Title: Consolidated USA Comments on CD-10747 (IDRP)

Source: Project Editor (C. A. Kunzinger)

References: 91-223, 91-225, 91-272, 91-273, 91-307, and 91-308

The attached set of consolidated comments is based on discussions that took place during the review of the referenced input contributions at our September 1991 meeting of X3S3.3.

IBM reconsidered some of its comments, as noted below. Rather than generate a separate document, they are tentatively included in this draft, as noted below.

Since the X3S3 meeting occurs on the Tuesday after our November meeting, I have tried to get the consolidated comments as close to final form as possible.

Notes:

1. In general, the contents of 91-223, -272, -273, -307 and -308 have been included, with minor amendments based on comments received during their review at our September meeting.
2. Contribution 91-225 has been superseded by a suggestion that we describe the cooperative inter-working between CD 10747 and IS 10589 in terms of Linkage MOs, and will recommend that the description be included within CD 10747. No text has yet been submitted to accomplish this.
3. At our last meeting, it was suggested that we consider dropping several IDRP attributes. I have received one response to this suggestion (our document 91-333).Therefore, I have tentatively included a new comment based on 91-333 that recommends removing all vestiges of CO/CL from CD 10747.
4. I have taken no action on 91-293R1--we will consider this at during November′s meeting.
5. IBM has revised the proposed text in Appendix A (Tie Breaking)-- same functions as previous version, but simplified and reorganized for clarity. Having revised this comment, IBM now recommends that it be reclassified as (m), rather than (M).
6. See the boxed text on item 4. Since the task group recommended a new comment (#13) to allow bit-wise granularity of NSAP address prefixes, IBM will withdraw comment #4 unless other members wish to retain it. The material that IBM recommends including in comment #2 is shown there in italics.
7. Based on the analysis of OPEN PDU size in contribution 91-334, IBM suggests a new USA comment (#14) on ″Permissible Sets of Distinguishing Attributes″. I tentatively included this comment for your review. We will discuss it at our November meeting.

The USA casts a vote of DISAPPROVAL of CD 10747, based upon the major technical comments in items 1 and 2. The remainder of the comments are not considered to be major. Upon satisfactory resolution of comments 1 and 2, the USA will change its vote to APPROVAL.

Notwithstanding its vote of DISAPPROVAL, the USA reiterates its strong support for progression of CD 10747. In developing the proposed solutions for items 1 and 2, the USA found that they have no adverse impact on other functions within CD 10747, and thus are believed to be non-contentious.

The USA has classified each of its comments as follows: (M)=major technical, (m)=minor technical, (E)=major editorial, and (e)=minor editorial.

1. ***Precedence of Match with Longest Prefix (M):***

   CD 10747 contains no normative text in 8.1.2.2 stating that a match of an NPDU with a longer prefix should take precedence over a match with a shorter prefix. The USA believes that this omission is an unintentional oversight, since the precedence of matches with longer prefixes is well-known and is an integral part of IS 10589. However, since we found no explicit normative text and the function is essential for correct and unambiguous operation of CD 10747, we classify this as a major technical comment, which can be satisfied by inserting the following sentence as a new last paragraph of clause 8.1.2.2:

   > In cases where a given NPDU matches several address prefixes, the match with the longest prefix shall take precedence.

   Further clarity can also be obtained by adding the following sentence to the first paragraph of related clause 9.4 (which describes how the matching process is used for forwarding NPDUS):

   > ...matches the NPDU-derived Distinguishing Attributes of the incoming NPDU. The incoming NPDU shall be forwarded based on the longest NSAP address prefix that matches (as in 8.1.2.2) the destination NSAP address of the incoming NPDU:

2. ***Handling of Overlapping Routes (M):***

   Within IDRP, a BIS can advertise a set of overlapping routes. Overlapping routes have the same distinguishing attributes, and the respective NLRI fields contain destinations in common. Since NLRI depicts destinations by means of NSAP address prefixes, this means that some prefixes, of different lengths, will be nested inside one another.

   Since the forwarding process selects a next hop for an NPDU based on matching the NPDU′s destination address with an address prefix, there is no ambiguity about what IDRP′s forwarding process will do if a given NPDU matches several address prefixes: IDRP will forward the NPDU based on the match with the longest prefix. However, IDRP contains no normative text to insure that the Decision Process will handle overlapping routes in a way that accurately portrays the actions that will be taken by the Forwarding Process. For example, nothing in CD 10747 prevents a BIS from accepting from a given neighbor BIS only the route with the shortest NSAP address prefix, while rejecting routes from the same neighbor that have longer nested prefixes.

   Therefore, we recommend that normative text should be added to CD 10747′s description of the Decision Process to define how a BIS will handle overlapping routes. The following text is proposed:

   > **8.15.3 Route Replacement**

If an UPDATE PDU is received carrying a route that matches an earlier route received from the same BIS (identical NLRI and distinguishing attributes), the new route replaces the old route, which becomes unfeasible and shall be deleted from the appropriate Adj-RIB-In.

If an UPDATE PDU is received carrying a route whose NSAP prefix is nested within the prefix of an earlier route advertised by the same adjacent BIS, and the path attributes of both routes are identical, then the newly advertised route shall be placed in the appropriate Adj-RIB-In, and no further actions will be necessary.

8.17.2.1 Overlapping Routes

A BIS may transmit overlapping routes to another BIS (routes with overlapping NLRI). NLRI overlap occurs when a set of destinations are identified in non-matching multiple routes with the same set of distinguishing attributes. Since IDRP encodes NLRI using prefixes, overlaps will always exhibit subset relationships. A route describing a smaller set of destinations (a longer prefix) is said to be *more specific* than a route describing a larger set of destinations (a shorter prefix).

When overlapping routes are transmitted from one BIS to another, the more specific routes shall take precedence, in order from most specific to least specific.

This precedence relationship effectively decomposes less specific routes into two parts:

- a set of destinations described only by the less specific route, and
- a set of destinations described by the overlap of the less specific and the more specific routes

The set of destinations described by the overlap comprise a feasible route that is not in use. If a more specific route is later advertised as being unreachable, the set of destinations described by the overlap will then be reachable using the less specific route.

If a BIS receives overlapping routes, the Decision Process shall not alter the semantics of the overlapping routes. In particular, a BIS shall not accept the less specific route while rejecting the more specific route, because the destinations represented by the overlap will not be forwarded along that route. Therefore, a BIS has the following choices:

1. Install both the less and more specific routes
2. Install the more specific route only
3. Install the non-overlapping part of the less specific route only
4. Aggregate the two routes and install the aggregated route
5. Install neither route

3. ***Breaking Ties (m)***

The procedure in clause 8.17.1 defines an unambiguous method for breaking ties within the Decision Process, but this method is inconsistent with the intended use of the MULTI-EXIT_DISC attribute described in 8.13.7. The text proposed in Appendix A, “Proposed Text for Breaking Ties” on page 12 eliminates this inconsistency by explicitly describing the role of the MULTI-EXIT_DISC attribute within the tie-breaking process.

The USA also notes that a need to break ties can also arise within IDRP’s internal update process (8.16.1). Therefore, Appendix A, “Proposed Text for Breaking Ties” on page 12 also outlines similar tie-breaking procedures for use in the internal update process.

4. ***Method to Eliminate Nested NSAP Address Prefixes (E)***

For Consideration

The original text of the proposed new annex is inconsistent with our new comment (#13, ″NSAP Prefix Encoding″), which recommends that granularity should be by bit rather than by semi-octet. Appendix B has been updated to reflect this.

However, when recast with a granularity of a single bit, the methods of this annex appear to be clumsy. Since the material is not normative, IBM offers to withdraw this comment, but would like to see the material in B.6 incorporated into our USA comment #2 (Overlapping Routes).

Since it is possible to replace a set of routes with nested NSAP prefixes by an equivalent set without nested prefixes, an informative annex describing an example procedure would be helpful. The USA suggests adding the following note at the end of the proposed new text for 8.17.2.1 (above), and then adding a new informative annex, using the text shown in Appendix B, "Transforming Nested Prefixes into Disjoint Prefixes" on page 14.

Note: Procedures exist to transform any set of routes with nested NSAP address prefixes into an equivalent set of routes with non-nested prefixes. A description of one such technique is given in Annex ___. A BIS is free to choose to do so as a local option, should it feel that such an approach is preferable to explicit storage and advertisement of routes with nested NSAP address prefixes.

5. ***Inputs to the Degree of Preference Function (m)***

Early discussions in SC6 on inter-domain routeing at the Ottawa meeting in May 1989 noted that if the feasibility of a given route depended on properties of other routes, then oscillations could occur in the route selection process. This premise is already accommodated in IDRP′s clause 8.17, which says that ″The selection process is formalized by defining a function that takes the attributes of a given path as an argument and returns a non-negative integer denoting the degree of preference for the path″.

This important concept can be further emphasized by adding the following statement to clause 8.17 after the second sentence of the second paragraph: ″The degree of preference function for a given path shall not use as its inputs any of the following: the existence of other routes, the non-existence of other routes, or the path attributes of other routes″.

Then, the next sentence should be changed as follows: ″Path selection then consists of individual application of the degree of preference function to each feasible path, followed by the choice of the one with the highest degree of preference.″

6. ***Stability of Routes (m)***

The method outlined in clause 8.11 is limited to the detection of a situation in which a BIS receives a route from an adjacent BIS whose RD_PATH attribute contains one of the RDIs associated with the local BIS. Such a condition typically arises when the advertising BIS is operating with outdated routeing information, and the situation will rectify itself as soon as the advertising BIS receives current routeing information.

Oscillation in the route selection process can occur if:

- The receiving BIS decides to advertise a new route in response to the received UPDATE PDU (that contained the outdated information), and

- The advertising BIS decides to change its selected route before the new UPDATE PDU arrives from the receiving BIS

These conditions can not persist indefinitely unless there is perfect synchronism between the UPDATE PDUs exchanged by the two BISs--that is, the UPDATE PDUs always cross in transit. However, CD 10747 specifies the addition of "jitter" (25%) to the timers used to control the propagation of UPDATE PDUS, thus providing protection against a long-lived situation where the UPDATE PDUs continuously cross in transit.

Thus, although not harmful, the methods of clause 8.11 appear largely superfluous. Hence, the USA recommends that clause 8.11 be deleted in its entirety. As a byproduct of its deletion, Annex K is no longer pertinent, and should hence be deleted as well.

7. ***External Updates (m)***

Clause 8.16.2 (External Updates) requires that the procedures of clause 8.16.1 (Internal Updates) must be done before propagating an UPDATE PDU out of a given routeing domain. In reviewing these procedures, the USA notes that there is no need to impose the restriction in 8.16.2 that internal advertisements must be *acknowledged* before the external advertisements can take place.

The second paragraph of 8.16.2 says that acknowledgement is needed "to insure that consistent information is will be propagated externally". However, since the protocol (see 8.15.2) specifies no corrective action to be taken even if an inconsistency is detected, the mere receipt of an acknowledgement (or lack thereof) does not guarantee consistency. Furthermore, the necessity to wait for an acknowledgement will actually slow down the convergence time of the protocol.

Since waiting for an acknowledgement offers no guarantee of consistent information but does slow down protocol convergence, the USA recommends that the second paragraph of clause 8.16.2 should be deleted in its entirety.

8. ***Destinations within the RD (m)***

IDRP contains mechanisms which make it counter-productive for a BIS's Decision Process to select a route to destinations within its own routeing domain if that route would leave the routeing domain and them later re-enter it. For example,

- The IDRP CLNS Forwarding Process will not forward NPDUs along such a route. or destinations within the RD, IDRP will hand the NPDU over to the intra-domain protocol for forwarding, rather forwarding it out of the routeing domain.
- If such a path were advertised externally, BISs in other RDs would detect a BISPDU looping error since the RDI of the advertising BIS's domain would appear twice. Therefore the advertised route can not be used by BISs located in other RDs.

Hence, the USA recommends that there should be a warning note inserted in clause 8.17 stating that selection of such routes is deprecated, as follows:

> **Note:** The decision process should not select a route to destinations located within the local routeing domain if that route would exit the local local routeing domain and later re-enter it. Such routes would be rejected by other RDs due to the existence of an RD-loop. Furthermore, the IDRP CLNS Forwarding Process will not forward NPDUs (destined to internal destinations) outside of the local RD, but will instead hand them over to the intra-domain routeing protocol.

9. ***Maximum PDU Size of OPEN PDU (m)***

   Until a given BIS receives a valid OPEN PDU from a peer BIS, it does not know the maximum size PDU that its neighbor is willing to accept. Therefore, it is possible that the first OPEN PDU is too large for the neighbor BIS to accept. To avoid this problem, it is suggested that IDRP should define a size for the OPEN PDU that every BIS must be able to handle. Recall that BISPDUs are encapsulated within ISO 8473 NPDUs, and that ISO 8473 can fragment the NPDU as appropriate for specific types of subnetworks.

   It is recommended that IDRP require every BIS to be able to handle all OPEN PDUs whose length is less than or equal to 3000 octets, regardless of the value that the BIS has chosen for its managed object **maximumPDUsize**.

10. ***Retransmission (m)***

    Clause 8.5.3 is deficient in not defining the condition upon which retransmission should be stopped and the BIS-BIS connection should be aborted. We suggest adding the following text:

    > However, if no acknowledgement is received within the time specified in the Hold Time field of the adjacent BIS′s OPEN PDU, then the local BIS shall issue a Stop Event, send a CEASE PDU, and enter the CLOSE-WAIT state.″

11. ***MinRouteSelection Interval (m)***

    The constant value of 30 minutes for the MinRouteSelectionInterval shown in Table 4 will not cause the protocol to fail, but it will result in slow convergence--on the average, there will be a 15 minute ″dead time″ before a newly selected route is advertised. Since inter-domain routes will typically be long-lived, the need to advertise a newly selected route will occur relatively infrequently. However, when a new route is selected, it is desirable to advertise that fact on a timely basis.

    Since there is a trade-off between the frequency of route selection and the convergence time of the protocol, the USA recommends that **routeuteSelectionInterval** should be a parameter rather than a constant. The value of the new parameter should be selectable from a range of values, with a requirement that the minimum acceptable value be 5 seconds. Appropriate GDMO descriptions of the new parameter should be generated.

12. ***IDRP ERROR PDU (m)***

    The ERROR PDU is difficult to parse because the presence or absence of the Error Subcode and Data Fields is context sensitive. Therefore, the Error Subcode field should be made mandatory, with a value of 0 defined as ″No_Error_Subcode″.

    An editorial correction is needed in 7.4 to change ″NOTIFICATION″ to ″IDRP ERROR PDU″ in the section describing ″Data″.

13. ***NSAP Prefix Encoding (m):***

    Although not currently the case, it is possible that future families of NSAP addresses may benefit from allowing prefixes that do not fall on semi-octet boundaries. To provide maximum generality now and avoid having to amend IDRP later, it is beneficial to specify the length of NSAP prefixes in units of bits, rather than semi-octets. To do this, the following changes should be made:

    - In clause 7.3, under header ″Network Layer Reachability Information″, subpart ″a″, replace current text with:

      > The length field indicates the length in bits of the following prefix. A length of zero indicates a prefix that matches all NSAPs.

> **Note:** Although IDRP can express prefixes with a granularity of bits, the use of a bit-level granularity may not be provide consistency with other protocols. For example, IS 10589 expresses its reachable address prefixes with a granularity of semi-octets.

- Strike clause 8.1.2.1 ("An NSAP prefix is said to be of length...")
- Replace item "a" of clause 8.1.2.2 with:

  a) If the encoded NSAP (or NSAP′) contains fewer bits than the NSAP address prefix, then there is no match.

- Replace item "b" of clause 8.1.2.2 with:

  b) If the NSAP (or NSAP′) contains at least as many octets as the NSAP address prefix, and all bits of the NSAP address prefix are identical to the corresponding leading bits of the encoded NSAP address (or NSAP′), there is a match. Otherwise, there is no match.

14. ***Permissible Sets of Distinguishing Attributes (m):***

The material in clause 8.12.2 is technically correct, but it does not present all the relevant information about permissible sets of distinguishing attributes. The constraints on permissible sets of distinguishing attributes arise from the header encodings of ISO 8473 NPDUs. Since the 8473 encodings limit the number and combinations of distinguishing attributes that can be encoded in a given NPDU, they also limit the distinguishing attributes that can appear in a given UPDATE PDU.

The USA suggests that the last two paragraphs of 8.12.2 be replaced with the following new text:

> The distinguishing path attributes map directly into the NPDU-derived Distinguishing Attributes which are obtained by parsing the header of an ISO 8473 NPDU (see 9.2). A permissible set of distinguishing attributes is one that can be legally encoded in the header of an ISO 8473 NPDU.
>
> The NPDU-derived distinguishing attributes are encoded in three parameters of the ISO 8473 header. For each parameter, the following constraints apply:
>
> - The QOS Maintenance Parameter of ISO 8473 is used to encode six distinguishing attributes: TRANSIT DELAY, RESIDUAL ERROR, EXPENSE, CAPACITY, SOURCE SENSITIVE QOS, and DESTINATION SENSITIVE QOS. The encoding of this parameter dictates that at most one of these attributes can be present in a given NPDU (see ISO 8473, clause 7.5.6). Therefore, a valid set of distinguishing attributes can contain at most one attribute from this group.
> - The distinguishing attribute PRIORITY is encoded in its own parameter field in ISO 8473 (see ISO 8473, clause 7.5.7).
> - The ISO 8473 Security Parameter is used to encode both SOURCE SENSITIVE SECURITY and DESTINATION SENSITIVE SECURITY. Since the encodings for these two attributes are mutually exclusive (see ISO 8473 clause 7.5.3), a permissible set of distinguishing attributes can contain only one of them.
>
> Therefore, a valid set of distinguishing attributes can contain at most three members: that is, one distinguishing attribute from each of the ISO 8473 header parameters used to encode the NPDU-derived distinguishing attributes.

15. ***Initializing FSM (m)***

    To insure that the lifetimes of all BISPDUs have expired before a BIS FSM is re-started after a ″system crash″, clause 6.3.1 should require that the BIS must wait for a time period equal to **CloseWaitDelay** after receipt of a Start Event before it enters the OPEN-Sent state.

16. ***Combining Routes (e)***

    The last paragraph of clause 6.6 presents only a partial picture of route aggregation, and hence can be confusing if read in isolation from other normative clauses of IDRP. Since this section of IDRP is not normative, it is suggested that this paragraph should be deleted entirely, and the following sentence be added at the end of the previous paragraph:

    > For example, it is possible under certain constraints to aggregate path attributes, NLRI, or entire routes, as described more fully in clauses 8.17.5 and its subclauses.

17. ***Version Negotiation (e)***

    For clarity, clause 8.7 should mention that the highest version number supported by the local BIS is contained in managed object **version**.

18. ***Handling of SSSEC and DSSEC (e)***

    The material in clause 8.12.3, items ″a″ and ″b″ speaks of constraints on the SSQOS and DSQOS attributes. However, the text does not provide equivalent constraints for SS SECURITY and DS SECURITY, which are treated exactly the same way in IDRP. Therefore, two new items should be added to the list, using the text of existing ″a″ and ″b′, with SECURITY substituted for QOS.

19. ***Removal of References to CO/CL and IFU (m)***

    At the Berlin meeting of SC6/WG2, it was decided that CD 10747 would address connectionless-mode operations only, routeing only ISO 8473 NPDUs. Therefore, the use of IFUs (see ISO TR 10172) will never be required, except possibly as Network Layer Relays. But in that mode, an IFU has no impact on the operation of CD 10747. Therefore, references in CD 10747 to ″CO/CL″ or ″IFU″ serve no useful purpose. The USA recommends that the editor remove all such references, and provides the following specific instances that we are aware of:

    - Delete reference to TR 10172 from clause 2
    - Delete clause 4.6 in its entirety
    - Delete the acronyms ″CONS″ and ″CO″ from clause5.3
    - Delete Type CO/CL 15 in figure 5
    - Delete the CO/CL attribute from Table 1
    - Delete references to X.25 and ISO 10030 from clause 8.13.3, third paragraph
    - Delete clause 8.13.15 in its entirety
    - Delete the item ″CO/CL″ from PICS (Table 13)

20. ***Annex L: Common Subnetworks (e)***

    It was decided in the Berlin WG2 meeting that IDRP will be a CLNP inter-domain routeing protocol, and will not interface directly to CONS services. Since IDRP interfaces directly to ISO 8473, using it as a SNICP, the material in Annex L is no longer relevant, and it should be deleted.

    However, believing it worthwhile to describe how one might ascertain that a pair of BISs are located on the same subnetwork, the USA proposes that the following note be added to clause 6.9.2:

Note: In the absence of an implementation specific method for ascertaining that a neighbor BIS listed in managed object **EXTERNAL-BIS-NEIGHBORS** is located on a common subnetwork with itself, a local BIS can list the neighbor BIS′s NET as the only entry in a complete source route list of the ISO 8473 NPDU that encapsulates its BISPDU. This assures that the neighbor BIS will not receive the BISPDU if the neighbor is located on a different subnetwork from the local BIS.

21. ***Clause 8.17.3 and Clause 8.17.6 (e)***

These clauses are logically out of order with respect to the remainder of clause 8.17. Clause 8.16.3 (Path Selection) should be moved forward, so that it occurs immediately after current clause 8.17; for consistency, the words ″path selection″ should be changed to ″route selection″; and finally, clause 8.17.6 (Interaction with Update Process) should be moved immediately after current 8.17.2, ″Updating the Loc-RIBs″. Then, the clauses will be in a more logical order of presentation: Decision Process, Route Selection, Breaking Ties, Updating the Loc-RIBs, and Interactions with Update Process.

22. ***Clauses 8.17.4 to 8.17.5.6 (e)***

The material presented in these clauses is not logically part of the Decision Process. They discuss ways to organize routeing information efficiently after it has been selected. Therefore, these clauses be located in a new 2nd level clause, to be entitled ″Efficient Organization of Routeing Information″.

23. ***Contents of Information Bases (e)***

Although the contents of IDRP′s routeing and forwarding information bases can be inferred from the text, this material is not presented in a single place within the standard. It is suggested that text should be added to existing clause 6.8 (Selecting the Information Bases), and that it should present the table shown in Table 1 on page 10, which collects and summarizes information about the RIBs and FIBs.

24. ***MD4 References (e)***

The reference to ″MD4″ in Figure 7 is inappropriate since the term ″MD4″ is not defined or mentioned anywhere in the IDRP text. To correct this, the following changes should be made:

- Change ″MD4 Algorithm″ to ″IDRP Checksum Algorithm″ in figure 7
- Insert a bibliographic reference to RFC 1186 in clause 3 (″Informative References″)
- Provide a reference in Annex B to the algorithmic description part of RFC 1186.

25. ***Attribute Numbering (e)***

There is inconsistent numbering of path attribute types in 7.4 and Table 1. The existing type numbers should be corrected as follows: SS SEC=17, DS SEC=18, CAPACITY=19, PRIORITY=20. The editor should also assure that the remainder of the text is also numbered consistently.

26. ***RDIs (e)***

The use of length fields with respect to RDIs is not consistent. In particular, the OPEN PDU uses a length in octets, while the DIST_LIST_INCL and DIST_LIST_EXCL attributes use semioctets.

Since RDIs must be valid NSAPs, they are always encoded as octets. Thus the description of the DIST_LIST_EXCL and DIST_LIST_INCL attributes on page 15 should be changed to say that the length is in octets.

Furthermore, the term ″RDI prefix″ occurs at least in these same two places. This term is incorrect; RDIs are not prefixes, nor are they ever abbreviated. The term ″RDI″ should replace ″RDI prefix″ wherever it occurs.

27. ***Header Length of BISPU (e):***

The length of the fixed header is 29 octets, but in several places in CD 10747, the value 31 is used in error. It should be changed in the following places:

- first sentence on page 18
- In 7.4, change ″...IDRP ERROR PDU is 32...″ to ″...IDRP ERROR PDU is 30...″.
- In 7.5, change 31 octets to 29 octets

Table 1. The IDRP Information Bases. The indexing variables and contents of the RIBs and FIBs are shown.

Information Base	***Indexed by...***	***Contains...***
Adj-RIB-In	• NET of adjacent BIS • RIB-Atts	• Path attributes • NLRI
Loc-RIB	• RIB-Atts	• Path attributes • NLRI
Adj-RIB-Out	• NET of adjacent BIS • RIB-Atts	• Path attributes • NLRI
FIB	• RIB-Atts • NLRI	• NET of next hop BIS • Output SNPA of local BIS • Input SNPA of next hop BIS

Notes:

1. As a local option, a BIS may elect to apply information reduction techniques to path attributes and NLRI information.
2. For each adjacent BIS, a given BIS maintains an Adj-RIB-In for each RIB-Att (including the Empty RIB-Att) that it supports.
3. A BIS maintains a separate Loc-RIB for each RIB-Att (including the Empty RIB-Att) that it supports.
4. For each adjacent BIS, a given BIS maintains an Adj-RIB-Out for each set of RIB-Atts (including the Empty RIB-Att) that it advertises to that neighbor.
5. A given BIS maintains a separate FIB for each set of RIB-Atts (including the Empty RIB-Att) that it has advertised to its neighbor BISs—that is, each FIB corresponds to an Adj-RIB-Out.

   To facilitate the forwarding process, a BIS can organize each of its FIBS into two conceptual parts: one containing information for NLRI located within its own RD, and another for NLRI located in other RDs (see clause 9). For external NLRI, a BIS can further organize the FIB information based on whether the next-hop-BIS is located within its own RD or in another RD (see clause 9.4, items ″a″ and ″b″). And finally, for those next-hop BISs located in its own RD, the local BIS can organize the information according to a specific forwarding mechanism (see clause 9.4, items ″b1″, ″b2″, and ″b3″).

- In 7.6, change 31 octets to 29 octets

28. ***GDMO Notation (E):***

    To align the GDMO description in clause 12 with both DIS 10733 and IS 10589, the USA suggests that clause 12 should be amended as shown by the ″gray background″ material in Appendix C, “Revised GDMO for CD 10747” on page 17.

29. ***Miscellaneous (e):***

    a. Add the words ″or routeing domain confederation″ to the end of the first sentence of 8.1.2.
    b. **MaxRIBIntegrityCHeck**= = = > **MaxRIBIntegrityCheck**
    c. In Annex N, change ″Set STATE=CLOSED″ to ″Set STATE=CLOSED-WAIT″ (pages 88, 90, 93)
    d. In Annex G, adjust notation to reflect that last RD traversed is listed last in the RD_PATH attribute (several places):
        - Page 76, 2nd paragraph: ″right to left″ = = > ″ left to right″ and ″leftmost″ = = > ″ rightmost″
        - Notation <X.*>==> <.*X>, several places on pages 76 and 77 for several values of ″X″

Appendix A. Proposed Text for Breaking Ties

The following text is suggested as a replacement for the existing text of 8.17.1:

8.17.1 Breaking Ties among Routes with Equal Degrees of Preference

In its Adj-RIBs-In, there may be several routes to the same destination that have the same degree of preference and also have an equivalent set of distinguishing attributes. The local BIS can select only one of these routes for inclusion in the associated Loc-RIB. The local BIS considers all equally preferable routes, both those received from BISs located in adjacent RDs and those received from other BISs located in the local BIS′s own RD.

Ties shall be broken according to the following rules:

1. If the candidate routes have identical path attributes or differ only in the NEXT_HOP attribute, select the route that was advertised by the BIS in an adjacent routeing domain whose NET has the lowest value. Otherwise, select the route that was advertised by the BIS in the local routeing domain whose NET has the lowest value.
2. If the candidate routes differ only in their NEXT_HOP and MULTI-EXIT_DISC attributes, and the local BIS′s managed object **Multiexit** is TRUE, select the route that has the lowest value of the MULTI-EXIT_DISC attribute.

   If the managed object **Multiexit** is false, select the route advertised by the BIS in an adjacent RD whose NET has the lowest value. Otherwise, select the route that was advertised by the BIS in the local routeing domain whose NET has the lowest value.
3. If the candidate routes differ in any path attributes other than NEXT_HOP and MULTI-EXIT_DISC, select the route that was advertised by the BIS whose NET has the lowest value.

For purposes of determining the lowest-valued NET, each binary-encoded NET shall be padded with trailing 0′s in order to bring its length up to 20 octets. The encoded (and possibly padded) NETs shall then be treated as unsigned binary integers.

To clarify the role of tie-breaking in the internal update process, the following changes are suggested in clause 8.16.1:

- Add a new item to the numbered list under item ″b″: ″the newly received route is selected as a result of breaking a tie between several routes, each of which have the highest degree of preference, the same destinations, and the same distinguishing attributes″
- Add the following new clause, numbering it as 8.16.1.1:

  8.16.1.1 Breaking Ties in the Internal Update Process

  When a local BIS has connections to several BISs in adjacent domains, there will be multiple Adj-RIBs-In associated with these neighbors. These Adj-RIBs-In might contain several equally preferable routes to the same destination, all of which have the same set of distinguishing attributes and all of which were advertised by BISs located in adjacent routeing domains. The local BIS shall select one of these routes, according to the following rules:

1. If the candidate routes differ only in their NEXT_HOP and MULTI-EXIT_DISC attributes, and the local BIS′s managed object **Multiexit** is TRUE, select the route that has the lowest value of the MULTI-EXIT_DISC attribute.
2. In all other cases, select the route that was advertised by the BIS whose NET has the lowest value.

For purposes of determining the lowest-valued NET, each binary-encoded NET shall be padded with trailing 0′s in order to bring its length up to 20 octets. The encoded (and possibly padded) NETs shall then be treated as unsigned binary integers.

Appendix B. Transforming Nested Prefixes into Disjoint Prefixes

(Informative)

B.1 Decomposition of Elementary Routes

An NSAP address prefix of length K bits can be decomposed into 2 prefixes, each of length $K + 1$ by appending the hexadecimal values 0 or 1 to the end of the original prefix.

Therefore, an elementary route can be decomposed into an equivalent set of 2 elementary routes by decomposing the NSAP address prefix of the original route and setting the path attributes of the decomposed routes equal to the path attributes of the original route.

B.2 Composition of Elementary Routes

A set of 2 NSAP address prefixes, each of length K, can be composed into a single NSAP address prefix prefix of length $K - 1$ if each bit in the range 0 to 1 is present as the last bit of one of the members of the set. If there is a set of 2 elementary routes with the same length NSAP prefixes, but possibly with different path attributes, and their NSAP address prefixes be composed into a single prefix, then the 2 original routes can be replaced by a smaller number of routes by the following procedure:

1. Find the largest subset from among the 2 routes whose members have the same path attributes.
2. Replace this subset with a single route whose NSAP address prefix is composed of the NSAP address prefixes of the 2 routes (and thus is one bit shorter), and whose path attributes are the same as the path attributes of the routes in the subset.
3. Keep all the remaining routes intact.

B.3 Constructing and Dismantling Prefix Trees

Each node of an NSAP address prefix tree (or *prefix tree*, for short) has a single prefix associated with it; furthermore, every non-leaf node has the property that the prefixes of all its children are nested within its own prefix.

If there are several routes with nested NSAP address prefixes, the semantics of this information can be determined by constructing and then dismantling a prefix tree for the nested NSAP address prefixes. This recursive procedure replaces a set of routes whose NSAP address prefixes are nested by an equivalent set of routes whose NSAP address prefixes are not nested. The procedure is as follows:

1. Starting at the root of the tree, decompose it into a forest of 2 trees (see B.1, “Decomposition of Elementary Routes”). Make the newly created nodes the roots of 2 new trees, and attach the chil-

dren of the original root node to the proper roots of the 2 new trees, such that each new tree is a valid prefix tree.

2. If a root of a new tree has only one child node, and that node′s prefix is the same as the prefix of its parent, remove the root (parent) node from the tree, and make the child node the root of the tree.
3. If, after applying steps ″1 on page 14″ and ″2″, there are trees in the forest that have more than 1 child node, arbitrarily select one of them and apply steps ″1 on page 14″ and ″2″ to it.
4. Terminate recursion when the forest contains only isolated nodes.

This procedure is similar to a depth-first search exploration of the tree, since only nodes with children are decomposed, while leaf nodes are not. Repeated application of the procedure replaces the original prefix tree with a forest of isolated nodes. That is, it provides a means to replace a set of routes having nested NSAP address prefixes by a set of routes having disjoint NSAP prefixes. Even though the disjoint prefixes may have different lengths, no two prefixes are nested within each other.

To select a next hop for a particular NPDU, itis sufficient to find a route whose prefix matches the destination NSAP address of the NPDU. Because the procedure above produces routes with disjoint NSAP address prefixes, at most one prefix can match the destination address of the NPDU. Thus, the semantics achieved as a result of these procedures is equivalent to the semantics of longest prefix matching.

B.4 Application of Construction and Dismantling of Prefix Trees

It is possible to have routes with nested NSAP address prefixes in the following information bases:

- in Adj-RIBs-In (either within a single RIB, across multiple RIBs, or both)
- in Loc-RIBs or the associated FIBs
- in Adj-RIBs-Out

These cases can be handled as shown below.

B.4.1 Nested Prefixes in Adj-RIBs-In

For each Adj-RIB-In, apply the procedures of B.3, "Constructing and Dismantling Prefix Trees" on page 14 in order to replace any nested NSAP prefixes with an equivalent set of disjoint prefixes. If a tree whose root has only one child is encountered during the dismantling phase, and the NSAP address prefix of that child is the same as the NSAP address prefix of the root (see step ″2″ of B.3, "Constructing and Dismantling Prefix Trees" on page 14), and the route associated with the child node can not be used by the local BIS (this may happen if the local RDI or one of the RDCIs is already in the RD_PATH), then replace this tree by a forest of trees, as follows: remove both the root and its only child, and make the children of that child (grandchildren of the root) the roots of the trees. At the completion of this phase for a single Adj-RIB-In, that Adj-RIB-In has no routes with nested NSAP address prefix.

After the procedures of B.3, "Constructing and Dismantling Prefix Trees" on page 14 are applied individually to all of the Adj-RIB-Ins, they are then applied to all the routes with nested NLRI that are present in all Adj-RIB-Ins. As a result of this, a BIS will have no routes with nested NSAP address prefixes, and the decision process can be applied to each individual route. The decision process may, as a matter of local routing policy, take into account the length of the prefix of the original route.

B.4.2 Nested Prefixes in Loc-RIBs and FIBs

Once decision process is completed, the selected routes can be installed in Loc-RIB/FIB. To conserve space it may be desirable to compose the routes that are about to be installed in Loc-RIB/FIB by applying the route composition procedure. If such procedure is performed, then the Loc-RIB and FIB are going to have routes with nested NSAP address prefix.

B.4.3 Nested Prefixes in Adj-RIBs-Out

If either all of the routes in the Loc-RIB whose NSAP address prefixes are nested, or none of the routes whose in the Loc-RIB whose NSAP address prefix are nested can be placed in a particular Adj-RIB-Out, then no further actions are necessary.

Otherwise, the procedure of B.3, "Constructing and Dismantling Prefix Trees" on page 14 should be applied to the sets of routes that have nested NSAP prefixes. If in step ″2 on page 15″, the sole child of the root can not be installed in the particular Adj-RIB-Out, then remove both the root and its child, and replace the tree with a forest of trees, where the root nodes of these trees are composed of the grandchildren nodes of the original root. After applying this procedure, all the routes that can not be installed in that particular Adj-RIB-Out are removed. The cardinality of the remaining set of routes can be reduced by applying the route composition procedure (see B.2, "Composition of Elementary Routes" on page 14). This can reduce the amount of information to be exchanged between the adjacent BISs, as well as the amount to be stored in the Adj-RIBs-Out.

B.5 Aggregating Routes with Nested NSAP Address Prefixes

If a BIS desires to aggregate a set of routes (either before installing them in Loc-RIB, or before installing them in Adj-RIB-Out), then it must check whether any of these routes were originally present in any of the Adj-RIB-Ins, or were created as a result of applying the procedures of B.3, "Constructing and Dismantling Prefix Trees" on page 14. In the latter case, it is necessary to find the original route (from one of the Adj-RIBs-In) whose decomposition created the route to be aggregated. Then, after constructing the path attributes of the aggregated route, path attributes of all other routes that were created from the decomposition of the original route should be replaced with the path attributes of the aggregated route.

B.6 Routes with Identical Path Attributes and Nested NSAP prefixes

If a BIS receives from an adjacent BIS an UPDATE BISPDU that contains an elementary route whose NSAP address prefix is nested within a route previously advertised by the same adjacent BIS, and the path attributes of both routes are identical, then the newly advertised route shall be placed in the appropriate Adj-RIB-In, and no further actions will be necessary.

Appendix C. Revised GDMO for CD 10747

(This Appendix will consist of our document 91-307.)