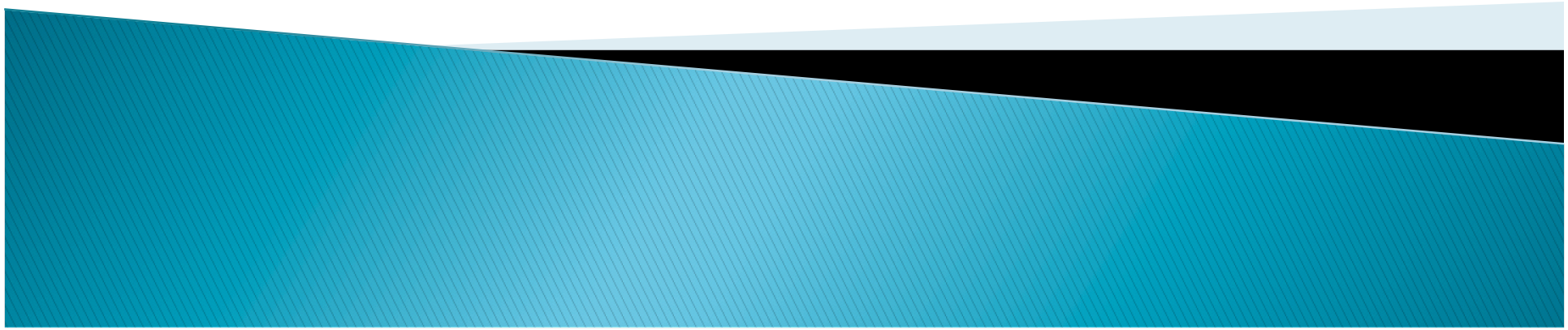
ExoGENI Racks status update

Ilia Baldine ibaldin@renci.org
Chris Heermann ckh@renci.org
Jonathan Mills jonmills@renci.org
Victor Orlikowski vjo@cs.duke.edu



ExoGENI Overview

ExoGENI Testbed

- 14 GPO-funded racks
  - Partnership between RENCI, Duke and IBM
  - IBM x3650 M4 servers (X-series 2U)
    - 1x146GB 10K SAS hard drive +1x500GB secondary drive
    - 48G RAM 1333Mhz
    - Dual-socket 8-core CPU
    - Dual 1Gbps adapter (management network)
    - 10G dual-port Chelseo adapter (dataplane)
  - BNT 8264 10G/40G OpenFlow switch
  - DS3512 6TB sliverable storage
    - iSCSI interface for head node image storage as well as experimenter slivering
- Each rack is a small networked cloud
  - OpenStack-based
  - EC2 node sizes (m1.small, m1.large etc)
- http://www.exogeni.net

ExoGENI Status

- 3 racks deployed
  - RENCI, GPO and NICTA
- 2 existing racks (not yet OpenFlow enabled)
  - Duke and UNC
- 2 more racks available by GEC14
  - FIU and UH
- Connected via BEN (http://ben.renci.org), LEARN and NLR FrameNet
- Partner racks
  - U of Alaska Fairbanks

Since GEC13

- Added ability to provision baremetal nodes
  - Via xCAT
- Added ability to attach VMs or baremetal nodes to mesoscale VLANs available at rack sites
- Monitoring channel from Nagios to GMOC
- Improved GENI AM API implementation with help from GPO
- Still Working on
  - Acceptance tests
  - Bringing ORCA closer to GENI AM API compliance (primarily RSpecs)
  - Providing slice information to GMOC to have correlate resource provisioning with slices/users/projects

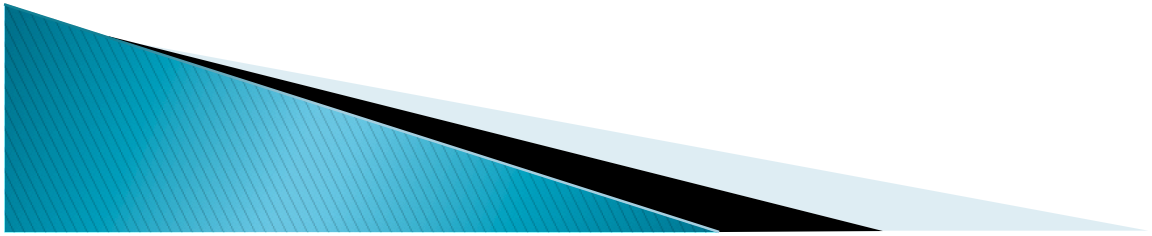
ExoGENI unique features

- ExoGENI racks are separate aggregates but also act as a single aggregate
  - Transparent stitching of resources from multiple racks
- ExoGENI racks are targeted at computing applications as well as experimentation
  - Already running HPC workflows linked to OSG and national supercomputers
  - Strong performance isolation is one of key goals
- A model for deeply reconfigurable federated compute/storage/network infrastructure for campuses and labs.

ExoGENI details

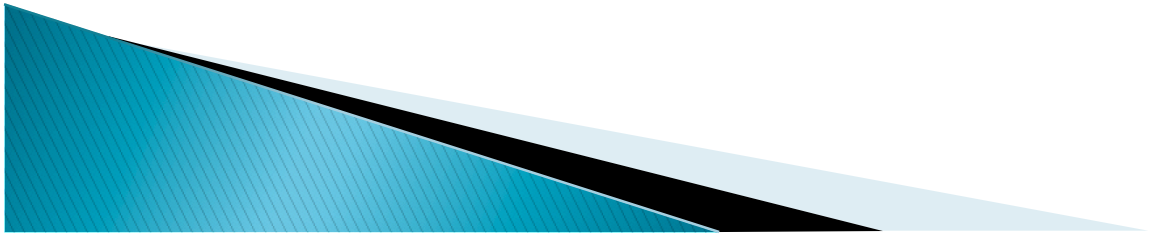
Rack Connectivity

- Rack has a management connection to campus network
- A connection to FrameNet or I2 ION
  - Direct peering
  - Via a pool of vlans with static tags through a regional
- It may have an optional connection to the OpenFlow campus network for experiments

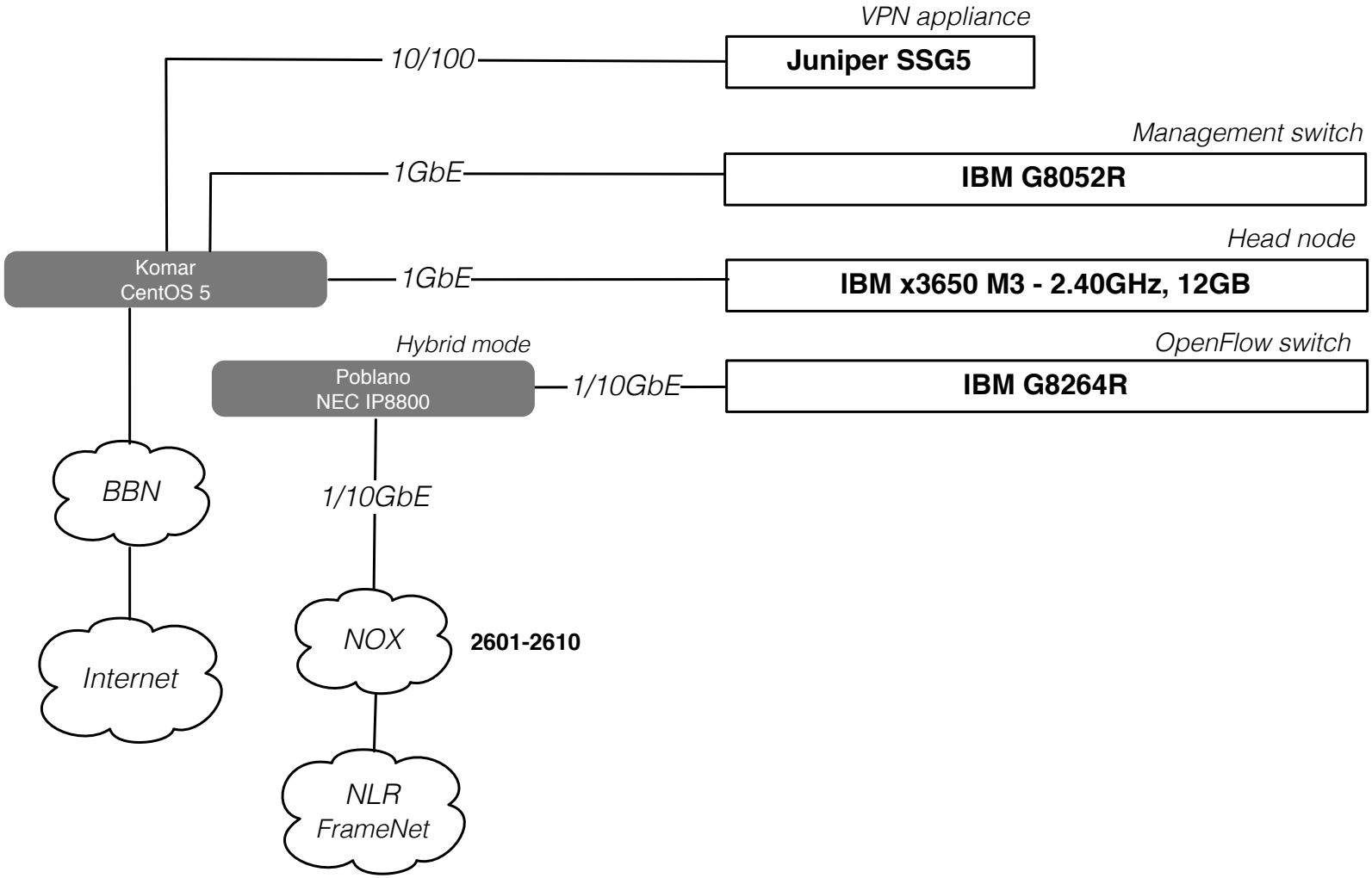


Rack IP address assignment

- /24 of publicly routable IP addresses is the best choice
- 2 are assigned to elements of the rack
  - Management/Head node
  - SSG5 VPN appliance (to create a secure mesh for management access between racks)
- The rest is used to assign IP addresses to experimenter instances
  - VMs and hardware nodes



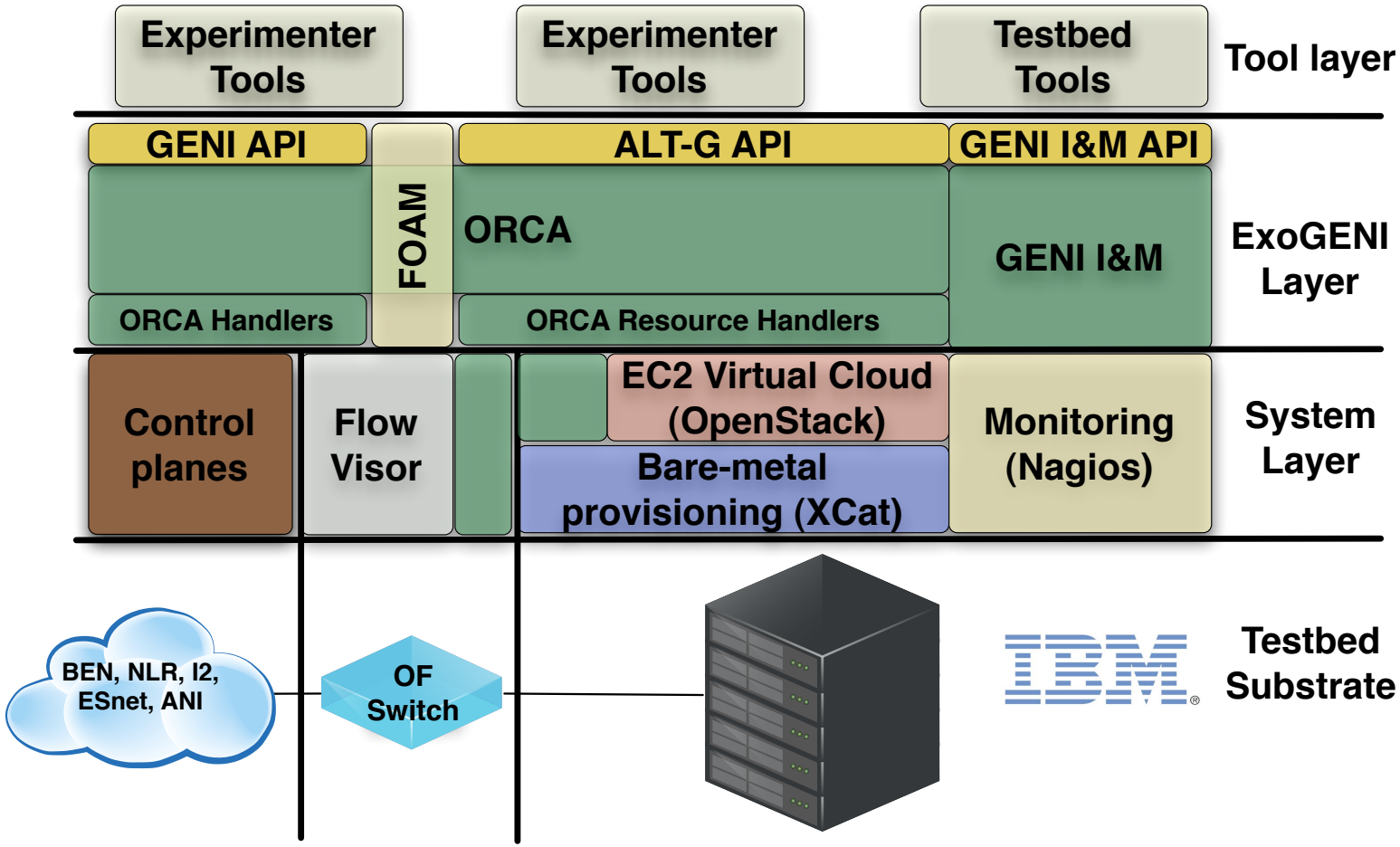
Example rack connection (GPO/ BBN)



Rack software

- CentOS 6.X base install
- Resource Provisioning
  - xCAT for bare metal provisioning
  - OpenStack + NEuca for VMs
  - FlowVisor
    - NOX used internally by ORCA
- GENI Software
  - ORCA for VM, baremetal and OpenFlow
  - FOAM for OpenFlow experiments
- Worker and head nodes can be reinstalled remotely via IPMI + KickStart
- Monitoring via Nagios (Check_MK)
  - ExoGENI ops staff can monitor all racks
  - Site owners can monitor their own rack
- Syslogs collected centrally

Rack Software Stack



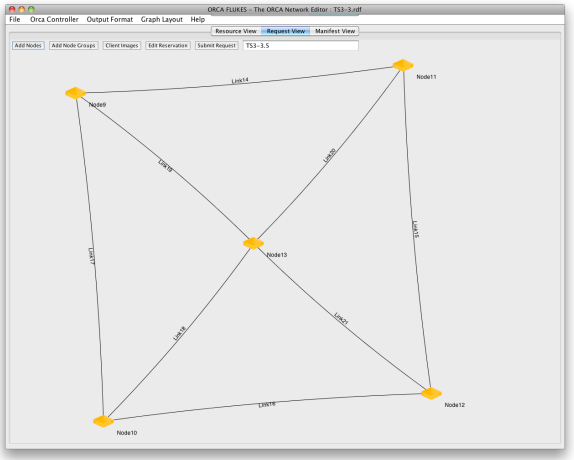
Rack installation

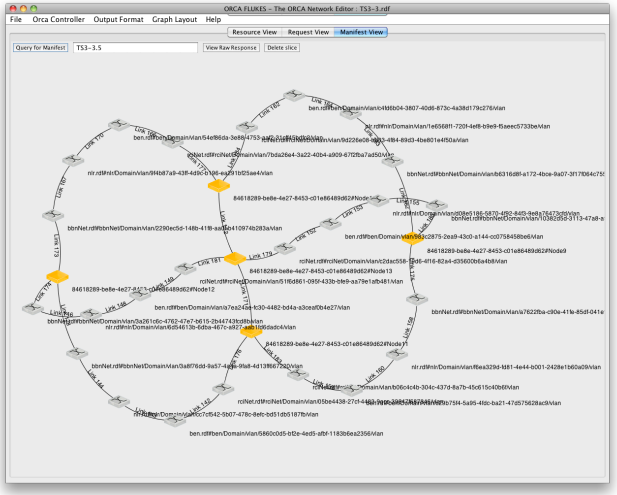
- Particulars:
  - Power options include (negotiated ahead of time)
    - 208V 3Phase
    - 208V 1Phase
    - 120V 1Phase
    - Total of ~10kW of power needed.
  - Space:
    - e1350 42U Rack Cabinet 79.5" H x 25.5" W x 43.5" D (2020 mm x 648 mm x 1105 mm)
- Racks arrive on-site pre-assembled and pre-tested by IBM with most software already pre-installed
  - IBM representative will need to come on-site to complete install and hookup
    - NBD hardware support
  - ExoGENI Ops finishes ORCA configuration
  - GPO acceptance testing

ExoGENI experiments

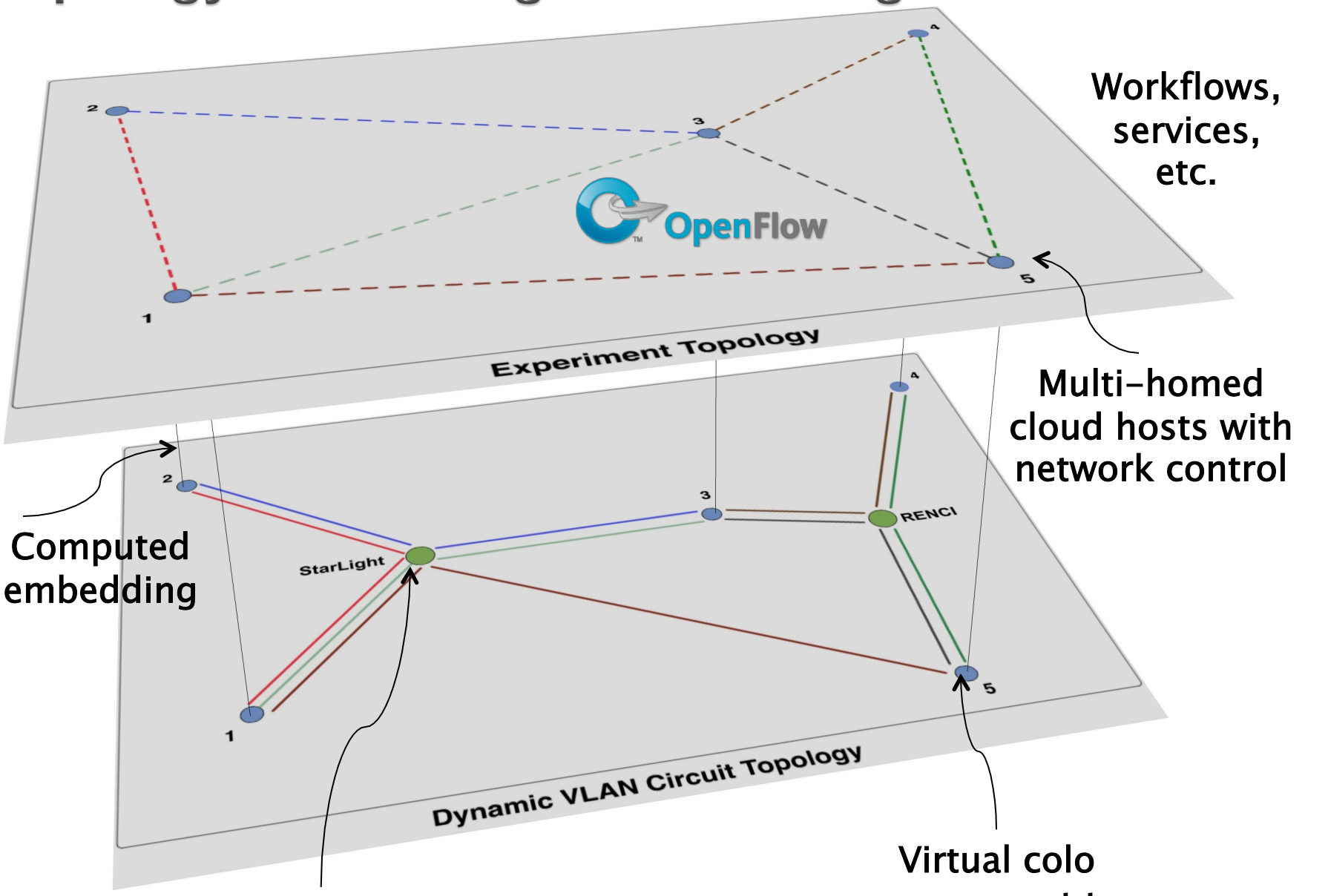
Experimentation

- Compute nodes
  - Up to 100 VMs in each full rack
  - A few (2) bare-metal nodes
- True Layer 2 slice topologies can be created
  - Within individual racks
  - Between racks
  - With automatic and user-specified resource binding and slice topology embedding
  - Dynamic circuit services (Sherpa, OSCARS) called where available
- OpenFlow experimentation
  - Within racks
  - Between racks
  - Include OpenFlow overlays in NLR (and I2)
  - On-ramp to campus OpenFlow network (if available)
- Experimenters are allowed and encouraged to use their own virtual appliance images



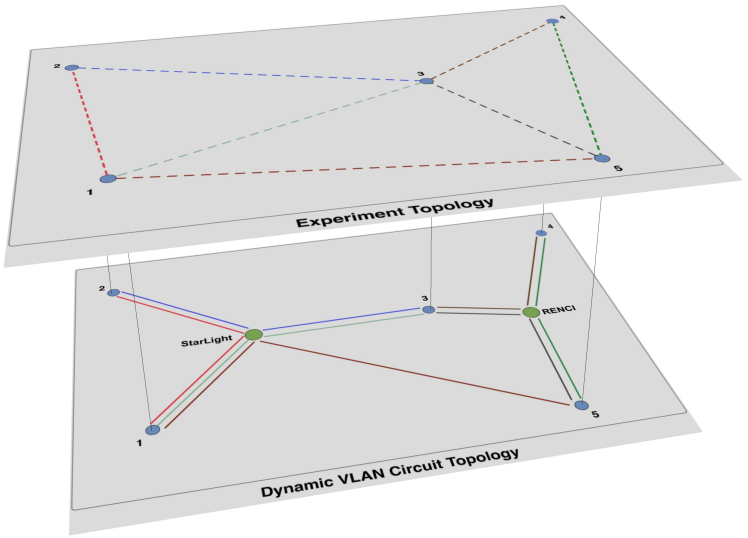


Topology embedding and stitching



ExoGENI slice isolation

- Strong isolation is the goal
- Compute instances are KVM based and get a dedicated number of cores
- VLANs are the basis of connectivity
  - VLANs can be best effort or bandwidth-provisioned (within and between racks)



ORCA control framework

ORCA Overview

- Originally developed by Jeff Chase and his students at Duke
- Funded as Control Framework Candidate for GENI
  - Jointly developed by RENCI and Duke for GENI since 2008.
- Supported under several current NSF and DOE grants to enable ORCA to run computational networked clouds
- Fully distributed architecture
- Federated with GENI
- Supports ORCA-native interface, resource specification and tools
  - Flukes
- Supports GENI AM API and GENI Rspec
  - Omni

ORCA Deployment in ExoGENI

- Each rack runs its own ORCA actor (SM) that exposes
  - ORCA native API
  - GENI AM API
- Rack-local SM
  - Can only create slice topologies with resources within that rack
- 'ExoSM' has global visibility
  - Has access to resources in all racks
  - Has access to network backbone resources for stitching topologies between racks
- Uniquely ExoGENI racks act both as
  - Independent GENI aggregates
  - A collective aggregate with intelligent topology embedding and stitching via dynamic or static circuits (NLR, ESnet, I2)

The team

- Grand Pooh-bah – Jeff Chase
- ExoGENI Ops
  - Chris Heermann (RENCI) – rack networking design
  - Victor Orlikowski (Duke) – software packaging and configuration
  - Jonathan Mills (RENCI) – operations and monitoring
- ORCA Development
  - Yufeng Xin (RENCI)
  - Aydan Yumerefendi (Duke)
  - Anirban Mandal (RENCI)
  - Prateek Jaipuria (Duke)
  - Victor Orlikowski (Duke)
  - Paul Ruth (RENCI)