

An OSPF Primer, edited from the Web by J. Scott, Feb 2007

Introduction to OSPF

Open Shortest Path First (OSPF) routing protocol is a Link State protocol based whose default metric is based on bandwidth and not hops. OSPF can support different sized subnet masks on different interfaces of the same router, which allows a more efficient utilization of available IP address space. Also, OSPF supports unnumbered point to point links and equal cost load balancing for up to 6 different paths.

Link State Advertisements

Each OSPF router is required to be in an OSPF area, and the backbone must be area 0. OSPF routers exchange only link state advertisements (LSA's) to other routers in their area, and not complete network information, as RIP does. The Link State Database (LSDB) contains the LSA's sent around the area the OSPF router is in. Each router in that area will contain an identical copy of this LSDB.

The LSA's are normally sent at time intervals called "hello intervals," but are also triggered as a result of a link change, such as when a link goes down or comes up. Using the LSA's and the LSDB, an OSPF router creates a Shortest Path First (SPF) tree using Dijkstra's algorithm. The routing table entry

for each network the router will have routes to will be derived from the SPF tree. Dijkstra's algorithm provides a provably optimal procedure for calculating the SPF tree from the LSA's. Building the SPF tree is CPU intensive, which is a drawback to using OSPF.

OSPF Networks

Within OSPF there can be Point-to-Point networks or Multi-Access networks. The Multi-Access networks could be one of the following:

- Broadcast Networks: A single message can be sent to all routers
- Non-Broadcast Multi-Access (NBMA) Networks: Has no broadcast ability, ISDN, ATM, Frame Relay and X.25 are examples of NBMA networks.
- Point to Multipoint Networks: Used in group mode Frame Relay networks.

Forming Adjacencies

Each router within an area maintains an identical LSDB by maintaining communications with other routers. The formation of an adjacency occurs between two routers A and B that are in the initial Down state as follows:

1. Init state: Hello packets are exchanged between routers A and B, in order to form a Neighbor Relationship. Then based on these packets they decide whether or not to become

adjacent. The Hello packet contains the router ID and the hello and dead intervals and is sent to the multicast address 224.0.0.5. In multi-access networks the hellos are sent every 10 seconds. The Dead Interval is normally 4 times the Hello interval and is the time waited before the router declares the neighbor to be down. The Hello packet also contains the 32-bit router ID, which is usually setup as a loopback address. Bi-directional communication is confirmed when the routers see each other in each other's hello packet. The Router Priority and the Designated Router (DR)/Backup Designated Router (BDR) addresses are also included and the routers have to agree on the Stub Area Flag and the Authentication Password.

2. Two-way state: The routers add each other to their Adjacencies database and they become neighbors.

3. DR and BDR Election:

Initially, on forming an adjacency, the router with the highest Router Priority, (normally the highest loopback address) becomes the DR. The router with the next highest ID becomes the BDR. The BDR just receives the same information as the DR but only performs the task of a DR when the DR fails. The BDR still maintains adjacencies with all routers. In a hub and spoke environment it is necessary to set all the spoke router priorities to '0' so that they never can become the DR or BDR and therefore become isolated from the other routers.

If a router with a higher priority is added to the network later on it does NOT take over the DR and no re-election takes place. It is possible for a router to be a DR in one network and a normal router in another at the same time.

4. After election the routers are in the Exstart state as the DR and BDR create an adjacency with each other and the router with the highest priority acts as the master and they begin creating their link-state databases using Database Description Packets.

5. The process of discovering routes by exchanging Database Description Packets (DBDs) is known as Exchange. These packets contain details such as the link-state type, the address of the advertising router, the cost of the link and the sequence number that identifies how recent the link information is. Unicasts are used to compare LSDBs to see which Link State Advertisements (LSAs) are missing or out of date.

6. Link State ACK: Once a DBD has been received a Link State ACK is sent containing the link-state entry sequence number. The slave router compares the information and if it is newer it sends a request to update.

7. Link State Request: In order to update its LSDB the slave router sends a Link State Request. This is known as the Loading state.

8. Link State Update: A Link State Update is sent in response to a Link State Request and it contains the requested LSAs.

9. Link State ACK: Once a Link State Update has been received a Link State ACK is sent again and the adjacency has been formed. At this point the databases are considered to be synchronous.

10. Full: In the Full state the routers can route traffic and the routers continue sending each other hello packets in order to maintain the adjacency and the routing information.

Maintaining the Routing Tables

Point-to-Point and Point-to-Multipoint links do not require a Designated Router (DR) or a Backup Designated Router (BDR) because adjacencies have to form with each other anyway. On a Point-to-Point and Point-to-Multipoint networks adjacencies are always formed between the two routers so there is no requirement for a DR or BDR, whilst on a multi-access network a router will form an adjacency with the Designated Router (DR) and the Backup Designated Router (BDR). In a broadcast or NBMA network it is not feasible for every router to form a full mesh of adjacencies with all the other routers. The Designated Router forms adjacencies with each of the other routers and performs the link-state information exchange thereby minimizing the traffic load and making sure that the information is consistent across the network.

On detection of a link state change, the OSPF router sends a Link State Update (LSU) to the multicast address 224.0.0.6 which will be processed by the DR and BDR routers for that area. The LSU contains several LSAs. After acknowledging the LSU, the DR Floods link-state information to *all* the OSPF routers on the OSPF multicast address 224.0.0.5. Each LSA is acknowledged separately with an LSAck. If the LSA is new, it is added to the Link State Database, otherwise the LSA is ignored.

Using the DR significantly reduces the amount of traffic that OSPF must put on the network. Here's how this works. For example, if there are 20 routers, and a link fails, an LSU is multcasted to the address 224.0.0.6. Only the DR and the BDR are listening on this address. Thus one packet goes out, and after the DR processes this information, it sends out the link state change to all the OSPF routers in the area by sending it to the multicast address 224.0.0.5. Thus, one LSU went to 224.0.0.6 and then one flood went to all routers which are listening on 224.0.0.5. DRs in other networks that are connected also receive the LSUs. On receipt of the new LSA the routers recalculate their routing tables.

The LSA has a 30 minute timer that causes the router to send an LSU to everyone on the network once it ages out. This verifies that the link is still valid. If a router receives an LSA with old information then it will send a LSU to the sender to update the sender with the newer information.

Important OSPF Parameters

The Retransmit Interval is the number of seconds between LSAs across an adjacency. The following settings are often recommended:

Broadcast network	5 seconds
Point-to-Point network	10 seconds
NBMA network	10 seconds
Point-to Multipoint network	10 seconds

The Hello Interval must be the same on each end of the adjacency otherwise the adjacency will not form. In a Point-to-Point network this value is 10 seconds whereas in a Non Broadcast Multi-access Network (NBMA) the Hello Interval is 30 seconds.

The Dead Interval is 40 seconds in a Point-to-Point network and 120 seconds in a Non Broadcast Multi-access Network (NBMA).

The Metric Cost can be related to line speed by using the formula $10^8 / \text{line speed (bps)}$

The following table gives some guidelines for costs:

Network Type	Cost
FDDI/Fast Ethernet	1
Token Ring (16Mbps)	6
Ethernet	10
E1	48
T1	64
64 kb/s	1562
56 kb/s	1785

These costs are used to calculate the metric for a link and thus determine the best route for traffic. The lowest cost to a destination is calculated using Dijkstra's Algorithm. The lowest cost link is used. If there are multiple equally low cost links, load balancing takes place between up to a maximum of 6 route entries.

RFC 2328 describes Dijkstra's Algorithm (also called the Shortest Path First (SPF) algorithm).

OSPF has a 5 second damper in case a link flaps. A link change will cause an update to be sent only after 5 seconds has elapsed so preventing routers locking up due to continually running the SPF algorithm and never allowing OSPF to converge. There is also a timer that determines the

minimum time between SPF calculations, the default for this is often 10 seconds.

A Password can be enabled on a per Area basis so providing some form of security and consistency in route information.

Types of Multi-access networks

As mentioned earlier these are typically Frame Relay, ATM or X.25 networks that have no broadcast capability but have many routers connected. There are three types:

- Hub and Spoke - a central router has links to other routers in a star arrangement. A spoke can only talk to other spokes via the hub.
- Full Mesh - each router has a link to every other router providing full resilience.
- Partial Mesh - not all routers have links to the central site.

Point-to-Point and Multipoint-to-Point networks have no need for DR/BDRs and form adjacencies with their neighbors automatically and quickly without the need for static neighbors being configured.

In a hub-spoke network operating in Broadcast mode the DR really needs to be the hub router in order for it to maintain contact with all the routers. It is therefore important to make sure that none of the other routers can become the DR by

setting their interface priorities to 0 or raising the hub router's interface priority to be the highest.

The Non-Broadcast Multi-Access (NBMA) network has all the router interfaces in the same subnet, in addition the neighbors have to be statically defined because there is no facility for broadcasts. You can also configure sub-interfaces to allow separate subnets and therefore separate NBMA networks to exist.

Rather than use a NBMA network where you have to statically configure the neighbors you can configure a Point-to-Multipoint network for Partial Mesh networks. In this case there is no DR and each link is treated as a separate Point-to-Point. A Point-to-Multipoint network can exist in one subnet.

There are some Point-to-Multipoint networks such as Classic IP over ATM that do not support broadcasts. For these networks you can configure a Point-to-Multipoint Non-broadcast mode that requires the configuration of static neighbors since they cannot be discovered dynamically.

OSPF Packet Types

Within the OSPF header the packet type is indicated by way of a type code as follows:

Type Code	Packet Type
1	Hello
2	Database Description
3	Link State Request
4	Link State Update
5	Link State Acknowledgment

OSPF Areas

Within a network, multiple Areas can be created to reduce CPU demands in the SPF calculations, memory use required for the LSDB, and the number of LSAs being transmitted. 60-80 routers are considered to be the maximum to have in one area. The Areas are defined on the routers and then interfaces are assigned to the areas. Area 0 is the default area, and you must have an Area 0, even if there is only one area in the whole network. As more areas are added, Area 0 will be your 'backbone area'. In fact, if you have one area on its own then it *could* be configured with a different area number than 0 and OSPF will still operate correctly, but this should really be a temporary arrangement. You may want to set up separate areas initially that are to be joined at a later date. Separate LSDBs are maintained one per area and networks outside of an area are advertised into that area, routers internal to an area have less work to do as only topology changes within an area affect a modification of the SPF specific to that area.

Another benefit of implementing areas is that networks within an area can be advertised as a summary so reducing the size of the routing table and the processing on routers external to this area. Creating summaries is made easier if addresses within an area are contiguous.

In a multiple area environment there are four types of router:

- **Internal router:** All its directly connected networks are within the same area as itself. It is only concerned with the LSDB for that area.
- **Area Border Router:** This has interfaces in multiple areas and so has to maintain multiple LSDBs as well as be connected to the backbone. It sends and receives Summary Links Advertisements from the backbone area and they describe one network or a range of networks within the area.
- **Backbone Router:** This has an interface connected to the backbone.
- **AS Boundary Routers:** This has an interface connected to a non-OSPF network which is considered to be outside its Autonomous System (AS). The router holds AS external routes which are advertised throughout the OSPF network and each router within the OSPF network knows the path to each ASBR.

A RIP network will look at any IP address within an OSPF network as only one hop away.

When configuring an area, authentication can be configured with a password which must be the same on a given network but (as in RIPv2) can be different for different interfaces on the same router.

There are seven types of Link State Advertisements (LSAs):

- Type 1: Router Links Advertisements are passed within an area by all OSPF routers and describe the router links to the network. These are only flooded within a particular area.
- Type 2: Network Links Advertisements are flooded within an area by the DR and describes a multi-access network, i.e. the routers attached to particular networks.
- Type 3: Summary Link Advertisements are passed between areas by ABRs and describes networks within an area.
- Type 4: AS (Autonomous System) Summary Link Advertisements are passed between areas and describe the path to the AS Boundary Router (ASBR). These do not get flooded into Totally Stubby Areas.
- Type 5: AS External Link Advertisements are passed between and flooded into areas by ASBRs and describe external destinations outside the Autonomous System. The areas that do not receive these are Stub, Totally Stubby and Not So Stubby areas. There are two types of External Link Advertisements, Type 1 and Type 2. Type 1 packets add the external cost to the internal cost of each link passed. This is useful when there are multiple ASBRs advertising the same route into an area as you

- can decide a preferred route. Type 2 packets only have an external cost assigned so is fine for a single ASBR advertising an external route.
- Type 6: Multicast OSPF routers flood this Group Membership Link Entry.
 - Type 7: NSSA AS external routes flooded by the ASBR. The ABR converts these into Type 5 LSAs before flooding them into the Backbone. The difference between Type 7 and Type 5 LSAs is that Type 5s are flooded into multiple areas whereas Type 7s are only flooded into NSSAs.

Stub Area

A stub area is an area which is out on a limb with no routers or areas beyond it. A stub area is configured to prevent AS External Link Advertisements (Type 5) being flooded into the Stub area. The benefits of configuring a Stub area are that the size of the LSDB is reduced along with the routing table and less CPU cycles are used to process LSA's. Any router wanting access to a network outside the area sends the packets to the default route (0.0.0.0).

Totally Stubby Area

This is a Stub Area with the addition that Summary Link Advertisements (Type 3/4) are not sent into the area, as well as External routes, a default route is advertised instead.

Not So Stubby Area (NSSA)

This area accepts Type 7 LSAs which are external route advertisements like Type 5s but they are only flooded within the NSSA. This is used by an ISP when connecting to a branch office running an IGP. Normally this would have to be a standard area since a stub area would not import the external routes. If it was a standard area linking the ISP to the branch office then the ISP would receive all the Type 5 LSAs from the branch which it does not want. Because Type 7 LSAs are only flooded to the NSSA the ISP is saved from the external routes whereas the NSSA can still receive them.

The NSSA is effectively a 'No-Mans Land' between two politically disparate organizations and is a hybrid stubby area. Over a slow link between the two organizations you would not normally configure OSPF because the Type 5 LSAs would overwhelm the link, so redistribution across RIP would be common. With NSSA, OSPF can still be maintained but by using less intensive Type 7 LSAs.

RFC 1587 describes the Not So Stubby Area.

Virtual Links

If an area has been added to an OSPF network and it is not possible to connect it directly to the backbone or two organizations that both have a backbone area have merged, then a virtual link is required. The link must connect two

routers within a common area called a Transit Area and one of these routers must be connected to the backbone. A good example of its use could be when two organizations merge and two Area 0s must be connected i.e. 'patching the backbone'.

Virtual links cannot be used to patch together a split area that is not the backbone area. Instead a tunnel must be used, the IP address of which is in one of the areas.

Summaries

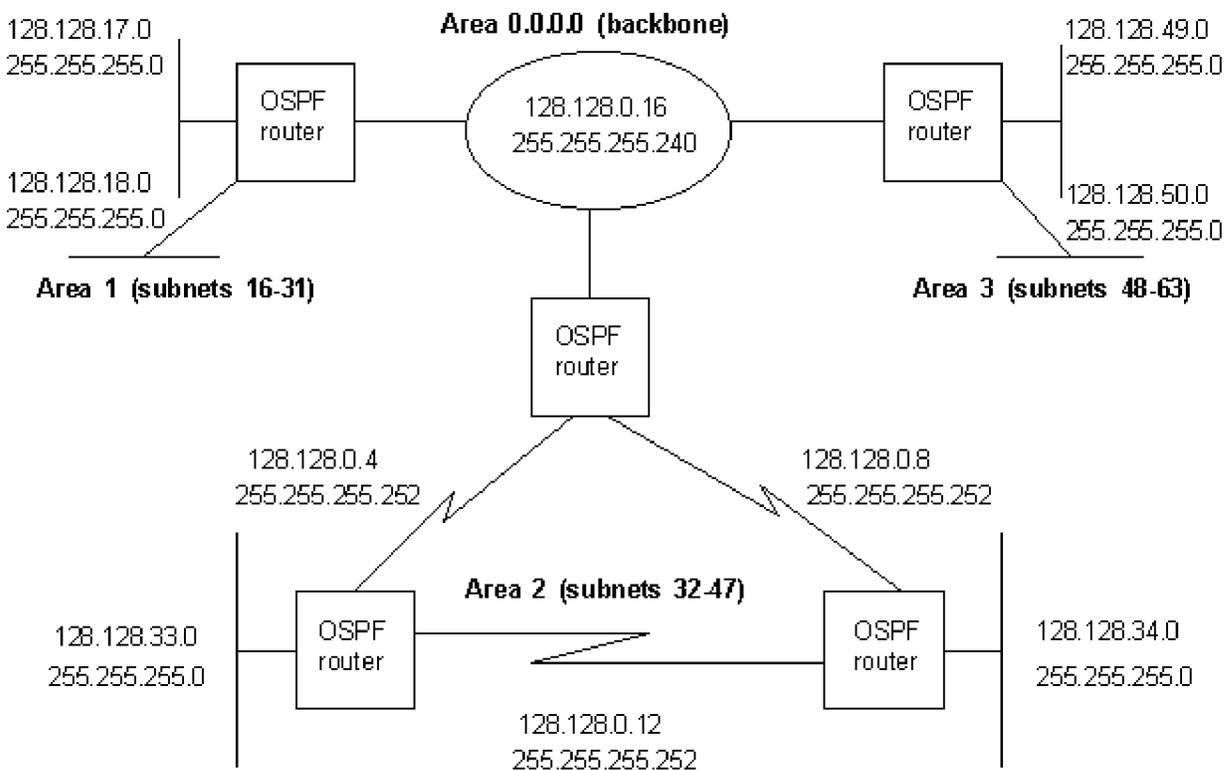
Summary Link Advertisements are sent by Area Border Routers. By default they advertise every individual network within each area to which it is connected. Networks can be condensed into a network summary to reduce the number of Summary Links Advertisements being sent. This also reduces the LSDB's of routers outside the area. In addition, if there is a network change then this will not be propagated into the backbone and other areas, which helps minimize the required SPF recalculation.

There are two types of summarizations:

- Inter-Area Route Summarization is carried out on ABRs and applies to routes from within each area rather than external routes redistributed into OSPF.
- External Route Summarization is specific to external routes redistributed into OSPF.

A summary is configured by defining a range within which the subnets that need to be summarized fall. The range is made up of an address and a summary mask, the address encompasses the range of subnetworks to be included within the summary and the mask describes the range of addresses.

Using the network in the following diagram, summaries can be created to illustrate the process:



Within Area 1: The summary address is 128.128.16.0 because of the way summarizing works. This forms the bottom of the range of addresses within the summary mask of 255.255.240.0 and gives available addresses up to 128.128.31.0, see below:

255.255.240.0	11111111 11111111 11110000 00000000
128.128.16.0	10000000 10000000 00010000 00000000
128.128.17.0	10000000 10000000 00010001 00000000
:	:
128.128.31.0	10000000 10000000 00011111 00000000

All the network possibilities from 16 to 31 are defined by the mask (third octet of 240), the existing networks can be added to. If 17 had been used as the summary address instead of 16, then the third octet would be 00010001, the problem here is that a subnet bit is set to '1' in the host area of the address. The system will not use bits that are set to '1', it only increments from '0' to '1', this means that subnet 19 would be ignored, and 21 etc. etc. The other areas can be summarized in a similar manner.

If an Area Border Router does not have an interface in area 0.0.0.0 then a virtual link needs to be created between an Area Border Router that is connected to the backbone and ends at an Area Border Router of the non-contiguous area. The virtual link is tied to the least-cost path through the 'Transit area' between the backbone and the non-contiguous area. An adjacency is formed between the two routers and the timers need to be identical.

External Routes

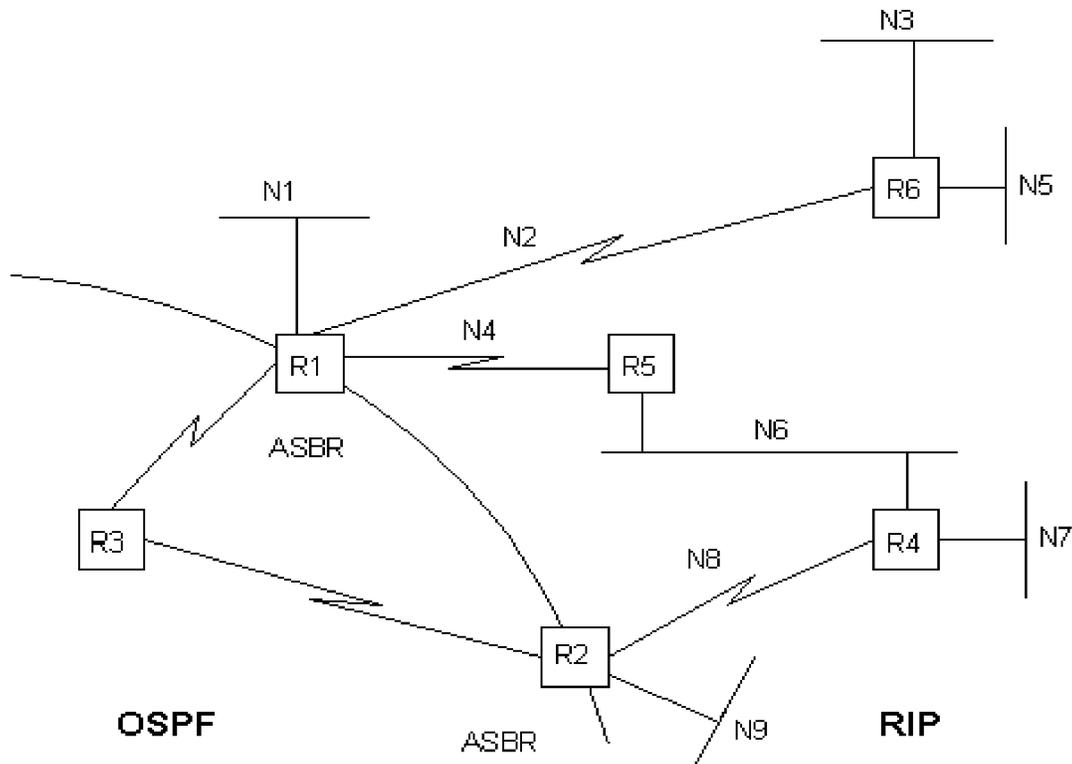
In order to make non-OSPF networks available to routers within an OSPF network, the router connected to the non-OSPF network needs to be configured as an AS Boundary Router (ASBR). As described earlier AS External Link Advertisements (one for each external route) are flooded into the OSPF network (except Stub networks). There are two types of metric for external destinations:

- Type-1 destination networks: The cost to an external network directly connected to the ASBR (close) plus the internal path cost within the OSPF network gives the total cost.
- Type-2 destination networks: The cost to a 'far away' network (i.e. not directly connected to the ASBR) is merely the number of hops from the ASBR to the external network.

If a number of routes to a network are advertised to an internal OSPF router, then the router picks the Type-1 route rather than the Type-2 route. If this router learns the route via different protocols then it decides which route to use based on firstly the preference value (configurable) and then on route weight (non-configurable).

OSPF Accept Policies

These can only be configured for external routes (Type-1 and Type-2) and can be set up on any router. Consider the following network:



An OSPF Accept Policy can be configured on R3 to prohibit R3 from forwarding IP datagrams to N1. N1 is learned as a Type-1 external route from R1 (since N1 is directly connected to R1 which is an ASBR) but N1 is also learned as a Type-2 external route from R2 (since N1 is now several networks away from R2). Because the routing table in R3 sees N1 as a Type-1 or Type-2 external route, an Accept Policy can be created to exclude these networks from R3's routing table, however other routers within the OSPF domain can still learn about N1 unless Accept Policies are also configured on these.

OSPF Announce Policies

Unlike OSPF Accept Policies, the OSPF Announce Policies can only be configured on an ASBR since they determine which

Type-1 and Type-2 external routes are advertised into the OSPF domain. Referring to Fig. 25c:

We want traffic from R3 to N6 to be routed via R2, and if R2 goes down then the traffic to go via R1. R3 learns about N6 after receiving Type-2 external LSAs from R2 and R1, the metric being 2. To force traffic through R2 we can create an announce policy on R1 that advertises N6 with a metric of 3.

Important parameters for both Accept and Announce Policies are Name (of Policy - this needs to describe what it actually does), precedence (out of a number of policies created, the one with the highest metric takes precedence) and route source (hexadecimal values indicating the non-OSPF protocols contributing to the route).

Just a final note to say that some items shown on the OSPF Announce Policy screen only actually apply to RIP Policies, the software has been lazily written.

The Achilles heel of OSPF is that all areas are connected to the backbone area. This limits the number of routers that can take part in OSPF to about 1000. The IS-IS routing is designed to be more scalable than OSPF.

RFC 1583 and RFC 2178 describe OSPF 2.