Washington University in St. Louis | Engineering

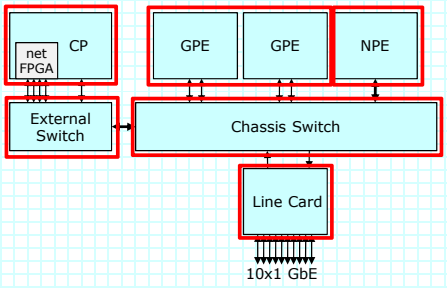
Prototype Deployment of Internet Scale Overlay Hosting

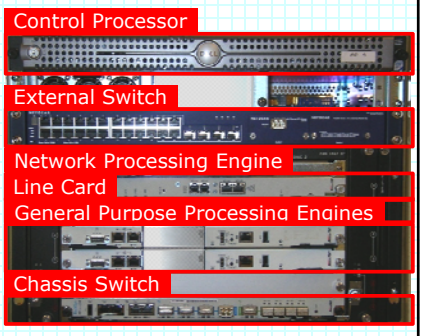
Patrick Crowley and Jon Turner
and John DeHart, Mart Haitjema Fred Kuhns, Jyoti Parwatikar, Ritun Patney, Charlie Wiseman, Mike Wilson, Ken Wong, Dave Zar
Computer Science & Engineering
Washington University

www.arl.wustl.edu

Washington University in St. Louis | Engineering

Project Objectives



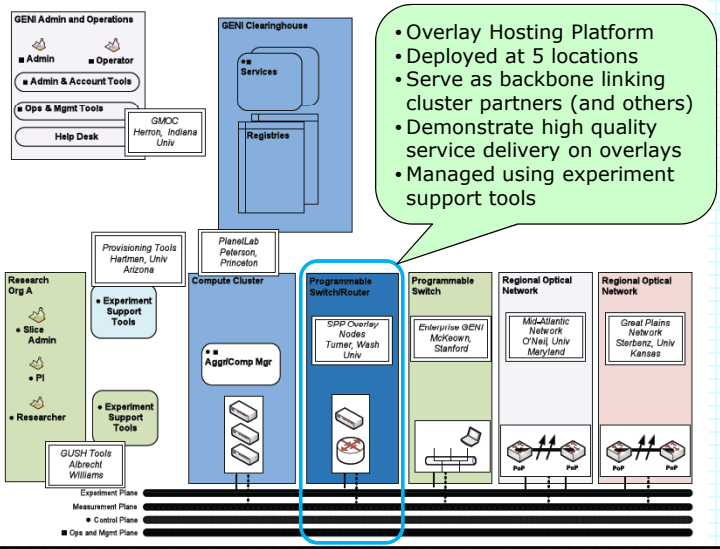


- Deploy five experimental *overlay hosting platforms*
  - » located at Internet 2 PoPs
  - » compatible with PlanetLab, moving to GENI control framework
  - » performance characteristics suitable for service deployment
    - integrated system architecture with multiple server blades
    - shared NP-based server blades for fast-path packet processing
- Demonstrate multiple applications

2

Washington University in St. Louis | Engineering

Role in Cluster B

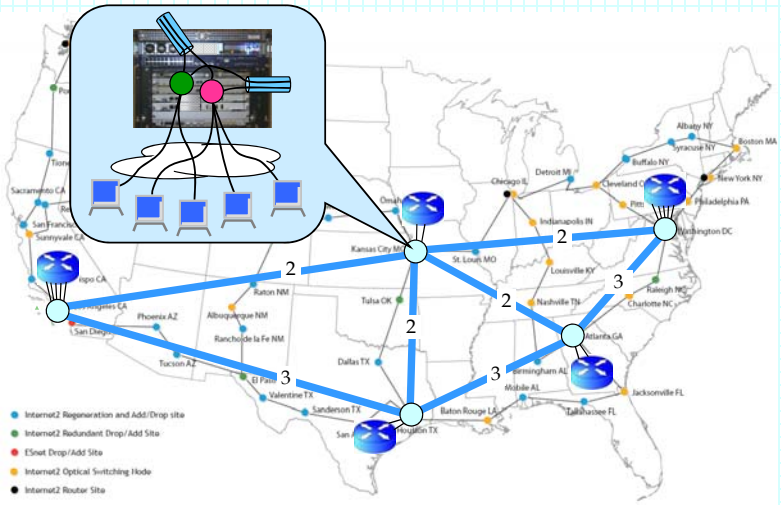


- Overlay Hosting Platform
- Deployed at 5 locations
- Serve as backbone linking cluster partners (and others)
- Demonstrate high quality service delivery on overlays
- Managed using experiment support tools

3

Washington University in St. Louis | Engineering

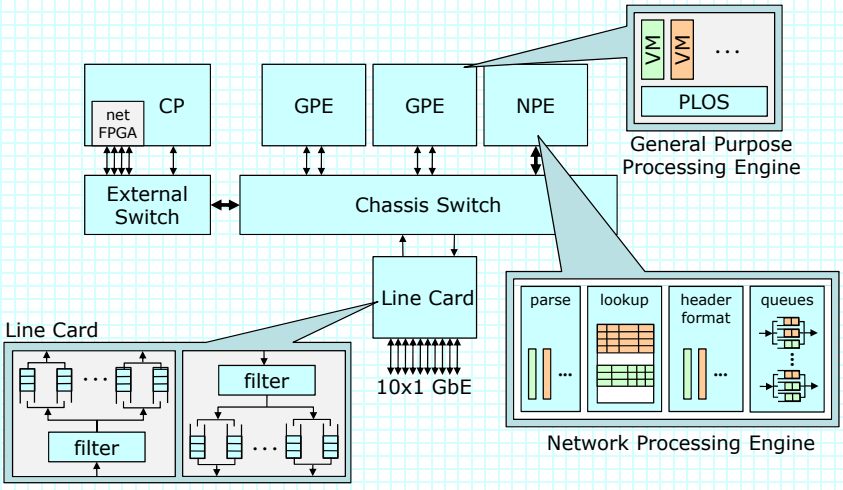
Target Internet 2 Deployment



4

Washington University in St. Louis Engineering

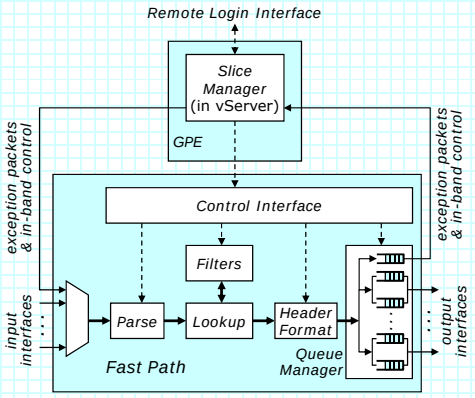
Hosting Platform Details



Washington University in St. Louis Engineering

Application Framework

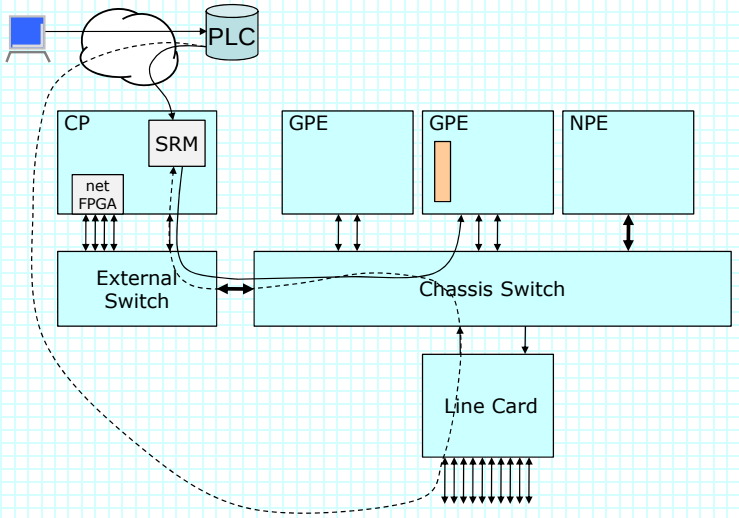
- Fastpath/slowpath
  - » fastpath mapped onto NPE
  - » slice manager in vServer on GPE
- Configurable elements
  - » code option – determines how packets processed by parse, header format
  - » logical interfaces
    - may be public or tunnel
    - guaranteed bandwidth
  - » TCAM filters
  - » Queues
    - length, bandwidth



- Slice manager can configure fastpath using provided library
  - » or manually, using command line interface

Washington University in St. Louis Engineering

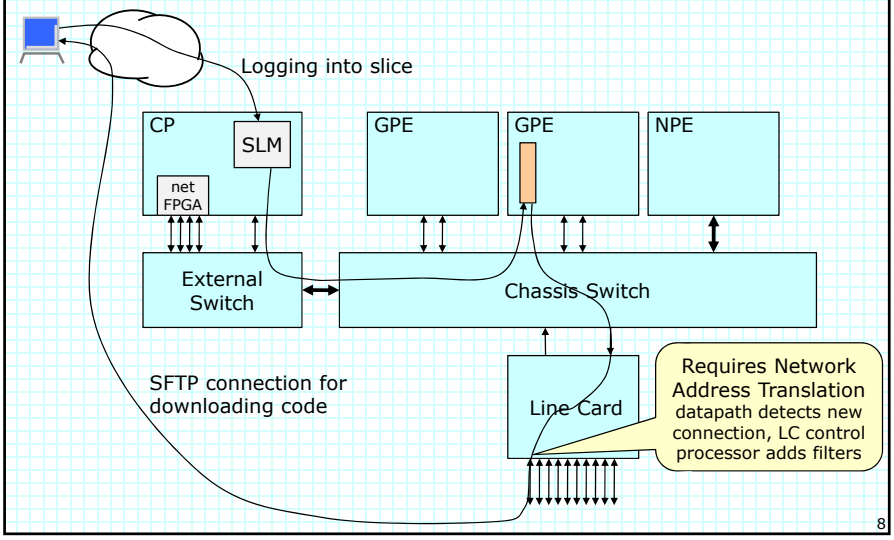
Creating a Slice



7

Washington University in St. Louis Engineering

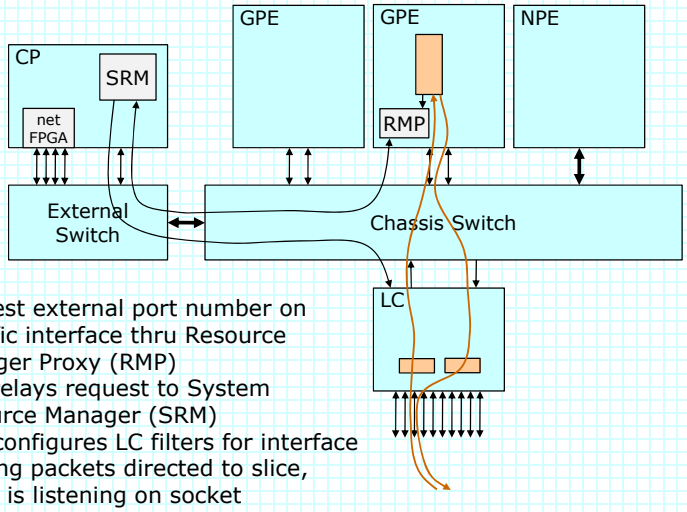
Starting up a Slice



8

Washington University in St. Louis Engineering

Configuring an External Port

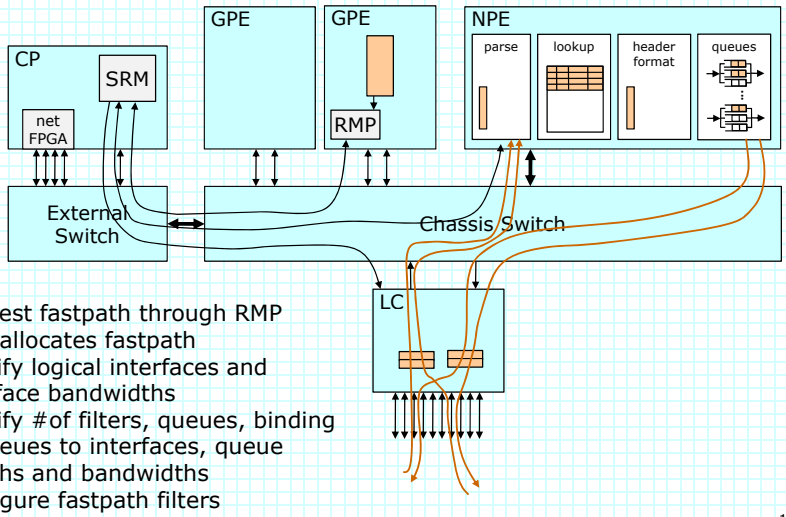


- Request external port number on specific interface thru Resource Manager Proxy (RMP)
- RMP relays request to System Resource Manager (SRM)
- SRM configures LC filters for interface
- Arriving packets directed to slice, which is listening on socket

9

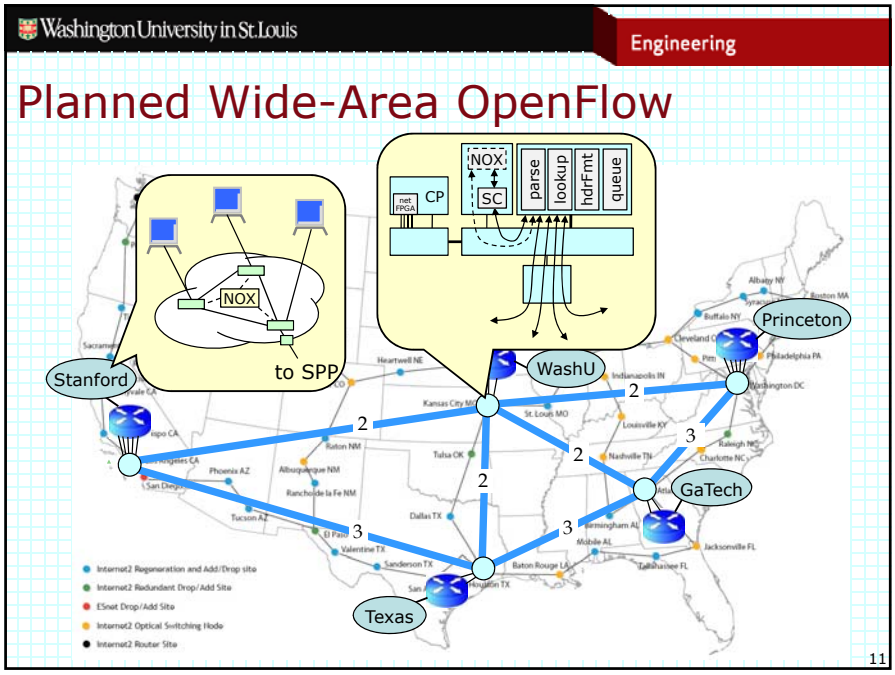
Washington University in St. Louis Engineering

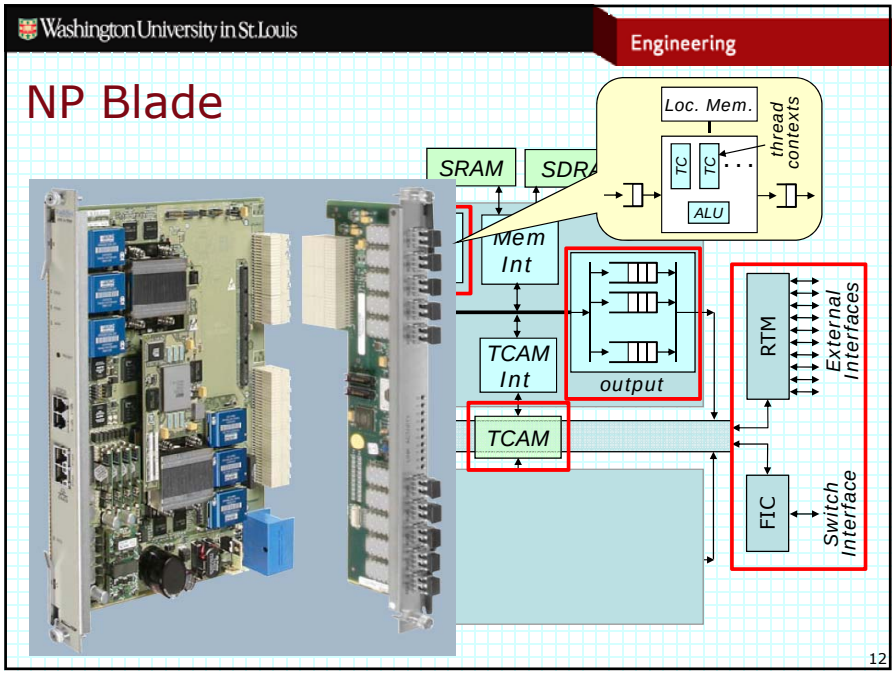
Setting Up a Fast Path



- Request fastpath through RMP
- SRM allocates fastpath
- Specify logical interfaces and interface bandwidths
- Specify #of filters, queues, binding of queues to interfaces, queue lengths and bandwidths
- Configure fastpath filters

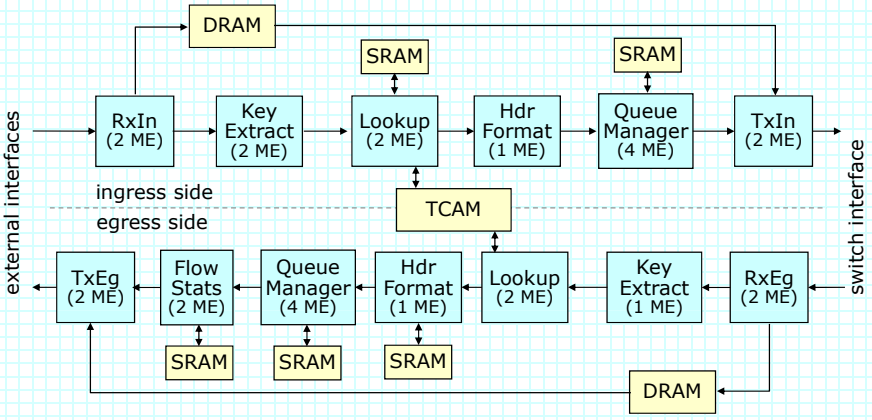
10


Washington University in St. Louis
Engineering
Planned Wide-Area OpenFlow
NOX
CP
SC
parse
lookup
hdrFmt
queue
NOX
to SPP
Stanford
WashU
Princeton
GaTech
Texas
2
2
2
3
3
3
Internet2 Regeneration and Add/Drop site
Internet2 Redundant Drop/Add Site
ESnet Drop/Add Site
Internet2 Optical Switching Node
Internet2 Router Site
11


Washington University in St. Louis
Engineering
NP Blade
Loc. Mem.
thread contexts
TC
TC
ALU
SRAM
SDR
Mem Int
TCAM Int
TCAM
output
RTM
External Interfaces
FIC
Switch Interface
12

Washington University in St. Louis Engineering

Line Card Datapath

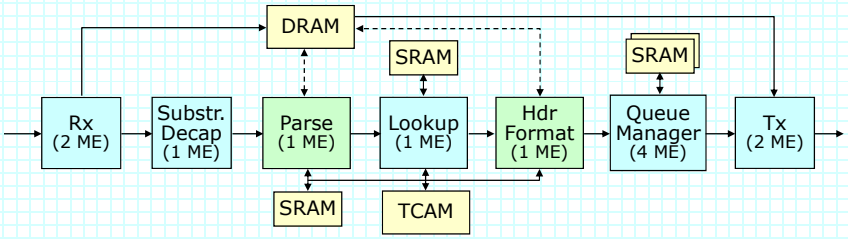


- Filter/route and rate-control traffic
- Network Address Translation for outgoing flows
- Record traffic statistics for all outgoing flows

13

Washington University in St. Louis Engineering

NPE Datapath (version 1)

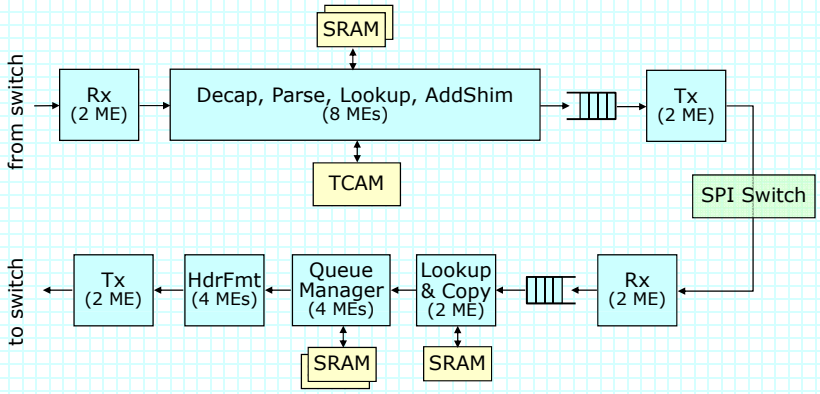


- Parse and Header Format include slice-specific code
  - » parse extracts header fields to form lookup key
  - » Hdr Format does any required post-lookup processing
- Lookup uses opaque key for TCAM lookup
- Multiple static code options can be supported
  - » multiple slices per code option
  - » each has own interfaces, filters, queues and private memory

14

Washington University in St. Louis Engineering

NPE Datapath (Version 2)



- Use both NPs, enabling 10 Gb/s throughput
- Integrated Decap,Parse,Lookup uses MEs more efficiently
- Multicast supported by substrate

15

Washington University in St. Louis Engineering

Project Plan Highlights

- Initial deployment
  - first two (maybe three) nodes deployed by mid 2009
  - shooting for v2 of NPE software
  - PlanetLab control with local resource allocation
- GENI-compatible control software
  - implement component manager interface
  - resource allocation using rspecs/tickets
- Working with users
  - online and hands-on tutorials
  - collaborating with users on new code options
- Completing deployment
  - final nodes deployed in late 2010
  - complete support for netFPGA

16

Looking Ahead

- Bad news
  - » slow market for ATCA means high cost, limited support
  - » Intel dropped IXP and Radisys discontinuing IXP blades
- Good news
  - » ATCA market now projected to grow rapidly and become more cost-competitive (10x growth over 3 years)
  - » new NPs and NP blades
    - Netronome 3200 – IXP successor with 40 microengines
    - Cavium Octeon, RMI XLR732 – MIPS-based, uses cache
  - » can also assemble systems from commodity components
    - 10 GbE switches now at $400-500 per port
    - conventional rack-mount servers with 8-16 processor cores
    - NPs and FPGAs on lower cost PCI-express cards

17