In reviewing the introductory clause of Clause 9 of P802.1AX-Rev/D0.3 I found significant problems with this mainly old text, possibly as a result of its incremental development and its derivation from presentations aiming at selling the big idea. Coming to it fresh I found that a high level view is mainly missing and that understanding many points depends on already knowing detail and terminology only introduced latter. Some goals are stated in terms of the detail of the solution, not in terms of what is to be achieved by the solution. I found it hard to determine what was actually being achieved because the points (ports in this case) where the consistent service is being maintained were not even named (they were simply said to be connected by an internal link to the (not externally visible) internal ports in the proposed model.

The degree of restatement and reorganization I believe necessary makes it very difficult to write a set of point changes that an editor would have a chance of reassembling into a coherent whole, and in any case I needed to write the intended result to debug the proposed changes.

The result follows. It is intended as a complete replacement for the introductory clauses, though I have not yet been able to complete the later subclauses. I have discovered quite a number of interesting points while preparing this text and it will have some effect on (or at least prompt further detail study of) the following detail clauses. However much of the following detail in Clause 9 should require only cosmetic changes (I hope).

I will (or possibly have by the time you read this) submit detailed point by point comments on what I think needed changing and why, though for the reasons described above I don’t think phrase by phrase substitution will fix the problems.

Mick

9. Distributed Resilient Network Interconnect

A Distributed Resilient Network Interconnect (DRNI) Portal, comprising two cooperating Portal Systems, can be used to terminate one end of a Link Aggregation Group (LAG). A system connected to the individual links at the other end uses the LAG, and the Link Aggregation Control Protocol (LACP), as if those links provide connectivity to a single Aggregation System. Portals can thus be deployed independently at either or both ends of any LAG to provide fault tolerance.

This clause provides an overview of the following:

a) DRNI goals (9.1)

b) DRNI operation (9.2)

c) The Intra-Portal Link (IPL) connecting the Portal Systems (9.3)

d) How DRNI meets its goals, protecting against link and system failures (9.4)

and specifies the operation of the following in detail:

e) The Distributed Relay functionality in each Portal System ()

f) The Distributed Relay Control Protocol (DRCP,)

Architectural concepts common to this and other IEEE 802.1 standards are used in this clause's specification of DRNI. The reader is encouraged to review Clause 7 of IEEE Std 802.1AC-2016.

The models of operation in this clause provide a basis for specifying the externally observable behavior of the operation, and are not intended to place additional constraints on implementations; these can adopt any internal model of operation compatible with the externally observable behavior specified. Conformance of equipment to this standard is purely with respect to observable protocol.

9.1 DRNI goals

Link Aggregation (as described in 6.1, 6.2, and 6.3) can aggregate the links that connect two Aggregation Systems. One or more System Functions in each of the Aggregation Systems can use the resulting LAG as if was a single link, but the LAG is resilient (i.e. provides fault tolerance) to the failure, addition, or removal of individual links in the LAG as long as it provides sufficient bandwidth for the conversations it carries.

DRNI is resilient to system failure while maintaining support for link resilience. Specifically DRNI meets the basic Link Aggregation goals described in 6.1.1, and provides the following:

a) **System redundancy**—An Aggregation System can be replaced by two cooperating DRNI Portal Systems that share access to a DRNI LAG, each attaching to some of the links in that LAG.

b) **System load sharing**—System functions that use the service provided by a DRNI LAG can be distributed between the Portal Systems, with conversations directed to one system or the other.

NOTE—A DRNI LAG provides the same Internal Sublayer Service (ISS, IEEE Std 802.1AC) as an individual link. Multiple system functions (and potentially different instances of a given function) can make use of a port providing that service using standardized protocol discrimination, addressing, and multiplexing functions.

c) **System and link conversation distribution**—Conversation-sensitive Collection and Distribution (CSCD, Clause 8) parameters apply independently to the allocation of conversations to links and to their allocation to each Portal System's system function instance. An Intra-Portal Link (IPL) allows a conversation associated with a system function instance in one Portal System to be carried by a DRNI LAG link attached to the other Portal System.

d) **Interoperability**—A DRNI Portal can support a LAG whose links provide connectivity to a single Aggregation System that meets the conformance requirements specified in 5.3 of this standard or specified in prior editions of this standard.

e) **Localized fault recovery**—The effects of DRNI link and system failure can be confined to the DRNI LAG and its attached Portal Systems which use DRCP to provide rapid recovery (see 9.4).

DRNI operates independently of the (possibly distributed) system functionality that it supports in each Portal System. It does not define or constrain ways of distributing arbitrary system functionality so that two systems mimic the behavior of one, a difficult task. DRNI can be used to attach a Portal providing end-station functionality to a network, with possible distributed system functions including support for VRRP (Virtual Router Redundancy Protocol, RFC xxxx) and File Server applications. Their specification is outside the scope of this Standard.

An explicit goal of DNRI is to support the connection of one network to another, where services (each comprising one or more potential conversations) are provisioned across one network, the DRNI LAG, and the second network. Each Portal System is typically (but not necessarily) a Provider Bridge, as specified in IEEE Std 802.1Q. The use of a DNRI LAG to connect the two networks provides the following:

f) **Administrative independence—**The connected networks can be administered separately, and can use different routing and fault recovery protocols.

g) **Distribution independence—**The frame distribution algorithm used to select the Portal System responsible for relaying frames for a given conversation to and from the attached network can differ from the algorithm used to select a link in the DRNI LAG, and can differ from the selection algorithm used by the Portal attached to the other end of the LAG to relay frames to and from its own network. Each algorithm’s assignment of frames to conversations can also differ.

h) **Internet-network fault isolation**—Failure or recovery of a link or node in one network can be hidden from the other network.

i) **DRNI fault isolation—**The failure or recovery of a link in a DRNI LAG can be hidden from both network’s control protocols. The failure of a Portal System attached to one network can be hidden from the other network.

Fault and fault recovery scenarios are described in 9.4.

9.2 DRNI Overview

When a LAG terminates in a single Aggregation System, frames transmitted and received on the LAG are accepted from and delivered to a single Aggregator Client attached to a single Aggregator Port, as shown in Figure 6-2 and described in 6.1, 6.2, and 6.3. When a LAG terminates in a DRNI Portal, frames can be accepted from and delivered to a single Client attached to a single DRNI Aggregator Port in one of the Portal Systems, but can also use Conversation-sensitive Collection and Distribution (CSCD) to support two DRNI Aggregator Ports, one in each Portal System. The CSCD parameters that allocate conversations to DRNI Aggregator Ports are independent of those that allocate conversations to individual links. An Intra-Portal Link (IPL) conveys frames passing between a DRNI Aggregator Port in one Portal System and an individual link in the other. If sufficient bandwidth is available, DRNI Aggregator Port conversation allocation can be independent of the failure of individual links or of a Portal System or IPL at the other end of the LAG.

Twin DRNI Aggregator Ports, one in each Portal System, can support a distributed system function that uses a single LAG.

Figure 9-1 shows the structure of an example DRNI Portal. Portal Systems A and B each have five external network connections (MACs a through b, and p through t, respectively). Their immediate client is the Link Aggregation sublayer (operating as specified in Clause 6) in each system. This supports an Aggregator Port that provides the Internal Sublayer Service (ISS) to the Portal System’s Distributed Relay function at an Aggregator Port, aggregating the links that system uses to participate in the DRNI LAG. In the figure these links are attached to MACs b and d in System A and MACs r and s in System B.

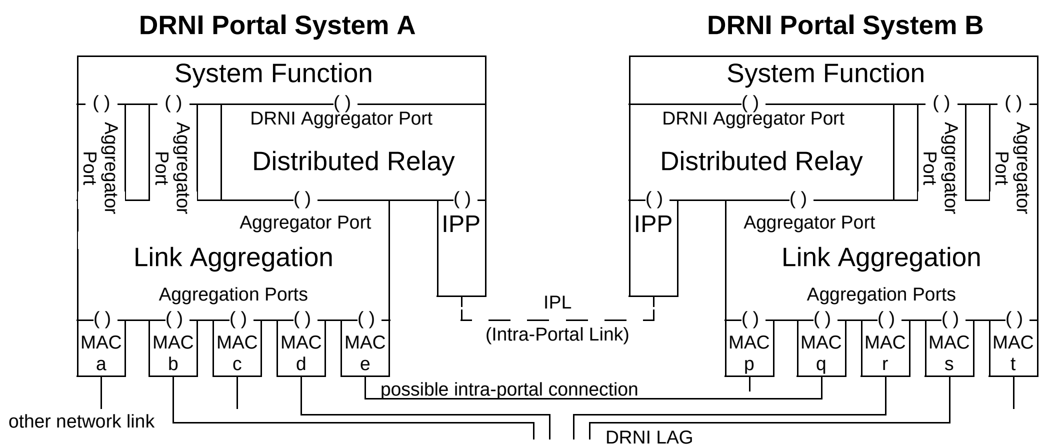

Figure 9-1—A DRNI Portal

The Distributed Relay uses its Aggregator Port and an Intra-Portal Port (IPP) to support a DRNI Aggregator Port that allows its system’s System Function to transmit and receive frames to and from any of the DRNI LAG links. Frames for conversations carried by links attached to that Portal System are passed to and accepted from the Distributed Relay’s Aggregator Port. Frames for conversations carried by links attached to the other Portal System are passed to and from the Distributed Relay’s IPP, and are relayed by that other system’s Distributed Relay to and from its Link Aggregation sublayer’s Aggregator Port. A Portal System’s Distributed Relay function thus receives both ‘Down’ frames (transmitted by the other Portal System’s System Function for DRNI LAG transmission one of its links) and ‘Up’ frames (received by one of its DRNI LAG links for the System Function in the other Portal System). See Figure 9-2 (frames to and from A’s System Function on the left of the figure, and to and from B’s on the right).

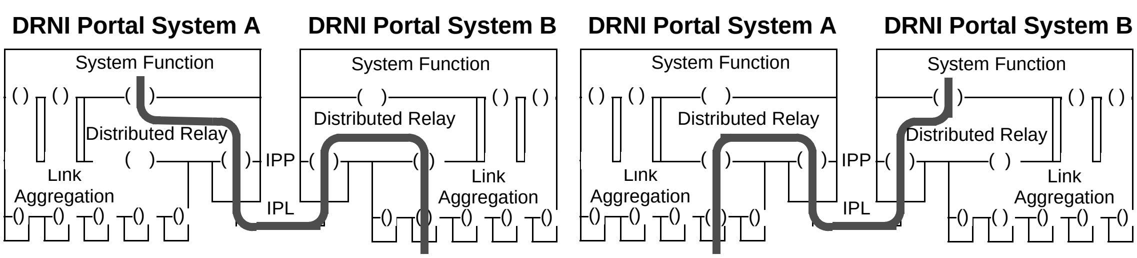

Figure 9-2—Transmission and reception via the IPL

The IPL is not necessarily a dedicated link between the Portal Systems (9.3). IPL frames do not carry additional information marking them as Up or Down or associating them with a DRNI Aggregator Port or an Aggregation Port, thus avoiding the need to support additional tagging and detagging operations and frame formats. Instead, DRCP synchronizes each system’s Distributed Relay assignment of each conversation to a DRNI Aggregator Port (its own or that of its peer Distributed Relay in the other system). If a system has changed that assignment since synchronization was last achieved, its Distributed Relay discards frames for that conversation that would otherwise be transmitted to or received from the IPL. Discarding these frames avoids accidentally returning an Up frame received from a DRNI LAG link attached to one Portal System to another link in the same DRNI LAG in the other Portal System, or forwarding a Down frame from one DRNI Aggregator Port to the other.

The Link Aggregation sublayer in each Portal System can support additional Aggregator Ports that allow the System Function to use other network links (individually or aggregated) directly (MACs a, c, e, p, q, and t in Figure 9-1). One or more of these (attached to MAC e and q in the figure) can be used to link the Portal Systems, supporting the IPL (see 9.3) and any System Function co-ordination required, though these communication requirements can also be supported by network relay of frames transmitted on other links.

A Portal System can include more than one Distributed Relay function, each with its own Aggregator Port, DRNI Aggregator Port, and IPP (providing IPL connectivity to a Distributed Relay function in its companion Portal System). Once IPL connectivity has been established, the LACPDUs transmitted on both Portal System’s Aggregation Ports whose operational Aggregation Key value matches that of a Distributed Relay’s Aggregator Port convey the same Actor System Identifier (comprising the Actor System and Actor System Priority fields as specified in 6.4.2). A single Aggregation System can thus aggregate links connecting to both Portal Systems. DRNI applications and deployments can constrain the assignment of administrative and operational Keys as specified below (9.4)

9.3 IPL Support

An IPL is a logical link that supports the ISS at the IPPs of paired Distributed Relay functions, one in each Portal System. DRCP is used to identify and verify IPL connectivity, guarding against physical or logical mis-wiring or administrative error. A Portal System can pair with at most one companion Portal System: the use of multiple Distributed Relay functions to pair with multiple Portal Systems is not supported.

In principle IPL connectivity can be provided by any method capable of supporting the ISS and a separate connectivity association (7.8 of IEEE Std 802.1AC-2016) for each pair of IPPs. Annex H discusses options. Selection and deployment of a particular method should take into account the potential for service disruption due to failure or reconfiguration of service supporting components. A Portal System that conforms to this standard shall be capable of using a dedicated link implementing an IEEE 802 media access control method, (e.g. as specified by IEEE Std 802.3).

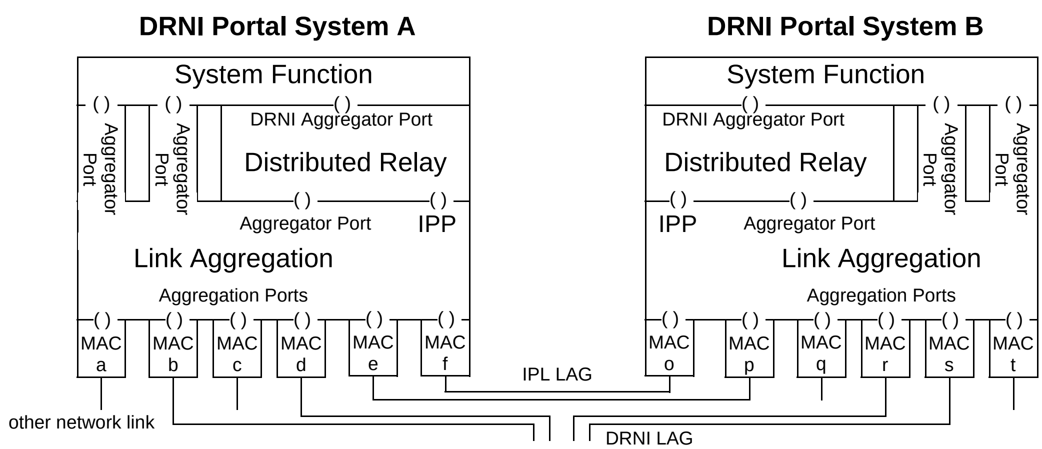

Figure 9-3—DRNI Portal with dedicated IPL LAG

Figure 9-3 shows a DRNI Portal that uses a dedicated, link aggregated, IPL. DRCP extensions to LACPDUs identify links (e-p and f-o in the figure) as potential IPL links, facilitating their attachment to the IPPs (the Aggregator Ports for this LAG).

9.4 Using DRNI

A DRNI LAG can be used to attach a pair of Portal Systems providing distributing end-station functionality to a network, as shown in figure

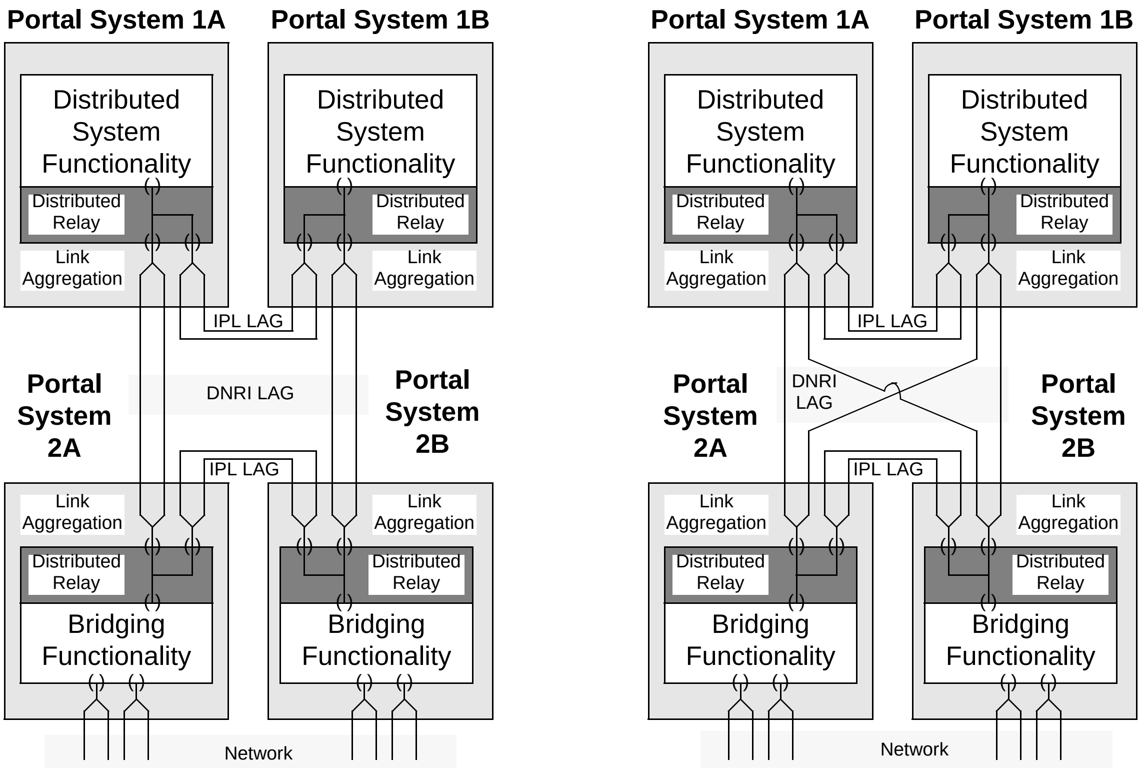

Figure 9-4—DRNI network attach

<<tbs, a discussion of the above figure showing how connectivity to a DRNI Aggregator Port for each conversation assigned to that port is maintained following link or system failure interrupting existing connectivity (or leaving inadequate bandwidth on a particular path)>>

Similarly, a DRNI LAG can be used to connect two networks....<<work in progress, stope reading here>>

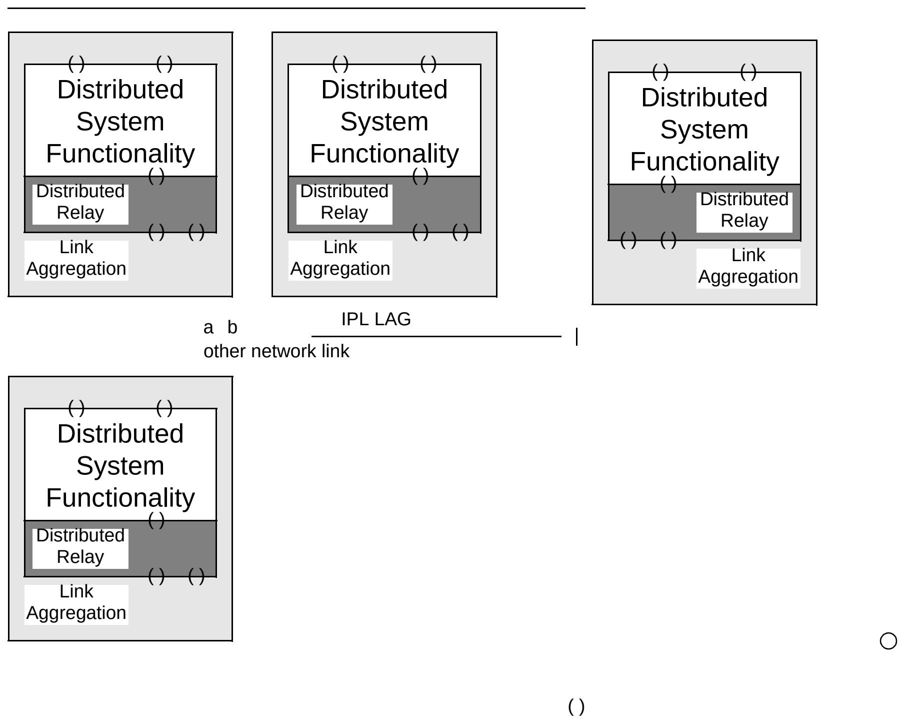
Distributed System Functionality
Distributed Relay
Link Aggregation
Distributed System Functionality
Distributed Relay
Link Aggregation
Distributed System Functionality
Distributed Relay
Link Aggregation
IPL LAG
a b
other network link
Distributed System Functionality
Distributed Relay
Link Aggregation

Figure 9-5—D