Title: Comments on DIS 10589 (Intra-domain Routeing)

Source: IBM

IBM votes ″NO″ on the DIS 10589 ballot for two principal reasons:

1. The normative material in DIS 10589 that addresses static inter-domain routeing is outside the scope of an intra-domain routeing information exchange protocol
2. DIS 10589 does not specify an interface between its forwarding process and the forwarding process of a complementary inter-domain routeing function.

Inter-domain routeing of any variety is outside the scope of DIS 10589, and should be removed. Furthermore, the static approach to inter-domain routeing in DIS 10589 compounds the problem because

- it differs significantly from the distributed dynamic approach that is under development in project JTC1 06.41.05 (SC6 N6387), and
- as presented in DIS 10589, it has several serious technical problems or omissions.

DIS 10589 has no normative text to describe its L2 forwarding process in detail. In particular, DIS 10589 is silent about the conditions under which responsibility for forwarding outbound NPDUs will be turned over from the intra-domain protocol to an inter-domain protocol. That is, DIS 10589 does not define an interface between its forwarding process and the forwarding process of a generic inter-domain routeing protocol. (In fact, DIS 10589 does not even provide normative text to define the operation of L2 forwarding with respect to its ″reachable address prefixes″ or its area addresses.) Without such an interface, it will not be possible to address the concerns raised in Sydney that the intra-domain protocol (DIS 10589) and the inter-domain protocol (SC6 N6387) must interwork smoothly.

If inter-domain routeing functions are kept in DIS 10589, IBM will not change its vote to ″YES″, even if current shortcomings are corrected. The IBM position is that DIS 10589 should not be used as a vehicle for implicitly standardizing an inter-domain routeing function. Inter-domain routeing protocols should be developed as separate standards under project JTC1 06.41.05. Therefore, the conditions for changing the ″NO″ vote to a ″YES″ are that all normative text relating to inter-domain routeing will be deleted from DIS 10589, and that new normative text will be added to define the L2 forwarding process in detail, with specific attention to how it will interface to the forwarding process of an inter-domain routeing protocols.

The detailed comments elaborate on these two concerns, and includes suggested text changes to resolve them. All comments should be viewed as directed towards the two concerns listed above, and should not be evaluated as a set of disjoint items. Several minor comments are also discussed.

Major Comments

IBM submits the following major comments:

1. ***Remove all references to inter-domain routeing functions and features from DIS 10589***

   This will involve changes throughout DIS 10589. The major change will be to provide normative text noting that DIS 10589 never delivers NPDUs out of the local routeing domain, but rather forwards them to an appropriate L3 IS located in its own domain. The major implication of this is the need to redefine the usage of ″address prefixes″ as described in a later comment on the L2 forwarding processes.

   Although this list is probably not exhaustive, IBM has identified the following specific portions of DIS 10589 that need to be modified:

   a. Clause 6.1, page 4 (M)

      Insert new text to define a L3 IS as a system which runs a complementary inter-domain routeing protocol. Thus, in the context of L2 route calculation, a L3 IS is a termination point (a leaf node) with respect to the level 2 SPF algorithm. Amend the ″Definitions″ and ″Abbreviations″ clauses accordingly.

   b. Clause 6.6, page 6 (M)

      Delete the item entitled ″Multiple Organisations″.

      Add the following new goal:

      Interaction with Inter-domain Routeing. NPDUs with destination addresses located outside the local routeing domain will be forwarded to L3 ISs in the local domain. Further routeing of such NPDUs is the responsibility of the complementary inter-domain routeing protocol. If a routeing domain contains no L3 ISs, then such NPDUs shall be discarded.

   c. Clause 6.6.1, page 7 (M)

      Add the following non-goal:

      Inter-domain paths. It relies on an inter-domain protocol to forward NPDUs out of the local routeing domain; it will not forward any NPDUs directly over links that leave the routeing domain.

   d. Clause 6.7, page 7 (M)

      Add new environmental requirements on the L3 ISs and their inter-domain routeing protocol: namely, that DIS 10589 relies on L3 ISs to forward NPDUs out of the local routeing domain, and it assumes that the inter-domain protocol running in the L3 ISs provides steady-state loop-free inter-domain paths in the presence of a stable inter-domain topology.

   e. Clause 6.4, page 5 (M)

      This clause is confusing because it attempts to describe the allowable structures of NSAPs and NETs by referring the reader to clause 7.1.1. However, 7.1.1 describes how an encoded NSAP (or NET) will be parsed, but places no constraints on the underlying content of its component fields. It would be more appropriate to refer the reader to 7.1.2, which lists the constraints on system deployment in terms of the content of the component fields of the NSAP (or NET).

Clause 6.4 also talks about end systems whose addresses do not conform to 7.1.1, noting that ″additional configuration″ may be required. However, the term ″additional configuration″ is not defined within the text of DIS 10589. IBM believes that in the original context of DIS 10589, ″additional configuration″ actually meant the representation of an ES (or a collection of ESs) as a level 2 reachable address prefix (that is, it represents the ES(s) as a collection of systems located in a different routeing domain which is reachable via level 2 routeing). This, in effect, would constitute an ″inter-domain routeing function″, and must therefore be deleted when DIS 10589 is revised.

The following new text is suggested as a replacement for the existing text in clause 6.4:

The NSAPs and NETs of all systems shall be structured in accordance with the requirements of ISO 8348/Add.2. Furthermore, within a routeing domain that implements this international standard, Intermediate systems and End systems must be deployed such that their NETs satisfy the guidelines given in clause 7.1.2.

DIS 10589 needs to define the new use of address prefixes which is described in more detail later in this paper. This can be accomplished by inserting the following new material as clause 7.1.3:

7.1.3 Use of Reachable Address Prefix Neighbours

A reachable address prefix is used by L2 ISs to associate the set of destinations denoted by the prefix with a given L3 IS. That is, if a L2 IS announces a reachable address prefix, it shall forward NPDUs matching that prefix to the adjacent (or co-located) L3 IS which is associated with that prefix.

f. Clause 8.1, page 35 (M)

The general concepts in this section are correct, but it needs to be recast to note that L2 routeing no longer forwards NPDUs across a domain boundary. It should state that L2 routeing forwards such NPDUs to L3 routers, and that it is the L3 routers that send information out of the domain. It should also not that the particular algorithm used by a L3 IS for calculating inter-domain paths is not within the scope of DIS 10589.

g. Annex C.3, page 113 (m)

Suitable changes should also be made to Annex C.3 (Forwarding Process) to bring it into alignment with the revised description of the level 2 forwarding process

2. ***DIS 10589 must include normative text to define the interaction of its forwarding process with the forwarding process of a cooperating inter-domain routeing protocol.***

The interface between the two complementary routeing protocols (intra- and inter-domain) should be implemented in their respective forwarding processes: that is, DIS 10589 must define when it will relinquish responsibility for further forwarding of an NPDU and will hand it over to the partner protocol. This is consistent with the hierarchical model of OSI routeing protocols that is used in TR 9575.

The approach outlined below for DIS 10589 is the dual of the approach taken in SC6 N6387′s forwarding process, thus allowing the two protocols to complement each other. Normative text should be developed for inclusion in DIS 10589 which is consistent with the following guidelines:

a. **Identification of Exit Systems**

It is necessary for DIS 10589 to specify a mechanism to identify all ISs in a given routeing domain which implement the complementary inter-domain routeing protocol. Call these systems ″L3 ISs″. An L3 router is a logical entity that may either reside in a system that implements the L2 functions, or it may be located in a physically distinct box that shares a common subnetwork with an L2 IS. (In this paper, the phrase ″adjacent″ is used to encompass both of these cases.)

Identity of L3 ISs enables the intra-domain forwarding process to know where to forward NPDUs whose destination lies outside the local routeing domain. This information can be provided by defining a new code point in the ″IS Type″ fields of the LSPs and IIHs:

- If the L3 and L2 functions are co-located in a single box, then this fact could be denoted by a suitable encoding of IS type for use in the L2 LSPs. This new type code would indicate that the box is actively running both the intra- and the inter-domain protocols.
- If the L2 and L3 functions reside in different boxes, then the L3 IS would need to inform its neighbour L2 Is of its existence and whether it is ″up″. In this case, something like a periodic exchange of L3-L2 Hellos could be used to indicate that the L3 IS is ″up″. The L2 IS could then use the L2 LSPs to report the L3 IS as an ″L3 Neighbour″: this would be a new field to be defined for 10589′s L2 LSP. (The non-colocated case merits further study, and we note that a similar case also applies to IDRP′s operations.)

Thus, as L2 LSPs are distributed throughout the routeing domain, each L2 will become aware of all L3 ISs in its domain: the L2 LSP can report that the IS itself implements L3 functions, or can report the existence of L3 neighbours.

b. **Inter-domain Links**

Only L3 ISs can support a link that exits the routeing domain. That is, unless a level 2 IS also runs an inter-domain routeing protocol, it can not deliver any NPDUs outside of its local routeing domain.

c. **Complete and Partial Source Routeing**

The description of the forwarding process in 7.4 is deficient because it fails to describe the actions to be taken when the 8473 NPDU contains a Source Route parameter. Normative text should be added to state that:

- if an NPDU contains a Complete Source route parameter, it shall be forwarded directly to the next system named in the source routeing list. If the next system in the source routeing list is not adjacent to the IS that received the NPDU, then the NPDU shall be discarded.
- if a partial Source route is specified, the NPDU shall be forwarded on the adjacency in the forwarding database which is associated with the next system named in the partial source routeing list.

d. **Forwarding of Outbound NPDUs**

Outbound NPDUs are those whose source address is in the local RD and whose destination address in a different RD. They will be forwarded according to the following sequence of events:

1) if the destination is in the local area, handle by L1 routeing
2) if the destination is not in the local area, forward to the nearest Level 2 IS
    a) if the destination is located in the local RD, handle by L2 routeing
    b) if the destination is located in a different RD, send the NPDU to nearest L3 IS

Appropriate normative text needs to be included in DIS 10589.

e. **Forwarding Inbound NPDUs**

Inbound NPDUs are those whose source address is in another routeing domain and whose destination address in the local routeing domain. A L2 IS will handle them as follows:

1) if the destination is in the L2′s area, handle by L1 routeing

2) if the destination is not in L2′s area, handle by L2 routeing

Appropriate normative text needs to be included in DIS 10589.

f. **Forwarding Transit NPDUs**

Transit NPDUs are those for which neither source nor destination address in the local routeing domain. DIS 10589 needs normative text stating that there are two options available:

1) The intra-domain protocol may elect to defer handling of such NPDUs to the complementary inter-domain protocol, or

2) The intra-domain protocol may handle them by means of an ″inter-/intra-domain protocol convergence function″. For example, at the discretion of the routeing domain administrator, an inter-domain routeing protocol may specify a set of detailed procedures by which it will make use of intra-domain resources in relaying transit NPDUs. In effect, this will be an ″inter-/intra-domain protocol convergence function″, and both protocols would need to support the convergence function. However, this convergence function is a separate protocol in its own right. For example, at the last meeting, we briefly discussed whether a TR might be the appropriate vehicle for defining a convergence function between DIS 10589 and IDRP.

g. **Choosing a L3 IS Exit Point**

It can be assumed that all L3s that implement an inter-domain routeing protocol act in concert—that is, as a default procedure, DIS 10589 can forward an outbound NPDU to the nearest L3 IS. (This is the case when the L3 IS is a BIS that runs IDRP.)

However, to accommodate cases where a local RD desires to explicitly select an exit L3 IS, we propose the following optional function, which is based on the original ″reachable address prefix″ and ″internal/external metric″ concepts of DIS 10589.

Note: There is a lack of normative text in DIS 10589 in regard to how the L2 forwarding process worked with ″address prefixes″, how it handles ″internal″ and ″external″ metrics, how it matched prefixes with NSAP addresses, etc. The following comments contain suggested normative text to remedy these omissions.

- Redefine the meaning of address prefixes. In this scheme, an address prefix will be associated with a link between a L2 IS and an L3 IS. When a L2 IS announces an address prefix, this means that it will forward outbound NPDUs destined for systems implied by the prefix directly to one of its adjacent L3 ISs for further handling by the inter-domain routeing protocol.

- The L2 forwarding process should be defined so that an outbound NPDU will be forwarded on the shortest path to an L2 that announces an address prefix that matches the destination address of the NPDU. If there is no match between the destination NSAP address and any address prefix in the RD, then the NPDU will be forwarded to the closest L3 IS.

- Matches shall be prioritized in order of prefix length, with matches to the longer prefixes taking precedence. For each prefix length, a match to a prefix with an internal metric shall take precedence over a match to a prefix with an external metric.

  Note: This step is a slight change on the method of DIS 10589. DIS 10589 seemed to discriminate first by categorization as ″internal″ or ″external″, and then subsequently by length. IBM believes that ″length″ should be considered before one discriminates on the type of metric. That is, since longer prefixes are more specific than shorter, they should be given greater preference.

- In DIS 10589, there is text stating that internal and external metrics are not directly comparable, but there is no text which explicitly states how they are to be used in selecting a path. Therefore, normative text needs to be written to define their use, and IBM suggests that it be based on the following functional principles:

  – For a given address prefix, the link between a L2 IS and an adjacent L3 IS can be characterized by an internal metric, an external metric, or both.

  – An internal metric for a given L2-L3 link may be combined directly with other internal metrics for L2-L2 links. For example, as described below, the combined metric for a path to a given L3 IS in some cases is obtained by adding the internal metric for the L2-L3 link to the combined internal metric for the path up to the ″desired L2 IS″.

  – An external metric may not be combined with (that is, added to) an internal metric.

  – For a given address prefix, if there are several L2 ISs in a routeing domain that support links with internal metrics to adjacent L3 ISs, then NPDUs whose destination NSAP address matches that prefix shall be forwarded to the L2 IS on the path that has the smallest combined internal metric.

    If there are several L2 ISs that have the same smallest metric, then a suitable tie-breaking mechanism must be defined, such as forwarding to the L2 IS with the numerically lowest NET, for example.

  – For a given address prefix, if there are several L2 ISs in a routeing domain that support links with external metrics, and no L2 ISs that support internal links, then the path selection entails two distinct parts:

    1) Selection of the L2 IS that has the smallest external metric (call this the ″desired L2 IS″)
    2) Selection of the next hop on the shortest path from the forwarding L2 IS to the desired L2 IS.

    NPDUs whose destination NSAP address matches that prefix shall be forwarded on a path to the L2 IS that has the smallest external metric (that is, to the desired L2 IS). The next hop for such NPDUs shall be an IS on the shortest path to the desired L2 IS, as measured by the smallest combined internal metric between the forwarding L2 IS and the desired L2 IS.

    If there are several L2 ISs that have the same smallest metric, then a suitable tie-breaking mechanism must be defined, such as forwarding to the L2 Is with the numerically lowest NET, for example. There could potentially be two ties that would need to be broken: a) when two or more L2 ISs have links with the same external metric, and b) when there are two or more internal paths to the desired L2 IS, and all paths have the same internal metric. Ties between equal external metrics shall be broken before ties between the internal paths.

Note: In summary, internal metrics are used to calculate the shortest path to the L2 IS that has the smallest external metric for a given address prefix. Note, however, that internal and external metrics are not added together in this process.

Because this function is optional and will only be activated by routeing domains that wish to optimize their internal routeing of outbound NPDUs, it is a local matter (on a routeing domain basis) to define which prefixes will be supported. If the choice of prefixes is coordinated (via a ″convergence function″) with the operations of the complementary inter-domain protocol, then optimum performance will result. However, even if the prefixes are chosen randomly, the inter-domain protocol will still function correctly because of the new environmental requirement which we added to clause 6.7. That is, the trade-offs between better local performance and the need for a larger RIB are strictly a local matter.

3. ***There is no normative text in DIS 10589 to define the prefix encoding and prefix matching processes***

The level 2 forwarding process of DIS 10589 depends upon matching an NSAP address with either an address prefix or an area address. Therefore, it is necessary for 10589 to be clear about several related issues, all of which must appear as normative text within 10589:

- How are address prefixes coded in 10589′s LSPs?

  10589 seems to address this subject in 7.1.4, but the text is somewhat confusing. It speaks of encoding address prefixes in two ways: sometimes the encoded form includes padding, but in other situation the use of padding is excluded. No rationale is given for this approach.

- How are area addresses encoded in 10589′s LSPs?

  10589 doesn′t define this explicitly, but it′s probably correct to assume that they are derived by dropping trailing octets from an encoded NSAP.

- What format is used for quantities which are being matched?

  10589 says that NSAPs are represented in their abstract syntax, but is silent on the format for area addresses and address prefixes. Since area addresses and address prefixes may or may not contain padding characters, it must be made clear if padding is to be part of the matching process.

- What is meant by the ″length″ of a prefix and the ″length″ of an area address in the context of the matching process?

  DIS 10589 has no normative text to address this, and the issue is confusing because of the possible presence of padding.

- What steps are involved in the matching process?

  DIS 10589 has no normative text to address this issue. The informative annex B.1 is more confusing than informative: it jumps back and forth between encoded forms and abstract syntax, implies but does not explicitly require that encoded NSAPs in the NPDUs must be converted back to the underlying abstract syntax before being matched, gives no guidance on how to proceed if the address prefix and the NSAP (in abstract syntax) are of different length, and does not define what is meant by the ″length″ of an address prefix (that is, does ″length″ include or exclude padding that may be present in the address prefix?). IBM is especially confused by the use of a special semi-octet ″F″ to mark the end of the IDP: we can see no reason why the comparison process needs to know where the IDP ends (or even if only part of the IDP is contained in the prefix).[1]

In order to correct these problems, the following suggestions are made:

a. **Provide all address encoding information in a single place in the standard**

This can be accomplished by replacing the existing clause 7.1.4 with the following replacement text. Note in particular that a new title is suggested for this clause since there is no such thing as a ″Level 2 Address″ defined within DIS 10589.

7.1.4 Encoding of Addressing Information

This international standard makes use of four types of address information: NETs, NSAPs, area addresses, and address prefixes. The encoding rules for each of them are given below.

1. NETs shall be encoded according to the preferred binary encoding specified in ISO 8348/Add.2.
2. NSAPs shall be encoded according to the preferred binary encoding specified in ISO 8348/Add.2.
3. The encoded form of an area address shall be obtained by dropping the last $n + 1$ octets of the preferred binary encoding of the corresponding NSAP, where n is equal to the length of the ID field used by the routeing domain.
4. The encoded form of an address prefix shall be derived as follows:
   a. Express the desired prefix as a string of m characters, using the decimal abstract syntax for the IDP and appropriate abstract syntax for the DSP, as required by the value of the AFI field.
   b. Encode the string as if it were a complete NSAP address, using the preferred binary encoding of ISO 8348/Add.2.

b. **Provide a new normative clause to define the matching process**

The following normative text is provided, and should be placed somewhere in the section on the Forwarding Process.

X.Y.Z Matching an NSAP Address with an Address Prefix or an Area Address

The destination NSAP address in an 8473 NPDU can be matched with either an area address or an address prefix. The matching procedures of this international standard operates in an identical fashion on both address prefixes and area addresses. The matching process does not operate on arguments expressed in the preferred binary encoding. All encoded arguments used by the matching process (that is, NSAPs, area addresses, or address prefixes) shall be converted to a decoded form, as described in the following procedure:

[1] We note that the example prefix given in B.1 would also match decoded addresses 371234AF and 37123AF, which do not use the special semi-octet.

1. All semi-octets of the address prefix (or area address) that were inserted as padding by the preferred binary encoding process shall be deleted. Call the resulting string of semi-octets *Target*.
2. The length of the address prefix (or area address) shall be defined as m, where m is the length in semi-octets of *Target*.
3. All semi-octets of the NSAP address that were inserted as padding by the preferred binary encoding process shall be deleted. Call the resulting string of semi-octets $NSAP_D$, the decoded NSAP.
4. If $NSAP_D$ contains fewer than m semi-octets, then the original NSAP does not match the prefix.
5. If $NSAP_D$ contains m or more semi-octets, and each of its first m semi-octets is identical to the corresponding semi-octet of *Target*, then the original NSAP matches the address prefix (or area address). Otherwise, the original NSAP does not match the prefix.

c. Delete Annex B.1 in its entirety

Minor Comments

1. Title (e)

   The title of the document should be amended from ″...routeing exchange protocol...″ to ″...routeing information exchange protocol....″

2. Clause 2.1, page 2 (e)

   Word processor artifact (″IPS-T&IEBS″) in the reference to ISO/IEC 10039 should be expanded to full English title.

3. Clause 6.3, page 5 (m/e)

   The word ″administrative domains″ in the last sentence of the first paragraph should be changed to ″routeing domains″. Since the inter-domain routeing function operates between routeing domains (not explicitly between administrative domains), the suggested change will be a more precise statement of the desired cooperation between the two routeing protocols.

4. Clause 6.5.1 (m/e)

   The discussion of the subnetwork independent functions mentions only ″full duplex NPDU transmission″ but does not touch upon the role of the subnetwork independent link state PDUs. Since discussion of the link state PDUs makes up the bulk of the material in clause 7 and since the routeing information is exchanged via LSPs, additional explanatory material should be added to 6.5.1.

5. Clause 6.8.1.4, page 9 (e)

   Change ″NPID″ to ″NLPID″.

6. Clause 7.1.2, page 10 (m)

   The requirement for all systems in a routeing domain shall to use same length for their ID fields should be included in 7.1.2. Although 7.1.1 defines how an IS shall parse the address information, 7.1.2 does not now contain a complementary requirement on the deployment of systems. The suggested change is to add the following text as a new item (number 4 under item ″a″):

   All systems located within a given routeing domain must have NETs or NSAPS whose ID fields are of equal length.

7. Clause 7.1.5, page 11 (e) In the last sentence, ″precedure″ = = > ″procedure″.

8. Clause 7.2.2, page 12 (e)

   The two paragraphs after the NOTE talk about ″reporting″ or ″not reporting″ a metric. For consistency with the LSP encodings of clause 9, it would be preferable to rephrase these sentences along the lines of ″supporting″ or ″not supporting″ a given metric in order to tie it in with the ″S″ bit of the LSP encodings.

   Also, to improve clarity, it may be desirable rename this bit to ″U″, since its encoding is that a value of ″1″ means ″unsupported″.

9. Clause 7.2.3, page 12 (e) In the 2nd line of the 4th paragraph: it′s = = > its

10. Clause 7.2.10.1, page 15 (e)

    The second sentence of this paragraph is worded in a way that could imply that there could be ″level 2 only routers″. It is suggested that it be reworded as follows: ″Participation in the partition repair process by a Level 2 Intermediate system is predicated on the fact that all L2 ISs also function as L1 ISs within their own area."

11. NOTE on page 18, 1st column (e)

    "and an event signalled..." = = > "and it should signal an event..."

12. Clause 7.3.11, page 22 (e)

    Delete "(i.e., the first 6 octets)"