Chapter 2 Internetworking

Topics covered:
Basic terminology. Principles of internetworking. Types of internetworking devices. Repeaters, hubs, bridges, routers, switches and gateways. Transparent and source-routing bridges. Multilayer switches. VLANs. Routing strategies. Addressing.

2.1 Terminology

Internetworking stands for connectivity and communication between two or more networks.

- Internetwork (internet): a collection of communication networks interconnected by bridges, switches and/or routers.

- Intranet: a corporate internet that provides key Internet applications. It is usually isolated and self-contained within an organization.

- End System (ES): a device attached to one of the networks.

- Intermediate System (IS): a device that connects two or more networks (e.g., switch, router). It is called sometimes an IWU (Internetworking Unit) or a relay.

2.2 Principles of Internetworking

2.2.1 Requirements for Internetworking

The overall requirements for an internetworking facility are:

1. Provide a link between networks. At minimum, a physical and link control connection is needed.
2. Provide for the routing and delivery of data between processes on different networks.
3. Provide an accounting service that keeps track of the use of the various networks and routers and maintains status information.
4. Provide the services just listed without requiring modifications to the networking architecture of constituent networks. This means accommodating the following differences:

    - Different addressing schemes: e.g., naming (DNS), DHCP.

    - Different maximum packet size: e.g., segmentation, ATM cells.

    - Different network access mechanisms: e.g., Ethernet, FDDI, ATM.

    - Different timeouts: longer with multiple networks.

- Different error recovery services: some networks will have it, others won't. Internetwork error recovery should be independent of individual networks.

- Different status reporting: how and whether this information can be shared.

- Different routing techniques: may depend on fault detection and congestion control techniques. Coordination is needed.

- Different user access control: authorization for use of the network.

- Connection-oriented vs. connectionless

Some of the above mentioned issues are dealt with in the IWUs.

It may be desirable for an internetwork service not to depend on the characteristics of individual networks.

By fulfilling these requirements, two important problems in interconnecting networks can be addressed:

- Heterogeneity of types of networks

- Scale of internetwork: routing and addressing issues with large growth.

2.2.2 Motivation for Internetworking

- Sharing of computer resources across a number of communications networks

- The use of multiple networks allows for network isolation when needed. This is critical to network performance as failure is contained within one network. Also, a network can be shielded from intrusion (Security).

- Contain the amount of traffic sent between the networks (e.g., Routing domains)

- Network Management that provides centralized support and troubleshooting capabilities in an internetwork.

2.2.3 Components of an Internetwork

- Campus Network: locally connected users in a building or group of buildings. It generally uses LAN technologies.

- Wide Area Networks (WANs): distant campuses connected together usually through connection providers such as a telephone company.

- Remote connections: linking branch offices and mobile users to a corporate campus. They are generally dial-up links or low bandwidth dedicated WAN links.

2.2.4 Routing domains

A routing domain is an administrative entity. Its goal is to establish boundaries for the dissemination of routing information.

- It is also useful for security administration.
- Provides accounting, billing, and revenue services (i.e., Accounting Management).
- Overcome the "flat network" problem by providing a routing hierarchy.

2.3 Internetworking Devices

Devices that interconnect LANs are known as relays and operate at one layer of the OSI model.

There are 5 common types of relays:

-
-
-
-
-

2.3.1 Repeaters (Hubs)

- Overcomes restrictions caused by single segment usage such as number of users, cable length.
- Amplifies or regenerates weak signals.
- Extends cable length.
- Connects LANs of similar type, but may use different media.
- Provides simple connection between adjacent LANs at the expense of increased network congestion.

2.3.2 Bridges

The bridge was designed for interconnection of LANs that use identical protocols at the MAC layer (i.e., layer 2). However, there are bridges capable of mapping between different MAC protocols (e.g., Ethernet and Token Ring).

A bridge main function is forwarding frames from one network to another. A bridge does the following:

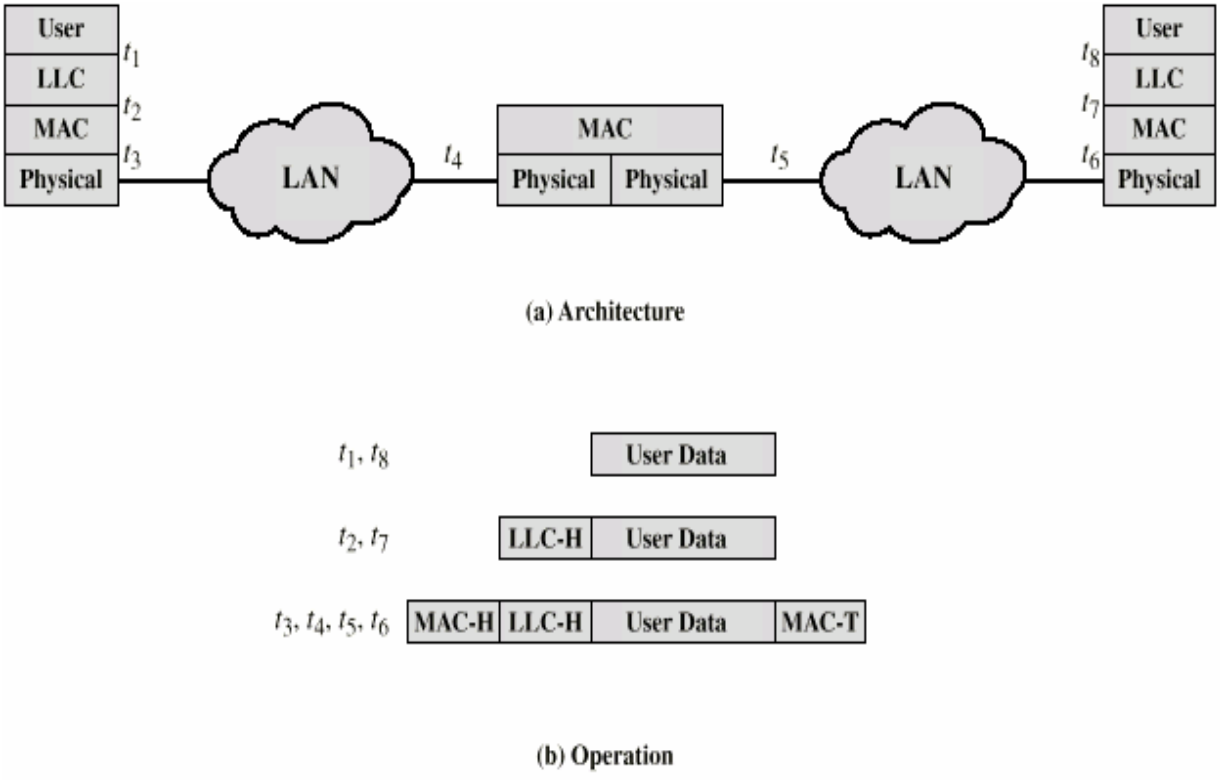


Figure 2-1: Connection of two similar LANs (Stallings)

Characteristics of bridges

- Interconnects two or more LANs (either similar or dissimilar) at the MAC level (e.g., Ethernet and Token Ring)
- Capable of deciding whether or not to forward a frame.
- Creates an extended network and keeps local traffic off.
- Can make minor changes to frame header.
- Does not inspect or modify the network layer packets inside frames.

Reasons for using bridges

- **Reliability**: fault is limited to the network where it happened.
- **Performance**: intra-network traffic stays within one network.
- **Security**: Types of traffic with different security needs are kept on physically separate media.
- **Geography**: LANs may need to be on separate locations.

Bridges have to make a routing decision

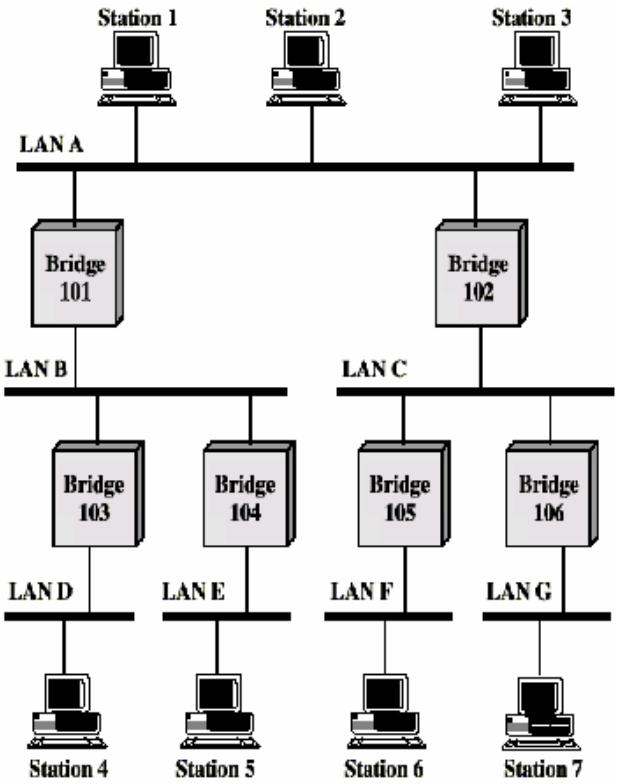


Figure 2-2: Multiple LANs (Stallings)

- S1 transmits a frame on LAN-A intended for S5. B1 and B2 will read the frame. Each one must make a decision of whether or not to retransmit the frame to other LANs. This continues until the frame reaches LAN-E where it is received by S5.

- The routing decision may not always be a simple one. If we add bridge B7 between LAN-A and LAN-E.

- B7 may fail.

-

Many routing strategies are used in bridges:

- Fixed routing
- Spanning tree routing (Transparent bridges)
- Source routing

2.3.2.1 Fixed routing

- A route is selected for each source-destination pair of LANs. If more are available, the one with the least number of hops is selected.
- A central routing matrix is created. It shows the identity of the first bridge on the route.

B1 table

From		From	
Dest	Next	Dest	Next

B2 table

From		From	
Dest	Next	Dest	Next

Advantages:

- Simplicity
- Minimal processing requirements

Disadvantages:

- Bridges can be dynamically added and failures may occur, so tables must change.

2.3.2.2 Spanning Tree Routing (Transparent bridges)

Transparent bridge characteristics:

- It is intended to interconnect LANs that satisfy any of the MAC standards without end stations being aware of its existence (i.e., transparent)

- The routing mechanism is the spanning tree algorithm

The bridge must map the content of the incoming frame into an outbound frame that conforms to the frame format for the outbound LAN, because MAC formats for the various LANs differ.

2.3.2.2.1 Frame Forwarding

A bridge maintains a filtering database. This information can be preloaded into the bridge (i.e., static routing).

2.3.2.2.2 Address Learning

The filtering database can be learned.

2.3.2.2.3 Spanning Tree Algorithm

Address learning is effective with a tree topology (i.e., no closed loop)

- **B1 and B2 know where S2 is:**

  1.

  2.

  3.

  4.

  5.

  6.

  7.

  Problem:

- **B1 and B2 do not know of S2 yet (→ worse problem):**

  1.

  2.

  3.

  4.

  5.

  Problem:

In graph theory: for any connected graph, consisting of nodes and edges connecting pairs of nodes, there is a spanning tree of edges that maintains the connectivity of the graph but contains no closed loops.

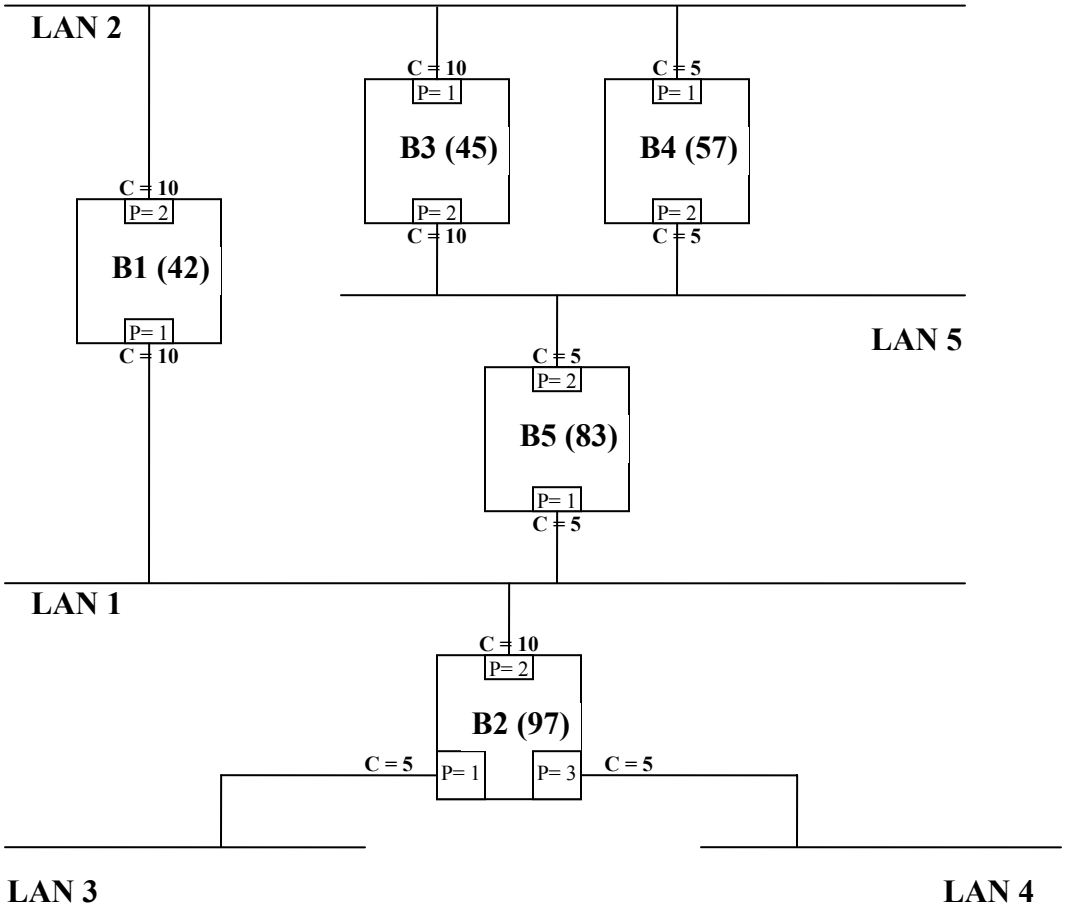
Algorithm:

- Each bridge is assigned a unique id

- A special group MAC address is used to send a frame to all bridges

- Each port of a bridge has a unique "port id"

- Each port of a bridge has an associated cost



The spanning tree is constructed as follows:

1. Determine the root bridge (RB) that is the bridge with the lowest id.

2. Determine the root port (R) on all other bridges. This is the port used for the first hop on the minimum cost path to the root bridge. The lower port number is selected if more than one port exists.

   - The root path cost (rpc) is the cost of the path to the root bridge with minimum cost.

3. Determine the designated port (D) on each LAN. This is the port with the minimum rpc. If more bridges have the same rpc, the one with the highest priority is chosen as designated bridge (i.e., lowest-numbered bridge identifier).

   - The designated bridge is the bridge that provides the minimum cost path to the root bridge.

4. Ports which are neither (R) nor (D) are Blocking (B).

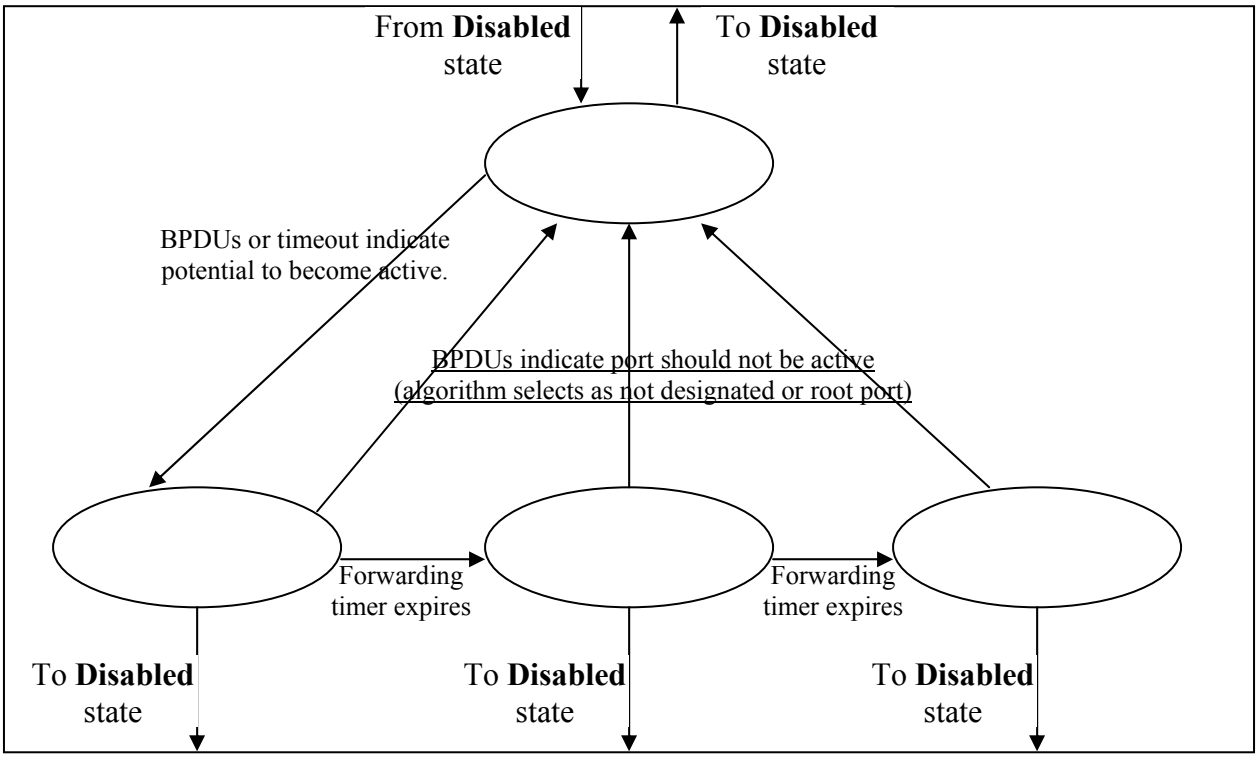
BPDUs (Bridge Protocol Data Units) are used to exchange information between bridges.

- BPDUs are sent by all the bridges each claiming to be the root bridge. B1 is elected as the root bridge.

- All other bridges determine the root port (R) and root path cost (rpc).

- Example: On LAN-5: B3, B4 and B5 send BPDUs claiming to be the designated bridge. B4 and B5 have the lowest RPC. B4 has a higher priority. B4 becomes the designated bridge (D).

2.3.2.2.4 Spanning Tree State Transition Diagram for a bridge port



The following is a table summarizing the actions taken by a bridge in each state.

	Receive BPDUs	**Transmit BPDUs**	**Learn addresses**	**Forward data frames**
Disabled				
Blocking				
Listening				
Learning				
Forwarding				

2.3.2.3 Source Routing Bridges

- Developed by IEEE 802.5 committee

The sending station determines the route to be followed by a frame and includes routing information with this frame.

Each frame includes the type of routing desired:

- Null: no routing desired.

- Nonbroadcast: the frame includes a single route using LANs and bridges.

- All-routes broadcast: the frame will reach each LAN (and the destination station) by all possible routes.

To avoid looping:

- Single-route broadcast: the frame will appear once on each LAN. The frame is forwarded to bridges on the spanning tree with source node as root. The spanning tree is built automatically or manually. The destination receives one copy.

All-routes broadcast and Single-route broadcast types of routing are used to discover route to destination. They are also used for group and all-stations addressing.

Route discovery and selection:

There are three options:

- Manually load information into each station.
    - Problem:

- Stations in the same LAN exchange routing information.
    - Problem:

- Dynamic route discovery procedure by stations

Two approaches are possible:

1.
    - Source station transmits an all-routes request to destination.
    - Destination sends back a nonbroadcast response on each discovered route.
    - Source uses one of these to send subsequent frames.

        - Problems:

2.
    - Source station transmits a single-route request.
    - Destination responds with an all-routes response.
    - Source chooses one for subsequent transmissions.

2.3.2.4 Spanning Tree vs. Source Routing

Characteristics	Transparent bridging	Source route bridging
Transparency		
Topology knowledge		
Frame format		
Frame forwarding		
Bridge mode		
Data Link operation		
Link utilization		
Configuration (LAN numbering, bridge numbering, spanning tree, etc)		
Performance		
Routing		

2.3.2.5 Source Routing Transparent (SRT) Bridges

A key problem is that both (transparent and source routing bridges) are incompatible. To allow the interconnections of LANs using a mixture of transparent and source routing bridges, a new standard was developed by the IEEE 802.5 committee, and that is the Source Routing Transparent (SRT) technique.

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 1918**, 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:

16 bits	8 bits	8 bits
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 ASes (e.g., BGP)

2.3.3.6 Routing Protocols

2.3.3.6.1 RIP (Routing Information Protocol)

- RFC 1058

RIP is:

- An IRP
- A distance-vector protocol
- A widely used protocol because of its simplicity and ease of use
- Based on the number of intermediate hops to destination
- Based on Bellman-Ford algorithm
- A distributed adaptive algorithm
- Maximum number of hops between a source and destination is 15
- 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:

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

2.3.3.6.2 OSPF (Open Shortest Path First)

- RFC 2328

OSPF:

- Is an IRP
- Is a link-state routing protocol
- Is based on Dijkstra's algorithm
- Is a distributive adaptive algorithm
- Routers send link state packets (LSPs) that include information about the cost of each of its links/interfaces
- 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.
- Uses flexible routing metrics: distance, delay, cost, etc.
- 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.
  - Is the routing protocol employed on the Internet.

- Challenges:
  - Lot of routing information to pass (~90,000 prefixes/routes in BGP routing tables.)
  - Autonomous nature of the domains (different than IRPs). Cost metrics are not the same and don't have the same meaning across ASes.
  - 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 ASes.
- How BGP works:
    - The administrator of each AS picks at least one node to be a "BGP speaker"
    - "BGP speakers" exchange reachability information among ASes.
    - BGP advertises complete paths as an enumerated list of ASes 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 ASes.
- 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.

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.

  Problems:

  Comparison:

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.

- Fast path:

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

- Because

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)

- Update lifetime Control and Checksum

- Fragmentation **is not** usually implemented in the fast path.

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, and reduce the effort of adds, moves, and changes:
    - 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.

- Some issues with VLANs:

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