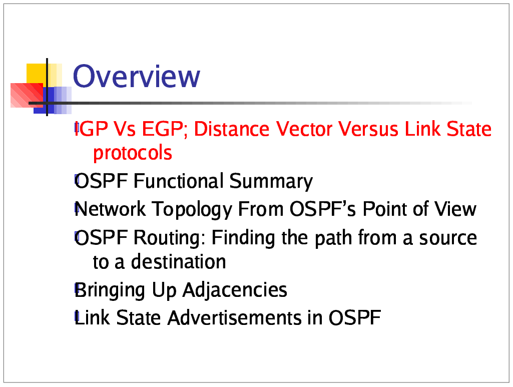
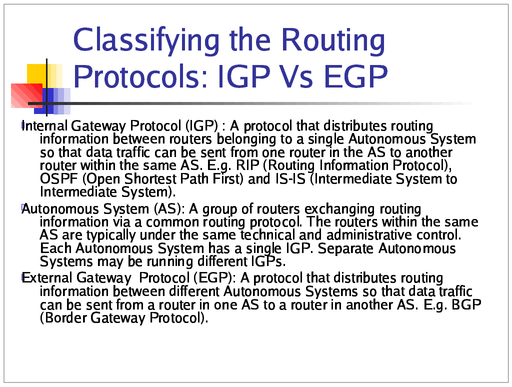
Open Shortest Path First – An Interior Gateway Routing Protocol

Mukul Goyal



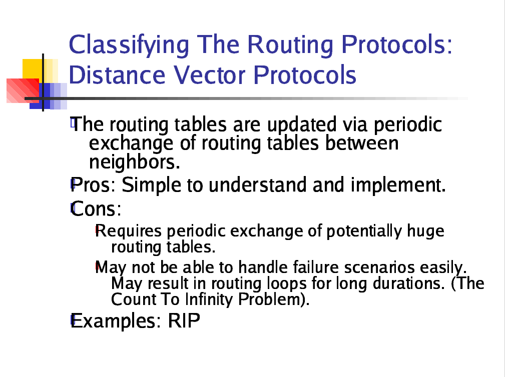
Overview

- IGP Vs EGP; Distance Vector Versus Link State protocols
- OSPF Functional Summary
- Network Topology From OSPF's Point of View
- OSPF Routing: Finding the path from a source to a destination
- Bringing Up Adjacencies
- Link State Advertisements in OSPF



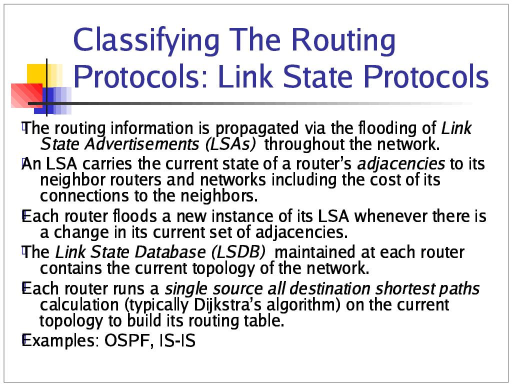
Classifying the Routing Protocols: IGP Vs EGP

- Internal Gateway Protocol (IGP) : A protocol that distributes routing information between routers belonging to a single Autonomous System so that data traffic can be sent from one router in the AS to another router within the same AS. E.g. RIP (Routing Information Protocol), OSPF (Open Shortest Path First) and IS-IS (Intermediate System to Intermediate System).
- Autonomous System (AS): A group of routers exchanging routing information via a common routing protocol. The routers within the same AS are typically under the same technical and administrative control. Each Autonomous System has a single IGP. Separate Autonomous Systems may be running different IGPs.
- External Gateway Protocol (EGP): A protocol that distributes routing information between different Autonomous Systems so that data traffic can be sent from a router in one AS to a router in another AS. E.g. BGP (Border Gateway Protocol).



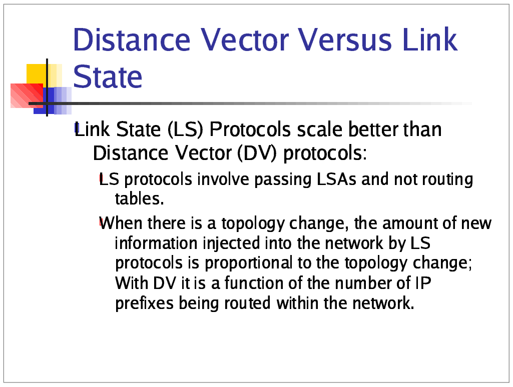
Classifying The Routing Protocols: Distance Vector Protocols

- The routing tables are updated via periodic exchange of routing tables between neighbors.
- Pros: Simple to understand and implement.
- Cons:
  - Requires periodic exchange of potentially huge routing tables.
  - May not be able to handle failure scenarios easily. May result in routing loops for long durations. (The Count To Infinity Problem).
- Examples: RIP



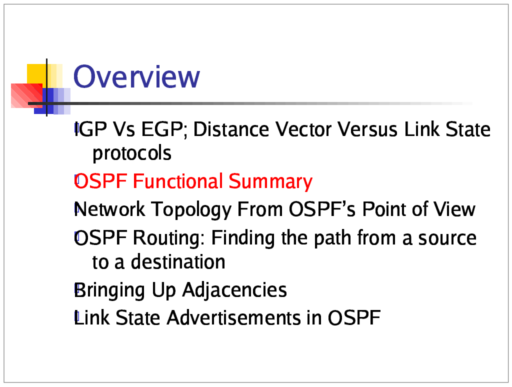
Classifying The Routing Protocols: Link State Protocols

- The routing information is propagated via the flooding of *Link State Advertisements (LSAs)* throughout the network.
- An LSA carries the current state of a router's *adjacencies* to its neighbor routers and networks including the cost of its connections to the neighbors.
- Each router floods a new instance of its LSA whenever there is a change in its current set of adjacencies.
- The *Link State Database (LSDB)* maintained at each router contains the current topology of the network.
- Each router runs a *single source all destination shortest paths* calculation (typically Dijkstra's algorithm) on the current topology to build its routing table.
- Examples: OSPF, IS-IS



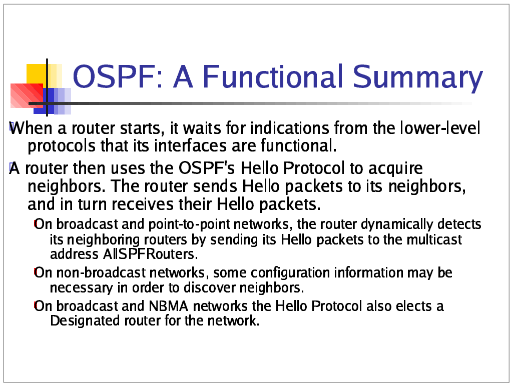
Distance Vector Versus Link State

- Link State (LS) Protocols scale better than Distance Vector (DV) protocols:
  - LS protocols involve passing LSAs and not routing tables.
  - When there is a topology change, the amount of new information injected into the network by LS protocols is proportional to the topology change; With DV it is a function of the number of IP prefixes being routed within the network.



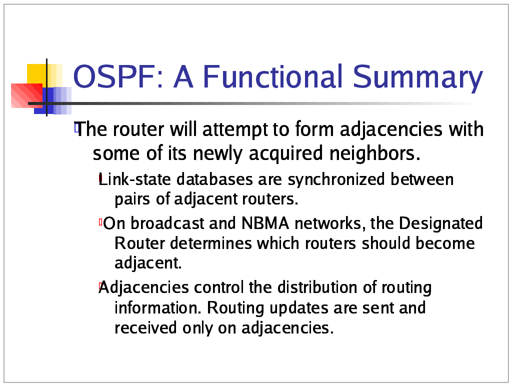
Overview

- IGP Vs EGP; Distance Vector Versus Link State protocols
- OSPF Functional Summary
- Network Topology From OSPF's Point of View
- OSPF Routing: Finding the path from a source to a destination
- Bringing Up Adjacencies
- Link State Advertisements in OSPF



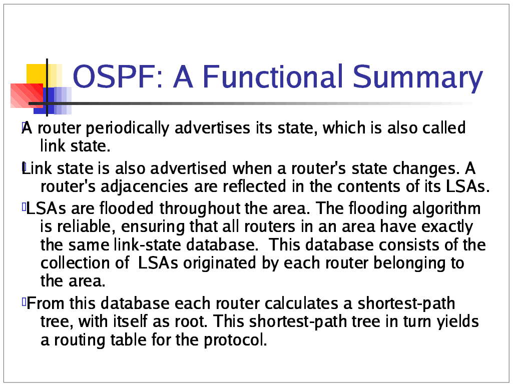
OSPF: A Functional Summary

- When a router starts, it waits for indications from the lower-level protocols that its interfaces are functional.
- A router then uses the OSPF's Hello Protocol to acquire neighbors. The router sends Hello packets to its neighbors, and in turn receives their Hello packets.
  - On broadcast and point-to-point networks, the router dynamically detects its neighboring routers by sending its Hello packets to the multicast address AllSPFRouters.
  - On non-broadcast networks, some configuration information may be necessary in order to discover neighbors.
  - On broadcast and NBMA networks the Hello Protocol also elects a Designated router for the network.



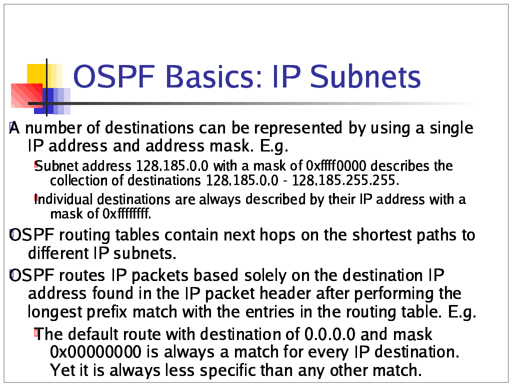
OSPF: A Functional Summary

- The router will attempt to form adjacencies with some of its newly acquired neighbors.
  - Link-state databases are synchronized between pairs of adjacent routers.
  - On broadcast and NBMA networks, the Designated Router determines which routers should become adjacent.
  - Adjacencies control the distribution of routing information. Routing updates are sent and received only on adjacencies.



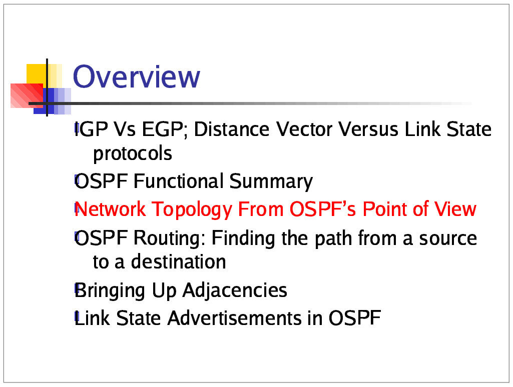
OSPF: A Functional Summary

- A router periodically advertises its state, which is also called link state.
- Link state is also advertised when a router's state changes. A router's adjacencies are reflected in the contents of its LSAs.
- LSAs are flooded throughout the area. The flooding algorithm is reliable, ensuring that all routers in an area have exactly the same link-state database. This database consists of the collection of LSAs originated by each router belonging to the area.
- From this database each router calculates a shortest-path tree, with itself as root. This shortest-path tree in turn yields a routing table for the protocol.



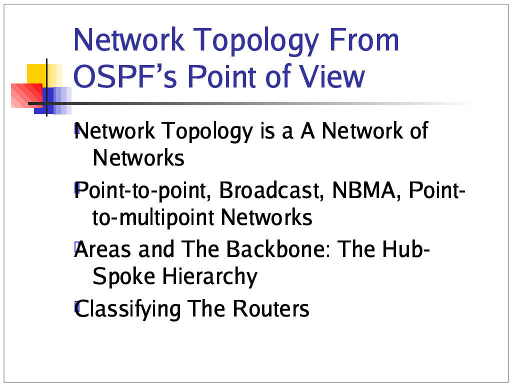
OSPF Basics: IP Subnets

- A number of destinations can be represented by using a single IP address and address mask. E.g.
  - Subnet address 128.185.0.0 with a mask of 0xffff0000 describes the collection of destinations 128.185.0.0 - 128.185.255.255.
  - Individual destinations are always described by their IP address with a mask of 0xffffffff.
- OSPF routing tables contain next hops on the shortest paths to different IP subnets.
- OSPF routes IP packets based solely on the destination IP address found in the IP packet header after performing the longest prefix match with the entries in the routing table. E.g.
  - The default route with destination of 0.0.0.0 and mask 0x00000000 is always a match for every IP destination. Yet it is always less specific than any other match.



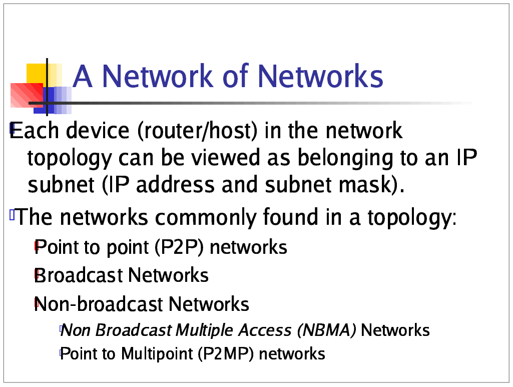
Overview

- IGP Vs EGP; Distance Vector Versus Link State protocols
- OSPF Functional Summary
- Network Topology From OSPF’s Point of View
- OSPF Routing: Finding the path from a source to a destination
- Bringing Up Adjacencies
- Link State Advertisements in OSPF



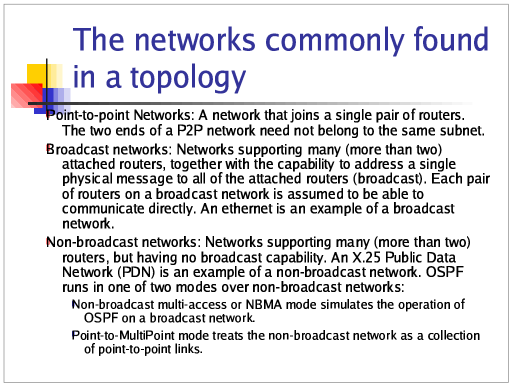
Network Topology From OSPF’s Point of View

- Network Topology is a A Network of Networks
- Point-to-point, Broadcast, NBMA, Point-to-multipoint Networks
- Areas and The Backbone: The Hub-Spoke Hierarchy
- Classifying The Routers



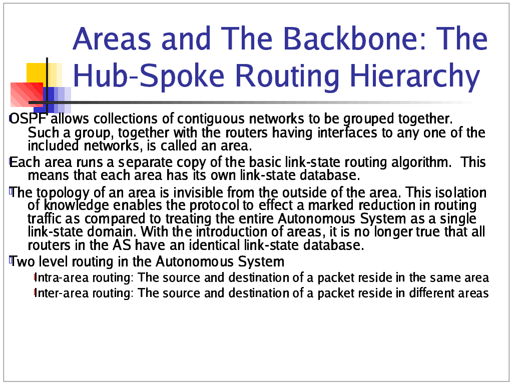
A Network of Networks

- Each device (router/host) in the network topology can be viewed as belonging to an IP subnet (IP address and subnet mask).
- The networks commonly found in a topology:
  - Point to point (P2P) networks
  - Broadcast Networks
  - Non-broadcast Networks
    - *Non Broadcast Multiple Access (NBMA)* Networks
    - Point to Multipoint (P2MP) networks



The networks commonly found in a topology

- Point-to-point Networks: A network that joins a single pair of routers. The two ends of a P2P network need not belong to the same subnet.
- Broadcast networks: Networks supporting many (more than two) attached routers, together with the capability to address a single physical message to all of the attached routers (broadcast). Each pair of routers on a broadcast network is assumed to be able to communicate directly. An ethernet is an example of a broadcast network.
- Non-broadcast networks: Networks supporting many (more than two) routers, but having no broadcast capability. An X.25 Public Data Network (PDN) is an example of a non-broadcast network. OSPF runs in one of two modes over non-broadcast networks:
  - Non-broadcast multi-access or NBMA mode simulates the operation of OSPF on a broadcast network.
  - Point-to-MultiPoint mode treats the non-broadcast network as a collection of point-to-point links.



Areas and The Backbone: The Hub-Spoke Routing Hierarchy

- OSPF allows collections of contiguous networks to be grouped together. Such a group, together with the routers having interfaces to any one of the included networks, is called an area.
- Each area runs a separate copy of the basic link-state routing algorithm. This means that each area has its own link-state database.
- The topology of an area is invisible from the outside of the area. This isolation of knowledge enables the protocol to effect a marked reduction in routing traffic as compared to treating the entire Autonomous System as a single link-state domain. With the introduction of areas, it is no longer true that all routers in the AS have an identical link-state database.
- Two level routing in the Autonomous System
  - Intra-area routing: The source and destination of a packet reside in the same area
  - Inter-area routing: The source and destination of a packet reside in different areas

Areas and The Backbone: The Hub-Spoke Routing Hierarchy

- The OSPF backbone is the special OSPF Area 0 (or Area 0.0.0.0) that connects different non-backbone areas (The "hub and spoke" hierarchy). The backbone is responsible for distributing routing information between non-backbone areas.
- The inter-area communication takes place via the backbone area.
- Each area needs to be adjacent to the Backbone either via a direct link or a "virtual" link.
  - A virtual link is the connection between two backbone routers that have an interface to a common non-backbone area.
  - A Virtual link belongs to the backbone.
  - The routing protocol traffic that flows along the virtual link uses intra-area routing of the area through which the virtual link passes.
  - The cost of the virtual link is the intra-area routing cost between its two ends.

Classification of routers

- Internal routers: A router with all directly connected networks belonging to the same area.
- Area border routers: A router that attaches to multiple areas.
  - Area border routers run multiple copies of the basic algorithm, one copy for each attached area.
  - Area border routers condense the topological information of their attached areas for distribution to the backbone. The backbone in turn distributes the information to the other areas.
- Backbone routers: A router that has an interface to the backbone area. This includes all routers that interface to more than one area (i.e., area border routers). However, backbone routers do not have to be area border routers.
- AS boundary routers: A router that exchanges routing information with routers belonging to other Autonomous Systems. Such a router advertises AS external routing information throughout the Autonomous System. AS boundary routers may be internal or area border routers, and may or may not participate in the backbone.

Overview

- IGP Vs EGP; Distance Vector Versus Link State protocols
- OSPF Functional Summary
- Network Topology From OSPF's Point of View
- OSPF Routing: Finding the path from a source to a destination
- Bringing Up Adjacencies
- Link State Advertisements in OSPF

Routing: Finding The Path From A Source To A Destination

- Intra-area Routing
- Inter-area Routing
- Routing Outside the AS.

Intra-area Routing

- In intra-area routing, the packet is routed solely on information obtained within the area; no routing information obtained from outside the area can be used. This protects intra-area routing from the injection of bad routing information.

Inter-area Routing

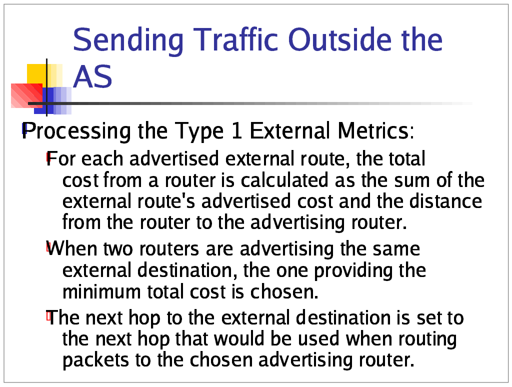
- When routing a packet between two non-backbone areas (say A1 and A2) the backbone is used. *The communication between spokes takes place via the Hub.*
- The path that the packet will travel can be broken up into three contiguous pieces:
  - an intra-area A1 path from the source to an area border router of area A1
  - a backbone path between the area border routers of areas A1 and A2.
  - an intra-area A2 path from A2's area border router to the destination.

Choosing the Area Border Router For Inter-area Routing

- Each area border router in an area summarizes for the area its cost to all networks external to the area.
- After the SPF tree is calculated for the area, routes to all inter-area destinations are calculated by examining the summaries of the area border routers.
- The backbone enables the exchange of summary information between area border routers. Every area border router hears the area summaries from all other area border routers. From this information, the router calculates paths to all inter-area destinations. The router then advertises (via summary LSAs) these paths along with their costs into its attached areas. This enables the area's internal routers to pick the best exit router when forwarding traffic to inter-area destinations.

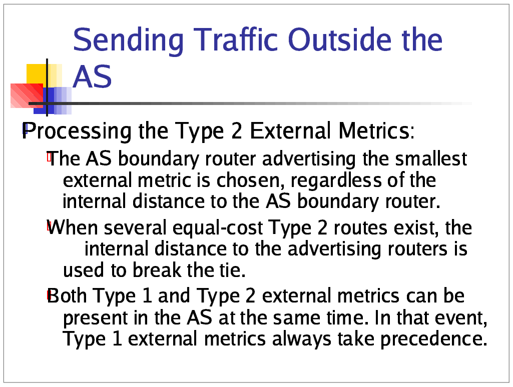
Sending Traffic Outside the AS

- Two kinds of external routing information:
  - The information imported from another routing protocol such as BGP. This information is flooded unaltered throughout the AS as *AS External LSAs.*
  - Static/default routes
- Two types of external metrics:
  - Type 1 external metrics: Have same unit as the ordinary link costs.
  - Type 2 external metrics: These are an order of magnitude larger; any Type 2 metric is considered greater than the cost of any path internal to the AS.



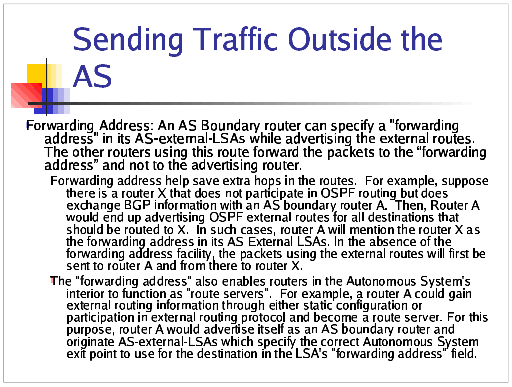
Sending Traffic Outside the AS

- Processing the Type 1 External Metrics:
  - For each advertised external route, the total cost from a router is calculated as the sum of the external route's advertised cost and the distance from the router to the advertising router.
  - When two routers are advertising the same external destination, the one providing the minimum total cost is chosen.
  - The next hop to the external destination is set to the next hop that would be used when routing packets to the chosen advertising router.



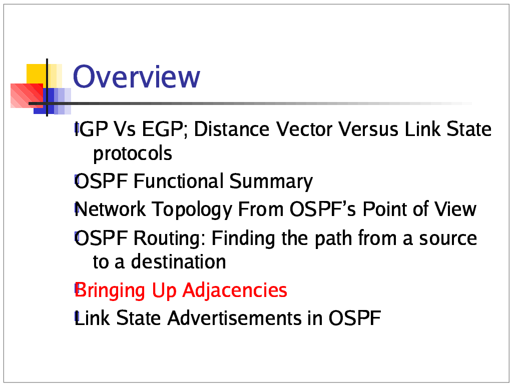
Sending Traffic Outside the AS

- Processing the Type 2 External Metrics:
  - The AS boundary router advertising the smallest external metric is chosen, regardless of the internal distance to the AS boundary router.
  - When several equal-cost Type 2 routes exist, the internal distance to the advertising routers is used to break the tie.
  - Both Type 1 and Type 2 external metrics can be present in the AS at the same time. In that event, Type 1 external metrics always take precedence.



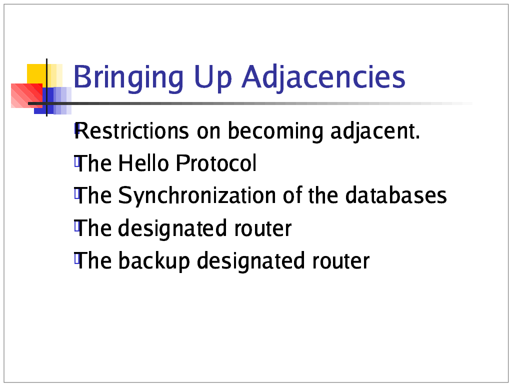
Sending Traffic Outside the AS

- Forwarding Address: An AS Boundary router can specify a "forwarding address" in its AS-external-LSAs while advertising the external routes. The other routers using this route forward the packets to the “forwarding address” and not to the advertising router.
  - Forwarding address help save extra hops in the routes. For example, suppose there is a router X that does not participate in OSPF routing but does exchange BGP information with an AS boundary router A. Then, Router A would end up advertising OSPF external routes for all destinations that should be routed to X. In such cases, router A will mention the router X as the forwarding address in its AS External LSAs. In the absence of the forwarding address facility, the packets using the external routes will first be sent to router A and from there to router X.
  - The "forwarding address" also enables routers in the Autonomous System's interior to function as "route servers". For example, a router A could gain external routing information through either static configuration or participation in external routing protocol and become a route server. For this purpose, router A would advertise itself as an AS boundary router and originate AS-external-LSAs which specify the correct Autonomous System exit point to use for the destination in the LSA's "forwarding address" field.



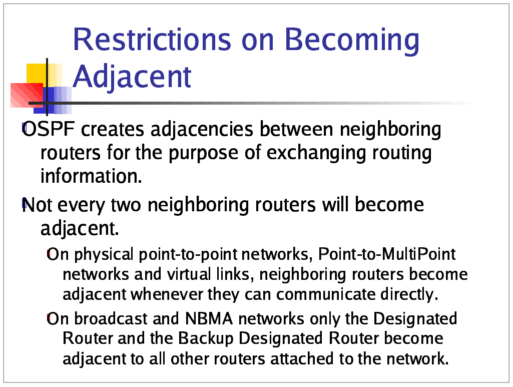
Overview

- IGP Vs EGP; Distance Vector Versus Link State protocols
- OSPF Functional Summary
- Network Topology From OSPF’s Point of View
- OSPF Routing: Finding the path from a source to a destination
- Bringing Up Adjacencies
- Link State Advertisements in OSPF



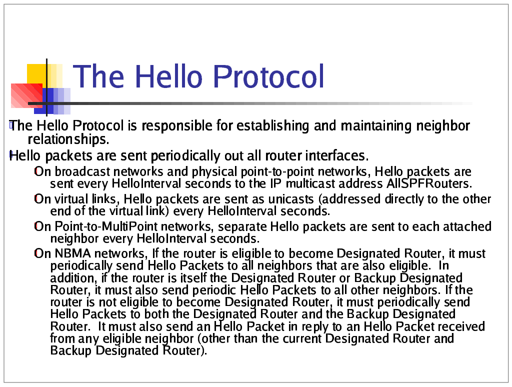
Bringing Up Adjacencies

- Restrictions on becoming adjacent.
- The Hello Protocol
- The Synchronization of the databases
- The designated router
- The backup designated router



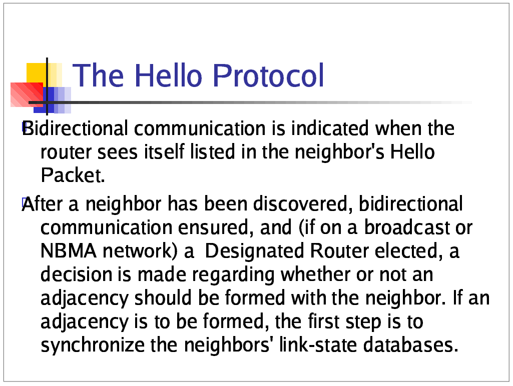
Restrictions on Becoming Adjacent

- OSPF creates adjacencies between neighboring routers for the purpose of exchanging routing information.
- Not every two neighboring routers will become adjacent.
  - On physical point-to-point networks, Point-to-MultiPoint networks and virtual links, neighboring routers become adjacent whenever they can communicate directly.
  - On broadcast and NBMA networks only the Designated Router and the Backup Designated Router become adjacent to all other routers attached to the network.



The Hello Protocol

- The Hello Protocol is responsible for establishing and maintaining neighbor relationships.
- Hello packets are sent periodically out all router interfaces.
  - On broadcast networks and physical point-to-point networks, Hello packets are sent every HelloInterval seconds to the IP multicast address AllSPFRouters.
  - On virtual links, Hello packets are sent as unicasts (addressed directly to the other end of the virtual link) every HelloInterval seconds.
  - On Point-to-MultiPoint networks, separate Hello packets are sent to each attached neighbor every HelloInterval seconds.
  - On NBMA networks, If the router is eligible to become Designated Router, it must periodically send Hello Packets to all neighbors that are also eligible. In addition, if the router is itself the Designated Router or Backup Designated Router, it must also send periodic Hello Packets to all other neighbors. If the router is not eligible to become Designated Router, it must periodically send Hello Packets to both the Designated Router and the Backup Designated Router. It must also send an Hello Packet in reply to an Hello Packet received from any eligible neighbor (other than the current Designated Router and Backup Designated Router).



The Hello Protocol

- Bidirectional communication is indicated when the router sees itself listed in the neighbor's Hello Packet.
- After a neighbor has been discovered, bidirectional communication ensured, and (if on a broadcast or NBMA network) a Designated Router elected, a decision is made regarding whether or not an adjacency should be formed with the neighbor. If an adjacency is to be formed, the first step is to synchronize the neighbors' link-state databases.

The Synchronization Of The Databases

- Two routers can become adjacent only when their link state databases are synchronized.
- Each router describes its database by sending a sequence of Database Description packets to its neighbor. Each Database Description Packet describes a set of LSAs belonging to the router's database.
- When the neighbor sees an LSA that is more recent than its own database copy, it makes a note that this newer LSA should be requested. After the Database Description is over, These LSAs are requested in Link State Request Packets.
- When the Database Description Process has completed and all Link State Requests have been satisfied, the databases are deemed synchronized. At this time the adjacency is fully functional and is advertised in the two routers' router-LSAs.
- The adjacency is used by the flooding procedure as soon as the Database Exchange Process begins. This simplifies database synchronization, and guarantees that it finishes in a predictable period of time.

The Designated Router

- Every broadcast and NBMA network has a Designated Router.
- The Designated Router performs two main functions:
  - The Designated Router originates a network-LSA on behalf of the network. This LSA lists the set of routers (including the Designated Router itself) currently attached to the network.
  - The Designated Router becomes adjacent to all other routers on the network.
- The Designated Router is elected by the Hello Protocol.
  - A router's Hello Packet contains its Router Priority.
  - When a router's interface to a network first becomes functional, it checks to see whether there is currently a Designated Router for the network. If there is, it accepts that Designated Router, regardless of its Router Priority. Otherwise, the router itself becomes Designated Router if it has the highest Router Priority on the network.
- In order to optimize the flooding procedure on broadcast networks, the Designated Router multicasts its Link State Update Packets to the address AllSPFRouters, rather than sending separate packets over each adjacency.

The Backup Designated Router

- In order to make the transition to a new Designated Router smoother, there is a Backup Designated Router for each broadcast and NBMA network.
- The Backup Designated Router is also adjacent to all routers on the network, and becomes Designated Router when the previous Designated Router fails.
- If there were no Backup Designated Router, when a new Designated Router became necessary, new adjacencies would have to be formed between the new Designated Router and all other routers attached to the network which can potentially take quite a long time. During this time, the network would not be available for transit data traffic.
- The Backup Designated Router is also elected by the Hello Protocol.

Overview

- IGP Vs EGP; Distance Vector Versus Link State protocols
- OSPF Functional Summary
- Network Topology From OSPF's Point of View
- OSPF Routing: Finding the path from a source to a destination
- Bringing Up Adjacencies
- Link State Advertisements in OSPF

The Link State Advertisements

- Different LSA Types:
  - Router LSA (Type 1)
  - Network LSA (Type 2)
  - Summary LSA (Type 3, 4)
  - AS External LSA (Type 5, 7)
  - Group Membership LSA (Type 6)
  - External Attributes LSA (Type 8)
  - Opaque LSA (Type 9, 10, 11)
- Aging LSAs

Router LSA (Type 1)

- Originated by all routers.
- This LSA describes the router's working connections (link/interfaces) to an area, mentioning the cost of using the link and the link type
  - Point to point link
  - Link to transit network
  - Link to stub network
  - Virtual link
- Flooded throughout a single area only.

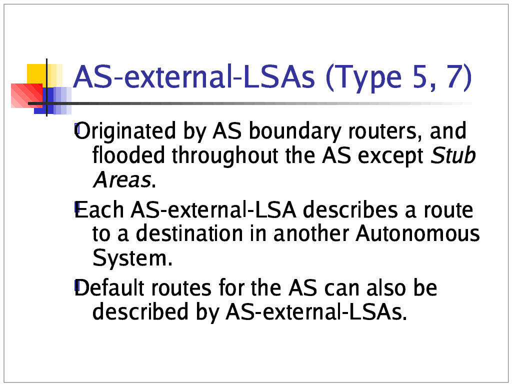
Network LSA (Type 2)

- Originated for *broadcast* and *NBMA* networks by the *Designated Router*.
- This LSA contains the list of routers connected to the network. Flooded throughout a single area only.

Summary-LSAs (Type 3, 4)

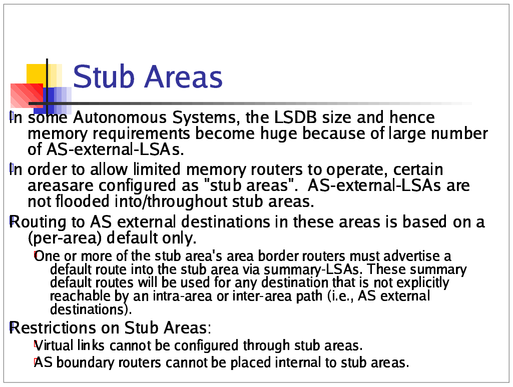
- Originated by area border routers, and flooded through-out the LSA's associated area.
- Each summary-LSA describes a route to a destination outside the area, yet still inside the AS (i.e., an inter-area route).
- In order to get better aggregation, area border routers may employ an area address range (similar to an IP subnet address) to describe a single route to many separate networks. The cost of the route is the maximum cost to any of the networks falling in the specified range.

- Type 3 summary-LSAs describe routes to networks.
- Type 4 summary-LSAs describe routes to AS boundary routers.
  - To utilize external routing information advertised by AS boundary routers, the path to these routers must be known throughout the AS. For that reason, the locations of these AS boundary routers are summarized by the area border routers.



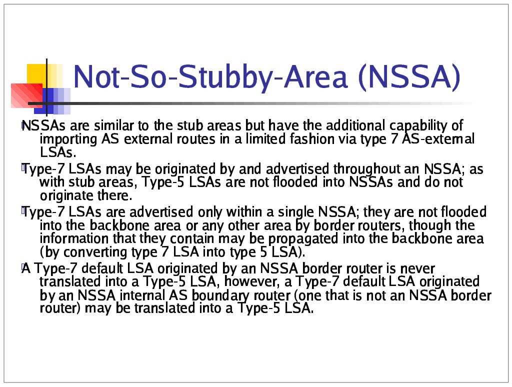
AS-external-LSAs (Type 5, 7)

- Originated by AS boundary routers, and flooded throughout the AS except *Stub Areas*.
- Each AS-external-LSA describes a route to a destination in another Autonomous System.
- Default routes for the AS can also be described by AS-external-LSAs.



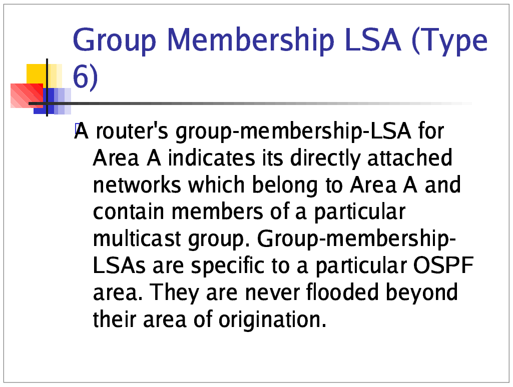
Stub Areas

- In some Autonomous Systems, the LSDB size and hence memory requirements become huge because of large number of AS-external-LSAs.
- In order to allow limited memory routers to operate, certain areasare configured as "stub areas". AS-external-LSAs are not flooded into/throughout stub areas.
- Routing to AS external destinations in these areas is based on a (per-area) default only.
  - One or more of the stub area's area border routers must advertise a default route into the stub area via summary-LSAs. These summary default routes will be used for any destination that is not explicitly reachable by an intra-area or inter-area path (i.e., AS external destinations).
- Restrictions on Stub Areas:
  - Virtual links cannot be configured through stub areas.
  - AS boundary routers cannot be placed internal to stub areas.



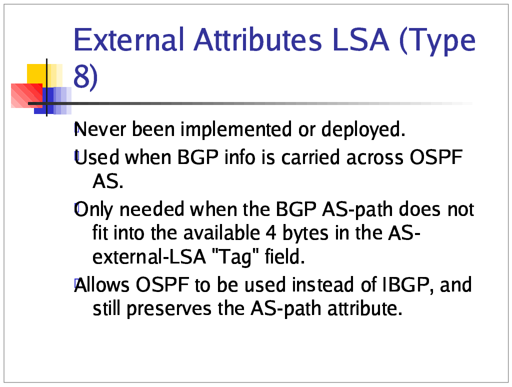
Not-So-Stubby-Area (NSSA)

- NSSAs are similar to the stub areas but have the additional capability of importing AS external routes in a limited fashion via type 7 AS-external LSAs.
- Type-7 LSAs may be originated by and advertised throughout an NSSA; as with stub areas, Type-5 LSAs are not flooded into NSSAs and do not originate there.
- Type-7 LSAs are advertised only within a single NSSA; they are not flooded into the backbone area or any other area by border routers, though the information that they contain may be propagated into the backbone area (by converting type 7 LSA into type 5 LSA).
- A Type-7 default LSA originated by an NSSA border router is never translated into a Type-5 LSA, however, a Type-7 default LSA originated by an NSSA internal AS boundary router (one that is not an NSSA border router) may be translated into a Type-5 LSA.



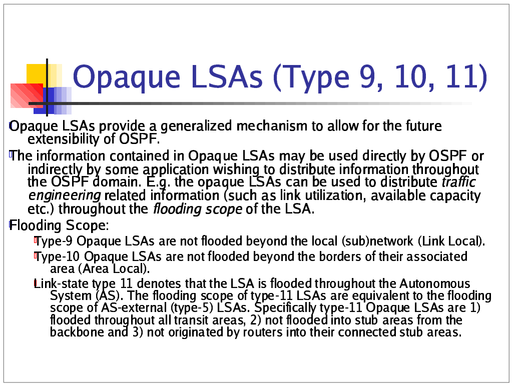
Group Membership LSA (Type 6)

- A router's group-membership-LSA for Area A indicates its directly attached networks which belong to Area A and contain members of a particular multicast group. Group-membership-LSAs are specific to a particular OSPF area. They are never flooded beyond their area of origination.



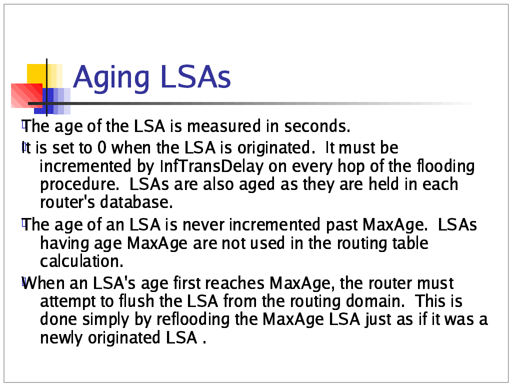
External Attributes LSA (Type 8)

- Never been implemented or deployed.
- Used when BGP info is carried across OSPF AS.
- Only needed when the BGP AS-path does not fit into the available 4 bytes in the AS-external-LSA "Tag" field.
- Allows OSPF to be used instead of IBGP, and still preserves the AS-path attribute.



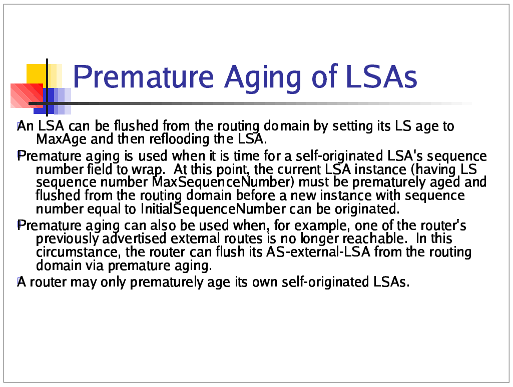
Opaque LSAs (Type 9, 10, 11)

- Opaque LSAs provide a generalized mechanism to allow for the future extensibility of OSPF.
- The information contained in Opaque LSAs may be used directly by OSPF or indirectly by some application wishing to distribute information throughout the OSPF domain. E.g. the opaque LSAs can be used to distribute *traffic engineering* related information (such as link utilization, available capacity etc.) throughout the *flooding scope* of the LSA.
- Flooding Scope:
  - Type-9 Opaque LSAs are not flooded beyond the local (sub)network (Link Local).
  - Type-10 Opaque LSAs are not flooded beyond the borders of their associated area (Area Local).
  - Link-state type 11 denotes that the LSA is flooded throughout the Autonomous System (AS). The flooding scope of type-11 LSAs are equivalent to the flooding scope of AS-external (type-5) LSAs. Specifically type-11 Opaque LSAs are 1) flooded throughout all transit areas, 2) not flooded into stub areas from the backbone and 3) not originated by routers into their connected stub areas.



Aging LSAs

- The age of the LSA is measured in seconds.
- It is set to 0 when the LSA is originated. It must be incremented by InfTransDelay on every hop of the flooding procedure. LSAs are also aged as they are held in each router's database.
- The age of an LSA is never incremented past MaxAge. LSAs having age MaxAge are not used in the routing table calculation.
- When an LSA's age first reaches MaxAge, the router must attempt to flush the LSA from the routing domain. This is done simply by reflooding the MaxAge LSA just as if it was a newly originated LSA .



Premature Aging of LSAs

- An LSA can be flushed from the routing domain by setting its LS age to MaxAge and then reflooding the LSA.
- Premature aging is used when it is time for a self-originated LSA's sequence number field to wrap. At this point, the current LSA instance (having LS sequence number MaxSequenceNumber) must be prematurely aged and flushed from the routing domain before a new instance with sequence number equal to InitialSequenceNumber can be originated.
- Premature aging can also be used when, for example, one of the router's previously advertised external routes is no longer reachable. In this circumstance, the router can flush its AS-external-LSA from the routing domain via premature aging.
- A router may only prematurely age its own self-originated LSAs.