Network Working Group
Request for Comments: 2966
Category: Informational

T. Li
Procket Networks
T. Przygienda
Redback
H. Smit
Procket Networks

October 2000

Domain-wide Prefix Distribution with Two-Level IS-IS

Status of this Memo

Copyright Notice

Abstract

This document describes extensions to the Intermediate System to Intermediate System (IS-IS) protocol to support optimal routing within a two-level domain. The IS-IS protocol is specified in ISO 10589, with extensions for supporting IPv4 (Internet Protocol) specified in RFC 1195 [2].

This document extends the semantics presented in RFC 1195 so that a routing domain running with both level 1 and level 2 Intermediate Systems (IS) [routers] can distribute IP prefixes between level 1 and level 2 and vice versa. This distribution requires certain restrictions to insure that persistent forwarding loops do not form. The goal of this domain-wide prefix distribution is to increase the granularity of the routing information within the domain.

1. Introduction

An IS-IS routing domain (a.k.a., an autonomous system running IS-IS) can be partitioned into multiple level 1 (L1) areas, and a level 2 (L2) connected subset of the topology that interconnects all of the L1 areas. Within each L1 area, all routers exchange link state information. L2 routers also exchange L2 link state information to compute routes between areas.

RFC 1195 [2] defines the Type, Length and Value (TLV) tuples that are used to transport IPv4 routing information in IS-IS. RFC 1195 also specifies the semantics and procedures for interactions between levels. Specifically, routers in a L1 area will exchange information within the L1 area. For IP destinations not found in the prefixes in the L1 database, the L1 router should forward packets to the nearest router that is in both L1 and L2 (i.e., an L1L2 router) with the "attached bit" set in its L1 Link State Protocol Data Unit (LSP).

Also per RFC 1195, an L1L2 router should be manually configured with a set of prefixes that summarizes the IP prefixes reachable in that L1 area. These summaries are injected into L2. RFC 1195 specifies no further interactions between L1 and L2 for IPv4 prefixes.

1.1 Motivations for domain-wide prefix distribution

The mechanisms specified in RFC 1195 are appropriate in many situations, and lead to excellent scalability properties. However, in certain circumstances, the domain administrator may wish to sacrifice some amount of scalability and distribute more specific information than is described by RFC 1195. This section discusses the various reasons why the domain administrator may wish to make such a tradeoff.

One major reason for distributing more prefix information is to improve the quality of the resulting routes. A well know property of prefix summarization or any abstraction mechanism is that it necessarily results in a loss of information. This loss of information in turn results in the computation of a route based upon less information, which will frequently result in routes that are not optimal.

A simple example can serve to demonstrate this adequately. Suppose that a L1 area has two L1L2 routers that both advertise a single summary of all prefixes within the L1 area. To reach a destination inside the L1 area, any other L2 router is going to compute the shortest path to one of the two L1L2 routers for that area. Suppose, for example, that both of the L1L2 routers are equidistant from the L2 source, and that the L2 source arbitrarily selects one L1L2 router. This router may not be the optimal router when viewed from the L1 topology. In fact, it may be the case that the path from the selected L1L2 router to the destination router may traverse the L1L2 router that was not selected. If more detailed topological information or more detailed metric information was available to the L2 source router, it could make a more optimal route computation.

This situation is symmetric in that an L1 router has no information about prefixes in L2 or within a different L1 area. In using the nearest L1L2 router, that L1L2 is effectively injecting a default route without metric information into the L1 area. The route computation that the L1 router performs is similarly suboptimal.

Besides the optimality of the routes computed, there are two other significant drivers for the domain wide distribution of prefix information.

When a router learns multiple possible paths to external destinations via BGP, it will select only one of those routes to be installed in the forwarding table. One of the factors in the BGP route selection is the IGP cost to the BGP next hop address. Many ISP networks depend on this technique, which is known as "shortest exit routing". If a L1 router does not know the exact IGP metric to all BGP speakers in other L1 areas, it cannot do effective shortest exit routing.

The third driver is the current practice of using the IGP (IS-IS) metric as part of the BGP Multi-Exit Discriminator (MED). The value in the MED is advertised to other domains and is used to inform other domains of the optimal entry point into the current domain. Current practice is to take the IS-IS metric and insert it as the MED value. This tends to cause external traffic to enter the domain at the point closest to the exit router. Note that the receiving domain may, based upon policy, choose to ignore the MED that is advertised. However, current practice is to distribute the IGP metric in this way in order to optimize routing wherever possible. This is possible in current networks that only are a single area, but becomes problematic if hierarchy is to be installed into the network. This is again because the loss of end-to-end metric information means that the MED value will not reflect the true distance across the advertising domain. Full distribution of prefix information within the domain would alleviate this problem as it would allow accurate computation of the IS-IS metric across the domain, resulting in an accurate value presented in the MED.

1.2 Scalability

The disadvantage to performing the domain-wide prefix distribution described above is that it has an impact to the scalability of IS-IS. Areas within IS-IS help scalability in that LSPs are contained within a single area. This limits the size of the link state database, that in turn limits the complexity of the shortest path computation.

Further, the summarization of the prefix information aids scalability in that the abstraction of the prefix information removes the sheer number of data items to be transported and the number of routes to be computed.

It should be noted quite strongly that the distribution of prefixes on a domain wide basis impacts the scalability of IS-IS in the second respect. It will increase the number of prefixes throughout the domain. This will result in increased memory consumption, transmission requirements and computation requirements throughout the domain.

It must also be noted that the domain-wide distribution of prefixes has no effect whatsoever on the first aspect of scalability, namely the existence of areas and the limitation of the distribution of the link state database.

Thus, the net result is that the introduction of domain-wide prefix distribution into a formerly flat, single area network is a clear benefit to the scalability of that network. However, it is a compromise and does not provide the maximum scalability available with IS-IS. Domains that choose to make use of this facility should be aware of the tradeoff that they are making between scalability and optimality and provision and monitor their networks accordingly. Normal provisioning guidelines that would apply to a fully hierarchical deployment of IS-IS will not apply to this type of configuration.

2. Proposed syntax and semantics for L2->L1 inter-area routes

This document defines the syntax of how to advertise level 2 routes in level 1 LSPs. The encoding is an extension of the encoding in RFC 1195.

To some extent, in IS-IS the level 2 backbone can be seen as a separate area itself. RFC 1195 defines that L1L2 routers can advertise IP routes that were learned via L1 routing into L2. These routes can be regarded as inter-area routes. RFC 1195 defines that these L1->L2 inter-area routes must be advertised in L2 LSPs in the "IP Internal Reachability Information" TLV (TLV 128). Intra-area L2 routes are also advertised in L2 LSPs in an "IP Internal Reachability Information" TLV. Therefore, L1->L2 inter-area routes are indistinguishable from L2 intra-area routes.

RFC 1195 does not define L2->L1 inter-area routes. A simple extension would be to allow a L1L2 router to advertise routes learned via L2 routing in its L1 LSP. However, to prevent routing-loops, L1L2 routers must never advertise L2->L1 inter-area routes that they

learn via L1 routing, back into L2. Therefore, there must be a way to distinguish L2->L1 inter-area routes from L1 intra-area routes. Draft-ietf-isis-traffic-01.txt defines the "up/down bit" for this purpose. RFC 1195 defines TLVs 128 and 130 to contain IP routes. TVLs 128 and 130 have a metric field that consists of 4 TOS metrics. The first metric, the so-called "default metric", has the high-order bit reserved (bit 8). Routers must set this bit to zero on transmission, and ignore it on receipt.

This document redefines this high-order bit in the default metric field in TLVs 128 and 130 to be the up/down bit. L1L2 routers must set this bit to one for prefixes that are derived from L2 routing and are advertised into L1 LSPs. The bit must be set to zero for all other IP prefixes in L1 or L2 LSPs. Prefixes with the up/down bit set that are learned via L1 routing, must never be advertised by L1L2 routers back into L2.

2.1 Clarification of external route-type and external metric-type

RFC 1195 defines two TLVs for carrying IP prefixes. TLV 128 is defined as "IP Internal Reachability Information", and should be used to carry IP prefixes that are directly connected to IS-IS routers. TLV 130 is defined as "IP External Reachability Information", and should be used to carry routes learned from outside the IS-IS domain. RFC 1195 documents TLV type 130 only for level 2 LSPs.

RFC 1195 also defines two types of metrics. Metrics of the internal metric-type should be used when the metric is comparable to metrics used to weigh links inside the ISIS domain. Metrics of the external metric-type should be used if the metric of an IP prefix cannot be directly compared to internal metrics. External metric-type can only be used for external IP prefixes. A direct result is that metrics of external metric-type should never be seen in TLV 128.

To prevent confusion, this document states again that when a router computes IP routes, it must give the same preference to IP routes advertised in an "IP Internal Reachability Information" TLV and IP routes advertised in an "IP External Reachability Information" TLV. RFC 1195 states this quite clearly in the note in paragraph 3.10.2, item 2c). This document does not alter this rule of preference.

> NOTE: Internal routes (routes to destinations announced in the "IP Internal Reachability Information" field), and external routes using internal metrics (routes to destinations announced in the "IP External Reachability Information" field, with a metric of type "internal") are treated identically for the purpose of the order of preference of routes, and the Dijkstra calculation.

However, IP routes advertised in "IP External Reachability Information" with external metric-type must be given less preference than the same IP routes advertised with internal-metric type, regardless of the value of the metrics.

While IS-IS routers must not give different preference to IP prefixes learned via "IP Internal Reachability Information" and "IP External Reachability Information" when executing the Dijkstra calculation, routers that implement multiple IGPs are free to use this distinction between internal and external routes when comparing routes derived from different IGPs for inclusion in their global RIB.

2.2 Definition of external IP prefixes in level 1 LSPs

RFC 1195 does not define the "IP External Reachability Information" TLV for L1 LSPs. However, there is no reason why an IS-IS implementation could not allow for redistribution of external routes into L1. Some IS-IS implementations already allow network administrators to do this. This document loosens the restrictions in RFC 1195, and allows for the inclusion of the "IP External Reachability Information" TLV in L1 LSPs.

RFC 1195 defines that IP routes learned via L1 routing must always be advertised in L2 LSPs in a "IP Internal Reachability Information" TLV. Now that this document allows "IP External Reachability Information" TLVs in L1 LSPs, and allows for the advertisement of routes learned via L2 routing into L1, the above rule needs a extensions.

When a L1L2 router advertises a L1 route into L2, where that L1 route was learned via a prefix advertised in a "IP External Reachability Information" TLV, that L1L2 router should advertise that prefix in its L2 LSP within an "IP External Reachability Information" TLV. L1 routes learned via an "IP Internal Reachability Information" TLV should still be advertised within a "IP Internal Reachability Information" TLV. These rules should also be applied when advertising IP routes derived from L2 routing into L1. Of course in this case also the up/down bit must be set.

RFC 1195 defines that if a router sees the same external prefix advertised by two or more routers with the same external metric, it must select the route that is advertised by the router that is closest to itself. It should be noted that now that external routes can be advertised from L1 into L2, and vice versa, that the router that advertises an external prefix in its LSP might not be the router that originally injected this prefix into the IS-IS domain. Therefore, it is less useful to advertise external routes with external metrics into other levels.

3. Types of IP routes in IS-IS and their order of preference

RFC 1195 and this document defines several ways of advertising IP routes in IS-IS. There are four variables involved.

1) The level of the LSP in which the route is advertised. There are currently two possible values: level 1 and level 2
2) The route-type, which can be derived from the type of TLV in which the prefix is advertised. Internal routes are advertised in IP Internal Reachability Information TLVs (TLV 128), and external routes are advertised in IP External Reachability Information TLVs (TLV 130).
3) The metric-type: Internal or External. The metric-type is derived from the Internal/External metric-type bit in the metric field (bit 7).
4) The fact whether this route is leaked down in the hierarchy, and thus can not be advertised back up. This information can be derived from the newly defined up/down bit in the default metric field.

3.1 Overview of all types of IP prefixes in IS-IS Link State PDUs

The combination IP Internal Reachability Information and external metric-type is not allowed. Also the up/down bit is never set in L2 LSPs. This leaves us with 8 different types of IP advertisements in IS-IS. However, there are more than 8 reasons for IP prefixes to be advertised in IS-IS. The following tables describe the types of IP prefixes and how they are encoded.

1) L1 intra-area routes

These are advertised in L1 LSPs, in TLV 128.
The up/down bit is set to zero, metric-type is internal metric.
These IP prefixes are directly connected to the advertising router.

2) L1 external routes

These are advertised in L1 LSPs, in TLV 130.
The up/down bit is set to zero, metric-type is internal metric.
These IP prefixes are learned from other IGPs, and are usually not directly connected to the advertising router.

3) L2 intra-area routes

These are advertised in L2 LSPs, in TLV 128.
The up/down bit is set to zero, metric-type is internal metric.
These IP prefixes are directly connected to the advertising router.
These prefixes can not be distinguished from L1->L2 inter-area routes.

4) L2 external routes

These are advertised in L2 LSPs, in TLV 130.
The up/down bit is set to zero, metric-type is internal metric.
These IP prefixes are learned from other IGPs, and are usually not directly connected to the advertising router. These prefixes can not be distinguished from L1->L2 inter-area external routes.

5) L1->L2 inter-area routes

These are advertised in L2 LSPs, in TLV 128.
The up/down bit is set to zero, metric-type is internal metric.
These IP prefixes are learned via L1 routing, and were derived during the L1 SPF computation from prefixes advertised in L1 LSPs in TLV 128. These prefixes can not be distinguished from L2 intra-area routes.

6) L1->L2 inter-area external routes

These are advertised in L2 LSPs, in TLV 130.
The up/down bit is set to zero, metric-type is internal metric.
These IP prefixes are learned via L1 routing, and were derived during the L1 SPF computation from prefixes advertised in L1 LSPs in TLV 130. These prefixes can not be distinguished from L2 external routes.

7) L2->L1 inter-area routes

These are advertised in L1 LSPs, in TLV 128.
The up/down bit is set to one, metric-type is internal metric.
These IP prefixes are learned via L2 routing, and were derived during the L2 SPF computation from prefixes advertised in TLV 128.

8) L2->L1 inter-area external routes

These are advertised in L1 LSPs, in TLV 130.
The up/down bit is set to one, metric-type is internal metric.
These IP prefixes are learned via L2 routing, and were derived during the L2 SPF computation from prefixes advertised in L2 LSPs in TLV 130.

9) L1 external routes with external metric

These are advertised in L1 LSPs, in TLV 130.
The up/down bit is set to zero, metric-type is external metric.
These IP prefixes are learned from other IGPs, and are usually not directly connected to the advertising router.

10) L2 external routes with external metric

These are advertised in L2 LSPs, in TLV 130.
The up/down bit is set to zero, metric-type is external metric.
These IP prefixes are learned from other IGPs, and are usually not directly connected to the advertising router. These prefixes can not be distinguished from L1->L2 inter-area external routes with external metric.

11) L1->L2 inter-area external routes with external metric

These are advertised in L2 LSPs, in TLV 130.
The up/down bit is set to zero, metric-type is external metric.
These IP prefixes are learned via L1 routing, and were derived during the L1 SPF computation from prefixes advertised in L1 LSPs in TLV 130 with external metrics. These prefixes can not be distinguished from L2 external routes with external metric.

12) L2->L1 inter-area external routes with external metric

These are advertised in L1 LSPs, in TLV 130.
The up/down bit is set to one, metric-type is external metric.
These IP prefixes are learned via L2 routing, and were derived during the L1 SPF computation from prefixes advertised in L2 LSPs in TLV 130 with external metrics.

3.2 Order of preference for all types of IP routes in IS-IS

Unfortunately IS-IS cannot depend on metrics alone for route selection. Some types of routes must always preferred over others, regardless of the costs that were computed in the Dijkstra calculation. One of the reasons for this is that inter-area routes can only be advertised with a maximum metric of 63. Another reason is that this maximum value of 63 does not mean infinity (e.g. like a hop count of 16 in RIP denotes unreachable). Introducing a value for infinity cost in IS-IS inter-area routes would introduce counting-to-infinity behavior via two or more L1L2 routers, which would have a bad impact on network stability.

The order of preference of IP routes in IS-IS is based on a few assumptions.

- RFC 1195 defines that routes derived from L1 routing are preferred over routes derived from L2 routing.
- The note in RFC 1195 paragraph 3.10.2, item 2c) defines that internal routes with internal metric-type and external prefixes with internal metric-type have the same preference.
- RFC 1195 defines that external routes with internal metric-type are preferred over external routes with external metric type.
- Routes derived from L2 routing are preferred over L2->L1 routes derived from L1 routing.

Based on these assumptions, this document defines the following route preferences.

1) L1 intra-area routes with internal metric
   L1 external routes with internal metric
2) L2 intra-area routes with internal metric
   L2 external routes with internal metric
   L1->L2 inter-area routes with internal metric
   L1->L2 inter-area external routes with internal metric
3) L2->L1 inter-area routes with internal metric
   L2->L1 inter-area external routes with internal metric
4) L1 external routes with external metric
5) L2 external routes with external metric
   L1->L2 inter-area external routes with external metric
6) L2->L1 inter-area external routes with external metric

3.3 Additional notes on what prefixes to accept or advertise

Paragraphs 4.1 and 4.2 enumerate all used IP route types in IS-IS. Besides these defined route types, the encoding used would allow for a few more potential combinations. One of them is the combination of "IP Internal Reachability Information" and external metric type. This combination should never be used when building an LSP. Upon receipt of an IP prefix with this combination, routers must ignore this prefix.

Another issue would be the usage of the up/down bit in L2 LSPs. Because IS-IS is currently defined with two levels of hierarchy, there should never be a need to set the up/down bit in L2 LSPs. However, if IS-IS would ever be extended with more than two levels of hierarchy, L2-only (or L1L2) routers will need to be able to accept L2 IP routes with the up/down bit set. Therefore, it is recommended that implementations ignore the up/down bit in L2 LSPs, and accept the prefixes in L2 LSPs regardless whether the up/down bit is set. This will allow for simpler migration once more than two levels of hierarchy are defined.

Another detail that implementors should be aware of is the fact that L1L2 routers should only advertise in their L2 LSP those L1 routes that they use for forwarding themselves. They should not unconditionally advertise into L2 all prefixes from LSPs in the L1 database.

Not all prefixes need to be advertised up or down the hierarchy. Implementations might allow for additional manual filtering or summarization to further bring down the number of inter-area prefixes they advertise in their LSPs. It is also recommended that the default configuration of L1L2 routers is to not advertise any L2 routes into L1 (see also paragraph 5.0).

4. Inter-operability with older implementations

The solution in this document is not fully compatible with RFC 1195. It is an extension to RFC 1195. If routers do not use the new functionality of external L1 routes, nor L2->L1 inter-area routes, older implementations that strictly follow RFC 1195 will be compatible with newer implementations that follow this document.

Implementations that do not accept the "IP External Reachability Information" TLV in L1 LSPs will not be able to compute external L1 routes. This could cause routing loops between L1-only routers that do understand external L1 routes for a particular destination, and L1-only routers that use the default route pointing the closest attached L1L2 router for that destination.

Implementations that follow RFC 1195 should ignore bit 8 in the default metric field when computing routes. Therefore, even older implementations that do not know of the up/down bit should be able to accept the new L2->L1 inter-area routes. These older implementations will install the new L2->L1 inter-area routes as L1 intra-area routes, but that in itself does not cause routing loops among L1-only routers.

However, it is vital that the up/down bit is recognized by L1L2 routers. As has been stated before, L1L2 routers must never advertise L2->L1 inter-area routes back into L2. Therefore, if L2 routes are advertised down into L1 area, it is required that all L1L2 routers in that area run software that understands the new up/down bit. Older implementations that follow RFC 1195 and do not understand the new up/down bit will threat the L2->L1 inter-area routes as L1 intra-area routes, and they will advertise these routes back into L2. This can cause routing loops, sub-optimal routing or extra routing instability. For this reason it is recommended that

implementations by default do not advertise any L2 routes into L1. Implementations should force the network administrator to manually configure L1L2 routers to advertise any L2 routes into L1.

5. Comparisons with other proposals

In [3], a new TLV is defined to transport IP prefix information. This TLV format also defines an up/down bit to allow for L2->L1 inter-area routes. [3] also defines a new TLV to describe links. Both TLVs have wider metric space, and have the possibility to define sub-TLVs to advertise extra information belonging to the link or prefix. The wider metric space in IP prefix TLVs allows for more granular metric information about inter-area path costs. To make full use of the wider metric space, network administrators must deploy both new TLVs at the same time.

Deployment of [3] requires an upgrade of all routers in the network and a transition to the new TLVs. Such a network-wide upgrade and transition might not be an easy task. In this case, the solution defined in this document, which requires only an upgrade of L1L2 routers in selected areas, might be a good alternative to the solution defined in [3].

6. Security Considerations

This document raises no new security issues for IS-IS.

7. References

[1] ISO 10589, "Intermediate System to Intermediate System Intra-Domain Routing Exchange Protocol for use in Conjunction with the Protocol for Providing the Connectionless-mode Network Service (ISO 8473)". [Also republished as RFC 1142.]

[2] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", RFC 1195, December 1990.

[3] Smit, H. and T. Li, "IS-IS Extensions for Traffic Engineering", Work in Progress.

8. Authors' Addresses

Tony Li
Procket Networks
1100 Cadillac Court
Milpitas, CA 95035-3025

EMail: tli@procket.com

Tony Przygienda
Redback
350 Holger Way
San Jose, CA 95134

EMail: prz@redback.com

Henk Smit
Procket Networks
1100 Cadillac Court
Milpitas, CA 95035-3025

EMail: henk@procket.com

9. Full Copyright Statement

Acknowledgement

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