International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Q.3721

(09/2022)

SERIES Q: SWITCHING AND SIGNALLING, AND ASSOCIATED MEASUREMENTS AND TESTS

Signalling requirements and protocols for SDN – Resource control protocols

Procedures for a programming protocol independent packet processor switch-based virtual border network gateway

Recommendation ITU-T Q.3721

ITU-T Q-SERIES RECOMMENDATIONS

SWITCHING AND SIGNALLING, AND ASSOCIATED MEASUREMENTS AND TESTS

SIGNALLING IN THE INTERNATIONAL MANUAL SERVICE	Q.1–Q.3
INTERNATIONAL AUTOMATIC AND SEMI-AUTOMATIC WORKING	Q.4–Q.59
FUNCTIONS AND INFORMATION FLOWS FOR SERVICES IN THE ISDN	Q.60–Q.99
CLAUSES APPLICABLE TO ITU-T STANDARD SYSTEMS	Q.100–Q.119
SPECIFICATIONS OF SIGNALLING SYSTEMS No. 4, 5, 6, R1 AND R2	Q.120–Q.499
DIGITAL EXCHANGES	Q.500–Q.599
INTERWORKING OF SIGNALLING SYSTEMS	Q.600–Q.699
SPECIFICATIONS OF SIGNALLING SYSTEM No. 7	Q.700–Q.799
Q3 INTERFACE	Q.800–Q.849
DIGITAL SUBSCRIBER SIGNALLING SYSTEM No. 1	Q.850–Q.999
PUBLIC LAND MOBILE NETWORK	Q.1000–Q.1099
INTERWORKING WITH SATELLITE MOBILE SYSTEMS	Q.1100–Q.1199
INTELLIGENT NETWORK	Q.1200–Q.1699
SIGNALLING REQUIREMENTS AND PROTOCOLS FOR IMT-2000	Q.1700–Q.1799
SPECIFICATIONS OF SIGNALLING RELATED TO BEARER INDEPENDENT CALL CONTROL (BICC)	Q.1900–Q.1999
BROADBAND ISDN	Q.2000–Q.2999
SIGNALLING REQUIREMENTS AND PROTOCOLS FOR THE NGN	Q.3000–Q.3709
SIGNALLING REQUIREMENTS AND PROTOCOLS FOR SDN	Q.3710–Q.3899
Resource control protocols	**Q.3710–Q.3739**
Network signalling and signalling requirements for services	Q.3740–Q.3779
TESTING SPECIFICATIONS	Q.3900–Q.4099
PROTOCOLS AND SIGNALLING FOR PEER-TO-PEER COMMUNICATIONS	Q.4100–Q.4139
SIGNALLING REQUIREMENTS AND PROTOCOLS FOR IMT-2020	Q.5000–Q.5049
COMBATING COUNTERFEITING AND STOLEN ICT DEVICES	Q.5050–Q.5069

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

Recommendation ITU-T Q.3721

Procedures for a programming protocol independent packet processor switch-based virtual border network gateway

Summary

Recommendation ITU-T Q.3721 specifies the architecture, interfaces and procedures for a programming protocol independent packet processor switch-based virtual border network gateway.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T Q.3721	2022-09-29	11	11.1002/1000/15045

Keywords

Procedures, programming protocol independent packet processor, switch-based.

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

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

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

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

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

NOTE

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

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

INTELLECTUAL PROPERTY RIGHTS

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

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

Table of Contents

Recommendation ITU-T Q.3721

Procedures for a programming protocol independent packet processor switch-based virtual border network gateway

1 Scope

This Recommendation specifies the architecture, interfaces and procedures for a programming protocol independent packet processor (P4) switch-based virtual border network gateway (vBNG).

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T Q.3719] Recommendation ITU-T Q.3719 (2019), *Signalling requirements for the separation of control plane and user plane in a virtualized broadband network gateway (vBNG).*

[IETF RFC 6241] IETF RFC 6241 (2011), *Network configuration protocol (NETCONF).*

[IETF RFC 7348] IETF RFC 7348 (2014), *Virtual extensible local area network (VXLAN): A framework for overlaying virtualized layer 2 networks over layer 3 networks.*

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 broadband network gateway (BNG) [b-ITU-T Q.3315]: The access point to the provider's IP network for wireline broadband services.

3.1.2 virtual BNG [b-ITU-T Q.3715]: The virtual BNG is the broadband network gateway of which all features or some features are directly implemented as virtual network function(s) (VNF(s)) running on the network function virtualization infrastructure (NFVI). It is used to either augment or replace the existing traditional BNG.

3.2 Terms defined in this Recommendation

None.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AAA Authentication Authorization Accounting

ACL Access Control List

AN Access Node

ASIC Application Specific Integrated Circuit

BNG	Broadband Network Gateway
CFI	Control Forwarding Interface
CGN	Carrier-Grade Network address translation
CHAP	Challenge Handshake Authentication Protocol
D	Dropped
DHCP	Dynamic Host Configuration Protocol
DPDK	Data Plane Development Kit
DS	Domain Specific
F	Flow
FPGA	Field Programmable Gate Array
gMNI	gRPC Network Management Interface
gRPC	Google Remote Procedure Call
I	Intermediate
ID	Identifier
IDL	Interface Define Language
INT	In-band Network Telemetry
INT-MD	In-band Network Telemetry – embed Data
INT-MX	In-band Network Telemetry – embed instructions
INT-XD	In-band Network Telemetry – export Data
IOAM	*In-situ* Operations, Administration, and Maintenance
IP	Internet Protocol
LCP	Link Control Protocol
MD	Metadata
NFVI	Network Function Virtualization Infrastructure
P4	Programming Protocol independent Packet Processor
PADI	PPPoE Active Discovery Initiation
PADO	PPPoE Active Discovery Offer
PADR	PPPoE Active Discovery Request
PADT	PPPoE Active Discovery Termination
PPP	Point-to-Point Protocol
PPPoE	Point-to-Point Protocol over Ethernet
Q	Queue
QoS	Quality of Service
Rsvd	Reserved
RG	Residential Gateway
SR-IOV	Single Root Input/Output Virtualization
TLV	Type Length Value

vBNG	virtual Border Network Gateway
vBNG-CP	Virtual Border Network Gateway-Control Plane
vBNG-UP	Virtual Border Network Gateway-User Plane
VNF	Virtual Network Function
VNFM	Virtual Network Function Management
VxLAN-GPE	Virtual Extensible Local Area Network-Generic Protocol Extension
YANG	Yet Another Next Generation

5 Conventions

In this Recommendation:

The phrase "is required" indicates a requirement that must be strictly followed and from which no deviation is permitted, if conformity to this Recommendation is to be claimed.

The phrase "is recommended" indicates a requirement that is recommended but which is not absolutely required. Thus, this requirement need not be present to claim conformity.

The phrase "can optionally" indicates an optional requirement that is permissible, without implying any sense of being recommended. This term is not intended to imply that the vendor's implementation must provide the option, and the feature can be optionally enabled by the network operator or service provider. Rather, it means the vendor may optionally provide the feature and still claim conformity with this Recommendation.

6 Overview

Network function virtualization has emerged as the leading transformative technology that will allow service providers to move to a truly virtualized infrastructure. A broadband network gateway (BNG) is deployed at the edge of the metro area network of an operator and aggregates user sessions from the access network, providing access control by interworking with an authentication authorization accounting (AAA) server. BNG is also evolving into vBNG, which can be implemented in different types of hardware including dedicated equipment based on a network processor, application specific integrated circuit (ASIC) or [b-ITU-T X.86]. In order to improve forwarding efficiency and maintain flexibility, acceleration technologies can be used, e.g., data plane development kit (DPDK), single root input/output virtualization (SR-IOV) and ASIC, field programmable gate array (FPGA) acceleration card or P4 switch.

P4 is a high-level language for data-forwarding programming, whose design is based on protocol independence, object independence and re-configurability. P4 is a southbound protocol, although its scope is large. It can not only guide the data stream to be forwarded, but also program the data processing flow of the forwarding device such as switch through software programming to guide how to process packets.

Traditional vBNG is suitable for low-traffic scenarios. Although it can decouple software and hardware, it enhances development flexibility greatly and achieves the desired performance with the support of acceleration technologies, e.g., DPDK, SR-IOV. However, in the scenario of high-throughput complex service, vBNG faces many challenges. It requires more hardware investment, but the performance does not increase linearly with the rise in hardware costs.

In order to solve these problems, P4 switch-based vBNG is proposed. Through P4 programming, iterative development can be completed in a short time, and device independence can be realized, which avoids the long period of development and commissioning to physical hardware necessary in the traditional vBNG forwarding plane.

P4 programming can be applied to vBNG forwarding plane devices of different manufacturers to control data-forwarding behaviour, completely solving the problem that the network device hardware is bundled by a few manufacturers. It can also support novel features of the new protocol or fix existing vulnerabilities through software upgrading only, and remove redundant forwarding plane functions according to the application scenario, reduce potential vulnerabilities and operation complexity, as well as improving performance.

7 Architecture for a P4 switch-based vBNG

7.1 General architecture

With regard to P4 switch-based vBNG, the architecture design should follow the following guidelines.

– The basic bit rate available to each user should meet current business requirements.

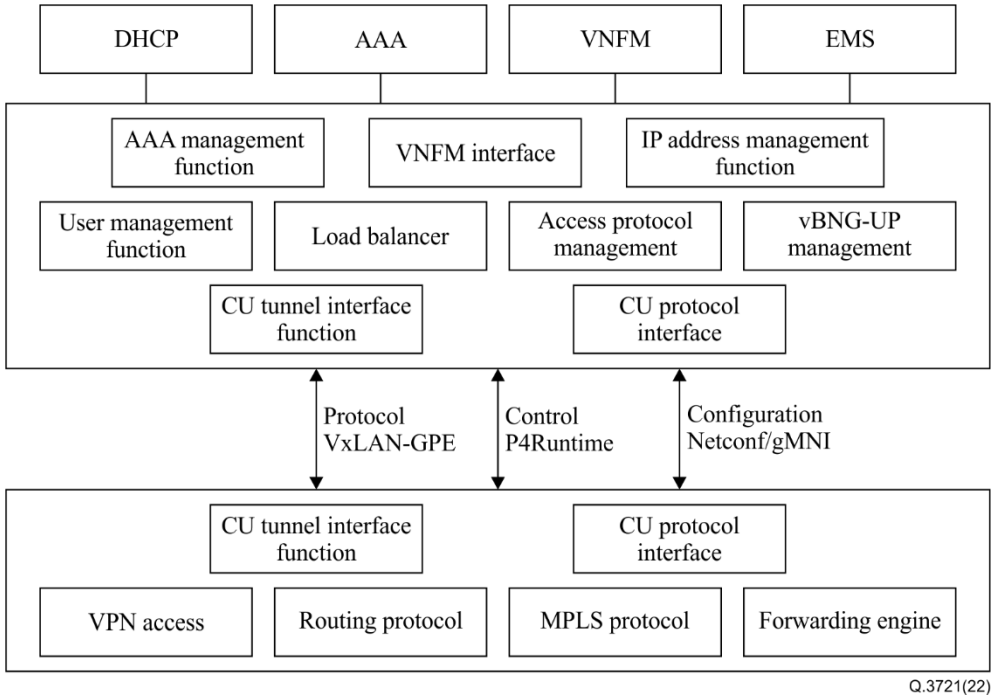
– The full virtualization of network functionalities should allow service providers to make a smooth transition to a cloud-based network, such as one with a fifth generation core.

– Carrier-grade network address translation (CGN) should be considered in the forwarding platform to deal with IPv4 address exhaustion.

– Some virtualized network functions (such as traffic detection function/deep packet inspection, CGN) should be implemented, so the signal gateway and BNG should be strengthened to achieve a unique virtualized platform, ensure ultra-low latency and improve performance.

– The implementation of network element equipment should be decoupled from hardware and software to eliminate dependence on proprietary hardware.

– The individual service requirements of each client should be required to have device-level granularity and in-depth analysis scalability.



gMNI: gRPC network management interface; VNFM: virtual network function management

Figure 7-1 – Architecture for P4 switch-based vBNG

This Recommendation focuses on the specification of the control and forwarding interface between the virtual border network gateway-control plane (vBNG-CP) and virtual border network gateway-user plane (vBNG-UP). The physical division of control and forwarding interfaces is the same as the separation of control plane and user plane in vBNG (specified in [ITU-T Q.3719), which is divided into interfaces for service, control and management. Logically, it also can be divided into five interfaces, including ones for service, control, management, maintenance and monitoring. The maintenance interface is required to be one for physical control, which shares the P4Runtime channel. The monitoring interface should be an optional logical interface, which can be carried by the user datagram protocol (UDP) tunnelling protocol.

– A *service interface* is used to transmit point-to-point protocol over Ethernet (PPPoE) and other protocol packets between the vBNG-CP and vBNG-UP by the virtual extensible local area network-generic protocol extension (VxLAN-GPE) tunnel.
– A *control interface* mainly includes two parts: one used by the vBNG-CP to deliver user forwarding entries (including user identification information and user policy information) to the vBNG-UP; and the other used by the vBNG-UP to report user-related information (e.g., statistics information) to the vBNG-CP. The control interface should be supported by P4Runtime.
– A *management interface* is used by the vBNG-CP to send configuration parameters to vBNG-UP, including quality of service (QoS) template, access control list (ACL) template, control channel and service channel configuration parameters, and other related parameters. The NETCONF protocol and the yet another next generation (YANG) model is used to deliver configuration parameters.
– A *maintenance interface* is used by the vBNG-CP to online edit and update the forwarding engine firmware in vBNG-UP, to real-time upgrade the control interface of the vBNG-UP and to implement the software defining the vBNG-UP. These should be implemented by P4Runtime.
– A *monitoring interface* is required to support visualization at the network packet-level, including data collection and reporting.

8 Interfaces for P4 switch-based vBNG

8.1 Service interface

The service interface is carried by VxLAN-GPE. The extension field needs to carry information (such as device number, port or virtual local area network) to identify the user identity to which the internal protocol message belongs. The vBNG-UP supports dynamic configuration of the VxLAN-GPE tunnel destination address and internal parameters, etc. The configuration can be implemented by the management interface. The vBNG-UP supports the VxLAN-GPE tunnel function, encapsulates user protocol packets (such as PPPoE) into VxLAN-GPE tunnel packets, and decapsulates user protocol packets encapsulated in the original packets by VxLAN-GPE.

8.2 Control interface

P4Runtime defines a set of standard interface define language (IDL) (proto) files for a Google remote procedure call (gRPC). The IDL abstractly defines the interfaces for adding, deleting, modifying and checking hardware entries, counter reading, packet-in/out, address notification and online upgrading. According to the specification of the overall architecture of vBNG, basic switching functions should be implemented in vBNG-UP, and user-related flow tables should be managed by the vBNG-CP, so P4Runtime is mainly used to control the user-side flow table.

Due to the diversification of P4 forwarding plane equipment, such as the coexistence of multiple vendors and multiple chips (FPGA, ASIC, etc.), a unified abstraction of the user-side forwarding table is required, so that the vBNG-CP can uniformly control different P4 forwarding planes. User

table attributes include virtual routing and forwarding, Internet protocol (IP), media access control, Session identifier (ID) (PPPoE) and rate limit. The vBNG-UP should support dynamic configuration of a user-side flow table through P4Runtime.

8.3 Management interface

The management interface is used by the vBNG-CP to manage the basic configuration attributes of vBNG-UP, such as templates for QoS and ACL. In addition, the specific parameters of the service and control channel are delivered to vBNG-UP through the management interface. Therefore, the management interface is necessary for the connection between the vBNG forwarding and control planes. It needs to be manually configured on the vBNG-UP; otherwise the vBNG-UP could actively detect the vBNG-CP. The channel is supported by the NETCONF protocol.

The vBNG-UP supports the configuration or active detection of the management interface. vBNG-UP supports the NETCONF protocol and the delivery of the YANG model. The specification of the YANG model refers to the OpenConfig standard.

The vBNG-UP supports the upgrade of the management interface.

8.4 Maintenance interface

The vBNG-UP supports online upgrade of the pipeline through P4Runtime. The scope of the online upgrade includes at least the user-side flow table, which serves as the maintenance interface of the control forwarding interface (CFI).

8.5 Monitoring interface

In-band network telemetry (INT) is a packet-level network monitoring method. It adds information to the message hop by hop through the switching devices the message passes through, so as to realize the collection of message-level path, delay and other information. The packet loss and congestion of the switching equipment are reported. The introduction of the P4 forwarding plane can support this feature through suitable P4 encoding. The P4 form of vBNG-UP supports the INT function that can send report messages to the monitoring server. The message format and content are determined by the telemetry report standard. The vBNG-UP based on the P4 switch supports dynamic configuration of the report message destination address and other parameters. The configuration can be implemented by the management interface.

9 Protocol requirement for P4 switch-based vBNG

9.1 P4Runtime protocol

The P4Runtime protocol supports the control of the user-side flow table between the vBNG-CP and vBNG-UP, reporting the traffic statistics information by the vBNG-UP, and updating the forwarding engine firmware.

The vBNG-CP manages the user information on the user side by controlling the user side flow table and distributes the access-related information in the form of the flow table after user authentication, as well as receiving the traffic statistics information reported by vBNG-UP. When the vBNG-CP delivers the user flow table, it carries user information, bandwidth limit information and billing information. The vBNG-CP periodically obtains the statistical value of the flow table through the read interface as the charging information after the user information is delivered. The user-side flow table model appears in Figure 9-1.

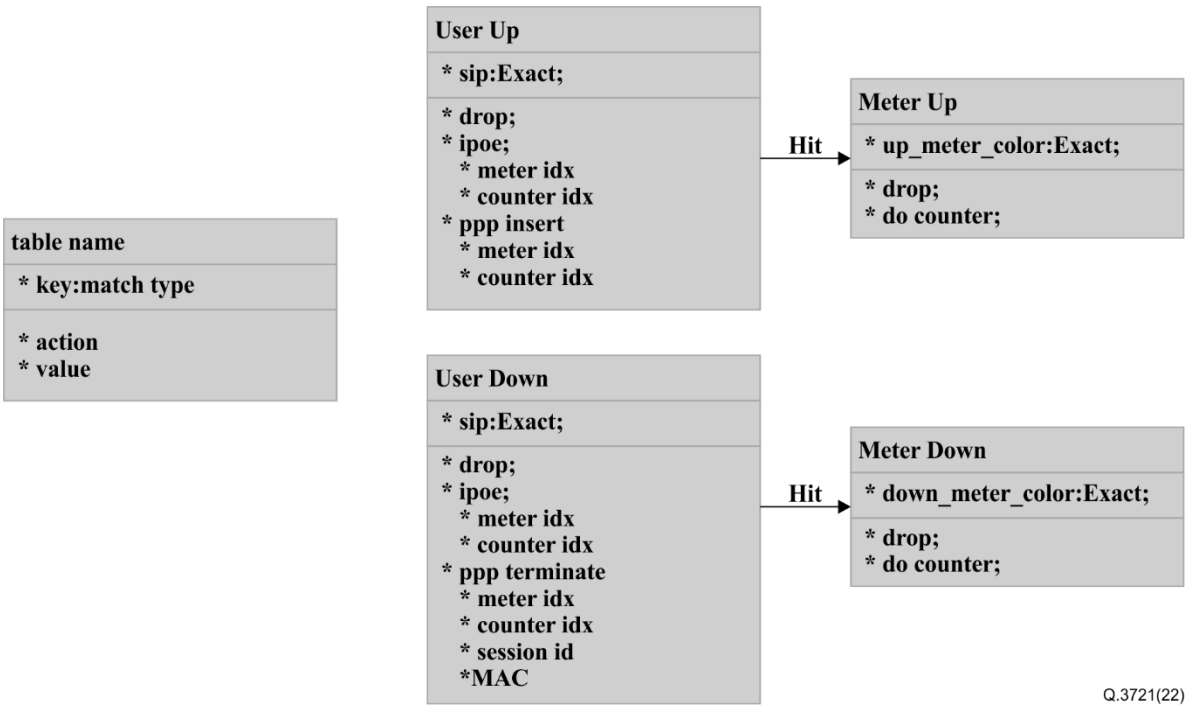


Figure 9-1 –User-side flow table model

table name：The user-side flow table name

match type：user up or user down

action：drop or ipoe or ppp insert/terminate

value：user information and counter table ID

The message format of vBNG-UP reporting traffic statistics appears in Figure 9-2.

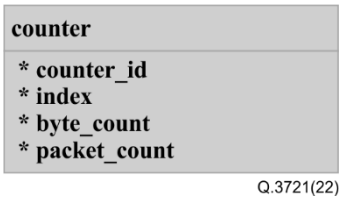


Figure 9-2 –Message format of vBNG-UP reporting traffic statistics

counter_id: counter table ID

index: the statistical index in the counter table

byte_count: the number of bytes counted

packet_count: the number of packets counted

The update message format of the vBNG-CP to the vBNG-UP forwarding engine firmware appears in Figure 9-3.

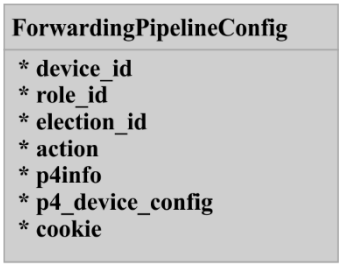


Figure 9-3 – Update message format of the vBNG-CP to the vBNG-UP forwarding engine firmware

device_id: Device (controller) ID

role_id: Controller role ID

election_id: number of elections,

action:

UNSPECIFIED = 0;

VERIFY = 1;

VERIFY_AND_SAVE = 2;

VERIFY_AND_COMMIT = 3;

COMMIT = 4;

RECONCILE_AND_COMMIT = 5;

p4info: description of firmware information

p4_device_config: firmware bin

9.2 Netconf protocol

The Netconf protocol is used for the configuration of parameters between the vBNG-CP and the vBNG-UP, such as the configuration of the virtual extensible local area network channel, interface, address segment, as well as templates for QoS and ACL. The implementation of the Netconf protocol should conform to the specification in [IETF RFC 6241].

9.3 Telemetry report protocol

It is recommended that the INT data collection between the vBNG-CP and vBNG-UP be carried by the telemetry report protocol, which is based on UDP and sends packet-level information collected by vBNG-UP to the vBNG-CP for visual analysis.

9.4 Tunnel protocol

The tunnel protocol should support VXLAN-GPE and its format should conform to the specification in [IETF RFC 7348].

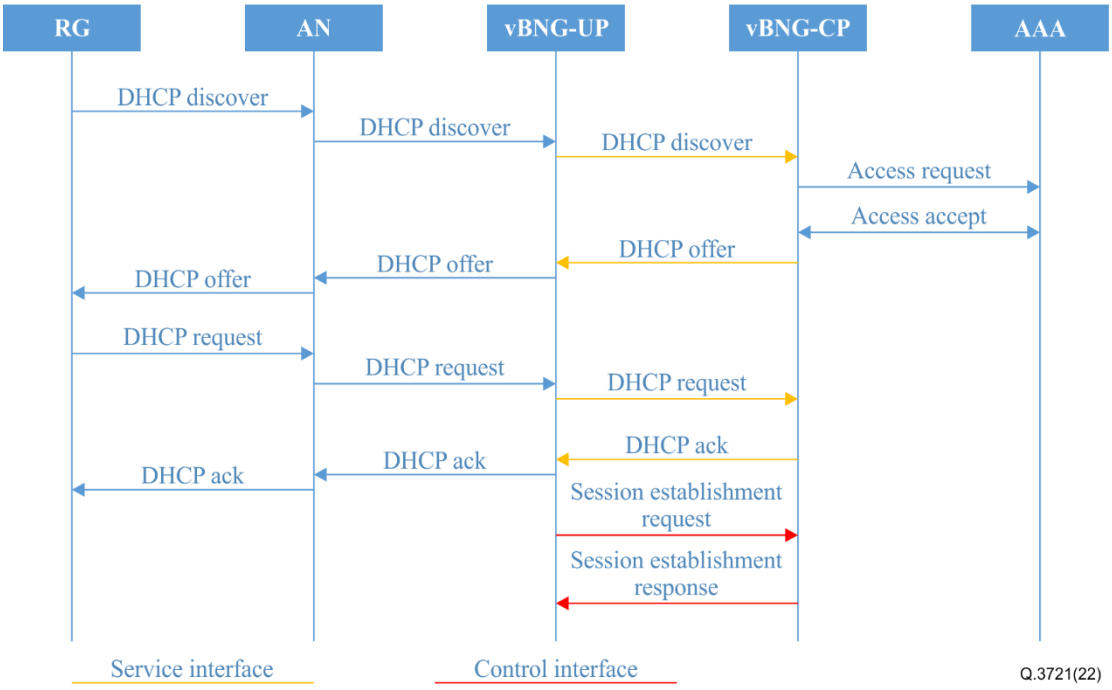
10 Procedures for P4 switch-based vBNG

Before starting the procedures, it is necessary to open the CFI configuration channel between the vBNG-CP and vBNG-UP. The configuration channel can be opened by directly configuring the vBNG-CP and vBNG-UP through an element management system. After opening, the configuration channel allows the vBNG-CP to configure the channels for protocol and control. The vBNG-CP can

realize the unified configuration of all vBNG-UPs it manages, including the configuration of the vBNG interface and QoS template.

10.1 IPoE DHCPv4 access procedure

The IPoE dynamic host configuration protocol version 4 (DHCPv4) access procedure appears in Figure 10-1.



AN: access node

Figure 10-1 – IPoE DHCPv4 access procedure

(1) The vBNG-UP sends a DHCP discovery message to the vBNG-CP, which triggers access request to authenticate the residential gateway (RG).

(2) The AAA successfully authenticates the RG and responds with an access accept message to the vBNG-CP.

(3) The vBNG-CP sends a DHCP offer message to the vBNG-UP through the service interface, and finally to the RG.

(4) The RG sends a DHCP request message that is forwarded to the vBNG-CP through the service interface through the vBNG-UP.

(5) The vBNG-CP sends a DHCP acknowledgement message to the RG to complete IPoE user access.

(6) The vBNG-CP obtains the IP address assigned to the RG, which can be assigned through a local address, AAA or DHCP server, and sends a session establishment request to the vBNG-UP.

(7) The vBNG-UP sends the session establishment response information to the vBNG-CP and informs the vBNG-CP that it is ready to forward the user IP address information.

10.2 IPoE DHCPv6 access procedure

See Figure 10-2.

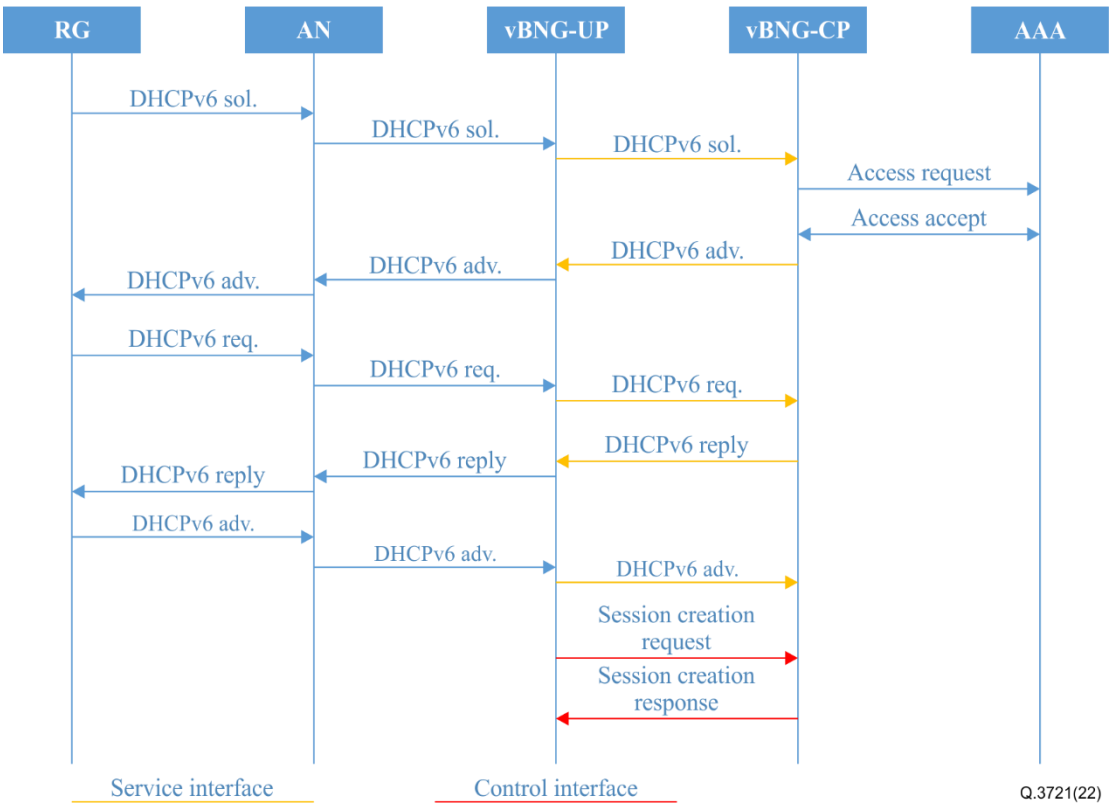


Figure 10-2 – IPoE DHCPv6 access procedure

(1) The vBNG-CP triggers an access request to the AAA for authentication of the RG.

(2) The AAA successfully authenticates the RG and sends an access reply message to the vBNG-CP.

(3) The vBNG-CP obtains the IP address assigned to the RG, which can be assigned through a server of the local address, AAA or DHCP type, and then sends a session request to the vBNG-UP.

(4) The vBNG-UP sends the session response message to the vBNG-CP, informing the vBNG-CP that it is ready to forward the user IP address information.

(5) The vBNG-CP sends a DHCPv6 advertisement message to the vBNG-UP through the service interface and finally to the RG.

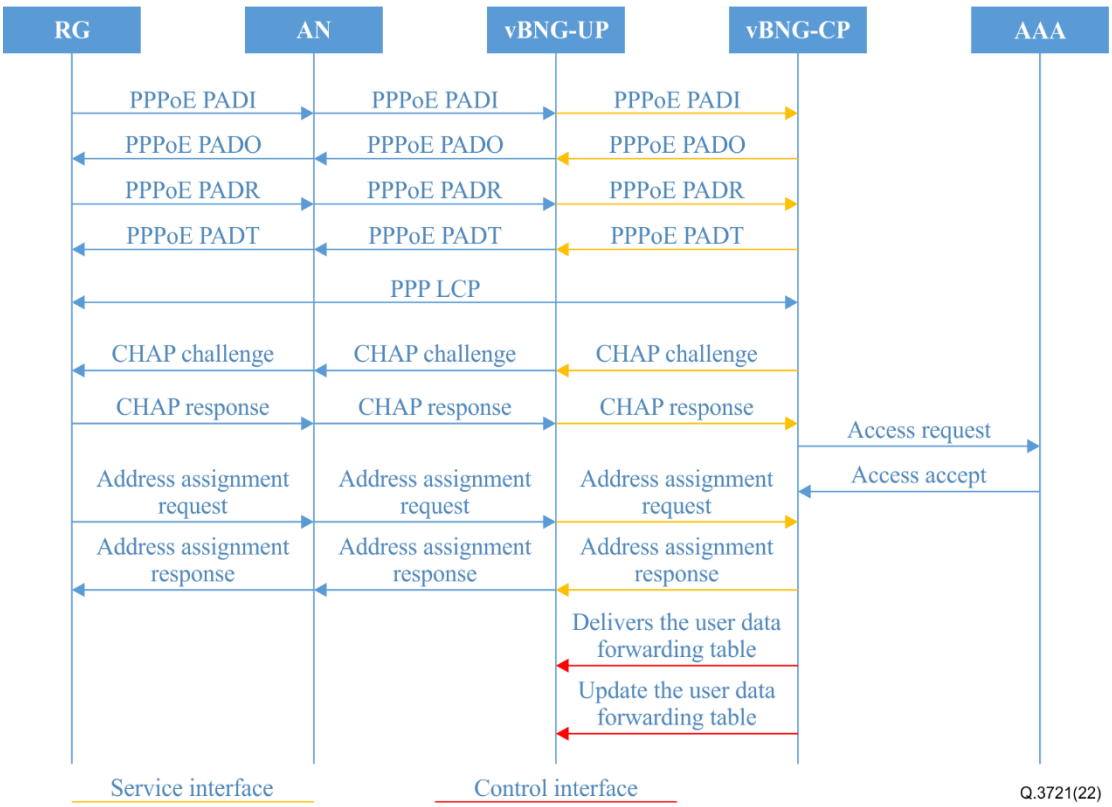
(6) The RG sends the DHCPv6 request message to the vBNG-UP through a customized session channel and to the vBNG-CP through a service interface.

(7) The vBNG-CP sends a DHCPv6 reply message to the RG.

(8) The vBNG-CP notifies the RG of the address information of the default gateway to complete IPoE user access.

10.3 PPPoE access procedure

See Figure 10-3.



CHAP: challenge handshake authentication protocol; LCP: Link Control Protocol; PADI: PPPoE active discovery initiation; PADO: PPPoE active discovery offer; PADR: PPPoE active discovery request; PADT: PPPoE active discovery termination; PPP: point-to-point protocol

Figure 10-3 – PPPoE access procedure

(1) The vBNG-UP receives the subscriber's PPPoE access message, and vBNG-UP sends it to the vBNG-CP through the protocol channel.

(2) The vBNG-CP processes the PPPoE state machine, triggers authentication by AAA and IP address allocation, and sends a response message to the vBNG-UP through the protocol channel. Then the vBNG-UP sends a response message back to the subscriber.

(3) The vBNG-CP creates a user forwarding table and synchronizes it to the user table database. Through the control channel, the vBNG-CP sends the user forwarding table and user address routing information to the vBNG-UP. The vBNG-UP runs the routing protocol, generates network forwarding table entries, and publishes the user routing to the network side, so that the downlink traffic directly returns to the vBNG-UP.

(4) The vBNG-UP collects user uplink and downlink traffic statistics, billing information and alarm information, and regularly sends them to the vBNG-CP through the control plane channel.

Annex A

Telemetry report format specification

(This annex forms an integral part of this Recommendation.)

This annex is derived from section 3 of [b-p4.org-TRF].

This annex specifies the packet format for telemetry reports.

A.1 Outer encapsulation

Telemetry reports are defined using a UDP-based encapsulation. Various outer encapsulations may be used to transport the UDP packets. Typically, this would simply be an Ethernet header, followed by an IPv4 or IPv6 header, followed by the UDP header. This specification does not preclude the use of different transport encapsulations.

The source IP address identifies the node that generates the telemetry report.

The destination IP address identifies a location in the distributed telemetry monitoring system that will receive the telemetry report.

In the case of IPv4, as is the case for any other IP packet, either the Don't Fragment (DF) bit must be set, or the IPv4 ID field must be set so that the value does not repeat within the maximum datagram lifetime for a given source address/destination address/protocol tuple.

A.1.1 UDP header (8 octets)

See Figure A.1.

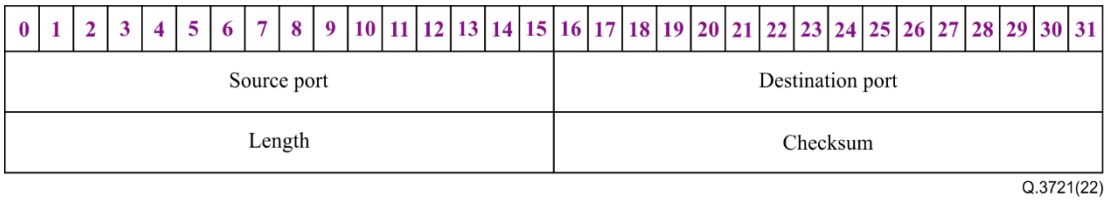


Figure A.1 – UDP header

The source port can optionally be used to carry flow entropy, e.g., based on a hash of the inner 5-tuple. Otherwise, it should be user configurable.

The destination port is user configurable. The expectation is that the same destination port value will be used for all telemetry reports in a particular deployment.

A.2 Telemetry report group header (version 2.0) (8 octets)

The telemetry report group header immediately follows the UDP header whose destination port identifies the contents as a telemetry report. This header contains the common fields in a telemetry report that optionally contains multiple coalesced individual reports, each corresponding to one data plane packet. There is at most one instance of the telemetry report group header in a packet. See Figure A.2.

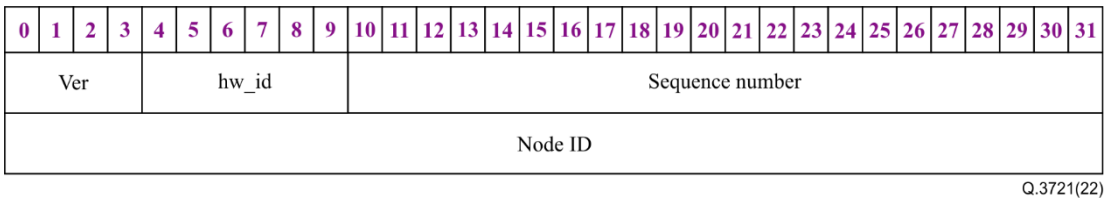


Figure A.2 – Telemetry report group header (version 2.0)

Ver (4 bits): Version

This specification defines **version 2**.

hw_id (6 bits): Hardware ID

Identifies the hardware subsystem within the node that generated this report. For example, in a chassis with multiple linecards, this could identify a specific linecard or a subsystem within a linecard. The hw_id is unique within the scope of a node ID.

Sequence number (22 bits): Sequence number

Reflects the sequence of reports from a specific combination of (node ID, hw_id) to a particular telemetry report destination. This can be used to detect loss of telemetry reports before they reach their intended destination.

Node ID (32 bits): Node ID

The unique ID of a node. This is generally administratively assigned. Node IDs must be unique within a management domain.

A.3 Individual report header (version 2.0) (4+ octets)

Each telemetry report packet contains one or more individual reports immediately following the telemetry report group header. Each report within the packet starts with the individual report header. The presence of multiple reports corresponding to multiple data plane packets, possibly from multiple flows, can be determined by comparing the report length in the individual report header with the length in the UDP header. See Figure A.3.

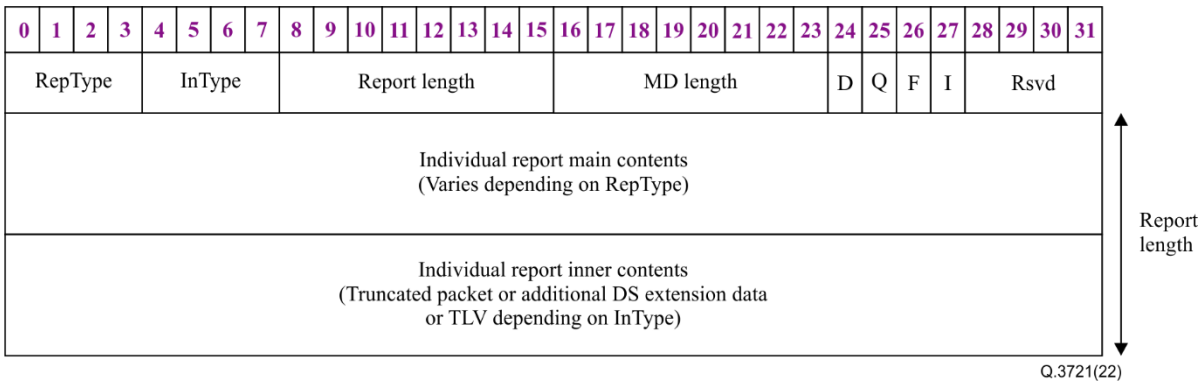


Figure A.3 – Individual report header (version 2.0)

RepType (4 bits): Report type

Type of the individual report:

- 0: Inner Only
- 1: INT
- 2: IOAM
- 3-15: Reserved.

InType (4 bits): Inner type

Type of data embedded after the *Individual Report Main Contents*:

- 0: None
- 1: TLV
- 2: Domain Specific Extension Data
- 3: Ethernet
- 4: IPv4
- 5: IPv6
- 6-15: Reserved.

Report length (8 bits): Report length

Indicates the length of the individual report header in a multiple of 4-byte words, including the *Individual Report Main Contents* and *Individual Report Inner Contents*, but excluding the length of the first 4-byte word (RepType, InType, Report Length, MD Length, D, Q, F, I, Rsvd).

For *RepType* codepoint 1 *INT*, the *Report Length* includes the lengths of *RepMdBits, Domain Specific ID, DSMdBits, DSMdstatus, Variable Optional Baseline Metadata* and *Variable Optional Domain Specific Metadata*.

The *Report Length* value 0xFF is a special value that indicates a length greater than or equal to 0xFF, extending to the end of the UDP payload, i.e., there are no subsequent individual reports in this telemetry report.

MD length (8 bits): Metadata length

Indicates the length of metadata (MD) included in this report in a multiple of 4-byte words. This may help the telemetry monitoring system determine where the *Individual Report Inner Contents* begins. Note that this does not include the length of the fixed portion of the *Individual Report Main Contents.*

For *RepType* codepoint 1 INT, this includes the length of the *Variable Optional Baseline Metadata* and *Variable Optional Domain Specific Metadata* in 4-byte words.

D (1 bit): Dropped

Indicates that at least one packet matching a watchlist was dropped.

Q (1 bit): Congested queue association

Indicates the presence of congestion on a monitored queue.

F (1 bit): Tracked flow association

Indicates that this telemetry report is for a tracked flow, i.e., the packet matched a watchlist somewhere (in the case of in-band network telemetry – embed data (INT-MD), in-band network telemetry – embed instructions (INT-MX) or *in-situ* operations, administration, and maintenance (IOAM)) or locally (in the case of in-band network telemetry – export data (INT-XD). The report might include INT-MD or IOAM metadata in the truncated packet. Other telemetry reports are likely to be received for the same tracked flow, from the same node and (in case of drop reports, INT-MX, INT-XD or path changes) from other nodes.

I (1 bit): Intermediate report

Indicates that a transit node sent this intermediate report for INT-MD.

Rsvd (4 bit): Reserved

Should be set to zero upon transmission and ignored upon reception.

Individual report main contents

The metadata that comprises this report, along with associated fields that assist in processing the metadata. The format varies depending on *RepType*.

When the *RepType* value is *Inner Only*, then the individual report main contents are empty. *MD Length* should be set to zero upon transmission and ignored upon reception.

The *INT* individual report main contents format was derived with INT 2.0/2.1 in mind, but it may be used with other INT versions as well. It is possible that other *RepType* codepoints and corresponding individual report main contents formats may be defined for future versions of INT.

The *IOAM* individual report main contents format is for further study.

Truncated packet

Layer 2/layer 3/ESP/layer 4 header of the IP packet for flow details. The presence of this field is indicated by *InType* codepoint 3, 4, or 5, which identifies the type of header at the beginning of the truncated packet. The length of the truncated packet can be determined as *Report Length* – ((fixed length of *Individual Report Main Contents*) + *MD Length*), unless the *Report Length* value is 0xFF.

Additional DS extension data

Additional domain specific (DS) extension data, whose format can be determined from the *Domain Specific* ID specified in the *Individual Report Main Contents*. For *RepType* codepoint 1 *INT*, this is additional domain specific data that is not associated with *DSMdBits*. The presence of this field is indicated by *InType* codepoint 2.

TLV

Type length value (TLV) format. Multiple TLV formatted data. The presence of this field is indicated by *InType* codepoint 1.

A.3.1 Individual Report Main Contents for RepType 1 (INT) (8+ octets)

See Figure A.4.

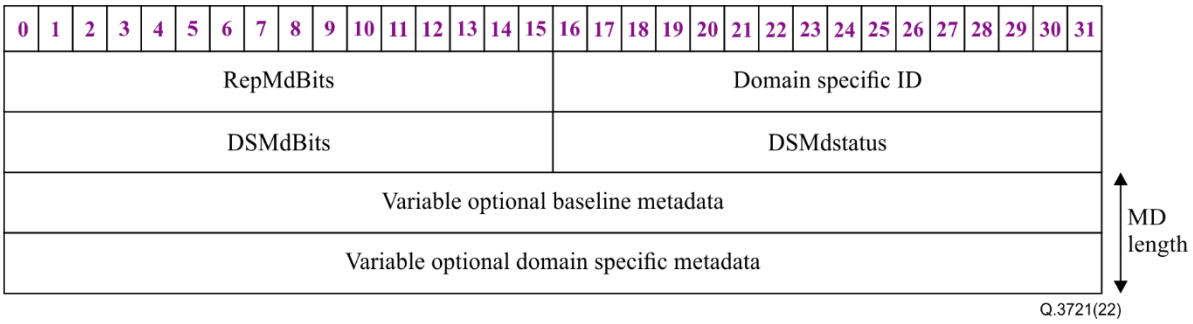


Figure A.4 – Individual report main contents for RepType 1 (INT)

RepMdBits (16 bit): Report metadata bits

Bitmap that indicates which optional baseline metadata is present in the telemetry report header. Each bit represents 4 octets of optional metadata, except for bits 4, 5 and 6 which represent 8 octets of optional metadata.

- bit 0 (MSB): Reserved
- bit 1: Level 1 Ingress Interface ID (16 bits) and Egress Interface ID (16 bit)
- bit 2: Hop Latency
- bit 3: Queue ID (8 bit) + Queue Occupancy (24 bit)
- bit 4: Ingress Timestamp (64 bit)

- bit 5: Egress Timestamp (64 bit)
- bit 6: Level 2 Ingress Interface ID (32 bit) + Egress Interface ID (32 bit)
- bit 7: Egress Port TX Utilization
- bit 8: Buffer ID (8 bits) + Buffer Occupancy (24 bit)
- bit 9-14: Reserved
- bit 15: Queue ID (8 bit) + Drop Reason (8 bit) + Padding (16 bit).

This specification defines the following metadata.

Drop reason

An enumeration that indicates the reason why a packet was dropped, e.g., as described in [b-GitHub].

See [b-p4.org-INT] for definitions of the remaining metadata.

Domain specific ID (16 bit)

The unique ID of the INT domain.

The *Domain Specific ID* value 0x0000 is the default, known to all nodes. For this value, all *DSMdBits* are treated as reserved. Operators can assign values in the range 0x0001 to 0xFFFF.

DSMdBits (16 bit): Domain specific Md bits

Bitmap that indicates which optional domain specific metadata is present in the telemetry report header. Each bit represents 4 octets or a multiple of 4 octets of domain specific optional metadata.

When using INT-MD or INT-MX, if the *Domain Specific ID* does not match any domain ID known to this node, then the node may either:

- set the telemetry report *DSMdBits* field to zero and rederive the telemetry report *MD Length* from *RepMdBits*; or
- not send any of its own metadata to the monitoring systems, doing any of the following:
  - not generating any Telemetry Report,
  - clear *RepMdBits* and *MD Length* as well as *DSMdBits* (this only makes sense for INT-MD) or
  - use *RepType* value *Inner Only* (this only makes sense for INT-MD).

DSMdstatus (16 bit): Domain specific Md status

Indicates the domain specific metadata status.

Variable optional baseline metadata

The metadata corresponding to RepMdBits, 4 octets for each bit, except 8 octets for bits 4, 5 and 6.

If a node receives an INT-MX or INT-MD packet with an instruction bitmap that requests one or more metadata values that are not available or reserved, then the node must ensure that the corresponding bit(s) in the telemetry report *RepMdBits* that specify the unavailable metadata are not set. The telemetry report *MD Length* must be derived based on the adjusted *RepMdBits* (and *DSMdBits*) values.

Variable optional domain specific metadata

The metadata corresponding to *DSMdBits*, 4 octets or a multiple of 4 octets for each bit.

If a node receives an INT-MX or INT-MD packet with a *DS Instruction* that requests one or more metadata values that are not available or reserved, then the node must ensure that the corresponding bit(s) in the telemetry report *DSMdBits* that specify the unavailable metadata are not set. The

telemetry report *MD Length* must be derived based on the adjusted *DSMdBits* (and *RepMdBits*) values.

A.3.2 Individual report inner contents for InType 1 (TLV) (4+ octets)

One or more TLVs, each following the format defined in this clause. The presence of multiple TLVs can be determined by comparing the *TLVLength* in the first TLV with the *Report Length* in the individual report header. See Figure A.5.

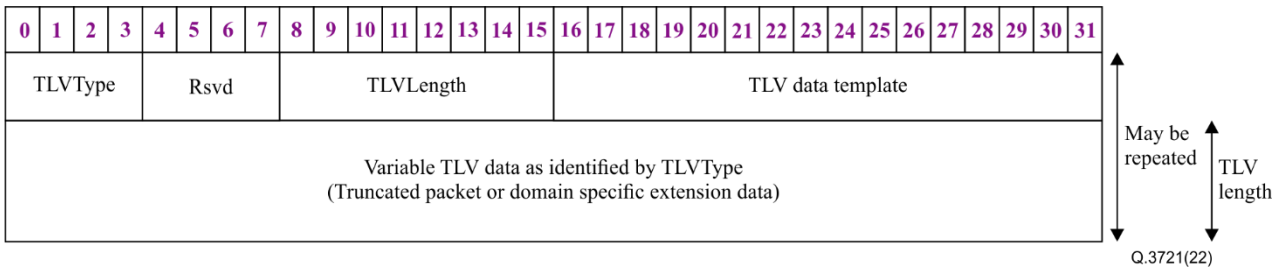


Figure A.5 – Individual report inner contents for InType 1 (TLV)

TLVType (4 bits): TLV data type

- 0: Domain Specific Extension Data
- 1: Ethernet
- 2: IPv4
- 3: IPv6
- 4-15: Reserved

Rsvd (4 bits) – Reserved

Should be set to zero upon transmission and ignored upon reception.

TLVLength (8 bits)

Indicates the length, in 4-byte words, of variable *TLV Data* as identified by *TLVType*. Note that this does not include the length of the first 4-byte word (*TLVType, Rsvd, TLVLength, TLV Data Template*).

TLV Data Template (16 bits)

Specifies the format of the *Variable TLV Data*. A non-zero *TLV Data Template* value specifies the template for *TLVType* codepoint of *Domain Specific Extension Data*. For *TLVType* code-points *Ethernet*, *IPv4*, and *IPv6*, the *TLV Data Template* value should be zero upon transmission and ignored upon reception.

Variable TLV data

Variable length data based upon *TLVType*. The following two fields are defined in this version.

Truncated packet

Layer 2/layer 3/ESP/layer 4 header of the IP packet for flow details. The presence of this field is indicated by *TLVType* codepoint 1, 2, or 3, which identifies the type of header at the beginning of the truncated packet.

Domain specific extension data

Domain specific extension data, whose format can be determined from the *Domain Specific ID* specified in the *Individual Report Main Contents* and the *TLV Data Template*. The presence of this field is indicated by *TLVType* codepoint 0.

For *RepType* codepoint 1 *INT*, this is additional domain specific data that is not associated with *DSMdBits*.

A.4 Embedded telemetry metadata in stacked reports

There may still be further telemetry metadata embedded within a truncated packet fragment. For example, this is typically the case when there is stacked telemetry metadata from hops prior to the node generating the report. The telemetry metadata will typically be encoded using a defined data plane format such as INT-MD or IOAM.

A node generating a telemetry report with stacked telemetry metadata may include its local telemetry metadata in any of the following:

- the embedded telemetry metadata in a truncated packet fragment;
- the stacked telemetry metadata in domain specific extension data;
- the *Individual Report Main Contents* in the same individual report header that contains the stacked telemetry metadata from previous hops, in either a truncated packet fragment or in domain specific extension data; or
- the *Individual Report Main Contents* in a separate report from the stacked telemetry metadata from previous hops – note that in this case the ingress timestamp (if present) will be the same in both reports.

If the *Tracked Flow Association (F)* bit is set to 0, then there will not be any embedded telemetry metadata in the report.

If the *Tracked Flow Association (F)* bit is set to 1, there may or may not be any embedded telemetry metadata in the report.

Bibliography

[b-ITU-T Q.3315] Recommendation ITU-T Q.3315 (2015), *Signalling requirements for flexible network service combination on broadband network gateway*.

[b-ITU-T Q.3715] Recommendation ITU-T Q.3715 (2018), *Signalling requirements for dynamic bandwidth adjustment on demand on broadband network gateway implemented by software-defined networking technologies.*

[b-ITU-T X.86] Recommendation ITU-T X.86 (2001), *Ethernet over LAPS.*

[b-GitHub] GitHub (2022). *p4lang/switch*. San Francisco, CA: Github. Available [2022-11-28] at: github.com/p4lang/switch.

[b-p4.org-INT] p4.org Applications Working Group (2020). *In-band network telemetry (INT) dataplane specification*, version 2.1. Available [viewed 2022-11-29] at: https://p4.org/p4-spec/docs/INT_v2_1.pdf

[b-p4.org-TRF] p4.org Applications Working Group (2020). *Telemetry report format specification*, version 2.0. Available [viewed 2022-11-29] at: https://p4.org/p4-spec/docs/telemetry_report_v2_0.pdf

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	**Switching and signalling, and associated measurements and tests**
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2023