (but can't be tampered with)
  - If you deploy 7 acceptors, you can get a MTTF of about 5 years (assuming a MTTF of 48 hours, and MTTR of 1 hour per acceptor)

Coallocation

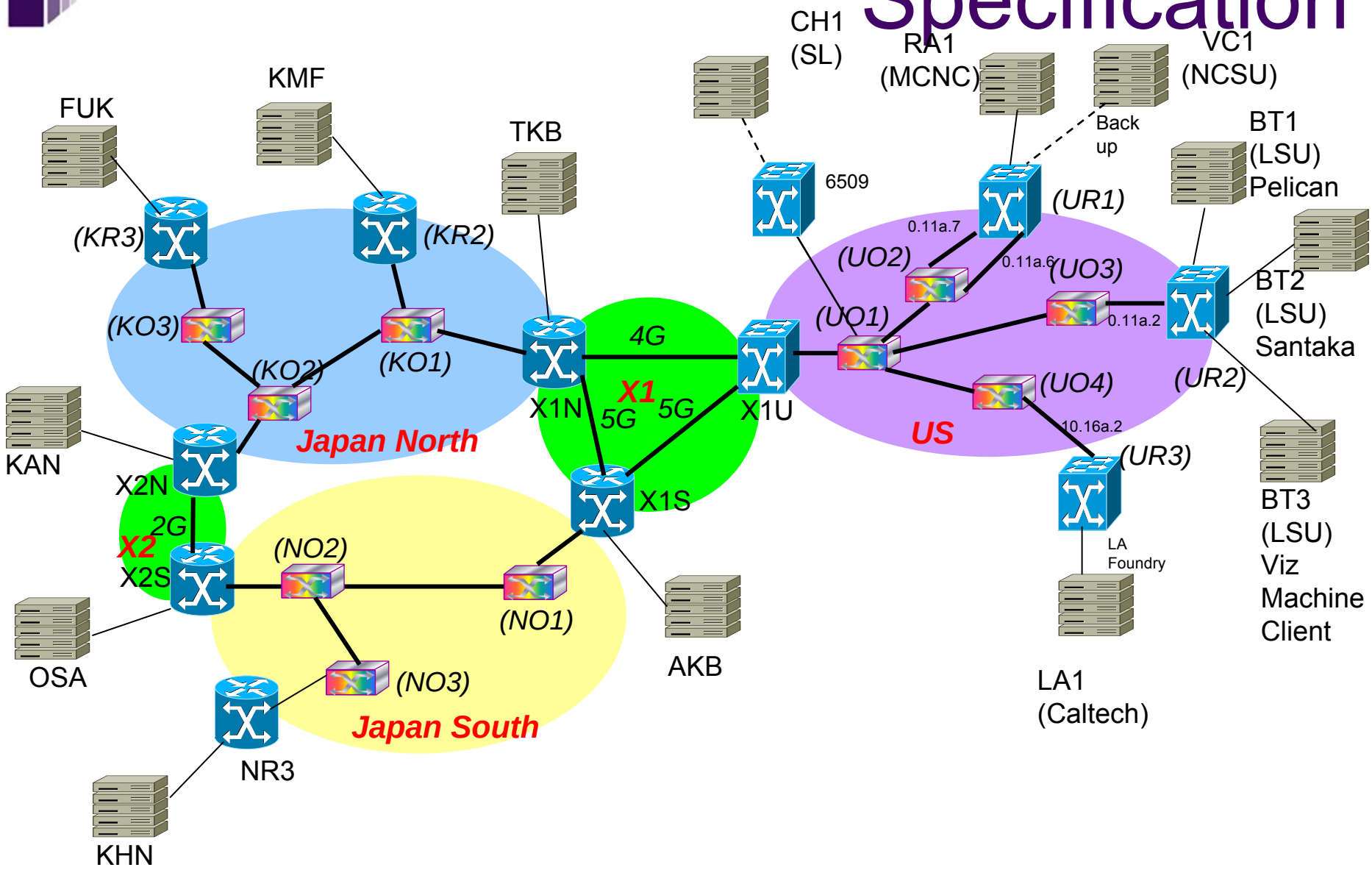
- Can support new types of RM
  - Can interface practically any reservation system
  - Without changing Acceptor code
  - Without changing/adding protocols
- Current RMs:
  - Compute resources (Batch queues, e.g. Torque/Moab, etc.)
  - Network resources (EnLIGHTened testbed) – the HARC NRM (includes simple scheduler)
- Future RMs:
  - Diary/Calendars (people/rooms)
  - VCL Cluster Reservation System

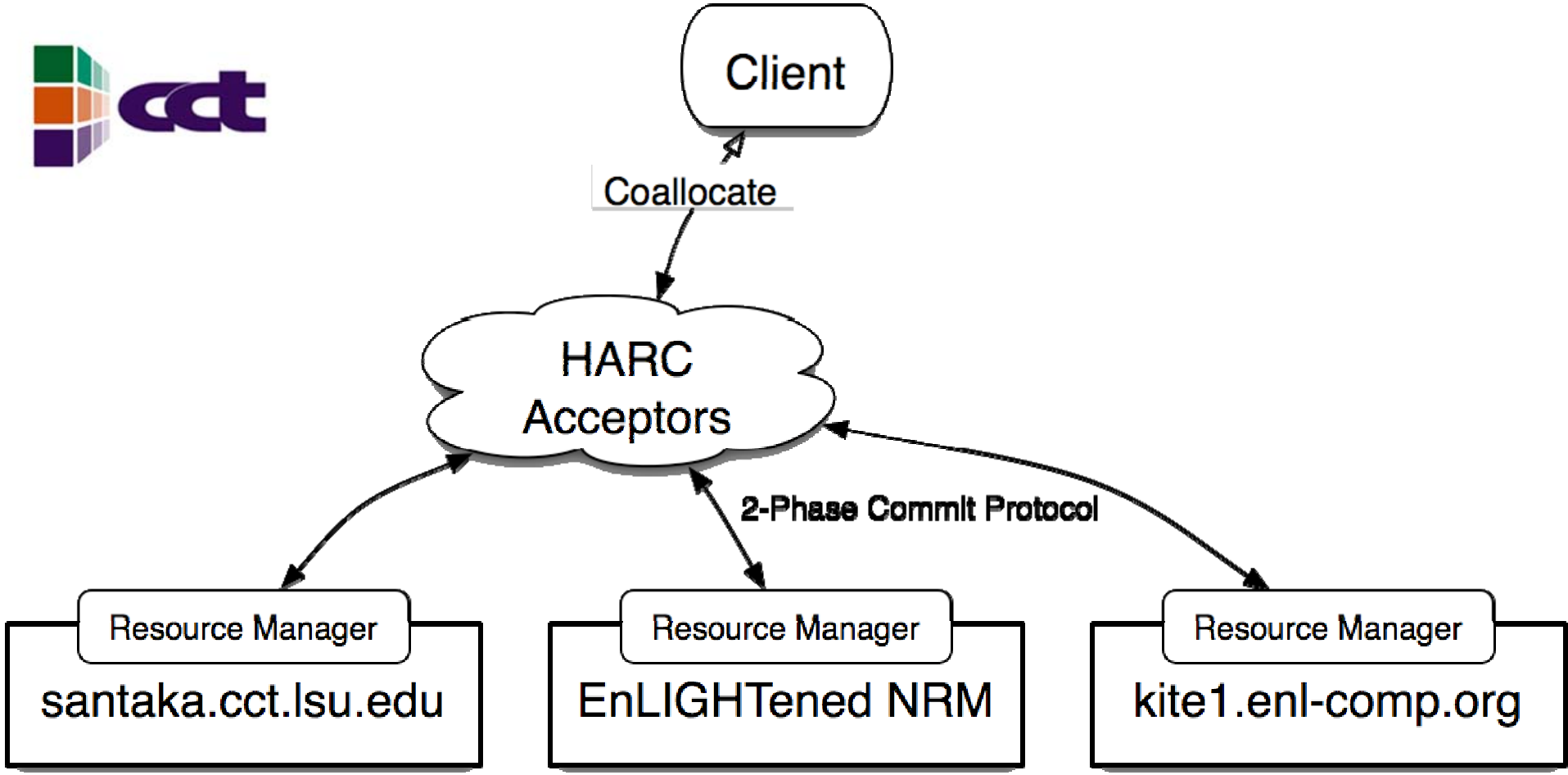
HARC NRM

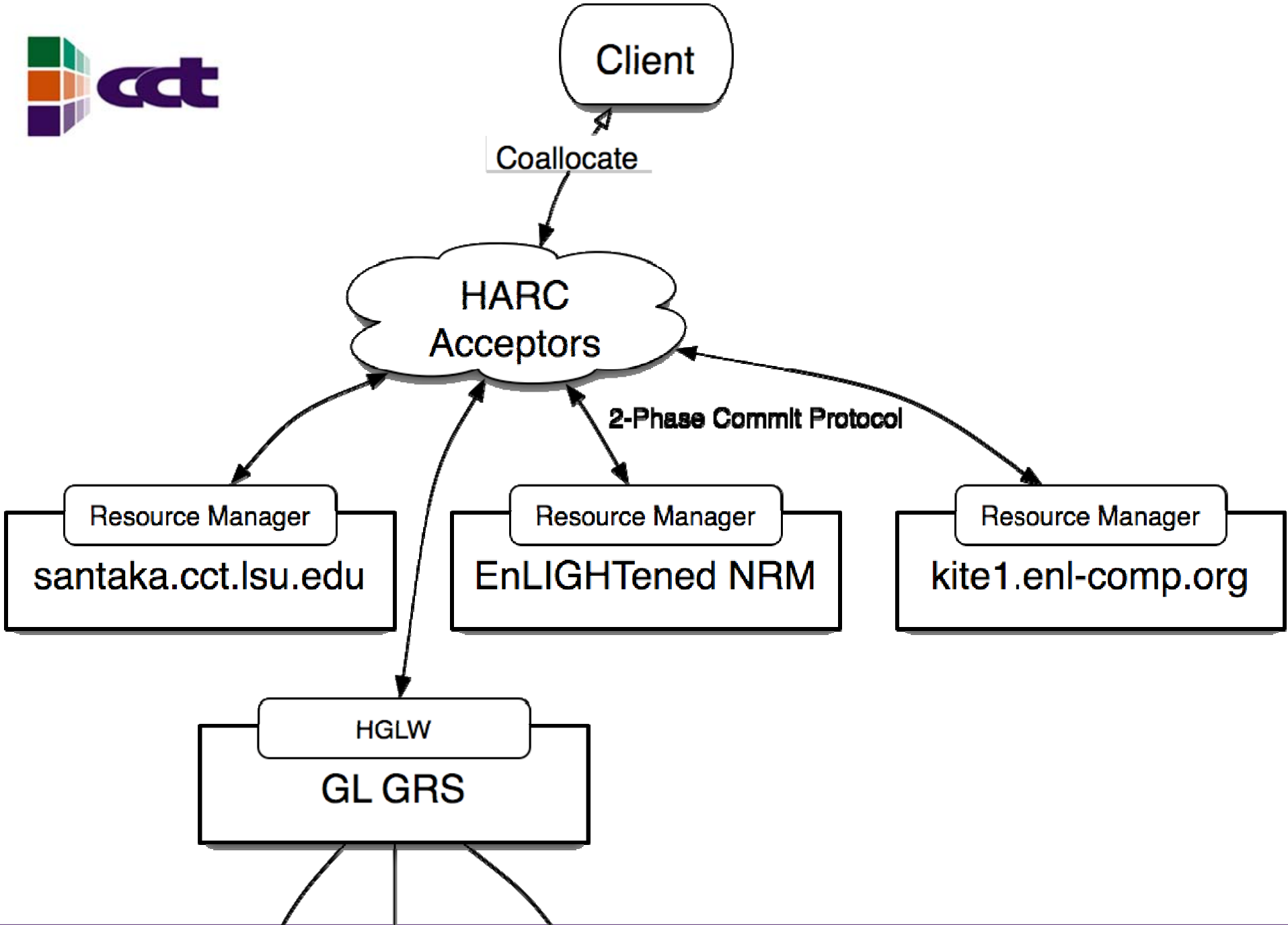
- Sends GMPLS Commands to the 4 Calient Diamondwave PXCs via TL1 commands in the EnLIGHTened testbed
- EROs are “computed” inside the NRM and sent to the switches as strict paths
- Contains a simple scheduler which maintains a centralized timetable of Trib/TE links
- No priority/preemption

- Replace lookup with simple computation for Supercomputing ‘06.
- Timetable should probably be distributed

Inter-domain Resource Specification

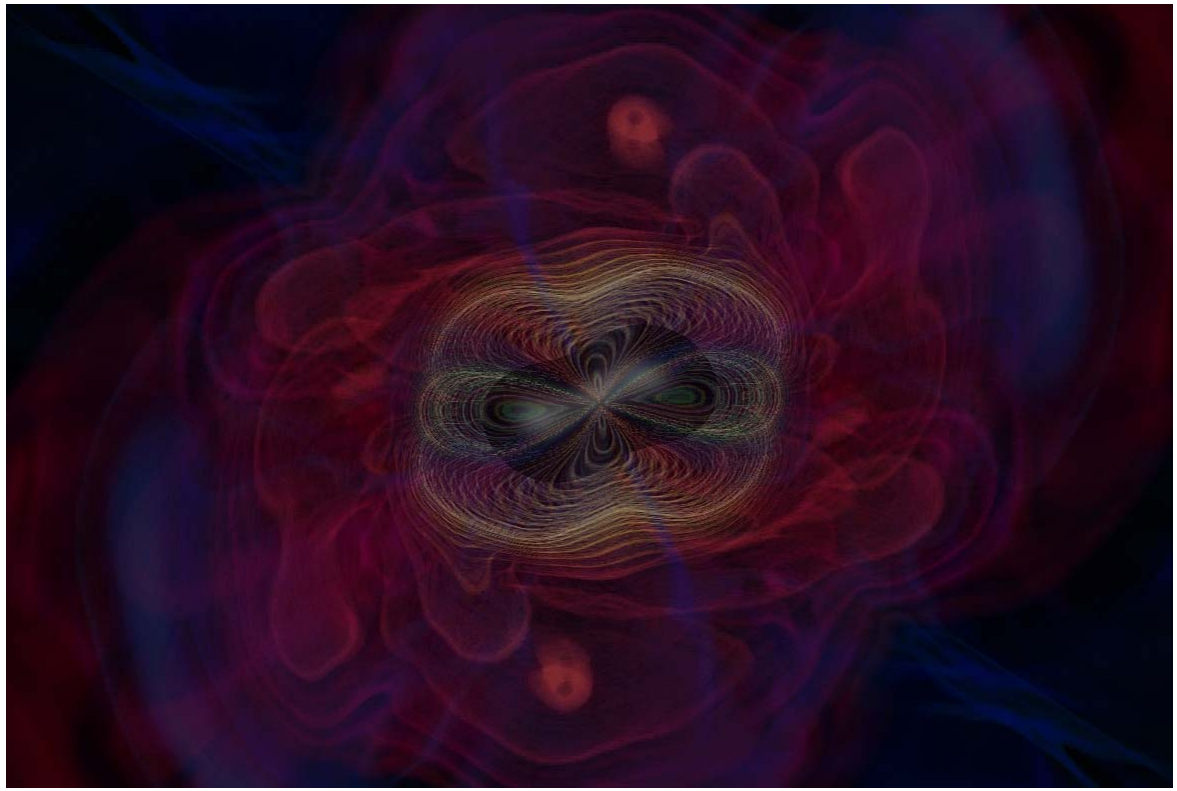



Client
Coallocate
HARC
Acceptors
2-Phase Commit Protocol
Resource Manager
santaka.cct.lsu.edu
Resource Manager
EnLIGHTened NRM
Resource Manager
kite1.enl-comp.org


Client
Coallocate
HARC
Acceptors
2-Phase Commit Protocol
Resource Manager
santaka.cct.lsu.edu
Resource Manager
EnLIGHTened NRM
Resource Manager
kite1.enl-comp.org
HGLW
GL GRS

Distributed Visualization

- Volume rendering using 3D textures of real component of Ψ_4 outgoing waves
- Optional Isosurfaces to show event horizon of merging black holes
- Positive values are blue while negative values appear reddish



LIVE DEMO

Inter-domain advance reservation of coordinated network and computing resources over the Pacific

A G-lambda & Enlightened collaboration

At 11th floor of THIS building
(AIST meeting room)

- Sep.11
  - 1:00PM-2:00PM
  - 6:00PM-
- Sep.12
  - 1:00PM-2:00PM
- Sep.13
  - 12:30PM-1:30PM

enλIGHTened computing

G-lambda