TECHNICAL REPORT

TR-101

Migration to Ethernet-Based Broadband Aggregation

Issue: 2

Issue Date: July 2011

Notice

The Broadband Forum is a non-profit corporation organized to create guidelines for broadband network system development and deployment. This Broadband Forum Technical Report has been approved by members of the Forum. This Broadband Forum Technical Report is not binding on the Broadband Forum, any of its members, or any developer or service provider. This Broadband Forum Technical Report is subject to change, but only with approval of members of the Forum. This Technical Report is copyrighted by the Broadband Forum, and all rights are reserved. Portions of this Technical Report may be copyrighted by Broadband Forum members.

This Broadband Forum Technical Report is provided AS IS, WITH ALL FAULTS. ANY PERSON HOLDING A COPYRIGHT IN THIS BROADBAND FORUM TECHNICAL REPORT, OR ANY PORTION THEREOF, DISCLAIMS TO THE FULLEST EXTENT PERMITTED BY LAW ANY REPRESENTATION OR WARRANTY, EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, ANY WARRANTY:

(A) OF ACCURACY, COMPLETENESS, MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, NON-INFRINGEMENT, OR TITLE;

(B) THAT THE CONTENTS OF THIS BROADBAND FORUM TECHNICAL REPORT ARE SUITABLE FOR ANY PURPOSE, EVEN IF THAT PURPOSE IS KNOWN TO THE COPYRIGHT HOLDER;

(C) THAT THE IMPLEMENTATION OF THE CONTENTS OF THE TECHNICAL REPORT WILL NOT INFRINGE ANY THIRD PARTY PATENTS, COPYRIGHTS, TRADEMARKS OR OTHER RIGHTS.

By using this Broadband Forum Technical Report, users acknowledge that implementation may require licenses to patents. The Broadband Forum encourages but does not require its members to identify such patents. For a list of declarations made by Broadband Forum member companies, please see http://www.broadband-forum.org. No assurance is given that licenses to patents necessary to implement this Technical Report will be available for license at all or on reasonable and non-discriminatory terms.

ANY PERSON HOLDING A COPYRIGHT IN THIS BROADBAND FORUM TECHNICAL REPORT, OR ANY PORTION THEREOF, DISCLAIMS TO THE FULLEST EXTENT PERMITTED BY LAW (A) ANY LIABILITY (INCLUDING DIRECT, INDIRECT, SPECIAL, OR CONSEQUENTIAL DAMAGES UNDER ANY LEGAL THEORY) ARISING FROM OR RELATED TO THE USE OF OR RELIANCE UPON THIS TECHNICAL REPORT; AND (B) ANY OBLIGATION TO UPDATE OR CORRECT THIS TECHNICAL REPORT.

Broadband Forum Technical Reports may be copied, downloaded, stored on a server or otherwise re-distributed in their entirety only, and may not be modified without the advance written permission of the Broadband Forum.

The text of this notice must be included in all copies of this Broadband Forum Technical Report.

Issue History

Issue Number	Issue Date	Issue Editor	Changes
1	April 2006	Amit Cohen, ECI Telecom Ed Shrum, BellSouth Telecommunications	Original
2	July 2011	Tom Anschutz, AT&T	Additional access technologies, errata, minor enhancements developed in related TRs.

Comments or questions about this Broadband Forum Technical Report should be directed to info@broadband-forum.org.

Editor	Tom Anschutz	AT&T
End to End Architecture WG Chairs	David Allan David Thorne	Ericsson BT
Vice Chair	Sven Ooghe	Alcatel-Lucent
Chief Editor	Michael Hanrahan	Huawei Technologies

Table of Contents

Table of Figures

Table of Tables

Executive Summary

This Technical Report outlines how an ATM aggregation network can be migrated to an Ethernet based aggregation network in the context of TR-25 and TR-59 based architectures. TR-101 provides an architectural/topological model of such an Ethernet based aggregation network that supports the business requirements in TR-058. In doing so it describes requirements for protocol translation and interworking, QoS, multicast, security, and OAM for a Broadband aggregation network. The architecture defined in TR-101 remains popular today for service providers with limited needs to support multiple service types on a single access. Therefore this revision was undertaken to bring the work “up to date” with respect to new access technologies it can support as well as to address a number of errata and minor improvements.

1. Introduction and Purpose

1.1 Document Scope

TR-101 outlines how an ATM aggregation network can be migrated to an Ethernet based aggregation network in the context of TR-25 and TR-59 based architectures. TR-101 provides an architectural/topological model of such an Ethernet based aggregation network that supports the business requirements in TR-058. In doing so it describes requirements for protocol translation and interworking, QoS, multicast, security, and OAM for a Broadband aggregation network.

TR-058 describes the marketing requirements for a multi-service architecture. These requirements include the following capabilities:

- Improved transport (the main focus of this document)
- Many-to-many access (multi-session)
- Differentiated services (including QoS and QoS on Demand)
- Bandwidth services (including Bandwidth on Demand)
- Content distribution (including multicast capabilities)
- Simpler provisioning
- Support for business services (e.g. Layer 2 VPN, high availability, higher bit rate services)

TR-101 does not provide details/requirements with respect to scale and performance of individual elements, but does focus on documenting a functional architecture and the requirements necessary to support it. Also note that TR-101 builds on the requirements defined in TR-092, Broadband Remote Access Server (BRAS), and TR-124, Functional Requirements for Broadband Residential Gateway Devices, as components of the architecture. This revision does not address multi-service business requirements defined in TR-144 nor its related architectures. Neither does this Technical Report provide guidance on the use of, or transition to, using IPv6. Those aspects are defined in TR-177 and in work under development within the Broadband Forum.

1.2 ATM Based Architectures

DSL deployments in the past have followed the architectural guidelines of TR-025 or the more advanced TR-059 (reference models are depicted in Figure 1 and Figure 2 respectively). Both architectures used ATM to aggregate the access networks into the regional broadband network. In such deployments the Access Node functions as an ATM aggregator and cross-connect, multiplexing user ATM PVCs from the U interface onto the V interface and de-multiplexing them back on the opposite direction (see Figure 1).

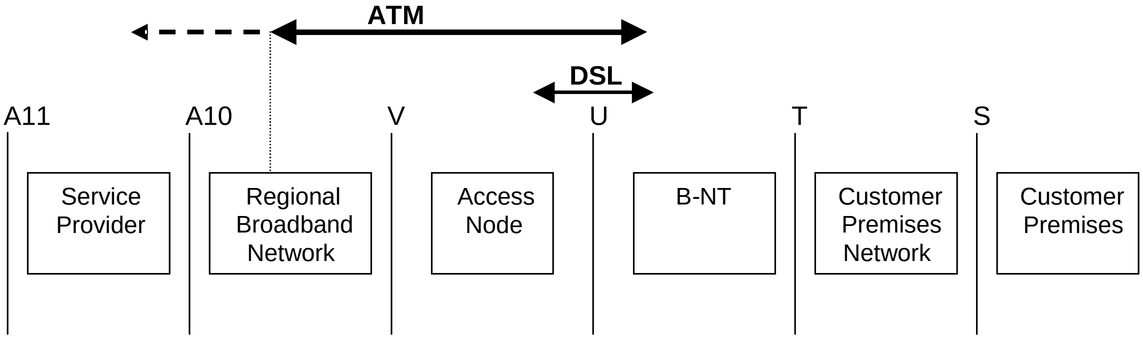


Figure 1 – TR-025 High Level Architectural Reference Model

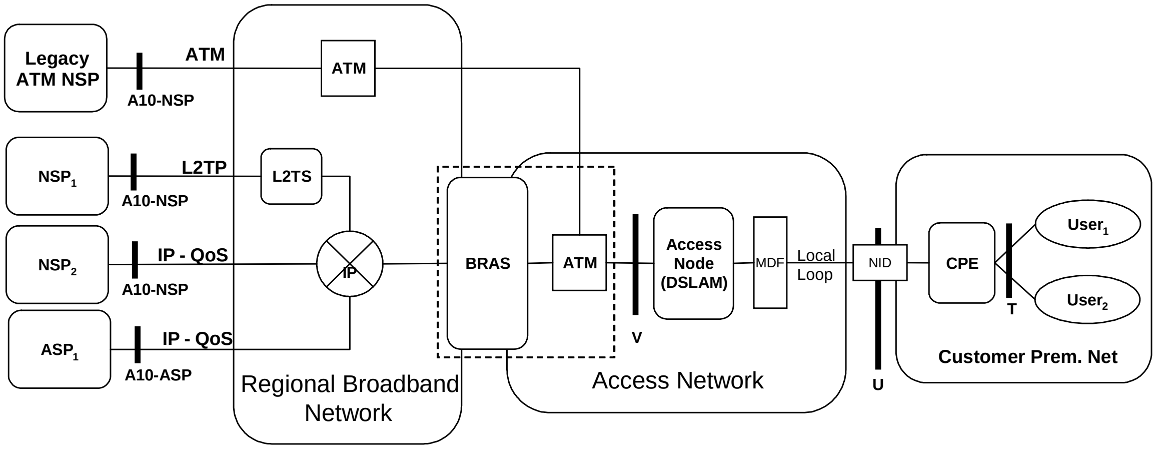


Figure 2 – TR-059 High Level Architectural Reference Model

The traffic aggregated from the Access Nodes is steered to an IP node, the BRAS. In TR-025 a BRAS could be physically located either in the regional network or in the service provider network and is mainly engaged in PPP termination and tunneling. In TR-059 the BRAS is located on the edge of the regional network and its functionality is enhanced to include subscriber management, advanced IP processing, including IP QoS, and enhanced traffic management capabilities, e.g. 5-layer hierarchical shaping.

TR-101 was originally released in 2006. It specified a new architecture based on Ethernet aggregation instead of ATM. TR-101 has been extremely well adopted. However, since the release of TR-101, several technologies have emerged that can be used instead of DSL in the Access with few or even no architectural changes. Further, DSL itself has undergone some updates and improvements. The TR-101 architecture remains desirable for service providers that are interested in triple-play capabilities, but may not wish to move to the more general multi-service architecture based on TR-144 and on work under development within the Broadband Forum. Significant interest, therefore, exists in re-issuing TR-101. This revision retains the business drivers and basic architecture of the original TR-101, but includes new technologies, and makes some clarifications and modifications based on the industries' experience with TR-101 over the past few years. The opportunity has also been taken to do a general update with regard to language, state of the industry, etc.

So as not to confuse the use of the term BRAS, this document has adopted the term Broadband Network Gateway (BNG). A Broadband Network Gateway may encompass what is typically referred to as a BRAS (as specified in TR-092), but this is not a requirement of this architecture (see the following Section 1.3).

For the purpose of clarity in this document we define the term *'aggregation network'* as the part of the network connecting the Access Nodes to the Broadband Network Gateway (i.e. this is the edge of the regional network according to TR-025 and part of the access network according to TR-059). In both TR-025 and TR-059 the aggregation network is ATM based.

TR-101 defines a new access network topology where the connectivity between the Access Node and the Broadband Network Gateway is Ethernet based rather than ATM. Access nodes can be directly connected or go through an aggregation layer(s) before reaching the Broadband Network Gateway.

1.3 Broadband Network Gateway Assumptions

TR-059 based architectures assume a single Broadband Remote Access Server (e.g. BRAS) where user services are performed.

TR-101, however, recognizes the potential for service segregation. Some applications, such as video, may have specific enough requirements from the network that they are best optimized separately from other types of traffic.

The architecture described in this document supports the possibility for a dual Broadband Network Gateway (BNG) scenario when required for video optimization. When used, the BNG dedicated to video is denoted as the 'video BNG'. Such an approach may have additional complexities not present with a single BRAS/Broadband Network Gateway arrangement described in TR-059. This is most notable with respect to traffic/bandwidth management and the resulting network efficiency as well as the additional operations overhead.

Furthermore, in dual node architectures, it is not mandated that *both* BNGs support all of the requirements detailed in this document. Specifically, the video BNG may not implement subscriber management functions (e.g. PPP termination, per user QoS) given that these functions are likely to be performed by the other BNG.

A multi-node deployment (more than 2) could follow the same recommendations described within this document, but will not be explicitly described here; this is the subject of TR-144 and of work under development within the Broadband Forum.

1.4 Motivation for Migration to Ethernet Based Broadband Aggregation

Access architectures are evolving from ADSL(1) based networks to a much wider range of access technologies including multiple types of xDSL (e,g, ADSL2+, VDSL, G.SHDSL), optical access (e.g. PON and point-to-point fiber) and even wireless (e.g. 3G, WiMAX and LTE). The new technologies can provide higher user bit rates, better support for services requiring QoS and multicast, and improved availability. The industry is also taking initial steps towards interworking with mobile networks. Ethernet still provides a technology foundation to meet the needs of the next generation broadband network through a maturing transport mechanism that supports higher connection speeds, packet based QoS, simpler provisioning, multicast, and redundancy in an efficient manner.

Broadband service providers are looking to support enhanced/managed services in conjunction with basic Internet access - including entertainment video services (Broadcast TV and VoD), video conferencing, VoIP, gaming, and business class services (e.g. Layer 2 VPN and IP VPN). Many of these services require significantly higher broadband data rates than are typically achieved in traditional ADSL deployments. The most reliable way to increase the maximum data rate for copper access is to reduce the distance between the ATU-C and the ATU-R, thus significantly changing the placement and density of xDSL Access Node deployments. Other ways to increase speeds include pair bonding, vectoring, and all-optical solutions. The number of Access Nodes deployed within a service provider's network will significantly increase if the Access Nodes are pushed further out (i.e. closer to the user's edge) or will decrease if the Access Nodes are more centralized through use of optical technologies and higher port densities. Their throughput capacity and ability to support multiple simultaneous access technologies is also expected to increase. Fast Ethernet, Gigabit Ethernet and GPON, as well as their 10G variants, provide highly efficient transport for delivering large amounts of bandwidth to a highly distributed Access Node topology, as well as provide the underlying QoS features needed by various applications.

1.5 Requirements

In this document, several words are used to signify the requirements of the specification. These words are always capitalized when used in their requirements sense.

MUST This word, or the adjective "REQUIRED", means that the definition is an absolute requirement of the specification

MUST NOT This phrase means that the definition is an absolute prohibition of the specification.

SHOULD This word, or the adjective "RECOMMENDED", means that there could exist valid reasons in particular circumstances to ignore this item, but the full implications need to be understood and carefully weighed before choosing a different course.

MAY This word, or the adjective "OPTIONAL", means that this item is one of an allowed set of alternatives. An implementation that does not include this option MUST be prepared to inter-operate with another implementation that does include the option.

1.6 Key Terminology

The following definitions apply for the purposes of this document:

Access Loop The physical connectivity between the Network Interface Device (NID), at the customer premises, and the Access Node.

Access Network The Access Network encompasses the elements of the broadband network from the NID at the customer premises to a Broadband Network Gateway. This network typically includes one or more types of Access Node and may include an Ethernet aggregation function.

Access Node (AN) The Access Node may implement one or more access technologies based on copper, fiber or wireless. It may also aggregate traffic from other access nodes. It can be placed in a variety of locations from climate controlled (central) offices to outside environments that require climate hardening of the equipment to avoid the need for additional cabinets or enclosures. In this specification, this node contains at least one standard Ethernet interface that serves as its uplink interface into which it aggregates traffic from broadband user ports. Access Nodes may be distributed systems, with distributed ONUs, shelves, or modules that attach to one another and present the management view of a single, distributed system.

Aggregation Network The part of the network stretching from the Access Nodes to the Broadband Network Gateway(s). In the context of this document the aggregation network is considered to be Ethernet based, providing standard Ethernet interfaces at the edges, for connecting the Access Nodes and Broadband Network Gateway(s), and some transport for Ethernet frames (e.g. Ethernet over SONET, MPLS, RPR, etc.) at the core.

Broadband Network Gateway (BNG) IP Edge Router where bandwidth and QoS policies may be applied. This term is used instead of BRAS to denote an Ethernet-centric IP edge node in this document.

Broadband Remote Access Server (BRAS) The BRAS is a Broadband Network Gateway and is the aggregation point for the user traffic. It provides aggregation capabilities (e.g. IP, PPP, Ethernet) between the access network and the NSP or ASP. Beyond aggregation, it can also support policy management and IP QoS in the access network. In this document, the term BRAS is used to describe the ATM-centric device described in TR-092.

C-Tag The innermost VLAN tag as defined in IEEE 802.1ad and having an EtherType value of 0x8100

C-VID The C-VLAN ID: The virtual LAN value of a C-Tag.

C-VLAN The virtual LAN indicated by a C-VID.

Downstream The direction of transmission from the regional network to the Access Node and from the Access Node toward the end user.

EFM Ethernet in the First Mile. This document uses a somewhat broader definition of EFM than the IEEE, who originally devised the term. Specifically, in this document EFM includes any access technology that supports native Ethernet framing on the access loop. Note, however, that EPON is covered by TR-200 and not here.

End User A broadband endpoint using any supported access technology.

Explicit Host Tracking A variant of an IGMP router function, where the IP address of each host sending an IGMP leave or join message is inspected and recorded. This enables the IGMP router to track, for each multicast group, the exact number and identity (i.e. IP address) of hosts receiving it on the IGMP router's segment. Enables support for immediate leave.

IEEE Priority Bits

The Ethernet priority bit field of the IEEE 802.1D bearer. Formally known as IEEE 802.1p and augmented in IEEE 802.1ad

IGMP Host

This is the system initiating IGMP operations in order to receive (or stop receiving) multicast flows. This is typically an IP host system.

IGMP Immediate Leave

A function associated with IGMP snooping or IGMP routing whereby the switch or router stops sending immediately the multicast stream when receiving an IGMP leave for the last member on this requesting interface, i.e. without sending one or more group specific queries and waiting for its timeout.

IGMP Proxy Routing

This is a scheme designed for simple tree topologies whereby a device uses IGMP/MLD rather than PIM/DVRMP to learn and forward multicast traffic. In this capacity the device has a single interface defined as an upstream interface, called the "Host interface" where it acts as a host IGMP/MLD entity and one or more downstream interfaces, called the “Router Interfaces" where it performs the IGMP router function. Reports, leaves sourced from the proxy function originate from the host address on the upstream interface. The implication of the proxy function being performed by a layer 3 forwarding device acting as IGMP querier on user ports and as IGMP host on the network port is that the device must be in two different subnets.

For more information on the IGMP proxy routing function please refer to RFC 4605.

IGMP Querier

Part of the IGMP router component that is responsible for maintaining the status of multicast groups on a particular interface. There is a single IGMP querier per subnet which uses an election process based on a querier or a newly started router sending an IGMP general query to the all-systems multicast group (224.0.0.1). The multicast router with the lowest IP wins the election process.

IGMP Router

This is the system processing IGMP operations in order to send (or stop sending) multicast flows towards clients. This is typically an IP router.

IGMP Snooping with Proxy Reporting

This macro-function can be decomposed in 3 elementary sub-functions:

- Report suppression: intercepts, absorbs and summarizes IGMP reports coming from IGMP hosts. IGMP reports are relayed upstream only when necessary, i.e. when the first user joins a multicast group, and once only per multicast group in response to an IGMP query.

- Last leave: intercepts, absorbs and summarizes IGMP leaves coming from IGMP hosts. IGMP leaves are relayed upstream only when necessary, i.e. when the last user leaves a multicast group.
- Query suppression: intercepts and processes IGMP queries, in such a way that IGMP specific queries are never sent to client ports, and IGMP general queries are relayed only to those clients ports receiving at least one multicast group

The IGMP proxy reporting function, when performing the above functions, may forward original IGMP messages generated by hosts and multicast routers but may also generate IGMP messages. In this case the traffic's source address will be 0.0.0.0 with the IGMP proxy reporting function's own unique MAC address.

IGMP Transparent Snooping

IGMP snooping optimizes the distribution of multicast within an IEEE 802.1 bridging domain so multicast traffic is only sent on bridge ports where there are known to be active receivers and/or multicast routers. IGMP snooping functionality resides on IEEE bridging devices that connect IGMP hosts to IGMP routers and consists of two main components. The first is the IGMP snooping control section which:

1) Monitors IGMP messages (and optionally other multicast router messages, such as PIM or DVMRP hello packets), to determine the port location of the multicast routers and active receivers within an IEEE bridged domain.

2) Builds per port, per VLAN multicast forwarding tables

3) Maintain basic IGMP membership state on non-router ports to determine when a forwarding entry should be removed.

The second function is the data forwarding section which:

1) Forwards packets in the 224.0.0.0/24 range which are not IGMP messages on all ports.

2) Forwards multicast packets with a destination IP address outside 224.0.0.0/24, which are not IGMP according to per VLAN, per port multicast forwarding tables.

This basic mode of operation is often referred to as "transparent IGMP snooping" and does not absorb, nor alter, nor generate IGMP messages when performing the above functions.

For more information on the IGMP snooping function please refer to RFC 4541. However, please note that the terms used within this document for various flavors of snooping do not directly match the terms used in the above RFC. More descriptive terms are used here to provide additional clarification.

Interworked PPP Session
A PPP session that is being converted by the Access Node from a PPPoA session on the U-interface to a PPPoE session on the V-interface.

Interworking Function (IWF)
The set of functions required for interconnecting two networks of different technologies. These functions include conversion of PDU framing, addressing schemes, priority mapping, security mechanisms, and OAM flows.

Layer-2 DHCP Relay Agent
RFC 3046 identifies the possibility that a layer-2 network element (a bridge) between the full DHCP relay agent and the client may add the relay agent information option, option 82, but not set the giaddr field. This is useful for access technologies with extensive layer-2 topology, where the device that has the information for constructing circuit-id and remote-id is a layer-2 element with no need for a layer-3 interface on each client subnet or VLAN. RFC 3046 gives no name to a network element performing this function. The Broadband Forum has adopted the term layer-2 DHCP relay agent for a network element implementing this function. This is described further in Section 3.8.1.

Multicast VLAN
Any VLAN (dedicated, shared, N:1, etc.) that carries multicast bearer traffic in response to IGMP snooping performed at the Access Node

Port
This term, when used to describe a user port, denotes a virtual (e.g. a PVC) or a physical (e.g. DSL or optical) port. The terms *DSL Port* and *DSL link* are the same as physical port.

PPPoE Intermediate Agent
A function performed on PPPoE discovery stage frames by an Access Node, mainly consisting of adding to the frames access loop identification and information.

Priority Tagged Frame
An Ethernet fame carrying a priority tag (i.e. VLAN ID Zero).

PVC Bundle
A multi-PVC user port where the PVCs are terminated at the Access Node and are used for ATM CoS on the U interface as well as indicating Ethernet priority mapping. Traffic associated with different VCs sharing the same bundle follows the same forwarding path across the V interface (i.e. same VLAN). This definition differs from the TR-059 term in that the PVCs are terminating at the Access Node instead of the BRAS.

Q-Tag
VLAN tag as described in IEEE 802.1Q-2005 and found at the U Interface.

Regional Broadband Network (RBN) The regional broadband network (‘regional network’ for short) interconnects the Network Service Provider's networks and the access networks. Typically more than one access network is connected to a common regional network.

S-Tag The outermost or single VLAN tag as defined in IEEE 802.1ad and having an Ethertype of 0x88a8.

S-VID The virtual LAN value of an S-Tag.

S-VLAN The virtual LAN indicated by an S-VID.

Untagged Frame An Ethernet frame without any VLAN or priority tagging.

Upstream The direction of transmission from the end user to the Access Node and from the Access Node towards the regional network.

User Same as End User

User Isolation In the context of this document, user isolation means that the user/subscriber does not have direct bi-directional connectivity at the Ethernet MAC layer to any other user/subscriber in the RBN. This needs to be enforced in all nodes between the U interface and the BNG(s).

VID VLAN ID. The number used to identify a specific VLAN.

VLAN Virtual Local Area Network.

VLAN Tagged Frame An Ethernet frame carrying a VLAN tag (VID different than zero).

1:1 VLAN Indicates a one-to-one mapping between user port and VLAN. The uniqueness of the mapping is maintained in the Access Node and across the Aggregation Network

N:1 VLAN Many-to-one mapping between user ports and VLAN. The user ports may be located in the same or different Access Nodes.

1.7 Acronyms

AAA – authentication, authorization and accounting

AAL – ATM adaptation layer

AF – assured forwarding

AIS - ARP – address resolution protocol

AN – access node

ASCII – American standard code for information interchange

ASM – any source multicast

ASP – application service provider

ATM – asynchronous transfer mode

ATU-C –ADSL termination unit-CO

ATU-R- ADSL termination unit- remote

B-NT – broadband network termination

BB - broadband

BBDLC – Broadband Digital Loop Carrier

BE – best effort

BNG – broadband network gateway

BRAS – broadband remote Access server

CBR – constant bit rate

CCM – continuity check message

CO – central office

CoS – class of service

CPE – customer premises equipment

CPID – connection point identifier

CPU – central processing unit

DEI – discard eligibility indicator

DHCP – dynamic host configuration protocol

DLC – Digital Loop Carrier

DSL – digital subscriber loop

DVMRP – distance vector multicast routing protocol

EAP – extensible authentication protocol

EF – expedited forwarding

EFM – Ethernet in the first mile

ETH – Ethernet

GPON – gigabit passive optical network

HP – hierarchical policing

HS – hierarchical scheduling

IANA – Internet Addressing and Numbering Authority

IEEE – Institute of Electrical and Electronic Engineers

IETF – Internet Engineering Task Force

ICMP – Internet Control Message Protocol

IGMP – Internet Group Management Protocol

IP – Internet Protocol

IPoE – IP over Ethernet

ITU – International Telecommunications Union

IVL – Independent VLAN Learning

IWF – Interworking Function

LAC – L2TP access concentrator

LAN – local area network

LB - loopback

LBM – loopback message

LBR – loopback reply

LCP – link control protocol

LE – lower effort

LLC – logical link control

LTM – link trace message

LTR – link trace reply

L2TP – layer 2 tunneling protocol

MAC – media access control

MBS – maximum burst size

MDF – main distribution frame

ME – maintenance entity

MEP – maintenance end point

MIP – maintenance intermediate point

MLD – multicast listener discovery

MP – maintenance point

MRU – maximum receive unit

NAS – network access server

NID – network interface device

NNI – network to network interface

NSP – network service provider

OSS – operational support system

PADI – PPPoE active discovery initiation

PADR – PPPoE active discovery request

PADS – PPPoE active discovery session-confirmation

PADT - PPPoE active discovery terminate

PCR – peak cell rate

PDU – protocol data unit

PHB – per hop behavior

PIM – protocol independent multicast

POP – point of presence

PPP – point to point protocol

PPPoA – PPP over ATM

PPPoE –PPP over Ethernet

PTA – PPP terminated access

PVC – permanent virtual circuit

QoS – quality of service

RADIUS – remote access dial in user services

RAM – Remote Access Module

RFC – request for comments

RG – residential gateway

RPR – resilient packet ring

SCR – sustained cell rate

SLA – service level agreement

SMI – structure of management information

SNAP – sub network access point

SSM – source specific multicast

STP – spanning tree protocol

TLS – transparent LAN service

TLV – type – length - value

UBR – unspecified bit rate

UNI – user to network interface

VB – virtual bridge

VBR-nrt – Variable bit rate – non-real time

VBR-rt – Variable bit rate – real time

VC – Virtual Circuit

VID – VLAN ID

VLAN – Virtual LAN

VoD – video on demand

VoIP – voice over IP

VPN – virtual private network

VSA – vendor specific attribute

WFQ –weighted fair queuing

2. Fundamental Architectural and Topological Aspects

The principle guiding this specification is to enable a smooth migration from an ATM based aggregation network to an Ethernet based one. The following requirements primarily focus on the aggregation network but also include some new requirements at the U interface and beyond. However, the functions specified in this document are designed to be compatible with all access protocols that were recommended by the Broadband Forum for usage over the U interface, i.e. TR-043. In addition to the access protocols, the specified functions are designed to allow service levels (e.g. QoS, privacy, security etc.) that are the same or better than those offered by an ATM based aggregation network. Several OAM and provisioning issues that are affected by replacing the ATM aggregation network by an Ethernet one are also addressed. In this Technical Report the term *port,* when used to describe a user port, typically denotes a physical (e.g. DSL) port. However, it may also be used to describe an ATM PVC in an arrangement where multiple ATM PVCs are used over a single physical port. The use of multiple ATM PVCs per physical port is not possible with many of the new access technologies, therefore it is strongly recommended to use a multi-VLAN arrangement on physical ports in cases where the functionality of multiple PVCs might have been desired.

Figure 3 depicts the network architecture considered in this document. The fundamental changes from initial deployment architectures are:

- V interface that supports Ethernet as the transport protocol with no ATM present.
- An Ethernet aggregation network
- Support for dual Broadband Network Gateways
- Higher user bit rates (e.g. by supporting Access Nodes nearer the end-user)
- Higher network availability (in support of application and business customer needs)
- U interface that can support direct Ethernet framing over various physical layers.

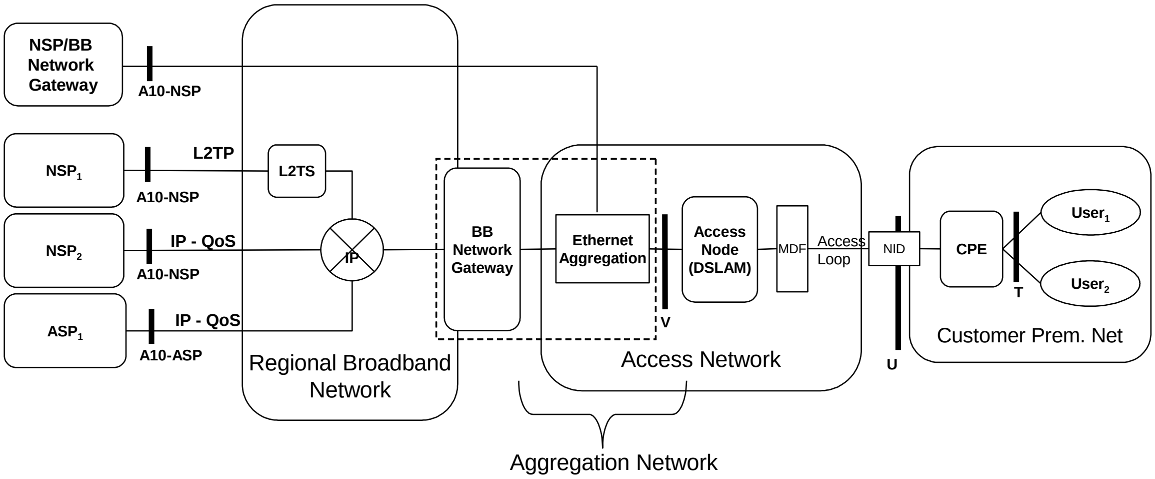


Figure 3 – Network architecture for Ethernet-based Broadband aggregation

Similar to the TR-059 architecture, the architecture depicted in Figure 3 uses a Broadband Network Gateway for providing IP QoS down to the user level. TR-059 also supported legacy ATM services that may not have traversed the Broadband Network Gateway. Similarly, the architecture described within this document also enables services to be injected into the network without going through a single Broadband

Network Gateway and allows native Ethernet services to be switched through the BNG or through other network elements. The QoS models supported are described in Section 2.9.

2.1 Residential Gateway

One goal of this architecture is to support new deployment models and applications. The requirements contained within this section are relevant to the support of these new deployment models.

It should be noted that Residential Gateways are not limited to use in consumer applications, and may also be suitable for use by small-medium business customers.

R-01 The RG MUST support sending the following frame types: untagged frames, priority-tagged frames and VLAN-tagged Ethernet frames in the upstream direction for stacks a, b, e, f and g in Figure 4

R-02 The RG used to support business customers SHOULD support sending double-tagged Ethernet frames in the upstream direction for stacks a, b, e, f and g in Figure 4

R-03 The RG MUST support setting the priority tag and VLAN ID values.

R-04 The RG MUST support receiving untagged and VLAN tagged Ethernet frames in the downstream direction, and MUST be able to strip the VLAN tagging from the ones received with tags.

RG Multicast requirements are captured in Section 6.

RG OAM requirements are captured in Section 7.

2.2 The U Interface

Stacks **a** to **d** of Figure 4 depict the access protocol stacks described in TR-43 for the U-interface.

Option **a** is denoted *IPoEoATM* and option **b** is denoted *PPPoEoATM*. Option **c** is denoted *IPoA* and option **d** is denoted *PPPoA*. The initial version of this document required the Access Node to translate option **c** and **d** access protocols to the protocol stacks supported on the V interface as illustrated in Figure 7. This interworking and bridging function is still described in Section 3.

Note: Using these capabilities should be avoided in new deployments.

Options **e** and **f** represent those scenarios when the access loop supports direct Ethernet encapsulation and are refered to as IPoE and PPPoE. Similarly, Option **g** represents a pure Ethernet service where there is no visibility above layer 2. Within this document both options **a** and **e** are commonly refered to as IPoE and options **b** and **f** are refered to as PPPoE. A port using IPoE or PPPoE access method is generally denoted a *bridged port*.

Those options where Ethernet is included in the protocol stack may also include 802.1Q headers to carry VLAN tags and priority markings.

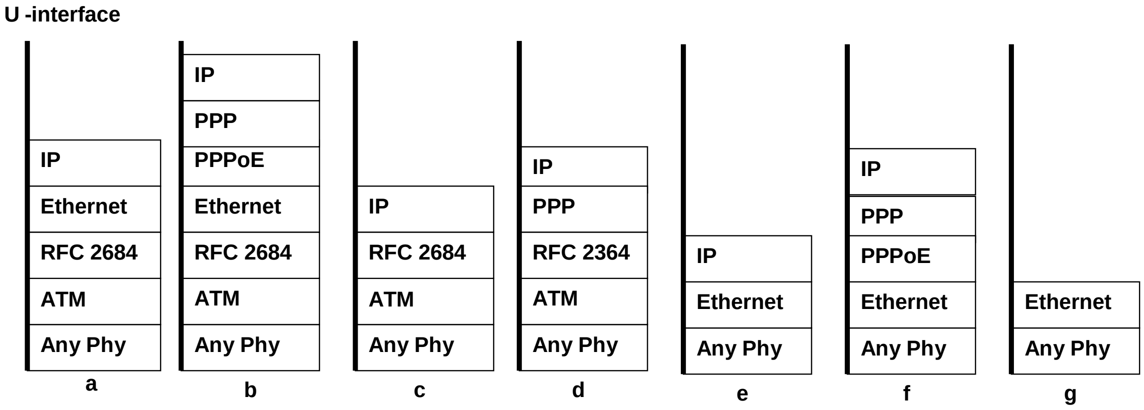


Figure 4 – Protocol stacks at the U interface

The physical layer of the U interface considered by this document includes, but is not limited to the following technologies:

- ADSL1 – ITU-T G.992.1
- ADSL2 - ITU-T G.992.3
- ADSL2plus - ITU-T G.992.5
- VDSL2 - ITU-T G.993.2
- G.SHDSL – ITU-T G.991.2
- Any point-to-point 802.3 Ethernet Physical layer
- Bonding of multiple DSL pairs – ITU-T G.Bond (ATM transport (G.998.1), and Ethernet transport (G.998.2))

Note: The use of GPON in TR-101-type architecture is specified in TR-156 and TR-167 and the use of EPON in TR-101-type architecture is specified in TR-200.

2.3 Access Node

The Access Node is the first aggregation point in the access network and in addition to terminating the physical layer signals it must support the following high level capabilities:

- The Access Node must be able to terminate the user ATM layer when it is present.
- The Access Node must have an Ethernet uplink providing connectivity to the aggregation network.
- When ATM is supported on the DSL line (U interface), the Access Node has to provide an interworking function (denoted IWF in Figure 5) between the ATM layer on the user side and the Ethernet layer on the network side, encompassing protocol translation, access loop identification, QoS, security and OAM issues. This may require the Access Node to snoop, modify or terminate protocols in layers above the AAL.
- Native Ethernet framing should be supported on the access loop.
- The Access Node must be able to support multicast deployments.
- The Access Node must support user isolation.

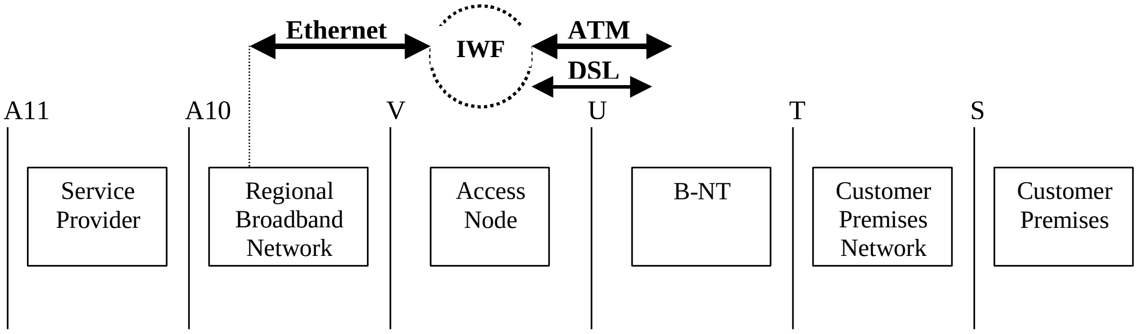


Figure 5 – ATM to Ethernet inter-working function

This specification is limited to AAL5 based services when ATM is present at the U interface.

This specification defines requirements for the Access Node addressing the need to: interwork between the ATM and Ethernet networks; provide compatibility with current deployment scenarios; and support multicast. This specification does not require BRAS functionalities such as LAC or PTA to be incorporated into the Access Node. The general intent is for the Access Node to behave as a Ethernet switch while also providing enhanced functionality for protocol interworking, multicast support, and customization for support of access networks (e.g. ARP and IGMP processing, user identification and isolation).

2.4 Access Node Deployment Options

Figure 6 depicts the typical Access Node scenarios and their associated attributes. A distinction is made with respect to scale and loop length at each of the Access Node ‘Levels’. For example, an Access Node placed in the CO may serve several thousand loops with lengths up to 18 kft or ~ 5.5km (or more), while a “Level 3 Remote” will serve several dozen loops with lengths of about 3 kft or ~ 900m. The table appended to Figure 6 provides definitions of the attributes of each ‘Level’ of this hierarchy.

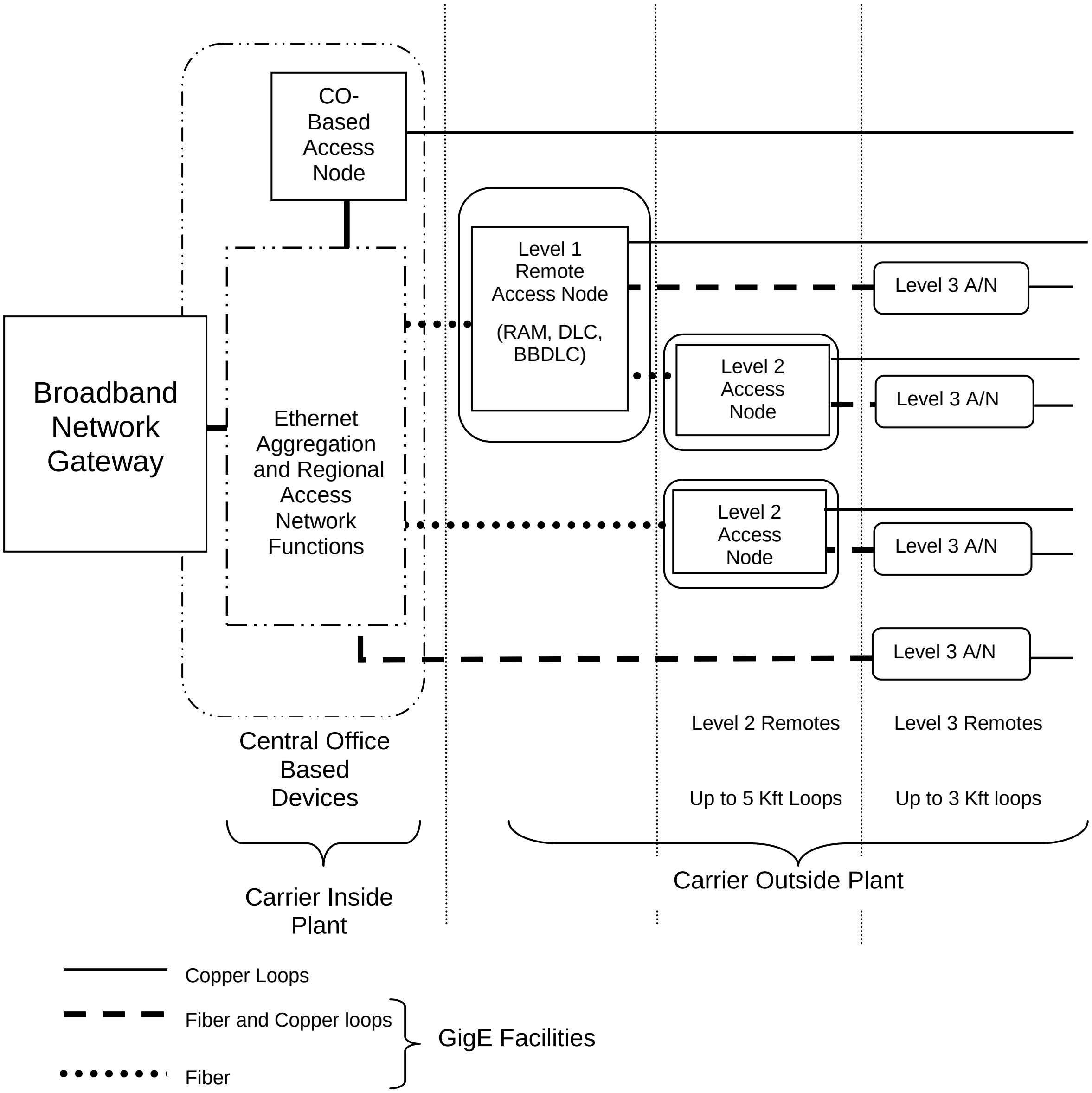
Remotes at a ‘higher level’ (i.e. a level 2 or 3 remote, vs. a level 1 remote) may optionally be subtended from a lower ‘level’ or may connect directly to an Ethernet aggregation function in the CO. This will allow a system that properly distributes the complexity of Ethernet Aggregation between the elements of the Access Network and Access Nodes themselves. This TR neither requires, nor precludes subtending architectures based on Ethernet transport to remotes. However this document specifically does not support the use of ATM to provide subtending connections between remotes. Ethernet aggregation functions may occur in Access Nodes of any ‘level’, may be centralized in an Ethernet Aggregation Node, or may be distributed between the Access Nodes, and other network elements. Uplinks of the lower “level” remotes are likely to be fiber based, while the ‘higher level’ Remote may use either fiber or bonded copper uplinks.

A distinction must be made with regard to requirements for the support of Multicast services at the Access Node between Transport plane capacity and Control plane capacity. A ‘higher level’ remote, when compared with one of ‘lower level’, will most likely require significantly less bandwidth to transport Multicast capabilities on the uplink from the remote and at the same time is likely to experience less saving on bandwidth usage due to several users viewing the same content simultaneously. However, even on the smaller ‘level 3” remotes the control message traffic for multicast services may still be considerable, especially when compared to the small scale size and complexity expected of the smaller remotes. This is because of the nature of user practice when selecting video programming. Channel surfing generates significant Control plane traffic from a user and the human factors expectations from these users with respect to response times are extremely stringent. The network architecture must be able to balance the need for simplicity in the level 2 and 3 remotes with issues of scale in centralized components that support the Multicast Control plane.

The redundancy expected of an Access Node at any particular ‘level’ is primarily a function of the number of user ports served. The ‘lower levels’ support large numbers of users and must support the level of redundancy and fault tolerance expected of major network elements. ‘Higher level’ Access Nodes are points of failure for few users and thus may have less stringent redundancy requirements placed on the element. Power efficiency requirements increase as ‘level’ increases. Level 3 remotes and smaller level 2 remotes may be loop powered and therefore require a high level of power efficiency.

It should be noted that ATM subtended Access Nodes are not included in this architecture as they introduce significant traffic engineering complexity and are in general inconsistent with the goals of this architecture. Subtending of Ethernet remotes is supported if the aggregating Access Node includes the Ethernet switch function (e.g. Access Node and Ethernet Aggregation Node are the same device). Additionally, any subtended device whose uplink facility capacity is less than the host device’s uplink facility introduces a potential congestion point where Hierarchical Scheduling may be required to ensure IP QoS is enforced as described in Section 2.9. Subtending arrangements where the facility capacities are equal or greater than those of the subtended devices pose no additional IP QoS requirements.

User isolation must also be maintained in subtended architectures. Such arrangements must not introduce undesirable bridging among user traffic. See Section 3 for more details.



1The CO based functions may be located in physically different COs
2The Ethernet Aggregation Function may be located in discrete devices, in the access nodes, or distributed throughout the Carriers Access Network.

Access Node 'Level'	CO Based	Level 1 Remote	Level 2 Remote	Level 3 Remote
Max Loop Length	Up to 18 Kft	Up to 12 Kft	Up to 5 Kft	Up to 3 Kft
Subs per device	High 100's to 1000's	100's to 1000's	High 10's to 100's	10 to High 10's
Multicast Capability Bandwidth Requirements	High	Medium	Medium	Low
Multicast Capability Control Plane Capacity	Very High	Very High	Medium	Medium

Redundancy Requirements	High	High	Medium	Low
Power Efficiency Requirements	Low (CO Powered)	Medium (Local Powered)	Medium to High (Local or Loop Power)	High (Loop Powered)

Figure 6 – Access Node deployment scenarios

2.5 The V interface

The fundamental requirements of the V interface are to provide the following capabilities:

1. Traffic Aggregation
2. Class of Service distinction
3. User isolation and traceability

Since the aggregation network is Ethernet-based and the Access Node and Broadband Network Gateway are both equipped with standard Ethernet interfaces, it is natural to try to address the above requirements with existing Ethernet layer mechanisms. The standard virtualization mechanism for Ethernet networks is VLANs. VLAN tagging was first defined in IEEE 802.1Q and then enhanced by 802.1ad. VLAN tagging can provide a means of aggregation by grouping a number of traffic flows under one VLAN with VID=X, and another group of traffic flows under another VLAN with VID=Y. VLAN tagging also enables marking of a frame's Class of Service (CoS) using the 3-bit priority field, thus providing a CoS distinction mechanism. The VLAN ID and priority assignment schemes are detailed in Section 2.5.1. VLAN tags can be used for user isolation and traceability by allocating a different VLAN ID to every Broadband port. Alternatively a VLAN tag may represent a certain group of users. In that case, other Layer 2 mechanisms for user isolation and traceability are used and those are discussed in the following sections.

The above considerations advocate that the V interface will be a tagged Ethernet Interface. This recommendation specifies the V interface to be an 802.1ad compliant provider network port. The main reasons are:

1. Scalability of the VLAN space. Allowing two levels of VLAN tags provides approximately 16 million combinations.
2. Support for L2 VPN services for business customers. The ability to encapsulate customer VLAN tagging with the provider's tagging is imperative for a transparent LAN service.

According to the 802.1ad terminology the outer (or only) VLAN tag on the V-interface is called an S-TAG (and similarly S-VID and S-VLAN) and the inner tag, when used, is called a C-TAG (and similarly C-VID and C-VLAN).

The overall tagging behavior of the Access Node may be seen as the 'Provider Edge Bridge' described in IEEE 802.1ad. The V-interface is a provider network port using S-VLANs. The U-interfaces are either customer edge ports (C-Tagged or untagged) or customer network ports (S-Tagged or untagged). Logically, a provider edge bridge is constructed from several C-VLAN aware components that are customer-facing, and those are internally connected to an S-VLAN aware component. The Aggregation Network nodes are described as 802.1ad S-VLAN bridges. Thus, C-Tags have significance only at the edges of the aggregation network, the BNG and the U-interface (at the Access Node side, and at the RG side depending on whether the U interface is tagged or not). At the BNG both S-Tags and C-Tags may be significant.

The protocol stack of the V interface is depicted in Figure 7.

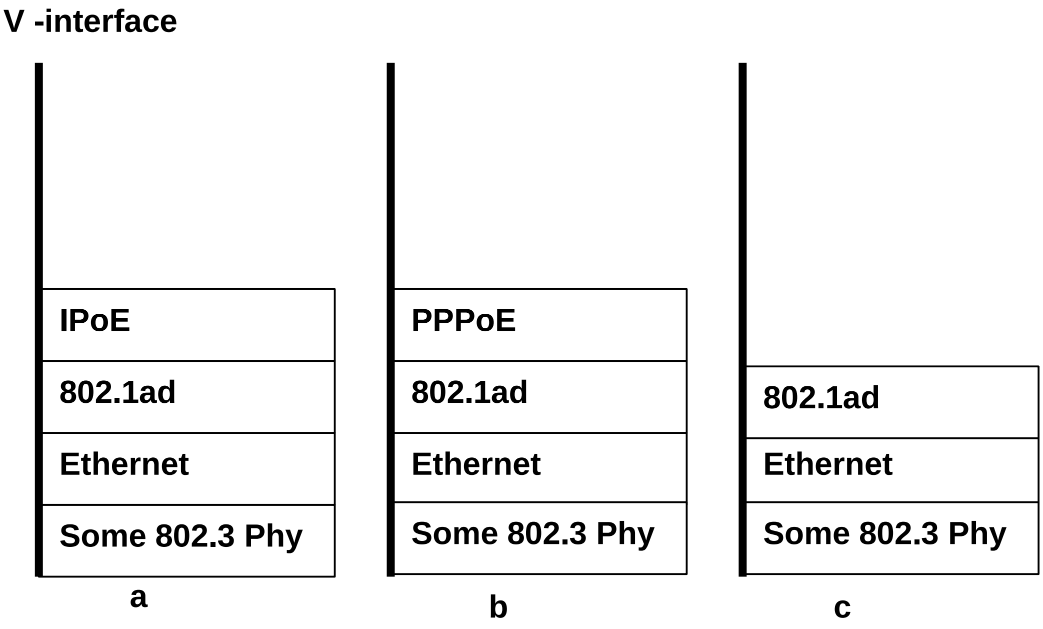


Figure 7 – Protocol stacks at the V interface

2.5.1 VLAN Architectures

While there are numerous VLAN assignment paradigms, the three fundamental approaches are described to reduce the scope of discussion. The following scenarios are described in the subsections below:

- ~~Multi-VCs DSL UNI~~ – ***Note: This should be avoided in new deployments***
- VLAN Tagged UNI
- Untagged/Priority Tagged UNI

In each case, both service connectivity and service prioritization are considered. Depending on the scenario, a user can be sending untagged frames, VLAN-tagged frames or priority-tagged frames. A priority-tagged frame is a tagged frame whose header carries priority information but carries no VLAN identification information. Depending on the type of user and the type of VLAN architecture used, the following VLAN configuration principles should apply:

For Business TLS users:

- Traffic may be received 802.1Q tagged on the user port
- Traffic must be S-Tagged for C-Tag transparency within the aggregation network
- The 802.1Q tag value must be preserved as the C-Tag
- The Access Node must apply an S-Tag if none is present, but the user may apply the S-Tag
- The S-Tag will be common to a TLS instance if switching is performed within the aggregation network.

For Other Business and Residential 1:1 VLAN users:

- The Access Node must apply at least an S-Tag

- The Access Node may apply a C-Tag for scalability reasons
- The S-Tag must not be shared across Access Nodes
- The S-Tag/C-Tag pair must be unique within Access Node

For Residential N:1 VLAN users:

- Traffic is single-tagged with an S-Tag throughout the aggregation network
- Each user port may have multiple sessions that can be carried in different VLANs
- The S-Tag may be common to more than one user port or user session. It can be shared by a large group of users as long as the Access Node and Access Network can maintain isolation between users. For example, an S-Tag may be common to:
    - All users attached to the same Access Node
    - Users sharing the same service
    - Traffic to/from a group of Access Nodes[1].

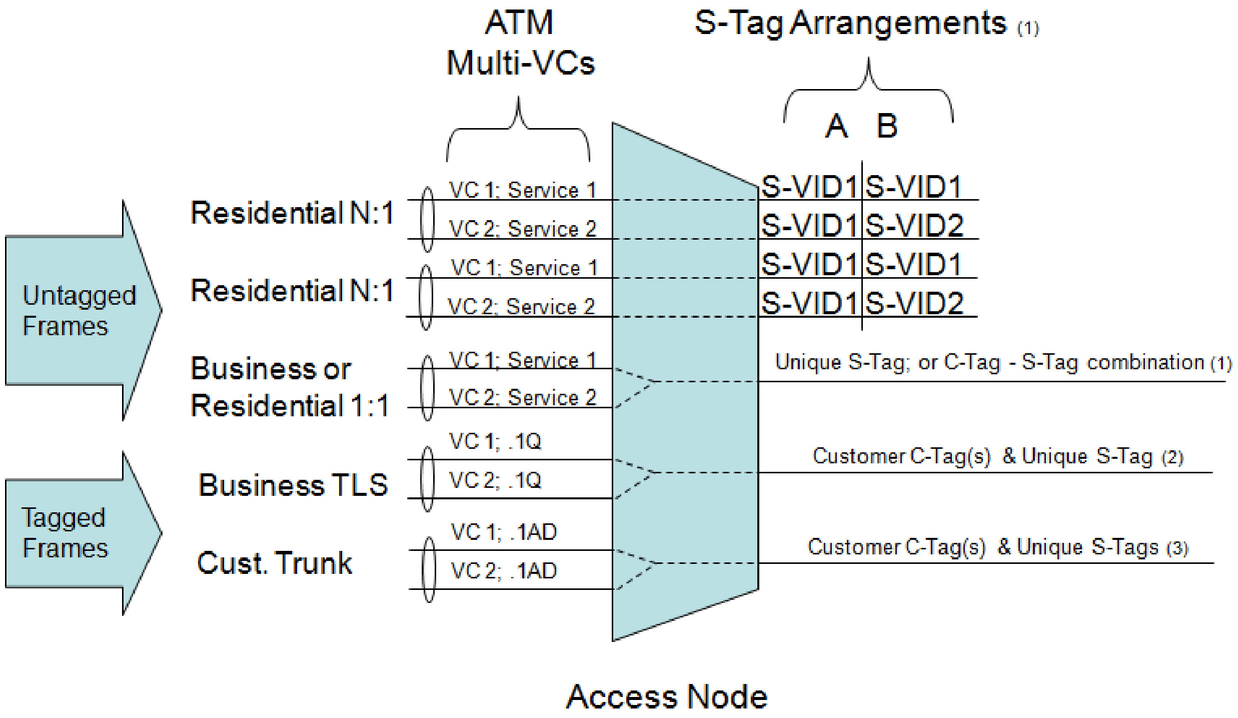
Note that this specification also allows for multipoint Layer 2 business service using N:1 VLANs. This option is not described in this section but is enabled by the requirements specified in the subsequent sections.

2.5.1.1 Multi-VC ATM Architecture

This section and others that describe the multi-VC architecture and requirements are provided as a historical reference only. The use of ATM VCs should be avoided in new deployments because they are only available on a small subset of the supported access protocols described by this document. As a consequence, Access Nodes that do not support Multi-VC access loop options need not support these functions and requirements.

> In a multi-service architecture, multi-VCs can be provisioned between the Access Node and RG. This assumes that the RG has the ability to map different traffic to different VCs. In such a case, the traffic is usually sent untagged as the Access Node can use the logical port (VC) on the customer-facing interface for VLAN assignment and priority marking.

[1] Grouping arbitrary topologies of access nodes within the same N:1 VLAN must be evaluated within the context of security, OAM, and QoS (hierarchal scheduling) complexities.



S-Tag Arrangements:

A – N:1 VLAN scenario where all subscribers are placed into a common VLAN

B – N:1 VLAN scenario where individual subscribers sessions are placed in a common VLAN based on service type (VC)

Notes:

(1) VC 1 & 2 may be mapped to different IEEE priority markings within the same VLAN or be mapped to separate VLANs (with or without frames being priority marked)

(2) VC 1 & 2 may be carrying the same or different VIDs resulting in 1 or more VLANs on V interface.

(3) VC 1 & 2 may be carrying the same or different S-VIDs resulting in multiple S-VIDs on the V interface.

Figure 8 – VLAN assignment in multi-VC architecture

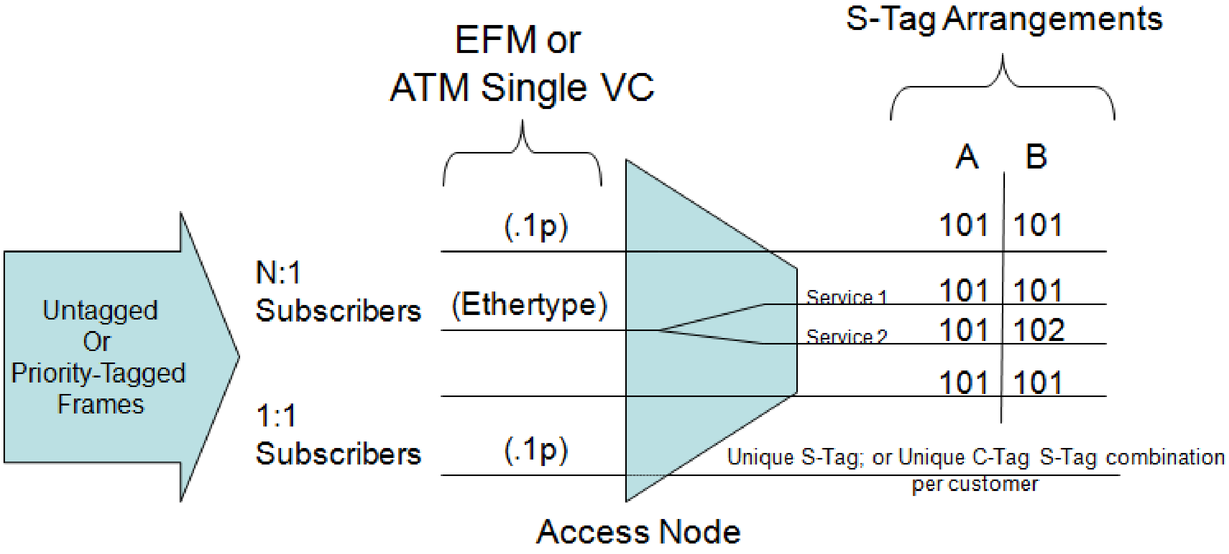
Note that when multiple VCs are used only for .1p CoS purposes (Figure 8, case A of the residential N:1, the residential 1:1 and the business TLS), there may be two or more VCs going from the same access loop to the same network VLAN. In such a case, the downstream forwarding at the Access Node can not be determined only by the VLAN or by the MAC forwarding table, but must also use the VLAN priority bits of the received frame. In the following, such a group of VCs is denoted a *PVC bundle*.

2.5.1.2 Untagged/Priority Tagged UNI

This refers to the case where untagged or priority tagged traffic is used between the Access Node and RG. If the access protocol is based on ATM, then a single ATM PVC is used – and typically with the same VPI/VCI value on every user port. The Access Node determines the VLAN assignment (most often related to service mapping) based on Ethertype or a static configuration on the Access Node so that all traffic is sent to the intended VLAN.

A decision based on the protocol/Ethertype of the incoming traffic is specifically geared towards supporting both PPPoE and IPoE sessions to the user where each 'connection' is mapped to a different VLAN based on the Ethertype value in the Ethernet header observed at the Access Node. In this scenario PPPoE could be used to carry typical data traffic and the IPoE traffic could be bridged into a N:1 video entertainment/multicast VLAN. This approach provides basic separation based on Ethertype.

Figure 9 illustrates the basic principle for VLAN assignment.



S-Tag Arrangements for N:1 Residential:

A – N:1 VLAN scenario where all subscribers are placed into a common VLAN

B – N:1 VLAN scenario where individual subscriber sessions are placed into a common VLAN based on service type

Figure 9 – VLAN assignment in untagged/priority-tagged architecture

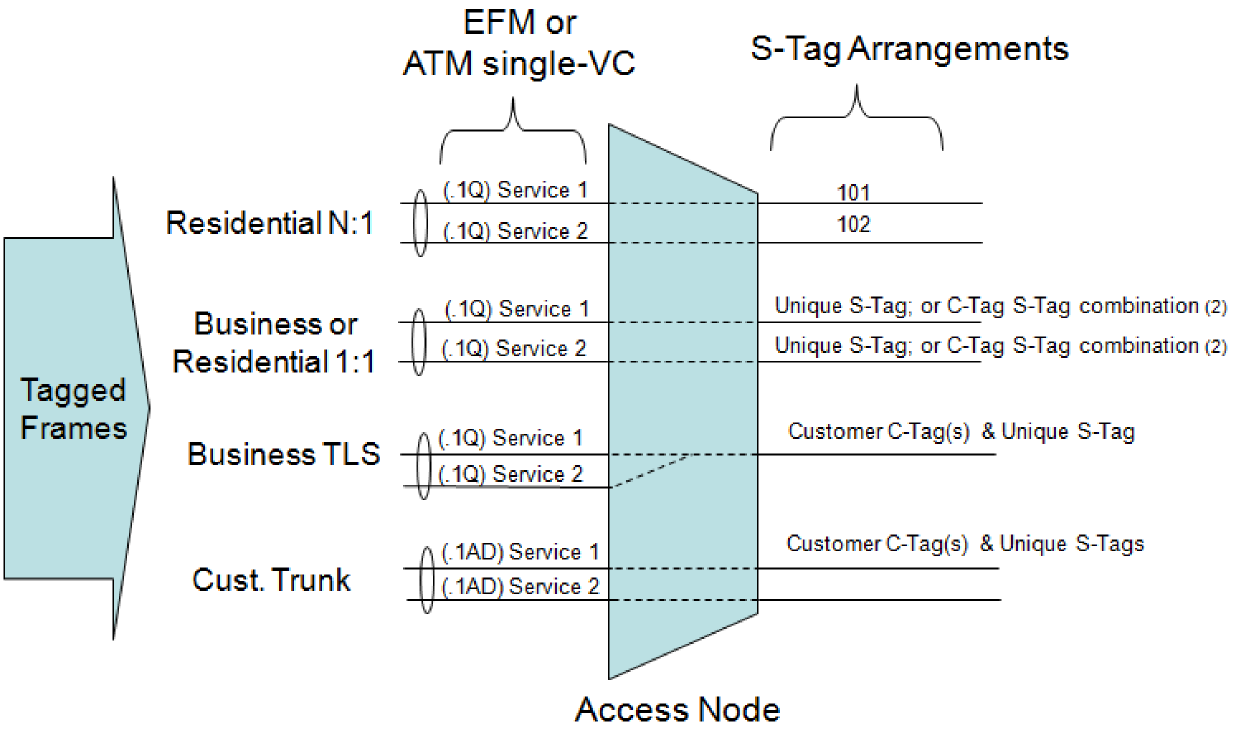
In Figure 9, the MAC layer over the ATM VC (or EFM encapsulation) may be carrying frames whose tag header contains a null VLAN ID and priority bits. A null VLAN ID is needed if priority indication is required but the VLAN ID itself has no importance or cannot be determined at the RG (i.e. classification is in the Access Node). In the case of N:1 with priority tagging, the priority tag is replaced with an S-Tag on the V-interface. The CoS indication of the priority tagged frame may be mapped to the CoS indication in the S-Tag.

In the 1:1 case, if the S-VLAN is not unique to this access loop, then the null VID has to be changed to a specific C-VID, thus making the (S-Tag, C-Tag) pair unique within the Access Node.

2.5.1.3 VLAN Tagged UNI

If more complex traffic separations are required than can be supported with untagged UNIs, then those separations should be done at the RG by setting 802.1Q tags. This approach allows the configuration of an Ethernet "trunk" between the Access Node and RG. The Access Node should be able to translate the access loop VLAN into a different C-Tag and/or S-Tag, allowing the same VLAN IDs to be used on the access loop for efficient configuration. Even more complex traffic separations can be supported using a Customer S-Tag Aware interface and double tagging at the U-interface.

Figure 10 illustrates the basic principle for VLAN assignment when a VLAN tagged UNI is used.



Notes:

(1) The non TLS scenarios above require VLAN (.1Q) based classification in the Access Node.

(2) Given that Aggregation Switches will only forward based on S-Tags, the unique C-Tag S-Tag combination refers to uniqueness of a S-Tag across ANs and the C-Tag S-Tag combination within an AN.

Figure 10 – VLAN assignment on tagged UNI

In the N:1 case, the Q-Tag received on the broadband port is replaced with an S-Tag on the V-interface.

2.6 Ethernet Aggregation Network

The Ethernet based aggregation network provides the following capabilities.

R-05 The aggregation network MUST provide a means to prioritize traffic in order to handle congestion at points of over-subscription.

R-06 The aggregation network MUST be able to support multicast deployments.

R-07 The aggregation network MUST provide support for high availability (e.g. by multi-homing).

R-08 The nodes within the aggregation network MUST be able to perform S-VLAN aware bridging, according to 802.1ad principles.

R-09 The aggregation network MUST maintain user isolation.

The aggregation network architecture must support the commonly deployed access solutions seen in Ethernet Broadband and metro Ethernet networks.

The topology of the Ethernet aggregation network is not specified in this document and may include several variations. The requirements included in this document are independent of the physical topology of the aggregation network. Some typical reference configurations are shown in the figure below.

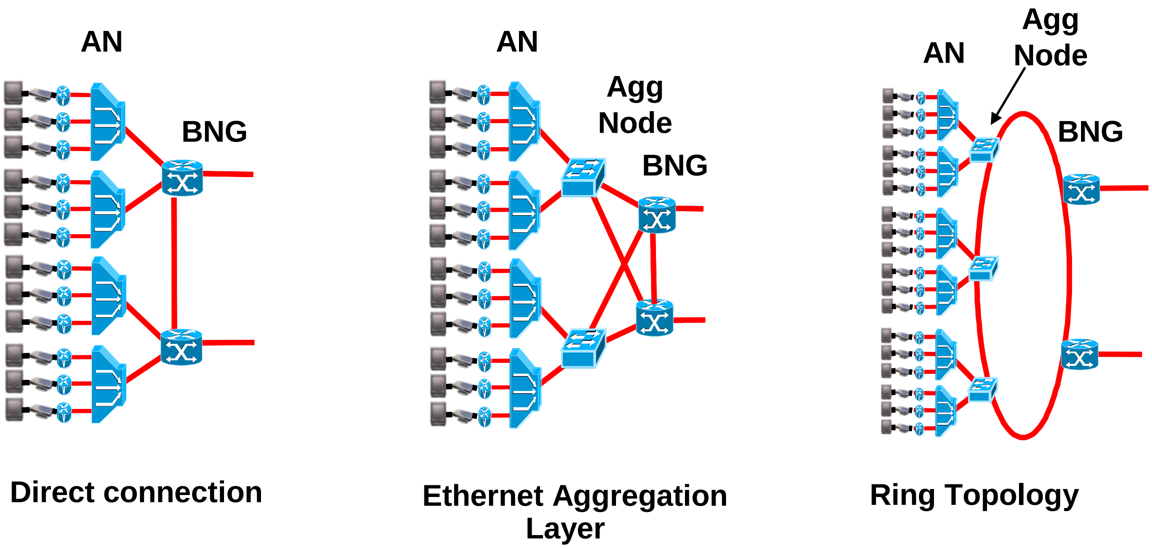


Figure 11 – Example aggregation architecture options

2.7 Broadband Network Gateways

The architecture described in this document supports a dual Broadband Network Gateway scenario when required for video optimization. Such an approach has additional complexities not present with a single Broadband Network Gateway arrangement. This is most notable with respect to traffic/bandwidth management and the resulting network efficiency as well as the additional operations overhead. In dual node architectures, it is not mandated that *both* BNGs support all of the requirements detailed in this document. Specifically, the video BNG does not need to implement subscriber management functions (e.g. PPP termination, per user QoS) given that these functions are likely performed by the other BNG. A multi-node (more than 2) deployment could follow the same recommendations described within this document, but will not be explicitly described.

The following are the high level Broadband Network Gateway requirements for this architecture:

R-10 The Broadband Network Gateway MUST be able to terminate the Ethernet layer and corresponding encapsulation protocols.

R-11 The Broadband Network Gateway MUST be able to implement the counterpart of the functions added to the Access Node for access loop identification, Ethernet-based QoS, security and OAM.

R-12 Following TR-059 QoS principles, the Broadband Network Gateway SHOULD be able to extend its QoS and congestion management logic (e.g. hierarchical scheduler) to address over-subscribed Ethernet-based topologies.

The approach to supporting hierarchy and the number of levels required may differ from that used in ATM based networks. It is not expected that the video Broadband Network Gateway would support this capability.

2.8 Multicast Architecture

One of the major drivers for migrating to an Ethernet network is the increased ability to provide entertainment video (broadcast and unicast). Therefore supporting multicast in a uniform way independent of the VLAN and physical arrangement employed by the service provider must be defined. The architecture described in this document aims to support:

- Efficient L2 replication in the Ethernet aggregation network through the use of N:1 VLANs
- Not requiring a separate multicast-VC on the last mile
- IGMP V2 and V3
- ASM and SSM models
- multiple content injection points in the network
- multiple multicast VLANs in the access network
- IP (and IGMP) packets being directly encapsulated in Ethernet frames
- IGMP hosts being connected to Access Node ports that are members of the multipoint VLAN that will carry (receive) the multicast frames
- IGMP packets being transmitted in the same VLAN from which the multicast packet will be received
- User ports being members of multiple VLANs.

2.9 QoS support

Three models for supporting QoS are provided below. It should be noted that these models can be used in combination within a given system and are not exclusive mechanisms.

1. Hard partitioning of bandwidth among the Broadband Network Gateways – rate limits can be established based on business or application characteristics of the Broadband Network Gateways that control each partition.

2. Distributed precedence and scheduling – mark services according to a Layer 2 precedence relationship so lower classes will be dropped under congestion. This model is applicable to single and dual Broadband Network Gateway deployments. This model can provide weighted fairness between classes of the same precedence but cannot provide fairness between users or instances of an application within a class. Precedence can be used to enhance option 1. In this case bursts above the partitioned rate can be allowed for discard-eligible traffic class(es) and priority used so that the burst traffic will only be delivered if there is excess bandwidth not being utilized by the other Broadband Network Gateway’s traffic. This approach is roughly the equivalent of allowing an ATM UBR traffic class to burst into the allocated, but unused bandwidth of other ATM traffic classes. Figure 12 provides a high level example of how this model could be used in a dual node environment.

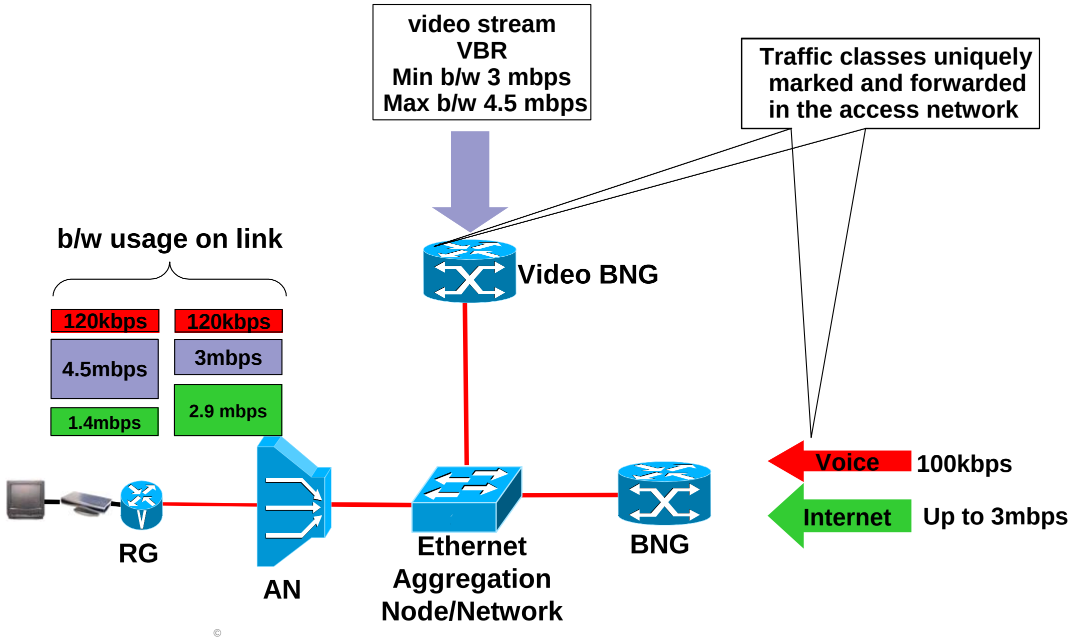


Figure 12 – Example distributed precedence and scheduling model with dual nodes

3. Hierarchical Scheduling (HS) within a BNG-allotted bandwidth partition. This option is applicable in the single BNG scenario (where the partition is equal to the access loop) or in the dual BNG scenario (where the partition is some fraction of the access loop's bandwidth). HS must be supported on the BNG with the possible exception of a BNG that only sources video. The HS must, at a minimum, support 3 levels modeling the BNG port, Access Node uplink, and broadband loop data rate. It should be noted that this option is required in order to meet the business QoS requirements set forth in TR-058. The rationale for the support of HS is as follows:
    a. Provide fairness of network resources within a class – including fairness across users and application instances.
    b. Better utilization of network resources. Drop traffic early at the Broadband Network Gateway rather than letting it traverse the aggregation network just to be dropped at the Access Node.
    c. Enable more flexible CoS behaviors other than only strict priority.
    d. The HS system could be augmented to provide per application admission control.
    e. Allow fully dynamic bandwidth partitioning between the various applications (as opposed to static bandwidth partitioning).
    f. Support "per user weighted scheduling" to allow differentiated SLAs (e.g. business services) within a given traffic class.

Hierarchal Policing (beyond the session and flow level) is not required given the Access Node topology options discussed in Section 2.4 and based on the significant asymmetry of bandwidth provisioned for the delivery of video traffic. Specifically, the access network will typically be provisioned with bi-directional Gigabit Ethernet links sized largely for the delivery of video traffic, therefore, leaving the upstream links with low utilization and unlikely to experience congestion. If a large number of symmetric physical access technologies are used, then that may invalidate these assumptions. In such cases it may be desirable to add support for hierarchical policing.

2.10 Business Services Support

Business class services need the following attributes:

- Availability – in order to make broadband a viable alternative to legacy data services, the overall availability of the access loop and the transport services provided must meet the performance levels expected from those legacy services. The overall availability can be increased by making the transport and/or network elements themselves highly available.
- Increased and symmetrical bandwidths – as more fiber is deployed the copper loop distance will reduced which coupled with advances in modulation schemes, will result in bitrates increasing to a level that is much more attractive to the business segment. Similarly, symmetric bandwidths can be offered which are more applicable to business usage (e.g. interconnecting remote sites).
- SLAs - Differentiated SLAs per site and per traffic class.
- Layer 2 VPN – Interconnecting remote sites in a seamless way at Layer 2 (Ethernet MAC) is a core requirement for some business customers. Such a service would provide a virtual extension to a customer’s own Ethernet networks across the service provider’s network including support for the customer’s VLAN and Ethernet priority tags.

2.11 Policy Management

The desire is to keep the dynamic policy management of the Access Node as simple as possible, enabling efficient highly distributed network deployment scenarios. Even without dynamic policy management on the Access Node, services like Bandwidth on Demand and QoS on demand can still be supported by using policy at the BNG and the RG (similar to the approach described in TR-059).

More sophisticated policy management capabilities are supported by TR-144 architectures.

3. Access Node Requirements

This section provides a set of requirements in support of the architecture defined in Section 2.

3.1 VLANs

As the aggregation network nodes are S-VLAN aware provider bridges, the Access Node and the BNG have to be S-VLAN aware. In some scenarios they may also have C-VLAN awareness, depending on the number of broadband ports and Access Nodes aggregated and the forwarding model (1:1 or N:1).

3.1.1 VLAN ID and Priority Assignment Capabilities

3.1.1.1 General Capabilities

R-13 The Access Node MUST support adding an S-Tag to untagged frames received on user ports in the upstream direction.

R-14 The Access Node MUST support adding an S-Tag and C-Tag to untagged frames received on user ports in the upstream direction.

R-15 The Access Node MUST support adding an S-Tag to C-Tagged frames received on user ports in the upstream direction.

R-16 The Access Node MUST support removing VLAN Tags from frames received from the aggregation network (i.e. the downstream direction) before sending them out on user ports. The options for removal are S-Tag only, or both S-Tag and C-Tag.

R-17 The Ethertype field for the 802.1ad tagging, i.e. S-Tags, MUST by default use the standardized value, 0x88a8. However, for backward compatibility, this field SHOULD be configurable per Port[2].

R-18 The Access Node MUST support configuring each U Interface port as one of:
Standard (see 3.1.1.2),
VLAN Transparent (see 3.1.1.3),
Customer S-Tag Aware (see 3.1.1.4).

3.1.1.2 Standard Ports

Standard Ports are used in most Internet Access applications.

R-19 The Access Node MUST support per port configuration of the acceptable frame types to one of the following values:
– C-VLAN tagged,
– Untagged,
– Priority-tagged, or
– accept-all (i.e. accepting VLAN-tagged, untagged and priority-tagged frames).

R-20 Frames not matching the configured acceptable frame types MUST be discarded.

[2] Note that care must be taken to ensure that S-Tags and C-Tags can be unambiguously discerned at any given port.

3.1.1.2.1 Handling Untagged and Priority Tagged Frames

R-21 For each port configured as untagged, priority-tagged or admit-all, the Access Node MUST support configuring the addition of an S-VID, or an S-VID, C-VID pair of tags to received untagged frames and the addition of an S-VID or an S-VID and setting the C-VID field in the priority-tagged frames.

R-22 For each port configured as untagged, priority-tagged or admit-all, the Access Node MUST support configuring whether it should copy the priority marking of the received upstream priority-tagged frame to the S-tag (and C-tag, if applicable) or whether it should overwrite it using an ingress to egress priority mapping (see Section 3.3.1).

R-23 For each port configured as untagged, priority-tagged, or admit all, the Access Node MUST support configuration of the following S-Tag parameters:

i. S-VID,

ii. S-Tag priority,

iii. DEI.

R-24 For each port configured as untagged, priority-tagged, or admit all, the Access Node MUST support configuring the following C-Tag parameters:

i. C-VID,

ii. C-Tag priority.

R-23 and R-24 provide flexibility in assigning VLAN priority to untagged frames, for example, assigning VLAN priority according to the [virtual] port's (i.e. PVC) ATM CoS (e.g. 6 for CBR, 5 for VBR-RT etc.). These capabilities are only available on ATM loops, and should not be used in new deployments.

The S-VID and the C-VID, if applicable, described in R-23 and R-24 above, are denoted the port's 'default tagging'.

R-25 Any untagged or priority-tagged frame received on port configured as 'untagged or priority-tagged' or 'admit all' MUST be tagged with the default tagging, unless it matches an Ethertype filter associated with this port.

R-26 The priority markings described in R-23 and R-24 (if applicable) MUST be applied to all received untagged frames.

R-27 Any frame destined for a given user port (i.e. in the downstream direction), carrying the port's default tagging MUST be forwarded as an untagged frame.

3.1.1.2.1.1 Protocol Based VLAN Assignment

The following requirements address the case where both bridged encapsulations, IPoE and PPPoE, are multiplexed over a single user port, but require different VLAN ID and/or priority assignment. The following describes a basic classification mechanism only applicable for untagged and priority-tagged frames (and thus for ports configured to receive those types).

R-28 The Access Node MUST support configuration of an Ethertype classifier to a given port.

At least the following types MUST be supported:

- PPPoE (Ethertype =0x8863 and 0x8864)
- IPoE (Ethertype=0x0800 and 0x86DD covering IPv4 and IPv6, respectively).
- ARP (Ethertype=0x0806)

R-29 Once a frame is classified by Ethertype the Access Node MUST support configuring the S-VID or S-VID and C-VID pair that will be used for tagging the frame.

R-30 The Access Node MUST support configuring whether to mark the Ethertype classified frames according to the received priority or the outcome of ingress to egress priority mapping (see Section 3.3.1), respective to the R-22 option selected.

The above VID(s) and priority are used by the Access Node to create the S-Tag and C-Tag (if applicable) that is added to the frame. In case of priority-tagged frames, the S-Tag replaces the received priority tag.

R-31 Any frame destined for a given user port (i.e. in the downstream direction), carrying a protocol-assigned VLAN MUST be sent out as an untagged frame.

3.1.1.2.2 Handling VLAN Tagged Frames

R-32 For each port configured to admit VLAN-tagged frames (i.e. acceptable frame type of C-VLAN tagged, or admit all), the Access Node MUST support configuration of a list of S-VIDs and C-VIDs, denoted as the port's *VLAN membership list*, that are acceptable for this port. In this case the Access Node MUST discard any VLAN-tagged frame received from a port with non-compliant C-VID, or S-VID.

R-33 For each port configured to admit VLAN-tagged frames, the Access Node MUST support configuring a VLAN translation table consisting of an entry for each VLAN in the port's VLAN membership list and VID value(s) to translate it to. This table can be used for:
a. Indicating an S-VID to replace the U-interface C-VID, if the C-Tag needs to be replaced with an S-Tag;
b. Indicating both a C-VID and an S-VID, if the U-interface C-VID has to be overwritten and the frame also needs an S-Tag to be attached.

R-34 For each C-VID in a given port VLAN membership list, the Access Node MUST support configuring whether to accept (i.e. forward 'as is') the received VLAN priority markings or rewrite the priority using an ingress to egress priority mapping (see Section 3.3.1).

R-35 In the downstream direction the Access Node MUST perform the reverse translation and required tag modification described in R-33 in order to reproduce the U-interface C-VIDs. The priority marking of the received downstream frame C-Tag, however, MUST NOT be modified.

Note that one must make sure that translations are not ambiguous.

3.1.1.3 VLAN Transparent Ports

A VLAN transparent port, a.k.a. TLS port, is a user facing port for which at least a portion of the traffic received on the U-interface (the TLS traffic) is forwarded without any modification of the original frame payload or header and without the Access Node being pre-configured with its VLAN identification information (if present).

A VLAN transparent port may have a mix of TLS and non-TLS traffic. The non-TLS traffic must be VLAN-tagged whereas the TLS traffic can be of any type (i.e. double-tagged, VLAN-tagged, untagged or priority tagged).

R-36 For ports that are provisioned VLAN Transparent the Access Node MUST support per port configuration of the acceptable frame types to one of the following values:
– C-VLAN tagged,
– Untagged,
– Priority-tagged, or
– accept-all (i.e. accepting VLAN-tagged, untagged and priority-tagged frames).

R-37 For each VLAN transparent port, the Access Node MUST support configuration of a list of C-VIDs, denoted as the port's *VLAN membership list*, that are allocated for non-TLS traffic. Any frames received on a user-facing TLS-enabled port that are untagged or do not match a C-VID from the port's VLAN membership list MUST be forwarded as TLS traffic.

R-38 For each VLAN transparent port, the Access Node MUST support configuration of one S-VID, denoted the 'TLS S-VID', used to encapsulate TLS traffic.

R-39 The Access Node MUST support using the same TLS S-VIDs among multiple ports belonging to the same TLS.

R-40 For each VLAN transparent port, the Access Node MUST support configuration of one of the following priority-marking options that will be used for marking the S-Tag encapsulating tagged TLS traffic:
1. Ingress to egress Priority mapping (see Section 3.3.1).
2. Copy ingress (802.1Q tag) priority to S-Tag.

R-41 The Access Node MUST support configuring the priority to be used for marking untagged ingress frames.

The S-VID and priority marking described in R-38, R-40 and R-41 are used for tagging all TLS frames received on the VLAN transparent port and sent out through the V interface.

R-42 For each configured VLAN transparent port, the Access Node MUST support configuring a VLAN translation table with an entry for each VLAN in the port's VLAN membership list. This table defines the translation between:
a. A U-interface C-VID and a V-interface S-VID,
b. A U-interface C-VID and a V-interface S-VID-C-VID pair.

R-43 For each C-VID or S-C-VID pair in a given port VLAN membership list, the Access Node MUST support configuring whether to accept (i.e. forward 'as is') the received VLAN priority markings or to rewrite the priority markings using an ingress to egress priority mapping (see Section 3.3.1).

R-44 In the downstream direction the Access Node MUST perform the reverse translation and required tag modification described in R-42 in order to reproduce the U-interface VIDs.

R-45 The priority marking of the received downstream frame MUST NOT be modified.

3.1.1.4 Customer S-Tag Aware Ports

A Customer S-Tag Aware Port, a.k.a. Trunking Port, is a U Interface port for which a customer may use a number of S-Tags with traffic in order to associate that data with one or more networking services including TLS. All traffic must be S-tagged at the U Interface.

The non-TLS traffic is single-tagged with an S-Tag that infers the service. For example, one value of an S-Tag may imply an Internet access service, and a different value may imply a managed VoIP service. For these non-TLS services, the S-Tag may be used as-is by the network, or it may be replaced (re-written) with an alternate S-Tag or a combination of C-Tag and S-Tag. The latter is the equivalent of double tagging of untagged traffic described under Standard Ports.

The TLS traffic must also have an S-Tag, but can utilize any of the C-Tag options previously described under VLAN Transparent Ports.

R-46 For each Trunking port, the Access Node MUST support configuration of a list of S-VIDs that it will accept, denoted as the port's *Service Membership List*.

R-47 For each configured Customer S-Tag Aware Port, the Access Node MUST support configuring a VLAN translation table with an entry for each VLAN in the port's Service Membership List. This table defines the translation between:
a. A U-interface and V-interface S-VID for any service type.
b. A U-interface S-VID and V-interface S-VID-C-VID pair for non-TLS service.
Note: For Non-TLS Service the CPE should not attach a C-Tag.

R-48 For each S-VID in a given port's Service Membership List, the Access Node MUST support configuring whether to accept (i.e. forward 'as is') the received VLAN priority markings or to rewrite the priority markings using an ingress to egress priority mapping (see Section 3.3.1).

R-49 In the downstream direction the Access Node MUST perform the reverse translation and required tag modification described in R-47 in order to reproduce the U-interface VIDs.

R-50 The priority marking of the received downstream frame MUST NOT be modified.

3.1.2 VLAN Allocation Paradigms

R-51 The Access Node MUST support configuring the same S-VID (and no C-VID) to a group of ports. This paradigm is denoted *N:1 VLAN* to indicate many-to-one mapping between ports and VLAN. Example criteria for grouping are same service and same service provider.

R-52 The Access Node MUST support configuring a unique VLAN identification to a port using either a unique S-VID or a unique (S-VID, C-VID) pair. The uniqueness of the S-VID MUST be maintained in the aggregation network.

Note: This paradigm is denoted *1:1 VLAN* to indicate a one-to-one mapping between port and VLAN.

3.2 Access Node Forwarding Mechanisms

3.2.1 General

R-53 The Access Node MUST support Link Aggregation according to 802.3ad for link resiliency.

R-54 The Access Node SHOULD support Link Aggregation according to 802.3ad for load balancing.

R-55 An Access Node SHOULD support Rapid Spanning Tree as per IEEE 802.1D (2004 edition).

R-56 An Access Node having multiple uplinks which are not members of the same 802.3ad link aggregation group SHOULD be able to perform VLAN aware bridging between its uplinks.

This requirement enables deploying Access Nodes in a ring topology.

R-57 The Access Node MUST support mini jumbo Ethernet frames of at least 1534 bytes total length (e.g. 1500 bytes PPP payload, PPPoE header and Ethernet framing with dual VLAN tags).

Following the N:1 and 1:1 paradigms, an Access Node needs to support two different forwarding schemes. These schemes will be detailed in the following sub-sections.

3.2.2 Forwarding in N:1 VLANs

According to this scheme, the Access Node is considered to be a VLAN aware bridge, where each N:1 S-VLAN is a separate Virtual Bridge (VB) instance. Each VB performs independent source MAC address learning and frame forwarding process as described in 802.1D and 802.1Q. Additional requirements, such as the use of specific filters, are specified in this document in order to form an 'access-optimized' forwarding process (e.g. L2 user isolation).

For a given N:1 VLAN , all ports bounded to it as well as the network uplink are considered as *bridge ports* of the corresponding VB.

R-58 The Access Node MUST support N:1 VLAN forwarding, i.e. determining the destination port according to the MAC address and the S-VID.

R-59 The Access Node MUST be able to prevent forwarding traffic between user ports (user isolation). If user isolation is provisionable, this behavior MUST be configurable per S-VID.

For forwarding on a PVC bundle, see Section 3.3.1.

3.2.3 Forwarding in 1:1 VLANs

Forwarding relies on the unique one-to-one binding between user port and a VLAN. In the upstream direction, a frame received on a port has one or two VLAN tags added and is sent to the network. In the downstream direction, a frame received from the network is stripped of one or two VLAN tags and sent to the corresponding port, after applying further framing adaptation if required. This 1:1 forwarding model can be mapped to the Independent VLAN Learning (IVL) model described in IEEE 802.1Q, using a unique C-VLAN per access port.

R-60 The Access Node MUST support 1:1 VLAN forwarding.

R-61 The Access Node MUST support a downstream mapping between S-VLAN and port.

R-62 The Access Node MUST support a downstream mapping between S-VLAN and C-VLAN pair and port.

R-63 The Access Node MUST be able to enable and disable MAC address learning for 1:1 VLANs.

For forwarding of a PVC bundle, see Section 3.3.1.

3.3 QoS

In an ATM-centric access node, every ATM PVC is assigned a traffic class (i.e. CBR, VBR-RT, VBR-nRT, or UBR) and a traffic profile (e.g. PCR, SCR, MBS etc.). In a typical implementation the above ATM traffic classes are scheduled with strict priority. Using Ethernet, the traffic class can be determined frame-by-frame by examining the VLAN tag priority field, providing 8 different priority values, which can map to a lower number of traffic classes (e.g. 2 to 4). According to IEEE 802.1D/2003, in a typical implementation, the above Ethernet traffic classes are scheduled with strict priority. A fundamental difference from ATM is that multiple Ethernet traffic classes may be multiplexed over a single VLAN (as opposed to a single ATM traffic class per ATM VC). Finally, it is desirable to retain the ATM and IP notions of marking drop-precedence of traffic for congestion management in the Access Node.

The following sub-sections discuss the required QoS implementation and the interworking between the different CoS methodologies.

R-64 The Access Node MUST support at least 4 traffic classes for Ethernet frames, and MUST support configurable mapping to these classes from the 8 possible values of the Ethernet priority field.

R-65 The Access Node SHOULD support at least 8 traffic classes for Ethernet frames, and MUST support configurable mapping to these classes from the 8 possible values of the Ethernet priority field.

R-66 The Access Node MUST support drop precedence within at least 2 traffic classes and MUST support configurable mapping to these classes and drop precedence from the 8 possible values of the Ethernet priority field.

R-67 The Access Node MUST support direct indication of drop precedence within all supported traffic classes based on the DEI bit value of the 802.1ad header.

R-68 The Access Node MUST support at least 4 queues per user facing port, one per traffic class.

Note: User-facing ports share a single set of queues across all VLANs. This does not imply 4 queues per VLAN.

R-69 The Access Node SHOULD support at least 8 queues per user facing port, one per traffic class.

R-70 The Access Node MUST support scheduling of user queues according to strict priority among at least 4 queues.

R-71 The Access Node SHOULD support scheduling of user queues according to their assigned priority and weight.

R-72 The number of priorities MUST be at least 4, however multiple queues may be assigned the same priority.

R-73 Queues assigned to the same priority MUST be scheduled according to a weighted algorithm (like WFQ) with weights assigned through provisioning.

This mechanism provides support for mapping diffserv PHBs (e.g. EF, AF, BE, LE) to the Ethernet queues. An example of a system that supports 4 queues is shown in the figure below. In this table, Queue 1 is scheduled at the highest priority, and since there are no other queues at that level, its weight is ignored. Queue 2 is similarly scheduled at priority 2. Once these two queues are exhausted, Queues 3 and 4 are scheduled with a weight ratio of 150:1. This approach is identical to the queuing arrangement standardized by the Broadband Forum for RGs.

Priority 1	**Queue 1 – 100**
Priority 2	**Queue 2 – 15000**
Priority 3	**Queue 3 – 15000**
	Queue 4 – 100
Priority 4	

Figure 13 – Example Scheduler

R-74 The Access Node MUST support at least 4 queues per network facing port, one per traffic class.

R-75 The Access Node SHOULD support at least 8 queues per network facing port, one per traffic class.

R-76 The Access Node MUST support scheduling of network queues according to strict priority among at least 4 queues.

R-77 The Access Node SHOULD support scheduling of network queues according to their assigned priority and weight. The number of priorities MUST be at least 4, however multiple queues may be assigned the same priority. Queues assigned to the same priority MUST be scheduled according to a weighted algorithm (like WFQ) with weights assigned through provisioning. This mechanism provides support for mapping diffserv PHBs (e.g. EF, AF, BE, LE) to the Ethernet queues.

R-78 The Access Node MUST support setting the maximum size/depth of all queues.

3.3.1 Traffic Classification and Class of Service Based Forwarding

R-79 The Access Node MUST be able to mark or re-mark the Ethernet priority bits based on the following classification criteria:

- User port (physical or logical)
- Ethertype (i.e. Ethernet Protocol ID)
- Received Ethernet priority bits
- IP protocol ID (specifically support classification of IGMP)

The following requirements address PVC bundles. Note that PVC bundles may be associated with either 1:1 or N:1 VLANs. In scenarios where a PVC bundle is used for .1p CoS purposes on the DSL line, that is, they are mapped to the same S-VLAN, the PVC bundle forms a single logical port. Downstream traffic destined for this port must be distributed over the individual VCs based on the .1p value of each frame.

R-80 For each VC belonging to a PVC bundle, the Access Node MUST support configuration of Ethernet priority values.

R-81 The Access Node MUST be able to distribute downstream traffic destined for a PVC bundle based on the Ethernet priority values of each frame.

NOTE: New deployments of TR-101 should avoid using PVC bundles because they are not available on many of the access technologies that this architecture supports.

3.4 Multicast Support

See Section 6.

3.5 Protocol Adaptation Functions

This section discusses the access protocols at the U interface, as described in TR-43, and how those are processed at the V interface.

R-82 Where the Access Node supports an ATM U Interface, it MUST be able to automatically sense the following protocol encapsulations to match the ADSL CPE modem's configuration.

1. PPPoE over ATM (RFC 2516/2684)
2. IPoE over ATM as per RFC 2684 bridge mode
3. ~~IP over ATM as per RFC 2684 routed mode~~
4. ~~PPP over ATM (RFC 2364)~~

R-83 The Access Node MUST be able to turn auto-sensing (described in R-82) on and off on a per port basis.

3.5.1 PPPoE over ATM (U-interface)

Figure 14 depicts the end-to-end protocol stacks associated with the PPPoE access method.

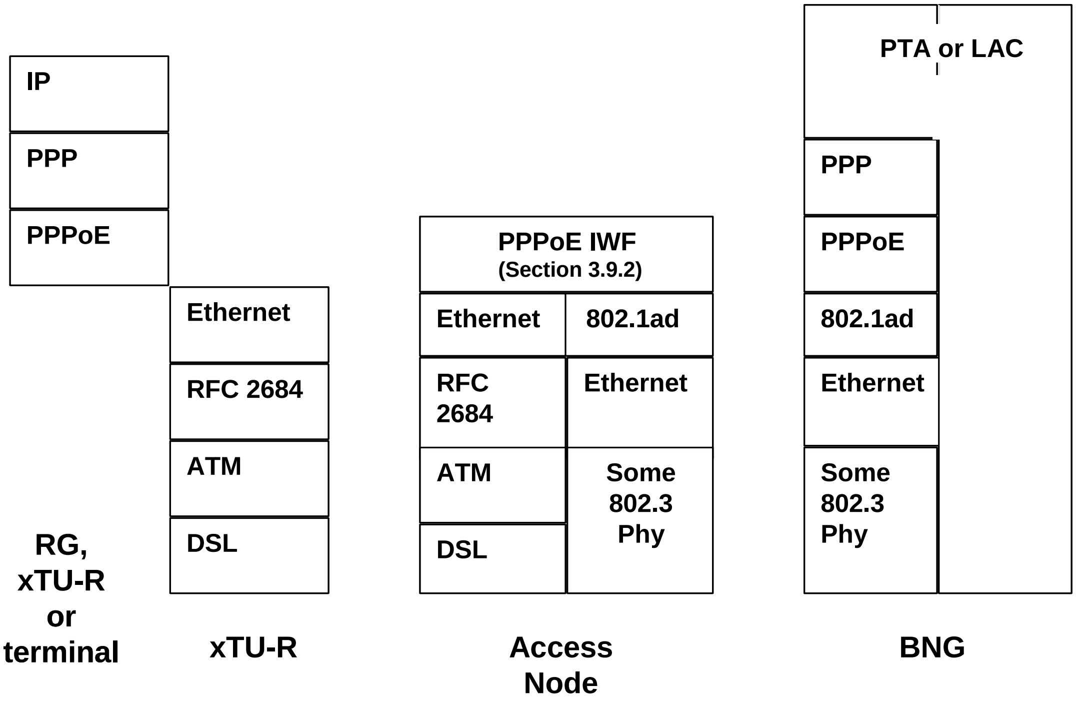


Figure 14 – End-to-end protocol processing for PPPoE access

3.5.2 IPoE over ATM (U-interface)

Figure 15 depicts the end-to-end protocol stacks associated with the IPoE access method.

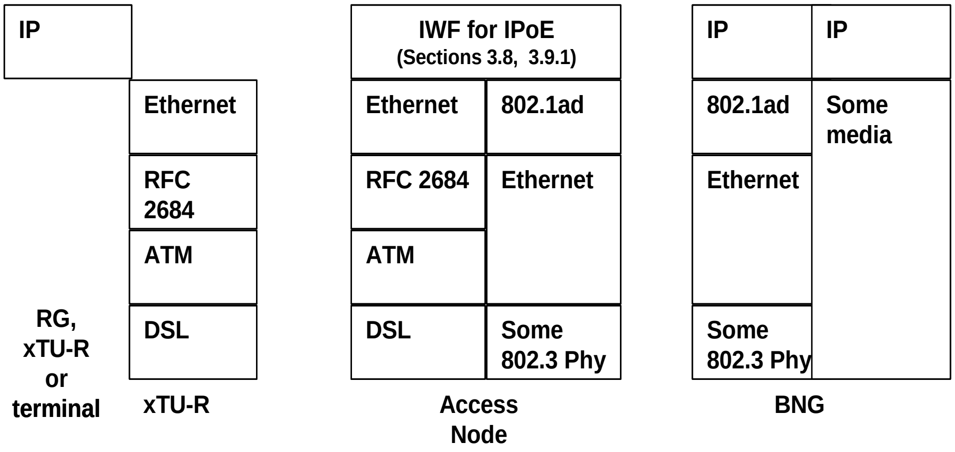


Figure 15 – End-to-end protocol processing for IPoE access

3.5.3 IP over ATM (U-interface)

NOTE: This access protocol should be avoided in new deployments.

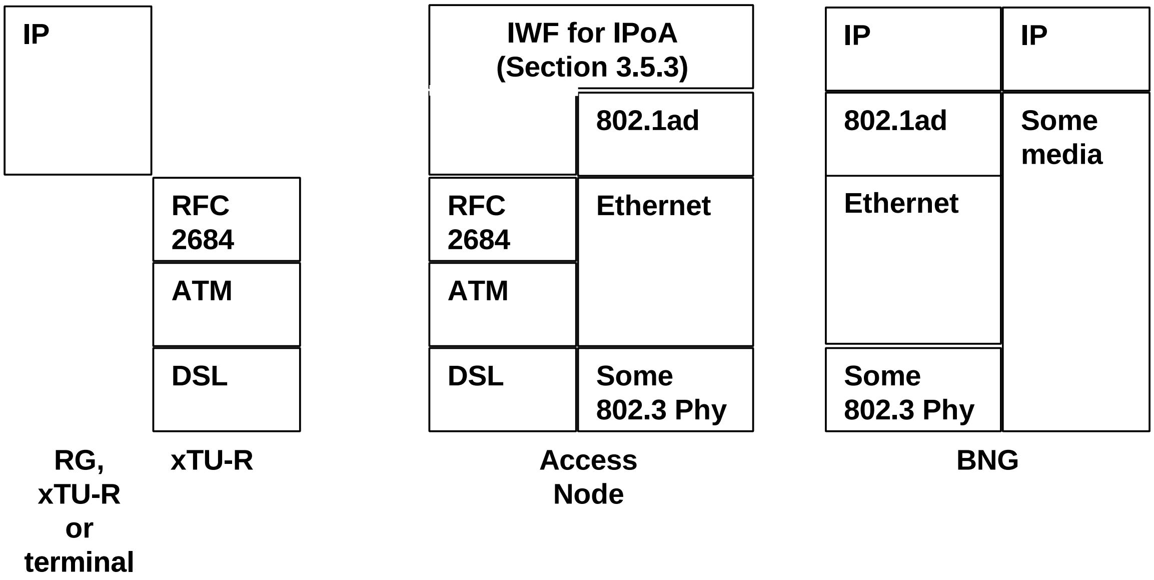


Figure 16 – End-to-end protocol processing for IPoA access

The use of IPoA encapsulation on the U-interface in legacy ATM access networks was predominantly applicable to business users. **This should now be avoided in new deployments**. The rest of this section is provided for historical reference. The business user typically made use of a /30 subnet between the RG and the BNG (cf. RFC 2225 “Classical IP and ARP over ATM”). IP addresses used in the customer network behind the RG were exchanged using routing protocols that run transparently over the ATM PVCs.

An IPoA Interworking Function provided initial continued support for this protocol stack in the original revision of this text. Figure 16 depicts the end-to-end protocol stacks associated with the IPoA access method. In line with the forwarding paradigm for business users in general, the IPoA IWF is based on the 1:1 VLAN paradigm, using a unique <C-VID,S-VID> pair per business user. With this model, the C-VID indicates the access loop and the S-VID indicates the Access Node. Since 1:1 VLANs are provisioned, routing protocols, if used between the RG and the BNG, will run transparently across the aggregation network up to the BNG.

The IPoA IWF converts the encapsulation from IPoA to IPoE, including adding the two VLAN tags identifying the port, and vice versa. When converting to IPoE, there is a need to use appropriate MAC source and destination addresses:

- The MAC source address of the packet is a MAC address under the control of the Access Node (e.g. the MAC address of the Access Node uplink).
- The MAC destination address of the packet depends on the packet type:
  - For unicast IP packets, the MAC destination address is the MAC address of a user-facing BNG interface that can be reached by the Access Node. The MAC destination address may be configured on the Access Node, or may be determined based on a pre-configured IP address of the BNG, which is then resolved using ARP;
  - For multicast and broadcast IP packets, the MAC destination address is automatically derived from the IP address, using the standard multicast/broadcast IP address to MAC address conversion algorithm.

In the downstream direction, packets are forwarded to the appropriate access loop based on the <C-VID,S-VID> pair. The Access Node remains transparent to the IP layer; the IP packet is extracted from the Ethernet encapsulation and then encapsulated into the selected ATM encapsulation.

In addition to this behavior, the BNG could send downstream ARP messages to resolve the MAC address of a user using IPoA-based access. In such a case, the IPoA IWF needs to respond with the MAC address that is used by the Access Node as the source address for upstream packets. This ensures that downstream packets will end up at the IPoA IWF, which can then forward them to the access loop as described earlier.

NOTE: The remaining requirements in this chapter only apply to Access Nodes that support loop technologies that use ATM. These capabilities are not usually required for new deployments.

R-84 The Access Node SHOULD support an IPoA IWF.

R-85 The IPoA IWF MUST be based on the 1:1 VLAN paradigm, using a unique <C-VID, S-VID> pair per business user; the C-VID indicates the access loop and the S-VID indicates the Access Node.

R-86 For upstream packets, the IPoA IWF MUST use a MAC source address that is under the control of the Access Node (e.g. the MAC address of the Access Node uplink).

R-87 For upstream unicast packets, the IPoA IWF MUST use a MAC unicast destination address of the BNG

R-88 For upstream multicast and broadcast packets, the IPoA IWF MUST derive the MAC destination address using the standard multicast/broadcast IP address to MAC address conversion algorithm.

R-89 Upon receiving ARP requests sent by the BNG, the IPoA IWF MUST be able to reply with the appropriate MAC address used as the source address for upstream packets.

3.5.4 PPP over ATM (U-interface)

NOTE that this access protocol should be avoided in new deployments.

The PPPoA access method is not layered on top of Ethernet and should be avoided in new deployments. The original version of this document required the Access Node to convert the PPP frames to a standard Ethernet protocol and framing. The protocol translation scheme is depicted in Figure 17. The approach taken was to perform conversion between PPPoA and PPPoE at the Access Node. This includes implementing a standard PPPoE control plane and setting-up a new PPPoE session with the BNG for each new PPPoA session, and encapsulating the PPP messages with PPPoE framing accordingly. Following this approach, the Access Node keeps state information for all configured PPPoA PVCs. Each connection may be in one of the following states:

1. **Disconnected** – In this state the Access Node blocks all traffic to and from the PVC. The Access Node ‘listens’ on the PVC and waits for an LCP Configure-Request message indicating the beginning of PPP session setup. When such a message is received, the Access Node stores it and sets up a new PPPoE session with the BNG, as described in RFC 2516. If the PPPoE session is successfully set-up (i.e. a session-id is received from the BNG) the Access Node transits to ‘connected’ state for this connection and the most recently received stored LCP frame is forwarded. When entering connected state the Access node records the mapping between the user PVC and the PPPoE session identification (i.e. PPPoE Session-ID, BNG MAC address, VLAN ID). If the PPPoE session setup has failed for any reason, the Access Node drops the LCP configure-request message and stays at ‘disconnected’ state.
2. **Connected** – Once a PPPoE session id is available (i.e. the PPPoE session stage), the Access Node can start forwarding upstream traffic received over the PVC to the network after adjusting the encapsulation, by adding Ethernet, VLAN and PPPoE headers and Ethernet FCS as a trailer. The Access Node can distinguish an interworked PPP session from native PPPoE traffic as the MAC destination of downstream PPPoE session packets equals an Access Node MAC Address. In the

downstream direction forwarding is performed using the mapping records mentioned above. The Access Node exits this state and returns to 'disconnected' in any of the following cases:

- Upon receipt of a PPPoE Active Discovery Terminate (PADT) for this session from the BNG (denoted in the following as graceful termination).
- Upon detection of non-graceful session termination
- Upon a DSL loss of synchronization.
- When the Access Node restarts.

One of the properties of this solution is that the Access Node does not have to go through the complexity of understanding and following the entire PPP state machine.

Figure 18 illustrates the state transitions as explained above and Figure 19 provides an example of end-to-end message flow (with graceful session termination). Note that the Access Node may be required to act as a PPPoE intermediate agent for the interworked sessions.

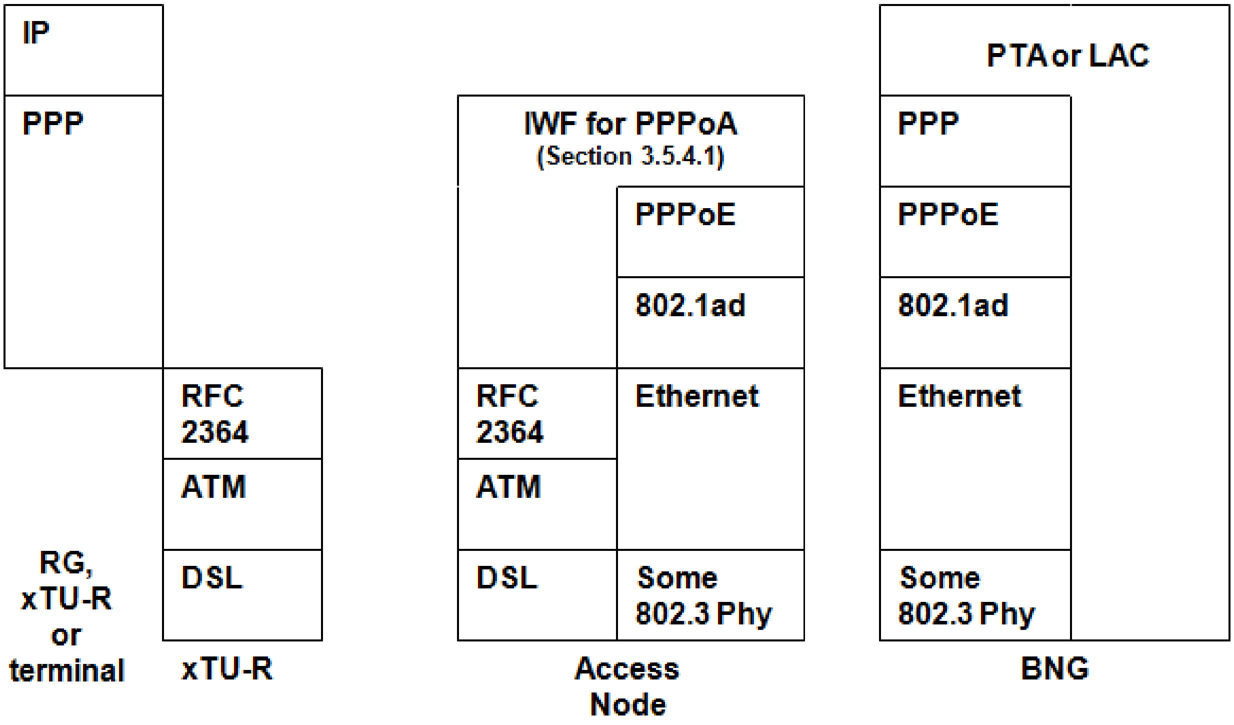


Figure 17 – End-to-end protocol processing for PPPoA access

The following requirements are for implementing the translation function. For clarity, both Access Node and BNG requirements are grouped into this section. Once again, requirements in Section 3.5.4 are not applicable to Access Nodes that do not support ATM on access loops.
This protocol should be avoided in new deployments.

3.5.4.1 Access Node Requirements for PPPoA IWF

R-90 The Access Node MUST set-up an interworked PPPoE session as per RFC 2516 to carry PPP frames received from an ATM VC configured on the access loop.

R-91 Once this interworked PPPoE session has been established, the Access Node MUST encapsulate all PPP frames it receives on this ATM VC into this interworked PPPoE session.

R-92 The Access Node MUST check Session-IDs received from a BNG to ensure the (Session-ID, BNG MAC address, Access Node MAC address, VLAN) 4-tuple is not already in use on the Access Node, and MUST remove the old session if it exists. This requirement aims to cover the case where a session was torn down at the BNG but is still considered active at the Access Node (i.e. when a PADT was lost but the session has not yet been timed out) and the BNG tries to 're-allocate' the same session-id to a new session.

R-93 The Access Node SHOULD store the latest LCP configure request until PPPoE negotiation completes successfully, at which point the Access Node forwards the most recent LCP configure request to the BNG.

R-94 The Access Node MUST support concurrent establishment of multiple interworked PPPoE sessions.

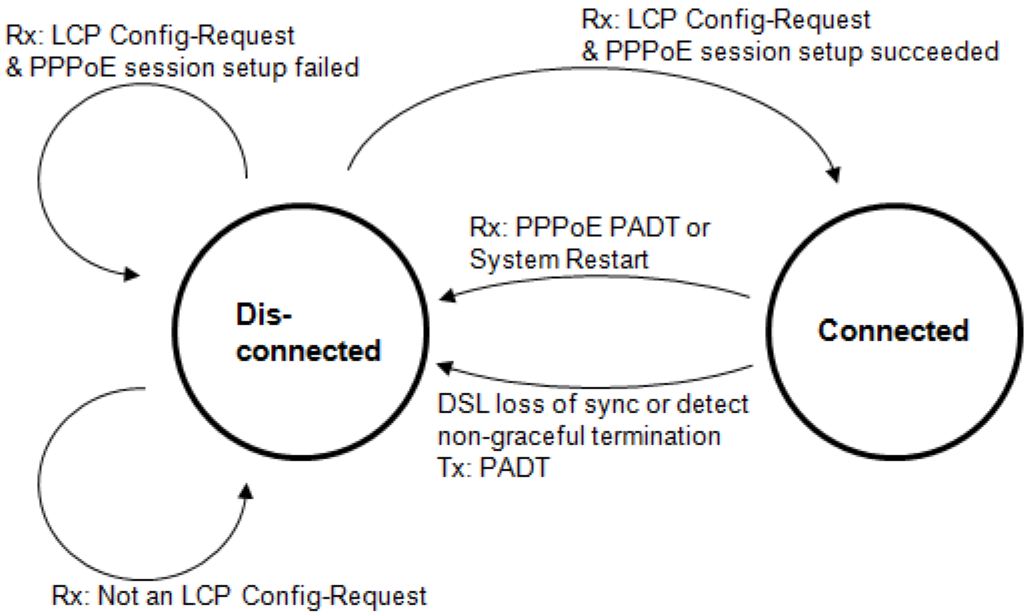


Figure 18 – State transition diagram for PPPoA IWF

R-95 If the Access Node initiates more than one PPPoE session using the same source MAC address and VLAN, the Access Node MUST distinguish between different users' sessions during the PPPoE discovery phase, for example by using either the Host-Uniq tag or the Relay-Session-ID tag. An Ethernet Aggregation Node MUST NOT add the Relay-Session-ID tag.

R-96 The Access Node MUST remove state for an interworked PPPoE session and send a PPPoE PADT message to the BNG upon a loss of connectivity to the customer; this can be indicated by loss of DSL synchronization on the associated customer line.

R-97 The Access Node MUST implement a mechanism to remove state when an interworked PPPoE session terminates. This mechanism MUST handle both signaled (graceful) termination and Access Node recovery following non-graceful teardown (e.g. PADT loss, BNG restart etc.).

One possible method to satisfy detection of non-graceful session teardown is described below.

The Access Node can identify stale interworked sessions by detecting inactivity. This approach assumes that the BNG will stop sending frames for non-existing stale sessions (the same also holds for temporary loss of Access Node-BNG connectivity). Hence, the Access Node should monitor the downstream direction (i.e. BNG to CPE) for inactivity. The time period to decide inactivity, denoted *inactivity timeout*, should be configurable on the Access Node. When no frame is received on the downstream direction of a given interworked session for a period of inactivity timeout, the session is considered disconnected and is cleared from the records. The Access Node should be able to simultaneously perform inactivity monitoring of all active interworked sessions.

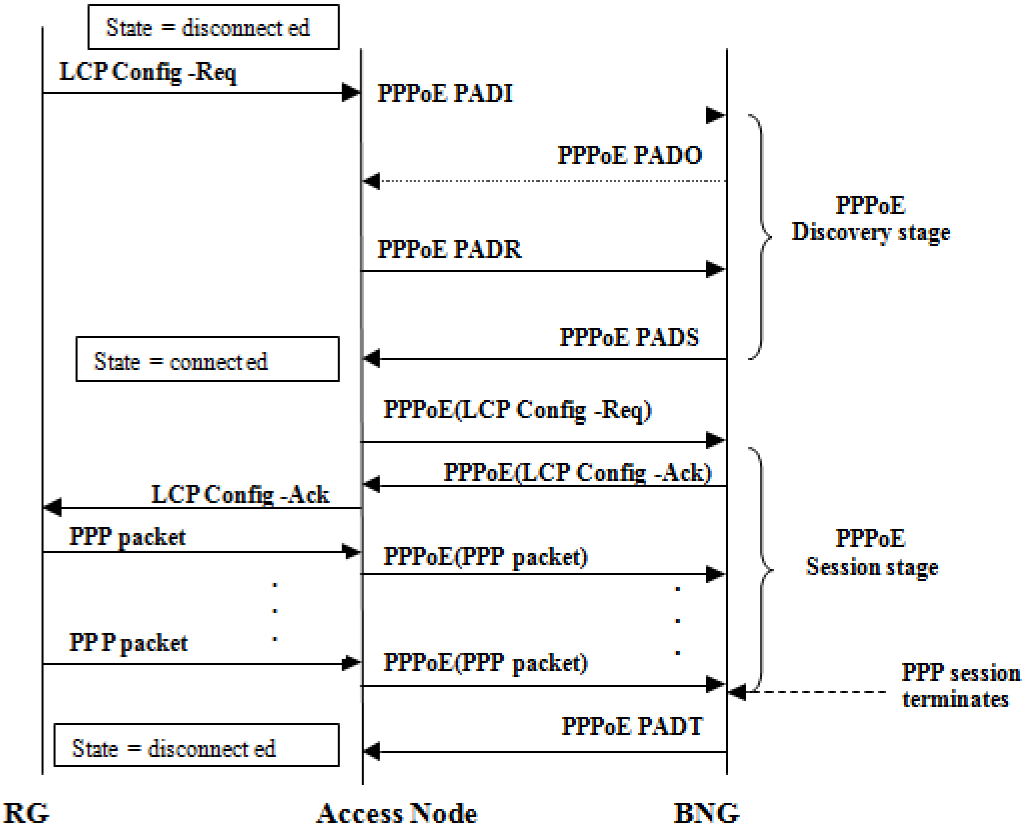


Figure 19 – Example message flow with PPPoA IWF

R-98 The Access Node SHOULD be able to mark PPPoE PADT packets with a higher VLAN priority than that used for best-effort PPPoE session packets.

R-99 The Access Node MUST support a PPP MRU of 1500 bytes to be negotiated for PPPoE sessions. The mechanism to support the MRU negotiation is defined in IETF RFC 4638. The Access Node SHOULD set the value of the PPP-Max-Payload tag to 1500 bytes.

3.5.4.2 BNG Requirements for PPPoA IWF

R-100 The BNG MUST be able to support more that one PPPoE session per source MAC address.

R-101 The BNG MUST allow a PPP MRU of 1500 bytes to be negotiated for PPPoE sessions. The mechanism to support the MRU negotiation is defined in IETF RFC 4638.

R-102 The BNG MUST send a PPPoE PADT message to the Access Node when it detects PPP session termination.

R-103 The BNG MUST ensure downstream packets are sent at a configurable minimum frequency, e.g. 60 seconds, per active PPPoE session to ensure the session is not timed out by the Access Node.

R-104 The BNG SHOULD be able to mark PPPoE PADT packets with a higher VLAN priority than that used for best-effort PPPoE session packets.

3.5.4.3 Signaling Interworked Sessions

It is required that the Access Node signal to the BNG that a given PPPoE session is going to carry interworked PPPoA traffic. This allows the BNG to modify its behavior for interworked PPPoE sessions. This could include configuring a limit on simultaneous PPPoE sessions or the rate of their establishment. The use of a universal flag for this purpose simplifies the decision process that might otherwise need to be implemented to identify a trusted session.

For this purpose the Interworked Session sub-option is defined. This is a sub-option of the Broadband Forum VSA Tag of PPPoE Discovery packets (described in Section 3.9.2).

R-105 The Access Node MUST insert the BBF IWF PPPoE Tag into the PPPoE Discovery packets when it is performing an IWF functionality. The sub-option code field is 0xFE and its length field is set to 0x00.

R-106 The Access Node MUST silently discard any end-user PPPoE discovery packet that contains the BBF IWF PPPoE Tag.

R-107 For each interworked session, the BNG MUST perform all multicast actions (e.g. leave, join) according to the IGMP messages received during the session (i.e. using IGMP/PPPoE encapsulation). The Access Node SHOULD NOT perform any IGMP processing or echo for interworked sessions.

3.6 Multi-session Support

R-108 The Access Node MUST support multiple PPPoE and IPoE sessions per logical/physical port on the U-interface.

3.7 L2 Security Considerations

This section provides security requirements for the Access Node. Many of these requirements are applicable to the N:1 VLAN configuration where user isolation is required, but several also apply in a 1:1 VLAN configuration. It should also be noted that some of the security features applicable for mass-market residential customers may not be applicable to some business customer configurations.

3.7.1 Broadcast Handling

R-109 The Access Node MUST protect the aggregation network and BNGs from broadcast and multicast storms at user and network port levels.

R-110 The Access Node MUST support filtering out broadcast and multicast frames in the downstream direction on a per VLAN basis. When this option is activated, protocol-specific interworking functions MUST be supported to handle downstream broadcasts (DHCP, ARP, IGMP, OAM).

3.7.2 MAC Address Spoofing

A malicious user might try using another user's MAC address (i.e. spoofing) in order to deny or disturb the other user's service or to 'hijack' some frames (when both users are on the same VLAN).

R-111 The Access Node SHOULD be able to provide service to users with duplicate MAC addresses.

Note: Guidance on methods for accomplishing this is a current work item at the BBF.

R-112 The Access Node SHOULD be able to deny service to users with duplicate MAC addresses.

In the case that the Aggregation Network forwards according to MAC learning, an even more harmful hazard might be a user spoofing the Broadband Network Gateway MAC address.

R-113 The Access Node SHOULD provide a mechanism to prevent Broadband Network Gateway MAC address spoofing.

3.7.3 MAC Address Flooding

This issue concerns the Access Node and the aggregation network in the case where it uses source MAC learning. A malicious user can intentionally generate traffic with alternating Ethernet source MAC address, thus exhausting the learning table resources of the Access Node or the aggregation network.

R-114 In order to prevent source MAC address flooding attacks, the Access Node MUST be able to limit the number of source MAC addresses learned from a given bridged port.

R-115 This limit MUST be configurable per user facing port.

R-116 The Access Node MUST support configuration of the MAC aging time.

3.7.4 Filtering

This section describes filtering capabilities that are not already described as requirements in other parts of the document. These capabilities are applicable to both N:1 and 1:1 VLAN configurations.

R-117 The Access Node SHOULD allow configuring the following filters and applying them to ports:

1. Source MAC address filter. This filter may be used in one of the following ways:
   i. *Allowing access from specific devices (i.e. MAC address).*
   ii. *Denying access from a specific MAC address.*
2. Destination MAC address filter. This filter may be used in one of the following ways:
   i. *Allowing access to specific destinations.*
   ii. *Denying access to specific destinations.*

R-118 The Access Node SHOULD provide filtering of reserved group MAC destination addresses (in the 01:80:C2 range)

This filter is applicable to frames received on a user port in the upstream direction.

R-119 When filtering on reserved group MAC addresses, any frame received with a reserved group MAC destination address MUST be handled as identified in the table below.

The default behavior may be overridden for some reserved group MAC destination address (as described in the optional behavior column). Note that the addresses in the range 01-80-C2-00-00-00 to 01-80-C2-00-00-0F are reserved for link-constrained protocols. As such, frames sent to this address are not forwarded by conformant MAC Bridges as its use is restricted to a single link.

MAC address	Application	Default behavior	Optional configurable behavior	Reference
01-80-C2-00-00-00	Bridge Group Address (BPDUs)	Block	None	IEEE 802.1D, Table 7-9
01-80-C2-00-00-01	PAUSE	Block	None	IEEE 802.3x
01-80-C2-00-00-02	Slow Protocols (LACP, EFM OAMPDUs)	Block	Peer	IEEE 802.3, Table 43B-1
01-80-C2-00-00-03	EAP over LANs	Block	Peer	IEEE 802.1X, Table 7-2
01-80-C2-00-00-04 - 01-80-C2-00-00-0F	Reserved	Block	None	IEEE 802.1D, Table 7-9
01-80-C2-00-00-10	All LANs Bridge	Block	None	IEEE 802.1D,

	Management Group Address			Table 7-10
01-80-C2-00-00-20	GMRP	Block	None	IEEE 802.1D, Table 12-1
01-80-C2-00-00-21	GVRP	Block	None	IEEE 802.1Q, Table 11-1
01-80-C2-00-00-22 - 01-80-C2-00-00-2F	Reserved GARP Application addresses	Block	Forward	IEEE 802.1D, Table 12-1
01-80-C2-00-00-30 - 01-80-C2-00-00-3F*	CFM	Forward	Block	IEEE 802.1ag-2007, Tables 8-9 & 8-10.

Table 1 – Default and/or configurable filtering behavior of reserved group MAC destination addresses

* The lowest-order nybble encodes the 8 MD levels for CCs and LTs, respectively as specified in IEEE 802.1ag 2007. In Table 1 the term “Peer” means that the Access Node will locally process the received frame and will send a reply to the sender, as specified by the indicated standard reference.

3.8 Additional IWF for IPoE based Access in N:1 VLANs

3.8.1 Layer 2 DHCP Relay Agent

A typical DHCP Relay Agent sets the *giaddr* of client requests in order to provide an indication to a DHCP server about the subnet on which a client is located and therefore the address that the server should provision.

The AN is well positioned to add the *circuit-id* and *remote-id* sub-options that a typical relay agent adds, but it does not have an IP interface on the subnet or VLAN where the client resides in order to properly set the *giaddr* field as an indication of the client subnet.

RFC 3046 identifies the possibility that a bridge positioned between the full DHCP relay agent and the client may add the relay agent information option, but not set the *giaddr* field. This approach is used in this architecture, where the AN that has the information for constructing *circuit-id* and *remote-id* is a Layer 2 element with no need for a Layer 3 interface on each client subnet or VLAN. Since RFC 3046 gives no name to a network element performing this function, it is called a *Layer 2 DHCP Relay Agent* in this Technical Report.

The use of a Layer 2 DHCP relay agent does not affect the need for a full DHCP relay agent upstream. Such an agent is still needed to set the *giaddr* in client requests and to manage multicast to unicast behavior. It does not need to add *circuit-id* and *remote-id*, but it does need to be able to discern trusted from untrusted interfaces and only honor those fields when they come from trusted interfaces.

3.8.1.1 Basic Operation

A core principal in RFC 3046 is the difference between trusted and untrusted circuits. This guides the operation of both the Layer 2 DHCP relay agent and the full DHCP relay agent. In the absence of a Layer 2 relay agent, the access network is untrusted, and the DHCP relay agent will discard packets arriving on this untrusted interface that already contain option-82. When a Layer 2 relay agent is present, the full DHCP relay agent must be configured to (and, in fact, be able to) trust the access network between them, and expect DHCP packets to arrive containing appropriate option-82 data. In this case, the access network downstream from the DHCP relay agent is trusted, and the access network downstream from the Layer 2 relay agent is untrusted.

It is possible that the Layer 2 relay agent may trust some access interfaces, or perhaps specific VLANs of those interfaces, if they are trunking, because of the presence of another Layer 2 relay agent downstream. For such an interface, option-82 addition, removal, and regulation will reflect the trusted nature of that interface, and the Layer 2 DHCP relay requirements apply to the downstream agent.

3.8.2 DHCP Processing

These requirements describe the DHCP processing on Access Nodes. This is part of the IPoE IWF.

R-120 The Access Node MUST be able to function as a Layer 2 DHCP Relay Agent on selected untrusted ports of a given VLAN.

R-121 The Access Node MUST be able to disable the Layer 2 DHCP Relay Agent on selected user-facing ports of a given VLAN.

R-122 The Access Node, when performing the function of a Layer 2 DHCP Relay Agent, MUST support configuration of downstream trusted interfaces from which it does not discard packets arriving with option-82 already present, and does not add or replace option-82 in these packets.

R-123 The Access Node, when performing the function of a Layer 2 DHCP Relay Agent, MUST NOT remove option-82 from response packets that are destined for downstream trusted interfaces.

R-124 The Access Node, when performing the function of a Layer 2 DHCP Relay Agent, MUST add option-82 with the '*circuit-id' and/or 'remote-id'* sub-options to all DHCP messages sent by the client before forwarding to the Broadband Network Gateway.

Note: see more details about the use of such sub-options in Section 3.9 Access Loop Identification and Characterization.

R-125 The Access Node, when performing the function of a Layer 2 DHCP Relay Agent, MUST remove option-82 information from all DHCP reply messages received from the Broadband Network Gateway before forwarding to untrusted interfaces.

R-126 A server-originated broadcast DHCP packet containing option 82 MUST NOT be bridged to untrusted user-facing ports by an Access Node.

R-127 An Access Node, when performing the function of a Layer 2 DHCP Relay Agent, MUST examine option-82 and/or the *chaddr* field, and only transmit these packets (after removal of option-82) to the untrusted interface for which it is intended.

R-128 The Access Node, when performing the function of a Layer 2 DHCP relay agent, MUST NOT convert the DHCP request from the client from a broadcast to a unicast packet at layer 2 or layer 3.

R-129 The Access Node, when performing the function of a Layer 2 DHCP relay agent, MUST NOT set the *giaddr* on the DHCP request from the client.

R-130 The Access Node, when performing the function of a Layer 2 DHCP relay agent, MUST be configurable per port to snoop all DHCP traffic and filter out those DISCOVER and REQUEST packets from the access loop that have nonzero *giaddr*, and unicast request packets with a zero *ciaddr*.

R-131 The Access Node, when performing the function of a Layer 2 DHCP relay agent, MUST discard any DHCP request packet containing option-82 or *giaddr* and received from an untrusted port.

R-132 The Access Node, when performing the function of a Layer 2 DHCP relay agent, MUST only forward DHCP requests to the upstream designated port(s) to prevent flooding or spoofing.

R-133 The Access Node, when performing the function of a Layer 2 DHCP relay agent, MUST be able to transparently forward any DHCP option information other than for option 82.

3.8.3 ARP Processing and IP Spoofing Prevention

R-134 The Access Node SHOULD inspect upstream and downstream DHCP packets, discover mapping of IP address to MAC address and access ports and populate its ARP table accordingly.

R-135 The Access Node SHOULD ensure that downstream broadcast ARP requests are not sent on access ports that do not have the associated requested IP address.

The above requirement is not made redundant by R-224 since it addresses the case of a second BNG (i.e. multiple BNG aggregation) that is not acting as a DHCP relay and is thus required to send ARP requests.

A malicious user might try using another user IP address (i.e. spoofing) in order to deny/disturb the other user or a network service or to gain unauthorized access to resources. The Access Node can be aware of the mapping between IP address, MAC address and port using the process described in R-134.

R-136 The Access Node SHOULD provide a mechanism to prevent user IP address spoofing.

R-137 The Access Node SHOULD be configurable with a list of IP addresses associated with user port and VLAN for users having static IP configuration.

3.9 Access Loop Identification and Characterization

In order to increase readability, this section includes requirements for both the Access Node and the Broadband Network Gateway.

In previous deployments that used ATM aggregation, access loop identification is facilitated by the typical one-to-one mapping between an access loop and ATM PVC between the Access Node and the Broadband Network Gateway. Based on this property, in a PPP scenario, the Broadband Network Gateway typically includes a NAS-Port-Id (or NAS-Port in some cases) attribute in RADIUS authentication and accounting packets sent to the RADIUS server(s). Such an attribute includes the identification of the ATM VC, which identifies the access loop.

Such access loop identification is useful because:

1. RADIUS authentication & accounting can be performed using such access loop identification. In a retail environment, the access loop identifier can be used to identify the user and determine which service parameters should be returned to the Broadband Network Gateway (via the RADIUS authorization process). In a wholesale environment, the access loop identifier can be used to check the domain-name in the user credentials and verify that it is compatible with the ISP (or list of ISPs), which require the establishment of a service (e.g. Internet access) on this access loop.
2. It can aid trouble shooting. For example, going through RADIUS logs based on an IP address or a user name, and finding out which access loop was used under certain circumstance.
3. The Broadband Network Gateway or policy management functions can have advanced processing based on the access loop identification (e.g. admission control logic, weighted queuing logic for hierarchical QoS, etc).

When the access protocol is IPoE combined with DHCP for IP address assignment, a similar mechanism applies. The Broadband Network Gateway typically implements a DHCP Relay Agent, and inserts the option 82 field defined in RFC 3046 containing the identification of the ATM VC. This is a way to identify the access loop just as before – will all the previously listed benefits.

When an Ethernet-based aggregation network is used, the Broadband Network Gateway can no longer derive the access loop identification from the ATM PVC and an alternative mechanism is needed.

A similar mechanism is available when there is 1:1 VLAN assignment for the access ports. This allows the access loop identification to be directly derived from the VLAN tagging, i.e. S-VID or S-VID plus C-VID, of the frames coming from this port.

R-138 The Broadband Network Gateway MUST support using the VLAN tagging (C-VID, S-VID, and optionally the priority bits) of a frame as access loop identification in RADIUS (e.g. NAS-Port-Id) and DHCP (option 82) messages.

R-139 The BNG MUST, when performing the function of a DHCP Relay Agent or proxy, add option 82 with the 'circuit-id' and/or 'remote-id' sub-options to all DHCP messages handled by the agent or proxy sent by the client before forwarding to the DHCP server.

Although feasible, the above-described solution might not be suitable for deployment scenarios where N:1 VLANs are used. The following sub-sections describe a more appropriate solution for cases where access ports are provisioned in an N:1 VLAN.

In addition to the port identification discussed above, the access loop sync rate and interleave delay is information of interest to the Broadband Network Gateway to enable policy decisions and advanced QoS enforcement The following sub-sections also describe a mechanism to report such parameters for both PPP and DHCP access sessions.

3.9.1 DHCP Relay Agent

The DHCP Relay Agent supports the IPoE access method and is based on locating the function in the Access Node. This function supports the same capabilities described in Section 3.8.1. The DHCP packet format is specified in RFC 2131. The DHCP Relay Agent Information option (option 82) format is specified in RFC 3046.

Two sub-options of option 82 are defined in RFC 3046:

- Agent Circuit ID (intended for circuits terminated by the system hosting the Relay agent)
- Agent Remote ID (intended to identify the remote host end of a circuit)

The exact syntax of the corresponding fields is left open by RFC 3046 to relay agent implementers.

R-140 The Access Node DHCP Relay Agent MUST be able to encode the access loop identification in the "Agent Circuit ID" sub-option (sub-option 1). The encoding MUST uniquely identify the Access Node and the access loop logical port on the Access Node on which the DHCP message was received. The Agent Circuit ID contains a locally administered ASCII string generated by the Access Node, representing the corresponding access loop logical port (U-interface). The actual syntax of the access loop identification in the Agent Circuit ID is mandated by this document in Section 3.9.3.

R-141 The Access Node DHCP Relay Agent MUST have the option to use the "Agent Remote ID" sub-option (sub-option 2) to further refine the access loop logical port identification. The Agent Remote ID contains a configurable string of 63 characters maximum that uniquely identifies the user on the associated access loop on the Access Node on which the DHCP discovery message was received. The actual syntax of the user identification in the Agent Remote ID is not specified in this document.

RFC 4243 defines a sub-option of the DHCP Relay Agent Information Option to carry vendor-identifying vendor-specific information. The sub-option allows a DHCP relay agent to include vendor-specific data in the DHCP messages it forwards as configured by the administrator. The use of this sub-option is detailed in Appendix B - DHCP Vendor Specific Options to Support Access Line Characteristics.

R-142 The Access Node DHCP Relay Agent MUST support inserting vendor specific information per RFC 4243.

3.9.2 PPPoE Intermediate Agent

The PPPoE Intermediate Agent supports the PPPoE access method and is a function placed on the Access Node in order to insert access loop identification. .

The PPPoE Intermediate Agent intercepts all upstream PPPoE discovery stage packets, i.e. the PADI, PADR and upstream PADT packets, but does not modify the source or destination MAC address of these PPPoE discovery packets. Upon receipt of a PADI or PADR packet sent by the PPPoE client, the Intermediate Agent adds a PPPoE TAG to the packet to be sent upstream. The TAG contains the identification of the access loop on which the PADI or PADR packet was received. If a PADI or PADR packet exceeds the Ethernet MTU after adding the access loop identification TAG, the Intermediate Agent must drop the packet, and issue the corresponding PADO or PADS response with a Generic-Error TAG to the sender.

R-143 The Access Node MUST implement a PPPoE intermediate agent as described above.

The Broadband Network Gateway implements the counterpart of this function as follows. The Broadband Network Gateway accepts PADI and PADR packets containing a TAG that is used to convey the access loop identification to the Broadband Network Gateway. The access loop information present in a TAG in the PADI and PADR packets may be used by the Broadband Network Gateway to check whether PPPoE discovery is allowed for the identified user line. This behavior is independent of the PPP authentication phase performed later-on. The Broadband Network Gateway then uses the access loop identification as described in Section 3.9.6.

R-144 The Broadband Network Gateway MUST NOT send the TAG used to convey the access loop identification in PADO, PADS and downstream PADT messages.

R-145 The Broadband Network Gateway MUST be able to support access loop identification carried over PPPoE as described above.

The required syntax for access loop identification is depicted in Figure 20. Note that RFC 2516 states that “To ensure inter-operability, an implementation MAY silently ignore a Vendor-Specific TAG.”

```
+--------------+--------------+--------------+--------------+
| 0x0105 (Vendor-Specific)    |         TAG_LENGTH          |
+--------------+--------------+--------------+--------------+
| 0x00000DE9 (3561 decimal, i.e. "BBF" IANA entry)          |
+--------------+--------------+--------------+--------------+
| 0x01         | length       | Agent Circuit ID value...   |
+--------------+--------------+--------------+--------------+
| Agent Circuit ID value (con't) …                          |
+--------------+--------------+--------------+--------------+
| 0x02         | length       | Agent Remote ID value...    |
+--------------+--------------+--------------+--------------+
| Agent Remote ID value (con't) …                           |
+--------------+--------------+--------------+--------------+
```

Figure 20 – PPPoE access loop identification tag syntax

The first four octets of the TAG_VALUE contain the vendor id. The low-order 3 octets are the SMI Network Management Private Enterprise Code for the Broadband Forum, 0x000DE9 (3561 decimal, the IANA "Broadband Forum" entry in the Private Enterprise Numbers registry). The high order octet of the vendor id is set to 0x00.The remainder of the TAG_VALUE is unspecified in RFC 2516. Note that the sub-options do not have to be aligned on a 32-bit boundary. For encoding the access loop identification, the same sub-option based encoding as used in DHCP option 82 is used, i.e. sub-options in a Type-Length-Value format (as defined by RFC 3046).

R-146 Both Access Node and Broadband Network Gateway MUST support the PPPoE access loop identification tag as specified above.

R-147 The Access Node MUST encode the access loop identification in the "Agent Circuit ID" sub-option (sub-option 1). The encoding MUST uniquely identify the Access Node and the access loop logical port on the Access Node on which the discovery stage PPPoE packet was received. The Agent Circuit ID contains a locally administered ASCII string generated by the Access Node, representing the corresponding access loop logical port (U-interface). The actual syntax of the access loop identification in the Agent Circuit ID is mandated by this document in Section 3.9.3.

R-148 The Access Node MUST have the option to encode the user identification in the "Agent Remote ID" sub-option (sub-option 2). The Agent Remote ID contains a configurable string of 63 characters maximum that uniquely identifies the user on the associated access loop logical port on the Access Node on which the PPPoE discovery packet was received. The actual syntax of the user identification in the Agent Remote ID is not specified in this document.

R-149 The Access Node MUST replace the Broadband Forum PPPoE vendor-specific tag with its own if the tag has also been provided by a PPPoE client.

3.9.3 Access Loop Identification Configuration and Syntax

In order to integrate with a diversity of existing OSS environments, the operator must be able to configure the Agent Circuit ID in a flexible way. A default syntax and automatic generation of the circuit IDs is defined to encourage consistency and automation in operator's practices, and is recommended. The default syntax may be overridden (per Access Node) by a more flexible circuit ID syntax to ease integration with legacy OSS environments, while keeping the principle of automatic generation of circuit-ids for all access loops. Finally, the circuit ID may be configured by the operator per individual access loop logical port, overriding the circuit ID being automatically generated.

The default syntax used for the Agent Circuit ID field by Access Nodes mimics a typical practice often used by BNG DHCP Relay Agents (using the Agent Circuit ID sub-option in DHCP option 82) and BNG RADIUS clients (using the NAS-Port-Id Attribute). An example of such typical practice is:

“Relay-identifier atm 3/0:100.33” (slot = 3, port = 0, vpi = 100, vci = 33)

R-150 The Agent Circuit ID field inserted by the Access Node DHCP Relay Agent and PPPoE Intermediate Agent MUST NOT exceed 63 characters.

R-151 The value of the Agent Circuit ID MUST be explicitly configurable, per individual access loop and logical port. When not explicitly configured, it MUST be automatically generated using the default or flexible syntax described in following requirements.

R-152 The Access Node DHCP Relay Agent and PPPoE Intermediate Agent MUST use the following default syntax to automatically generate the Agent Circuit ID field, identifying access loop logical ports as follows:

“Access-Node-Identifier atm slot/port:vpi.vci” (when ATM/DSL is used)

“Access-Node-Identifier eth slot/port[:vlan-id]” (when Ethernet[/DSL] is used)

In this syntax, Access-Node-Identifier MUST be a unique ASCII string (not using character spaces). The Access-node-identifier, L2 type (ATM, ETH) field and the slot/port fields are separated using a single space character. The slot identifier MUST NOT exceed 6 characters in length and the port identifier MUST NOT exceed 3 characters in length and MUST use a ‘/’ as a delimiter[3]. The vpi, vci and vlan-id fields (when applicable) are related to a given access loop (U-interface) [4].

R-153 The value of Access-Node-Identifier MUST be configurable per Access Node, using an element management interface. The Access-Node-Identifier MAY be derived automatically from an already defined object ID (e.g. IP address of management interface).

R-154 It MUST be possible to override the default syntax of circuit IDs, and support configuration of a more flexible syntax for the Agent Circuit ID, with flexibility in the choice of elements used in the automated generation of circuit-IDs. Such syntax is unique per Access Node.

R-155 The flexible syntax MUST allow the concatenation of 2 types of elements:

- Configured strings of ASCII characters. This will typically include characters used as separators between variable fields (usually # . , ; / or space)
- Variable fields whose content is automatically generated by the Access Node. The minimum list of those variable fields is given in the following table. Fields should include information which does not vary over time for a given access loop.

Description of the variable	Possible name for the variable	Type of variable and max length	Range of values for the variable
Logical name of the Access Node.	Access_Node_ID	Variable. Note that total length of the	

[3] The exact way to identify slots is implementation-dependent. In some cases, the slot field may convey some additional semantics (e.g. the “705” value could mean rack #7 and slot #5). Concepts like chassis (for a multi-chassis system), racks or shelves may also be captured in the same way (e.g. “9-9-99” for a rack-shelf-slot construct) by further structuring the slot field.

[4] In other words, in the ATM case, vpi/vci will always be used. In the EFM case, *if the DHCP or PPPoE message is received with a VLAN tag, the received VLAN ID will be appended to the string*.

		overall agent-circuit-id must not to exceed 63 bytes	
Chassis number in the access node	Chassis	Char(2)	“0”..”99”
ONU number (Port)	ONUID	Char(3)	“0”..”999”
Rack number in the access node	Rack	Char(2)	“0”..”99”
Frame number in the rack	Frame	Char(2)	“0”..”99”
Slot number in the chassis or rack or frame	Slot	Char(2)	“0”..”99”
Sub-slot number	Sub-slot	Char(2)	“0”..”99”
Port number in the slot	Port	Char(3)	“0”..”999”
VPI on U interface in case of ATM over DSL	VPI	Char(4)	“0”..”4095”
VCI on U interface in case of ATM over DSL	VCI	Char(5)	“0”..”65535”
VLAN ID on U interface (when applicable)	Q-VID	Char(4)	”0”..”4095”
S-VLAN ID on V interface	S-VID	Char(4)	“0”,”4095”
C-VLAN ID on V interface	C-VID	Char(4)	“0”,”4095”
Ethernet Priority bits on V interface	Ethernet Priority	Char(1)	“0”..”7”

Table 2 – Circuit ID Syntax

3.9.4 Access Loop Characteristics

This solution is designed as an extension of the role of a Layer2 DHCP Relay Agent or a PPPoE Intermediate Agent in an Access Node, inserting the appropriate access loop characteristics (e.g. sync rate and interleaving delay) values while forwarding DHCP or PPPoE messages[5].

For convenience and readability, this section includes requirements for Access Nodes and Broadband Network Gateways.

The rates being discussed are “raw” layer2 bearer data rates.

R-156 The Access Node MUST be able to insert the access loop characteristics via its PPPoE intermediate agent and/or via its layer2 DHCP Relay agent. It MUST be possible to enable/disable this function per port, depending on the type of user.

R-157 The Broadband Network Gateway MUST be able to receive access loop characteristics information, and share such information with AAA/policy servers as defined in Section 3.9.6.

R-158 When sending the access line characteristics in DHCP / PPPoE messages, the Access Node MUST be configurable to either send only the actual bitrate information, or include the full set of access line characteristics.

R-159 In all cases (PPPoE intermediate agent, DHCP-Relay), the access loop characteristics information MUST be conveyed with a loop characteristics field structured with type-length-value sub-fields as described in RFC 4243 and again in Appendix A - PPPoE Vendor-Specific BBF Tags and

[5] Similar function can be performed by a Routing Gateway in pure ATM aggregation scenarios, e.g. based on the TR-59 architecture where the Access Node is transparent to DHCP and PPPoE. However, this is out of the scope of this specification.

Appendix B - DHCP Vendor Specific Options to Support Access Line Characteristics. Sync data rate values MUST be encoded as 32-bit binary values, describing the rate in Kbps. Interleaving delays MUST be encoded as 32-bit binary values, describing the delay in milliseconds. The complete set of sub-options is listed in the following table:

subopt.	Message Type	Information	Reference
0x81	Actual data rate Upstream	Actual data rate of an access loop	ITU-T G.997 Section 7.5.2.1
0x82	Actual data rate Downstream	Actual data rate of an access loop	ITU-T G.997 Section 7.5.2.1
0x83	Minimum Data Rate Upstream	Minimum data rate at which the loop is set to operate	ITU-T G.997 Section 7.3.1.1.1
0x84	Minimum Data Rate Downstream	Minimum data rate at which the loop is set to operate	ITU-T G.997 Section 7.3.1.1.1
0x85	Attainable Data Rate Upstream	Maximum data rate that can be achieved.	ITU-T G.997 Section 7.5.1.12 and 7.5.1.13
0x86	Attainable Data Rate Downstream	Maximum data rate that can be achieved.	ITU-T G.997 Section 7.5.1.12 and 7.5.1.13
0x87	Maximum Data Rate Upstream	Maximum data rate at which the loop is set to operate	ITU-T G.997 Section 7.3.2.1.3
0x88	Maximum Data Rate Downstream	Maximum data rate at which the loop is set to operate	ITU-T G.997 Section 7.3.2.1.3
0x89	Minimum Data Rate Upstream in low power state	Minimum data rate at which the loop is set to operate during the low power state (L1/L2).	ITU-T G.997 Section 7.3.2.1.5
0x8A	Minimum Data Rate Downstream in low power state	Minimum data rate at which the loop is set to operate during the low power state (L1/L2).	ITU-T G.997 Section 7.3.2.1.5
0x8B	Maximum [Interleaving] Delay Upstream	Maximum one-way interleaving delay	ITU-T G.997 Section 7.3.2.2
0x8C	Actual [interleaving] Delay Upstream	Value in milliseconds which corresponds to the interleaver setting.	ITU-T G.997 section 7.5.2.3
0x8D	Maximum [Interleaving] Delay Downstream	Maximum one-way interleaving delay	ITU-T G.997 Section 7.3.2.2
0x8E	Actual [interleaving] Delay Downstream	Value in milliseconds which corresponds to the interleaver setting.	ITU-T G.997 section 7.5.2.3

Table 3 – Access loop characteristics sub-options

Note: Although the references point to parameters used for ADSL2/2plus, equivalent use should also be provided for VDSL2 and SHDSL-based DSL systems and some Ethernet systems. When DSL bonding is used, the reported metrics reflect the aggregate properties of the bonded ports.

The following requirements describe how the loop characteristics are conveyed in the various cases.

R-160 Access Nodes MUST support all sub-options applicable to their access technology.

R-161 Access Nodes MUST be configurable to send only sub-options 0x81 and 0x82, or to send the entire list.

R-162 In the DHCP Relay case, the access loop characteristics information MUST be conveyed by the DHCP option-82 field, with a vendor-specific sub-option, encoded according to RFC 4243, with the enterprise number being the Broadband Forum enterprise code, i.e. 3561 in decimal (0x0DE9 in hexadecimal), corresponding to the IANA “Broadband Forum” entry in the Private Enterprise Numbers registry. Sub-options codes are described in Table 3.

R-163 In the PPPoE case, the access loop characteristics information MUST be conveyed by an extension of the Broadband-Forum vendor-specific PPPoE tag defined in Section 3.9.2, using additional sub-options with codes as described in Table 3. See Appendix A - PPPoE Vendor-Specific BBF Tags for more detailed sub-option encoding.

3.9.5 Signaling the Access Loop Encapsulation

In a TR-59 based architecture, the BRAS is responsible for shaping downstream traffic to the DSL line rate, or the service rate (which may be less than the line rate). If this is not done, then significant and indiscriminate packet loss can result. In the Ethernet aggregation scenario described by this recommendation, the BNG will be shaping at the IP level, but the Layer 2 encapsulation added at the Access Node can increase packet overhead to the point where the physical line rate is exceeded. If the BNG knows the nature of this encapsulation, then the IP shaping rate can be adjusted accordingly. This section describes sub-options for use with the PPPoE VSA Tag / DHCP option-82 to signal the access loop encapsulation from the Access Node to the BNG.

R-164 The PPPoE intermediate agent and DHCP relay agent on the Access Node SHOULD insert the following sub-option (in the same manner described in R-160 and R-163) to signal to the BNG the data-link protocol and the encapsulation overhead on the Access Loop.

```
+--------+--------+--------+--------+--------+--------+
| Sub-option type | Length |Data Lnk|Encaps 1|Encaps 2|
|      0x90       |  0x03  | 1 byte | 1 byte | 1 byte |
+--------+--------+--------+--------+--------+--------+
```

Data link ∈ {

ATM AAL5 = 0,

Ethernet = 1}

Encaps 1 ∈ {

NA = 0,

Untagged Ethernet = 1

Single-tagged Ethernet = 2

Double-tagged Ethernet = 3}

Note: The Double-tagged encap value should be used whenever it is possible to send 2 tags on the U-Interface.

Encaps 2 ∈ {

NA = 0,

~~PPPoA LLC = 1,~~

~~PPPoA Null = 2,~~

~~IPoA LLC = 3,~~

~~IPoA Null = 4,~~

Ethernet over AAL5 LLC w FCS = 5,

Ethernet over AAL5 LLC w/o FCS = 6,

Ethernet over AAL5 Null w FCS= 7,

Ethernet over AAL5 Null w/o FCS=8}

3.9.6 BNG to RADIUS Signaling of Broadband Line Characteristics

This section is provided to define the BNG requirements for sharing the Broadband line information, received from the Access Node, with AAA/Policy servers if the RADIUS protocol is used for such communication.

The intention is for the BNG to use the Broadband Forum Private Enterprise Number as the Vendor-ID in all RADIUS Accounting-Request VSA messages that share Access Node Broadband Line Characteristics with a RADIUS server. This will remove the necessity of all vendors defining individual VSA codes for the same information.

R-165 The BNG MUST be able to configure the Broadband Forum Private Enterprise Number, hexadecimal 0x000DE9, (3561 decimal, the IANA “Broadband Forum” entry in the Private Enterprise Numbers registry) as the Vendor-ID in RADIUS Vendor Specific Attributes.

```
0                   1                   2                   3
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    VSA (26)   |  Length       |        Vendor-Id (3561)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    Vendor-Id (cont)           |  TLV...
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
```

R-166 The BNG MUST be able to send to a AAA/policy server the TLVs defined in Table 4, in a RADIUS Vendor Specific Attribute, using the encoding as specified in R-165.

R-167 The BNG MUST be able to send to a AAA/policy server the TLVs defined in Table 4, in a RADIUS Vendor Specific Attributes, using RADIUS Access-Request and Accounting-Request messages as defined in Table 5.

Type	Length	Value	Value type
0x01	Length	Agent Circuit ID	String max. 63 chars
0x02	Length	Agent Remote ID	String max. 63 bytes
0x81	4	Actual data rate Upstream in kbps.	32 bit binary value
0x82	4	Actual data rate Downstream in kbps.	32 bit binary value
0x83	4	Minimum Data Rate Upstream in kbps.	32 bit binary value
0x84	4	Minimum Data Rate Downstream in kbps.	32 bit binary value
0x85	4	Attainable Data Rate Upstream in kbps.	32 bit binary value
0x86	4	Attainable Data Rate Downstream in kbps.	32 bit binary value
0x87	4	Maximum Data Rate Upstream in kbps.	32 bit binary value
0x88	4	Maximum Data Rate Downstream in kbps.	32 bit binary value
0x89	4	Minimum Data Rate Upstream in low power state in kbps.	32 bit binary value
0x8A	4	Minimum Data Rate Downstream in low power state in kbps.	32 bit binary value
0x8B	4	Maximum Interleaving Delay Upstream in millisec.	32 bit binary value
0x8C	4	Actual interleaving Delay Upstream in millisec.	32 bit binary value
0x8D	4	Maximum Interleaving Delay Downstream in millisec.	32 bit binary value
0x8E	4	Actual interleaving Delay Downstream in millisec.	32 bit binary value
0x90	3	Access loop encapsulation	24 bit binary values as explained in Appendix A - PPPoE Vendor-Specific BBF Tags
0xFE	0	PPPoA/oE IWF session flag	Empty (indication)

Table 4 – TLVs for New Broadband Forum RADIUS VSAs.

VSA Type	Sent in Access-Request	Received in Access-Accept	Sent in Accounting-Req
0x01, agent circuit id	Yes	No	Yes
0x02, agent remote id	Yes	No	Yes
0x81, actual upstream	Yes	No	Yes
0x82, actual downstream	Yes	No	Yes
0x83, min. upstream	Yes	No	Yes
0x84, min. downstream	Yes	No	Yes
0x85, attainable upstream	No	No	Yes
0x86, attainable downstream	No	No	Yes
0x87, max. upstream	No	No	Yes
0x88, max. downstream	No	No	Yes
0x89, min. upstream (LP)	No	No	Yes
0x8A, min. downstream (LP)	No	No	Yes
0x8B, max. interleave up	No	No	Yes
0x8C, actual interleave up	No	No	Yes
0x8D, max. interleave down	No	No	Yes
0x8E, actual interleave down	No	No	Yes
0x90, access loop encapsulation	No	No	Yes
0xFE, IWF session flag	Yes	No	Yes

Table 5 – Mapping new Broadband Forum RADIUS VSAs to RADIUS message types

3.10 OAM

For Access Node OAM requirements refer to Section 7

4. Ethernet Aggregation Node Requirements

4.1 VLAN Support

R-168 The Aggregation Node MUST be an S-VLAN bridge as per 802.1ad (i.e. no C-Tag awareness).

R-169 The Aggregation Node MUST support user isolation. This behavior MUST be configurable on a per S-VLAN basis.

4.2 QoS

R-170 The Aggregation Node MUST support at least 4 traffic classes for Ethernet frames, and MUST support configurable mapping to these classes from the 8 possible values of the Ethernet priority field.

R-171 The Aggregation Node SHOULD support 8 traffic classes for Ethernet frames, and MUST support configurable mapping to these classes from the 8 possible values of the Ethernet priority field.

R-172 The Aggregation Node SHOULD support drop precedence within at least 2 traffic classes and MUST support configurable mapping to both the classes as well as drop precedence from the 8 possible values of the Ethernet priority field.

R-173 The Aggregation Node MUST support direct indication of drop precedence within all supported traffic classes based on the DEI bit value of the Ethernet header.

R-174 The Aggregation node(s) SHOULD support scheduling of the interface queues according to their assigned priority and weight. The number of priorities MUST be at least 2; however multiple queues MUST be able to be assigned to the same priority. Queues assigned to the same priority MUST be scheduled according to a weighted algorithm (like WFQ) with weights assigned through provisioning. This mechanism provides support for mapping diffserv PHBs (e.g. EF, AF, BE, LE) to the Ethernet queues.

R-175 The Aggregation Node MUST support at least 4 queues per interface, one per traffic class.

R-176 The Aggregation Node SHOULD support at least 8 queues per interface, one per traffic class.

R-177 The Aggregation Node MUST support scheduling of queues according to strict priority with the number of priority levels being at least 2.

R-178 The Aggregation Node MUST support setting the maximum size/depth of all queues.

4.3 Multicast

See Section 6.

4.4 Forwarding Information and Loop Detection (Spanning Tree)

R-179 The Aggregation Node MUST support a minimum of two instances of Multiple Spanning Tree, as per IEEE 802.1Q (2005 edition).

R-180 The Aggregation Node MUST support interworking with the Common Spanning Tree often found in Access Nodes according to IEEE 802.1Q (2003 edition).

R-181 The Aggregation Node MUST support Rapid Spanning Tree as per IEEE 802.1D (2004 edition).

R-182 The Aggregation Node MUST support Link Aggregation as specified by IEEE 802.3ad to allow link resilience.

R-183 The Aggregation Node MUST support load balancing over IEEE 802.3ad Aggregated Links.

R-184 The Aggregation Node SHOULD support a method to detect unidirectional L2 links between devices.

R-185 The Aggregation Node MUST be able to prioritize BPDUs in the data plane (e.g. by providing dedicated queues) and in the control plane (e.g. by providing dedicated CPU queues for BPDUs).

R-186 The Aggregation Node MUST be able to drop BPDUs if those BPDUs have a root bridge identifier which is lower (better) than the current Spanning Tree root. This function MUST be configurable on a per port basis.

R-187 The Aggregation Node MUST be able to drop BPDUs regardless of the BPDU content. This function MUST be configurable on a per port basis.

R-188 The Aggregation Node MUST support the disabling of MAC learning on a per VLAN basis.

R-189 The Aggregation Node MUST be able to use IGMP Snooping to install bridge table entries for multicast entries while MAC learning is disabled in the Multicast VLAN.

For further Aggregation Node requirements concerning multicast delivery refer to Section 6.

4.5 OAM

For Aggregation Node OAM requirements refer to Section 7.

5. Broadband Network Gateway Requirements

This section contains many of the Broadband Network Gateway (BNG) requirements. For the sake of readability, additional BNG requirements are located in Sections 3.5.4 (PPPoA IWF), 3.9 (Broadband line identification and characterization), 6 (multicast), 7 (OAM) and 8 (network management).

5.1 VLAN Support

R-190 The Broadband Network Gateway MUST be able to attach a single S-Tag to untagged frames in the downstream direction.

R-191 The Broadband Network Gateway MUST be able to double-tag frames (S-C-VID pair) in the downstream direction.

R-192 The Broadband Network Gateway MUST be capable of associating one or more VLAN identifications with a physical Ethernet aggregation port. These may be S-VIDs or S-C-VID pairs.

R-193 The Broadband Network Gateway MUST support a one-to-one mapping between an S-VID or S-C-VID pair and a user PPPoE or IPoE session.

R-194 The Broadband Network Gateway MUST support a one-to-many mapping between an S-VID or S-C-VID pair and a user PPPoE or IPoE sessions, where multiple PPPoE and/or IPoE sessions from the same user are within the same S-VID or S-C-VID pair.

R-195 The Broadband Network Gateway MUST support a one-to-many mapping between a S-VID or S-C-VID pair and users sessions, where multiple PPPoE and/or IPoE sessions from multiple users are within the same S-VID or S-C-VID pair.

5.2 QoS – Hierarchical Scheduling

To support QoS enabled IP services, TR-059 introduced the concept of Hierarchical Scheduling (HS). This capability provides IP QoS support across a QoS unaware ATM aggregation network. As described in TR-059, Hierarchical Scheduling provides 5 levels of hierarchy corresponding to Session, VC, Group of VCs, VP and port. Similarly, HS can be applied in the case of an Ethernet aggregation network which otherwise might provide insufficient traffic management capabilities.

For hierarchical scheduling to be effective, any bandwidth bottlenecks must be determined in the access network topology and element architectures and should be canonically laid out. The most efficient hierarchical scheduling occurs when each “aggregation-and-speedup” (or slowdown) bottleneck in the access topology can be represented by a node on the scheduler. This is the optimum solution. Less efficient hierarchical scheduling can still be accomplished when one or more “aggregation-and-speedup” steps are scheduled together at the lowest rate of the group. This is less optimum and should be avoided.

Given these constraints, any given topology might be supported up to the limits of the number of nodes in the scheduler. The minimum number of network queuing points to be considered is three and four is required under some typical deployment scenarios.

The minimum number of supported levels normally corresponds to the BNG port, access node uplink, and access loop synch rate. If there are multiple logical ports or sessions to the user, then a 4th level may be required to represent the logical port or session level. It should also be noted that hierarchical scheduling may not be required for multicast router / BNG elements that comprise part of a multi-edge architecture.

R-196 The Broadband Network Gateway MUST be able to perform at least 3-level HS towards the Ethernet aggregation network.

R-197 The Broadband Network Gateway SHOULD be able to perform 4-level HS towards the Ethernet aggregation network.

Some levels have a natural relationship to a well-known entity. For example, it is expected that the root of the HS would be based on the physical port of the BNG. However, other levels may be identified in a number of ways, depending on the deployment topology used, so requirements for these levels are flexibly defined as any grouping of logical or physical ports/connections that make sense. Specifically:

R-198 The Broadband Network Gateway MUST be able to identify the root level by a single physical port.

R-199 The Broadband Network Gateway SHOULD be able to identify the root level by a group of physical ports.

R-200 The Broadband Network Gateway MUST be able to identify the second level (and potentially the third) by either 1 or 2 below.

1. A group of one or more S-VIDs.
2. A group of one or more C-VID ranges that share a S-VID (this is for the case that S-VIDs are shared between access nodes).

R-201 The Broadband Network Gateway SHOULD identify the second level (and potentially the third) by a combination of 1 and 2 above.

R-202 The Broadband Network Gateway MUST identify the access loop by: a single C-VID, S-VID or S-C-VID pair, or by the User Line Identification (described in Section 3.9).

R-203 The Broadband Network Gateway MUST identify the logical port or session by a C-VID, S-VID or S-C-VID pair, by the User Line Identification, by IP address, or by PPPoE session.

R-204 The Broadband Network Gateway MUST be able to map between IP traffic classes and the Ethernet priority field.

R-205 The Broadband Network Gateway MUST support marking Ethernet drop precedence within at least 2 traffic classes and MUST support configurable mapping from both the classes as well as drop precedence to the 8 possible values of the Ethernet priority field.

Note: Using P bits to indicate drop precedence should be avoided in new deployments.

R-206 The Broadband Network Gateway MUST support marking Ethernet direct indication of drop precedence within all supported traffic classes based on setting the DEI bit value of the S-Tag header.

R-207 The Broadband Network Gateway, when receiving information about Broadband line rate parameters through PPP or DHCP, MUST NOT apply the information in an additive fashion when multiple sessions are active on the same Broadband line (the underlying rate is shared by all the sessions on a given line although each session will report the rate independently).

5.2.1 Policing

R-208 The Broadband Network Gateway MUST support the application of ingress policing on a per user basis.

R-209 The Broadband Network Gateway MUST support the application of Ingress policing on a per C-VID, S-VID or S-C-VID pair basis.

R-210 The Broadband Network Gateway SHOULD support the application of ingress policing of a group of sessions or flows for a given user.

5.3 Multicast

See Section 6.

5.4 ARP Processing

R-211 For a given IP interface (say in subnet Z), the Broadband Network Gateway MUST be able to work in 'Local Proxy ARP' mode: routing IP packets received from host X on this interface to host Y (X and Y are in subnet Z) back via the same interface. Any ICMP redirect messages that are usually sent on such occasions MUST be suppressed.

R-212 The Broadband Network Gateway MUST respond to ARP requests received on this interface for IP addresses in subnet Z with its own MAC address. This requirement refers to both N:1 VLANs as well as to several 1:1 VLANs sharing the same IP interface on the BNG.

The above requirements allow two users located on the same IP subnet and on the same or different Access Nodes (and VLANs) to establish IP communication between them through the BNG.

5.5 DHCP Relay

These requirements describe the DHCP processing on Broadband Network Gateways.

R-213 The Broadband Network Gateway MUST be able to function as a DHCP Relay Agent as described in RFC 951 "BOOTP", RFC 2131"DHCP" and RFC 3046 "DHCP Relay Agent Information Option" on selected untrusted interfaces.

R-214 The Broadband Network Gateway MUST be able to disable the DHCP Relay Agent on selected interfaces.

R-215 The Broadband Network Gateway MUST be able to function as a DHCP relay agent on selected trusted interfaces, from which it does not discard packets arriving with option-82 already present, and does not add or replace option-82 in these packets.

R-216 The Broadband Network Gateway MUST be able to function as a DHCP relay agent on selected trusted interfaces and MUST NOT strip out option-82 from the corresponding server-originated packets it relays downstream.

R-217 The Broadband Network Gateway, when functioning as a DHCP Relay Agent, MUST discard any DHCP packets with non-zero 'giaddr' in the DHCP request from the client.

R-218 The Broadband Network Gateway, when functioning as a DHCP Relay Agent, MUST send the DHCP packets downstream as Layer 2 unicast or Layer 2 broadcast, according to the broadcast bit in the request.

R-219 The Broadband Network Gateway, when functioning as a DHCP relay agent, MUST be able to transparently forward any DHCP option information other than for option 82.

5.6 OAM

For BNG OAM requirements refer to Section 7.

5.7 Security Functions

5.7.1 Source IP Spoofing

R-220 The Broadband Network Gateway MUST only respond to user ARP requests when they originate with the proper IP source address and are received on the appropriate 802.1q VLAN, or 802.1ad stacked VLAN.

A malicious user might try to spoof an IP address by sending ARP messages (both ARP requests and replies) indicating the binding of its MAC address to the spoofed IP address.

R-221 The Broadband Network Gateway MUST be able to detect and discard ARP requests and reply messages with ‘sender protocol address’ other than the one assigned (i.e. spoofed). Specifically, the Broadband Network Gateway MUST NOT update its ARP table entries based on received ARP requests.

R-222 The DHCP relay agent in the Broadband Network Gateway MUST inspect downstream DHCP ACK packets, discover mapping of IP address to MAC address and populate its ARP table accordingly.

R-223 The DHCP relay agent in the Broadband Network Gateway SHOULD follow the lease time and lease renewal negotiation, and be able to terminate any user sessions and remove the corresponding ARP table entry when the lease time has expired.

R-224 Having the knowledge of MAC to IP mapping (achievable by following R-222 and R-223), the Broadband Network Gateway MUST NOT send broadcast ARP requests to untrusted devices (i.e. RGs).

6. Multicast

6.1 Methodology

A set of baseline requirements is outlined based on existing Ethernet multicast deployments (e.g. no restrictions on VLAN models used, no use of PPP, no multi-node tracking of IGMP). Additional scenarios applicable to Broadband deployments are then described with the necessary additional requirements. The baseline requirements must be supported as described while the use of the subsequent requirements are needed if a given deployment scenario dictates.

6.2 Baseline Multicast Description

There is desire to deploy IP video services (broadcast and unicast) over the broadband infrastructure. Multicast functionality is required to make efficient use of network resources when delivering broadcast content. Figure 21 depicts the reference architecture and specifically indicates the scope of the multicast architecture within this document. The figure also highlights the points where multicast optimization should be defined and at what layer a given device is operating (Layer 2, Ethernet or Layer 3, IP).

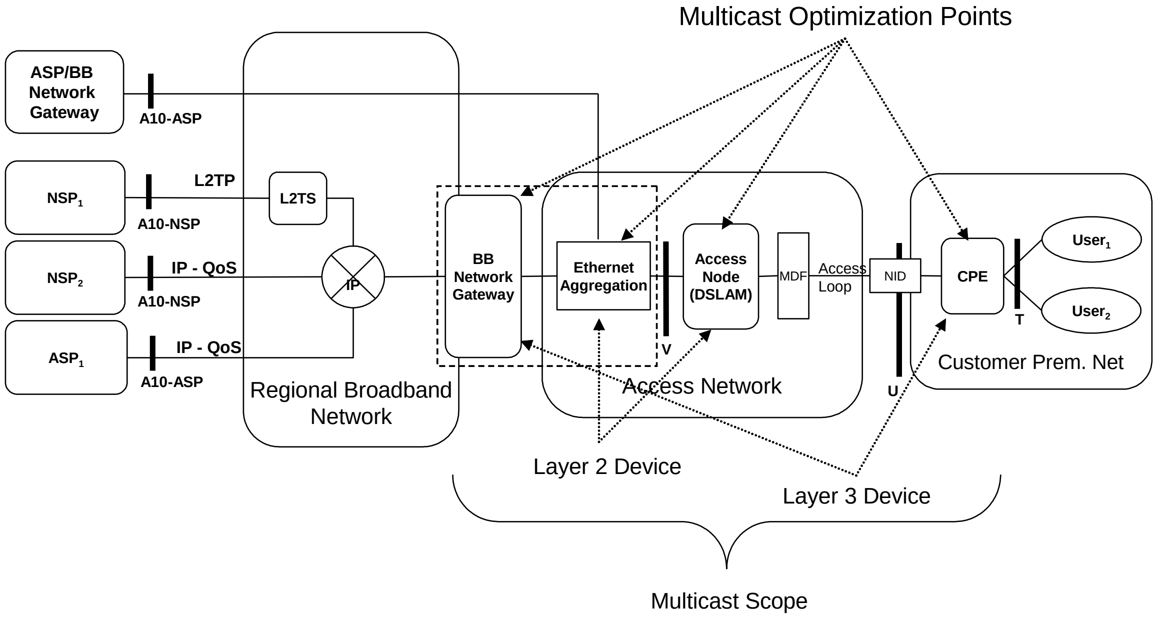


Figure 21 – Multicast Reference Model

The goal is to support multicast optimization by controlling the flooding of Ethernet multicast frames by making use of IGMP agents in intermediate (L2) devices. Such agents can locally set filters on the device such that packets are replicated only on those ports (physical and logical) that have requested the associated multicast group. Specifically, the data plane uses Multicast MAC-address based filters which link L2 multicast groups to egress ports on bridging devices. These bridges include both aggregation switches and Access Nodes. The net effect is that a bridge, upon receiving a packet destined to a certain multicast group will limit the flooding of that packet to the list of ports attached to the filter. Note that these filters are applied to reduce bandwidth. Without them a bridge will flood multicast out on all ports that are part of that multicast VLAN.

In order to automate the setup of these filters, and as the multicast traffic is IP multicast, IGMP snooping will be used. A bridge forwarding engine will redirect IGMP packets to its controlling function. Based on the requested IP multicast group, the bridge will set up a L2 multicast filter entry that allows or prevents packets to flow to the port on which it received the IGMP report. Ports that have routers attached (either directly or via other bridges) are automatically discovered based on IGMP General Queries received on them. So in essence IGMP messages in the upstream direction (from user to router) on a certain VLAN will set up state (MAC-address filters) in bridges to limit the flooding of multicast data in the downstream direction.

N:1 VLANs will be used in order to efficiently forward multicast traffic. It should be noted that other types of traffic (data, voice, unicast video) could be delivered via N:1 VLANs as well. Different traffic types (multicast/unicast video, data, voice) could be delivered over the same VLAN, or this VLAN can be dedicated to multicast traffic.

Out of this reasoning we extract two models:

a) Dedicated Multicast VLAN model.
This is a model where a dedicated N:1 VLAN is used to send some multicast groups from a multicast router / BNG over an aggregation network to one or several Access Nodes.,. Other traffic is sent across different VLANs, where these VLANs could be 1:1 or N:1. The Access Node is responsible for forwarding multicast traffic to a user port, based on the reception of control plane traffic (IGMP) on the user port.

b) Integrated Multicast VLAN model
This is a model where multicast traffic is inserted into one of the N:1 VLANs that are terminated at a user port, or alternatively .1q trunked to the RG. This effectively means multicast and unicast share a VLAN.

Note: the multicast VLAN may be tagged or untagged. Tagging options of a multicast VLAN are as described in Section 3.1.1.

In the description and requirements below, when the term ‘multicast VLAN’ is used it refers to either option a) or b) above.

The baseline attributes described are modeled on existing deployments (e.g. no restrictions on VLAN models used, of use of PPP, multi-node tracking of IGMP). Such a model has the following attributes:

- Efficient L2 replication in the Ethernet aggregation network through the use of N:1 VLANs
- Support for IGMP V2 and V3
- Support for ASM and SSM models
- Support for multiple content injection points in the network
- Support for multiple multicast VLANs in the access network
- IP (and IGMP) packets are directly encapsulated in Ethernet frames
- IGMP hosts are connected to Access Node ports that are members of the multipoint VLAN that will carry (receive) the multicast frames
- IGMP packets are transmitted in that same VLAN from which the multicast packet will be received
- User ports can be members of multiple VLANs.

Finally, there is an underlying assumption that the Ethernet elements in this architecture (the Access and Aggregation Nodes) will support IGMP v3 in the control plane, but not in the forwarding plane. This assumption yields 2 observations:

1. There are 2^{28} IP multicast addresses available (the first four bits of the IP address identify the multicast range, i.e. 224/4). When mapping a multicast IP address to a multicast MAC address, the lower 23 bits of the IP multicast address are mapped to the lower 23 bits of the Ethernet multicast address (01:00:5E:00:00:00 through 01:00:5E:7F:FF:FF). This means that the mapping ratio is 2^5:1 or 32:1. Thus, the network operator will need to take precautions not to use multicast groups that map to the same Ethernet multicast address on the same VLAN.

2. Because the Ethernet multicast-forwarding plane is not aware of the IP source address of the multicast groups, there is no discernment between (SSM) multicast groups that differ only in their source address. Once again the network operator will need to ensure that multicast groups that share the same group address but have different source addresses are not forwarded over the

same VLAN at the same time. Tools are provided in TR-101 to aid this task. Classifiers and filters are provided in the Ethernet elements so that IGMP may be assigned and/or filtered with respect to its proper multicast VLAN.

Based on these characteristics a model can be constructed for supporting Broadband customers.

6.2.1 RG Requirements

The RG is a layer 3 router, which uses IPoE to receive multicast from the access network and forward it to the home network. It also provides local DHCP and NAT capabilities into the home network environment - including the multicast host equipment. With the addition of multicast hosts in the home the RG needs additional mechanisms to forward IGMP reports to the access network and distribute the multicast within the home environment. Therefore, the RG needs to provide an IGMP proxy-routing function. This is a function typically provided on layer 3 devices and can be summarized by the statement: the RG provides the multicast router function into the home subnets and appears to the access network as a host device. As typical for RG deployments, multiple sessions must be supported.

The RG requirements are as follows:

R-225 The RG MUST support an IGMP Proxy-Routing function.

R-226 The RG MUST support IGMP version 3 as per RFC 3376.

R-227 The RG MUST support IGMP forwarding with local NAT and firewall features including establishing any pin-holes in the firewall for the multicast streams received (after join).

R-228 When the RG is configured with multiple WAN-facing IP interfaces (e.g. PPP or IPoE), the IGMP Proxy-Routing function MUST be able to multicast upstream IGMP messages to all or a configured subset of those WAN interfaces.

R-229 The RG MUST be configurable as to which interfaces IGMP messages should be forwarded to in the upstream direction. The default behavior MUST be to forward messages to all provisioned interfaces.

R-230 When the RG receives an IGMP membership query on a given WAN-facing IP interface, the IGMP Proxy-Routing function MUST only send a corresponding membership report on this specific interface.

R-231 The RG SHOULD be able to classify IGMP requests according to source IP/MAC address or incoming LAN physical port on the RG to distinguish between multicast services (e.g. IPTV and some other Best Effort (BE) Internet IGMP application).

R-232 The RG MUST be able to suppress the flooding of multicast on selected ports, either through dedicated ports connecting to IGMP hosts or IGMP Proxy-Routing.

R-233 The RG SHOULD be able to configure which ports are allowed to have IGMP hosts.

R-234 The RG MUST support IGMP immediate leave with explicit host tracking.

R-235 The RG MUST NOT forward UPNP multicast messages to its WAN interface.

6.2.2 Access Node Requirements

The Access Node provides the first point of aggregation and can make multicast more efficient. The Access Node is primarily a layer 2 device and as such needs to implement an IGMP snooping function in order to inspect IGMP packets and adjust multicast MAC forwarding filters. The Access Node receives the multicast streams via point to multipoint (N:1) multicast VLANs[6].

Within the context of this document the following procedures are assumed to take place at the Access Node:

1. Identify IGMP (configurable per port).

2. Process IGMP and create filters (configurable per VLAN).

3. Forwarding and further processing (if required e.g. proxy reporting) of IGMP (configurable per VLAN).

The Access Node requirements are as follows.

6.2.2.1 Per User-facing Port and VLAN Requirements

R-236 The Access Node MUST support the identification and processing of user-initiated IGMP messages. When this function is disabled on a port and/or VLAN, these messages are transparently forwarded.

Note that transparent forwarding of IGMP messages in N:1 VLANs might result in network flooding and is therefore discouraged. Hence, this function ought not be disabled.

R-237 The Access Node MUST support dropping of all IGMP messages received on a user port and/or VLAN.

R-238 The Access Node MUST support matching groups conveyed by IGMP messages to the list of groups (R-253) corresponding to a multicast VLAN associated with this port. When there is no match, the IGMP message MUST be either forwarded as regular user data or dropped. This behavior MUST be configurable. When there is a match, the IGMP message MUST be forwarded within a multicast VLAN, and enter the IGMP snooping function. Note that transparent forwarding of IGMP messages in N:1 VLANs might result in network flooding and is therefore discouraged.

IGMP V3 report messages may carry membership information for multiple multicast groups. Therefore, a single IGMP report message may carry membership information on groups ‘matching’ a multicast VLAN as well as on groups ‘not matching’ a multicast VLAN.

R-239 Upon receipt of an IGMP v3 report carrying information on a mix of ‘matching’ and ‘non-matching’ multicast groups (as described above), the Access Node SHOULD be able to copy the frame to the IGMP snooping function as well as forward it as user data (or drop it, as configured).

R-240 The Access Node MUST be configurable per port and/or VLAN to stop user ports injecting multicast traffic to the aggregation network.

R-241 The Access Node MUST be able to discard IGMP queries received from user-facing ports on a multicast VLAN.

R-242 The Access Node MUST be able to rate limit IGMP messages received from user-facing ports on a multicast VLAN.

R-243 The Access Node MUST support an IGMP v3 (as per RFC 3376) transparent snooping function. This feature MUST be configurable on a per VLAN basis. Note: V3 includes support of earlier

[6] Delivery of multicast content to the access node in a point-to-point VLAN (1:1) is not precluded and has no affect on the requirements included in this section.

versions of IGMP. Specifically, this function is responsible for configuring multicast filters such that packet replication is restricted to those user ports that requested receipt.

R-244 The Access Node's IGMP v3 transparent snooping function MUST support the capability to snoop the multicast source IP address and destination IP group address in IGMP packets and to set the corresponding MAC group address filters as specified in R-245.

R-245 The Access Node's IGMP v3 transparent snooping function MUST be able to dynamically create and delete MAC-level Group Filter entries to enable/disable selective multicast forwarding from network-facing VLANs to user-facing ports.

R-246 The Access Node MUST support IGMP immediate leave as part of the IGMP transparent snooping function.

R-247 Upon detecting topology changes (e.g. VLAN membership change, port being disabled by STP or network port changing state), the Access Node MUST be able to issue an IGMP proxy query solicitation, i.e. an IGMP Group Leave with group address '0.0.0.0'. This will indicate to the BNG it immediately needs to send Group Specific queries, which will populate the L2 multicast filters in the Access Node, in order to speed up network convergence. For reference see RFC4541, chapter 2.1.1 section 4.

R-248 For security purposes, the Access Node MUST drop any user-initiated IGMP Leave messages for group '0.0.0.0'.

R-249 The Access Node MUST support marking, in the upstream direction, user-initiated IGMP traffic with Ethernet priority bits.

R-250 The Access Node MUST support forwarding user initiated IGMP messages to a given multicast VLAN to which that user is attached.

R-251 The Access Node SHOULD provide the following statistics.
Per VLAN, per multicast group counters:

1. Total number of currently active hosts

Per - port, per multicast VLAN counters:

1. Total number of successful joins[7]
2. Total number of leave messages
3. Total number of general queries sent to users
4. Total number of group-specific queries sent to users
5. Total number of invalid IGMP messages received

Per multicast VLAN counters:

1. Current number of active groups
2. Total number of joins sent to network
3. Total number of successful joins from users
4. Total number of leave messages sent to network
5. Total number of leave messages received from users
6. Total number of general queries sent to users
7. Total number of general queries received from network
8. Total number of group-specific queries sent to users
9. Total number of group-specific queries received from network

[7] Successful join is defined as a valid IGMP join request for which the device has set the MAC filters for the delivery of the multicast group and not violated any IGMP processing thresholds (e.g. number of simultaneous groups, joins sent per second)

10. Total number of invalid IGMP messages received

6.2.2.2 Access Node Configuration Requirements

R-252 The Access Node MUST support configuring which user ports are members of a multicast VLAN.

R-253 The Access Node MUST allow the configuration of IP multicast groups or ranges of multicast groups per multicast VLAN based on:

- Source address matching
- Group address matching

R-254 The Access Node MUST be able to configure per - port the maximum number of simultaneous multicast groups allowed.

Note: This allows the Access Node to protect against denial of service attacks.

R-255 The Access Node MUST support enabling IGMP snooping on a per VLAN basis.

R-256 The Access Node MUST drop IGMP v1 messages.

6.2.3 Aggregation Node Requirements

R-257 The Aggregation Node MUST support an IGMP v3 transparent snooping function on a per VLAN basis. Note: V3 includes support for previous versions of IGMP.

R-258 The Aggregation Node MUST provide the following statistics.
Per VLAN, per multicast group counters:

1. Total number of currently active hosts

Per port, per multicast VLAN counters:

1. Total number of successful joins
2. Total number of leave messages
3. Total number of general queries sent
4. Total number of specific queries sent
5. Total number of invalid IGMP messages received

Per multicast VLAN counters:

1. Current number of active groups
2. Total number of joins sent
3. Total number of joins received
4. Total number of successful joins
5. Total number of leave messages
6. Total number of general queries sent
7. Total number of general queries received
8. Total number of specific queries sent
9. Total number of specific queries received
10. Total number of invalid IGMP messages received

R-259 The Aggregation Node's IGMP v3 snooping function MUST be able to dynamically create and delete MAC-level Group Filter entries, to enable/disable, selective multicast forwarding from network-facing VLANs to user-facing ports.

R-260 Upon detecting topology changes, the Aggregation Node MUST be able to issue an IGMP proxy query solicitation, i.e. an IGMP Group Leave with group address '0.0.0.0'. This will indicate to the BNG it immediately needs to send Group Specific queries, which will populate the L2 multicast filters in the Aggregation Node, in order to speed up network convergence. For reference see RFC4541, chapter 2.1.1 section 4.

6.2.4 BNG Requirements

The BNG includes an IGMP router function that must support the following requirements:

R-261 The Broadband Network Gateway MUST support multicast routing capabilities per TR-092 Appendix A "Multicast Support."

R-262 The Broadband Network Gateway MUST support IGMPv3. Note: IGMP v3 includes support for endpoints using earlier IGMP versions.

R-263 The Broadband Network Gateway MUST support IGMPv2 group to source address mapping for IGMP v2 to PIM/SSM compatibility.

R-264 The Broadband Network Gateway MUST provide the following statistics.
Per VLAN, per multicast group counters:

1. Total number of currently active hosts

Per Access Network facing port, per multicast VLAN counters:

1. Total number of successful joins
2. Total number of leave messages
3. Total number of general queries sent
4. Total number of specific queries sent
5. Total number of invalid IGMP messages received

Per multicast VLAN counters:

1. Current number of active groups
2. Total number of joins received
3. Total number of successful joins
4. Total number of leave messages received
5. Total number of general queries sent
6. Total number of specific queries sent
7. Total number of invalid IGMP messages received

R-265 The Broadband Network Gateway MUST support forwarding the multicast traffic on the same Layer 2 interface on which it receives the IGMP joins.

R-266 The Broadband Network Gateway MUST support the following configurable parameters per port (i.e. physical or logical port (VLAN), but not per end user). This allows the Broadband Network Gateway to enforce service level agreements in real-time.

- Maximum number of simultaneous multicast groups allowed for the port
- Maximum accumulated bandwidth allowed for multicast services

R-267 The Broadband Network Gateway MUST support IGMP immediate leave as part of the IGMP router function.

R-268 The Broadband Network Gateway MUST immediately send Group Specific Queries out of an interface if it receives an IGMP query solicitation message (i.e. a Group Leave for group '0.0.0.0').

6.3 Specific Broadband Considerations

6.3.1 Goals

The requirements described in Section 6.2 will support the deployment of multicast services in a Broadband environment; however, it does not provide coverage for all expected deployment scenarios.

This section details other deployment options and the requirements necessary to support them. Specifically, this section will extend the baseline framework to support:

- Existing PPP deployments
- Hierarchical Scheduling
- Multiple BNG deployment models

Figure 22 illustrates the decision tree for determining which features are needed in addition to the baseline (Section 6.2)[8]. The key questions shown within the decision tree are described in more detail below. The order of these questions as well as the need to answer a given question is determined based on the path taken through the decision tree.

Question 1: Single or Dual node deployment? Very simply put: does the service provider have a single or dual injection point for multicast and other data traffic?

Single – All traffic to and from a given user flows through a single BNG

Dual – Video and other data traffic are sourced from two separate BNGs in the network

Question 2: If a single node deployment is used, is PPPoE used for the transport of unicast (where there is no replication between the BNG and the user) traffic?

Yes – PPPoE encapsulation is the default route for forwarding traffic from the user premises to the network/BNG. Since all IGMP messages will be received at the BNG, per user IGMP statistics and hierarchical scheduling can be supported.

- Additional forwarding features for IGMP must be enabled to allow for multicast replication.

No – IPoE encapsulation is leveraged for forwarding traffic from the user premises to the network/BNG. This scenario follows what is described in the baseline.

Question 3: As PPPoE is the default forwarding behavior from the customer premises, all traffic (including IGMP) will be forwarded in that manner. To enable multicast replication IGMP messages must also be sent outside of the PPP session. The key question at this point is: Will an RG based approach be used or will an Access Node approach be used?

RG Approach – Capabilities within the RG ‘fork’ IGMP messages so that they are forwarded up as both PPPoE as well as IPoE packets.

Access Node Approach – Capabilities within the Access Node observe IGMP messages within the PPPoE session and generate IPoE IGMP messages.

Question 4: If a dual node deployment is chosen, are BNG-based per-user IGMP statistics (collected at a node that does not source the multicast) and/or dynamic hierarchal scheduling of bandwidth available to both nodes required based on observation of IGMP?

Yes – Per-user IGMP messages must be received by the BNG and not be aggregated within the network. Note that for dual node deployments support of this capability is out the scope of this document.

[8] Based on the deployment strategies envisioned by participating service providers within the BBF

No – Per-user IGMP messages need not be received by the BNG and may be aggregated within the network.

Question 5: Is aggregation of IGMP messages desired?

Yes – Proxy reporting features are required in the network to support this capability.

No – No additional features are required.

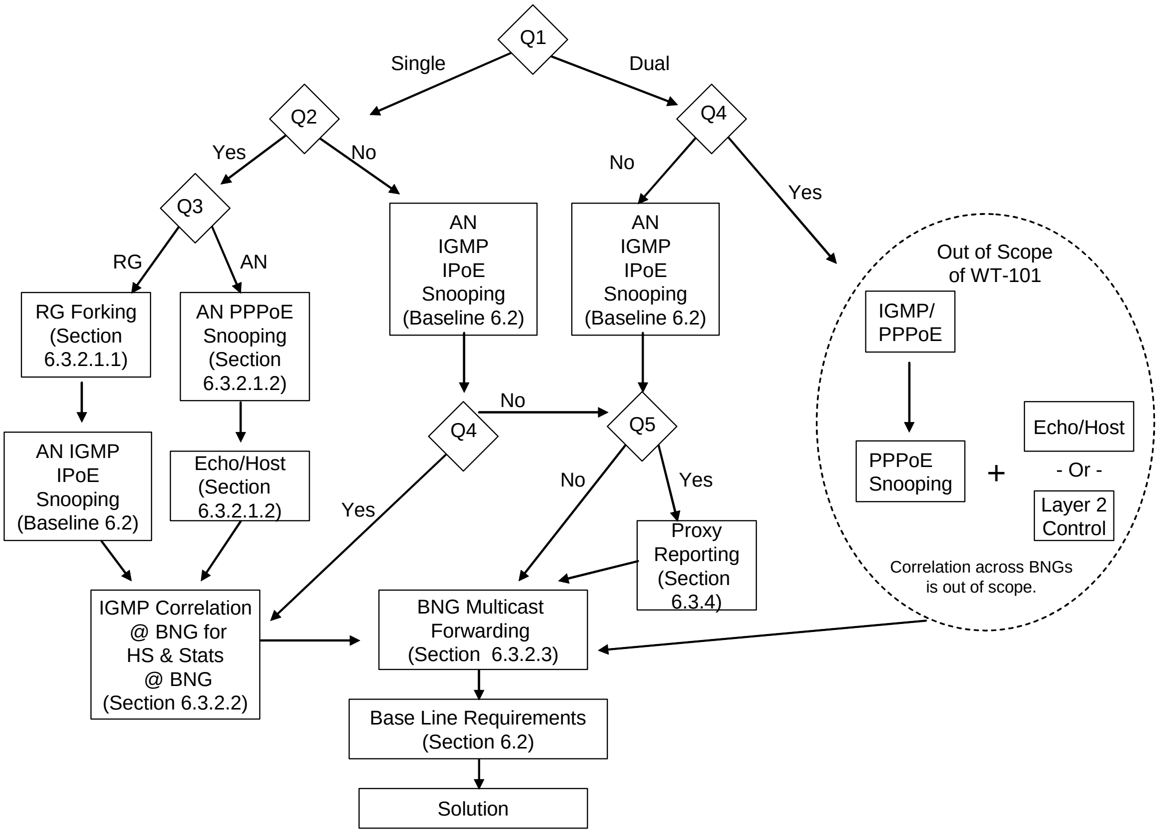

Q1
Single
Dual
Q2
Q4
Yes
No
No
Yes
Q3
RG
AN
AN IGMP IPoE Snooping (Baseline 6.2)
AN IGMP IPoE Snooping (Baseline 6.2)
Out of Scope of WT-101
IGMP/ PPPoE
RG Forking (Section 6.3.2.1.1)
AN PPPoE Snooping (Section 6.3.2.1.2)
No
Q4
Q5
AN IGMP IPoE Snooping (Baseline 6.2)
Echo/Host (Section 6.3.2.1.2)
Yes
No
Yes
PPPoE Snooping
+
Echo/Host
- Or -
Layer 2 Control
Proxy Reporting (Section 6.3.4)
Correlation across BNGs is out of scope.
IGMP Correlation @ BNG for HS & Stats @ BNG (Section 6.3.2.2)
BNG Multicast Forwarding (Section 6.3.2.3)
Base Line Requirements (Section 6.2)
Solution

Figure 22 – Multicast deployment decision tree

6.3.2 Single Node Deployments

6.3.2.1 PPPoE Deployments

In current Broadband deployments PPPoE is often used for providing connectivity to users. PPP, as the name implies, is designed to establish a point-to-point connection between two endpoints and does not inherently provide support for multicast injection at intermediate points along the connection. So, if IGMP is to be forwarded in the same path as the multicast groups, the IGMP messages must be sent in an IPoE encapsulation. This can be supported in two ways:

1. IGMP messages received at the RG are forwarded as both PPPoE and IPoE packets to the network (referred to as RG-based approach).
2. IGMP messages are sent from the RG only as PPPoE and must be observed from within the PPPoE session and replicated as IPoE and forwarded in the multicast VLAN by the Access node (referred to as Access Node based approach).

6.3.2.1.1 RG Based Approach

R-269 It MUST be possible to configure an RG WAN-facing IP interface with an IPoE encapsulation and no IP address visible by the access network.

R-270 The RG MUST be able to receive downstream multicast traffic on the interface described in R-269, independent of whether upstream IGMP is sent on this interface or not.

R-271 The RG IGMP Proxy-Routing function MUST be able to send upstream IGMP traffic on the interface described in R-269, using a null (0.0.0.0) IP source address.

6.3.2.1.2 Access Node Based Approach

R-272 The Access Node MUST support an IGMP/PPPoE transparent snooping function. This capability will use the methods described for classification and establishment of group address filters based on the baseline requirements (Section 6.2.2).

R-273 For those IGMP packets observed within PPPoE the Access Node MUST be able to trigger a local IGMP Host function (a.k.a "echo client") when a group is joined or left by a user-facing port. The Access Node IGMP Host function MUST then locally generate IGMP/IPoE messages (e.g. membership report/leave) and locally reply to IGMP membership queries to reflect the groups whose delivery to the Access Node is needed. The IGMP Host function MUST be triggered in the context of the multicast VLAN.

Notes:

- This requires support for R-269 and R-270 at the RG as well.
- The Host function in the Access Node may be optionally followed by an Access Node proxy/aggregation function.

R-274 Triggering of the local IGMP Host at the Access Node MUST be configurable per multicast VLAN and user port.

6.3.2.2 IPoE Deployments

R-275 For IPoE IGMP packets observed within the regular 1:1 VLAN user data path, the Access Node SHOULD be able to trigger a local IGMP Host function (a.k.a "echo client") when a group is joined or left by a user-facing port. Using R-238, the host function is triggered in the appropriate multicast VLAN. The Access Node IGMP Host function MUST then locally generate IGMP/IPoE messages (e.g. membership report/leave) and locally reply to IGMP membership queries to reflect the groups whose delivery to the Access Node is needed. The Host function in the Access Node SHOULD be followed by an Access Node proxy/aggregation function or filtering function.

6.3.2.3 IGMP Processing at the BNG for HS and User Statistics

The requirements within this section assume that the BNG receives all IGMP messages sent by RGs unaltered.

These messages may then be used for:

- Gathering statistics
- Adjusting dynamic IP Hierarchical Schedulers (HS).

IGMP is sent upstream on both the PPPoE as well as the IPoE multicast VLAN. This is by design – as the multicast VLAN requires the IGMP in order to set the multicast switching filters in the Ethernet fabrics, and the PPPoE connection may require the IGMP messages in order to deliver them to an ISP, or to use them for the statistics and HS adjustments. In this architecture IGMP is received over both the IPoE multicast VLAN as well as the PPPoE session, and therefore two main BNG implementation strategies are possible.

- The first approach is that a BNG uses the IGMP over the multicast VLAN interface and correlates statistics and HS adjustments to the proper PPPoE session's interface by matching the source address. Note that since the IP source address may be 0.0.0.0 - this matching may require comparison of MAC source address of the IGMP with that used for the PPPoE session.
- The second approach is essentially the reverse: the BNG may use the IGMP from the PPPoE session to correlate to the appropriate multicast traffic levels present in the IPoE multicast VLAN interface. Once again, the source address may be required to determine from which (of potentially many) multicast VLAN the traffic level should be metered. And once again, this source address may be a MAC address.

Note that MAC address correlation is also needed to shape multiple PPPoE sessions correctly when they originate from the same access line. This is the TR-059 multi-session requirement.

Also, TR-101 specifies a PPPoE multicast architecture that would allow deployment of either implementation of IGMP correlation in a BNG – or even deployment of both implementations – in a single network design without changes in design.

6.3.2.3.1 BNG Requirements

R-276 A Broadband Network Gateway supporting hierarchical scheduling MUST support dynamic adjustment of the user-facing QoS shapers to reflect changes in the number of multicast groups joined by a user. (These adjustments would be inclusive of all levels of the hierarchy).

R-277 A Broadband Network Gateway supporting hierarchical scheduling MUST be able to trigger dynamic adjustment of the user-facing QoS shapers based on the tracking of IGMP messages received on both regular user-facing interfaces as well as on the appropriate multicast VLAN, and also based on local knowledge of the peak-rate of multicast streams. The correlation mechanism to identify the proper scheduler node with an associated multicast group or groups is an implementation option of the BNG. The PPPoE session VLAN and IPoE multicast VLAN may or may not be the same.

R-278 A Broadband Network Gateway supporting hierarchical scheduling SHOULD debit the amount of traffic offered by a given multicast group from the user-facing QoS shapers based on a provisioned association between a multicast group and a peak information rate.

R-279 A Broadband Network Gateway supporting hierarchical scheduling MAY debit on a packet by packet basis the amount of traffic offered by a given multicast group from the user-facing QoS shapers on a real time basis.

6.3.2.4 BNG Multicast Forwarding Requirements

R-280 The Broadband Network Gateway MUST support configuration of a map of multicast groups, indicating which groups should be handled using special functions, e.g. this map (or set of maps) can be used to specify the multicast groups that must be delivered by the regular (user-facing) IP interface, as well as the multicast groups that must be delivered by specific (and possibly different) multicast VLAN interfaces, or the map can be used for static or dynamic debits of HS schedulers without affecting replication of multicast traffic. The map(s) MUST be configurable based on specific source and multicast group addresses as well as ranges of source and group addresses. IGMP requests corresponding to groups that are not in the map are assumed to be requests for multicast from the NSP and MUST follow standard IGMP processing.

6.3.3 Dual Node Deployments

The dual edge scenario assumes IGMP is IPoE encapsulated as per the baseline. The use of PPPoE with dual nodes and the need to support user statistics and dynamic hierarchical scheduling at a secondary node are out of scope for this document. Dual node deployments (as well as single node deployments) can also be supported using the distributed precedence and scheduling or static partitioning QoS models as described in Section 2.9 with no new requirements.

6.3.4 Proxy Reporting Support

An Access Node providing snooping with proxy-reporting function has to support the following requirements in addition to those specified in Section 6.2.2, except for R-281 below which replaces R-243. An Aggregation Node providing snooping with proxy-reporting function has to support the following requirements in addition to those specified in Section 6.2.3, except for R-281 below which replaces R-257.

R-281 The Access and Aggregation Nodes MUST support IGMP v3 snooping with proxy reporting. This feature MUST be configurable on a per VLAN basis.

R-282 The Access and Aggregation Nodes MUST allow selection between transparent snooping and snooping with proxy reporting on a per-VLAN basis.

R-283 The IGMP snooping with proxy reporting function MUST support IGMP proxy query functions.

R-284 The Access Node proxy-reporting function MUST support marking IGMP traffic it initiates with Ethernet (VLAN) priority bits.

7. OAM

7.1 Ethernet OAM

The intent of OAM mechanisms for Ethernet is to maintain similar Layer 2 end-to-end OAM capabilities that were available when the aggregation was ATM-based, via ATM OAM as specified in ITU-T (I.610).

This section is based on IEEE 802.1ag-2007, and ITU Y.1731. Devices are typically capable of supporting Ethernet OAM basic functionality. This is required in order to successfully manage the Ethernet network.

IEEE 802.1ag and ITU-T Y.1731 define the concept of a Maintenance association End Point (MEP) and Maintenance association Intermediate Point (MIP), which are configured on a per port, per VLAN and per Maintenance Domain Level. MEPs initiate Connectivity Fault Management (CFM) OAM[9] messages and are configured at the far end of the service perimeter or S-VLAN (e.g. in the BNG and Access Nodes). MIPs are configured across the path of the S-VLAN (e.g. in Ethernet Aggregation Nodes). Various Domain Maintenance Entity (ME) Levels can be configured, allowing the network administrator to divide the network into multiple administrative OAM domains and to allow nesting of OAM domains, where an ME Level corresponds to an OAM domain. An example of three Maintenance Domain Levels in a Broadband network is shown in Figure 23. The “Customer” domain extends between the BNG and the RG. The “Carrier” domain spans the carrier network, from the edge of the Access Node to the BNG. The “Intra-Carrier” domain goes from the carrier side of the Access Node to the BNG.

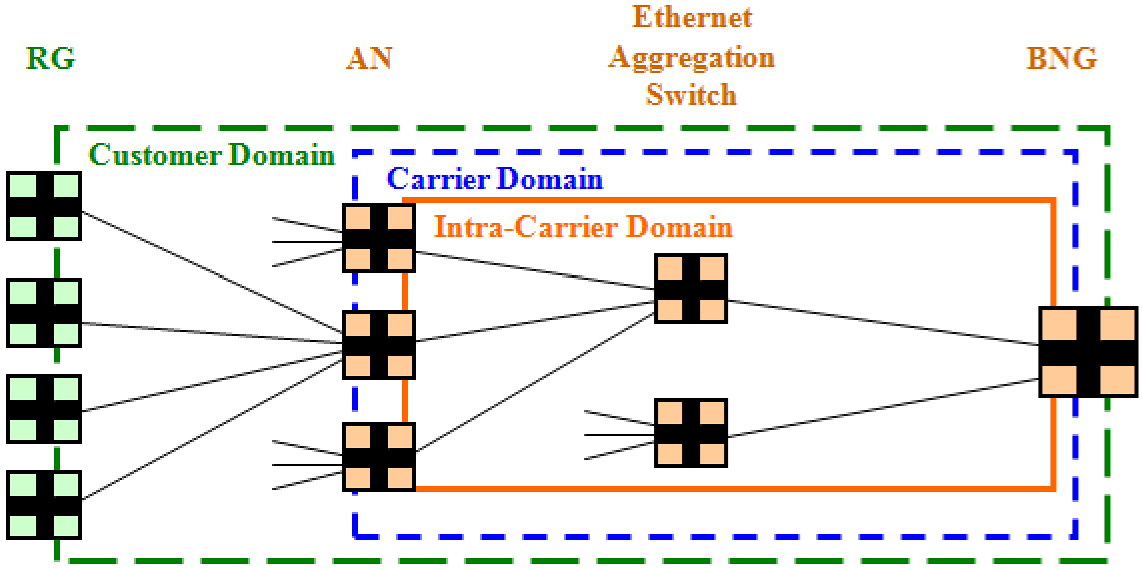


Figure 23 – Example of Ethernet OAM maintenance domains in a Broadband network

CFM Ethernet OAM defines, amongst others, the following message types, allowing the various MEPs and MIPs to communicate with one another:

- Continuity Check Message (CCM): This is a multicast message from a MEP that is received by all MEPs, or a specific MEP, in the same service instance on the same ME level. In this way all MPs at a given level have visibility into MEPs active at that level.

[9] Ethernet OAM frames are referred to as Connectivity Fault Management frames in IEEE 802.1ag.

- Loopback Message (LBM): This is usually a unicast message sent to a MP's MAC address, but can be a multicast message which flows to all remote MEP(s)[10]. Note that the multicast version of the LBM can at times be used to learn the MAC address of a remote MEP, or could be used if there is only one remote MEP within the service instance and/or VLAN; i.e., there is no need to know the MAC-address of the remote MEP.
- Loopback Reply (LBR): This is a unicast message sent from the receiving MP to the MAC address of the MEP that originated the LBM.
- Link Trace Message (LTM): A multicast message that is relayed to all MEPs in the Maintenance Association, and inspected by every MIP along the path to determine whether it can be forwarded by the MIP. If able, the receiving MIP forwards the LTM towards its target MAC and sends a Link Trace Reply (LTR) back to the originator of the LTM. The MEP at the target MAC terminates the LTM message. If the target MAC is unreachable from a MIP, the Link Trace is terminated and an LTR is initiated back to the originating MEP.
- Link Trace Reply (LTR): A unicast message sent from a MEP/MIP upon receiving and forwarding a LTM. MIPs with knowledge of the target MAC address will forward the LTM on the appropriate bridge port and generate a successful LTR toward the originator of the LTM. A MIP with no knowledge of the target MAC address will issue an LTR message.

7.2 Ethernet OAM Model for Broadband Access

IEEE 802.1ag and Y.1731 define 8 possible maintenance domain ME levels for Ethernet OAM, numbered 0-7. In a Broadband network, there are four distinct levels that deserve attention. Numerical values are not specified by this document for all levels, but rather in this section they will be referred to as the Customer, Carrier, Intra-Carrier, and Access Link.[11] The primary constraint is that the Customer level is “higher” or superior to (numerically greater than) the Carrier, which in turn is higher or superior to (numerically greater than) the Intra-Carrier, and Intra-Carrier is superior to the Access Link. In any specific implementation, a carrier should assign specific numerical values to those levels, although defaults are specified for levels that flow to RGs.

Figure 24 illustrates these concepts indicating where the BNG is operated by the same access provider. Note that there are “shorter” versions of the Carrier and Intra-Carrier levels shown in Figure 25. These short levels are relevant for a carrier that offers an Ethernet access service to a separate carrier that operates the BNG.

[10] Y.1731 allows a LBM to be sent to a multicast address. To support interworking, 802.1ag-2007 specifies that a MEP that receives a LBM with a multicast DA must respond to it.

[11] The Customer, Carrier, and Intra-Carrier levels are depicted as the Green, Blue, Orange levels respectively.

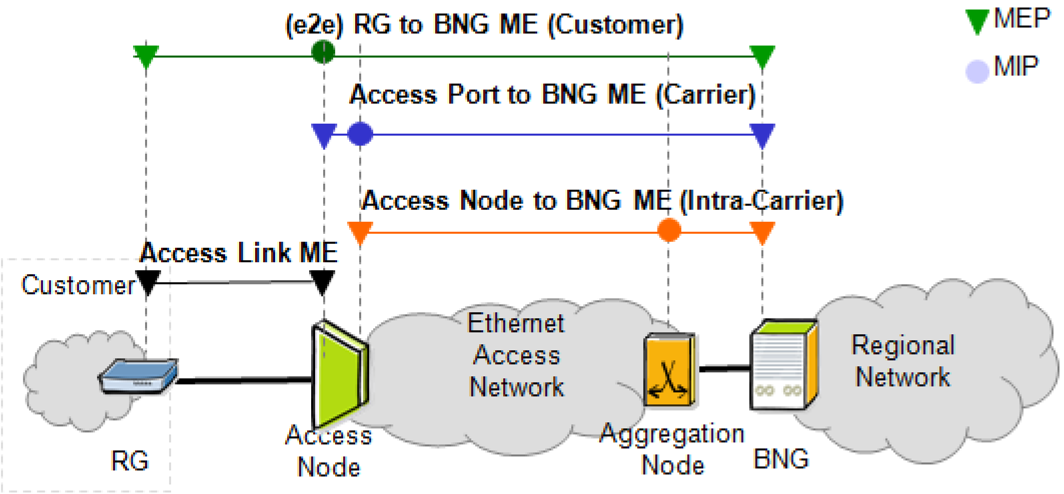


Figure 24 – Ethernet OAM model for broadband access

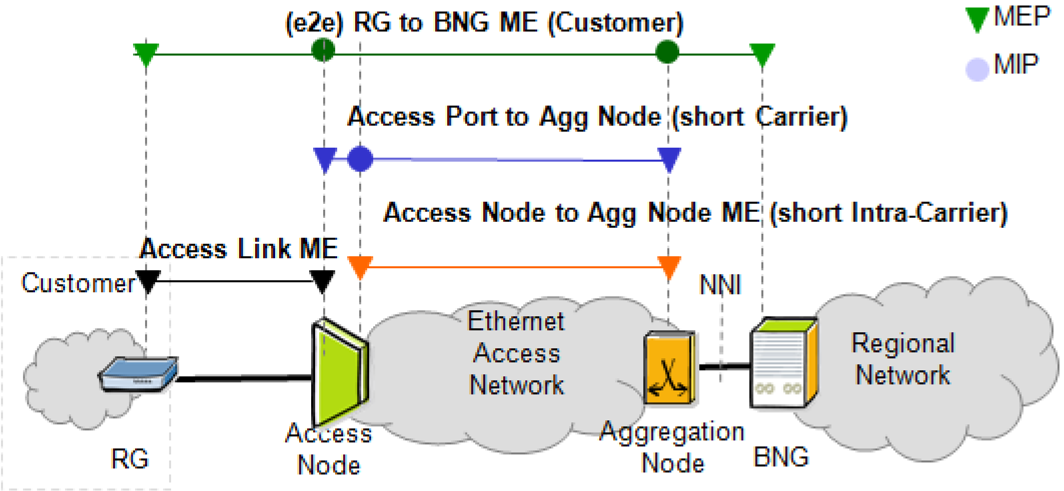

(e2e) RG to BNG ME (Customer)
MEP
MIP
Access Port to Agg Node (short Carrier)
Access Node to Agg Node ME (short Intra-Carrier)
Access Link ME
Customer
NNI
Ethernet Access Network
Regional Network
RG
Access Node
Aggregation Node
BNG

Figure 25 – Ethernet OAM model for broadband access – wholesale "Ethernet bit-stream" services model

Several other terms will be used in the following requirements, and require definition:

- Inward-facing MEP: A MEP that faces toward the bridge device. In the figure above, the endpoint of the Carrier level at the AN is an inward-facing MEP. This is called an "Up" MEP in 802.1ag.
- Outward-facing MEP: A MEP that faces away from the bridge device. In the figure above, all MEPs at the BNG are outward-facing MEPs. This is called a "Down" MEP in 802.1ag.

  Note that only MEPs have directionality associated with them. MIPs have a diagnostic function in both directions.
- Bridge port model: The 802.1ag standard has been written such that MEPs and MIPs are resident on bridge ports and have a unique MAC-address within the maintenance domain.
- Bridge brain or Master port model: This is a model where multiple MEPs and MIPs on a bridge share a MAC-address, i.e. that of the CPU on the 'master-port' of the bridge.

It should be noted that although 802.1ag dictates that MPs are only addressable by a MAC-address, there is no need to know the MAC-address of the remote MEP in a 1:1 VLAN construct. ITU-T Y.1731 defines this as VLAN OAM and documents that one could use a multicast address rather than the remote MEP's MAC-address to target that remote MEP. If a VLAN has more than two MEPs (i.e. in the N:1 VLAN case), the MAC-addresses of the remote MEPs have to be known to all MEPs. Usually this happens by having the MEPs send continuity checks, thereby announcing their MAC address. In a broadband environment however, other methods of resolving the remote MEPs MAC addresses may have to be used.

7.3 Ethernet OAM Requirements

7.3.1 RG Requirements

7.3.1.1 Customer Maintenance Level

The RG requirements only apply to RGs that support Ethernet OAM. For such devices, only the Customer and Access Link levels are relevant.

R-285 The RG MUST support a Maintenance association End Point (MEP) on a per VLAN basis.

R-286 The RG MUST support a default ME level value of 5 for the Customer level.

Of course, a different value than the default could be selected if needed.[12]

R-287 The RG SHOULD support a Loopback Message (LBM) function that can generate a Multicast LBM towards its peer MEP(s). This requirement allows the RG to dynamically learn the MAC address of the BNG MEP (i.e. from the LBR) and could also be used to announce the RG MEP MAC address to the BNG upon request.

Upon receiving a LBM, the RG needs to respond by initiating a LBR. In other words, it must support the LBM sink function and the LBR source function.

R-288 The RG MUST support a Loopback Reply (LBR) function towards its peer MEP(s) in response to both unicast and multicast LBMs.

R-289 The RG MUST support a Link Trace Reply (LTR) function towards the source address of a received LTM.

R-290 For business customers and/or premium customers requiring proactive monitoring, the RG SHOULD support generating Continuity Check Messages (CCMs).

R-291 The RG MUST support turning off the sending of CCMs, while keeping the associated MEP active.

R-292 The RG MUST support receiving AIS messages.

R-293 The RG SHOULD trigger the appropriate alarms for Loss of Continuity.

7.3.1.2 Access Link Maintenance Level

This level supports the testing of the Ethernet layer from the RG to the AN.

R-294 The RG MUST support a Maintenance association End Point (MEP) on a per VLAN basis.

[12] Note that the Metro Ethernet Forum (MEF) has defined a "Test" Maintenance Entity Group to allow a carrier to test connectivity to the customer (e.g., to send Loopbacks to the subscriber location), and uses level 5 for that purpose. If a service is also a MEF-compliant service, then ME level 6 or 7 should be used for the Customer level.

R-295 The RG MUST support a default ME level value of 1 for the Access Link level. [13]

R-296 The RG SHOULD support a Loopback Message (LBM) function that can generate a Multicast LBM towards its peer MEP(s). This requirement allows the RG to dynamically learn the MAC address of the AN MEP, and test the connectivity to that MEP.

Note that the ability of the RG to generate a multicast LBM at the Customer level and at the Access Link level are sufficient to test connectivity to the near edge of the carrier's network and to the BNG, which are the only two points that are visible to the RG. A Link Trace initiation capability would provide no added value.

R-297 The RG MUST support a Loopback Reply (LBR) function towards its peer MEP(s), in response to both unicast and multicast LBMs.

7.3.2 Access Node Requirements

The Access Node can have Maintenance Points (MPs) at the Customer, Carrier, Intra-Carrier and Access Link levels.

7.3.2.1 Customer Maintenance Level

R-298 The Access Node MUST support a Maintenance association Intermediate Point (MIP) function on a per-user-port and per-VLAN basis.

R-299 The Access Node MUST support a Link Trace Reply (LTR) function for each MIP.

R-300 The Access Node MUST support a Loop Back Reply (LBR) function for each MIP.

R-301 The Access Node SHOULD support filtering CFM Ethernet OAM messages arriving on a user port. Specifically, the Access Node SHOULD support discarding LTMs arriving on a user port.

There is a concern that users could intentionally or unintentionally launch an attack on a carrier's network using Ethernet OAM messages. The Access Node must be able to protect against such events by limiting the number of Ethernet OAM messages that the Access Node will receive during a given time period. Those that exceed the limit would be discarded.

R-302 The Access Node MUST support rate limiting of CFM Ethernet OAM messages arriving on a user port. The rate MUST be configurable per port.

7.3.2.2 Carrier Maintenance Level

The MEP at the Carrier level in the AN is of great interest in a broadband network. In a typical Metro Ethernet network, it could be assumed that the interface to a customer has a MAC at that node. However, a MAC address may not always be available at a broadband port, particularly when it is DSL and supports ATM. This complicates the discussion of testing the Carrier's network to its edge.

If the AN port does not have a MAC, it remains desirable to test the data path between the port and the BNG.

Consider the AN as having two parts: the Ethernet part (shown in white in Figure 26 below), and the non-Ethernet (e.g., ATM) part (shown in blue/shaded).

One method would be to use the AN to initiate a test within the AN between the "first" Ethernet OAM point and the specified DSL port. This could be done through various proprietary methods within the AN, as

[13] The "lowest" value of 0 needs to be reserved to account for the case where there is a customer bridge device between the RG and the AN.

long as the (non-Ethernet) data path can be verified. In Figure 26, this is shown by the dashed line path to point "*". Then a test can be made of the path between the "first" Ethernet OAM point in the AN and the BNG, using Ethernet OAM methods.

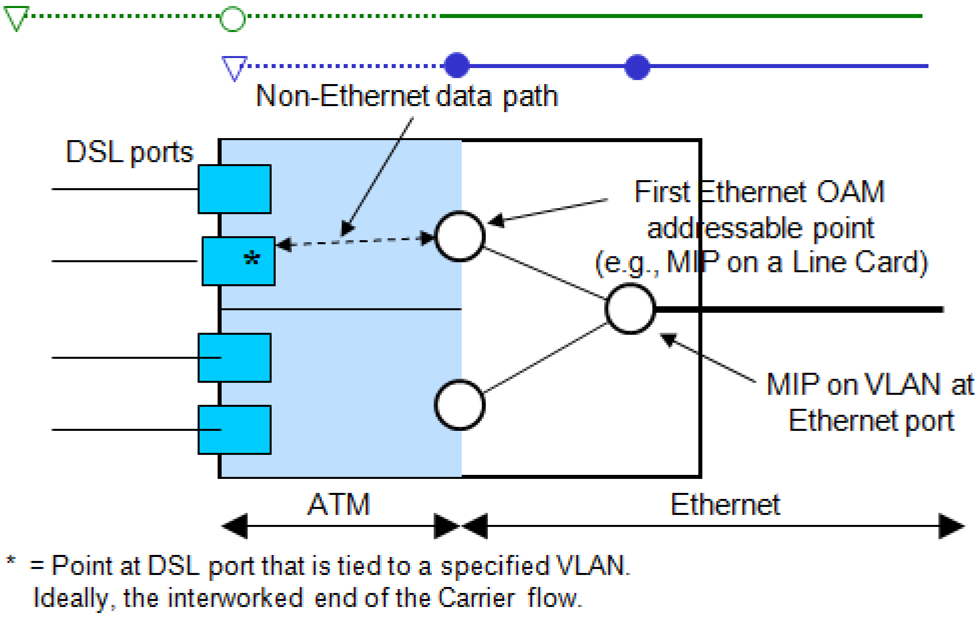

Non-Ethernet data path
DSL ports
*
First Ethernet OAM addressable point (e.g., MIP on a Line Card)
MIP on VLAN at Ethernet port
ATM
Ethernet
* = Point at DSL port that is tied to a specified VLAN. Ideally, the interworked end of the Carrier flow.

Figure 26 – The Ethernet and non-Ethernet flow within the AN

One way to initiate the test of the Carrier level is first for the AN Management System to initiate the test of the non-Ethernet data path. Then the BNG or AN can initiate a Loopback or Link Trace that flows between the BNG and the last Ethernet OAM addressable point.

Another method is to execute these functions through a Maintenance Channel (ETH-MCC). The BNG would initiate a Maintenance Channel request that includes the (non-MAC) identity of the AN port of interest. The AN would then perform a check of the path to the port to verify its operation, as described above. Finally, it would send a report of the success or failure of the check to the BNG, over the Maintenance Channel.

Yet another approach is to use a proxy function, often referred to as Virtual MEP, or vMEP. A vMEP behaves exactly like a MEP from the point of view of the other MEPs in the network, but runs inside the CPU of the access node on a per-port basis, rather than on the port itself. Note that the vMEP construct needs a separate MAC address per port, although they are all running on the same 'master-port' (CPU/bridge Brain). A vMEP is a MEP augmented to use receipt of Ethernet OAM transactions to trigger lower layer OAM functionality on the loop.

R-303 The AN MUST support the ability to receive a request from the management plane to test the user plane between the ATM DSL port and the first point that is Ethernet OAM addressable, where these do not coincide. The user plane test MUST as a minimum include checking the configured connectivity between a DSL port and the associated Ethernet VLAN(s) on the Access Node uplink.

R-304 The AN SHOULD support the ability to receive a request from the BNG to test the user plane between the ATM DSL port and the first point that is Ethernet OAM addressable, where these do not coincide. Notice that the above requirement does not specify whether that method uses a vMEP approach, a Maintenance Channel approach, or some other type.

R-305 In this maintenance level the Access Node has an "inward-facing" MEP on every user port (i.e. access loop termination). The Access Node MUST support an inward-facing Maintenance association End Point (MEP) on a per user port and per VLAN basis. In a basic implementation this could be achieved via the "Bridge Brain/Master Port" model; in this case it should be noted that this MEP will not be used to test the actual data path through the switch fabric of the Access Node.

R-306 The Access Node SHOULD support a Maintenance association Intermediate Point (MIP) as follows:

Per 1:1 VLAN: a MIP on a per network port and per C-VLAN basis (The S-VLAN appended at the network port is allocated a MEP at the Intra-Carrier level),

Per N:1 VLAN: a MIP on a per network port and per S-VLAN basis.

R-307 The Access Node SHOULD support initiating a Loopback Message (LBM) towards its peer MEP(s) and receiving the associated Loopback Reply (LBR), for the MEP on the user port.

R-308 The Access Note MUST support receiving a Loopback Message (LBM) from its peer MEP(s) and initiating the associated Loopback Reply (LBR), for the MEP on the user port.

R-309 The Access Node MUST support a Loop Back Reply (LBR) function for each MIP.

R-310 The Access Node SHOULD support initiating a Link Trace Message (LTM) (toward the multicast address) and receiving the associated Link Trace Reply (LTR) messages, for the MEP on the user port.

R-311 The Access Node MUST support the function of receiving a Link Trace Message (LTM) from a peer MEP and initiating the associated Link Trace Reply (LTR), for the MEP on the user port.

R-312 The Access Node MUST support receiving a Link Trace Message (LTM) from a peer MEP and initiating the associated Link Trace Reply (LTR), for the MIP on the network port.

R-313 The Access Node SHOULD support populating the table with the <MEP name, MAC address> associations for its peer MEP(s) via management.

Note that the MEP name is defined in 802.1ag as an optional field in CCMs. It uses the values defined in 802.1ab, i.e. a chassis ID and a port ID.

R-314 For business customers and/or premium customers requiring proactive monitoring, the Access Node SHOULD support generating Continuity Check Messages (CCMs) for the MEP on the user port.

R-315 The Access Node MUST support turning off sending CCMs (i.e. CCM source function disabled and sink function enabled) for the MEP on the user port, while keeping the associated MEP active.

R-316 The Access Node MUST support receiving AIS messages on the MEP on the user port (at a so-called inferior Maintenance Level) and send out an AIS message at the next-superior Maintenance Level (i.e. towards the RG).

R-317 For monitoring of customers that are using 1:1 VLANs but have an RG that does not support 802.1ag Ethernet OAM, the Access Node SHOULD support using a “Server MEP” function (defined in Y.1731) to report failure of a Server layer (e.g. ATM layer) on the access loop and send out an AIS message at the next-superior Maintenance Level (i.e. towards the BNG). This behavior MUST be configurable per port. Typically, it will only be activated for business customers.

The Access Node Server MEP function could also be used to generate Server layer AIS messages towards the RG (e.g. native ATM OAM) upon receiving an AIS message on the MEP on the user port. This would require interworking.

R-318 The Access Node SHOULD trigger the appropriate alarms for Loss of Continuity.

7.3.2.3 Intra-Carrier/Short Intra-Carrier Maintenance Level

R-319 The Access Node MUST support an outward-facing Maintenance association End Point (MEP) on a per-network-port and per-S-VLAN basis.

R-320 The Access Node SHOULD support initiating a Loopback Message (LBM) towards its peer MEP(s) and receiving the associated Loopback Reply (LBR), for the MEP(s) on the network port.

R-321 The Access Node MUST support receiving a Loopback Message (LBM) from its peer MEP(s) and initiating the associated Loopback Reply (LBR) for the MEP(s) on the network port.

R-322 The Access Node SHOULD support initiating a Link Trace Message (LTM) function (toward the multicast address) and receiving the associated Link Trace Reply (LTR) messages, for the MEP(s) on the network port.

R-323 The Access Node MUST support receiving a Link Trace Message (LTM) and initiating the associated Link Trace Reply (LTR) towards the source MEP, for the MEP(s) on the network port.

R-324 The Access Node SHOULD support populating the table with the <MEP name, MAC address> associations for its peer MEP(s) via management.

R-325 The Access Node SHOULD support generating Continuity Check Messages (CCMs) for the MEP(s) on the network port.

R-326 The Access Node MUST support turning off sending CCMs for the MEP(s) on the network port, while keeping the associated MEP active.

R-327 The Access Node MUST support receiving AIS messages on the MEP(s) on the network port (at a so-called inferior Maintenance Level) and send out an AIS message at the next-superior Maintenance Level.

R-328 The Access Node SHOULD trigger the appropriate alarms for Loss of Continuity.

7.3.2.4 Access Link Maintenance Level

This level supports the testing of the Ethernet layer to the RG, and, more importantly, discovery of the RG's MAC address by the AN.

R-329 The AN MUST support a Maintenance association End Point (MEP) on a per VLAN basis.

R-330 The AN MUST support a default ME level value of 1 for the Access Link level.

R-331 The AN MUST support a Loopback Message (LBM) function that can generate a Multicast LBM towards the RG. This requirement allows the AN to discover the MAC address of the RG, and also tests the connectivity to that MEP.

R-332 Upon receiving a LBR, the AN MUST return the RG's MAC address to the requestor.

R-333 The AN MUST support a Loopback Reply (LBR) function towards its peer MEP(s), in response to a unicast or multicast LBM.

R-334 The Access Node SHOULD trigger the appropriate alarms for Loss of Continuity.

7.3.3 Aggregation Node Requirements

The Aggregation Node can have Maintenance Points (MPs) at the Intra-Carrier and Carrier levels.

R-335 The Aggregation Node MUST support rate limiting of received CFM Ethernet OAM messages arriving on all maintenance levels.

7.3.3.1 Intra-Carrier Maintenance Level

R-336 The Aggregation Node MUST support a Maintenance association Intermediate Point (MIP) on a per port and per S-VLAN basis.

R-337 The Aggregation Node MUST support a Link Trace Reply (LTR) function for each MIP.

R-338 The Aggregation Node MUST support a Loop Back Reply (LBR) function for each MIP.

R-339 The Aggregation Node SHOULD support receiving AIS messages from an inferior Maintenance Level MEP(s) and send out an AIS message at the appropriate MIP level.

R-340 The Aggregation Node SHOULD support using a "Server MEP" function (defined in Section 5.3.1/Y.1731) to report failure of a Server layer and send out an AIS message at the next-superior Maintenance Level. This is required in network deployments that do not make use of the Spanning Tree Protocol.

R-341 The Aggregation Node SHOULD trigger the appropriate alarms for Loss of Continuity.

7.3.3.2 Short Intra-Carrier Maintenance Level

This maintenance level may be used for 'Ethernet bit-stream' services.

R-342 The Aggregation Node MUST support an inward-facing Maintenance association End Point (MEP) on a per NNI port and per S-VLAN basis.

R-343 The Aggregation Node MUST support initiating a Loopback Message (LBM) towards its peer MEPs and receiving the associated Loopback Reply (LBR), for the MEP on the NNI port.

R-344 The Aggregation Node MUST support receiving a Loopback Message (LBM) from its peer MEPs and initiating the associated Loopback Reply (LBR), for the MEP on the NNI port.

R-345 The Aggregation Node MUST support initiating a Link Trace Message (LTM) toward the multicast address and receiving the associated Link Trace Reply (LTR) messages, for the MEP on the NNI port.

R-346 The Aggregation Node MUST support receiving a Link Trace Message (LTM) from its peer MEPs and initiating the associated Link Trace Reply (LTR), for the MEP on the NNI port.

R-347 The Aggregation Node SHOULD support generating Continuity Check Messages (CCMs) towards its peer MEPs for the MEP on the NNI port.

R-348 The Aggregation Node MUST support turning off sending CCMs for the MEP on the NNI port, while keeping the associated MEP active.

R-349 The Aggregation Node SHOULD be able to be configured to assume continuity exists from a remote MEP while not receiving CCMs from this MEP.

R-350 The Aggregation Node SHOULD support a means to determine the MAC address of a remote MEP without relying on the reception of CCMs from this remote MEP. One possible way to accomplish the above is via the Multicast LBM.

R-351 The Aggregation Node MUST support receiving AIS messages on the MEP on the NNI port (at a so-called inferior Maintenance Level) and sending out an AIS message at the next-superior Maintenance Level across the NNI.

R-352 The Aggregation Node SHOULD trigger the appropriate alarms for Loss of Continuity.

7.3.3.3 Short Carrier Maintenance Level

In addition to the requirements specified in 7.3.3.2, Short Intra-Carrier Maintenance Level, the following requirements apply to the shortened Carrier maintenance level.

R-353 The Aggregation Node SHOULD be configurable to assume continuity exists from a remote MEP while not receiving CCMs from this MEP.

Note: Short Carrier Maintenance Level only works with single-tagged VLANs.

7.3.4 BNG requirements

The BNG can have Maintenance Points (MPs) at the Intra-Carrier, Carrier, and Customer levels.

7.3.4.1 Intra-Carrier Maintenance Level

R-354 The BNG MUST support an outward-facing Maintenance association End Point (MEP) on a per user-facing port and per S-VLAN basis.

R-355 The BNG MUST support initiating a Loopback Message (LBM) towards its peer MEPs and receiving the associated Loopback Reply (LBR), for the MEP(s) on the user-facing port.

R-356 The BNG MUST support receiving a Loopback Message (LBM) from its peer MEPs and initiating the associated Loopback Reply (LBR), for the MEP(s) on the user-facing port.

R-357 The BNG MUST support initiating a Link Trace Message (LTM) towards its peer MEPs and receiving the associated Link Trace Reply (LTR) messages, for the MEP(s) on the user-facing port.

R-358 The BNG MUST support receiving a Link Trace Message (LTM) from its peer MEPs and initiating the associated Link Trace Reply (LTR), for the MEP(s) on the user-facing port.

R-359 For business customers and/or premium customers requiring proactive monitoring, the BNG SHOULD support generating Continuity Check Messages (CCMs) towards its peer MEPs for the MEP(s) on the user-facing port.

R-360 The BNG MUST support turning off sending CCMs for the MEP(s) on the user-facing port, while keeping the associated MEP active.

R-361 The BNG SHOULD be configurable to assume continuity exists from a remote MEP while not receiving CCMs from this MEP.

R-362 The BNG MUST support receiving AIS messages on the MEP(s) on the user-facing port.

R-363 The BNG SHOULD trigger the appropriate alarms for Loss of Continuity.

7.3.4.2 Carrier Maintenance Level

In addition to the requirements specified in 7.3.4.1, Intra-Carrier Maintenance Level, the following requirement applies to the Carrier maintenance level.

R-364 For 1:1 VLANs, the BNG MUST support using a Multicast LBM towards its peer MEP.

Note: continuity with the MEP on the Access Node’s user port does not necessarily imply continuity to the RG.

7.3.4.3 Customer Maintenance Level

At the Customer Maintenance level of the BNG, support of R-358 (replying to LTMs) is not required.

At the Customer Maintenance level of the BNG, support of R-356 is modified slightly as follows:

R-365 The BNG MUST support receiving a *unicast or multicast* Loopback Message (LBM) from its peer MEPs and initiating the associated Loopback Reply (LBR), for the MEP(s) on the user-facing port.

At the Customer Maintenance level of the BNG, support of R-361 is clarified as follows:

R-366 The BNG SHOULD be able to be configured to assume continuity exists from a remote *RG* MEP while not receiving CCMs from this MEP.

In addition to the remaining requirements specified in 7.3.4.2, Carrier Maintenance Level, the following requirements apply to the Customer maintenance level.

R-367 The BNG MUST be able to be configured to discard all incoming LTMs on a per user-facing port, per S-VLAN and per C-VLAN basis.

There is a concern that users could intentionally or unintentionally launch an attack on a carrier's network using Ethernet OAM messages. The BNG must be able to protect against such events by discarding Ethernet OAM messages that exceed a specified rate.

R-368 The BNG MUST support rate limiting of received CFM Ethernet OAM messages arriving on a per user-facing port.

7.4 Interworking between Ethernet and ATM OAM

Testing end-to-end connectivity between a BNG and RG requires end-to-end visibility between both network elements. Using ATM between BNG and RG fulfills this requirement by providing end-to-end visibility in the data path. When part of the ATM is replaced by Ethernet, end-to-end visibility between BNG and RG is lost. Restoring the lost visibility can be performed by a data path or control path mechanism e.g. CLI, EMS, Layer 2 Control, or 802.1ag with an interworking function. All solutions require the presence of an approach-dependent interworking function which provides the translation of a trigger message to DSL port specific ATM OAM messaging.

The actual scope of functionality of I.610 within this document's context is very limited. It is expected that I.610 functionality will only have utility in the absence of tools based upon the emerging Y.1731 or IEEE 802.1ag standards and will merely exercise more of the NT than the DSL PHY. This requires an interim mechanism at the access node to remotely invoke I.610 Loopback (either CLI or control plane action).

R-369 The RG MUST support receiving an ATM Loopback, and initiating the reply.

R-370 The RG SHOULD support initiating an ATM Loopback, and receiving the reply.

R-371 The RG MUST support an ATM connection endpoint for all VCs.

R-372 The RG MUST provide a default CPID of all 1s (FFFF).

R-373 The AN MUST support an ATM connection endpoint.

R-374 The AN MUST support receiving an ATM Loopback, and initiating the reply.

R-375 The AN MUST be able trigger the following ATM Loopback test functions:

- Insertion of F5 e-t-e Loop Back cells per VC,
- Insertion of a number of LB cells,
- Receipt of all LB cells and check against the number of emitted cells.

There is a subset of AN functions, which can be initiated by the AN Management System request, that would be more conveniently initiated by the BNG in some networks. In other to implement such an approach, a communication mechanism is needed between the BNG and AN. There are several options, for example:

- A Management Channel, as described in Y.1731 (ETH-MCC).
- A Layer 2 Control mechanism as specified in TR-147.

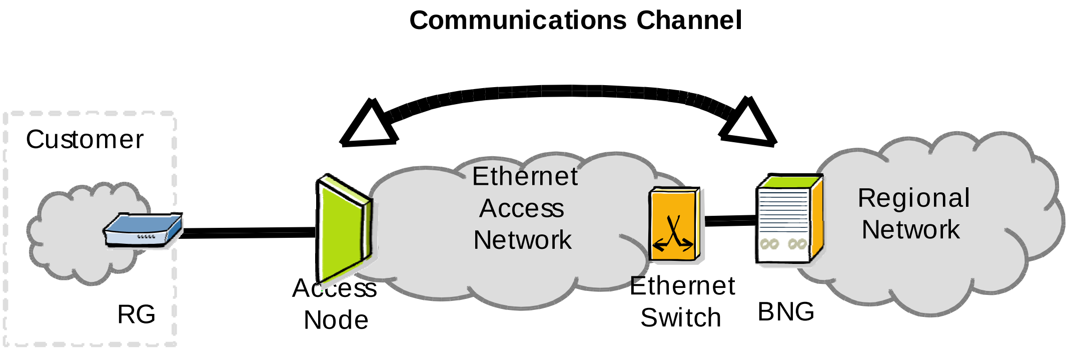


Figure 27 – Communication channel between BNG & AN

Messages on the selected communication channel need to support functions including the following:

- The ability for the BNG to check a Broadband line's status. The Circuit ID/ Agent ID can be used to identify the line.
- The ability for the BNG to trigger the AN to initiate an ATM loopback to an RG, on a specified VPI/VCI on a specific access line, when the RG does not support Ethernet OAM.
- The ability for the BNG to trigger the AN to initiate an ETH-LBM to an RG, at a specific access line.
- The ability for the AN to report the results to the BNG.

8. Network Management

8.1 Access Node Requirements

Access Node provisioning can be simplified by exploiting commonalities among groups of users. Such groups are characterized by identical or similar characteristics on the access line, in the forwarding processing to be performed by the Access Node, and in the association to VLANs in the aggregation network.

R-376 The Access Node MUST support configuring all access loops, or groups of access loops using pre-defined profiles.

> Example profiles include configurations for users connected to the same service or service provider. In the case of a single service for all users, profiles may be used to pre-configure all ports of an Access Node at installation time, i.e. to perform bulk provisioning.

R-377 The Access Node MUST support bulk pre-configuration including line-specific settings irrespective of whether the line-cards are present in the device.

R-378 The capability MUST exist to pre-configure access loops on the Access Node (i.e. support for bulk provisioning).

> As an example, the Access Node could be pre-configured with an ATM UBR profile for each port, terminated by the Access Node, and associated with a unique C-VLAN stacked on top of the same (shared) S-VLAN for the network-facing interface.

R-379 The access node MUST support a network management interface with the ability to perform configuration of logical or physical entities (e.g. via scripts or via a GUI) in order to support bulk provisioning. This requires the device’s interface to be able to provide an acknowledgement indicating that a specific configuration request has been completed.

8.2 BNG Requirements

The requirements described in this section add to the requirements already established by TR-092.

R-380 A Broadband Network Gateway SHOULD be able to globally auto-sense S-VLAN-tagged Ethernet frames. Such a Broadband Network Gateway MUST be able to automatically build the corresponding interface stack.

R-381 The Broadband Network Gateway MUST be able to globally auto-sense C-VLAN-tagged frames within a pre-provisioned or auto sensed S-VLAN. The Broadband Network Gateway MUST be able to automatically build the corresponding interface stack.

R-382 The dynamically created interfaces, even if becoming inactive, SHOULD NOT be deleted unless corresponding resources need to be reclaimed to satisfy another user session or until an inactivity timer elapses.

R-383 The Broadband Network Gateway MUST be able to mix statically configured VLAN interfaces with dynamically created VLAN interfaces on a given Ethernet port.

R-384 The Broadband Network Gateway SHOULD be able to define constraints on the range of VLAN values allowed for dynamic VLAN interface creation on a given Ethernet port.

R-385 The Broadband Network Gateway MUST be able to auto-sense PPP/PPPoE as well as IPoE frames. The Broadband Network Gateway MUST be able to automatically build the corresponding interface stack, as well as delete such dynamic interfaces when the corresponding user session is terminated.

Appendix A - PPPoE Vendor-Specific BBF Tags

All of the new PPPoE Vendor-Specific Tags introduced by the BBF will be formatted as the following:

```
+--------------+--------------+--------------+--------------+
| 0x0105 (Vendor-Specific)    |         TAG_LENGTH          |
+--------------+--------------+--------------+--------------+
| 0x00000DE9 (Vendor-id) Broadband Forum - IANA entry       |
+--------------+--------------+--------------+--------------+
|SUB-TAG-Number|  SUB-TAG-Len |         SUB-TAG-Value       |
+--------------+--------------+--------------+--------------+
|SUB-TAG-Value Cont’                                        ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Tag-name: Vendor-Specific

Tag-value: 0x0105

Tag-length: 2 octets unsigned number in network byte order indicating the length in octets of the TAG_VALUE, including the IANA vendor ID and the total sum of the SUB-TAG values (Tag, Length and Value).

Vendor-id: 4 octets unsigned number in which the high-order octet of the vendor id is 0x00 and the low-order 3 octets are the Broadband Forums SMI Network Management Private Enterprise Code of the Vendor in network byte order. 0x0DE9 or 3561 decimal.

Sub-Tag-Number: 1 octet unsigned number indicating the BBF Sub-Tag-Number assignment for the TAG.

Sub-Tag-Len: 1 octet unsigned number indicating the length of Sub-Tag-Value.

Sub-Tag-Value: Content of the value

8.3 PPPoE Tag - Circuit ID and Remote ID

The Access-Node MUST encode and send the Circuit ID and Remote ID as a TAG in PPPoE discovery Packet in the format described below:

```
+--------------+--------------+--------------+--------------+
| 0x01         |    LENGTH    |(63)Byte(Char)Circuit ID     |
+--------------+--------------+--------------+--------------+
| Circuit ID value (con’t)                                  |
+--------------+--------------+--------------+--------------+
| 0x02         |    LENGTH    |(63)Byte(Char)Agent Remote ID|
+--------------+--------------+--------------+--------------+
| Remote ID value (con’t)                                   |
+--------------+--------------+--------------+--------------+
```

8.4 PPPoE Tag - Broadband Line characteristics

The current set of values and the corresponding encoding is summarized below:

```
+--------------+--------------+--------------+--------------+--------------+
| 0x81         |   LENGTH     |(4)bytes(Bin)Actual Data Rate Upstream      |
+--------------+--------------+--------------+--------------+--------------+
| 0x82         |   LENGTH     |(4)bytes(Bin)Actual Data Rate Downstream    |
+--------------+--------------+--------------+--------------+--------------+
| 0x83         |   LENGTH     |(4)bytes(Bin)Minimum Data Rate Upstream     |
+--------------+--------------+--------------+--------------+--------------+
| 0x84         |   LENGTH     |(4)bytes(Bin)Minimum Data Rate Downstream   |
+--------------+--------------+--------------+--------------+--------------+
| 0x85         |   LENGTH     |(4)bytes(Bin)Attainable DataRate Upstream   |
+--------------+--------------+--------------+--------------+--------------+
| 0x86         |   LENGTH     |(4)bytes(Bin)Attainable DataRate Downstream|
+--------------+--------------+--------------+--------------+--------------+
| 0x87         |   LENGTH     |(4)bytes(Bin)Maximum Data Rate Upstream     |
+--------------+--------------+--------------+--------------+--------------+
| 0x88         |   LENGTH     |(4)bytes(Bin)Maximum Data Rate Downstream   |
+--------------+--------------+--------------+--------------+--------------+
| 0x89         |   LENGTH     |(4)bytes(Bin)Min DataRate Upstream in low   |
|              |              |             power state                    |
+--------------+--------------+--------------+--------------+--------------+
| 0x8A         |   LENGTH     |(4)bytes(Bin)Minimum Data Rate Downstream   |
|              |              |              in low power state            |
+--------------+--------------+--------------+--------------+--------------+
| 0x8B         |   LENGTH     |(4)bytes(Bin)Max Interleaving Delay         |
|              |              |             Upstream                       |
+--------------+--------------+--------------+--------------+--------------+
| 0x8C         |   LENGTH     |(4)bytes(Bin)Actual Interleaving Delay      |
|              |              |             Upstream                       |
+--------------+--------------+--------------+--------------+--------------+
| 0x8D         |   LENGTH     |(4)bytes(Bin)Maximum Interleaving Delay     |
|              |              |             Downstream                     |
+--------------+--------------+--------------+--------------+--------------+
| 0x8E         |   LENGTH     |(4)bytes(Bin)Actual Interleaving Delay      |
|              |              |             Downstream                     |
+--------------+--------------+--------------+--------------+--------------+
| 0x90         |   0x03       |(1)byte       +|(1)byte      + (1) byte     |
|              |              |  data link   + encaps 1     + encaps 2     |
+--------------+--------------+--------------+--------------+--------------+
| 0xFE         |   0x00       |           empty                            |
|              |              |                                            |
+--------------+--------------+--------------+--------------+--------------+
```

Appendix B - DHCP Vendor Specific Options to Support Access Line Characteristics

This Appendix describes how Access Line Characteristics are mapped to a DHCP message.

The Vendor-Specific Suboption
The Vendor-Specific suboption takes the following form:

```
0                   1                   2                   3
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Code      |    Length     |        Enterprise Number1     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                               |  DataLen1     |               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+               +
\                         Suboption Data1                       \
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Enterprise Number2
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  DataLen2     |             Suboption Data2                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
\                                                               \
.                                                               .
.                                                               .
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Code 9 for the DHCP suboption

Length >= 4
The one-byte Length field is the length of the data carried in the suboption, in bytes. The length includes the length of the first Enterprise Number; the minimum length is 4 bytes.

Enterprise Number1 “3561” the Broadband Forum IANA entry
The value is a four-byte integer in network byte-order.

DataLenN The length of the data associated with the Enterprise Number.

Suboption Data RFC4243 defines the Suboption as an opaque sequence of bytes allowing the Vendor to make use of the Suboptions to define its own specification.

Broadband Line Characteristics DHCP Vendor-Specific Suboption Data format

The sub option data format is shown below. The fields are transmitted from left to right. The Broadband Line Characteristics are to be transmitted in a single request as multiple Type/Length/Values (TLVs). The TLVs follow the same method as the currently defined PPPoE and Radius TLV. The exception to this is Circuit-Id and Remote-Id. Circuit-Id and Remote-Id are already defined and transmitted via Option 82 and do not need to be retransmitted.

```
 0                   1                   2
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
|     Type1     |    Length1    |  Value1 ...
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
|     Type2     |    Length2    |  Value2 ...
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
|     Type ...
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
```

TypeN The Type field is one octet. The following values are reserved for the type field and each is explained in a later section.

LengthN The one-byte Length field is the length of the data carried in the suboption, in bytes. The length is the length of the data carried in the Value.

ValueN The Value field is zero or more octets and contains information specific to the Attribute. The format and length of the Value field are determined by the Type and Length fields.

Broadband Line Characteristics DHCP Type Definitions

Broadband Line Characteristics DHCP Type Definition			
Type	Length	Value	Value Type
0x81	4	Actual data rate Upstream in kb/s.	32 bit binary value
0x82	4	Actual data rate Downstream in kb/s.	32 bit binary value
0x83	4	Minimum Data Rate Upstream in kb/s.	32 bit binary value
0x84	4	Minimum Data Rate Downstream in kb/s.	32 bit binary value
0x85	4	Attainable Data Rate Upstream in kb/s.	32 bit binary value
0x86	4	Attainable Data Rate Downstream in kb/s.	32 bit binary value
0x87	4	Maximum Data Rate Upstream in kb/s.	32 bit binary value
0x88	4	Maximum Data Rate Downstream in kb/s.	32 bit binary value
0x89	4	Minimum Data Rate Upstream in low power state in kb/s.	32 bit binary value
0x8A	4	Minimum Data Rate Downstream in low power state in kb/s.	32 bit binary value
0x8B	4	Maximum Interleaving Delay Upstream in millisec.	32 bit binary value
0x8C	4	Actual interleaving Delay Upstream in millisec.	32 bit binary value
0x8D	4	Maximum Interleaving Delay Downstream in millisec.	32 bit binary value
0x8E	4	Actual interleaving Delay Downstream in millisec.	32 bit binary value
0x90	3	Access-Loop-Encapsulation	24 bit binary value

End of Broadband Forum Technical Report TR-101