Title: Draft USA Comments on IDRP (SC6 N6387)

To: X3S3.3, for approval and forwarding to X3S3

Source: C. Kunzinger (Project Editor)

Reference: SC6 N6387 (WD on Inter-domain Routeing Protocol)

Task group X3S3.3 has reviewed the referenced working draft, and recommends that the USA support its progression to CD-ballot status out of the Berlin SC6/WG2 meeting. X3S3.3 also requests that the the attached comments be approved for forwarding to the SC6/WG2 secretariat, in accordance with SC6 Sydney Resolution 29 (2.4):1990, for discussion at the July 1991 SC6/WG2 meeting in Berlin.

During its review of SC6 N6387, X3S3.3 developed the following detailed comments, which are organized into three parts:

1. Responses to questions posed in Editor's notes in SC6 N6387
2. Comments on Technical Issues
3. Comments of an Editorial Nature (Clarifications, Additional Informative Material, Typographical Errors, etc.)

For ease of reference, comments are numbered consecutively across all three parts of the attachment.

DRAFT! DRAFT! DRAFT!!

As you know, we did not get a chance to discuss IDRP in detail at February's meeting. Because comments are due to the SC6/WG2 secretariat by May 1, we must approve our USA comments at our April meeting. To facilitate the approval process, I have assembled this TENTATIVE DRAFT. I used my best judgment in formulating these comments, but they must still be discussed and approved by X3S3.3 at our April meeting before they can be forwarded to X3S3.

I tried to include all the comments that I know of, and I tried to capture what I believe was the sense of the task group. If I omitted anyone's comment, if you have a new substantial technical comment, or if you have strong disagreement with anything in this paper, please let me know before March 11. This will allow me to fold in any omissions, and complete a revision before the deadline for the last postal mailing for the April mailing. Your feedback will also help me to plan the agenda for our IDRP discussions in April.

Part I: Responses to Calls for Comment

In document SC6 N6387, there are several ″Editor′s Notes″ to which member bodies were asked to respond. The USA response to each of these items is listed below:

1. **Information about Systems Inside a Routeing Domain (Clause 7.3, page 20):**

   The protocol needs this information to operate correctly, but correct operation does not depend critically on the manner in which it is obtained. This information is used simply to list all the systems located in a given routeing domain, but not whether each system is ″up″ or ″down″. Since it conveys system identity, not operational status, it is not expected to change at a very rapid rate. Therefore, the current text in SC6 N6387 is sufficient, and there is no need to expand upon it within SC6 N6387.

   However, the USA notes that further work on the viability of acquiring this information dynamically would be extremely useful, but does not properly fall within the scope of SC6 N6387.

2. **Authentication Methods (Clause 7.9, page 27):**

   It is appropriate to expand upon the material in Annex B and make appropriate updates in the body of the text to define BISPDU encodings and appropriate usage rules. Annex B should then be deleted, and the revised material should be included in the normative sections of SC6 N6387.

3. **CO/CL Attribute (Clause 7.11.9, page 31):**

   Since CO/CL interworking function units are not within the scope of the OSI Reference Model, and are only described in a TR, the USA recommends that this attribute be dropped from SC6 N6387.

4. **TRANSIT DELAY Attribute (Clause 7.11.10, page 31):**

   A single value of TRANSIT DELAY per routeing domain is sufficient. The benefits from introducing ″path specific delay″ for each possible path between each pair of BISs in a routeing domain would be minimal, and would be far outweighed by the complexity that would be added.

5. **RESIDUAL ERROR Attribute (Clause 7.11.11, page 32):**

   A single RDLRE value per routeing domain is sufficient. The benefits from introducing ″path specific error rate″ for every possible path across a routeing domain would be minimal, and would be far outweighed by the complexity that would be added.

6. **EXPENSE Attribute (Clause 7.11.12, page 32):**

   A single EXPENSE value per routeing domain is sufficient. The benefits from introducing ″path specific expense″ for every possible path across a routeing domain would be minimal, and would be far outweighed by the complexity that would be added.

7. **HIERARCHICAL RECORDING Attribute (Clause 7.11.15, page 33):**

   The USA has developed a detailed technical approach for usage of the HIERARCHICAL RECORDING attribute, as described in [our document X3S3.3/91-005]. The USA recommends that this approach be incorporated into SC6 N6387.

8. **Support for Forwarding of 8208 CR Packets (Clause 7.16, page 42):**

   Although it would be technically possible to route CR packets (under the assumption that the problem of transient looping is soled in some other protocol), the USA has found no justification for

providing this function. The USA recommends that ″Forwarding for CONS″ be deleted from SC6 N6387.

9. **Authentication (Annex C, page 54):**

   As noted above, the material from Annex B should be expanded and included in the normative sections of the standard. Then, Annex B itself should be deleted from SC6 N6387.

10. **Information Exchange between Intra-domain and Inter-domain Routeing Protocols (Annex D, page 58):**

    Development of such material will be extremely valuable. However, it should not be an integral part of SC6 N6387. The USA recommends that this material be presented in a Type 2 Technical Report, and submits [our document X3S3.3/91-31] as suggested draft text.

Part II: Technical Comments

11. **Source Routeing of 8473 NPDUs**:

    The text on the forwarding process should be expanded to address the source route parameters of 8473 NPDUs. For example, there should be normative text stating that a complete source route (if present) takes precedence over the next IS contained in IDRP's FIB.

12. **NEXT_HOP attribute:**

    NEXT_HOP is listed as a well-known mandatory attribute whose value is the NET of the BIS that originates an UPDATE PDU (see clause 7.11.3). Since the NET is known at the time the BIS-BIS connection is established or may be ascertained from the NUNITDATA.INDICATION primitive, the presence of the NEXT_HOP attribute in the UPDATE PDU is redundant. We recommend that this attribute be deleted in its entirety.

13. **Use of ranges:**

    The USA did not see that support for ranges provided significant benefit, but was concerned about the additional complexity that its support would entail. Therefore, the USA recommends that ranges should not be supported in IDRP. However, support for prefixes should be retained in order to provide consistency with the intra-domain routeing protocol (DIS 10589), and to allow for convenient exchange of information between these two protocols.

14. **Handling of the 8473 Security parameter in NPDUs:**

    SC6 N6387 should be expanded to address handling of the 8473 Security parameter: a new distinguishing attribute and associated usage rules are needed.

    An approach is outlined below:

    > The usage rules should be similar to those already defined for SS-QOS and DS-QOS routeing: that is, it can parallel the text of clause 7.11.13-14 of SC6 N6387. The new path attributes would be type-value specific, and the matching process of attributes to the parameters in the 8473 NPDU would be based on the methods of IDRP's clause 7.15.3. It would be necessary to expand clause 7.15.2 so that NPDU-derived Distinguishing Attributes could be obtained from the 8473 security parameter as well as from its QOS-Maintenance parameter.

15. **Supported RIB_Atts**:

    The coding of the OPEN PDU's *RIB-ATTsSET* field depicts "all *RIB_Atts* that the local BIS is willing to support when communicating with the remote BIS." The value of 0 (for number of RIB-Atts) is used to denote that the local BIS will accept (from the remote BIS) only routes that are associated with the Default Attribute. It is suggested that the value "FF" be used to denote that the local BIS will accept routes (from the remote BIS) based on any distinguishing attribute.

16. **Controlling Inter-Domain Routeing Traffic Overhead**

    To ensure the overall stability of the Inter-Domain routeing in the global OSIE, and to avoid potentially excessive overhead associated with the Inter-Domain routeing traffic, IDRP should take certain measures that limit the amount of routeing traffic (that is, BISPDUs) generated by BISs participating in the IDRP. These measures can constrain both the link bandwidth that may potentially be consumed by the Inter-Domain routeing traffic, and the BIS's processing time for handling the Inter-Domain routeing traffic. Several measures are discussed below:

- Frequency of Route Selection

  To limit the amount of Inter-Domain routeing traffic generated by a given BIS as a result of receiving new routeing information about feasible routes from BISs located in other routeing domains, the BIS is allowed to advertise better route (as determined by the local BIS) no more often than once per period, where the period is specified by an architectural constant **MinRouteSelectionInterval**.

  To insure fast convergence within an RD, this limit does not apply to advertisement of better routes that were received from other BIS′s located in the given BIS′s RD. To avoid long-lived black holes, it does not apply to advertisement of previously selected routes which have become unreachable. In both of these situations, the local BIS must advertise such routes immediately.

  If a given BIS has selected new routes that were received from BISs in adjacent RDs, and has not yet advertised the new routes because the **MinRouteSelectionInterval** has not yet expired, the reception of any routes from other BISs in its own RD forces **MinRouteSelectionInterval** timer to expire, and triggers a new selection process that includes both routes received from other BISs in the same RD and from other BISs in adjacent RDs.

- Frequency of Route Origination

  To limit the amount of routeing traffic generated by a BIS as a result of changes to the information about the systems located within that BIS′s routeing domain, there should be a minimum amount of time that must elapse before a change to that information can be advertised. It is suggested that this can be accomplished by defining a new architectural constant **MinRDOriginationInterval** whose value specifies the minimum time interval suggested above.

- Introduction of Jitter

  When BISPDUs are transmitted by a BIS as a result of either receiving UPDATE PDUs or changes in the routeing information internal to the RD that the BIS belongs to, there is a danger that the BIS generated traffic distribution will contain peaks. Where there are a large number of BISs, this can cause overloading of both the transmission medium and the BISs.

  To prevent this from occurring, we propose that ″jitter″ (as defined in DIS 10589) be imposed upon **MinRouteSelectionInterval**, **MinRDOriginationInterval**. Note that a given BIS will apply the same ″jitter″ to each of these quantities regardless of the destinations to which the updates are being sent: that is, jitter is not applied on a ″per peer″ basis.

17. **Length Field in UPDATE PDUs:**

    The additional complexity and the level of indirection involved in the use of an ″Extended Length″ bit in the path attributes flags (see clause 6.3, page 13) seems unnecessary. The USA recommends that the ″Extended Length″ bit be deleted, and that the ″Length″ filed itself be a fixed length of 2 octets.

18. **LOCAL_PREF Attribute:**

    The usage rules for this attribute (clauses 7.11.8 and 7.12.7) should be expanded to state that if this attribute is present in an UPDATE PDU received from a BIS located in another routeing domain, then the receiving BIS shall ignore this attribute.

19. **Sequence Number Rollover:**

    Clause 7.4.2 should be expanded to state that the sequence numbers wrap, and they are represented modulo 2^{32}. There should also be text saying that when a new BIS-BIS connection is established, the first sequence number used should be ″1″. Finally, there should be a requirement that if the next sequence number to be used would result in a rollover, then the originating BIS should wait for a period of 2 minutes and 10 seconds before using it. This will insure that the lifetimes of any BISPDUs that contain numerically higher sequence numbers will have expired before the rolled-over sequence numbers are used.

    It is suggested that a note be inserted to explain that the interval of 2 minutes 10 seconds was chosen because BISPDUs are always transmitted as the data portion of an 8473 NPDU. 8473 NPDUs have a maximum permissible lifetime of 128 seconds, the remaining 2 seconds is a safety factor.

20. **8473 ER PDUs**

    Clause 7.5 needs to contain material that describes how a BIS will handle an incoming 8473 ER PDU.

21. **Finite State Machines**

    Material should be added to clause 7.5.3.3 that defines the actions to be taken when a BIS receives an OPEN PDU while it is in the ESTABLISHED state: if a valid OPEN_PDU is received, the BIS shall generate a Stop Event.

22. **Internal Updates**:

    Clause 7.12.10 describes how a BIS can propagate updates to other BISs located in its own routeing domain. It discusses updates received from BIS located in adjacent RDs, but doesn′t provide sufficient detail for received from other BISs located in its own routeing domain. Therefore, the content of this clause should be expanded, and in particular there should be a normative requirement:

    > If a given BIS receives an UPDATE PDU from another BIS located in its own routeing domain, the receiving BIS shall not re-distribute the routeing information contained in that UPDATE PDU to other BISs located in its own routeing domain.

    Note that this does not apply to updates that are originated by a BIS: it applies only to updates received from other BISs located in its own routeing domain.

23. **Rejection of ″Packet Bombs″**:

    SC6 N6387 should define mechanisms for rejecting ″packet bombs″, which are defined to be bogus BISPDUs delivered to a BIS from an improper source. IDRP can provide this function by requiring that the receive process in clause 7.14 be expanded. The required additional check can be accomplished by amending the second dashed item under ″c)″, as follows (new text is shown in italics):

    > If the SPI identifies IDRP *and the source address of the outer 8473 NPDU identifies any of the systems listed in either managed object* **INTERNAL-BIS** *or managed object* **EXTERNAL-BIS-NEIGHBORS**, then the inner BISPDU shall be extracted....
    >
    > *However, if the source address of the outer 8473 NPDU does not identify a system listed in these managed objects, then the NPDU shall be rejected by the receiving BIS.*

24. **Breaking Ties for Routes with Equal Preference:**

    It is conceivable that there will be routeing policies that will assign the same degree of preference to several routes to a given destination. Thus, a several routes to the same destination could have the same value of the LOC-PREF attribute. To provide determinable performance, the tie can be broken by requiring a BIS to select the route that was advertised by the BIS that has the lowest value NET. (If the local BIS also has a route to the same destination, and its own NET is lowest, then its route will be the one that is selected.) Appropriate normative text should be added to clause 7.13.1.

25. **RDCs and Distribution Lists:**

    The DIST_LIST_INCL and DIST_LIST_EXCL attributes contain lists of RDIs. When confederations are supported, a given RDI could be either the name of a routeing domain or the name of a confederation. When a BIS is a member of a routeing domain which is a member of a confederation, it is possible to associate several RDIs with that BIS: namely, the RDI of its routeing domain and the RDIs of all confederations of which its routeing domain is a member. Note that all of this information is available to a BIS located in an adjacent RD because it is contained in the OPEN PDU which was used to set up the BIS-BIS connection.

    To accommodate the case when several RDIs are associated with a given BIS, the rules for DIST_LIST_INCL should allow the update to be sent to an adjacent BIS whenever one or more of its RDIs are contained in the list. Similarly, the rules for DIST_LIST_EXCL should prevent an UPDATE PDU from being sent to an adjacent BIS whenever one or more of its RDIs are contained in the list.

    Terminology

    Comments 26 through 30 use the terms *Loc-RIB*, *Adj-RIB-In*, and *Adj-RIB-Out* as they are defined in comment 33 on page 14. Use of this conceptual model serves to clarify the specific RIBs which are the subject of each of these comments.

26. **Abbreviated form of an unfeasible route:**

    SC6 N6387 advertises an unfeasible route by attaching the UNREACHABLE attribute to the previously feasible route. This implies that all the path attributes that were present in the feasible route must also be present in the unfeasible route.

    However, in IDRP an Adj-RIB-In or an Adj-RIB-Out holds at most one route to a particular destination as specified in the Network Layer Reachability Information, and the distinguishing attributes of the route unambiguously identify a particular Adj-RIB. Since the combination of the distinguishing attributes and the Network Layer Reachability Information unambiguously identifies a previously feasible route the following shortcut is possible: to announce that some previously feasible route has become unfeasible, it is sufficient to supply only the Network Layer Reachability Information and the distinguishing attributes (rather than all the attributes) of the previously feasible route.

27. **Interaction between Decision and Update Processes:**

    SC6 N6387 should require the Decision Process to run to completion based on currently available routes before any newly arrived routes are used in the computation. The following text takes an approach similar to that of DIS 10589, and the USA suggests that it be incorporated into SC6 N6387:

    > ″Since the Adj-RIBs-In are used both to receive inbound UPDATE PDUs and to provide input to the Decision Process, care must be taken that their contents are not modified while the Decision Process is running. That is, the input to the Decision Process shall remain stable while a computation is in progress.

There are two approaches that could be taken:

1. The Decision Process can signal when it is running. During this time, any incoming UPDATE PDUs will be queued and will not be written into the Adj-RIBs-In. If more UPDATE PDUs arrive than can be fit into the allotted queue, they will be dropped and will not be acknowledged.

2. A BIS can maintain two copies of the Adj-RIBs-In—one used by the Decision Process for its computation (call this the Comp-Adj-RIB) and the other to receive inbound UPDATE PDUs (call this the Holding-Adj-RIB). Each time the Decision begins a new computation, the contents of the Holding-Adj-RIB will be copied to the Comp-Adj-RIB: that is, the a snapshot of the Comp-Adj-RIB is used as the input for the Decision Process. The contents of the Comp-Adj-RIB remain stable until a new computation is begun.

The advantage of the first approach is that it takes less memory; the advantage of the second is that inbound UPDATE PDUs will not be dropped. This international standard does not require that either of these methods be used. Any method that guarantees that the input data to the Decision Process will remain stable while a computation is in progress and that is consistent with the conformance requirements of this international standard may be used.″

28. **Checksum Considerations**

According to clause 7.8, a BIS sends a separate CHECKSUM PDU to each of its neighboring BISs, and computes the checksum over the portion of its Loc-RIB(s) that has been advertised to the adjacent BIS. In terms of the more precise conceptual model of comment 33 on page 14, this is equivalent to stating that a ″BIS computes a separate over the Adj-RIBs-Out that have been advertised to the adjacent BIS″. SC6 N6387 assumes a direct correspondence between the Adj-RIBs-Out of the sending BIS and the Adj-RIBs-In of the receiving BIS: that is, it expects the receiving BIS to maintain multiple Adj-RIBs-In, one for each of the Adj-RIBs-Out which its partner BIS used when it generated the CHECKSUM PDU. If the receiving BIS does not maintain the requisite Adj-RIBS, then the database integrity check will fail.

However, it is conceivable that the receiving BIS, for reasons of its own (such as local policy or memory constraints) may not wish to support multiple Loc-RIBs. In such a situation, there would then be no need for it to maintain the associated Adj-RIBs-In, since it would never utilize any of the information contained in them. However, if these Adj-RIBs-In are not present (even though their contents will not be used for route computation), it will not be possible under the current methodology to compute a checksum that agrees with the one contained in the CHECKSUM PDUs received from its partner BIS.

To allow each BIS the freedom to support only the Adj-RIBs-In and Loc-RIBs that actually will be used for route computation, we propose that a BIS should compute individual checksums over the portion of each of its Adj-RIBs-Out that has been advertised to a neighbor BIS, and that each checksum be passed to the partner BIS separately (either in a single CHECKSUM PDU, or in multiple CHECKSUM PDUs).

In SC6 N6387, each RIB (Adj-RIB-In, Adj-RIB-Out, and Loc-RIB) is unambiguously identified by its RIB-Att, which consists of a set of distinguishing attributes (see SC6 N6387, clause 7.12.3). Therefore, to receiving BIS to identify the Adj-RIB-In to which a given checksum is related, the format of the CHECKSUM PDU should be associate each checksum value with the RIB-Att of the Adj-RIB-Out over which the checksum was computed.

The proposed format allows several <RIB-Att, checksum> pairs to be included in a single CHECKSUM PDU; each RIB-Att field is encoded exactly the same as in an OPEN PDU.

Fixed Header
First RIB-Att
First Checksum (2 octets)
Second RIB-Att
Second Checksum (2 octets)
....
Last RIB-Att
Last Checksum (2 octets)

The USA also notes that SC6 N6387 does not provide clear procedural rules for computing the checksums. The following rules are suggested, and should be included in the normative sections of SC6 N6387.

When a BIS computes a checksum over an individual information base (Adj-RIB-Out when the BIS is generating a CHECKSUM PDU, or an Adj-RIB-In when a BIS is computing a checksum to be compared with the value contained in an incoming CHECKSUM PDU), the following rules shall be observed:

1. A sequence number shall be associated with each route in the information base: it shall be the sequence number of the BISPDU used to transmit (for Adj-RIB-Out) or receive (for Adj-RIB-In) the UPDATE PDU that contains the given route.
2. When a checksum is computed over the routes contained in a single information base, those routes shall be sorted in a non-decreasing order of their sequence numbers.
3. Within each route, path attributes shall be sorted in a non-decreasing order based on their type codes.
4. Within each route, Network Layer Reachability Information shall be sorted:
   a. prefixes shall be placed ahead of ranges
   b. prefixes shall be sorted in lexicographical order, based on their binary values
   c. ranges shall be sorted in lexicographical order, based on the binary value of the lower limit of the range.
   d. If several ranges have the same lower limit, then they shall be further sorted in lexicographical order based on the binary value of their upper limits.
5. A checksum shall be computed according to the ISO 8473 algorithm. This algorithm shall be applied to the data as sorted by the previous rules, and the sorted data shall be treated as a sequence of octets.

29. **Handling BIS Overload**

Due to misconfiguration or certain transitory conditions, it is possible that there may be insufficient resources available at a particular BIS to correctly implement the procedures of IDRP. In this situation we say that the BIS becomes overloaded.

There are two distinct overload conditions: memory overload and CPU overload. For IDRP-based information exchange methods, CPU overloads are potentially more harmful than memory overloads because the IDRP route computation phase must precede its route distribution phase. That is, in a CPU-overload situation, the UPDATE process should be halted and priority should be given to the completion of the Decision Process's computation.

Since the remedies suggested for memory overload could be construed as enactment of local policies, this material is advisory in nature, and should be located in a new informative annex. However, the material dealing with CPU overload should be part of the normative text of SC6 N6387.

- Memory Overload:

  We say that a BIS becomes memory overloaded when there is not enough memory to store both the Adj-RIBs-In (that are used to store the routeing information as received from other BISs), and the Loc-RIBs (that are derived by the decision process, using the information in the Adj-RIBs-In as one of the inputs).

  Since the Loc-RIBs form a subset of the Adj-RIBs-In, the amount of memory needed to store the Adj-RIBs-In is greater than or equal to the amount of memory needed to store the Loc-RIBs. Therefore, the first step to alleviate the memory overload condition would be to reduce the amount of information that is stored in Adj-RIBs-In. That can be accomplished by removing routes that are not in the Loc-RIBs (that is, those routes that have not been selected by the Decision Process for advertisement to other BISs).

  Clearly, all routes in Adj-RIBs-In to destinations that are not in the Loc-RIB may be removed without no negative impact. Even if Adj-RIBs-In have routes to destinations that are in the Loc-RIBs as well, it still may be possible to remove some of these routes. Since these routes may potentially be used as a fallback routes (if the current route that is in the Loc-RIB becomes unfeasible), removing them from the Adj-RIBs-In may cause at some point in the future suboptimal connectivity.

  If several Adj-RIBs-In (that have the same RIB attribute) have routes to the same destination, then routes with higher degree of preference (as computed by the local BIS) should be retained, while routes with lower degree of preference may be deleted, thus reducing the amount of memory needed for the Adj-RIBs-In. To ensure routeing consistency within an RD, the above procedure may be applied only to the Adj-RIBs-In associated with BISs in adjacent RDs.

  (Note that if a BIS unilaterally deletes a route, then it will not be able to compute a correct checksum for comparison to the CHECKSUM PDU received from a neighbor. Note also that any solicted refresh (see comment 30 on page 11) will only serve to reinstate the deleted route. Hence, if the condition persists, the memory-overloaded BIS should close the IDRP connection, and then take corrective action, such as re-opening it with an OPEN PDU that indicates support for a smaller **RIB-ATTsSet**, for example.

  A more drastic measure would be to terminate one or more of the IDRP sessions with other BISs. That would result in releasing the memory that was previously used to store the Adj-RIB-Ins and the Adj-RIBs-Out associated with that BIS. To ensure routeing consistency within an RD this measure may be applied only to the IDRP sessions with BISs in adjacent RDs. If the above measures do not alleviate the memory overload condition, the local BIS terminates all of its IDRP sessions.

- CPU Overload:

We say that a BIS becomes CPU overloaded when there is not enough CPU processing power to process incoming BISPDUs received from other BISs. In this situation BIS must continue to update the Adj-RIBs-In with information contained in BISPDUs received from other BISs, but may not run the Decision Process using this information except for routes received with the UNREACHABLE path attribute.

If a route received in the UPDATE_PDU has the UNREACHABLE path attribute, the local BIS checks whether this route is currently installed in one of the Loc-RIBs; if so, it removes it from the appropriate Loc-RIB, updates the appropriate FIB, and generates (if necessary) an UPDATE-PDU to inform other BIS's of the change in its Loc-RIBs and its Adj-RIBs-Out. The Decision Process on the local BIS does not select another to replace the one that becomes unfeasible.

Since this procedure decreases the size of the Loc-RIB, persistence of a CPU overload condition can eventually deplete the entire Loc-RIB, thus making the BIS unavailable as an intermediate system. If the CPU overload condition disappears, then the Decision Process and Update Process should be run over all the new routes that were installed into the Adj-RIBs but have not yet been processed by the Decision Process. If the CPU overload condition persists for more than the predefined architectural constant **MaxCPUOverloadPeriod**, the local BIS terminates its IDRP sessions.

The order of termination of the IDRP sessions is significant. First the BIS may terminate one or more of the IDRP sessions with BISs in adjacent RDs. If after terminating IDRP sessions with all of the BISs in adjacent RDs the CPU overload still persists, the BIS terminates the rest of its IDRP sessions (with all the BISs within its own RD).

30. **Solicited and unsolicited refresh of Adj-RIBs:**

In certain situations (e.g. memory overload) a BIS may have to purge some of the routeing information stored in its Adj-RIBs-In. If such purges occur, the Database integrity scheme (CHECKSUM PDU) will not work correctly. Therefore, we propose to add the solicited and unsolicited Adj-RIB-In refresh capability to the IDRP. Addition of such capability requires a new type of BISPDU, called the RIB-REFRESH PDU, with the following format:

Fixed Header
OpCode (1 octet)
Variable part

Currently defined OpCode are:

1 - RIB-Refresh-Request
2 - RIB-Refresh-Start
3 - RIB-Refresh-End

- **Solicited Refresh**

A BIS may request the refresh of one or more of its Adj-RIBs-In by sending a RIB-REFRESH PDU that contains the OpCode for RIB-Refresh-Request, and it can restrict the scope of refresh by specifying the RIB-Att of the Adj-RIB-In that it wants to refresh.

When a BIS receives a RIB-REFRESH PDU with OpCode RIB-Refresh-Request, it sends back RIB-REFRESH PDU with OpCode RIB-Refresh-Start, followed by a sequence of UPDATE PDUs that contain the information in its Adj-RIBs-Out that have been advertised to the requesting BIS.

The completion of the refresh procedure is indicated by sending the RIB-REFRESH PDU with OpCode RIB-Refresh-End.

- **Unsolicited Refresh**

  A BIS may initiate an unsolicited refresh by sending a RIB-REFRESH PDU with OpCode RIB-Request-Start, followed by a sequence of UPDATE PDUs that contain the information in its Adj-RIBs-Out that been advertised to a given BIS. The completion of the refresh is indicated by sending the RIB-REFRESH PDU with OpCode RIB-Refresh-End.

If the refreshing BIS receives a RIB-Refresh-Request while it is in the middle of refresh (after sending RIB-REFRESH PDU with OpCode RIB-Refresh-Start, but before sending RIB-REFRESH PDU with OpCode RIB-Refresh-End), then the current refresh is aborted and the new refresh is initiated.

If the BIS being refreshed receives a RIB-Refresh-Start in the middle of refresh (after receiving a RIB-REFRESH PDU with OpCode RIB-Refresh-Start, but before receiving the RIB-REFRESH PDU with OpCode RIB-Refresh-End), then the current refresh is aborted and the new refresh is initiated. That is, a refresh cycle may be terminated either by receipt of RIB-Refresh-End or by receipt of a new RIB-Refresh-Start.

If the OpCode is RIB-Refresh-Request, then the variable part of the RIB-REFRESH PDU contains the RIB-Atts of the Adj-RIBs for which a refresh is being requested. For all other OpCodes the Variable Part is empty.

Part III: Editorial Comments

31. **More Descriptive Title**:

The current title for SC6 N6387 can in many cases lead to confusion on the part of new readers: it is long, awkwardly constructed, and almost identical to the title of DIS 10589 (″inter″ vs. ″intra″). Since SC6 N6387 deals primarily with the protocol for exchanging routeing information, and it defines the rules which must be followed for advertising locally selected routes to other systems that are participating in an instance of IDRP, the USA suggests that the title of SC6 N6387 be changed to *Protocol for the Exchange of Inter-domain Routeing Information among Intermediate Systems*.

32. **Usage of the MULTI_EXIT_DISC attribute**

The MULTI_EXIT_DISC attribute can be used as a means for exchanging routeing information between the inter-domain protocol and an intra-domain protocol. However, SC6 N6387 provides no clarifying examples for the use of the MULTI_EXIT_DISC attribute. The USA submits the following illustrative example, and suggests that it be incorporated in a new informative annex to SC6 N6387:

EXAMPLE OF MULTI-EXIT_DISC USAGE

The MULTI-EXIT DISC attribute can be used to provide a limited form of multi-path (load-splitting), as is shown in the following examples.

- Example 1 (see Figure 1 on page 14):

  Consider the case when a BIS A located in routeing domain RD-A has two adjacent BISs (B1 and B2) that belong to the routeing domain RD-B. Assume that RD-B has Network Layer Reachability information about NSAPs N1, N2, ... Nk, and it wants to advertise this information to RD-A. By using the MULTI-EXIT_DISC attribute RD-B may do selective load splitting (based on NSAP addresses) between B1 and B2.

  For example, BIS B1 advertises to BIS A Network Layer Reachability information N1, N2, ... Nm with the MULTI_EXIT_DISC set to X, and advertises N(m+1), ... Nk with the MULTI_EXIT_DISC set to X + 1.

  Similarly, BIS B2 advertises to BIS A Network Layer Reachability information N1, N2, ... Nm with the MULTI_EXIT_DISC set to X + 1, and advertises N(m+1), ... Nk with the MULTI_EXIT_DISC set to X.

  As a result, traffic from BIS A that destined to N1, N2, ... Nm will flow through BIS B1, while traffic from BIS A that destined to N(m+1), ... Nk will flow through BIS B2. This scenario illustrates the simplest way of doing limited multipath with IDRP.

- Example 2 (see Figure 2 on page 15):

  Next consider more complex case where there is a multihomed routeing domain RD-A that has only slow speed links. RD-A is connected at several points to a transit routeing domain RD-B that has only high speed links; BIS A1 is adjacent to BIS B1, and BIS A2 is adjacent to BIS B2. RD-A wants to minimize the distance that incoming NPDUs addressed to certain ESs—say ES(1) through ES(k)—will have to travel within RD-A.

  One way of doing this is by making BIS A1 to announce to BIS B1 destinations ES(1) – ES(k) with a lower MULTI_EXIT_DISC, as compared to the MULTI_EXIT_DISC that BIS A2 will use when announcing the same destinations to the BIS B2. Similarly, BIS A2 would announce to

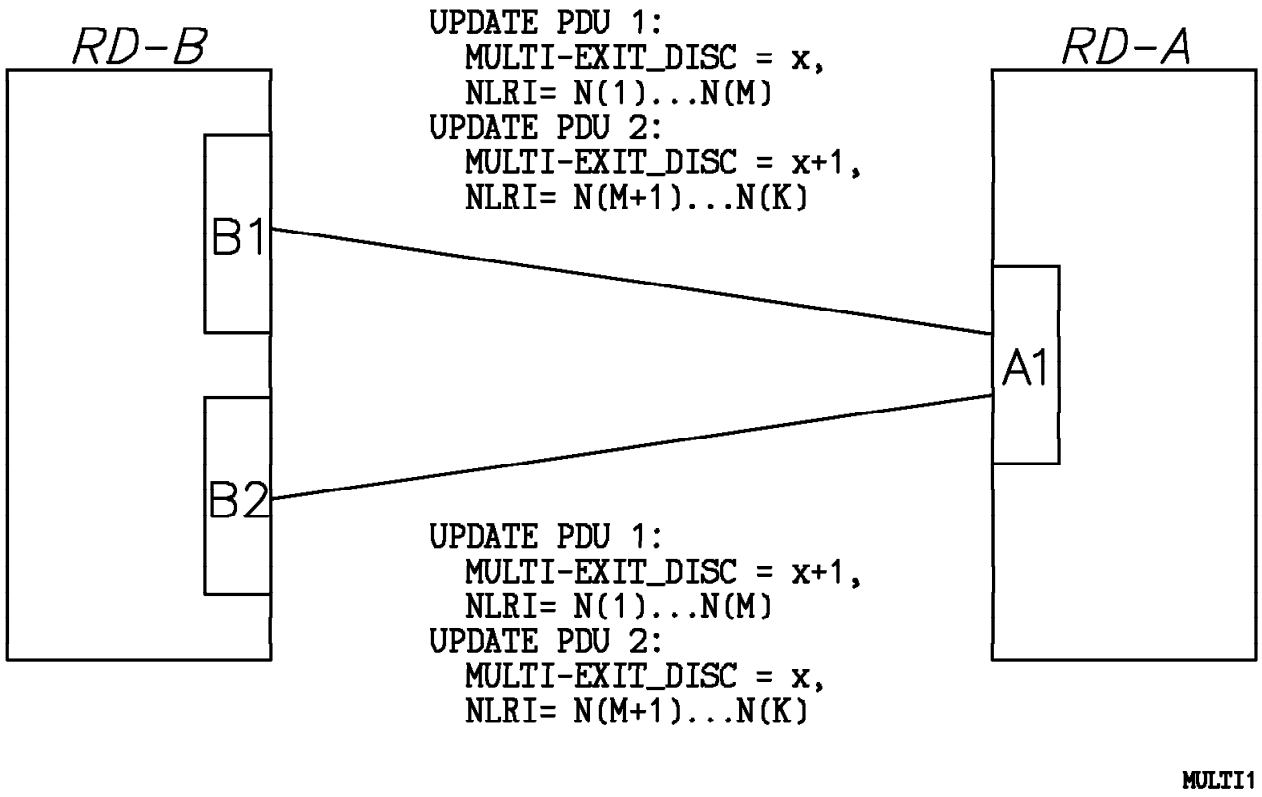

Figure 1. Example 1 Configuration

BIS B2 destinations ES(k=1)– , ES(n) within the RD-A that are closer to the BIS A2 (than to the BIS A1) with the lower MULTI_EXIT_DISC, as compared to the MULTI_EXIT_DISC that the BIS A1 will use when announcing the same destinations to the BIS B1.

When traffic that destined to some ES within RD-A enters RD-B on its way to RD-A via BIS X, X picks up the exit BIS that has the lowest MULTI_EXIT_DISC value for that destination. For example, X may pick up BIS A2 as an exit, even if the distance between A2 and X is greater than the distance between A1 and X.

33. **Conceptual Model for RIB(s)**

SC6 N6387 describes a routeing information base as being composed of two conceptual parts: the Loc-RIB(s) and the Adj-RIB(s). In this model, the Adj-RIB(s) are the repository for routeing information received from other BISs.

Note: The use of a conceptual model which distinguishes between Adj-RIBs-IN, Adj-RIBs-Out, and Loc-RIBs in no way implies that an implementation must maintain 3 separate sets of buffer hardware. The choice of implementation (3 separate buffers, one buffer with pointers to different parts, etc.) is not constrained by this standard.

Additional clarity can be obtained by revising the conceptual model so that it contains three parts:

- Loc-RIB(s): This part of the model remains unchanged from that of SC6 N6387.

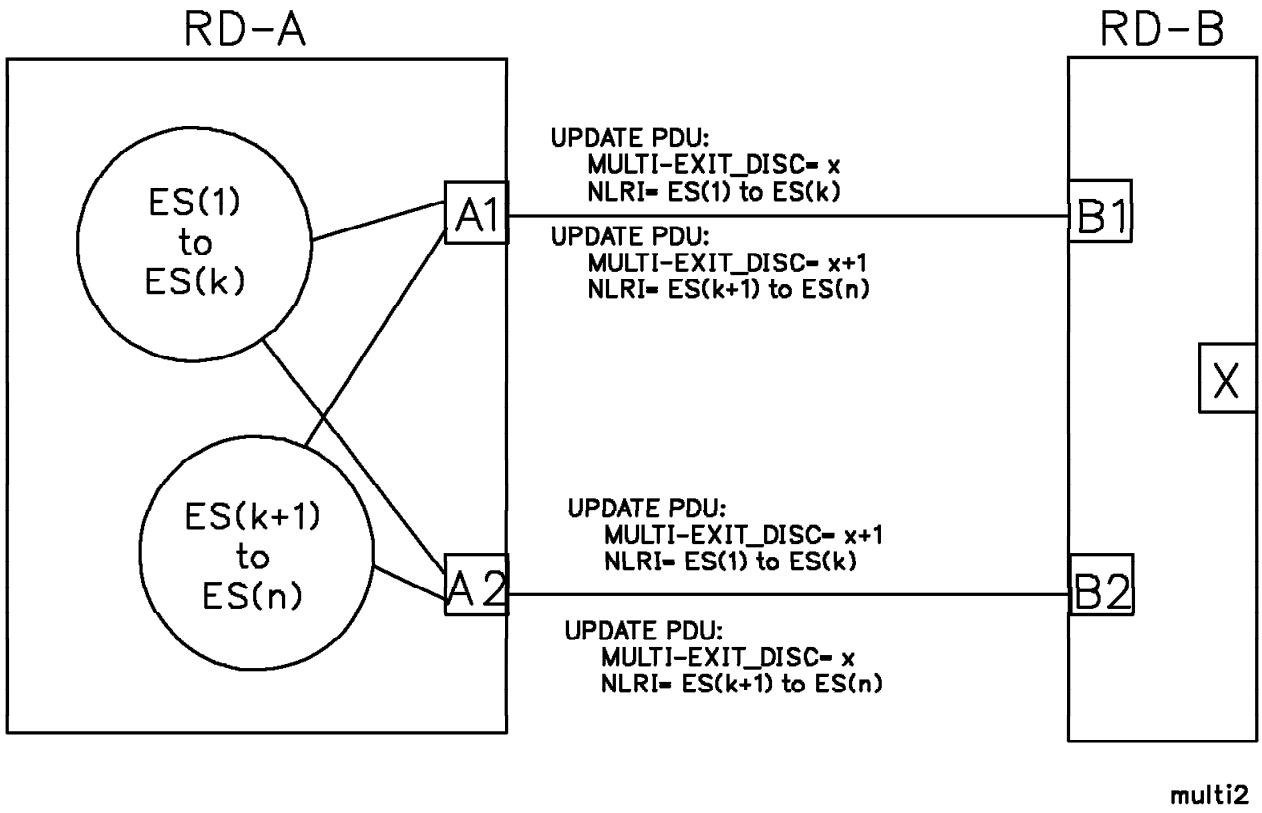

Figure 2. Example 2 Configuration

- Adj-RIB(s)-In: This part of the model identifies the storage used to hold incoming routeing information received from other BISs. (It corresponds to ″Adj-RIB(s)″ of SC6 N6387.)
- Adj-RIB(s)-Out: This is the new part of the model. It identifies the portion of storage used to hold routes which will be advertised to other BISs. Note from the picture that it is legitimate, for example, to have only one Adj-RIB-Out for a given neighbor even if there are several Loc-RIBs. This could occur, for example, if the receiving BIS only supported the Default Attribute while the advertising BIS supported the Default Attribute plus others.

Figures 4 and 8 of SC6 N6387 would be replaced with the Figure 3 on page 16 and Figure 4 on page 17

34. **Conformance-related Language:**

The following comments list several sections of SC6 N6387 in which the language needs to be amended to be consistent with the conformance requirements for the standard:

a. Clause 6.3: The ″must″ and ″may″ should be changed to ″shall″ in the descriptions of the various flag bits of the Path Attribute field.

b. Page 14, last sentence under DIST_LIST_INCL: ″can not″-->″shall not″

c. Clause 6.4: In items ″a″, ″b″, and ″c″ of the first paragraph: ″will″--> ″shall″

d. Clause 7.5.4, next-to-last paragraph: ″will″--> ″shall″.

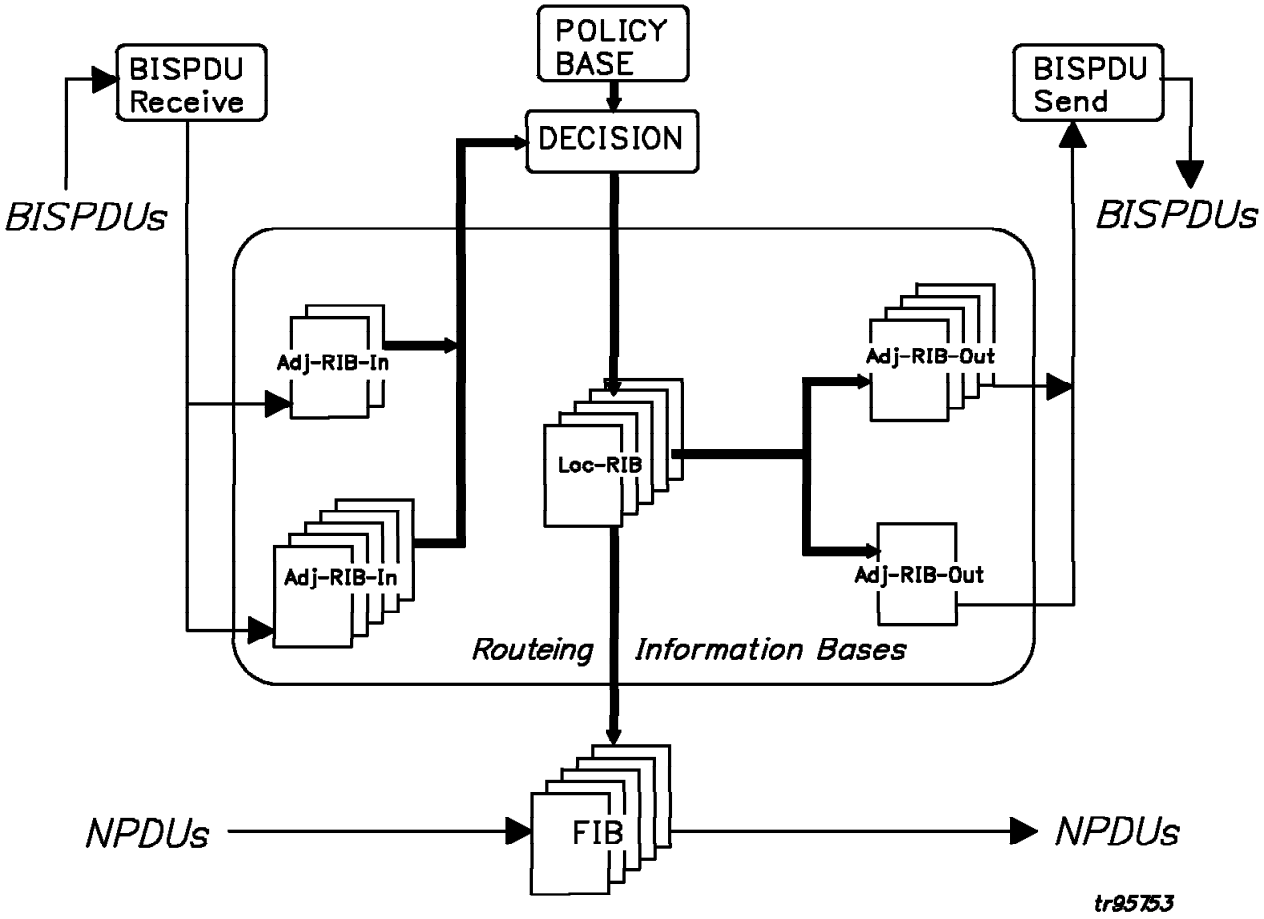

Figure 3. Replacement for SC6 N6387 Figure 4

e. Clause 7.9, first sentence:″must be all ones″--> ″shall consist of all one bits″.

f. Clause 7.11.2, first sentence (2 occurrences), ″must″-->″shall″.

g. Clause 7.12.1, second paragraph, ″must be able″ --> ″shall″ (2 places)

35. **Use of ″Notes″ for Parenthetical Material:**

There are several sections where the written material should be presented as a ″Note″ rather instead of being incorporated directly into the body of the standard:

a. Clause 7.1.3: The parenthetical material at the end of this clause should be presented as a Note.

b. The last paragraph of this 7.4.4 should be presented as a NOTE, because no particular algorithm is specified within IDRP.

c. The very last paragraph of 7.4.5 should be presented as a NOTE.

d. Clause 7.6.4: The end of this paragraph, beginning with ″Any error...″ should be presented as a NOTE.

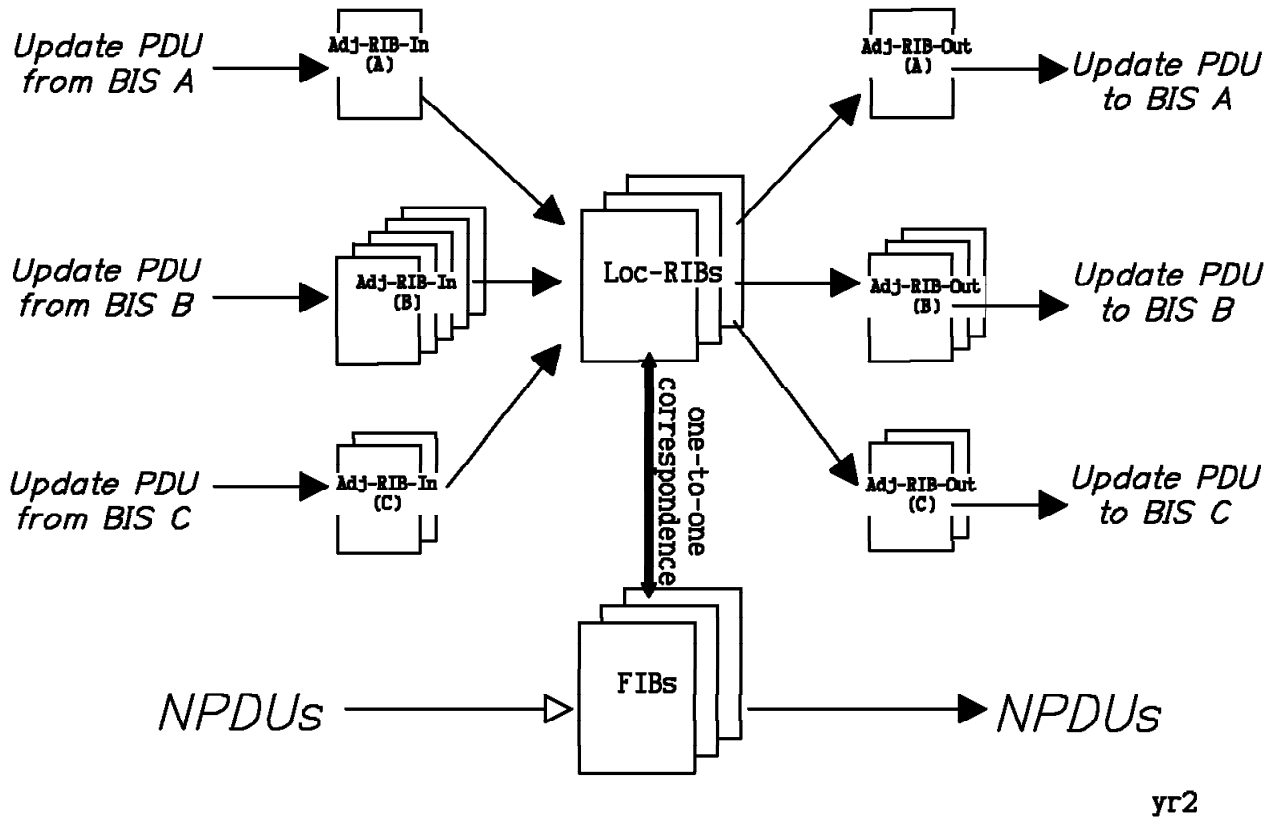

Figure 4. Replacement for SC6 N6387 Figure 8

e. Clause 7.11.1, last paragraph (top of page 29) should be presented in a NOTE.

36. **Miscellaneous:**

a. Add a definition for ″Policy Information Base″ to clause 3.7.

b. Clause 5.2; Reword the first two sentences: ″The direct exchange of policy information is outside the scope of IDRP. Instead, IDRP communicates ...in its UPDATE PDUs which reflect the effects....″

Remove the words ″own local″ from the first sentence of the second paragraph.

c. Clause 5.4: Delete the last sentence under bullet ″a″.

d. Clause 6.1: Page 11, first table: 2 octet==> 2 octets.

Delete the word ″currently″ from last sentence describing the TYPE field.

e. Page 15, MULTI-EXIT_DISC: The words ″unsigned non-negative″ in the first sentence under MULTI-EXIT_DISC are redundant, and should be removed.

f. Page 15,16, Source and Destination QOS: The last sentence of the descriptions of the address fields speaks of ″the indicated QOS type″. References to clauses 7.11.13 and 7.11.14 would serve to clarify what this phrase actually means.

g. Clause 6.4: Provide a reference to clause 7.4.2 to clarify how sequence numbers are chosen for this PDU.

h. To avoid confusion with management notifications, it would be desirable to change the name NOTIFICATION PDU″ to ″IDRP Error PDU″.

i. The material in clause 7.1.1 belongs in clause 7.1, not in a subclause.

j. Clause 7.1.3.1: characters--> octets.

k. Clause 7.2.1, second item: For clarity, reword as follows: ″support inter-domain links to at least two different routeing domains...″

l. Clause 7.2.2, item ″a″: second clause of first sentence, change to ″must reside in that routeing domain.″

m. Clause 7.3, item ″a″ on page 20: change ″physical box″ to ″open system″.

n. Clause 7.4.3: For clarity, reword the first sentence of second paragraph: ″Acknowledgements can be carried in the headers of any type of BISPDU.″

In the third sentence, ″last correctly received″ is sloppy wording. For example, it doesn′t consider the problem of mis-sequencing.

In the middle of the second paragraph, the word ″prior″ should be defined, especially if the sequence numbers are permitted to wrap.

o. Clause 7.4.4, first and second sentences: ″must″--> ″shall″

p. Clause 7.4.5: Define ″left″ and ″right″.

Note that the third paragraph doesn′t apply to BISPDUs that don′t increment a sequence number.

q. Clause 7.5: Change the first sentence: ″The protocol described...on the underlying Network layer protocol to establish...between each pair of BISs.″

r. Clause 7.5.4: third item: delete ″then″, add ″state″ as the very last word.

s. Clause 7.6.2: first item: ″largest locally supported version number″ --> ″highest supported version number″.

first item in right hand column: ″different than″-->″different from that″

In the case of authentication failure, management should also be notified.

t. Clause 7.6.3, second item: NOTIFICATION PDU contains″--> ″IDRP Error PDU shall contain″

The note in the right hand column could be reworded for clarity, by deleting the current last sentence, and inserting the following as a new first sentence: ″It is permissible for an UPDATE PDU to contain neither the DIST_LIST_INCL nor the DIST_LIST_EXCL attributes.″

u. Clause 7.9: It would be preferable not to use zero as an authentication code. Any other value would be suitable.

Right-hand column, second dashed item: ″will define″--> ″specifies″

Last paragraph, last sentence: Change to ″Path attributes are summarized in Table 1; their encoding is described....″

v. Clause 7.10.2, item ″c″, page 28; Define ″equivalent″.

w. Clause 7.11.7: The text should state that the MULTI-EXIT_DISC attribute should be applied on a ″per QOS″ basis.

x. Clause 7.11.14, second paragraph: ″SOURCE SPECIFIC″-->″DESTINATION SPECIFIC″

y. Clause 7.12.8, second line at top of page 37: ″router″-->″BIS″

z. Clause 7.12.9, first paragraph: In last two sentences: ″will perform″-->″performs″, ″will be advertised″-->″is advertised″, ″will aggregate″-->″may aggregate″.

aa. Clause 7.12.10: In several places, change ″external inter-domain links″ to ″inter-domain links″.

ab. Clause 7.13.2: Text is needed to describe what is meant by ″replacement route″, and how it would be distinguished from a new route.

ac. Clause A.3.4.1, first paragraph: change ″appliesory,...″ to ″applies—mandatory,...″.