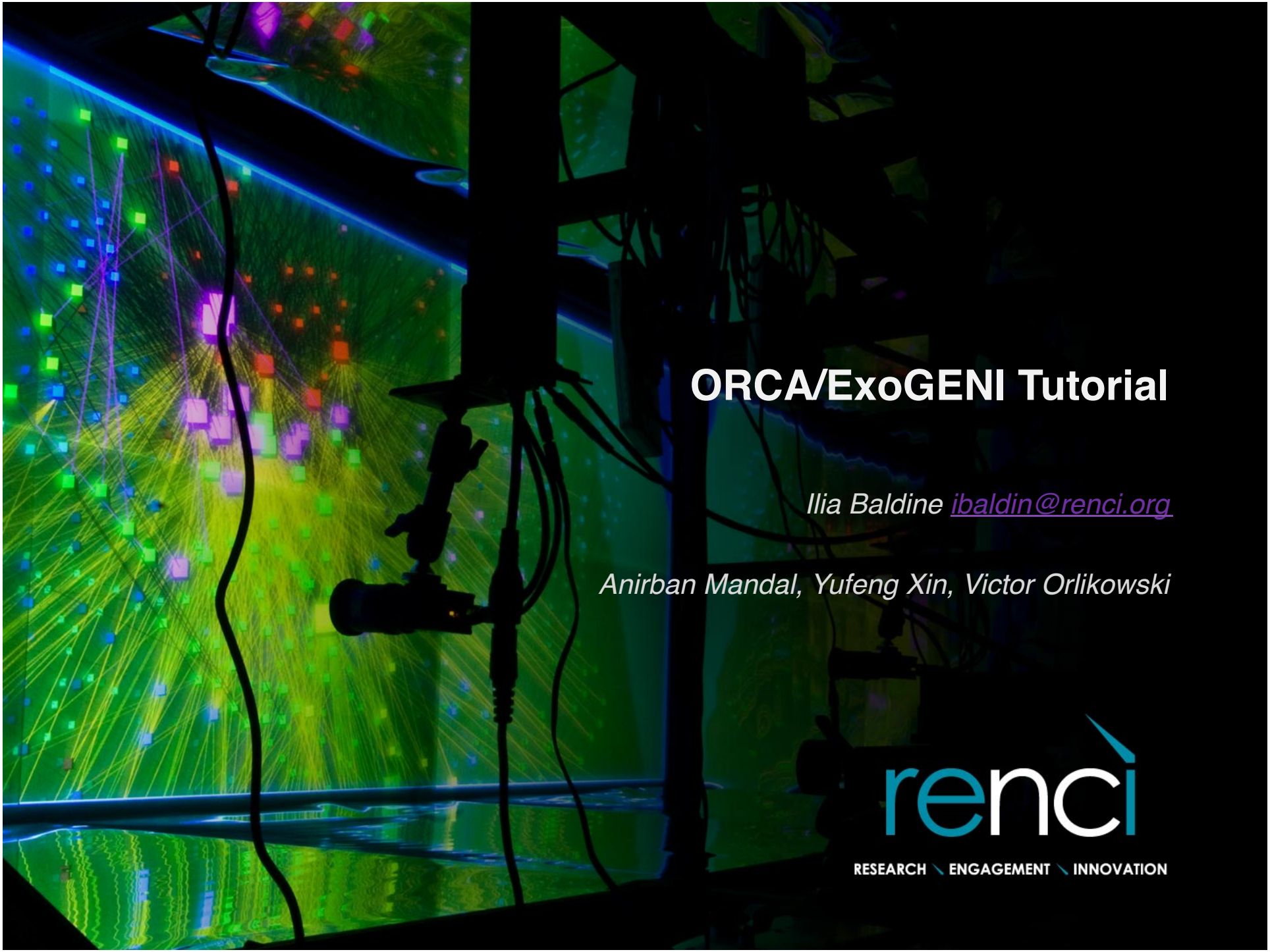


ORCA/ExoGENI Tutorial

Ilia Baldine *ibaldin@renci.org*

Anirban Mandal, Yufeng Xin, Victor Orlikowski

renci

RESEARCH \ ENGAGEMENT \ INNOVATION

Tutorial sections

- ExoGENI and Orca Overview
- Flukes Overview
- Creating slices with Flukes
  - Intra-rack slices (everyone)
  - Inter-rack slices (demo only)
- OpenFlow slices on ExoGENI (demo only)
- Creating slices with Omni and GENI AM API (everyone)
- Tutorial page:
  - http://groups.geni.net/geni/wiki/GEC13Agenda/ORCATutorial
  - Please open in your browser
  - Please open the presentation

renci

Section: ExoGENI and ORCA Overview

ExoGENI Testbed

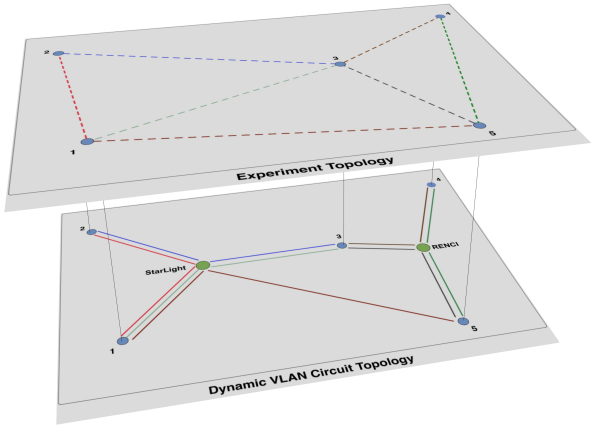
- 14 GPO-funded racks
  - Partnership between RENCI, Duke and IBM
  - IBM x3650 M3/M4 servers
    - 1x146GB 10K SAS hard drive +1x500GB secondary drive
    - 48G RAM 1333Mhz
    - Dual-socket 6-core Intel X5650 2.66Ghz CPU
    - Dual 1Gbps adapter
    - 10G dual-port Chelseo adapter
  - BNT 8264 10G/40G 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 (some older racks run Eucalyptus)
  - EC2 nomenclature for node sizes (m1.small, m1.large etc
  - Interconnected by combination of dynamic and static L2 circuits through regionals and national backbones
- http://www.exogeni.net

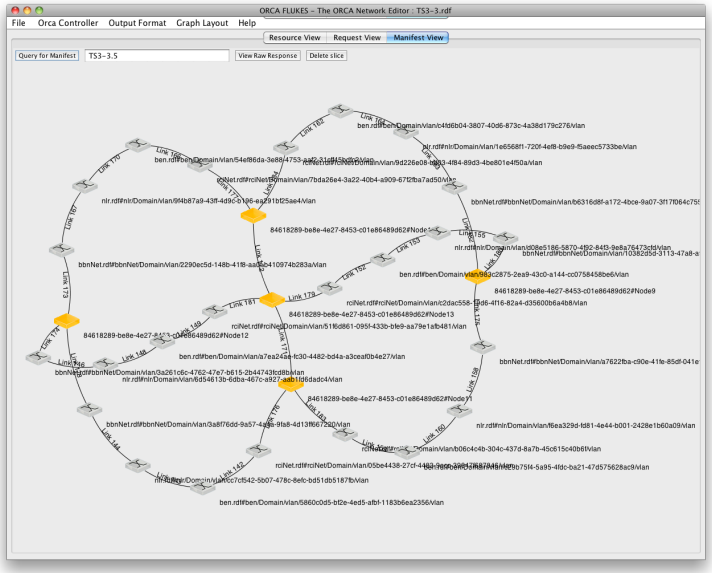
ExoGENI Status

- 2 new racks deployed
  - RENCI and GPO
- 2 existing racks
  - Duke and UNC
- 2 more racks available by GEC14
  - FIU and UH
- Connected via BEN (http://ben.renci.org), LEARN and NLR FrameNet, (eventually I2)
- Partner racks
  - NICTA (under construction)
  - U of Alaska Fairbanks

ExoGENI slice isolation

- Strong isolation is the goal
- Compute instances are KVM based and get a dedicated number of cores
  - Caveat: currently all instances get 1 core (different RAM and disk). Will be remedied by Summer/Fall 2012
- VLANs are the basis of connectivity
  - VLANs can be best effort or bandwidth-provisioned (within and between racks)
  - Caveat: current hardware in the racks allows best-effort VLANs only – will be remedied by Fall 2012 with support from the vendor





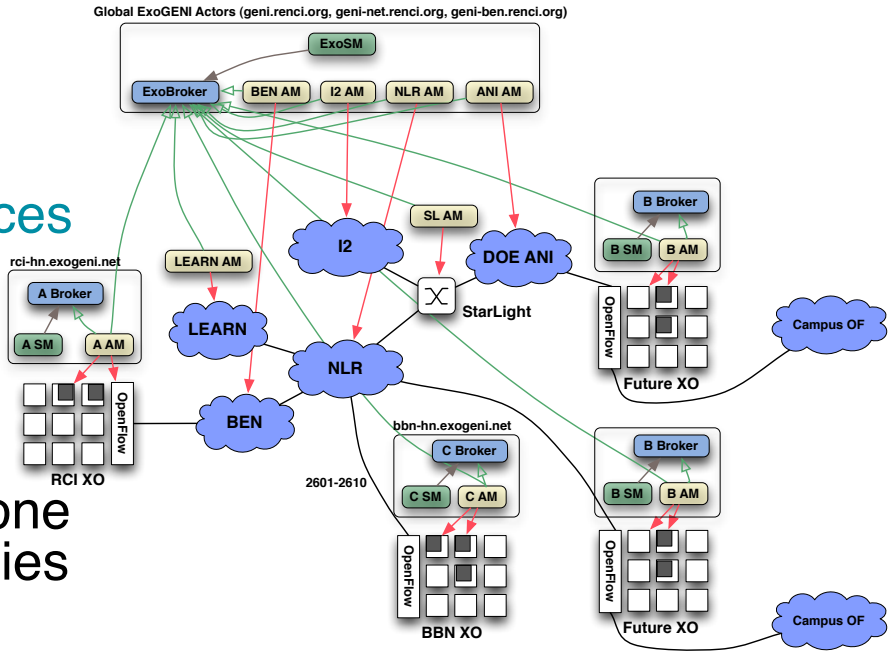
ORCA Overview

- 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
  - We do not run SA's or issue GENI credentials
  - We honor GPO and Emulab-issued credentials
- Supports ORCA-native interface, resource specification and tools
  - Flukes
- Supports GENI AM API and GENI Rspec
  - Omni
- (Almost) compatible with Gush

renci

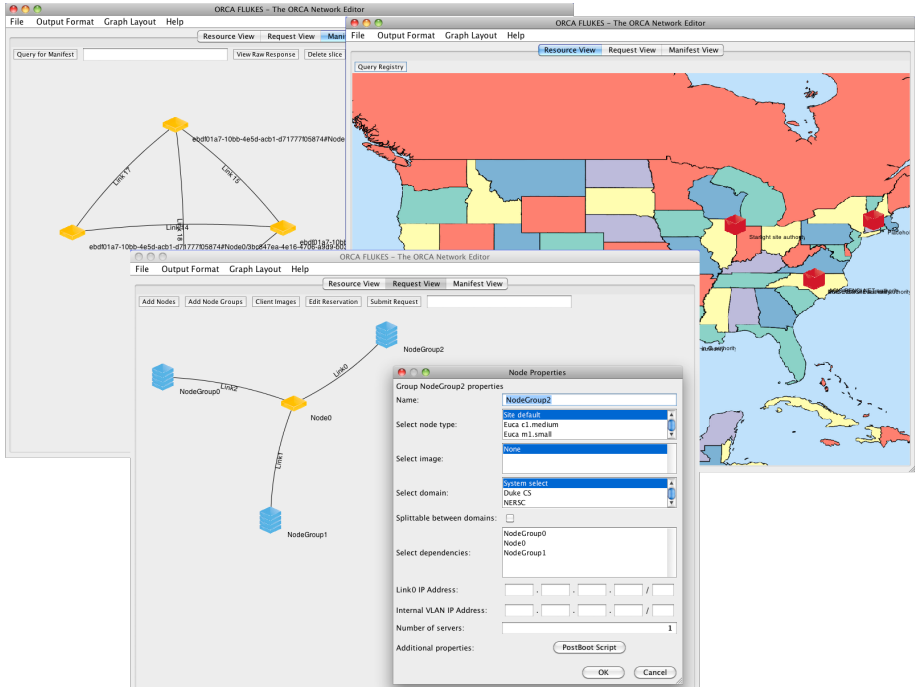
ORCA Deployment in ExoGENI

- Each rack runs its own SM actor that acts as GENI AM and exposes
  - ORCA native API
  - GENI AM API
- Rack-local SM can only create slices 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
- ExoSM
  - https://geni.renci.org:11443/orca/xmlrpc



Section: Flukes Overview

- Graphical tool for creating and managing slice topologies in ORCA
  - JAVA (JNLP)



Section Overview

- Configuring Flukes prior to launch
- Launching Flukes
  - GUI Overview
  - Nodes, NodeGroups and Link parameters
  - Node-level vs. reservation level options
- Building slice request topologies
- Launching slice requests
- Inspecting slice manifests
- Logging into nodes in the slice
- NEuca-py tools

renci

Configuring Flukes

- All user properties are under $HOME/.flukes.properties – it is a text file
  - Edit $HOME/.flukes.properties
    - Open in an editor and replace *EVERY* occurrence of XX with your index (including leading zero)
  - Inspecting keystore file (make note of key alias ('tutorialXX')
    - $ cd Tutorials/ORCA/orcatuXX
    - $ keytool -list -keystore ssh/orcatu01.jks
    - NOTE: your key name is tutorialXX, and your key and keystore password is tut0rialXX

renci

Launch Flukes!

- Double-click Flukes icon on your desktop
- Permanent stable version link
  - http://geni-images.renci.org/webstart/flukes.jnlp

renci

GUI Overview

- Tabs
  - Resources, Request, Manifest
- Menus
  - Current properties
  - Overwriting properties ($HOME/.flukes.properties)
- Mouse modes
- Buttons
- Adding nodes, nodegroups and links

renci

Node and Link parameters

- Create a single node
- Right-click on the Node
  - Look at properties
  - Edit properties
    - Node type (size)
    - VM image
    - Domain (binding)
    - PostBoot script
- Create another node, link the two together
- Right-click and open properties again
  - Specify IP address on the link
  - Node functional dependencies
- Right-click on links
  - Inspect and edit link properties
  - Note only bandwidth is currently respected (and not everywhere due to hardware limitations)

renci

NodeGroup parameters

- Create a single unattached node group
- Right-click to inspect and edit properties
  - Group sizes
  - Internal VLANs
  - Splittable groups

Nodes and NodeGroups

- Node is an individual compute element
  - Typically a VM or a hardware node
  - IP address(es) on links, size, image, site binding, post boot script
  - Can I control management IP address assignment? NO!
- NodeGroup is a group of identically configured nodes
  - A lot like a node except
    - May have an internal VLAN
    - PostBoot script is templated using Velocity template engine
      - https://geni-orca.renci.org/trac/wiki/flukes
    - IP address assignment is semi-automatic (starting with a user-specified address)
    - Node groups can be splittable between sites

renci

What do I get when I …?

- Create a standalone node?
  - You get a single compute element at one of the sites with a single network interface to the management network through which you can SSH into the node
  - Management interface is always eth0
- Connect two nodes together?
  - You get two compute elements each with two network interfaces – one for management access and one for the link between two nodes.
  - User-controlled interfaces start with eth1
  - You can control IP address assignment on the interfaces linking the two nodes (suggested range: 172.16.0.0/16)

renci

What do I get when I …?

- Create a standalone NodeGroup?
  - You get some number of nodes (specified in the group size) each with a single interface to the management network (eth0)
  - Nodes typically will be within the same rack
  - If node group is marked splittable nodes may be split across sites
- Create a node group with a private VLAN?
  - You get some number of nodes each with two interfaces – one on the management network and one on the VLAN interconnecting the members of the group.
  - If you want IP addresses automatically configured, specify the first address in Flukes; all others will be configured sequentially
  - Example: NodeGroup with a private VLAN size 10, specified IP address 172.16.100.1/24. Nodes will get addresses 172.16.100.1, 2, 3, 4, 5, 6, 7, 8, 9 and 10
- Connect a node group to a node or another node group?
  - All nodes within the group and the adjacent node (or all nodes in both groups) have interfaces on a common VLAN. They also have management interfaces (eth0)
  - IP address is specified similarly to private VLAN
  - Beware of address clashes! (ie here is a piece of rope, feel free to shoot yourself in the foot)
- Connect a node group (A) that has a private VLAN to another node or a group (B)
  - Nodes within group A will have 3 interfaces – one for management (eth0), one on the private VLAN with members of this group, one on a VLAN shared with members of this group and the adjacent node or group B.

renci

Can I tell which interface in the node will be eth1, eth2 etc?

- No, nor should you need to. Interfaces are identified by links they belong to.
- Note that different OSs name interfaces differently

renci

Illustrations

renci

Node-level vs. Reservation-level options

- Reservation-level options overwrite node-level options
  - VM image
  - Domain binding
- OpenFlow slice parameters can currently only be specified at reservation level

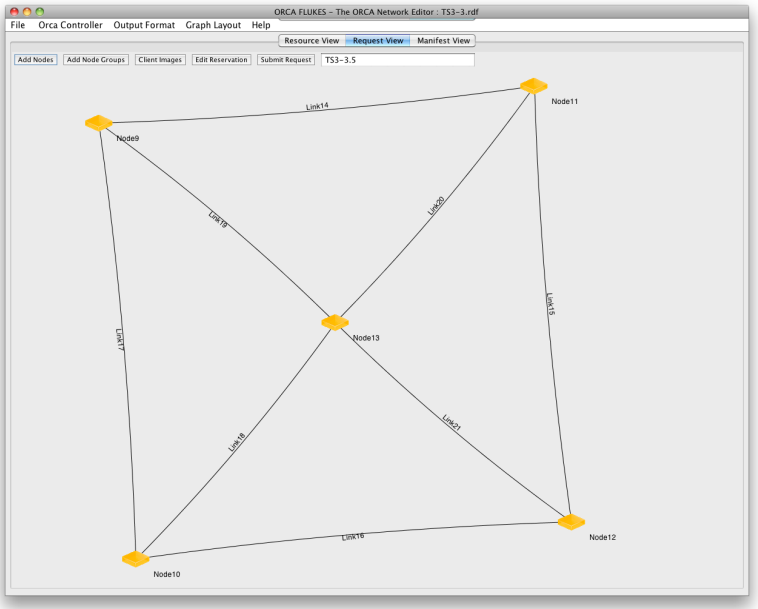
renci

Section: Creating slices with Flukes

renci

Launching a slice

- Click ‘File I New’ to clear the request
- Use 3 or 4 Nodes and create a topology
  - Use regression image and m1.small instance type
  - Assign IP addresses on links
    - Using subnets 172.16.0.0/16
  - Specify a post-boot script



Untitled

```
#!/bin/bash

DATE=`date`
echo 'This is my very first boot script and today is $DATE' > /tmp/out.txt
```

Launch a slice

- These will be unbound slices
  - We will let ORCA select sites with available resources
- Specify slice duration (click ‘Edit Reservation’)
- Fill in slice name (must be unique)
- Click ‘Submit Request’
  - Type in the alias of the key in the keystore (‘tutorialXX’)
  - Type in the password (‘tut0rialXX’)
  - ‘OK’
- Inspect the output window
  - Mainly a debugging tool. Will go away in the future.

renci

Inspect slice manifest

- Cut and paste slice name into the ‘Manifest View’ tab
- Click ‘Query for Manifest’
  - Inspect raw output if interested
  - Inspect the state of slice elements by right clicking on each element (usually ‘Ticketed’)
  - If you see ‘Failed’ you have a problem
- Poll by clicking ‘Query for Manifest’
  - Topology should materialize
  - All states should report ‘Active’
- Play around with layouts to get something pleasing

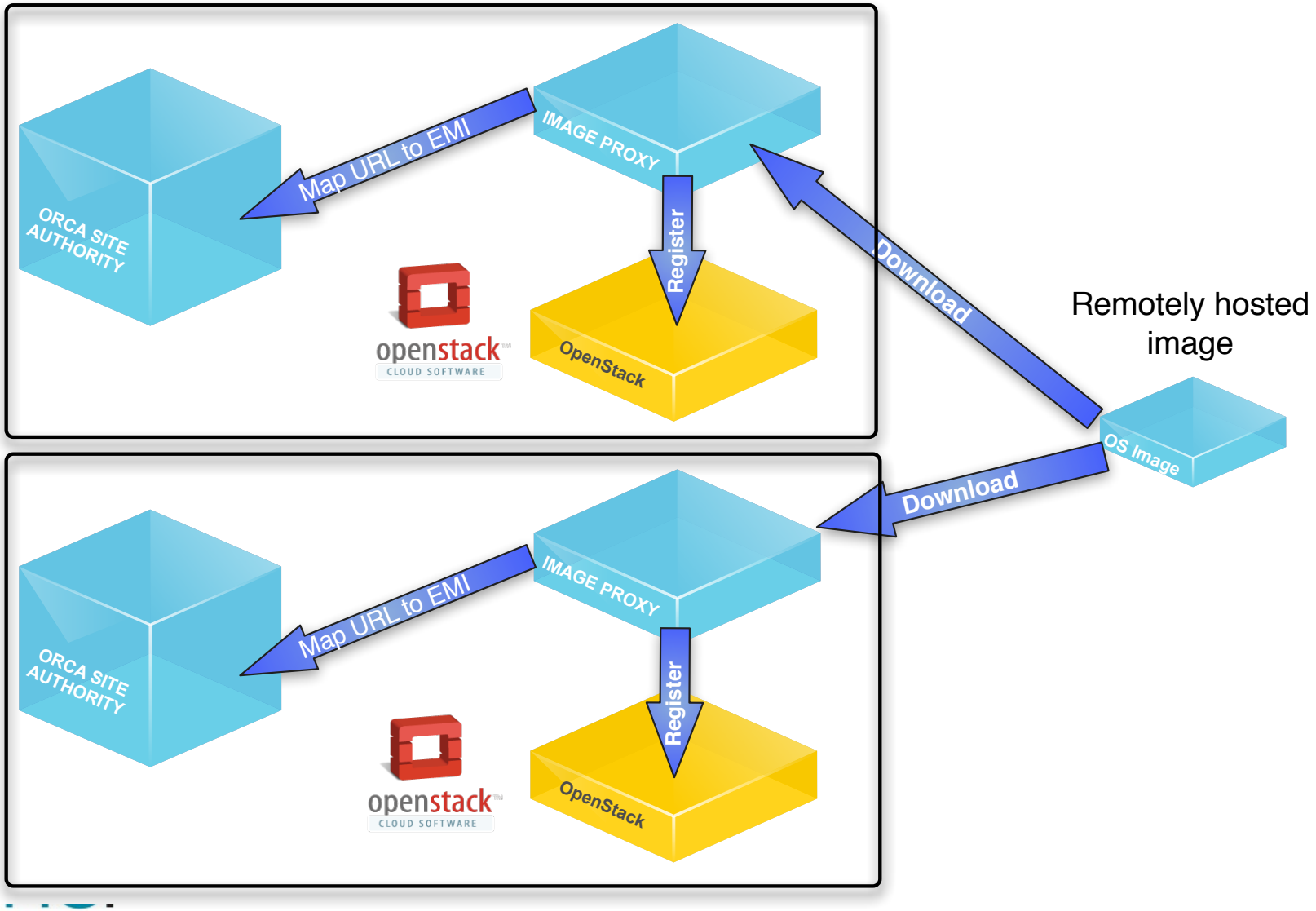
renci

VM images

- Creating your own image
- Specifying your own image for ORCA
- Delays
  - Images are downloaded and registered with the site at the time of slice creation
    - If you repeatedly use the same image and the site already has it, this step is skipped
    - Images may be cached-out causing longer delays (to download and re-register)
- Are there examples of known good images?
  - Yes, visit https://geni-orca.renci.org/trac/wiki/neuca-images

renci

VM Images


ORCA SITE AUTHORITY
Map URL to EMI
IMAGE PROXY
Register
openstack
CLOUD SOFTWARE
OpenStack
Download
Remotely hosted image
OS Image
Download
ORCA SITE AUTHORITY
Map URL to EMI
IMAGE PROXY
Register
openstack
CLOUD SOFTWARE
OpenStack

Example image metafile

```
<images>
    <image>
        <type>ZFILESYSTEM</type>
        <signature>b54ed5a42cd99475c3d5d7c7a9839b69cf2076d5</signature>
        <url>http://geni-images.renci.org/images/workflows/pegasus/images/
     pegasus-4.0-v0.3.sparse.img.tgz</url>
    </image>
    <image>
        <type>KERNEL</type>
        <signature>f8a64d3bc429e8fb46c94ff3b11a932a27c142bc</signature>
        <url>http://geni-images.renci.org/images/workflows/debian-squeeze-kernel/
     vmlinuz-2.6.28-11-generic</url>
    </image>
    <image>
      <type>RAMDISK</type>
      <signature>6225968f43299aa40f6b1491360f3ce080bd16c4</signature>
      <url>http://geni-images.renci.org/images/workflows/debian-squeeze-kernel/
     initrd.img-2.6.28-11-generic</url>
    </image>
</images>
```

How does ORCA refer to an image

- URL of a metafile (can be same or different webserver as the image)
- SHA1 checksum of the metafile (to ensure it has not been modified)
- Workflow
  - Create filesystem, kernel ramdisk
  - Place on webserver
  - Take SHA1 signatures of each file
  - Generate metafile
  - Take SHA1 signature of metafile and its URL and add it to .flukes.properties or put it in Rspec
  - Try on a small slice (one node) to make sure it boots
- Can I put my image on your server?
  - Sorry, no.
- Will my image always remain cached at the racks?
  - No, depending on the use, your image may be cached out.

renci

Doman binding

- Leaving domain as ‘System Select’ leaves ORCA to pick the domain that has available resources
- You can explicitly bind to specific domains
- If nodes or groups in slice request belong to different domains, appropriate inter-domain links will be provisioned on demand

renci

Inspect slice manifest

- Cut and paste slice name into the ‘Manifest View’ tab
- Click ‘Query for Manifest’
  - Inspect raw output if interested
  - Inspect the state of slice elements by right clicking on each element (usually ‘Ticketed’)
  - If you see ‘Failed’ you have a problem
- Poll by clicking ‘Query for Manifest’
  - Topology should materialize
  - All states should report ‘Active’
- Play around with layouts to get something pleasing

renci

Logging into nodes

- Right click on node
- Select ‘Login to node’
  - In terminal window type in SSH key password (‘gec13’)
- Inspect uptime

  $ uptime
- Inspect the output of your boot script
- Inspect interfaces

  $ ifconfig
- Try to ping node neighbors

renci

NEuca-py tools

- NEuca tools are loaded in the image
  - They configure network interfaces at boot time
  - They execute the post boot script
  - An image with NEuca tools will do neither of those things
    - You can still configure interfaces manually
- Allow you to inspect the VM configuration
- Run ‘neuca’ to get the list of neuca tools

  $ neuca
- Run ‘neuca-user-data’
  - Note your boot script
- If you create your own VM image you are strongly encouraged to install NEuca tools on it
  - Visit https://geni-orca.renci.org/trac/wiki/NEuca-guest-configuration for instructions

renci

Please delete slices!

renci

Inter-domain slice

- Simple inter-domain slice (RCI to BBN)
  - Uses NLR FrameNet
- Compare latencies on two links

renci

Section: OpenFlow slices in ExoGENI

OpenFlow slices

- Start OpenFlow controller
  - Using FloodLight 0.82
- Create a slice of one NodeGroup
  - More than one NodeGroup possible
  - Explicitly bind to one of XO racks
    - Automatic binding for OpenFlow is not yet available
- Declare reservation as ‘OpenFlow’ in reservation level properties and submit slice
  - Any link (including a private VLAN within a group) within the request that goes through the OpenFlow switch will be controlled by your controller

renci

Section: Creating slices with Omni

renci

Creating slices with Omni and GENI AM API

- ORCA's AM API is mostly v2 compliant
  - SliverStatus does not return individual resource URNs
  - ListResources may return multiple RSpec documents
- Use the RSpec file linked to this tutorial webpage:
  - http://groups.geni.net/geni/wiki/GEC13Agenda/ORCATutorial

renci

Sample simple RSpec

```
<?xml version="1.0" encoding="UTF-8"?>
<rspec type="request"
xsi:schemaLocation="http://www.protogeni.net/resources/rspec/2
                    http://www.protogeni.net/resources/rspec/2/request.xsd"
    xmlns:flack="http://www.protogeni.net/resources/rspec/ext/flack/1"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns="http://www.protogeni.net/resources/rspec/2">
<node client_id="geni1">
 <sliver_type name="m1.small">
   <disk_image
name="http://geni-images.renci.org/images/regression/regression-deb5-i386.xml"
version="ea80af6601a2a000ec5b050d7e7701f26db096fc" />
 </sliver_type>
 <interface client_id="geni1:0">
    <ip address="172.16.1.1" netmask="255.255.255.0" />
 </interface>
</node>
<node client_id="geni2">
 <sliver_type name="m1.large">
   <disk_image
name="http://geni-images.renci.org/images/regression/regression-deb5-i386.xml"
version="ea80af6601a2a000ec5b050d7e7701f26db096fc" />
 </sliver_type>
 <interface client_id="geni2:0" >
   <ip address="172.16.1.2" netmask="255.255.255.0" />
 </interface>
</node>
<link client_id="center">
  <interface_ref client_id="geni1:0" />
  <interface_ref client_id="geni2:0" />
</link>
</rspec>
```

renci

Issuing OMNI commands

- Create a slice with GPO SA
  - omni.py -c omni_config -a https://geni.renci.org:11443/orca/xmlrpc createslice orcav2-test3
- CreateSliver with given RSpec
  - omni.py -c omni_config -a https://geni.renci.org:11443/orca/xmlrpc -n createsliver orcav2-test3 two-node.rspec
- Query sliver status
  - omni.py -c omni_config -a https://geni.renci.org:11443/orca/xmlrpc sliverstatus orcav2-test3
- List resources within the slice (IP addresses)
  - omni.py -c omni_config -a https://geni.renci.org:11443/orca/xmlrpc listresources orcav2-test3
- Login to nodes
- Delete sliver
  - omni.py -c omni_config -a https://geni.renci.org:11443/orca/xmlrpc deletesliver orcav2-test3

renci

RSpec conventions

- Install and execute service tags are respected
  - tar.gz, tar.Z, tar.bz, deb, rpm and zip recognized
- Domain binding is possible by specifying component id or component manager id

```
<node component_id="urn:publicid:IDN+uncvmsite+node+vm"
      component_manager_id="urn:publicid:IDN+uncvmsite+authority+cm"
      client_id="pc175"
      exclusive="true">
   <sliver_type name="raw-pc" />
</node>
```

- No NodeGroups
- No post boot scripts (yet)
- No OpenFlow slices (yet)
- See https://geni-orca.renci.org/trac/wiki/orca-and-rspec for updated information

renci

Thank you for attending the tutorial!

- More Orca information
  - http://geni-orca.renci.org
- More ExoGENI informatino http://
  - http://www.exogeni.net

renci