Network Working Group
Request for Comments: 1940
Category: Informational

D. Estrin
USC
T. Li
Y. Rekhter
cisco Systems
K. Varadhan
D. Zappala
USC

May 1996

Source Demand Routing: Packet Format and Forwarding Specification (Version 1).

Status of this Memo

1. Overview

The purpose of SDRP is to support source-initiated selection of routes to complement the route selection provided by existing routing protocols for both inter-domain and intra-domain routes. This document refers to such source-initiated routes as "SDRP routes". This document describes the packet format and forwarding procedure for SDRP. It also describes procedures for ascertaining feasibility of SDRP routes. Other components not described here are routing information distribution and route computation. This portion of the protocol may initially be used with manually configured routes. The same packet format and processing will be usable with dynamic route information distribution and computation methods under development.

The packet forwarding protocol specified here makes minimal assumptions about the distribution and acquisition of routing information needed to construct the SDRP routes. These minimal assumptions are believed to be sufficient for the existing Internet. Future components of the SDRP protocol will extend capabilities in this area and others in a largely backward-compatible manner.

This version of the packet forwarding protocol sends all packets with the complete SDRP route in the SDRP header. Future versions will address route setup and other enhancements and optimizations.

2. Model of operations

An Internet can be viewed as a collection of routing domains interconnected by means of common subnetworks, and Border Routers (BRs) attached to these subnetworks. A routing domain itself may be composed of further subnetworks, routers interconnecting these subnetworks, and hosts. This document assumes that there is some type of routing present within the routing domain, but it does not assume that this intra-domain routing is coordinated or even consistent.

For the purposes of this discussion, a BR belongs to only one domain. A pair of BRs, each belonging to a different domain, but attached to a common subnetwork, form an inter-domain connection. By definition, packets that traverse multiple domains must traverse BRs of these domains. Note that a single physical router may act as multiple BRs for the purposes of this model.

A pair of domains is said to be adjacent if there is at least one pair of BRs, one in each domain, that form an inter-domain connection.

Each domain has a globally unique identifier, called a Domain Identifier (DI). All the BRs within a domain need to know the DI assigned to the domain. Management of the DI space is outside the scope of this document. This document assumes that Autonomous System (AS) numbers are used as DIs. A domain path (or simply path) refers to a list of DIs such as might be taken from a BGP AS path [1, 2, 3] or an IDRP RD path [4]. We refer to a route as the combination of a network address and domain paths. The network addresses are represented by NLRI (Network Layer Reachability Information) as described in [3].

This document assumes that the routing domains are congruent to the autonomous systems. Thus, within the content of this document, the terms autonomous system and routing domain can be used interchangeably.

An application residing at a source host inside a domain, communicates with a destination host at another domain. An intermediate router in the path from the source host to the destination host may decide to forward the packet using SDRP. It can do this by encapsulating the entire IP packet from the source host in an SDRP packet. The router that does this encapsulation is called the "encapsulating router."

2.1 SDRP routes

A component in an SDRP route is either a DI (AS number) or an IP address. Thus, an SDRP route is defined as a sequence of domains and routers, syntactically expressed as a sequence of DIs and IP addresses. Thus an SDRP route is a collection of source routed hops.

Each component of the SDRP route is called a "hop." The packet traverses each component of the SDRP route exactly once. When a router corresponding to one of the components of the SDRP route receives the packet from a router corresponding to the previous component of the SDRP route, the router will process the packet according to the SDRP forwarding rules in this packet. The next component of the SDRP route that this router will forward the packet to, is called the "next hop," with respect to this router and component of the SDRP route.

An SDRP hop can either be a "strict" source routed hop, or a "loose" source routed hop. A strict source route hop is one in which, if the next hop specified is a DI, refers to an immediately adjacent domain, and the packet will be forwarded directly to a route within the domain; if the next hop specified is an IP address, refers to an immediately adjacent router on a common subnetwork. Any other kind of a source route hop is a loose source route hop.

A route is a "strict source route" if the current hop being executed is processed as a strict source route hop. Likewise, a route is a "loose source route" if the current hop being executed is processed as a loose source route hop.

It is assumed that each BR participates in the intra-domain routing protocol(s) (IGPs) of the domain to which the BR belongs. Thus, a BR may forward a packet to any other BR in its own domain using intra-domain routing procedures. Forwarding a packet between two BRs that form an inter-domain connection requires neither intra-domain nor the inter-domain routing procedures (an inter-domain connection is a common Layer 2 subnetwork).

It is also assumed that all routers participate in the intra-domain routing protocol(s) (IGPs) of the domain to which they belong.

While SDRP does not require that all domains have a common network layer protocol, all the BRs in the domains along a given SDRP route are required to support a common network layer. This document specifies SDRP operations when that common network layer

protocol is IP ([5]).

While this document requires all the BRs to support IP, the document does not preclude a BR from additionally supporting other network layer protocols as well (e.g., CLNP, IPX, AppleTalk). If a BR supports multiple network layers, then for the purposes of this model, the BR must maintain multiple Forwarding Information Bases (FIBs), one per network layer.

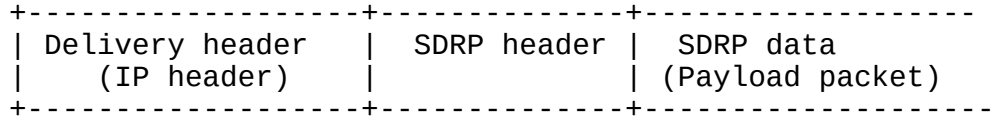
2.2 SDRP encapsulation

Forwarding an IP packet along an SDRP route is accomplished by encapsulating the entire packet in an SDRP packet. An SDRP packet consists of the SDRP header followed by the SDRP data. The SDRP header carries the SDRP route constructed by the domain that originated the SDRP packet. The SDRP data carries the original packet that the source domain decided to forward via SDRP.

An SDRP packet is carried across domains as the data portion of an IP packet with protocol number 42.

This document refers to the IP header of a packet that carries an SDRP packet as the delivery IP header (or just the delivery header). This document refers to the packet carried as SDRP data s the payload packet, and the IP header of the payload packet is the payload header.

Thus, an SDRP Packet can be represented as follows:



Each SDRP route may have an MTU associated with it. An MTU of an SDRP route is defined as the maximum length of the payload packet that can be carried without fragmentation of an SDRP packet. This means that the SDRP MTU as seen by the transport layer and applications above the transport layer is the actual link MTU less the length of the Delivery and SDRP headers. Procedures for MTU discovery are specified in Section 9.

2.3 D-FIB

It is assumed that a BR participates in either BGP or IDRP. A BR participating in SDRP augments its FIBs with a D-FIB that contains routes to domains. A route to a domain is a triplet <DI, Next-Hop, NLRI>, where DI depicts a destination domain, Next-Hop

depicts the IP address of the next-hop BR, and NLRI depicts the set of reachable destinations within the destination domain. D-FIBs are constructed based on the information obtained from either BGP, IDRP, or configuration information.

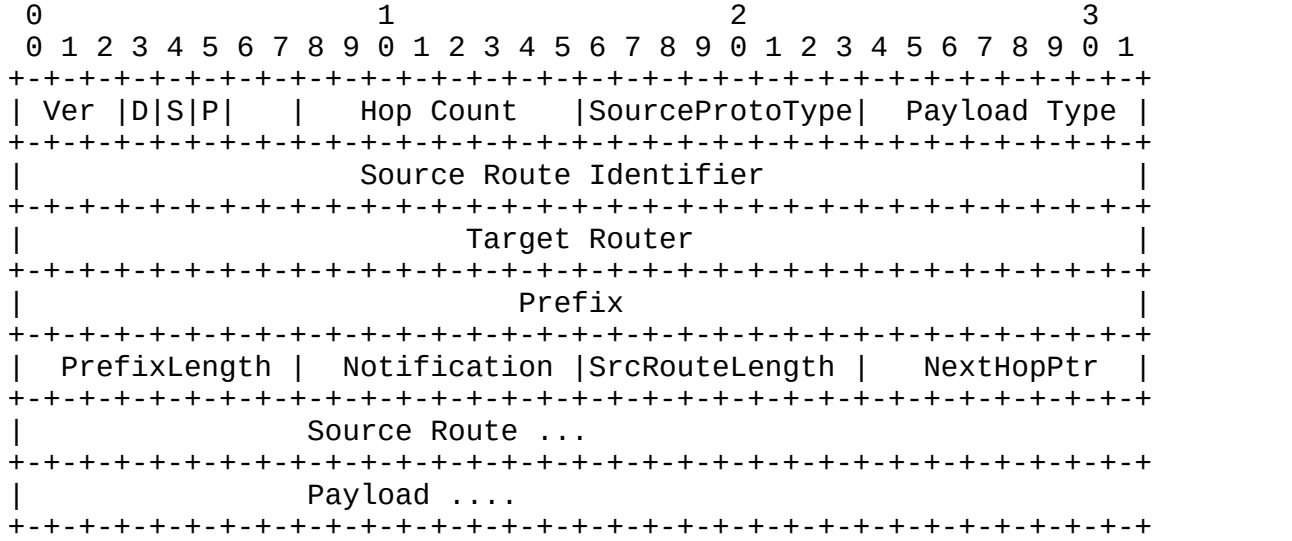
An SDRP packet is forwarded across multiple domains by utilizing the forwarding databases (both FIBs and D-FIBs) maintained by the BRs.

The operational status of SDRP routes is monitored via passive (Error Reporting) and active (Route Probing) mechanisms. The Error Reporting mechanism provides the originator of the SDRP route with a failure notification. The Probing mechanism provides the originator of the SDRP route with confirmation of a route's feasibility.

3. SDRP Packet format

The total length of an SDRP packet (header plus data) can be determined from the information carried in the delivery IP header. The length of the payload packet can be determined from the total length of an SDRP packet and the length of its SDRP Header.

The following describes the format of an SDRP packet.



Version and Flags (1 octet)

The SDRP version number and control flags are coded in the first octet. Bit 0 is the most significant bit, bit 7 is the least significant bit.

Version (bits 0 through 2)

The first three bits contain the Version field indicating the version number of the protocol. The value of this field is set to 1.

Flags (bits 3 through 7)

Data packet/Control packet (bit 3)

If the bit is set to 1, then the packet carries data.

Otherwise, the packet carries control information.

Loose/Strict Source Route (bit 4)

The Loose/Strict Source Route indicator is used when making a forwarding decision (see Section 5.2). If this bit is set to 1, it indicates that the next hop is a Strict Source Route Hop. If this bit is set to 0, it indicates that the next hop is a Loose Source Route.

Probe Indicator (bit 5)

The Probe Indicator is used by the originator of the route to request verification of the route's feasibility (see Sections 4 and 7.1). If this bit is set to 1, it indicates that the originator is probing the route. This bit should always be set to 0 for control packets.

Hop Count (1 octet)

The Hop Count field carries the maximum number of routers an SDRP data packet may traverse. It is decremented by 1 as an SDRP data packet traverses a router which forwards the packet using SDRP forwarding. Once the Hop Count field reaches the value of 0, the router should discard the data packet and generate a control packet (see Section 5.2.6). A router that receives a packet with a Hop Count value of 0 should discard the data packet, and generate a control packet (see Section 5.2.6).

Source Route Protocol Type (1 octet)

The Source Route Protocol Type fields indicates the type of information that appears in the source route. The value 1 in this field indicates that the contents of the source route are as described in this document and indicates an Explicit Source

Route. The value 2 in this field indicates a Route Setup. The syntax of the source route for this value is identical to a value of 1, but also has additional semantics which are defined in other documents.

Payload Protocol Type (1 octet)

The Payload Protocol Type field indicates the protocol type of the payload. If the payload is an IP datagram, then this field should contain the value 1.

Note that this Payload Protocol Type is not the same as the IP protocol type[5,7].

Source Route Identifier (4 octets)

The BR that originates the SDRP packet should insert a 32 bit value in this field which will serve as an identifier for the source route. This value needs to be unique only in the context of the originating BR.

Target Router (4 octets)

This field is meaningful only in control packets.

The Target Router field contains one of the IP addresses of the router that originated the SDRP packet that triggered the control packet to be returned.

Prefix (4 octets)

The Prefix field contains an IP address prefix. Only the number of bits specified in the Prefix Length are significant. The Prefix field is used to prevent routing loops when using BGP or IDRP to route to the next AS in a loose source route (see Section 4).

Prefix Length (1 octet)

The Prefix Length field indicates the length in bits of the IP address prefix. A length of zero indicates a prefix that matches all IP addresses.

Notification Code (1 octet)

This field is only meaningful in control packets. In data packets, this field is transmitted as zero, and should be ignored on receipt.

This document defines the following values for the Notification Code:

1 - No Route Available

2 - Strict Source Route Failed

3 - Transit Policy Violation

4 - Hop Count Exceeded

5 - Probe Completed

6 - Unimplemented SDRP version

7 - Unimplemented Source Route Protocol Type

8 - Setup Request Rejected

Source Route Length (1 octet)

The Source Route Length field indicates the length in 32 bit words of the domain level source route carried in the SDRP Header.

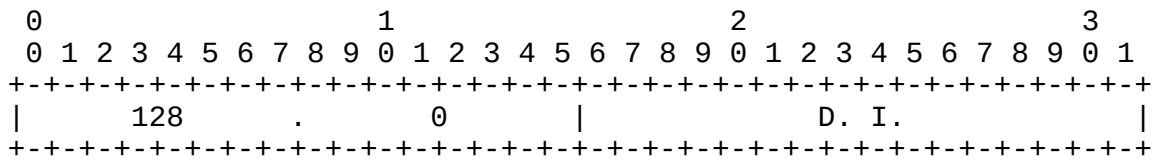
Next Hop Pointer (1 octet)

The Next Hop Pointer field indicates the offset of the high-order byte of the next hop along the route that the packet has to be forwarded. This offset is relative to the start of the Source Route field; so if the value of the Next Hop Pointer field equals the value of the Source Route Length field, then the entire source route has been completely traversed. All other source routes are said to be incompletely traversed.

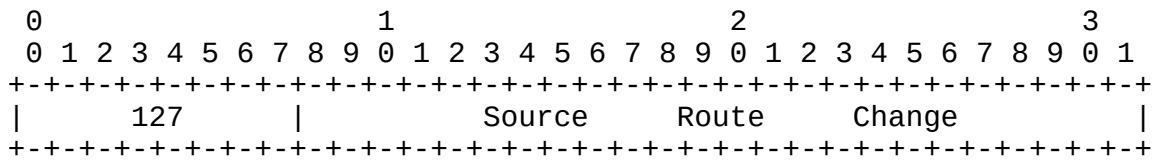
Source Route (variable)

The components of the source route are syntactically IP addresses.

An IP address from network 128.0.0.0 is used to encode a next hop that is a domain. The least significant two octets contain the DI, which is an Internet Autonomous System number.



An IP address from the network 127.0.0.0 is used to encode characteristics of the source route. The least significant three octets are used as a Source Route Change field.



Source Route Change (3 octets)

Loose/Strict Source Route Change (bit 1)

The Loose/Strict Source Route Change bit reflects a new value of the Loose/Strict Source Route bit in the SDRP header. The value of the Loose/Strict Source Route Change bit is copied into the Loose/Strict Source Route bit in the SDRP header when a Source Route Change field is encountered in processing an SDRP packet.

The rest of the Source Route Change field is transmitted as zero, and should be ignored on receipt.

Payload (variable)

The Payload field carries the datagram originated by the end-system within the domain that constructed the SDRP packet. The Payload field forms the data portion of the SDRP packet. In a control packet this field may be empty or may carry the payload header of the packet that triggered the control message (see 5.2.5). Note that there is no padding between the Source Route and the Payload, and that the Payload may start at any arbitrary octet boundary.

4. Originating SDRP Data packets

This document assumes that a router that originates SDRP packets is preconfigured with a set of SDRP routes. Procedures for constructing these routes are outside the scope of this document. SDRP packet forwarding may be deployed initially without additional routing protocol support.

An application on a source host generates packets that must be delivered to a given destination. The packet traverses the Internet by following normal hop-by-hop routing information. An intermediate router in the path between the source host and the destination host may decide to forward some of these packets via SDRP.

When this router receives an IP datagram, the router uses the information in the datagram and the local criteria to determine whether the datagram should be forwarded along a particular SDRP route. Associated with each set of criteria is a set of one or more SDRP routes that should be used to route matching packets. The exact nature of the criteria is a local matter. The only restrictions this document places on the applicability of SDRP routes is that an IP datagram that contains a strict source route should not be forwarded along an SDRP route, that SDRP encapsulation should never be applied to an SDRP packet, and that if SDRP is used with inter-domain routes, the destination domain must also run SDRP.

If the router decides to forward a datagram along a particular SDRP route, the router constructs the SDRP packet by placing the original datagram into the Payload field of the SDRP packet and constructing the SDRP header based on the selected SDRP route. The Next Hop pointer is set to 0 (the first entry in the Source Route field of the SDRP packet). The value of the Time To Live field in the payload header should be copied into the Hop Count field of the SDRP header.

Even if we assume that interior routing is loop free, it is possible, either due to the state of inter-domain routing or due to other SDRP routers, that a domain level source route that does not terminate with the intended destination domain may lead a packet into a routing loop. Originating SDRP routers that wish to insure that this does not occur should include a final domain level hop of the destination's domain, i.e. specify the SDRP route as <DI1, DI2, DI3> instead of <DI1, DI2>, if the destination host is in domain DI3. The means for determining the DI of the destination domain is outside of the scope of this document.

Similarly, when using SDRP for interior routing, it is possible that the source route does not coincide with IGP routing. In this case, one means of preventing a loop is to specify the last hop router's IP

address as the last address within the source route. The encapsulating router can do this by specifying the source route to reach destination host IP3 as <IP1, IP2, IP3> instead of <IP1, IP2>.

The source address field in the delivery header should contain an IP address of the router. The value of the Don't Fragment flag of the delivery header is copied from the Don't Fragment flag of the payload header. The value of the Type Of Service field in the delivery header is copied from the Type Of Service field in the payload header. If the payload header contains an IP security option, that option is replicated as an option in the delivery header. All other IP options in the payload header must be ignored.

If the SDRP route that is used is learned from IDRP, then the TOS corresponding to this route is copied into the TOS field in the delivery header.

The resulting SDRP packet is then forwarded as described in Section 5.2.2.

If the encapsulating router decides to forward a datagram along a particular SDRP route that has an MTU smaller than the length of the datagram, then if the payload header has the Don't Fragment flag set to 1, the router should generate an ICMP Destination Unreachable message with a code meaning "fragmentation needed and DF set" in accordance with [6]. The ICMP message must be sent to the original source host. The router should then discard the original datagram.

If a router has learned an MTU for a particular SDRP route, either via ICMP messages or via configuration information, and it determines that an SDRP packet must be fragmented before transmission, then it first calculates the the effective MTU seen by the payload packet. If the effective MTU is greater than or equal to 512 bytes, the router SHOULD first fragment the payload packet using normal IP fragmentation. SDRP packets are then constructed for each fragment, as describe above. Otherwise, the router should first form the SDRP packet, and then fragment it.

A router may use locally originated SDRP packets to verify the feasibility of its SDRP routes. To do this the router sets the value of the Probe Indicator field in the SDRP packet to 1. Receipt of an SDRP control packet by the originating router with the "Probe Completed" Notification Code (see Section 7.1) indicates feasibility of the SDRP route. Persistent lack of SDRP control packets with the "Probe Completed" Notification Code should be used as an indication that the associated SDRP route is not feasible.

5. Processing SDRP packets

We say that a router receives an SDRP packet if the destination address field in the delivery header of the packet arriving at the router contains one of the IP addresses of the router.

When a router receives an SDRP packet, the router extracts the Source Route Protocol field from the SDRP header.

5.1 Supporting Transit Policies

A router may be able to verify that a packet that it is given to forward does not violate any of the transit policies that may exist, of the domain to which the router belongs. Specific verification mechanisms are a matter that is local to the router and are outside the scope of this document.

The restriction on the verification mechanisms is that they may take into account only the contents of the SDRP header, the payload header, and transport protocol header of the payload packet.

With SDRP a domain may enforce its transit policies by applying filters based on the information present in the IP Header. For example a router may initially carefully filter all SDRP traffic from all possible sources. A filter that allows certain SDRP traffic from selected sources to pass through the router could then be installed dynamically to pass similar types of traffic. Thus, by caching appropriate filtering information, a transit domain can efficiently support transit policies. Other mechanisms for supporting transit policy and implementation techniques are not precluded by this document.

If the router detects that the SDRP packet violates a domain’s transit policy it sends back an SDRP control packet to the encapsulating router and discards the violating packet.

SDRP control packets are not subject to transit policies.

If a router does not discard an SDRP packet due to a transit policy violation, then the router attempts to forward it as specified in Section 5.2.

5.2 Forwarding SDRP packets

Procedures for forwarding of an SDRP packet depend on

a) whether the router has the routing information needed to forward the packet;

b) whether the SDRP route has been completely traversed;
c) whether the SDRP route is strict or loose, and
d) whether the packet is a data or control packet.

When forwarding an SDRP packet (either data or control) a router should not modify the following fields in the delivery header:

a) Source Address
b) Don’t Fragment flag

If the Source Route Protocol Type of a packet indicates a Route Setup and the router does not or cannot support setup, the router MAY send the encapsulating router a control packet with a Notification Code of Setup Request Rejected. It MAY then modify the data packet so that the Source Route Protocol Type is Explicit Source Route and the Probe Indicator bit is 0, then forwards the packet as described below. The router MAY send notification of a failed setup request only periodically. Alternately, a router MAY silently drop the Route Setup packet.

5.2.1 Forwarding algorithm pseudo-code

The following pseudo-code gives an overview of the SDRP forwarding algorithm. Please consult the text below for more details.

Let LOCAL_DI be the DI of the domain of the local system, let NEXT_HOP be the next hop in the source route if the source route has not been completely traversed, let NEXT_DI be the DI portion of NEXT_HOP if NEXT_HOP is from network 128.0.0.0, and let NEXT_ROUTER be the IP address of the next router if the packet is to be forwarded using SDRP. We say that NEXT_DI is adjacent if the local domain is adjacent to the domain that has NEXT_DI as its DI, and we say that NEXT_ROUTER is adjacent if it represents an IP address of a router that shares a link with the current router. Normal IP forwarding refers to forwarding that can be accomplished using FIBs constructed via BGP, IDRP or one or more IGPs.

The pseudo code requires sending control messages in a number of places. All such control messages must be sent to the encapsulating router, which is indicated in the source address of the delivery header. Note too that all intermediate SDRP routers that process an SDRP packet must ensure that the source address of the delivery header is left untouched, since this source address is the address of the encapsulating router to which any control messages must be sent.

```
if the packet is a control packet begin
  if the Target Router equals an address assigned to the
    local router begin
    remove the delivery header
    process information carried in the control packet
    return
  end if
  if the packet can be forwarded using normal IP forwarding begin
    set Next Hop Pointer to Source Route Length
    forward the packet using normal IP forwarding
    return
  end if
end if

if the version field is not 1 begin
  if the packet is a data packet begin
    generate a control packet with "Unimplemented SDRP version"
  end if
  discard the packet
  return
end if

if the source route protocol type is not 1 begin
  if the packet is a data packet begin
    generate a control packet with "Unimplemented source route
      protocol type"
  end if
  discard the packet
  return
end if



if the Hop Count field is greater than 0 begin
  decrement the Hop Count field
end if
if the Hop Count field is 0 begin
  if the packet is a data packet begin
    generate a control packet with "Hop Count Exceeded"
 end if
  discard the packet
  return
end if


if the packet is a data packet begin
  if the packet violates transit policy begin
    generate a control packet with "Transit Policy Violation"
```

```
        discard the data packet
        return
      end if
    end if

    set mode to NONE
    set advanced to FALSE
    if Next Hop Ptr does not equal Source Route Length begin
      set NEXT_HOP to the next hop in the source route
      while mode equals NONE begin
        if NEXT_HOP is from network 127.0.0.0 begin
          set the Loose/Strict Source Route bit equal to
              the Loose/Strict Source Route Change bit
        else if NEXT_HOP is from network 128.0.0.0 begin
          set NEXT_DI to the least significant two octets of NEXT_HOP
          if NEXT_DI is not equal to LOCAL_DI begin
            set mode to DOMAIN
          end if
        else if NEXT_HOP does not equal an address assigned to the
          local router begin
          set mode to LOCAL
        end if
        if mode equals NONE begin
          set advanced to TRUE
          increment the Next Hop Pointer field
          if Next Hop Pointer equals Source Route Length begin
            set mode to COMPLETE
          else
            set NEXT_HOP to the next hop in the source route
          end if
        end if
      end while
    end if


    if mode equals DOMAIN begin
      set route to NONE
      if the source route is loose begin
        if not advanced begin
          find the route, if any, based on Prefix and Prefix Length
          if the route is an aggregate formed at the local router begin
            set route to NONE
          end if
        end if
        if route equals NONE begin
          select a BGP or IDRP route, if any, with a path that includes
            NEXT_DI and is not an aggregate formed at the local router
          if route equals NONE begin
```

```
            if the packet is a data packet begin
             generate a control packet with "No Route Available"
            end if
            discard the packet
            return
          end if
          copy the NLRI from the route to the Prefix and Prefix Length
        end if
        if the route is an IDRP route begin
          set appropriate TOS in delivery header
        end if
        set NEXT_ROUTER from the route
      else
        set NEXT_ROUTER from the routing information for NEXT_DI
          using the D-FIB
        if route equals NONE begin
          if the packet is a data packet begin
            generate a control packet with "No Route Available"
          end if
          discard the packet
          return
        end if
        if NEXT_DI is not adjacent begin
          if the packet is a data packet begin
            generate a control packet with "Strict Source Route Failed"
          end if
          discard the packet
          return
        end if
      end if
      end if
    end if


    if mode equals LOCAL begin
      set NEXT_ROUTER equal to NEXT_HOP
      if the source route is strict and NEXT_ROUTER is not
        adjacent begin
        if the packet is a data packet begin
          generate a control packet with "Strict Source Route Failed"
        end if
        discard the packet
        return
      end if
    end if

    if mode equals LOCAL or mode equals DOMAIN begin
      set the destination address of the delivery header equal
```

```
        to NEXT_ROUTER
      checksum the delivery header
      route packet to NEXT_ROUTER using normal IP forwarding
      return
    end if

    if the packet is a control packet begin
      discard the packet
    end if
    remove the delivery header and the SDRP Header
    if there is no normal IP route to the payload destination begin
      generate a control packet with "No Route Available"
      discard the data packet
      return
    end if
    forward the payload using normal IP forwarding
    if the probe bit is set begin
      generate a control packet with "Probe Completed"
    end if
```

5.2.2 Handling an SDRP control packet.

An SDRP control packet is indicated by 0 in the Data packet/Control packet bit in the Flags field in the SDRP Header.

If the Target Router field of the received SDRP packet contains an IP address that is assigned to the router that received this SDRP packet, then the router should use the information carried in the Notification Code field, the Source Route Identifier field and the information carried in the Payload field to update the status of its SDRP routes. Details of such procedures are described in Section 7.

Otherwise, the router checks whether it can forward the packet to the router specified in the Target Router field by using the routing information present in its local FIB. If forwarding is possible then the local system sets the destination address of the delivery header to the address specified in the Target Router field, and hands the packet off for normal IP forwarding. If normal IP forwarding is impossible then the packet may be forwarded in the same manner as an SDRP data packet (described below) but with the following exceptions.

- Control packets are not subject to transit policies.
- In no case should a control packet be generated in response to an error caused by a control packet.
- If the source route is completely traversed and the packet still cannot be forwarded via normal IP routing, the packet should be silently dropped.

5.2.3 Handling an SDRP data packet.

An SDRP data packet is indicated by a one in the Data packet/Control packet bit in the Flags field in the SDRP Header.

An SDRP data packet is forwarded by sending the packet along the source route in the SDRP Header. When the source route is completely traversed and the packet has reached the destination domain, the payload may be removed from the data packet and forwarded normally. Further details are described below.

5.2.4 Checking the SDRP version number

An SDRP packet that has a version number other than 1 should be discarded. If the SDRP packet was a data packet, then a control packet with the Notification Code "Unimplemented SDRP version" should be generated as specified in section 6.

5.2.5 Checking the Source Route Protocol Type

This document describes Source Route Protocol Type 1. An SDRP router may support multiple Source Route Protocol Types; however an SDRP router is NOT required to support all defined Source Route Types. Any packet that has a Source Route Protocol Type which is not supported should be discarded. If the SDRP packet was a data packet, then a control packet with the Notification Code "Unimplemented Source Route Protocol Type" should be generated as specified in section 6.

5.2.6 Decrementing and checking Hop Count

If an SDRP packet is to be forwarded and the Hop Count field is non-zero, the Hop Count field should be decremented. If the resulting value is zero and the packet was a data packet, then a control packet with the Notification Code "Hop Count Exceeded" should be generated and sent to the encapsulating router as specified in section 6, and the packet should be discarded. If the resulting value is zero and the packet was a control packet, the packet should be discarded. The payload of the control packet should carry the payload header followed by 64 bits of the payload data of the data packet.

5.2.7 Upholding transit policies

It is not a goal of SDRP to create a security routing system. Therefore, we need to qualify our use of the term "upholding transit policy". It is assumed that transit policies have the nature of a "gentleperson’s agreement", and are upheld by all the participants. In other words, it is assumed that there will be no malicious

attempts to violate transit policies and that parties will rely on auditing and post facto detection of violations. When a security architecture is developed for IP or other network protocols then it may be applied to increase the assurance of transit policy enforcement. These issues are beyond the scope of this document.

A router may examine any data packet to verify if it complies with local transit policies, as described in section 5.1. If the verification fails, the router generates a control packet. If the verification referred to only the contents of the SDRP header, then the payload field of the control packet should be empty. If the verification referred to both the contents of the SDRP header and the payload header, then the payload field of the control packet should carry the payload header. If the verification referred to the transport protocol header, then the payload field of the control packet should carry the payload header and the transport header.

The Notification Code field of the SDRP header in the control packet is set to Transit Policy Violation. The procedures for constructing the rest of the SDRP Header of the control packet are specified in Section 6.

5.2.8 Partially traversed source routes

If a router receives an SDRP packet with a partially traversed source route, it extracts the next hop of the source route from the Source Route field. The router locates the high-order byte of the appropriate hop by using the Next Hop Pointer field as a 32 bit word offset relative to the start of the Source Route field. The next hop is always four octets long. The following procedure is used to interpret the next hop.

Syntactically, each element in the source route appears as an IP address. There are three encodings for the next hop:

a) The next hop is an address in network 127.0.0.0. In this case, the Loose/Strict Source Route field is set equal to the Loose/Strict Source Route Change bit. Then the Next Hop Pointer is incremented, the next hop is read from the Source Route field, and these three cases are examined again.

b) The next hop is an address in network 128.0.0.0. In this case, the DI of the next domain is extracted from the least significant two octets of the next hop. If the extracted DI is the same as the DI of the local domain, then the Next Hop Pointer is incremented, the next hop is read from the Source Route field, and these three cases are examined again. Otherwise, if the extracted DI is different from the DI of the local domain, the next hop is the extracted DI, and the

forwarding process may proceed.

c) The next hop is any other IP address. If the next hop is equal to any IP address assigned to the local router, the Next Hop Pointer is incremented, the next hop is read from the Source Route field, and these three cases examined again. Otherwise, the next hop is the IP address of the next router in the source route and the forwarding process may proceed.

The above procedure for interpreting the next hop in the source route finishes when the next hop is either a router other than the local router or an encoded DI that is not the local DI or a completed source route.

If upon termination of this procedure the source route is completely traversed, see section 5.2.9.

5.2.8.1 Finding a route to the next hop

If the next hop is not a DI, then the destination address in the delivery header is replaced by the next hop address and the resulting packet can then be forwarded using normal IP forwarding. Otherwise, a DI was extracted from the next hop in the source route, and the following procedure is used to find a route to the next domain.

Given the DI of the next domain, the router next consults its D-FIB. If no entry exists in the D-FIB for the next domain, then the packet should be discarded. If the packet was a data packet, a control message with Notification Code "No Route Available" should be generated as specified in Section 6. No other actions are necessary.

If there is a D-FIB entry, the router next examines the SDRP header to determine if the packet specified a strict source route. If so, and the next domain is not adjacent to the local domain, then a control packet with the Notification Code "Strict Source Route Failed" should be generated, as specified in section 6, and the original packet should be discarded. No other actions are necessary.

If source route is loose, then BGP or IDRP information must be used to insure that there is no loop in reaching the next hop. If the Next Hop Pointer was incremented when determining the next hop, then the router must select a BGP or IDRP route with a path that includes the extracted DI, and the NLRI for this route is copied into the Prefix Length and Prefix fields.

Otherwise, the Next Hop Pointer was not incremented, and the router should use the information carried in the Prefix and Prefix Length as an index into its BGP or IDRP routing table. If it finds a matching

route then it must select the corresponding D-FIB entry. If the route was formed locally by aggregation, then the router must consult its D-FIB and select any route with a path that includes the extracted DI. The NLRI for this route should be copied into the Prefix Length and Prefix fields.

In either case, the D-FIB entry includes the IP address of the next SDRP-speaking router to which the SDRP packet should be routed. The destination address in the delivery header is replaced by this address. The resulting packet can then be forwarded using normal IP forwarding.

5.2.8.2 Last Hop Optimization

A small optimization can be performed if there is only a single DI or IP address in the source route that has not been traversed.

In this case, if the next hop in the SDRP route is a DI, that DI is adjacent to the router processing this packet, the route has a route to the destination address in the payload header in its FIB, and this FIB route passes through the adjacent domain, then the source route may be considered completely traversed and processing may proceed as in section 5.2.9.

If the next hop in the SDRP route is an IP address, that IP address is adjacent to the router processing this packet, the router has a route to the destination address in the payload header in its FIB, and this FIB route passes through the adjacent IP address, then the source route may be considered completely traversed and processing may proceed as in section 5.2.9.

Since the last hop optimization may only be done if the last hop is directly adjacent, and reachable, it is irrelevant whether the SDRP route specifies that this is a strict source route or a loose source route hop.

5.2.9 Completely Traversed source routes

If the SDRP packet received by a router with a completely-traversed source route is a control packet and if the Target Router field carries an IP address assigned to the router, then the packet should be processed as specified in Section 7. Otherwise, if the SDRP packet is a control packet, and the packet cannot be forwarded via either SDRP or normal IP forwarding, the packet should be silently dropped.

The Hop Count field has already been decremented when processing the SDRP header. The Hop Count field should now be copied from the SDRP

header into the IP TTL field in the payload header. The resulting payload packet is then forwarded using normal IP forwarding. If there is no FIB entry for the destination, then the packet should be discarded and a control message with Notification Code "No Route Available" should be generated as specified in Section 6. If the packet can be forwarded and if the Probe Indication bit is set to one in the SDRP header, then a control message with Notification Code "Probe Completed" should be generated as specified in section 6. If a control packet is generated, then it must be sent to the encapsulating router. The payload of the control packet should carry the first 64 bits of the SDRP header and the payload header.

6. Originating SDRP control packets

A router sends a control packet in response to either error conditions, or to successful completion of a probe request (indicated via Probe Indication in the Flags field).

The Data Packet/Control Packet field is set to indicate Control Packet. The following fields are copied from the SDRP header of the Data packet that caused the generation of the Control packet:

- Loose/Strict Source Route
- Source Route Protocol Type
- Source Route Identifier
- Source Route Length field
- Payload Protocol Type

A Control packet should not carry a Probe Indication field.

A router should never originate a Control packet as the result of an error caused by a control packet.

The Target Router is copied from the source IP address of the delivery header of the SDRP Data packet. This causes the control packet to be returned to the encapsulating router.

The router generating a control packet checks its FIB for a route to the destination depicted by the Target Router field. If such a route is present, then the value of the Destination Address field in the delivery header is set to the Target Router, the Source Address field in the delivery header is set to the IP address of one of the interfaces attached to the local system, and the packet is forwarded via normal IP forwarding.

If the FIB does not have a route to the destination depicted by the Target Router field, the local system constructs the Source Route field of the Control packet by reversing the SDRP route carried in

the Source Route field of the Data packet, sets the value of the Next Hop Pointer to the value of the Source Route Length field minus the value of the Next Hop Pointer field of the SDRP data packet that caused generation of the Control Packet. All Loose/Strict Source Route change bits in the new source route should be set to 0 (loose source route).

The contents of the Payload field depends on the reason for generating a control packet.

The resulting packet is then handled via SDRP Forwarding procedures described in Section 5.2.

7. Processing control information

A router participating in SDRP may receive control information in two forms, SDRP control packets from other routers and ICMP messages from routers that do not participate in SDRP, but are involved in forwarding SDRP packets.

7.1 Processing SDRP control packets

Most control packets carry information about some SDRP routes used by the router. To correlate information carried in the SDRP control packet with the SDRP routes used by the router, the router uses information carried in the SDRP header of the control packet, and optionally in the SDRP payload of the control packet (if present).

In general, receipt of any SDRP control packet that carries one of the following Notification codes

- No Route Available
- Strict Source Route Failed
- Unimplemented SDRP Version
- Unimplemented Source Route Probe Type

indicates that the corresponding SDRP route is presently not feasible, and thus should not be used for packet forwarding. The router must mark the affected routes as not feasible, and may use alternate routes if available.

The router may at some later point attempt to use an SDRP route that was marked as infeasible. The criteria used for retrying routes is outside the scope of this document and a subject of further study. It need not be standardizes and can be a matter of local control.

Receipt of an SDRP control packet that carries "Probe Completed" Notification code indicates that the corresponding SDRP route is feasible.

Receipt of an SDRP control packet that carries the "Transit Policy Violation" Notification Code shall be interpreted as follows:

- If the control packet carries no payload data then the corresponding SDRP route violates transit policy regardless of the content of the payload packet carried along that route.
- If the control packet carries only the payload header, then the corresponding SDRP route violates transit policy due to the content of the payload header.
- If the control packet carries the payload header and the transport header, then the corresponding SDRP route violates transit policy for the particular combination of payload and transport header contents.

If a router receives an SDRP control packet that carries "Hop Count Exceeded" Notification Code, the router should use the information in the payload of the Control packet to construct an ICMP Time Exceeded Message with code "time to live exceeded in transit" and send the message to the host indicated by the source address in the Payload Header.

7.2 Processing ICMP messages

To correlate information carried in the ICMP messages with the SDRP routes used by the router, the router uses the portion of the SDRP datagram returned by ICMP. This must contain the Source Route Identifier of the SDRP route used by the router.

ICMP Destination Unreachable messages with a code meaning "fragmentation needed and DF set" should be used for SDRP MTU discovery as described in Section 9.

All other ICMP Unreachable messages indicate that the associated route is not feasible.

8. Constructing D-FIBs.

A BR constructs its D-FIB as a result of participating in either BGP or IDRP. A BR must advertise a route to destinations within its domain to all of its external peers (BRs in adjacent domains), via BGP or IDRP. In BGP and IDRP, a BR must advertise a route to destinations within its domain to all of its external peers (BRs in adjacent domains).

If a BR receives a route to an adjacent domain from a BR in that domain and selects that route as part of its BGP or IDRP Decision Process, then it must propagate this route (via BGP or IDRP) to all other BRs within its domain. A BR may also propagate such a route if it depicts an autonomous system other than the adjacent domain.

Since AS numbers are encoded as network numbers in network 128.0.0.0, it is possible to also advertise a route to a domain in BGP or IDRP.

9. SDRP MTU Discovery

To participate in Path MTU Discovery ([6]) a router may maintain information about the maximum length of the payload packet that can be carried without fragmentation along a particular SDRP route.

SDRP provides two complimentary techniques to support MTU Discovery.

The first one is passive and is based on the receipt of the ICMP Destination Unreachable messages (as described in Section 7.2). By combining information provided in the ICMP message with local information about the SDRP route the local system can determine the length of a payload packet that would require fragmentation.

The second one is active and employs the Probe Indicator bit. If an SDRP data packet that carries the Probe Indicator bit in the SDRP header and Don't Fragment flag in the delivery header triggers the last router on the SDRP route to return an SDRP Control packet (with the Notification Code "Probe Completed"), then the information carried in the payload header of the control packet can be used to determine the length of the payload packet that went through the SDRP route without fragmentation.

10. Acknowledgments

The authors would like to thank Scott Bradner (Harvard University), Noel Chiappa (Consultant), Joel Halpern (Newbridge Networks), Christian Huitema (INRIA), and Curtis Villamizar (ANS) for their comments on various aspects of this document.

Security Considerations

Security issues are not discussed in this memo.

Authors' Addresses

Deborah Estrin
USC/Information Sciences Institute
4676 Admiralty Way
Marina Del Rey, Ca 90292-6695.

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

Tony Li
cisco Systems, Inc.
1525 O'Brien Drive
Menlo Park, CA 94025

Phone: +1 415 526 8186
EMail: tli@cisco.com

Yakov Rekhter
Cisco systems
170 West Tasman Drive
San Jose, CA, USA

Phone: +1 914 528 0090
Fax: +1 408 526-4952
EMail: yakov@cisco.com

Kannan Varadhan
USC/Information Sciences Institute
4676 Admiralty Way
Marina Del Rey, Ca 90292-6695.

Phone: +1 310 822 1511 x 402
EMail: kannan@isi.edu

Daniel Zappala
USC/Information Sciences Institute
4676 Admiralty Way
Marina Del Rey, Ca 90292-6695.

Phone: +1 310 822 1511 x 352
EMail: daniel@isi.edu

References

[1] Lougheed, K., and Y. Rekhter, "A Border Gateway Protocol 3 (BGP-3), RFC 1267, October 1991.

[2] Rekhter, Y., and P. Gross, "Application of the Border Gateway Protocol in the Internet", RFC 1268, October 1991.

[3] Rekhter, Y., and T. Li, "A Border Gateway Protocol 4 (BGP-4)", RFC 1654, July 1994.

[4] Hares, S., "IDRP for IP", IDR Working Group, 1994. Work in Progress.

[5] Postel, J., "Internet Protocol - DARPA Internet Program Protocol Specification", STD 5, RFC 791, September 1981.

[6] Mogul, J., and S. Deering, "Path MTU Discovery", RFC 1191, November 1990.

[7] Reynolds, J., and J. Postel, "ASSIGNED NUMBERS", STD 2, RFC 1700, October 1994.