Title: Additional Comments on CD 10747 (IDRP)

Source: IBM and Merit Computer

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

   In reviewing the text, it was discovered that there is no normative text in the Forwarding Process that explicitly states that an NPDU shall be forwarded based on the longest NSAP address prefix match, even though it is common knowledge that this is how the process is intended to work.

   Therefore, the first paragraph of clause 9.4 should be changed as follows:

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

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

   Within the framework of IDRP, it is possible for a BIS to advertise a set of overlapping routes. All routes in such a set have the same distinguishing attributes, and the respective NLRI contain destinations in common. Since NLRI depicts destinations by means of NSAP address prefixes, this means that some of the prefixes will be nested inside each other.

   Since the forwarding process selects a next hop for an NPDU based on longest NSAP address prefix that matches the NPDU′s destination address field, there is no ambiguity even if a BIS has advertised several routes with nested NSAP prefixes. 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 suggested:

   > **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 the old route shall be deleted from the appropriate Adj-RIB.
   >
   > **8.17.2.1 Overlapping Routes**
   >
   > A BIS may transmit overlapping routes to another BIS (routes with overlapping NLRI). NLRI overlaps 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. ***Elimination of Nested NSAP Address Prefixes (E)***

Since it is always possible to replace a set of routes having nested NSAP prefixes by an equivalent set without nested prefixes, IBM suggests that an informative annex describing such a procedure would be helpful. To accomplish this, we suggest adding the following note at the end of the new suggested text for 8.17.2.1 (above), and then adding a new informative annex, using the text shown in Appendix Z:

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 Z. 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.

4. ***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″.

5. ***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.

6. ***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.

7. ***Suggested text in Appendix A of 91-223***

Note: This is not a comment on IDRP; it is a comment aimed at improving the text suggested in prior IBM comment 91-223, Appendix A.

The comparison process called out by the suggested replacement text (″pick the NET with lowest value″) is underspecified, because encoded versions of NETs can be of variable length. To handle such situations unambiguously, it is recommended that it be required, for purposes of comparison, to pad the NET with trailing semi-octets up to the maximum encoded length of an NET, namely 20 octets.

Appendix Z. Transforming Nested Prefixes into Disjoint Prefixes

(Informative)

Z.1 Decomposition of Elementary Routes

An NSAP address prefix of length K semi-octets can be decomposed into 16 prefixes, each of length $K + 1$ by appending the hexadecimal values 0 through F to the end of the original prefix.

Therefore, an elementary route can be decomposed into an equivalent set of 16 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.

Z.2 Composition of Elementary Routes

A set of 16 NSAP address prefixes, each of length K, can be composed into a single NSAP address prefix prefix of length $K - 1$ if each semi-octet in the range 0 to F is present as the last semi-octet of one of the members of the set. If there is a set of 16 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 16 original routes can be replaced by a smaller number of routes by the following procedure:

1. Find the largest subset from among the 16 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 16 routes (and thus is one semi-octet 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.

Z.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 16 trees (see Z.1). Make the newly created nodes the roots of 16 new trees, and attach the children of the original root node to the proper roots of the 16 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″ and ″2″, there are trees in the forest that have more than 1 child node, arbitrarily select one of them and apply steps ″1″ 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.

Z.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.

Z.4.1 Nested Prefixes in Adj-RIBs-In

For each Adj-RIB-In, apply the procedures of Z.3 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 Z.3), 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 Z.3 are applied individually to all of the Adj-RIB-Ins, they are then applied to all the routes with nested NRLI 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.

Z.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.

Z.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 Z.3 should be applied to the sets of routes that have nested NSAP prefixes. If in step ″2″, 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 Z.2). 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.

Z.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 Z.3. 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.

Z.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.