

BS ISO/IEC/IEEE 8802-1AX:2016



BSI Standards Publication

Information technology — Telecommunications and information exchange between systems — Local and metropolitan area networks — Specific requirements

Part 1AX: Link Aggregation

bsi.

...making excellence a habit.
AVA Documents™
Document Sharing Hub

National foreword

This British Standard is the UK implementation of ISO/IEC/IEEE 8802-1AX:2016.

The UK participation in its preparation was entrusted to Technical Committee IST/6, Data communications.

A list of organizations represented on this committee can be obtained on request to its secretary.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

© The British Standards Institution 2016.
Published by BSI Standards Limited 2016

ISBN 978 0 580 90843 9

ICS 35.240.80

Compliance with a British Standard cannot confer immunity from legal obligations.

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 31 January 2016.

Amendments/corrigenda issued since publication

Date	Text affected
------	---------------

INTERNATIONAL
STANDARD

**ISO/IEC/
IEEE
8802-1AX**

First edition
2016-01-15

**Information technology —
Telecommunications and information
exchange between systems — Local and
metropolitan area networks — Specific
requirements**

**Part 1AX:
Link Aggregation**

*Technologies de l'information — Télécommunications et échange
d'information entre systèmes — Réseaux locaux et métropolitains —
Exigences spécifiques —*

Partie 1AX: Agrégation de lien



Reference number
ISO/IEC/IEEE 8802-1AX:2016(E)

AVA Documents
Document Sharing Hub
© IEEE 2014



COPYRIGHT PROTECTED DOCUMENT

© IEEE 2014

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from ISO, IEC or IEEE at the respective address below.

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org
Published in Switzerland

Institute of Electrical and Electronics Engineers, Inc.
3 Park Avenue, New York
NY 10016-5997, USA
E-mail stds.ipr@ieee.org
Web www.ieee.org

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

IEEE Standards documents are developed within the IEEE Societies and the Standards Coordinating Committees of the IEEE Standards Association (IEEE-SA) Standards Board. The IEEE develops its standards through a consensus development process, approved by the American National Standards Institute, which brings together volunteers representing varied viewpoints and interests to achieve the final product. Volunteers are not necessarily members of the Institute and serve without compensation. While the IEEE administers the process and establishes rules to promote fairness in the consensus development process, the IEEE does not independently evaluate, test, or verify the accuracy of any of the information contained in its standards.

The main task of ISO/IEC JTC 1 is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is called to the possibility that implementation of this standard may require the use of subject matter covered by patent rights. By publication of this standard, no position is taken with respect to the existence or validity of any patent rights in connection therewith. ISO/IEEE is not responsible for identifying essential patents or patent claims for which a license may be required, for conducting inquiries into the legal validity or scope of patents or patent claims or determining whether any licensing terms or conditions provided in connection with submission of a Letter of Assurance or a Patent Statement and Licensing Declaration Form, if any, or in any licensing agreements are reasonable or non-discriminatory. Users of this standard are expressly advised that determination of the validity of any patent rights, and the risk of infringement of such rights, is entirely their own responsibility. Further information may be obtained from ISO or the IEEE Standards Association.

ISO/IEC/IEEE 8802-1AX was prepared by the LAN/MAN of the IEEE Computer Society (as IEEE 8802-1AX-2014). It was adopted by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 6, *Telecommunications and information exchange between systems*, in parallel with its approval by the ISO/IEC national bodies, under the “fast-track procedure” defined in the Partner Standards Development Organization cooperation agreement between ISO and IEEE. IEEE is responsible for the maintenance of this document with participation and input from ISO/IEC national bodies.

IEEE Standard for Local and metropolitan area networks— Link Aggregation

IEEE Computer Society

Sponsored by the
LAN/MAN Standards Committee

IEEE
3 Park Avenue
New York, NY 10016-5997
USA

IEEE Std 802.1AX™-2014
(Revision of
IEEE Std 802.1AX-2008)

IEEE Std 802.1AX™-2014

(Revision of
IEEE Std 802.1AX-2008)

IEEE Standard for Local and metropolitan area networks— Link Aggregation

Sponsor
LAN/MAN Standards Committee
of the
IEEE Computer Society

Approved 10 December 2014
IEEE SA-Standards Board

Abstract: MAC-independent Link Aggregation capability and general information relevant to specific MAC types are defined in this standard. Link Aggregation allows parallel full-duplex point-to-point links to be used as if they were a single link and also supports the use of multiple links as a resilient load sharing interconnect between multiple nodes in two separately administered networks.

Keywords: Aggregated Link, Aggregator, Distributed Resilient Network Interconnect, DRNI, IEEE 802®, IEEE 802.1AX™, interconnect, Link Aggregation, Link Aggregation Group, local area network, management, Network-Network Interface, NNI

The Institute of Electrical and Electronics Engineers, Inc.
3 Park Avenue, New York, NY 10016-5997, USA

Copyright © 2014 by The Institute of Electrical and Electronics Engineers, Inc.
All rights reserved. Published 30 December 2014. Printed in the United States of America.

IEEE and 802 are registered trademarks in the U.S. Patent & Trademark Office, owned by The Institute of Electrical and Electronics Engineers, Incorporated.

Print: ISBN 978-0-7381-9448-6 STD20052
PDF: ISBN 978-0-7381-9449-3 STDPD20052

IEEE prohibits discrimination, harassment, and bullying.

For more information, visit <http://www.ieee.org/web/aboutus/whatis/policies/p9-26.html>.

No part of this publication may be reproduced in any form, in an electronic retrieval system or otherwise, without the prior written permission of the publisher.

Important Notices and Disclaimers Concerning IEEE Standards Documents

IEEE documents are made available for use subject to important notices and legal disclaimers. These notices and disclaimers, or a reference to this page, appear in all standards and may be found under the heading “Important Notice” or “Important Notices and Disclaimers Concerning IEEE Standards Documents.”

Notice and Disclaimer of Liability Concerning the Use of IEEE Standards Documents

IEEE Standards documents (standards, recommended practices, and guides), both full-use and trial-use, are developed within IEEE Societies and the Standards Coordinating Committees of the IEEE Standards Association (“IEEE-SA”) Standards Board. IEEE (“the Institute”) develops its standards through a consensus development process, approved by the American National Standards Institute (“ANSI”), which brings together volunteers representing varied viewpoints and interests to achieve the final product. Volunteers are not necessarily members of the Institute and participate without compensation from IEEE. While IEEE administers the process and establishes rules to promote fairness in the consensus development process, IEEE does not independently evaluate, test, or verify the accuracy of any of the information or the soundness of any judgments contained in its standards.

IEEE does not warrant or represent the accuracy or content of the material contained in its standards, and expressly disclaims all warranties (express, implied and statutory) not included in this or any other document relating to the standard, including, but not limited to, the warranties of: merchantability; fitness for a particular purpose; non-infringement; and quality, accuracy, effectiveness, currency, or completeness of material. In addition, IEEE disclaims any and all conditions relating to: results; and workmanlike effort. IEEE standards documents are supplied “AS IS” and “WITH ALL FAULTS.”

Use of an IEEE standard is wholly voluntary. The existence of an IEEE standard does not imply that there are no other ways to produce, test, measure, purchase, market, or provide other goods and services related to the scope of the IEEE standard. Furthermore, the viewpoint expressed at the time a standard is approved and issued is subject to change brought about through developments in the state of the art and comments received from users of the standard.

In publishing and making its standards available, IEEE is not suggesting or rendering professional or other services for, or on behalf of, any person or entity nor is IEEE undertaking to perform any duty owed by any other person or entity to another. Any person utilizing any IEEE Standards document, should rely upon his or her own independent judgment in the exercise of reasonable care in any given circumstances or, as appropriate, seek the advice of a competent professional in determining the appropriateness of a given IEEE standard.

IN NO EVENT SHALL IEEE BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE PUBLICATION, USE OF, OR RELIANCE UPON ANY STANDARD, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

Translations

The IEEE consensus development process involves the review of documents in English only. In the event that an IEEE standard is translated, only the English version published by IEEE should be considered the approved IEEE standard.

Official statements

A statement, written or oral, that is not processed in accordance with the IEEE-SA Standards Board Operations Manual shall not be considered or inferred to be the official position of IEEE or any of its committees and shall not be considered to be, or be relied upon as, a formal position of IEEE. At lectures, symposia, seminars, or educational courses, an individual presenting information on IEEE standards shall make it clear that his or her views should be considered the personal views of that individual rather than the formal position of IEEE.

Comments on standards

Comments for revision of IEEE Standards documents are welcome from any interested party, regardless of membership affiliation with IEEE. However, IEEE does not provide consulting information or advice pertaining to IEEE Standards documents. Suggestions for changes in documents should be in the form of a proposed change of text, together with appropriate supporting comments. Since IEEE standards represent a consensus of concerned interests, it is important that any responses to comments and questions also receive the concurrence of a balance of interests. For this reason, IEEE and the members of its societies and Standards Coordinating Committees are not able to provide an instant response to comments or questions except in those cases where the matter has previously been addressed. For the same reason, IEEE does not respond to interpretation requests. Any person who would like to participate in revisions to an IEEE standard is welcome to join the relevant IEEE working group.

Comments on standards should be submitted to the following address:

Secretary, IEEE-SA Standards Board
445 Hoes Lane
Piscataway, NJ 08854 USA

Laws and regulations

Users of IEEE Standards documents should consult all applicable laws and regulations. Compliance with the provisions of any IEEE Standards document does not imply compliance to any applicable regulatory requirements. Implementers of the standard are responsible for observing or referring to the applicable regulatory requirements. IEEE does not, by the publication of its standards, intend to urge action that is not in compliance with applicable laws, and these documents may not be construed as doing so.

Copyrights

IEEE draft and approved standards are copyrighted by IEEE under U.S. and international copyright laws. They are made available by IEEE and are adopted for a wide variety of both public and private uses. These include both use, by reference, in laws and regulations, and use in private self-regulation, standardization, and the promotion of engineering practices and methods. By making these documents available for use and adoption by public authorities and private users, IEEE does not waive any rights in copyright to the documents.

Photocopies

Subject to payment of the appropriate fee, IEEE will grant users a limited, non-exclusive license to photocopy portions of any individual standard for company or organizational internal use or individual, non-commercial use only. To arrange for payment of licensing fees, please contact Copyright Clearance Center, Customer Service, 222 Rosewood Drive, Danvers, MA 01923 USA; +1 978 750 8400. Permission to photocopy portions of any individual standard for educational classroom use can also be obtained through the Copyright Clearance Center.

Updating of IEEE Standards documents

Users of IEEE Standards documents should be aware that these documents may be superseded at any time by the issuance of new editions or may be amended from time to time through the issuance of amendments, corrigenda, or errata. An official IEEE document at any point in time consists of the current edition of the document together with any amendments, corrigenda, or errata then in effect.

Every IEEE standard is subjected to review at least every ten years. When a document is more than ten years old and has not undergone a revision process, it is reasonable to conclude that its contents, although still of some value, do not wholly reflect the present state of the art. Users are cautioned to check to determine that they have the latest edition of any IEEE standard.

In order to determine whether a given document is the current edition and whether it has been amended through the issuance of amendments, corrigenda, or errata, visit the IEEE-SA Website at <http://ieeexplore.ieee.org/expel/standards.jsp> or contact IEEE at the address listed previously. For more information about the IEEE-SA or IOWA's standards development process, visit the IEEE-SA Website at <http://standards.ieee.org>.

Errata

Errata, if any, for all IEEE standards can be accessed on the IEEE-SA Website at the following URL: <http://standards.ieee.org/findstds/errata/index.html>. Users are encouraged to check this URL for errata periodically.

Patents

Attention is called to the possibility that implementation of this standard may require use of subject matter covered by patent rights. By publication of this standard, no position is taken by the IEEE with respect to the existence or validity of any patent rights in connection therewith. If a patent holder or patent applicant has filed a statement of assurance via an Accepted Letter of Assurance, then the statement is listed on the IEEE-SA Website at <http://standards.ieee.org/about/sasb/patcom/patents.html>. Letters of Assurance may indicate whether the Submitter is willing or unwilling to grant licenses under patent rights without compensation or under reasonable rates, with reasonable terms and conditions that are demonstrably free of any unfair discrimination to applicants desiring to obtain such licenses.

Essential Patent Claims may exist for which a Letter of Assurance has not been received. The IEEE is not responsible for identifying Essential Patent Claims for which a license may be required, for conducting inquiries into the legal validity or scope of Patents Claims, or determining whether any licensing terms or conditions provided in connection with submission of a Letter of Assurance, if any, or in any licensing agreements are reasonable or non-discriminatory. Users of this standard are expressly advised that determination of the validity of any patent rights, and the risk of infringement of such rights, is entirely their own responsibility. Further information may be obtained from the IEEE Standards Association.

Participants

The following individuals were officers and members of the Higher Layer LAN Protocols Working Group at the beginning of the Working Group ballot. Individuals may have not voted, voted for approval, disapproval, or abstained on this standard.

Glenn Parsons, *Working Group Chair*
John Messenger, *Working Group Vice Chair*
Stephen Haddock, *Chair, Interworking Task Group*
Michael Seaman, *Chair, Security Task Group*
Michael Johas Teener, *Chair, Time Sensitive Networking Task Group*
Pat Thaler, *Chair, Data Center Bridging Task Group*
Maximilian Riegel, *Chair, OmniRAN Task Group*
Eric Gray, *Recording Secretary*

Ting Ao
Christian Boiger
Paul Bottorff
David Chen
Feng Chen
Weiyang Cheng
Diego Crupnicoff
Rodney Cummings
Patrick Diamond
Aboubacar Kader Diarra
Janos Farkas
Norman Finn
Geoffrey Garner
Anoop Ghanwani
Mark Gravel
Craig Gunther

Hitosh Hayakawa
Jeremy Hitt
Rahil Hussain
Anthony Jeffree
Peter Jones
Hal Keen
Marcel Kiessling
Yongbum Kim
Philippe Klein
Jouni Korhonen
Jeff Lynch
Ben Mack-Crane
Christophe Mangin
James McIntosh
Eric Multanen
Donald Pannell

Karen Randall
Dan Romascanu
Jessy V. Rouyer
Panagiotis Saltsidis
Behcet Sarikaya
Daniel Sexton
Johannes Specht
Kevin B. Stanton
Wilfried Steiner
Vahid Tabatabaee
Jeremy Touve
Karl Weber
Yuehua Wei
Brian Weis
Jordon Woods
Juan-Carlos Zuniga

The following members of the individual balloting committee voted on this standard. Balloters may have voted for approval, disapproval, or abstention.

Thomas Alexander	James Innis	Michael Newman
Richard Alfvin	Osamu Ishida	Nick S.A. Nikjoo
Butch Anton	Akio Iso	Paul Nikolich
Jacob Ben Ary	Atsushi Ito	Satoshi Obara
Nancy Bravin	Raj Jain	Maximilian Riegel
William Byrd	Anthony Jeffree	Dan Romascanu
Juan Carreon	Michael Johas Teener	Jessy V. Rouyer
Keith Chow	Peter Jones	Panagiotis Saltsidis
Charles Cook	Shinkyō Kaku	Peter Saunderson
Patrick Diamond	Piotr Karocki	Michael Seaman
Yezid Donoso	Stuart Kerry	Kapil Sood
Sourav Dutta	Yongbum Kim	Thomas Starai
Donald Eastlake, 3rd	Bruce Kraemer	Rene Struik
Richard Edgar	Geoff Ladwig	Walter Struppler
Andrew Fieldsend	Mark Laubach	William Taylor
Yukihiro Fujimoto	John Lemon	Dmitri Varsanofiev
Devon Gayle	Ru Lin	Prabodh Varshney
Anoop Ghanwani	Elvis Maculuba	George Vlantis
Randall Groves	Roger Marks	Hung-Yu Wei
Chris Guy	Jeffery Masters	Andreas Wolf
Stephen Haddock	Brett McClellan	Chun Yu
Werner Hoelzl	Jonathon McLendon	Charles Wong
Rita Horner	Richard Mellitz	Michael Wright
Victor Hou	John Messenger	Oren Yuen
Noriyuki Ikeuchi	Charles Moorwood	Daidi Zhong
	Jose Morales	

When the IEEE-SA Standards Board approved this standard on 10 December 2014, it had the following membership:

John Kulick, *Chair*
Jon Walter Rosdahl, *Vice Chair*
Richard H. Hulett, *Past Chair*
Konstantinos Karachalios, *Secretary*

Peter Balma
Farooq Bari
Ted Burse
Clint Chaplain
Stephen Dukes
Jean-Phillippe Faure
Gary Hoffman

Michael Janezic
Jeffrey Katz
Joseph L. Koepfinger*
David J. Law
Hung Ling
Oleg Logvinov
T. W. Olsen
Glenn Parsons

Ron Peterson
Adrian Stephens
Peter Sutherland
Yatin Trivedi
Phil Winston
Don Wright
Yu Yuan

*Member Emeritus

Also included are the following nonvoting IEEE-SA Standards Board liaisons:

Richard DeBlasio, *DOE Representative*
Michael Janezic, *NIST Representative*

Michelle Turner
IEEE-SA Content Publishing

Kathryn Bennett
IEEE-SA Technical Community Programs

Introduction

This introduction is not part of IEEE Std 801.AX™-2014, IEEE Standard for Local and metropolitan area networks—Link Aggregation.

This standard contains state-of-the-art material. The area covered by this standard is undergoing evolution. Revisions are anticipated within the next few years to clarify existing material, to correct possible errors, and to incorporate new related material. Information on the current revision state of this and other IEEE 802® standards can be obtained from:

Secretary, IEEE-SA Standards Board
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
USA

Contents

1.	Overview.....	1
1.1	Scope.....	1
1.2	Purpose.....	1
1.3	State diagram conventions	2
2.	Normative references	3
3.	Definitions	4
4.	Acronyms and abbreviations	7
5.	Conformance.....	8
5.1	Requirements terminology.....	8
5.2	Protocol Implementation Conformance Statement.....	8
5.3	Link Aggregation requirements	8
5.3.1	Link Aggregation options	9
5.4	Distributed Resilient Network Interconnect requirements	9
5.4.1	Distribution Resilient Network Interconnect options	9
6.	Link Aggregation.....	11
6.1	Overview.....	11
6.1.1	Goals and objectives	11
6.1.2	Positioning of Link Aggregation within the IEEE 802 architecture.....	12
6.1.3	LLDP Parser/Multiplexer	13
6.1.3.1	LLDP Parser state diagram	14
6.1.3.1.1	LLDP Parser Function	14
6.1.3.1.2	Constants	14
6.1.3.1.3	Variables.....	14
6.1.3.1.4	State diagram	14
6.2	Link Aggregation operation.....	15
6.2.1	Principles of Link Aggregation.....	15
6.2.2	Service interfaces.....	16
6.2.3	Frame Collector	17
6.2.3.1	Frame Collector state diagram.....	17
6.2.3.1.1	Constants	17
6.2.3.1.2	Variables.....	17
6.2.3.1.3	Messages.....	17
6.2.3.1.4	State diagram	17
6.2.4	Frame Distributor.....	18
6.2.4.1	Frame Distributor state diagram	19
6.2.4.1.1	Variables.....	19
6.2.4.1.2	Messages.....	19
6.2.4.1.3	State diagram	19
6.2.5	Marker Generator/Receiver (optional).....	19
6.2.6	Marker Responder.....	20
6.2.7	Protocol Parser/Multiplexer	20
6.2.7.1	Protocol Parser state diagram	20
6.2.7.1.1	Functions	20

	6.2.7.1.2	Variables	21
	6.2.7.1.3	Messages.....	21
	6.2.7.1.4	State diagram	21
6.2.8	Aggregator Parser/Multiplexer		22
6.2.8.1	Aggregator Parser state diagram.....		22
	6.2.8.1.1	Constants	22
	6.2.8.1.2	Variables.....	22
	6.2.8.1.3	Messages.....	23
	6.2.8.1.4	State Diagram	23
6.2.9	Aggregator		24
6.2.10	Control Parser/Multiplexer		24
6.2.10.1	Control Parser state diagram.....		24
	6.2.10.1.1	Control Parser Function.....	24
	6.2.10.1.2	Constants	24
	6.2.10.1.3	Variables.....	25
6.2.11	Addressing		25
	6.2.11.1	Source address (SA).....	25
	6.2.11.2	Destination address.....	25
6.3	Link Aggregation Control.....		26
6.3.1	Characteristics of Link Aggregation Control.....		27
6.3.2	System identification		28
6.3.3	Aggregator identification.....		28
6.3.4	Port identification		28
6.3.5	Capability identification		29
6.3.6	Link Aggregation Group identification		30
	6.3.6.1	Construction of the Link Aggregation Group Identifier	30
	6.3.6.2	Representation of the Link Aggregation Group Identifier.....	31
6.3.7	Selecting a Link Aggregation Group		31
6.3.8	Agreeing on a Link Aggregation Group		32
6.3.9	Attaching a link to an Aggregator.....		32
6.3.10	Signaling readiness to transfer user data.....		32
6.3.11	Enabling the Frame Collector and Frame Distributor		33
6.3.12	MAC_Operational status		33
6.3.13	Monitoring the membership of a Link Aggregation Group.....		33
6.3.14	Detaching a link from an Aggregator		34
6.3.15	Configuration and administrative control of Link Aggregation		34
6.3.16	Link Aggregation Control state information		34
6.4	Link Aggregation Control Protocol		35
6.4.1	LACP design elements.....		35
6.4.2	LACPDU structure and encoding		35
	6.4.2.1	Transmission and representation of octets.....	35
	6.4.2.2	Encapsulation of LACPDU in frames	36
	6.4.2.3	LACPDU structure	36
	6.4.2.4	Version 2 TLVs	40
	6.4.2.4.1	Port Algorithm TLV	40
	6.4.2.4.2	Port Conversation ID Digest TLV.....	41
	6.4.2.4.3	Port Conversation Mask TLVs.....	41
	6.4.2.4.4	Port Conversation Service Mapping TLV	44
6.4.3	LACP state machine overview		44
6.4.4	Constants.....		46
6.4.5	Variables associated with the System.....		46
6.4.6	Variables associated with each Aggregator		47
6.4.7	Variables associated with each Aggregation Port.....		48
6.4.8	Variables used for managing the operation of the state machines.....		50

6.4.9	Functions.....	52
6.4.10	Timers	54
6.4.11	Messages	54
6.4.12	Receive machine	54
6.4.13	Periodic Transmission machine	56
6.4.14	Selection Logic	57
6.4.14.1	Selection Logic—Requirements	58
6.4.14.2	Selection Logic—Recommended default operation	59
6.4.15	Mux machine	60
6.4.16	Transmit machine	64
6.4.17	Churn Detection machines.....	64
6.4.18	Long LACPDU machine	65
6.5	Marker protocol	67
6.5.1	Introduction.....	67
6.5.2	Sequence of operations	67
6.5.3	Marker and Marker Response PDU structure and encoding	68
6.5.3.1	Transmission and representation of octets.....	68
6.5.3.2	Encapsulation of Marker and Marker Response PDU in frames.....	68
6.5.3.3	Marker and Marker Response PDU structure.....	68
6.5.4	Protocol definition	70
6.5.4.1	Operation of the marker protocol.....	70
6.5.4.2	Marker Responder state diagram	70
6.5.4.2.1	Variables.....	70
6.5.4.2.2	Messages.....	71
6.6	Conversation-sensitive frame collection and distribution	71
6.6.1	Conversation-sensitive collection and distribution state diagrams.....	72
6.6.1.1	Conversion-sensitive collection state diagram	72
6.6.1.1.1	Variables.....	72
6.6.1.1.2	Variables associated with each Aggregation Port	73
6.6.1.1.3	Functions	73
6.6.1.1.4	Messages.....	73
6.6.1.1.5	State diagram	73
6.6.2	Conversation-sensitive LACP state diagrams.....	74
6.6.2.1	Per-Aggregator Variables	74
6.6.2.2	Variables associated with each Aggregation Port.....	76
6.6.2.3	Variables used for managing the operation of the state diagrams	78
6.6.2.4	Functions.....	78
6.6.2.5	Timers	81
6.6.2.6	Messages	81
6.6.2.7	State diagrams.....	81
6.7	Configuration capabilities and restrictions	87
6.7.1	Use of system and port priorities	87
6.7.2	Dynamic allocation of operational Keys	87
6.7.3	Link Aggregation on shared-medium links	88
6.7.4	Selection Logic variants.....	88
6.7.4.1	Reduced reconfiguration.....	88
6.7.4.2	Limited Aggregator availability.....	89
7.	Management.....	90
7.1	Overview.....	90
7.1.1	Systems management overview	90
7.1.2	Management model.....	91
7.2	Managed objects	91

7.2.1	Introduction.....	91
7.2.2	Overview of managed objects.....	92
7.2.2.1	Text description of managed objects	92
7.2.3	Containment.....	93
7.2.4	Naming.....	93
7.2.5	Capabilities	94
7.3	Management for Link Aggregation	98
7.3.1	Aggregator managed object class	98
7.3.1.1	Aggregator attributes	99
7.3.1.1.1	aAggID	99
7.3.1.1.2	aAggDescription	99
7.3.1.1.3	aAggName	100
7.3.1.1.4	aAggActorSystemID	100
7.3.1.1.5	aAggActorSystemPriority	100
7.3.1.1.6	aAggAggregateOrIndividual	100
7.3.1.1.7	aAggActorAdminKey	100
7.3.1.1.8	aAggActorOperKey	101
7.3.1.1.9	aAggMACAddress	101
7.3.1.1.10	aAggPartnerSystemID	101
7.3.1.1.11	aAggPartnerSystemPriority	101
7.3.1.1.12	aAggPartnerOperKey	102
7.3.1.1.13	aAggAdminState	102
7.3.1.1.14	aAggOperState.....	102
7.3.1.1.15	aAggTimeOfLastOperChange.....	102
7.3.1.1.16	aAggDataRate.....	103
7.3.1.1.17	aAggOctetsTxOK	103
7.3.1.1.18	aAggOctetsRxOK	103
7.3.1.1.19	aAggFramesTxOK.....	103
7.3.1.1.20	aAggFramesRxOK	104
7.3.1.1.21	aAggMulticastFramesTxOK	104
7.3.1.1.22	aAggMulticastFramesRxOK	104
7.3.1.1.23	aAggBroadcastFramesTxOK.....	104
7.3.1.1.24	aAggBroadcastFramesRxOK	105
7.3.1.1.25	aAggFramesDiscardedOnTx	105
7.3.1.1.26	aAggFramesDiscardedOnRx	105
7.3.1.1.27	aAggFramesWithTxErrors	105
7.3.1.1.28	aAggFramesWithRxErrors	106
7.3.1.1.29	aAggUnknownProtocolFrames	106
7.3.1.1.30	aAggPortList.....	106
7.3.1.1.31	aAggLinkUpDownNotificationEnable.....	106
7.3.1.1.32	aAggCollectorMaxDelay	106
7.3.1.1.33	aAggPortAlgorithm	107
7.3.1.1.34	aAggPartnerAdminPortAlgorithm.....	107
7.3.1.1.35	aAggConversationAdminLink[]	107
7.3.1.1.36	aAggPartnerAdminPortConversationListDigest	107
7.3.1.1.37	aAggAdminDiscardWrongConversation.....	108
7.3.1.1.38	aAggAdminServiceConversationMap[]	108
7.3.1.1.39	aAggPartnerAdminConvServiceMappingDigest	108
7.3.1.2	Aggregator Notifications	108
7.3.1.2.1	nAggLinkUpNotification.....	108
7.3.1.2.2	nAggLinkDownNotification.....	109
7.3.2	Aggregation Port managed object class.....	109
7.3.2.1	Aggregation Port Attributes.....	109
7.3.2.1.1	aAggPortID.....	109

7.3.2.1.2	aAggPortActorSystemPriority	109
7.3.2.1.3	aAggPortActorSystemID	109
7.3.2.1.4	aAggPortActorAdminKey	110
7.3.2.1.5	aAggPortActorOperKey	110
7.3.2.1.6	aAggPortPartnerAdminSystemPriority	110
7.3.2.1.7	aAggPortPartnerOperSystemPriority	110
7.3.2.1.8	aAggPortPartnerAdminSystemID	110
7.3.2.1.9	aAggPortPartnerOperSystemID	111
7.3.2.1.10	aAggPortPartnerAdminKey.....	111
7.3.2.1.11	aAggPortPartnerOperKey.....	111
7.3.2.1.12	aAggPortSelectedAggID	111
7.3.2.1.13	aAggPortAttachedAggID	111
7.3.2.1.14	aAggPortActorPort	112
7.3.2.1.15	aAggPortActorPortPriority	112
7.3.2.1.16	aAggPortPartnerAdminPort.....	112
7.3.2.1.17	aAggPortPartnerOperPort.....	112
7.3.2.1.18	aAggPortPartnerAdminPortPriority	112
7.3.2.1.19	aAggPortPartnerOperPortPriority	113
7.3.2.1.20	aAggPortActorAdminState.....	113
7.3.2.1.21	aAggPortActorOperState	113
7.3.2.1.22	aAggPortPartnerAdminState	113
7.3.2.1.23	aAggPortPartnerOperState	114
7.3.2.1.24	aAggPortAggregateOrIndividual.....	114
7.3.2.1.25	aAggPortOperConversationPasses	114
7.3.2.1.26	aAggPortOperConversationCollected	114
7.3.2.1.27	aAggPortLinkNumberID	114
7.3.2.1.28	aAggPortPartnerAdminLinkNumberID	115
7.3.2.1.29	aAggPortWTRTime.....	115
7.3.2.2	Aggregation Port Extension Attributes.....	115
7.3.2.2.1	aAggPortProtocolDA.....	115
7.3.3	Aggregation Port Statistics managed object class	115
7.3.3.1	Aggregation Port Statistics attributes	116
7.3.3.1.1	aAggPortStatsID	116
7.3.3.1.2	aAggPortStatsLACPDU Rx	116
7.3.3.1.3	aAggPortStatsMarkerPDURx	116
7.3.3.1.4	aAggPortStatsMarkerResponsePDURx	116
7.3.3.1.5	aAggPortStatsUnknownRx.....	116
7.3.3.1.6	aAggPortStatsIllegalRx	117
7.3.3.1.7	aAggPortStatsLACPDU Tx	117
7.3.3.1.8	aAggPortStatsMarkerPDUTx	117
7.3.3.1.9	aAggPortStatsMarkerResponsePDUTx	117
7.3.4	Aggregation Port Debug Information managed object class	117
7.3.4.1	Aggregation Port Debug Information attributes	117
7.3.4.1.1	aAggPortDebugInformationID.....	117
7.3.4.1.2	aAggPortDebugRxState.....	118
7.3.4.1.3	aAggPortDebugLastRxTime	118
7.3.4.1.4	aAggPortDebugMuxState.....	118
7.3.4.1.5	aAggPortDebugMuxReason	119
7.3.4.1.6	aAggPortDebugActorChurnState	119
7.3.4.1.7	aAggPortDebugPartnerChurnState.....	119
7.3.4.1.8	aAggPortDebugActorChurnCount	119
7.3.4.1.9	aAggPortDebugPartnerChurnCount	120
7.3.4.1.10	aAggPortDebugActorSyncTransitionCount	120
7.3.4.1.11	aAggPortDebugPartnerSyncTransitionCount	120

	7.3.4.1.12	aAggPortDebugActorChangeCount	120
	7.3.4.1.13	aAggPortDebugPartnerChangeCount.....	120
	7.3.4.1.14	aAggPortDebugActorCDSChurnState	120
	7.3.4.1.15	aAggPortDebugPartnerCDSChurnState.....	121
	7.3.4.1.16	aAggPortDebugActorCDSChurnCount.....	121
	7.3.4.1.17	aAggPortDebugPartnerCDSChurnCount	121
7.4		Management for Distributed Resilient Network Interconnect.....	121
7.4.1		Distributed Relay Managed Object Class	121
	7.4.1.1	Distributed Relay Attributes	122
	7.4.1.1.1	aDrniID	122
	7.4.1.1.2	aDrniDescription	122
	7.4.1.1.3	aDrniName.....	122
	7.4.1.1.4	aDrniPortalAddr	122
	7.4.1.1.5	aDrniPortalPriority	122
	7.4.1.1.6	aDrniThreePortalSystem	123
	7.4.1.1.7	aDrniPortalSystemNumber.....	123
	7.4.1.1.8	aDrniIntraPortalLinkList	123
	7.4.1.1.9	aDrniAggregator	123
	7.4.1.1.10	aDrniConvAdminGateway[]	123
	7.4.1.1.11	aDrniNeighborAdminConvGatewayListDigest	124
	7.4.1.1.12	aDrniNeighborAdminConvPortListDigest	124
	7.4.1.1.13	aDrniGatewayAlgorithm	124
	7.4.1.1.14	aDrniNeighborAdminGatewayAlgorithm	124
	7.4.1.1.15	aDrniNeighborAdminPortAlgorithm.....	125
	7.4.1.1.16	aDrniNeighborAdminDRCPState	125
	7.4.1.1.17	aDrniEncapsulationMethod	125
	7.4.1.1.18	aDrniIPLEncapMap	125
	7.4.1.1.19	aDrniNetEncapMap	126
	7.4.1.1.20	aDrniDRPortConversationPasses	126
	7.4.1.1.21	aDrniDRGatewayConversationPasses.....	126
	7.4.1.1.22	aDrniPSI	126
	7.4.1.1.23	aDrniPortConversationControl.....	127
	7.4.1.1.24	aDrniIntraPortalPortProtocolDA	127
7.4.2		IPP Managed Objects Class	127
	7.4.2.1	IPP Attributes.....	127
	7.4.2.1.1	aIPPID	127
	7.4.2.1.2	aIPPPortConversationPasses	127
	7.4.2.1.3	aIPPGatewayConversationDirection	128
	7.4.2.1.4	aIPPAAdminState.....	128
	7.4.2.1.5	aIPPOperState.....	128
	7.4.2.1.6	aIPPTimeOfLastOperChange	128
7.4.3		IPP Statistics managed object class	129
	7.4.3.1	IPP Statistics attributes	129
	7.4.3.1.1	aIPPStatsID.....	129
	7.4.3.1.2	aIPPStatsDRCPDUsRx	129
	7.4.3.1.3	aIPPStatsIllegalRx	129
	7.4.3.1.4	aIPPStatsDRCPDUsTx.....	129
7.4.4		IPP Debug Information managed object class	129
	7.4.4.1	IPP Debug Information attributes	130
	7.4.4.1.1	aIPPDebugInformationID.....	130
	7.4.4.1.2	aIPPDebugDRCPRxState	130
	7.4.4.1.3	aIPPDebugLastRxTime	130
	7.4.4.1.4	aIPPDebugDifferPortalReason.....	130

8.	Frame distribution and collection algorithms	131
8.1	Conversation Identifiers.....	131
8.2	Per-service frame distribution.....	131
8.2.1	Goals and objectives	131
8.2.2	Overview.....	131
8.2.3	Port Conversation Identifiers	132
9.	Distributed Resilient Network Interconnect	133
9.1	Goals and objectives	133
9.2	Distributed Relay	134
9.3	Distributed Relay operation and procedures.....	136
9.3.1	Portal Topology	139
9.3.2	Intra-Portal Link	140
9.3.2.1	Network / IPL sharing by time	140
9.3.2.2	Network / IPL sharing by tag.....	141
9.3.2.3	Network / IPL sharing by encapsulation.....	141
9.3.3	Protocol Identification	142
9.3.4	DR Function state machines	142
9.3.4.1	Service interfaces.....	143
9.3.4.2	Per-DR Function variables.....	143
9.3.4.3	Per-IPP Intra-Portal Port variables	144
9.3.4.4	Functions.....	144
9.3.4.5	Messages	145
9.3.4.6	DR Function: Aggregator Port reception state machine	145
9.3.4.7	DR Function: Gateway distribution state machine	145
9.3.4.8	DR Function: IPP N reception state machine	146
9.4	Distributed Relay Control Protocol	147
9.4.1	Establishing the Portal and Distributed Relay	149
9.4.2	DRCPDU transmission, addressing, and protocol identification	149
9.4.2.1	Destination MAC Address.....	149
9.4.2.2	Source MAC Address	150
9.4.2.3	Priority	150
9.4.2.4	Encapsulation of DRCPDUs in frames.....	150
9.4.3	DRCPDU structure and encoding.....	150
9.4.3.1	Transmission and representation of octets.....	150
9.4.3.2	DRCPDU structure	151
9.4.3.3	Conversation Vector TLVs.....	158
9.4.3.3.1	2P Gateway Conversation Vector TLV	158
9.4.3.3.2	3P Gateway Conversation Vector-1 TLV	159
9.4.3.3.3	3P Gateway Conversation Vector-2 TLV	159
9.4.3.3.4	2P Port Conversation Vector TLV	160
9.4.3.3.5	3P Port Conversation Vector-1 TLV	160
9.4.3.3.6	3P Port Conversation Vector-2 TLV	161
9.4.3.4	Network/IPL sharing TLVs	161
9.4.3.4.1	Network/IPL Sharing Method TLV	162
9.4.3.4.2	Network/IPL Sharing Encapsulation TLV	163
9.4.3.5	Organization-Specific TLV	163
9.4.4	DRCP Control Parser/Multiplexer.....	164
9.4.4.1	Control Parser state diagram.....	164
9.4.4.1.1	Control Parser Function.....	164
9.4.4.1.2	Constants	164
9.4.4.1.3	Variables.....	164

9.4.5	DRCP state machine overview	165
9.4.6	Constants.....	166
9.4.7	Variables associated with the Distributed Relay	167
9.4.8	Per-DR Function variables.....	167
9.4.9	Per-IPP Intra-Portal Port variables	170
9.4.10	Variables used for managing the operation of the state machines.....	176
9.4.11	Functions.....	178
9.4.12	Timers	191
9.4.13	Messages.....	191
9.4.14	DRCPDU Receive machine.....	191
9.4.15	DRCP Periodic Transmission machine.....	194
9.4.16	Portal System machine.....	195
9.4.17	DRNI Gateway and Aggregator machines	196
9.4.18	DRNI IPP machines.....	197
9.4.19	DRCPDU Transmit machine	198
9.4.20	Network/IPL sharing machine	199
Annex A (normative) Protocol Implementation Conformance Statement (PICS) proforma		201
A.1	Introduction.....	201
A.1.1	Abbreviations and special symbols	201
A.1.2	Instructions for completing the PICS proforma	202
A.1.3	Additional information	202
A.1.4	Exceptional information.....	202
A.1.5	Conditional items	203
A.1.6	Identification	203
A.1.6.1	Implementation identification.....	203
A.1.6.2	Protocol summary	203
A.2	PICS proforma for Clause 6.....	204
A.2.1	Major capabilities/options	204
A.2.2	LLDP Port connectivity	205
A.2.3	Protocol Parser/Multiplexer support	205
A.2.4	Frame Collector.....	205
A.2.5	Frame Distributor	206
A.2.6	Marker protocol	206
A.2.7	Aggregator Parser/Multiplexer.....	206
A.2.8	Control Parser/Multiplexer.....	207
A.2.9	System identification.....	207
A.2.10	Aggregator identification	207
A.2.11	Port identification.....	208
A.2.12	Capability identification.....	208
A.2.13	Link Aggregation Group identification.....	208
A.2.14	Detaching a link from an Aggregator.....	208
A.2.15	LACPDU structure	209
A.2.16	Version 2 LACPDU	209
A.2.17	State machine variables	209
A.2.18	Receive machine	210
A.2.19	Periodic Transmission machine	210
A.2.20	Selection Logic.....	210
A.2.21	Mux machine.....	211
A.2.22	Transmit machine.....	211
A.2.23	Churn Detection machines	212
A.2.24	Marker protocol.....	212
A.2.25	Management	214

A.2.26	Per-Service Frame Distribution.....	217
A.2.27	Conversation-sensitive frame collection and distribution	218
A.2.28	Configuration capabilities and restrictions.....	218
A.2.29	Link Aggregation on shared-medium links.....	219
A.2.30	Distributed Resilient Network Interconnect.....	219
A.2.31	DRCPDU structure.....	220
A.2.32	Bridge specific support.....	221
Annex B (informative)	Collection and distribution algorithms.....	222
B.1	Introduction.....	222
B.2	Port selection.....	223
B.3	Dynamic reallocation of conversations to different Aggregation Ports	223
B.4	Topology considerations in the choice of distribution algorithm	224
Annex C (informative)	LACP standby link selection and dynamic Key management	226
C.1	Introduction.....	226
C.2	Goals	226
C.3	Standby link selection.....	227
C.4	Dynamic Key management.....	227
C.5	A dynamic Key management algorithm	227
C.6	Example 1	229
C.7	Example 2	229
Annex D (normative)	SMIPv2 MIB definitions for Link Aggregation	231
D.1	Introduction.....	231
D.2	SNMP Management Framework	231
D.3	Security considerations	231
D.4	Structure of the MIB module	232
D.4.1	Relationship to the managed objects defined in Clause 7	233
D.4.2	MIB Subtrees.....	238
D.4.2.1	The dot3adAgg Subtree.....	238
D.4.2.2	The dot3adAggPort Subtree.....	238
D.4.2.3	The dot3adAggNotifications Subtree.....	238
D.4.2.4	The dot3adDrni Subtree	238
D.4.2.5	The dot3adIPP Subtree.....	238
D.5	Relationship to other MIBs	238
D.5.1	Relationship to the Interfaces MIB	238
D.5.2	Layering model	239
D.5.3	ifStackTable	239
D.5.4	ifRcvAddressTable.....	239
D.6	Definitions for Link Aggregation MIB.....	239
Annex E (informative)	Distributed Bridge	306
E.1	Distributed VLAN Bridge	306
E.2	Higher Layer Entities in a Distributed Bridge	310

Annex F (normative) Link Layer Discovery Protocol TLVs 311

 F.1 Link Aggregation TLV 311

 F.1.1 aggregation status 311

 F.1.2 aggregated Port ID 312

 F.1.3 Link Aggregation TLV usage rules 312

 F.1.4 Use of other TLVs on an Aggregator or Aggregation Link 312

Annex G (normative) Network / IPL sharing by time—MAC Address synchronization 314

 G.1 Address synchronization—design goals 315

 G.2 Address synchronization—non-goals 315

 G.3 Protocol summary 315

 G.4 Address Synchronization Description 315

 G.5 ASPDU transmission, addressing, and protocol identification 317

 G.5.1 Destination MAC Address 317

 G.5.2 Source MAC Address 317

 G.5.3 Priority 317

 G.5.4 Encapsulation of ASPDUs in frames 317

IEEE Standard for Local and metropolitan area networks— Link Aggregation

IMPORTANT NOTICE: IEEE Standards documents are not intended to ensure safety, health, or environmental protection, or ensure against interference with or from other devices or networks. Implementers of IEEE Standards documents are responsible for determining and complying with all appropriate safety, security, environmental, health, and interference protection practices and all applicable laws and regulations.

This IEEE document is made available for use subject to important notices and legal disclaimers. These notices and disclaimers appear in all publications containing this document and may be found under the heading “Important Notice” or “Important Notices and Disclaimers Concerning IEEE Documents.” They can also be obtained on request from IEEE or viewed at <http://standards.ieee.org/IPR/disclaimers.html>.

1. Overview

1.1 Scope

Link Aggregation provides protocols, procedures, and managed objects that allow the following:

- One or more parallel instances of full-duplex point-to-point links to be aggregated together to form a Link Aggregation Group (LAG), such that a MAC Client can treat the LAG as if it were a single link.
- A resilient interconnect using multiple full-duplex point-to-point links among one to three nodes in a network and one to three nodes in another, separately administered, network, along with a means to ensure that frames belonging to any given service will use the same physical path in both directions between the two networks.

This standard defines the MAC-independent Link Aggregation capability and general information relevant to specific MAC types that support Link Aggregation. The capabilities defined are compatible with previous versions of this standard.

1.2 Purpose

Link Aggregation allows the establishment of full-duplex point-to-point links that have a higher aggregate bandwidth than the individual links that form the aggregation, and the use of multiple systems at each end of the aggregation. This allows improved utilization of available links in bridged local area network (LAN) environments, along with improved resilience in the face of failure of individual links or systems. In

applications connecting separately administered networks, the networks are isolated from each other's fault recovery events.

1.3 State diagram conventions

This document uses the state diagram conventions of IEEE Std 802.1Q™-2011, Annex E.¹

Should a conflict exist between a state diagram and either the corresponding global transition tables or the textual description associated with the state machine, the state diagram takes precedence.

¹Information on references can be found in Clause 2.

2. Normative references

The following referenced documents are indispensable for the application of this document (i.e., they must be understood and used, so each referenced document is cited in text and its relationship to this document is explained). For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments or corrigenda) applies.

IEEE Std 802[®], IEEE Standard for Local and Metropolitan Area Networks: Overview and Architecture.^{2, 3}

IEEE Std 802.1AC[™], IEEE Standard for Local and Metropolitan Area Networks Media Access Control (MAC) Service Definition.

IEEE Std 802.1D[™], IEEE Standard for Local and metropolitan area networks Media Access Control (MAC) Bridges.

IEEE Std 802.1Q[™]-2011, IEEE Standard for Local and metropolitan area networks: Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks.

IEEE Std 802.3[™]-2012, IEEE Standard for Ethernet.

IETF RFC 1213 (IETF STD 17), Management Information Base for Network Management of TCP/IP-based internets: MIB-II, McCloghrie K., and M. Rose, Editors, March 1991.⁴

IETF RFC 1321, The MD5 Message-Digest Algorithm, R. Rivest. April 1992.

IETF RFC 2578 (STD 58), Structure of Management Information Version 2 (SMIv2), K. McCloghrie, D. Perkins, J. Schoenwaelder. April 1999.

IETF RFC 2579 (STD 58), Textual Conventions for SMIv2, McCloghrie, K., Perkins, D., Schoenwaelder, J., Case, J., Rose, M. and S. Waldbusser, April 1999.

IETF RFC 2580 (STD 58), Conformance Statements for SMIv2, McCloghrie, K., Perkins, D., Schoenwaelder, J., Case, J., Rose, M. and S. Waldbusser, April 1999.

IETF RFC 2863, The Interfaces Group MIB, K. McCloghrie, F. Kastenholz, June 2000.

IETF RFC 3410, Introduction and Applicability Statements for Internet-Standard Management Framework, J. Case, R. Mundy, D. Partain, B. Stewart. December 2002.

IETF RFC 3414 (STD 62), User-based Security Model (USM) for Version 3 of the Simple Network Management Protocol (SNMPv3), U. Blumenthal, B. Wijnen. December 2002.

IETF RFC 3415 (STD 62), View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP), B. Wijnen, R. Presuhn, K. McCloghrie, December 2002.

ISO/IEC 10165-4:1992, Information technology—Open Systems Interconnection—Structure of management information—Part 4: Guidelines for the definition of managed objects.⁵

²IEEE publications are available from The Institute of Electrical and Electronics Engineers (<http://standards.ieee.org/>).

³The IEEE standards or products referred to in this clause are trademarks of The Institute of Electrical and Electronics Engineers, Inc.

⁴IETF documents (i.e., RFCs) are available for download at <http://www.rfc-archive.org/>.

⁵ISO/IEC publications are available from the ISO Central Secretariat (<http://www.iso.org/>). ISO publications are also available in the United States from the American National Standards Institute (<http://www.ansi.org/>).

3. Definitions

For the purposes of this document, the following terms and definitions apply. The *IEEE Standards Dictionary Online* should be consulted for terms not defined in this clause.⁶

Actor: The local entity in a Link Aggregation Control Protocol exchange.

agent: A term used to refer to the managed nodes in a network. Managed nodes are those nodes that contain a network management entity, which can be used to configure the node and/or collect data describing operation of that node. The agent is controlled by a network control host or manager that contains both a network management entity and network management application software to control the operations of agents. Agents include Systems that support user applications as well as nodes that provide communications services such as front-end processors, bridges, and routers.

Aggregation Key: A parameter associated with each Aggregation Port and with each Aggregator of an Aggregation System identifying those Aggregation Ports that can be aggregated together. Aggregation Ports in an Aggregation System that share the same Aggregation Key value are potentially able to aggregate together.

Aggregation Link: A LAN connecting a pair of Aggregation Systems.

Aggregation Port: A Service Access Point (SAP) in an Aggregation System that is supported by a single MAC entity.

Aggregation System: A uniquely identifiable entity comprising (among other things) an arbitrary grouping of one or more Aggregation Ports for the purpose of aggregation. An instance of an aggregated link always occurs between exactly two Aggregation Systems. A physical device may comprise a single Aggregation System or more than one Aggregation System.

Aggregator: A logical MAC, bound to one or more Aggregation Ports, through which the Aggregator Client is provided access to the physical media.

Aggregator Client: The layered entity immediately above the Link Aggregation Sublayer, for which the Link Aggregation sublayer provides an instance of the Internal Sublayer Service (ISS).

Aggregator Port: A Service Access Point (SAP) in an Aggregation System that is supported by an Aggregator.

bridge: A layer 2 interconnection device that conforms to IEEE Std 802.1D and/or IEEE Std 802.1Q.

conversation: A set of frames transmitted from one end station to another, where all of the frames form an ordered sequence, and where the communicating end stations require the ordering to be maintained among the set of frames exchanged.

Conversation ID: An identifier using values in the range of 0 through 4095 to identify conversations.

data terminal equipment (DTE): Any source or destination of data connected to the local area network.

⁶IEEE Standards Dictionary Online subscription is available at:
http://www.ieee.org/portal/innovate/products/standard/standards_dictionary.html.

Distributed Relay: A functional entity, distributed over a Portal by a DR Function in each of the Systems comprising a Portal, that distributes Down frames from Gateways to Aggregators, and distributes Up frames from Aggregators to Gateways.

Distributed Resilient Network Interconnect (DRNI): Link Aggregation expanded to include at least one end of the Link Aggregation Group attached to a Portal such that independence and isolation is provided between the interconnected network segments.

NOTE—Link Aggregation as specified in Clause 6.⁷

Down frame: A frame entering a Portal through a Gateway.

DR Function: That part of a Distributed Relay residing within a single Portal System.

end station: A System attached to a local area network (LAN) that is an initial source or a final destination of frames transmitted across that LAN.

frame: A unit of data transmission on a local area network (LAN) that conveys a MAC Protocol Data Unit (MPDU) and can cause a service indication with, at a minimum, a Destination Address, Source Address, a MAC Service Data Unit (MSDU), and priority, or an MPDU that is the result of a service request with those parameters.

full duplex: A mode of operation of a network, data terminal equipment (DTE), or Medium Attachment Unit (MAU) that supports duplex transmission as defined in the *IEEE Standards Dictionary Online*.

NOTE—Within the scope of this standard, this mode of operation allows for simultaneous communication between a pair of stations, provided that the Physical Layer is capable of supporting simultaneous transmission and reception without interference. (See IEEE Std 802.3.)

Gateway: A connection, always virtual (not a physical link between Systems) connecting a Distributed Relay to a System, consisting of a Gateway Link and two Gateway Ports.

Gateway Conversation ID: The Conversation ID value that is used within the Distributed Relay to identify frames passing through a Gateway.

Gateway Vector: The operational Boolean vector, indexed by Gateway Conversation ID, indicating whether the indexed Gateway Conversation ID is enabled to pass through a Portal System's Gateway (FALSE = blocked). There is one Gateway Vector per Portal System in a Portal.

half duplex: A mode of operation of an CSMA/CD local area network (LAN) in which data terminal equipment (DTEs) contend for access to a shared medium. Multiple, simultaneous transmissions in a half-duplex mode CSMA/CD LAN result in interference, requiring resolution by the Ethernet protocol. (See IEEE Std 802.3.)

Internal Sublayer Service (ISS): An augmented version of the MAC Service, defined in IEEE Std 802.1AC.

Intra-Portal Link (IPL): A logical link used to connect the Distributed Relay (DR) Functions comprising a Distributed Relay.

Intra-Portal Port (IPP): A port to an Intra-Portal Link.

⁷Notes in text, tables, and figures of a standard are given for information only and do not contain requirements needed to implement this standard.

Key: *See: Aggregation Key.*

link: *See: Aggregation Link.*

Link Aggregation Group (LAG): A group of links that appear to an Aggregator Client as if they were a single link. A LAG can connect two Aggregation Systems, an Aggregation System and a Portal, or two Portals. One or more conversations may be associated with each link that is part of a LAG.

Long LACPDU: A Version 2 or higher version Link Aggregation Control Protocol Data Unit (LACPDU) of a frame size that is larger than 128 octets.

Management Information Base (MIB): A repository of information to describe the operation of a specific network device.

MAC: The data link sublayer that is responsible for transferring data to and from the Physical Layer.

Partner: The remote entity in a Link Aggregation Control Protocol exchange.

Port Conversation ID: The Conversation ID value that is used to select frames passing through an Aggregation Port.

Portal: One end of a Distributed Resilient Network Interconnect (DRNI); one to three Systems, each with physical links that together comprise a Link Aggregation Group. The Portal's Systems cooperate to emulate the presence of a single Aggregation System to which the entire Link Aggregation Group is attached.

Portal System: A System that participates in a Portal.

Portal System Number: An integer from 1 through 3, inclusive, uniquely identifying a Portal System within its Portal.

selection algorithm: The algorithm used to assign frames to Conversation IDs and Conversation IDs to Aggregation Ports and, in the context of Distributed Resilient Network Interconnect (DRNI), also to Gateways.

Service ID: A value extracted from the header of a frame (VID, I-SID, etc.), received from an Aggregation Link, that identifies the service instance with which that frame is associated.

service instance: A set of Service Access Points (SAPs) such that a Data.Request primitive presented to one SAP can result in a Data.Indication primitive occurring at one or more of the other SAPs in that set. In the context of operators and customers, a particular customer is given access to all of the SAPs of such a set by the operator.

System: *See: Aggregation System.*

Type/Length/Value (TLV): A variable length encoding of an information element consisting of sequential type, length, and value fields where the type field identifies the type of information, the length field indicates the length of the information field in octets, and the value field contains the information itself. The type value is locally defined and needs to be unique within the defined protocol.

Up frame: A frame entering a Portal through an Aggregation Port.

4. Acronyms and abbreviations

This standard contains the following acronyms and abbreviations:

ANSI	American National Standards Institute
CID	Company ID ⁸ (see also OUI)
DA	destination address
DRCP	Distributed Relay Control Protocol
DRCPDU	Distributed Relay Control Protocol Data Unit
DRNI	Distributed Resilient Network Interconnect
DTE	data terminal equipment
FCS	frame check sequence
IEC	International Electrotechnical Commission
IPL	Intra-Portal Link
IPP	Intra-Portal Port
ISO	International Organization for Standardization
ISS	Internal Sublayer Service
LACP	Link Aggregation Control Protocol
LACPDU	Link Aggregation Control Protocol Data Unit
LAG	Link Aggregation Group
LAG ID	Link Aggregation Group Identifier
LAN	local area network
LLC	logical link control
LLDP	Link Layer Discovery Protocol
LLID	logical link identifier
MAC	Medium Access Control
MAN	metropolitan area network
MIB	management information base
NTT	Need To Transmit
OUI	organizationally unique identifier
PDU	Protocol Data Unit
PICS	protocol implementation conformance statement
PSI	Portal System Isolated
SA	source address
TLV	Type/Length/Value
TPMR	Two-Port Media Access Control (MAC) Relay
UCT	unconditional transition

⁸ See <http://standards.ieee.org/develop/regauth/tut/eui.pdf>.

5. Conformance

This clause specifies the mandatory and optional capabilities provided by conformant implementations of this standard.

5.1 Requirements terminology

For consistency with existing IEEE and IEEE 802.1™ standards, requirements placed upon conformant implementations of this standard are expressed using the following terminology:

- a) **Shall** is used for mandatory requirements;
- b) **May** is used to describe implementation or administrative choices (“may” means “is permitted to,” and hence, “may” and “may not” mean precisely the same thing);
- c) **Should** is used for recommended choices (the behaviors described by “should” and “should not” are both permissible but not equally desirable choices).

The protocol implementation conformance statement (PICS) proformas (see Annex A) reflect the occurrences of the words “shall,” “may,” and “should” within the standard.

The standard avoids needless repetition and apparent duplication of its formal requirements by using **is**, **is not**, **are**, and **are not** for definitions and the logical consequences of conformant behavior. Behavior that is permitted but is neither always required nor directly controlled by an implementer or administrator, or whose conformance requirement is detailed elsewhere, is described by **can**. Behavior that never occurs in a conformant implementation or System of conformant implementations is described by **cannot**. The word **allow** is used as a replacement for the phrase “support the ability for,” and the word **capability** means “can be configured to.”

In this clause the term **conformant System** refers to a System that, for all ports for which support is claimed, conforms to the provisions of this standard for Link Aggregation (5.3) or Distributed Resilient Network Interconnect (DRNI)(5.4).

5.2 Protocol Implementation Conformance Statement

The supplier of an implementation that is claimed to conform to this standard shall provide the information necessary to identify both the supplier and the implementation, and shall complete a copy of the PICS proforma provided in Annex A.

5.3 Link Aggregation requirements

A conformant System, shall:

- a) Support the Link Aggregation Sublayer and conform to the state machines and procedures in 6.2;
- b) Support the Link Aggregation Control Protocol (LACP) and conform to the state machines and procedures in 6.3 and 6.4;
- c) Transmit and receive version 1 Link Aggregation Control Protocol Data Unit (LACPDUs) in the formats specified in 6.4.2;
- d) Implement the Marker Responder as specified in 6.5.4.2 and respond to received Marker PDUs as specified in 6.5.1.

5.3.1 Link Aggregation options

A conformant System, may:

- a) Support the Marker Protocol and conform to the state machines and procedures in 6.5 and transmit and receive required Marker and Marker Response PDUs in the formats specified in 6.5.3;
- b) Support the management functionality for Link Aggregation as specified in Clause 7;
- c) Support SMIPv2 management information base (MIB) modules for the management of Link Aggregation capabilities (Annex D);
- d) Support Aggregation Port Debug Information package support as specified in 7.3 and in this case shall support the Churn Detection machine as specified in 6.4.17;
- e) Conform to the required specifications of the Link Layer Discovery Protocol (LLDP) of IEEE Std 802.1AB. A conformant System that supports LLDP shall provide an LLDP Parser/Multiplexer to attach zero or more instances of LLDP to each Aggregation Port as specified in 6.1.3;
- f) Implement the Protocol Parser/Multiplexer (6.2.7) for a protocol not explicitly specified in 6.2.7;
- g) Support Per-service frame distribution as specified in 8.2. A conformant System that supports Per-service frame distribution shall support:
 - 1) Conversation-sensitive frame collection and distribution state diagrams as specified in 6.6.1; and
 - 2) Classification by Customer VLAN Tag for C-tagged interfaces and classification by Service VLAN tag for S-tagged and B-tagged interfaces, and may support:
 - 3) Conversation-sensitive LACP operation as specified in 6.6.2;
 - 4) Classification by Backbone Service Instance Tag as specified in 8.2.2;
 - 5) Classification by other methods not specified by this standard;
- h) Transmit and receive version 2 LACPDUs in the formats specified in 6.4.2. A conformant System that supports version 2 LACPDUs shall support the Long LACPDU machine as specified in 6.4.18.

5.4 Distributed Resilient Network Interconnect requirements

A conformant System, shall:

- a) Conform to the requirements specified for Link Aggregation as specified in 5.3;
- b) Support a Portal consisting of two Portal Systems as specified in Clause 9. A conformant System that supports a Portal consisting of two Portal Systems shall:
 - 1) Support a DR Function having a single Intra-Portal Link (IPP) and conform to the state machines and procedures in 9.2 and 9.3 that are applicable to a Portal System supporting a single IPP;
 - 2) Support the Distributed Relay Control Protocol (DRCP) and conform to the state machine and procedures applicable to a Portal System supporting a single IPP as specified in 9.4;
 - 3) Transmit and receive required Distributed Relay Control Protocol Data Units (DRCPDUs) in the formats specified in 9.4.3 that are applicable to a Portal consisting of two Portal Systems;
- c) Support using separate physical links for the Intra-Portal Link (IPL) and the network links as specified in 9.3.2.

5.4.1 Distribution Resilient Network Interconnect options

A conformant System, may implement any of the options specified for Link Aggregation (5.3.1), and may:

- a) Support a Portal consisting of three Portal Systems as specified in Clause 9. A conformant System that supports a Portal consisting of three Portal Systems shall:

- 1) Support a DR Function having two IPPs and conform to the state machines and procedures in 9.2 and 9.3 that are applicable to a Portal System supporting two IPPs;
 - 2) Support the DRCP and conform to the state machine and procedures applicable to a Portal System supporting two IPPs as specified in 9.4;
 - 3) Transmit and receive required DRCPDUs in the formats specified in 9.4.3 for three Portal Systems;
- b) Support any of the methods in 9.3.2, by which the Systems can distinguish frames on a network link from frames on a particular IPL. A System that supports Network / IPL sharing by time (9.3.2.1) may also support MAC address synchronization as specified in Annex G.
- c) Transmit and receive DRCPDUs in the formats specified in item b3) in , including any additional DRCP Type/Length/Values (TLVs) from Table 9-8.

6. Link Aggregation

6.1 Overview

This clause defines an optional Link Aggregation sublayer for use with links offering the IEEE 802.1AC Internal Sublayer Service (ISS). Link Aggregation allows one or more links to be aggregated together to form a Link Aggregation Group (LAG), such that an Aggregator Client can treat the LAG as if it were a single link. To this end, Link Aggregation specifies the establishment of data terminal equipment (DTE)-to-DTE logical links, consisting of N parallel instances of full-duplex point-to-point links.

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.

6.1.1 Goals and objectives

Link Aggregation, as specified in this clause, provides the following:

- a) **Increased bandwidth**—The capacity of multiple links is combined into one logical link.
- b) **Linearly incremental bandwidth**—Bandwidth can be increased in unit multiples as opposed to the order-of-magnitude increase available through Physical Layer technology options (10 Mb/s, 100 Mb/s, 1000 Mb/s, etc.).
- c) **Increased availability**—The failure or replacement of a single link within a LAG need not cause failure from the perspective of an Aggregator Client.
- d) **Load sharing**—Aggregator Client traffic may be distributed across multiple links.
- e) **Automatic configuration**—In the absence of manual overrides, an appropriate set of LAGs is automatically configured, and individual links are allocated to those groups. If a set of links can aggregate, they will aggregate.
- f) **Rapid configuration and reconfiguration**—In the event of changes in physical connectivity, Link Aggregation will quickly converge to a new configuration, typically on the order of milliseconds for link down events and 1 s or less for link up events.

NOTE 1—Timing out LACPDU exchanges is the method of last resort for detecting link failures and shifting data flows to other physical links. The MAC_Operational status generated by the link hardware is the first choice for link failure detection. IEEE Std 802.1Q-2014, Clause 18, provides a method for detecting link failures (also indicated via the MAC_Operational status) when hardware means are not applicable.

- g) **Deterministic behavior**—Depending on the selection algorithm chosen, the configuration can be made to resolve deterministically; i.e., the resulting aggregation can be made independent of the order in which events occur, and be completely determined by the capabilities of the individual links and their physical connectivity.
- h) **Low risk of duplication or misordering**—During both steady-state operation and link (re)configuration, there is a low probability that frames are duplicated or misordered.
- i) **Support of existing MAC Clients**—No change is required to existing higher-layer protocols or applications to use Link Aggregation.
- j) **Backwards compatibility with aggregation-unaware devices**—Links that cannot take part in Link Aggregation—either because of their inherent capabilities, management configuration, or the capabilities of the devices to which they attach—operate as normal, individual links.
- k) **Accommodation of differing capabilities and constraints**—Devices with differing hardware and software constraints on Link Aggregation are, to the extent possible, accommodated.
- l) **No change to frame formats**—Link aggregation neither adds to, nor changes the contents of, frames exchanged between Aggregator Clients.

- m) **Network management support**—The standard specifies appropriate management objects for configuration, monitoring, and control of Link Aggregation.
- n) **Dissimilar MACs**—Link Aggregation can be used over any physical or logical medium offering the ISS.

Link Aggregation, as specified in this clause, does not support the following:

- o) **Multipoint Aggregations**—The mechanisms specified in this clause do not support aggregations among more than two Systems. (See Clause 9 for aggregations among two Portals each consisting of up to three Systems.)
- p) **Half-duplex operation**—Link Aggregation is supported only on point-to-point links with MACs operating in full-duplex mode.

Aggregation of links of different data rates is not prohibited nor required by this standard. Determining how to distribute traffic across links of different data rates is beyond the scope of this standard.

NOTE 2—Previous versions of this standard restricted the data rate of the aggregated links. However, the specified mechanisms did not depend on whether the links operated at the same rate or not.

6.1.2 Positioning of Link Aggregation within the IEEE 802 architecture

Link Aggregation comprises an optional sublayer between an Aggregator Client and the MAC. Figure 6-1 depicts the positioning of the Link Aggregation sublayer in the IEEE 802 layer architecture, and the relationship of that architecture to the Data Link and Physical Layers of the OSI Reference Model. The figure also shows the ability of the Link Aggregation sublayer to combine a number of individual links in order to present a single MAC interface to the Aggregator Client. (See also IEEE Std 802.1Q-2014, Figure 22-8.)

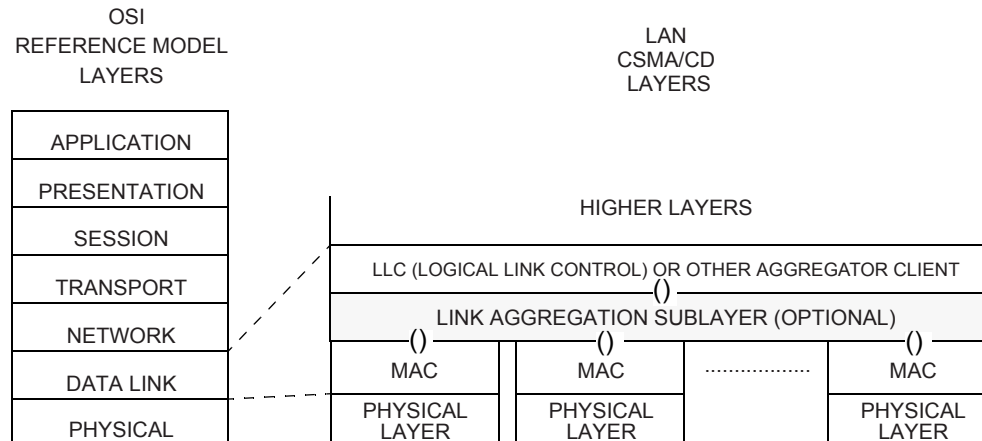


Figure 6-1—Architectural positioning of Link Aggregation sublayer

Figure 6-2 depicts the major blocks that form the Link Aggregation sublayer and their interrelationships.

It is possible to implement the optional Link Aggregation sublayer for some ports within a System while not implementing it for other ports; i.e., it is not necessary for all ports in a System to be subject to Link Aggregation. A conformant implementation is not required to be able to apply the Link Aggregation sublayer to every port.

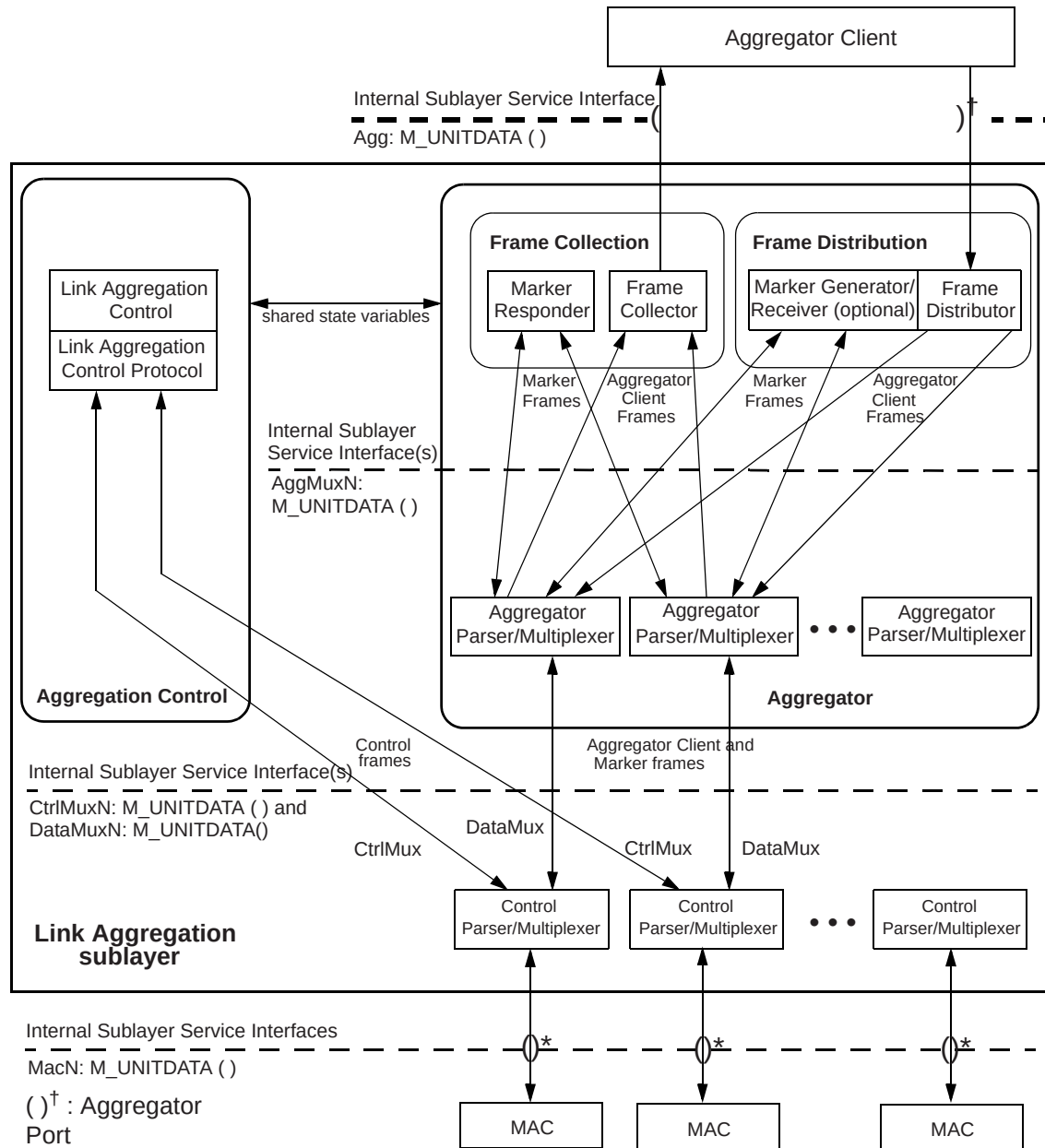


Figure 6-2—Link Aggregation sublayer block diagram

NOTE—Earlier editions of IEEE Std 802.1AX and IEEE Std 802.3, Clause 43, used the term “MAC Client” instead of the term “Aggregator Client,” and placed Link Aggregation between the IEEE 802.3 MAC sublayer and the IEEE 802.3 MAC Control sublayer. This standard repositions Link Aggregation immediately above the MAC sublayer. This rearrangement of the abstract layering diagram is intended to align the standard with common industry practice; it does not cause an implementation that is conformant to earlier editions of the standard to become non-conformant with this edition.

6.1.3 LLDP Parser/Multiplexer

Not shown in Figure 6-1 is the fact that some protocols require that the higher layers be able to access physical ports directly, without passing through the Link Aggregation sublayer. Some protocols specify a

means whereby this access is accomplished. See, for example, IEEE Std 802.1X™-2010, Figure 6-2, the “SecY.”

In order to satisfy the needs of other protocols, the Protocol Parser/Multiplexer (6.2.7) is provided. In particular, if IEEE Std 802.1AB™ is implemented in a System supporting Link Aggregation, an LLDP Parser/Multiplexer shall be provided to attach zero or more instances of LLDP to each Aggregation Port as configured by the system administrator.

There are N LLDP Parser/Multiplexers, one for each Aggregation Port. Each LLDP Parser/Multiplexer is an instance of the Protocol Parser/Multiplexer described in 6.2.7 and is inserted as a shim layer, either between the Control Parser/Multiplexer and the MAC layer, or between the Aggregator Parser/Multiplexer and the Control Parser/Multiplexer, at the discretion of the implementer. Specifically:

- a) The *DownPort* of Protocol Parser/Multiplexer N is either the *DataMuxN* of Control Parser/Multiplexer N (6.2.10) or the *MacN* of MAC N, as determined by the position of the LLDP Parser/Multiplexer in the protocol stack.
- b) The *ControlPort* of Protocol Parser/Multiplexer N is utilized by the LLDP instances.
- c) The *DataPort* of Protocol Parser/Multiplexer N is either the *DataMuxN* of Aggregator Parser/Multiplexer N (6.2.8) or the *MacN* of Control Parser/Multiplexer N, as determined by the position of the LLDP Parser/Multiplexer in the protocol stack.
- d) The *IsControlFrame* function is defined in 6.1.3.1.1.

NOTE—Any functional entity that operates above the Aggregator is under the control and the specification of the System implementing Link Aggregation (in Bridges, for example, this is provided by the Transmit and Receive functions specified in 8.5 of IEEE Std 802.1Q-2014). The discussion of the position of any Protocol Parser/Multiplexer entity that is above the Aggregator is not in scope of this standard.

6.1.3.1 LLDP Parser state diagram

6.1.3.1.1 LLDP Parser Function

IsControlFrame

Returns Boolean value: (Length/Type == LLDP_Type && DA == LLDP_Multicast address)

6.1.3.1.2 Constants

LLDP_Multicast address

The value of the LLDP DA field. (See IEEE Std 802.1AB-2009, 7.1.)

LLDP_Type

The value of the LLDP Length/Type field. (See IEEE Std 802.1AB-2009, Figure C.1.)

6.1.3.1.3 Variables

DA

The value of the DA field in a received frame.

Length/Type

The value of the Length/Type field in a received frame.

6.1.3.1.4 State diagram

The LLDP Parser state diagram, when supported, shall implement the function specified in Figure 6-5 with its associated parameters as specified in 6.1.3.

6.2 Link Aggregation operation

As depicted in Figure 6-2, the Link Aggregation sublayer comprises the following functions:

- a) *Frame Distribution*. This block is responsible for taking frames submitted by the Aggregator Client and submitting them for transmission on the appropriate Aggregation Port, based on a frame distribution algorithm employed by the Frame Distributor. Frame Distribution includes a *Frame Distributor* and an optional *Marker Generator/Receiver* used for the Marker protocol. (See 6.2.4, 6.2.5, and 6.5.)
- b) *Frame Collection*. This block is responsible for passing frames received from the various Aggregation Ports to the Aggregator Client. Frame Collection includes a *Frame Collector* and a *Marker Responder*, used for the Marker protocol. (See 6.2.3, 6.2.6, and 6.5.)
- c) *Aggregator Parser/Multiplexers*. On transmit, these blocks simply pass frame transmission requests from the Frame Distributor, Marker Generator, and/or Marker Responder to the appropriate Aggregation Port. On receive, these blocks distinguish among Marker Request, Marker Response, and MAC Protocol Data Units (MPDUs), and pass each to the appropriate entity (Marker Responder, Marker Receiver, and Frame Collector, respectively).
- d) *Aggregator*. The combination of Frame Distribution and Frame Collection, along with the Aggregator Parser/Multiplexers, is referred to as the Aggregator.
- e) *Aggregation Control*. This block is responsible for the configuration and control of Link Aggregation. It incorporates a *Link Aggregation Control Protocol* (LACP) that can be used for automatic communication of aggregation capabilities between Systems and automatic configuration of Link Aggregation.
- f) *Control Parser/Multiplexers*. On transmit, these blocks simply pass frame transmission requests from the Aggregator and Control entities to the appropriate Aggregation Port. On receive, these blocks distinguish Link Aggregation Control PDUs from other frames, passing the LACPDUs to the appropriate sublayer entity, and all other frames to the Aggregator.

6.2.1 Principles of Link Aggregation

Link Aggregation allows an Aggregator Client to treat a set of one or more Aggregation Ports as if it were a single port. In doing so, it employs the following principles and concepts:

- a) An Aggregator Client communicates with a set of Aggregation Ports through an Aggregator, which presents a standard ISS interface to the Aggregator Client. The Aggregator binds to one or more Aggregation Ports within a System.
- b) It is the responsibility of the Aggregator to distribute frame transmissions from the Aggregator Client to the various Aggregation Ports, and to collect received frames from the Aggregation Ports and pass them to the Aggregator Client transparently.
- c) A System may contain multiple Aggregators, serving multiple Aggregator Clients. A given Aggregation Port will bind to (at most) a single Aggregator at any time. An Aggregator Client is served by a single Aggregator at a time.
- d) The binding of Aggregation Ports to Aggregators within a System is managed by the Link Aggregation Control function for that System, which is responsible for determining which links may be aggregated, aggregating them, binding the Aggregation Ports within the System to an appropriate Aggregator, and monitoring conditions to determine when a change in aggregation is needed.
- e) Such determination and binding may be under manual control through direct manipulation of the state variables of Link Aggregation (e.g., Keys) by a network manager. In addition, automatic determination, configuration, binding, and monitoring may occur through the use of a Link Aggregation Control Protocol (LACP). The LACP uses peer exchanges across the links to determine, on an ongoing basis, the aggregation capability of the various links, and continuously provides the maximum level of aggregation capability achievable between a given pair of Systems.

- f) Frame ordering has to be maintained for certain sequences of frame exchanges between Aggregator Clients (known as *conversations*, see Clause 3). The Frame Distributor ensures that all frames of a given conversation are passed to a single Aggregation Port. For any given Aggregation Port, the Frame Collector is required to pass frames to the Aggregator Client in the order that they are received from that Aggregation Port. The Frame Collector is otherwise free to select frames received from the Aggregation Ports in any order. Since there are no means for frames to be misordered on a single link, this guarantees that frame ordering is maintained for any conversation.
- g) Conversations may be moved among Aggregation Ports within an aggregation, both for load balancing and to maintain availability in the event of link failures. Means are provided (6.5, 6.6) for maintaining frame ordering when such movement takes place.
- h) This standard does not impose any particular distribution algorithm on the Frame Distributor. Whatever algorithm is used should be appropriate for the Aggregator Client being supported.
- i) Each Aggregation Port is assigned a MAC address, unique over the LAG and the IEEE 802.1Q Bridged LAN (if any) to which the LAG is connected. This MAC address is used as the source address (SA) for frame exchanges that are initiated by entities within the Link Aggregation sublayer itself (i.e., LACP and Marker protocol exchanges).

NOTE—The LACP and Marker protocols use a multicast destination address (DA) for all exchanges, and do not impose any requirement for an Aggregation Port to recognize more than one unicast address on received frames.

- j) Each Aggregator is assigned a MAC address, unique over the LAG and the IEEE 802.1Q Bridged LAN (if any) to which the LAG is connected; this address is used as the MAC address of the aggregation from the perspective of the Aggregator Client, both as a SA for transmitted frames and as the destination address (DA) for received frames. The MAC address of the Aggregator may be one of the MAC addresses of an Aggregation Port in the associated LAG (see 6.2.11).

6.2.2 Service interfaces

The Aggregator Client communicates with the Aggregator using the ISS specified in IEEE Std 802.1AC. Similarly, Link Aggregation communicates internally (between Frame Collection/Distribution, the Aggregator Parser/Multiplexers, the Control Parser/Multiplexers, and Link Aggregation Control) and with its bound Aggregation Ports using the same service interface. No new interlayer service interfaces are defined for Link Aggregation.

Since Link Aggregation uses four instances of the ISS, it is necessary to introduce a notation convention so that the reader can be clear as to which interface is being referred to at any given time. A prefix is therefore assigned to each service primitive, indicating which of the four interfaces is being invoked, as depicted in Figure 6-2. The prefixes are as follows:

- a) *Agg:*, for primitives issued on the interface between the Aggregator Client and the Link Aggregation sublayer.
- b) *AggMuxN:*, for primitives issued on the interface between Aggregator Parser/Multiplexer N and its internal clients (where N is the Port Number associated with the Aggregator Parser/Multiplexer).
- c) *CtrlMuxN:*, for primitives issued on the interface between Control Parser/Multiplexer N and Link Aggregation Control (where N is the Port Number associated with the Control Parser/Multiplexer).
- d) *DataMuxN:*, for primitives issued on the interface between Control Parser/Multiplexer N and Aggregator Parser/Multiplexer N (where N is the Port Number associated with the Control Parser/Multiplexer).
- e) *MacN:*, for primitives issued on the interface between underlying MAC N and its Control Parser/Multiplexer (where N is the Port Number associated with the underlying MAC).

Aggregator Clients may generate *Agg:M_UNITDATA.request* primitives for transmission on an aggregated link. These are passed by the Frame Distributor to an Aggregation Port selected by the distribution algorithm. *MacN:M_UNITDATA.indication* primitives signifying received frames are passed unchanged from an Aggregation Port to the Aggregator Client by the Frame Collector.

In order to satisfy the needs of higher layers to access physical ports directly, the Protocol Parser/Multiplexer (6.2.7) is provided. In particular, if IEEE Std 802.1AB is implemented in a System supporting Link Aggregation, the LLDP Parser/Multiplexer (6.1.3) will enable to attach zero or more instances of LLDP to each Aggregation Port.

6.2.3 Frame Collector

A Frame Collector is responsible for receiving incoming frames (i.e., AggMuxN:M_UNITDATA.indications) from the set of individual links that form the LAG (through each link's associated Aggregator Parser/Multiplexer) and delivering them to the Aggregator Client. Frames received from a given Aggregation Port are delivered to the Aggregator Client in the order that they are received by the Frame Collector. Since the Frame Distributor is responsible for maintaining any frame ordering constraints, there is no requirement for the Frame Collector to perform any reordering of frames received from multiple links. If a System uses conversation-sensitive frame collection and distribution, then the Frame Collector will discard frames received on Aggregation Ports whose Conversation IDs are not in the list supplied by the Frame Distributor for that Aggregation Port (6.6).

The Frame Collector shall implement the function specified in the state diagram shown in Figure 6-3 and the associated definitions contained in 6.2.3.1.

6.2.3.1 Frame Collector state diagram

6.2.3.1.1 Constants

CollectorMaxDelay

In tens of microseconds, the maximum time that the Frame Collector may delay the delivery of a frame received from an Aggregator Parser to its Aggregator Client. Value is assigned by management or administration policy.

Value: Integer

6.2.3.1.2 Variables

DA

SA

mac_service_data_unit

priority

The parameters of the M_UNITDATA.indication primitive.

BEGIN

A Boolean variable that is set to TRUE when the System is initialized or reinitialized, and is set to FALSE when (re-)initialization has completed.

Value: Boolean

6.2.3.1.3 Messages

Agg:M_UNITDATA.indication

AggMuxN:M_UNITDATA.indication

The service primitives used to pass a received frame to a client with the specified parameters.

6.2.3.1.4 State diagram

The architectural models of ISS and Link Aggregation do not make any provision for queuing of frames between the link and the Aggregator Client. However, practical implementations of Link Aggregation will typically incur both queuing and delay in the Frame Collector. In order to ensure that frame delivery is not delayed indefinitely (which could cause a frame ordering problem when moving conversations from one

link to another), the Frame Collector shall, upon receiving a frame from an Aggregator Parser, either deliver the frame to its Aggregator Client, or discard the frame within a CollectorMaxDelay time. The Frame Distributor (within the Partner System at the other end of the link) can assume that all frames transmitted on a given link have been either received by its Partner's Aggregator Client or discarded after a CollectorMaxDelay plus the propagation delay of the link. The use of CollectorMaxDelay is further discussed in B.3.

NOTE—Because frame discard due to CollectorMaxDelay is a function of queuing and other entities not specified in this document, the discard action is a requirement of the system, not just of the Frame Collector, and is therefore not shown in Figure 6-3.

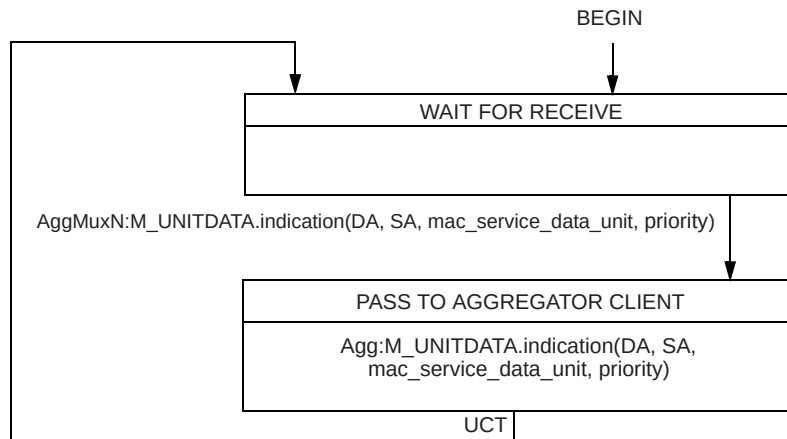


Figure 6-3—Frame Collector state diagram

6.2.4 Frame Distributor

The Frame Distributor is responsible for taking outgoing frames from the Aggregator Client and transmitting them through the set of links that form the LAG. The Frame Distributor implements a distribution function (algorithm) responsible for choosing the link to be used for the transmission of any given frame or set of frames.

This standard does not mandate any particular distribution algorithm(s); however, any distribution algorithm shall ensure that, when frames are received by a Frame Collector as specified in 6.2.3, the algorithm shall not cause the following:

- a) Misordering of frames that are part of any given conversation, or
- b) Duplication of frames.

The preceding requirement to maintain frame ordering is met by ensuring that all frames that compose a given conversation are transmitted on a single link in the order that they are generated by the Aggregator Client; hence, this requirement does not involve the addition (or modification) of any information to the MAC frame, nor any buffering or processing on the part of the corresponding Frame Collector in order to reorder frames. This approach permits a wide variety of distribution and load balancing algorithms to be used, while also ensuring interoperability between devices that adopt differing algorithms.

NOTE—The subject of distribution algorithms and maintenance of frame ordering is discussed in Clause 8 and Annex A.

The Frame Distributor shall implement the function specified in the state diagram shown in Figure 6-4 and the associated definitions contained in 6.2.4.1.

6.2.4.1 Frame Distributor state diagram

6.2.4.1.1 Variables

DA

SA

mac_service_data_unit

priority

The parameters of the M_UNITDATA.request primitive.

BEGIN

A Boolean variable that is set to TRUE when the System is initialized or reinitialized, and is set to FALSE when (re-)initialization has completed.

Value: Boolean

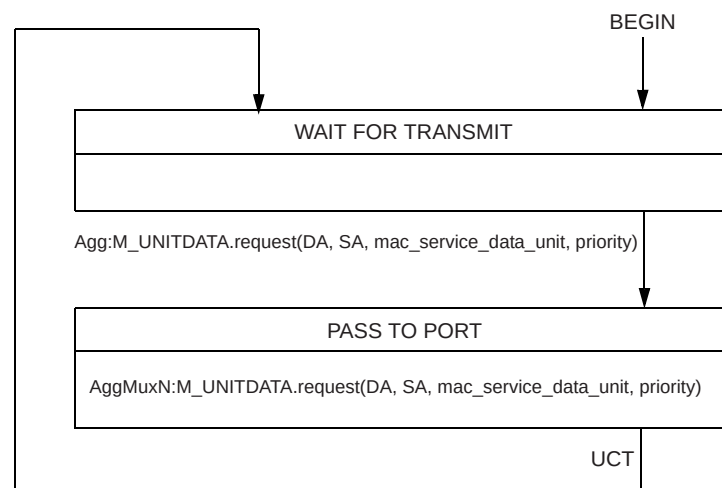
6.2.4.1.2 Messages

Agg:M_UNITDATA.request

AggMuxN:M_UNITDATA.request

The service primitives used to transmit a frame with the specified parameters.

6.2.4.1.3 State diagram



NOTE—The algorithm that the Frame Distributor uses to select the value of N in AggMuxN:M_UNITDATA.request for a given frame is unspecified.

Figure 6-4—Frame Distributor state diagram

6.2.5 Marker Generator/Receiver (optional)

The optional Marker Generator is used by the Marker protocol, as specified in 6.5. When implemented and so requested by the distribution algorithm, the Marker Generator shall issue an AggMuxN:M_UNITDATA.request primitive, with a mac_service_data_unit containing a Marker PDU as defined in 6.5.3, to the Aggregation Port associated with the conversation being marked, subject to the timing restrictions for Slow Protocols specified in IEEE Std 802.3-2008, Annex 57A.

The optional Marker Receiver is used by the Marker protocol, as specified in 6.5. It receives Marker Response PDUs from the Aggregator Parser.

6.2.6 Marker Responder

The Marker Responder is used by the Marker protocol, as specified in 6.5. The Marker Responder receives Marker PDUs (generated by a Partner System's Marker Generator), and transmits a Marker Response PDU through the same Aggregation Port from which the Marker PDU was received. While implementation of the Marker Generator/Receiver is optional, the ability to respond to a Marker PDU (the Marker Responder) is mandatory. An implementation conformant to this clause that is not under a Distributed Relay function (DR Function, see 9.2, 9.3) shall implement the Marker Responder as specified in 6.5.4.2, thus ensuring that implementations that need to make use of the protocol can do so.

A Marker Responder that is under a DR Function is not required to respond to a Marker PDU.

NOTE—Since this standard lacks any means for coordinating Marker PDU responses among the Portal Systems comprising a Portal, it is safer to not respond to a Marker PDU, and allow the sender to time out, rather than to return a Marker PDU and run the risk of delivering frames out of order.

6.2.7 Protocol Parser/Multiplexer

A Protocol Parser/Multiplexer is a function of which there are three examples in this standard, as follows:

- a) The LLDP Parser/Multiplexer (6.1.3)
- b) The Control Parser/Multiplexer (6.2.10)
- c) The DRCP Control Parser/Multiplexer (9.4.4)

A Protocol Parser/Multiplexer has three service interfaces, each an instance of the ISS, as follows:

- d) One *DownPort* makes use of an instance of the ISS to pass data and control frames to and from lower layers in the protocol stack.
- e) One *DataPort* offers an instance of the ISS to a higher level data function.
- f) One *ControlPort* offers an instance of the ISS to higher level control functions.

The specification of a given instance of the Protocol Parser/Multiplexer, e.g., the Control Parser/Multiplexer (6.2.10), ties these three notional service interfaces to specific service interfaces, e.g., those in 6.2.2, and provides a definition of the *IsControlFrame* function ().

On transmission, the Protocol Multiplexer shall provide transparent pass-through of frames from the ControlPort or the DataPort to the DownPort.

On receipt of a frame from its DownPort, the Protocol Parser uses the *IsControlFrame* function to determine whether the frame is or is not a control frame. It passes control frames up through the ControlPort, and passes all other frames up through the DataPort. The Protocol Parser shall implement the function specified by the state diagram shown in Figure 6-5 and the associated definitions contained in 6.2.7.1.

6.2.7.1 Protocol Parser state diagram

6.2.7.1.1 Functions

IsControlFrame

Given an input frame (DA, SA, mac_service_data_unit, priority), returns TRUE if the frame is a control frame for the ControlPort, or FALSE if it is a data frame for the DataPort.

Value returned: Boolean

6.2.7.1.2 Variables

DA

SA

mac_service_data_unit

priority

The parameters of the M_UNITDATA.indication primitive.

BEGIN

A Boolean variable that is set to TRUE when the System is initialized or reinitialized, and is set to FALSE when (re-)initialization has completed.

Value: Boolean

6.2.7.1.3 Messages

ControlPort:M_UNITDATA.indication

DataPort:M_UNITDATA.indication

DownPort:M_UNITDATA.indication

The service primitives used to pass a received frame to a client with the specified parameters.

6.2.7.1.4 State diagram

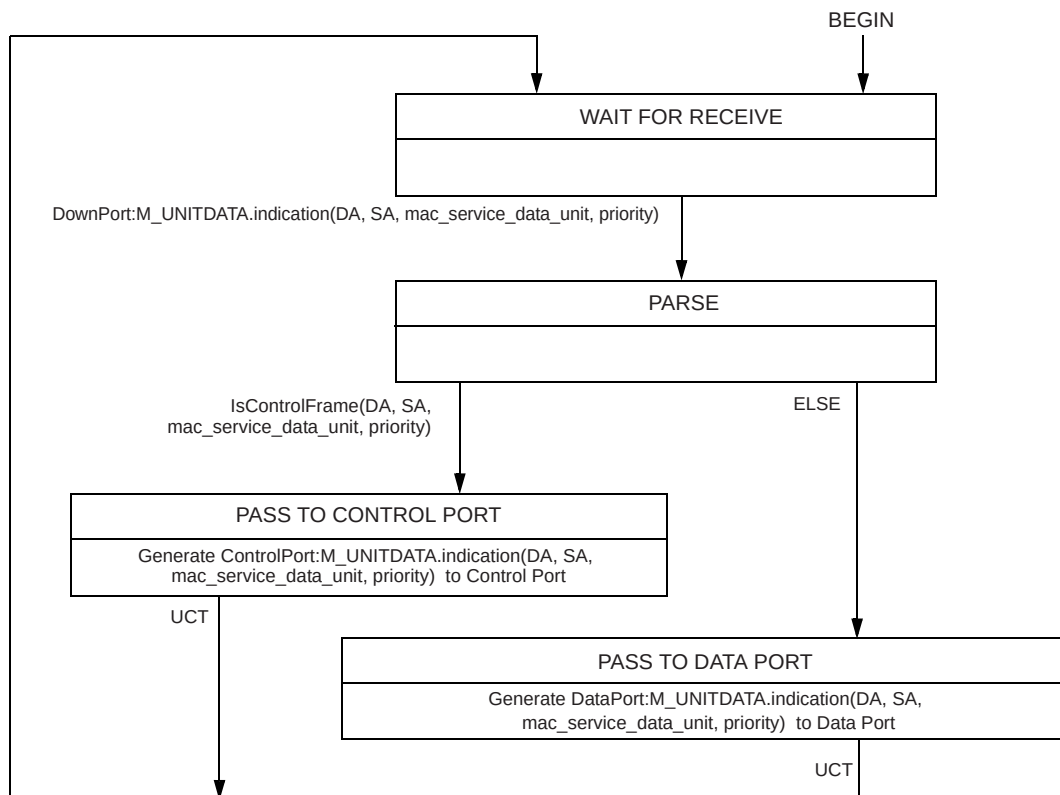


Figure 6-5—Protocol Parser state diagram

6.2.8 Aggregator Parser/Multiplexer

On transmission, the Aggregator Multiplexer shall provide transparent pass-through of frames submitted by the Marker Responder and optional Marker Generator to the Aggregation Port specified in the transmission request. The Aggregator Multiplexer shall provide transparent pass-through of frames submitted by the Frame Distributor to the Aggregation Port specified in the transmission request only when the Aggregation Port state is Distributing (see 6.4.15); otherwise, such frames shall be discarded.

On receipt, the Aggregator Parser decodes frames received from the Control Parser, passes those frames destined for the Marker Responder or Marker Receiver to the selected entity, and discards frames with invalid Slow Protocol subtype values (see IEEE Std 802.3-2008, Table 57A–2). The Aggregator Parser shall pass all other frames to the Frame Collector for passage to the Aggregator Client only when the Aggregation Port state is Collecting (see 6.4.15); otherwise, such frames shall be discarded. The Aggregator Parser shall implement the function specified in the state diagram shown in Figure 6-6 and the associated definitions contained in 6.2.8.1.

6.2.8.1 Aggregator Parser state diagram

6.2.8.1.1 Constants

Slow_Protocols_Type

The value of the Slow Protocols Length/Type field. (See IEEE Std 802.3-2008, Annex 57A.)

Marker_subtype

The value of the Subtype field for the Marker protocol. (See 6.5.3.)

Value: Integer

2

Marker_Information

The encoding of the Marker Information TLV_type field. (See 6.5.3.)

Value: Integer

1

Marker_Response_Information

The encoding of the Marker Response Information TLV_type field. (See 6.5.3.)

Value: Integer

2

6.2.8.1.2 Variables

DA

SA

mac_service_data_unit

priority

The parameters of the M_UNITDATA.indication primitive.

Protocol_DA

One of the addresses selected from Table 6-1 determined by the setting of the aAggPortProtocolDA managed object (7.3.2.2.1). A particular instance of link aggregation shall use the same DA for LACP as it does for the Marker protocol.

Value: 48 bits.

Length/Type

The value of the Length/Type field in a received frame.

Value: Integer

Subtype

The value of the octet following the Length/Type field in a Slow Protocol frame.
(See IEEE Std 802.3-2008, Annex 57A.)

Value: Integer