2.3.3 Routers

2.3.3.1 Motivation

Bridges do not stop broadcast traffic. This can lead to broadcast storms (e.g., more than 100 non-unicast frames/sec) which can be catastrophic. This can bring the network down.

Some sources of broadcast traffic:

- Address resolution (e.g., ARP, RARP, BOOTP)
- RIP (Routing Information Protocol)
- DHCP (Dynamic Host Configuration Protocol)
- IPX (Internet Packet eXchange) generates broadcast traffic to advertise services and routes
- Netware clients rely on broadcast to find services
- Appletalk: Route discovery protocol

To contain/reduce broadcast traffic, we need to reduce the size of the network (i.e., LAN).

Two approaches are used to do this:

- Use routers to subnet the LAN
- Use VLANs (Virtual LANs)

2.3.3.2 Characteristics

- A router separates traffic of different networks. It does not flood packets.
- Routers route packets at the network layer (layer 3)
- Routers route packets based on the contents of a routing table.
- Routing tables contain a mapping of a destination to a port. They can be static or dynamic.
- Routers “learn” their routing table entries by communicating with their routing peers.
- Routing protocols are used to implement routing (RIP, OSPF, BGP, PNNI)

- Routers perform routing decisions on the basis of the Network ID part of the destination IP address.

- The Host ID part of the destination address is used by the destination router to determine the destination station.

2.3.3.3 IP Addressing

2.3.3.3.1 IP Address Structure

IP address = Network ID + Host ID (32 bits)

- **Class A:**

0	Network ID	Host ID
	7 bits	24 bits

- Address range: **1.0.0.1 → 126.255.255.254**
- Max. number of networks: **126**
- Max. number of hosts: **16,777,214**

- **Class B:**

1 0	Network ID	Host ID
	14 bits	16 bits

- Address range: **128.0.0.1 → 191.255.255.254**
- Max. number of networks: **16,384**
- Max. number of hosts: **65,534**

- **Class C:**

1 1 0	Network ID	Host ID
	21 bits	8 bits

- Address range: **192.0.0.1 → 223.255.255.254**
- Max. number of networks: **2,097,152**
- Max. number of hosts: **254**

- **Class D:**

1 1 1 0	Multicast address
	28 bits

- Address range: **224.0.0.0 → 239.255.255.255**

- **Class E:**

1 1 1 1 0	Reserved for future use
	27 bits

- Address range: **240.0.0.0 → 247.255.255.255**

- ***Note:*** The Internet Network Information Center (InterNIC: www.internic.net) assigns IP addresses

Private allocations:

- In **RFC 1597**, several IP addresses have been allocated for private addressing. An organization can use these addresses if they are not registered with the Internet. Systems are available that translate private, unregistered addresses to public, registered addresses.

Class A addresses:	10.x.x.x → 10.x.x.x	⇨ 1 network
Class B addresses:	172.16.x.x → 172.31.x.x	⇨ 16 networks
Class C addresses:	192.168.0.x → 192.168.255.x	⇨ 256 networks

2.3.3.3.2 Address Resolution

Address Resolution Protocol (ARP) and the relationship between IP and MAC addresses:

2.3.3.3.3 Subnetting

Subnet Address Structure:

Example of Class B network:

Network ID	Subnet ID	Host ID

Subnet mask: 11111111 11111111 11111111 00000000

1s: identify the network address portion of the IP address.
0s: identify the host address portion of the IP address.

IP routing algorithms are modified to support subnet masks (subnet addressing)

- One problem is how to store, maintain and access many network addresses in one routing table. → The Internet establishes a scheme whereby multiple networks are identified by one address entry in the routing table.

Address aggregation:

Address aggregation is used to reduce the size of the routing tables.

How is subnet mask interpreted?

IP address(Class B)	**128.**	**1.**	**17.**	**1**
Mask	255.	255.	240.	0
IP address (binary)	10000000	00000001	00010001	00000001
Mask (binary)	11111111	11111111	11110000	00000000
Result (Logical AND)	10000000	00000001	00010000	00000000
Logical address	**128.**	**1.**	**16.**	**0**

This subnet address is **128.1.16.0/20** (with 16 bits Network ID, 4 bits Subnet ID, and 12 bits Host ID).

2.3.3.3.4 CIDR - Classless InterDomain Routing ("Supernetting")

- **RFCs: 1518, 1519, 1466, 1447**. (http://www.rfc-editor.org/)

It permits networks to be grouped together logically, and to use one entry in a routing table for multiple class C networks.

2.3.3.4 Key Routing Strategies

2.3.3.4.1 Fixed Routing

A single, permanent route is configured for each source-destination pair of nodes in the network (A least-cost routing algorithm could be used to configure routes). Link costs are based on static variables such as expected traffic or capacity.

Problem:

2.3.3.4.2 Flooding

A packet is sent by a source node to every one of its neighbors and each node retransmits it again to its neighbors (similar to "all-routes broadcast" in source routing bridges). The flooding technique has three properties:

- All possible routes are tried, and there is always a backup route (good for emergency messages)
- One copy of the packet will reach destination by following a minimum-hop route (can be use to setup virtual circuits)
- All nodes are visisted (disseminate information to all nodes)

Problem:

2.3.3.4.3 Random Routing

A node selects only one outgoing path chosen at random for retransmission of an incoming packet.

Problem:

2.3.3.4.4 Adaptive Routing

Routing decisions that are made are updated as conditions on the network change (e.g., failure, congestion). Information about the state of the network must be exchanged.

Problems:

- More complex routing decision.
- Information exchanged is itself a load
- Reaction to changes can be too quick or too slow.

However:

- Adaptive routing can improve performance from the user perspective.
- Adaptive routing can aid in congestion control, because it tends to balance load.

2.3.3.5 Definitions

- Autonomous System (AS):
  - Consists of a group of routers exchanging info via a common routing protocol.
  - A set of routers and networks managed by a single organization.
  - Is connected (i.e., a path exists between any 2 nodes) except in time of failure.

- Interior Router Protocol (IRP, IGP)
  - Passes routing information between routers within an AS (e.g., RIP, OSPF).

- Exterior Router Protocol (ERP/EGP)
  - Passes routing information between routers in different Ass (e.g., BGP)

2.3.3.6 Routing Protocols

2.3.3.6.1 RIP (Routing Information Protocol)

- RFC 1058

RIP is:

- o An IRP
- o A distance-vector protocol
- o A widely used protocol because of its simplicity and ease of use
- o Based on the number of intermediate hops to destination
- o Based on Bellman-Ford algorithm
- o A distributed adaptive algorithm
- o Maximum number of hops between a source and destination is 15
- o Routing information is sent every 30 seconds to all adjacent routers using broadcast frames.

A distance of **1** means a directly connected network, and a distance of **16** means unreachable network.

Some major problems with RIP are:

- o "Count to infinity" and there are several partial solutions to this problem such as "Split Horizon"
- o Update of changes in the network is very slow.

2.3.3.6.2 OSPF (Open Shortest Path First)

- RFC 2328

OSPF:

- o Is an IRP
- o Is a link-state routing protocol
- o Is based on Dijkstra's algorithm
- o Is a distributive adaptive algorithm
- o Routers send link state packets (LSPs) that include information about the cost of each of its links/interfaces
- o Relies on two mechanisms:
  - Reliable flooding: the newest information must be flooded to all nodes as quickly as possible, while old information must be removed from the network.
  - Route Calculation: Each node gets a copy of the LSP from all nodes and computes a complete map for the network topology. Then, it decides the best route to each destination.
- o Uses flexible routing metrics: distance, delay, cost, etc.
- o Allows for scalability

- Uses multiple paths to allow for load balancing
- Supports security measures

2.3.3.6.3 BGP (Border Gateway Protocol)

- RFC 1771 (BGP-4)

- BGP:
  - Is a replacement for EGP (Exterior Gateway Protocol). EGP had limitations that include forcing a tree-like topology onto the network.
  - Provides inter-domain routing.
  - Is more concerned with reachability than optimality.

- Challenges:
  - Lot of routing information to pass (~50,000 prefixes)
  - Autonomous nature of the domains (different than IRPs). Cost metrics are not the same and don't have the same meaning across ASs.
  - Trust between different providers (e.g., wrong configuration in an AS, competitors, etc.)

- BGP operates with networks with looped topologies.

- It runs on a reliable transport layer protocol (e.g., TCP).

- Each AS is identified by an AS number.

- BGP considers the Internet as a graph of ASs.

- How BGP works:
  - The administrator of each AS picks at least one node to be a "BGP speaker"

  - "BGP speakers" exchange reachability information among ASs.

  - BGP advertises complete paths as an enumerated list of ASs to reach a particular network.

  - Each AS has one or more border gateways.

- BGP prevents the establishment of looping paths (because it uses the complete AS path)

- BGP supports CIDR and address aggregation.

- BGP supports negative advertisement (i.e., withdrawn route) to cancel path(s).

- EBGP: operates between ASs.

- IBGP: is used to tunnel a user’s traffic through a transit (pass-through) AS.

- BGP uses policy-based metrics. (RFC 1655: BGP policy-based architecture). Policies include various routing preferences and constraints, such as economic, security, or political considerations. (e.g., preference of internal routes over external routes).

2.3.4 Switches

Switching combines advanced microprocessor technology with the concept of a layer-2 bridge.

Whatever we have said about bridges apply to switches (i.e., a switch is a bridge is a switch).

Sometime the difference between a bridge and a switch is looked at as a marketing distinction rather than a technical one.

A switch has bridge’s functionality:

- Learning (generally dynamic)
- Address table (forwarding table) including timers.
- Flooding when destination is unknown.

It can be said that a switch is a high-speed multi-port bridge. A large switch can have more than 100 interfaces.

2.3.4.1 Types of Switches

- **Port switches:** repeaters

- **Switches:** operate at layer 2. They leverage transparent bridging. Typically one port provides a high speed uplink to the backbone.

- **Layer-3 switches (i.e., multilayer switches):** include properties of layer-2 switches and some layer-3 capabilities (i.e., routing capabilities). They use the philosophy of “Switch (bridge) where you can, route where you must”.

- **Layer-4 switches:** It does not implement layer-4 functionality, but it prioritizes certain classes of application traffic. Applications are identified using TCP port number.

2.3.4.2 Inside a switch

Switching fabric refers to the hardware and software design of the switch. ASICs (Application Specific Integrated Circuits) and DSPs (Digital Signal Processors) are used to implement switching fabrics.

Switch Fabrics:

- **Shared memory:**
  - Buffers and output queues are used.

- **Shared bus:**
  - Uses a common bus as the exchange mechanism of frames between ports.

- **Crosspoint matrix:**

Two methods of switch operation:

- **"Store-and-forward" switches:**
  - Buffer data.
  - Check for CRC (Cyclic Redundancy Check) errors.
  - Filter out frames

  Problem:

- **"Cut-through" switches:**
  - Frame header is read.
  - Data is switched without being buffered.
  - Only works if both the input and output ports operate at the same data rate.

  Problem:

  Comparison:

Two switch architectures:

- **"Blocking":** Waits when a particular data path is busy.

  Problem:

- **"Non-blocking":** Handles moving data from input port to output port without delay, i.e., it switches at line speed.

Parameters in switches:

- **Backplane speed:** Internal capacity of a switch. It must exceed the summation of all ports capacities, otherwise blocking and frame dropping will occur.

- **Memory:** Used for buffering data. If it is not enough, then frames dropping will occur.

Switch features:

- **Filtering:** Switches, in contrast to traditional bridges, can filter traffic (i.e., forward traffic conditionally) by interpreting the frame beyond the SA (Source Address) and DA (Destination Address). E.g., layer-3 switches.

  Filters can be complex and may result in performance degradation.

- **Forwarding table:** If the size of this table is exceeded constantly, entries are deleted prematurely and lots of flooding of frames will happen.

- **Oversubscription:** where aggregate bandwidth at the leaves exceeds that of the trunk.

2.3.4.3 Layer-3 Switches

They carry the image of switching as high-performance, cost-effective, hardware-based internetworking, together with the feature set associated with network-layer protocols.

(See the internetworking product timeline in table 4.1 of “The Switch Book”.)

Operation:

The switch architecture can be optimized for functions that must be performed in real-time, for the majority of packets, known as the **fast path** of the flow.

A layer-3 switch needs to implement only this fast path in hardware, e.g., implement hardware-based routing for IP.

Other protocols can be implemented in software.

Exception conditions can also be implemented in software.

The IP fast path:

- Subnet mask represented using 5 bits: used for high-speed routing table lookup operations.
- Packet parsing and validation.
- Routing table lookup.
- Mapping the destination to a local data link address (ARP mapping)
- Fragmentation
- Update lifetime Control and Checksum

2.3.4.4 Virtual Local Area Networks (VLANs)

- VLANs enable the creation of logical groups of network devices across a network.

- Bandwidth Preservation: The broadcast traffic is contained within each VLAN

- LAN Security: VLANs allow for traffic isolation.

- User Mobility: VLANs allow for more flexibility in the positioning of end stations and servers:

    - They can be placed physically anywhere in the building and still remain in the same logical LAN (i.e., VLAN).

    - They can be placed physically in the same location but move to a new logical LAN.

- VLANs are used to partition a flat bridged network using of these techniques:

    - **MAC Address Grouping:** VLAN membership is determined by the device MAC address.

    - **Port Grouping:** A VLAN is a collection of ports across one or more switches. A device attached to one of these ports is a member of this VLAN.

    - **Protocol Grouping:** A VLAN group is based on protocol type (e.g., IP) or on network address.

2.3.5 Brouters and Gateways

- **Brouters:** another name for layer-3 switches.

- **Gateways:** more complex as they interface between two dissimilar networks (operates above layer-3). They are necessary when two networks do not share the same network layer protocol.

2.4 References

1. "Data and Computer Communications" by William Stallings, 6th Edition, Prentice Hall, 2000

2. "Computer Networks - A Systems Approach" by Peterson and Davie, 2nd Edition.

3. "Local & Metropolitan Area Networks" by William Stallings, 6th Edition, Prentice Hall, 2000

4. “The Switch Book” by Rich Seifert. John Wiley & Sons Inc., 2000.

5. “Computer Networks” by Andrew S. Tannenbaum, 4th Edition, Prentice Hall, 2003

6. “LAN Technologies Explained” by Philip Miller and Michael Cummins. Digital Press, 2000