Backbone Node: Requirements and Architecture

GDD-06-26

GENI: Global Environment for Network Innovations

November 1, 2006

Status: Draft (Version 1.0)

Note to the reader: this document is a work in progress and continues to evolve rapidly. Certain aspects of the GENI architecture are not yet addressed at all, and, for those aspects that are addressed here, a number of unresolved issues are identified in the text. Further, due to the active development and editing process, some portions of the document may be logically inconsistent with others.

This document is prepared by the Backbone Node Hardware Working Group.

Editors:

Dan Blumenthal, *University of California, Santa Barbara*

Nick McKeown, *Stanford University*

Contributing workgroup members:

Member 1, *Affiliation*

We'd also like to acknowledge comments and suggestions from Paul Morton and members of the GENI Optical Network Subcommittee.

This work is supported in part by NSF grants CNS-0540815 and CNS-0631422.

Revision History:

Version	Changes log	Date
0.1	Document Outline	4/4/2006
0.2	Transfer of initial draft into this format	4/4/2006
0.5	Substantive writing of document	7/27/2006
0.6	Corrected typos after daylong meeting at UCSB	7/27/2006
0.7	Restructured & assigned ownership to sections	9/7/2006
0.75	Expanded Section 3	9/27/2006
0.76	Reorder and re-definition of sections 4-7	10/25/2006
0.77	New Section 7; minor edits throughout.	10/25/2006
0.78	Auto figure numbering inserted. Section 7 updated and reduced. Material moved to new document.	10/26/06
0.90	1st release to limited audience. (Formatting in GENI style to happen in next release).	10/27/06
1.0	1st Formatting in GENI style. Addressed questions and comments from Jen Rexford and Paul Morton. 1st release to broader audience.	11/1/06

Table of Contents

1 Executive Summary

This document describes the high-level functional requirements of the GENI Backbone node, and the architecture. In this first draft, the basic subsystems and their requirements are described. The purpose of this document is to provide a high-level description of the backbone optical node and link to specific documents for each subsystem. At the highest level, the document describes how a node enables the data-rate to be sliced, how multiple network architectures can be supported by different experiments, and how the node supports multiple communication capabilities at different layers. The subsystems include a Packet Processing System (aka Programmable Router), Programmable Framer, Fast Circuit Switch, Wavelength Selective Switch and photonic switch.

2 GENI Backbone Network Requirements

The GENI Backbone Network is designed to satisfy the requirement that researchers can build and run an experiment over a *slice* of the backbone network with maximum flexibility, but without affecting other simultaneous experiments. In general, an experimenter's slice will consist of more than the backbone network – it will include the access network, tail circuits and end-hosts. GENI slices are described in "Overview of the GENI Architecture." Larry Peterson (Ed.) [G0611]. This document explains the requirements and architecture of the backbone network so that it can support slices in a manner consistent with the GENI requirements and other components of GENI connected to the backbone.

In the backbone network, a slice consists of a topology of links (virtual or physical). To explain what a slice is in the backbone network, we'll start with some illustrative examples; then, we'll define slices more precisely, to show the multi-layer flexibility that they offer the experimenter.

Our first simple example of a slice is illustrated in Figure 1. In this example, the experimenter has a dedicated router in each GENI backbone node. The router can be a process running on a general purpose CPU, an NPU, or some dedicated hardware. Packets arriving to a backbone node are de-multiplexed and directed to the experimenter's router. The routers are connected by a topology of links defined by the experimenter – in this example the experimenter is allocated a dedicated fraction of each physical link between nodes. In this way, the experimenter could create a network of IPv4 or IPv6 routers with standard features; or could create a new router that processes and routes packets in completely new ways.

Our second example shows how a slice's topology can be different from the topology of the backbone network. The topology of the slice in Figure 1 is a strict subset of the overall topology, and – in this example – is created by connecting a slice's incoming link to an outgoing link at a lower layer (for example, at an electrical or optical circuit-switched layer, just as an add-drop multiplexer does today). The slice's links can traverses multiple physical links between routers, allowing a large set of possible topologies.

It is a requirement that an experimenter's slice can include a mixture of stable and experimental layers (or "substrates") in its topology. Not only can the router be replaced by any element that processes packets, including elements that are far removed from what we call a router today;

the way that packets are framed in circuits can also be replaced by a new framing scheme, the circuits themselves can be static or dynamic, electrical or optical, and the optical transmission itself can, for example, be replaced by an experimental modulation or coding scheme. The backbone node design will allow network researchers to run experiments with any combination of stable (or non-stable if they choose) building blocks.

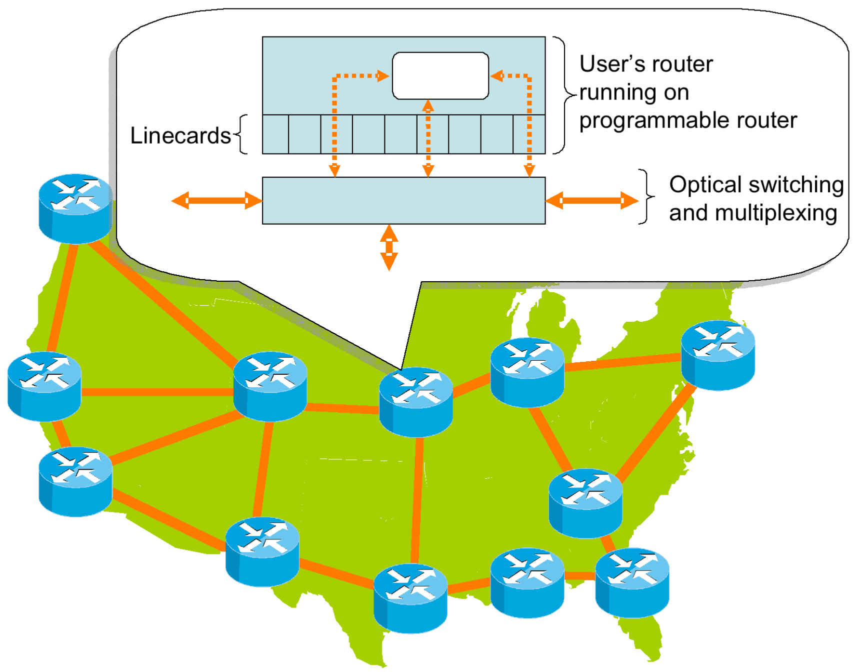

Figure 1. Topology of network showing links and users routers running on programmable routers.

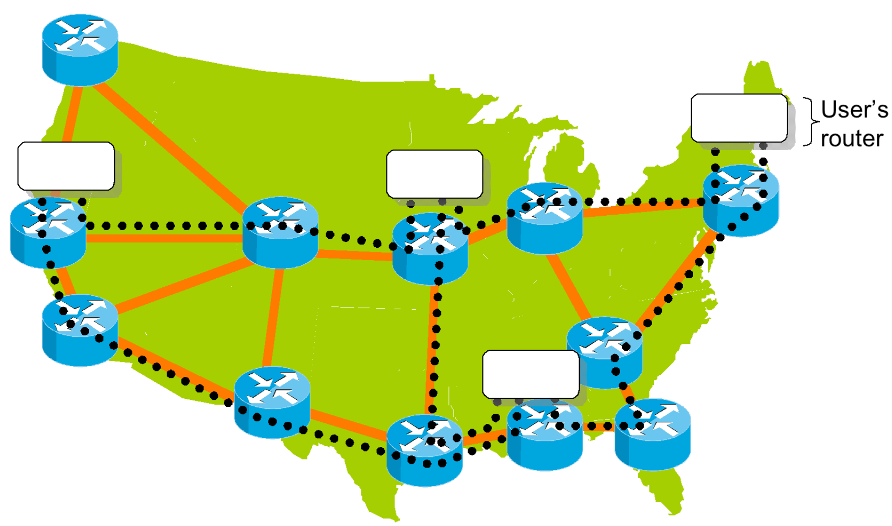

Figure 2. Network is configured so that the experimenter has a set of virtual links (shown by dashed lines) running over physical links (in solid lines) between a set of experimenter's routers.

In our next, and most detailed example, imagine an experiment of fast, dynamic optical circuit switching to create capacity-on-demand between a set of boundary routers. Experimental boundary routers create dynamic circuits on demand between each other, using experimental framing structures and experimental optical circuit switches. For example, a boundary router might monitor the amount of traffic destined to another boundary router. If demand increases, it might choose to rapidly deploy new capacity, within milliseconds. In our example, imagine that the network researcher is primarily interested in the network architecture and algorithms, but wants to exploit new technology that allows optical circuits to be switched at a granularity smaller than a wavelength. The researcher might build a slice that consists of experimental routers, over a stable (but novel and previously experimental) framing structure and fine-granularity circuit. In this manner, an experiment at one level can build on experiments and building blocks from other levels, allowing synergies and innovations in ways not conceivable today.

The GENI Backbone must also allow experiments at a higher level (for example, those supported by PlanetLab today) to exploit new technologies and substrates. For example, imagine an experimental content distribution network that exploits dedicated per-flow circuits to end-users. It might be accomplished by a new application-layer service that is able to communicate with, and exploit, a previously experimental, but now stable, fast circuit service.

We offer these examples only to describe some experiments that cannot be performed in the Internet today; we don't mean to argue the merits of a particular experiment, nor should it limit the scope of other experiments. We pick these examples in part because of our own limited imagination. GENI needs to support experiments that the broad community of researchers can and will think of, and not be limited to ones a small group of researchers can enumerate now. The key point is that GENI must allow innovation from the top down or the bottom up, but without sacrificing the reliability of a default substrate. In this manner, unknown results from innovation will have a high probability of evolving GENI into a network that is very different than what exists today.

It is important to state that allowing flexibility must not come at the expense of dependability for the researcher (i.e. stability of the substrate on which they are building their experiment on). At each layer, the network must provide a stable and reliable substrate using well-tested, stable technology. On the other hand, the network must allow experiments with new substrates to take place in parallel, without violating the requirements of the stable substrate that it might, eventually, influence, co-exist with, or replace.

2.1 Summary of Requirements to Support Multiple Layers of Research

The GENI Backbone Network must support research at multiple layers. Here, we categorize different experimenters and how they use might GENI.

Type 1. Network researchers who want a stable network of standard routers (e.g. IPv4, IPv6 routers with standard features) operating over a topology of their choosing, but using links that are statistically shared with other users and experiments.

Type 2. Network researchers who want a network of stable, standard routers (e.g. IPv4, IPv6 routers with standard features) operating over a topology with dedicated and private bandwidth (e.g. a topology of circuits), with stable (possibly standard/default) framing.

Type 3. Network researchers who want to deploy their own packet processing element and protocols in a private, or shared, slice; running over shared or dedicated bandwidth links within a topology. The experimenter has complete control of how data passes over the network (including framing and packet format).

Type 4. Network researchers who want specific bandwidths on demand within a topology. E.g. a topology with precise bandwidths between nodes, and where bandwidth can be setup and removed dynamically.

Type 5. Researchers who want access to raw optical wavelengths with no framing, protocol or transport constraints.

Type 6. Researchers who want access to raw fiber bandwidth. E.g. new transmission, modulation, coding and formats.

From day one, researchers of Types 1-3 must be able to run experiments stably and uninterrupted.

While it is important for GENI to enable experiments of Types 4-6, it needs to be done with care. On one hand, experiments of Types 4-6 should be supported so as to allow exploration of new technologies and substrates that might eventually become alternative stable substrates for experiments of Types 1-3. On the other hand, technology experiments should not be enabled at the expense of, or by preventing stable experiments of Types 1-3.

2.2 Concurrent Substrates

The GENI Backbone Network must support concurrent network substrates, where a substrate is defined as the set of stable underlying layers used by an experiment.

Concurrent substrates provide simultaneous access for researchers to investigate and develop new stable underlying technologies (e.g. protocols, techniques, subsystems, components) for lower layers without affecting the stability of existing stable substrates. This notion of concurrent substrates allows evolution over time as GENI is deployed in stages. An approach to staging concurrent substrates in GENI is shown in the example in Figure 3. In this example the early stage of GENI backbone node deployment is shown on the left with the grey area indicating the stable substrate and the white area indicating the experimental research that can operate on top of that substrate. The numbers used for project phase and network layer are arbitrary (not intended to be linked to any plan or model like the OSI model). The key to this model, it that in the next stage, where the stable substrate sits one layer lower, the old stable substrate is not taken down, hence the two are decoupled and exist at the same time. At the end of the GENI backbone node construction cycle the substrate exists as simultaneous, decoupled substrates shown as the grey staircase.

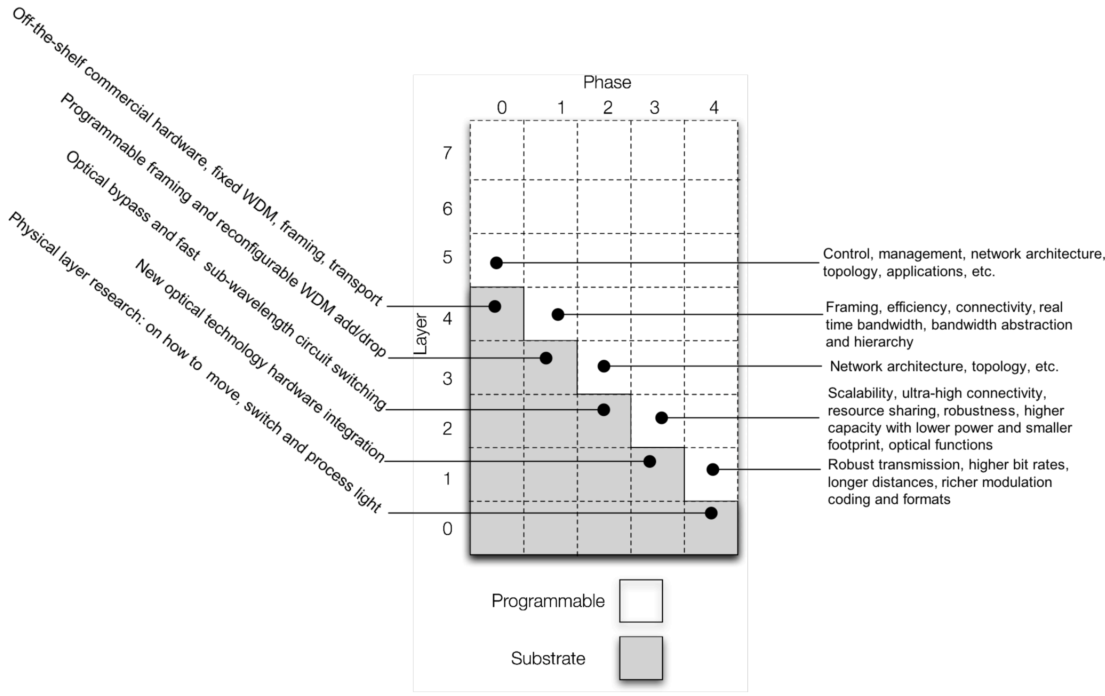

Figure 3. An example of concurrent substrates (grey) and overlay research experiments (white).

2.3 Examples of data flow

1. How two packet processors communicate over Gigabit Ethernet. One packet processor creates a 1GE packet; the framer packs it into a null frame, which is part of a 1 gigabit TDM circuit that is time division, multiplexed onto a 10Gb/s wavelength.

2. Loopback at any layer: Data entering on a wavelength can be switched directly to another wavelength (without passing through any other layers) or can be switched at the TDM layer, or passed through the packet processor for packet level processing.

3. Connections to outside world: These can take place at any layer. For example, a non-packet processor node (e.g. commercial router, or another network) can connect directly to GENI through the programmable framer.

3 Definition and Straw Design of the GENI Backbone Node

The GENI Backbone Node will be deployed in several stages so as to provide stable substrates for researchers at the various network layers, even while the node evolves. The initial stage (Stage 1) will make maximum use of existing commercial products, allowing experiments to start as soon as possible. **Error! Reference source not found.** Shows the five primary layers of the backbone node architecture. An example of the GENI backbone node that meets these requirements is shown in Figure 5.

1. The Fiber Switch (FS)

2. The Wavelength Selective Switch (WSS)
3. The Fast Circuit Switch (FCS)
4. The Programmable Framer (PF)
5. The Packet Processor (PP)

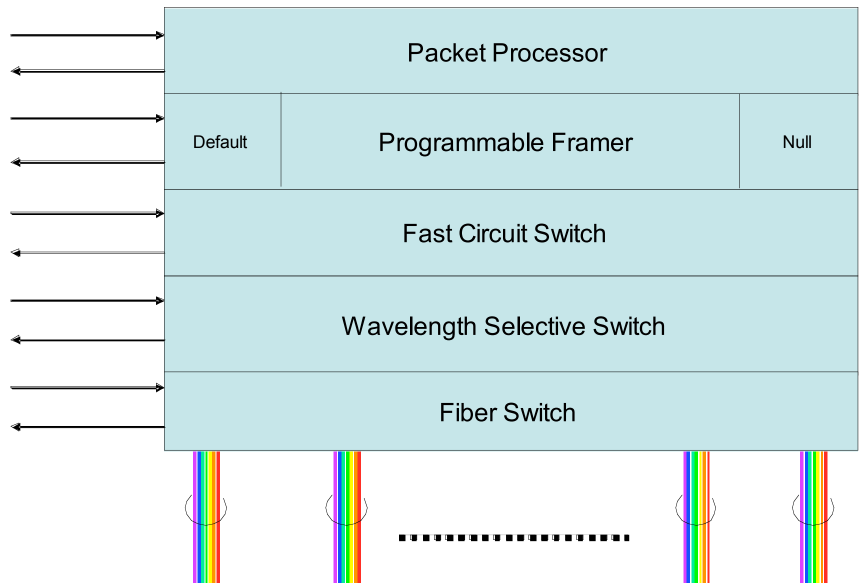

Figure 4. The five primary layers of the GENI backbone node.

Fiber Switch (FS): This switch interconnects raw fibers from network input to network output without any modification of wavelengths or data within the fibers. Fibers in the plant can be physically allocated to the primary GENI nodes, next generation GENI nodes scheduled for hitless rollover, and fibers used for fiber optical physical layer research. The FS provides complete isolation at all levels between the running GENI infrastructure and future GENI nodes and technologies.

Wavelength Selective Switch (WSS): Data on one 10Gb/s wavelength can be switched to another wavelength, or delivered to the Fast Circuit Switch. Data carried on a wavelength is not interpreted in any way by the WSS. The WSS will connect to the FS via optical fiber. User research equipment can connect directly to the WSS (shown on the left hand side of the figure) using spare optical ports on the WSS.

Fast Circuit Switch (FCS): Circuits are time-division-multiplexed onto a 10Gb/s wavelength. The minimum granularity will be 1Mb/s, and the circuit switch should be capable of establishing a new circuit in less than 1ms. Initially, the FCS will be electronic. Over time, optical FCS's are possible. Virtual circuits of any bandwidth with granularity of 1Mb/s can be established. Individual circuits can be assembled from multiple basic slots within and across wavelengths. The FCS will connect to the WSS via optical fiber. User research equipment can connect directly to the FCS (shown on the left hand side of the figure) using spare electrical ports on the FCS.

Programmable Framer (PF): The framer will frame packets inside circuits. SONET is one obvious choice for a default framing format. The framer should have a null framing format in cases where the packets themselves carry sufficient information for recovery at the destination. The PF will connect to the FCS via electrical links or short-reach optics. User research equipment can connect directly to the PF (shown on the left hand side of the figure) using spare ports on the PF.

Packet Processor (PP): This is what is frequently referred to as the Programmable Router [G0609]. It can process packets in any way it chooses, at layers higher than we traditionally think of as routing. E.g. it could transcode packets from one format to another, or do deep application level packet inspection, or even terminate flows as an end-point. The PP will connect to the PF via electrical links or short-reach optics. It is likely that the PP and PF will be co-located in the same piece of programmable equipment.

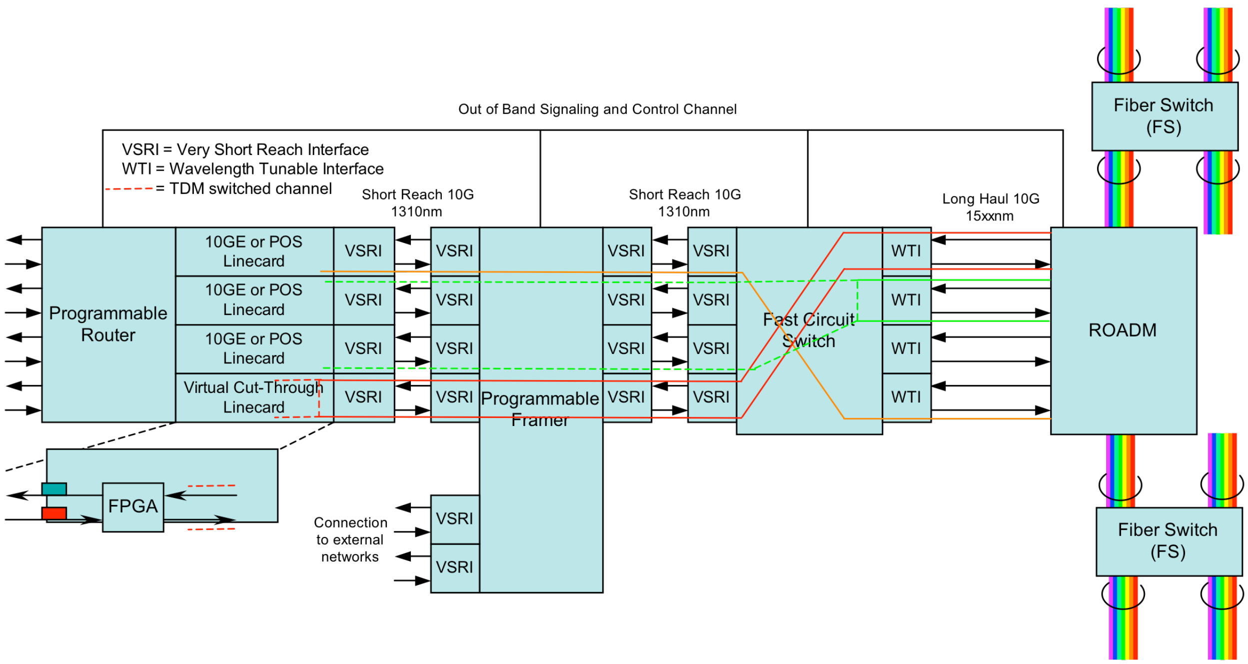

Figure 5. Straw example of GENI backbone node subsystems.

3.1 Packaging

The five layers represent logical functional blocks, and will not necessarily be implemented as five distinct pieces of equipment. The implementation and packaging of the five layers will be determined by the commercial equipment available at the time of first deployment – and will depend on the details of the staged deployment. For example, the programmable framer and packet processor might be packaged together, such as the ATCA chassis described in [G0609]; and the fiber switch and fast circuit switch might be packaged together as a dynamic optical circuit switch.

3.2 Layers and Experiments

Different experiments will use each layer in different ways. Examples include:

- The researcher uses the packet processor to directly communicate over an optical wavelength using proven commercial equipment. For example, a researcher can use one packet processor to creates a 1GE packet that is multiplexed and framed using an off-the-shelf SONET card. Alternatively, the 10GE router blade might encapsulate the packets in SONET in which case the framer packs it into a null frame. The fast circuit switch connects the SONET streams to the required 10 Gb/s wavelength that is added to the physical link by the wavelength selective switch.

- The node provides the unique function of loopback at any layer. Data entering on a wavelength can be switched directly to another wavelength (without passing through any other layers) or can be switched at the TDM layer where only a portion of a wavelength is passed to the framer or packet processor and the remainder of the wavelength is added directly back to the network, with the possibility of new data added to the empty TDM frame. Or a complete wavelength can be passed through the circuit switch through the programmable framer (or the null or default) directly to the packet processor for packet level processing. One of the blades in the packet processor itself could provide the function of loopback. So the research has control of loopback and fast circuit switching functions at any layer in the node.

- The Programmable framer in addition to providing default (SONET) and null functions also allows new framing protocols to be implemented in FPGA blades. Researchers who want to study the performance improvements of new transport protocols can utilize the programmability at this layer. Additionally, researchers on other networks may want to connect to GENI and the FPGAs can be used as protocol translators that link the two networks (similar to MPLS). As these new framing protocols become stable researchers may choose to use these new protocols for their own experiments.

- The Fast Circuit Switch enables the researcher to access sub-wavelength granularity, or for bandwidth to be abstracted to the researcher in near arbitrary multiples of a basic unit (e.g. 10Mbps).

- Connections to outside world: These can take place at any layer. For example, a non-packet processor node (e.g. commercial router, or another network) can connect directly to GENI through the programmable framer.

4 Meeting the requirements of different user communities

There are multiple communities of users who will use, connect to, and/or run experiments on GENI. Within the community of GENI researchers, we can broadly identify two sub-communities: (1) The networking research community – typified by Types 1-3 above – who will need a stable substrate over which to perform their experiments; and (2) The networking physical layer research community – typified by Types 4-6 above – who will invent and explore new ways to provide future stable substrates for experiments of Types 1-3. In this manner, GENI can continue to evolve and take advantage of advances in the lower layers.

It is natural to ask how the requirements of both communities can be met simultaneously. On one hand, the GENI Backbone Node must provide access to the layers needed for lower layer research, while on the other hand not jeopardizing (by complexity or cost) the layers needed for networking research.

We make it a requirement that both goals be met simultaneously, by providing a layered set of functional blocks inside the GENI Backbone Node, starting from packet processing at the highest layer, and fiber plant at the lowest layer. For every layer there will be a set of stable substrates, including a default, upon which experiments can run at a higher layer. Yet at the same time, a stable substrate must be able run alongside experimental and unstable substrates.

Different researchers will connect in to the GENI Backbone Node at different layers, and in some cases might provide their own equipment that operates on top of that layer. For example, if a network researcher wants to introduce a new router, they can do so by connecting to the Programmable Framer. This is necessary so as to allow GENI to evolve beyond the initial deployment of programmable routers. Similarly, an optical network researcher who wishes to perform experiments on top of a raw fiber – for example to experiment with new coding and multiplexing schemes – can attach lab equipment directly to a terminated fiber at the Fiber Switch.

Other communities will connect to the GENI network, including commercial operators and other experimental networks. While some will connect by routing packets using standard framing formats, others will need conversion between framing formats using the Programmable Framer. The fiber switch will allow these communities to run over the same physical footprint over a separate fiber plant so there is no interference between experimental physical networks and the operation of GENI. The node hardware for these experiments will be located in laboratories or facilities distinct from GENI nodes, and the users will connect to the GENI footprint via their own fiber connected the GENI fiber switch. New device, transmission, subsystem and other technologies can be explored over the GENI footprint making future migration of these technologies into GENI more likely to occur. An example experimental network is shown in Figure 6.

5 Staged Deployment of GENI Backbone Node

The GENI Backbone Node will be deployed in several stages so as to provide stable substrates for researchers at the various network layers, even while the node evolves. The initial stage will make maximum use of existing commercial products, allowing experiments to start as soon as possible, including management and control software. The work breakdown structure (WBS) document, not repeated here, describes in more detail the staged deployment. The WBS uses terminology that combines the FS, WSS and FCS, referred to as the Dynamic Optical Switch (D-OS). The plan is to procure and deploy five (5) D-OSs during the first year of GENI construction, as well as procurement, construction and deployment of 17 D-OS switches by the end of year 4, and control software procurement and deployment by the end of year 5.

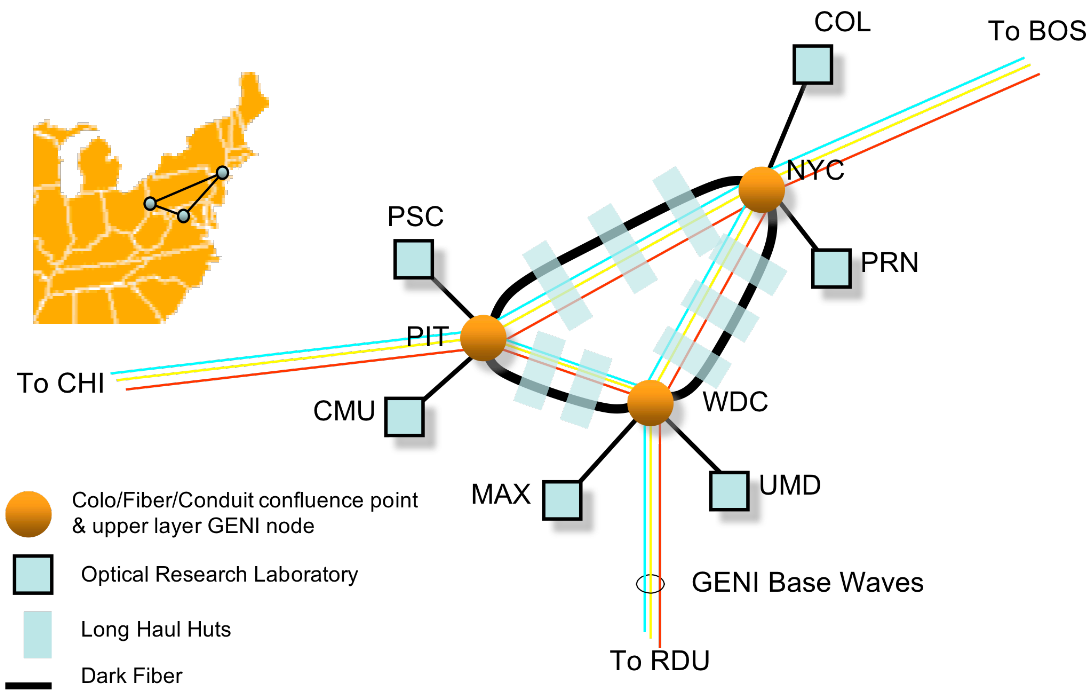

Figure 6. An example experimental network operating independent from GENI but over the GENI footprint using auxiliary ports on the node fiber switches.

6 Fiber Plant

Prior to deployment, it will be necessary to specify the Physical Medium Dependent requirements for the optical fiber plant, including the type of fiber, wavelengths, power levels, jitter specification, connector types, etc. These are expected to follow best practice at the time of deployment, or be dictated by an existing fiber infrastructure, and so do not need to drive the architectural requirements at this stage. Each wave will be running at 10Gbps, so current commercial transmission equipment will be utilized. However, the use of cascaded ROADMs in the network raises issues related to transmission engineering for various path configurations. This issue of dynamic transmission path engineering will have to be addressed in the backbone design. An example nationwide GENI fiber infrastructure with GENI nodes is show in Figure 7.

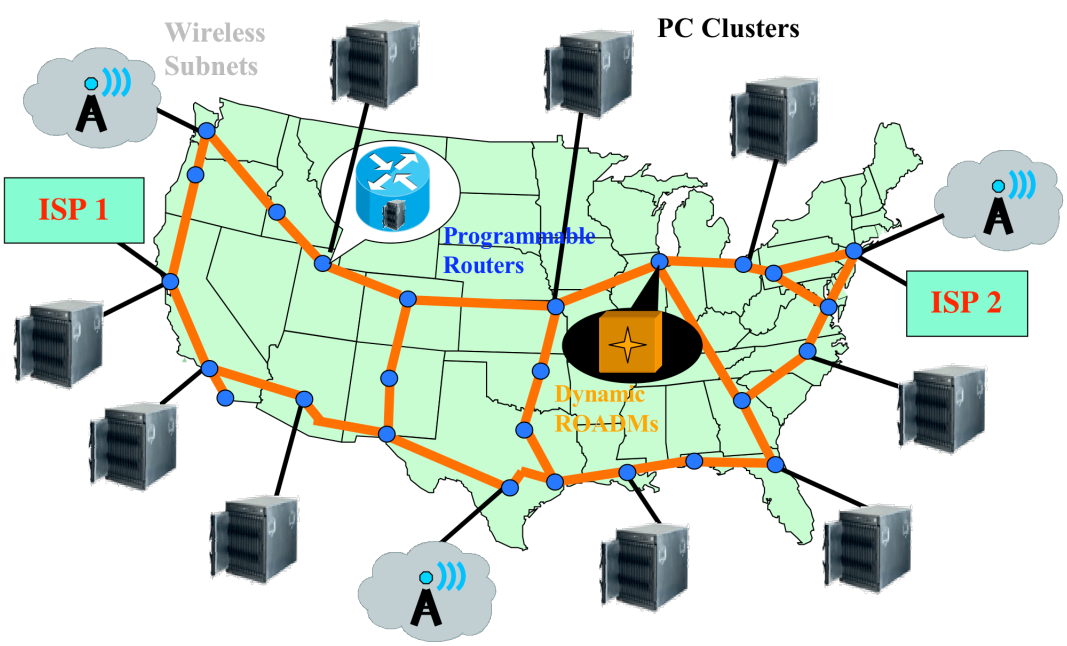

Figure 7. Example nation-wide WDM network built with GENI backbone nodes based on ROADM and fiber switch technology.

In a separate document on the high level requirements for the backbone network that utilizes the nodes described in this document, the following issues will be addressed:

- The backbone optical design starting at the network level, showing the different nodes and interconnections of the nationwide network, including capacities for links and key components of the design.
- Key requirements for the WDM terminals to show where there are WDM signals and where there are short reach (e.g. 1310nm) interfaces, and how the overall network fits together.
- Embedded in the network description will be reference to this document in order to highlight the node design. WDM terminals within a node in that description will also be included in the backbone node HLR.
- Transmission requirements for access to WDM signals over separate fiber plants will be described in further detail.
- The issue of 'span design' that occurs when moving from point to point transmission to an all-optical nationwide mesh network.
- The issue of engineering each point-to-point connection, which occurs in current networks, expands to engineering for all potential routes that each individual wavelength can take within a large mesh network. This in itself is an optical networking research project, which can start with a basic network infrastructure

and be expanded as research progress on the experimental fiber on the GENI footprint is advanced.

- How the ROADMs and fiber switches in the backbone node will be utilized. For example, will the network be constructed out of rings and meshes virtually configured, or will there be actual optical meshes in the backbone design.

7 Mechanical layout

(The following information in the next 3 sections is placeholders – to be determined during deployment; and might not belong in this document).

23″ Shelf 22.7″ x 21.3″ x 11.5″

19″ Shelf 22.7″ x 18.5″ x 11.5″

Weight

23″ Shelf <186 lb

19″ Shelf <160 lb

Power input -48V DC

8 Power consumption

Dual redundant -48VDC connections are supplied to each F

Optical switch/hub shelf <325W (1108 BTU/hr)

OADM shelf <240W (820 BTU/hr)

Transponder shelf 23″ <440W (1500 BTU/hr)

Fujitsu Flashwave 7500 (example)

9 Environmental considerations

Temperature 0 -50C (32-122F)

Humidity 5 – 95% (non condensating)

NEBS Level 3 compliant

References

[G0611] Larry Peterson, John Wroclawski (Eds), "**Overview of the GENI Architecture**," *GENI Design Document 06-11*, Facility Architecture Working Group, September 2006.

[G0609] Jonathan Turner, "**A Proposed Architecture for the GENI Backbone Platform**," *GENI Design Document 06-09*, March 2006.