

Panel Discussion

Future Internet:
Challenges in Virtualization and Federation

Future Internet

Future Internet

- Revolutionary reinvention of the network

Future Internet

- Revolutionary reinvention of the network
- To replace the ossified Internet

Future Internet

- Revolutionary reinvention of the network
- To replace the ossified Internet
- Clean-slate approach

Future Internet

- Revolutionary reinvention of the network
- To replace the ossified Internet
- Clean-slate approach
- Testbed for concurrent wild ideas

Future Internet

- Revolutionary reinvention of the network
- To replace the ossified Internet
- Clean-slate approach
- Testbed for concurrent wild ideas
 - with global scale

Future Internet

- Revolutionary reinvention of the network
- To replace the ossified Internet
- Clean-slate approach
- Testbed for concurrent wild ideas
 - with global scale
- Virtualization and Federation

Virtualization

Virtualization

- Resources

Virtualization

- Resources
 - Network, Storage, Computing, ...

Virtualization

- Resources
 - Network, Storage, Computing, ...
- Concurrent trial of competing ideas

Virtualization

- Resources
 - Network, Storage, Computing, ...
- Concurrent trial of competing ideas
- Coexistence with the Internet

Virtualization

- Resources
 - Network, Storage, Computing, ...
- Concurrent trial of competing ideas
- Coexistence with the Internet
- Possibly, inherent meta-architecture for FI

Virtualization

- Resources
 - Network, Storage, Computing, ...
- Concurrent trial of competing ideas
- Coexistence with the Internet
- Possibly, inherent meta-architecture for FI
- Challenges

Federation

Federation

- Of multiple virtualization platforms

Federation

- Of multiple virtualization platforms
- Federated usage of virtualized resources

Federation

- Of multiple virtualization platforms
- Federated usage of virtualized resources
 - across different domains

Federation

- Of multiple virtualization platforms
- Federated usage of virtualized resources
 - across different domains
- Even greater challenges

Federation

- Of multiple virtualization platforms
- Federated usage of virtualized resources
 - across different domains
- Even greater challenges

Federation

- Of multiple virtualization platforms
- Federated usage of virtualized resources
 - across different domains
- Even greater challenges
- Issues, Visions, Strategies, Experiences

Panelists

Panelists

- Laurent Mathy: Lancaster University / UK

Panelists

- Laurent Mathy: Lancaster University / UK
- Mauro Campanella: GARR / Italy

Panelists

- Laurent Mathy: Lancaster University / UK
- Mauro Campanella: GARR / Italy
- Rick Summerhill: Internet2 / USA

Panelists

- Laurent Mathy: Lancaster University / UK
- Mauro Campanella: GARR / Italy
- Rick Summerhill: Internet2 / USA
- James Williams: Indiana Univ. / USA

Panelists

- Laurent Mathy: Lancaster University / UK
- Mauro Campanella: GARR / Italy
- Rick Summerhill: Internet2 / USA
- James Williams: Indiana Univ. / USA
- Shinji Shimojo: NICT / Japan

Panelists

- Laurent Mathy: Lancaster University / UK
- Mauro Campanella: GARR / Italy
- Rick Summerhill: Internet2 / USA
- James Williams: Indiana Univ. / USA
- Shinji Shimojo: NICT / Japan
- DaeYoung Kim: Chungnam Univ. / Korea

Platform for high-performance network virtualization

Future Internet: Challenges in Virtualization and Federation

Laurent Mathy

laurent@comp.lancs.ac.uk

In collaboration with

N. Egi, M. Hoerd (Lancs)

A. Greenhalgh, M. Handley (UCL)

F. Huici (NEC Europe)

- **Traditionally, high performance networking devices:**
 - built as custom hardware
 - Lengthy and expensive process
 - Made of various specialized “CPUs”
 - Either “real” CPUs or specialized silicon functions (ICs)
 - NP
 - Specialized multi-cores
 - **Software routers several orders of magnitude slower**
 - **BUT recent advances in commodity HW architecture**
 - Multi/many core (with very high clock frequency)
 - Buses (usual bottlenecks) are disappearing
- ⇒ How viable are SW routers in this new context?**

Lowdown on SW routers

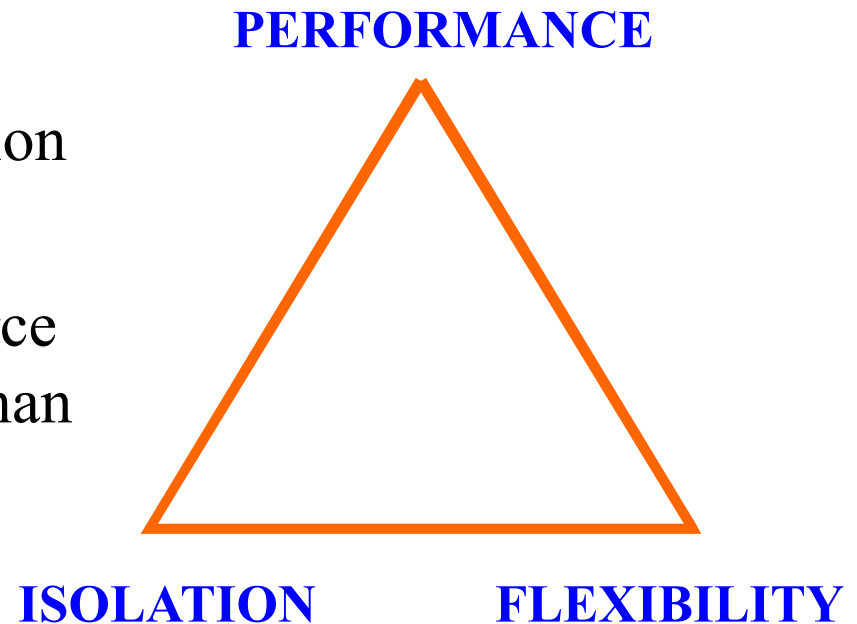
The bottleneck is memory access latency

- **Plenty of spare CPU cycles**
 - Surplus scales with number of cores
 - Performance scales with number of memory controller
 - NUMA
- ⇒ **What about running several (virtual) routers on a single box?**

Virtual Routers

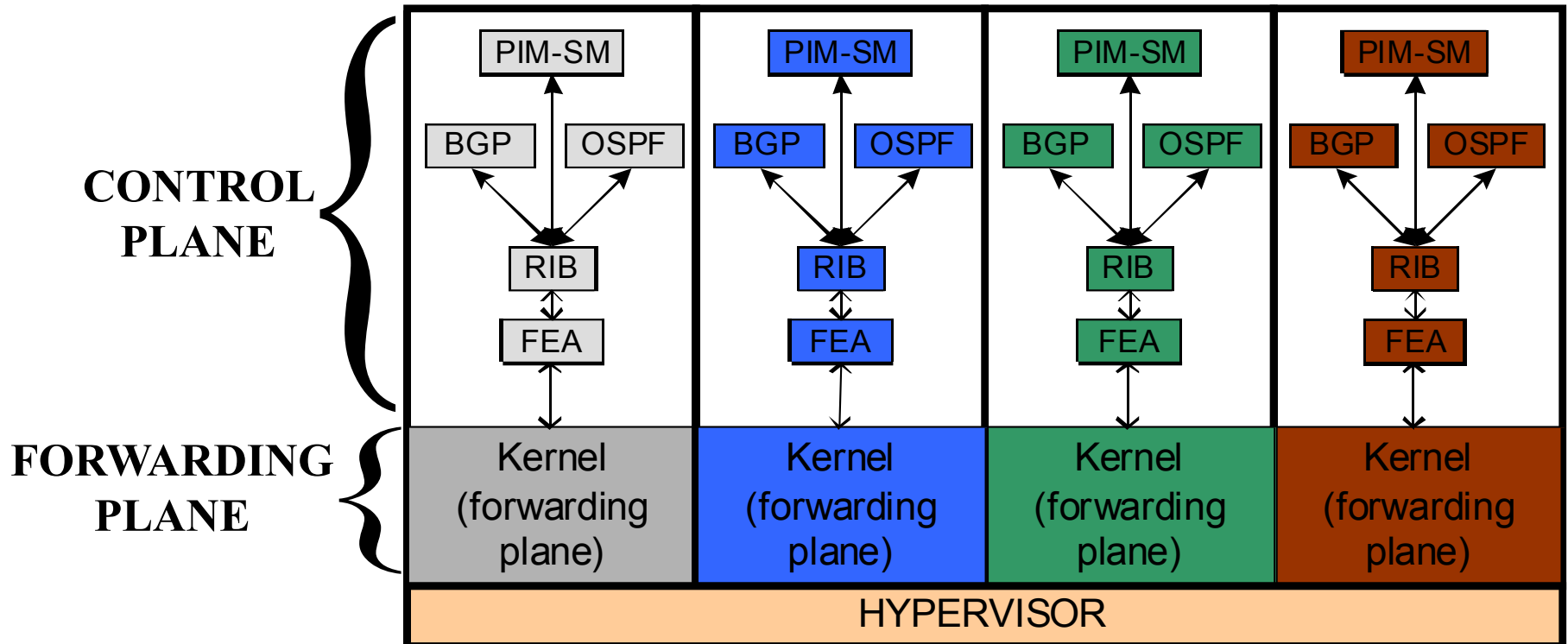
Challenges:

- Isolating the VRs from each other
- Low overhead of the virtualization
- Fair scheduling, despite the scarce resource being memory rather than CPU cycles
- Mapping cores to VR in a performant manner



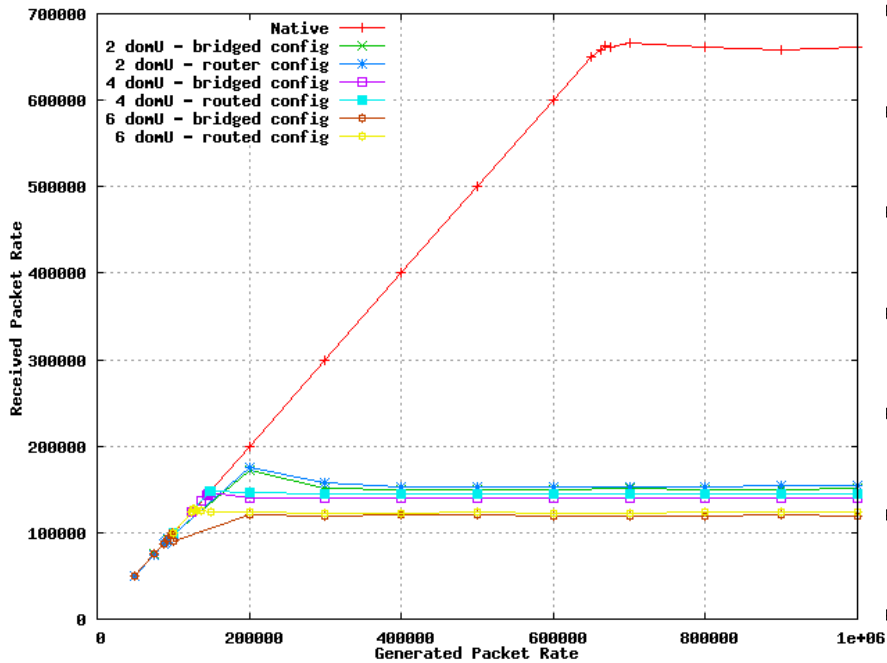
Virtual Routers: naïve implementation

Can conventional techniques for server virtualization be applied to network router virtualization?

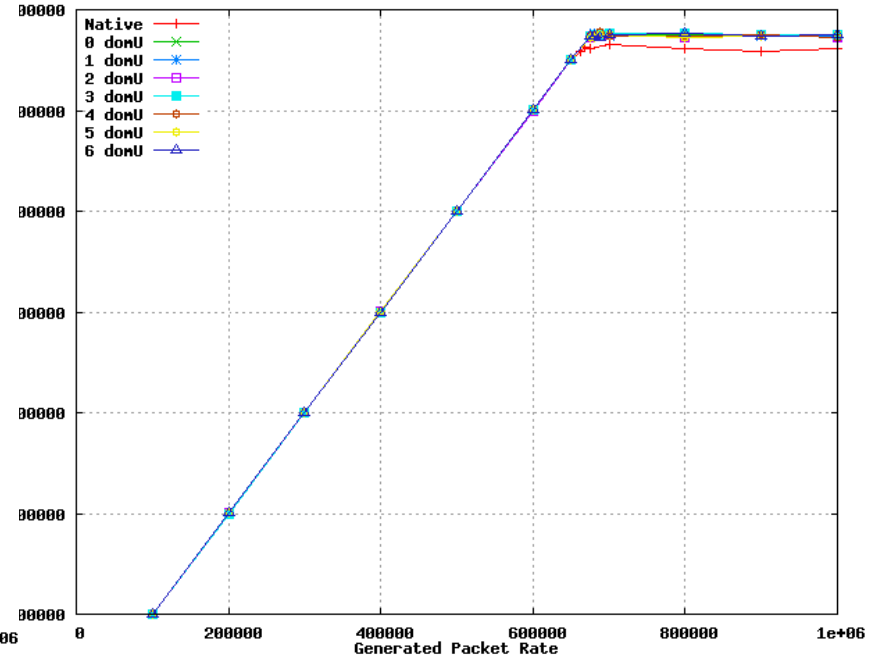


Virtual Routers: naïve implementation performance

Can conventional techniques for server virtualization be applied to network router virtualization? **Not really.**



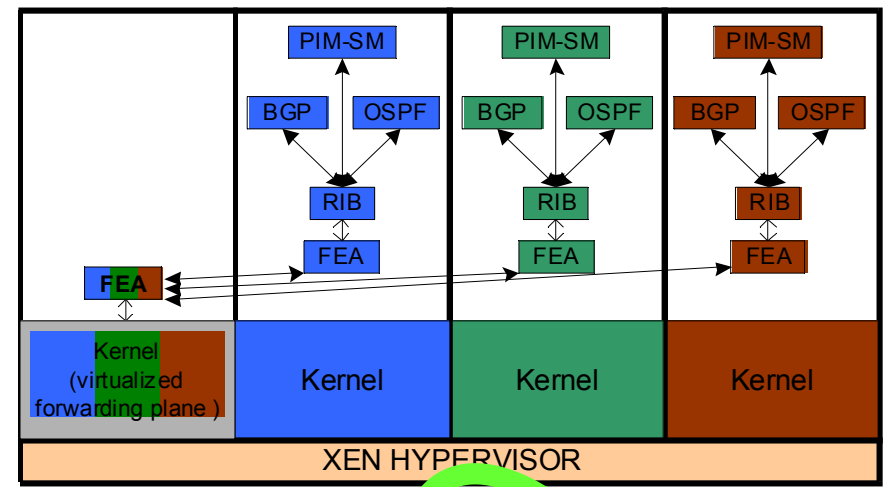
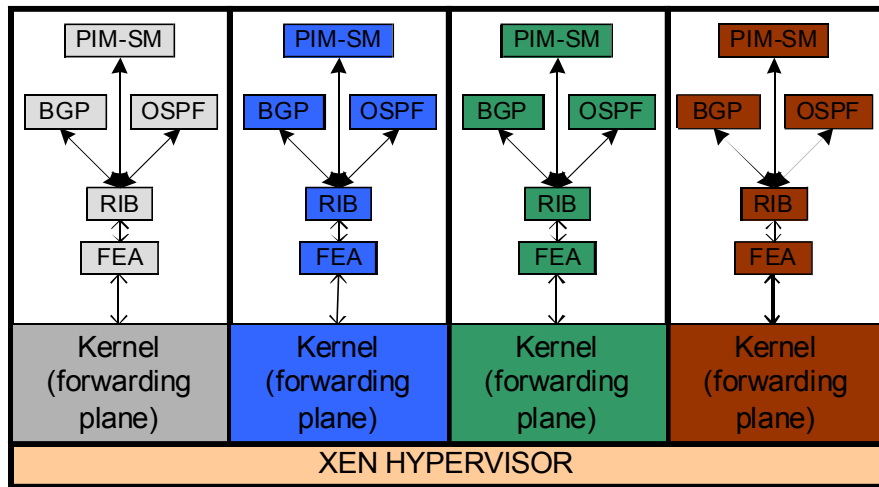
Forwarding in the DomUs



Forwarding in Dom0 only

Virtual Routers: architectural principle

What to do then?



PERFORMANCE

PERFORMANCE

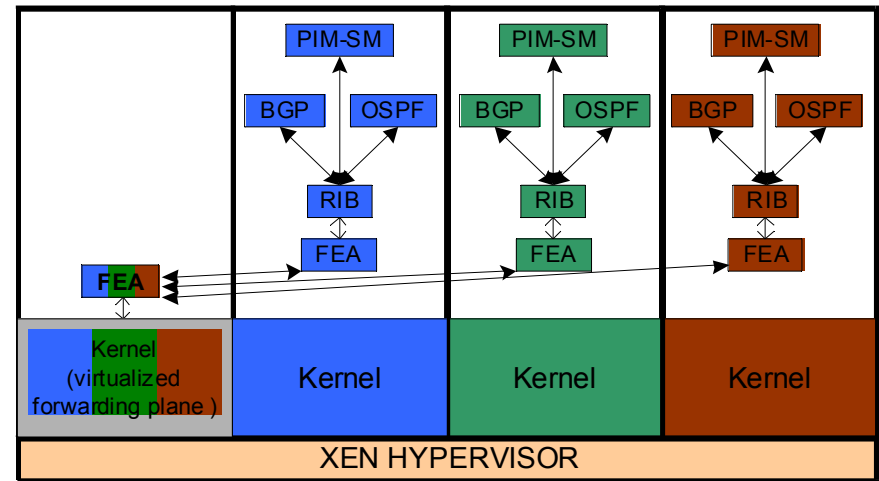
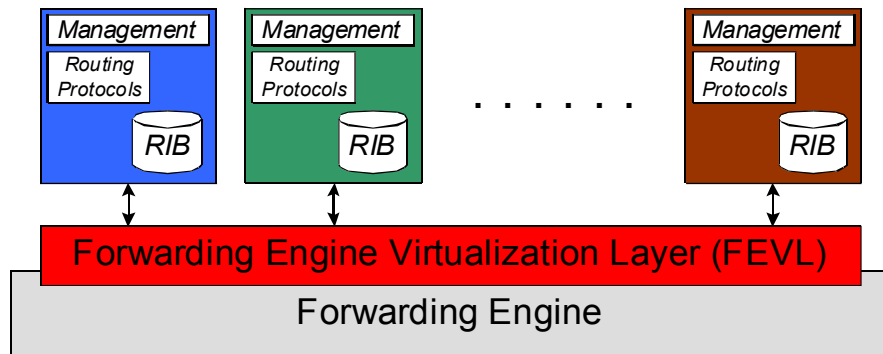
ISOLATION

FLEXIBILITY

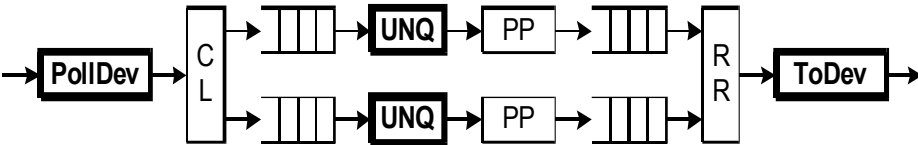
ISOLATION

FLEXIBILITY

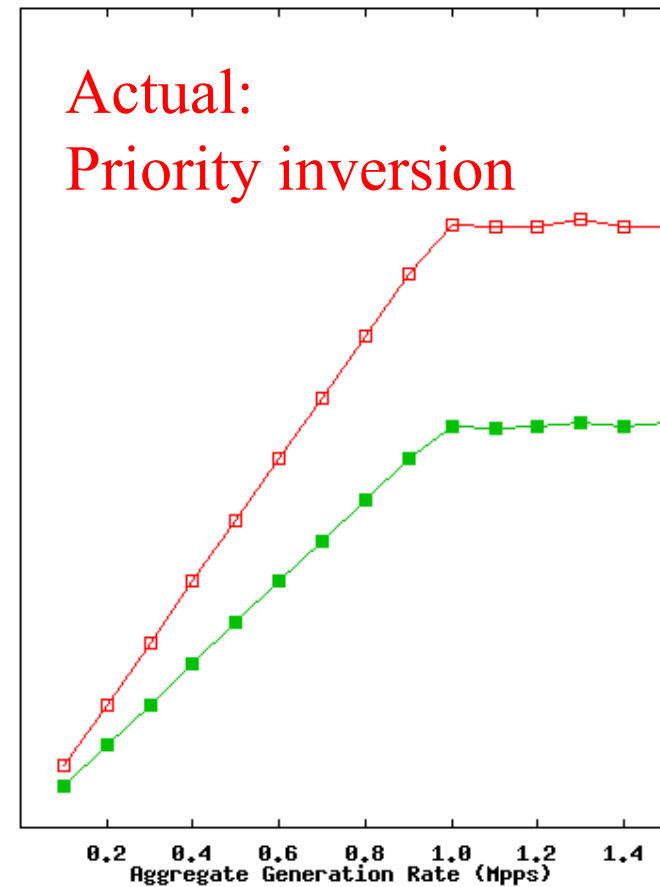
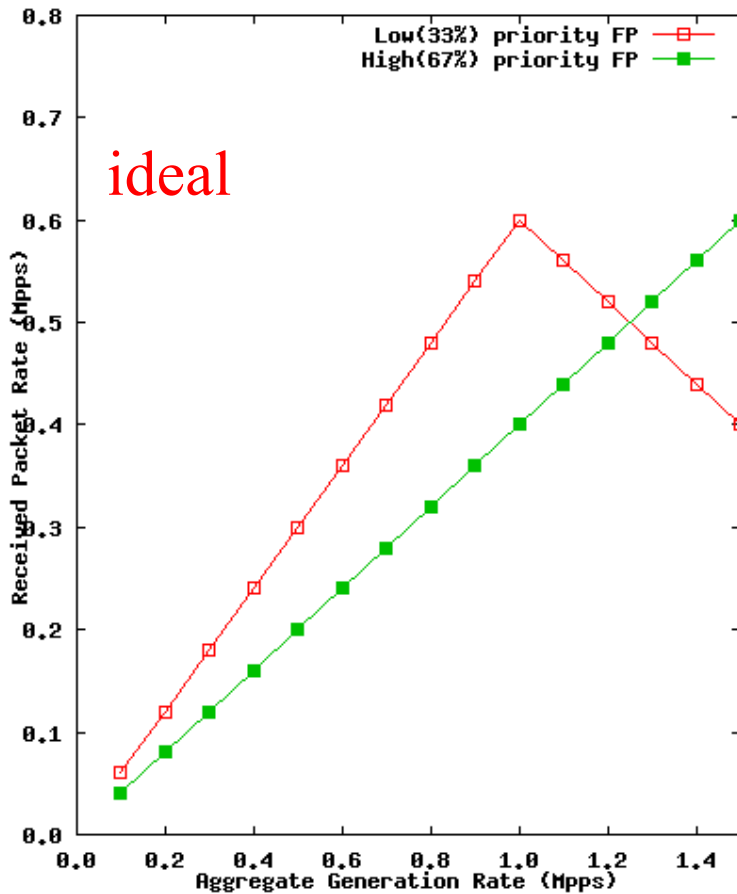
Virtual Routers is forwarding engine virtualization



Sharing NICs: software demux



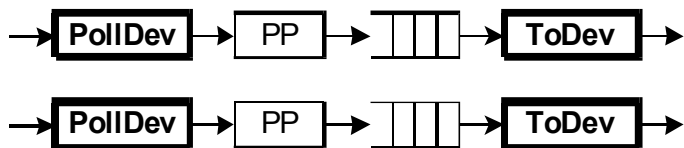
- 2:1 prioritization
- 60% traffic is low priority



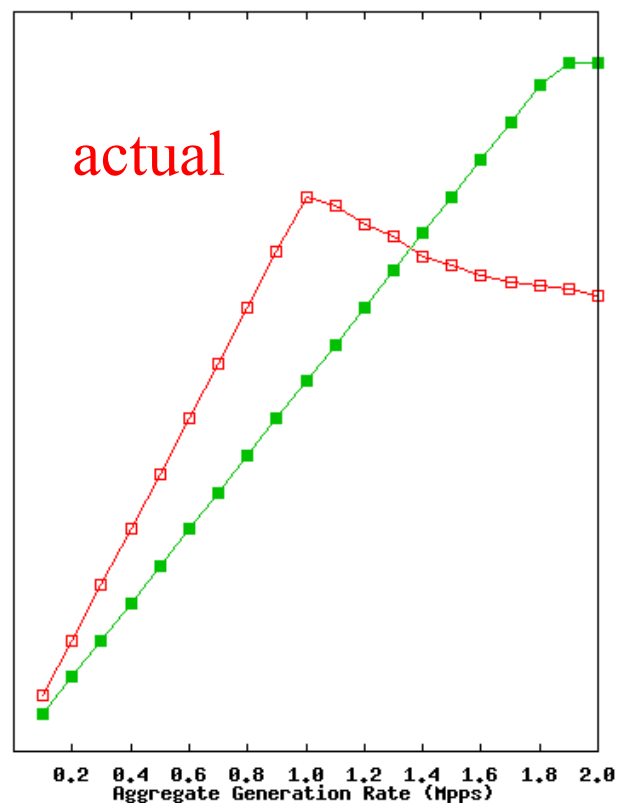
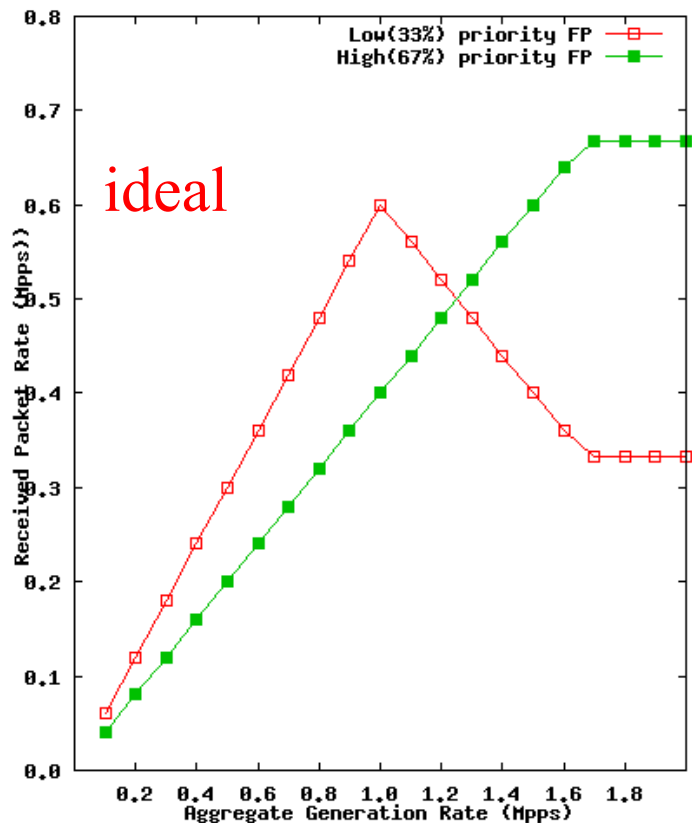
Sharing NICs: software demux (2)

- **Scheduling dictated by arrival rate**
- **Once polled, i.e. copied in memory, a packet has passed the bottleneck**
- **We could re-queue and do weighted fair queueing, but...**
- **Dropping excess packets is very wasteful as main cost is polling in the first place**
 - Dropping low priority packets is NOT going to improve aggregate forwarding performance much
- **What we need is hardware demux**
 - VMDQ
 - RSS

Sharing NICs: hardware demux

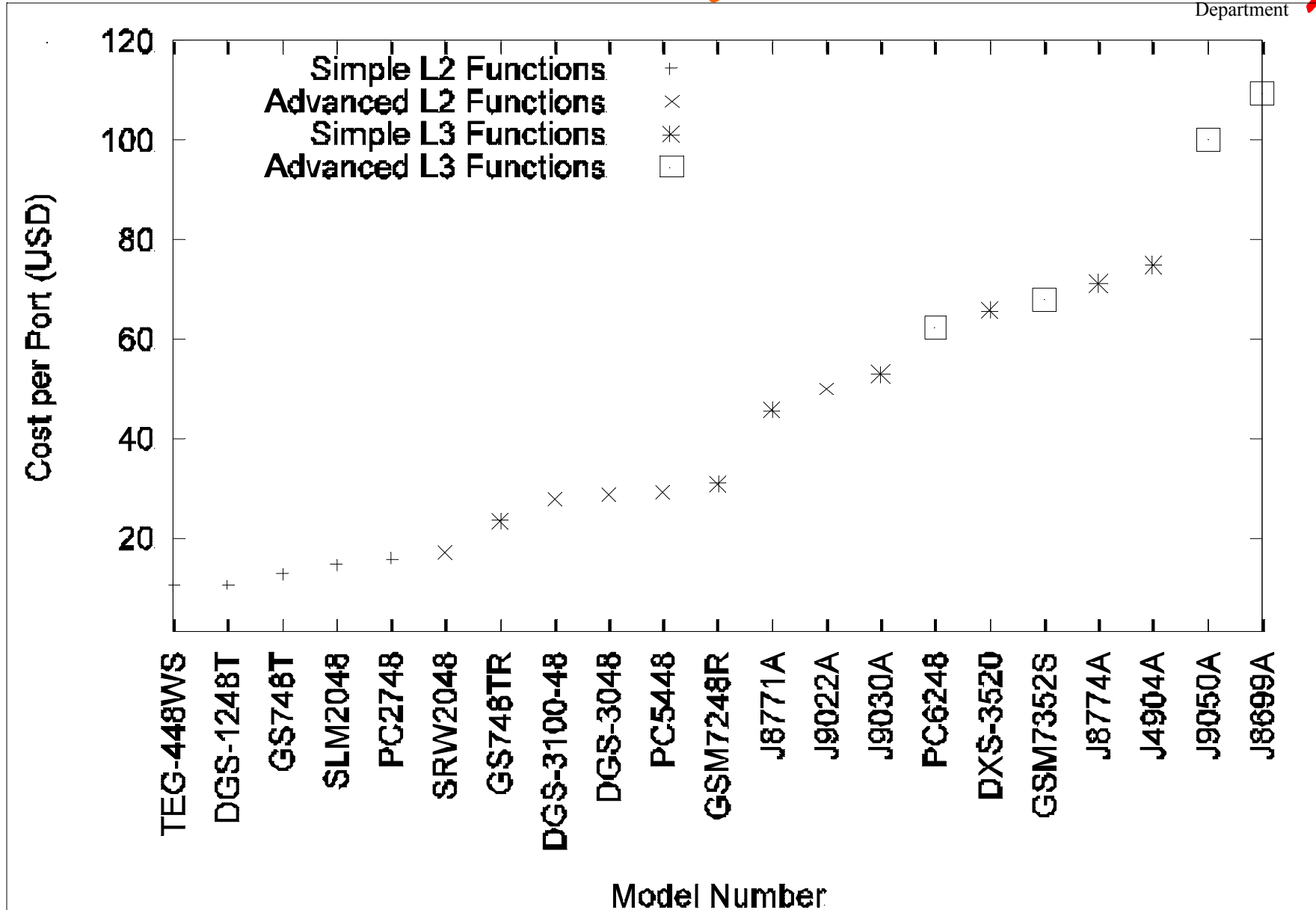


Emulate HW queues as multiple NICs



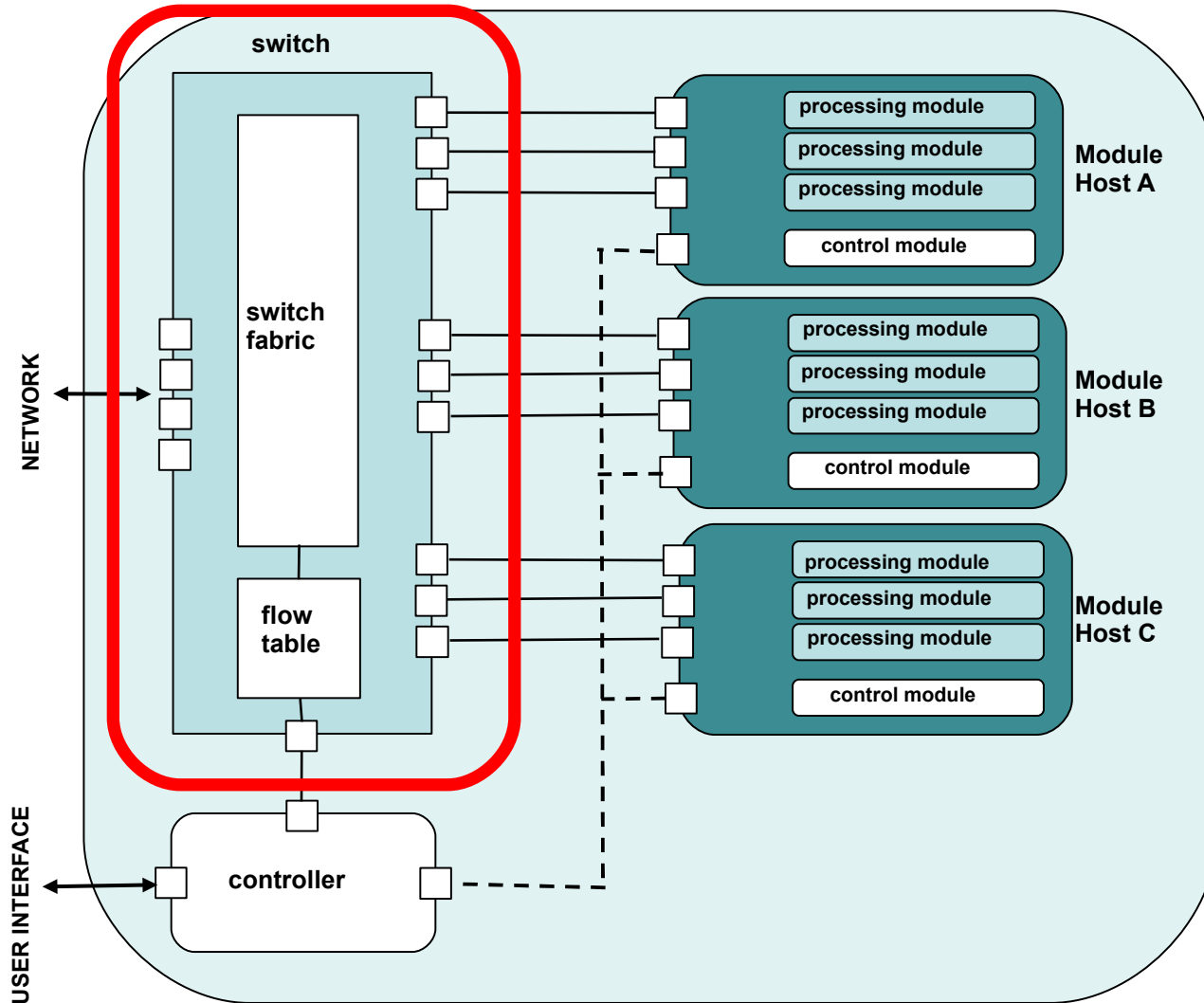
- Fairness not perfect
 - saturation point (partly) depends on traffic distribution
- Close enough for most purposes

Switches are commodity too

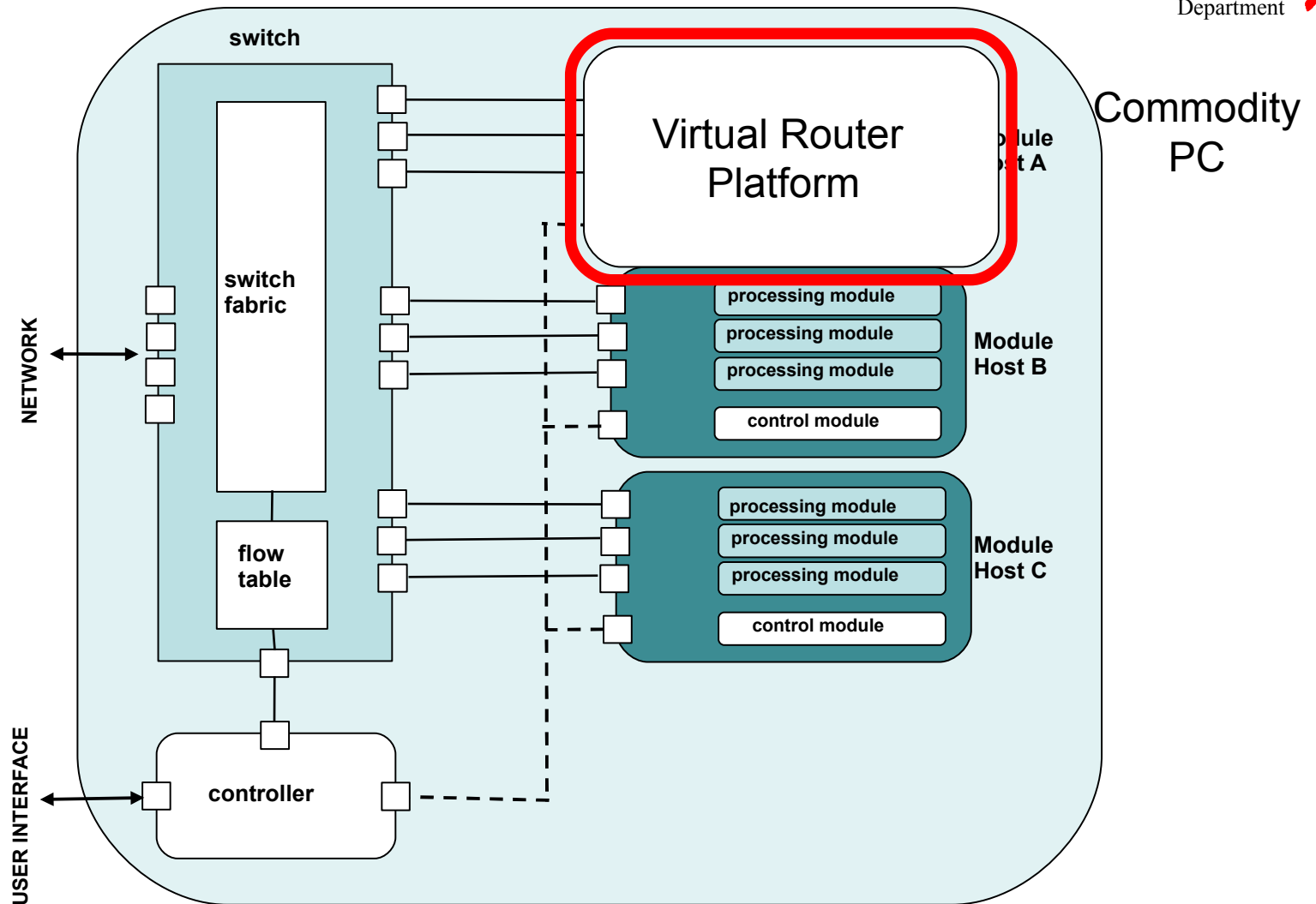


A platform for virtual router

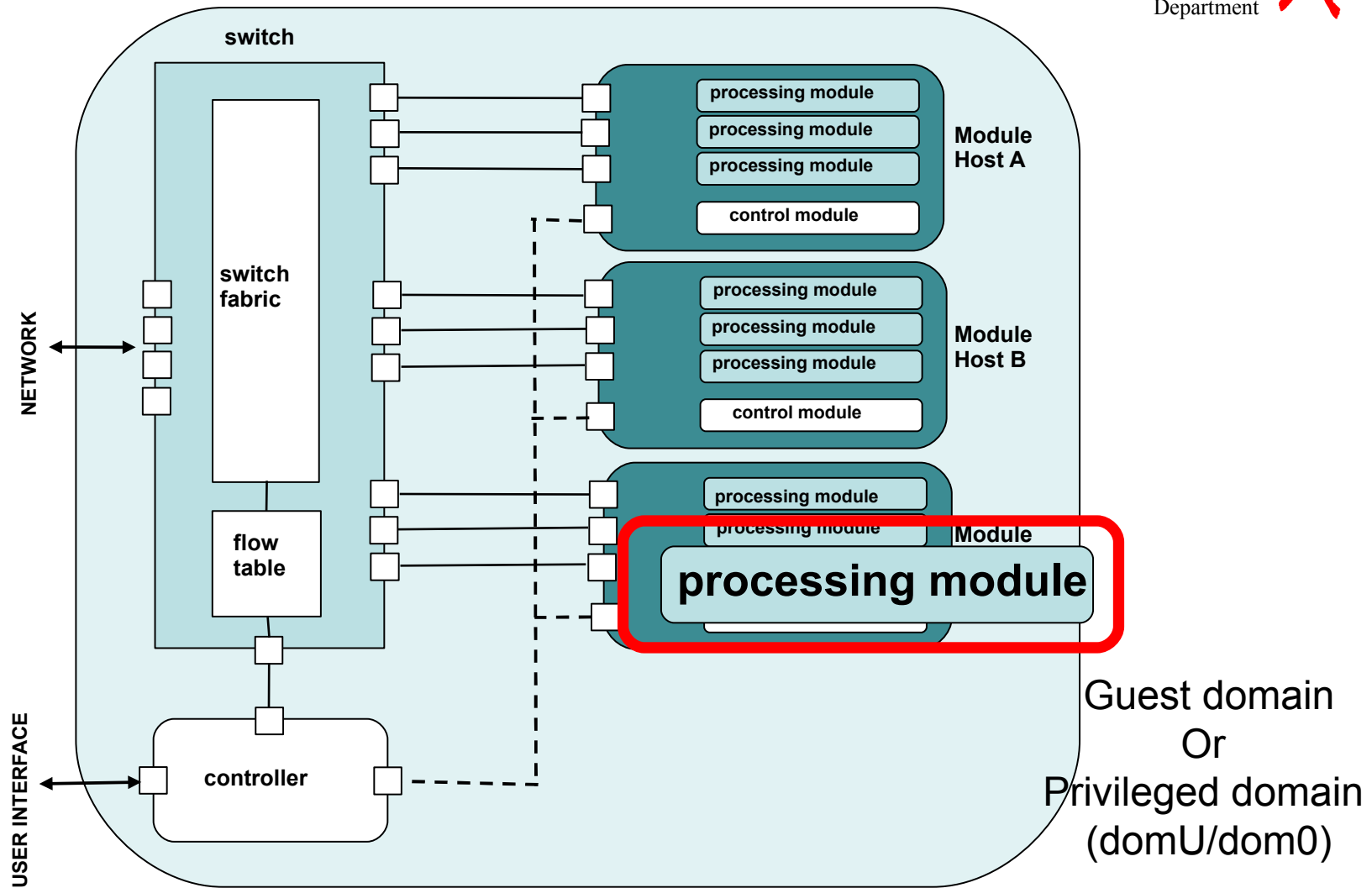
Commodity switch with Openflow software.



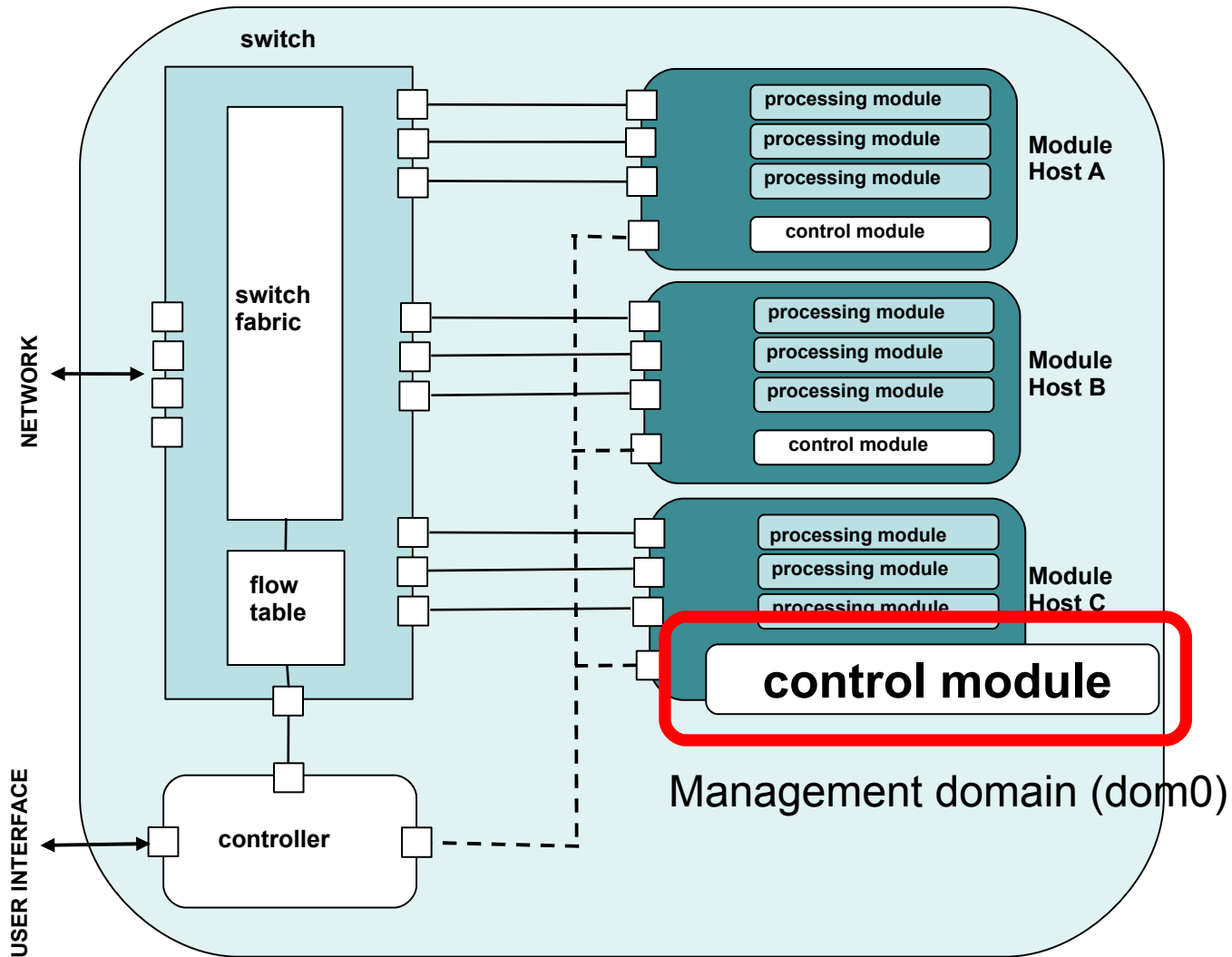
A platform for virtual router (2)



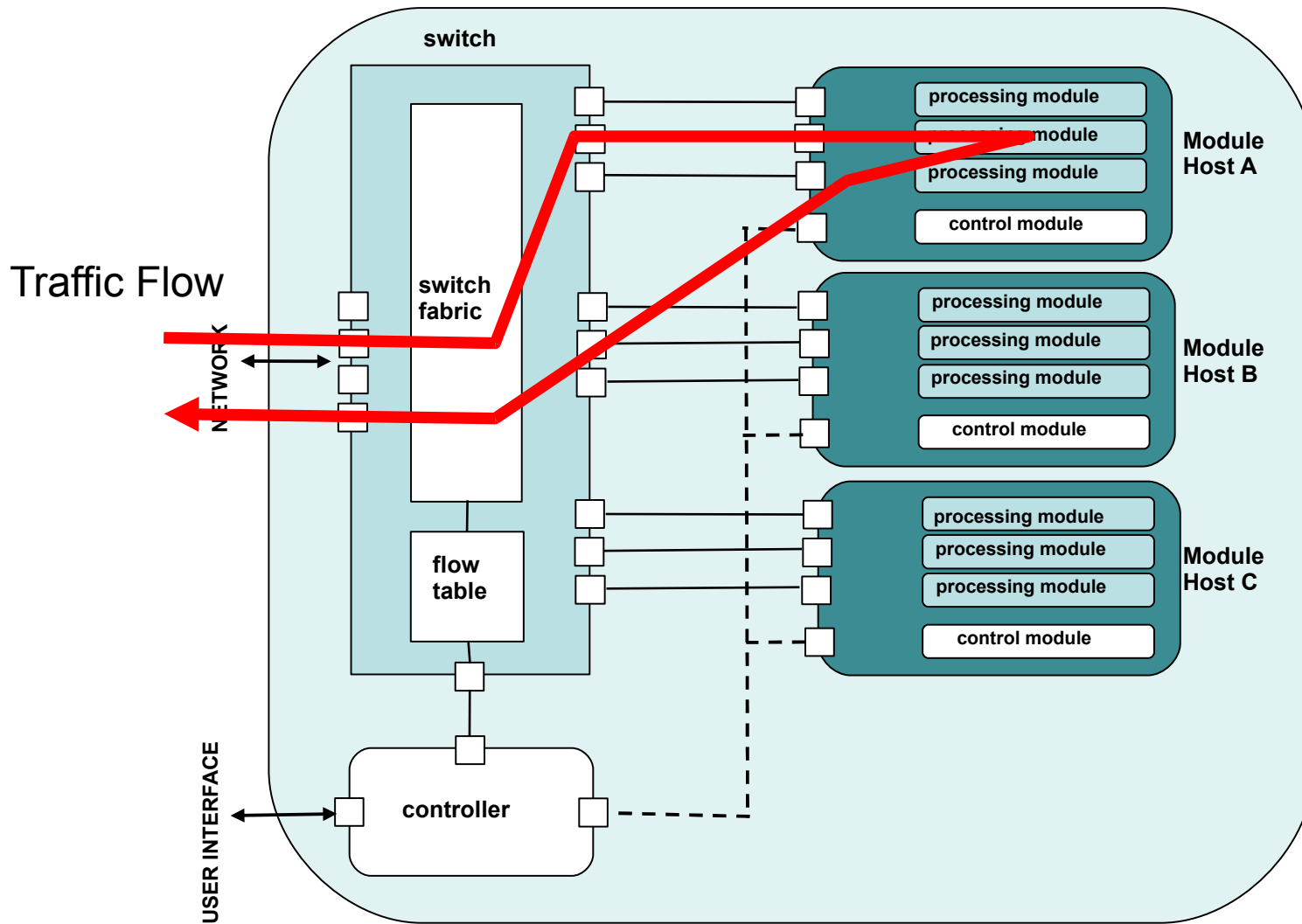
A Platform for virtual router (3)



A Platform for virtual router (4)



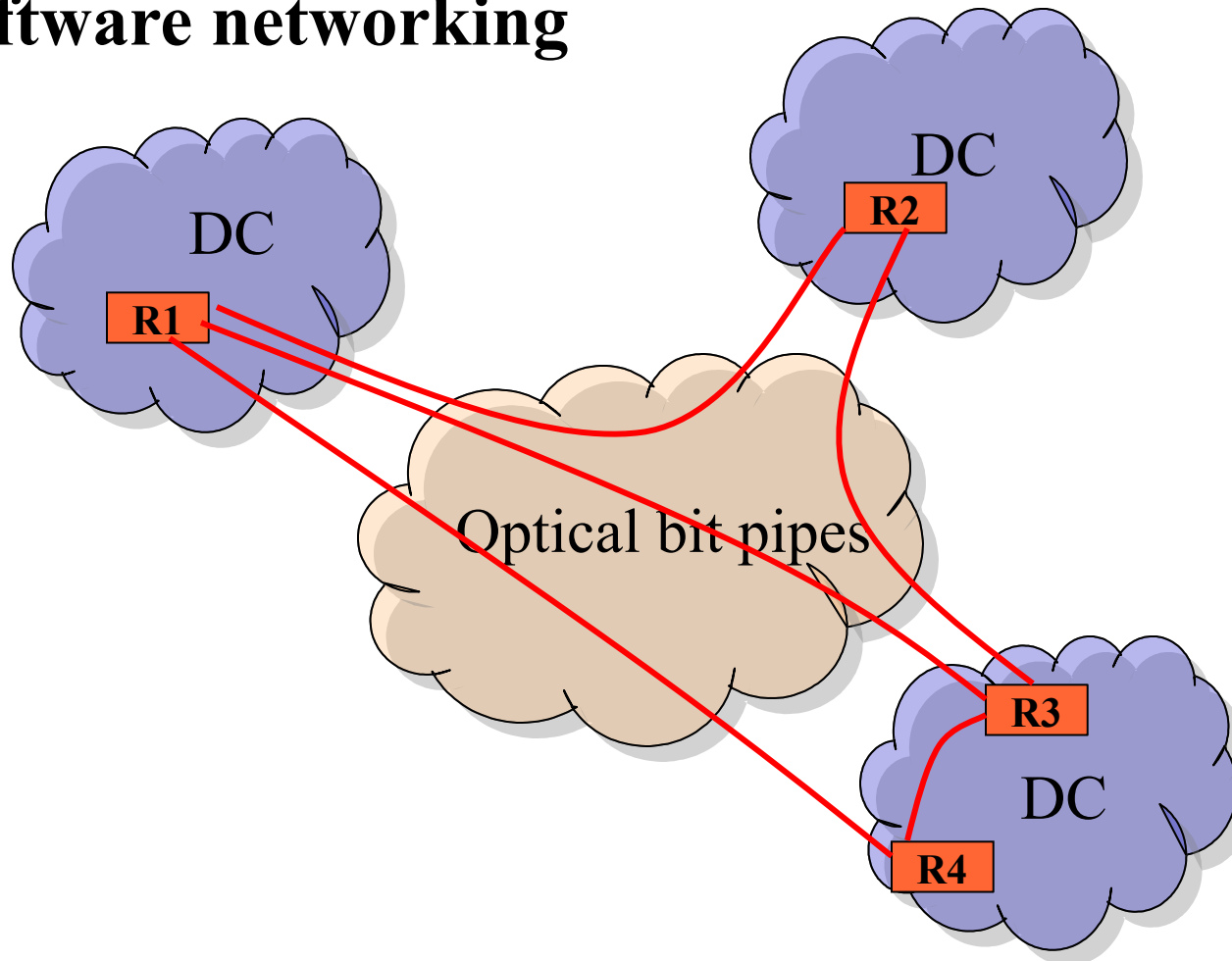
A Platform for virtual router (5)



→ Push this into the cloud

Virtual networking

- **Software networking**



Conclusions

- **Memory latency is bottleneck in SW (VR) routers**
 - Scales in terms of number of memory controllers, not cores!
- **High performance SW VR**
 - is forwarding engine virtualization
 - needs multi-queue NICs
 - **Commodity HW + familiar OS = better scope for programmability**
 - Deployment of new protocols
- **Virtual networking could be software networking**
- **Need to sort out**
 - Resource mapping
 - Routing
 - Etc.

Virtualization and Federation for Future Internet the FEDERICA experience

Mauro Campanella GARR

Mauro.campanella@garr.it

AICT 2009, May 25th 2009 - Venice/Mestre, Italy

Virtualization : FEDERICA Vision

An **e-Infrastructure** based on **virtualization** in both computers and network is a fundamental tool for researchers on **Future (and current) Internet**.

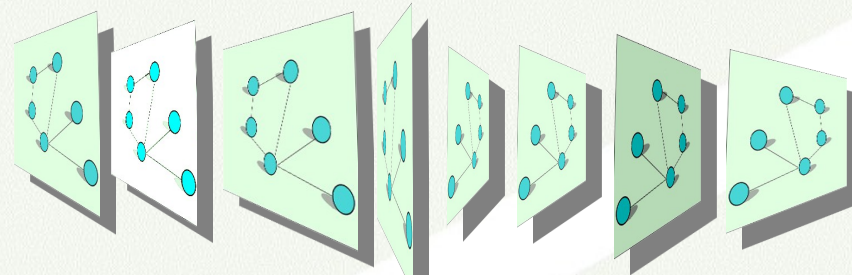
The facility should allow researchers a **complete control** of their set of resources in a “slice”, enabling **disruptive experiments** at all communication layers.




Particular care should be placed in **reproducibility** of the experiments and in avoidance of **complexity**.

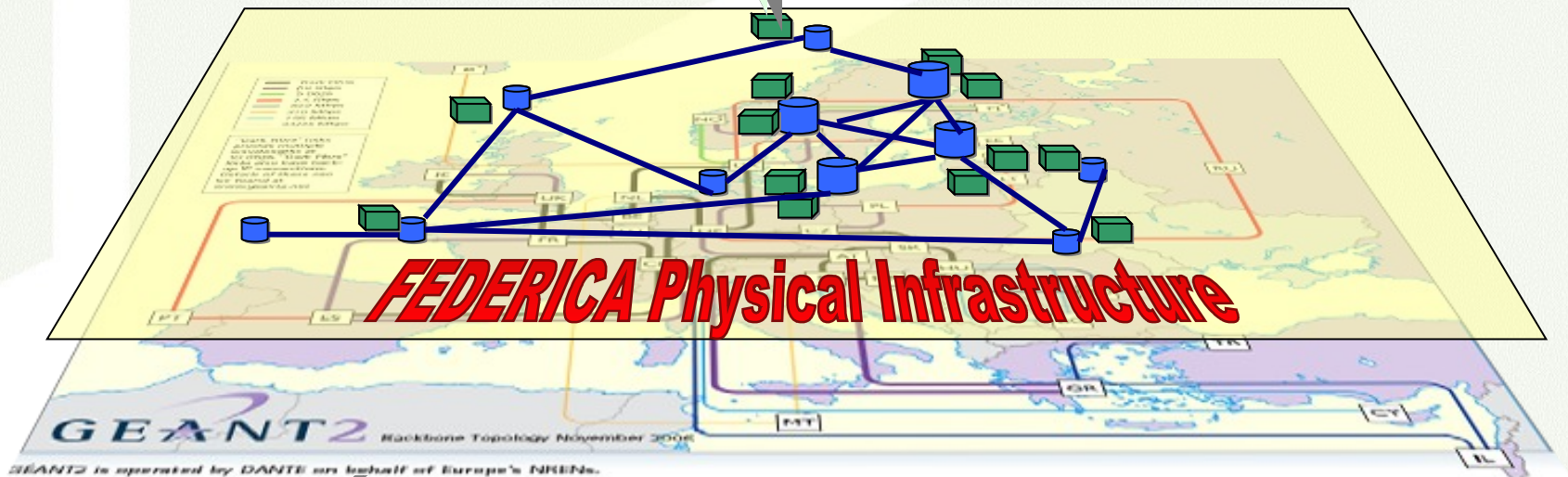
Such e-Infrastructure can be built on **existing infrastructures** using a **practical approach** and should be able to “federate” with other facilities.

An e-Infrastructure over other e-Infrastructures

Virtual Infrastructures (slices)



-  Router/Switch
-  Host for Virtual nodes
-  Raw Ethernet 1Gbps (Fiber later)



GEANT2 and NRENs Infrastructure

FEDERICA at a glance

What: European Community co-funded project in its 7th Framework Program in the area “Capacities - Research Infrastructures”, 3.7 MEuro EC contribution, 461 PMs

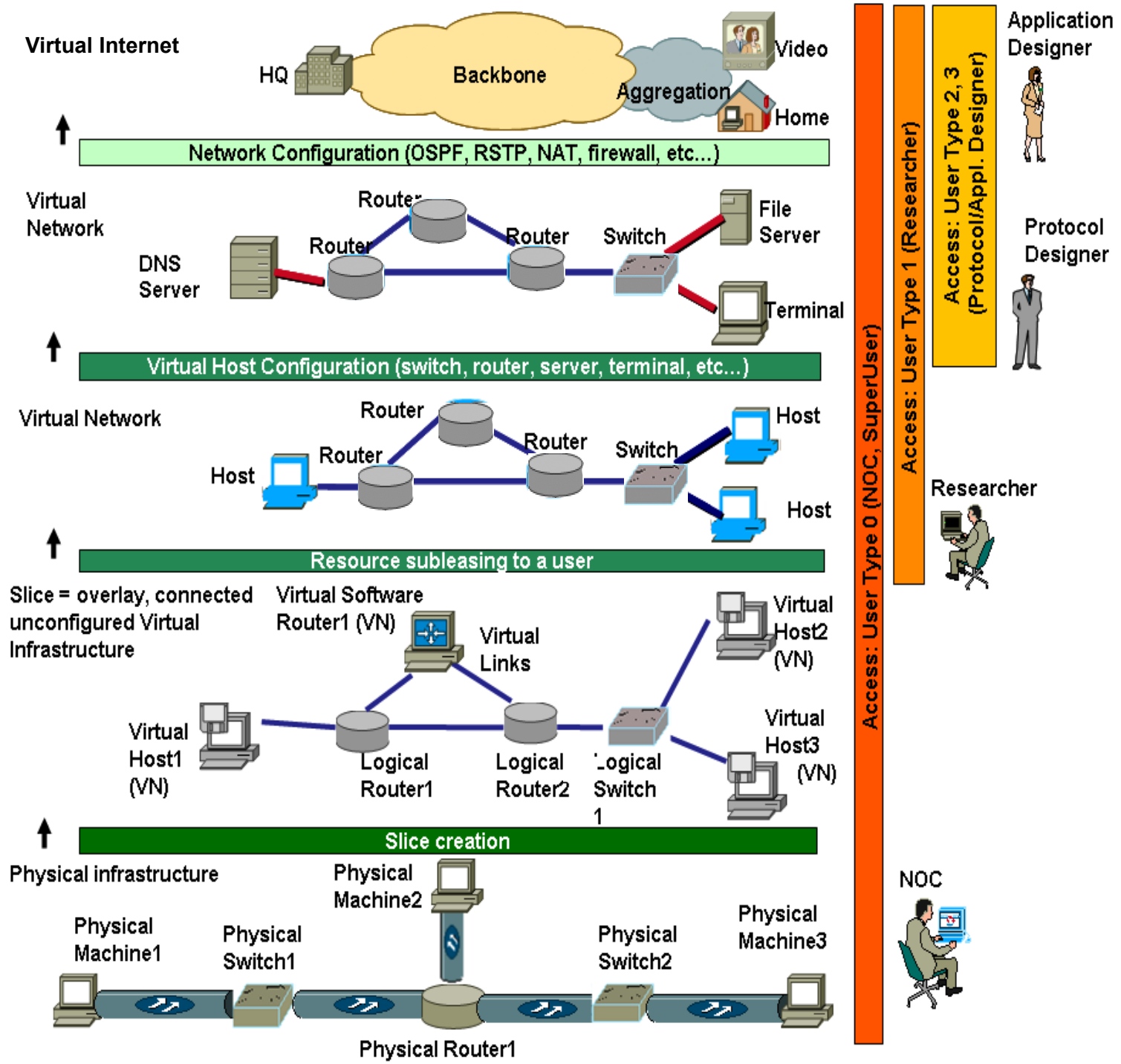
When: 1st January 2008 - 30 June 2010 (30 months)

Who: 20 partners, based on stakeholders on network research and management:

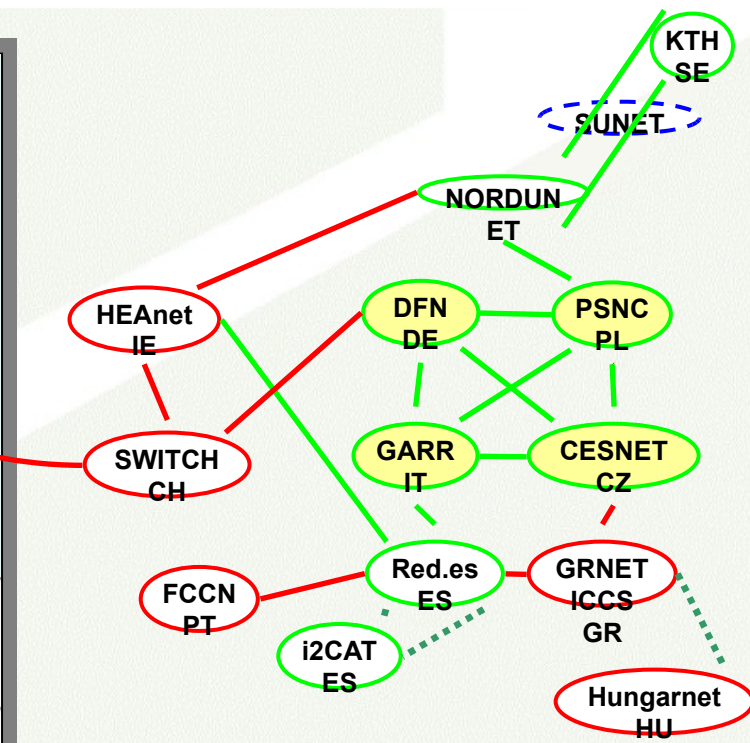
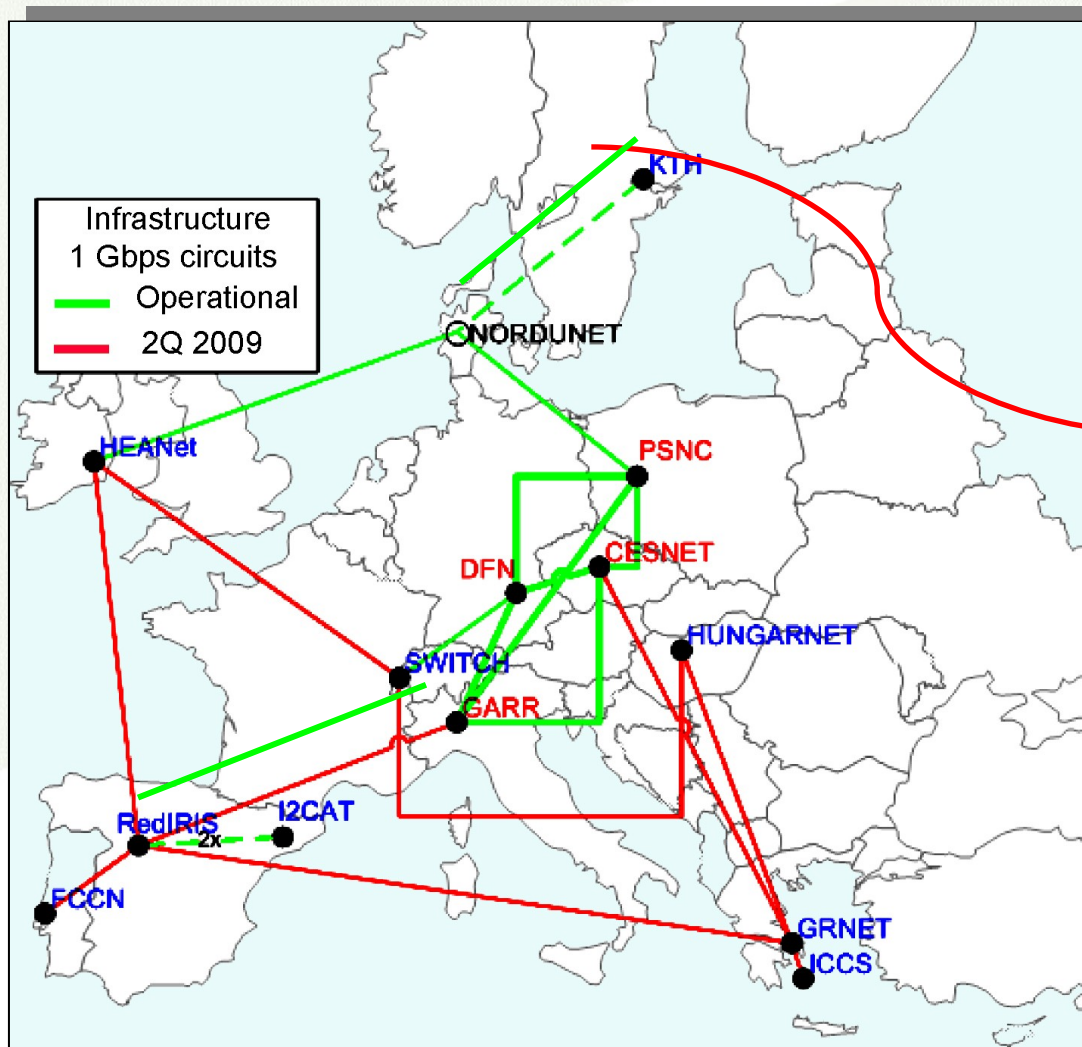
11 National Research and Education Networks, DANTE (GÉANT2), TERENA, 4 Universities, Juniper Networks, 1 small enterprise (MARTEL), 1 research centre (i2CAT) - Coordinator: GARR (Italian NREN)

Where: Europe-wide e-Infrastructure, open to external connections

From physical to virtual



Infrastructure Status: Operational



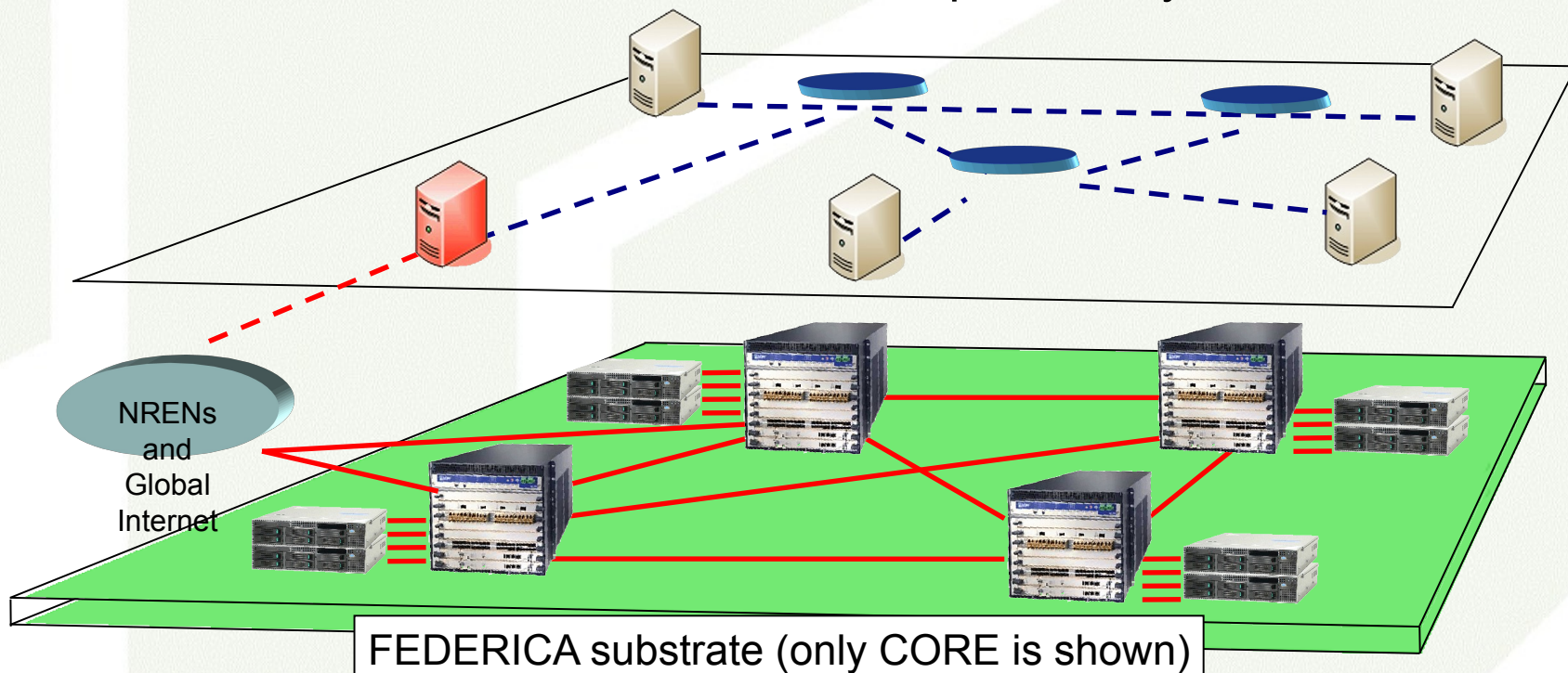
— 1 Gbps Ethernet

Each core PoP is equipped with a switch/router (Juniper) and two or more V-Nodes

Pictorial of creation of a Slice

The user requests an Infrastructure made of L2 circuits, un-configured virtual nodes, to test a new BGP version. Creation of:

1. user credentials and authentication and a “Slice”
2. Virtual Gateway (in red) to bridge the user from outside into the slice
3. Create resources and connect them as specified by the user



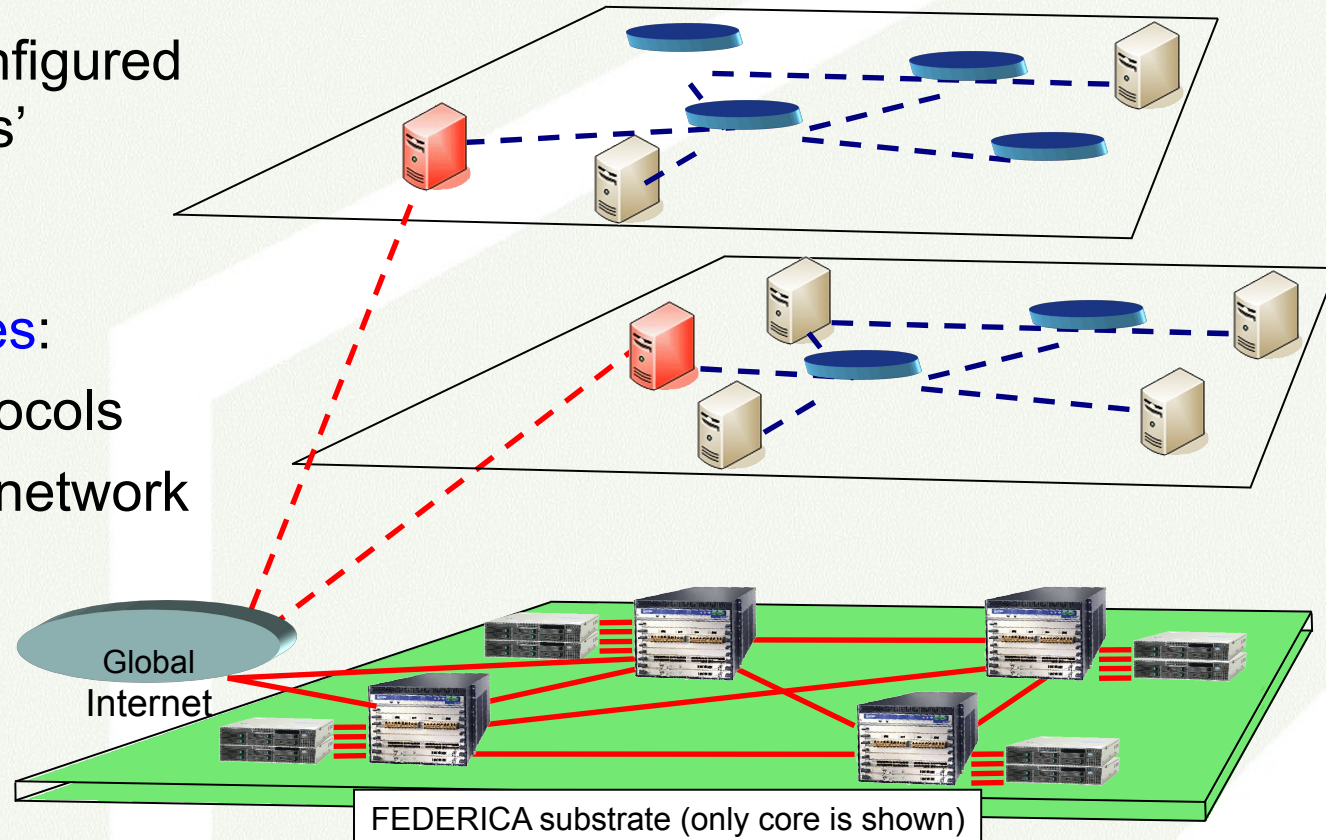
Offering “Slices” for “any” Research

Using **Virtualization technologies** the FEDERICA e-Infrastructure creates “slices” composed by virtual resources (**circuits, nodes, routers**)

The slices are configured according to users’ requests

Possible **use cases**:

- new **routing** protocols
- behavior on the network of **distributed applications**
- **Inter-domain services**



Federating FEDERICA

We should differentiate between various forms of federation:

1. **integrated** (the facilities can be used as one with a inter-domain common control plane)
2. **partially integrated** (only part of the control is exchanged, e.g. calendar, AAA information)
3. **overlay** (each facility just uses the services of the other without a common control plane, just a data plane, there is an exchange of information related to monitoring, faults, and so on)

All these possibilities present challenges ...

Federating FEDERICA (cont)

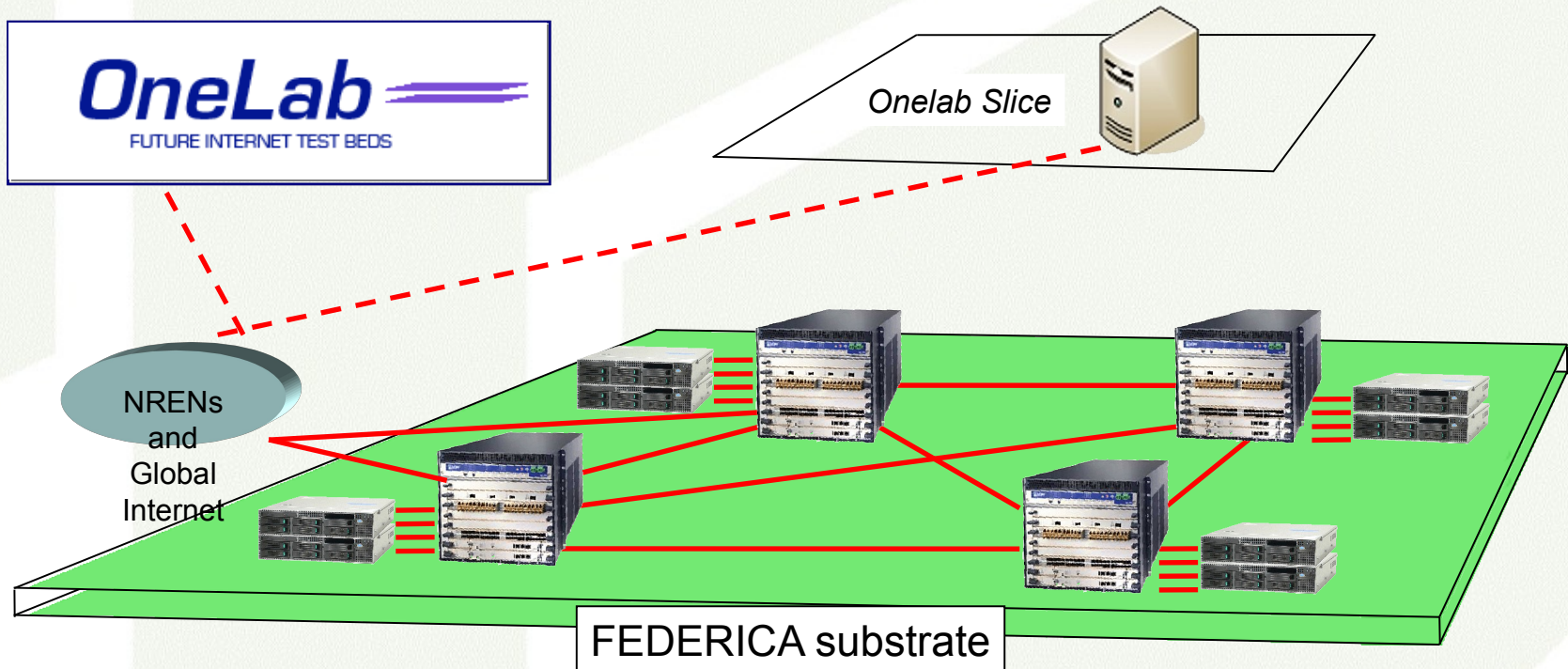
Having a common control plane in a multidomain environment is very difficult, as it places many constraints to each facility (in time, technology and developments).

Need to develop standard **resources representation schemas** for virtual resources and virtual resource sets to exchange **services**. The **inter**-facility exchange of information and synchronization between facilities has to scale gracefully.

The **Intra**-facility **control plane** is complex, due to resource scheduling and resources mapping from virtual topology (slice) to physical topology, especially if reproducibility or guarantees are mandated.

FEDERICA - Onelab2 pre-federation

OneLab nodes can be hosted in a slice. Those node have full control of their network interface and circuits up to the egress from FEDERICA into General Internet. The slice can contain also a “OneLab router”



Conclusions (for panel debate)

An infrastructure based on **virtualization, network and computing resources** can offer a very useful platform for **innovation and research** at many (if not all) communication layers and can be realized **on existing facilities**.

The **physical resources** functionalities and infrastructure **engineering** should ensure the functionality to guarantee the “**quality**” of the virtualization for reproducibility.

Federation is of outmost importance (e.g. to access new technologies and to ensure a larger facility), but has many facets and complexities. A practical approach through a “supportive” or “pre”-federation is advised.

Internet2's Involvement in the GENI Project – Federation and Virtualization

Rick Summerhill
Chief Technology Officer, Internet2
AICT Meeting
Venice, Italy
25 May, 2009

(With some slides developed by the GPO)

Outline

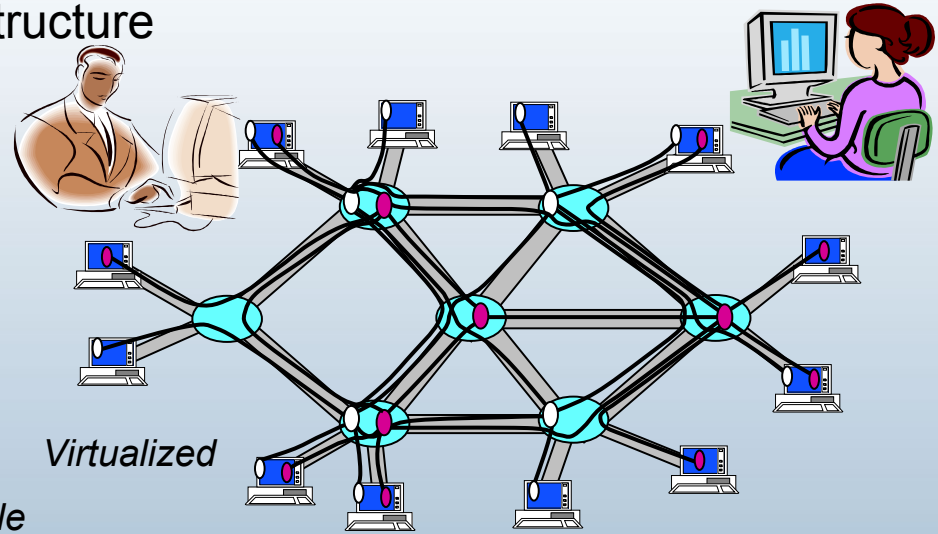
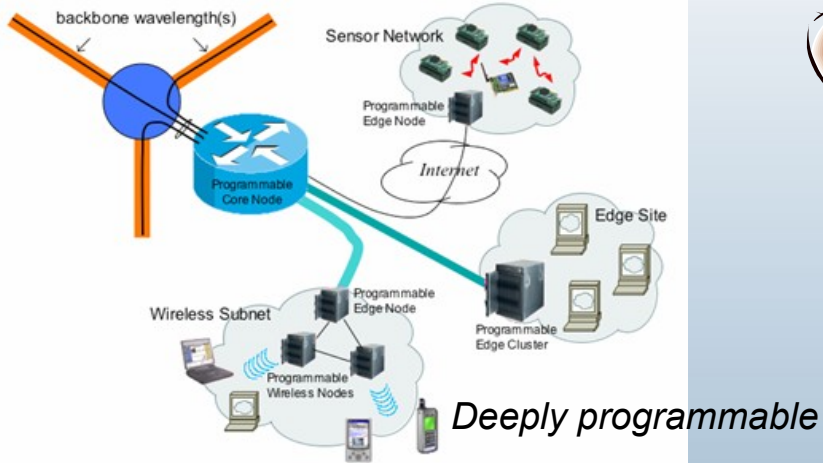
- Federation and Virtualization
- The GENI Project Vision
- The GENI system concept
- Internet2 Participation in GENI
- Examples

Federation and Virtualization

- Must occur at multiple levels – infrastructure, project, research, usage
- Federation
 - "Resource federation permits the interconnection of independently owned and autonomously administered facilities in a way that permits owners to declare resource allocation and usage policies for substrate facilities under their control, operators to manage the network substrate, and researchers to create and populate slices, allocate resources to them, and run experiment-specific software in them."
- Virtualization
 - The ability to create resources that appear to be physical in nature, and indeed sit atop such resources, but allow independent control as if it were a separate physical system.

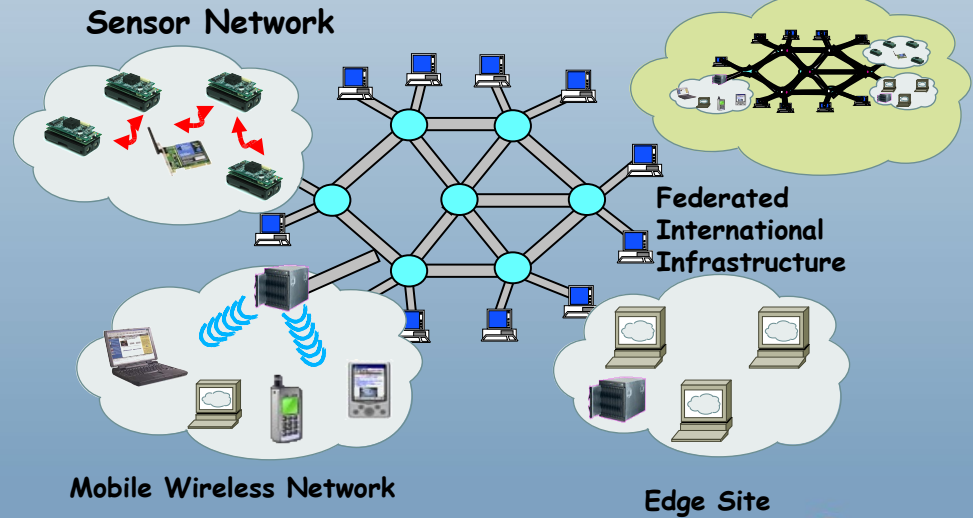
The GENI Vision

A national-scale suite of facilities to explore radical designs for a future global networking infrastructure



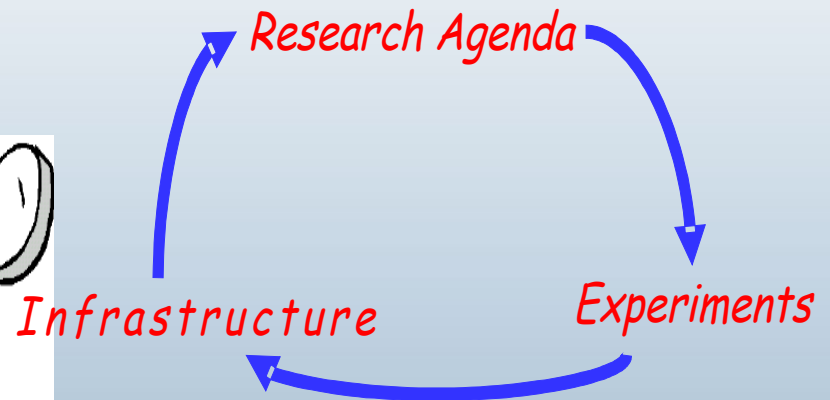
Programmable & federated, with end-to-end virtualized “slices”

Heterogeneous, and evolving over time via spiral development



Research Agenda to Experiments to Infrastructure

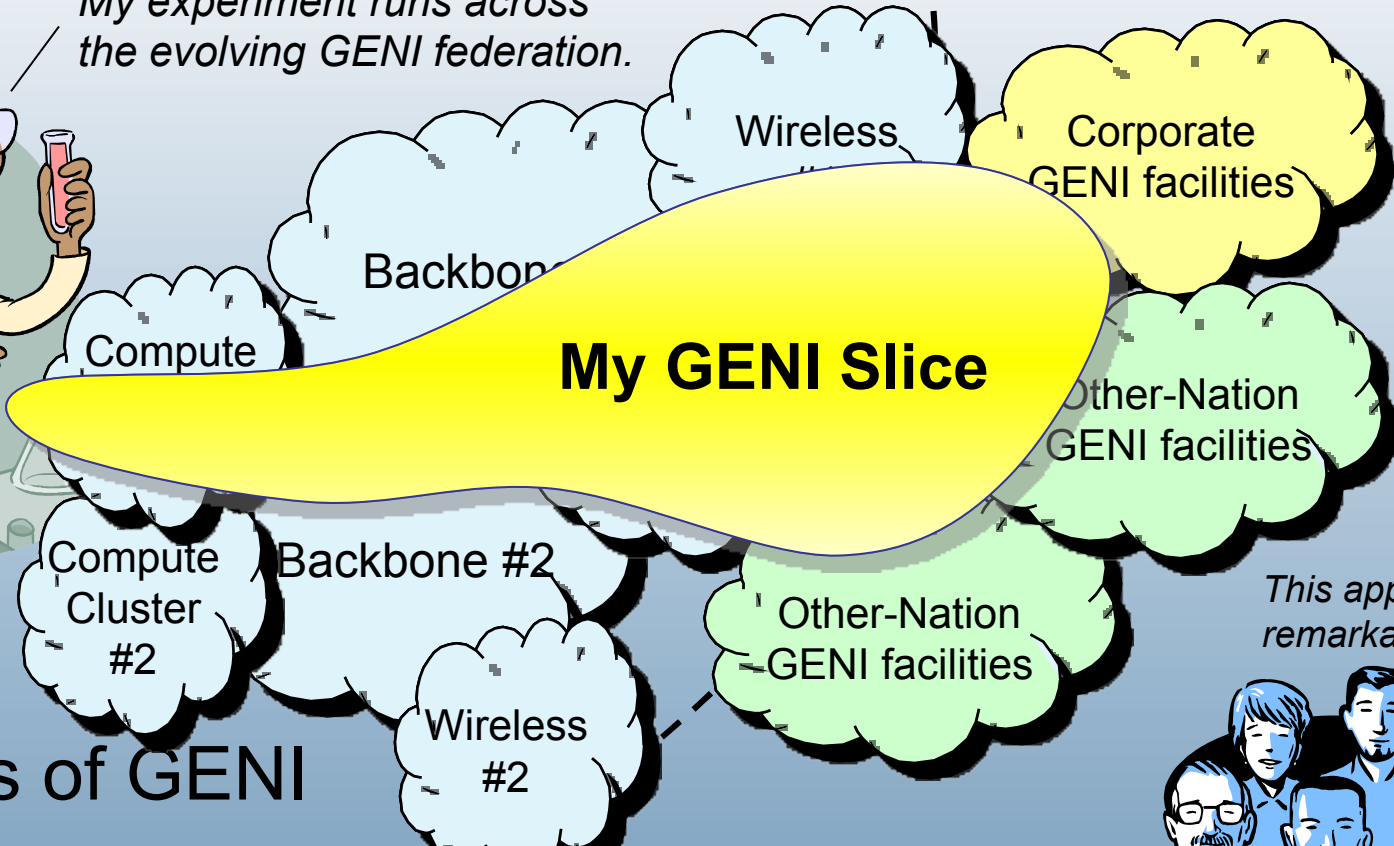
- Research agenda
 - Identifies fundamental questions
 - Drives a set of experiments to validate theories and models
- Experiments & requirements
 - Drives what infrastructure and facilities are needed
- Infrastructure could range from
 - Existing Internet, existing testbeds, federation of testbeds, something brand new (from small to large), federation of all of the above, to federation with international efforts
 - No pre-ordained outcome



Federation

GENI grows by “gluing together” heterogeneous facilities over time

My experiment runs across the evolving GENI federation.



This approach looks remarkably familiar . . .



SF parts of GENI

Goals: avoid technology “lock in,” add new technologies as they mature, grow quickly by incorporating existing facilities into the overall “GENI ecosystem”

Internet2 Participation in GENI

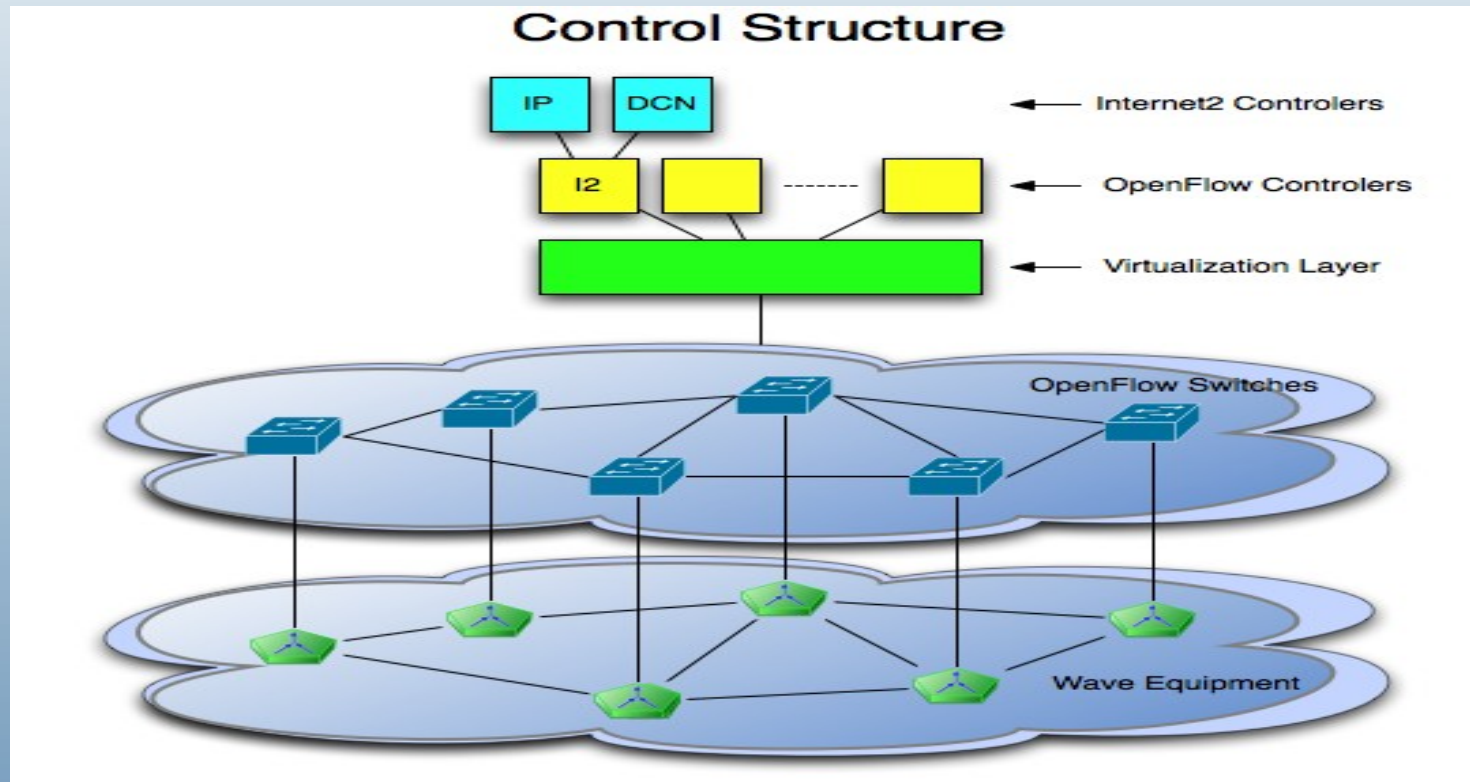
- Submitted initial proposal with the University of Utah as lead - a response to the first solicitation
 - Uses a 10 Gbps Wave on the full Internet2 footprint
 - Ethernet switch based
 - PCs having 2 netFPGA cards
 - Can be programmed in a variety of ways
- Proposal was successful
- Other Projects will use the wave - SPP

Internet2 Participation in GENI

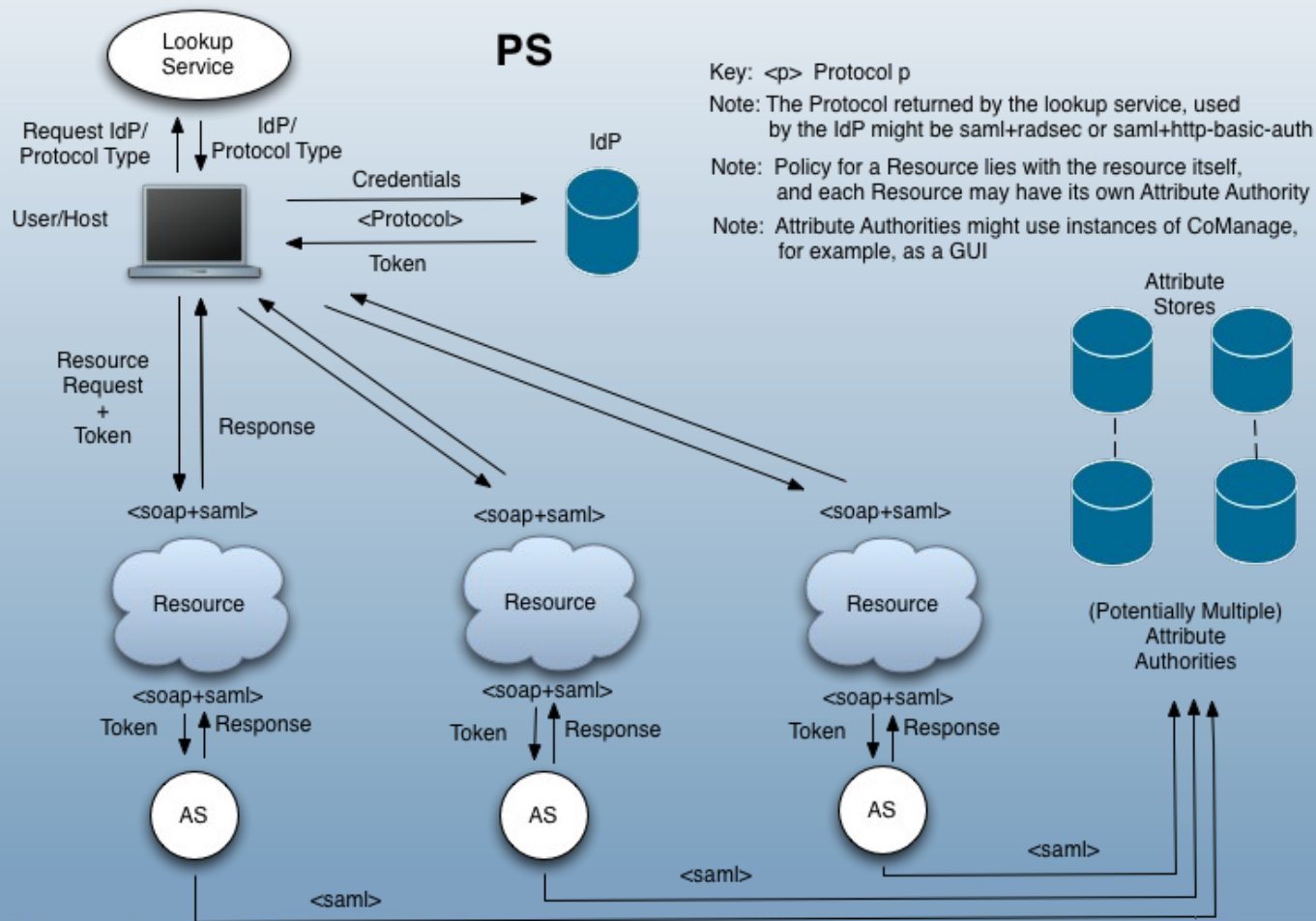
- Subsequent discussions with the GPO lead to Internet2 contributing a 10 Gbps wave on the full Internet2 Network footprint to the GENI project
 - Dedicated totally to GENI
 - Additional Proposals Submitted to the Second Round
- Challenge for Infrastructure
 - How do these projects “fit” together?
 - How do we insure reproducibility?

Example: Virtualized Control - OpenFlow

- Research Projects Coexisting with Production Services



Example: Federated Identity and Authorization



References

GENI: <http://www.geni.net>

Internet2: <http://www.internet2.edu>

OpenFlow: <http://www.openflowswitch.org/>

InCommon:

<http://www.incommonfederation.org/>

www.internet2.edu

INTERNET[®]
2

The GENI Project and the GMOC

Federated Management Infrastructure and
Implications Beyond GENI

James G. Williams
Indiana University
Bloomington, Indiana USA

The GENI Project

Global Environment for Network Innovations

Funded by the US National Science Foundation with contributions from public/private organizations

Operated by BBN through the GENI Project Office (GPO)

Initially a three-year view forward

Five “control plane framework” projects with virtualized components

Some cross-framework activities

GENI Meta Operations Center (GMOC) is one of these

See: <http://groups.geni.net/geni/wiki/SpiralOne>

The GENI Meta Operations Center (GMOC)

Cross-cutting activity

Goal: To start to help develop the datasets, tools, formats, & protocols needed to **share operational data among GENI constituents**

Initial projects:

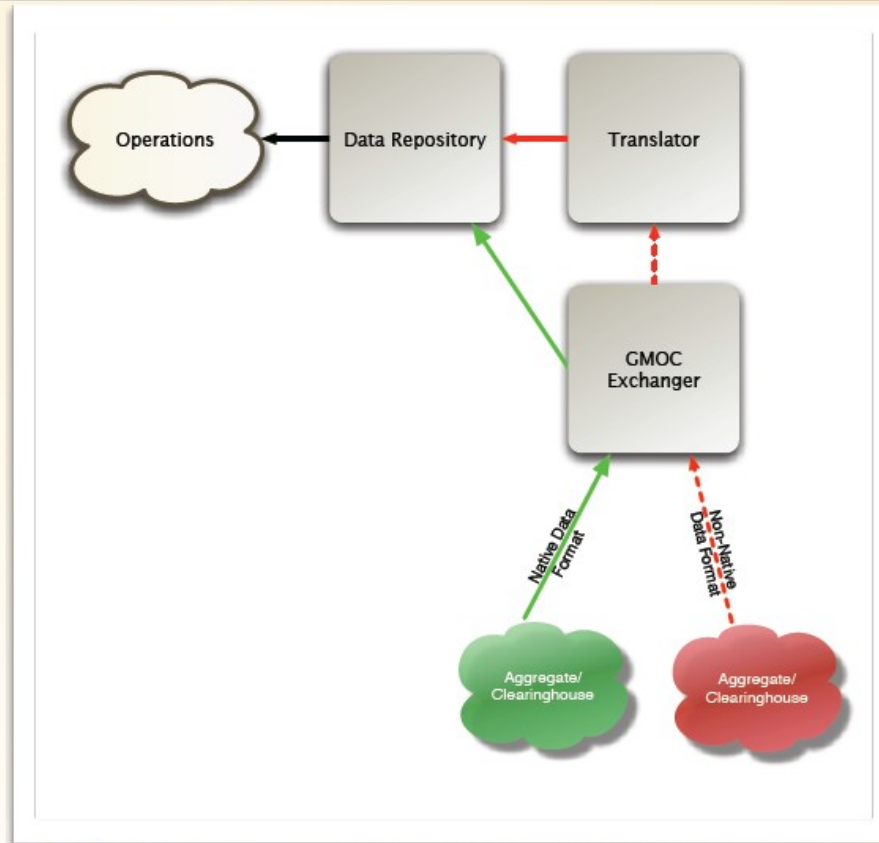
Provide GENI operational data views (both GENI-wide and researcher specific) <Need for a common operational dataset>

Emergency STOP – find out-of-control virtual slices and isolate or shut them down

Lead by: Jon-Paul Herron (PI) and Luke Fowler (co-PI) Indiana University

See: <http://groups.geni.net/geni/wiki/GENIMetaOps>

GENI Meta-Operations Center



Implications beyond GENI

(a dynamic circuit example)

Consider an experiment spanning international network domains of control [Italy -> DANTE -> Internet2/NLR ->Texas]

Imagine this experiment needs dedicated bandwidth for a limited time (4 hours) on a regular basis (once a week) which will be provided by a dynamic circuit.

From the viewpoint of the experimenter, how can he/she regularly determine the end-to-end performance characteristics of this dynamic network connection.

From the viewpoint of a network operator (one of several involved here), how can the operator determine the status of this network connection, beyond simply saying “My part of the connection looks good”.

This is the general case of the specific GMOC activity.

How will network operators exchange the data necessary to allow end-to-end troubleshooting of cross-domain circuits?

How will network operators exchange data to create a end-to-end view (user view or operator view) of cross-domain circuits?

How will network security concerns be taken into account?

We believe the GMOC activity represents **one possible path forward** in addressing these complex cross-domain issues.

Actively Searching for Collaborators

Indiana University is the network operator for many R/E networks in the US including Internet2 and NLR. See: <http://globalnoc.iu.edu/>

We are deeply involved in the GENI research activities, specifically leading the GMOC Project. We also actively participate in perfSONAR development activities and distribution, especially in Asia.

We are actively seeking a diverse set of partners to discuss the possibility of the extension of the GMOC ideas and federated network management in general to production R/E networks.

Contacts:

James Williams williams@indiana.edu

Jon-Paul Herron jph@grnoc.iu.edu

Towards A service Platform for
Virtualization and Federation :
a case of JGN2Plus

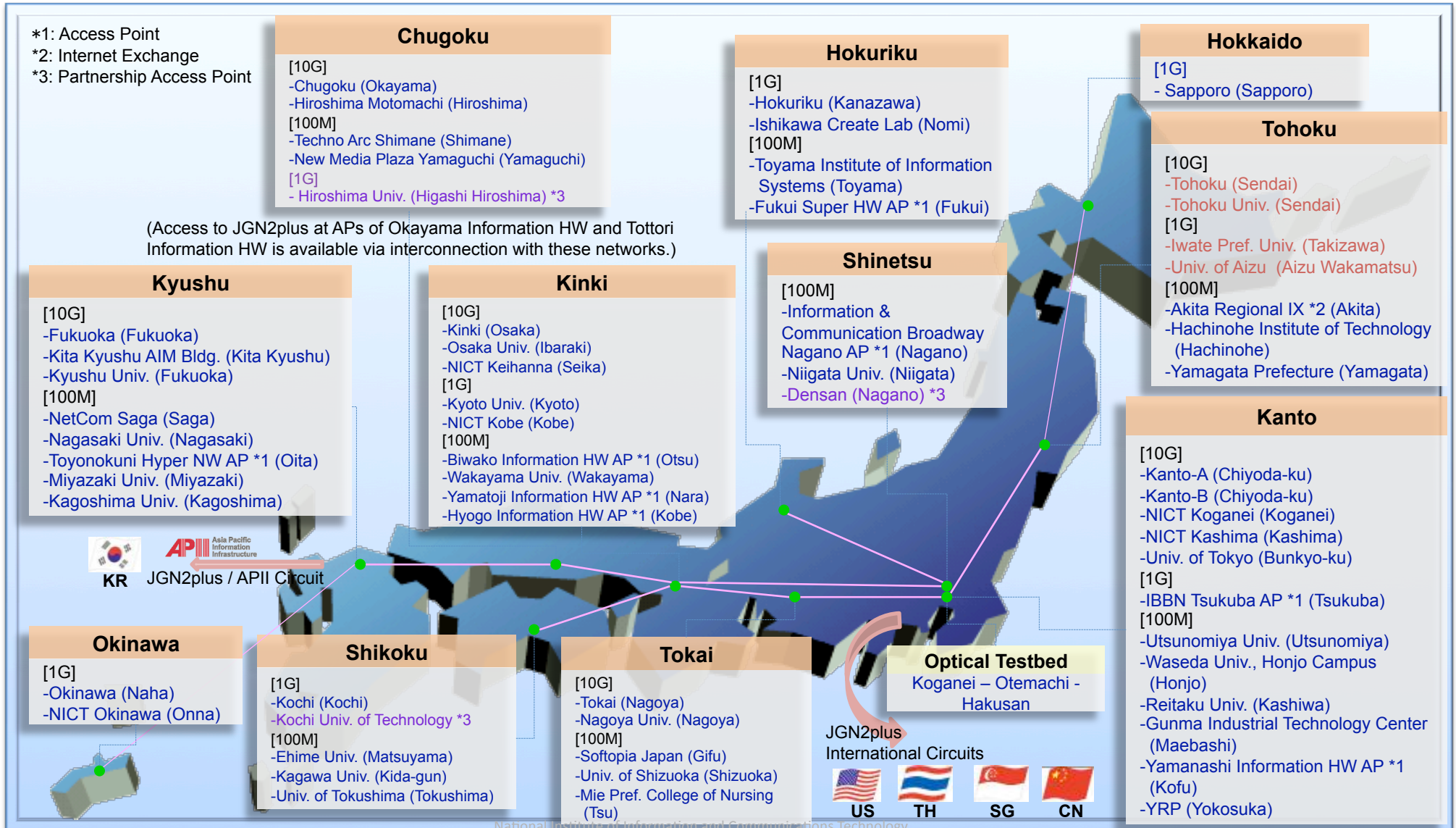


Shinji Shimojo
(NICT) May 2009

**The National Institute of Information and
Communications Technology
Collaborative Research Department
Network Testbed Research Promotion Group**
National Institute of Information and
Communications Technology

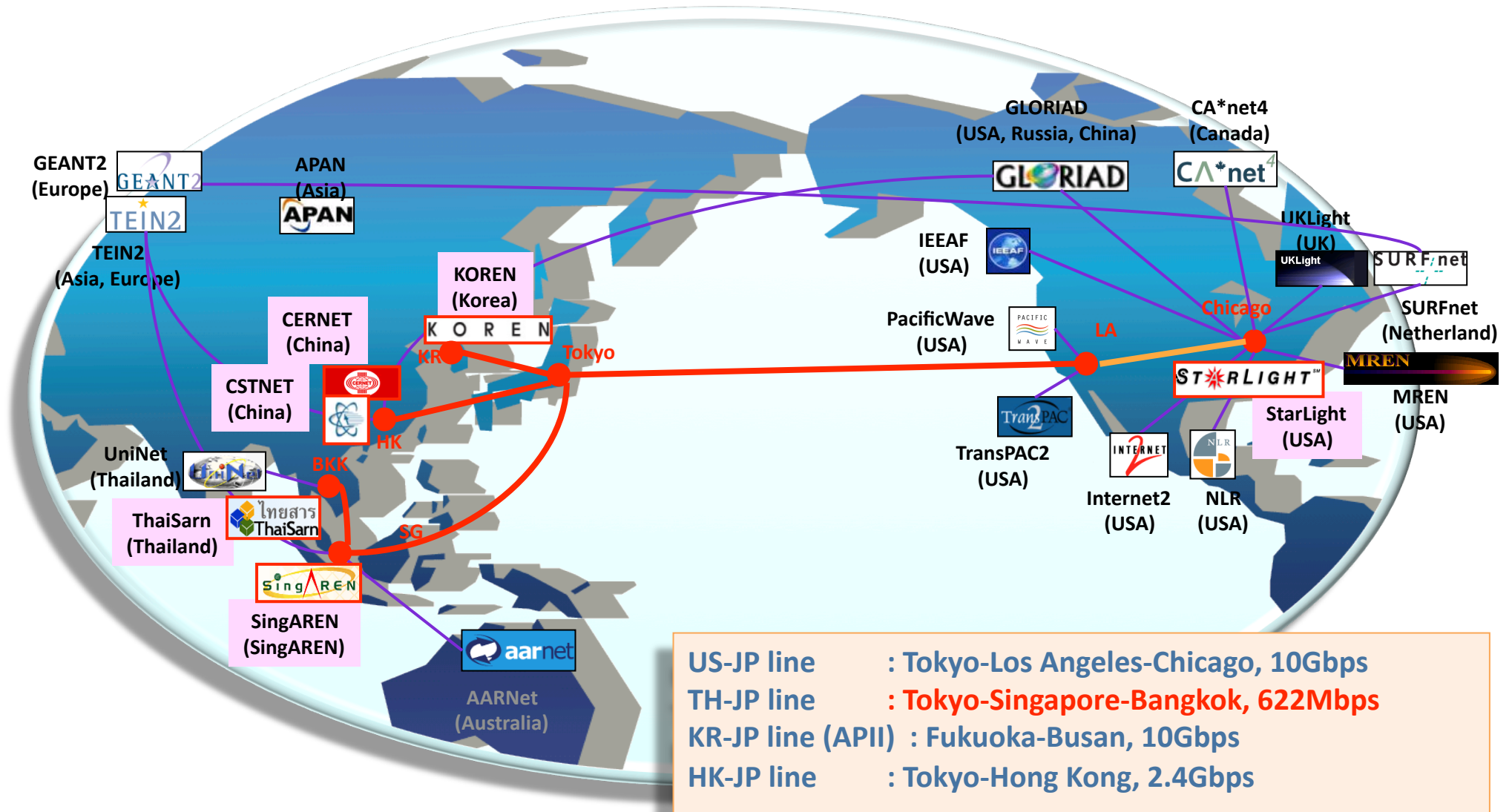
JGN2plus Services

2. JGN2plus Network Outline



JGN2plus Services

JGN2plus International Circuits (L2/L3)

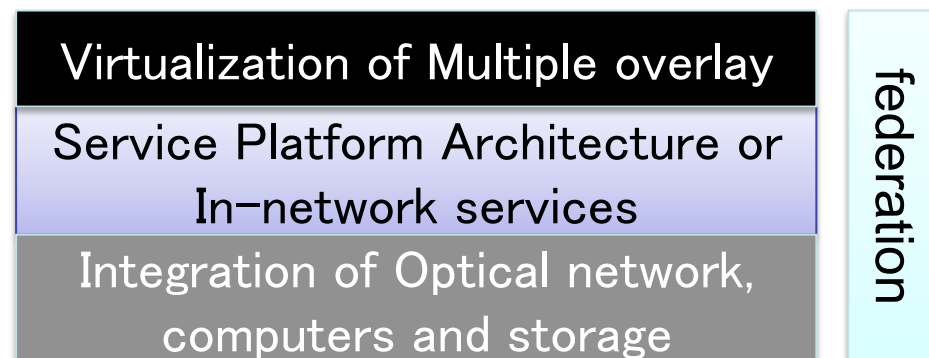


Testbed for NwGN

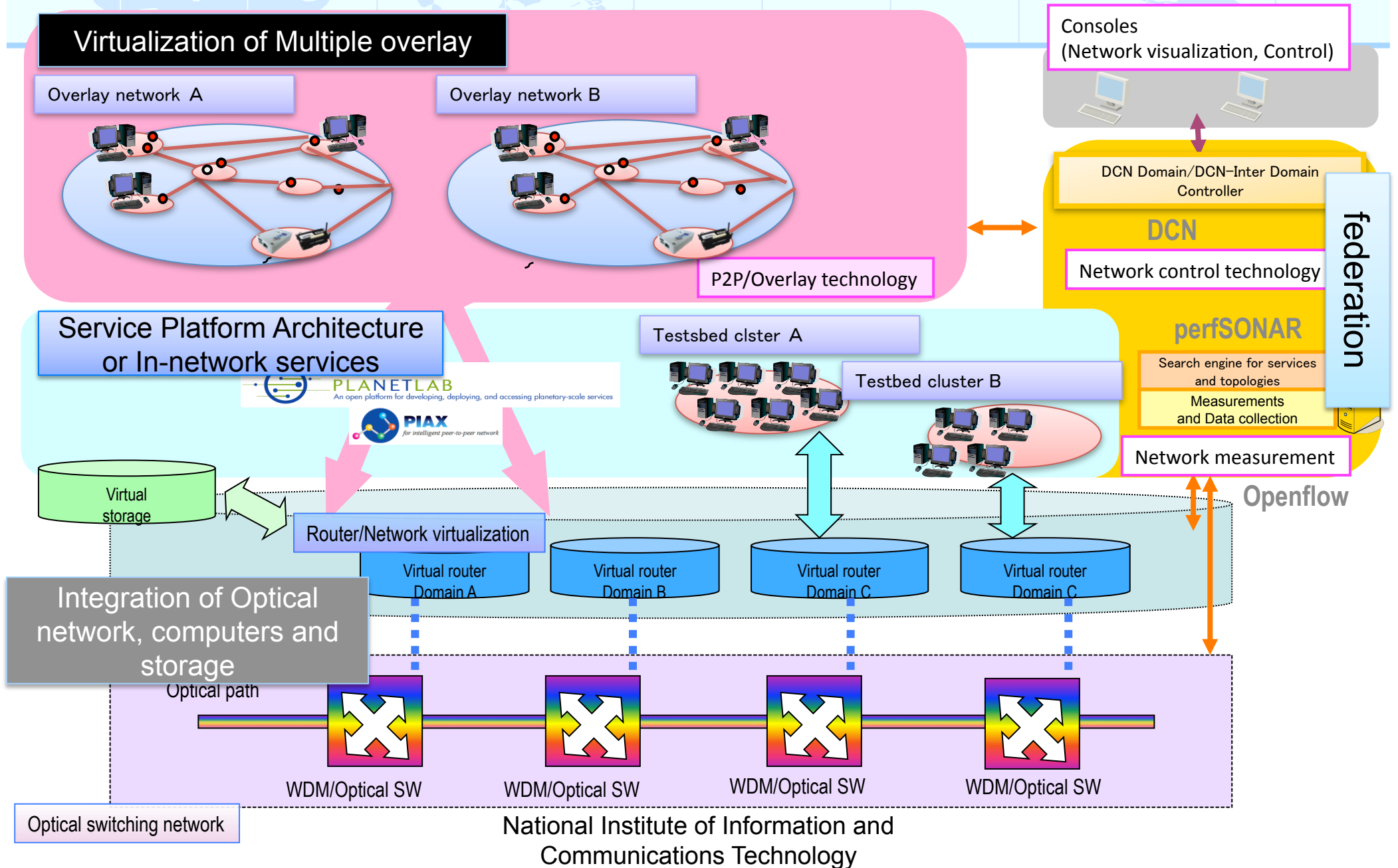
- JGN has been providing cyberinfrastructure as well as an advanced networking R&D infrastructure for wide variety of applications such as ...
 - Optical networking (optical switch, GMPLS, G-lambda)
 - Next generation TCP
 - Advanced Broadcasting applications (IPv6, various type of multicasting, HD, 4K)
 - E-sciences (e-VLBI, WINDS, TDW, art)
 - Other applications (remote lecture/telemedicine etc)
- Testbed is used for NwGN research of
 - Proof of concept
 - International Collaboration and standardization
 - Social implication
 - Human resource development



- Keeping production service
 - Existing activities in cyberinfrastructure
- Encouraging new experimental R&D activities for NwGN inside testbed
 - Introducing new functionality inside the network
 - which affect network functionality.
- supporting evolution of technology
- And make network to rely on existing and operational technology



Detailed view of the service platform in JGN2plus (Near Future)



■ Multiple Overlay – PLANETLAB, PIAX, IMS

■ SPARC (In-network service)



- Provisioning (Dynamic circuit network)
- Measurement (PerfSONAR)
- Openflow



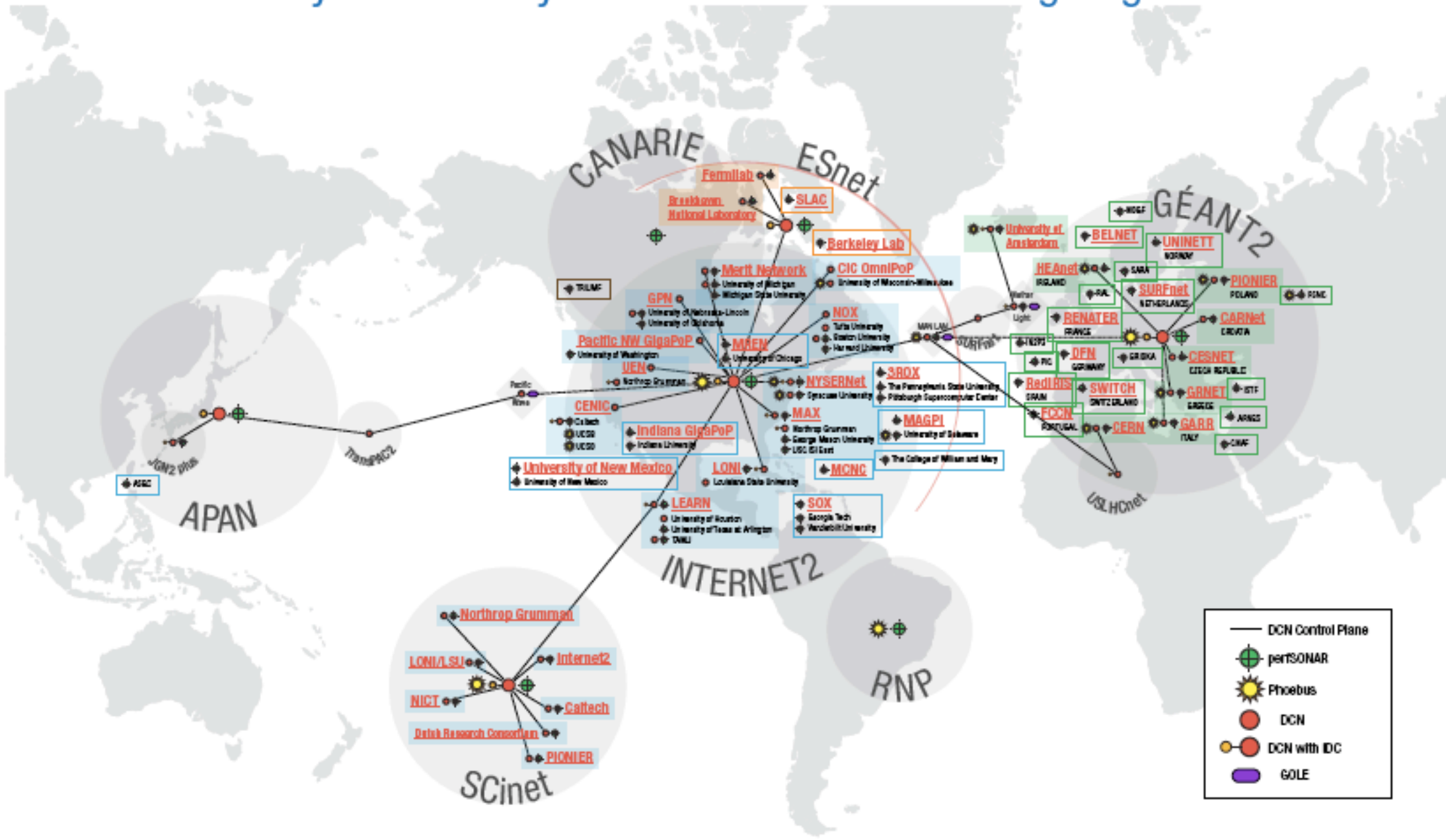
■ Integration of Optical network, computers and storage

- Cisco/Juniper Commercial Virtual Management Router(3 CSRs and 3 T1600 with JCS)
- Optical Network core (in future)

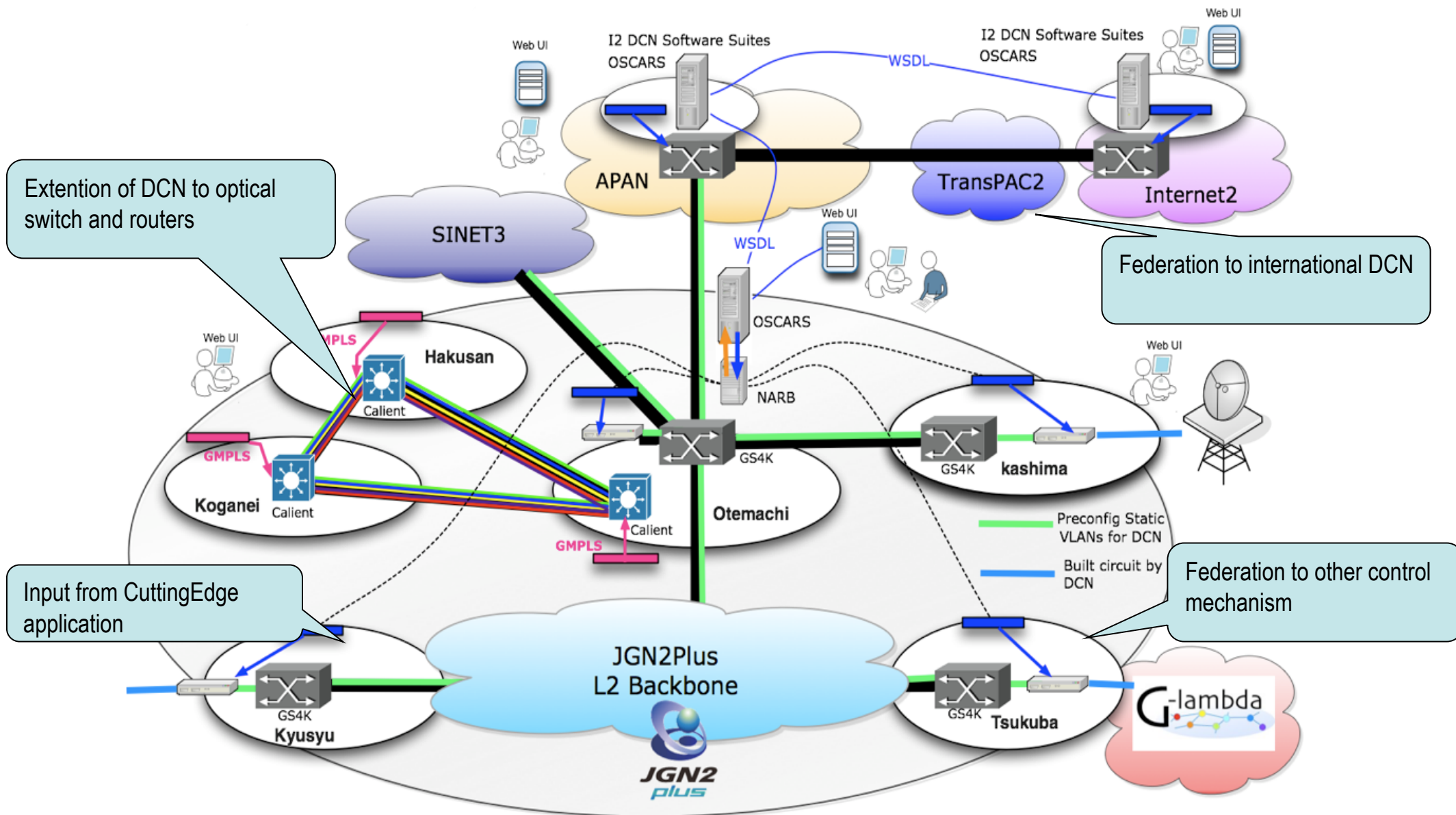
■ Federation of In-network services

- DCN, PerfSONAR

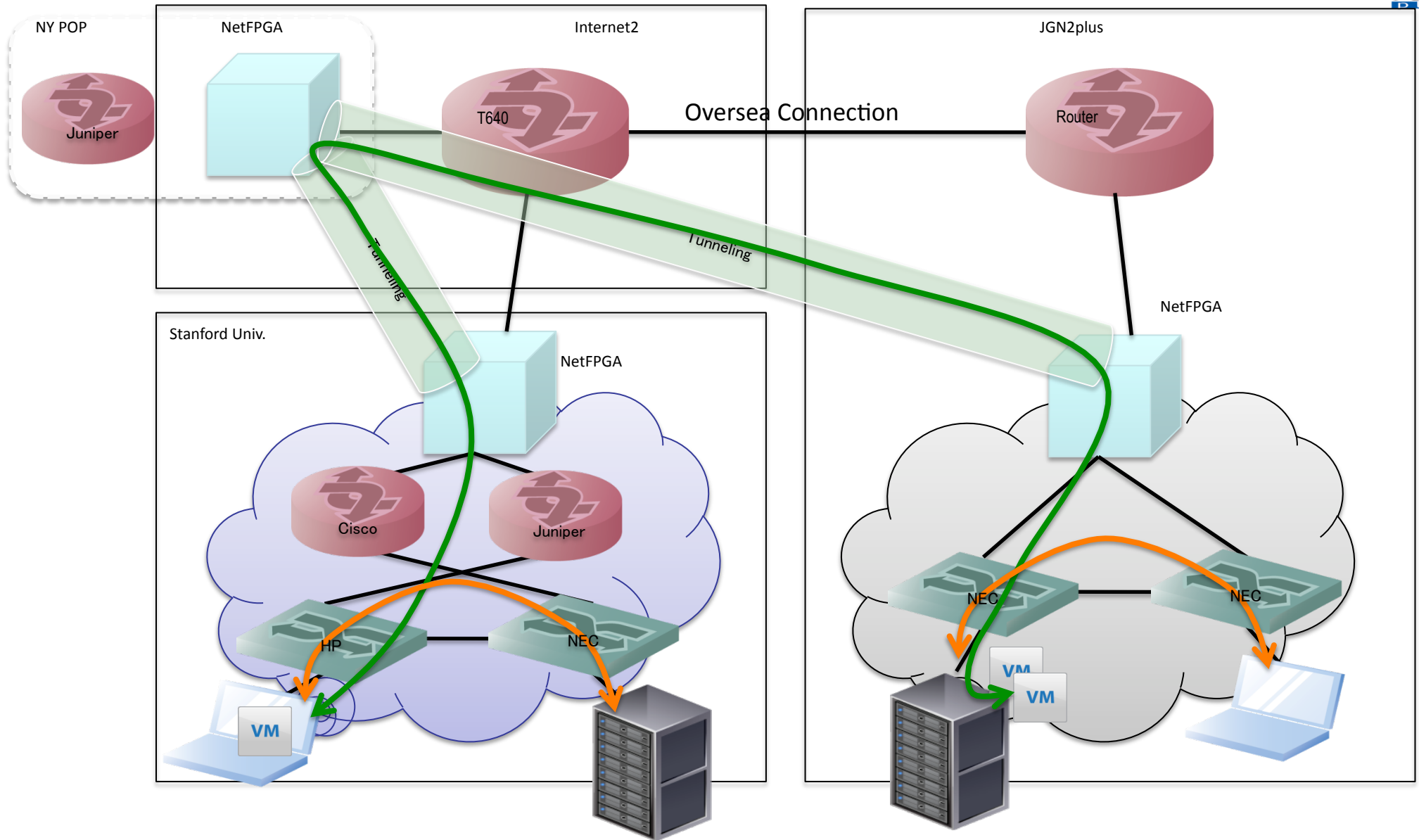
Community-Powered Cyberinfrastructure for Leading-Edge Research



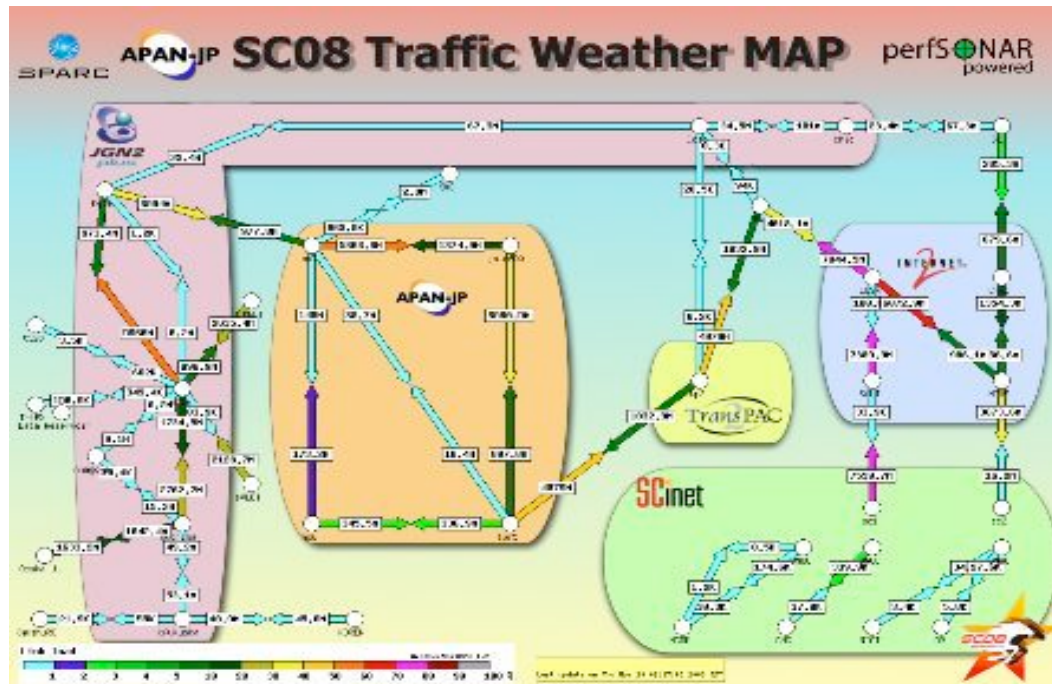
Federation of DCN



Openflow Demonstration at GEC



Courtesy of NEC



- perfSONAR tested at SC08, with great success
- perfSONAR placed at 6 locations in Japan now

DCN Live traffic usage by perfSONAR perfSONAR powered



Source Ikeda, Kitamura

Multiple service overlay on 2009 Sapporo Snow Festival

■ Live Broadcast by DTV(HD-SDI) Video Transfer System

- Measurement of how the broadcast system will affect the contents which is displayed on the DTV
- Measurement of the impact of video quality and latency (packet loss, delay, MTU size, jitter, etc) from the network's quality
- Experiment of the video quality by difference of network protocol (UDP/TCP/etc)
- Experiment of recovering time on failures
- Broadcast from virtual storage system (DELL2950-EMC CX4)

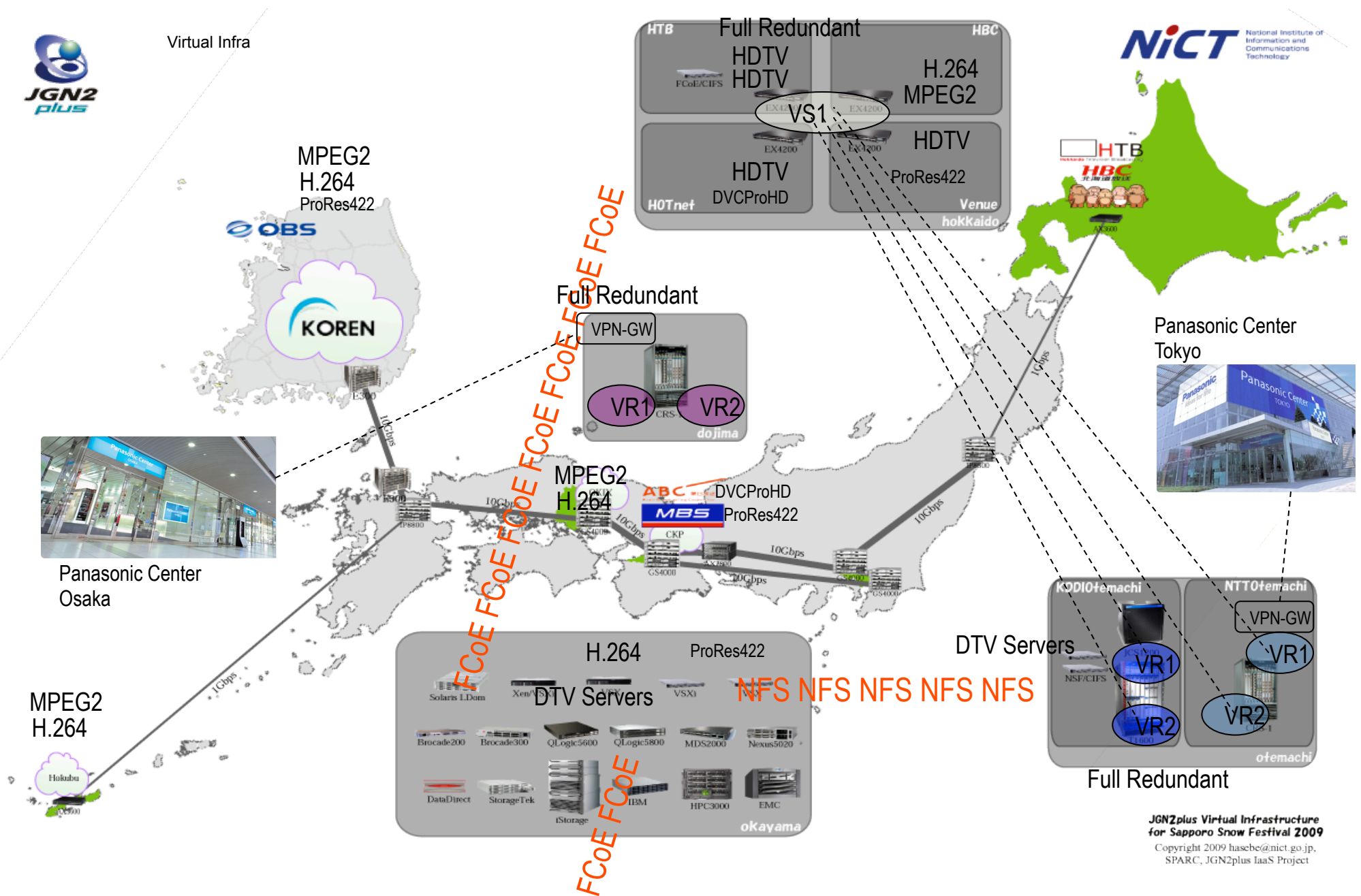
2/5 HD live broadcast from snow festival (HBC, Korea OBS)
2/9~11 Japan baseball Tigers camp from Nago MM HD(MBS)
2/21or22? SD live broadcast from Nago Stadium (Gaora)
2/late Osaka Obayashi Building Live Camera (MBS)

■ Local One-Segment TV based on IP Network Experiment

- Using IPv6 Multicast to send local one-segment TV content to experiment its quality, and to measure video broadcast via virtual network
- Broadcast from virtual storage system (SUN T5240 – DataDirect)



Multiple service overlay for Snow Festival



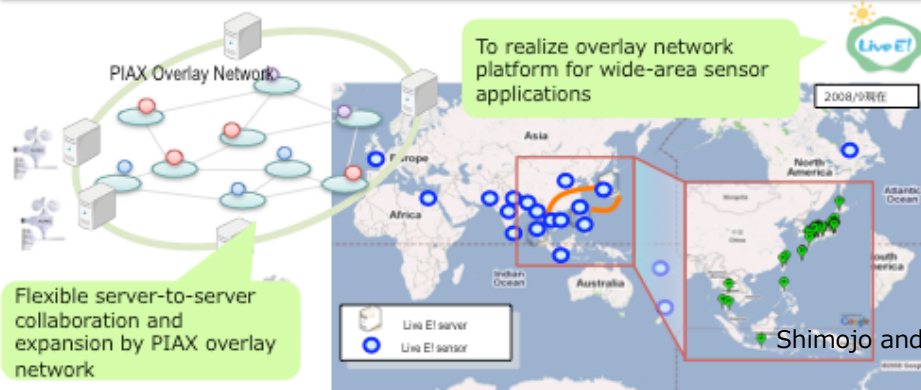
Overlay Platforms



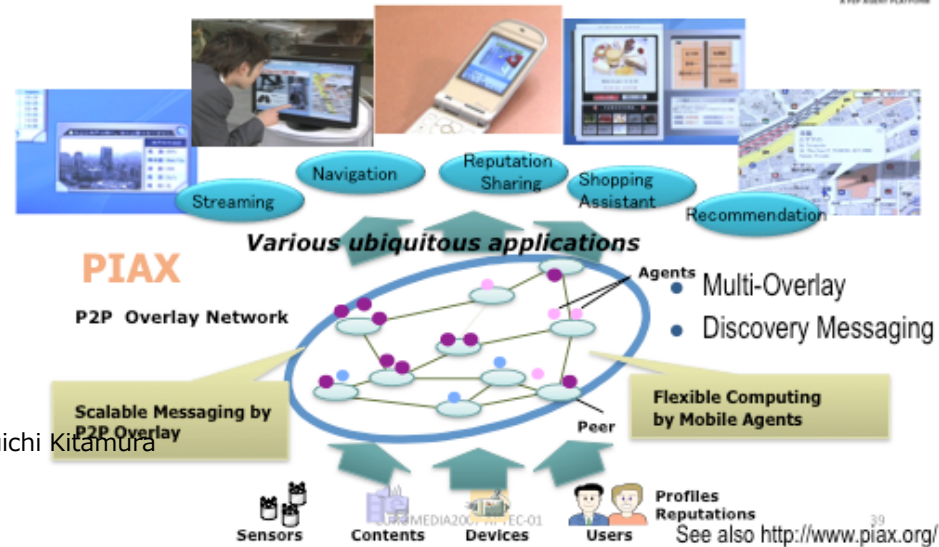
Live E!/PIAX Collaboration and Expansion



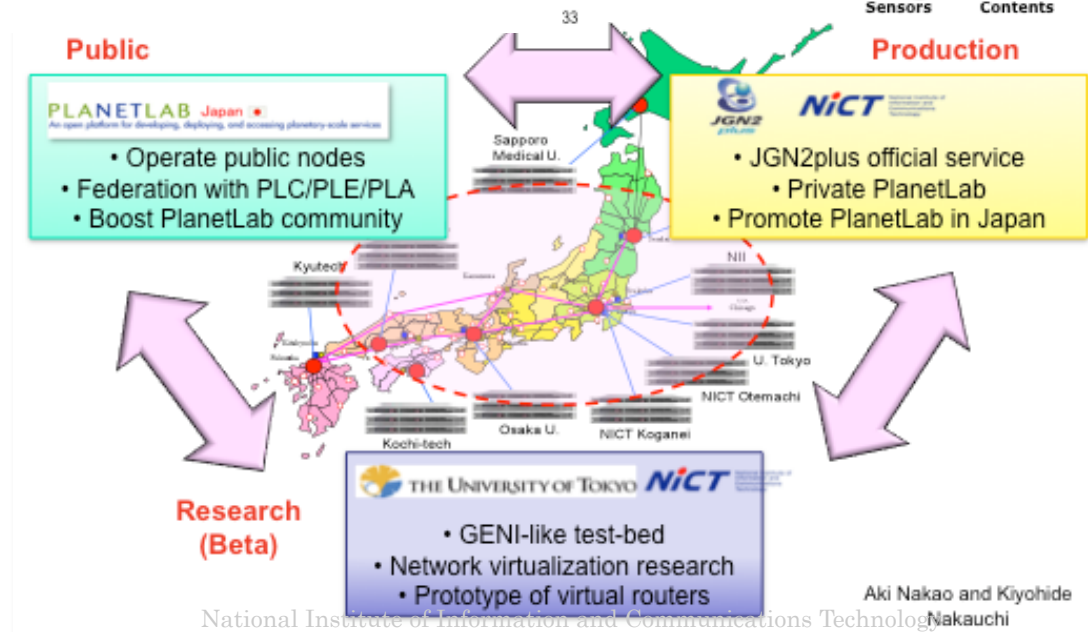
- Distributed servers of LiveE! project added with PIAX agent search plane to collaborate with PIAX overlay network
- Evaluation and Development of overlay network platform for wide-area sensor applications



PIAX: A P2P mobile Agent Platform



Shimojo and Yasuichi Kitamura



- Virtualization of network, storage, and computer is a key for having multiple experiments.
- And its federation is important
- But, it is a nightmare for network operator
- We are moving to service platform provider not just as a network provider
- We should have sophisticated management tools along with virtualization technology.

- Interoperability and federation is sometime difficult.
- In the field of virtualization and control plane, there will be many way and on going.
 - Vmware/hypervisor, Xen, Vserver, etc
 - DCN, MANTICORE, G-lambda, GMPLS, MPLS, etc.
- I like the way try first, then decide
- How much flexibility we can have while keeping integrity ?