Specific Routing Protocols And MORE

Ben Schultz

UNH InterOperability Lab

June, 2001

Overview, Part 1

- Review of IP and Routing Protocol Types
- IP Fragmentation
- BOOTP and DHCP
- NAT
- Router Architecture
- Gateway Protocols, Interior and Exterior
- Autonomous Systems
- More on Distance Vector Protocols

Review

The IP addressing Scheme contains 4 Byte Source and Destination addresses.

Subnetting Concepts

- Subnets are statically configured on host machines and routers.
- Routers share this information with routing tables

Domain Name System

Routing Protocol Types

Distance Vector

- Passes routes by next hop and path cost or metric.
- Takes up little memory
- Ease of implementation

Link State

- Each router meets all others and passes information about the links that attach it to the network
- Each router contains complete information about the topology and from this information uses Dijkstra's Algorithm to calculate forwarding decisions
- Faster Convergence time
- Uses more memory, complex to implement

IP Fragmentation

MTU = Maximum Transmission Unit

Different physical and data link layer networking technologies have different sized packets. Routers support the capability to fragment and reassemble packets that are too large to be transmitted on some media.

Fragmentation Offset , More Fragments, and Don't Fragment fields help routers perform this task.

BOOTP

The Bootstrap Protocol (BOOTP) was designed for hosts that had minimal computing power. It allows these systems to attach to a network and complete the boot process.

- The client sends a BOOTREQUEST to a server that includes all the boot information that is available.
- The server sends a BOOTREPLY which contains the requested information for the client. A TFTP process is then initiated to transfer the boot files to the client.

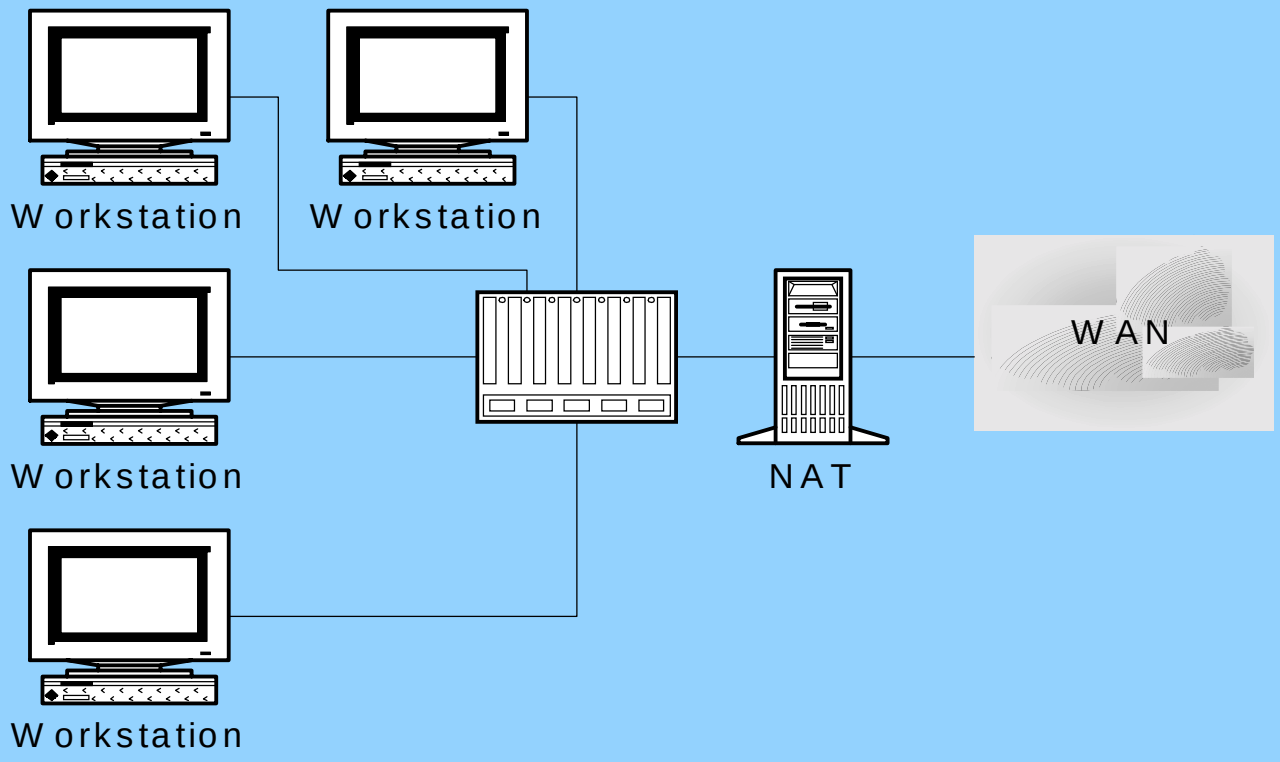
DHCP

Dynamic Host Configuration Protocol uses the same frame format and transport mechanism as BOOTP. It is supposed to provide a complete set of parameters to a host that queries the server. The neat new capability that DHCP adds is that it can assign addresses and reuse them.

- BOOTP : assigns a host an address it can use “forever”
- DHCP : loans an address to a host and is available again if the host does not renew.

NAT

A Network Address Translator sits on a network and translates IP addresses of multiple stations into one address viewable by the outside world.



The Translator

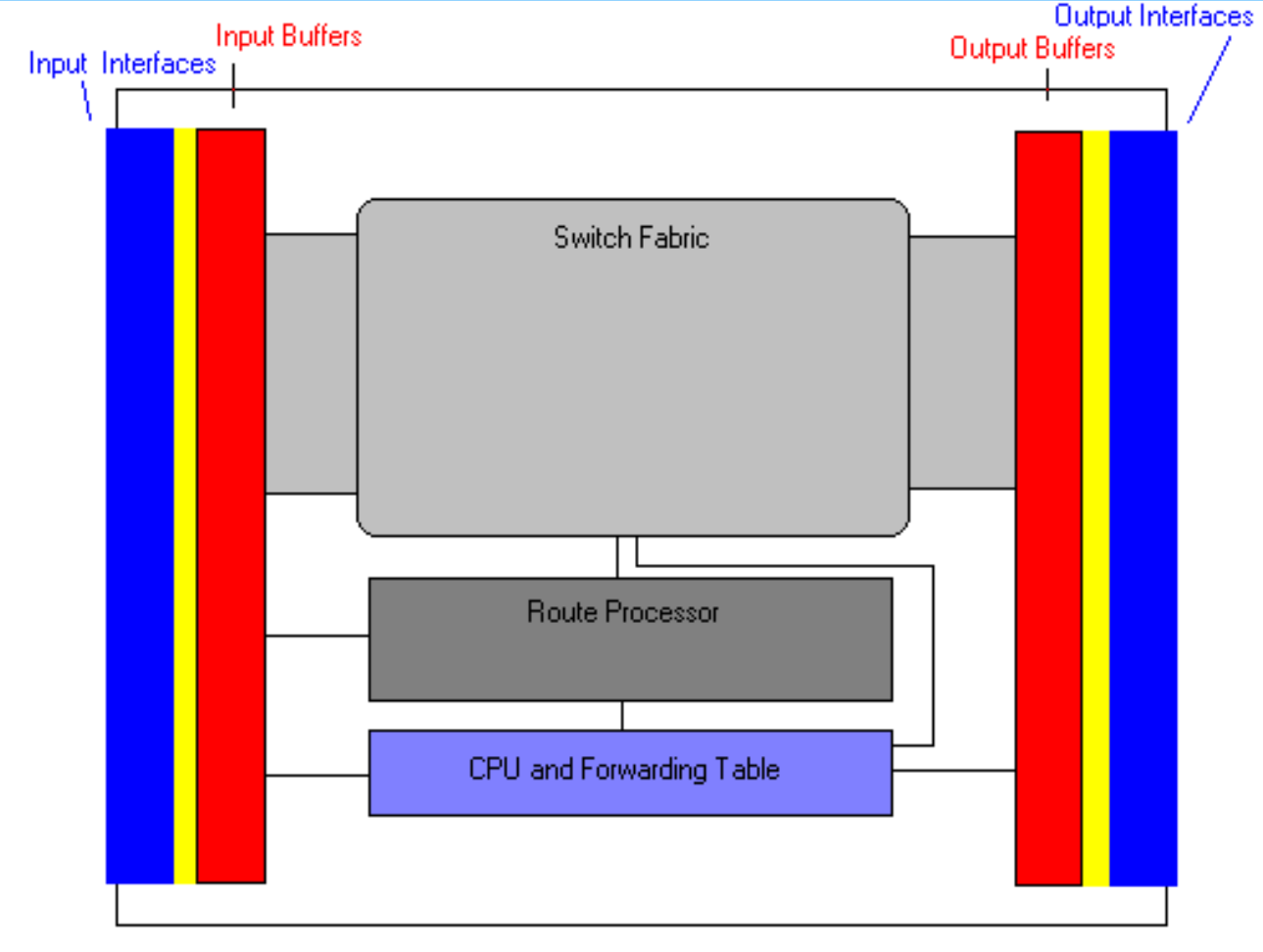
The NAT has a set of one or more globally unique IP addresses that it can assign to nodes in the masked network.

If the NAT has a pool of globally unique IP addresses that is less than the number of nodes in the masked network, it can do Network Address Port Translation (NAPT). This translates between address and port pairs, allowing thousands of connections through the translator.

NAT and NAPT have helped delay the deployment of IPv6. These protocols can get ugly when encryption is used.

Router Architecture

Routers contain several key pieces. A diagram of how a router is structured is below.



Data Plane and Control Plane

The Data Plane does the data packet forwarding and the Quality of Service provisioning and enforcement

The Control Plane runs the routing protocols, processing and advertises the routing information. It also builds the Forwarding database from the routing information obtained from the routing protocols.

Physical Interfaces and Buffers

All routers have Physical Interfaces and Buffers associated with those interfaces.

- The Buffers store packets that are either input too fast to be processed or are queued for transmission.

The CPU and Forwarding Database

The CPU does the address lookups in the Forwarding Database, controlling the forwarding mechanism in the Switching Fabric.

The CPU is also responsible for traffic scheduling and queue management.

The Route Processor

The Route Processor manages the routing protocol information and properly updates the Forwarding Database.

The Route Processor also makes sure that changes in link status or policy are updated and sent to other routers through the routing protocols.

Autonomous Systems

An Autonomous System (or AS) is a set of routers under a single technical administration, like an internet service provider. The administration of an AS appears to other ASs to have a single coherent interior routing plan and presents a consistent picture of what destinations are reachable through its network.

Examples of an AS are Sprint, Qwest, and MCIWorldCom

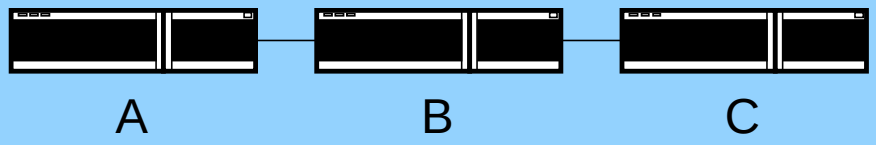
Gateway Protocols

There are 2 types of Gateway Protocols

1. Interior Gateway Protocols are used within Autonomous Systems
2. Exterior Gateway Protocols are used between Autonomous Systems

A Problem with Distance Vector Protocols

Counting to infinity: the LONG convergence time.



Assume each link has a metric of one. A knows that its cost to get to C is 2, B knows that its cost to get to C is 1.

C Crashes!

B discards its distance vector from C and recalculates, using the advertisement of 2 from router A and incrementing it one. Router A receives a metric of 3 to C from router B and changes its distance vector to 4.

This Process continues until routers A and B determine that the metric to C is infinity.

This problem is called counting to infinity.

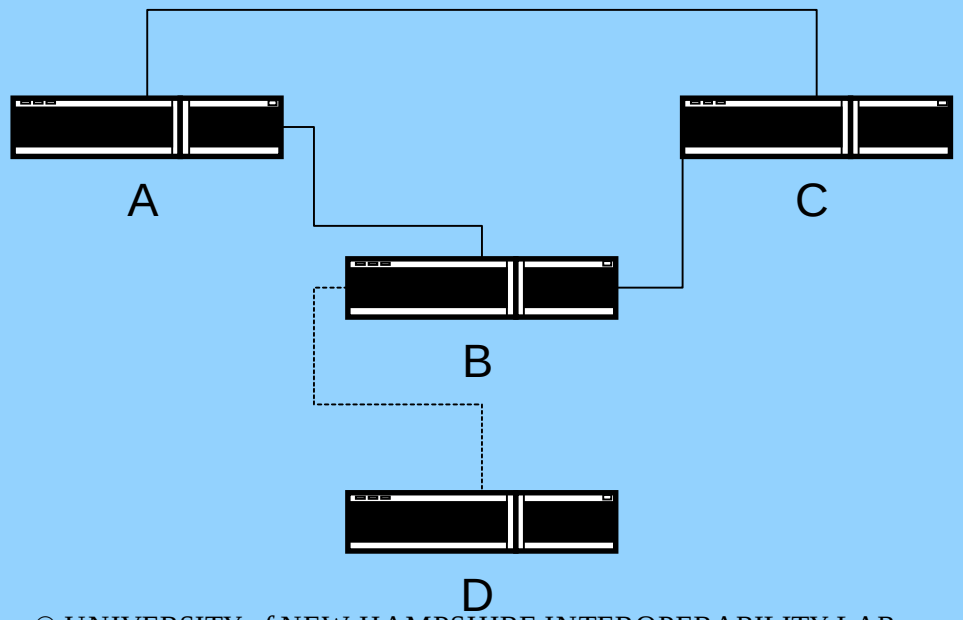
Solutions

- Hold Down: When a route that is in use goes down, the connected router advertises that path as infinity to the rest of the network. Once a period of time passes, the connected router finds an alternative path. The issue here is that it slows down convergence time and doesn't always work.
- Report the entire path: this is expensive and you might as well use a link state algorithm
- Split Horizon: When Router C Crashes, Router B stops advertising a route to C to Router A. Problem solved… somewhat.

Split Horizon

Split Horizon still does not solve the count to infinity problem in the topology below if the link to router D goes down. Router B will stop advertising the route to D, but routers A and C will continue to advertise routes to D, counting to infinity.

Counting to infinity doesn't break the protocol, it just slows network convergence time.



Poison Reversal

Poison Reversal is used with Split Horizon.

Instead of not advertising the route, the router advertises the route with a metric of infinity. This solves the counting to infinity problem.

BREAK

Overview, Part 2

- CIDR
- Routing Information Protocol
- Open Shortest Path First
- Intermediate System to Intermediate System
- Border Gateway Protocol
- VRRP
- Multicast Management and DVMRP

“Classful” Inter Domain Routing

In the early days of the internet, the network masks were inferred from the class of IP address to the destination network. They were not passed around by the routing protocols. In this age network masks were not needed by routers, as they could implicitly determine the network mask by looking at the first byte of the IP address in question.

It was assumed that Class A netmasks were 255.0.0.0, Class B netmasks were 255.255.0.0 and Class C netmasks were 255.255.255.0.

Classless Inter Domain Routing

CIDR routing protocols pass subnet information with their routing information, eliminating the need to infer an address class.

Subnets are thus no longer limited to the classes. They could now be any arbitrary length that the network manager configured (within reason).

More on CIDR

- The concept of address “classes” goes away. A 16 bit address block is the same as an old class B network.
- Modern networks can be arbitrarily grouped (multiple class C sized networks, into a class B sized network), or divided (one class B sized network into any set of arbitrary smaller sized networks).
- A single routing advertisement can cover a block of old style addresses. This makes the size of routing tables smaller.
- Larger blocks of addresses can be divided and allocated, increasing the lifetime of IPv4.

Yesterday's Question

Thus subnet masks are configured both on the host station and the router, and the router shares the subnet masks with the routes.

CIDR makes this possible.

Routing Information Protocol

RIP is an Interior Gateway Protocol that uses the Distance Vector approach to routing. The most primitive version (1) was a class oriented routing protocol. RIP Version 2 adds support for subnet masks and authentication.

RIP works just as I have explained Distance Vector routing protocols.

RIP uses split horizon with poison reversal.

More on RIP

- RIP is designed for smaller simple networks, as the infinity metric is 16 hops. Thus the protocol is limited to networks whose longest path (the network's diameter) is 15 hops with a metric of 1. RIP should not be used in larger networks.
- Like other Distance Vector routing protocols, RIP counts to infinity to resolve unusual situations.
- RIP is not appropriate for situations where routes need to be chosen based on real-time parameters such a measured delay, reliability, or load.

Why Use RIP?

Because of the simplicity of the protocol:

- There are many good, interoperable implementations.
- These implementations have a minimal number of bugs.
- There is minimal configuration.

Open Shortest Path First

OSPF is an Interior Gateway Protocol that uses the Link State approach to routing.

In OSPF, each router maintains a database describing the Autonomous System's topology.

- This database is referred to as the link-state database.
- Each participating router has an identical database.
- Each individual piece of this database is a particular router's local state (e.g., the router's usable interfaces and reachable neighbors).
- The router distributes its local state throughout the Autonomous System by flooding.

Hello Protocol

Routers discover other OSPF capable routers through the Hello Protocol.

Once 2 routers have detected each other, a partial adjacency has been formed. They can now share link state information through the exchange of database description packets.

During and after the Database Exchange Process, each router has a list of those LSAs for which the neighbor has more up-to-date instances. Requests are sent until the database is updated, now the routers are fully adjacent.

What happens when there is more than one router on the subnet?

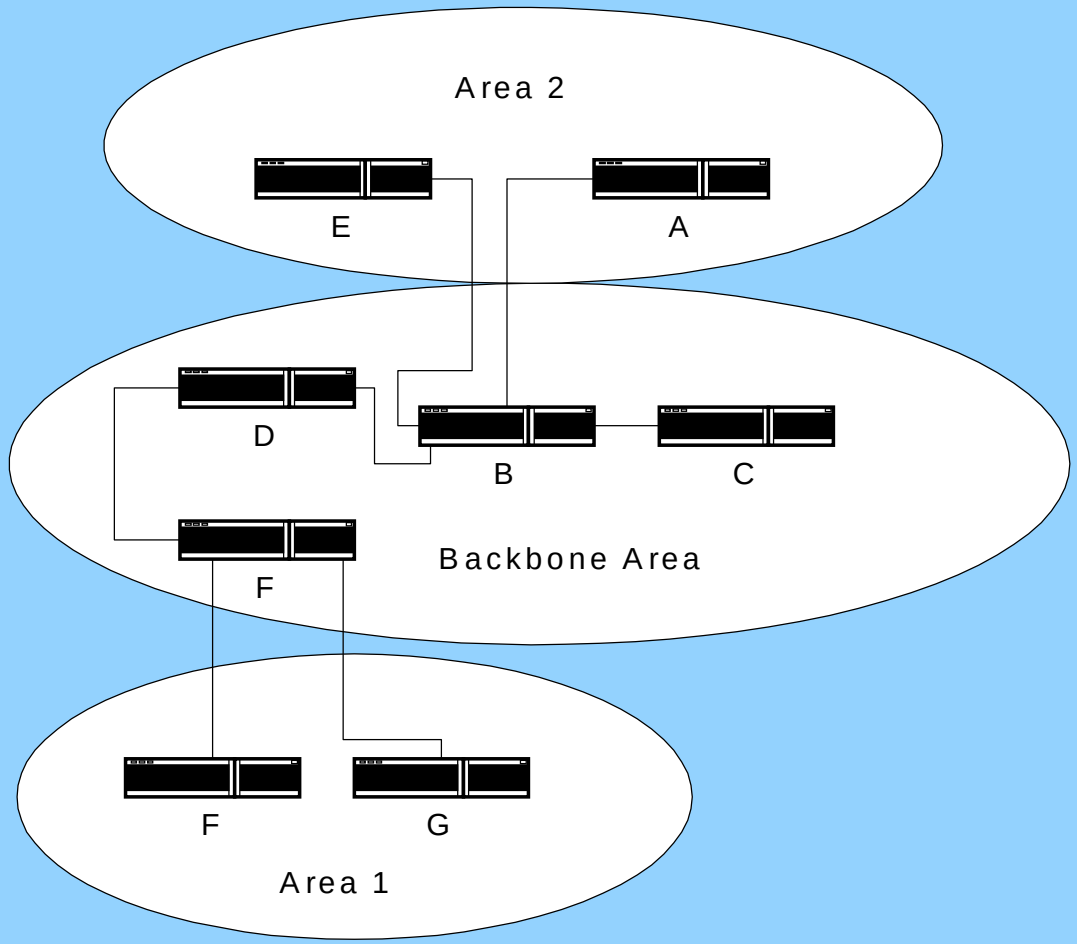
- In this situation, the Hello Protocol has the capability to elect a designated router.
- The router that is first to initialize becomes designated router
- Or if 2 routers initialize at the same time, DR is determined by priority or router ID.
- The designated router is the only router on that specific network that shares the database with other routers on that network.

The Calculation

- As described in the general Link State talk yesterday, OSPF uses Djikstra's Algorithm to construct a tree of shortest path routes across an autonomous system.
- This is performed by all routers on the network in parallel.
- The route costs or metrics are configured by the network administrator.
- The tree that is calculated determines the entire path, but the router only uses this to determine forwarding of data packets to the next hop router.

Hierarchy

OSPF has a 2 level hierarchy. All data traveling between areas must travel through the backbone area.



A Map of the entire topology? Not Really.

Link State Advertisements are flooded throughout the area, each router in an area then creates a SPF tree.

The Area Border Routers transmit summary link state advertisements (LSAs) across area boundaries.

Intra-Area Routing

Routers that have links in 2 areas are known as Area Border Routers. Area Border Routers are connected to the backbone area, and thus have a complete topological map of the backbone, and summaries of other OSPF areas.

From this information, the router calculates paths to all inter-area destinations. The router then advertises these paths into its attached areas. This enables the area's internal routers to pick the best exit router when forwarding traffic inter-area destinations.

Intermediate System to Intermediate System

IS-IS is an Interior Gateway Protocol that uses the Link State approach to routing.

But OSPF does too. Why is there an alternative? IS-IS was defined before OSPF and carriers (like Verizon and AT&T) used implementations in their networks. Instead of switching over – which would be a painful process, they continued to use IS-IS.

IS-IS “Areas”

In IS-IS, the network is partitioned into routing domains (like OSPF areas). The boundaries of routing domains are defined by the network administrator, by setting some links to be "exterior links".

Like OSPF, IS-IS uses a 2 level routing approach.

Unlike OSPF, IS-IS areas can have multiple addresses. This allows graceful address migration and the merging of 2 areas into one.

IS-IS Hierarchy

IS-IS has 2 types of routers:

A level 1 router has the area portion of its address manually configured. It will refuse to become a neighbor with a node whose area addresses do not overlap its area addresses.

A level 2 router will accept another level 2 router as a neighbor, regardless of area address. However, if the area addresses do not overlap, the link would be considered by both routers to be "level 2 only", and only level 2 LSPs would flow on the link. A level 2 router is similar to an OSPF router that is attached to the backbone. Unlike OSPF, if a level one area becomes partitioned, IS-IS has the option of a repair using level 2 routes.

Border Gateway Protocol

BGP is an Exterior Gateway Protocol that uses the Distance Vector approach to routing… Well sort of. Technically it is referred to as a Path Vector protocol. Instead of receiving a distance to a destination, you receive a sequence of AS numbers that compose a path to that destination.

Assumptions

BGP operates on the assumption that Autonomous Systems want to remain independent. Reasons:

- Each AS is independently funded
- When traffic transits an AS, resources are being consumed
- Each AS should have the flexibility to route the traffic they choose
- Each AS has a different charging policy, policies can minimize the cost
- Some ASes do not trust others to have correctly implemented routing protocols.
- There are legal and administrative rules as to how traffic can be routed.

What BGP Does

- Allows paths to be edited before they are passed to neighbors
- Destinations can be configured that are not allowed to be advertised to neighbors
- Routing preferences can be configured that don’t allow specific ingress or egress traffic

BGP overview

BGP is:

- Configuration Intensive
- Likely to function in strange ways, because there are a large number of policies
- Convergence is dependent on policies and the order in which they are configured

Virtual Redundancy Protocol

2 routers are connected in parallel on a subnet. The routers exchange advertisements and through this process one becomes master, and the other becomes backup. A specific priority value determines the outcome of this process.

VRRP in action

The master router continues to send advertisements, while the backup router listens. The master uses a virtual MAC address and IP address that the backup router is also configured for.

If the master fails to transmit an advertisement for approximately 3 * advertisement interval, the backup router assumes the master router role. The new master router will forward data packets sent to the virtual router MAC and IP address and reply to ARP requests to that address.

Multicast Routing

Multicast hosts register with their local router through a protocol called Internet Group Management Protocol.

There are several routing protocols that can route IP multicast packets.

Distance Vector Multicast Routing Protocol creates source routed tree structures.

A Special Thanks To:

- Radia Perlman, Interconnections Second Edition
- Pete Loshin, TCP/IP clearly explained
- Frank Cappellari, Agilent Corporation
- Y. Rekhter, IBM Corp and T. Li, Cisco Systems. RFC 1771, BGP-4
- Thomas A. Maufer, Deploying Multicast In The Enterprise
- G. Malkin, Bay Networks. RFC 2453, RIP Version 2
- J. Moy, Ascend Communications, Inc. RFC 2328, OSPF Version 2
- R. Callon, DEC. RFC 1195, Use of OSI IS-IS for Routing in TCP/IP and Dual Environments