Title: Liaison response to IETF regarding Proposed IETF BFD WG work on Ethernet LAG

Date: 15 March 2012

Location: Kona, Hawaii

From: IEEE 802.1

Contacts: Stephen Haddock, Chair, Interworking Task Group (shaddock@stanfordalumni.org)

Tony Jeffree, Chair, IEEE 802.1 (tony@jeffree.co.uk)

To: Stewart Bryant, IETF Routing Area Director <stbryant@cisco.com>

Dear Stewart

Thank you for your liaison letter. Closer cooperation will improve the output from both of our organizations. We recognize that your communication reports an early stage of the process, and we would appreciate being kept informed of IETF decisions and progress with respect to this proposed work item.

We understand and agree that relying on LACPDU exchange as a link failure detection mechanism does not result in acceptable response and recovery times. LACP was never intended to serve as a means for link failure detection, except as a last resort.

Since LACP was originally done in IEEE 802.3, the Ethernet "hardware" group, it is intended to depend upon the 802.3 physical layer error detection schemes to detect link failure. The Link Aggregation state machines do NOT wait for LACP to detect a failure; they react instantly to a hardware failure signal. We understand and agree with the desire to have frame-based link failure detection in addition to link-level hardware detection. We use Connectivity Fault Management (CFM, 802.1Q-2011 clauses 18-22) for monitoring the individual links for this purpose, but we understand that in environments already using BFD for monitoring the aggregation, it may be preferable to use BFD for monitoring the individual links.

CFM can operate below the Link Aggregation layer, on the individual physical links. CFM forms a layer in the RFC2863 Interface Stack. Built in to the CFM model is the idea that a link is not visible (the MAC_Operational signal is FALSE) until CFM has determined that the link is operational. MAC_Operational is, in essence, equivalent to ifOperStatus in the Interface MIB.

802.1AX Link Aggregation is currently under revision. There are two purely editorial changes anticipated that may be of interest to you:

1. The revised 802.1AX will make the relationship of LACP and MAC_Operational clear, which it is not in the current edition. This will make the fact that Link Aggregation responds quickly to link failures more obvious.

2. Certain protocols, 802.1AB Link Layer Discovery Protocol and 802.1AX Port Authentication, in particular, operate over physical links, even when those links are part of an Aggregation. The next revision of 802.1AX will clarify the fact that higher layers have access to individual links even after they have been aggregated.

We believe that these two clarifications are sufficient for CFM or any other frame-based link failure detection mechanism to interwork with Link Aggregation without having to modify Link Aggregation state machines.

In reading your draft, we feel that the suggested interaction between LACP and BFD could be greatly simplified. Our interpretation is that your intention is to somehow pause the LACP state machines during the bring-up phase, in order that BFD can verify the link before it is used by the aggregation. It is worth noting that LACP does provide an essential failure detection service during link bring-up. LACP ensures that a link is not only bi-directionally operational, but that both directions are connected to the same MAC (i.e., a fiber pair is not split across two ports) before accepting the link into an aggregation. Thus, there should be no need to delay bring-up of the link. There are at least two ways to interact with LACP so that there is no need to modify the LACP state machines:

a) Start BFD ahead of time, running on the physical link. LACP doesn't even start until BFD has determined the link is available. If BFD determines the link is not available, set ifOperStatus to DOWN (reset MAC_Operational) and the link is immediately withdrawn from the aggregation. In other words, use BFD exactly the same way CFM is currently used.
b) Have an inactive BFD layer in the interface stack below LACP. After LACP comes up, activate the BFD layer (start BFD exchanges) under LACP. Keep the time interval, between LACP indicating the aggregation is available and starting BFD, as short as you want to. You can observe the LACP status to start (or stop) the BFD operation, and use BFD failure to signal link failure to LACP via MAC_Operational/ifOperStatus.

We must point out Link Aggregation is a (lower) Layer 2 construct, not a Layer 3 construct. It is not uncommon to connect a router to a bridge via an aggregated link. In this case, it is not clear how one uses a Layer 3 protocol to support the aggregation, or how to achieve interoperability when two different protocols are used (i.e. BFD and CFM) for the same purpose for the two connected systems.

We would be very interested in continued dialog to ensure that your frame-based link detection interacts well with Link Aggregation, as well as on other subjects that will undoubtedly arise.

With best regards,

Tony Jeffree

IEEE 802.1 WG Chair