IS-IS Detailed IP Reachability Extension

draft-swallow-isis-detailed-reach-00.txt

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/1id-abstracts.html

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

Abstract

This document defines a means for IS-IS to carry detailed host reachability information along with summarized IP reachability. In particular it defines a new sub-TLV of the extended IP reachability TLV.
Contents

1  Introduction .............................................. 3
1.1  Conventions ............................................ 3
1.2  Terminology ............................................ 3
2  Background ................................................ 3
3  Overview .................................................. 4
4  Detailed Reachability Sub-TLV ............................ 5
4.1  Backward Compatibility ................................. 6
5  Semantics of detailed reachability ....................... 6
6  Open issues .............................................. 6
7  Security Considerations .................................. 7
8  IANA Considerations ...................................... 7
9  References ............................................... 7
9.1  Normative References .................................. 7
9.2  Informative References .................................. 8
10  Authors’ Addresses ....................................... 8
1. Introduction

The IS-IS protocol is specified in ISO-10589 [1], with extensions for supporting IPv4 specified in RFC1195 [2]. The extended IP reachability TLV is specified in RFC3784 [3]. This document defines a sub-TLV of that TLV to allow detailed host reachability information to be carried along with summarized IP reachability.

1.1. Conventions

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].

1.2. Terminology

<table>
<thead>
<tr>
<th>ABR</th>
<th>Area border router</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGP</td>
<td>Interior gateway protocol</td>
</tr>
<tr>
<td>RIB</td>
<td>Routing information base</td>
</tr>
<tr>
<td>VPN</td>
<td>Virtual private network</td>
</tr>
</tbody>
</table>

2. Background

IS-IS advertises routing/reachability information in link-state packets within a domain. Currently no distinction is made between routing and reachability. In the case of a host-route (/32 addresses in the case of IPv4) this is not a problem as there can be no ambiguity between routing and reachability. If a host is advertised as reachable, then there is (except during a convergence period or in very unusual circumstances) a routed path to that address. However, when shorter prefixes are advertised as reachable, reachability to a specific host address is hidden.

When reachability is summarized as it often is between levels, detailed reachability information is lost. Such summarization is critical to the scaling and convergence of the forwarding plane.

However, various control plane elements require host reachability information (usually to PE loopback addresses) either for correct action or to speed convergence. This level of detail very often is not needed in the forwarding plane. But the current all-or-nothing behavior of the IGP leaves a network operator with a choice of missing the benefits of summarization for IGP scalability or losing the benefits of detailed reachability information.
Among the control plane elements that could benefit from detailed host-reachability information are BGP next-hop tracking and PIM.

The Border Gateway Protocol (BGP) advertises routes that are external to the domain by associating them with a BGP next-hop address that is known within the domain. Often multiple next-hops are available to reach a particular prefix. If a prefix becomes unreachable, then BGP will withdraw the route. Such withdrawals take time. In particular if the advertising router goes down the withdrawal may be delayed until the BGP TCP session times out.

In order to speed convergence routers employ a technique called next-hop tracking. In next-hop tracking the reachability of the BGP next-hop is tracked. If a next-hop becomes unreachable, BGP route selection is run. External routes that are reachable through a known alternative next-hop are then installed.

Currently if next-hop tracking is to be performed, the above mentioned host-routes cannot be summarized. The proposed extension allows the IGP routes to be summarized while distributing the detailed reachability information needed for next-hop tracking.

PIM depends on the IGP reachability to the source of an (S, G) state to determine its RPF interface. When PIM installs an (S, G) state for the first time, it registers with the RIB for being notified of any route change to S. Later on, if the route to S changes, RIB immediately sends a notification to PIM.

3. Overview

In IS-IS IP reachability information may be carried in the extended IP reachability TLV. The TLV carries an IP prefix and a prefix length. This enables routes to be summarized to cover $2^n$ routes where $n$ is the difference between 32 and the prefix length. A consequence of this summarization is that detailed reachability is hidden.

This document defines a means to carry detailed reachability information along with a summarized IP prefix. Host reachability information is carried via a bit vector of $2^n$ bits. For example, if an area that had 10.0.1.0/25 assigned as its address range and had routes with loopbacks as follows

$$10.0.1.1 - 10.0.1.27$$
$$10.0.1.46$$
$$10.0.1.74 - 10.0.1.87$$
then the bit mask encoding would advertise a summary route to 10.0.1.0/25 with an associated 128-bit vector (shown in network order) like this:

<table>
<thead>
<tr>
<th>0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0</td>
</tr>
<tr>
<td>+---------------------------------------------------------------+</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>+---------------------------------------------------------------+</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>+---------------------------------------------------------------+</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>+---------------------------------------------------------------+</td>
</tr>
</tbody>
</table>

4. Detailed Reachability Sub-TLV

The detailed reachability sub-TLV is defined as a sub-TLV of the extended IP reachability TLV. Its type is sub-TLV type [to be assigned]. The sub-TLV length is the minimum number of octets required to contain a bit vector with a length equal to the number of IP addresses covered by the prefix contained in the parent extended IP reachability TLV. If \( L \) stands for the sub-TLV length and \( p \) stands for the prefix length then \( L = \text{ceiling}(2^{(32-p)}/8) \). The maximum length of the value field of any sub-TLV is 247 octets. Since the bit-vectors are always powers of 2 in length, the maximum bit-vector that will fit is 128 octets. This is sufficient to handle a prefix of 22 bits. Shorter prefixes cannot be expressed directly. Instead they may be expressed by advertising as many 22 bit prefixes as are contained within the longer prefix.

The value field encodes the bit vector where the high-order bit of the first octet corresponds to zero, the low-order bit to seven, the high-order bit of the second octet to eight and so forth. Each of these values taken as a binary number and used as the low-order bit of a 32 bit address formed with the prefix as the high-order bits indicates a particular host route. A bit value of one indicates that the associated host is reachable. A bit value of zero indicates that the associated host is not reachable.
4.1. Backward Compatibility

As defined in RFC 3784, a sub-TLV which is not understood, is to be ignored. Thus a router which does not understand the new sub-TLV will behave as if it had simply received the summary route.

5. Semantics of detailed reachability

As stated above, detailed reachability is determined by the setting of the bit associated with a specific host. When a router receives the same summary route from two or more different routers the following considerations apply.

If some of the advertisements contain the detailed reachability sub-TLV while others do not, then detailed reachability MUST be determined based solely on those advertisements containing the detailed reachability sub-TLV.

If the advertisements contain differing bit vectors, then the router SHOULD base reachability on the logical OR of the bits. In some situations where aggressive recovery from failures is needed, reachability MAY be based on the logical AND of the bits.

[Note this section is far from complete - a full description of how bit-vectors are to be combined will involve considerations of metric values and partition repair. See the open issues section below.]

6. Open issues

A number of issues remain open which, with input from the community, will be refined in future versions of this draft.

- The IPv4 encoding defined in this draft allows for a minimum prefix length of 22 bits due to the limitations of the sub-TLV length. Shorter prefixes may be handled by advertising multiple 22-bit-prefixes. Is this sufficient for the community?

- The section on combining bit-vectors needs to be discussed in greater detail. That discussion needs to include metrics and partition repair.
Beyond its other uses, the bit-vector opens the possibility of automatic partition detection and repair. Two means of partition repair have been discussed among the authors. These discussions are preliminary and a full proposal will be put forward in a future draft.

One potential means of partition repair consists of an L1L2 router auto-generating host routes for those hosts seen by that router but not by all of the L1L2 routers serving the same area.

The other consists of building tunnels between the set of L1L2 routers serving an area and "shunting" packets destined to hosts that are not seen by an L1L2 to a L1L2 router that is (via its bit-vector) advertising reachability to that host.

7. Security Considerations

The detailed reachability sub-TLV does not change the information that IS-IS can share with other routers, nor does it change the set of routers to which the information is sent. It does RECOMMEND that a router treat the information differently, delivering the detailed reachability to the control plane while using the summary to scale the forwarding plane. These changes however are not mandated. Thus this extension to IS-IS poses no new security threats.

8. IANA Considerations

[to be written]

9. References

9.1. Normative References


[KEYWORDS] Bradner, S., "Key words for use in RFCs to Indicate
9.2. Informative References


10. Authors’ Addresses

Clarence Filsfils
Cisco Systems, Inc.
Email: cfilsfil@cisco.com

Stefano Previdi
Cisco Systems, Inc.
Email: sprevidi@cisco.com

George Swallow
Cisco Systems, Inc.
Email: swallow@cisco.com

Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.
The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Full Copyright Notice

Copyright (C) The IETF Trust (2007). This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET 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.