X3S3.3/91-293R2

From: Paul F. Tsuchiya, Bellcore

To: X3S3.3

Topic: Comments on IDRP

Revision 2 of 91-293 is different from revision 1 in that the constraints as to where SA-route updates can go has been changed. Revision 1 allowed SA-route updates to go only to RDs that had the right RIB configuration. This revision changes the distribution so that the sending RD can only send SA-route updates to RDs that are on the sending RD's path to the NLRI listed in the SA-route attribute. This prevents certain situations where the actual path taken was not the path advertised to a source RD. In addition to this change, the encoding of the SA-route attribute now includes only one prefix. A few other minor changes were also made. All of the changes in this revision are attributable (either directly or indirectly) to Yakov Rekhter.

Revision 1 of 91-293 is different from 91-293 in that it removes the S-ROUTE PDU, and therefore the ability for a source to set up its own route. Now, backbones must establish source-address routes on behalf of its subscribers, and all RDs involved in the S-route must pre-configure the S-route. While this removes one of the features of source address based routing, it also greatly simplifies the mechanism. The text changes in this contribution have gone from 15 pages to 5 pages!

Revision 1 also added introductory text that 1) explains the need for the sorts of policy routing that are best implemented with source-address routing, 2) gives a general description of source address based routing, and 3) compares source-address routing with QOS routing and partial source routing and explains the advantages and disadvantages of source-address routing. Note that the term S-route has been changed to SA-route (for Source Address route) in this revision.

The remainder of the contribution is written as though it were a contribution from the USA to ISO.

Overview

This contribution adds source address based routing (SA routing) to IDRP. SA routing is choosing among multiple paths to a destination NSAP address based on the contents of the source NSAP address. The intent of SA routing is to give efficient and simple operation of relatively common policies while providing simple but less efficient operation of certain less common policies. While it is possible to operate these same policies using source specific QOS, SA routing is generally advantageous to source specific QOS.

The common policies mentioned in the last paragraph can be succinctly described as "choosing a local backbone transit domain". That is, an end routing domain (stub) that can derive service from multiple directly connected or nearby transit routing domains (backbones) will wish to choose

among those backbones on a packet by packet basis, for both outgoing and incoming traffic. This type of policy is ubiquitous in the USA telephone system, and is more and more commonly seen in the IP internet. It is expected that as various commercial backbone types emerge, this policy type will become even more common.

SA routing takes advantage of the facts that 1) for scaling purposes, NSAP addresses will be hierarchically formed based on the hierarchical routing domain topology, and 2) hierarchical NSAP addresses influence the path taken by 8473 NPDUs.

For instance, assume two backbones A and B. Assume further that, for scaling purposes, all NSAP prefixes for stubs connected to the two backbones are derived from the NSAP prefix assigned to the backbone. Consider a stub that is connected to both backbones A and B. Because of the backbone-derived prefix assignment, the stub has two prefixes, and all systems in the stub have two NSAP addresses.

Now, consider the path traversed by an 8473 NPDU routed to a system in the stub. If an NSAP address with A's prefix is used, the NPDU will be routed through backbone A. Likewise, if an NSAP address with B's prefix is used, the NPDU will be routed through backbone B. In other words, the destination address controls the route taken by 8473 NPDUs.

Unfortunately, the same effect is not seen for NPDUs departing from the stub. These NPDUs will be routed based only on destination address, which may be for a system that is under neither A nor B. There will be nothing in the destination address that would indicate which backbone should be traversed. The routers in the stub, however, could discriminate between the two paths through A and B by looking at the source address (which would contain either A's prefix or B's prefix). And, this could be done without requiring the source specific QOS field.

This contribution specifies the means by which IDRP can provide BISs with the information needed to discriminate between multiple paths to destination NSAP addresses based on the contents of the source NSAP address. This contribution also describes the source address based forwarding function.

Since it is important that any new mechanisms in IDRP are useful to a large number of systems, this contribution first discusses common policy requirements. It then gives an overview of the operation of SA routing in IDRP, showing how the common policy requirements are efficiently satisfied. It then compares the use of SA routing with source specific QOS routing. Finally, it gives the modifications to IDRP necessary for SA routing.

Common Policy Routing Requirements (choosing local backbone)

The purpose of policy routing is to find paths that satisfy the policies of source, intermediate, and destination routing domains. The policy criteria for choosing a route are 1) the service offered by the route, 2) the monetary cost of the route, and 3) any restrictions placed on the route by source, backbone, and destination routing domains.

Since many possible routes may be available through an internet, and since backbones may place arbitrary restrictions of their usage, these policy criteria seem to imply an arbitrarily complex decision process. However, real constraints on topologies, offered services, and charging policies tend to reduce the number of meaningful policy routing choices required by most routing domains. The result of these real constraints, which are discussed below, is that most policy routing can be satisfied through the act of a stub domain choosing its directly connected or "one-hop-away" backbone (that is, a backbone reachable through a regional "access" transit domain). We call such a backbone the "local" backbone.

In support of this claim, consider that backbone networks can typically be classified as belonging to one type of backbone or another. That is, a backbone will be an X.25 backbone, or an IP research backbone, or an SMDS backbone, or a voice backbone, or an ATM backbone, and so on. Typically, these backbones connect to backbones of like type. (They may also connect to backbones of different type, via routers, but this is in addition to their connections to backbones of like type.) If a stub domain chooses as part of its route a particular local backbone, the implication is that the service provided by this backbone is adequate, and that therefore a route including backbones of the same type is probably acceptable. But a backbone will typically prefer routes through backbones of like type over routes through backbones of different type. (These preferred routes are maintained using the RD_PATH attribute of IDRP.) Therefore, by choosing a given local backbone, and indirectly the service provided by that backbone, a route that upholds that service will typically be found.

Pricing policies are also typically satisfied by choosing the local backbone. This is because stub domains most often do not maintain billing relationships with remote backbones—rather, they maintain billing relationships with local backbones. (The local backbone maintains its own billing relationship to the backbones it connects to, but this is transparent to the stub domain.) Therefore, the primary influence most stubs will have on monetary cost is that of choosing their local backbone.

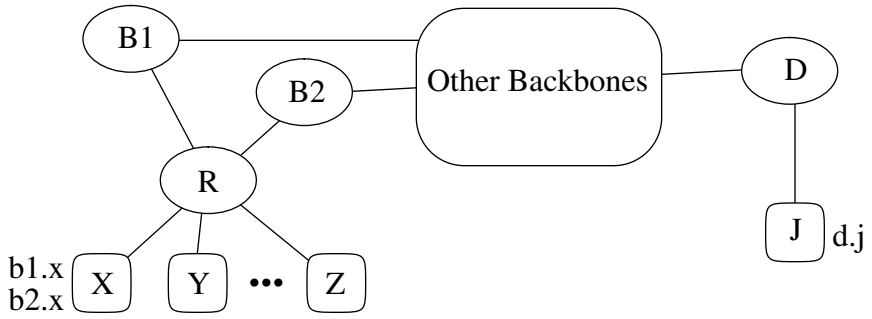
Of course, there will be exceptions to the above rule. In some cases, a stub may want to control more of the path than just its local backbone. SA-routing (and QOS routing) allows for this additional control. But controlling routes beyond the local backbone is less efficient than picking only the local backbone because of the amount of source-specific routing information that must be maintained.

Overview of SA Routing

The basic mechanism behind SA routing is a new distinguishing attribute—the SA-ROUTE DISCRIMINATOR. While there are some specific differences between the SA-route distinguishing attribute and the other distinguishing attributes, the SA-route distinguishing attribute behaves basically like the other distinguishing attributes. It describes a route (RD_PATH) to a destination, and indicates that only 8473 NPDUs that contain the given distinguishing attributes will follow that route. In the case of SA routing, the distinguishing attribute is a source address with a particular prefix. The main difference between the SA-route distinguishing

attribute and the other distinguishing attributes is that the SA-route distinguishing attribute can be originated by an RD other than the one whose NLRI is being advertized.

For example, consider a stub domain X attached to a regional backbone R, which in turn is attached to two backbones B1 and B2. Assume that X has two address prefixes derived from the two backbones; b1.x and b2.x. Suppose that a BIS in a stub domain X receives two UPDATES to destinations with prefix d. One has no distinguishing attributes, and has RD_PATH R-B1-...-D. The other has an SA-route distinguishing attribute for source addresses with prefix b2, and has RD_PATH R-B2-...-D. Based on these two UPDATEs, the BIS knows that if an 8473 NPDU is sent with destination address prefix d and a source address with prefix b2, it will take route R-B2-...-D. If an 8473 NPDU with destination address d and a source address without prefix b2 is sent, it will take route R-B1-...-D.



It is important to note that if the source address on the outgoing NPDU has prefix B2, then the resulting incoming NPDU will return via backbone B2 (assuming that, for scaling purposes, B2 advertises only prefix B2, and X's B2-based prefix is not advertised through B1).

Unlike the other distinguishing attributes, the SA-route distinguishing attribute does not need to be originated by the destination RD. A transit RD can add an SA-route distinguishing attribute to UPDATEs originated by other RDs. For instance, in the above example, it is not necessary for any backbones other than R to know of the SA-route. In fact, it is unnecessary overhead for any other backbones to know of the SA-route. Therefore, R can add the SA-route distinguishing attribute to the UPDATEs received from B2 and advertise this, and its UPDATE from B1, to X. The UPDATES are confined to the RDs that care about them by configuring the appropriate RIBs in those RDs.

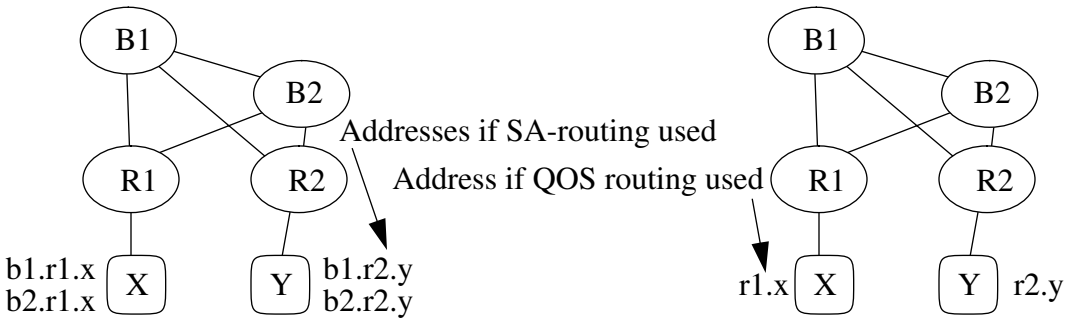
Comparison with Source-specific QOS

Anything that can be done with SA-routing can also be done with source-specific QOS. However, SA routing and source-specific QOS have different characteristics, and in many situations, SA routing is advantageous.

The most obvious difference between the two is that SA routing has a smaller header.

Another set of differences has to do with the fact that the destination NSAP address, when assigned with topology significance (backbone-oriented) already provides some level of policy routing. Since the destination NSAP address becomes the source NSAP address for return NPDUs, the use of source NSAP address as a distinguishing attribute is generally more natural (in the context of current standards) than using the source-specific QOS. This naturalness exhibits itself in several ways, discussed below.

First, consider the following common topology, where single-homed stubs have access to multiple top-level backbones via an access backbone. The two views show which addresses might be used for SA-routing and QOS routing. Note that although the SA-routing addresses scale better than the QOS routing addresses, both scale adequately well for the topologies emerging in the current internet.

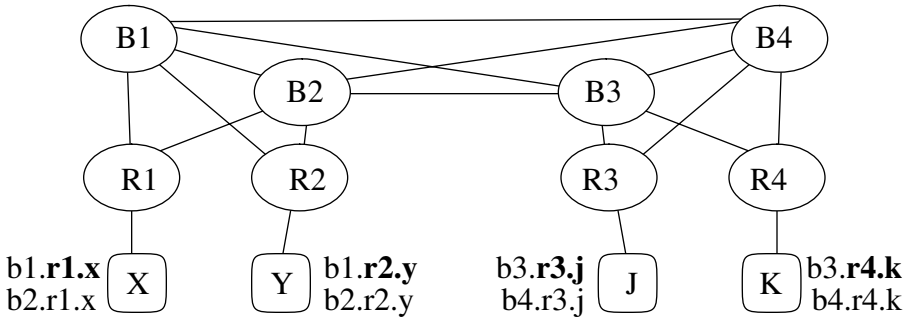


In the left-hand example, we give X and Y two prefixes each, one for each local backbone they wish to discriminate between. The hierarchical prefixes nominally provide scaling, since each backbone only need advertise one prefix. However, they also serve to provide policy routing, since paths will go through B1 and B2 if prefixes b1 and b2 are used respectively (even without the use of SA-routing, since in this example, both source and destination are under the same backbones).

For the right-hand example, the address is not sufficient to indicate which top-level backbone to go through. Therefore, QOS routing could be used. Assume for the moment that X wishes to use B1, and Y doesn't care which backbone is used. NPDUs from X have the source-specific QOS field with r1 as the prefix in the QOS field, and a value indicating B1 as the backbone. When the ES in Y attempts to return the NPDU, however, it cannot simply exchange source and destination address, remove the QOS field, and transmit the NPDU, because the NPDU may not return via B1. Instead, the ES in Y must create a destination-specific QOS, perhaps by using the same prefix and value in the source-specific QOS (if symmetrical values have been chosen a priori). This is not a standard use of 8473, however.

Now, consider the following topology, where the source and destination do not necessarily share a common top-level backbone The part of the prefix in bold shows the prefix that would be used in the case of QOS routing.

Here, destination address alone is not sufficient for finding the desired paths. For instance, if X and J exchange NPDUs using addresses b1.r1.x and b3.r3.j, the NPDU from X to J will correctly



go through B3, but will not necessarily go through B1 (depending on whether R1 chose B1 or B2 for its path to B3). Therefore, it is necessary for X (or R1 or B1) to install an SA-route for X. With the SA-route, NPDUs would go through B1 on the way to J, and would return through B1 on the return path because the ES in J would put the source address of the incoming NPDU into the destination address field of the outgoing NPDU.

Now, consider using the QOS field to route NPDUs between X and J through B1 and B3. In order for X to cause a departing NPDU to go through B1, a source specific QOS field is needed. In order for the same NPDU to go through B3, a destination specific QOS field is also needed. ISO 8473, however, allows only a source specific or destination specific QOS, not both. An alternative is to use a single QOS field (source or destination specific) that implied both B1 and B3. This, however, results in an explosion of QOS values, because now one is needed for each combination of source and destination backbone.

One solution to the problem of using QOS fields is to form hierarchical addresses in the same fashion as with the SA-route example, and let the destination address indicate the destination-side backbone, while letting the source specific QOS field indicate the source-side backbone. This solution, however, has no advantage to SA-routing (it still requires multiple addresses), while suffering the disadvantage of larger headers. In other words, to the extent that the use of multiple hierarchical addresses reflects the desired choice of local backbones, QOS routing has no advantages over SA routing. To the extent that addresses do not reflect the desired choice of local backbones, QOS routing, at least in the context of current 8473, is problematic.

Proposed Modifications to IDRP

Section 4.7 (Additional Definitions)

Add the following definition:

SA-route: An SA-route is one of multiple routes to a set of reachable address prefixes (destination NSAP addresses) that can be chosen based on the source NSAP address of the ISO 8473 header.

Section 6.7 (Distinguishing Path Attributes and RIB-Atts):

Add the following paragraphs after the first paragraph:

There are two types of RIB-Atts, those that do not distinguish on source address, called basic RIB-Atts, and those that do, called SA-route RIB-Atts. The basic RIB-Atts are defined by system administration before a BIS establish adjacencies. SA-route RIB-Atts, on the other hand, can be created dynamically based on the reception of IDRP PDUs. Each SA-route RIB-Att must have the same non-SA-route distinguishing attributes as one of the basic RIB-Atts.

Change the first two sentences of the second paragraph to:

While the number of distinct permitted combinations of distinguishing attributes forming a basic RIB-Atts is large, and the theoretical number of resulting SA-route RIB-Atts is unbounded, the number of RIB-Atts will be limited in practice by the number of useful combinations of these attributes.

Section 7.3 (UPDATE PDU)

Add the following attribute after the PRIORITY attribute:

SA-ROUTE DISCRIMINATOR (Type code 20)

This variable length type-specific distinguishing attribute is associated with SA-routes. It contains an NSAP address prefix, which is encoded as follows:

NSAP prefix length (1 octet)
NSAP prefix (variable)

The use and meaning of the fields is as follows:

a) NSAP prefix length:

Gives the length, in semi-octets, of the NSAP prefix

b) NSAP prefix:

Contains the NSAP prefix, encoded according to clause 8.1.2.1. If an ISO 8473 source NSAP address matches the NSAP prefix, and the ISO 8473 destination NSAP address matches one of the NLRI addresses, then the ISO 8473 NPDU is routed along the RD Path. If the ISO 8473 source NSAP address matches NSAP prefixes from more than one SA-ROUTE DISCRIMINATOR, then the match with the longest prefix is used.

An UPDATE PDU with the SA-ROUTE DISCRIMINATOR attribute is called an SA-route UPDATE PDU. Usage of this attribute is defined in clause 8.13.21.

Section 8.10.1 (Identifying an Information Base):

Second paragraph, change the first instance of "RIB-Atts" to "basic (non S-Route distinguished) RIB-Atts". Change the remaining instances of "RIB-Atts" to "basic RIB-Atts"

Add the following paragraph after the second paragraph:

For each of the basic RIB-Atts, multiple additional RIB-Atts can be dynamically created by further distinguishing each basic RIB-Att by an SA-route. That is, the SA-route RIB-Att will have exactly the same distinguishing attributes as one of the basic RIB-Atts, plus an SA-route distinguishing attribute. Because of nesting of NSAP prefixes, any individual source NSAP address may indicate multiple SA-route RIB-Atts. However, the match with the longest prefix must be chosen. Therefore, any individual source NSAP address will indicate a single SA-route RIB-Att.

In the third and fourth paragraphs, change the two instances of "RIB-Att" to "basic RIB-Att".

Table 1 (Path Attribute Characteristics):

Add the following entry:

SA-ROUTE DISCRIMINATOR	well-known discretionary	20	variable	Yes

Section 8.12.2 (Handling of Distinguishing Attributes):

Change the first sentence in the second paragraph to "All non-SA-route distinguishing path attributes shall only be created by the BIS that originates the routing information. SA-route distinguishing path attributes may be created by any RD. All path attributes can be updated by any BIS that receives an UPDATE PDU that contains them."

Add the following after paragraph a), and re-letter the subsequent paragraphs as appropriate:

b) A permissible set of Distinguishing Attributes can not include both a SOURCE SPECIFIC QOS attribute and an SA-ROUTE DISCRIMINATOR attribute.

c) A permissible set of Distinguishing Attributes can not include both a SOURCE SPECIFICSECURITY attribute and an SA-ROUTE DISCRIMINATOR attribute.

Section 8.12.3 (Equivalent Distinguishing Attributes):

Change "and Destination Specific Security" to "Destination Specific Security, and SA-Route".

Section 7.2 (UPDATE PDU)

Before the note, prepend to the second bullet: “except for the SA-ROUTE DISCRIMINATOR, “

After the note, add the following third bulleted paragraph:

- The SA-ROUTE DISCRIMINATOR type-value-specific distinguishing attribute is encoded as a triple <type. length, value>, where the value field is the entire contents of the value field as specified in clause 7.3 for the SA-ROUTE DISCRIMINATOR attribute.

After Section 8.13.20 (PRIORITY):

8.13.21 SA-ROUTE DISCRIMINATOR

SA-ROUTE DISCRIMINATOR is a well-known discretionary, type-value-specific distinguishing attribute. It allows the source NSAP address of an ISO 8473 NPDU to be used to choose among multiple routes to the destination NSAP address. The finest granularity of this control is a single End System. A BIS shall include this attribute in its UPDATE PDU to indicate that it supports SA routing based on the reported address prefix, and that it maintains Adj-RIBs and a Loc-RIB distinguished by the indicated source address prefixes.

An UPDATE PDU with an SA-ROUTE DISCRIMINATOR is called an “SA-route UPDATE PDU”. A RIB with an SA-ROUTE distinguishing attribute is called an “SA-route RIB”. A BIS may not create an SA-route RIB unless it has a corresponding RIB such that 1) the corresponding RIB has no source address distinguishing attribute, 2) every other distinguishing attribute of the corresponding RIB exactly matches those of the SA-route RIB, and 3) the corresponding RIB has neither SOURCE SPECIFIC QOS attribute nor a SOURCE SPECIFIC SECURITY attribute. This corresponding RIB is called the “corresponding non-SA-route RIB”.

Unlike all other distinguishing attributes, an SA-route UPDATE PDU can be originated by an RD other than the destination RD. In order to determine whether or not a BIS can originate an SA-route UPDATE PDU, a BIS will associate the following managed objects with the RIB corresponding to the SA-route. Note that in order for a BIS to originate an SA-route UPDATE PDU, it must be configured with the corresponding SA-route RIB (the RIB will not be dynamically created).

a) originateAllowed: This managed object is 1 if the BIS can originate an SA-route UPDATE PDU, and 0 otherwise.

b) transitRD: This is an RD that must appear in the RD_PATH of the UPDATE PDU.

A BIS may advertise an SA-route UPDATE PDU only if all of the following criteria are true. The BIS must subsequently advertise a corresponding unreachable SA-route UPDATE PDU if any of the following criteria becomes false.

1. The BIS has advertised or could correctly advertise a corresponding non-SA-route UPDATE PDU (that is, an UPDATE PDU identical to the SA-route UPDATE PDU except for the SA-ROUTE DISCRIMINATOR).

2. The current path (that is, the one installed in the FIB) to some or all of the NSAP addresses indicated by the prefix in the SA-ROUTE DISCRIMINATOR is via the adjacent BIS. If there are multiple paths (via multiple FIBs), any one can be used, although in the absence of a specific preference, the path via the default FIB should be used. (This criteria does not apply to the case where the NSAP addresses indicated by the prefix in the SA-ROUTE DISCRIMINATOR are members of the BIS's RD. In this case, the BIS may need to advertise the SA-route UPDATE PDU to all internal BISs, for instance if intra-domain routing cannot route based on source address.

A BIS may originate an SA-route UPDATE PDU only if all of the additional criteria are also true. The BIS must subsequently advertise a corresponding unreachable SA-route UPDATE PDU if any of the following criteria becomes false.

3. The originateAllowed managed object associated with the SA-route RIB is 1.

4. The RD identified by the transitRD managed object associated with the SA-route RIB is listed in the RD_PATH of the non-SA-route UPDATE PDU.

When a BIS advertises a reachable SA-route UPDATE PDU to an adjacent BIS, all of the NSAP addresses indicated by the prefix in the SA-ROUTE DISCRIMINATOR must be reachable via the adjacent BIS. Therefore, a BIS may need to make the prefix in the advertised SA-ROUTE DISCRIMINATOR longer than that of the received SA-ROUTE DISCRIMINATOR or the RIB SA-route attribute (if the BIS originated the SA-route UPDATE PDU).

If the path to the NSAP addresses indicated by the prefix in the SA-ROUTE DISCRIMINATOR changes, then an unreachable SA-route UPDATE PDU must be sent to the BIS that received the previous reachable SA-route UPDATE PDU. If the path for only some of the NSAP addresses indicated by the prefix in the SA-ROUTE DISCRIMINATOR changes, then an unreachable SA-route matching the previous reachable SA-route UPDATE PDU must be sent, followed by a reachable SA-route UPDATE PDU indicating the smaller set of NSAP addresses in the SA-ROUTE DISCRIMINATOR.

If a BIS advertises a non-SA-route UPDATE PDU, and subsequently chooses not to advertise a corresponding received SA-route UPDATE PDU that satisfies criteria 1 and 2 above, and the RD_PATHs of the two UPDATE PDUs are different, then the BIS must either:

a) advertise the non-SA-route as unreachable, or

b) change the RD_PATH or the non-SA-route UPDATE PDU into an RD_SET, and include in the RD_SET the combined RDs of the RD_PATHs of the UPDATE PDUs.

> NOTE—A non-SA-route UPDATE PDU is essentially an SA-route UPDATE PDU with an SA-ROUTE DISCRIMINATOR prefix of length 0. That is, the source addresses of an SA-route UPDATE PDU are nested in those implied by a non-SA-route UPDATE PDU. If a non-SA-route UPDATE PDU is advertised without a corresponding SA-route PDU, the source will understand the RD_PATH to be that of the non-SA-route UPDATE PDU when in fact the RD_PATH will follow that of the SA-route. Note also that a BIS may advertise an SA-route UPDATE PDU without advertising the corresponding non-SA-route UPDATE PDU.

Section 8.16.2 (External Updates):

Last paragraph, after “path attributes”, add: “(excluding the SA-ROUTE DISCRIMINATOR)”

Section 9 (Forwarding Process for CLNS):

Add the following parenthetical to the third sub-bullet (c) “(possibly discriminated by the source address field of the NPDU)”.

In the second paragraph, the description has the BIS determining if the destination system is located in its own RD before determining the RIB-Atts of the NPDU. Since the reachability of a system depends on the RIB-Atts, it is inappropriate to determine that a destination system is reachable within an RD without first looking at the RIB-Atts.

Therefore, sub-paragraph a) and the first paragraph of sub-paragraph b) should be removed. The second paragraph of clause 9 will read:

“Having determined the system to which a path is needed, the BIS shall perform the following actions:

1) It shall determine.....”

Change sub-paragraph ii) to read as follows:

ii) If there is a match indicating that the destination system is located in its own RD, the local BIS shall proceed as defined in clause 9.1.

Add the following sub-paragraph:

iii) If there is a match indicating that the destination system is located in another RD, the local BIS shall proceed as defined in clause 9.4.

Section 9.2 (Determining the NPDU-derived Distinguishing Attributes):

Change “of the priority” in the last sentence of the first paragraph to “of the source address, priority”

Add the following bulleted item before the first bulleted item:

- The 8473 source address field corresponds to the SA-Route path attribute

Section 9.3 (Matching RIB-Att to NPDU-derived Distinguishing Attributes):

Paragraph a) is redundant given the following two paragraphs. Remove paragraph a) and re-letter the remaining paragraphs accordingly.

Add the following bulleted item under paragraph c):

- SA-route

Add before the first comma of the first sentence of the last paragraph under c): "(including the SA-route attribute)".

Add the following note after the last paragraph under c):

> Note: With respect to matching longer prefixes, a RIB without an SA-Route attribute is considered identical to a RIB with an SA-Route attribute indicating "all NSAP address". In other words, if two RIBs are identical except that one has an SA-Route attribute and the other does not, and the source address of an 8473 NPDU matches that of the SA-Route attribute, then the RIB with the SA-Route attribute is considered the better match.

Section 9.4 (Forwarding to External Destinations):

In paragraph b)2), add the following sentence after the first sentence:

This may be necessary in cases where multiple routes to a destination are discriminated by distinguishing attributes such as QOS or SA-route, but where intra-domain routing does not recognize the distinguishing attributes.

In paragraph b)2), change "(inner) NPDU." to "(inner) NPDU provided that the intra-domain routing protocol can properly route in the presence of the QOS parameters."

System Management, GDMO, PICS, etc.

Of course, all of the above comments on SA-routes require these things. If the comments are excepted, the filler will be forthcoming.

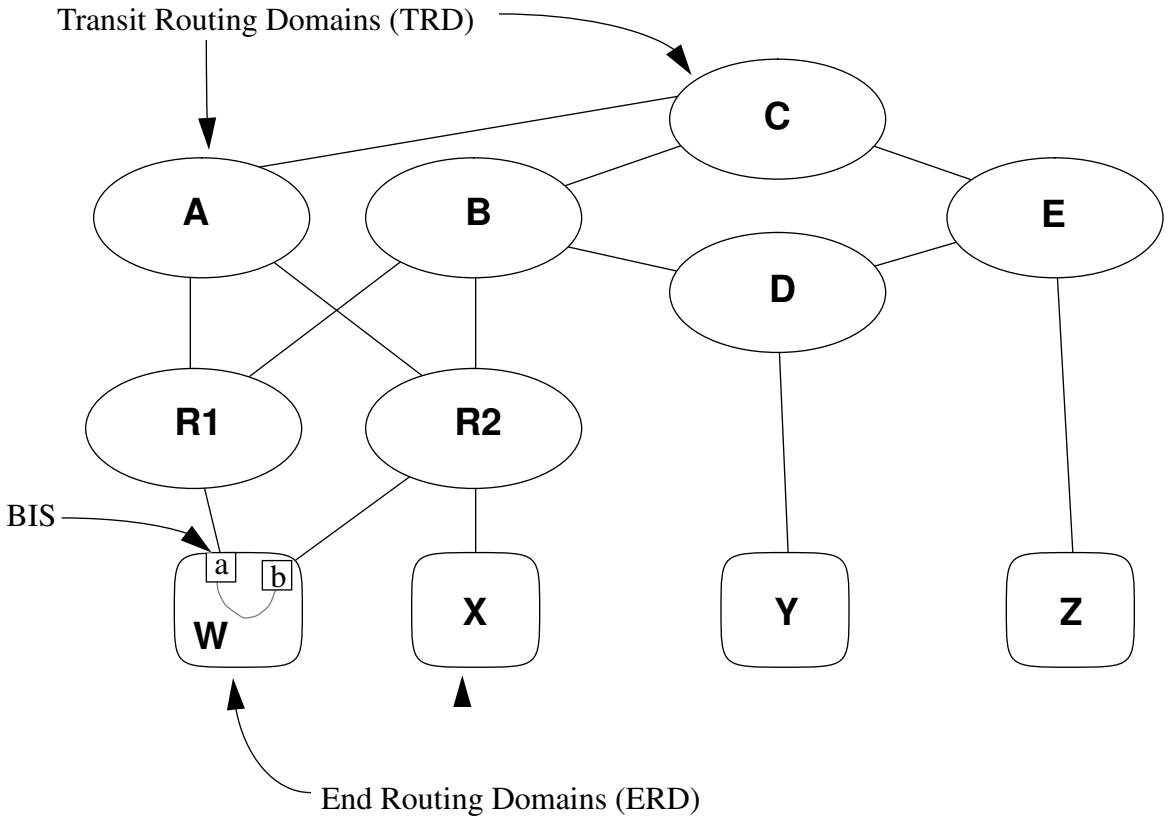
Tutorial annex

Add the following annex:

Annex X: Use of SA-routes

(Informative)

This informative annex describes the SA-route function and its characteristics by giving examples of its use. All examples refer to figure X1.



Example X.1

In this example, systems in End Routing Domain (ERD) W wish to be able to send 8473 NPDUs through either Transit Routing Domain (TRD) R1 or R2. Assume that for scaling purposes each ERD has an address prefix that is derived from that of the TRDs it is connected to. An example of the notation for this is r1.w, where r1 is the prefix common to all or most ERDs attached to R1, and r1.w is the prefix for W.

> Note: IDRP would most commonly take advantage of this hierarchical address format by forming an RDC consisting of R1 and the ERDs that formed an address using R1's prefix, and then R1 would aggregate the NLRI of all the SA-route UPDATEs from the ERDs into one SA-route UPDATE.

To install the SA-route, the BISs in W that have adjacencies in R1 or R2 must be configured with corresponding SA-route RIBs, one describing an SA-route for source addresses with the prefix r1.w and a TRD of R1 (if attached to R1), and the other describing an SA-route for source addresses with the prefix r2.w and a TRD of R2 (if attached to R2). When any UPDATEs are received from R1 or R2, the BISs with external links to R1 or R2 originate SA-route UPDATE

PDUs internally. As a result, BISs in W know that they can reach various destinations via either R1 or R2, depending on which source address is used.

Consider 8473 NPDUs sent between W and Z using source address r1.w and destination address e.z. The 8473 NPDU to Z is routed to R1 by BISs in W because of the SA-route. The BISs in R1 do not have or need an SA-route for this 8473 NPDU because the source doesn't care whether A or B is used (if it did, it would include A or B in its address), and so sends the 8473 NPDU to A based on destination address only. Likewise, the 8473 NPDU gets forwarded to C, E, and on to Z.

The return 8473 NPDU from Z to W is routed based entirely on the destination address (r1.w). Note, however, that because the address is hierarchical, BISs in Z, E, C, and A view the 8473 NPDU as being destined for R1, and so the return 8473 NPDU is routed to R1 via E, C, and A. Therefore, the desire that the 8473 NPDUs go through R1 was satisfied for both outgoing and incoming 8473 NPDUs.

This example demonstrates the best use of SA-routes. Scaling is achieved because of the use of hierarchical addresses, and because the only RD that required source-based routing is the one that had multiple paths (W). Moreover, the hierarchical address itself was used to full advantage in that it influenced the return path from Z to W. This example also demonstrates what will likely be the most common use of SA-routing—that is, an ERD choosing between multiple directly attached TRDs.

Note that intra-domain routing in W may not route based on both source and destination address. Therefore, ISs in W will not know whether to deliver a given 8473 NPDU to BIS a or BIS b. In this case, encapsulation may be necessary to tunnel the 8473 NPDU from the BIS chosen by intra-domain routing to the correct exit BIS. For instance, if for this example intra-domain routing had routed the 8473 NPDU to BIS b, BIS b would encapsulate the 8473 NPDU in another 8473 NPDU addresses to BIS a, which would then decapsulate the NPDU and forward it onto R1.

Note finally that, as long as both R1 and R2 can reach the large majority of possible destinations, it is not even necessary for BISs in W to maintain RIB entries for every destination. Instead, BISs in W can maintain "default" (everything is reachable) entries for both R1 and R2, and use SA-routing to pick between them.

Example X.2

In this example, systems in W wish to choose between A and B, but don't care whether R1 or R2 is used. In this case, W should have addresses a.w and b.w. It is necessary for R1 and R2 to originate SA-route UPDATE PDUs "on behalf of" W. For this to happen, the BISs in R1 and R2 that have external adjacencies with A or B must be configured with RIBs indicating SA-routes through A and B for source address prefixes a.w and b.w (or more likely, in order to provide policy routing for ERDs other than just W, source address prefixes a and b).

In this case, W would receive SA-route UPDATE PDUs from R1 and from R2 (and optionally the non-SA-route UPDATE PDUs). For instance, R1 might choose A to reach some destinations (say those in Z) and B to reach other destinations (say those in Y).

Assume for this example that W chooses paths through R1 only. Now, outgoing 8473 PDUs from W to Z using addresses a.w and e.z will go W-R1-A-C-E-Z. The return 8473 PDUs take the reverse path.

If all links between R1 and W crash, the BISs in W will replace the SA-routes via R1 with SA-routes via R2.

Example X.3

In this example, hosts in X wish for their 8473 NPDUs to go through backbone C. Assume that X's only prefix is r2.x. That is, X has not derived a prefix from the backbone that it wishes to send traffic through.

Assume that C reaches W and X through R2 and B, reaches Y through B and D, and reaches Z through E. Assume also that both A and B reach W and X through R2.

Assume that the BISs in C that have adjacencies with BISs in A and B have SA-route RIBs for sources in X. (Alternatively, the BISs in A and B with external adjacencies to BISs in C could have SA-route RIBs configured for RD_PATHs through C.) However, since the only destinations that C reaches without going through B is Z, and since B is on C's path to X, B receives only one SA-route UPDATE PDU from C—that with NLRI in Z. Since C does not reach X via A, A receives zero SA-route UPDATE PDUs from C. Only X's 8473 NPDUs to Z will actually go through C.

Now assume that C is reconfigured to reach X via A, but W (and all other RDs under R1 and R2) via B. In addition, C is reconfigured to reach Y via D and E. Now, A will receive SA-route UPDATEs from C for all destinations except X, and these will be passed on to X via R2. Now, all of X's 8473 NPDUs, even those to W, will go through C (NPDUs to W would go X-R2-A-C-B-R1-W). Assuming that since A did not previously advertise to R2 a non-SA-route paths to W via C, A could have chosen not to have advertised the corresponding SA-route to R2. In this case, the path from X to W would have been X-R2-W.

Note that return NPDUs to X will not necessarily go through C. For instance, even if the path from X to W is X-R2-A-C-B-R1-W, the return path from W to X would be W-R2-X, unless a return SA-route or perhaps a destination-specific QOS route was setup. In general, the further away the transit backbone(s) of the SA-route are from the source, particularly when the source address is not derived from the transit backbone, the less likely the return packet is to go through the transit backbone, and the greater the overhead (i.e., more BISs must maintain the SA-route information). In other words, the benefits from SA-routing decrease as the desired scope of control gets further away from the source.