Network Virtualization

Jong-won Lee

School of CSEE
Handong University

Outline

- Background
- Virtualization & Network Virtualization
  - Definition
  - Current available techniques
- Overlay Networks
  - OverQoS, *i3* (Internet Indirection Infrastructure)
  - Interaction Effect of Overlay Networks
- Approaches to Network Virtualization
  - Carbo
  - SPP
  - OpenFlow Switch
- Concluding Remarks

Background

- Internet state
  - "Internet is under stress"
  - "Internet is in an Impasse"
  - "Internet has become ossificated"
  - "Internet is unable to integrate new ideas, new architectures, and to provide paths for future integration"
- Need a large-scale, realistic testbed for evaluating a new architecture and protocols
  - Support multiple experiments running in parallel
  - Carry real traffic

⇒ **Virtualized Network**

Virtualization

- The definition of virtualization in computing
  - A broad term that refers to the abstraction of computer resources
  - A technique for hiding the physical details of computing resources from the way in which other systems, applications, or end users interact with those resources.
  - Ex.
    - Virtual memory,
    - Vmware ESX 3.5, Xen (open standard hypervisor)
    - Cf.) Hypervisor
      - A virtualization platform that allows multiple operating systems to run on a host computer at the same time.

Virtualization

- The rapidly changing demands of the modern business require a flexible and highly adaptable IT infrastructure.
- The virtualization of resources plays a key role in achieving the required degree of adaptability.
- Therefore, the term virtualization is heard in many areas, including the virtualization of servers, applications, storage devices, security appliances, and, not surprisingly, the network infrastructure.

Network Virtualization

- An architectural approach to providing a separate logical networking environment for each group
- These logical environments are created over a single shared network infrastructure.
- Each logical network provides the corresponding user group with full network services similar to those provided by a traditional non-virtualized network.
  - Sharing and isolation

Network Virtualization

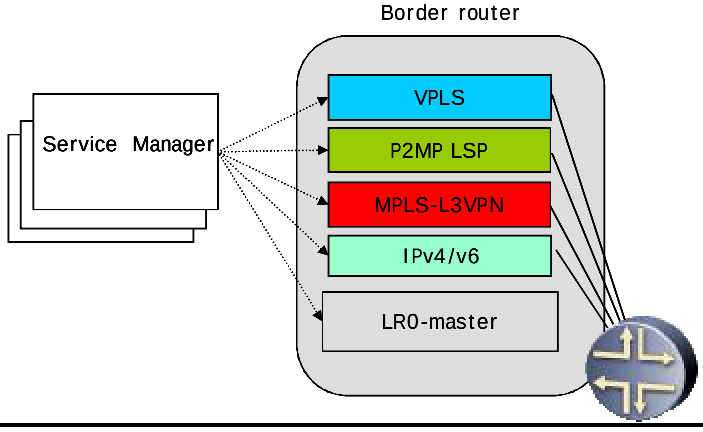
- Data-path virtualization
  - The virtualization of the interconnection between network devices, that is, the traffic separation enforced across a network path.
  - Ex.: An Ethernet link can be virtualized by means of 802.1q VLAN.
  - Ex.: For frame-relay or ATM, separate VCs provide data-path virtualization.
- Network device virtualization
  - Includes the virtualization of all processes, DB, tables, and interfaces
  - Two planes to be virtualized: control plane and forwarding plane

Virtual Networks: currently available techniques

- VLAN (virtual LAN)
  - One single switched LAN can be made to look like separate LANs (virtual LANs)
  - for security and performance
- VPN (Virtual Private Networks)
  - Networks perceived as being private networks by customers using them, but built over shared infrastructure owned by service provider
  - for privacy, security, and cost

Virtual Networks: currently available techniques

- Logical Router (LR) of Juniper Networks
  - A separate administration plane for each LR.
  - A dedicate routing process for each LR.
  - A converged forwarding plane for uplink
- Multi-Layer Service Network (MLSN)



Overlay Networks

Overlay Networks

- Overlay network
  - A network built on top of one or more existing networks
    - Without deploying entirely new networking equipments
  - Adds an additional layer
  - Changes properties in one or more layers of underlying network
  - All the networks after PSTN have begun as overlay networks
    - The Internet is also an overlay network.

Overlay Networks

- Overlay networks are popular in the Internet
  - To provide additional functions that the Internet does not support
  - Ex.:
    - For security: VPN
    - For multicast: M-Bone
    - For QoS: OverQoS
    - For Resilient routing: RON
    - Etc.

OverQoS

§ Background

- Today's Internet provides only best-effort service.
- Difficult to change IP infrastructure
  - Underlying routers should be equipped with QoS mechanisms such as Intserv or Diffserv.

§ Approach

- An overlay based architecture for enhancing Internet QoS
- Using a CLVL (controlled loss virtual link) abstraction

L. Subramanian and Ion Stoica et. al., "OverQoS: An Overlay based Architecture for Enhancing Internet QoS," CCR, vo;.33, no.1, 2003.

OverQoS

§ Architecture

- Pre-determined placement of overlay nodes
- Fixed end-to-end paths between overlay nodes

§ QoS Enhancements

- Smoothing losses
- Packet prioritization
- Statistical loss and bandwidth guarantee

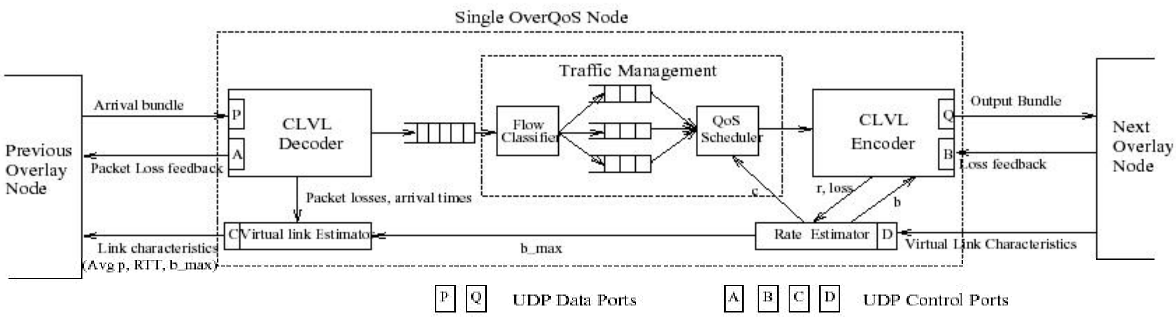
OverQoS

§ Principles to achieve the goals

- **w** Bundle loss control :
  - A statistical target loss rate by a hybrid FEC/ARQ
    - Constraint: the aggregated arrival rate of the bundle should not exceed its available BW.
- **w** Resource management within a bundle
  - The statistical BW guarantee is given a flow with a higher priority.
    - Empirical data for min available BW ($P(c < c_{min}) < 0.01$)
      - 160kbps (Sweden Lulea-Korea), 269kbps (Boston-SF), 420kbps (Netherlands-Intel(SF))

OverQoS

§ Node architecture



- **w** Via UDP socket
- **w** CLVL encoder/decoder implement the CLVL.
  - The decoder feedbacks the characteristics of a virtual link.
- **w** The traffic management module implements per-flow resource management.

OverQoS

§ Evaluation

- Audio streaming application
  - Smoothing bursty losses
  - ARQ-based CLVL
  - PESR (perceptual evaluation of speech quality)

		Sample 1	Sample 2
Mazu-Korea	Without OverQoS	4.25 ± 0.3	4.27 ± 0.5
Mazu-Korea	With OverQoS	4.46 ± 0.4	4.45 ± 0.3
Intel-Lulea	Without OverQoS	4.04 ± 0.2	4.13 ± 0.3
Intel-Lulea	With OverQoS	4.19 ± 0.3	4.31 ± 0.3

– 0.15 – 0.2 : a reasonable improvement in the audio quality

OverQoS

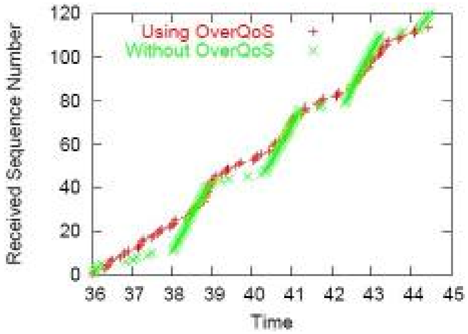
§ Evaluation

- MPEG streaming
  - Packet prioritization (I frame: high priority,)
  - ARQ-based CLVL
  - PSNR (Peak SNR)

		5% PSNR	Median PSNR
Mazu-Korea	Without OverQoS	15.27	22.33
Mazu-Korea	Using OverQoS	17.4	24.95
Intel-Lulea	Without OverQoS	14.68	21.59
Intel-Lulea	Using OverQoS	16.21	24.7

OverQoS

- Evaluation
  - Conterstrike : a team-based multi-player game
    - Under the environment of an high loss rate of 10%
    - An FEC+ARQ-based CLVL
    - Higher priority to control packets
    - Smoothly drop data packets
    - Received sequence number



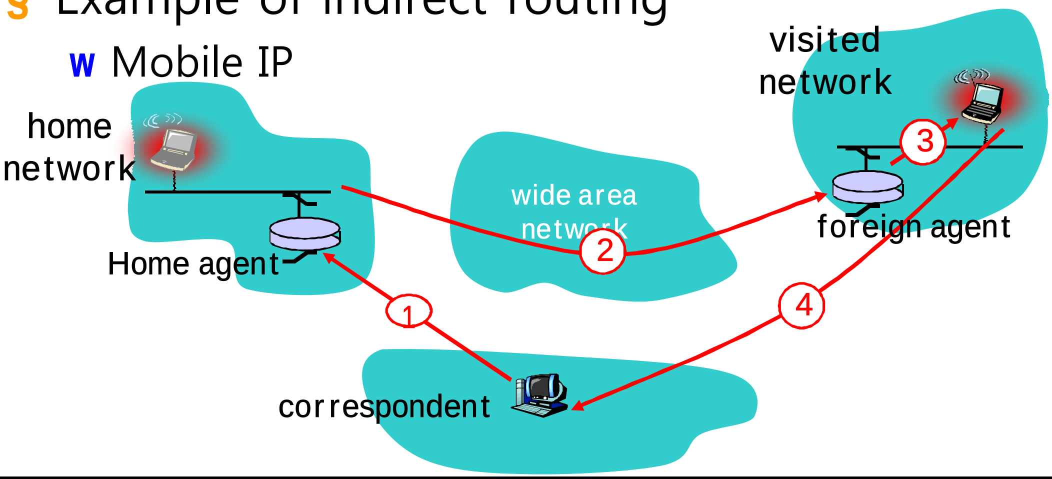
i3: Motivation

Ion Stoica et. al., "Internet Indirection Infrastructure," Proc. of SIGCOMM, 2002.

- Original Internet architecture
  - Unicast point-to-point communication
    - Send packets from a host 'A' to a host 'B'
    - One receiver at a fixed and well-known location
  - Not well suited for applications that requires
    - Mobility (one to anywhere)
    - Multicast (one to many)
    - Anycast (one to any)
- Change communication abstraction
  - Use the concept of indirection
  - Decouple sending from receiving
  - Use a rendezvous-based commun. abstraction

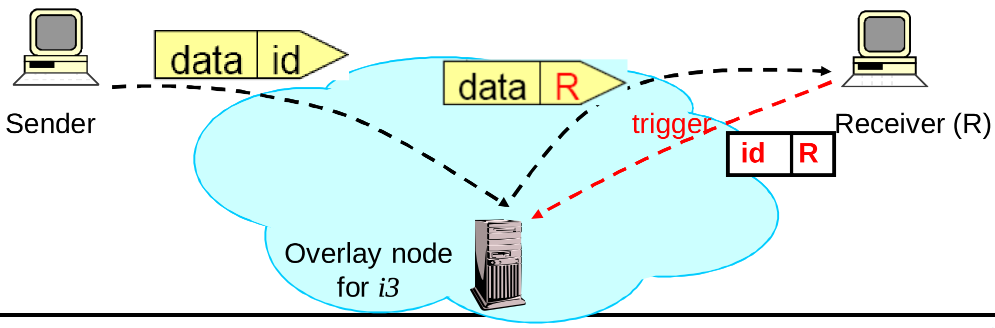
i3: Indirection

- Indirection
  - Rather than reference an entity directly, reference it indirectly via another entity, which in turn can access the original entity
- Example of indirect routing
  - Mobile IP



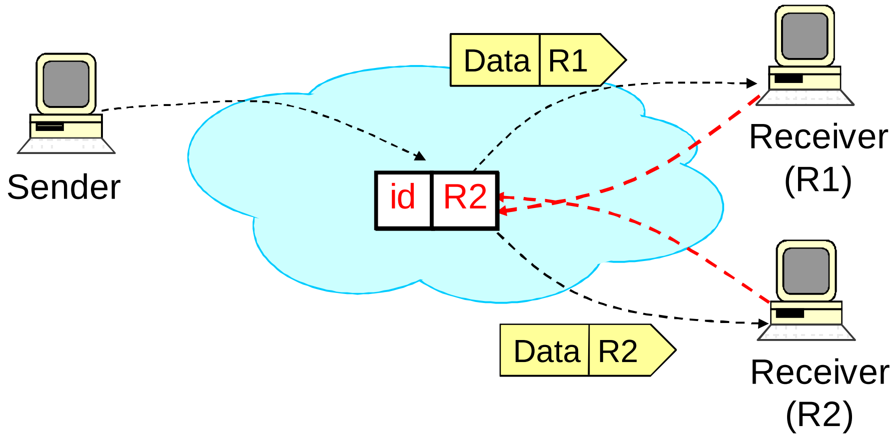
i3: Overview

- Communication model
  - Send packets with ID instead of a dest. add.
  - To receive the packets with ID, a receiver issues a trigger (id, address) into a network.
- Add an indirection layer on top of IP
  - Use an overlay network to implement i3
  - Application layer publish-subscribe infrastructure



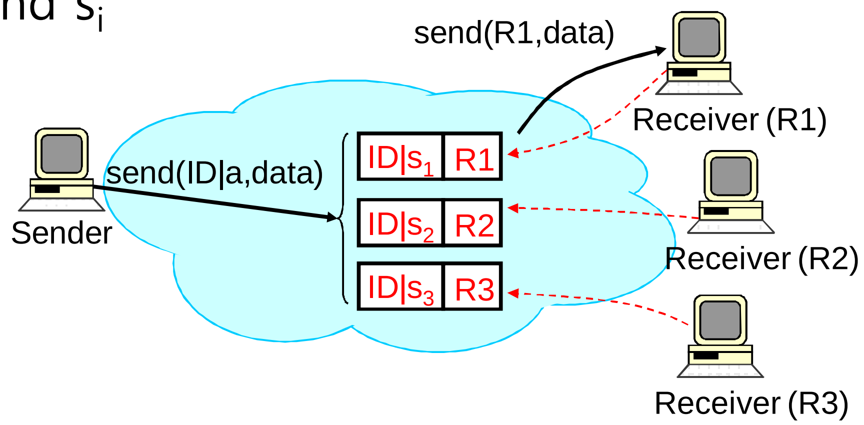
i3: Mobility

- A receiver just needs to update its trigger as it moves from one subnet to another
  - Transparent to a sender



i3: Anycast

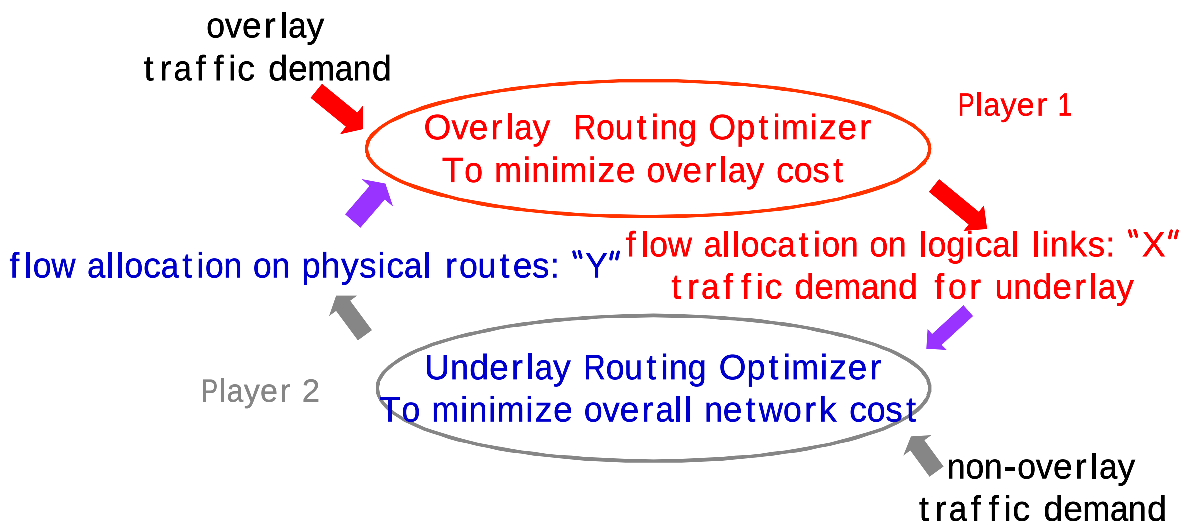
- route to any one in set of receivers
- receiver *i* in anycast group inserts same ID, with anycast qualifications
- route to receiver with best match between a and s_i



Interaction Effect in Overlay Networks

- IP routing is
  - Optimized for system-wide criteria (e.g., minimize maximum link utilization)
  - Often sub-optimal in terms of user performance
    - Because of policy routing, etc.
- Application-level control
  - For its own (selfish behavior)
  - On top of existing network control
  - Try to improve performance for individual
- Q: How does the overlay routing
  - affect overall network performance?
  - affect non-overlay traffic performance?

Interaction between Overlay and IP-layer



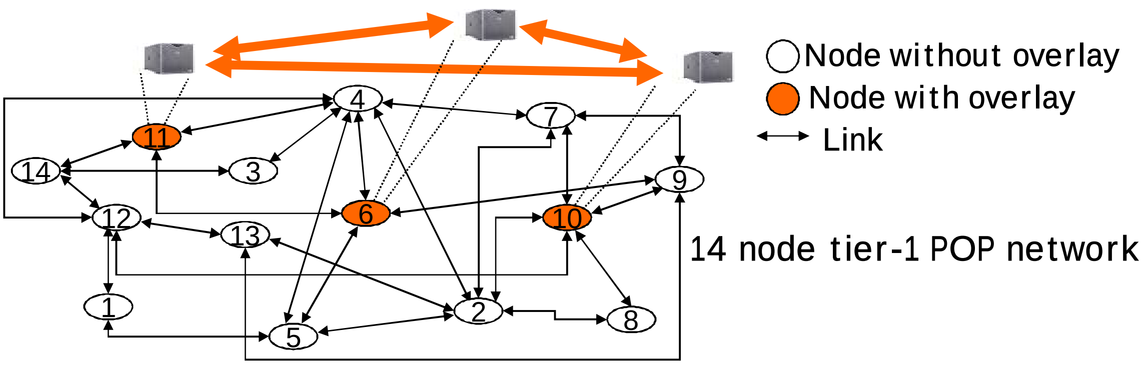
Iterative Dynamic Process

- equilibrium: existence? uniqueness?
- dynamic process: convergence? oscillations?
- performance of overlay and underlay traffic?

Interaction between Overlay and IP-layer

§ Simulation study

- Routing models:
  - Optimal underlay routing (minimize total delay for all network traffic)
  - Optimal overlay routing (minimize total delay for all overlay traffic)
- Topology



Interaction between Overlay and IP-layer

§ Simulation result (10% overlay traffic)

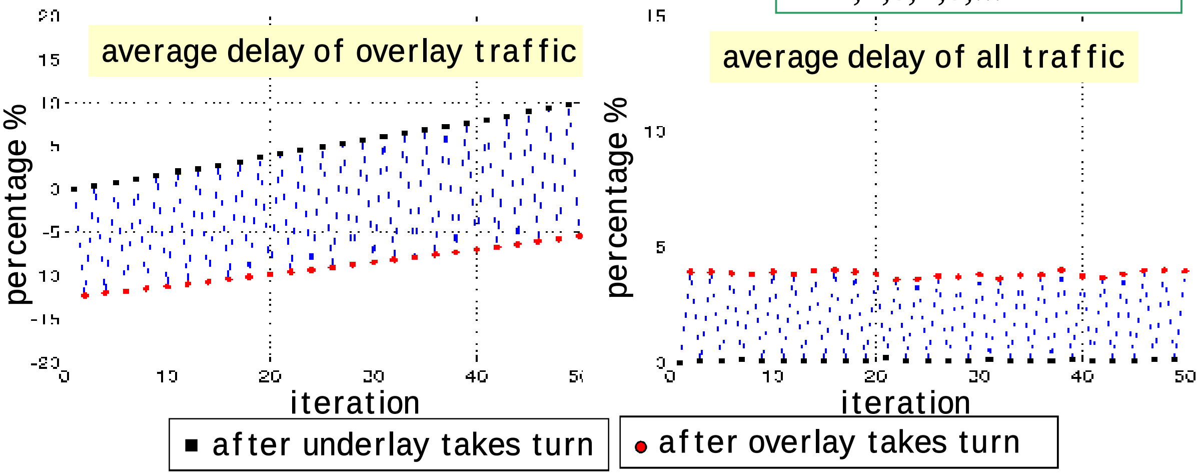
Iterative process

§ Underlay takes turn at step 1, 3, 5, ...

§ Overlay takes turn at step 2, 4, 6, ...

$$\frac{\text{Delay}(k) - \text{Delay}(1)}{\text{Delay}(1)} 100\%$$

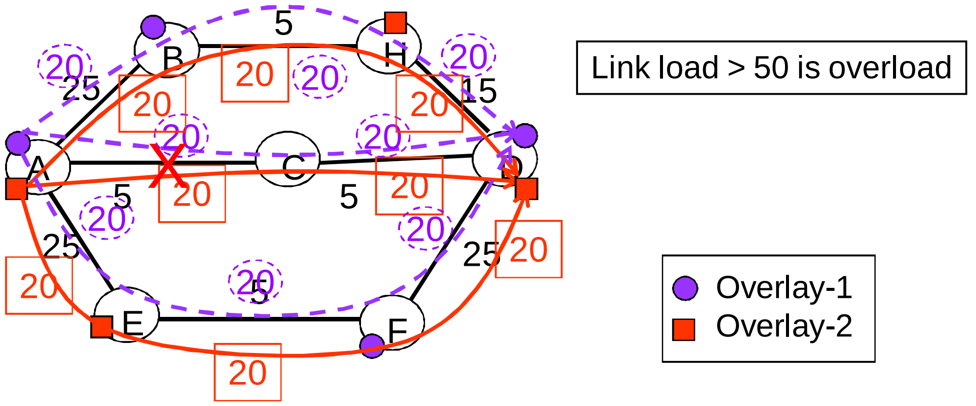
$k = 1,2,3,4,5,\ldots$



[Liu'05] Y. Liu, H. Zhang, W. Gong, D. Towsley, "On the Interaction Between Overlay Routing and Underlay Routing," Infocom 2005.

Interaction between multiple overlay

- Multiple overlays can get synchronized (horizontal interactions)
  - Can impact both overlay and non-overlay traffic
  - leading to oscillations



R. Keralapura, N. Taft, C-N. Chuah, and G. Iannaccone, "Can ISPs take the heat from Overlay Networks?" HotNets-III, November 2004

Interaction Effect in Overlay networks

- Selfish overlay routing can degrade performance of network as a whole

- Interactions between blind optimizations at two levels may lead to lose-lose situation

- Multiple overlay can cause traffic to oscillate.

Approaches to Network Virtualization

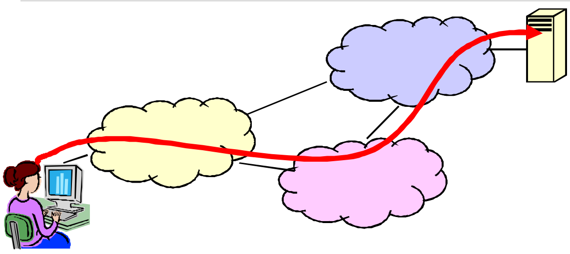
- Cabo
- SPP
- OpenFlow Switch

Cabo (Concurrent Architectures are Better than One)

§ The Internet inside is quite difficult to change

- Ex.: IPv6, QoS, multicast etc.
- Ex.:
  - multi-ISP VPN,
  - path with end-to-end performance guarantee

Competing ISPs with different goals must coordinate



Cabo

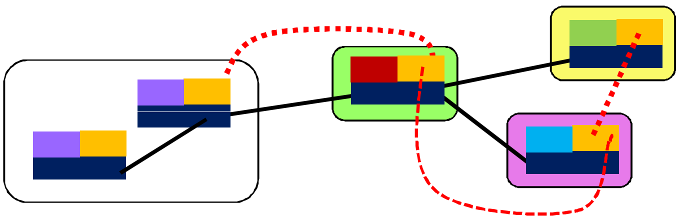
- To facilitate the deployment of new protocols and architectures
  - Decouple infrastructure provider from service provider
  - Infrastructure provider:
    - deploy and maintain physical infrastructure, that is, links and routers.
  - Service provider:
    - deploy network protocols and offer end-to-end service to users
    - An organization that composes network services and protocols on top of physical infrastructure

Today: ISPs try to play both roles, and cannot offer *end-to-end* services

Cabo

- Cabo as a new architecture
  - Ex.: multi-ISP VPN,
    - path with end-to-end performance guarantee

Single service provider controls end-to-end path

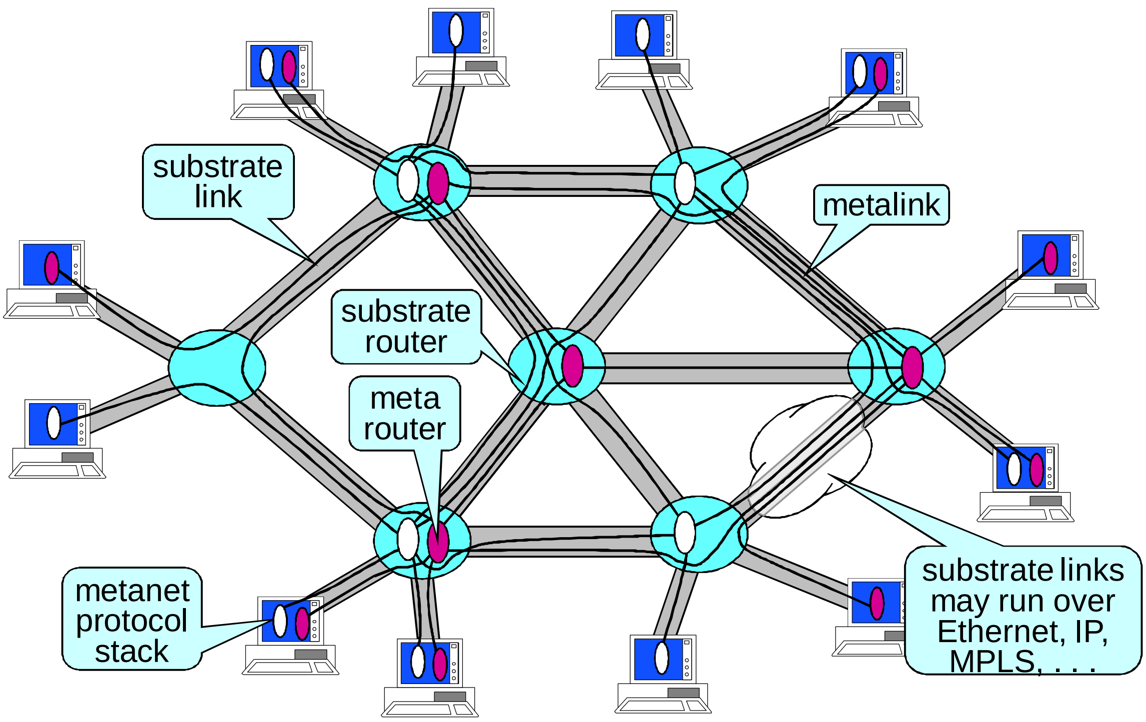


  - Economic refactoring
    - Infrastructure and service providers
  - Virtualization
    - Multiple virtual networks (virtual nodes and virtual links)

Supercharged PlaetLab Platform (SPP)

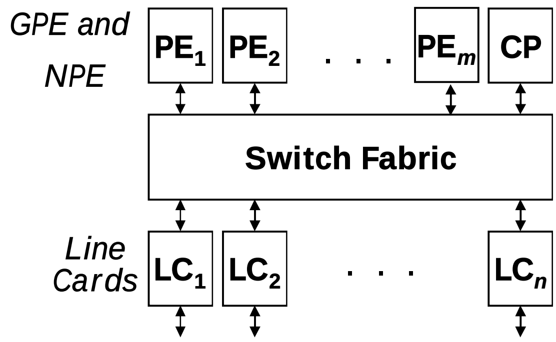
- "An Architecture for a Diversified Internet" project in GENI
- At Washington Univ. in St. Louis
- Objectives
  - Design an experimental PlanetLab platform capable of much higher levels of performance
  - To support high performance overlay hosting services
- High performance platform
  - General-purpose servers and high performance network processor subsystems

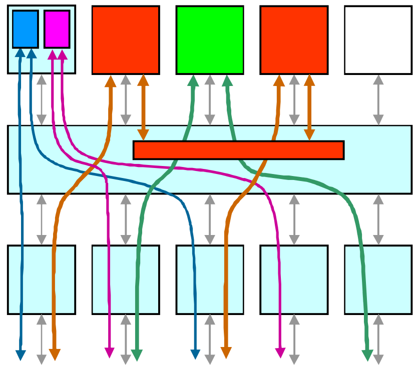
SPP: As a substrate router for virtualized net.



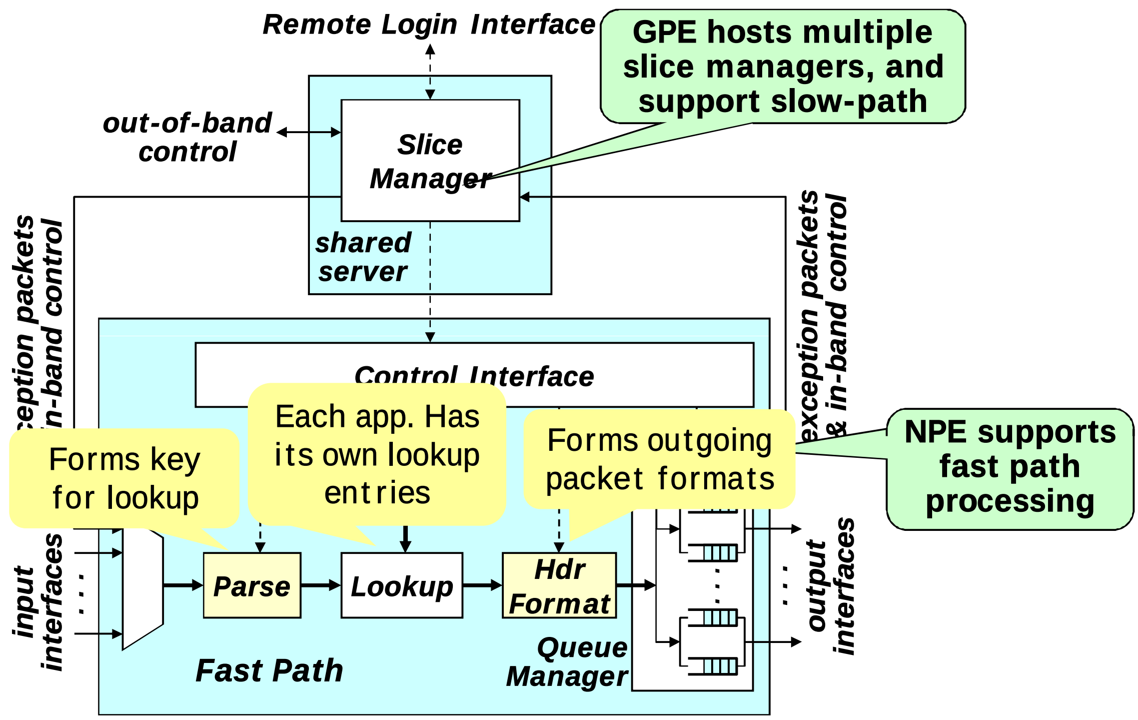
SPP Architecture

- Line cards support multiple meta-lines.
- Processing engines (GPEs and NPEs) are used to implement meta-routers
- Non-blocking Switch fabric guarantees traffic islolation between meta-routers





SPP: data-path



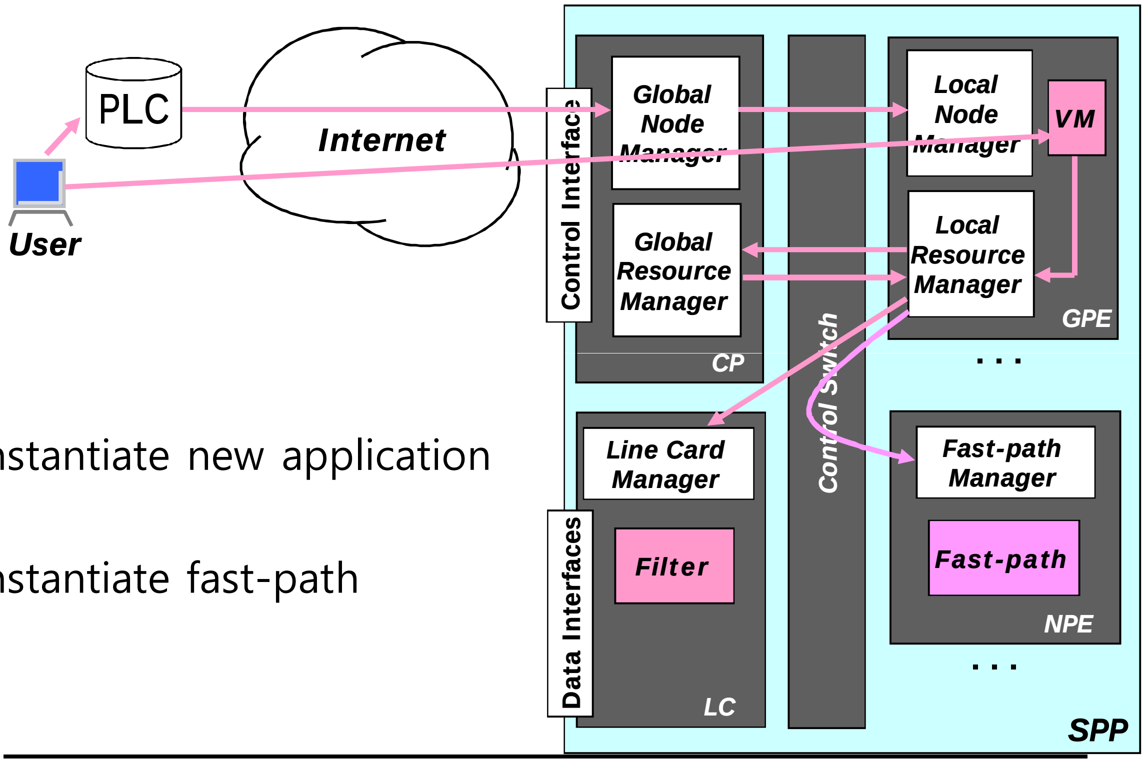
SPP: An example

- Based on ATCA
  - 6 slot shelf
- 2 NP blades
  - 2 IXP2850 NPs/blade
  - 1 NP blade for line card
  - 1 NP blade for supporting fast path
- 1 Switching fabric blade
  - 10GbE and 1GbE switching fabric
  - Supports VLAN
- 2 Intel server blades
- Power supply & CM

SPP

- Based on NP
  - High performance
    - 4.8 M packets/sec for IPv4 forwarding app.
  - Expensive
  - Virtualization of NP resources
    - Max. 8 threads/ME
      - It has HW threads which are operated in a round-robin fashion
    - 16 MEs for IXP2850 (8 MEs for IXP2400)
    - A small program memory/ME
      - 8k for IXP2800, 4k for IXP2400
    - Provide a dedicated FIFO, called next neighbor FIFO, between consecutive pairs of MEs

SPP: Control



- Instantiate new application
- Instantiate fast-path

OpenFlow Switch

- EASI (Enable And Scale Innovation) project of the Stanford Clean Slate Program
- EASI goal
  - enabling fundamental changes to the Internet architecture
  - lowering barrier-to-entry for scalable service deployment
- Strategy
  - A pragmatic compromise
    - *Allow researchers to run experimental protocols in their network…*
    - *…without requiring vendors to expose internal workings*

OpenFlow Switch

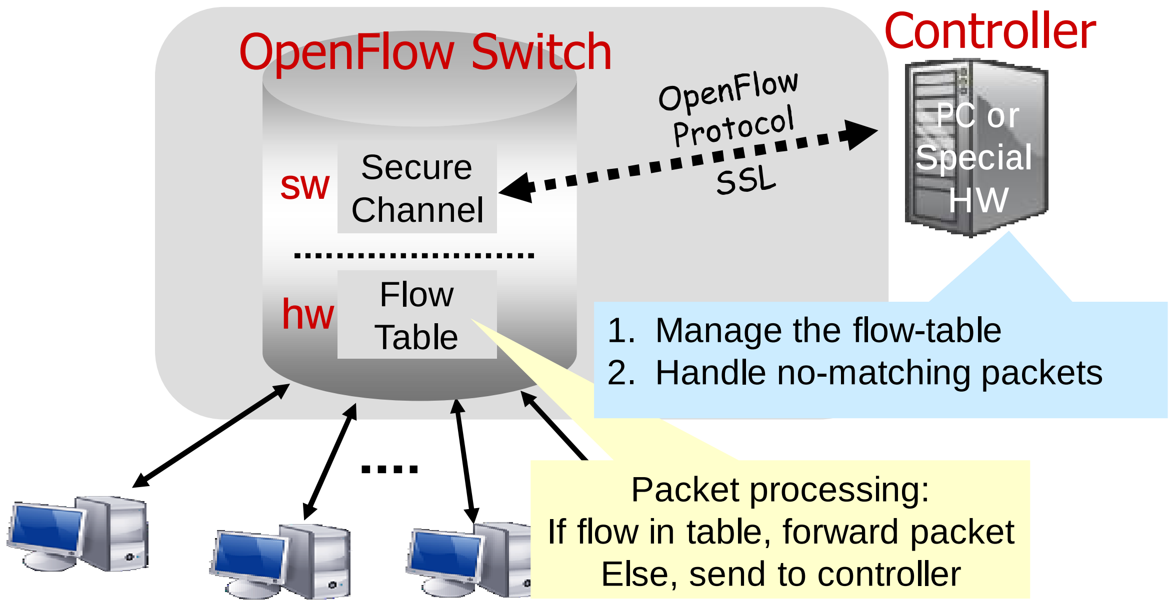
- background
  - Virtualized programmable networks could lower the barriers to entry for new ideas
  - The research platforms
    - Insufficient performance (open software paltforms)
    - Too expensive (SPP of Washington Univ.)
    - Too small number of ports (NetFPGA of Stanford Univ.)
  - Commercial vendors will not provide an open, programmable, virtualized platform on their swithces and routers
    - Complexity of support
    - Market protection and barrier to entry

OpenFlow Switch

- An OpenFlow switch consists of
  - A flow table, which is used for packet lookup and forwarding, and
  - A secure channel to external controller
- The flow table
  - Contains a set of flow entries, activity counters, and a set of actions
  - Is used for matching and forwarding packets
- The exteranl controller
  - Handles the packets without valid flow entries
  - Manages the switch flow table by adding and removing flow entries

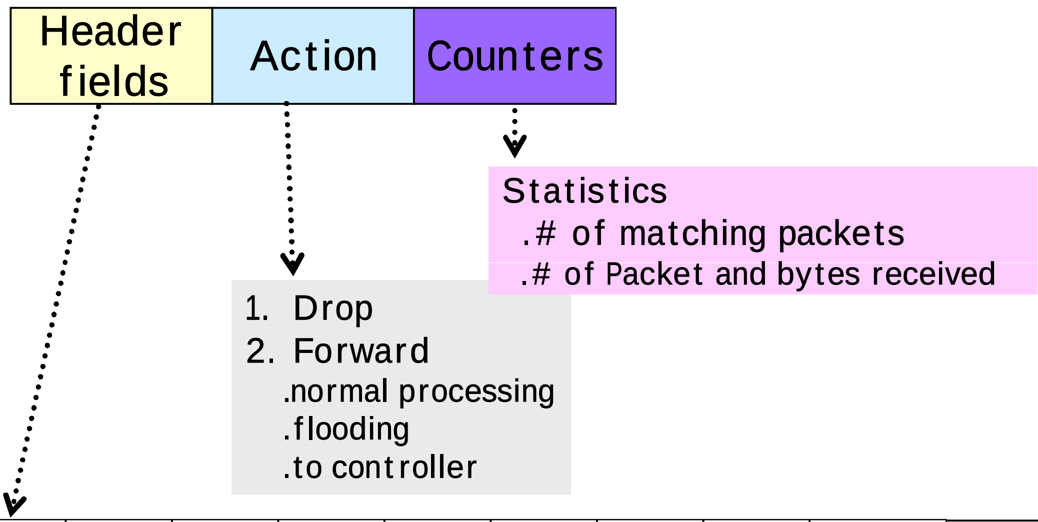
OpenFlow Switch

- The conceptual architecture



OpenFlow Switch

- The flow table entry (for type 0)



Ingress Port	MAC DA	MAC SA	Eth type	VLAN ID	IP SA	IP DA	IP Prot	TCP/UDP SrcPort	TCP/UDP DstPort

OpenFlow Switch

- Simple interface
  - Vendors do not need to expose the internal workings of their switches or routers
- Can be used for experiments
  - at the flow level
  - at the packet level
  - for non-IP protocols

OpenFlow Switch

- Current status
  - Plan to deploy at the Stanford CS and EE buildings
  - Reference switches
    - based on NetFPGA and 48-port 1GE switch based on Broadcom reference design
  - Commercial Ethernet switches and routers
    - Working with six vendors to add to existing products
    - Expect OpenFlow "Type 0" to be available in 2008-09

Concluding remarks

- Network virtualization is a mean or an end?
- Network progammability and virtualization
  - Interesting to network device manufacturers?
  - Interesting to service provider?
    - Real challenge is how to quickly and safely deploy new services.
- Multiple levels/granularity of virtualization
  - Depending on experiments
  - Depending on technology
- Should consider economical and operational issues (incremental deployment, upgrade etc.) to be successful.

References

- L. Peterson, Thomas Anderson, Scott Shenker, Jonathan Turner, “Overcoming the Internet Impasse through Virtualization,” IEEE Computer, April 2005.
- GENI Planning Group, “GENI Design Principles’” IEEE Computer, Sep. 2006.
- N. Niebert, et. al., “Network Virtualization: A Viable Path towards the Future Internet,” Wireless Personal Commun. Pp. 511-520, 2008.
- D. G. Andersen, H. Balakrishnan, M. Frans Kaashoek, R. Morris, "Resilient Overlay Networks," Proc. 18th ACM SOSP, Oct 2001
- Ion Stoica et. al., “Internet Indirection Infrastructure,” Proc. of ACM SIGCOMM’02.
- Lakshminarayanan Subramanian and Ion Stoica et. al., OverQoS: An Overlay based Architecture for Enhancing Internet QoS, CCR, vo;.33, no.1, 2003.
- Y. Liu, H. Zhang, W. Gong, D. Towsley, “On the Interaction Between Overlay Routing and Underlay Routing,” INFOCOM 2005.

References

- R. Keralapura, N. Taft, C-N. Chuah, and G. Iannaccone, "Can ISPs take the heat from Overlay Networks?" HotNets-III, November 2004
- Nick Feamster, Lixin Gao, Jeniffer Rexford, “How to Lease the Internet in Your Spare Time,” ACM SIGCOMM Computer Communications Review, p. 61-64, January 2007.
- Jon Tuner, et. al., Superchaging PlanetLab-A High Performance, Multi-Application, Overlay Network Platform,” Proc. of SIGCOMM’07.
- Ben Wun, Jon Tuner, et. al., Virtulaizing Network Processors
- Nick McKeown et. Al., OpenFlow: Enabling Innovations in Csmpus Networks. Available at http://OpenFlowSwitch.org
- The OpenFlow Switch Specification. Avaliable at http//OpenFlowSwitch.org

Thank you for your attention!