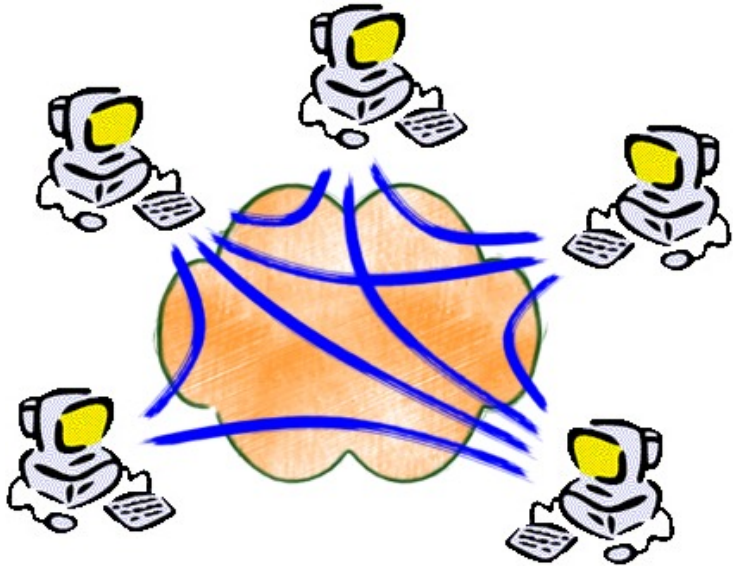
Overlays can do more

… if not everything

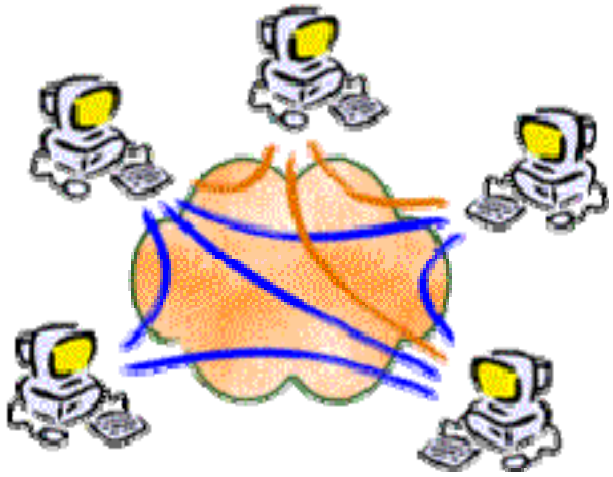
Jorg Liebeherr
University of Toronto
jorg@comm.utoronto.ca



Acknowledgements

- **Contributors**:
  - *Bhupinder Sethi, Tyler Beam, Burton Filstrup, Mike Nahas, Dongwen Wang, Konrad Lorincz, Jean Ablutz, Haiyong Wang, Weisheng Si, Huafeng Lu, Josh Zaritsky, Jianping Wang, Guimin Zhang, Guangyu Dong, Greg Mattes, Zhi An, Shaoyi Chen, Wittawat Tantisiriroj,* Majid Valipour, Jon Lei

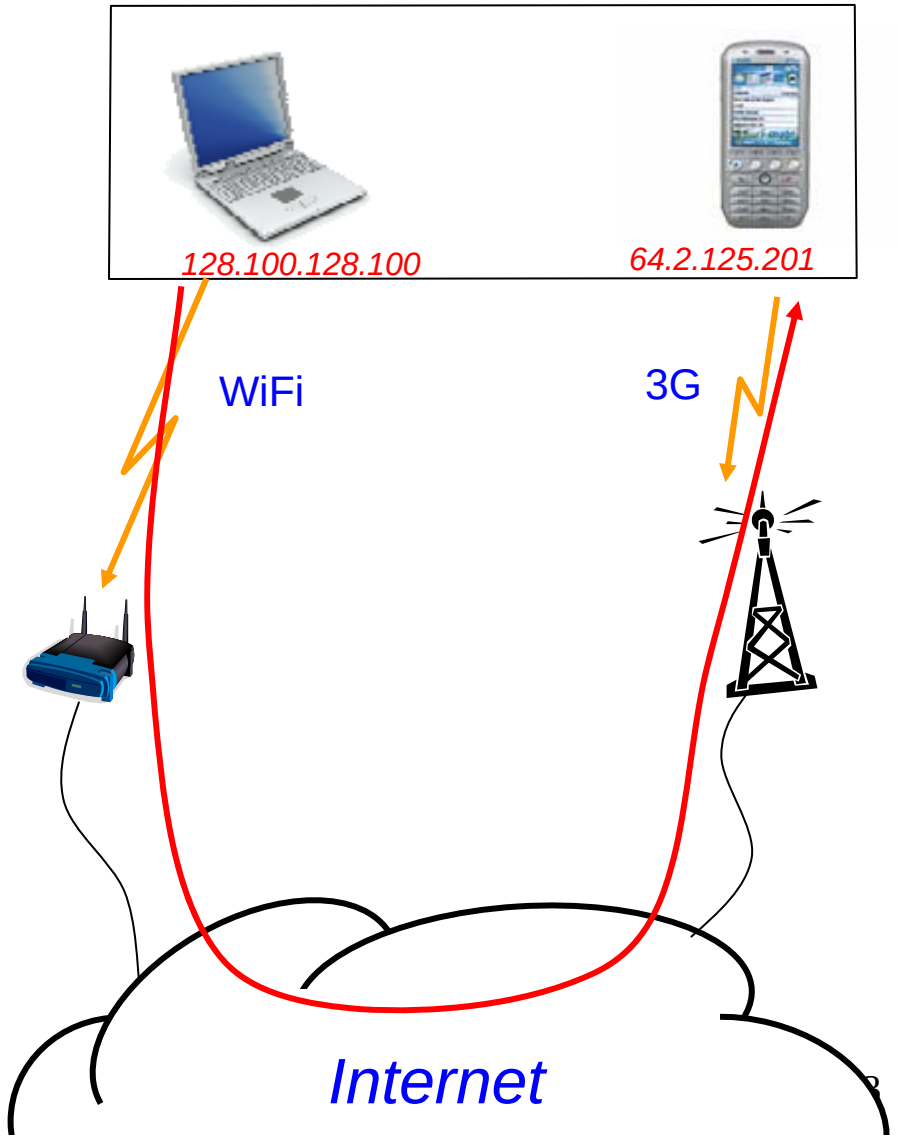
- **Funding:** NSF, DOD, NSERC



Internet-centric Networking

Things that are not needed:

- Access to infrastructure
- Allocation of global addresses
- Resulting vulnerabilities



An Analogy

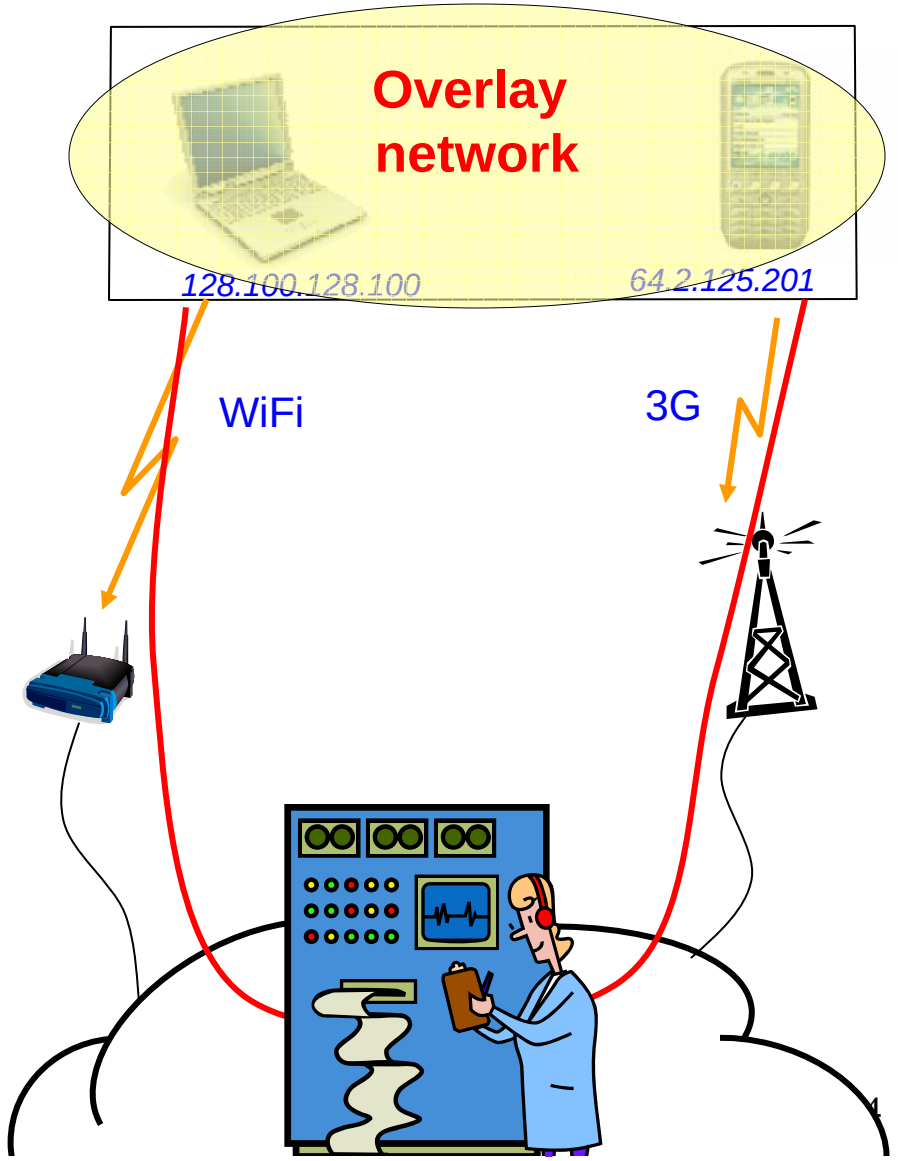
Internet today
appears like
Mainframe computing of 1970s

All *networking*
~~Many computing~~ needs can be

Internet
met without ~~mainframe computers~~

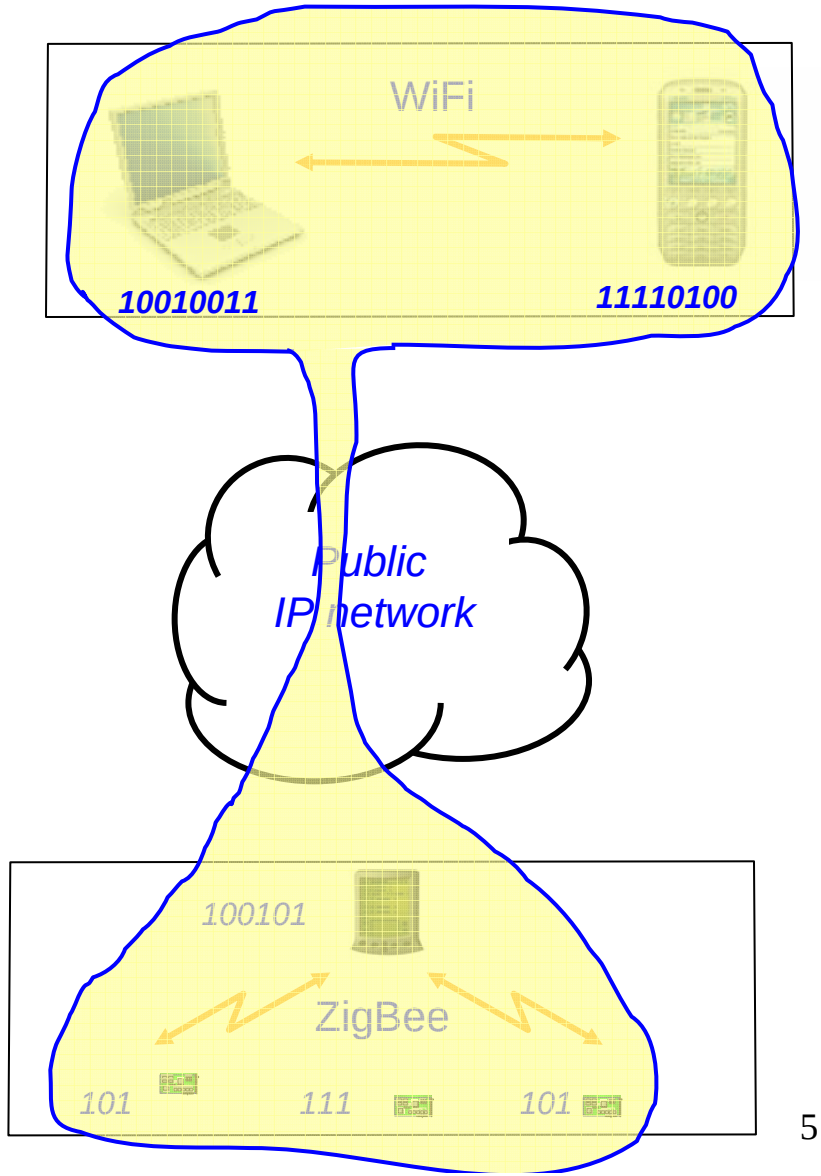
Overlays
~~Smaller computers~~ can replace

Internet
~~mainframes~~



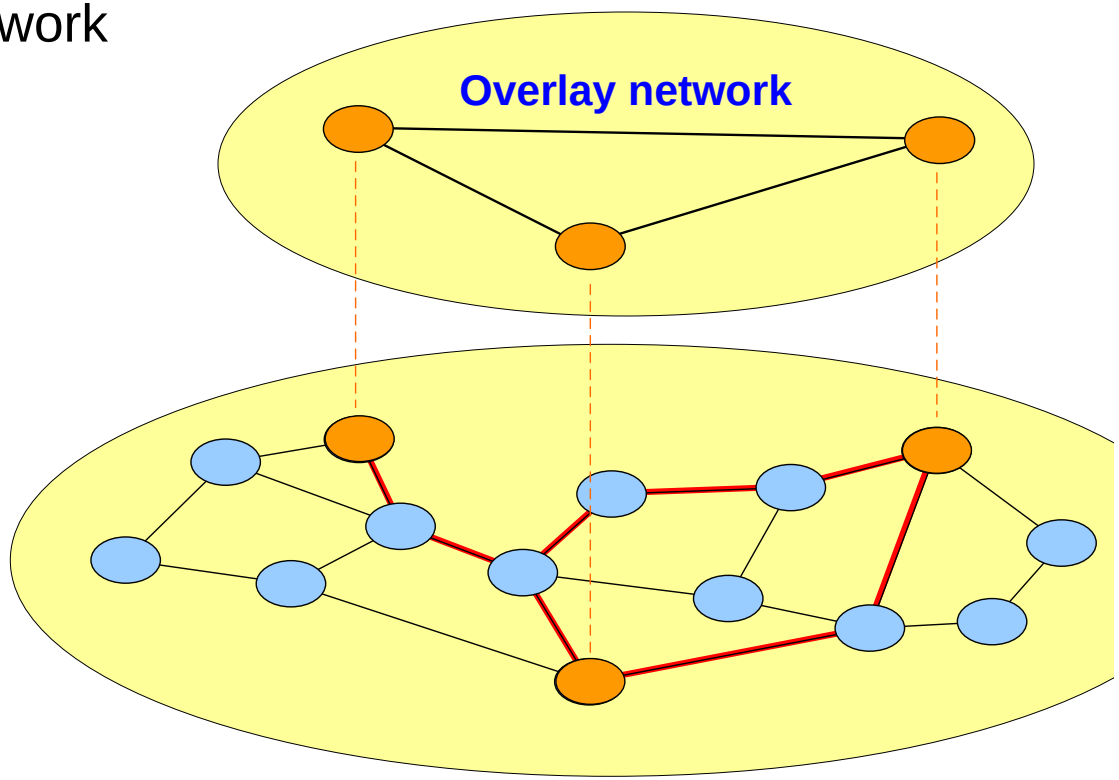
Overlay-based networking

- Applications/devices self-organize as a network
- Application networks define their own address space
- Access infrastructure as needed



What is an overlay anyway?

- An overlay network is a virtual network of nodes and logical links built on top of an existing network
- A virtual link in the overlay corresponds to a path in the underlay



Overlays Everywhere

- Overlays are a main architectural element for building complex networks

- **4 reasons to build overlay networks**
  1. **Virtualize network resources**
  2. **Deploy new network services**
  3. **Create sub-groups for information exchange**
  4. **Interconnect substrate networks**

Application Layer Overlays

late 1990s

- **Application-layer overlay**
  on top of the Internet

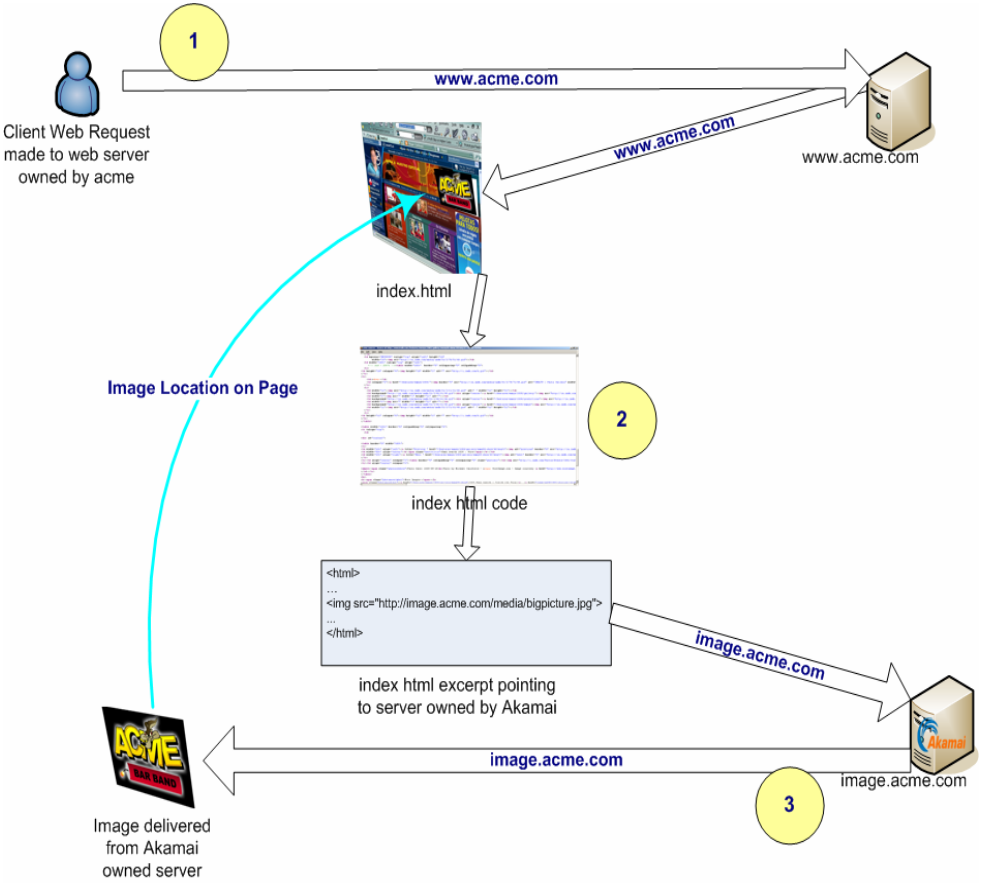
- Made popular by
  - Application-layer multicast
  - Distributed lookup (DHT)

- Also:
  - Security
  - Intrusion detection
  - Replication
  - QoS

Content distribution networks

1998

- AKAMAI overlay for content distribution:
  - 15,000 servers, in 69 countries, within 1,000 networks.

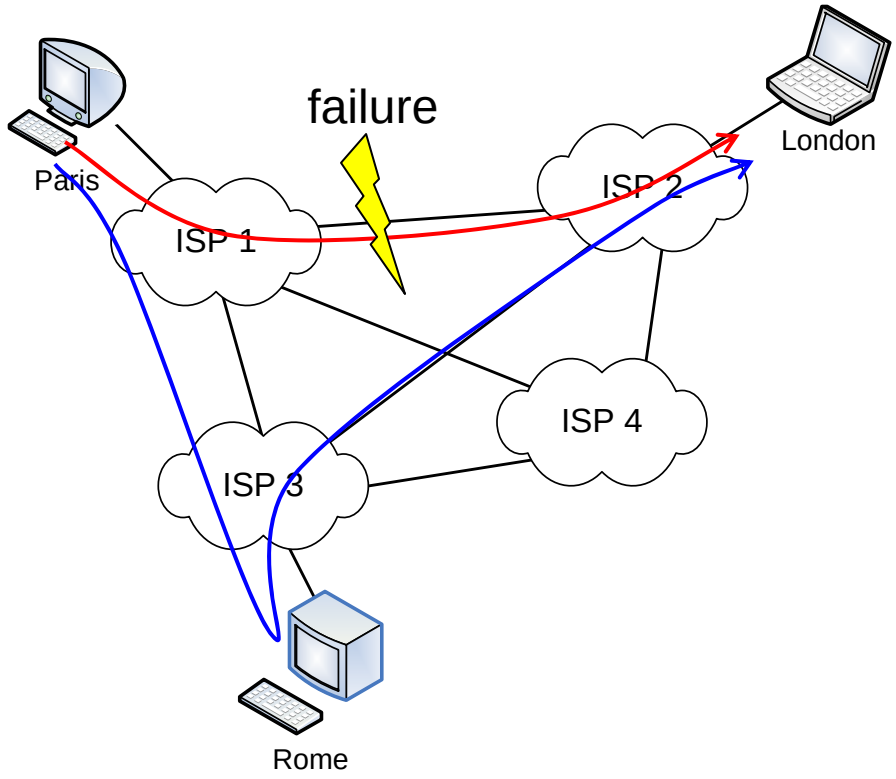


Source: Wikipedia

RON (resilient overlay network)

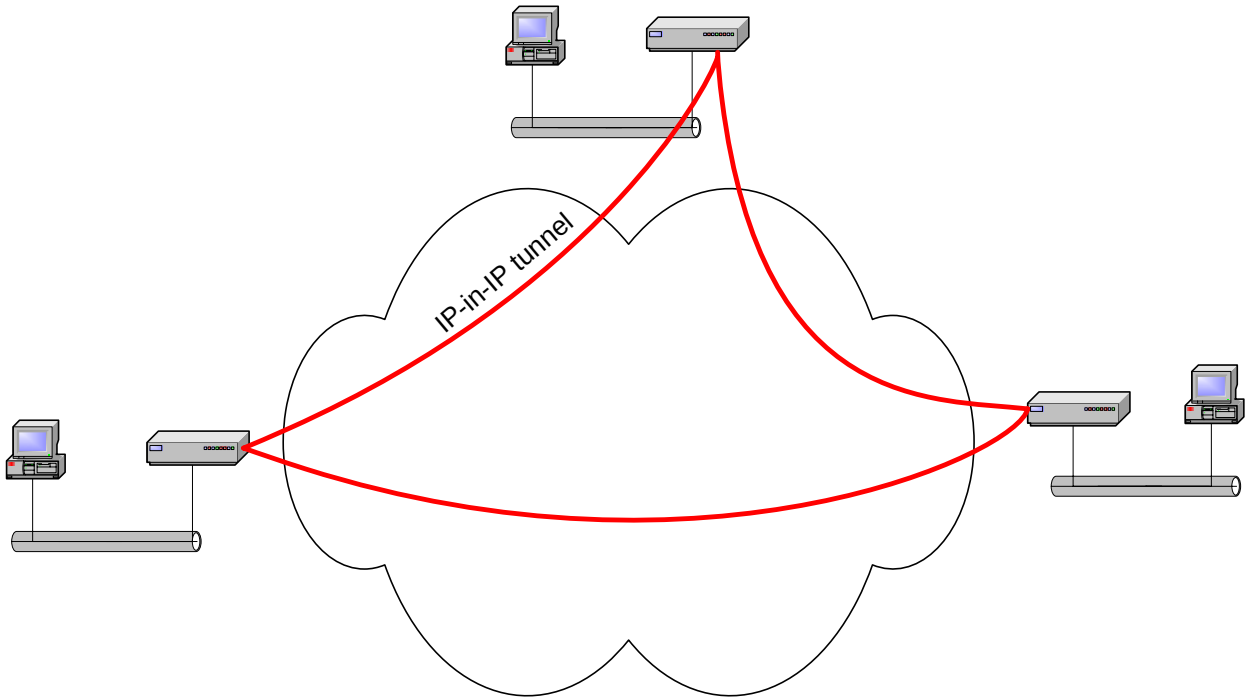
2001

- Detect outages
- Upon outage, relay data to a peer node



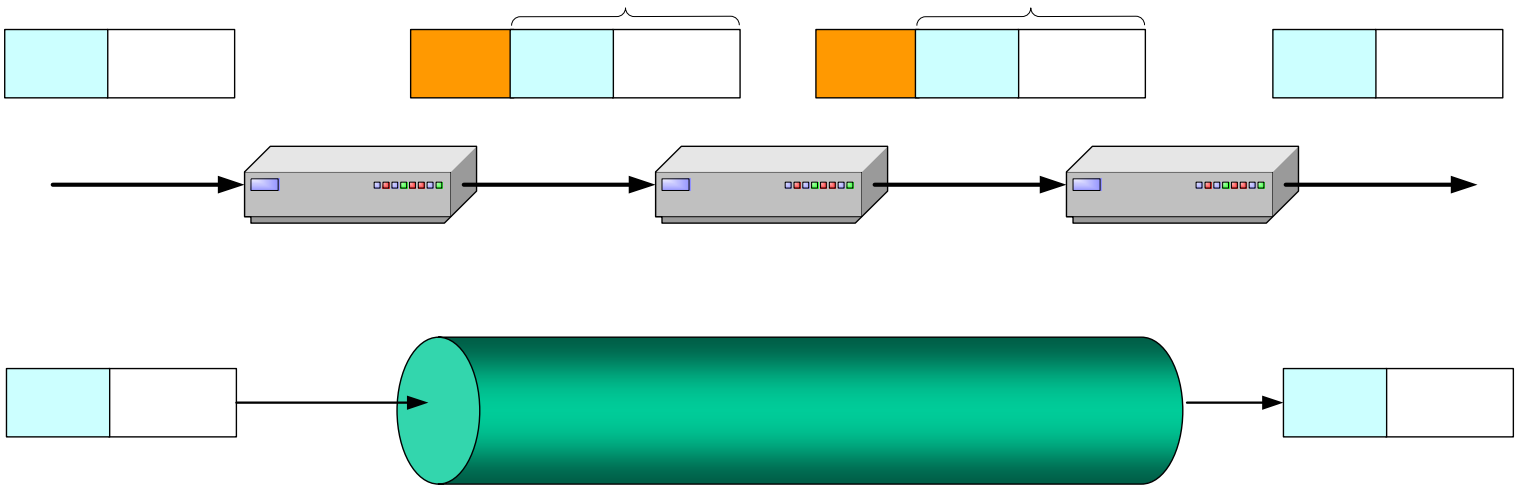
MBone (1992), VPN Overlays (1994), 6Bone (2000)

- Deploy new services in IPv4 Internet
  - Multicast (MBone)
  - IPv6 (6Bone)
  - Security (VPN)



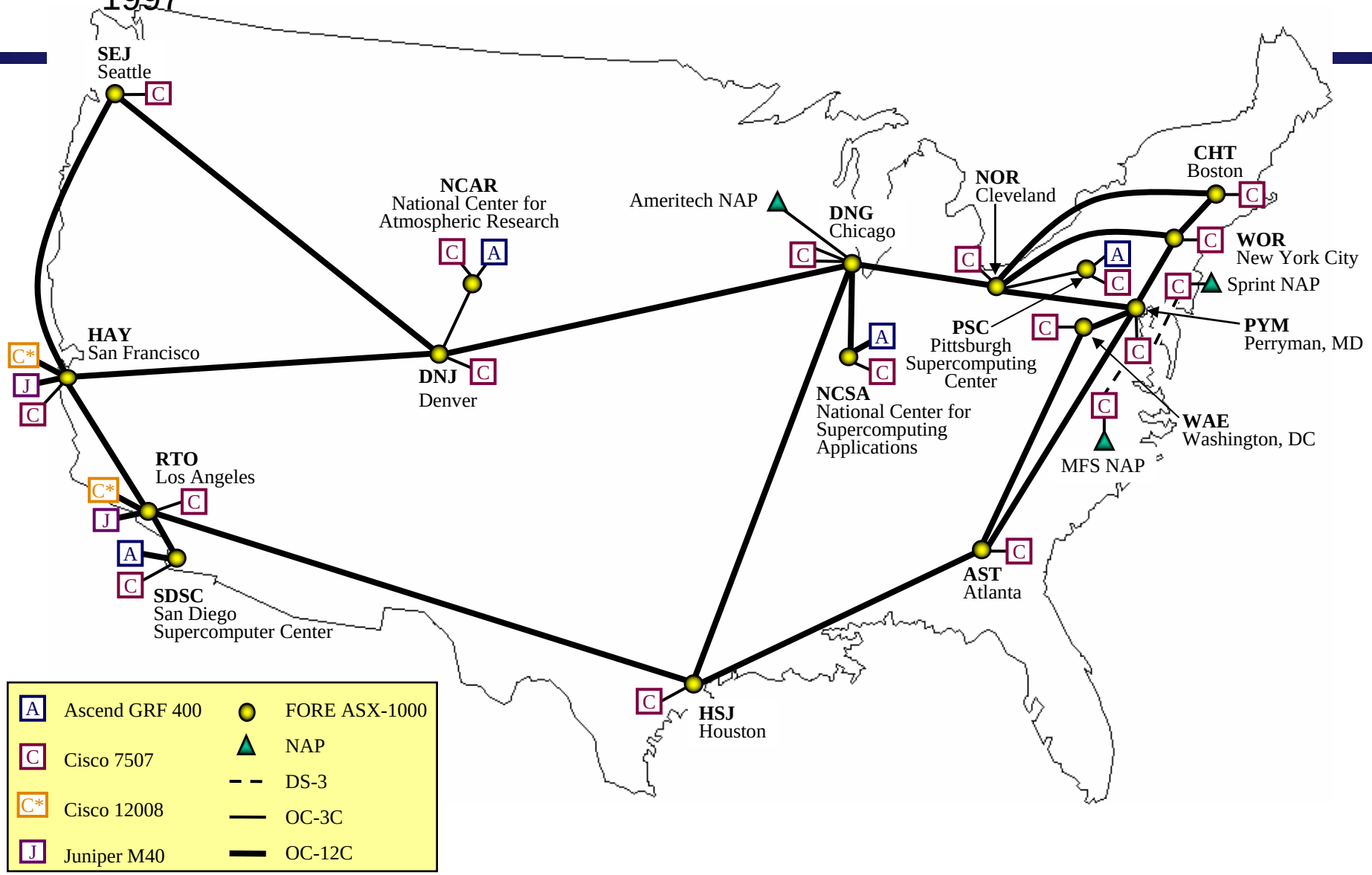
MBone (1992), VPN Overlays (1994), 6Bone (2000)

- Network-layer overlays
  - Principle: IP-in-IP encapsulation



vBNS Backbone Network Map

1997

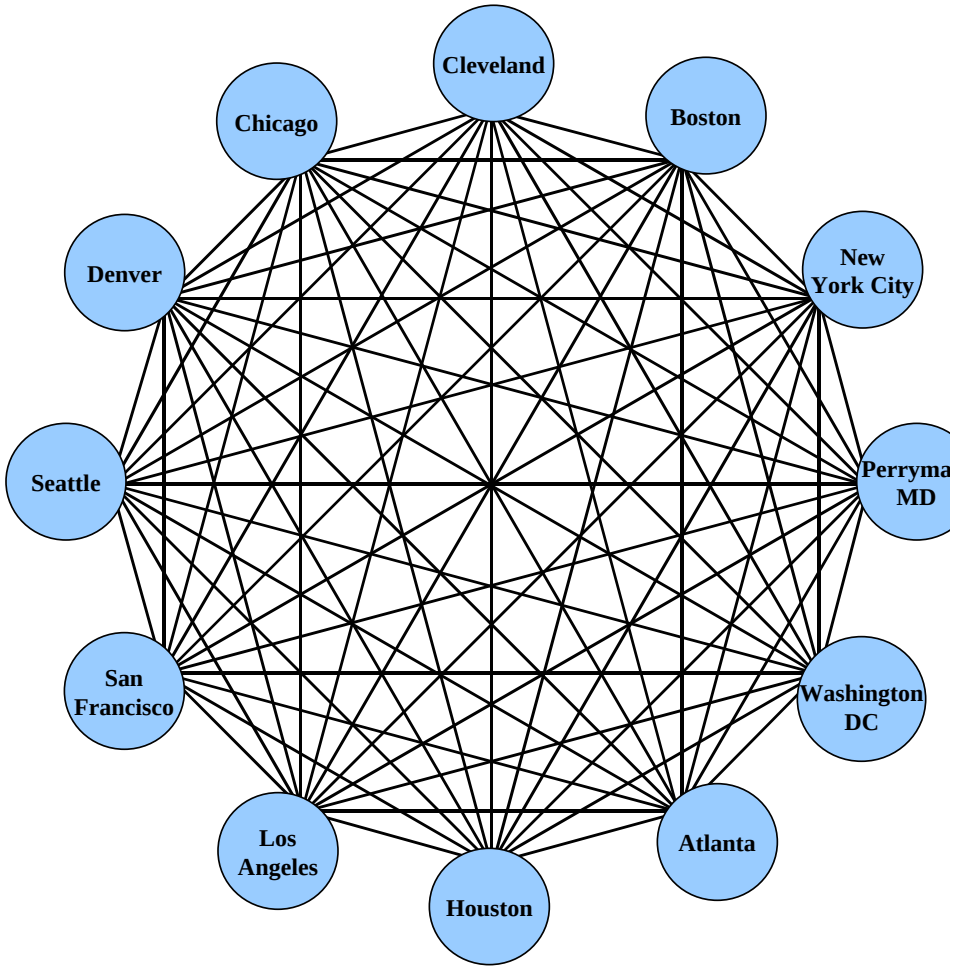


Source: Internet archive, vbns.net, 2000

vBNS: IP Topology

1997

- Each vBNS node has an ATM switch and at least one IP router
- ATM switches build a PVP mesh between all nodes
- IP routers see full mesh topology

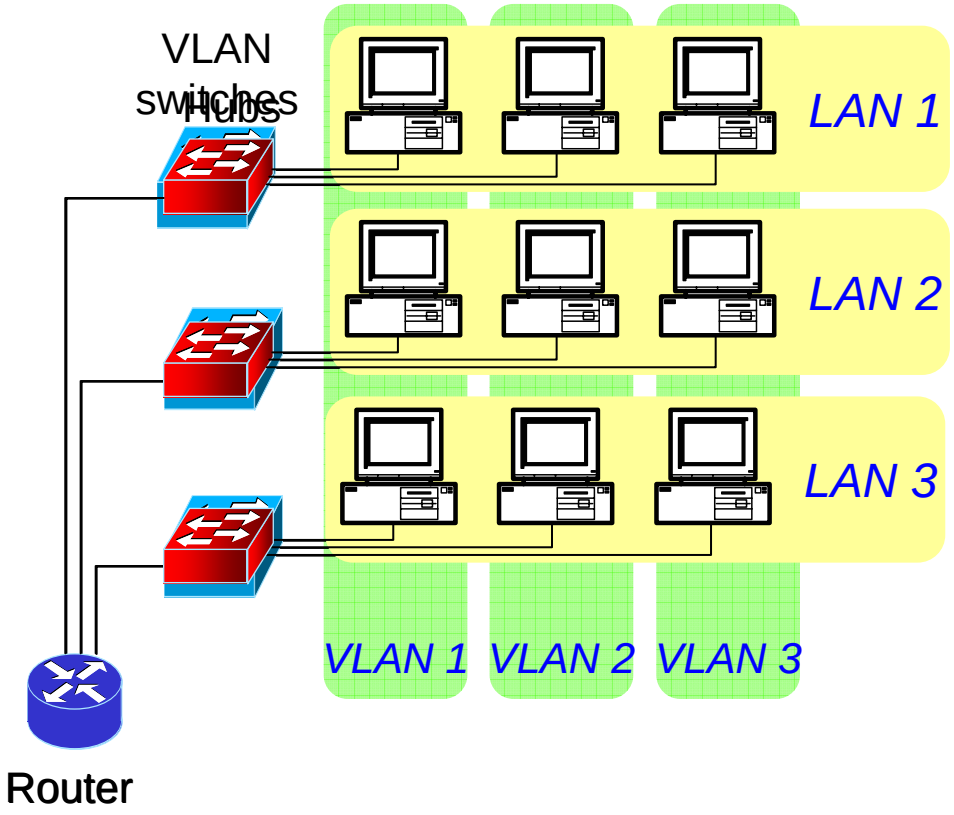


Source: Internet archive, vbns.net, 2000

VLANs

1998

- VLAN overlays separate broadcast domain from location of hosts

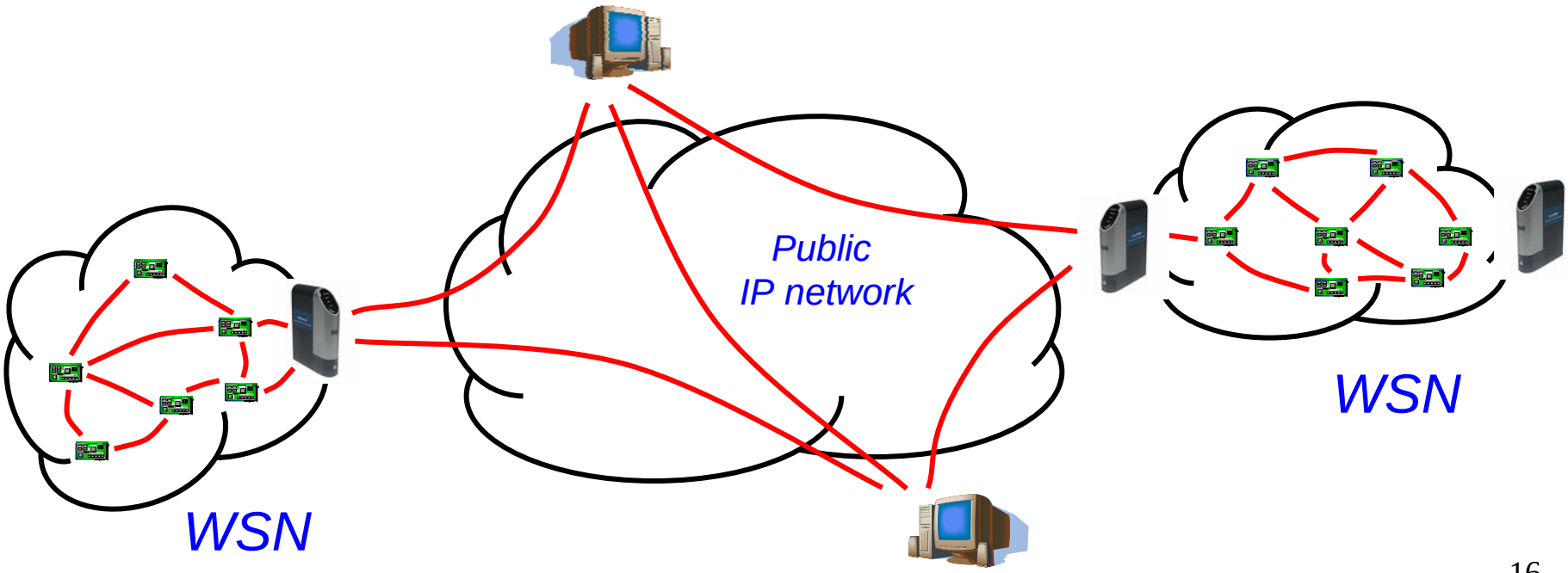


Source: Cisco

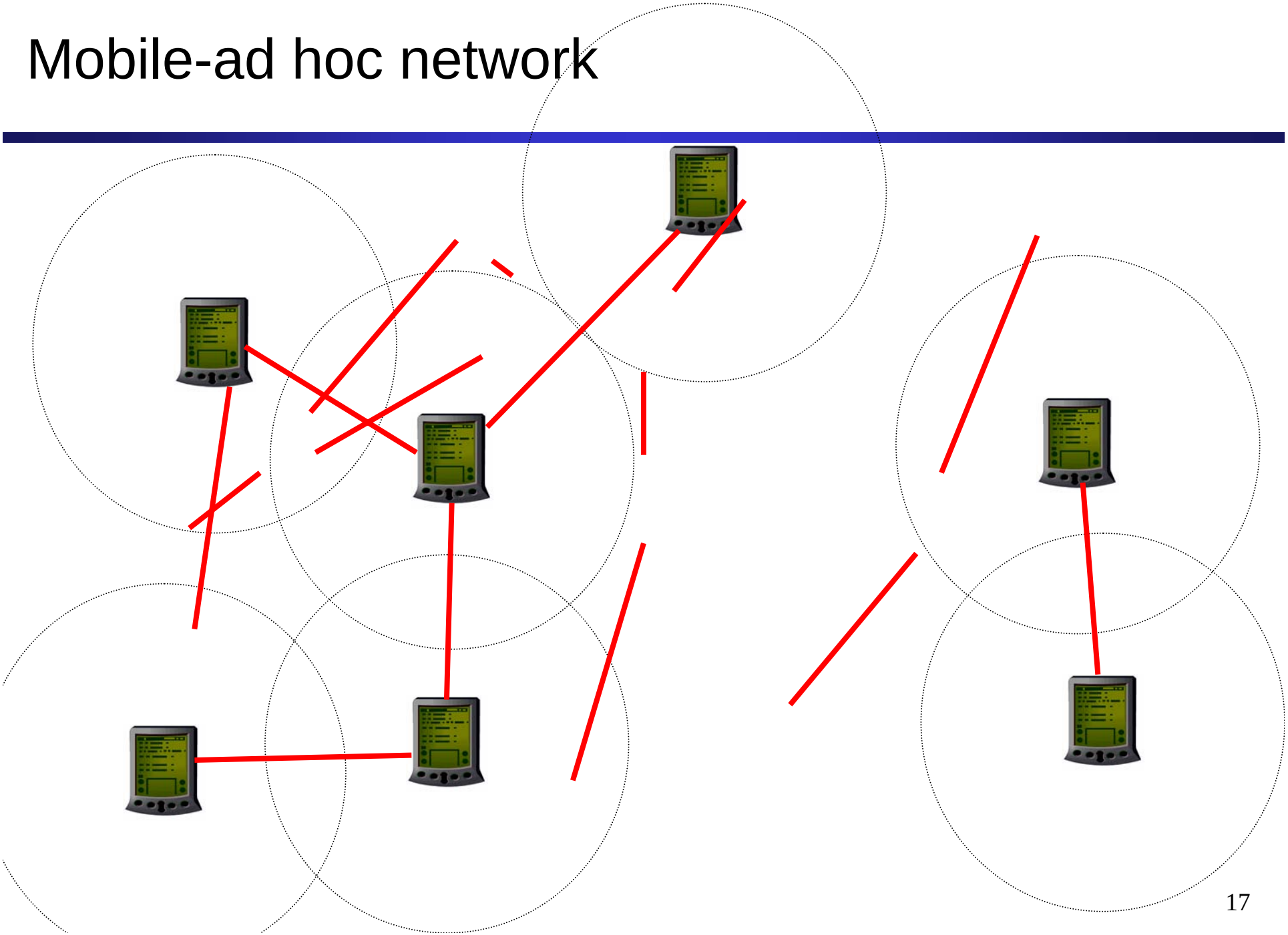
Heterogeneous Sensor Network

since 2005

- Interconnect remote wireless sensor networks
- Extend overlay to sensor nets

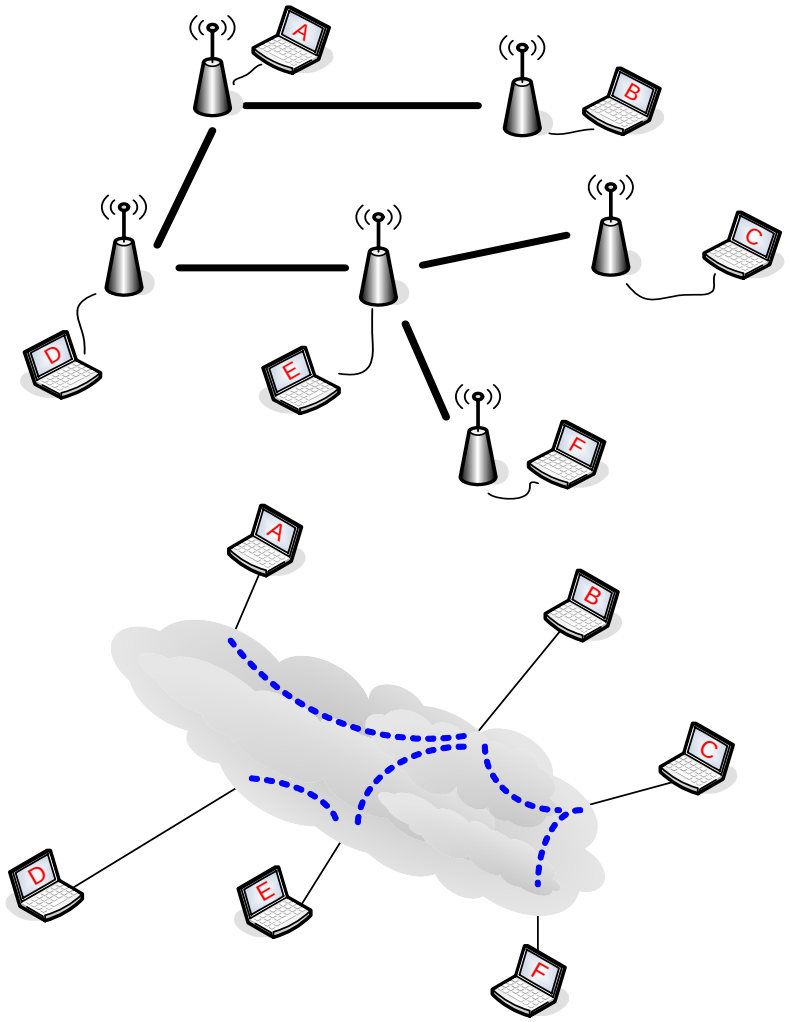


Mobile-ad hoc network



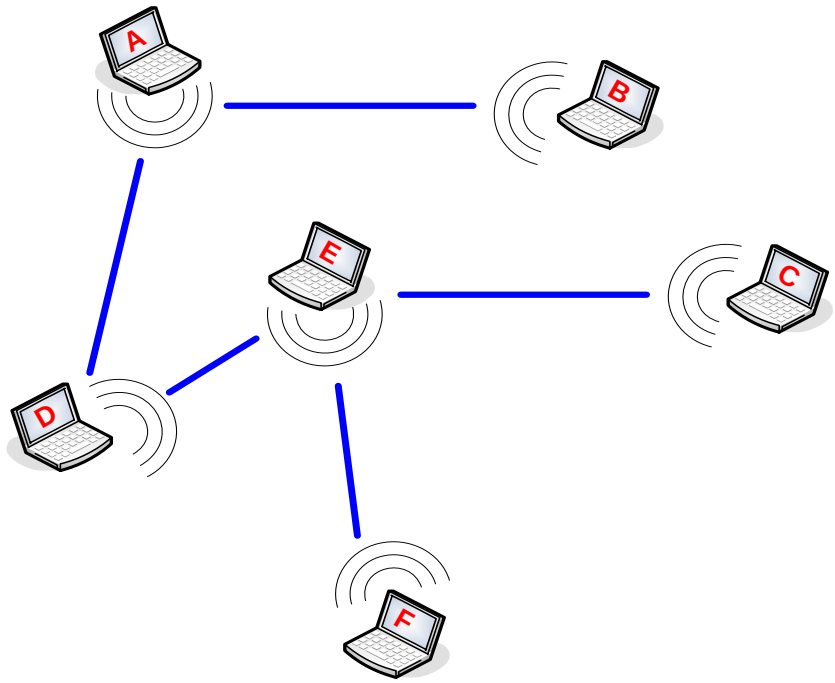
Layered Overlay MANET

- **Lower Layer:** Mesh network is built by mesh radios
- **Higher Layer:** Overlay is created on top of the mesh network



Integrated Overlay MANET

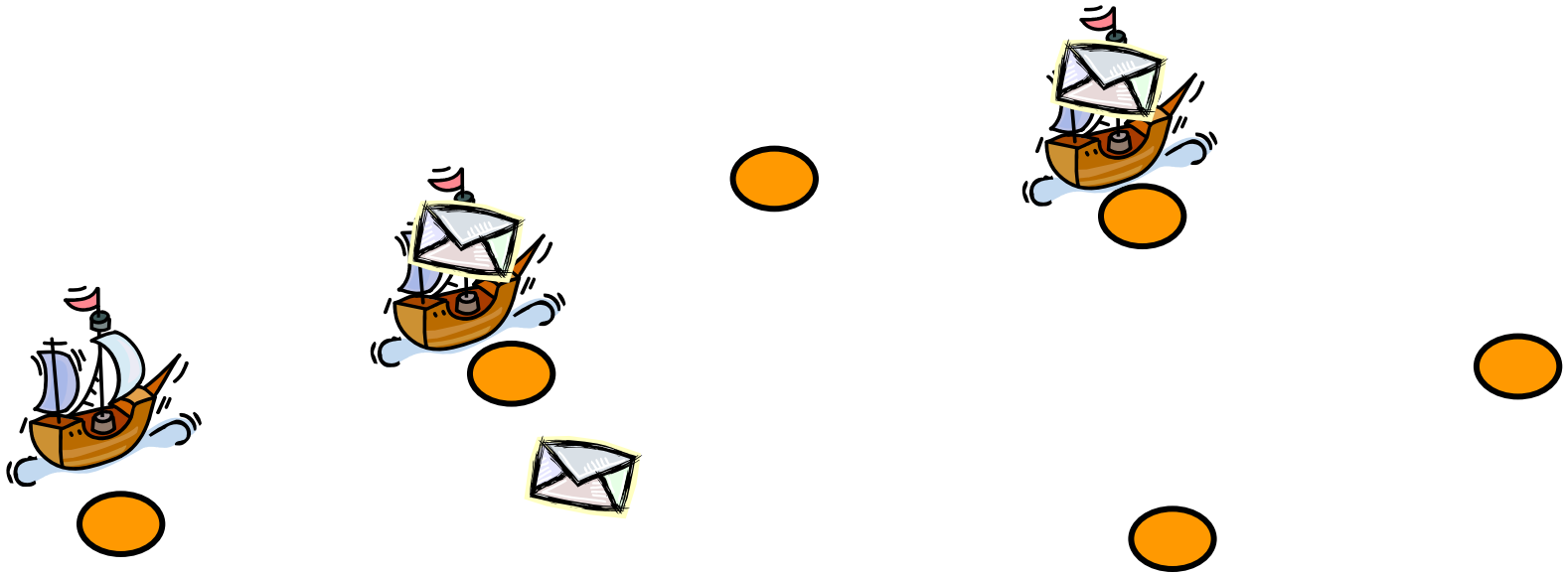
Overlay is responsible for creating and maintaining a mesh topology



Disruptive Tolerant Networks

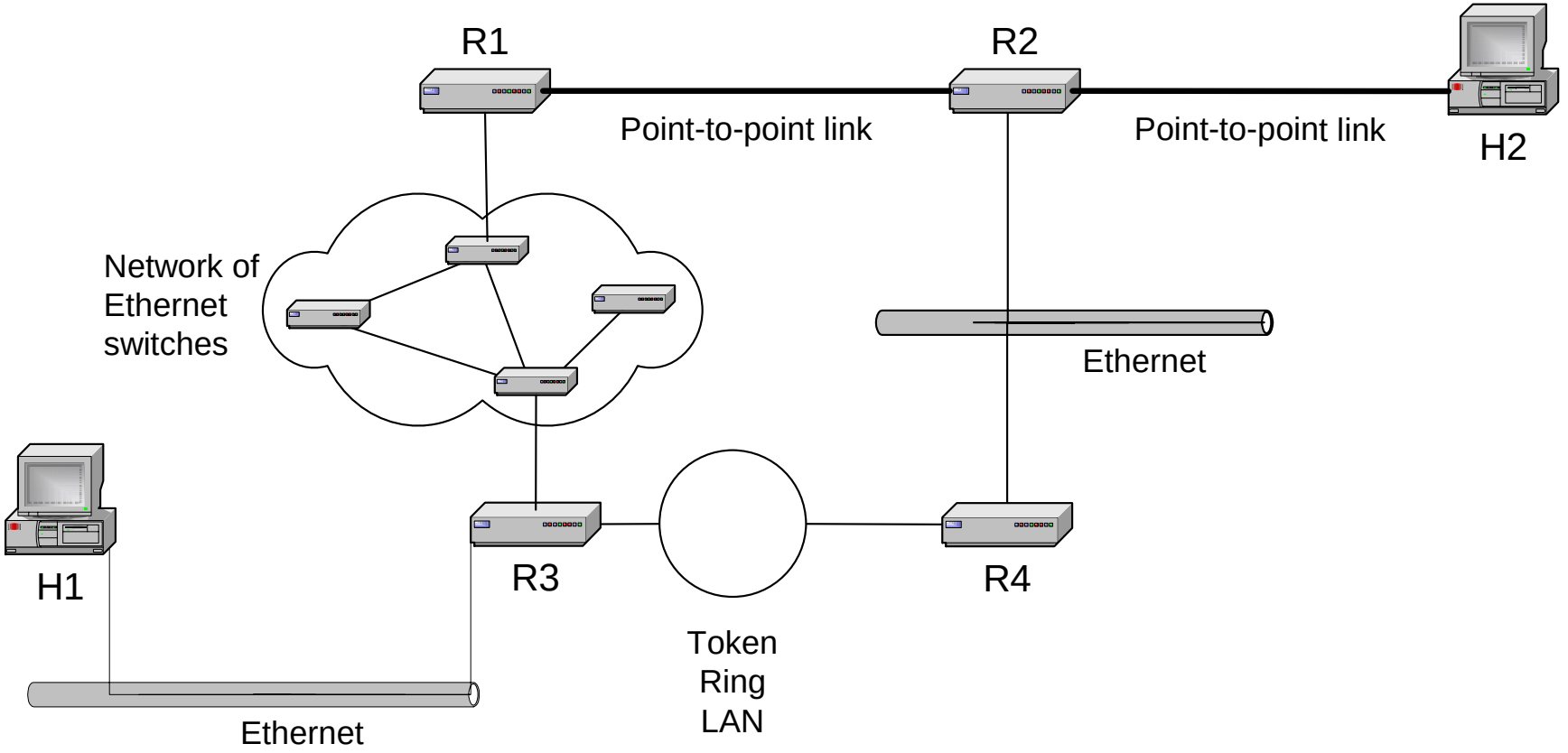
since 2000

- “Message Ferrying”
- Multihop data path, but whole path may not exist at one point in time.



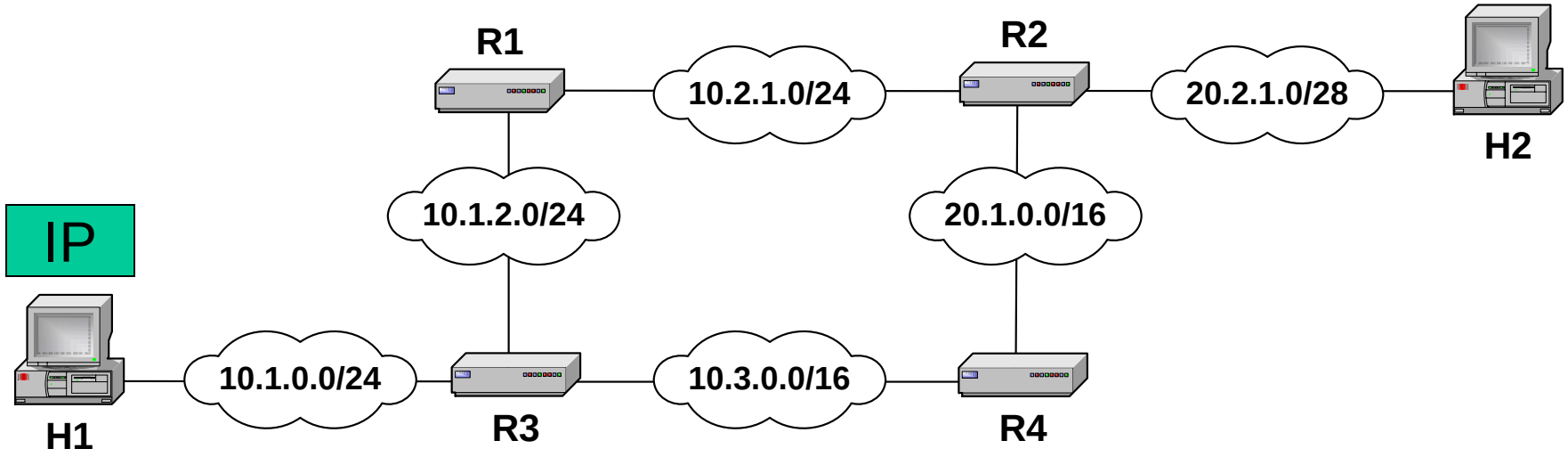
… and of course … The Internet

- View at the data link layer layer:



… and of course … The Internet

- View at the IP layer:



	Virtualization	New Service	Interconnection	Exchange Groups	Forwarding	Self-organization
IP over ATM	✓	✓	✓	×	✓	×
CDN (Akamai)	✓	✓	×	×	×	×
MBONE, VPN, 6Bone	✓	✓	×	?	✓	×
RON	✓	✓	×	×	✓	×
ALM	✓	✓	×	✓	✓	✓
VLAN	✓	✓	×	✓	✓	✓
WSNs	✓	✓	✓	?	✓	✓
DTN	?	✓	?	?	✓	✓
MANET	?	✓	?	×	✓	✓
IP Network	✓	✓	✓	×	✓	×

Overlay Network Research

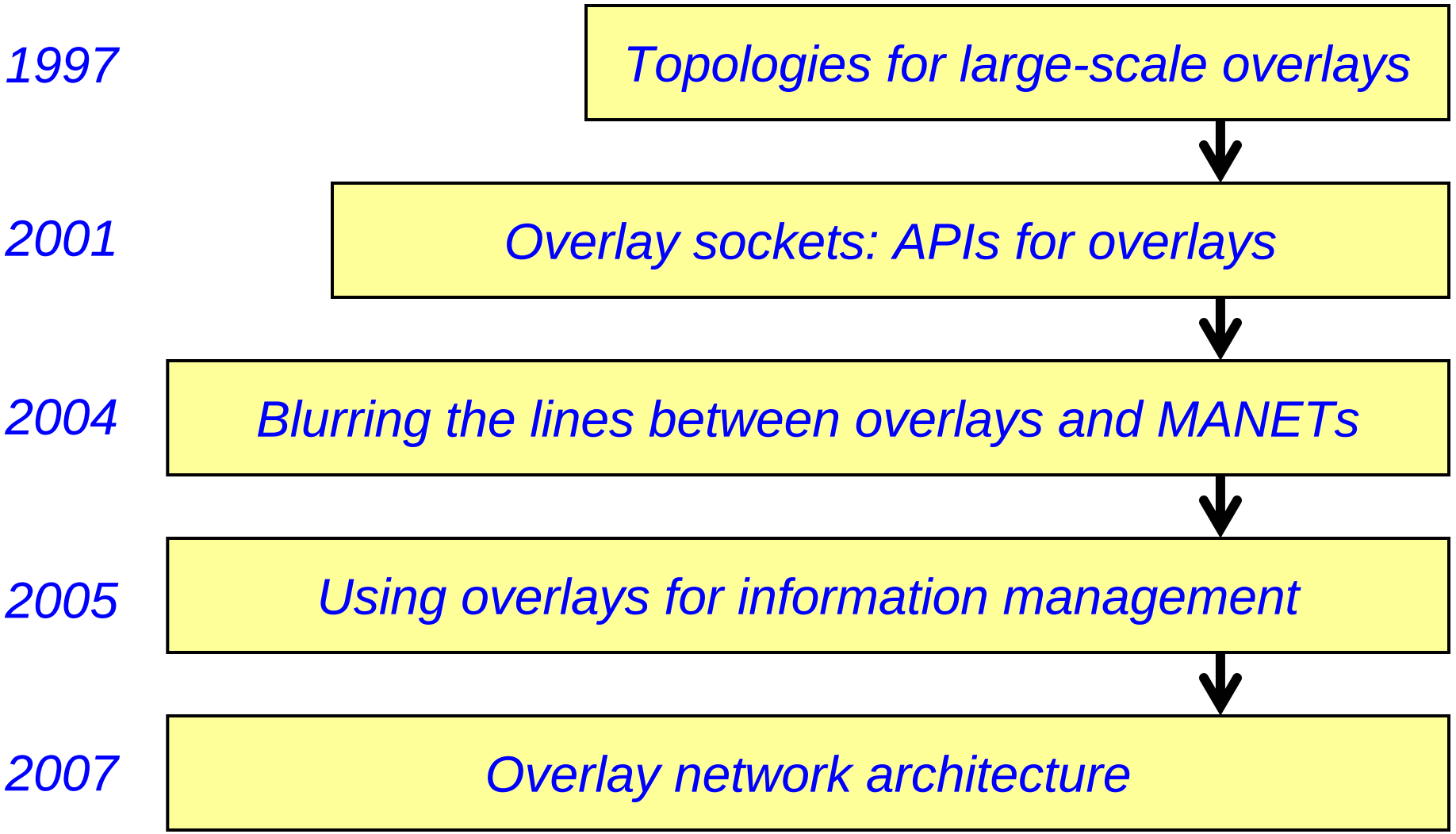
Past 10 years:

- Mostly application-layer
- Mostly single-substrate networks
- Mostly structured topology
- Distinct from mobile networks (MANET, WSN, DTN)

Future:

- All layers
- Multisubstrate networks
- Structured topology if possible
- Mobile overlays are integral

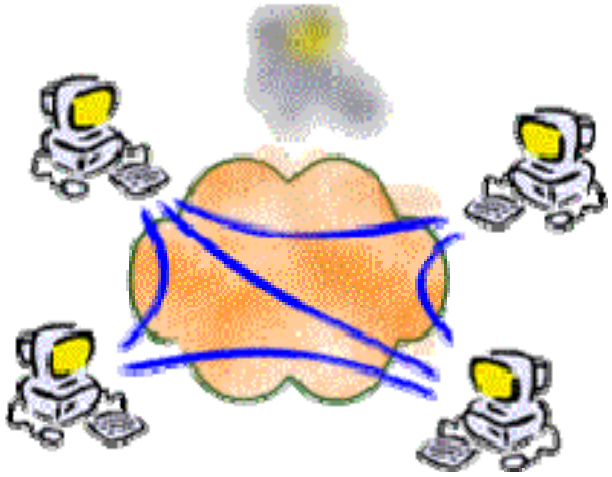
Rest of Talk:
How I came to think Overlays can do more



My point of Departure

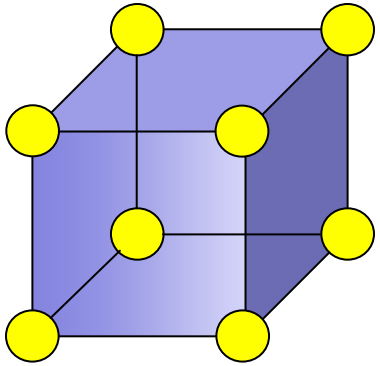
- **HyperCast** (www.hypercast.org) is a set of protocols for large-scale self-organizing overlay networks

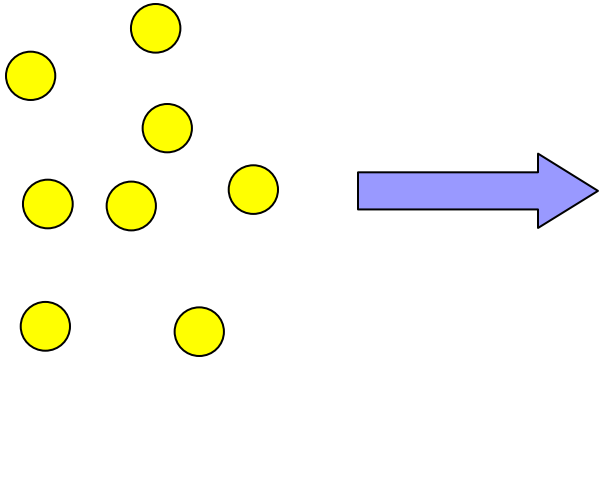
- ~100,000 lines of Java
- Available on Sourceforge
- Socket inspired APIs
- Security, Monitor and control, …
- Some applications



HyperCast Overlay Networks

- Applications self-organize to form a given overlay topology
- Data is forwarded along the edges of the overlay topology

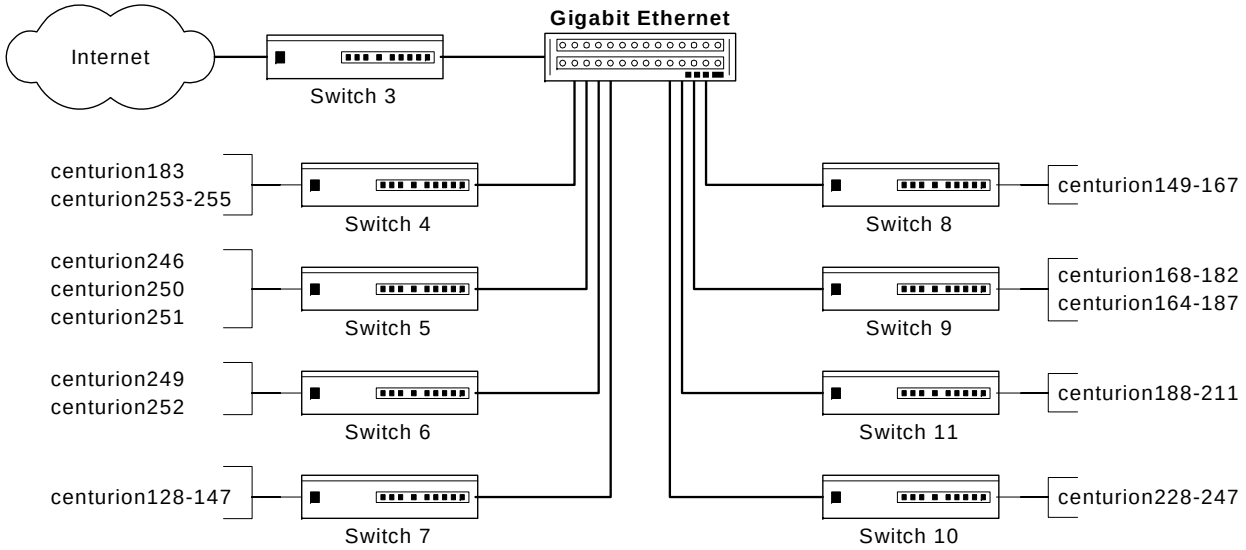




Local Area Experiments

2002

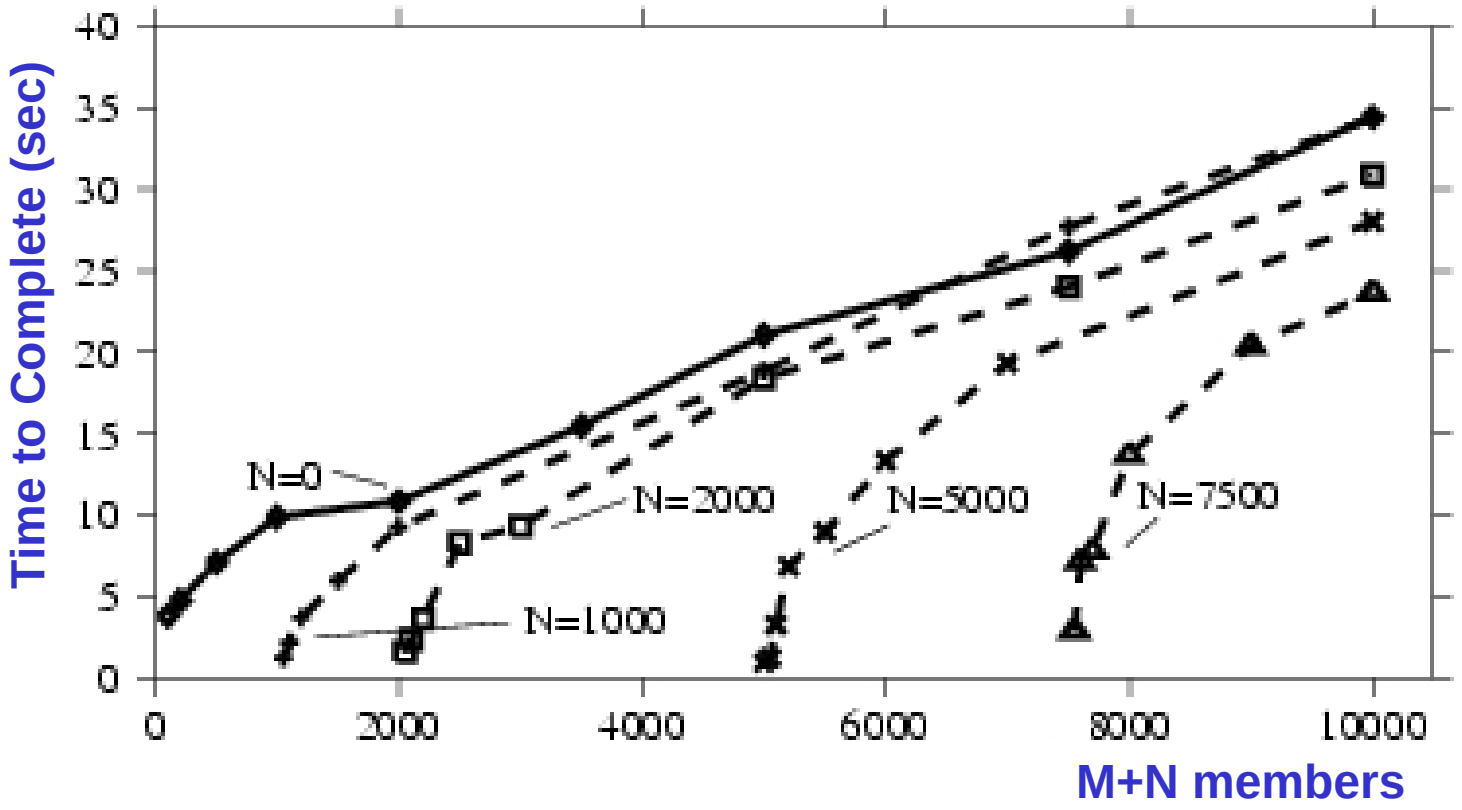
- **Experimental Platform:**
  Centurion cluster at UVA (cluster of 300 Linux PCs)
  - **up to 100 PCs**
  - **Up to 100 members per PC**
    - **→ up to 10,000 overlay nodes**



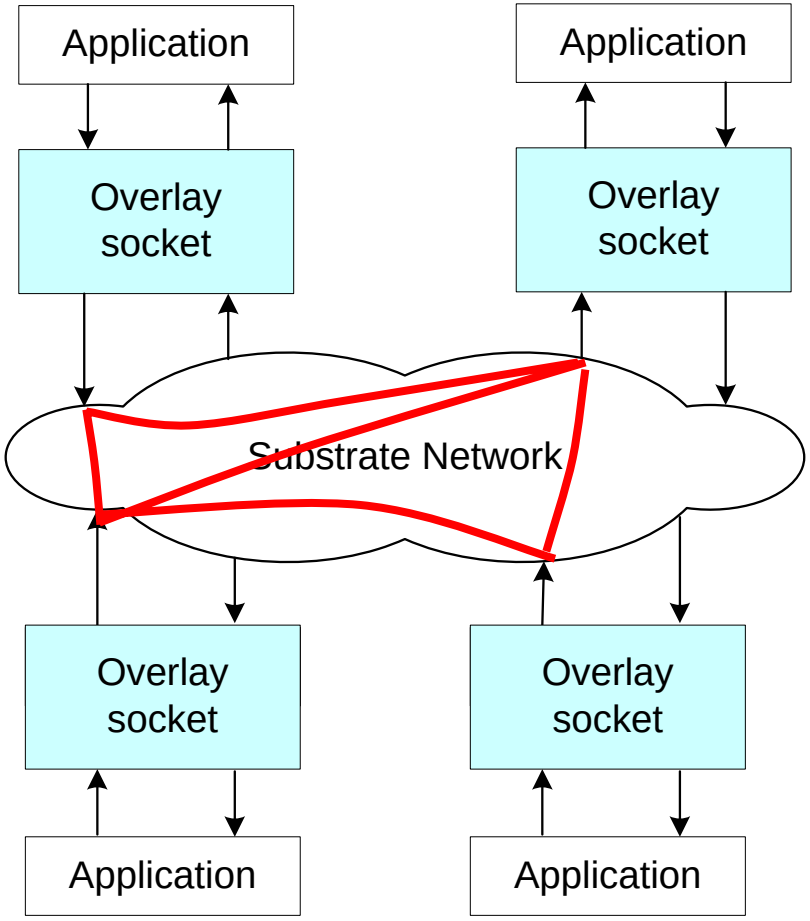
Experiment: Scalability

Topology: Delaunay Triangulation

Measurement: How long does it take to add M members to an overlay network of N members ?

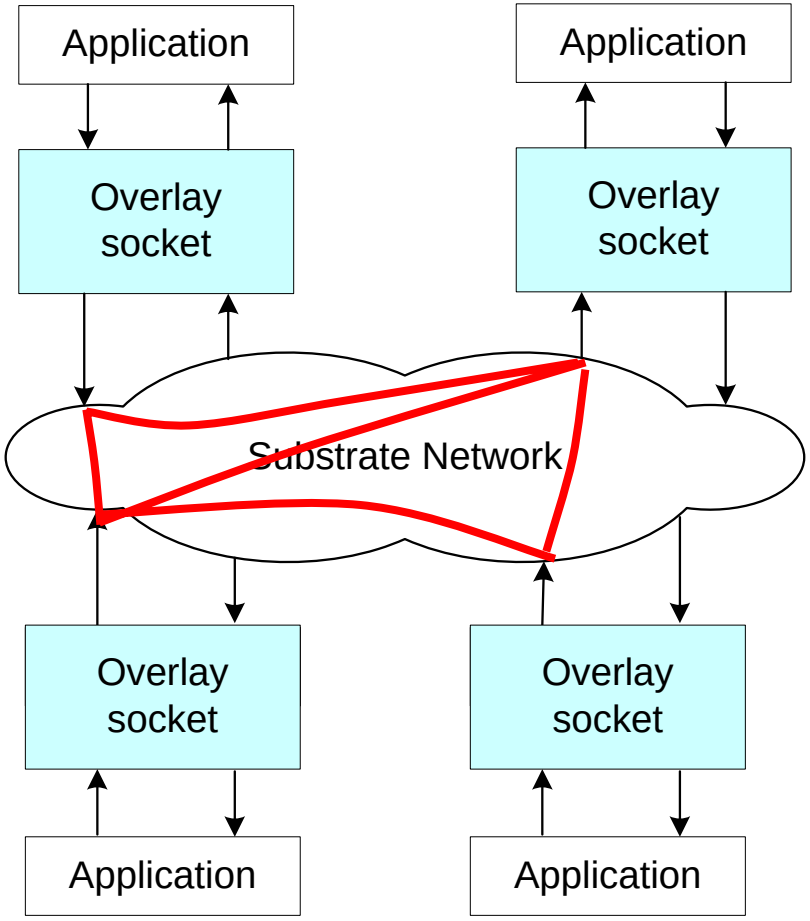


Overlay Sockets

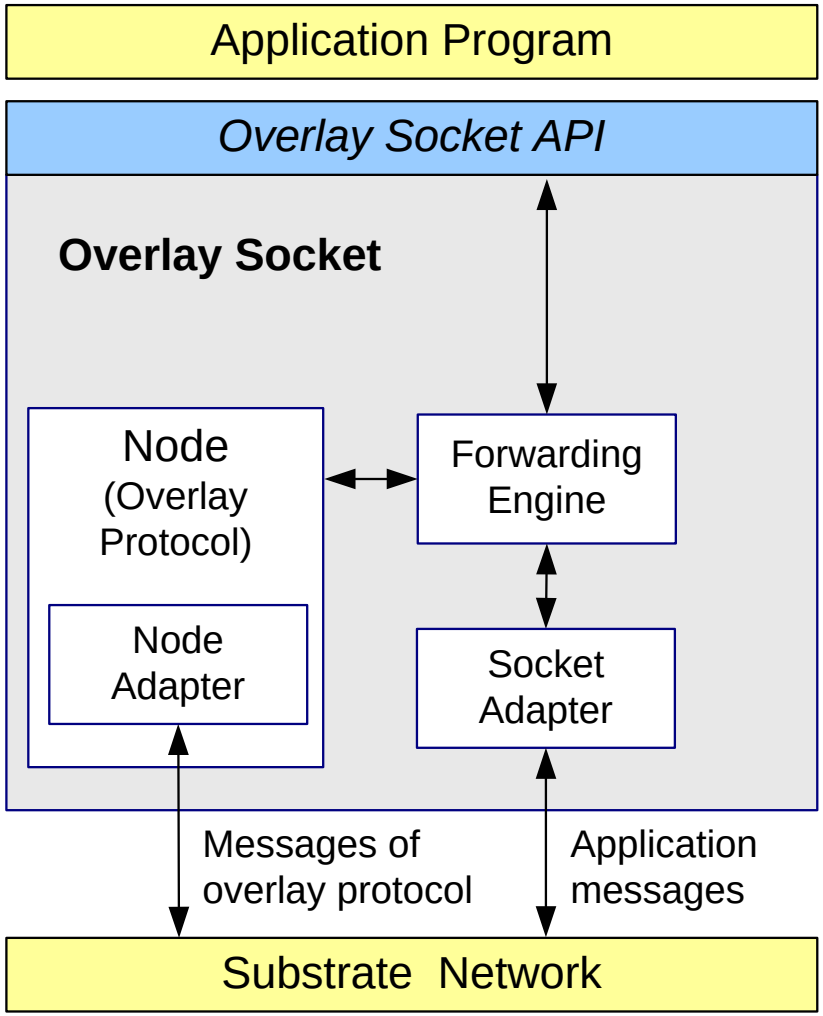


(b) Overlay Network
(Collection of overlay sockets)

Overlay Sockets



(b) Overlay Network
(Collection of overlay sockets)

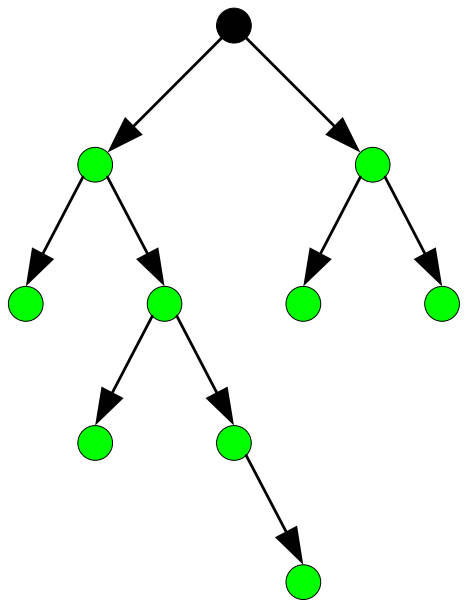


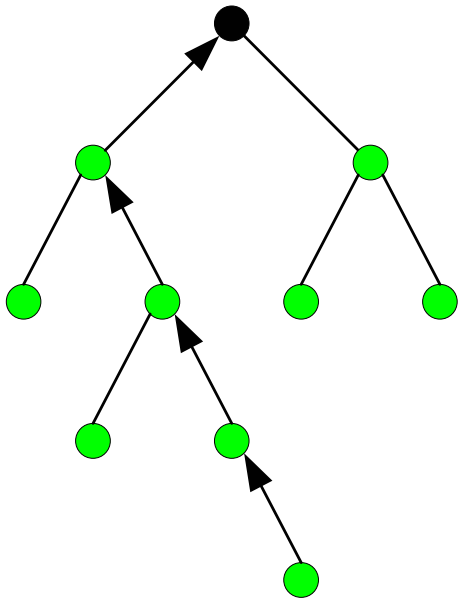
(a) Overlay socket

Forwarding

- Data is forwarded along embedded trees

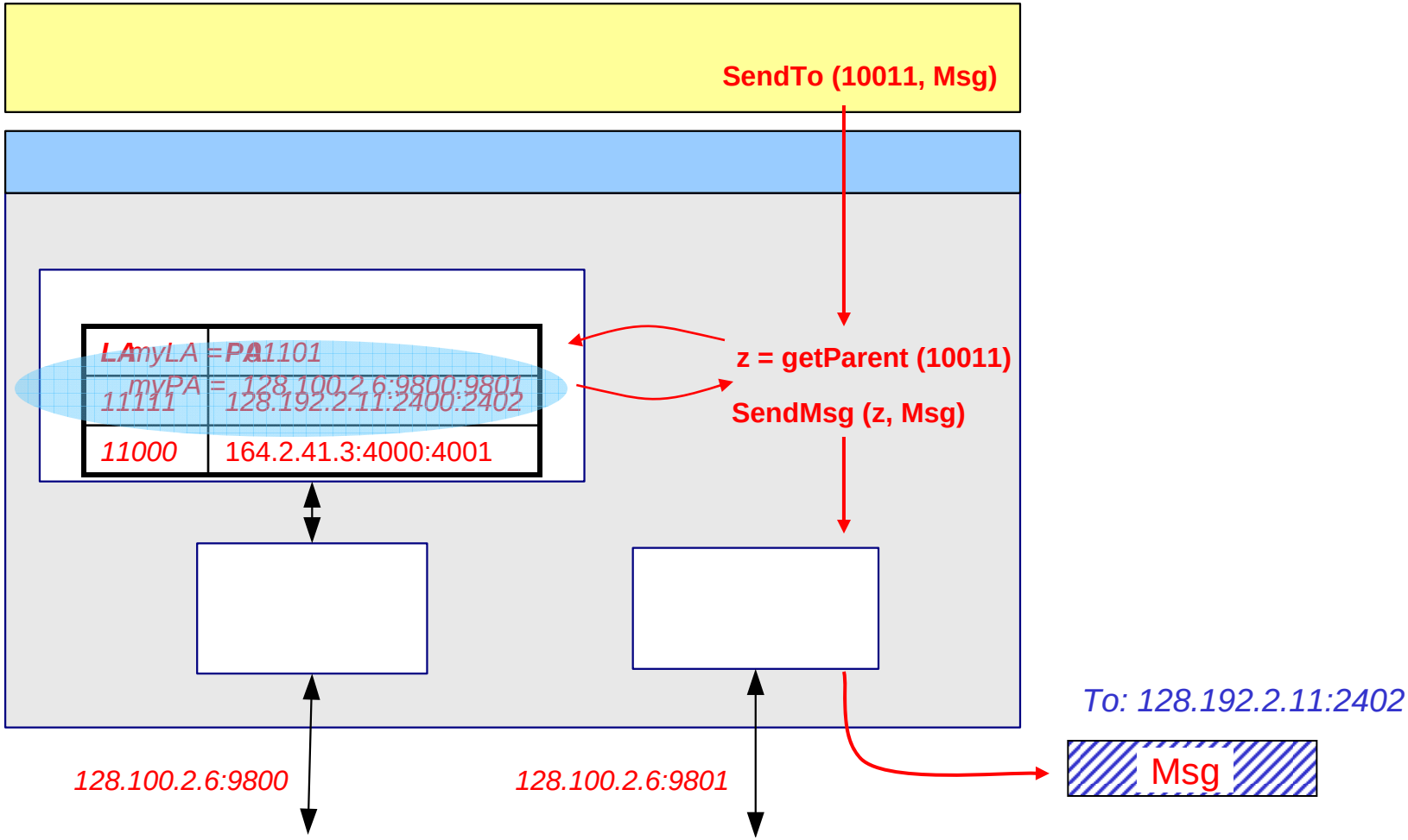
- Operations:
  - **`sendToChildren()`**
  - **`sendToParent()`**





Forwarding

- **Logical address (LA):** overlay specific address of an overlay socket
- **Physical address (PA)**: address of the overlay socket in substrate network



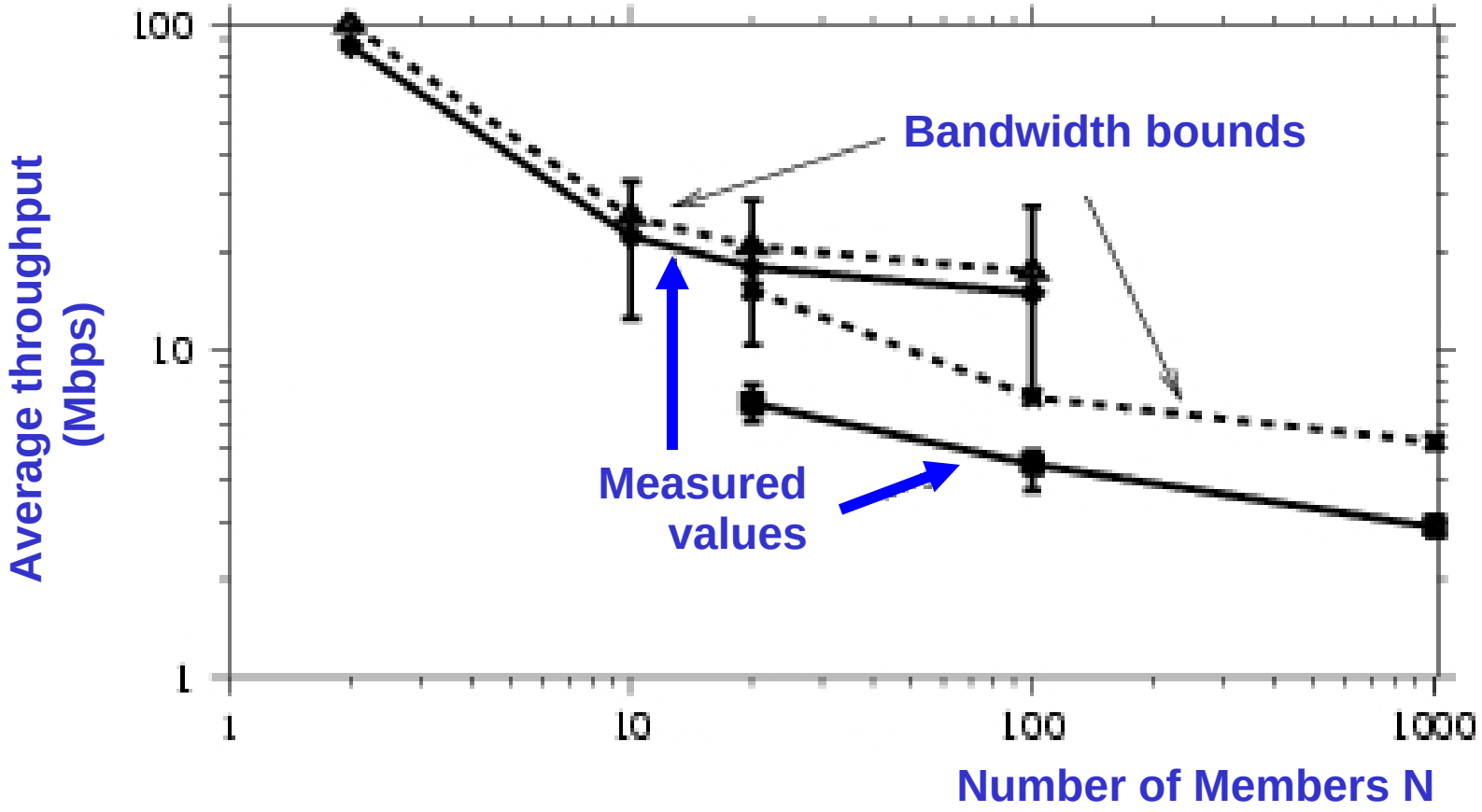
Writing Programs with Overlay Sockets

- Similarities to Berkeley Socket API
- Programs are independent of overlay topology and substrate network
- Arbitrary number of overlay sockets per applicaiton

```
//Generate the configuration object
OverlaySocketConfig ConfObj =
          OverlaySocketConfig.createOLConfig("hypercast.xml");

//Create an overlay socket
 I_OverlaySocket socket=ConfObj.createOverlaySocket(null);

//Join an overlay
socket.joinOverlay();

//Create a message
OL_Message msg = socket.createMessage(byte[] data);

//Send the message to all members in overlay network
socket.sendToAll(msg);

//Receive a message from the socket
OL_Message msg = socket.receive();

//Extract the payload
byte[] data = msg.getPayload();
```

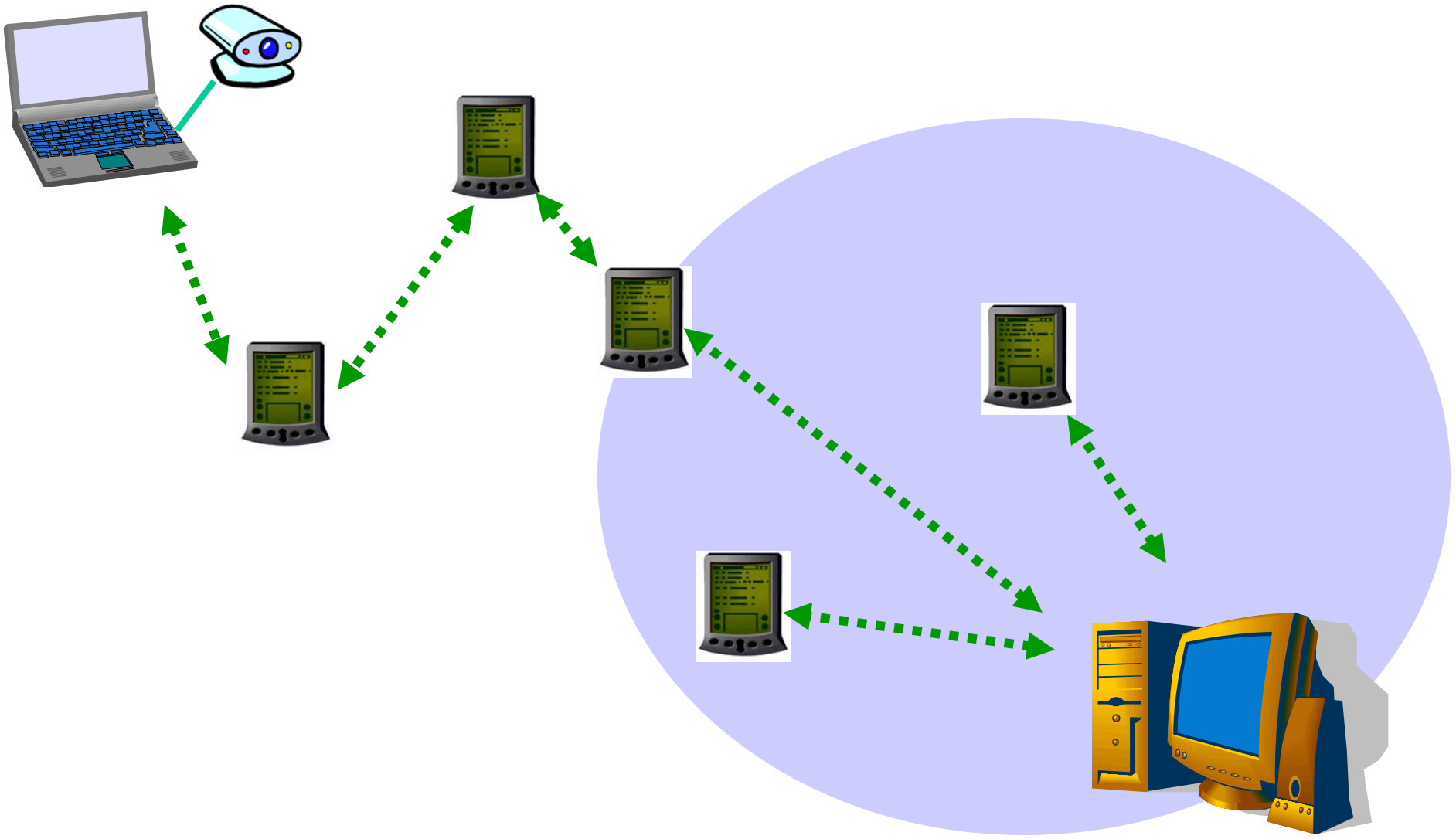
Experiment: Throughput of Multicasting

100 MB bulk transfer for N=2-100 members (1 node per PC)
10 MB bulk transfer for N=20-1000 members (10 nodes per PC)

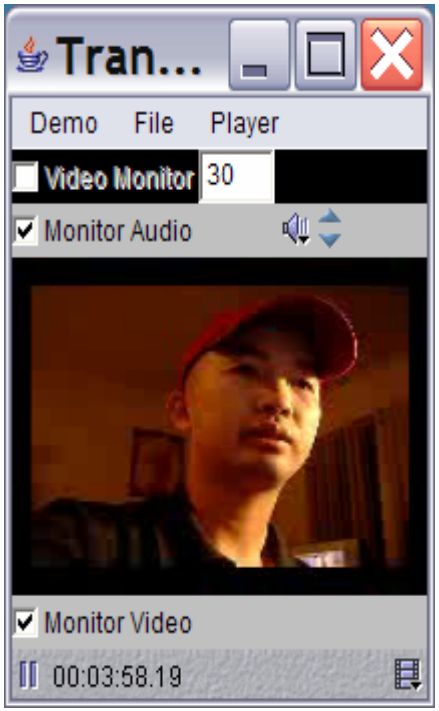


Hypercast Application:

Video-streaming in ad-hoc network



Snapshot


Tran...
Demo
File
Player
Video Monitor
30
Monitor Audio
Monitor Video
00:03:58.19

Wireless Ad-hoc Measurements
(2005)

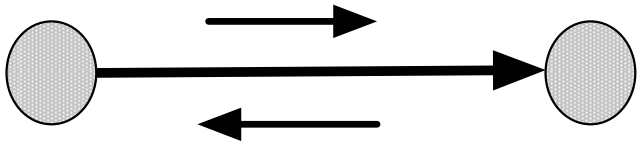
- **HP iPAQ 5555 PDA (x 8)**
  - 400MHz Intel XScale CPU, 128MB
  - Jeode JVM
  - WiFi 802.11b

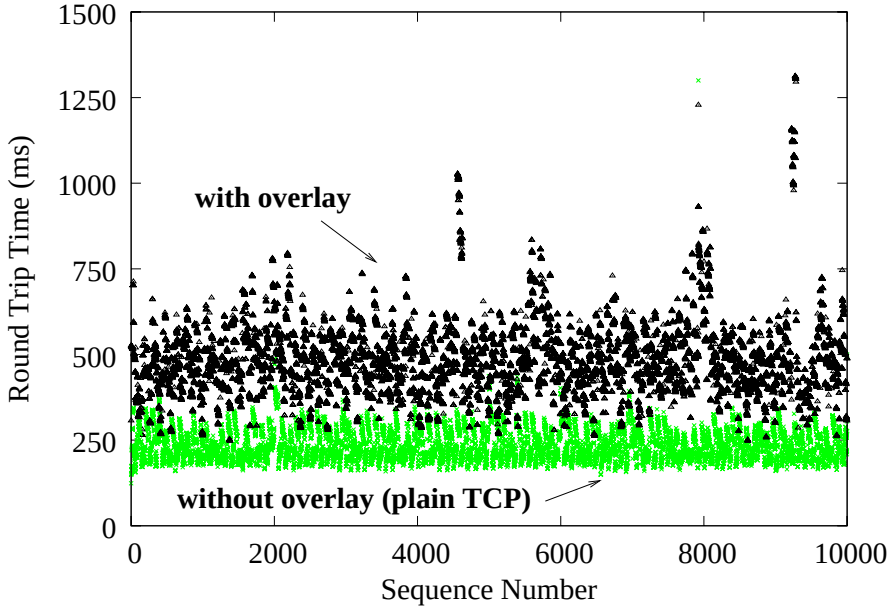
- **Toshiba M200 Tablet PC (used as monitor)**
  - 1.5GHz Intel Centrino CPU, 512MB
  - WiFi 802.11g

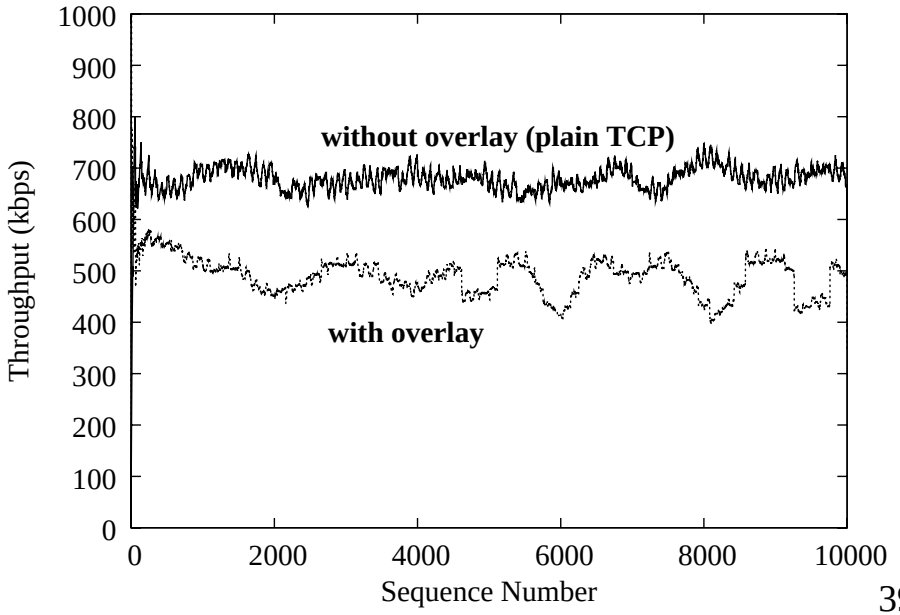
Performance of Overlay in Ad hoc network

Single hop, point-to-point:

- One sender, one receiver
- Indoor, approx. 30ft distance
- TCP is used as underlay protocol
- 10,000 messages (@512 bytes)
- Short ACK (32 bytes) is sent for each message to measure RTT
- Overlay socket is compared to plain TCP socket



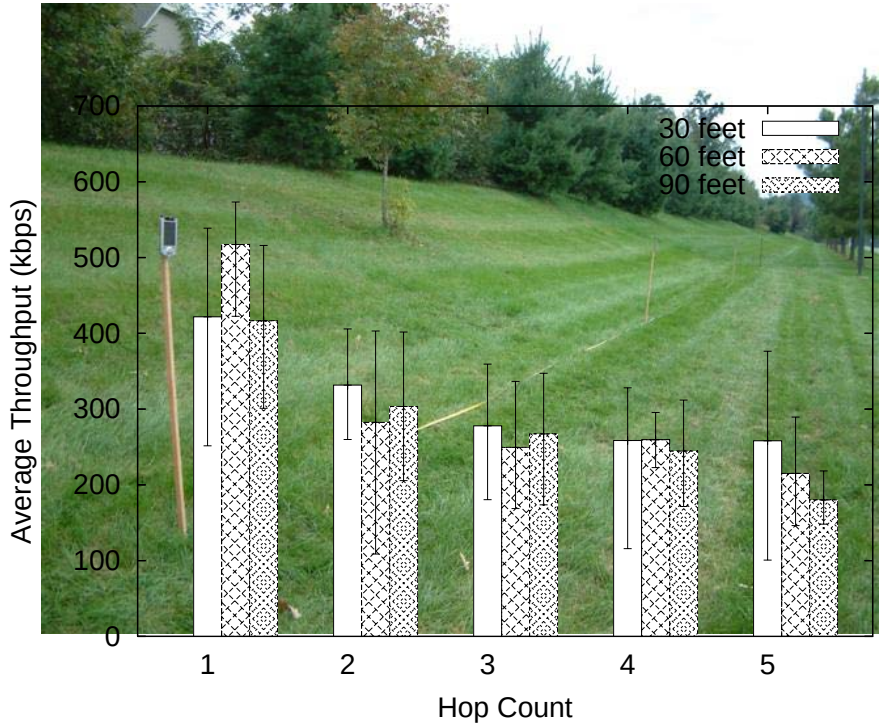




Performance of Overlay in Ad hoc network

Multi-hop:

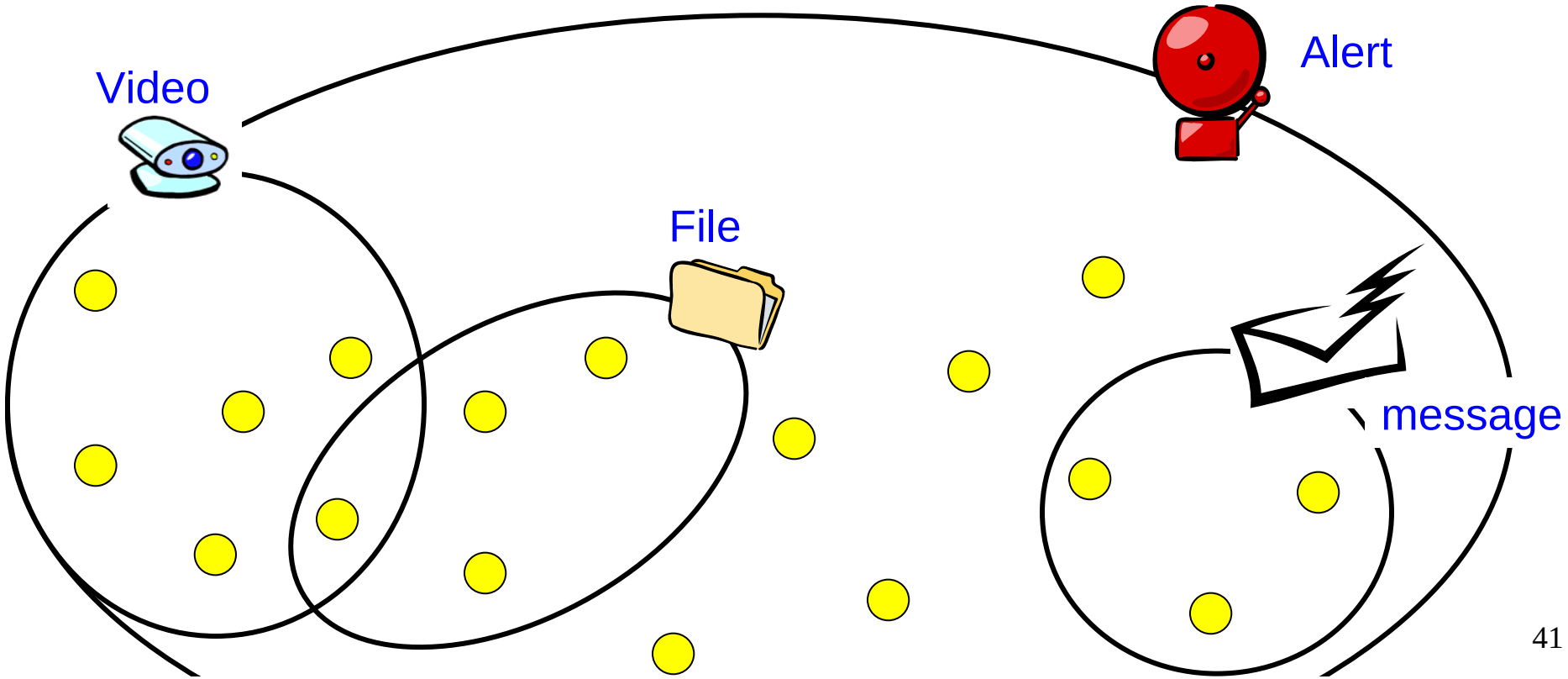
- Six iPAQ PDAs in a line topology
- Distance between PDAs is 30, 60, 90 ft
- Outdoor measurement (taken over several days)
- 10,000 messages (@ 512 byte)
- TCP adapter
- Fixed topology



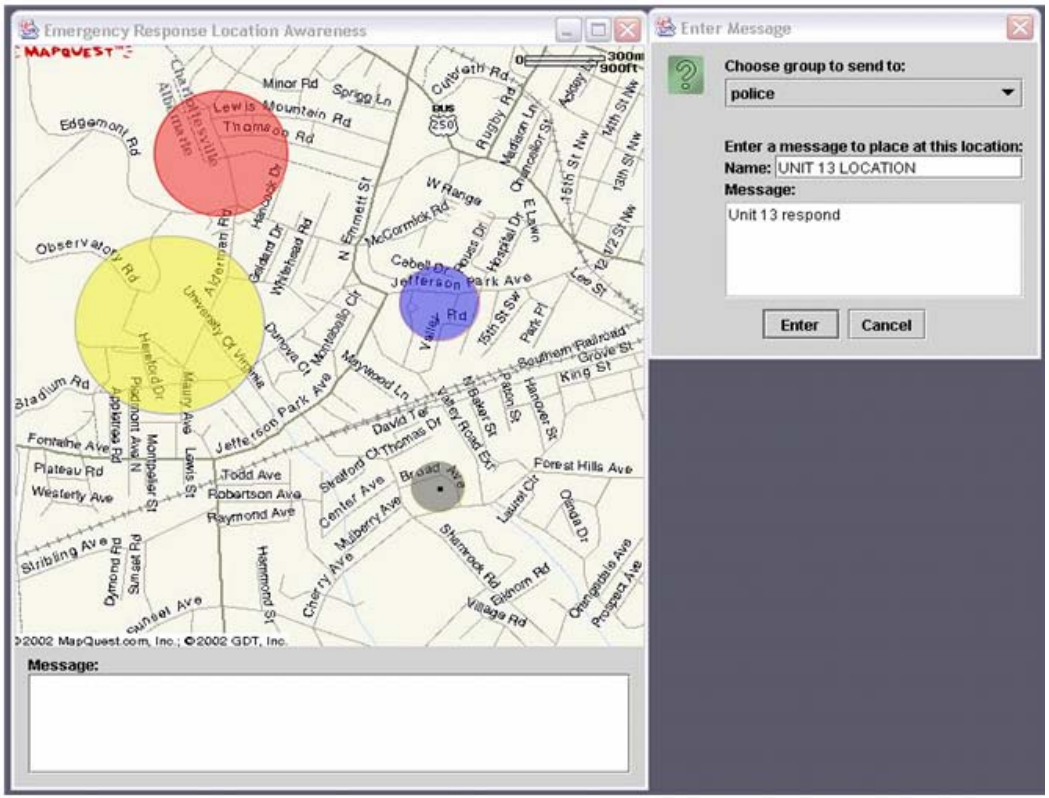
Information Management with Overlay Networks

Concept:

- Each resource or user can be a member of arbitrarily many peer networks
- Access to information is provided through dynamically created overlay networks

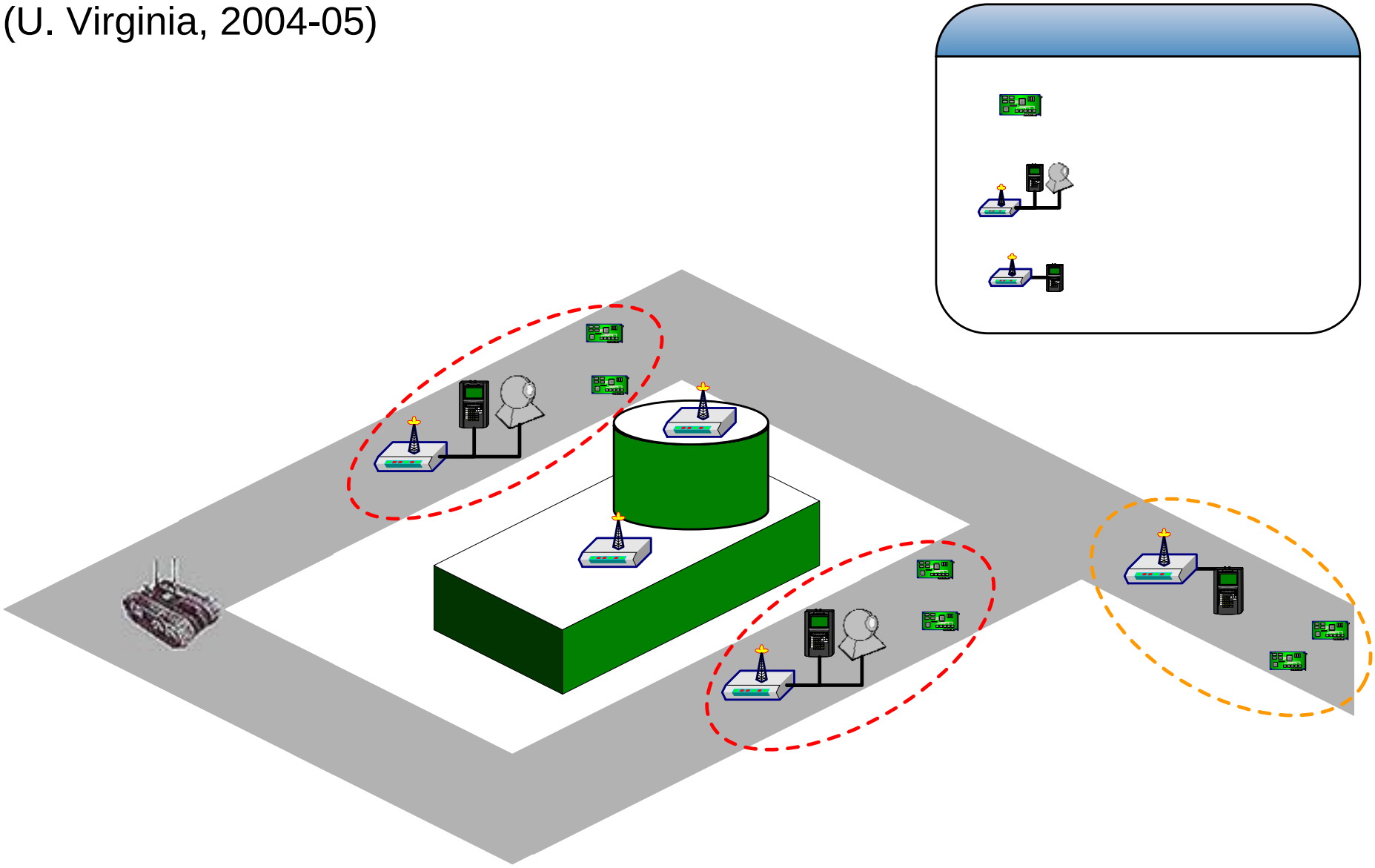


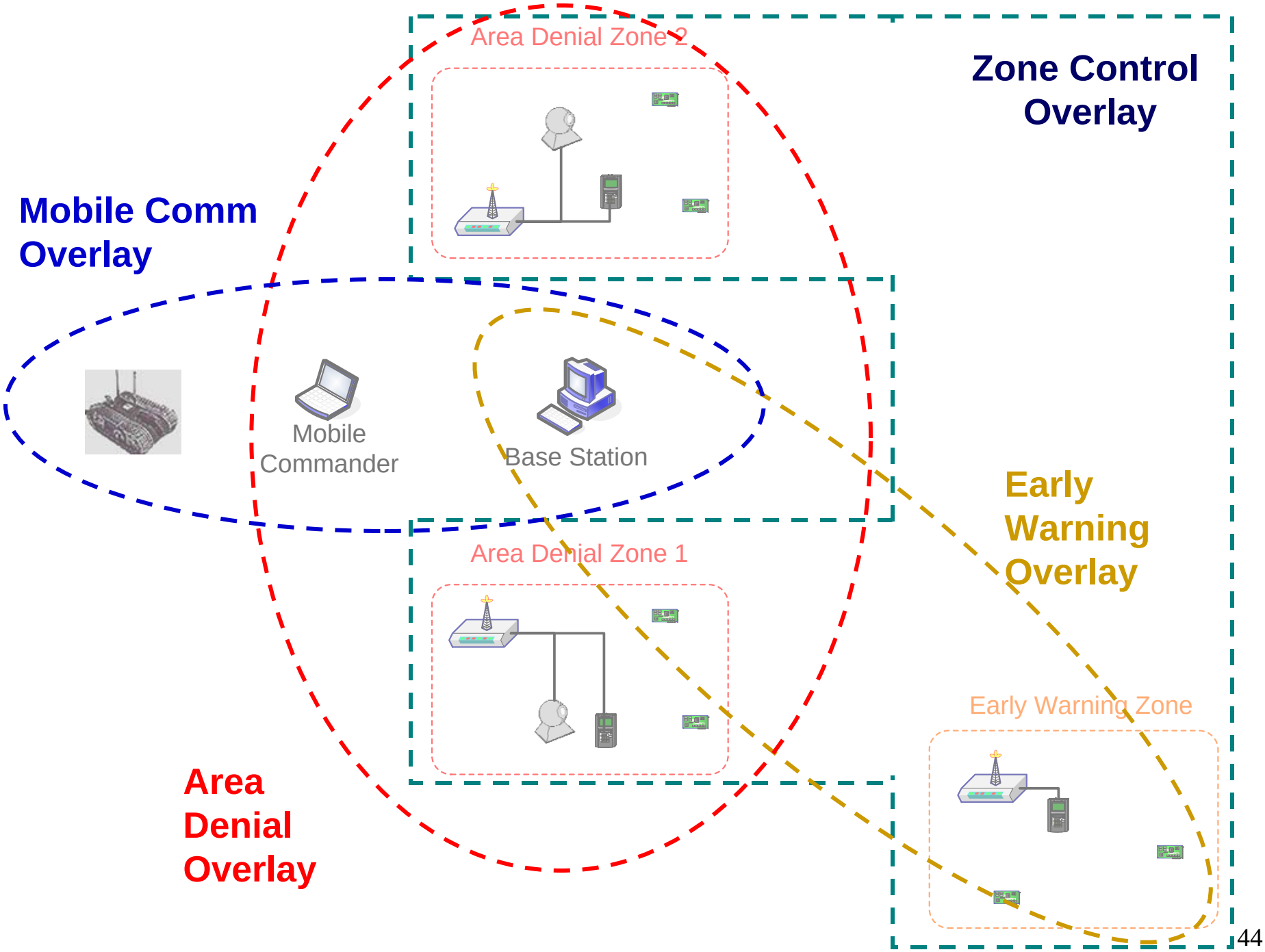
HyperCast Application:
Location Aware Messaging System
(2003)



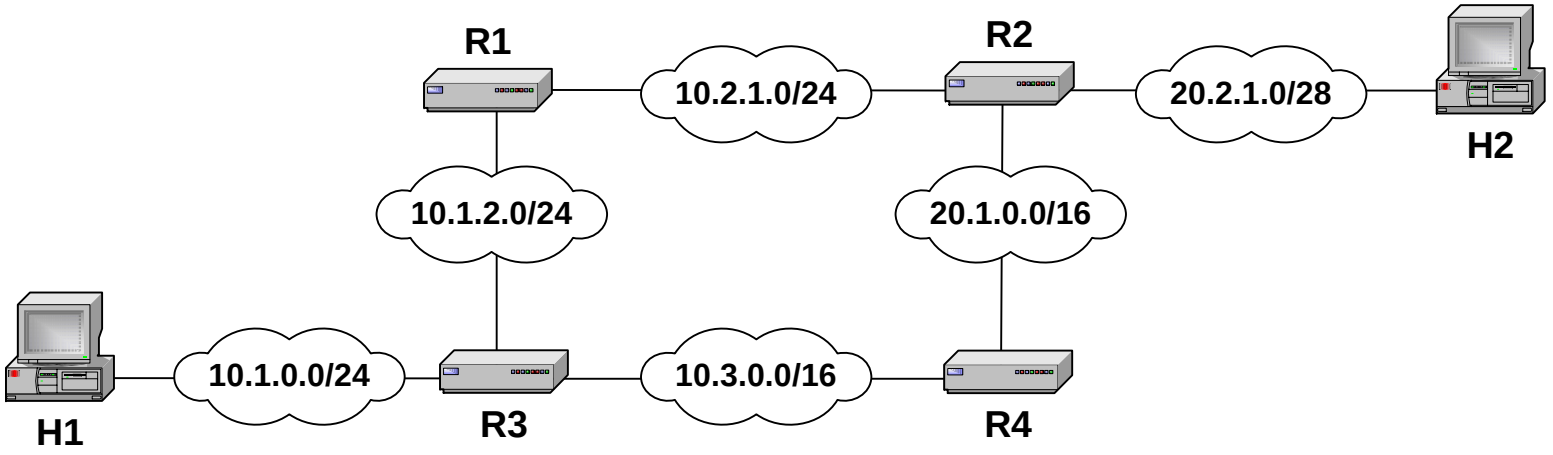
HyperCast Application: Area Protection

(U. Virginia, 2004-05)




Area Denial Zone 2
Zone Control Overlay
Mobile Comm Overlay
Mobile Commander
Base Station
Early Warning Overlay
Area Denial Zone 1
Early Warning Zone
Area Denial Overlay

30 years ago ….



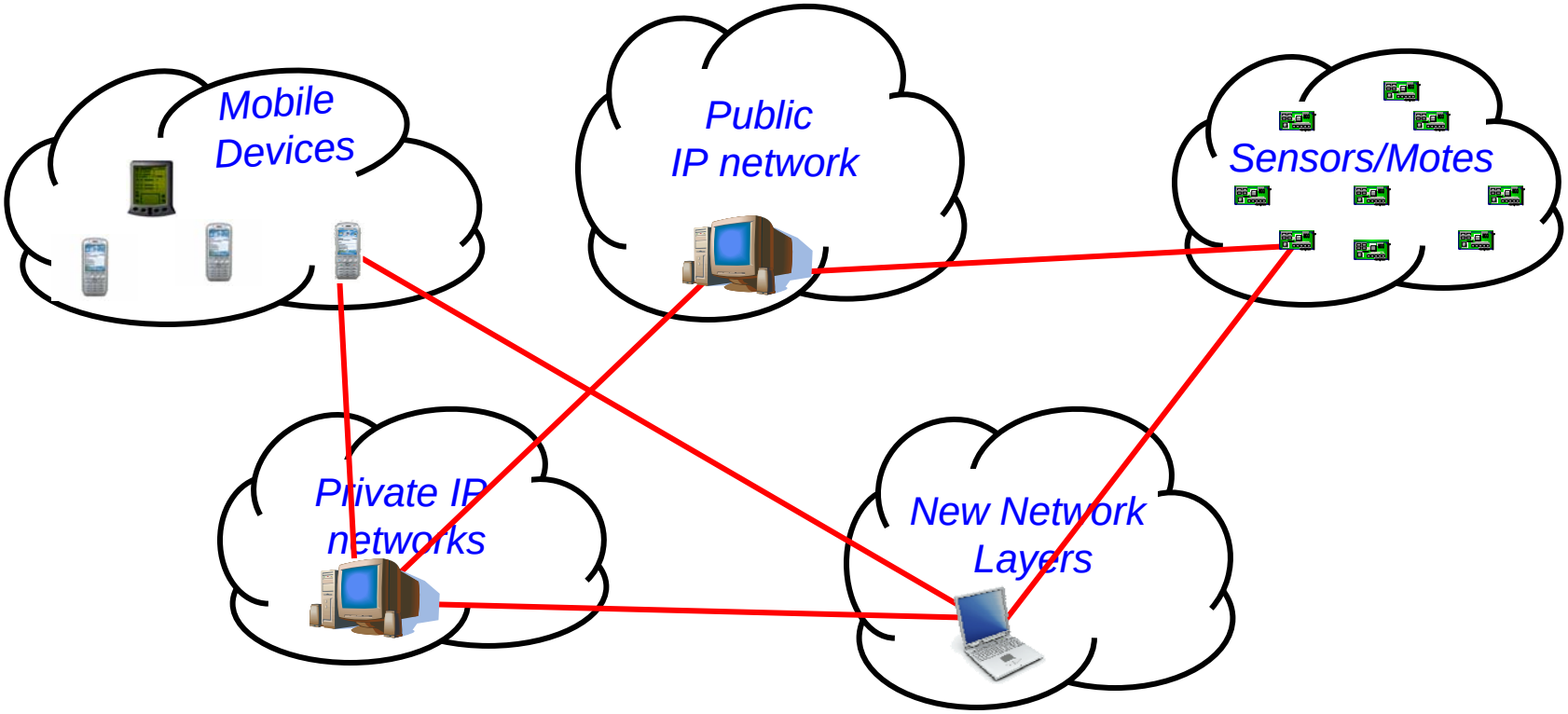
- The Internet emerges as an overlay network:
  - Collection of networks form a single internetwork
  - New service is end-to-end packet delivery

Toward an Overlay Network Architecture

Objective:

Create a network architecture entirely based on the concept of self-organizing overlay networks

Today or soon …



- Connection to a single infrastructure/address space not feasible or desirable
- **Objective: Connect applications via self-organizing overlays**
- Each overlay creates its own address space

Toward an Overlay Network Architecture

Objective:

Create a network architecture entirely based on the concept of self-organizing overlay networks

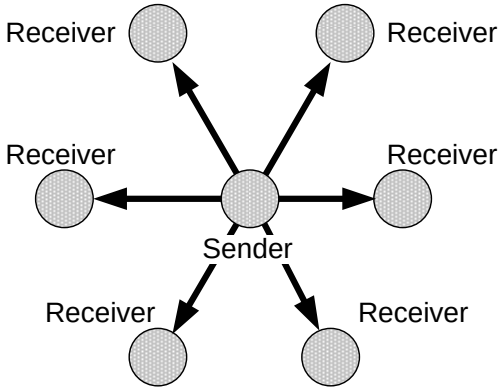
Challenges:

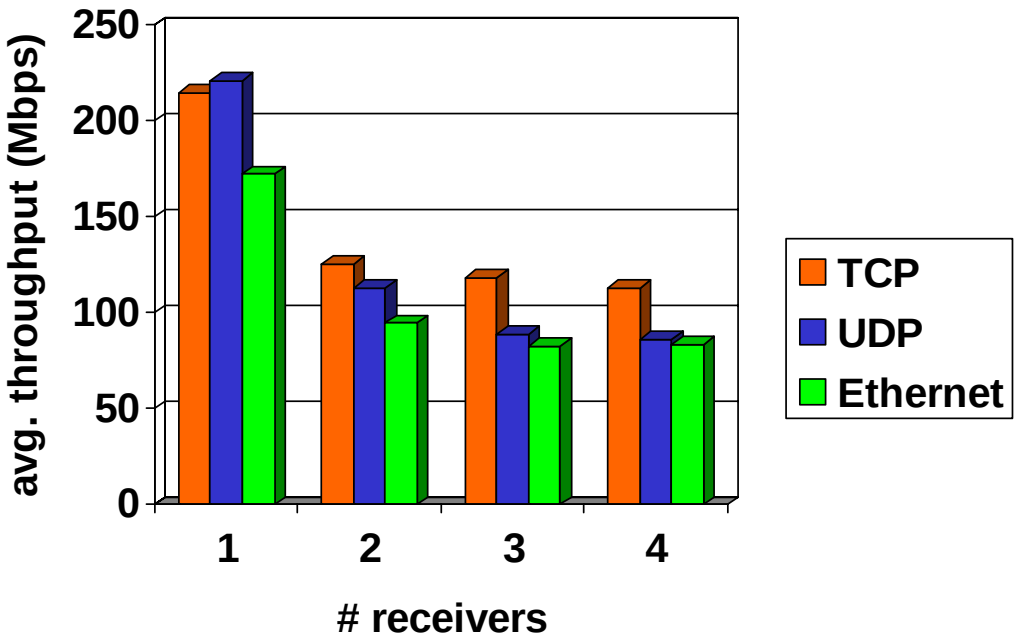
- Build overlay networks across any mix of layer-1, layer-2, layer-3, and application layer networks
- Support a virtually unlimited number of overlay networks
- Build very large overlay networks very quickly

Performance of Overlay-over-Ethernet
(2007)

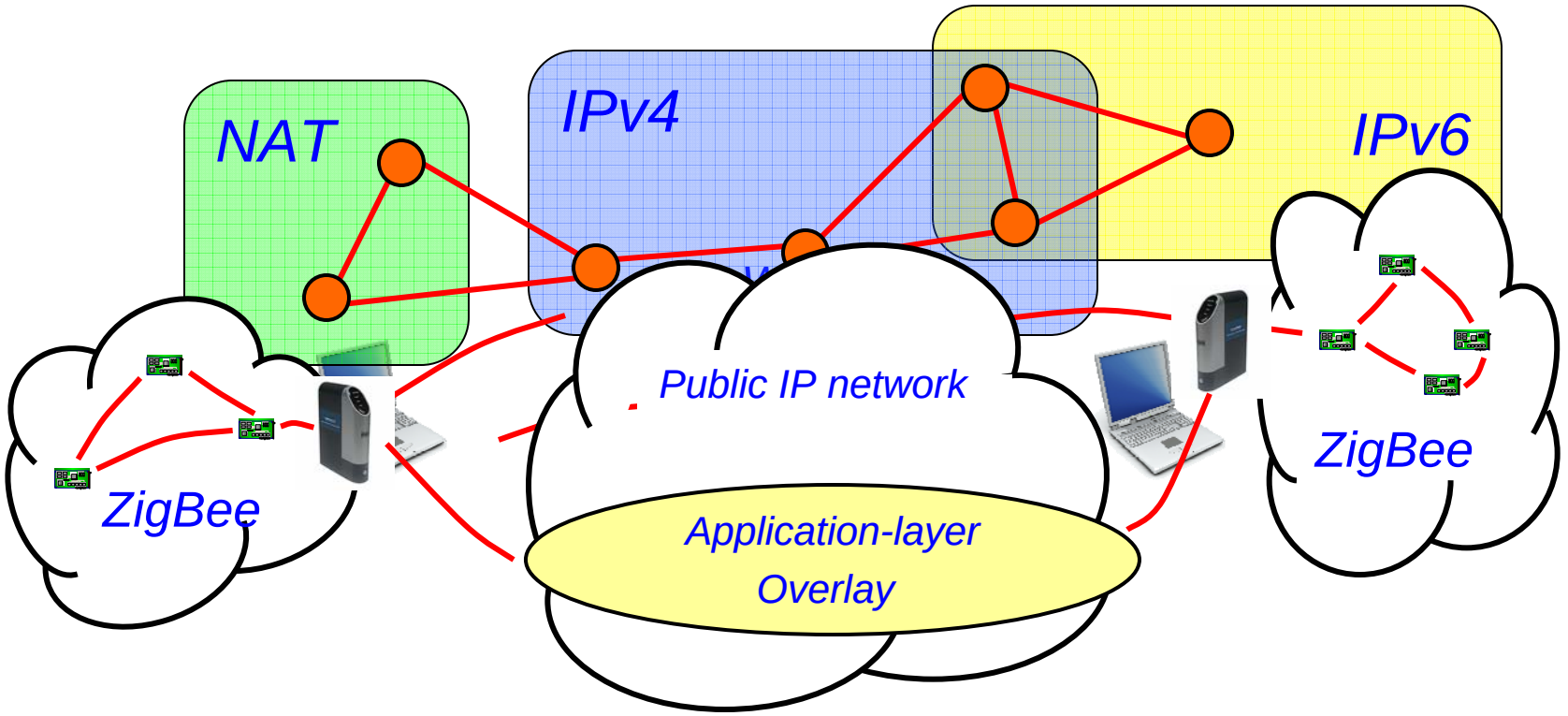
Single Hop, Multicast:

- Single overlay
- 1 sender, 1-4 receivers
- Shared Ethernet (GigE) LAN
- Single-substrate:
  - UDP
  - TCP
  - Ethernet (no IP)
- 8,192 messages (@ 1200 bytes)
- Max. transmission rate

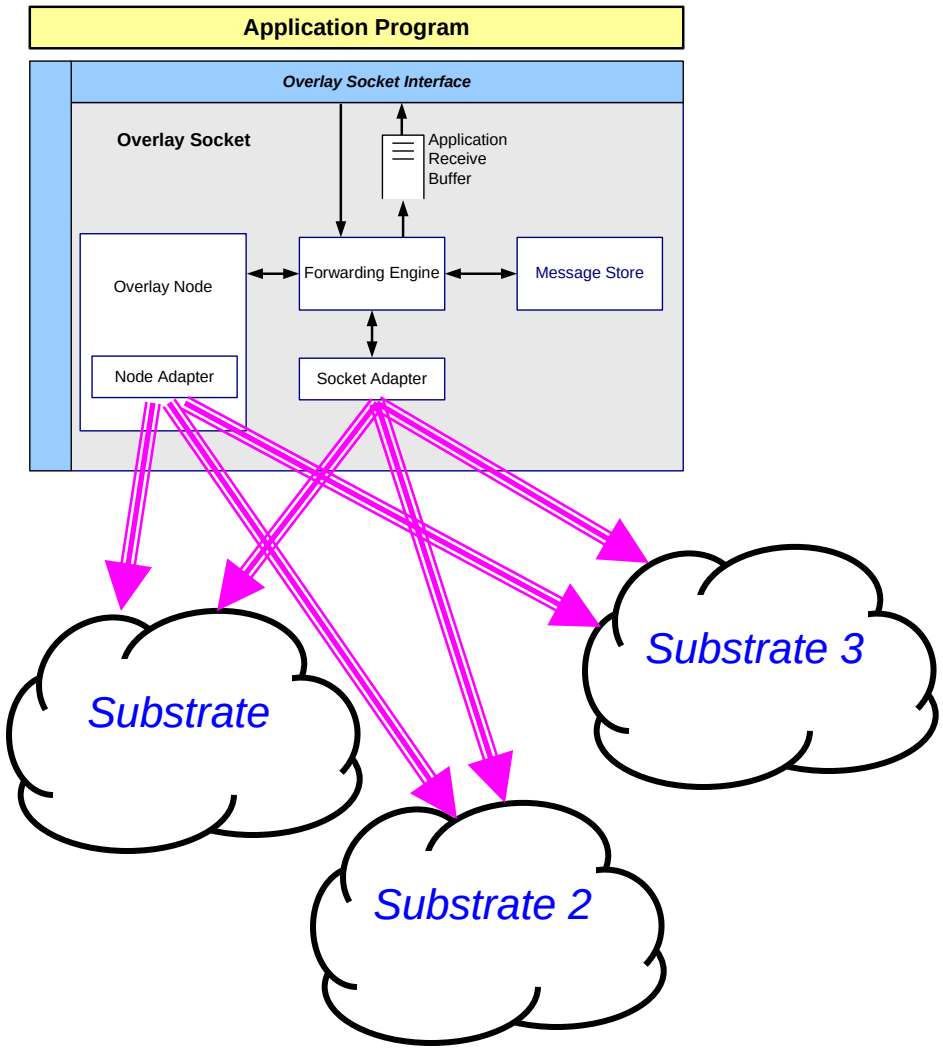




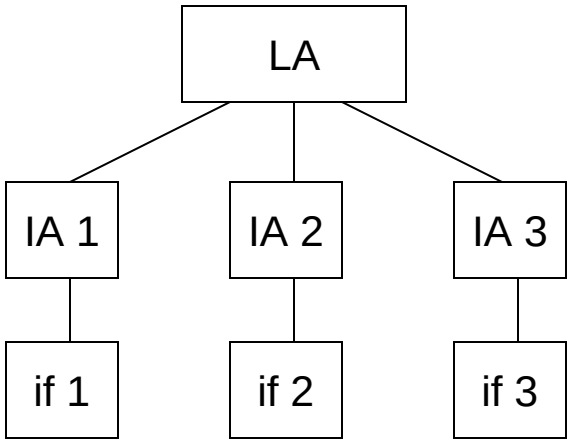
Multi-substrate networks



Problem to solve: Multiple substrate networks



The strawman



Logical addresses
(e.g., "10110001")

Interface addresses
(e.g., "128.100.243.2/3444")

interfaces

node

adapter

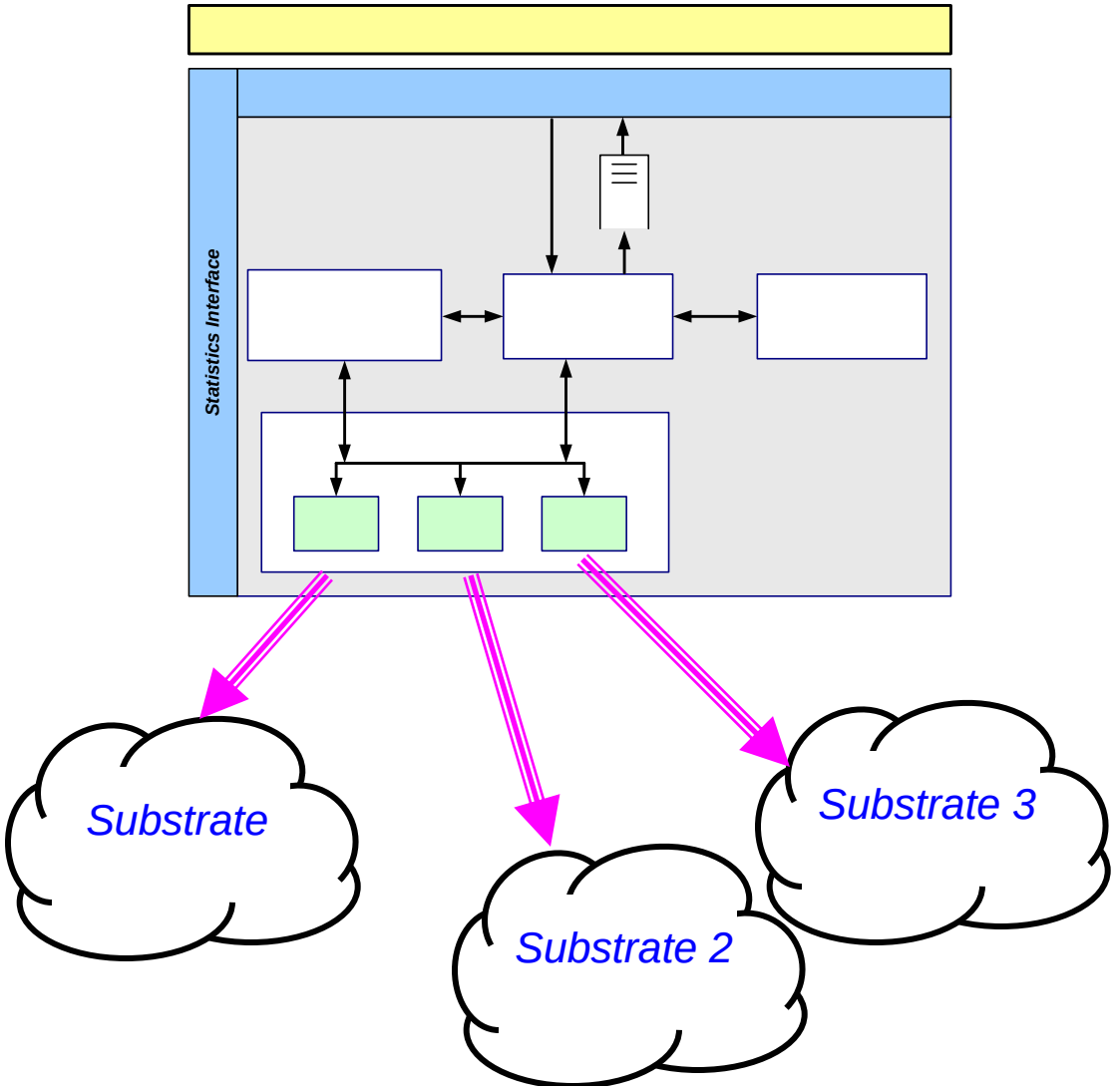
- Solution:
  - One "logical address" per socket
  - One "interface address" per substrate network
  - Each adapter has one interface for each substrate.

Note:

IP : *one IP address per layer-2 interface*

Here: *same logical address for all interfaces*

Multi-substrate overlay socket



Adapter has one interface for each substrate.

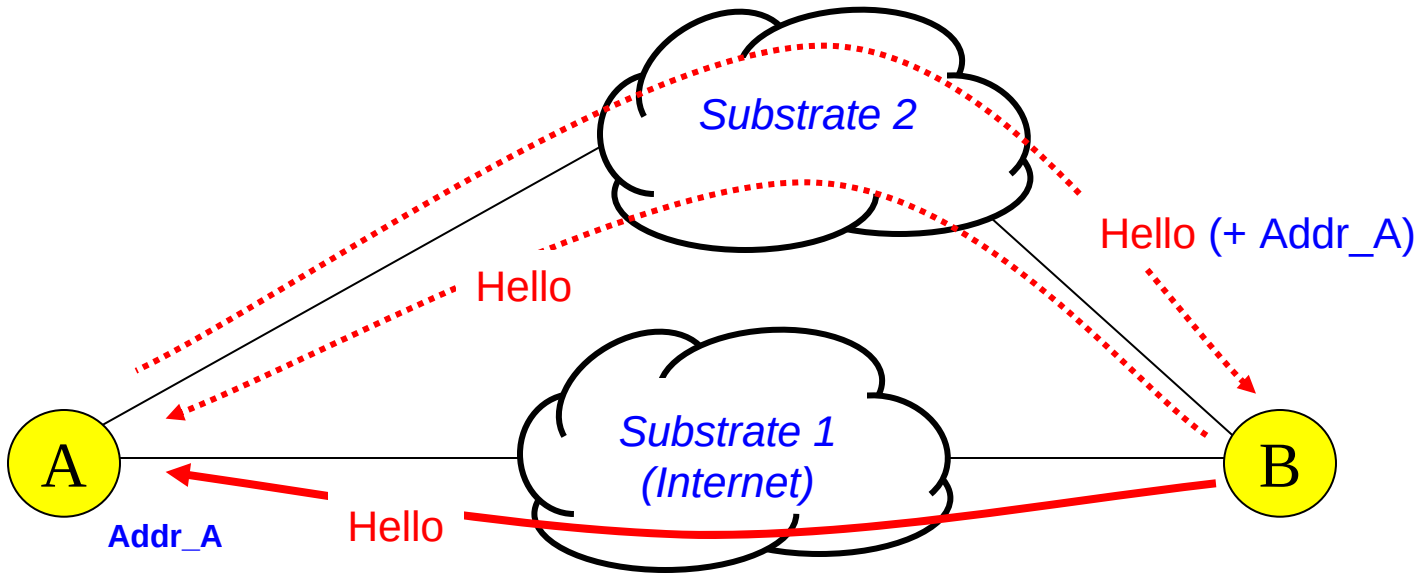
Issues in multi-substrate overlays

1. **Cross-substrate advertising**
2. **Need for and design of “relays”**

Cross-substrate advertising

Example 1:

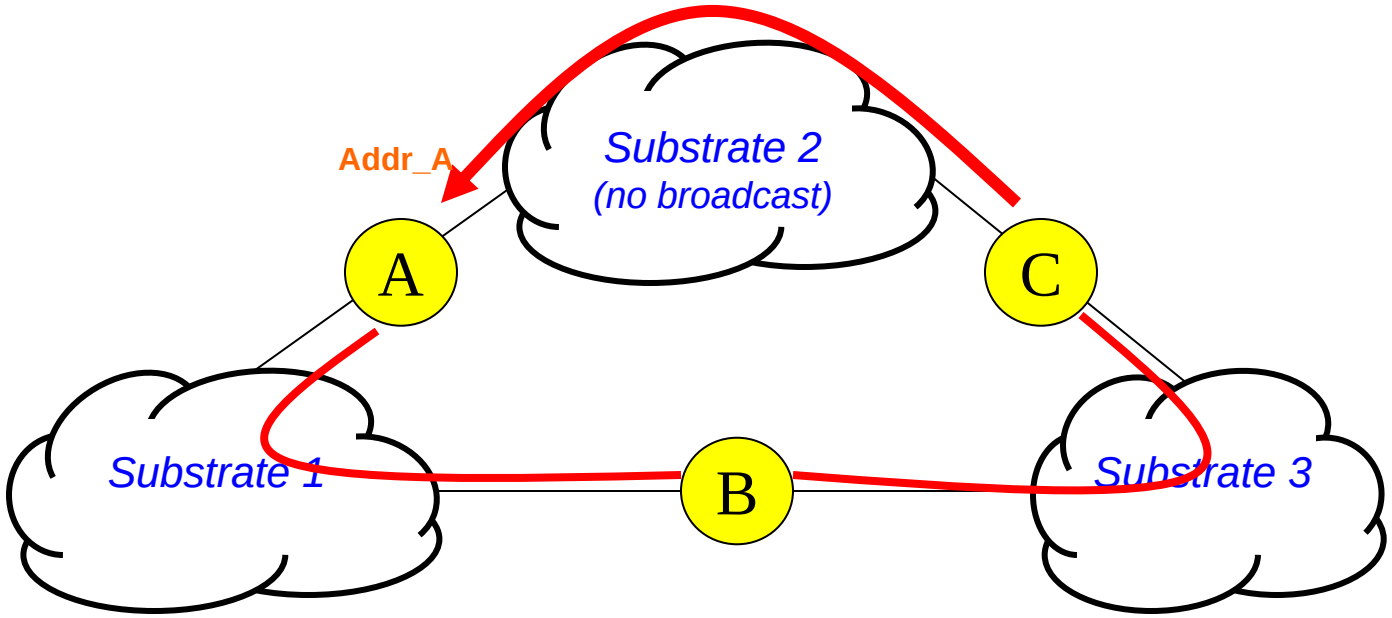
- A and B prefer Substrate 1, but does not have address of A
- Use Substrate 2 for rendezvous, and exchange address information about Substrate 1



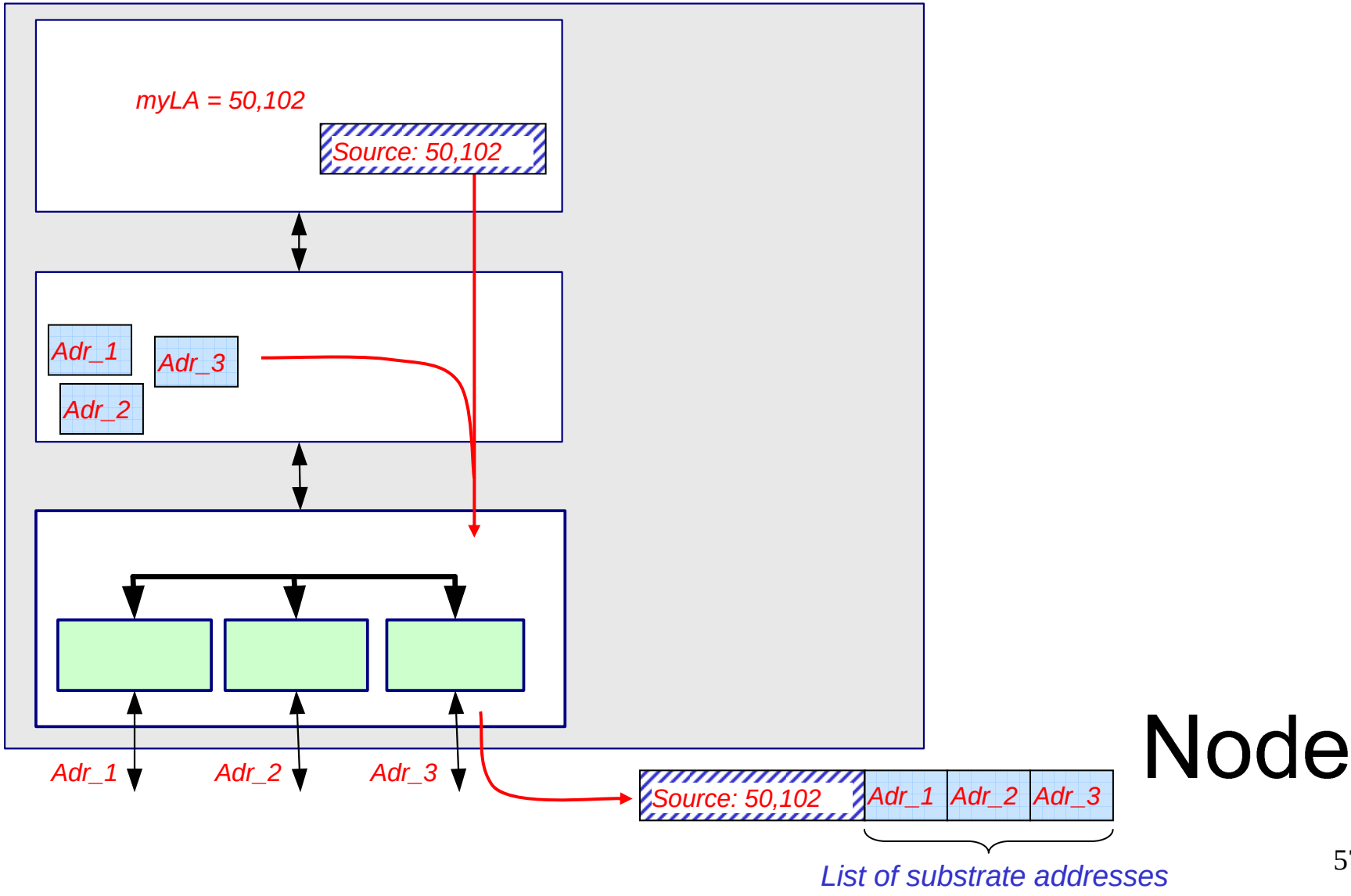
Cross-substrate advertising

Example 2:

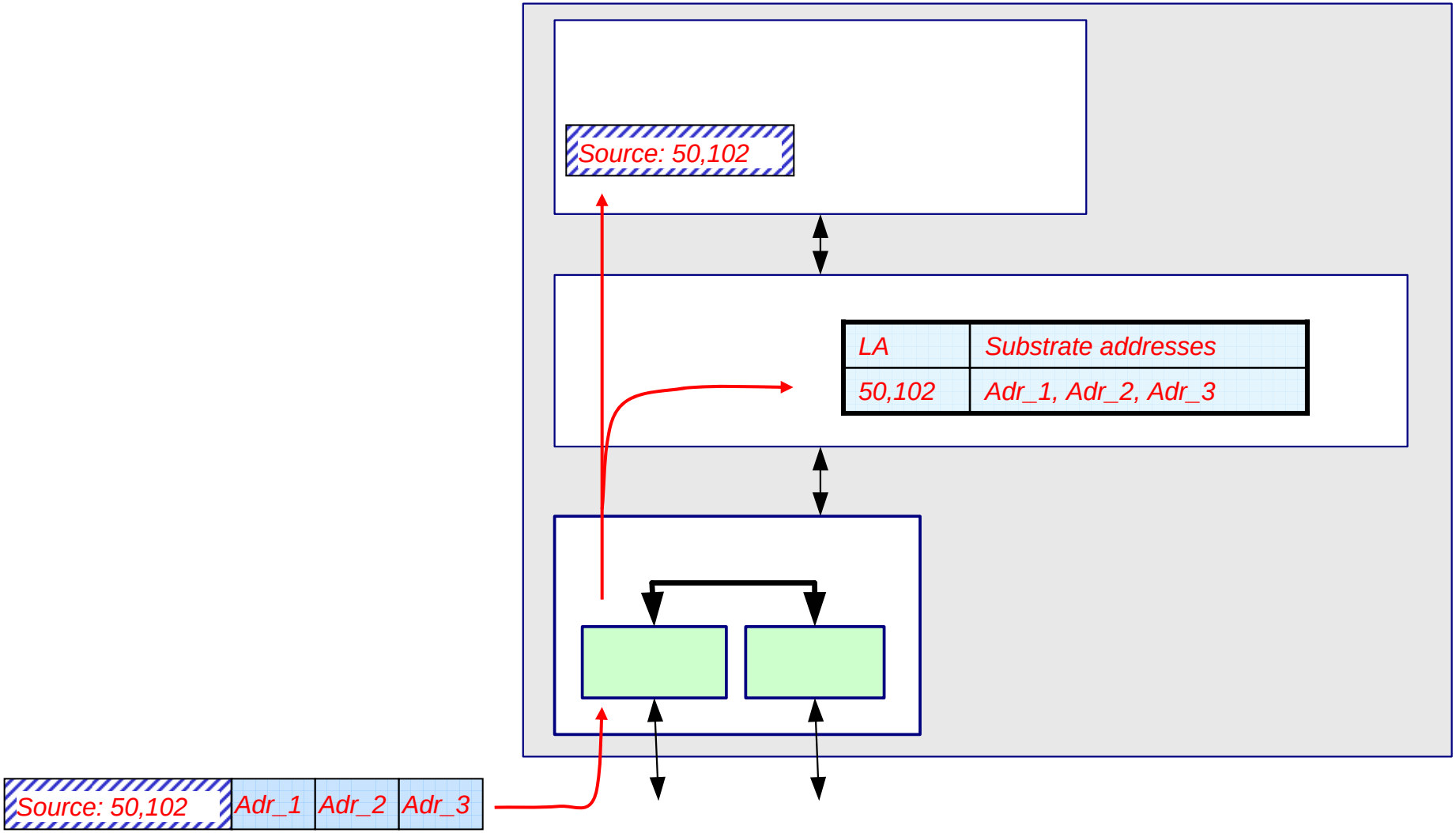
- C joins overlay via Substrate 3 with B as neighbor
- C prefers A as neighbor in the overlay
- C detects that it can talk to A via Substrate 2



Cross-substrate advertising
- Outgoing -



Cross-substrate advertising
- Incoming -

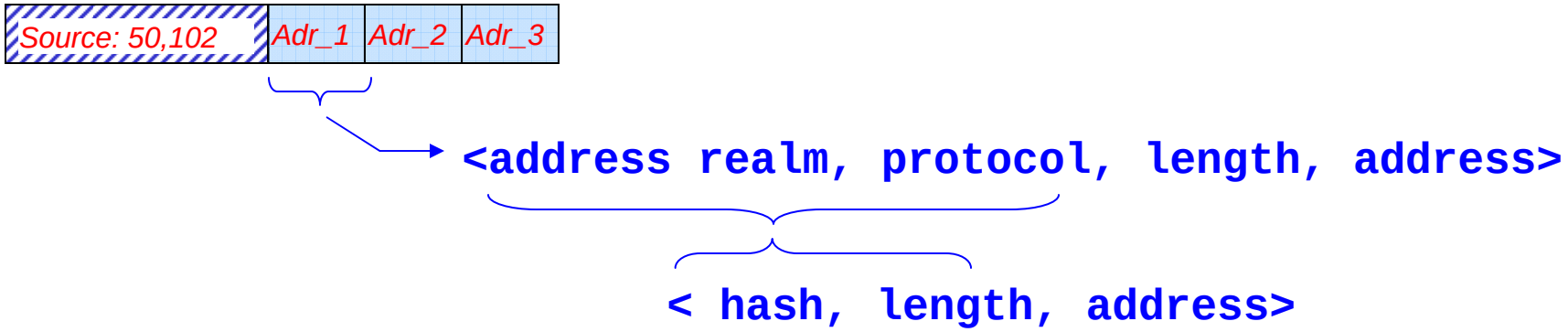


Cross-substrate advertising (CSA)
- Features -

- Independent of overlay protocol (topology)
- CSA component need not interpret addresses

- CSA messages are:
  - standalone, or
  - encapsulate protocol message
- CSA messages can be sent by:
  - Explicit request -
  - Broadcast
  - Gossiping
- Consider “preferences”
  - E.g., create topology over WiFi, data over GigE

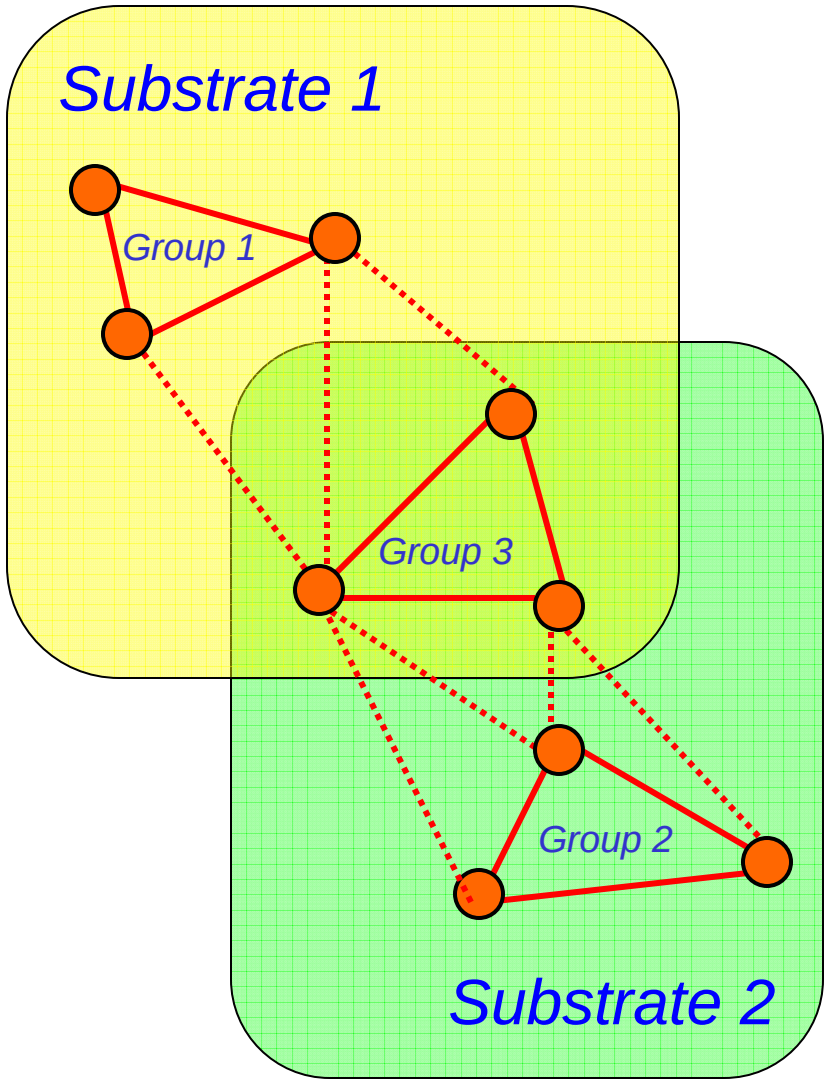
Address entries: Format



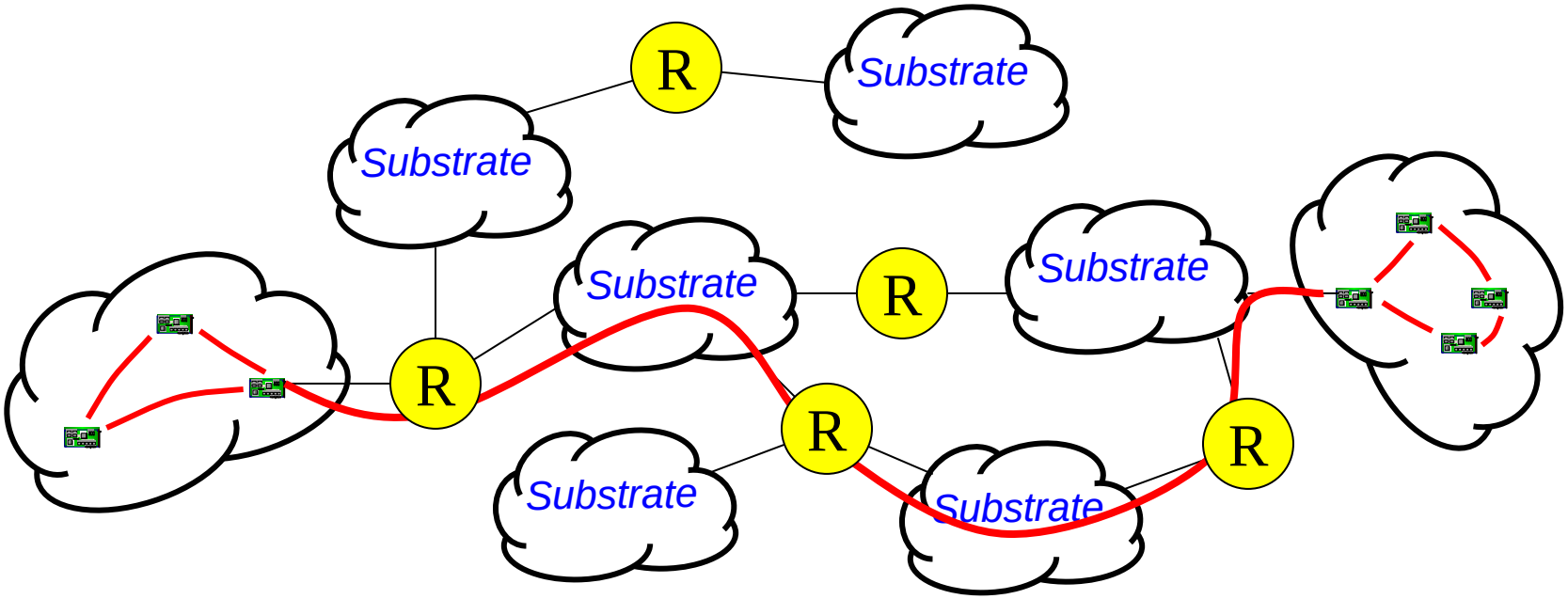
Substrate address	Address list entry
Public Internet 128.100.100.128 TCPUnicast, port 8001	<public,tcp,128.100.100.128:8001>
Private IP: private1 192.168.0.10 UDPMulticast, port 8001, Multicast address:24.228.19.78:9999	<private1,udp,192.168.0.10:8001>
MACEthernet FF:FA:34:09:41 Mux number=1234	<ethernet ,multiplexing_ethernet, FF:FA:34:09:41/1234>
OverlayInterface overlay id= 1000 protocol = DT protocol logical address =(100,100)	<1000,overlay_DT,(100,100)>

Demo: Cross-substrate advertisement

- Two substrates:
  - <udp1>
  - <udp2>
- 3 groups of overlay sockets:
  - Group 1 and 2: one interface to different substrates
  - Group 3: two interfaces
- One overlay network: DT protocol



Relays



- **Relays** provide connectivity between substrate networks
  - Can serve as rendezvous points across substrates
  - Help with getting data across substrates

Relays

- **Questions:**
  - Are relays part of the overlay network? Should they be transparent?
  - How do nodes find relay and how do relays find each other?

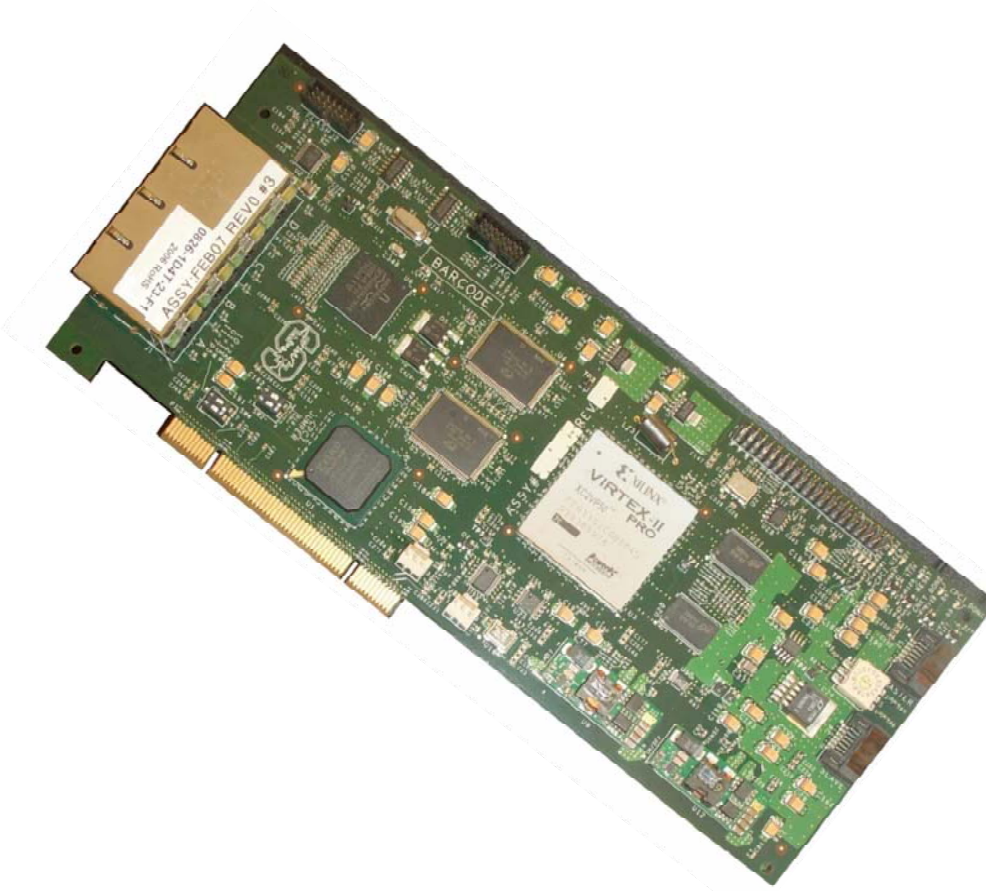
Requirements:

- Support many (unlimited) overlay networks
- Virtualization of the data plane:
  - Instantiate a separate data plane for each overlay
  - Create forwarding state upon request

Next: Overlay software on FPGA

- **Current project (2008):**
  - Overlay router (“relay”)
  - Platform:
    NetFPGA from Stanford U.
    - Xilinx Virtex-II Pro 50
    - 4 GigE ports

- Native support of (HyperCast) overlay protocols
- Arbitrary number of overlays:
  - Each overlay is a separate data plane



Source: www.netfpga.org

Summary

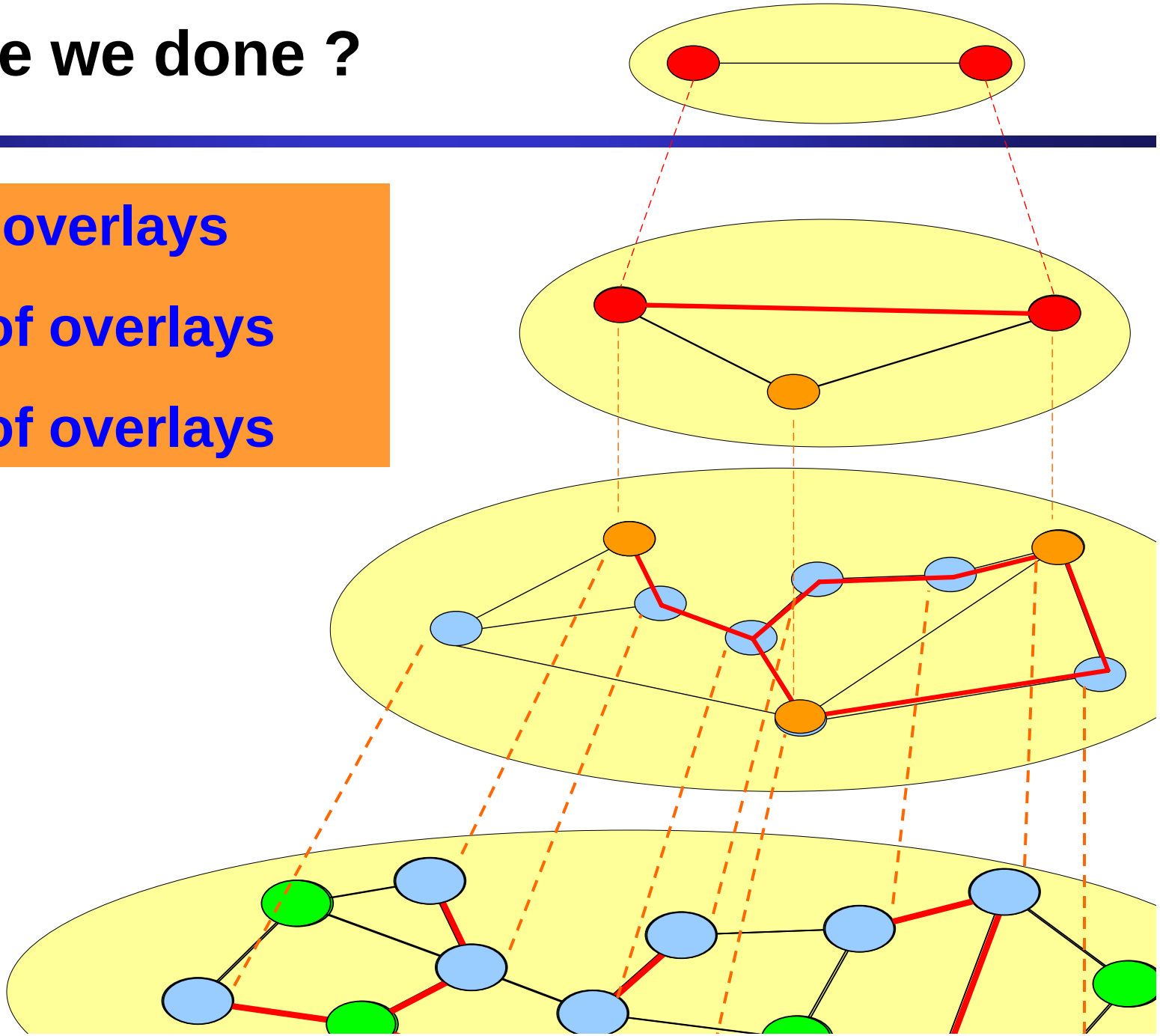
- 1960s-1990s: Overlays are key architectural element for building networks
  - *… but rarely self-organizing*
- Last 10 years: Many new insights on self-organizing application-specific overlays
  - *… but mostly single substrate (Internet)*
  - *… mostly application-layer*
- Next: New network architecture based on principles of self-organizing overlays:
  - *… each networked application builds network with own address space*
  - *… arbitrary substrates*
  - *… less dependent on infrastructures*
  - *… less need for management*
  - *… mobilility not an issue*

When are we done ?

Putting overlays

on top of overlays

on top of overlays



Some Guidance from Monty Python's Flying Circus

"Society for Putting Things on Top of Other Things" (ca. 1970)

There are many things, and I cannot emphasize this too strongly, not on top of other things.

. . .

For, we must never forget that if there was not one thing that was not on top of another thing our society would be nothing more than a meaningless body of people that had gathered together for no good purpose.