Internet Draft
Expires August 27, 1995
draft-ietf-sdr-route-construction-01.{ps,txt}

Kannan Varadhan
Deborah Estrin
Steve Hotz
USC/ISI
Yakov Rekhter
IBM

February 27, 1995

SDRP Route Construction

Abstract

SDRP packet forwarding and setup have been described in another document, and a prototype implementation completed. However, we are presently relying on manual mechanisms to construct the source routes. This document explores several design options for automating the process of SDRP route construction and thereby making it more dynamic. The proposed approaches make use of IDRP RIB information, and protocol mechanisms.

Status of this Memo

Contents

[0] The work of K. Varadhan and D. Estrin was supported by the National Science Foundation under contract number NCR-92-06418. The work of S. Hotz was supported by ARPA Graduate Fellowship, administered through UMIACS contract no. Z984114. Systems research at USC is supported through NSF infrastructure grant, award number CDA-9216321. The work of Y. Rekhter was supported by the National Science Foundation, under contract NCR-92-19216. Views and conclusions expressed in this paper are not necessarily those of the National Science Foundation or of the Advanced Research Projects Agency.

1 Introduction

SDRP[3, 2] forwarding and setup have been implemented, but currently we rely on hand configured routes. This approach obviously does not scale and will limit the use of SDRP in the long (and maybe even short) term. The motivation for this white paper, and the work proposed herein, is to support automatic route construction for SDRP.

We view the problem as one of obtaining routes or route fragments from various sources, and concatenating them to derive complete and valid source routes. We can then install the source routes in the kernel and use them appropriately. This document addresses the problem of obtaining these route fragments[1].

We first note that the RIBs and FIBs used in IDRP[5] route exchange are an instantiation of route fragments, i.e., each entry contains a path vector. We therefore at a minimum can use RIFs[13] to query remote BGP/IDRP peers for FIB and RIB information. For example, by peering with a remote IDRP speaker and obtaining its FIB, a domain client can obtain routes directly from the selected provider; in particular, routes that are not otherwise propagated/advertised by domains in between the selected provider and the client domain. This effects a relatively trivial policy, referred to as provider selection.

To support the wider range of policies for which SDRP was designed, we propose additional mechanisms: (1) a generalization of the IDRP query called RIB Query (Section 3.1), and (2) initiation of on-demand IDRP computations, called Path Explorer (Section 3.2).

The various requirements for route computation, and the applicable techniques are summarized in table 1.

Source of Route Information	Computation Mechanisms	
	Pre-Computed	On-Demand
Remote	Provider Selection	RIB Query
Local	IDRP/N^2	Path Explorer

Table 1: Classification of the techniques for route computation

2 A First Step: Using SDRP to Extend IDRP for Provider Selection Flexibility

Consider a service provider, "SP" connecting to 2 backbone providers, as figure 1. A stub client "C" connects to SP. We assume that all the domains either speak IDRP or are capable of doing so.

The problem is that in the HbH paradigm of IDRP, C is constrained to use the default provider "DBP" that SP is using.

[1] Rules for concatenating these fragments automatically will be discussed elsewhere. you put in anything more sensible here?

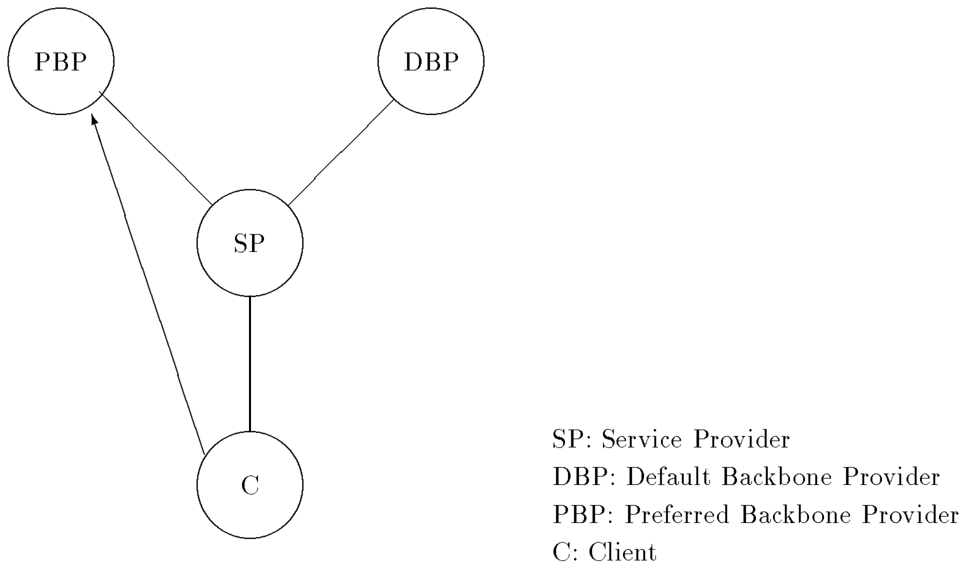

Figure 1: Provider Selection in the Internet today, using the unified routing architecture

If domain C can speak IDRP directly to its preferred provider "PBP," it can use SDRP to loose source route its traffic via PBP. This permits C to establish a "virtual adjacency" with PBP.

The following protocol mechanisms are needed to support provider selection.

- Protocol specifications to overcome the common subnetwork restrictions in IDRP.
- Mechanisms to install SDRP routes in the FIB. This mechanism will be useful for all of the route construction techniques we discuss in this document.

There is another technique called direct provider selection documented in [6], where a host in the domain requests a choice of provider other than that chosen by the BR (or RS) for that domain. Allowing hosts to select routes directly, instead of relying on a mapping by their BRs, is an important but orthogonal issue and will not be addressed further in this document.

3 Medium Term Goals

Provider selection only extends available routes to those to which the selected provider has access. When the routes provided by the local domain and by the selected providers do not meet the policy or service requirements of the client, we require mechanisms to construct alternate routes. We discuss two such mechanisms: RIB Query and Path Explorer. In the first, transits are queried as in provider selection, however RIB information for select destinations is requested, instead of complete FIB information for all destinations; moreover the request is done on demand, and the information is not continuously updated. In the second, an IDRP computation is initiated on-demand for a select destination.

3.1 RIB Query

Routes in an IDRP speaker's RIBs contain route fragments. Barring policy constraints, a client can connect to a remote peer, request and accumulate their RIBs, and build explicit routes.

A source domain attempting to reach a destination may prefer transit domains other than those that are chosen by its neighbouring domains. In this situation, the domain must either use the transits chosen by its neighbours, or acquire the RIBs from a remote peer. The mechanism by which a source domain can acquire RIBs from a remote peer and install explicit routes is called RIB Query. In the general policy-rich situation, the domain will have to find at least two route fragments, one from the preferred transit network to itself, and another from the transit to the destination.

The complete source route can be derived by concatenating the route fragments appropriately.

Another mechanism for computing route fragments (or complete source routes when the system is small) is described in the following section.

The following machinery is required in order for a source domain to perform a successful RIB Query:

- IDRP protocol modifications to permit retrieval of RIBs from a peer.
- Methods for source domains to specify transit domain preferences.
- We also need heuristics on whom to query.

3.2 Path Explorer

We define a new mechanism for computing Route Fragments called Path Explorer. Path Explorer was first proposed by Brijesh Kumar(UCL). Path Exploration is an IDRP distributed route computation initiated on-demand in which intermediate nodes: (1) do not exert their own route selection preferences (but they do apply transit policies), (2) propagate but do not store the particular routes. This has the added advantage of permitting the destination to influence the set of choices available to the source.

This path exploration can be done in any number of different ways. The simplest of such methods is for the source to request the destination to compute a path from it to the source; i.e.

explore(start-node)

This "brute force" *explore* computes end to end routes but is too expensive because the number of routes propagated grows exponentially with path length. To constrain route propagation we can exploit the following:

- Routes are already constrained by specified attributes in IDRP such as QoS etc.
- A hop count will constrain route propagation and frequently very long paths are not desired anyway.
- The source can specify a ranking of qualifying routes, and a cut-off, which the destination can use to limit the number of routes propagated.

- If the source knows of a small set of providers via which it will route its packets, it can request a path explorer from the destination to those intermediate proxy points, instead of all the way back to itself. For example:

  proxy-explore(start-node, intermediate-node, originator)

  or

  multi-proxy-explore(start-node, (set of intermediate-nodes), originator)

Requirements for Path Explorer

- Protocols to initiate the path explore and proxy.
- Specifications of the preference functions.
- Modifications in IDRP to perform the following:

  to propagate certain routes without applying local selection criteria.

  to not propagate routes that violate transit policies.

  to not store these routes in the longer term, either in the RIBs or the FIB.

  To store these on demand routes in a short lived cache for the purpose of evaluating viable routes to the cached route to decide which routes to propagate.

  modifications to the IDRP route selection process to use the source specified preference function.
- Implementations of path explorers.
- Mechanisms to constrain path explorers.

3.3 Path Verifier

We expect the results of RIB queries and Path Explorers to be cached. Caching implies that there is potential for the data to become invalid when it is actually used, or re-used. The data could become invalid either because one or more of the links used in the source route are down; or because the characteristics of the path changed from when they were queried for and computed, to when they are being used. Therefore, we need additional tools to verify the paths that are computed, either by RIB query, or path explorer. We call this tool Path Verifier.

4 Integrating Route Construction into the NSFnet Routing Architecture

The new NSFnet architecture consists of a very high speed backbone service (the vBNS); other Network Service Providers (NSPs); Network Attachment Points (NAPs) where the NSPs connect to the vBNS and to each other; and the Routing Arbiter (RA) to manage the routing process for the US portion of the global Internet.

The RA architecture[1] consists of the following sub-components:

- Routing Registry Database (RRDB)

  The RRDB contains the configured topology of the Internet. As part of this data, it contains domain level connectivity information, transit and route selection policies.

- Route Servers (RSs) on each NAP

  Each RS contains configurations computed by the RRDB. These configurations govern the exchange of routing information with the different routers connected to the RSs' NAP.

- RA-NOC

  The RA-NOC is the entity that monitors routing in the Internet. As such it does not itself influence route computation. We will not discuss this component in the remainder of the document.

In the following section, we will see how each of the techniques discussed can leverage off the RRDB and the RSs.

4.1 Application of the Route Construction Mechanisms

4.1.1 Provider Selection

Provider Selection is the technique of querying a particular domain for the routes it has pre-computed for itself, i.e. its FIB. By using these routes, a client is selecting the particular provider over others that were selected by other BGP/IDRP speakers closer to the client.

Any client domain that chooses to use a particular provider at the NAP can use the RS at that NAP to acquire the provider's routes. This makes it a single point of contact for the client domain to do provider selection.

A provider can defer connection and authentication issues for a class of client domains to the RS. The RS will be configured using information registered by the provider in the RRDB. Since the ability to hand out routes for provider selection is explicitly handed out through the RRDB, the provider can still retain control over the set of clients domains that can use it as a transit provider. For instance, and provider may be required to carry traffic for any NSF specified site, and may not wish to authenticate every single NSF sponsored site individually. On the other hand, the provider might wish to control the set of authentication and route exchange for some of its business clients, would prefer not to delegate this control to the RS.

The RRDB has relatively static information, and hence is not relevant in provider selection.

4.1.2 RIB Query

The RIB Query mechanism is used to obtain the domain level connectivity information available to the domain. This could be obtained in one of three ways: (1) by querying the RRDB directly for the appropriate connectivity, (2) by querying a particular domain directly for its RIBs, or (3) by querying the RS that serves that particular domain, for that domain's RIBs.

Contacting the RRDB is attractive because it refers only to a single point of contact. However it has the disadvantage that it is static information and does not reflect the active status of the topology at any time. Moreover, if the RRDB is contacted, a better option would be to have the RRDB compute the entire route. We discuss this further in section 4.3 below.

Contacting the domain provides more relevant information about the topology, but may require multiple queries to multiple domains to compute a complete route. Using an RS to make the query simplifies the computation.

The FIB for one domain is a subset of the RIBs from all its neighbour domains. Given that the RS at a NAP contains the FIB for all domains incident on the NAP, the RS has some subset of the RIBs for all domains incident on the NAP.

The RIB query mechanism can directly connect to the RS, and ask for a subset of the RIBs for a single domain of interest. By definition then, the RS can provide the RIBs for the domains that peer with it. Note that this may not be the complete set of RIBs for the domain, especially if it has RIBs from other domains not directly attached to the NAP.

The querying client can evaluate the amount of information available to it from the RS by comparing this information to the domain level connectivity information available to it from the RRDB.

Further evaluation is needed to determine whether performing the RIB query through an RS will be sufficient to compute routes in most instances.

4.1.3 Path Explorer

The source requests the destination domain to start an IDRP computation of routes with certain characteristics.

The source can request the RSs incident on a NAP that the domain is connected to, to start the appropriate IDRP flood for either a proxied, or multi-proxied path explorer. The source can additionally use one or more of the RSs as proxy agents.

4.1.4 Path Verifier

This tool may be used operationally prior to actually using a computed route to verify that the route chosen will be usable to the instant it is being used. By itself, this tool will have very little impact for the RA architecture. Hence if the RRDB computes full source routes on demand, this would be a much needed tool.

4.2 Requirements for the mechanisms

There are a number of issues to be resolved to compute routes in the RA architecture. Some of these requirements are:

- Policies need to be established regarding who shall and shall not connect to the RS and the RRDB to compute on-demand routes.
- We need protocols to query the RS and RRDB for various route requests and compute requests.
- Implementations to perform the query.

4.3 Computing Routes Via the RRDB

Since the RRDB is intended contain most of the domain-level topology of the Internet, and the RRDB is used to constructed configuration files for the RSs, it follows that the RRDB could also compute valid source routes that satisfy TOS requirements as well as route selection and transit policies.

Some of the advantages and disadvantages of using the RRDB to compute source routes is a set of tradeoffs discussed below:

- The RRDB is an ideal central location to obtain all information.

  The tradeoff with this is that this may be a hot spot deluged by requests to compute a large number of special case routes.
- The RRDB has complete information about domain level connectivity of the Internet.

  It should be noted that this complete information is necessarily static, and does not reflect the current status of the links and/or domains about which information is provided.
- The RRDB could pre-compute a large number of source routes for various policies and TOS.

  It is not clear in what range of cases this solution of pre-computing special purpose routes will be tractable.

The tradeoffs must be carefully considered, and the scalability of this approach needs to be investigated carefully.

5 Current Status

1. The Inter-Domain Routing WG(IDR-WG) at the IETF is working on removing the common subnetwork restrictions in BGP-4[9, 10, 7, 8], IDRP for IP[4], and IDRP for IPv6[11, 12].
2. Routing Information Filters (RIFs)[13] allow a client to perform a RIB Query to acquire the RIBs and/or FIB from a remote peer. A RIF update contains tuples of the form ⟨Action, NLRI, Scope, Base⟩. **Action** indicates to the remote peer how the RIF should be manipulated, and can be one of ALLOW, RESTRICT, or REMOVE. **Scope** constrains the set of routes that the remote peer must consider for NLRI. Scope can be one of EXACT, REFINE, NORMAL, or REACHABILITY. **Base** indicates whether the routes must be returned from the remote peer's RIBs or its FIB.

   The RIF is sent as part of the BGP update message in BGP-4, or in the OPEN or RIB REFRESH messages in IDRP for IPv6.

A Interfaces with IDRP

The following outlines existing IDRP constraints that should be modified for easier inter-operation.

1. IDRP requires peers to share a common subnetwork. This requirement should be relaxed to allow a remote peer to either choose a preferred provider, as outlined in section 2, or to query a remote peer for RIBs, as described in section 3.1.
2. IDRP must provide its Adj-RIB-In on demand to support mechanisms such as RIB query (section 3.1).
3. IDRP must provide mechanisms for a client to specify source specific preference functions, which it must then apply in lieu of its standard selection criteria for the requested path.
4. IDRP must provide a mechanism for a peer to advertise a set of routes as being on-demand routes.
5. IDRP must provide a mechanism by which a peer that gets a route marked as being an on-demand route must not apply any local selection criteria.
6. An IDRP peer need not store on-demand routes in its local RIB.
7. IDRP peers may maintain a short-lived cache to compare viable on-demand routes and decide which ones they should propagate.

Since a route that violates transit criteria is not a valid route, an IDRP peer continues to apply its transit criteria in all of the techniques outlined in this paper.

Security Considerations

Security Considerations are not discussed in this memo.

Author's Present Addresses

Kannan Varadhan
USC/Information Sciences Institute
4676, Admiralty Way,
Marina Del Ray, Ca 90272.
Phone: +1 310 822 1511 x 402
email: kannan@isi.edu

Deborah Estrin
USC/Information Sciences Institute
4676, Admiralty Way,
Marina Del Ray, Ca 90272.

Phone: +1 310 822 1511 x 253
email: estrin@isi.edu

Steve Hotz
USC/Information Sciences Institute
4676, Admiralty Way,
Marina Del Ray, Ca 90272.
Phone: +1 310 822 1511
email: hotz@isi.edu

Yakov Rekhter
T.J. Watson Research Center, IBM Corporation
P.O. Box 218
Yorktown Heights, NY 10598
Phone: +1 914 945 3896
e-mail: yakov@watson.ibm.com

References

[1] D. Estrin, J. Postel, and Y. Rekhter. Routing arbiter architecture. *conneXions*, 8(8):2–7, August 1994.

[2] D. Estrin, Y. Rekhter, and S. Hotz. Scalable inter-domain routing architecture. In *Proceedings of the ACM SigComm*, pages 40–52, Baltimore MD, August 1992. ACM.

[3] D. Estrin, D. Zappala, T. Li, Y. Rekhter, and K. Varadhan. Source Demand Routing: Packet format and forwarding specification (version 1). *Internet Draft, SDR Working Group*, March 1994. This document denotes work in progress.

[4] S. Hares. IDRP for IP. *Internet Draft, IDR Working Group*, 1994. This document denotes work in progress.

[5] ISO. *Protocol for Exchange of Inter-domain Routeing Information among Intermediate Systems to Support Forwarding of ISO 8473 PDUs*, ISO/IEC/JTC1/SC6 IS10747 edition.

[6] Y. Rekhter. Selecting a direct provider, 1994. This document denotes work in progress.

[7] Y. Rekhter and P. Gross; Editors. Application of the Border Gateway Protocol in the Internet. *RFC 1655*, July 1994. USC/Information Sciences Institute.

[8] Y. Rekhter and P. Gross; Editors. Application of the Border Gateway Protocol in the Internet. *Internet Draft, IDR Working Group*, December 1994. This document denotes work in progress.

[9] Y. Rekhter and T. Li; Editors. A Border Gateway Protocol 4 (BGP-4). *RFC 1654*, July 1994. USC/Information Sciences Institute.

[10] Y. Rekhter and T. Li; Editors. A Border Gateway Protocol 4 (BGP-4). *Internet Draft, IDR Working Group*, December 1994. This document denotes work in progress.

[11] Y. Rekhter and P. Traina. IDRP for IPv6. *Internet Draft, IDR Working Group*, January 1995. This document denotes work in progress.

[12] Y. Rekhter and P. Traina. IDRP for IPv6. *Internet Draft, IDR Working Group*, January 1995. This document denotes work in progress.

[13] Y. Rekhter, K. Varadhan, and C. Villamizar. Extensions for selective updates in Inter-Domain routing. *Internet Draft, IDR Working Group*, February 1995. This document denotes work in progress.