Communication Protocols
and Internet Architectures

Harvard University
CSCI S-1

Lectures #5 and #6

Instructor: Len Evenchik
(evenchik@fas.harvard.edu)

© 1998-2005 L. Evenchik

Lecture Agenda

• Logistics and Q & A
• IP Design and Implementation
• Subnets and CIDR
• Routers and Routing
• Virtual LANs
• Three Minute Wrap-Up

© 1998-2005 L. Evenchik

(C) 2005 L. Evenchik
Typical Network Topology

Host A

Layer
4
3
2
1

TH
IP
Frame 1
(802.3/LLC)

Frame 1
(802.3/LLC)

Frame 2a
FRelay

Frame 2b
FRelay

Frame 3
(FDDI)

Ethernet

Mesh Network

Router 1 (Gateway)

Router 2 (Gateway)

Host B

TH
IP

TH
IP

TH
IP

TH
IP

Frame 3
(FDDI)

Ring Network

(C) 2005 L. Evenchik
Internet Protocol (IP)

- IP protocol grew out of the DOD ARPAnet need for a network layer protocol that could handle the growth and variety of the underlying networks.
- IP assumes the presence of a mostly reliable substrate, but does not define what that substrate must be.
- Originally defined in a series of "Requests for Comments" (RFCs).
- Additional protocols were defined in concert with IP to handle reliable packet delivery, routing, address resolution, network problem management, etc., etc., etc..
- Many protocol developers have learned from IP and other protocols of that era (and continue to do so today.)
Services Provided by IP

- Connectionless, datagram service only. No delivery acknowledgment or delivery guarantee is offered.
- Datagram service is another way to say unreliable.
- IP datagram service is usually used in conjunction with a connection-oriented, reliable transport layer protocol such as TCP (given that a reliable service is required.)
- Transport layer performs end-to-end sequencing and error detection and correction to compensate for IP's unreliability. (IP can lose and mis-order packets.)

IP Addressing

- Internet addresses specify a host's connection(s) to a network. A host may have more than one IP address - to the same or to multiple networks.
- Internet addresses are 32 bits long and include both a network ID and a host address. “Classic” addresses are of four different forms: classes A, B, C and D. (This was the original approach, it has since been updated.)
- IP addresses are written with dotted decimal notation (for example, 128.192.5.3)
- Internet routing is based on network ID.
- Internet addresses must be mapped to physical network addresses (for example, by the Address Resolution Protocol on LANs).
“Classic” Classful IP Address Format

Class A
- NET: 0
- HOST: 24 bits
- Class A: 1.0.0.0 through 126.0.0.0

Class B
- NET: 1 0
- HOST: 16 bits
- Class B: 128.1.0.0 through 191.255.0.0

Class C
- NET: 1 1 0
- HOST: 8 bits
- Class C: 192.0.1.0 through 223.255.255.0

Class D
- NET: 1 1 1 0

Address Resolution

- Address resolution is a general problem of mapping Internet (global) addresses to physical (local) addresses.
- Both static and dynamic address resolution algorithms are used in various networks.
- ARP (RFC 826) is a low-level protocol that dynamically maps IP addresses to physical addresses. ARP requests require network broadcasts.
- An address resolution cache is important to reduce network overhead. Cache entries should be updated by all stations on the network.
- RARP is used by diskless workstations to obtain their IP addresses.
**ARP Format and Encapsulation**

**ARP Protocol Format**

<table>
<thead>
<tr>
<th>Hardware</th>
<th>Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLEN</td>
<td>PLEN</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ARP encapsulated in Ethernet frame

| Ethernet Header | Ethernet Payload - encapsulated ARP message |

Source: ARP RFCs  © 1998-2005 L. Evenchik

**Ethernet ARP Procedure**

Who is 18.12.121.43: Ethernet broadcast

© 1998-2005 L. Evenchik
ARP Cache Information

fas% arp -a

admin.fas.harvard.edu (140.247.30.41) at 00-00-f8-09-24-fe
ns3.fas.harvard.edu (140.247.30.30) at 00-00-f8-03-ba-8a
core4-33.fas.harvard.edu (140.247.33.24) at 00-60-6d-21-11-b1
sc-mr-gw-vl30.fas.harvard.edu (140.247.30.1) at 00-d0-d3-28-f2-58
is01.fas.harvard.edu (140.247.30.101) at 00-00-f8-71-ad-e7
is05.fas.harvard.edu (140.247.30.105) at 00-00-f8-09-24-f6

..............
(Note that the above is an abridged display of the output from arp -a)

Services Provided by IP

• Connectionless, datagram service only. No delivery acknowledgment or delivery guarantee is offered.
• Datagram service is another way to say unreliable.
• IP datagram service is usually used in conjunction with a connection-oriented, reliable transport layer protocol such as TCP (given that a reliable service is required.)
• Transport layer performs end-to-end sequencing and error detection and correction to compensate for IP's unreliability. (IP can lose and mis-order packets.)
Sequencing, Flow Control, Option Negotiation

- Since IP does not perform sequencing, IP does not guarantee the arrival sequence of packets. That's left to a higher layer protocol (if needed.)
- Absence of connections and sequence numbers means no mechanism for flow control. The system must rely on the Data Link layer and/or the transport layer for flow control (if needed.)
- No connection means no opportunity to negotiate options.
- Options field in IP header available to indicate option request; receiver either accepts or ignores datagram according to its ability to honor option selection
Error Control

- IP assumes a generally error-free Data Link layer, and assumes that the higher layers will handle data loss and sequencing errors.
- Since there is only datagram service, error handling method within IP is to discard datagrams that cannot be processed.
- IP header does contain a header checksum field but it is of limited usefulness today.

Fragmentation and Reassembly

- Datagram size is limited by the underlying physical network.
- Once a datagram is fragmented it should remain fragmented until it reaches its final destination.
- All hosts must be prepared to accept a minimum datagram size of 576 octets.
- IP fields necessary for fragmentation and reassembly are address fields, length fields, ID, MF flag, and fragment offset.
- Some datagrams should not be fragmented.
- Fragmentation can be important for networks that carry Voice over IP traffic.
Addressing, Subnets and CIDR

“Classic” Classful IP Address Format

Class A

Class B

Class C

Class A: 1.0.0.0 through 126.0.0.0
Class B: 128.1.0.0 through 191.255.0.0
Class C: 192.0.1.0 through 223.255.255.0
Class D: 224.0.0.0 through 239.255.255.255
Determining IP Address Information

- UNIX command is `ifconfig –a` (many options)
- Windows command is `ipconfig`
- Sample output from command:
  
  Host Name .......... fas.harvard.edu
  IP Address.........140.247.34.98
  Subnet Mask.......255.255.255.0
  Default Gateway...140.247.34.1

- Other fields can be displayed for hardware and DHCP information

Special IP Addresses

- 0.0.0.0 this host on this net
- 0.0.0.hostid specified host on this net
- 127.0.0.1 loopback address
- 255.255.255.255 limited broadcast, this subnet
- *.*.*.255 net directed broadcast

The value 0 typically means this host or this net.
Subnet Addressing

- Original IP address space not optimized for large numbers of small networks
- Subnetting allows a single network, which is used by an organization, to be split into multiple networks for internal use.
- The organization still appears as a single network to the outside world
- Subnets require subnet masks to separate “networks” from “hosts”

Simplified Subnet Topology

As seen from the Internet, Harvard is known by its class B address 140.247.y.z

The Harvard network’s IP address is 140.247.y.z

Routing within the Harvard network is done using a three byte network number (netmask 255.255.255.0 aka /24)
Regional IP Addresses

- 194.000.000.000 to 195.255.255.255
  Europe
- 198.000.000.000 to 199.255.255.255
  North America
- 200.000.000.000 to 201.255.255.255
  Central and South America
- 201.000.000.000 to 203.255.255.255
  Asia and Pacific

Private IP Addresses

- RFC 1918 - private network address
- 10/8 10.0.0.0 to 10.255.255.255
- 172.16/12 172.16.0.0 to 172.31.255.255
- 169.254/16 169.254.0.0 to 169.254.255.255
- 192.168/16 192.168.0.0 to 192.168.255.255
Addressing using Network Prefixes

- Address descriptions using the Network Prefix Length format are now the standard (but many people don’t know this.)
- Notation is: address / <prefix length>
  - Class A, networks 1 - 126, /8 prefix, 34.2.3.4/8
  - Class B, networks 128 - 191, /16 prefix, 140.247.30.33/16
  - Class C, networks 192 - 223, /24 prefix, 198.3.4.23/24
- Prefix approach is also used with subnetting.
- Consider 140.247.198/23 as an example

CIDR – Classless InterDoman Routing

- CIDR is also known as “Supernetting”
- CIDR requires the use of network prefixes
- CIDR provides a solution to the “3 bears problem” and the resulting growth of internet routing tables.
- CIDR provides hierarchical routing and route aggregation via grouping of network addresses
- Routers must evaluate longest address match for proper routing decisions
- Consider 140.247.198/23 as an example
Routers and Routing (part 1)

Repeater versus Bridge/Switch versus Router

- **Repeater/Hub**
  - Improved distance
  - End stations see one physical LAN
  - Single Broadcast domain, single collision domain

- **Bridge/Switch/Ethernet Switch**
  - End stations see one logical LAN
  - Protocol insensitive
  - Single broadcast domain, multiple collision domains

- **Router/L3 Switch**
  - Protocol sensitive
  - Traffic isolation
  - End stations see multiple LANs
  - Multiple broadcast and collision domains

© 1998-2005 L. Evenchik
**Generic Router Functionality**

- Operate at layer 3 and specific layer 3 protocols are supported (or not supported)
- Provide multiple collision and broadcast domains
- Router services are requested explicitly by the end stations on the network. End stations send packets to the router if they need to leave the local network.
- Supports IP fragmentation and other protocol specific network layer functionality such as Novell SAP filtering, Apple zones, etc.
- WAN links are easily integrated with routers
- Access lists are typically implemented
- Supports large complex network topologies
- Installation planning and network management is required

---

**What Does a Router Do?**

- Forwards packets based on some understanding of “next best hop”
- Talks to other routers to update its table of “next best hop” information
- Implements management and control functions
Two Network Routing Behavior

- What is the behavior to reach hosts on the local network?
- What is the behavior to reach hosts on another network?
- How is ARP used in this topology?

Simplified Router Schematic

© 1998-2005 L. Evenchik
Ethernet Headers are Built from Scratch on Each Side on a Router

Simplified Router Schematic
1 LAN and 1 WAN Port
Simplified Router Forwarding Table

<table>
<thead>
<tr>
<th>Network #</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>

Multiple Network Simplified Routing Behavior

What is in the routing tables of routers A, B and C?
Routing in networks with subnets

As seen from the Internet, Harvard is known by its class B address 140.247.y.z

The Harvard network's IP address is 140.247.y.z

Routing within the Harvard network is done using a three byte network number (netmask 255.255.255.0 aka /24)
Subnet Addressing

- Original IP address space not optimized for large numbers of small networks
- Subnetting allows a single network to be split into multiple networks for internal use but appear as a single network to the outside world
- Subnets require subnet masks to separate “networks” from “hosts”

Network Routing with Subnets

- What is the behavior to reach hosts on the local network?
- What is the behavior to reach hosts on another Harvard network?
- What is the behavior to reach hosts on the Internet?
Simplified Router Forwarding Table
What changes do we need to make?

Network #

Routers and Routing (part 2)
**Simplified Router Forwarding Table**

<table>
<thead>
<tr>
<th>Network #</th>
<th>Distance (or cost)</th>
<th>Outgoing Port #</th>
<th>Next Hop IP Address</th>
<th>Etc.,</th>
</tr>
</thead>
</table>

**The Basic Routing Decision**

The basic question is where Router A sends the packets destined for stations located off of Router D.

© 1998-2005 L. Evenchik
Routing Approaches

• Both centralized and distributed approaches are used
• Routing algorithms can be broadly classified as distance-vector or link-state
• Distance vector - good example is RIP
  – Cost function is simply based on the number of hops
  – No subnet masks
• Link state - good example is OSPF
  – “Complex” cost function for links
  – Routers flood their link state within AS
  – Packets include IP address mask information

Routing Information Protocol (RIP)

• RIP is a most widely used IGP, based on work originally done at Xerox PARC for XNS
• Berkeley UNIX version is known as “routed”
• Implements distance vector routing
• The complete routing database is broadcast from each router every 30 seconds
• Network ID and hop-count sent for each entry in the database
• Each router computes its own routing table
• Maximum hop count is 15, infinity equals 16
• RIP doesn’t detect loops
• Hop count does not define delay
**RIP Encapsulation**

![Diagram of RIP Encapsulation]

**RIP Packet Format**

<table>
<thead>
<tr>
<th>Field</th>
<th>Length</th>
<th>Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Command</td>
<td>1 bit</td>
<td></td>
</tr>
<tr>
<td>Version</td>
<td>8 bits</td>
<td></td>
</tr>
<tr>
<td>Address Family of Net 1 (IP=2)</td>
<td>16 bits</td>
<td>Must be zero</td>
</tr>
<tr>
<td>IP Address of Net 1</td>
<td>32 bits</td>
<td></td>
</tr>
<tr>
<td>Must be zero</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Must be zero</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Must be zero</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance to Net 1</td>
<td>16 bits</td>
<td></td>
</tr>
<tr>
<td>Address Family of Net 2</td>
<td>32 bits</td>
<td>Must be zero</td>
</tr>
<tr>
<td>IP Address of Net 2</td>
<td>32 bits</td>
<td></td>
</tr>
</tbody>
</table>

...Consistent layout continued for Net 2, Net 3, etc.....
Open Shortest Path First (OSPF)

- Uses “complex” cost functions for link state
- Link state updates sent when required
- Subnet masks are associated with each advertised route
- Uses IP directly (does not use UDP or TCP)
- OSPF supports multiple classes of routers and routes
- Updates are authenticated
- Networks may be partitioned into areas for easier management
Simplified Router Forwarding Table
What has to be added to support OSPF?

<table>
<thead>
<tr>
<th>Network #</th>
<th>Distance (or cost)</th>
<th>Outgoing Port #</th>
<th>Next Hop IP Address</th>
<th>Etc.,</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Routing Protocol Families

- Concept of Autonomous System provides a way to manage internet complexities.
- Exterior gateway protocols advertise routes between autonomous systems
- Interior gateway protocols are used to exchange routing information within an autonomous systems
Routers and Routing in WANs

Network with Private Line
Simplified Router Schematic
1 LAN and 1 WAN Port

Ethernet (Layers 1 and 2)

Private Line Interface (Layers 1 and 2)

Filter/Forward Logic
Routing Table Logic

IP Layer

Frame Relay Network

Frame Relay Cloud

Router Q

Router R

© 1998-2005 L. Evenchik

© 1998-2005 L. Evenchik

(C) 2005 L. Evenchik
Encapsulation Change: Ethernet to Frame Relay

Ethernet (Layers 1 and 2) to Frame Relay Protocol Interface (Layers 1 and 2)

Multiple Site WAN Using Private Lines
4 Port Router Schematic
1 LAN and 3 WAN

Virtual LANs (VLANs)
802.1P & 802.1Q
Very Simple Ethernet Switch Topology

Simple LAN Architecture
Rational for Virtual LANs

- Groups of users within an organization are typically separated into their own IP networks (typically subnets) for network management, performance, security and other policy reasons.
- Users on LANs should be grouped by their community of interest (sales dept., engineering, accounting), not by their location in the building.
- However, users within a single community of interest are rarely located in the same part of a building.
- “Ethernet switches are easy, routers are hard.”
- Given all the above, how can we separate users via switches? What are the benefits to users and network administrators?

© 1998-2005 L. Evenchik

Ethernet Switch and VLAN Topology

© 1998-2005 L. Evenchik
Virtual LANs

- “Switches are easy, routers are hard.” Given this, how can we separate users via switches? What are the benefits?
- Virtual LANs provide separate collision and broadcast domains for groups of users.
- Users are assigned to one or more VLANs automatically or via a management system.
- VLANs can span multiple switches and sites
- How do users on different VLANs talk to each other?

Virtual LANs

- VLANs are LAN segments (in the classic sense of the word) that can span multiple ethernet switches.
- VLANs provide separate collision and broadcast domains for each group of users.
- Users are assigned to one or more VLANs automatically or via a management system.
- Potential advantages of VLANs include:
  - Better isolation between groups of users: however it is incorrect to think that VLANs significantly improve network security.
  - Improved performance: the specific LAN performance requirements of each group can be met more easily.
  - Improved performance: VLANs provide multiple broadcast domains
  - Provides for more sophisticated network administration

© 1998-2005 L. Evenchik
IEEE 802.1Q Frame Format

<table>
<thead>
<tr>
<th></th>
<th>Dst Address</th>
<th>Source Addr</th>
<th>TAG</th>
<th>Length</th>
<th>Data/Payload</th>
<th>CRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>802.1Q</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>0 - 1500</td>
<td>4</td>
</tr>
</tbody>
</table>

VLAN Protocol ID (8100)

Priority (3 bits) | CFI | VLAN Identifier (12 bits)

© 1998-2005 L. Evenchik
Virtual LANs – 802.1p and 802.1q

- Ethernet priority and VLANs are distinct concepts, but they are intertwined by the technology.
- VLANs are identified by a 12 bit VLAN Identifier.
- Frame priority is marked by a 3 bit field, 0 to 7. This is known as Class of Service.
- Switches can and do, write or re-write, the priority field based on:
  - Port on the frame was received
  - MAC address of the sending station
  - Protocol – IP, IPX, etc.
  - IP Precedence field or DSCP
  - Other IP and/or TCP information
  - Combination of the above

Virtual LANs – some questions

- Can VLANs span multiple sites, such as different buildings on a campus?
- What is the best way to assign users to VLANs?
- How do users on different VLANs talk to each other?
Three Minute Wrap-Up

- Please write down the three or four major points that were discussed during the lecture.
- Note whether the material was presented clearly and/or how it should have been done differently.
- Ask any questions that I should address next time.
- Please do a Wrap-Up at the end of lecture and hand it in as you leave or fill out the form on the web site.
- Do not sign your form. (The form on the web site is also anonymous.)
- Thank you!