

International Telecommunication Union

**ITU-T**

TELECOMMUNICATION  
STANDARDIZATION SECTOR  
OF ITU

**G.998.2**  
**Amendment 2**  
(12/2007)

SERIES G: TRANSMISSION SYSTEMS AND MEDIA,  
DIGITAL SYSTEMS AND NETWORKS

Digital sections and digital line system – Access networks

---

Ethernet-based multi-pair bonding

**Amendment 2**

Recommendation ITU-T G.998.2 (2005) –  
Amendment 2



ITU-T G-SERIES RECOMMENDATIONS  
TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100–G.199
GENERAL CHARACTERISTICS COMMON TO ALL ANALOGUE CARRIER-TRANSMISSION SYSTEMS	G.200–G.299
INDIVIDUAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON METALLIC LINES	G.300–G.399
GENERAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON RADIO-RELAY OR SATELLITE LINKS AND INTERCONNECTION WITH METALLIC LINES	G.400–G.449
COORDINATION OF RADIOTELEPHONY AND LINE TELEPHONY	G.450–G.499
TRANSMISSION MEDIA AND OPTICAL SYSTEMS CHARACTERISTICS	G.600–G.699
DIGITAL TERMINAL EQUIPMENTS	G.700–G.799
DIGITAL NETWORKS	G.800–G.899
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900–G.999
General	G.900–G.909
Parameters for optical fibre cable systems	G.910–G.919
Digital sections at hierarchical bit rates based on a bit rate of 2048 kbit/s	G.920–G.929
Digital line transmission systems on cable at non-hierarchical bit rates	G.930–G.939
Digital line systems provided by FDM transmission bearers	G.940–G.949
Digital line systems	G.950–G.959
Digital section and digital transmission systems for customer access to ISDN	G.960–G.969
Optical fibre submarine cable systems	G.970–G.979
Optical line systems for local and access networks	G.980–G.989
<b>Access networks</b>	<b>G.990–G.999</b>
QUALITY OF SERVICE AND PERFORMANCE – GENERIC AND USER-RELATED ASPECTS	G.1000–G.1999
TRANSMISSION MEDIA CHARACTERISTICS	G.6000–G.6999
DATA OVER TRANSPORT – GENERIC ASPECTS	G.7000–G.7999
PACKET OVER TRANSPORT ASPECTS	G.8000–G.8999
ACCESS NETWORKS	G.9000–G.9999

*For further details, please refer to the list of ITU-T Recommendations.*

# **Recommendation ITU-T G.998.2**

## **Ethernet-based multi-pair bonding**

### **Amendment 2**

#### **Summary**

Amendment 2 to Recommendation ITU-T G.998.2 includes new functionality for discovery/aggregation and pair management at the bonding layer.

#### **Source**

Amendment 2 to Recommendation ITU-T G.998.2 (2005) was approved on 7 December 2007 by ITU-T Study Group 15 (2005-2008) under Recommendation ITU-T A.8 procedure.

## FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

## NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g. interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

## INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2008

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

## CONTENTS

	<b>Page</b>
Annex C – Frame-based protocol for discovery/aggregation and link management.....	2
C.1    Bonding control architecture .....	3
C.2    Operational overview of bonding aggregation control protocol .....	4
C.3    Protocol operation .....	7
C.4    Bonding aggregation control protocol frame formats .....	15



## Recommendation ITU-T G.998.2

### Ethernet-based multi-pair bonding

#### Amendment 2

This amendment includes new functionality for discovery/aggregation and pair management at the bonding layer. New Annex C describes a frame-based protocol (the bonding aggregation control protocol (BACP)) that exchanges information allowing two systems to seamlessly control mapping of physical media entities to physical medium attachment aggregation functions.

*In clause 2, add the following reference to IEEE 802.3-2005, the standard that replaced IEEE 802.3 (2002) and IEEE 802.3ah (2004). Original references to IEEE 802.3 (2002) and IEEE 802.3ah (2004) have been retained and, when explicitly identifying the 2005 version, this amendment refers to IEEE 802.3-2005:*

- [11] IEEE 802.3 (2005), *Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications.*

NOTE – Original references to IEEE 802.3 (2002) and IEEE 802.3ah (2004) have been retained and, when explicitly identifying the 2005 version, this Recommendation refers to IEEE 802.3-2005.

*Amend the definition of physical media entity:*

**3.10 physical media entity (PME):** In IEEE 802.3-2005, a reference to an individual transceiver contained within an aggregate group where a transceiver carries only one packet stream. This Recommendation also supports operation over transceivers carrying multiple (at gamma interface logically separated) packet streams. In this Recommendation, a PME is a reference to one logical channel out of a set of one or more channels on the same individual transceiver. Each such PME carries a single packet stream and is contained within one aggregate group.

NOTE 1 – In case pre-emption is used on a bearer channel, two PMEs each reference one of the two logical channels on the same bearer channel. In case pre-emption is not used on a bearer channel, a PME defaults to a reference to the one bearer channel.

NOTE 2 – Although such a PME cannot be physically isolated within a transceiver carrying multiple packet streams, the naming PME is maintained for consistency with IEEE 802.3-2005 and for consistency in case the transceiver carries only one packet stream.

*Add the following text to the end of clause 6.2.2:*

The network side uses the PME\_ID of each link in the bonding group to retrieve management data through Ethernet OAM (clause 30/57 of IEEE 802.3). Each G.994.1 CLR message shall include the 48-bit remote discovery register and the 32-bit PME aggregate register. The network side may derive the PME\_IDs of each of the links in the bonding group by adding the links one-by-one to the bonding group during PME aggregation, and observing the change in the PME aggregate register with each addition. However, to allow parallel initialization of the links in the bonding group and, hence, to allow shortening of the bonding group initialization time, the remote side should also include the 5-bit PME\_ID in each CLR message. Corresponding G.994.1 codepoints are defined for the PME\_ID as a supplement to IEEE 802.3 clause 61.

*Replace clause 8 with the following:*

#### **8 Handshake**

Clause 61.4 of IEEE 802.3-2005 defines the handshake operation for the IEEE standards 2BASE-TL and 10PASS-TS. That clause defines handshake procedures for discovering which pairs

are provisioned in which aggregate groups using the PME aggregation and PME aggregation discovery codepoints of the IEEE handshake tree.

The procedures of that clause are followed in this Ethernet bonding Recommendation with the following exceptions.

- There is no "PAF enable" parameter as described in clause 61.4 of IEEE 802.3-2005 [11]. That parameter is used to define whether bonding is performed. As this Recommendation references bonded operation only, that parameter is not required.
- The "PME identification" Spar(2) bit indicates that PME identification is included in the CLR message. The corresponding Npar(3) field contains the five-bit value (0 to 31) of the PME ID.
- The "BACP support" NPar(2) bit indicates support for the BACP protocol specified in Annex C. If this bit is set to ZERO in the CLR or CL message, then the bit shall also be set to ZERO in the MS message. If this bit is set to ONE in both the CLR and CL message, then this bit shall also be set to ONE in the MS message.
- Table 2 describes the meaning of the "BACP support" bit in the MS message.

**Table 2 – Interpretation of MS message bits for BACP**

<b>BACP support</b>	<b>Meaning</b>
0	Use the G.994.1-based discovery/aggregation. Do not use BACP discovery/aggregation/management
1	Use BACP for discovery/aggregation/management. Do not use G.994.1-based discovery/aggregation

Unlike the handshake operations of IEEE 802.3-2005 [11], this Recommendation defines the bonding layer to operate independently of the TPS-TC. Each line may select its TPS-TC independently of the others, as long as the maximum differential latency of the pairs is maintained as specified in this Recommendation.

NOTE – This Ethernet bonding Recommendation uses the PME aggregation and PME aggregation discovery codepoints in ITU-T Rec. G.994.1. The definition of these variables is in clause 45.2.3.20 (PME aggregation) and 45.2.6.8 (PME aggregation discovery) of IEEE 802.3-2005. The transactions used to exchange these codepoints (e.g., "Get", "Set if Clear", "Clear if Same") are defined in clause 61.4 of IEEE 802.3-2005. Note that an example of the aggregation discovery procedure is contained in clause 61A of IEEE 802.3-2005.

*Add the following new Annex C:*

## **Annex C**

### **Frame-based protocol for discovery/aggregation and link management**

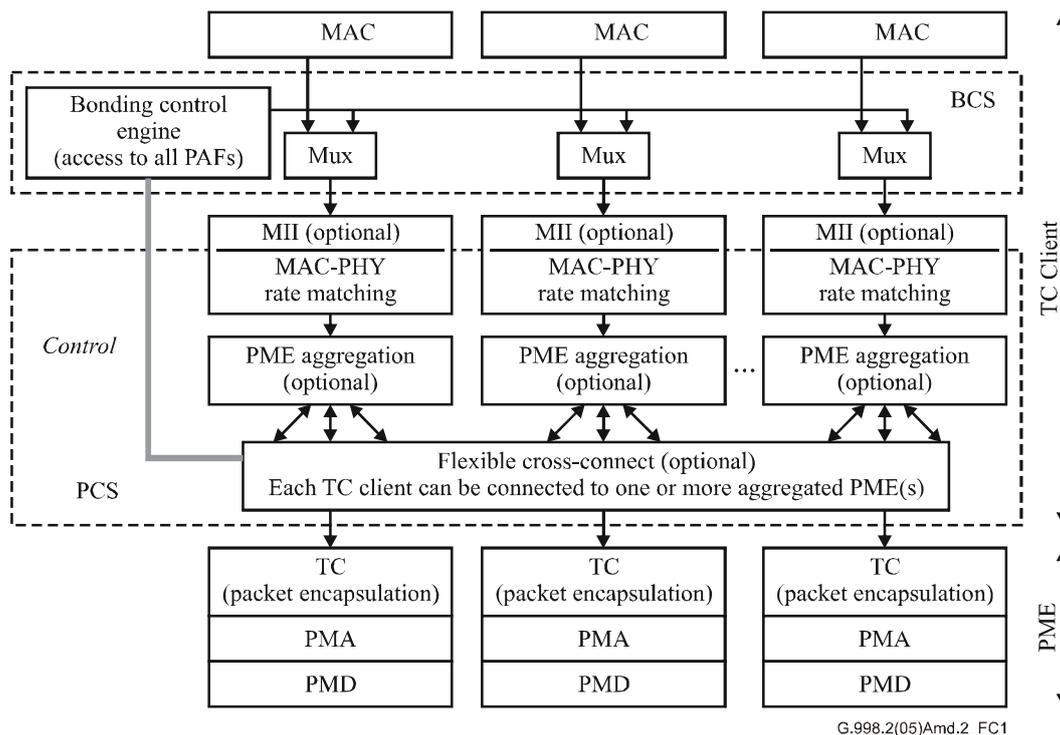
This annex describes a frame-based protocol, referred to as the bonding aggregation control protocol (BACP), that exchanges information allowing two systems to control which PMEs are mapped to which PAFs in a simple and seamless manner. In this annex, a "system" is defined as a piece of equipment, such as a DSLAM or a CPE device, associated with a set of PAFs and PMEs. This protocol is used in lieu of the G.994.1 discovery/aggregation process defined in clause 8.

This protocol is backward compatible with existing Ethernet bonding implementations and standards.

### C.1 Bonding control architecture

The bonding control architecture is shown in Figure C.1. A bonding control engine (BCE) sends and receives BACP data units (BACPDUs). The BACPDUs are Ethernet frames that are transmitted and received through a standard PAF function. A mux function between the MAC and MII interface (see Note) is required to insert and remove BACPDUs into the data path. BACPDUs are identified by the headers of the Ethernet frame as discussed in clause C.4.

NOTE – Architecturally the BCE is in a bonding control sublayer (BCS) similar to the Ethernet OAM sublayer shown in Figure 57-3 of IEEE 802.3-2005.



**Figure C.1 – Bonding control architecture**

The bonding control engine uses BACPDUs to exchange information with a BCE on a remote system. The information in these frames is then used by the BCE to control the flexible cross-connect shown in Figure C.1, mapping specific PMEs to specific PAFs. All PMEs on which BACP is enabled are 'PAF enabled', meaning that the PAF function is always in use, and the PME carries fragments rather than full Ethernet frames. This ensures that the BCE simply controls the flexible cross-connect, and does not have to change encapsulations on the PMEs.

Because the BCE essentially moves PMEs from one PAF to another, systems that implement BACP should be careful in that moving PMEs from PAFs that are carrying user (higher layer) data, such as an Internet service, since they would have a clear effect on that service. Implementations should take care in protecting higher layer services as the protocol operates and PME aggregation or removal occurs.

In order to achieve seamless insertion and removal of PMEs into a bonded group, the cross-connect and PAF must support the ability to separately control the receive and transmit connectivity

between a PME and a PAF. For example, it must (temporarily) be possible for a PME to be mapped to a PAF receive (reassembly) process, but not to the PAF transmit (segmentation) process.

It is also assumed that the BCE has knowledge of when PMEs are activated and deactivated. Once a PME is activated (i.e., able to transmit and receive fragments), BACP can be used to control to which PAF the PME is connected, and provide smooth transitions of a PME into or out of a PAF.

Because BACP uses standard Ethernet frames transmitted through the PAF, the BCE does not control over which PME(s) the BACPDU's are sent. Thus BACP cannot be used for per-PME link monitoring.

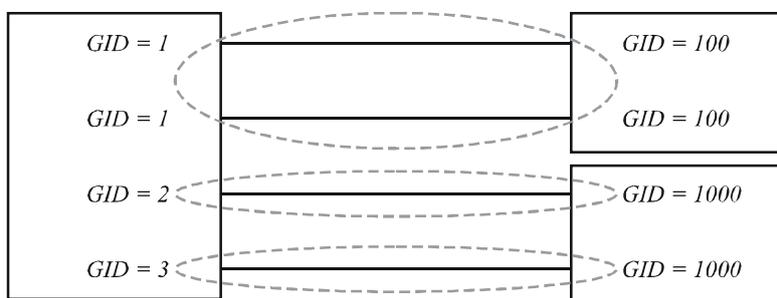
## C.2 Operational overview of bonding aggregation control protocol

This clause provides an overview of the operational behaviour of the BCE and BACP protocol. Additional details are supplied in later clauses.

### C.2.1 Group identification and bonding eligibility

In order to determine which PMEs are eligible to be aggregated with other PMEs, a system assigns a group ID (GID) to each PME. The GID parameter is specified in clause C.3.1.1.

Two PMEs are eligible to be aggregated if they have the same GID on both systems to which they are attached. Specifically, if the GIDs associated with PME1 on the local system and remote systems are G1local and G1remote, respectively, and if the GIDs associated with PME2 are G2local and G2remote, then PME1 and PME2 are eligible to be aggregated if G1local = G2local and G1remote = G2remote. In the typical environment, a CPE device would assign the same GID to all of the PMEs attached to it, in order to assert that all of its PMEs should be in the same bonded group.



G.998.2(05)Amd.2\_FC2

**Figure C.2 – Group eligibility example**

An example of determining group eligibility is given in Figure C.2. In this example, the top two PMEs both have GIDs of (1, 100), and are thus eligible to be aggregated. The bottom two PMEs have unique GIDs ((2, 1000) and (3, 1000)) and are thus not eligible for aggregation with any of the other PMEs.

If each system uses a single GID across all PMEs, then all PMEs between the systems will be eligible to be aggregated together. Alternatively, one could configure "fixed" aggregation groups (most likely on a CO system), each with their own GID. For example, PMEs 1-4 could be given one GID, PMEs 5-8 a different GID, and PMEs 9-15 yet a different GID. The system would then ensure that no aggregation can occur across PMEs in these three different groups based on this GID assignment policy. These are expected to be two common methods of using the GIDs.

Management operations can be used to configure group eligibility by means of the GID based on local policy and configuration.

GIDs define *group eligibility*. GIDs do not define actual bonding groups. Group make-up is determined by the current connectivity of the cross-connect and the PME status. PMEs with different GIDs shall never be connected to the same PAF. PMEs with the same GID may be connected to different PAFs.

### **C.2.2 PME identification**

The PMEs require identification within 2 different contexts for BACP: within the Group ID context (eligibility to be a member of a bonding group), and within the PAF context (role as actual member of a bonding group).

The PME identification (PME ID) parameter identifies the PME within the context of a particular PAF. It is specified in clause C.3.1.3. PME identification is required for multiple reasons, e.g., when adding/removing PMEs from an aggregate it is vital that the correct PME be identified for that operation. In case the implementation supports IEEE 802.3 clause 57 OAM, and when the implementation desires to retrieve statistics or status on a particular PME within an aggregate, the PME ID is required to access the information on a particular PME. In this context, the PME ID is an integer in the range 0..31, and identifies which of the 32 potential PMEs in this aggregate the OAM request is targeting.

The stream ID is an identifier that is unique within a group ID (i.e., PMEs with the same GID). The stream ID parameter is specified in clause C.3.1.2. Each PME is assigned a stream ID that uniquely identifies that PME within that group ID. The use of the stream ID allows group eligibility to extend beyond 32 streams. This is important in the following respect. Although any PAF is limited by this Recommendation (and IEEE 802.3) to at most 32 streams in a PAF, many implementations in the market allow those 32 streams to be selected from a much larger set. For example, a line card may have 48 ports and any of those ports could be aggregated with any other (up to at most 32 in any one PAF). Or a system may support hundreds of ports with a centralized bonding capability such that any stream can be aggregated with any other. Generalized discovery across a large number of potential streams is supported in BACP via the combination of the group and stream identifiers. This is useful for systems that are supposed to be plug-and-play interconnectable in a true grid network (e.g., on enterprise or campus scale) and where group eligibility is not restricted through network operator configuration to a maximum of 32 pairs.

### **C.2.3 PME and PAF initialization**

PMEs are initialized as the only PME in their PAF. When the PAF is able to transmit and receive fragments, the bonding control engine can begin the transmission of BACPDU s that advertise the local GID, PME ID and stream ID for that PME through the PAF for that PME. BACPDU s are transmitted through every PAF on which the protocol is enabled as specified in clause C.3.2.5. When the BCE receives a BACPDU on a single-PME PAF, it knows the GID, PME ID and stream ID for that PME as defined by the remote system. This allows the local BCE to determine the group eligibility for that PME (e.g., can this PME be aggregated into some other PAF). This absolute PME identification mechanism requires that the size of a newly initialized PAF be limited to a single PME.

After determining eligibility, the BCE may decide to initiate a PME transfer to another PAF. A PME transfer allows a PME in a single PAF to be transferred seamlessly to another PAF. The BCE is not required to initiate a PME transfer, and can reject or postpone a transfer request from a remote system.

For example, a BCE may decide not to initiate a PME transfer because the desired target PAF already has a sufficient number of PMEs. This may be because of a physical limit of the PAF

(e.g., it only supports 4-pair bonding and already has 4-pair in it), or because of a logical limitation (it was provisioned to only allow 4-pair in a PAF and the target PAF has 4-pair already). Other reasons a BCE may not initiate a PME transfer request after determining that the PME is eligible include that the PME is not of sufficient quality (too many errors), the PME is not of sufficient bandwidth (would violate the 4x rate differential), or that differential delay with other PMEs in the target PAF would be too great (violate the differential delay requirements of a bonded group).

It is up to the local BCE policies to determine if a BCE transfer should be initiated, and up to the remote BCE policies to accept and progress a transfer request.

#### **C.2.4 Transferring a PME from a single-PME PAF to an existing PAF**

Transferring a PME from one PAF to another PAF is a seamless process in which no data loss is incurred on the target PAF, but which can only be executed for a source PAF that is a single-PME PAF. The source PAF will of course lose its only PME. The source PAF PME is expected not to be carrying data traffic, but if it is, then that data traffic will all be dropped once the PME transfer begins.

The following process outlines the procedure for transferring a PME from its single PAF to another PAF that has one or more existing PMEs. Within the scope of the source and target PAF, PMEs and streams have a one-to-one mapping. By transferring a PME, only the one related stream is transferred. Other streams carried on the same transceiver but aggregated through other PAFs remain unaffected by this process.

- 1) The BCE decides to initiate a transfer of PME A in PAF X (a single-PME PAF) to PAF Y (which already has one or more PMEs).
- 2) The BCE selects a PME ID for the PME A within the target PAF Y. The BCE then transmits a BACPDU indicating the selected PME ID (as well as other parameters) through the target PAF Y. The BCE waits for the far end to acknowledge the local PME ID selection, and acknowledges the far end PME ID selection. At this point, both systems have selected a PME ID that is applicable to the target PAF Y, and both systems have received the selected PME ID from the remote system.
- 3) When the PME IDs have been exchanged and confirmed through PAF Y, the BCE disconnects PME A from its current PAF X. At this point, no frames will be transmitted through PAF X. This is signalled to the remote system via a BACPDU transmitted through PAF Y, and the local system waits for the confirmation that the remote system also has disconnected PME A.
- 4) Once both sides have disconnected PME A from the transmit path of any PAF, the receive side can be switched to PAF Y. When the local system has switched the receive path to PAF Y, it signals this status change via a BACPDU transmitted through PAF Y. At this point, PME A is ready to receive traffic on PAF Y, but neither side is transmitting on it.
- 5) After the local system receives confirmation that PME A is connected to the receive path of PAF Y on the remote system, it is free to connect PME A to the transmit path of PAF Y.

At the end of this process, PME A is connected to PAF Y both locally and remotely. And because the receive path was activated before the transmit path was activated, there was no data loss on PAF Y because of the addition of PME A.

#### **C.2.5 Transferring a PME from an existing PAF to a single-PME PAF**

Transferring a PME from a PAF with multiple PMEs is basically the reverse of the earlier process for adding a PME. The PME is transferred from its existing PAF to a PAF in which it is the only PME.

Transferring a single PME out of an existing PAF is a seamless process in which no data loss is incurred on the current PAF if the offered traffic rate is supported by the remaining PMEs. Within the scope of the source and target PAF, PMEs and streams have a one-to-one mapping. By transferring a PME into a newly created PAF, only the one related stream is transferred. Other streams carried on the same transceiver but aggregated through other PAFs remain unaffected by this process.

The following process outlines the procedure for removing a PME from an aggregate PAF to its single PAF:

- 1) The BCE decides to initiate a transfer of PME A out of PAF Y (which has two or more PMEs) to an empty (no PME) PAF (PAF X).
- 2) The BCE disconnects PME A from the transmit path of PAF Y, and signals this status change to the remote BCE on PAF Y. The PME A remains connected for the receive path of PAF Y.
- 3) When the BCE receives confirmation from the remote side that it also has disconnected the transmit path for PME A in PAF Y and that the remote side has received the notification sent by the local BCE, the BCE disconnects the receive side of PME A from PAF Y and connects the transmit and receive paths of PME A to an empty PAF (PAF X).
- 4) The BCE signals the remote BCE on PAF X that the transmit and receive paths have been connected to PAF X.
- 5) When the BCE receives confirmation that the remote side has also connected PME A to PAF X, and that the remote side has received the notification sent by the local BCE, then PME A has been successfully transferred into its own PAF.

NOTE – If the transfer was initiated because the PME has excessive errors on the related stream, such confirmation may not be received. If the transceiver is to be reinitialized, other streams carried on the same transceiver should be transferred to a single-PME PAF first to maintain seamlessness.

At this point, PME A has been seamlessly removed from PAF Y and transferred into its own single-PME PAF X.

### **C.3 Protocol operation**

The following clauses describe BACP operation in more detail.

#### **C.3.1 Protocol variables and attributes**

This clause identifies the primary variables and attributes of BACP. These attributes control the operation of the protocol, and the variables are communicated via BACPDUs. These parameters are necessary only within the context of this protocol. PMEs that are not eligible to be bonded, or PMEs on which this protocol is not enabled, do not require these parameters to be configured.

##### **C.3.1.1 Group ID (GID)**

The group ID is a universally unique 6-octet identifier assigned by each system that identifies a set of PMEs which are eligible to be aggregated on the local system. When two PMEs have the same GID on a system, they are eligible to be aggregated on that system.

Clause C.2.1 describes how GIDs can be used to determine group eligibility. This value is initialized by management.

The GID of a PME shall not be changed while that PME is aggregated with any other PME. The GID may be changed while the PME is in a single-PME PAF, either during BACP initialization (clause C.3.2.1) or after a PME has been transferred out of a bonded group (clause C.3.2.3).

If a transceiver carries multiple logically separated data streams over the gamma reference point (e.g., in the case of dual latency operation), then each of those streams shall be assigned to a different group ID.

### C.3.1.2 Stream ID

As discussed in clause C.2.2, the Stream ID is a 2-octet identifier that uniquely identifies a particular PME within a group ID (i.e., PMEs with the same GID). A group + stream ID is required for each PME under the control of this BCE. Each stream is assigned a stream ID that uniquely identifies that stream within its group ID, so the combination of GID and stream ID is a universally unique identifier for that PME. The allocation of the stream ID values is implementation-specific and no particular structure is specified.

The stream ID shall not be changed while that PME is aggregated with any other PME. The stream ID may be changed while the PME is in a single-PME PAF, either during BACP initialization (clause C.3.2.1) or after a PME has been transferred out of a bonded group (clause C.3.2.3).

If a transceiver carries multiple logically separated data streams over the gamma reference point (e.g., in the case of dual latency operation), then each of those streams may be assigned the same or a different stream ID.

This value is initialized by management.

### C.3.1.3 PME ID

The PME ID is an integer in the range 0..31, and identifies a particular PME out of the 32 potential PMEs in a PAF. This is the same value that would be used by IEEE 802.3 clause 57 OAM to retrieve statistics or status on a particular PME within an aggregate. This relative PME identifier (e.g., an identifier that has context only within a particular PAF) is referred to as the PME ID.

This value is initialized by the BCE when the PME is selected to be mapped or moved to a specific PAF, and this value identifies this PME within the PAF over which the BACPDU is sent or received.

The allocation of the PME ID values is implementation-specific and no particular structure is specified.

If a transceiver carries multiple logically separated data streams over the gamma reference point (e.g., in the case of dual latency operation), then each of those streams may be assigned the same or a different PME\_ID.

### C.3.1.4 PME status

Each of the 32 potential PME positions within a PAF (representing PME IDs 0 to 31) has an associated PME status value. The valid PME status values are:

- **Unknown (0):** Information about this PME is not yet known.
- **Unassigned (1):** The PAF does not have a PME in this position.
- **Assigned (2):** The PME has been assigned a PME ID for the PAF but is not connected to the PAF in either the transmit or receive direction.
- **Moving (3):** The PME is in the process of moving into this PAF. It is currently not connected to any PAF on the local side, and BCE is awaiting confirmation that the PME is not connected to any PAF on the remote side.
- **RxOnly (4):** The PME is connected to the advertising PAF reassembly process, but is not connected to the transmit path of any PAF.
- **TxRx (5):** The PME is fully connected to the transmit and receive paths of the advertising PAF. This is the normal fully operational status.

The local PME status shall not take the value "Unknown". Only the remote PME status may take the value "Unknown" when the local BCE has not yet received information from the remote BCE on the PAF, which is then to be resolved through a TLV exchange.

The PME status is carried in a 4-bit field as an unsigned integer (e.g., RxOnly(4) is carried as 0100).

### C.3.1.5 PME status array

The PME status is advertised as an array of 32 4-bit fields (16-octets) in the PME status array field of the local info TLV. The PME ID for a specific PME is used as the index into this array to identify the status of that specific PME.

0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1
PME 0	PME 1	PME 2	PME 3	PME 4	PME 5	PME 6	PME 7	PME 8	PME 9	PME 10	PME 11	PME 12	PME 13	PME 14	PME 15	PME 16	PME 17	PME 18	PME 19	PME 20	PME 21	PME 22	PME 23	PME 24	PME 25	PME 26	PME 27	PME 28	PME 29	PME 30	PME 31

**Figure C.3 – PME status array**

### C.3.1.6 Remote group ID

Each PAF has a remote group ID. The remote GID is the GID value in the local info TLV most recently received from the remote BCE on that PAF. This value is included in the remote group ID field of the remote info TLV in each BACPDU transmitted on that PAF.

### C.3.1.7 Remote stream ID

Each stream under the control of the BCE has an associated remote stream ID. The remote stream ID is the stream ID value in the assignment TLV from the remote BCE for this stream as discovered during the Initialization process. When an assignment TLV is received from a remote BCE, the association between the remote stream ID and local stream ID is stored. The BCE must be able to associate the PME ID received in that assignment TLV and the PAF over which it was received with that remote stream ID.

Thus, from just its local stream ID, the BCE can associate a remote stream ID, the PME ID of that stream as identified by the remote system, and the PAF in which that stream is aggregated.

#### C.3.1.7.1 Remote PME ID

Each stream under the control of the BCE has an associated remote PME ID. The remote PME ID is the PME ID value in the assignment TLV most recently received from the remote BCE for this stream. When an assignment TLV is received from a remote BCE, the association between the remote PME ID and local stream ID is stored.

### C.3.1.8 Remote PME status array

Each PAF has an associated remote PME status array. This array reflects the status of the PMEs as last communicated by the remote BCE and received by the local BCE. The remote PME ID for a specific PME is used as the index into the remote PME status array to identify the remote status of that specific PME.

## C.3.2 Protocol operation

### C.3.2.1 BACP initialization

Each PME shall be initialized into a PAF with only that PME. The local BCE has no preceded information about the remote BCE, and BACP itself provides the discovery process.

BACP initialization is described below. At the completion of the BACP initialization phase, the local BCE knows the stream ID, group ID and PME ID used by the remote BCE for each PME in the initialized PAF. At the completion of the BACP initialization, data flow can begin immediately on the PAF.

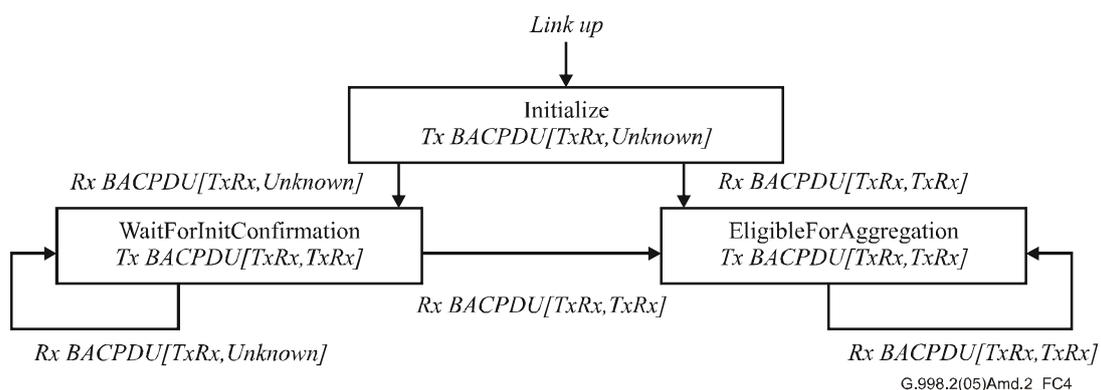
The BACP initialization process happens separately for each PME. There is at most one initialization happening on a PAF because the PMEs are initialized as a single-PME PAF, and multiple PAFs may be initialized in parallel.

A PME is initialized into its own PAF as the only PME in that PAF. The PME's initial local status is TxRx and its remote status is Unknown with that PAF and BACPDU's can be sent through the PAF over that single PME. The BCE transmits BACPDU's over that PAF until it receives BACPDU's from the remote entity that confirm it has received the local information. Each BACPDU frame contains the local information to be communicated to the remote BCE, as well as a copy of the latest received information from the remote BCE, so that the remote BCE can confirm that the local BCE has received its information.

Whenever there is a PME in the Initialize or WaitForInitConfirmation state, the assignment TLV must be included in the BACPDU for that PME. The local TLV includes the group ID and the assignment TLV includes a stream ID and PME ID for the single PME. The group ID and stream ID are critical to the protocol operation as their combination is the only constant identifier for this stream in a multi-PME PAF. While a PME is being moved from one PAF to another (as defined in clauses C.3.2.2 and C.3.2.3), neither the group ID nor the stream ID is allowed to change as they are used to identify which PME is being moved. Only while a PME is part of a single-PME PAF, may its group ID and stream ID be changed (resulting in sending TLVs to notify the far-end BCE).

Whenever an assignment TLV is received in a single-PME PAF, the BCE must store the relationship between the stream ID and that PME. The stream ID might be received in BACPDU's on any PAF, and the local BCE must be able to associate the stream ID with this PME.

In the following state machines, [X,Y] is used to represent the local information about the PME status, where X is the current local status of the PME, and Y is the most recently received status of the PME on the far end. For example, [TxRx, Unknown] represents when the local system has the PME connected to its PAF in both the transmit and receive directions, and it has not yet heard the status of the PME on the remote BCE.



**Figure C.4 – Single-PME PAF state machine**

The operation  $Tx \text{ BACPDU}[TxRx, Unknown]$  represents that the local BCE is transmitting BACPDU's with the PME status information [TxRx,Unknown] for the single PME in this PAF. In the BACPDU, other local PME status values are set to Unassigned and other remote PME status

values are also set to Unknown. The transitions  $Rx\ BACPDU[TxRx, TxRx]$  occur when the local BCE receives a BACPDU with the PME status information  $[TxRx, TxRx]$ .

The target state for this phase is for the PME to be in the EligibleForAggregation state. At this point, the BCE is free to transition the PME into another PAF.

### C.3.2.2 PME transfer into a bonded group

Transferring a PME into a bonded group consists of taking the PME through a series of phases, where each phase is complete when both the local and remote sides have confirmed, and received confirmation, that the status transition has been accomplished.

All referenced BACPDU transmissions occur on the target PAF to which the PME is being transferred.

#### C.3.2.2.1 Assigning a PME ID for the stream on the target PAF

Once the BCE decides to transfer a PME to a specific target PAF, the PME ID for that PME within the target PAF must be assigned and communicated with the remote BCE.

Whenever a BCE transmits a BACPDU with the local PME status of Assigned, an assignment TLV must also be included. This TLV contains the stream ID of the transferring PME as well as the relative PME ID for that PME with this target PAF.

Whenever a BCE receives an assignment TLV, it stores the relationship between the stream ID and PME ID as indicated by the remote BCE. The inclusion of the stream ID by the remote BCE allows the local BCE to know which PME the remote BCE is trying to transfer to this PAF. The local BCE can then find that PME and continue the PME transfer process if desired.

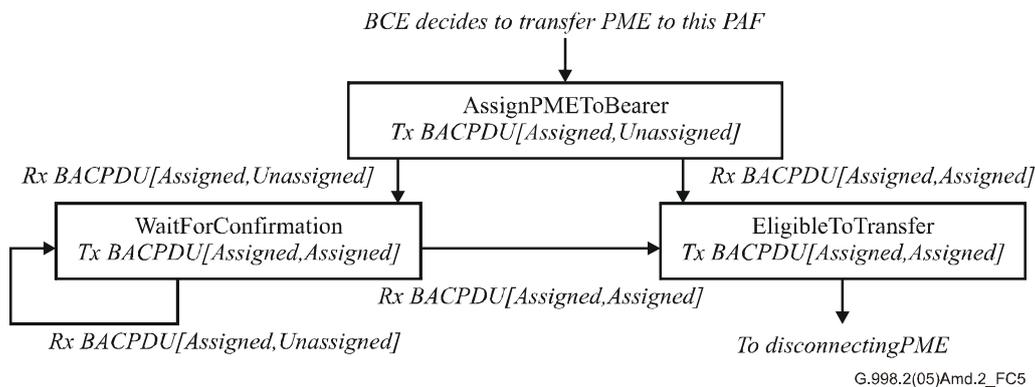
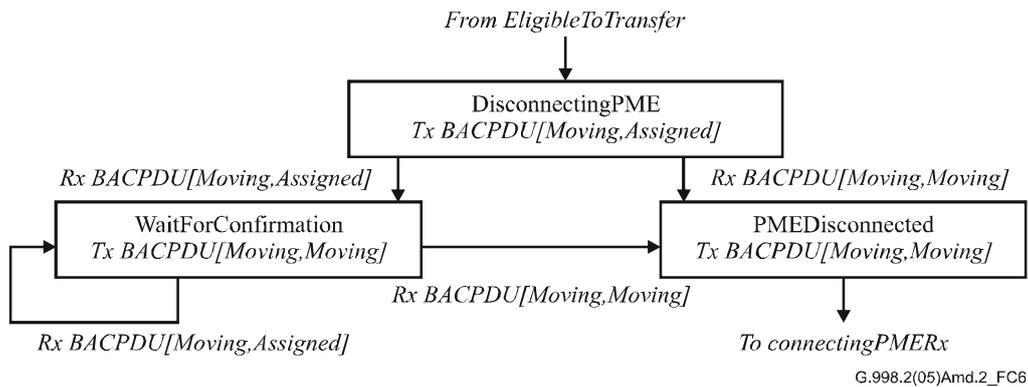


Figure C.5 – PME ID assignment state machine

This assignment process is complete when the BCE has received confirmation that the remote system has received its transmitted assignment TLV information.

#### C.3.2.2.2 Disconnecting the PAF from the single-PME PAF

Once both the local and remote BCEs know the target PME to be moved as well as how that PME shall be identified within the target PAF, the BCE can begin to move the PME to this PAF. The first step in this process is for both the local and remote systems to disconnect the PME from its current PAF. Once the PME is disconnected locally from the local PAF, the PME moves to the DisconnectingPME state. This process is shown in Figure C.6.

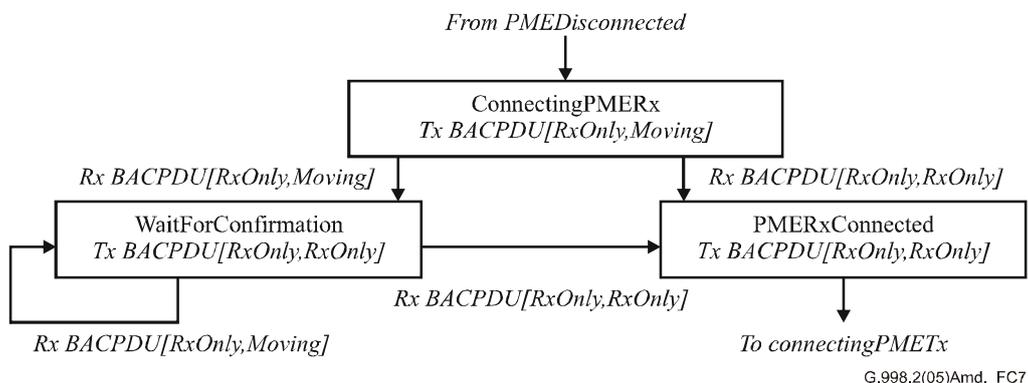


**Figure C.6 – PME disconnect state machine**

Disconnecting the PME from the single-PME PAF is complete when the BCE has received confirmation that the remote system received the locally transmitted status information. At this point the PME is in the PMEDisconnected state.

### C.3.2.2.3 Connecting the PME Rx to the target PAF

Once the PME has been disconnected from its previous PAF, the PME can be attached to the receive path of the target PAF. This basically sets up the receive process for this PME before the transmit process is established. By doing so, the PME transfer can be made in a seamless manner. Once the BCE locally connects the PME receive path to the target PAF, the PME can transfer into the ConnectingPMERx state. The process is shown in Figure C.7.

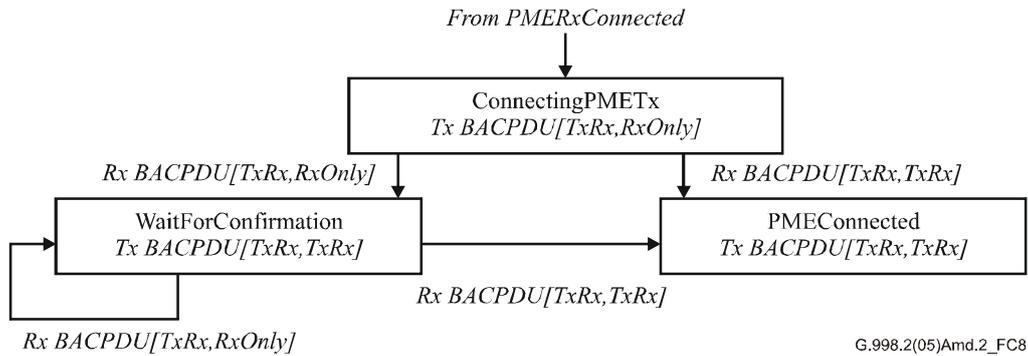


**Figure C.7 – PME Rx connect state machine**

Connecting the PME receive path to the target PAF is complete when the BCE has received confirmation that the remote system received the locally transmitted status information and has performed a similar connection.

### C.3.2.2.4 Connecting the PME Tx to the target PAF

The final phase of the PME transfer is enabling the transmit path on the target PAF for the transferring PME. The BCE begins the process by locally connecting the transmit path of the PME to the target PAF. It then signals this change to the remote BCE and awaits confirmation that the remote BCE has made a similar connection and received the status information transmitted by the local BCE.



**Figure C.8 – PME Tx connect state machine**

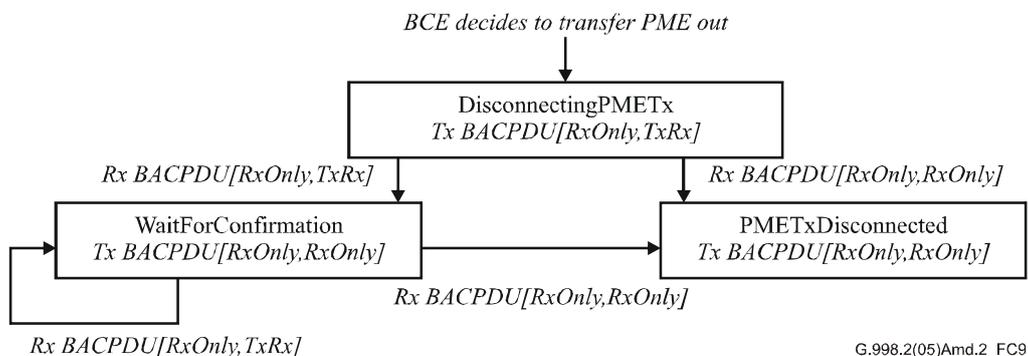
Connecting the PME transmit path to the PAF is complete when the BCE has received confirmation that the remote system received the locally transmitted status information.

### C.3.2.3 PME transfer out of a bonded group

This clause details the process for seamlessly removing a PME from a bonded group. As with the process of transferring a PME into a bonded group, the basic philosophy is that the local BCE takes an action, communicates that action to the remote BCE via the status array in BACPDU, and waits for the remote BCE to acknowledge that change and make a similar change.

#### C.3.2.3.1 Disconnecting the PME Tx from the current PAF

The first step in transferring a PME out of an existing PAF is to disconnect the transmit path of the PME from the existing PAF so that data no longer flows over this PME. Once the PME is locally disconnected from the transmit path of the PAF, the BCE enters the DisconnectingPMETx state for this PME and follows the state machine of Figure C.9.

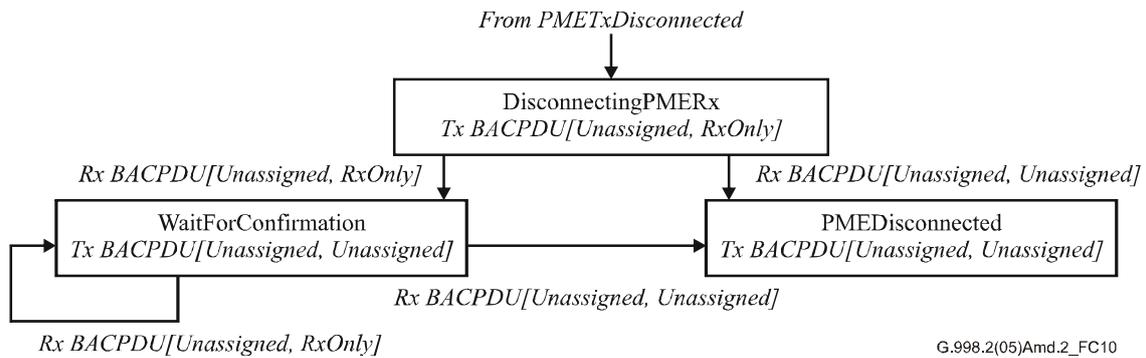


**Figure C.9 – PME Tx disconnect state machine**

At the end of this phase, the PME transfer path is not connected to any PAF.

#### C.3.2.3.2 Disconnecting the PME Rx from the existing PAF

Once the PME is no longer connected to any transmit path, the BCE can switch the receive path from the existing PAF to any 'empty' PAF. This is actually done in two parts – disconnecting the PME receive path from the current PAF, and then re-connecting it to a new (empty) PAF. For the disconnect process, the BACPDU are transferred over the existing PAF, where the PME is switched from the RxOnly status to the Unassigned status. Once this status change has been communicated and confirmed on the existing PAF, the PME is no longer in the PAF. This is shown in Figure C.10.



G.998.2(05)Amd.2\_FC10

**Figure C.10 – PME Rx disconnect state machine**

When this phase is complete, the PME is not connected to any PAF. In particular, the PME ID associated with the removed PME is no longer in use or relevant in the PAF from which it was removed, and that PME ID may be used for other purposes.

A PME shall not be disconnected if it is part of a single-PME PAF.

### C.3.2.3.3 Connecting the PME to an empty PAF

Connecting the PME to an empty PAF consists of the BCE connecting the PME to the receive and transmit paths of any empty PAF, and starting with the initialization process as if this PME was just initialized into a single-PME PAF (clause C.3.2.1).

### C.3.2.4 Termination

If the transceiver associated with a PME which is a member of an active PAF reinitializes, then after a vendor-discretionary timeout, the PME status shall be changed to "Unassigned" (i.e., removal from the active PAF, with notification TLVs sent to the far end BCE) and BACP initialization shall occur for this PME as specified in clause C.3.2.1. If the PME is able to transmit and receive fragments and the remote PME status has not been changed to "Unassigned" within the timeout period, then the local PME status remains as TxRx, without BACP reinitialization. To minimize data loss, the PAF should not send fragments to this PME while the PME is not able to transmit fragments on the line.

### C.3.2.5 BACPDUs re-transmissions and frequency

It is the responsibility of the BCE to ensure that all status changes are acknowledged by the remote BCE. If a BCE does not receive an acknowledgement of a transmitted status change, it should re-transmit that status several times until it becomes acknowledged, or until the BCE gives up on making that transition.

It is recommended that, by default, a BCE retransmit three times and provide a window of one second for an acknowledgement to be received.

There is no requirement to transmit BACPDUs when the local and remote BCEs have synchronized information. BACPDUs need only be exchanged to synchronize status change information.

The local BCE shall send a BACPDU only when its local status changes (to indicate that local status change to the remote end) or when it receives a BACPDU with changed remote status (to indicate to the remote end it knows about that remote status change) or when the remote status is Unknown.

## C.4 Bonding aggregation control protocol frame formats

The bonding control engine transmits and receives frames over a PAF via the standard MII interface to the PAF as shown in Figure C.1. Bonding aggregation control protocol PDUs must therefore conform to the standard definition of an Ethernet frame.

This clause describes the format in which BCEs communicate information in the BACPDUs.

### C.4.1 Slow protocol format

BACPDUs are in fact an instance of an IEEE 802.3 slow protocol (see Annex 43B of IEEE 802.3-2005). ITU has been allocated an organizationally unique identifier (OUI), and BACP is a subtype allocated from that ITU OUI. BACP frames are subject to pause frames, i.e., stop and resume their transmission with pause frames just like other data or protocol traffic.

It is the responsibility of the BCE to ensure that the BACPDUs conform to the transmission rate requirements of slow protocols. No more than 10 BACPDUs can be sent over any PAF in one second. BACPDUs can, in general, be sent immediately upon status changes as reflected in the earlier state descriptions, but must still overall live within the 10 BACPDUs-per-second limitation imposed by the slow protocols encapsulation.

The BACPDU frame content is shown in Figure C.11. A BACPDU shall be padded to meet the minimum frame size restrictions for Ethernet frames.

BACP PDU header
Variable number of TLVs
Ethernet FCS

**Figure C.11 – Generic BACPDU content**

BACPDUs consist of a fixed BACPDU header followed by a variable number of TLVs. A more detailed look at the BACPDU header is given in Figure C.12.

Destination MAC 01:80:C2:00:00:02
Source MAC (local)
Slow protocols Ethertype = 0x8809
Organization specific subtype = 0x0A
ITU OUI = 0x0019A7
ITU subtype = 0x01 (BACP)
BACP version = 0x01
BACP timestamp (32-bits)

**Figure C.12 – BACP header – Slow protocols subtype**

The fields in the BACP header (beyond the fixed MAC header) are:

- BACP version. The BACP version number must be set to 0x01 on transmit. BACP frames received with version numbers other than 0x01 must be discarded.
- BACP timestamp. The BCE may choose to timestamp some or all BACPDUs. When timestamps are not in use, this field is all zeros (0x00000000). When timestamps are in use, this field contains the value of a local clock expressed in 0.1-ms units.

Each TLV contains a one-octet type field and a one-octet length field, followed by the value parameter. The length field includes the length of the entire TLV, including the type and length fields. The rules for processing TLVs are as follows:

- a) Detection of the TLV type NULL TLV indicates that there are no additional TLVs to parse. The length and value of the NULL TLV can be ignored.
- b) Other TLVs with lengths 0x00 or 0x01 must be ignored and considered invalid. The BACP frame should be considered to have no additional TLVs after the detection of a TLV of length 0x00 or 0x01.
- c) TLVs with unknown types should be ignored.
- d) If the length of the TLV is less than expected for a known type, the TLV should be ignored.
- e) If the length of a TLV is greater than expected for a known type, the TLV is assumed to have been extended and the expected fields of that TLV can be processed in the expected locations. The data in the area greater than the expected length but less than the actual length can be ignored.
- f) If a length field indicates that the TLV extends beyond the length of the actual frame, the TLV should be ignored and the frame discarded.

The TLV types defined in this annex are given in Table C.1. A BACPDU shall contain, at a minimum, a local info TLV and a remote info TLV and shall be terminated by a NULL TLV.

**Table C.1 – Defined TLV Types**

<b>TLV Type</b>	<b>TLV Name</b>	<b>TLV length</b>	<b>TLV description</b>
0x00	NULL TLV	0	End-of-TLVs marker.
0x01	Local info TLV	24	Contains the status of all PME's related to this PAF as well as the group ID for the PME as known by the transmitting BCE.
0x02	Remote info TLV	24	This TLV contains a copy of last received local info TLV from the remote BCE (thus reflecting a copy of the information back to the remote BCE, allowing the remote BCE to confirm that its information had been received).
0x03	Assignment TLV	8	Contains the stream ID and PME ID for this PME. The stream ID has context across all PAFs, while the PME ID is relative only to the current PAF.
0xFF	Organization-specific TLV	N/A	Organization-Specific TLV

## C.4.2 BACP TLVs

This clause details the format of the TLVs that may be contained in a BACPDU. If a remote GID, a remote PME ID or a remote stream ID is not known, then that field shall be set to all ones.

### C.4.2.1 NULL TLV

The NULL TLV is used to indicate that this is the last TLV in the BACPDU. TLV processing for the frame terminates when this TLV is processed. There is no length or value field for this TLV.

The NULL TLV is required in every BACPDU. BACPDU without NULL TLVs should be discarded.

### C.4.2.2 Local info TLV

The local info TLV contains two pieces of information as known by the local (transmitting) BCE: the group ID (see clause C.3.1.1) and PME status array (see clause C.3.1.5). These are, respectively, 6-octet and 16-octet fields.

0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1
Type = 0x01						Len = 24						group ID																			
Group ID (continued)																															
PME status array																															
PME status array (continued)																															
PME status array (continued)																															
PME status array (continued)																															

Each PME known to this PAF has its status reflected in the PME status array. The PME assigned PME ID 'n' has its status carried in the nth location of the PME status array (n = 0..31), where each location identifies a 4-bit status field.

### C.4.2.3 Remote info TLV

The remote info TLV is included to echo back to a BCE the values that the remote BCE is receiving from the local BCE on this PAF. This TLV is used to confirm that local status changes communicated by one BCE have been received by the other BCE.

This TLV has the same format as the local info TLV.

0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1
Type = 0x02						Len = 24						Remote group ID																			
Remote group ID (continued)																															
Remote PME status array																															
Remote PME status array (continued)																															
Remote PME status array (continued)																															
Remote PME status array (continued)																															

When no information has been received from the remote BCE on this PAF, all remote PME status values are set to Unknown, i.e., the remote PME status array is returned as all zeros.

### C.4.2.4 Assignment TLV

The assignment TLV is used to associate a stream with a relative PME ID within a specific PAF. The assignment TLV contains four values, the local stream ID and PME ID of that stream within this PAF, and the remote stream ID and PME ID of that stream within this PAF. It is used to communicate the association of these four values to the remote BCE. The format of the assignment TLV is given below.

0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9 0 1
Type = 0x03	Len = 8	Stream ID
Remote stream ID		PME ID Remote PME ID

The assignment TLV is included in a BACPDU in two situations. First, whenever the PME is first initializing with BACP (see clause C.3.2.1). Second, when a PME is being transferred into a new PAF, the assignment TLV is used in the BACPDUs on that new PAF to associate a known stream with a PME ID for that PAF.

#### C.4.2.5 Organization-specific TLV

Organizations may extend this protocol by including additional TLV information in BACPDUs. This Recommendation does not restrict or control organization-specific TLVs – they are controlled by the organization identified within the TLV. The format of organization-specific TLVs is given below:

0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9 0 1
Type = 0xFF	Len = ?	OUI
OUI (cont)	Specified by organization...	

Organization-specific TLVs can be included in any BACPDU. They can be of any length. An OUI is used to identify the organization that specified the syntax and semantics of the rest of the value field of the TLV.



## **SERIES OF ITU-T RECOMMENDATIONS**

Series A	Organization of the work of ITU-T
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
<b>Series G</b>	<b>Transmission systems and media, digital systems and networks</b>
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems