Interdomain Routing Working Group
Internet Draft
<draft-ietf-idr-idrp-v4v6-02.txt>

Yakov Rekhter
cisco Systems
Paul Traina
cisco Systems
January 1996

IDRP for IP v4 and v6

Status of this memo

This document is an Internet Draft. 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. Internet Drafts may be updated, replaced, or obsoleted by other documents at any time. It is not appropriate to use Internet Drafts as reference material or to cite them other than as a ``working draft'' or ``work in progress.''

Please check the 1id-abstracts.txt listing contained in the internet-drafts Shadow Directories on nic.ddn.mil, nnsc.nsf.net, nic.nordu.net, ftp.nisc.sri.com, or munnari.oz.au to learn the current status of any Internet Draft.

1 Overview

IDRP [5] is defined as the protocol for exchange of Inter-Domain routing information between routers to support forwarding of ISO 8473 (Connectionless Network Layer Protocol (CLNP))[6] packets.

The network reachability information exchanged via IDRP provides sufficient information to detect routing loops and enforce routing decisions based on performance preference and policy constraints as outlined in RFC 1104 [1]. In particular, IDRP exchanges routing information containing full domain-level paths and enforces routing policies based on configuration information.

IDRP may be viewed as an extension of BGP-4 ([9], [10]) that provides (among other things) much better scaling with respect to support for routing information aggregation based on CIDR ([2], [11]), as well as stronger capabilities for policy based routing (e.g. ability to impose control over transit traffic). Enhanced scaling capabilities

are provided via the concept of Routing Domain Confederations (RDCs), that allow to express both topology and policy information in terms of aggregates (confederations) rather than individual entities (domains). IDRP also provides capability to carry reachability and

forwarding information associated with multiple network layer protocols (e.g. IPv6, IPv4).

This document contains the adaptation of the IDRP protocol definition that enables it to be used as a protocol for the exchange of inter-domain system routing information among routers to support the forwarding of IPv6 packets across multiple domains. We refer to IDRP with this adaptation as "IDRP for IPv6". While this document doesn't cover use of IDRP to support routing for other network layer protocols (e.g. IPv4), it is expected that IDRP for IPv6 will be able to operate in a multiprotocol environment as well.

2 Terminology

This document assumes that the reader is familiar with the following documents:

IPv6 protocol specification [3], IPv6 Addressing Architecture [4], and IDRP specification (IS 10747) [5].

A few definitions are in order to aid the reader:

BIS - a Boundary Intermediate System (or border router)

BISPDU - an IDRP message exchanged between a pair of BISs

ES - End System (host)

FIB - Forwarding Information Base (IP forwarding table)

IS - Intermediate System (router)

NET - Network Entity Title (a network layer address for a router)

NLRI - Network Layer Reachability Information (set of reachable

destinations)

NPDU - an IPv6 packet

NSAP - Network Service Access Point (a network layer address)

PDU - a packet

SNPA - subnetwork point of attachment (Data Link address)

It is expected that the above definitions should be adequate for understanding of IDRP. Familiarity with any of the documents listed in the normative references of the protocol specifications (section 2 of [5]) is not required.

Unless stated otherwise here, any reference to the above terms in [5]

should be interpreted based on the above definitions.

3 The Adaptation Layer

The Inter-Domain Routing Protocol (IDRP) or, more formally,

"The Protocol for the Exchange of Inter-Domain Routing information among Intermediate Systems to support Forwarding of ISO 8473 PDUs (IDRP)"

is the inter-domain routing protocol defined to support the forwarding of Connectionless Network Layer Protocol (CLNP) [6] packets that traverse multiple routing domains.

IDRP document [5] covers both the protocol specifications and the usage issues (which is in contrast to BGP-4 documentation that has a separate document that defines the protocol [10], and a separate document that describes the protocol's usage [9]).

While IDRP was developed within ISO, it makes few, if any, ISO-specific assumptions. In particular, it does not require

participating domains to support any specific ISO Intra-Domain protocol, such as IS-IS [7], nor does it require participating routers to run ES-IS [8].

The only requirements imposed by the protocol on the participating routers is that the protocol information can be exchanged among them over a connectionless network layer (which in the case of OSI is CLNP), and that the network layer connectivity between routers within a single routing domain should be provided by means outside of IDRP (e.g., via some intra-domain routing protocol). IDRP does not place any restrictions on the structure of reachability information, as long it can be expressed as an arbitrary set of variable length address prefixes.

Since IPv4 and IPv6 can provide connectionless service between routers, and since reachable IPv4/IPv6 destinations can be expressed as IP address prefixes, IDRP can be easily adapted to be an inter-domain routing protocol which can be used in the IP Internet.

The adaptation described in this document consists of: specifying the parts of the protocol that are not needed, specifying modifications/clarifications to certain parts of the protocol to reflect IP specifics and operational experience with BGP-4, adding new features to reflect operational experience with BGP-4.

4 Features in IDRP which shall not be implemented

The following lists the functions that shall not be implemented by

IDRP for IPv4 an IPv6 (all references are with respect to [5]):

Support for distinguishing path attributes according to sections 5.7, 7.11.2 and 7.11.3 Expense according to section 7.12.10 Security according to section 7.12.14 Priority according to section 7.12.16 Procedures for detecting inconsistent routing decisions, according to section 7.15.1 Forwarding CLNP packets according to section 8 The interface to CLNP according to section 9 support of the Network Management information described in the IDRP GDMO according to section 11

All the material presented in the sections listed above may be ignored.

5 Features in IDRP which shall be implemented

An implementation of IDRP for IPv4 and IPv6 shall contain all mandatory features
of IDRP, except those mentioned in section 4 of this document. In addition, a BIS for IDRP for IPv4 and IPv6 shall implement the following (all references are with respect to this document):

an interface to the IPv4 and IPv6 protocol, as described in section 5.1 Modifications to the encoding of reachability and forwarding information, as well as the ability to identify and extract IPv4 and IPv6 reachability and forwarding information as described in sections 5.2 and 5.3 Modifications to the ROUTE_SEPARATOR and MULTI_EXIT_DISCRIMINATOR path attributes, as described in section 5.4 Support for the ATOMIC_AGGREGATE path attribute, as described in section 5.5 Modifications to the tie-breaking procedures, as described in sections 5.6 Modifications to handling Hold Time, as described in section 5.7 Constructing forwarding address (next hop), as described in section 5.8 Modifications to the UPDATE PDU format, as described in section 5.9 Modifications to the OPEN PDU format, as described in section 5.10 Modifications to the RIB REFRESH PDU format, as described in section 5.11 New Error Subcodes, as described in section 5.12

Naming and addressing conventions discussed in sections 5.10, 5.11 and 7.1 of [5] do not apply to IDRP for IPv4 and IPv6, and thus should be ignored. Section 6 of this document contains the material that covers naming and addressing conventions for IDRP for IPv4 and IPv6.

Deployment guidelines for IDRP for IPv4 and IPv6 are specified in section 7 of this document. These guidelines supersede the material presented in section 7.2 of [5].

Domain configuration information for IDRP for IPv4 and IPv6 is defined in section 8 of this document. The material of that section supersedes the material presented in section 7.3 of [5].

5.1 An interface to IP

This sections supersedes the material in section 7.5 of [5].

IDRP information is carried between a pair of BISs in the form of BISPDUs. For IDRP for IPv6 these BISPDUs are carried in the data field of IP packets of protocol type 45.

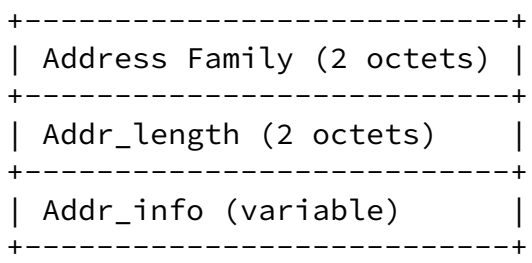
IDRP relies on IP to perform the initial processing of incoming BISPDUs. The IP protocol machine shall process inbound packets according to the appropriate IP functions.

If a fixed header of an IP packet contains a protocol type that identifies IDRP, and the packet's source address identifies any system listed in managed objects internalBIS or externalBISNeighbor, then the packet contains a BISPDU. The BISPDU shall be passed to the IDRP finite state machine defined in section 7.6.1 of [5].

5.2 Encoding IP reachability information

The text in this section supersedes the material presented in section 6.3.2 of [5].

The Network Layer Reachability information is a variable length field that contains a list of reachable destinations encoded as zero or more triples of the form <Address Family, Addr_length, Addr_info>, whose fields are described below:

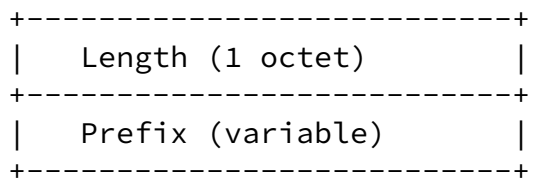


The use and meaning of these fields are as follows:

Address Family:
This field carries the identity of the protocol associated with the address information that follows. Presently defined values for this field are specified in RFC1700. A conformant implementation of IDRP for IPv6 may ignore any address information indicating other than IPv6. A conformant implenetation of IDRP for IPv4 may ignore any address information indicating other than IPv4. Address Family.

Addr_Length:
This field specifies the total length in octets of the address information that follows.

Addr_Info:
This is a variable length field that contains a list of IP address prefixes for the routes that are being advertised. Each IP address prefix is encoded as a 2-tuple of the form <Length, Prefix>, whose fields are described below:



The use and the meaning of these fields are as follows:

a) Length:

The Length field indicates the length in bits of the IP address prefix. A length of zero indicates a prefix that matches all IPv4 or IPv6 (as specified by the address family) addresses (with prefix, itself, of zero octets).

b) Prefix:

The Prefix field contains IP address prefixes followed by enough trailing bits to make the end of the field fall on an octet boundary. Note that the value of trailing bits is irrelevant.

5.3 Encoding IP forwarding information

IPv6 forwarding information is carried in the NEXT_HOP path attribute. As specified in [5], the attribute has a Proto_type, Proto_Length and Protocol fields which indicate the protocol family

for the address of the NEXT_HOP (see section 6.3.1.4 of [5]). This document replaces these three fields (Proto_type, Proto_Length, and Protocol) with a single field -- Address Family. This 2-octets field carries the identity of the protocol associated with the address information that follows. Presently defined values for this field are specified in RFC1700. A conformant implementation of IDRP for IPv6 may ignore any address information indicating other than IPv6 Address Family. A conformant implementation of IDRP for IPv4 may ignore any address information other than the IPv4 Address Family.

An implementation of IDRP for IPv4 or IPv6 shall have the following values in the NEXT_HOP field:

IPv6:
Length of NET: 16

NET of Next Hop: an IPv6 unicast address

SNPA information: as appropriate for the subnetwork type in use

IPv4:
Length of NET: 4

NET of Next Hop: an IPv4 unicast address

SNPA information: as appropriate for the subnetwork type in use

All other fields of the NEXT_HOP attribute remains as specified in [5].

5.4 Modification to the existing path attributes

To facilitate operations, IDRP for IPv6 modifies the following path attributes:

LOCAL_PREF field in the ROUTE_SEPARATOR attribute (see section 6.3.1.1) is changed from 1 octet to 4 octets. The ROUTE-ID field in the ROUTE_SEPARATOR attribute is eliminated. As a result the length

of the ROUTE_SEPARATOR attribute is changed from 5 to 4 octets. The length of the MULTI_EXIT_DISCRIMINATOR attribute is changed from 1 octet to 4 octets.

Semantics, as well as handling of the modified attributes is left intact.

5.5 New path attributes

IDRP for IPv6 defines the following new attribute:

AGGREGATOR (Type Code 17):

AGGREGATOR is an optional transitive attribute of length 32. The attribute contains the last RDI that formed the aggregate route (encoded as 16 octets), followed by the IP address of the BIS that formed the aggregate route (encoded as 16 octets, IPv4 addresses are prefixed with 12 octets of zeros). The BIS that formed the aggregate route may decline to encode its address and instead insert a value of all zeros into that field.

The attribute may be included in routes which are formed by route aggregation. A BIS that performs the aggregation may add the AGGREGATOR attribute which shall contain BIS's own RDI and IPv6 address.

ATOMIC_AGGREGATE (Type Code 18):

ATOMIC_AGGREGATE is a well-known discretionary attribute of length 0. It is used by a BIS to inform other BISs that the local system selected for advertisement a less specific route without selecting a more specific route which is included in it.

If a BIS, when presented with a set of overlapping routes from one of its peers, selects the less specific route without selecting the more specific one, then the local system shall attach the ATOMIC_AGGREGATE attribute to the

route when propagating it to other BISs (if that attribute is not already present in the received less specific route). A BIS that receives a route with the ATOMIC_AGGREGATE attribute shall not remove the attribute from the route when propagating it to other BISs. A BIS that receives a route with the ATOMIC_AGGREGATE attribute shall not make any NLRI of that route more specific when advertising this route to other BISs. A BIS that receives a route with the ATOMIC_AGGREGATE attribute needs to be cognizant of the fact that the actual path to destinations, as specified in the NLRI of the route, while having the loop-free property, may traverse domains/confederations that are not listed in the RD_PATH attribute.

5.6 Modifications to tie-breaking procedures for phase 2

This section supersedes the material in section 7.16.2.1 and 7.16.1.1 of [5].

In its Adj-RIBs-In a BIS may have several routes to the same destination that have the same degree of preference. The local BIS can select only one of these routes for inclusion in the associated Loc-RIB. The local BIS considers all equally preferable routes, both those received from BISs located in adjacent RDs, and those received from other BISs located in the local BIS's own RD.

Ties shall be broken according to the following algorithm:

a) If the local BIS is configured to take into account MULTI_EXIT_DISC, and the candidate routes differ in their MULTI_EXIT_DISC attribute, select the route that has the lowest value of the MULTI_EXIT_DISC attribute. If the local BIS is configured to take into account MULTI_EXIT_DISC, but that attribute is not present, a locally defined "default" MULTI_EXIT_DISC may be assumed as a basis for performing tie-breaking.

b) Otherwise, if the local BIS can ascertain the cost of a path to the entity depicted by the NEXT_HOP attribute of the candidate route, select the route with the lowest cost (interior distance) to the entity depicted by the NEXT_HOP attribute of the route. If there are several routes with the same cost, then the tie-breaking

shall be broken as follows:

- if at least one of the candidate routes was advertised by the BIS in an adjacent RD, select the route that was advertised by the BIS in an adjacent RD whose address has the lowest value among all other BIS in adjacent RDs;

- otherwise, select the route that was advertised by the BIS whose address has the lowest value.

5.7 Modifications to handling Hold Time

Upon receipt of an OPEN BISPDU, a BIS must calculate the value of the Hold Timer by using the smaller of its configured Hold Time and the Hold Time received in the OPEN BISPDU.

IDRP for IPv6 requires the value of the Hold Time field carried in the OPEN BISPDU to be either zero or at least 3 seconds. An implementation must reject Hold Time values of one or two seconds. An implementation may reject any proposed Hold Time. An implementation which accepts a Hold Time must use the negotiated value for the Hold Time. If the negotiated Hold Time interval is zero, then periodic KEEPALIVE messages shall not be sent.

In addition to the OPEN PDU error handling procedures specified in section 7.20.2 of [5] this document specifies that if the Hold Time field of the OPEN message is unacceptable, then the Error Subcode shall be set to Unacceptable Hold Time.

5.8 Determining the forwarding address (Next Hop)

Next hop forwarding information information associated with a particular route shall be derived from the NEXT_HOP attribute in the UPDATE BISPDU that carries the route. If that attribute is not present, the next hop (forwarding address) shall be derived from the source IPv6 address of the IPv6 packet that carries the UPDATE BISPDU containing the route.

In addition to the procedures for handling the NEXT_HOP attribute specified in section 7.12.4 of [5], IDRP for IPv4 and IPv6 specifies the following:

A BIS must never advertise an address of a peer to that peer as a

NEXT_HOP, for a route that the speaker is originating. A BIS must never install a route with itself as the next hop. When a BIS advertises the route to a BIS located in its own domain, the advertising BIS should not modify the NEXT_HOP attribute associated with the route. When a BIS receives the route from an internal neighbor BIS, it may use the NEXT_HOP address as the forwarding address, provided that the address is on a common subnet with the

local BIS.

5.9 Modifications to the UPDATE PDU

This document specifies that NLRI of a route, rather than the Route-ID of the route, shall be used to withdraw a previously advertised route from service.

The Withdrawn Routes field in the UPDATE PDU is specified as a variable length field that contains a list of NLRIs (rather than the list of Route-IDs) for the routes that are being withdrawn from service. Each NLRI is encoded as specified in Section 5.2 of this document. An UPDATE PDU can list multiple routes to be withdrawn from service. Each such route is identified by its NLRI, which unambiguously identifies the route in the context of the BIS-BIS connection in which it had been previously been advertised.

Eliminating Route-ID is also reflected in the encoding of the ROUTE_SEPARATOR attribute (see Section 5.4 of this document).

5.10 Modifications to the OPEN PDU

Since IDRP for IPv6 doesn't support any Distinguishing Attrbutes, the RIB-AttsSet field is eliminated from the OPEN PDU. (PST--bring back DA's?)

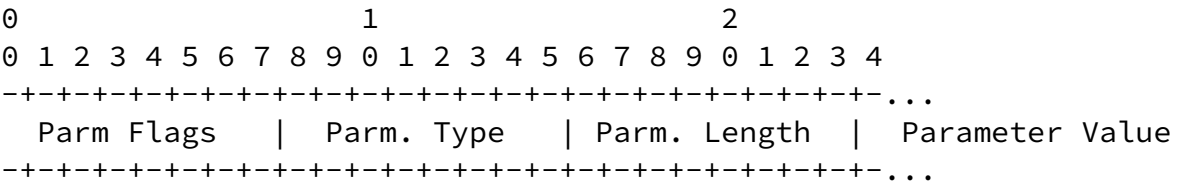
The last two fields of the OPEN PDU message, Authentication Code and Authentication Data, are replaced with the following two fields:

Optional Parameters Length:

This 2-octet unsigned integer indicates the total length of the Optional Parameters following this field in octets. If the value of this field i s zero, no Optional Parameters are present.

Optional Parameters:

This field may contain a list of optional parameters, where each parameter is encoded as a <Parameter Flags, Parameter Type, Parameter Length, Parameter Value> vector.



(PST: grab BGP/4 flags text and talk to yakov about making

length extensible
just like attribute length in BGP)

Parameter Flags is a one octet field that (PST: grab text from BGP-4)

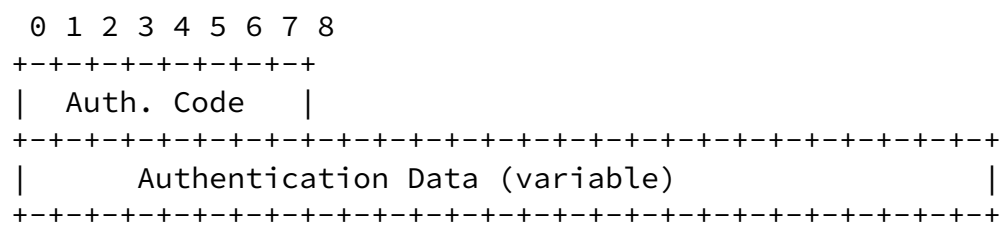
Parameter Type is a one octet field that unambiguously identifies individual parameters. Parameter Length is a one octet field that contains the length of the Parameter Value field in octets. Parameter Value is a variable length field that is interpreted according to the value of the Parameter Type field.

This document defines the following Optional Parameters:

a) Authentication Information (Parameter Type 1):

This optional parameter may be used to authenticate a BIS peer. The Parameter Value field contains a 1-octet Authentication Code followed by a variable length

Authentication Data.



The syntax and semantics of these two field is left unchanged (as specified in section 6.2 of [5]).

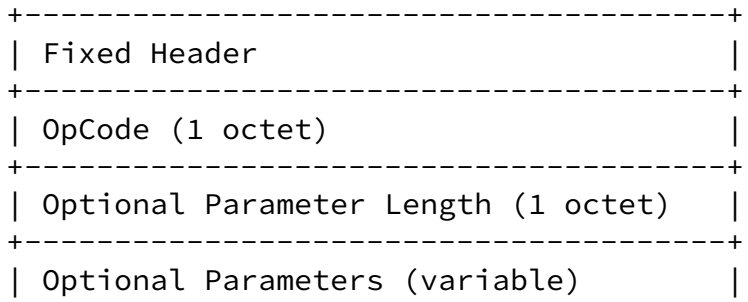
Absence of any Authentication Information in an OPEN PDU shall be treated as if the PDU carries Authentication Information with Authentication Type 1 (see section 7.1.1 of [5]).

In addition to the OPEN PDU error handling procedures specified in section 7.20.2 of [5] this document specifies that if one of the Optional Parameters in the OPEN message is not recognized, then the Error Subcode is set to Unsupported Optional Parameters.

5.11 Modifications to the RIB REFRESH PDU

This sections supersedes the material in section 6.7 of [5].

The RIB REFRESH PDU is used to allow a BIS to send a refresh of the routeing information in an Adj-RIB-Out to a neighbor BIS, or to solicit a neighbor BIS to send a refresh of its Adj-RIB-Out to the local BIS. The RIB REFRESH PDU contains a fixed header and also the additional fields shown below:



```
   +--------------------------------------+
```

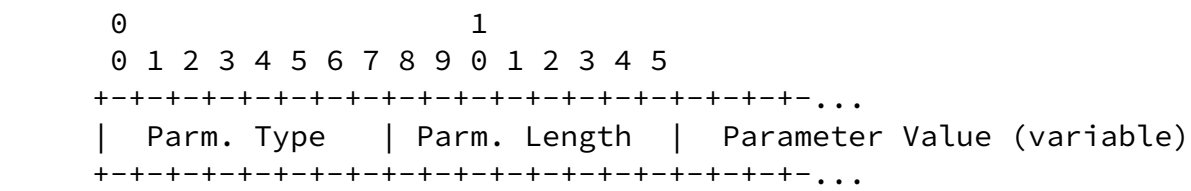
The use and meaning of these fields is as follows:

There are three OpCode values defined:

Code	Operation
1	RIB Refresh Request
2	RIB Refresh Start
3	RIB Refresh End

Optional Parameters Length:
This 1-octet unsigned integer indicates the total length of the Optional Parameters field in octets. If the value of this field is zero, no Optional Parameters are present.

Optional Parameters:
This field may contain a list of optional parameters, where each parameter is encoded as a <Parameter Type, Parameter Length, Parameter Value> triplet.



Parameter Type is a one octet field that unambiguously identifies individual parameters. Parameter Length is a one octet field that contains the length of the Parameter Value field in octets. Parameter Value is a variable length field that is interpreted according to the value of the Parameter Type field.

When a BIS receives a RIB REFRESH PDU that contains one or more Optional Parameters, and the BIS doesn't support or does't recognize at least one of the parameters, the BIS processes the PDU as if it wouldn't have any Optional Parameters. This document doesn't specify any Optional Parameters.

Usage of RIB REFRESH PDU is defined in 7.10.3 of [5].

5.12 Additional Error Subcodes

In addition to the Error subcodes defined in section 5.4 of [5], this document defines the following OPEN PDU Error subcodes:

8 - Unacceptable Hold Time (see Section 5.7 of this document)

9 - Unsupported Optional Parameter (see Section 5.12 of this document)

6 Naming and addressing conventions

This section supersedes the material of sections 5.10, 5.11 and 7.1 of [5].

IDRP for IPv4 and IPv6 does not assume or require any particular structure for IP addresses. That is, as long as the domain administrator assigns addresses that are consistent with the deployment constraints of section 7 of this document, the protocol will operate correctly.

IP address prefixes provide a compact way for identifying groups of systems that reside in a given domain or confederation. A prefix may have a length that is either smaller than, or the same size as the IP address (an IPv4 or IPv6 address is a special case of an address prefix). The length of an encoded prefix is specified in bits.

Each routing domain and routing domain confederation whose BIS(s) implement IDRP for IPv4 and IPv6 shall have an unambiguous routing domain identifier (RDI), which is an IPv4 or IPv6 address prefix. In the case of IPv4 address prefixes, the prefix value shall be prepended with 12 octets of zeros.

An RDI is assigned statically and does not change based on the operational status of a routing domain. An RDI identifies routing domain or confederation uniquely, but does not necessarily convey any information about policies or identities of its members.

7 Deployment guidelines

This section supersedes the material in section 7.2 of [5].

Hosts and routers may use any IP unicast addresses, provided that these addresses are globally unambiguous. However correct and

efficient operation of this protocol can only be guaranteed if the address assignment reflects the actual topology -- addresses are topologically significant. One possible architecture for IPv6 address assignment that satisfies this requirement is described in [12].

8 Domain Configuration Information

Correct Operation of IDRP described in [5] assumes that a minimum amount of information is available to both the inter-domain and intra-domain routing protocols. This information is static in nature, and is not expected to change frequently. This document assumes that this information is supplied via IDRP MIB. While the following in phrased in terms of MIB, this document allows alternative mechanisms (e.g. configuration files) as well.

The information required by a BIS that implements the IDRP for IPv4 and IPv6 protocol is:

Location and identity of adjacent Intra-Domain routers:

The MIB table IntraIS lists the IP addresses of the routers to which the local BIS may deliver an inbound NPDU whose destination lies within the BIS's routing domain. These routers listed in the IntraIS table support the intra-domain routing protocol of this domain, and share at least one common subnet with the BIS.

In particular, if the local BIS participates in both the inter-domain routing protocol (IDRP) and the intra-domain routing protocol, then the IP address of the local BIS will be listed in the IntraIS table.

Location and identity of BISs in the BIS's domain:

This information permits a BIS to identify all other BISs located within its routing domain. This information is contained in the MIB table InternalBIS, which contains a set of IPv6 addresses which identify the BISs in the domain.

Location and identity of BISs in adjacent domains:

Each BIS needs information to identify the IP address of each BIS located in an adjacent RD and reachable via a single subnetwork hop. This information is contained in the IDRP MIB table externalBISNeighbor, which is a table of IPv6 addresses.

IP network address information for all systems in the routing domain:

This information is used by the BIS to construct its network layer reachability information. This information is contained in the MIB table internalSystems, which lists NLRI (expressed as address prefixes) of the systems within the routing domain.

Local RDI:

This information is contained in managed object localRDI; it is the RDI of the routing domain in which the BIS is located.

RDC-Config:

This information identifies all the routing domain confederations (RDCs) to which the RD of the local BIS belongs, and it describes the nesting relationships that are in force between them. It is contained in the MIB table rdcConfig.

Note that since a domain is not required to belong to a confederation this information is optional and needs to be present only at BISs of the domains that are part of one or more of RDCs.

9 Multiple IDRP sessions between the same pair of routers

An IP router may have multiple IP addresses, one for each interface. In contrast, an OSI Intermediate System has only one Network Entity Title (network address). An OSI BIS thus may not have multiple IDRP sessions with another BIS, since the NET is unique and there is no mechanism for multiplexing sessions. However, an IP router may potentially have multiple IDRP sessions with another router, since each BIS may have multiple IP addresses, and one BIS may not be able to ascertain that those addresses correspond to the same BIS. Multiple IDRP sessions between BISs may not be efficient, but they are not illegal, nor do they impact the robustness of the IDRP for IP protocol; they will simply appear as multiple paths to the same neighboring domain. One possible way of avoiding multiple parallel IDRP sessions between a pair of BISs within a single domain is to bind all source addresses of outgoing BISPDUs to the IPv6 address of a particular interface (either physical or logical) of the BIS. Likewise, for a pair of BISs located in adjacent domains, binding the source addresses to a single address of an interface attached to a common subnetwork allows for the elimination of multiple parallel sessions.

10 Required set of supported routing policies

Policies are provided to IDRP in the form of configuration information. This information is not directly encoded in the protocol. Therefore, IDRP can provide support for very complex routing policies (an example of such policy is presented in Annex K of [5]). However, it is not required that all IDRP implementations support such policies.

We are not attempting to standardize the routing policies that must be supported in every IDRP implementation; we strongly encourage all implementors to support the following set of routing policies:

IDRP implementations should allow a domain to control announcements of IDRP-learned routes to adjacent domains. Implementations should also support such control with at least the granularity of a single address prefix. Implementations should also support such control with the granularity of a domain, where the domain may be either the domain that originated the route, or the domain that advertised the route to the local system (adjacent domain). Care must be taken when a BIS selects a new route that can't be announced to a particular

external peer, while the previously selected route was announced to that peer. Specifically, the local system must explicitly indicate to the peer that the previous route is now infeasible. IDRP implementations should allow a domain to prefer a particular path to a destination (when more than one path is available). At the minimum an implementation shall support this functionality by allowing to administratively assign a degree of preference to a route based solely on the IP address of the neighbor the route is received from. The allowed range of the assigned degree of preference shall be between 0 and 2^(31) - 1. IDRP implementations should allow a domain to ignore routes with certain domains in the RD_PATH path attribute. Such function can be implemented by assigning "infinity" as "weights" for such domains. The route selection process must ignore routes that have "weight" equal to "infinity".

11 Operations over Switched Virtual Circuits

When using IDRP for IPv4 and IPv6 over Switched Virtual Circuit (SVC) subnetworks it may be desirable to minimize traffic generated by IDRP. Specifically, it may be desirable to eliminate traffic associated with periodic KEEPALIVE messages. IDRP for IPv4 and IPv6 includes a mechanism for operation over switched virtual circuit (SVC) services which avoids keeping SVCs permanently open and allows it to eliminates periodic sending of KEEPALIVE messages.

This section describes how to operate without periodic KEEPALIVE messages to minimize SVC usage when using an intelligent SVC circuit manager. The proposed scheme may also be used on "permanent" circuits, which support a feature like link quality monitoring or echo request to determine the status of link connectivity.

The mechanism described in this section is suitable only between the BISs that are directly connected over a common virtual circuit.

11.1 Establishing an IDRP Connection

The feature is selected by specifying zero Hold Time in the OPEN BISPDU.

11.2 Circuit Manager Properties

The circuit manager must have sufficient functionality to be able to

compensate for the lack of periodic KEEPALIVE BISPDU:

It must be able to determine link layer unreachability in a predictable finite period of a failure occurring. On determining unreachability it should: start a configurable dead timer (comparable to a typical Hold timer value). attempt to re-establish the Link Layer connection.

If the dead timer expires it should: send a deactivate indication to IDRP FSM. If the connection is re-established it should: cancel the dead timer. transmit any queued BISPDUs.

11.3 Combined Properties

Some implementations may not be able to guarantee that the IDRP process and the circuit manager will operate as a single entity; i.e. they can have a separate existence when the other has been stopped or has crashed.

If this is the case, a periodic two-way poll between the IDRP process and the circuit manager should be implemented. If the IDRP process discovers the circuit manager has gone away it should close all relevant BIS-BIS connections. If the circuit manager discovers the IDRP process has gone away it should close all its BIS-BIS connections associated with the IDRP process and reject any further incoming BIS-BIS connections.

12 Modifications to the conformance clause

To reflect the list of functions that shall not be implemented (see section 4 of this document) the following items in the IDRP conformance clause (section 12.1 of [5]) shall not be implemented:

clause (d): Transit Delay, Residual Error, Expense, clause (m) clause (r) clause (s) clause (t)

13 Modifications to PICS

The PICS (Protocol Implementation Conformance Statement) provides a convenient and concise mechanism to define which function need and need not be implemented for IDRP for IPv4 and IPv6. All references

in this section are with respect to [5].

All items with PICS Status as Optional need not be implemented in IDRP for IPv4 and IPv6. In addition, IDRP for IPv4 and IPv6 should not support the following items (even if some of the items are listed as Mandatory):

Table A.4.3:
MGT

Table A.4.5:
INCONS

Table A.4.8:
PSRCRT, DATTS, MATCH

Table A.4.11:
TDLY, RERR, EXP, LQOSG, SECG, PRTY

Table A.4.12:
TDLYP, RERRP, EXPP, LQOSP, SECP, PRTYP

Table A.4.13:
TDLYR, RERRR, EXPR, LQOSR, SECR, PRTYR

Implementation of all other items with Optional Status not listed in the previous paragraph is optional.

14 Navigating through IDRP

Here is the list of sections in [5] that are relevant to the IDRP for IPv6 implementation: chapters 1, 3, 4, 5 (except 5.10 and 5.11), 6, 7 (except for 7.1, 7.2, 7.3, 7.4, 7.12.8, 7.12.9, 7.12.10, 7.12.11 and 7.12.16), 10. The rest of the material in [5] could be safely ignored.

15 Security Considerations

Security issues are not discussed in this document.

16 Acknowledgements

Large parts of this document are borrowed from the BGP Protocol specifications and BGP Usage documents ([9], [10]).

We would like to thank Susan Hares (MERIT) and John Scudder (MERIT) for their work on IDRP for IPv4. Portions of this document are borrowed from their work.

We would like to thank Tony Li (cisco Systems) for his review of this document.

Finally we would like to thank the whole Inter-Domain Routing (IDR) Working Group for their contribution to this document.

---

17 References

[1] Braun, H-W., "Models of Policy Based Routing", RFC 1104, Merit/NSFNET, June 1989.

[2] Fuller, V., Li, T., Yu, J., Varadhan, K., "Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy", RFC 1519, September 1993

[3] Deering, S., Hinden, B., "Internet Protocol, Version 6 (IPv6) Specification", RFC1883, January 1996

[4] Hinden, B., Deering, S., "IP Version 6 Addressing Architecture", RFC1884, January 1996

[5] ISO/IEC IS 10747 - Information Processing Systems - Telecommunications and Information Exchange between Systems - Protocol for Exchange of Inter-domain Routing Information among Intermediate Systems to Support Forwarding of ISO 8473 PDUs, 1993 ftp://networking.raleigh.ibm.com/pub/standards/idrp/is10747.ps

ftp://networking.raleigh.ibm.com/pub/standards/idrp/is10747.txt

[6] ISO 8473 - Information Processing Systems - Data Communications - Protocol for Providing the Connectionless-mode Network Service, 1988.

[7] ISO/IEC 10589 - Information Processing Systems - Telecommunications and Information Exchange between systems - Intermediate System to Intermediate System Intra-Domain routing information exchange protocol for use in conjunction with the Protocol for providing the Connectionless-mode Network Service (ISO 8473), 1992.

[8] ISO 9542 - Information Processing Systems - Telecommunications and information exchange between systems - End system to Intermediate system routing exchange protocol for use in conjunction with the Protocol for providing the connectionless-mode network service (ISO 8473)

[9] Rekhter, Y., Gross, P., ``Application of the Border Gateway Protocol in the Internet'', RFC1655, July 1994

[10] Rekhter, Y., Li, T., ``A Border Gateway Protocol 4 (BGP-4)'', RFC1654, July 1994

[11] Rekhter, Y., Li, T., "An Architecture for IP Address Allocation with CIDR", RFC1518, September 1993

[12] Rekhter, Y., Li, T., "An Architecture for IPv6 Unicast Address Allocation", RFC1887, January 1996

Authors' Addresses

Yakov Rekhter
cisco Systems, Inc.
170 W. Tasman Dr.
San Jose, CA 95134

email: yakov@cisco.com

Paul Traina
cisco Systems, Inc.
170 W. Tasman Dr.
San Jose, CA 95134
email: pst@cisco.com