Network Working Group
Internet Draft
Expires: January 2006
Informational

Manav Bhatia
Riverstone Networks
Vishwas Manral
SiNett Corp.
Yasuhiro Ohara
Keio University

IS-IS and OSPF Difference Discussions

draft-bhatia-manral-diff-isis-ospf-01.txt

Status of this Memo

Copyright Notice

Abstract

The increasing popularity of IS-IS [IS-IS] and OSPF [OSPF] over the years has drawn significant attention to the relative merits and de-merits of one with respect to the other. This draft presents an elaborate comparison between the two routing protocols to explain how the features and functionalities of one differs from the other. Wherever applicable the differences between OSPFv2 and OSPFv3[OSPFv3] have also been pointed out.

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED","MAY", and "OPTIONAL" in this

document are to be interpreted as described in RFC 2119 [KEYWORDS]

Table of Contents

1. Terminologies

Since both these routing protocols originated in different standard bodies, IS-IS in ISO and OSPF in the IETF, there exists some difference in the terminologies used.

IS-IS - OSPF

End System - Host
Intermediate System - Router
Circuit - An adjacency on one link
SNPA Address - Data link Address
Protocol Data Unit (PDU) - Packet
Designated Intermediate System (DIS) - Designated Router (DR)
IS to IS Hello PDU (IIH) - Hello Packet
Not Applicable - Backup Designated Router (BDR)
Link State Packet(LSP) - Link State Advertisement (LSA)
Link State Packet - Link State Update
Complete Sequence Number Packet(CSNP) - Database Description packet
Partial Sequence Number Packet(PSNP) - Link state ACK or Request Packet
Routing Domain - AS
Level 2 Subdomain - Backbone Area
Level 1 Area - Non Backbone Area
Level 1/2 IIH PDU - Simple Hello Packet
Level 1/2 LSP - No Distinction
L1L2 router - ABR
System ID - Router ID
Link State Packet ID(LSPID) - Link State ID
Pseudonode LSP - Network LSA

Router LSAs, Summary LSAs, Network LSAs, ASBR Summaries, AS-external LSAs are equivalent of TLVs carried in LSPs in IS-IS. The difference is that each LSA has its own header whereas the TLVs share a common header.

IS-IS Terms with no OSPF equivalent:
TLV - Type-Length-Value tuple. These carry most of the information in
IS-IS PDUs.

OSPF Terms with no IS-IS equivalent:

Advertising Router - Router that originated the advertisement. In IS-IS, this is the LSP's originator.

Backup Designated Router - Router which takes over in case the DR goes down. In IS-IS, there is no Backup DIS and the DIS election takes place again in case the former goes down or is no more available.

Backbone Area - In IS-IS, L2 routers appear in all areas, but must all be interconnected to form a backbone (the L2 subdomain)..

2. Acknowledgements

This document is a result of the extensive discussions in the diff-ospf-isis list and the following people have co-authored and contributed to this draft, either directly or indirectly:

Danny McPherson, Jeff Learman, Jonathan Sadler, Radia Perlman, Philip Christian, J.J. Syed, Satish Dattari, Sina Mirtorabi, Nabendu Das, Russ White, Alex Zinin and Venkata Naidu.

3. Evolution of the protocols

Both Integrated IS-IS and OSPF were specified in the latter part of the 1980s.

In 1987 OSI adopted DECnet Phase V's routing algorithm with some modifications and named it IS-IS. Around 1988, the NSFnet deployed an IGP loosely based on an early draft of IS-IS. Around the same time, development on OSPF started which took most of the basic concepts from this early version of IS-IS but was designed to support only IPv4. In October 1989 the version 1 of OSPF was released as RFC 1131 and around the same time in December 1990, Integrated IS-IS was released and published as RFC 1195.

Version 2 of OSPF was first published in July 1991 as RFC 1247 and CISCO started shipping it. It released its implementation for Dual IS-IS in 1992. Till now numerous ISPs had deployed OSPF and very few IS-IS. In 1994 there were significant improvements done to CISCO's IOS implementation for in conjunction with support for Network Link Service Protocol (Novell's IPX protocol).

These enhancements improved the performance, resilience and robustness of CISCO's implementation which made a lot of ISPs to shift to IS-IS.

By 1995 most of the major ISPs had started deploying IS-IS. What helped this further was US government's interest in ISO CLNS suite, which was reflected in a requirement for CLNP routing support in the

NSFnet project by the NSF. Interest in Dual IS-IS continued to grow, and most ISPs that sprung up in Europe chose to deploy ISO standards based on IS-IS instead of OSPF.

Unlike IS-IS which started as an ISO protocol, OSPF was inherently designed to support only IPv4 and was promoted by IETF as the referred IGP for IP networks. Additionally, because IS-IS support was not available on some major routers (noticeably Bay and 3com routers), OSPF automatically became the standard de-facto IGP for the reasonably large sized networks with multi-vendor platforms. An active IETF WG and evolving specifications also went a long way to help promote OSPF; and thus it started becoming more popular and more widely adopted compared to IS-IS [MARTEY].

There has been no major standardization effort in the ITU for a while, so ISO 10589 and RFC 1195 still remain the authoritative complete standards for IS-IS. The IETF IS-IS WG has been opened recently which is now working on standardizing newer applications like MPLS, Traffic Engineering, IPv6, etc for IS-IS.

To summarize, both the protocols have prevailed through the test of time and have established themselves as the IGPs of choice for ISPs. New extensions such as, MPLS TE, IPv6, have been deployed over the past 3 years, and with active working groups for either protocol in IETF, they continue to evolve in lock-step fashion.

4. Interface Types Supported

OSPF models networks as
- Broadcast links
- Point to Point (P2P)
- Point to Multi-Point (P2MP)
- Non-Broadcast multi-access Networks (NBMA)

IS-IS models networks as
- P2P
- Broadcast
- Unnumbered Broadcast

The key differences are the way OSPF provides support for NBMA networks and inherent protocol support for unnumbered broadcast by IS-IS

4.1 Support for NBMA Networks

IS-IS has no direct support for connecting ISs over a NBMA network and it must be modeled as a LAN or treated as a set of P2P links.

Modeling it as the latter involves a lot of configuration and if full

connectivity is not configured, multiple hops might be required for traversing the NBMA cloud.

Experience with ATM LAN emulation has proven un-scalable and insufficiently reliable because of the single point where replication takes place to emulate multicast.

The best alternative for IS-IS is thus to treat each PVC as a point-to-point link. All PVC failures are handled by the protocol since each PVC is visible to the protocol. IS-IS mesh groups [MESH] may be used to address the scaling issues which may result from redundant flooding in the highly meshed environments.

In OSPF there is a "NBMA mode" in the original specification which makes the protocol aware that it is on a NBMA network.

Neighbours are discovered initially through configuration which is restricted to the ones eligible for the DR election. To make administration easier and to reduce the HELLO traffic, most of the other routers attached to the NBMA subnet are assigned a router priority of zero. It thus involves quite a bit of administration overhead and is prone to mis-configuration. Also the network will malfunction if one of the nodes loses its link to the DR.

In this mode, each node in the NBMA must have a PVC to the DR and BDR.
Since adjacencies between non-DR nodes is not mandated, the order of the number of adjacencies is O(2n), rather than O(n^2) as required when running OSPF without NBMA mode.

NBMA networks are thus only as robust and reliable as the underlying data-link service. If for example, a PVC fails or is mis-configured or if an SVC cannot be established, due to capacity or policy reasons, routing over NBMA subnet will fail. And, unfortunately, often the reason for the failure will not be immediately obvious to the network operator.

The P2MP can be applied to rectify these problems, although at some loss of efficiency.

4.2 Point-to-Multipoint model

This model can be used on any data link technology that the NBMA model can be used on. In addition, the P2MP model doesn't require all the participating routers to be able to communicate directly to model a partial PVC mesh as a single P2MP networks. Dropping the full mesh requirement also allows the modeling of more exotic data link technologies, such as packet radio, as P2MP networks [Moy].

So if an Operating system can't support virtual interfaces or if there's too much overhead involved in generating separate sub interfaces to each of the 500 ATM circuits then P2MP is good and can be handy that way.

However, when operating a full mesh Frame Relay or ATM network in P2MP mode, the work involved in neighbor maintenance, flooding, and database representation increases as O(n^2), where n is the number of OSPF routers attached to the subnet, instead of O(n)behavior that can be achieved with the original NBMA model.

4.3 Unnumbered broadcast

IS-IS supports unnumbered broadcast interfaces; however, most implementations do not. The protocol provides all necessary routing information without the aid of ARP [ARP], but doing this requires that each FIB entry contain a next-hop (circuit, SNAP address) pair for each path to a destination, and many routers are designed with FIB entries that contain only next-hop IP addresses instead, to reduce the size of the FIB and perhaps as a simplification.

For this reason, many implementations won't interoperate with an unnumbered broadcast interface, and won't interoperate with an implementation that doesn't support ARP.

5. Encapsulation

IS-IS runs directly over the data link alongside IP. On Ethernet, IS-IS packets are always 802.3 frames, with LSAP value 0xFEFE while IP packets are either Ethernet II frames or SNAP frames identified with the protocol number 0x800. OSPF runs over IP as protocol number 89.

IS-IS runs directly over layer 2 and hence

- cannot support virtual links unless some explicit tunneling is implemented

- packets are intentionally kept small so that they don't require hop-by-hop fragmentation

- uses ATM/SNAP encapsulation on ATM but there are hacks to make it use VcMux encapsulation

- some operating systems that support IP networking have been implemented to differentiate Layer 3 packets in kernel. Such OSs require a lot of kernel modifications to support IS-IS for IP routing.

- can never be routed beyond the immediate next hop and hence shielded from IP spoofing and similar Denial of Service attacks.

- need to provide code points of access for each data link protocol types (Frame Relay, Ethernet, ATM, PPP [PPP], etc.)

- don't need to rely on network layer protocols (like ARP) to communicate with the neighboring systems. Some implementations however, do rely on ARP or static routing to communicate with neighbors on LAN.

OSPF runs over IP and hence

- can support virtual links
- can use IP fragmentation services
- can use VcMux encapsulation on ATM
- if an OS already supports IP, no changes are necessary to support OSPF
- can be routed to a destination multiple hops away and thus vulnerable to Denial of Service attacks and IP spoofing
- transmitted with additional IP header information, thereby increasing some packet overhead

5.1 IP Fragmentation

LSPs in IS-IS, unlike as in OSPF, are not regenerated hop-by-hop and so they must be small enough that they are guaranteed to be able to cross *any* media in the network and the value of the maxsized LSP should thus not be greater than the minimum link MTU size in the area.
If a router has more than maxsized LSP bytes of information to advertise into IS-IS, then this originating router must fragment its LSP before flooding.

One area of the concern regarding the scalability of the link state routing protocols is the flooding and it is believed that preventing fragmentation during flooding is the reason why IS-IS fragments only at the originating router.

OSPF does not provide any explicit fragmentation/reassembly support. When fragmentation is necessary, IP fragmentation/reassembly is used.
OSPF protocol packets have been designed so that large protocol packets can be generally be split into several smaller protocol packets.

5.2 ATM Encapsulation

OSPF can run over ATM using VcMux encapsulation (which essentially
assumes that all the packets carried are IP) while IS-IS requires LLC/SNAP encapsulation where ATM layer can distinguish between multiple Layer 3 protocols over the same VC. The disadvantage of

using the LLC/SNAP encapsulation is that it has some additional bytes
for the LLC-SNAP header which results in a packet size > 40 bytes. Thus a simple TCP ACK message of 40 bytes along with the LLC-SNAP header adds enough bytes so that a single TCP ACK won't fit into one ATM cell.

Much bandwidth is thus wasted because now each TCP ACK requires 2 ATM
cells. An IETF draft proposes a workaround to this issue in which both IS-IS and IP packets can be sent over an ATM VC using Vc Mux encapsulation by reading into the first byte of the L3 header to distinguish between IP and ISO family packets, such as IS-IS, CLNS and ES-IS. However this did not gain popularity because of the demise
of ATM cores in the largest ISPs (which were also among the few running IS-IS).

[*] The first two fields in the IP header are the 4-bit version number and the 4-bit header length. The value of the first byte is normally 0x45. If there are IP header options attached to the IP header, the first byte can be between 0x46 and 0x4F. The first byte in an IS-IS packet is always 0x83. Thus by looking at the first byte of an incoming packet, the receiver can separate IP and IS-IS packets.
Because of this feature one does not need to depend on the ATM layer anymore to help with the de-multiplexing. Routers an now send and receive both IS-IS and IP packets using Vc Mux encapsulation and thus
avoid the ATM cell tax. [*]

6. Designated Router (DR) concept

The DR concept is used by both IS-IS and OSPF on the broadcast media to limit the amount of LS information exchanged between the routers on such media. It helps to reduce the number of adjacencies formed on
broadcast media to O(n) instead of O(n^2), where n is the number of nodes.

IS-IS

- DR election is deterministic
- No concept of backup DIS
- A new DIS is elected when the current goes down.

OSPF

- DR election is non-deterministic.
- Elects DR and BDR to conduct flooding on a LAN.
- All routers on the LAN are only synchronized with the DR and BDR.
- DRship is sticky

6.1 DR election deterministic/non-deterministic

In IS-IS, deterministic DIS election makes the possibility of predicting the router that will be elected as DIS from the same set of routers. The router advertising the numerically highest priority wins, with numerically highest MAC address breaking the tie. In IS-IS, DIS can be pre-empted at any time by a router with higher priority coming alive.

In OSPF, the DR election is sticky meaning that after a router has been elected, no other router can take over the position unless the original DR goes down. When a router comes up, it accepts the DR regardless of its own priority if a DR is already there. Otherwise the router itself becomes DR if it has the highest priority on the network. The above scheme makes it harder to predict the identity of the DR, but ensures that DR changes less often.

The rationale behind this sticky nature of DRship in OSPF is that it is disruptive to have DR changes as DR keeps track of which nodes have acknowledged which link state information and it would require a lot of time and protocol messages for another router to take over in case the DR went down.

Both the sticky and deterministic mechanisms of DR/DIS elections in OSPF and IS-IS can be modified to provide the functionality of the other with some simple modifications in the implementations.

6.2 Backup Designated Router/Intermediate System

A backup DIS is redundant in IS-IS because all the routers are synchronized with each other and also because the shorter Hello interval used by the DIS allows for faster detection of failures and subsequent replacement of the DIS.

The presence of BDR in OSPF makes the replacement of the DR transparent in case the DR goes down. All routers on the LAN are only adjacent and synchronized with DR and BDR; and backup DR is fully synchronized with the DR. Forming adjacencies with only the DR/BDR is done to reduce the complexity of data exchange and minimize flooding.

7. Areas/Hierarchy

This is required primarily for scalability issues wherein instabilities inside one small section of the network are hidden from the rest of the network. This also helps in reducing the size of the routing tables, etc. Both the protocols establish a two level hierarchy among the areas.

IS-IS

- Divides the whole routing domain into small areas and uses logical hierarchy based on routing levels called Level 1 and Level 2

- Level 1 routing is within the area and L2 is between the areas.

- Original spec called for Level 1 routers to know only the topology inside their area and they were unaware of routers/destinations outside of their area. They simply forwarded all their traffic for outside their area to the nearest Level 2 router

- Level 2 routers knew only the Level 2 topology and didn't know any topology inside the area. This forced strict hierarchal routing between the areas where all inter-area data traffic originating from one area followed a default route to the Level 2 sub-domain, where it was forwarded by L2 routing to the destination area.

- This has now changed and a recent draft in IETF allows leaking L2 information inside L1 for more optimal routing.

- There was some work done in IS-IS for multi-level hierarchies but it wasn't all that useful and was dropped in between. The idea was that if the networks use IDRP as well along with IS-IS then the 2 levels may not be enough.

- IS-IS routers are associated with a single area and the whole router then belongs to that particular area.

- Area boundaries intersect on links

- can be extended to support higher levels of hierarchy based on the way routes are leaked in between the levels by setting the up/down bit, when routes are propagated down the hierarchy.

OSPF

- Divides the routing domain into regular areas and a backbone area that is designated as area 0.0.0.0 and all packets going from one area to the other must traverse through this backbone.

- The spec calls for the backbone to be contiguous and to be connected to all the areas through an ABR. There is however a provision to work with disconnected physically disparate backbone areas using virtual links [Refer to section 13 for more details]

- Can be attached to multiple areas as its designed around links and uses a links based addressing scheme. It's the links which are assigned to the areas and not the routers themselves.

- Areas intersect on routers.

8. Checks on Hellos for adjacency formation

The HELLO protocol is responsible for formation of adjacencies. Forming adjacencies is an integral part of link state routing protocols as all protocol packets other than hellos are flooded only over these adjacencies. The rules for formation of such adjacencies however differ between IS-IS, OSPF v2 and OSPF v3. The main points are: -

IS-IS

Besides the basic checks to verify the integrity of the packet, IS-IS has a few checks to verify before formation of adjacencies when receiving hellos.

- The IS-IS protocol allows multiple area-address to be configured on a router. During the hello exchange the adjacency is formed only if atleast one of the area address matches. The advantage of having multiple areas is given in section 22. However Level 2 only adjacencies can be formed even if the area addresses are not matching.

- Besides to prevent the LSP's and CSNP's being dropped due to different values for originatingLSPBufferSize and ReceiveLSPBufferSize, all HELLOs are padded till the adjacency comes up again. This check verifies consistent settings between the adjacent routers. This is however not a sufficient check.

- Adjacencies are formed without regard to interface addressing or asymmetric in HOLD timer values. Values of HELLO interval are not sent in HELLO packets. While the IS-IS protocol provides sufficient routing information for relaying packets between adjacent routers, many implementations nonetheless require ARP support to do this. These implementations typically refuse to form an adjacency unless the neighbour interface IP address is on the local interface's IP subnet.

- IS-IS can carry addressing information of different protocols inT TLV's. However, the protocol supported field must be sent in Dual[RFC1195] and IP-Only routers. RFC1195 specifies no checks for the protocol supported field for adjacency formation. It places topology restrictions on multi-protocol networks. In networks that conform to these restrictions, neighboring routers will always have a protocol in common. Therefore, it does not state whether adjacency formation should take protocols supported into account. Many implementations however, do not form an adjacency with a neighbor unless they have at least one protocol in common [as described in ITU-T G.7712 and draft-ietf-IS-IS-auto-encap-02.txt.]

- Not matching hold timer values has advantages wherein the administrator can set different hold times for different routers. This helps in cases where the going down of a DIS or some router needs to be detected faster. For such routers the hold timer can be set to a lower value.

OSPFv2

The checks for formation of adjacencies are stricter in OSPFv2 than IS-IS.

- The area-id of the received packet should always match the incoming interface (with the exception of virtual links). Area type is strictly checked by checking the E-bit (not set for non-default areas) and the N- bit (not-set for non-NSSA areas).

- The values of the HELLO interval, the Router Dead Interval and network mask received in HELLOs are matched with those on the configured interface. Any mismatch in the values causes the HELLO packet to be dropped and hence prevents formation of adjacencies. The disadvantages of this approach is that Hello Interval and Router Dead Interval changes need to be done within the Router Dead Interval, to prevent breaking adjacencies. The advantage is we would not form adjacency in case there is a router that has been mis-configured with a large value and which could cause problems later. The network mask check however does not apply to point to point links. That allows the two ends of a Point-to-Point link to have different addresses.

- MTU check is not done in the hellos. It is done in the during the DB Exchange process.

OSPFv3

Most of the checks for OSPFv3 are similar to that of OSPFv2. The main points of differences are: -

- OSPFv3 runs on a per link basis instead of a per subnet basis. The check for network mask is not done.

- Instance ID field (non-existent in OSPFv2) on the link is matched with the incoming ID in Hellos. The adjacency is formed only if the Instance-ID matches. This allows multiple instances of OSPF to run on a single link.

9. Database Exchange and Flooding

9.1 Initial Database Exchange

For the SPF algorithm to work properly, all routers in the area should have the same database information on which the SPF algorithm works. The process of synchronization includes the "Initial Database Exchange" which is done when the adjacency is coming up and the asynchronous flooding when the Adjacencies are up.

OSPF

- A master-slave relation is established to do the database exchange.
Besides the MTU is exchanged in the database description packets before any database exchange starts.

- The database exchange begins once the adjacency state reaches Exstart. On a broadcast links, the DR and BDR form adjacencies with all other routers on the network.

- Only one DB Description packet can be unacknowledged at a time that
is, the window size is 1. Each DB Description packet from the master is acknowledged by the slave. The slave sends its own DB Description packet with similar identifiers as the masters.

- DB description packets containing the summary of LSA's at each end are exchanged. Only when the entire summary is received by the neighbour can it tell which instance of the LSA is not there in the senders database.

- An adjacency in OSPF is declared FULL/UP, when the entire database exchange is completed.

- OSPF does not allow routers to resynchronize their link state database in the steady state. It is only done during the initial database synchronization or when network topology changes. However, there are techniques to do that. One such way is described in "OSPF Out-of-band LSDB resynchronization" [OOB]

IS-IS

- The MTU check is done at the hello exchange time itself.

- CSNP's are sent by the DIS on a broadcast link. On a point-to-point
link both the neighbours exchange CSNP's with each other.

- On point-to-point link all the LSP's SRM flag is also set for the circuit, to indicate the LSP's have to be sent over the circuit.

- The CNSP's are sent to reduce the actual flooding of all the LSP's between the neighbours.

- Multiple CSNP's can be sent together. CSNP's unlike DB Descriptions in OSPF are not acknowledged.

- As the CSNP's have a range of LSP-ID's, and contain all the LSP's in the database falling in that range. A neigbour on receiving a CSNP can know which LSP's in the neighbour are newer, which older and which are absent. Based on this the neighbour can send newer LSP's to the neighbour.

- Link state database is continuously refreshed and synchronized because of the periodic CSNPs that are announced.

9.2 Asynchronous Flooding

Whenever any information in an the database changes, the information is to be exchanged with all other routers in the network. This is done by the flooding process: -

OSPF

- Uses reliable flooding mechanism for all link types.

- Changed LSA's are packed in LS Update packets and send over adjacencies to the neighbour, which unpacks the LSA's. LS Acknowledgement packets are sent by the receiver, which informs the sender that the receiver has received the LSA.

- The sender retransmits the LSA's after the re-transmission interval if it does not get acknowledgements for them.

- On a broadcast link LSUpdate packets are sent only to all-DR routers multicast address. The DR floods the LSUpdate packets to All-SPF-Routers.

- Whenever a new DR/BDR is elected, it has to form adjacencies with all other routers in the network.

- There is no difference in the asynchronous flooding procedures between OSPFv2 and OSPFv3.

IS-IS

- LSP's are flooded as is across the area. They are not packed inside any other packet.

- On broadcast links flooding is not done reliably. A changed LSP is

flooded to all IS-IS routers, however no retransmissions occur.

- The reliability in database exchange on a broadcast link is achieved by periodic database exchange. This is done as CSNP's are sent periodically by the DIS, which initiates the entire database exchange process all over again.

- As the DIS sends periodic CSNP, nothing different needs to be done when a new DIS is exchanged.

- On a point-to-point link flooding is done reliably. LSP's are flooded to the neighbour and if CSNP entry for the LSP is not received in a particular time interval, the LSP is re-flooded to that neighbour.

10. Flushing LSA/LSP

An LSA/LSP is flushed (purged) when the contents carried by the LSA/LSP are no longer valid. In OSPF when an LSA is flushed the age is set to MaxAge and the LSA is flooded. In IS-IS when an LSP is purged (flushed) the header alone is flooded with the Remaining Lifetime set to 0, and the value of checksum set to 0. OSPF only allows self originated LSA to be flushed, IS-IS spec allows in certain cases for non-self originated the LSP to be purged (though new implementations don't support this and the update draft has changed it) which can lead to problems.

In OSPF a flushed LSA is not removed unless the LSA is not on any of the retransmit lists and none of the adjacencies on the router are in state Exchange or loading. This ensures that an LSA that an LSA is flooded to all its neighbors before it is removed from the domain. In IS-IS an LSP purged is kept for ZeroAge lifetime if the LSP purged is a self originated LSP and the LSP is kept for MaxAge if the LSP is non self-originated before the LSP is deleted.

When purging an IS-IS LSP the header and authentication data is kept while purging (certain OSPF implementations do the same). However for those LSP's that don't support authentication, because the checksum is set to 0 for purged LSP's, the integrity of the contents cannot be verified. In OSPF the entire content of the LSA is intact while flushing leading to unnecessary data sending.

11. SPF Calculation

Both the protocols use Shortest Path First (SPF) algorithm to calculate the best path to all known destinations based on the information in their link state database. The SPF algorithm works by building the shortest path tree from a specific root node to all other nodes in the area and thereby computing the best route to

every
known destination from that particular source/node.

IS-IS

- SPF for a given level is computed in a single phase by taking all IS-IS LSP's TLV's together.

- IP routing is integrated into IS-IS by adding some new TLVs which carry IP reachability information in the LSPs. All IP networks are considered externals, and they always end up as leaf nodes in the shortest path tree when IS-IS does a SPF run.

- Performs only the less CPU intensive Partial Route Calculation (PRC) when network events do not affect the basic topology but only the IP prefixes.

- Used narrow (6 bits wide) metrics which helped in some SPF optimization. However such small bits proved insufficient for providing flexibility in designing IS-IS networks and other applications using IS-IS routing (MPLS-TE). "IS-IS extensions for Traffic Engineering" [X] draft introduced new TLVs which defined wider metrics to be used for IS-IS thus taking away this optimization.
But then CPU are fast these days and there are not many very big networks anyway.

OSPF

- SPF is calculated in three phases. The first is the calculation of intra-area routes by building the shortest path tree for each attached area. The second phase calculates the inter-area routes by examining the summary LSAs and the last one examines the AS-External-
LSAs to calculate the routes to the external destinations.

- Is built around links, and any IP prefix change in an area will trigger a full SPF.

- Only changes in interarea and external routes result in partial SPF
calculations and thus IS-IS's PRC is more pervasive than OSPF's partial SPF. This difference allows IS-IS to be more tolerant of larger single area domains whereas OSPF forces hierarchical designs for relatively smaller networks. However with the route leaking from L2 to L1 [RFC 2966] incorporated into IS-IS the apparent motivation for keeping large single area domains too goes away.

12. Area Types

IS-IS: Leaking between levels/areas(how it is controlled) OSPF: NSSA/stub/default

12.1 Area Partitions

With hierarchical routing (look at Areas/Hierarchy), it is possible for an area to partition so that level 1 routing cannot connect the partitions. If both partitions contain level 2 routers, and the level 2 network is connected, the network as a whole is not physically partitioned. There is a path between the partitions of the area. The path is level 2 path.

The symptoms of a partitioned area can be difficult to diagnose and annoying for the users. Not only is communication impossible between nodes that should be in the same area, but are currently in different partitions of the area, but communication between members of the area and nodes outside the area can be disrupted since the traffic into the area might enter the wrong partition and be undeliverable.

IS-IS has mechanisms in which level 2 routers residing in a partitioned area automatically detect and repair the partition by utilizing the level 2 path as a level 1 link. Routing control messages as well as data packets are encapsulated with a network layer header and transmitted over the virtual link. To the rest of the nodes in the area, the area is no longer partitioned and level 1 routing proceeds normally within the area.

OSPF does not have any standard explicit area repair mechanisms. If an area splits in such a way that a ABR in one partition announces an address summary that includes an address reachable in a different partition, then routing will not work, since a packet may be delivered to the incorrect partition.

There are two methods by which OSPF can accomplish this:
- Someone might notice that the area has partitioned, and manually reconfigures the ABR in the area, so ABRs in each partition do not contain summary addresses for addresses reachable in other partitions.

- No summary address were used, and each ABR reports each IP address individually. If summary addresses are not used, areas do not become partitioned, they merely break into multiple areas.

However an on demand tunnel [TUNNEL] adjacency mechanism has been recently proposed in the IETF which solves this problem by choosing an inter-area path over an intra-area path.

12.2 Level 2 Partitions (Backbone Area Connectivity)

IS-IS requires a connected level 2 network. This means there must be a path from every level 2 router to every other level 2 router that traverses only level 2 routers [RADIA].

OSPF similarly requires a connected backbone (level 2) area, but

allows a link between a pair of backbone routers to consist of a

manually configured virtual link" that consists of a path through a non-backbone area. Communication over a virtual link between backbone routers A and B can be done in two ways:

- A can encapsulate traffic being forwarded to B in a network layer header giving B as the destination.
- A can assume all non-backbone routers on the path towards B know enough to forward traffic to the destination towards B.

Virtual link uses the second approach, this requires that all non-backbone routers in the transit area know about all destinations in the backbone area, so they will be able to forward backbone traffic in case they windup in the path of a virtual link. In other words summarization of backbone area into the transit area is ignored.

Tunnel adjacency uses the first approach, further it can used for on demand partition so that the adjacency will be established dynamically once the backbone is partitioned.

Because of the possibility of manually configured virtual links in OSPF, IS-IS has a topological restriction that OSPF does not.

12.3 Injection of Level 2 Information

In IS-IS, level 1 routers only know information about their own area. If a level 1 router R receives a packet with an address not reachable within the area, R forwards the packet to the level 2 router nearest to R. In OSPF, level 2 information is fed into the non-backbone areas. Suppose there is an area A in some AS such that:

- n IP destination addresses are reachable within the AS, but outside the area A
- m IP destinations are reachable outside the AS
- k ABRs in area A
- j ASBRs in the AS

Each of the "k" ABRs reports their own distance to the "n" IP destination addresses and the "j" ASBRs. This information is O(k*(j+n)). Each of the "j" border routers also reports its distance to each of the "m" IP destinations reachable outside the AS. This information is O(j*m).

Giving level 2 information to level 1 routers enables the routers to choose the exit level 2 router that will give the best path to the destination.

Thus, OSPF yields more optimal interarea routes than IS-IS. The cost of providing more optimal routing is increased bandwidth usage by

the
   routing algorithm and increases memory and CPU requirements in level

1 routers. Aside from increased bandwidth, CPU, and memory usage, there is an additional issue raised as a result of the OSPF requirement for level 1 routers to store level 2 information. In IS-IS where an area is independent of the rest of the network, database sizes in level 1 routers can be calculated based on the size of the area. If the area never changes, the level 1 routers will continue to function. In contrast, as the entire network grows in OSPF, demand on level 1 routers increases. One small area with small routers, cannot be sheltered from the growth of the rest of the network.

12.4 Stub Area

There is an option in OSPF, called "Stub Area." If an area is a stub area, the information concerning destinations outside the AS is not flooded into the area, saving O(j*m). Information about destinations within the AS, but outside the area are still flooded within an area, even if the area is configured as a stub area.

In other words, an OSPF stub area is a compromise between a nonstub OSPF and an IS-IS area. OSPF stub areas require significantly less storage than nonstub OSPF areas. Like IS-IS, OSPF does not attempt to optimize the route from a stub area to a destination outside the AS, but unlike IS-IS, OSPF does attempt to optimize routes from a stub area to destinations within the AS, but outside the area.

In IS-IS, none of this information is seen by the level 1 routers. The cost of not storing, propagating, and computing this information in IS-IS is that some routes to other ASs will be less optimal than those used in OSPF.

12.5 Not So Stub Area (NSSA)

"not-so-stubby" area (or NSSA), which has the capability of importing external routes in a limited fashion.

The OSPF specification defines two general classes of area configuration. The first allows Type-5 LSAs to be flooded throughout the area. In this configuration, Type-5 LSAs may be originated by routers internal to the area or flooded into the area by area border routers. These areas are distinguished by the fact that they can carry transit traffic. The backbone is always a Type-5 capable area. The second type of area configuration, called stub (described in section 10.4) does not allow Type-5 LSAs to be propagated into/throughout the area and instead depends on default routing to external destinations.

NSSAs are defined in much the same manner as existing stub areas. Type-7 LSAs provide for carrying external route information within

an
   NSSA. Type-7 LSAs have virtually the same syntax as Type-5 LSAs with

the obvious exception of the link-state type. Both LSAs are considered a type of OSPF AS-external-LSA. There are two major semantic differences between Type-5 LSAs and Type-7 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.

In order to allow limited exchange of external information across an NSSA border, NSSA border routers will translate selected Type-7 LSAs received from the NSSA into Type-5 LSAs. These Type-5 LSAs will be flooded to all Type-5 capable areas. NSSA border routers may be configured with address ranges so that multiple Type-7 LSAs may be aggregated into a single Type-5 LSA. The NSSA border routers that perform translation are configurable. In the absence of a configured translator one is elected.

IS-IS does not have such capability of an area being a Not-So-Stubby Area (NSSA).

13. Architectural Values

13.1 Architectural Constants

OSPF does have a large number of tunable parameters that can make configuration seem complicated. However, most of these parameters should be set to default values in an OSPF implementation.

13.2 Synchronized Parameter Setting

In OSPF, there are several parameters that must be configured identically in routers, or else the router will refuse to communicate with each other. This creates a problem because it is virtually impossible to change the parameter setting via network management. Once a router's parameter setting is changed, it is cut off from the rest of the network since no other routers will be able to communicate with it. In contrast, there is always a way in IS-IS to migrate from one setting to another by configuring routers one at a time while the network is running.

The parameters in OSPF that must be set identically in neighboring routers are the HelloTime and the DeadTime

IS-IS reports only DeadTime in its Hello messages (not HelloTime). As a result, the ratio between DeadTime and HelloTime is fixed in IS-IS, but can be configured in different ways by OSPF. IS-IS uses the information solely to determine how long to wait between receipt of Hellos from a particular neighbor before declaring the link to that neighbor down. There is no necessity for neighboring nodes to have the same value.

Being able to change these timers in a running network is important. As a LAN becomes larger it might be decided that the overhead from hellos is too great. It also might be important in some configurations to be able to run with different hello timers for different routers. There might be some routers for which quick deletion of failure would be very desirable, whereas for other routers quick deletion of failure might not be as important. To lower overhead these routers might be configured with a longer HelloTime. This cannot be done in OSPF since all routers must have identical timers.

- Stub Area Flag:

OSPF requires every router in an area to be configured with a flag indicating whether the area is a stub area. If a level 2 router has a stub area flag set, it will not flood type 5 LSPs into the area. The "Stub Area" flag is reported in OSPF Hello messages. If a router disagrees with a neighbor as to the setting of the "stub area" flag, it will bring the link to the neighbor down. IS-IS has no such parameter.

- Authentication Password:

Both OSPF and IS-IS have the optional feature of providing authentication. In OSPF, there is a single password per link. The password a router transmits is the same as the password it will accept on the link. IS-IS allows configuration of multiple receive passwords so it is possible to migrate from one password to another without disrupting the operation.

14. Virtual Links

IS-IS

- IS-IS allows a Level-1 Area which is partitioned to be automatically repaired, by electing Partition Designated Level 2 routers and having a virtual link between them. The mechanism is not often implemented and requires an explicit tunnelling mechanism."

- Used in ISO IS-IS for connecting partitions of Level 1 Area over the Level 2 backbone.

OSPF

- Used for connecting physically separate area zeroes (0.0.0.0) to maintain contiguity of the backbone

- Used for connecting remote areas to the backbone through other areas if direct physical connectivity is not possible. This enables an OSPF packet to be sent from one part of an remote isolated site to the main OSPF network.

- For Virtual links to work, OSPF accepts packets which are have originated more than one hop away. This can lead to security concerns if the packets at the edge of the domain are not properly filtered.

15. Packet Alignment/Extensibility

IS-IS

- Does not require any particular alignment of packet fields.

- Uses TLV (Tag-Length-Value) encoded packets to advertise routing information

- TLVs not supported/recognized are ignored by IS-IS routers

- LSPs are flooded intact with unrecognized TLV information making it very extensible. Ipv6 support is provided by simply adding a few more TLVs.

- TLVs can be nested as sub-TLVs providing even more flexibility for future extensions. Though the base spec does not use them but the newer drafts have started using them (TE extensions, etc).

OSPFv2

- Uses fixed format packets with all fields aligned at 32-bit boundaries for faster processing of the OSPF packets (doesn't really matter anymore as the CPUs are really fast these days!). This was also primarily done because OSPF was meant to be an IPv4 only protocol.

- The downside is that the packet formats are not at all extensible.

- It uses LSAs for advertising the routing information and the original spec called for dropping any unrecognized LSA type.

- LSAs of type 9, 10 and 11 (Opaque LSAs) have been introduced for advertising other application-specific information and enough

vendors

now support this so that they are likely to get from one side of the network to the other.

- Since the unrecognized LSA types are not flooded to neighbors it makes it very difficult to extend. It in turn means that all the OSPF routers must be upgraded network-wide to make the new extensions work.

- The new drafts (TE, GMPLS extensions, etc) written for OSPF now support TLV encoding.

OSPFv3

- Exhibits implicit opaque LSA behaviour i.e. unrecognized LSA types are flooded to the neighbors making it more extensible that OSPFv2

- Designed in a way which makes it easily extensible to any other layer 3 protocol suite.

16. MTU Limitations

The MTU of a sub-network is the largest size packet or frame, specified in octets that can be sent over it. Both OSPF and IS-IS require communicating routers to have matching MTU sizes in order to form adjacencies. This is needed so that routers will not advertise packets larger than a neighbor can receive and process. However, each protocol uses a different mechanism to check against MTU mismatch. For this discussion the term MTU is used for a links Maximum Receive Unit (MRU) too.

IS-IS

- IS-IS works over the link layer, which does not provide for fragmentation and reassembly.

- Hello's are sent padded to MTU size till an adjacency comes up. If there is an MTU mismatch, the side having the lesser MTU would drop the bigger than MTU hello. This would not allow adjacencies to be formed between interfaces having different MTU's.

- The hello MTU match is an insufficient condition for IS-IS as LSP's are flooded as is and not packed into any other packets. For the LSP's to be successfully synchronized across the subdomain, all LSP's need to be of a size lesser than the smallest link MTU in the subdomain, else the flooding of the LSP on the link will fail resulting in inconsistent routing tables.

- Mis-configuration of the maximum packet size that a router sends out can cause problems across the subdomain as there is no way to

check the value between routers that are not adjacent.

OSPF

- OSPF works over IP, so the fragmentation and reassembly of any OSPF packet is taken care by the IP layer. However for some link technologies where MTU is configurable but not negotiated, we can have packet black-holes whenever packets larger than the receiving sides MTU are sent.

- The MTU is exchanged in the database description packets. If the value of MTU received in the first DB description packet is greater than that can be accepted on an interface, the packet is rejected and the adjacency is not formed. Retransmissions of DB description packets occur because the packets are never acknowledged. The adjacency therefore gets stuck in EXstart state.

- As LS Update's are assembled in each router, the MTU of another link does not affect the size of the LS Update packet.

- As the MTU match is done at the database exchange state after the DR election has been completed, in case the DR itself cannot form adjacencies with the rest of the routers, it can cause the network to become a stub.

17. Security/Authentication Issues

OSPF: Replay protection/KeyId field

IS-IS: HMAC MD5/checksum not in all PDU's(optional)/ need to dig into
PDU's to find TLV/ LSP's checksum does not cover length field/
purging done with 0 checksum (contents can't be verified)

Both protocols have a field indicating the "type" of authentication. There are however differences in the two protocols. In IS-IS, the data associated with the authentication is a variable length. In OSPF it is fixed at 64 bits. 64 bits is sufficient for a password scheme, but would not suffice for a public key signature scheme, which would need a field several hundreds of bits long.

In OSPF there is a single password per link. A router is configured with a password for each link to which it is attached. It transmits that password when it transmits OSPF messages on that link. It expects all OSPF messages it receives on that link to have that password. In IS-IS, a router is configured with a transmit password on a link, which is the password it uses when it transmits IS-IS messages, as well as a set of acceptable receive passwords.

On a P2P link a password scheme in which the receive and transmit passwords are different offers some security. If the passwords are the same, the intruder need only wait for the other router to transmit first, and the intruder will find out the password. Even with two passwords, an intruder can, with effort, discover the passwords.

The reason IS-IS configures routers with a set of acceptable receive passwords, rather than a single receive password, is so that a link, such as a LAN, can be migrated from one password to another without disrupting the network. Since OSPF has single password per link, it is not possible to change the password in an operational network. The routers would all have to be brought down and locally reconfigured.

One of the brought up issue with IS-IS proponents is apparently the big advantage that IS-IS has over OSPF from a security point of view as IS-IS protocol packets cannot be routed beyond the immediate next hop or can never be sourced by non-border routers. This is claimed, can prevent a variety of potential DoS attacks as anyone can launch OSPF packet bombs in the others network. This apparent vulnerability to DoS attacks is because OSPF rides over IP rather than directly running on the link layer.

Since all OSPF packets can be authenticated using MD5, all spurious OSPF packets can be dropped. But there can be times when MD5 can itself be a part of a problem because it takes significant CPU to check signatures and discard the packets. This is partly true but it is to be noted however that even if OSPF encapsulation is changed to L2, we would still have to support IP encapsulation for virtual links, so we would still have to do MD5.

Moreover the system administrator can filter on the edges of the network to pry away all the OSPF messages coming from the edges. This will of course be done in addition to cryptography.

18. IS-IS/OSPF for IPv6

IS-IS

- Designed to be protocol-agnostic using TLV encoding.

- Distinct TLVs used to encode topology information and reachability (address prefix) information. As a direct consequence, extending ISIS to support IPv6 is just a matter of introducing some new TLVs. The existing TLVs continue to be used to advertise topology information

- An extension to ISIS has been proposed that calculated Ipv4 and IPv6 topologies separately. This would still use a single instance of

ISIS for each network protocol. There are proposals to extend ISIS to

enable multiple instances for each network layer protocol, thereby applying the "Ships in the Night" model for ISIS.

OSPF

- All routing information is advertised using LSAs, which are identified by the LS Type, LS Identifier and the advertising Router.

- Adapting this to support IPv6 was difficult for the following reasons:

Many fields (LS Identifier, the DR/BDR field in the HELLO Message, etc) in the OSPF packets are IPv4 specific. Thus adapting OSPFv2 to support IPv6, which has an expanded address space, becomes impossible.

- OSPFv2 inherits IPv4's "subnet" restriction. Thus an OSPFv2 Router denies to form an adjacency if the neighboring router's IPv4 address does not match the router's IPv4 subnet. Further, OSPFv2 can calculate only one IPv4 prefix for a LAN segment. These "subnet" restrictions were removed in IPv6 specification, which makes OSPFv2 even more difficult to adapt to IPv6.

- Presents a "ship in the night" solution during the IPv6 migration. This means that the operator needs to run OSPFv2 for IPv4 routing and OSPFv3 for IPv6, as against an integrated solution provided by ISIS. If using OSPF, then OSPFv2 and OSPFv3 will independently calculate their network topology, routes, etc. This can lead to some redundancy and duplication when IPv4 network topology is identical to the IPv6 topology. This leads to greater CPU, memory and bandwidth utilization because of double computation and advertisement.

ISIS on the other hand, presents an integrated solution in the presence of IPv4 and IPv6 network protocols. Since ISIS can calculate IPv4 and IPv6 routes simultaneously it is relatively efficient with respect to the utilization of resources.

However, most of the networks deploying IPv4 and IPv6 simultaneously typically have different topologies and IPv4 and IPv6 networks are constructed separately. This avoids a breakdown of one network because of the failure in the other.

OSPFv3

- Instead of putting hacks in OSPFv2 to support IPv6, OSPFv3 (also referred to as "OSPF for IPv6") was laid out by the OSPF WG.

- The packet format was changed, calculation and representation of address prefix information was separated from the topology information.

- OSPFv3 provides native support for opaque LSAs

- Other fundamental mechanisms of OSPF, like database synchronization, etc remain unchanged. The DR/BDR field in the Hello packet described above was simply changed to contain Router-ID of the DR/BDR.

- Extensions have been proposed to adapt OSPFv3 for an "Integrated model" where OSPFv3 would be extended to calculate IPv4 routes

19. Current Deployments

Both the protocols have been currently deployed in large scale IP networks.

IS-IS

- used in most Tier 1 ISP networks and in single area configurations

- initally most large ISPs adopted IS-IS as it had a stable implementation, coupled with U.S. government's mandate to support ISO CLNS alongside IP.

OSPF

- more widespread from medium to large IP networks.

- deployed in most IP based enterprise networks

20. Metrics Size

Each interface in the link state protocols in given a metric, which is advertised with the link state information in LSP/LSA. The SPF algorithm uses this metric to calculate the cost and the nexthop to a destination. Metrics used are generally the inverse of bandwidth. A larger bandwith capacity link would have a lesser metric.

IS-IS

- ISO10589 specifies metric 6 bit in size. Therefore the metric value can range from 0-63. The information is carried in neighbor reachbility TLV and the IP reachability TLV. This is called the Narrow metric. The maximum path metric MaxPathMetric supported is 1023. This in theory brought the complexity of the SPF from O(nlog n) to O(n). But this isn't significant any more as the CPUs are really

fast these days. The metric size was kept small to optimize search

while doing SPF. It also allows two types of metrics External and Internal.

- The Narrow metric range was however found to be too small for certain networks. New TLV's(Extended IP and Extended neighbor reachability TLV's) to carry larger metrics was added as part of the traffic engineering document[IS-IS-TE]. This is called Wide Metrics. The MaxLinkMetric value is 0xFFFFFFand the MaxPathMetric is 0xFE00000.

The Extended IP reachability TLV allows for a 4 byte metric, while the Extended Neighbor reachability TLV allows for 3 bytes metric size.
This is to enable the metric summarized across levels/domains to be as large as 0xFFFFFFFF while the link metric itself is no larger than
0xFFFFFE. If a metric value of 0xFFFFFF is used the prefix is not used in SPF calculations.

- Four kinds of narrow metrics are defined however only the default metric is used in networks.

OSPFv2

- OSPFv2 allows a link to have a 2 byte metric feild in the Router LSA. This implies the maximum metric of 0xFFFF.

- The Summary, Summary-ASBR, AS-External and NSSA LSA's have a 3 byte
metric value. A cost of 0xFFFFFF (LSInfinity) is used to tell the destination described in the LSA is unreachable.

- AS-External and NSSA LSA's allow two metric types, Type-1 and Type-
2 which are equivalent to IS-IS Internal and External metrics. The type 1 considers the cost to the ASBR in addition to the advertised cost of the route while the latter uses just the advertised cost while calculating the routes.

- The scheme thus allows for links to be configured with a metric no larger than 0xFFFF, while allowing cost of destinations injected across areas/levels to be as large as 0xFFFFFE.

OSPFv3

- OSPFv3 allows similar metric size for the Router LSA's as in OSPFv2.

- OSPFv3 allows similar metric sizes for Intra Area Prefix LSA, Inter
Area Prefix LSA, AS-External LSA and NSSA LSA as in OSPFv2. The value
and significance of LS Infinity is valid here.

21. Database Granularity

This section compares how the two protocols hold their routing information in their link state databases. The way these protocols encode the routing information in their database, affects their behavior in how they flood/distribute the change of routing information.

OSPF

- Organization of Routing Information

OSPF encodes the routing information into small chunks, which it calls Link State Advertisement (LSA). Each LSA has its own 20-byte header in order to be identified uniquely. This header is called the LSA Header. There is no limitation on the size of a LSA, though the actual LSA size is limited by IP packet size limitation: 65,535 bytes minus the LSA Header size and IP packet header size. The database access in OSPF is per LSA basis.

In OSPF routing, the information within an area is described by type 1 and type 2 LSAs (known as Router-LSA and Network-LSA respectively). These LSAs can become big depending upon the number of adjacencies to be advertised and prefixes to be carried inside an area. In other words, the routing information with respect to a single node (either router or network node) is encoded inside a single LSA. On the other hand, each inter-area or external prefix is advertised in a separate LSA (AS-External LSA).

An OSPFv2 router may originate only one Router-LSA for itself, while in OSPFv3, a router is allowed to originate multiple Router-LSAs. A router may originate a Network-LSA for each IP subnet on which the router acts as a DR. A router may originate one LSA for each inter-area and external prefix, with no limitations on the number of LSAs that it may originate.

- Consequences

Originating a new and a unique LSA for each inter-area route and an external prefix implies that there is a LSA Header overhead involved while the information is kept in the database or is flooded to the neighbors. There is thus some extra memory and bandwidth consumed in total.

- Carrying Routing Information

LSAs are carried in Link State Update packets (called LS Updates or LSUs). Each LS Update packet has its own header, consists of a 24-byte OSPF protocol header, and a 4-bytes field indicating the number of LSAs contained in the packet. Thus multiple LSAs can be packed

into a single LS Update packet. Some implementations may not do this as its considered difficult achieving this during flooding.

- Consequences

In the face of network changes, OSPF floods only the updated LSAs. Therefore, even if an implementation does not pack multiple LSAs into a single LS Update packet (and so bandwidth is consumed by LS Update header for each update of a single LSA), the bandwidth consumption for each network change can be considered adequately small.

IS-IS

- Organization of the Routing Information

In IS-IS, protocol packets are called Protocol Data Units or PDUs. IS-IS encodes the link state information into the set of Type-Length-Value tuples (called TLVs), and packs these TLVs into one or more Link State PDUs (LSPs). The size limit of a LSP is configurable. The Routing database consists of these PDUs and the access to the database is per PDU basis. The original IS-IS specification places an upper bound on the number of LSPs a router can originate to 255. There are however techniques which enable a router to originate more than 255 LSPs, by using multiple system-id's for itself.

- Consequences

Since routing information in IS-IS for each router is packed in fewer LSPs, the memory consumed for bookkeeping of the routing data within the database is less and is more efficient.

- Carrying Routing Information

Each LSP is flooded independently, without being modified all the way from the originator through the routers till the very end. This results in all the routers having the same LSPs as that originated by the first router.

- Consequences

Since LSPs are not modified in any way and are not allowed to be fragmented, in order to be flooded successfully over all links existing in the IS-IS network, great care must be ensured when configuring the size limit of LSP that routers can originate and receive. [INTEROP] If the size limit of the LSP is set without taking into account the minimum value of the MTUs throughout the network,

or

if the size limit of LSPs conflict among some the routers in the network, the database synchronization may not be achieved, and this can result in routing loops and/or blackholes.

When a change occurs to a LSP, the whole LSP needs to be flooded, and therefore the bandwidth usage can be non-optimal. There is however a solution which exists in theory. If an implementation finds some of the entities to be flapping, then they may be packed into smaller LSPs or may be isolated from the other stable entities. This way one needs to only advertise the unstable LSP/LSPs.

Database granularity also affects when two routers need to synchronize their databases. In OSPF, because of its high database granularity there are a lot of items which it needs to synchronize and that process is somewhat complicated with a lot of DBD packets being exchanged back and forth. This is simpler in case of IS-IS and there isn't any FSM that the neighbors need to go through to synchronize their databases. It just uses it regular flooding mechanism (a couple of CSNPs describe their entire topology information) to exchange its entire database.

22. Separation of TE and topology information

Traffic Engineering (TE) is defined as the aspect of Internet Network Engineering concerned with the performance optimization of traffic handling in operational networks. The Link State Routing protocols transport traffic engineering information reliably by flooding mechanisms, thus helping in TE.

IS-IS

- TE information is carried in Extended IS reachability TLV's which are also used in normal routing table calculations. TE information is carried as subTLV's.

- A new Router-Id TLV is defined for TE purposes.

- The Value field of the TLV length can only be 255 bytes, because of the limitations SRLG is defined in a seperate TLV.

OSPF

- TE extensions information is carried in TE LSA's. A TE LSA is an opaque type-10 LSA [OPAQUE], with the first 8 bits of the LSA-ID field value being 1 and the remaining 24-bits being used for type-specific data [OSPF-TE].

- The payload of the TE LSA consists of TLV's. There are two top level TLV's defined though any LSA can carry only one TLV. The TLV's defined are Router address TLV and Link Address TLV.

- The length of the value field is 16 bits, hence the maximum length of the Value field in the TLV can be 2^16.

- The Router-Id field used for OSPF is used to identify the other end of a point-to-point link. This Router-Id field is the same field used for normal SPF calculations.

23. Convergence and Scalability Issues

IS-IS

- Is limited by the maximum number of LSPs that each IS-IS router can issue. This is 256 as its LSP ID is 1 octet long. The total number of IP prefixes carried by IS-IS can be easily computed which comes to O(31000). For actual calculations refer to the [APPENDIX]

This seems to be a reasonable number for any sane IS-IS deployment and it will not run out of space unless someone actually injects the entire BGP feed into the IGP. In that case we will run out of space at about 20% of the way into redistribution and not be able to advertise the rest. It is for this reason that this practice has now been deprecated and the RFC 1745 which lays down the rules for BGP-OSPF interaction moved to the HISTORICAL status [RFC1745].

- 8 bits are used for defining a pseudonode number in the LSPID which means that a router can be DIS for only 256 LANs. Additionally there is also a limitation on the number of routers that can be advertised in pseudonode LSP of the DIS.

- There is however a recent IETF draft [256LSP] which describes a mechanism that allows an IS-IS router to originate more than 256 LSP fragments and RFC 3373 [3WAY] which proposes a method for new TLV HELLO packets that increase the number of p2p adjacencies.

- The "Remaining lifetime" field which gives the number of seconds before LSP is considered expired is 16 bits wide.

This gives the life time of the LSP as 2^16/60/60 Hrs = 18.7 Hrs

Thus each LSP needs to be refreshed after every 18.7 Hrs.

OSPF

- In theory, OSPF topology is limited by the number of links that can be advertised in the Router LSA as each router gets only one Router LSA and it cant be bigger than 64K which is the biggest an IP packet

can be. The same constraint applies to the Network LSA also.

Each link in the router can take up at most 24 bytes. Thus, number of links which can be supported is given by (64 * 1024) / 24 = 2370

However, if we take the minimum link size per link (12 bytes) then the maximum is about 2 * 2370 = O(5000) links

To be more specific, we can have O(2300) p2p and p2mp links (not considering virtual links, etc) and O(5000) broadcast/NMBA links

Thus each Router LSA can carry some 5000 links information in it. It is hard to imagine a router having 5000 neighbors but there are already routers with 400 neighbors in some ISPs, and it doesn't take long to reach the order of the magnitude limited by OSPF.

- Network LSAs are generated by the DR for each broadcast network it is connected to. To have scaling problems it should have 2730 * 6 times neighbors on that interface. This is even less probable and hence there are no scalability problems with OSPF per se.

- All other LSAs apart from Type 1 and Type 2 hold single prefixes. Because there is no limit to the number of such LSAs, a large number of inter-area and externals can be generated depending upon the memory resources of the router.

- Each LSA has an LS Age field which is counted upwards starting from zero. Its life is an architectural constant which says one hour. When an LSA's LS age field reaches MaxAge, it is reflooded in an attempt to flush the LSA from the routing domain. One hour seems like a long time but if one originates 50,000 LSAs then OSPF will be refreshing on an average of just 36ms

Total number of LSAs to be refreshed = 50,000

Time by which all the LSAs must be refreshed = LSRefreshTime = 30mins = 1800 secs

Rate at which the LSAs need to be refreshed = 1800/50000 = 36ms

However, if the refreshes are perfectly spread out across time and perfectly batched, the actual update transmission rate may be on the order of one packet per second.

There is however a "do-not-age" LSA [DEMAND] which in theory can be pressed into service and which never gets aged. However, such LSAs will be eventually purged from the LS database if they become stale after being held for at least 60 minutes and the originator not reachable for the same period. Moreover it is not backward compatible and if one deploys that in the network today with some routers not

supporting this then the network can really get weird. So there isn't
really much that can be done using these unless the whole network is changed.

Both the routing protocols are scalable and there should not be any scalability issues with any one of them if implemented properly. Both
have similar stability and convergence properties.

24. Area Id Change Functionality

Changing area-id for an area is useful for link state routing protocols in order to merge two areas into one or to split an area into several areas.

IS-IS

- An area address is a variable length quantity.

- An area can have multiple area addresses. Neighboring IS's will not
form an adjacency unless they have a single area address in common. This is quite useful for IP networks that are transitioning from one area address to another, merging two areas into one or even to split an area into several pieces.

- Seamless transition of area addresses for an area is easier in IS-IS, e.g. initially an area can have area adress A, then the set {A, B} and when the new area address B is recognized by all the routers in the area, old area address A can be removed.

OSPF

- In OSPF each area has a single ID, a 4-byte quantity.

- OSPF does not have the ability to merge and split areas dynamically
as IS-IS has, though partitioned backbone can be repaired by using virtual link. But it should be ensured that the area through which virtual link is configured is having full routing information, i.e. it should not be a stub area.

- Area-id can not be changed dynamically in case of OSPF.

25. Backward Compatibility

For a protocol to be extensible, it should have mechanisms to allow changes in the protocol packets, without affecting backward compatibility. OSPFv2, OSPFv3 as well as IS-IS allow for extending the protocol in a backward compatible manner.

IS-IS

- All IS-IS packets contain TLV's. Unrecognized TLV's are ignored or receipt, this allows TLV types to be extended in a backward compatible manner.

- TLV's can signal more information between neighbors than can option bits. It is for this reason IS-IS was able to allow IS-IS for IP extensions without any backward compatibility being lost.

OSPFv2

- OSPFv2 has options bit in the Hello, Database description packets as well as the LSA header filed, which can be used to signal to its capabilities of the neighbor. Any change of capability can be signaled and decision to form adjacency as well as the LSA's to exchange can be based on the option bits

- There are only 8 bits in the options header most of which have already been utilized. To allow for further extensions OSPF allows the LLS option [LLS]. However this is not widely supported in commercial routers.

- Any unrecognized LSA received is dropped. This does not allow new LSA types to be defined and prevents OSPFv2 to be really extensible.

- Some fields in the OSPFv2 packets contain IPv4 specific information. It is for this reason a different protocol for OSPF for IPv6 was required.

OSPFv3

- OSPFv3 also allows options field like OSPFv2, however the options field have been expanded to 24-bits allowing for more options to be signaled. The options have been removed from LSA header and been added into LSA body for Router, Network, Inter-area-router and link LSA.

- OSPFv3 LSA have a flooding scope in the upper three bits of the LSA type field. Unrecognized LSA's are not ignored but flooded based on the flooding scope of the 3 bits. This allows new LSA types to be flooded in the domain

26. Hitless Restart Mechanisms

If the control and forwarding functions in a router can be separated independently, it is possible to maintain a router's data forwarding capability intact while the router's control software is

restarted/reloaded. This functionality is termed as "graceful restart" or "non-stop forwarding".

IS-IS

- Restarting router does not re-compute its own routes until it has achieved database synchronization with its neighbors [GRACE-IS-IS].

- IS-IS uses new type of TLV (restart TLV) in IIH to obtain the graceful restart functionality. Grace period is decided as the minimum of the Remaining times of received IIHs containing a restart TLV with RA bit set.

- During grace period, restarting router does not transmit self-originated LSPs and self-LSPs are not purged or modified. These restrictions are necessary to prevent premature removal of an own LSP and hence churn in other routers.

- Restart mechanism in IS-IS allows to establish adjacency without cycling through the normal operation of adjacency state machine.

- Proper database synchronization is achieved in situations where the neighboring routers of the restarting router do not support the restart TLV.

OSPF

- OSPF routers can play either of two roles during graceful restart - as a restarting router or as a helper neighbor [GRACE-OSPF].

- Restarting OSPF router originates new type of Grace-LSAs (link local Opaque-LSA) specifying the 'grace period'.

- During graceful restart, the restarting router neither originates LSAs with LS types 1-5,7 nor does modify or flush received self-originated LSAs.

- Router as helper neighbor advertises the restarting router in their LSAs as if it were fully adjacent during the grace period and also detects network topology changes.

- OSPF automatically reverts back to standard OSPF restart from graceful restart if topological changes are detected or if one or more of the restarting router's neighbors do not support graceful restart.

27. Demand Circuits

Demand circuits are network segments whose costs vary with usage; charges can be based both on connect time and on bytes/packets transmitted. Examples of demand circuits include ISDN links, X.25 SVCs, dial-up lines,etc. It is thus desirable to use them only for the user traffic and minimal control traffic.

IS-IS

- ISO 10589 provides very limited support for demand circuits called "dynamically assigned circuits" wherein it supports sending data traffic over them, but does not support running the routing protocol over them. Thus there are no HELLO suppression/DNA schemes in IS-IS for such circuits.

OSPF

- A new optional capability is described in RFC 1793 which modifies OSPF for supporting such circuits. In this, a router will set the DC bit in the options field if it supports this capability. Routers that support the capability will also set the high bit (known as the do-not-age bit), of the LS age field to indicating that the LSA should not be aged. OSPF running on such circuits suppresses periodic HELLOs and LSAs, but a topology change will still activate the demand circuit since LSA updates will be sent which are required to keep the LS database accurate [DEMAND].

- Demand circuits are generally defined in stub areas which have limited topology database thus shielding them from frequent topology changes.

- There is however a problem in detecting inactive OSPF neighbors over such links as HELLO exchange is suppressed on these circuits. To work out a solution for this there are solutions suggested in a recent IETF draft [PROBE] which addresses this problem by the using neighbor probing" mechanisms.

28. IANA Considerations

This document introduces no new security concerns to either of the specifications referenced in this document.

29. References

[OSPF]
J. Moy, "OSPF Version 2", RFC 2328, April 1998

[OSPFv3]

R. Coltun, D. Ferguson and J. Moy, "OSPF for IPv6", RFC 2740, December 1999

[MARTEY]
A. Martey, "IS-IS Network Design Solutions", CISCO Publications, February 2002

[Moy]
John T. Moy, "OSPF: Anatomy of an Internet Routing Protocol", Addison Wesley, February 1998

[MESH]
R. Balay, D. Katz and J. Parker, "IS-IS Mesh Groups", RFC 2973, October 2000

[ARP]
D. C. Plummer, "Ethernet Address Resolution Protocol: or Converting Network Protocol Addresses to 48.bit Ethernet Addresses for Transmission on Ethernet Hardware", RFC 826, November 1982

[PPP]
W. Simpson, "The Point-to-Point Protocol (PPP)", RFC 1661, July 1994

[OOB]
A. Zinin, A. Roy and L. Nyugen, "OSPF Out-of-band LSDB resynchronization", Work in Progress

[TUNNEL]
S. Mirtorabi, P. Psenak, "OSPF Tunnel Adjacency", Work in Progress

[RADIA]
R. Perlman, "A comparision between two routing protocols: OSPF and IS-IS", IEEE Network, vol. 5, no. 5, pp. 18, 24, September 1991

[OPAQUE]
R. Coltun, "The OSPF Opaque LSA Option", RFC 2370, July 1998

[OSPF-TE]
D. Katz, K. Kompella and D. Yeung, "Traffic Engineering Extensions to OSPF Version 2", RFC 3630, September 2003

[INTER-OP]
J. Parker, "Recommendations for Interoperable Networks using Intermediate System to Intermediate System (IS-IS)", RFC 3719, February 2004

[IS-IS-TE]
H. Smit and T. Li, "Intermediate System to Intermediate System (IS-IS) Extensions for Traffic Engineering (TE)", RFC 3784, June 2004

[256LSP]
A. Hermelin, S. Previdi and M. Shand, "Extending the Number of Intermediate System to Intermediate System (IS-IS) Link State PDU (LSP) Fragments Beyond the 256 Limit", RFC 3786, May 2004

[3WAY]
D. Katz and R. Saluja, "Three-Way Handshake for Intermediate System to Intermediate System (IS-IS) Point-to-Point Adjacencies", RFC 3373, September 2002

[RFC 1745]
K. Varadhan, S. Hares and Y. Rekhter, "BGP4/IDRP for IP---OSPF Interaction", RFC 1745, December 1994

[LLS]
A. Zinin, B. Friedman, A. Roy, L. Nguyen and D. Yeung, "OSPF Link-local Signaling", Work in Progress

[GRACE-IS-IS]
M. Shand and L. Ginsberg, "Restart Signaling for Intermediate System to Intermediate System (IS-IS)", RFC 3847, July 2004

[GRACE-OSPF]
J. Moy, P. Pillay-Esnault and A. Lindem, "Graceful OSPF Restart", RFC 3623, November 2003

[DEMAND]
J. Moy, "Extending OSPF to Support Demand Circuits", RFC 1793, April 1995

[PROBE]
S. Rao, A. Zinin and A. Roy, "Detecting Inactive Neighbors over OSPF Demand Circuits (DC)", RFC 3883, October 2004

30. Author's Addresses

Vishwas Manral
SiNett Corp,
Embassy Icon Annexe,
2/1, Infantry Road,
Bangalore, India

Email: vishwas@sinett.com

Manav Bhatia
Riverstone Networks,
3/1, J.P. Techno Parks,

Millers Road,
Bangalore, India

Email: manav@riverstonenet.com

Yasuhiro Ohara
Keio University, Shonan Fujisawa Campus
5322 Endo, Fujisawa
Kanagawa, Japan 252-8520

Phone: +81-(0)466-47-5111

Email: yasu@sfc.wide.ad.jp

31. Appendix

The maximum size of an LSP is 1492 bytes.

Available space = 1492 - 27 (Header) = 1465 bytes for TLVs.

Thus an IS-IS router has theoretically up to 256*1465 of space to pack IP reachability TLVs.

The following calculation enables us to determine the number of IP prefixes that can be advertised in an LSP.

The following constraints are to be considered in the calculation:

The maximum size (maxLSPsize) of an LSP is 1492 bytes.
The LSP header (lspHeadersize) is 27 bytes.
The maximum length of a TLV (maxTLVlength) is 255 bytes.

Each TLV 128 consists of type (1 byte), length (1 byte), and IP prefixes (n x 12 bytes) up to total of 255 bytes. The maximum number of fragments of an LSP (maxLSPfragments) is 256.

The number of fragments is determined from the 1-byte LSP Number field in the LSP identifier.

The first fragment contains other TLVs, and the remaining 255 fragments are packed with only TLV 128.

The actual calculation is as follows:

The total space available for TLVs in an LSP is
TLVSpace = maxLSPsize - lspHeadersize = 1492 - 27 = 1465 bytes

The number of TLVs that can fit into TLVSpace is 1465/255 = 5.7, approximately 6

Assuming a 1-byte Type field and 1-byte Length field, overhead for 6 TLVs is 6 x 2 = 12 bytes.

Actual space available for prefixes is 1465 - 12 bytes overhead = 1453 bytes

Number of prefixes, each 12 bytes (address + subnet mask + metric) that can fit into TLVSpace is 1453/12 = 121.08 (approximately 121 IP prefixes per LSP)

Considering that few other TLVs can be generated by the router, the number of IP prefixes that can be supported per IS-IS router is 256 fragments, each containing 121 prefixes, for a total of 30,976 prefixes.

32. Intellectual Property Notice

33. Disclaimer of Validity

ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE

INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

34. Full Copyright Notice

35. Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.