Software Defined Networks: Engineering GENI

Heidi Picher Dempsey and Tim Upthegrove
BBN Technologies
NANOG62 Meeting
October 6-8, 2014
www.geni.net
GENI provides a virtual lab for networking and distributed systems research and education

- GENI started with exploratory, rapid prototyping 5 years ago
- GENI design assumes federation of *autonomously owned and operated* systems
- Yearly prototyping cycle for an idea: develop, integrate and *operate*
- Experimenters use the testbed *while we are building it out*
- Even prototypes have “activist” users, and must evolve to satisfy those users or fade away. Two of five original design frameworks predominate now.
  - “Horizontal” dataplane slicing as a service (or sometimes just engineered)
  - “Vertical” control plane APIs to negotiate and allocate resources
GENI provides compute, network, and wireless resources that can be connected in experimenter-specified Layer 2 topologies.
GENI provides compute resources that can be connected in experimenter specified Layer 2 topologies.
Multiple GENI Experiments run Concurrently

Resources can be shared between slices

Experiments live in somewhat isolated “slices”
GENI is “Deeply Programmable”

I install software I want throughout my network slice (into routers, switches, …) or control switches using OpenFlow

Experimenters can set up custom topologies, protocols and switching of flows
Access to GENI

- Over 2500 users (September, 2014)
- Experimental resources at 52 campuses, 11 regional networks, 10 WiMAX/LTE wireless sites
- GENI credentials and management based on Shibboleth single sign on and InCommon
- GENI experiments run continuously
- Operations support from six groups in different US locations
Software: Clearinghouse and Aggregates

- **Clearinghouse**: manages users, projects and slices
  - Standard credentials shared via custom API or new Common CH API
  - GENI supported accounts: GENI Portal/CH, PlanetLab CH, ProtoGENI CH
- **Aggregate**: provides resources to GENI experimenters
  - Typically owned and managed by an organization
  - Speaks the GENI Aggregate Manager API (AM API)
  - [http://groups.geni.net/geni/wiki/GAPI_AM_API_V3](http://groups.geni.net/geni/wiki/GAPI_AM_API_V3) most recent version
  - [http://trac.gpolab.bbn.com/gcf](http://trac.gpolab.bbn.com/gcf) download reference implementation (gcf), OMNI command line client
  - Examples: PlanetLab, Emulab, GENI racks on various campuses
Engineering for Layer2 SDN

- Experimenters run their own SDN controllers
- Each network aggregate may run their own controller (many don’t)
- SDN switches and endpoints are configured with VLAN ranges that can be used for SDN experiments. Supported configurations:
  - One VLAN per experiment with/without OpenFlow controller
  - One shared VLAN with multiple OpenFlow controllers (per-experiment addressing and controllers mediated by GPO and GENI software)
  - One multipoint VLAN with one service (e.g. wireless network experiments)
- Experimenters can choose software or hardware switches—this talk is about hardware switches
- GENI Aggregate Manager (AM) software negotiates and coordinates resource access
  - AM API includes VLAN “stitching”
    http://groups.geni.net/geni/wiki/GAPI_AM_API_V3
  - OpenFlow site/network access AM
    http://groups.geni.net/geni/wiki/OpenFlow/FOAM
Network Engineering Requirements for Shared Services

- L2 dataplane engineering
  - campuses, regional, core and international networks
  - many vendors and technologies
  - 1-100GbE interfaces (GENI shares with other R&E projects)
  - Shared or exclusive experimenter VLANs on interfaces, depending on experiment (mostly exclusive)
- SDN (OpenFlow 1.0) switches with experimenter’s and sometimes R&E network’s controllers (many vendors, varying implementation of standards)
- Standard Internet control plane
- Internet2 AL2S cross-connects and ION

http://groups.geni.net/geni/wiki/GENIOESSTopologiesPerformance-IONtoAL2SPerformance
GENI Interoperable SDN

• Network aggregates operate various switches
  – Brocade
  – IBM
  – HP
  – Pica8
  – Cisco
  – NEC
  – Juniper
  – Dell
  – Open vSwitch (software-only switch)

• Experimenters and network engineers develop various controllers, based on open source projects
  – Floodlight
  – POX (replaced NOX)
  – OpenDaylight

• Operators develop additional open source tools to support resource sharing and monitoring (several—see www.geni.net)
Deployment engineering for SDN

- Disable Spanning Tree Protocol (not just on SDN switches)
- Disable MAC learning
- Coordinate IP address ranges to avoid duplication, especially with shared VLANs
- Monitor for loops and load, use external limits if needed
- Compare firewall rules to SDN traffic profiles
- Separate control plane from SDN data plane
- Beware partial OpenFlow specification implementations
SDN Operations Example

- Two shared VLANs for ops monitoring
- Routinely send small amounts of OF data on dataplane slices using local controller
- low CPU use
Operations Traffic and CPU Monitoring

Known Causes of High CPU Usage
- High control plane activity
- Many unmatched packets arriving on switch
- Forwarding loops etc.
- Experimenters may cause intentionally (flood, rewrite)
Reported Errors

Denver Switch Symptoms:
- Most monitoring traffic failing
- Many flows installed to Seattle
- High CPU use
- Experimenters report problems

Chicago Switch Symptoms:
- Some monitoring traffic failing
- Few flows
- Relatively low CPU use
Ops Debug Workflow

1. **Monitoring Failure**
   - Investigate

2. **Review OF controller traffic towards control plane traffic**
   - Nothing interesting

3. **Review Switches flow tables**
   - Output action OFPP_IN_PORT
     - Determine flow owner from sliced IP range
       - BBN owns the flow

4. **Change BBN Controller to Reactive and review ALL traffic**
   - Does not help
     - Chicago is source according to Denver

5. **Review Chicago flows, but none found**
   - Firmware bug causing phantom flow!!

6. **Restart Denver OpenFlow Switch**
   - Does not help

7. **Review BBN Controller Logic**
   - Does not help

8. **Restart BBN OpenFlow Controller**
   - Offending flows cleared

   **Success!!!**
SDN Basic Tools

Floodlight Table Stats

Floodlight Flow Stats

Wireshark OF dissector
Real Life Flow Matches—Only One Vendor

Flow match on v2 modules

<table>
<thead>
<tr>
<th>Flow type</th>
<th>VLAN ID</th>
<th>VLAN Ptl</th>
<th>In_Port</th>
<th>Ethernet Type</th>
<th>Source MAC</th>
<th>Destination MAC</th>
<th>Source IP</th>
<th>Destination IP</th>
<th>IP ToS</th>
<th>IP Prot.</th>
<th>Source Port</th>
<th>Destination Port</th>
<th>V2 module flow location</th>
</tr>
</thead>
<tbody>
<tr>
<td>VLAN ID</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>hardware</td>
</tr>
<tr>
<td>VLAN PCP</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>hardware</td>
</tr>
<tr>
<td>In_Port</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>hardware</td>
</tr>
<tr>
<td>Ethertype IP^d</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>IP</td>
<td>c</td>
<td>c</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>hardware</td>
</tr>
<tr>
<td>Ethertype IP^e</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>IP</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>software</td>
</tr>
<tr>
<td>Ethertype Non-IP^f</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>Non-IP</td>
<td>b</td>
<td>b</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>hardware</td>
</tr>
<tr>
<td>Ethertype Non-IP^g</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>Non-IP</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>software</td>
</tr>
<tr>
<td>No Ethertype</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>c</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>software</td>
</tr>
</tbody>
</table>

^a A flow that matches the VLAN-ID, VLAN-PCP and IN_PORT with all other fields being blank will be in hardware.

^b **Wildcard** — It does not matter if this field is specified or not in the flow.

^c **Blank** — This field MUST NOT be present in the flow or is not applicable.

^d If the Ethertype is IP, the MAC address fields must not be specified for the flow to be in hardware.

^e If the Ethertype is IP and any MAC address fields is specified, the flow will be in software.

^f If the Ethertype is non-IP, the flow can match against MAC address fields also in hardware provided the IP address fields are not specified.

^g If the Ethertype is non-IP and any of the IP fields are specified, the flow will be in software.

^h If the Ethertype fields is blank and any of the MAC address fields or IP address fields are specified, the flow will be in software.
GENI SDN Evolution

• Switch support for hybrid networking (non-OF and OF on same switch)
• Separating network slicing from SDN control
• OpenFlow 1.0 to 1.3 migration
• OpenFlow policies on a distributed network
• Keeping networks interoperable
• SDN for broadband and home networks
• SDN Exchange points
• Cross-domain SDN monitoring
SDN Operations Requirements

- Site confirmation tests with logs and RSPECs
  [http://groups.geni.net/geni/wiki/GENIRacksHome/InstageniRacks/ConfirmationTestStatus](http://groups.geni.net/geni/wiki/GENIRacksHome/InstageniRacks/ConfirmationTestStatus)
  [http://groups.geni.net/geni/wiki/GENIRacksHome/ExogeniRacks/ConfirmationTestStatus](http://groups.geni.net/geni/wiki/GENIRacksHome/ExogeniRacks/ConfirmationTestStatus)

- Emergency Stop and Legal, Law Enforcement and Regulatory Event Coordination (GMOC at Indiana University)

- Shared monitoring infrastructure and shared operations (6 major ops groups)
SDN Ops Deployment Requirements (cont)

- Standard installation processes
  http://groups.geni.net/geni/wiki/GENIRacksHome/RacksChecklistStatus
- System Acceptance Testing
  - Production: InstaGENI, ExoGENI
  - Provisional: OpenGENI (Dell), Cisco
- Shared site resource and access details
  http://groups.geni.net/geni/wiki/GeniAggregate