Internet Engineering Task Force
INTERNET DRAFT

A. Patel/T. Przygienda
Bell Labs
15 February 1999

L1/L2 Optimal IS-IS Routing
<draft-ietf-isis-l1l2-00.txt>

Abstract

This draft describes an optional extension within IS-IS [Cal90a, Cal90b, ISO90] for leaking level 2 IP prefixes into level 1. IS-IS is an interior gateway routing protocol developed originally by OSI and used with IP extensions as IGP. This draft describes how to allow for optimal routing in L1/L2 per destination prefix and to support BGP [RL95] MEDs derived from level 1 and level 2 IGP metric when using ISIS.

1. Introduction

In IS-IS extensions described in RFC 1195 [Cal90b] all level 1 routers are equivalent to ``stub'' routers which translates into the fact that no level 2 routes are being leaked actively into level 1. Globally optimal routing across levels is hard since routers in

level 1 are forced to route to the closest level 2 gateway due to the lack of more specific information than just the default route. For scalability and management reasons it is preferable to divide the topology into level 1 and level 2 from a certain size on. With the extension proposed, globally optimal routing is possible that does route per destination prefix to the appropriate level 2 gateway. Moreover, beside the scalability and optimality reasons, given a case where an ISP desires to advertise MEDs to their customer based on IGP metric to BGP next hops [LMJ99], it is not possible or at least misleading to use today's metrics. The metric consists of the cost to traverse the area to the closest level 2 router and a default level 2 cost which is inaccurate. This documents proposes to use existing TLVs and extended processing rules to allow for routing where such cost is adequately computed.

It is important for the understanding of this draft to properly differentiate between level 1 routes, level 2 routes, level 2 external routes with internal metrics (1) and level 2 external routes with external metrics.

2. Description

We extend the usual preference within IS-IS with a new type of routes called level 1 external route which is not defined within RFC 1195 [Cal90b]. To advertise such routes, as an optional capability described in this RFC, IP external reachability information (TLV 130) is allowed within level 1 TLVs. Level 2 routes or level 2 external routes with internal metric are leaked using internal metric translated into an external level 1 metric. Level 2 external routes with external metrics MUST NOT be leaked. (TLV 130) with internal metrics in level 1 are undefined and MUST be ignored.
At this point, we introduce a simple topology in Figure 1 to discuss scenarios encountered. Lines between routers indicate physical point-to-point networks. RT1 is a level-1 router deploying the proposed extension. RT2 is a conventional level-1 router. RT7 is a level-2-only router. RT3, RT4 and RT6 are level-1-2 routers and also participate in level-2 to level-1 leaking. Links between RT3 and RT4 and RT0 and RT3 are level-2-only links. Naturally, links between RT4

---

1. which is equivalent to a (TLV 130) in level-2 with the I/E bit not set.

and RT7 and RT6 are level-2-only as well. A cost for each physical point-to-point network is being assumed as having the cost of 1, except between RT3 and RT2 having a cost of 3. On RT0, RT3 and RT4 externally derived data (e.g., BGP-learned routes) are leaked into level-2 as an external route with internal cost of 1. Therefore, network N-8 will be present in RT3's level-2 LSP as external route with internal cost of 1 and within the level-1 as external route with external cost of 1.

An implementation that does not supports the proposed extension and receives such a TLV MAY ignore it. Traditional RFC 1195 [Cal90b] implementations ignoring this TLV can form routing loops if deployed in a level-1-only domain mixed with level-1-only routers supporting this capability. This happens since routers can disagree on the best possible level-2 gateway for a destination for which no level-1 internal route exists. No routing loops can be formed if traditional RFC 1195 routers are run in level-1-2 or level-2 only mixed with routers deploying the proposed capability.
To see why routing loops in mixed level-1 deployment are possible, consider RT1 that sees an level 1 external route for N-8 with external cost 1 from RT3 and cost 2 from RT0. To route towards N-8 it will choose RT2 as its next hop. RT2 does not understand level 1 external routes and will therefore try to forward towards the closest level-2 gateway, which happens to be RT0.

An implementation that supports handling of the (TLV 130) at level 1 MUST not leak level 1 external prefixes into level 2 since otherwise persistent routing loops are possible if metrics conversions are not executed carefully. To understand why level 1 external prefixes must not be leaked into level 2 consider again the simple topology given in Figure 1. We assume that RT4 leaks N-8 as level 1 external with external cost of 2 to RT5. If RT6 does not leak N-8 into level-1 but would re-advertise level 1 N-8 again into level 2 as external with internal cost of less or equal to 3, RT4 would form a persistent routing loop (2)

---

2. **it would be theoretically possible to leak level 1 external routes (that always have internal metric) into level** 2 external routes with external metrics.

3. Order of Preference of Routes

In order to ensure correct inter-operation of different implementations, it is necessary to specify the order of preference of routes in the forwarding decision which is an extension of the one used today.
For routers participating in level 1 and level 2 and leaking level 2 into level 1, the routes are preferred in the following sequence:

1. Amongst all routes, if the specified destination address matches more than one [IP address, subnet mask] pair, then the most specific address match (the one with more "1" bits in the mask) is preferred.

2. Among the routes with equal address match the preference is determined by the type in the following sequence:

   - level 1

   - level 2

   - level 2 external with internal metric type

   - level 1 external with external metric type (3)

   - level 2 external with external metric type

3. Amongst routes of the same type with equal cost, multi-path load balancing may be performed.

To visualize the concept, consider again Figure 1. After ISIS converges, RT4 will see following entries for network N-8.

- level-2-external route with internal metric 2 with next-hop towards RT3 generated by RT3.

- level-1-external route with next-hop towards RT5 with external metric 3 obtained from RT6 that leaked it from within L2 into L1 domain.

---

3. observe again that level 1 external with internal metric are not allowed.

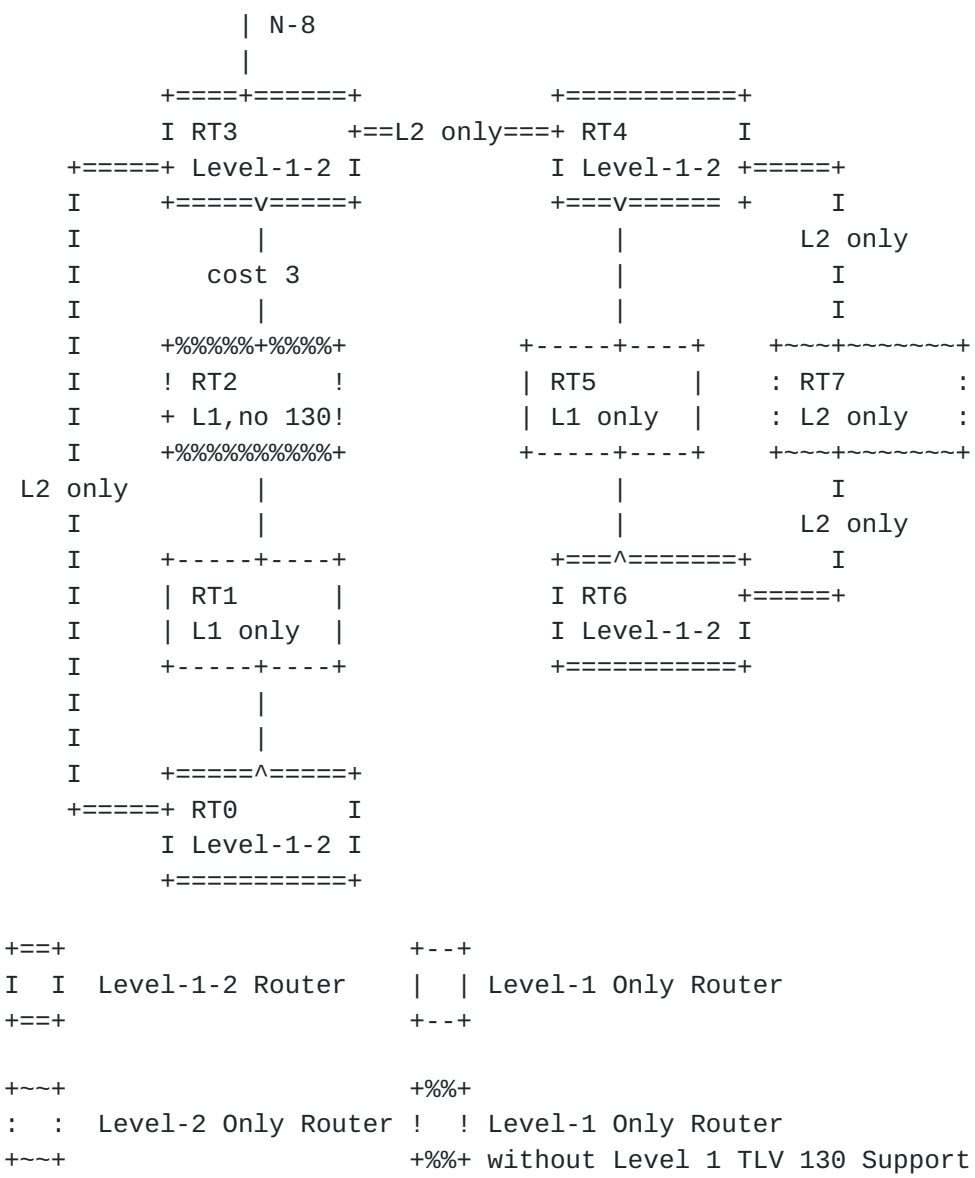

Figure 1: Topology used in our examples

From those entries, route towards N-8 must be chosen according to preferences specified above. Following the rules, level 2 external route through RT3 with internal metric 2 must be preferred, otherwise

a stable loop through RT6 would exist if e.g. level-1 external would be given preference over level-2 external route with internal metric.

4. Acknowledgments

Rohit Dube reviewed the draft carefully and helped to clarify it.

5. Security Consideration

ISIS security applies to the work presented. No specific security issues with the proposed solutions are known.

References

[Cal90a] R. Callon. OSI ISIS Intradomain Routing Protocol. INTERNET-RFC, Internet Engineering Task Force, February 1990.

[Cal90b] R. Callon. Use of OSI ISIS for Routing in TCP/IP and Dual Environments. INTERNET-RFC, Internet Engineering Task Force, December 1990.

[ISO90] ISO. Information Technology - Telecommunications and Information Exchange between Systems - Intermediate System to Intermediate System Routing Exchange Protocol for Use in Conjunction with the Protocol for Providing the Connectionless-Mode Network Service. ISO, 1990.

[LMJ99] C. Labovitz, G. Malan, and F. Jahanian. Origins of internet routing instability. In Proceedings of Infocomm'99 Conference, New York, USA, 3 1999.

[RL95] Y. Rekhter and T. Li. A Border Gateway Protocol 4 (BGP-4), RFC 1771. Internet Engineering Task Force, March 1995.

Authors' Addresses

Ajay Patel
Bell Labs, Lucent Technologies
101 Crawfords Corner Road
Holmdel, NJ 07733-3030
ajayp@dnrc.bell-labs.com

Tony Przygienda
Bell Labs, Lucent Technologies
101 Crawfords Corner Road
Holmdel, NJ 07733-3030
prz@dnrc.bell-labs.com