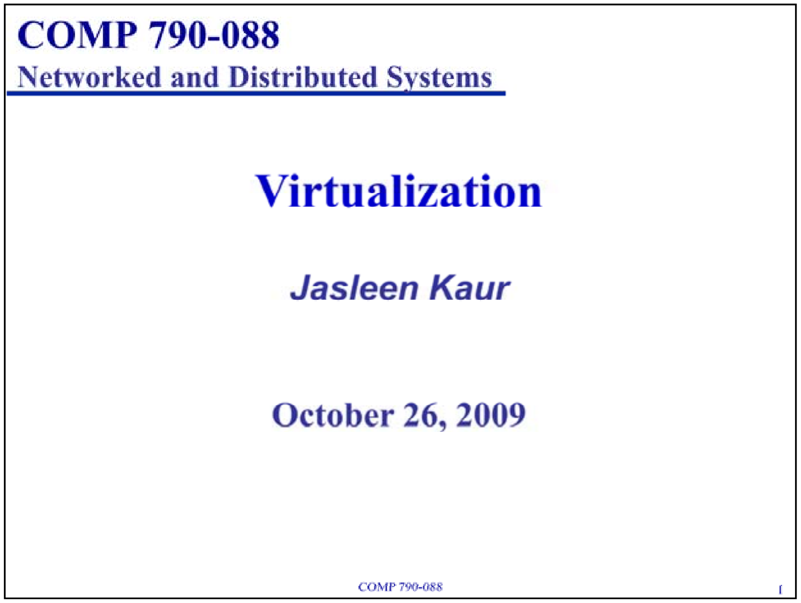


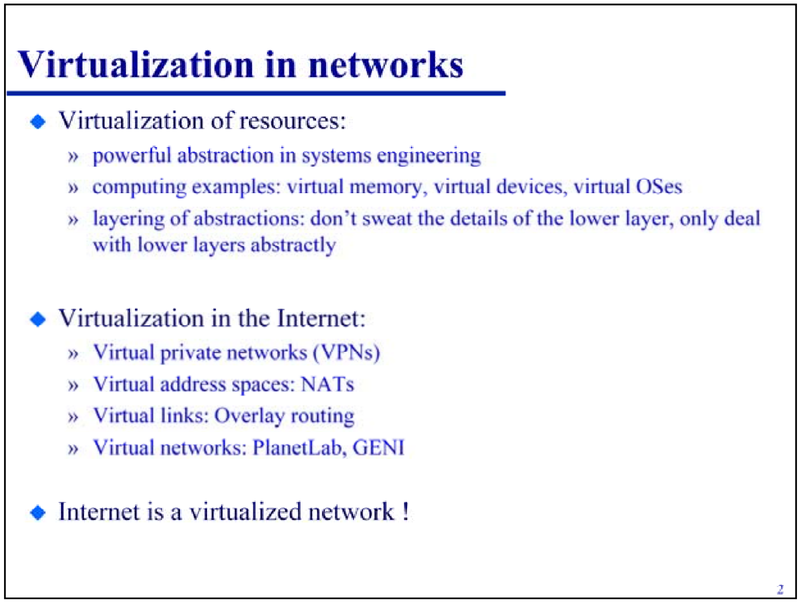
COMP 790-088

Networked and Distributed Systems

Virtualization

Jasleen Kaur

October 26, 2009



Virtualization in networks

- Virtualization of resources:
  - » powerful abstraction in systems engineering
  - » computing examples: virtual memory, virtual devices, virtual OSes
  - » layering of abstractions: don't sweat the details of the lower layer, only deal with lower layers abstractly
- Virtualization in the Internet:
  - » Virtual private networks (VPNs)
  - » Virtual address spaces: NATs
  - » Virtual links: Overlay routing
  - » Virtual networks: PlanetLab, GENI
- Internet is a virtualized network !

The Internet: Virtualizing Local Networks

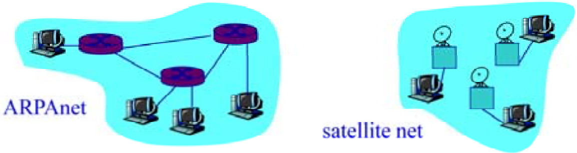
1974: multiple unconnected networks

- » ARPAnet
- » data-over-cable networks
- » packet satellite network (Aloha)
- » packet radio network

.. differing in:

- » addressing conventions
- » packet formats
- » error recovery
- » routing

Cerf & Kahn: Interconnecting Two Networks



- "…interconnection must preserve intact the internal operation of each network."
- " ..the interface between networks must play a central role in the development of any network interconnection strategy. We give a special name to this interface that performs these functions and call it a GATEWAY."
- ".. prefer that the interface be as simple and reliable as possible, and deal primarily with passing data between networks that use different packet-switching strategies
- "…address formats is a problem between networks because the local network addresses of TCP's may vary substantially in format and size. A uniform internetwork TCP address space, understood by each GATEWAY and TCP, is essential to routing and delivery of internetwork packets."

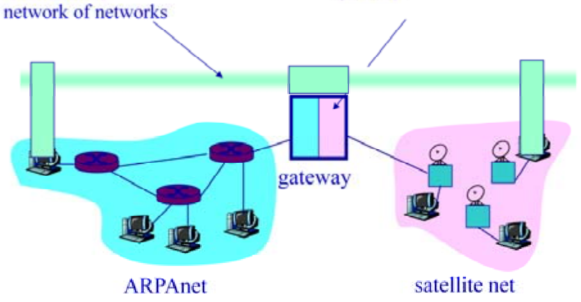
Cerf & Kahn: Interconnecting Two Networks

Internetwork layer:

- addressing: internetwork appears as a single, uniform entity, despite underlying local network heterogeneity
- network of networks

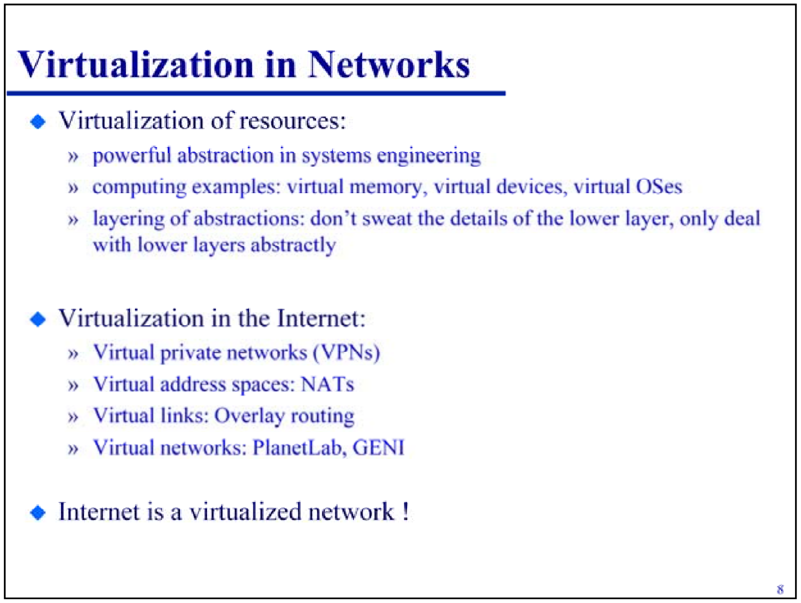
Gateway:

- "embed internetwork packets in local packet format or extract them"
- route (at internetwork level) to next gateway



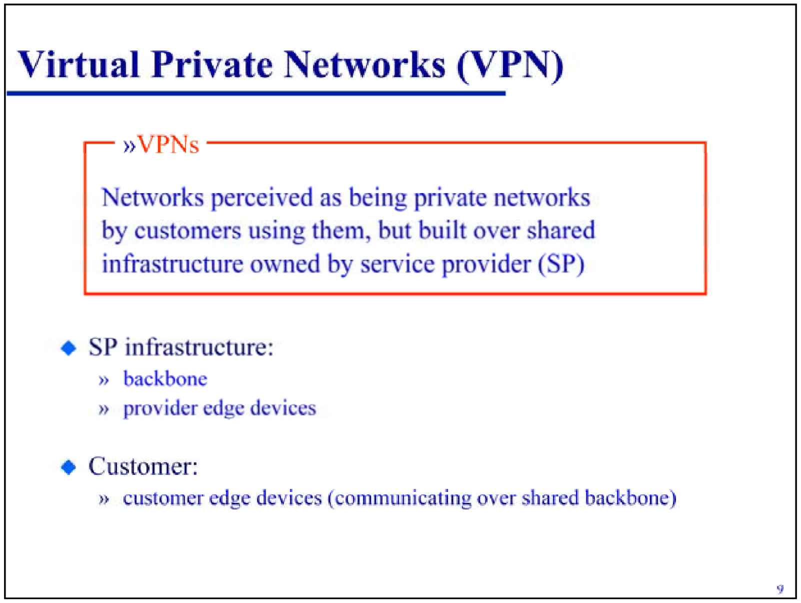
Cerf & Kahn's Internetwork Architecture

- What is virtualized?
  - » two layers of addressing: internetwork and local network
  - » new layer makes everything homogeneous at internetwork layer
  - » underlying local network technology (cable, satellite, 56K modem) is "invisible" at internetwork layer



Virtualization in Networks

- Virtualization of resources:
  - » powerful abstraction in systems engineering
  - » computing examples: virtual memory, virtual devices, virtual OSes
  - » layering of abstractions: don’t sweat the details of the lower layer, only deal with lower layers abstractly
- Virtualization in the Internet:
  - » Virtual private networks (VPNs)
  - » Virtual address spaces: NATs
  - » Virtual links: Overlay routing
  - » Virtual networks: PlanetLab, GENI
- Internet is a virtualized network !



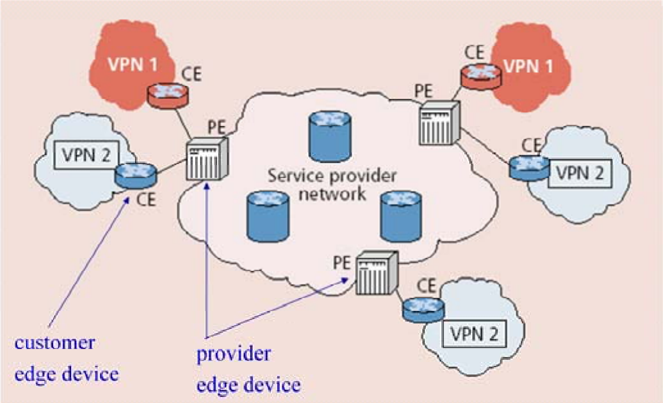
Virtual Private Networks (VPN)

»VPNs

Networks perceived as being private networks by customers using them, but built over shared infrastructure owned by service provider (SP)

- SP infrastructure:
  - » backbone
  - » provider edge devices
- Customer:
  - » customer edge devices (communicating over shared backbone)

VPN Reference Architecture

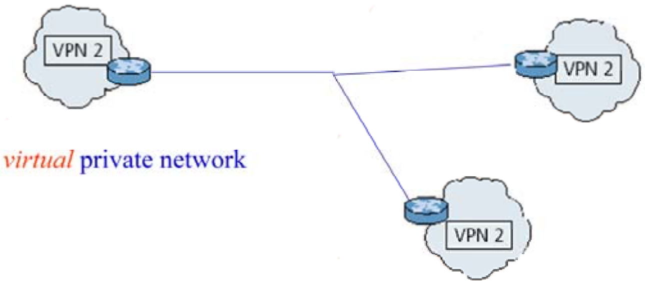


customer edge device

provider edge device

10

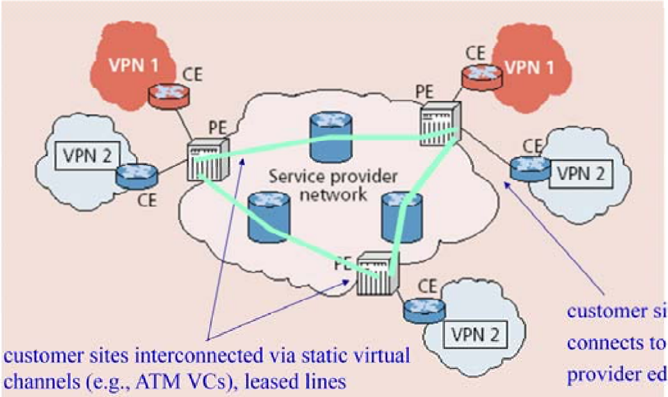
VPN: Logical View



virtual private network

11

Leased-line VPN

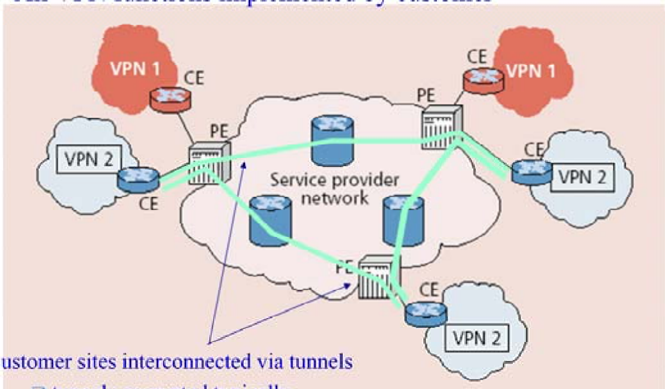


customer site connects to provider edge

customer sites interconnected via static virtual channels (e.g., ATM VCs), leased lines

Customer Premise VPN

All VPN functions implemented by customer



customer sites interconnected via tunnels

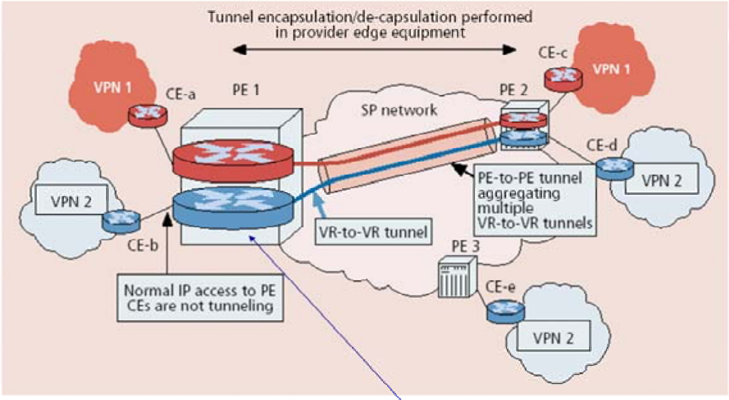
- ❑ tunnels encrypted typically
- ❑ SP treats VPN packets like all other packets

Drawbacks

- Leased-line VPN: configuration costs, maintenance by SP: long time, much manpower
- CPE-based VPN: expertise by customer to acquire, configure, manage VPN

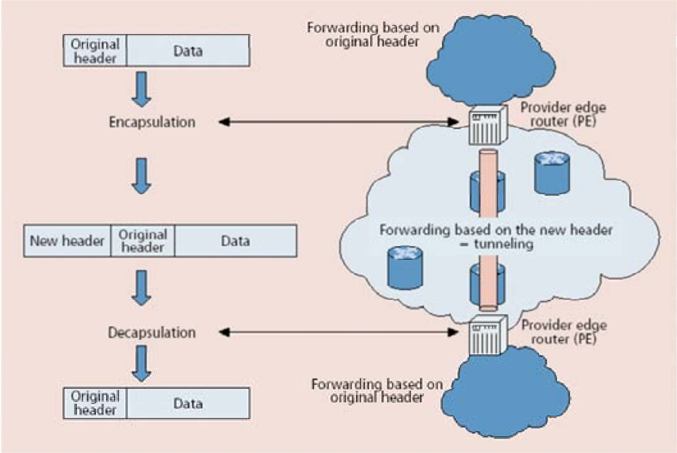
- Network-based VPN
  - » customer's routers connect to SP routers
  - » SP routers maintain separate (independent) IP contexts for each VPN
    - sites can use private addressing
    - traffic from one VPN cannot be injected into another

Network-based Layer 3 VPNs


Tunnel encapsulation/de-capsulation performed in provider edge equipment
VPN 1
CE-a
PE 1
SP network
PE 2
CE-c
VPN 1
CE-d
VPN 2
PE-to-PE tunnel aggregating multiple VR-to-VR tunnels
VPN 2
CE-b
VR-to-VR tunnel
PE 3
CE-e
Normal IP access to PE CEs are not tunneling
VPN 2

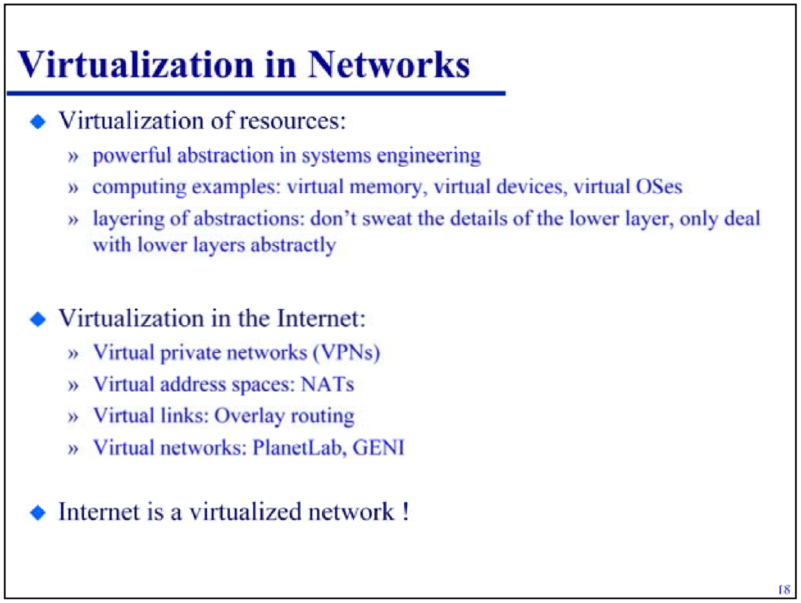
multiple virtual routers
in single provider edge device

Tunneling


Original header
Data
Encapsulation
New header
Original header
Data
Decapsulation
Original header
Data
Forwarding based on original header
Provider edge router (PE)
Forwarding based on the new header = tunneling
Provider edge router (PE)
Forwarding based on original header

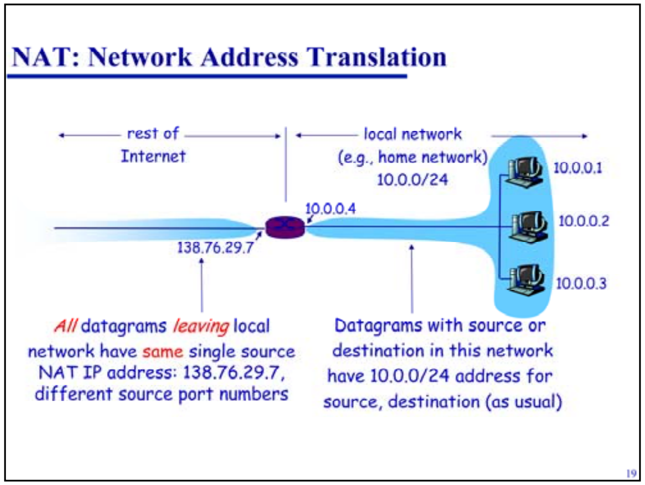
VPNs: Why?

- Privacy
- Security
- Works well with mobility (looks like you are always at home)
- Cost: newer forms of VPNs are cheaper than leased line VPNs
  - ability to share at lower layers (even though logically separate) lowers cost
  - exploit multiple paths, redundancy, fault-recovery in lower layers
  - need isolation mechanisms to ensure resources shared appropriately
- Abstraction and Manageability: all machines with addresses that are “in” are trusted no matter where they are



Virtualization in Networks

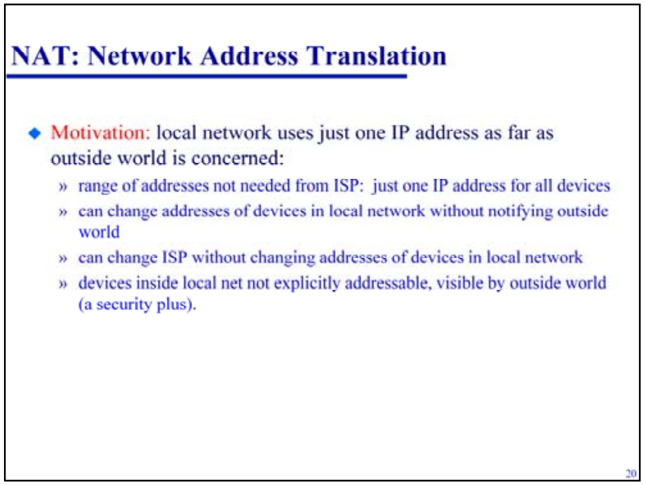
- Virtualization of resources:
  - » powerful abstraction in systems engineering
  - » computing examples: virtual memory, virtual devices, virtual OSes
  - » layering of abstractions: don't sweat the details of the lower layer, only deal with lower layers abstractly
- Virtualization in the Internet:
  - » Virtual private networks (VPNs)
  - » Virtual address spaces: NATs
  - » Virtual links: Overlay routing
  - » Virtual networks: PlanetLab, GENI
- Internet is a virtualized network !



NAT: Network Address Translation

All datagrams *leaving* local network have same single source NAT IP address: 138.76.29.7, different source port numbers

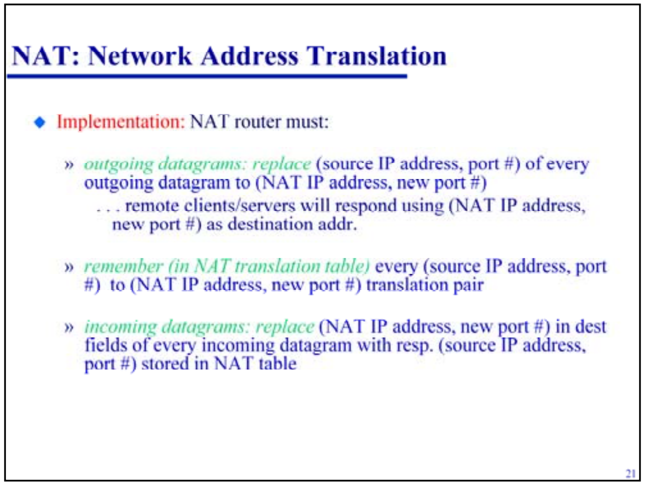
Datagrams with source or destination in this network have 10.0.0/24 address for source, destination (as usual)



NAT: Network Address Translation

- Motivation: local network uses just one IP address as far as outside world is concerned:
  - » range of addresses not needed from ISP: just one IP address for all devices
  - » can change addresses of devices in local network without notifying outside world
  - » can change ISP without changing addresses of devices in local network
  - » devices inside local net not explicitly addressable, visible by outside world (a security plus).

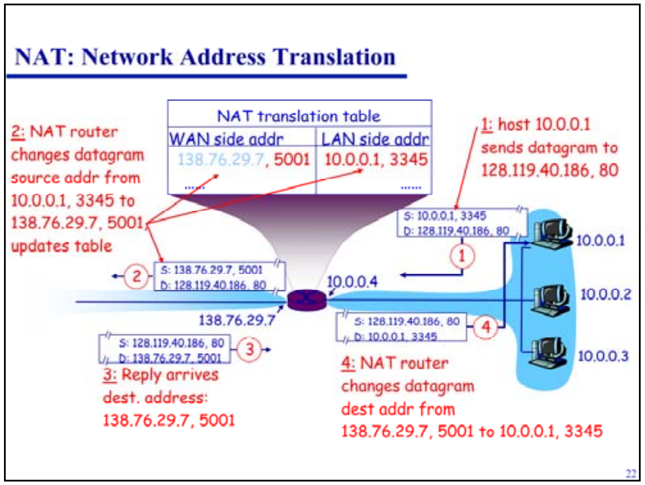
20



NAT: Network Address Translation

- Implementation: NAT router must:
  - » *outgoing datagrams: replace* (source IP address, port #) of every outgoing datagram to (NAT IP address, new port #)
    . . . remote clients/servers will respond using (NAT IP address, new port #) as destination addr.
  - » *remember (in NAT translation table)* every (source IP address, port #) to (NAT IP address, new port #) translation pair
  - » *incoming datagrams: replace* (NAT IP address, new port #) in dest fields of every incoming datagram with resp. (source IP address, port #) stored in NAT table

21



NAT: Network Address Translation

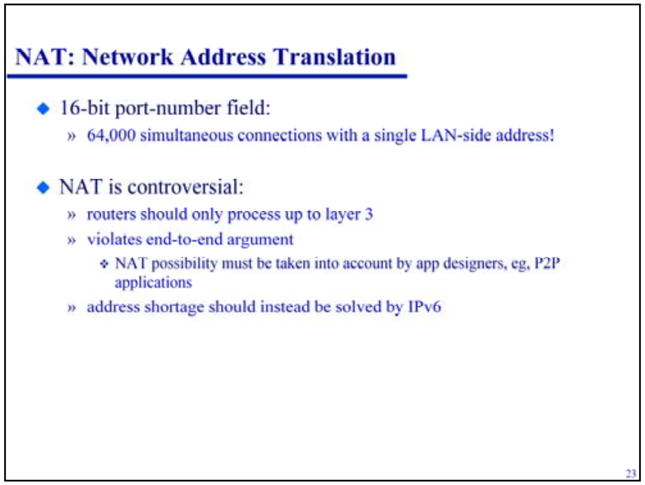
NAT translation table	
WAN side addr	LAN side addr
138.76.29.7, 5001	10.0.0.1, 3345
……	……

1: host 10.0.0.1 sends datagram to 128.119.40.186, 80

2: NAT router changes datagram source addr from 10.0.0.1, 3345 to 138.76.29.7, 5001, updates table

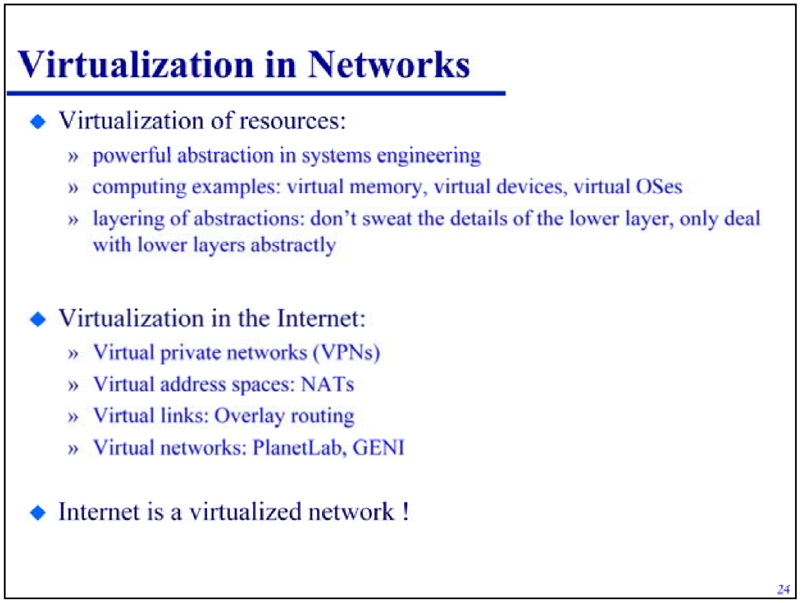
3: Reply arrives dest. address: 138.76.29.7, 5001

4: NAT router changes datagram dest addr from 138.76.29.7, 5001 to 10.0.0.1, 3345



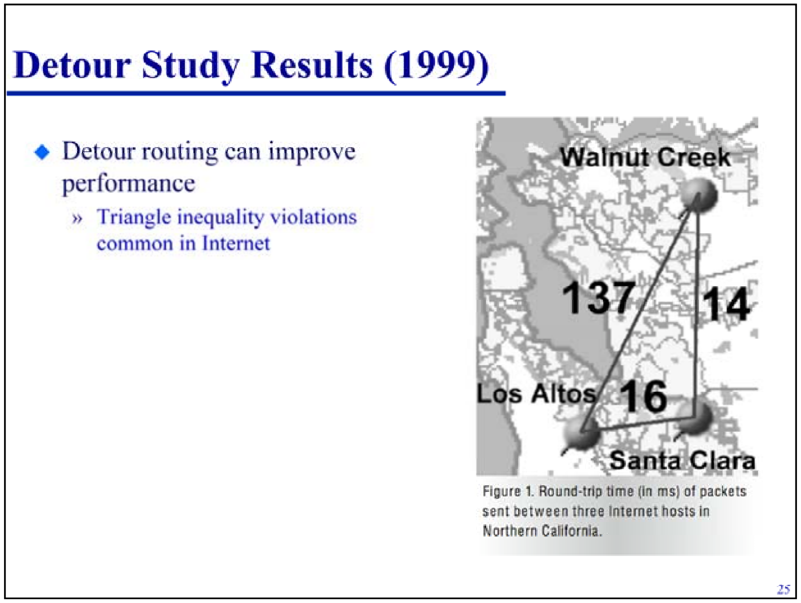
NAT: Network Address Translation

- 16-bit port-number field:
  - » 64,000 simultaneous connections with a single LAN-side address!
- NAT is controversial:
  - » routers should only process up to layer 3
  - » violates end-to-end argument
    - NAT possibility must be taken into account by app designers, eg, P2P applications
  - » address shortage should instead be solved by IPv6



Virtualization in Networks

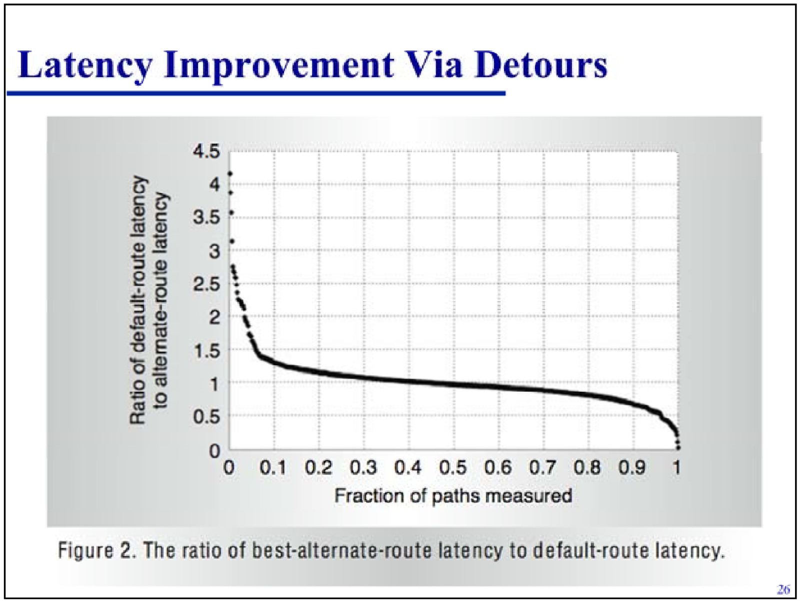
- Virtualization of resources:
  - » powerful abstraction in systems engineering
  - » computing examples: virtual memory, virtual devices, virtual OSes
  - » layering of abstractions: don't sweat the details of the lower layer, only deal with lower layers abstractly
- Virtualization in the Internet:
  - » Virtual private networks (VPNs)
  - » Virtual address spaces: NATs
  - » Virtual links: Overlay routing
  - » Virtual networks: PlanetLab, GENI
- Internet is a virtualized network !



Detour Study Results (1999)

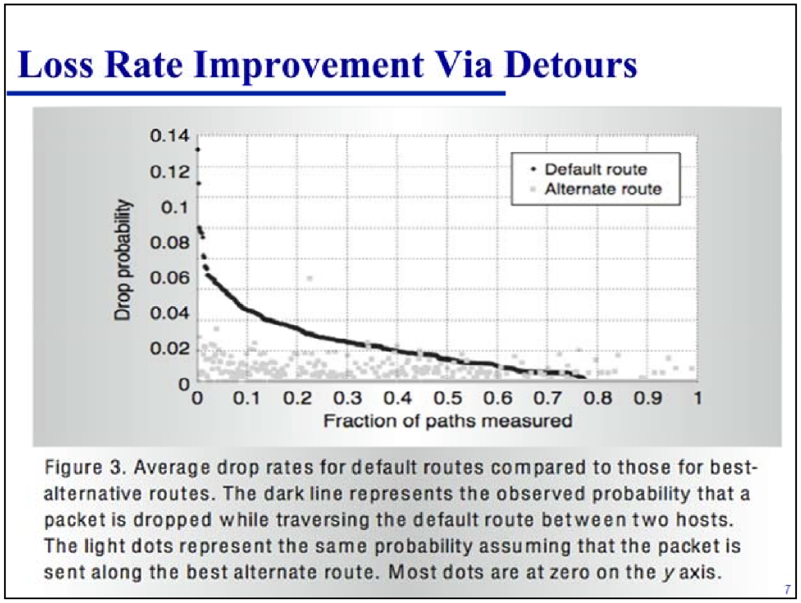
- Detour routing can improve performance
  - » Triangle inequality violations common in Internet

Figure 1. Round-trip time (in ms) of packets sent between three Internet hosts in Northern California.



Latency Improvement Via Detours

Figure 2. The ratio of best-alternate-route latency to default-route latency.



Loss Rate Improvement Via Detours

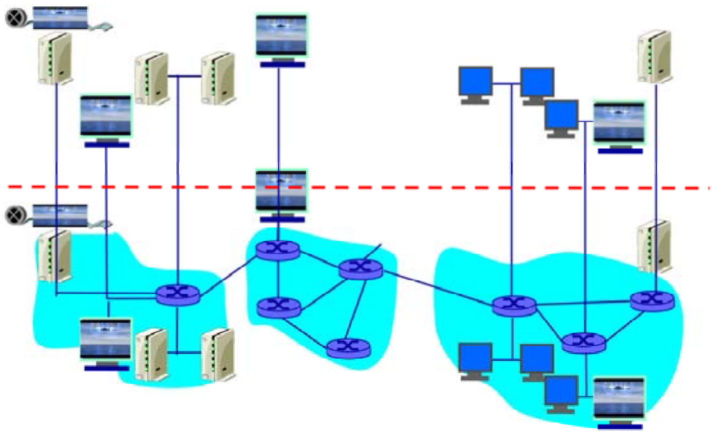
Figure 3. Average drop rates for default routes compared to those for best-alternative routes. The dark line represents the observed probability that a packet is dropped while traversing the default route between two hosts. The light dots represent the same probability assuming that the packet is sent along the best alternate route. Most dots are at zero on the *y* axis.

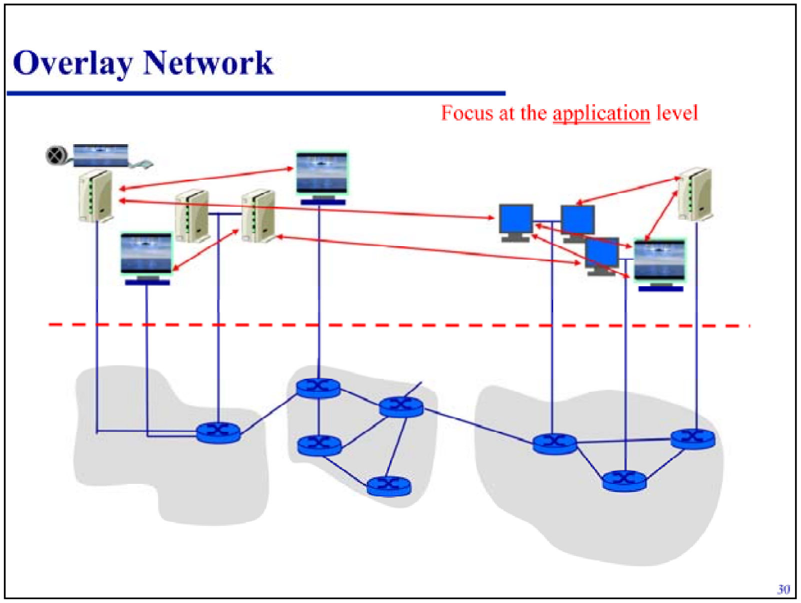
Resilient Overlay Networks

Overlay network:

- Applications, running at various sites as “nodes” on an application-level network
- Create “logical” links (e.g., TCP or UDP connections) pairwise between each other
- Each logical link: multiple physical links, routing defined by native Internet routing

Overlay Network

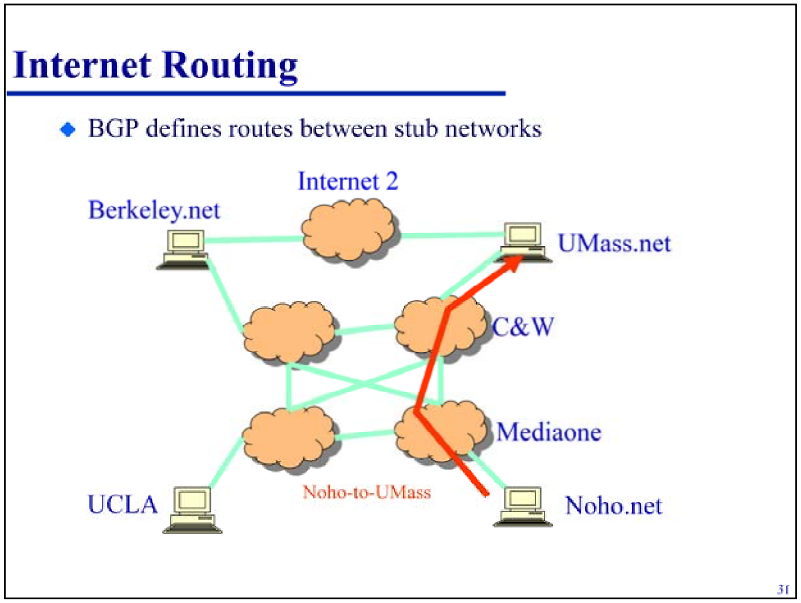




Overlay Network

Focus at the application level

30



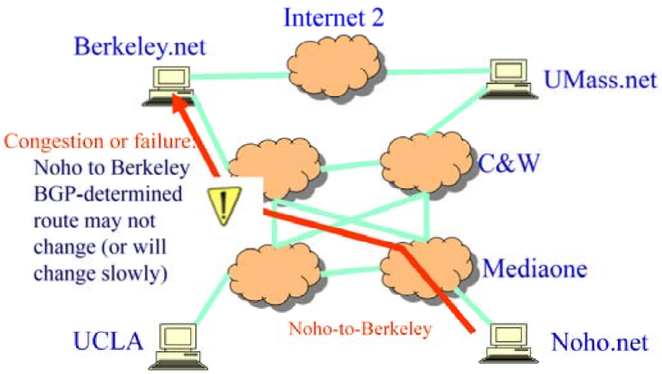
Internet Routing

- BGP defines routes between stub networks

31

Internet Routing

- BGP defines routes between stub networks



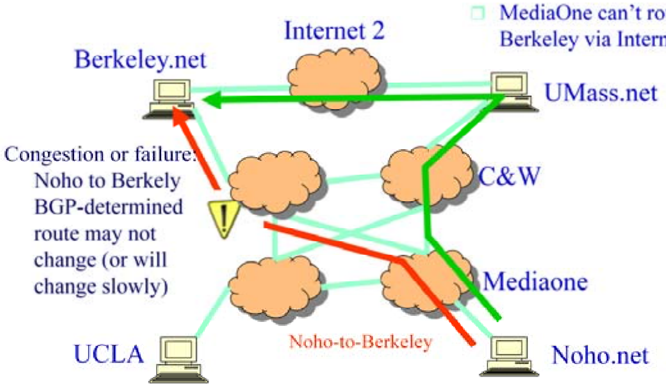
Congestion or failure:
Noho to Berkeley
BGP-determined
route may not
change (or will
change slowly)

32

Internet Routing

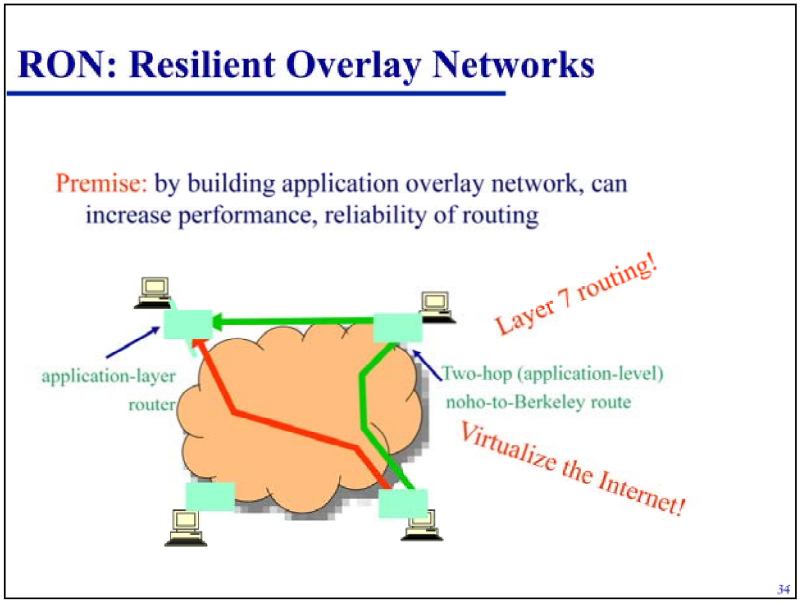
Noho to UMass to Berkeley

- route not visible or available via BGP!
- MediaOne can't route to Berkeley via Internet2



Congestion or failure:
Noho to Berkely
BGP-determined
route may not
change (or will
change slowly)

33

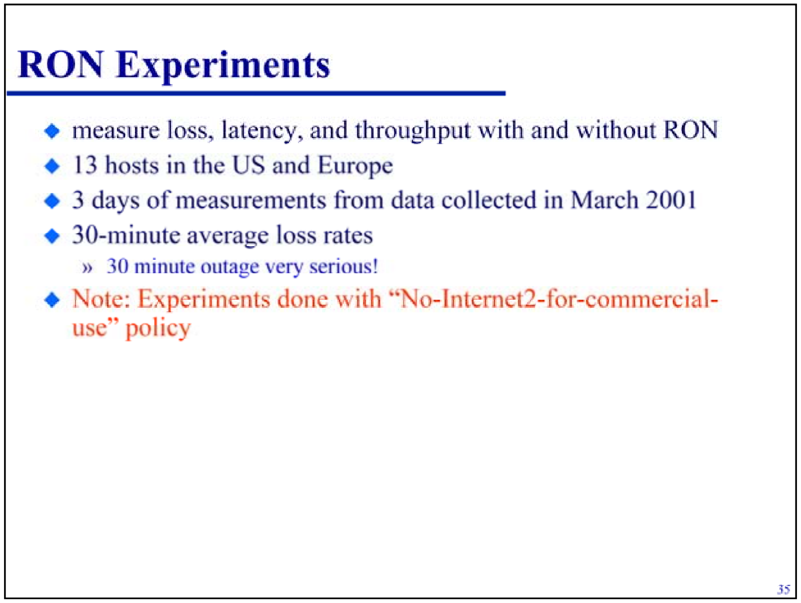


RON: Resilient Overlay Networks

Premise: by building application overlay network, can increase performance, reliability of routing

Layer 7 routing!
application-layer router
Two-hop (application-level) noho-to-Berkeley route
Virtualize the Internet!

34

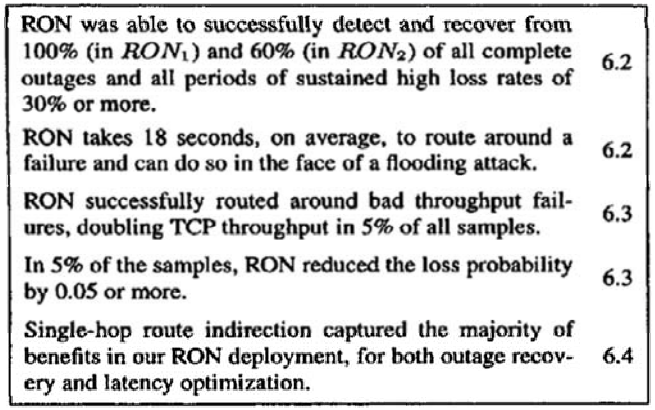


RON Experiments

- measure loss, latency, and throughput with and without RON
- 13 hosts in the US and Europe
- 3 days of measurements from data collected in March 2001
- 30-minute average loss rates
  - » 30 minute outage very serious!
- Note: Experiments done with "No-Internet2-for-commercial-use" policy

35

RON Study Results (2001)

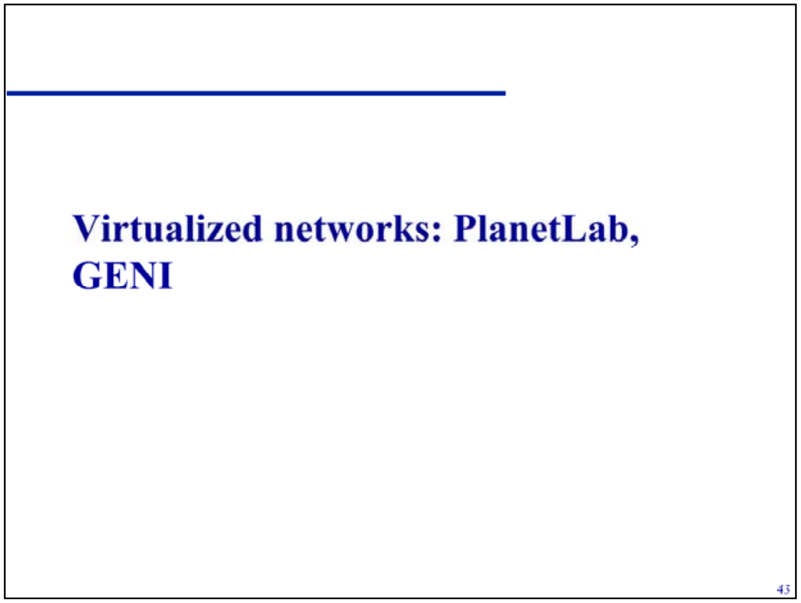


RON was able to successfully detect and recover from 100% (in RON_1) and 60% (in RON_2) of all complete outages and all periods of sustained high loss rates of 30% or more.	6.2
RON takes 18 seconds, on average, to route around a failure and can do so in the face of a flooding attack.	6.2
RON successfully routed around bad throughput failures, doubling TCP throughput in 5% of all samples.	6.3
In 5% of the samples, RON reduced the loss probability by 0.05 or more.	6.3
Single-hop route indirection captured the majority of benefits in our RON deployment, for both outage recovery and latency optimization.	6.4

Table 1: Major results from measurements of the RON testbed.

RON Research Issues

- how to design overlay networks?
  - measurement and self-configuration
  - fast fail-over
  - sophisticated metrics
  - application-sensitive (e.g., delay versus throughput) path selection
- effect of RON on underlying network
  - if everyone does RON, are we better off?
  - interacting levels of control (network- and application-layer routing)



Virtualized networks: PlanetLab, GENI