To: Tom Huber, Han Li, ITU-T Q9/SG15
From: Tony Jeffree, IEEE 802.1 WG Chair
Date: Thursday, 06 March 2014
Reference: COM 15 – LS 076 – E

Thank you for your liaison COM 15 – LS 076 – E, continuing our dialog on the subject of Multi Domain Segment network Protection. IEEE 802.1 Interworking Task Group assigned a subcommittee to formulate answers to the questions you present in LS 076, and the IEEE 802.1 Working Group as a whole has reviewed the answers and approved this reply.

Our objective has been to create a standard that applies to any kind of interface, including certainly the UNI, INNI and ENNI. It is not limited to operation by IEEE 802.1 bridges; it is intended to be applicable to routers or ITU-T Ethernet switching devices, as well.

1) Could DRNI cause segmentation of the networks it is interconnecting under failure conditions?

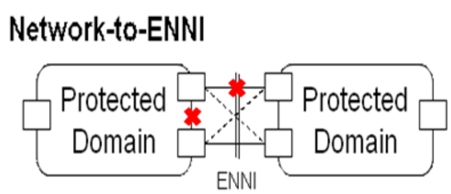

Figure 1

In this specific example, the DRNI would continue to pass all conversations between the two protected domains in the presence of one or both of the indicated errors, whether or not the additional cross-links are present.

More specifically:

1) If the upper link connecting the two domains fails (Aggregation Link), the associated upper nodes will use other available links to pass conversations between the two domains: the additional cross-links if present, or the link connecting the nodes in the same domain if not.
2) If the link connecting the two nodes in the left hand side protected domain fails (Intra-Portal Link – IPL), the node having the highest priority (a preconfigured value) will take over the conversation exchange with the protected domain at the right hand side.
3) If both links fail, the lower node at the left hand side will take over the conversation exchange to the protected domain at the right hand side, if no cross-links. With cross-links, either of the left hand nodes will take over the conversations, as described, below.

Note also that the allocation of particular conversations to physical links after the two failures does not depend upon the order in which the failures occur.

2) With respect to Figure 2 [not copied in this liaison]:

   a. Is DRNI's transfer time T_t 50ms or less? If yes, under which conditions?

The process, starting with the recognition of the failure, includes the generation, transmission, reception, and processing of certain protocol data units (PDUs) and also includes time to make adjustments to the data forwarding mechanisms in the various nodes in order to change the paths along which data frames are sent. DRNI does not mandate specific requirements for timing on these actions, but on the other hand rapid configuration and reconfiguration has been one of the main goals and objectives in the DRNI protocol design. In the event of changes in the physical connectivity, DRNI will quickly converge to a new configuration, typically on the order of milliseconds for link down events and 1 second or less for link up events.

The DRNI protocol, upon detecting a link failure, immediately updates the local conversation assignments and triggers transmission of the updated conversation assignments to the other systems in the DRNI, in order to check the consistency across the DRNI. These operations are done in parallel by the affected systems and correspondingly the restoration of all conversations across the DRNI is not expected to involve repeated PDU transmissions.

In theory, slow transmission rates, long wires, and poor implementation choices could result in a recovery time that exceeds 50 ms. In practice, bridge implementations have not found it difficult to perform almost identical actions in well under 50 ms in bridged networks with similar topologies.

2) With respect to Figure 2 [not copied in this liaison]:

   b. What is DRNI's recovery time (T_5) in the cases where the transfer time is 50ms or less?

One answer is that, between the time that the failure first occurs (before it has been detected) and the time that the data plane has been adjusted to the recovered state, inter-domain data frames for some flows will be discarded. Once the recovered state has been achieved, the worst-case time to deliver frames will be simply the worst-case propagation delay for the rerouted conversations. In Figure 1, this would clearly be the propagation time for the two hops from the upper node of the right Portal, to the lower node of that Portal, to the destination node in the left Portal.

There is a second answer to question 2b. In addition to the propagation delay, and if all of the following conditions apply:

a. A given conversation is passing into the network (domain) from one node of a Portal;
b. That node fails, so that the conversation must pass through the other node of the Portal;
c. The operator has chosen to configure the network (domain) so that the conversation in question is delivered, normally, only the failed node; and
d. The operator depends on automatic mechanisms, available in IEEE 802.1Q, to signal to the network that the conversation needs to be delivered to the remaining node after the failure;

Then there will be a delay in re-establishing the flow until that automatic process completes. That process is not bounded, but 100s of milliseconds would be typical. This, however, is a choice made by the operator to balance the bandwidth required to deliver data to both nodes vs. the response time to a failure. The operator can also choose to eliminate this delay, either at the expense of delivering the data to both nodes at all times, or by using a mechanism such as PBB-TE to control the data in the network.

3) Could operator commands such as those specified in ITU-T Recommendation G.8031 be supported with DRNI?

The primitives provided by DRNI support any combination of actions with regard to the assignment of conversations to physical links or to Gateways (the points at which a conversation can enter or exit a domain). Every conversation is configured with its own prioritized list of Gateways and physical links. The Portal Nodes adjust their forwarding tables to pass any given conversation between the highest-priority operational Gateway and the highest-priority operational physical link for that conversation. How the G.8031 commands would be translated into changes to these configured priority lists is not specified in IEEE 802.1AX-REV, but IEEE 802.1 believes that this translation is straightforward, and that ITU-T could define it for management stations, network nodes, or both.

4) What is DRNI's expected behavior when the Intra-Portal Link (IPL) fails in a Portal, i.e. when the link between the two Portal Systems fail?

In DRNI terms, the pair of nodes in the left-hand domain of your Figure 1 (between which there is a failure) constitute a two-node Portal, as do the pair of nodes in the right-hand domain. When a Portal is divided by the loss of the Intra-Portal Link (IPL), the Portal is dissolved until the IPL can be re-established. That is, instead of advertising their ports, via the Link Aggregation Control Protocol (LACP), as being able to form a single Aggregator, the left-hand Portal nodes revert to advertising themselves as belonging to two independent Key Groups. Upon seeing this, the nodes in the right-hand Portal select one of the left-hand nodes to partner with, and reject the other. Thus, both of the nodes in the divided Portal continue to function, and both attempt to become the carrier for all of the data, but only one of them is able to partner with the other Portal and only that one actually transmits and receives the data.

In the specific example in Figure 1, and in the absence of the cross-links, all traffic would flow through the lower of the two inter-domain links. If the cross-links are present, the selection of either the upper or the lower left-hand node to carry all of the traffic would depend on the details of the configuration of those nodes.

We hope that this liaison answers your questions, and look forward to further cooperation.

Regards,

Tony Jeffree
IEEE 802.1 Working Group Chair