ISO/IEC JTC 1/SC 6 N 7535

Project: JTC1.06.41
Date: 1992-07-21

ISO/IEC JTC 1/SC 6
TELECOMMUNICATIONS AND INFORMATION EXCHANGE BETWEEN SYSTEMS

Secretariat: U.S.A. (ANSI)

Title: Interactions and Routeing Information Exchange between ISO 10589 (ISIS) and ISO 10747 (IDRP)

Source: Routeing Group of SC6 /WG2

Address reply to:
Secretariat ISO/IEC JTC 1/SC 6 – American National Standards Institute, 11 West 42nd Street, New York, NY 10036

Tel: 212 642-4931; TX 42 42 96 ANSI UI; FAX 212 302-1286

Interactions and Routeing Information Exchange between ISO 10589 (ISIS) and ISO 10747 (IDRP)

1 Introduction

At the July 1992 SC6/WG2 meeting in San Diego the Routeing ad hoc group developed this technical proposal for how routeing information gets *exported* from Intra-domain Routing for use by Inter-domain routeing, and how routeing information exported by Inter-domain routeing gets *imported* into Intra-domain routeing to facilitate the transit of interdomain traffic through a routeing domain.

The technical design contained herein was accepted as the appropriate way to proceed on this important area of work. In addition to the general technical design, the following principles for the progression of the work were also agreed:

1) To the extent possible, all interactions between the two routeing protocols should be expressed through the representation of the information as GDMO management data structures, in order to avoid one protocol needing to directly process internal state information maintained by the other protocol. This approach requires enhancement to the management definitions of both ISO 10589 and DIS 10747.

2) The design should be incorporated into the texts of ISO 10589 and ISO 10747, respectively, and the creation of a third document that would need to be referenced by implementers avoided, if possible.

3) The progression of the work is expected to be accomplished by Ammendment of the two standards. Further progression of DIS 10747 to IS status should not be delayed in order to incorporate this work.

The intention of the routeing group is have base text and NP proposals for the two necessary ammendments prepared in time for the next meeting of SC6/WG2

2 Technical Issues covered by this proposal

This document covers the following technical issues concerning the interaction and routeing information exchange between ISIS and IDRP:

1) Exportation of Intra-domain Routeing information (specifically Area Addresses and Reachable Address prefixes) from ISO 10589 for importation by IDRP as NLRI (Network Layer Reachability Information).
    a) Policy on using the ISIS internal metric to create the IDRP multi-exit discriminator.
    b) Handling of statically-configured Inter-domain routeing information expressed as ISO 10589 Reachable Address Prefixes.

2) Tunneling (encapsulation) of packets BIS-to-BIS through an RD using ISIS

3) Importation of Routeing information from IDRP's NLRI into ISIS's Reachable Address Prefixes.

a) Policy for which and how many NSAP address prefixes in IDRP's NLRI will be imported into ISIS as Reachable Address Prefixes.

b) Policy for whether NLRI is used by ISIS with internal or external metrics.

4) BIS Discovery using ISIS

5) Decision on whether/how to handle partitions of a Routeing Domain.

Other possible areas for work are not addressed here. They include:

1) Possible piggybacking of IDRP on ISIS L2 LSPs, thus using the ISIS flooding mechanism for all BIS-BIS communication inside an RD. This might be a performance win on transit RDs with few interior ISs.

3 Exportation of ISIS Information for Importation and Summarization by IDRP

In order for a BIS running IDRP to route to destinations inside its local routeing domain, IDRP needs some source of information which represents the dynamic state of routeing domain, i.e. what destinations exist in the domain and of those, which are currently reachable. In the case of a system which is both an ISO 10589 Level 2 IS *and* an IDRP BIS, IDRP can obtain this information directly from the routeing information maintained by ISIS. ISIS provides this routeing information in the form of Area Addresses and Reachable Address prefixes carried in the Level 2 LSPs of ISIS. IDRP can import this information to construct NLRI to represent destinations in the local routeing domain.[1]

3.1 Policy for Importing Routeing information into IDRP

In order to control the volume of routeing information about the internal topology of the local routeing domain that gets propagated to other routeing domains, IDRP needs to have a policy about what information to summarize in its NLRI, and how.

The policy variable is represented as an attribute of the MO containing the IDRP global management parameters. A possible name for this attribute could be intradomainSummarizationPolicy. It has three values:

a) Automatic summarization

b) Pre-configured Summarization

c) No Summarization

If the value is "No Summarization", then IDRP imports all of the Area Addresses in the RD from the destinationArea managed objects of the IS and imports them as IDRP NLRI.[2]

The default value of intradomainSummarizationPolicy should be "No Summarization" for safety of not reporting routes to destinations that are in fact not in the local routeing domain.[3]

[1] In the case of area addresses, this information is already abstracted as the GMDO destinationArea managed object .The reachable address information is only available as fields in the L2 LSPs. Making this information available through management could be done by directly modelling the information the LSP databases. This is the initial approach that will be pursued in completing the design

[2] The destinationnArea MO of ISIS currently exists for both internal and external destinations. IDRP needs to import the internal destinations, but also needs to import *some* of the external destinations. Extreme care has to be taken here so that IDRP does not re-import the same information it previously exported to ISIS. Further, when importing external destinations into IDRP, the protocol must tag the information as *externally learned*, since it came from statically configured Reachable Address prefixes in some level 2 IS in the RD.

[3] If you don't summarize, the protocol will be able to reach most internal destinations even if the RD is partitioned but this is not deemed sufficient reason to avoid summarization.

If the value of intradomainSummarizationPolicy is "Automatic", IDRP uses the RDI for the local routeing domain (the localRDI attribute of the iDRPConfig MO) as a template for what destinations to import as NLRI. IDRP scans all of the destinationArea MOs of ISIS and imports:

1) a prefix equal to the longest common prefix of all of area addresses which match the localRDI of the routeing domain. This covers the common case of a routeing domain which has area addresses taken from a common addressing assignment from one authority, and uses one of its these addresses as its own RDI.

2) In addition, any destinationAreas or externalDestinations which do not match the localRDI are imported individually.

If the value of intradomainSummarizationPolicy is "preconfigured", IDRP uses a set of preconfigured prefixes which it is willing to import as NLRI. These can be represented as a set of destinationsToImport managed objects, contained within the iDRPConfig MO[4].

If any destinations in the set of destinationAreas or externalDestinations matches one of the destinationsToImport, then IDRP imports that prefix; otherwise it does not. This deals with the case where a RD has address assignments taken from a number addressing authorities (e.g. a corporate network with area addresses taken from the French and Botswanan addressing authorities).

3.2 Tradeoffs between summarization and route optimality

In addition to announcing the prefixes obtained via automatic or preconfigured summarization, it is possible to also announce individual area addresses. This enables IDRP to direct the entry of traffic into the RD for areas that are "close" to the entry BIS. This is accomplished by defining a set of intradomainCostThreshold attributes (one for each routeing metric potentially supported by tthe ISIS implementation).

If the value of the corresponding pathCost attribute a destinationArea falls below the threshold, IDRP imports the area address individually, in order to optimize the entry of traffic. If the metric falls above the threshold, IDRP does not import that individual area address and lets it be covered by the summarization prefix[5]. Longest match routeing will ensure that the traffic comes in the better way. This solves the well-known "east coast/west coast" problem illustrated below in figure 1. A, B, and G are BISs each in its own area *A*, *B*, and *G* respectively. C, D, E, and F represent both L2 ISs and the areas in which they reside. If the intradomainCostThreshold is set to 7 on each of the BISs,

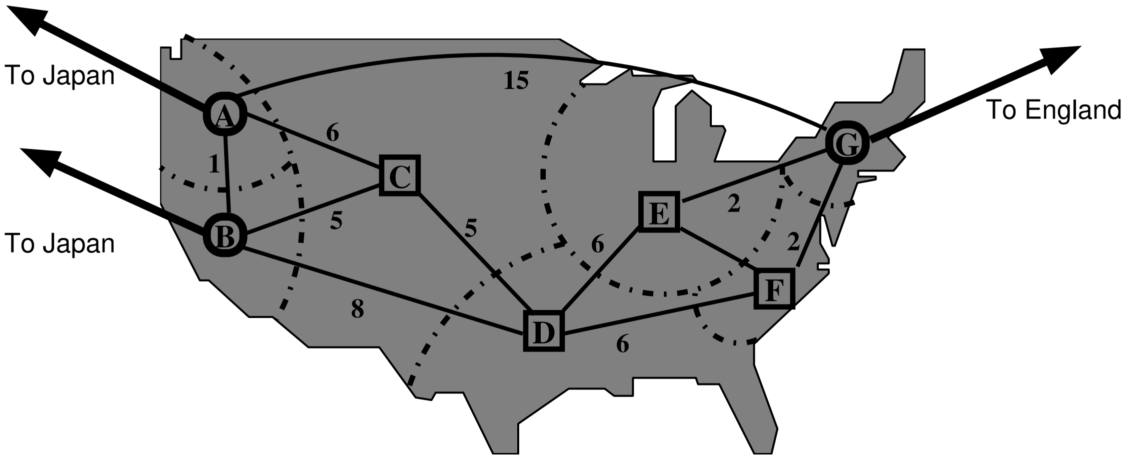

Figure 1 - Route Advertisement using Level 2 Cost Threshold

[4]The destinationsToImport contains, as a minimum, the prefix to match against a candidate destination to import. In addition, it may be useful to also have an attribute which specifies constraints on the matching. For example, one could specify the candidate set of Intra-domain protocols to scan for a match and an indication of whether to treat the information as “externally learned”.

[5]If L2 Cost Threshold is set to zero, then only the area in which the BIS resides is announced individually. Some encoding needs to be defined to completely disable the use of this feature if announcement of the BISs own area is not desired.

then the following announcements will be made in IDRP NLRI (considering only the case of the ISIS default routeing metric):

BIS	Areas Announced
A	$\mathcal{A}, \mathcal{B}, \mathcal{C}$
B	$\mathcal{A}, \mathcal{B}, \mathcal{C}$
G	$\mathcal{G}, \mathcal{E}, \mathcal{F}$

Traffic inbound to the U.S. from Hong Kong might come in either to San Francisco (BIS A) or Los Angeles (BIS B) via Japan. Assuming the Inter-domain paths are of the same preference, then by announcing area $\mathcal{E}$ individually, traffic to Akron is directed through Boston rather than through San Francisco or Los Angeles. Conversely, by not announcing $\mathcal{D}$ traffic inbound to Dallas shows no preference for any of the individual entry points.

When announcing the individual imported areas, BISs A and B will include the MULTI_EXIT_DISC path attribute, so that the Japanese RD can discriminate between the inbound paths through San Francisco and Los Angeles. In the above example, the Japanese RD would choose BIS B over BIS A for traffic inbound to area $\mathcal{C}$ since it has the lower Intra-domain metric value. When reporting the MULTI_EXIT_DISC path attribute, IDRP multiplies the L2 ISIS metric by 4 to account for the difference in dynamic range of the two metrics. Note that the additional information may be constrained to only the adjacent routeing domain.

3.3 Handling statically configured ISIS Routes

IDRP is also permitted to imports routeing information from ISIS for destinations that were statically configured via ISIS Reachable Address Prefixes. This is necessary, if only for the graceful introduction of IDRP into Routeing Domain, but may be needed until such time as all L2 ISIS implementation platforms support IDRP. Statically configured Reachable Address Prefixes from ISIS are imported into IDRP as follows[6]:

1) If an ISIS L2 LSP indicates that the IS is *not* a BIS (see BIS Discovery, below), then IDRP imports the prefixes in the L2 LSP. If the LSP indicates that the L2 IS *is* a BIS, the information in the reachable address prefixes is not imported.[7]

2) The BIS sets the IDRP Path Attribute EXTERNAL_INFORMATION in indicate that the information describes external destinations not originating from IDRP.

3) If the BIS supports multiple routeing metrics, a separate path is imported for each of the supported routeing metrics. The degree of preference for each is set to the value of the corresponding ISIS metric assigned to the Reachable Address Prefix.

4 Tunneling of Interdomain Traffic Through the RD

By default, IDRP and ISIS are totally independent of one another, in the sense that neither relies on the other to ensure correct routeing of data PDUs. IDRP BISs in different routeing domains are by definition connected to each other over a shared subnetwork. On the other hand, BISs in the same local routeing domain do not necessarily share a subnetwork and use the Intra-domain routeing protocol to establish communication amongst themselves. IDRP therefore needs some way to get NPDUs from one BIS to another in the local RD without involving ISIS[8]. One way

[6] The representation of the ISIS static routes within IDRP can be accomplished by creating a pseudo-Adj-RIB-In and placing the NLRI there as if it had come from a real BIS. Other implementation techniques are of course possible.

[7] The reason is that both ISIS and IDRP in one BIS obtain statically configured routing information from the same Managed Objects — the Reachable Address MOs. Hence any information in the L2 LSP Reachable Address Prefix fields would be redundant with information already obtained via IDRP BIS-BIS protocol exchange.

[8] ISIS would not be able to forward these NPDUs because the destination NSAP is outside the local routeing domain and only IDRP has the necessary routeing information to determine the next Inter-domain hop for the NPDU.

to accomplish this is to have IDRP provide the necessary routeing information to ISIS: this is discussed in clause 5. In many circumstance however, it is best to forward NPDUs through a routeing domain by encapsulating them inside another NPDU destined for the exit BIS as the network layer address. This form of Inter-domain forwarding is known as *tunneling*.

Tunneling is done using encapsulation as described in the current IDRP specification, with the exception that text needs to be added to IDRP to ensure that if you are using ISIS to do the tunneling, that you follow the procedures in ISIS for "encapsulation of traffic for partition repair".

IDRP needs to manipulate the ISIS management information to control the forwarding of packets through the domain, such that ISIS and IDRP can have a unified management representation of the FIBs used by the forwarding machine(s). The ISIS "Adjacency", "Circuit", "Virtual Adjacency" and "Reachable Address" managed objects are used to contain the shared state between ISIS and IDRP as follows:

1) Inter-RD BIS-BIS links are distinguished from Intra-RD links by marking the Circuit MO for Inter-RD links as "external domain". To handle the case of "DMZ" subnetworks (multi-access subnetworks with some internal and some external neighbors — DMZ is an acronym meaning “de-militarized zone”), the network manager can create multiple Circuit MOs and point them all at the same SNPA managed object.

2) The Circuit MO has an Adjacency MO for each BIS neighbor on the circuit. These can be either manually created via configuration information (to control the Inter-RD topology directly), or could be created automatically by running ESIS on the circuit. (This latter method hasn’t been worked out yet — it isn’t clear if it would work properly on DMZ subnets.)

3) For each BIS reachable in the local RD which does not have a real adjacency (i.e. is not an ISIS neighbor of the BIS), IDRP creates a "Virtual Adjacency" MO to represent a tunnel to that BIS. (This is done even if the intent is to import all IDRP NLRI as ISIS Reachable Address Prefixes since under some conditions it may be necessary to stop importing the NLRI and use tunnels — see the discussion of ISIS overload below).[9]

4) IDRP create/updates ISIS Reachable Address MOs each time it recomputes its FIBs. It creates one or more Reachable Address MOs for each Inter-RD circuit, and puts in the "Address Prefix" attribute the prefixes for the destinations which are being forwarded to over that circuit. The Reachable Address MO is left "disabled" for destinations for which tunneling is being done. It is Enabled by IDRP when paths to the external destinations are to be computed by ISIS (see below for the details on how to decide whether to tunnel or import the NLRI). Note that this technique allows the existing management capability of ISIS to be used directly to model the complete forwarding state of the IS, since the union of the destinations in the Adjacency, Destination Area, and Reachable Address MOs represents the entire contents of the IS’s FIBs.

5 Importation of NLRI from IDRP into ISIS Reachable Address Prefixes

In order to avoid the processing and bandwidth overhead imposed by tunneling, it is desirable to allow IDRP to export routeing information to ISIS. To accomplish this, ISIS must become aware of destinations outside the local routeing domain through a more robust means than statically-configured reachable address prefixes.

IDRP creates an ISIS Reachable Address MO as described above for each NLRI to be exported to ISIS — the L2 decision process of ISIS then simply computes Intra-domain routes to these destinations as with any other reachable address prefix. These Intra-domain routes allow the NPDUs to reach the proper exit BIS. The issue is to decide which NLRI to export for importation by ISIS. This is done as follows.

[9] It may also be useful to use the Adjacency or Virtual Adjacency MO as the management handle for BIS-BIS connection state, by attaching an IDRP conditional package to the Virtual Adjacency MO.

5.1 Resource limits by ISIS on the Importation of Routeing information

There is an ISIS MO attribute maximumImportedRoutes which enforces an upper bound on the amount of routeing information ISIS is willing to import. The limit is expressed in terms of the number of Reachable Address Prefix MOs created by automatic protocol operation which may be simultaneously enabled.[10] If IDRP wishes ISIS to import some NLRI and the limit is exceeded, IDRP discovers this when the enable operation fails with “insufficient resources”. Under these circumstances it is a local decision what action IDRP takes.

A BIS may export NLRI only for routes that it received from external BISs and that are present in the BIS’s Loc-RIB. Once a BIS determines (by means of IDRP) that a previously supplied NLRI is no longer available, the NLRI needs to be withdrawn from ISIS by disabling the reachable address managed object.

5.2 Policy for Exportation of NLRI by IDRP

The BIS must have preconfigured information for potential NLRI to export for importation by ISIS. This is in keeping with the IDRP philosophy of using explicitly specified policies for controlling the propagation and summarization of all Inter-domain routeing information.

The preconfigured information is in the form of the nLRIToExport and exportPolicy managed objects (contained in the iDRPConfig MO), The nLRIToExport MO has the following attribute:

- addressPrefix — the prefix representing the NLRI IDRP is willing to export.

Each nLRIToExport MO has an instance of the exportPolicy managed object for each supported Intra-domain metric. This managed object has the following attributes:

- exportLonger — a boolean value indicating whether or not the exported information can include more specific NLRI than that indicated by the addressPrefix. If exportLonger is TRUE, the BIS may export any NLRI whose prefix matches the configured prefix. If supplyLonger is FALSE, the BIS may supply only NLRI whose prefix is identical in length (and value) to the configured prefix.

- metricType is an enumeration {internal, external}. See clause 5.4 for how this attribute is used.

- metricValue is a legal ISIS metric value.

- useMultiExitMetric is a boolean indicating whether or not to use the metric present in the IDRP MULTI_EXIT_DISC path attribute.

5.3 Default Routes

The BIS has a boolean-valued attribute exportDefaultRoute which indicates to the BIS whether or not to export a route covering all possible destinations for importation by ISIS. If the value of this attribute is TRUE, and the BIS can reach at least one other routeing domain, the the BIS creates a Reachable Address Prefix containing the null address prefix and enables it.

5.4 Mapping of IDRP metrics when exporting routes

When exporting NLRI for importation by ISIS, the ISIS metricType in the Reachable Adddress MO is set according the metricType attribute of the exportPolicy managed object for the corresponding ISIS metric.

The metric value for each supported ISIS metric is set based on the value of the nLRIToExport and its contained exportPolicy MOs, as follows:

[10] Reachable Address Prefix MOs created by IDRP are distinguished by using a different name binding from those created directly by system management for the purpose of instantiating static routes.

1) If the IDRP attribute multiExit is TRUE, and the IDRP route contains the MULTI_EXIT_DISC path attribute, and the useMultiExitMetric attribute is TRUE, then the metric value is set to the MULTI_EXIT_DISC value divided by 4 and rounded (reflecting the fact that the IDRP value has four times the dynamic range of the ISIS metric value).

2) Otherwise, the ISIS metric value is set to the value of the metricValue attribute.

Note that supplying overlapping NLRI with inconsistent metric types (internal vs. external) may result in ISIS computing a sub-optimal exit point for the traffic, thus forcing IDRP to tunnel the traffic to the correct exit point.

The metric value is set as follows:

1) If multiExit[11] is TRUE, then the metric value is taken from the MULTI_EXIT_DISC path attribute of the NLRI being exported. When setting the ISIS metric value, the MULTI_EXIT_DISC value is divided by 4 and rounded to reflect that the IDRP value has four times the dynamic range of the ISIS metric value.

2) Otherwise, the metric value is set to the value of the metricValues attribute of the nLRIToExport MO.

6 Policies for Deciding dynamically which NLRI to export for importation into ISIS

Amongst the routes in the RD, the BIS has to choose which ones, up to maximumImportedRoutes to export to ISIS. One approach is to sort the prefixes by length and prefer either shorter or longer prefixes. A second is to assign to each prefix a preference value and sort by preference. Neither of these techniques is coupled to the actual traffic matrix, however, and the preference value technique requires extra configuration information and traffic history analysis by the network manager in order to be more effective than a random selection technique. A more difficult, but adaptive technique is to actually observe the traffic to decide which NLRI to supply. This could be done as follows:

1) When a packet arrives from an Intra-RD circuit to be forwarded outside the RD, it is handed to the forwarding machinery, which does a FIB lookup to find the next hop. The FIB entry has a "back pointer" to the reachable address MO containing the longest matching prefix representing this destination (this needs to be known one way or another in order to guarantee longest match routeing as mandated already by ISIS).

2) Each reachable address MO has an LRU timer attribute added to it.

3) If the PDU arrived directly (i.e. not over a virtual adjacency) forwarding proceeds normally, with the extra step of resetting the LRU timer.

4) If the PDU arrived over a virtual adjacency and was decapsulated, then look at the state of the corresponding reachable address MO. If the reachable address is currently enabled, reset the LRU timer as described in (3).

5) If the reachable address is currently disabled, the value of maximumImportedRoutes is not exceeded, and the overload state is not set (see below), then enable the reachable address prefix. This will cause the routeing information for this destination to be propagated through ISIS, and once ISIS converges subsequent PDUs will arrive directly and not over the virtual adjacency.

6) To garbage collect inactive NLRI, when the LRU timer exceeds the garbage collection threshold, the reachable address MO is disabled by ISIS to remove the inactive routeing information.

[11] multiExit is an existing attribute of the iDRPConfig managed object.

7 Dealing with ISIS Memory Overload

It is possible that the union of all of the imported routes from all of the BISs causes some L2 IS in the Routeing Domain to become overloaded. (The overload might be due to some other transient/permanent problem, but the IS can't distinguish *why* the overload happens, so we assume that the imported NLRI are at least part of the problem). In this case it is desirable that the IS reduce the memory load on ISIS by reducing the number of reachable address prefixes enabled. The overload is detected by an interface between ISIS and IDRP to report the overload. A good way of doing this is to add an attribute l2NetOverload to the ISIS package of the cLNS MO which is set true if a scan of the L2 LSP database by ISIS detects that at least one IS in the RD is overloaded, and to have ISIS signal IDRP when it detects the condition. At this point ISIS does the following:

1) Immediately removes some imported routeing information by disabling one or more of the Reachable address MOs. ISIS will consequently reissue the L2 LSP removing the Reachable Address prefix information corresponding to the imported NLRI.

2) While l2NetOverload is TRUE, ISIS refuses to allow Reachable address prefixes to be enabled, by causing the enables to fail with an insufficientResources error.

3) IDRP monitors the value of l2NetOverload. If l2NetOverload becomes False, IDRP may again attempt to enable the Reachable Address MOs to start supplying the NLRI again.

An interesting design issue is just how elaborate and adaptive this machinery ought to be. The simplest approach is to disable all the reachable address MOs, and hence revert to tunneling for all destinations. Like all dynamic schemes, oscillation can result. The frequency of the oscillation is damped by the "Waiting Time" timer, which is sufficiently long to ensure that ISIS converges and some useful routeing occurs over the RD. A simple extension is, of course to first disable all Reachable Address Prefixes routes whose LRU timer (see above) is longer than a minute or so.

One possibility for a more elaborate adaptive scheme is to use the "Multiplicative Decrease/Additive Increase" technique defined to handling congestion control. If this is employed, the BIS, instead of immediately exporting all of the NLRI it is permitted to export (according to the maximumImportedRoutesattribute), uses a "slow start" technique of supplying NLRI for one (or a few) destination(s), waiting for a while (a timer shorter than "Waiting Time" but longer than the L2 ISIS convergence time), and then additively supplying more NLRI. If the overload is detected, the BIS multiplicatively reduces the number of destinations it supplies, using the algorithm described above, and choosing the Reachable Address Prefixes with the oldest LRU timers first. The value of the decrease factor should probably be smaller than the .875 used by congestion control, since the effects of overload are *much* more serious than the effects of larger-than-optimum transport windows. It may need to be .5 or smaller, but we'd need analysis and simulation studies to pick a good value.

The standard need not mandate any particular mechanism, as this can easily be left as a local implementation matter, but some way of capturing the design tradeoffs might be useful as non-normative material.

8 BIS Discovery

Auto-configuration of BISs in the same routeing domain can be done by piggybacking the knowledge of which L2 ISs in the domain are BISs on the normal ISIS flooding machinery. This is accomplished by defining a new option field in the level 2 LSPs to indicate that the reporting IS is a BIS.

One logical way to instantiate this option is to incorporate into ISIS the proposed mechanism in DIS 10747 for reporting the protocols that a BIS supports. If this option is present in a L2 LSP and contains the protocol identifier for IDRP, then the receiving IS knows which other L2 ISs in the routeing domain are BISs, and what their Network entity titles are.

On receipt of a L2 LSP with the "I'm a BIS" option, the BIS creates an "adjacentBIS" managed object for that BIS if one does not exist already, and sets the bisNET attribute from the IS's source ID and lowest area number in the LSP. Then, open processing is initiated for that BIS, if necessary.

9 Routeing Domain Partitions

There are two sub-problems one could address:

1) Getting traffic from outside to the "correct" part of a partitioned routeing domain.

2) Using Inter-RD routes to heal an RD partition

There is general agreement, long-standing, that solving problem (2) is not necessary and would involve breaching the firewall between Intra- and Inter-domain routeing, which we don’t want to do.

Solving problem (1) is potentially desirable, but you need first to be able to detect the partition. This is a hard problem. Therefore, all bets are off if an RD gets partitioned, except in the special case above where IDRP does no summarization.