ITUPublications
Recommendations

International Telecommunication Union
Standardization Sector

Supplement

ITU-T G Suppl. 79 (12/2023)

SERIES G: Transmission systems and media, digital systems and networks

Supplements to ITU-T G-series Recommendations

Latency control and deterministic capability over a PON system

ITU-T G-SERIES RECOMMENDATIONS

Transmission systems and media, digital systems and networks

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100-G.199
GENERAL CHARACTERISTICS COMMON TO ALL ANALOGUE CARRIER-TRANSMISSION SYSTEMS	G.200-G.299
INDIVIDUAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON METALLIC LINES	G.300-G.399
GENERAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON RADIO-RELAY OR SATELLITE LINKS AND INTERCONNECTION WITH METALLIC LINES	G.400-G.449
COORDINATION OF RADIOTELEPHONY AND LINE TELEPHONY	G.450-G.499
TRANSMISSION MEDIA AND OPTICAL SYSTEMS CHARACTERISTICS	G.600-G.699
DIGITAL TERMINAL EQUIPMENTS	G.700-G.799
DIGITAL NETWORKS	G.800-G.899
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900-G.999
MULTIMEDIA QUALITY OF SERVICE AND PERFORMANCE – GENERIC AND USER-RELATED ASPECTS	G.1000-G.1999
TRANSMISSION MEDIA CHARACTERISTICS	G.6000-G.6999
DATA OVER TRANSPORT – GENERIC ASPECTS	G.7000-G.7999
PACKET OVER TRANSPORT ASPECTS	G.8000-G.8999
ACCESS NETWORKS	G.9000-G.9999

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

Supplement 79 to ITU-T G-series Recommendations

Latency control and deterministic capability over a PON system

Summary

Supplement 79 to ITU-T G-series Recommendations describes the latency control and deterministic capability of passive optical network (PON) systems. It reviews feasible technologies of latency control and optimization technologies over a PON and discusses possible future expansion on a PON system including use cases and requirements, technologies for latency control and deterministic improvement, and possible extensions in the future.

History *

Edition	Recommendation	Approval	Study Group	Unique ID
1.0	ITU-T G Suppl. 79	2023-12-01	15	11.1002/1000/15852

Keywords

Latency, PON.

* To access the Recommendation, type the URL https://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID.

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

This is an informative ITU-T publication. Mandatory provisions, such as those found in ITU-T Recommendations, are outside the scope of this publication. This publication should only be referenced bibliographically in ITU-T Recommendations.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this publication 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 standards development process.

As of the date of approval of this publication, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this publication. 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

Supplement 79 to ITU-T G-series Recommendations

Latency control and deterministic capability over a PON system

1 Scope

This Supplement collects the use cases requiring low latency and deterministic capability over a passive optical network (PON) system. The reference functional models with technologies enabling latency control and supporting deterministic capability are analysed.

2 References

[ITU-T G.989.3]	Recommendation ITU-T G.989.3 (2021), *40-Gigabit-capable passive optical networks (NG-PON2): Transmission convergence layer specification*.
[ITU-T G.9804.2]	Recommendation ITU-T G.9804.2 (2021), *Higher speed passive optical networks – Common transmission convergence layer specification*.
[ITU-T G.Sup.71]	ITU-T G.sup.71 (2023), *Optical line termination capabilities for supporting cooperative dynamic bandwidth assignment*.
[ETSI GR F5G 002]	ETSI GR F5G 002 V1.1.1 (2021), *Fifth Generation Fixed Network (F5G); F5G Use Cases Release #1*.
[ETSI GS F5G 003]	ETSI GS F5G 003 V1.1.1 (2021), *Fifth Generation Fixed Network (F5G); F5G Technology Landscape*.
[IEEE 802.1]	IEEE 802.1 (n.d), *Time-Sensitive Networking (TSN) Task Group*.
[Christodoulopoulos 2023]	Christodoulopoulos, K., Bidkar, S., Pfeiffer, T., and Bonk, R. (2023), *Deterministically Scheduled PON for Industrial Applications*, Optical Fiber Communication Conference (OFC).
[Zhang 2022]	Zhang, D., Luo, Y., and Jin, J. (2022), *Highspeed 50 Gb/s Passive Optical Network (50G-PON) Applications in Industrial Networks*.

3 Abbreviations and acronyms

This Supplement uses the following abbreviations and acronyms:

BWmap	BandWidth Map
CBR	Constant Bit Rate
CO DBA	Cooperative DBA
DBA	Dynamic Bandwidth Assignment
DBRu	Dynamic Bandwidth Report, upstream
FEC	Forward Error Correction
HQ	High-Quality
IIoT	Industrial Internet of Things
M2M	Machine to Machine
ODN	Optical Distribution Network
OLT	Optical Line Terminal

OMCI	ONU Management and Control Interface
ONU	Optical Network Unit
OT	Operation Technology
PLC	Programmable Logic Controller
PLOAM	Physical Layer Operations, Administration and Maintenance
PMD	Physical Medium Dependent
PON	Passive Optical Network
PSBd	Downstream Physical Synchronization Block
PSBu	upstream Physical Synchronization Block
PTP	Precision Time Protocol
SNI	Service Node Interface
TC	Transmission Convergence
T-CONT	Transmission Container
TM	Traffic Management
TML	Transmission Medium Layer
TSN	Time Sensitive Networks
XGEM	10-Gigabit-capable PON Encapsulation Method

4 Conventions

4.1 Determinism capability in a PON context

In a PON system, determinism capability provides predictable and guaranteed data paths with bounds on latency, jitter, and packet loss, through behaviour control actions including scheduling, processing, and forwarding.

5 Introduction

Figure 5-1 shows a reference functional model for the delivery of traffic between devices separated by intermediate networks. Device B (e.g., a client) is shown to be connected via a PON-based access network. Device A (e.g., a server) is shown to be at some remote location deeper in the network. It introduces a new reference point RefTML (transmission medium layer) at the access node optical line terminal (OLT) side and the optical network unit (ONU) side, between the PON physical medium dependent (PMD) and transmission convergence (TC) functions and the service-related functions in resp. OLT and ONU. The service mapping into 10-Gigabit-capable PON encapsulation method (XGEM) frames is done in the transmission medium layer (TML) function as per the PON TC recommendations (see e.g., clause 9.4 in [ITU-T G.9804.2]).

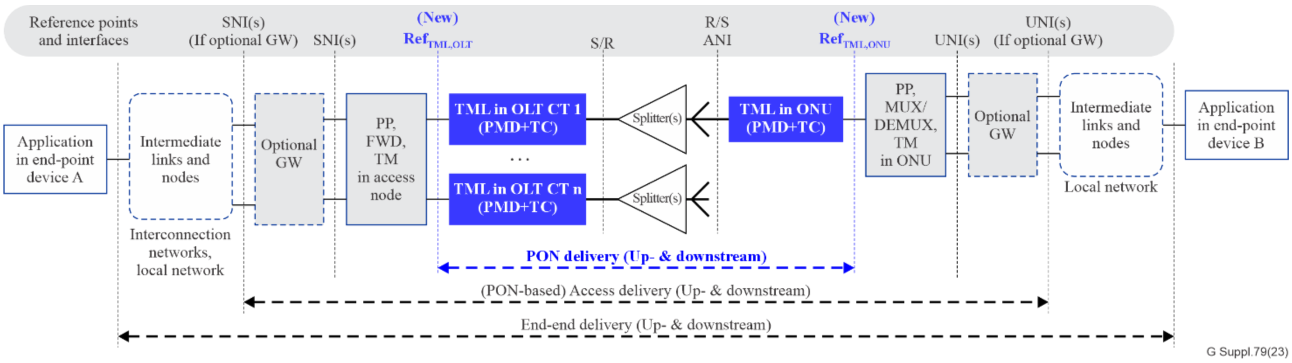

Figure 5-1 – Reference functional model between a PON-connected device and a remote device

Figures 5-2a and 5-2b show the reference model for the case of (e.g., user-to-user) communication between devices A and B when both are connected via a PON-based access network (resp. leg A and leg B).

Both devices could either be connected to different access nodes or the same access nodes.

When on different access nodes the end-to-end connectivity between A and B runs through both service node interfaces (SNIs) across the interconnection network (Figure 5-2a). When on the same access node, the end-to-end connectivity can either run through both SNIs (e.g., when a service needs to pass through an edge node) or run through an internal connectivity point inside the common access node (Figure 5-2b).

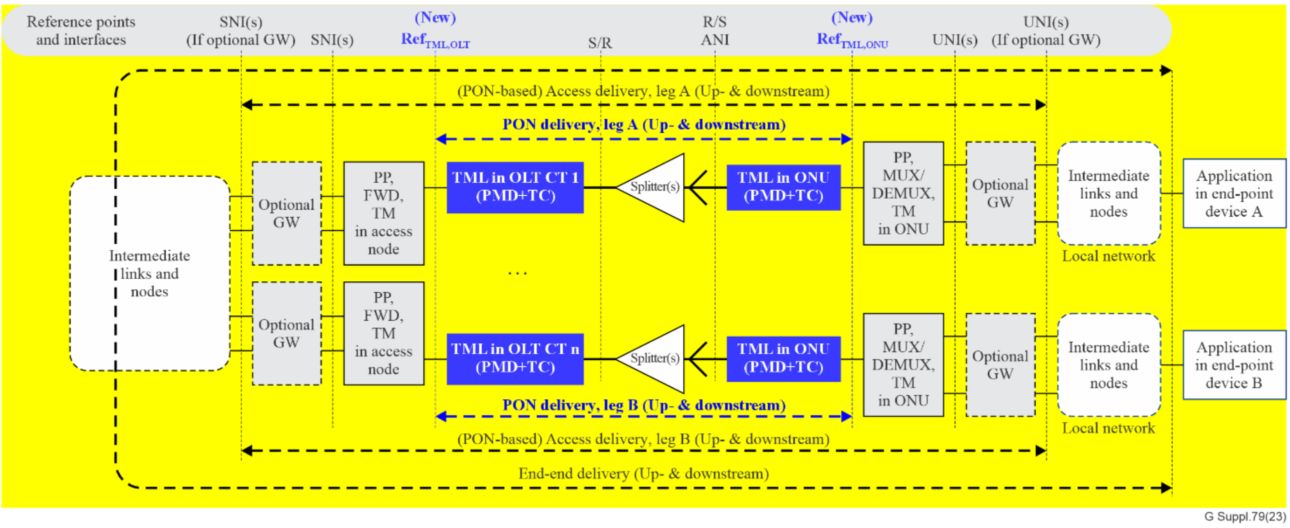

Figure 5-2a – Reference functional model between two PON-connected devices interconnected through the SNIs

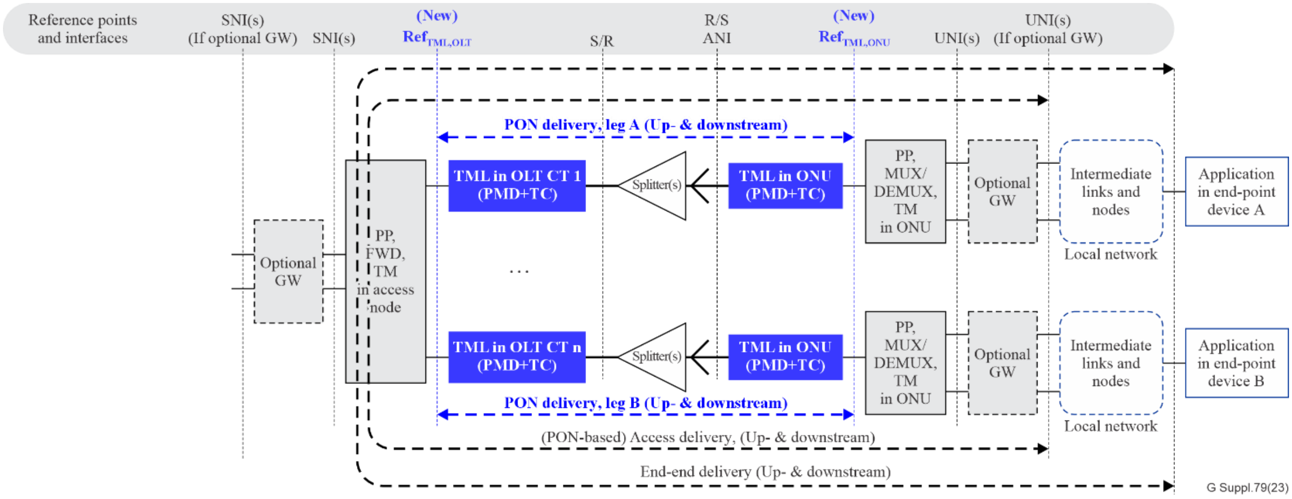

Figure 5-2b – Reference functional model between PON-connected devices interconnected inside the access node

The performance of the end-to-end delivery of traffic depends on many factors along the end-to-end path. This Supplement focuses on those PON-related mechanisms that influence the performance. Unless stated otherwise, the parameters mentioned in the Supplement refer to the delivery over the PON which comprises the OLT TML (PMD and TC layers), the optical distribution network (ODN), and the ONU TML (PMD and TC layers) (indicated in blue in Figure 5-1, Figure 5-2a and Figure 5-2b).

The OLT and ONU devices contain additional functions such as packet processing (PP), packet forwarding (FWD) in OLT or multiplexing/demultiplexing (MUX/DEMUX) in ONU, and traffic management (TM) such as queueing and scheduling. There can also optionally be a gateway function (GW) to e.g. shape the ingress and egress traffic to accommodate application-specific time patterns. The gateway function can be implemented as a stand-alone or as part of the OLT or ONU node. The focus of this Supplement is on the PON related functions (TML) at OLT and ONU side, the other functions (indicated in grey in Figure 5-1, Figure 5-2a, and Figure 5-2b) are not the focus of this Supplement.

The Supplement does not exclude any use case with additional functionality at the ONU side (e.g., single family residential deployments with integrated residential gateway, multi-dwelling deployments, mobile x-haul, business deployments, etc.).

6 Use case and requirements

6.1 Use case 1 – Manufacturing industrial PON

6.1.1 Description

Recent technologies such as the industrial Internet of things (IIoT), industrial cloud computing and analytics, machine-to-machine (M2M) communication, and augmented/virtual reality can be integrated into production facilities and operations, driving conventional industrial networks into modern and automatic industry 4.0 networks, realizing intelligent decision-making and manufacturing control.

Due to the variety of different industrial machines and equipment and the diversity of industrial protocols, a flexible and integrated network solution is required. As an advanced passive optical fibre-based communication technology, PON is the major upgrade direction for manufacturing industrial networks, which not only has the basic communication advantages but also supports cloud-based connectivity between fundamental services and enables real-time data transmission between various interfaces from various intra-plant machines and devices in the manufacturing industry.

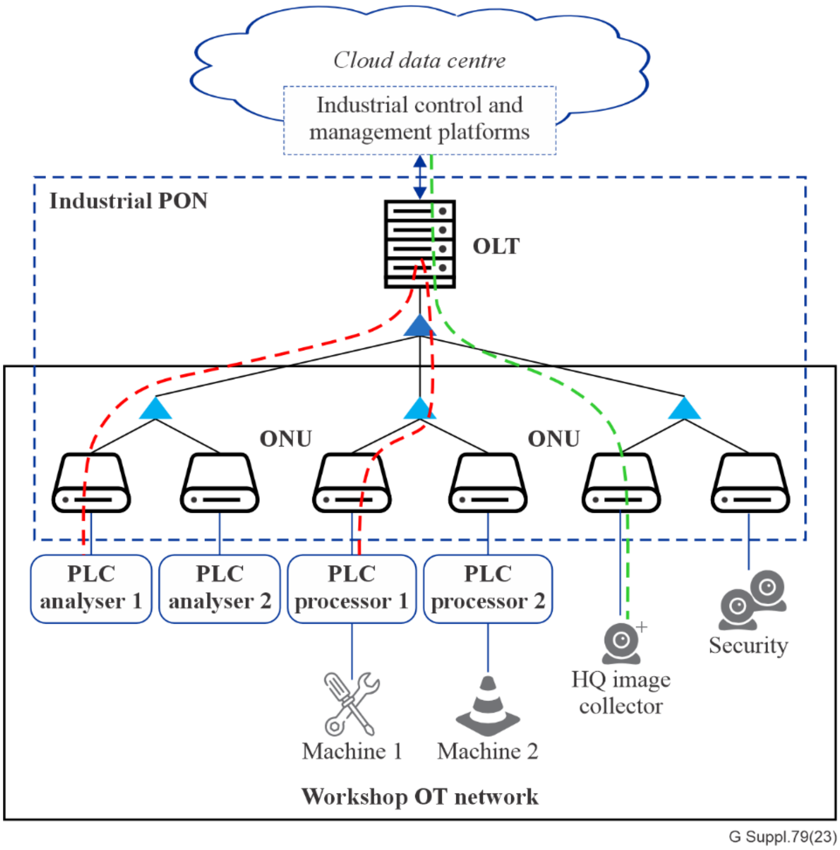

Figure 6-1 – The illustration of a typical industrial PON system

The illustration of a typical industrial PON system is given in Figure 6-1.

The basic production manufacturing system is operating in the workshop operational technology (OT) network. The traditional programmable logic controller (PLC) is supported to provide general manufacturing functions such as data analysis and control data processing. The high-quality (HQ) image collector is used to collect the image information of the manufacturing process and the production for intelligent management and control.

The industrial PON provides optical transmission for north-south intelligent management and control traffic, and low-latency interconnections between intra-plant machines and devices with various industrial physical interfaces. With the help of the broadband optical access network, the cloud platform can provide integrated and unified control and management for the global manufacturing industry system.

The manufacturing industrial PON can significantly enhance industrial production efficiency with intelligent system control and operation while retaining the capabilities to support basic manufacturing industrial applications.

6.1.2 Expectation of latency/jitter control of PON system

As illustrated in Figure 6-1, the interconnection between the PLC analyser and PLC processor indicates the traffic type which demands ultra-low latency and jitter for real-time smart production manufacturing, while the communication between HQ image collector and industrial control and management platform indicates the traffic type which demands high bandwidth but a relaxed latency and jitter.

The intelligent manufacturing process requires a certain level of latency and jitter. Different types of applications in the manufacturing industrial network require different latency and jitter performance. Examples are listed in the following:

1) For synchronous and asynchronous traffic: the latency/jitter can be less than 500 µs [Zhang 2022] and [ETSI GS F5G 003].

2) For network control traffic: the latency/jitter can be 10 ms to 1 s level [Zhang 2022] and [ETSI GR F5G 002].

To provide deterministic performance when forwarding time-sensitive traffic, the underlying industrial PON should be able to identify the time-sensitive traffic and provide a low latency forwarding path in the PON system, with the satisfied latency/jitter requirements.

6.2 Use case 2 – PON system in context of a TSN network

A time sensitive network (TSN) consists of TSN ethernet switches with features described by a set of IEEE TSN standards (see [IEEE 802.1]). A PON system can be used to connect end-side devices or end-side TSN networks to a LAN TSN network, as shown in Figure 6-2 with resp. ONU A and B. Time-sensitive flows can run in the east-west direction (e.g., between device A on ONU A and device B on ONU B) and the north-south direction (e.g., between the industrial application server and device C on ONU C). By using QoS awareness the same infrastructure is also capable of transporting non-time sensitive flows locally or to the wider network (e.g., a terminal on ONU D connected to the Internet via a WAN gateway).

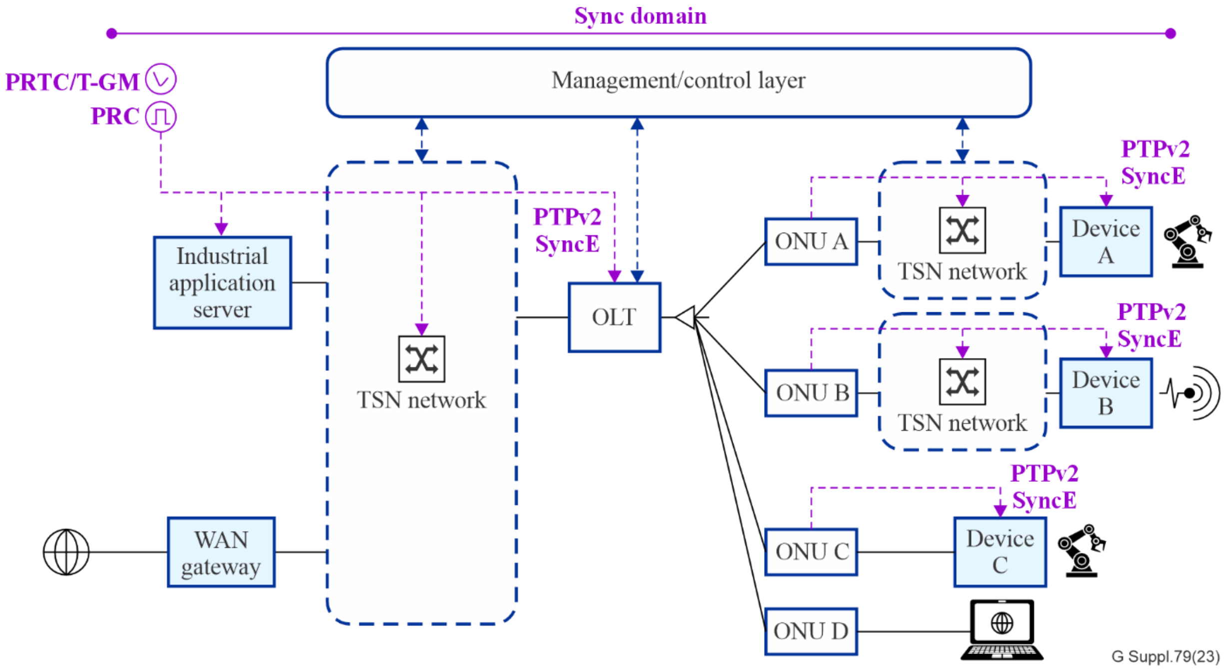

Figure 6-2 – PON system in the context of the TSN network

In case information about time-sensitive industrial flows is available to the PON system from the traffic source, management/control layer, or from the TSN network (e.g., via management or control messages), this information can be used for the OLT's internal traffic management of these flows. Such information should contain a timing description of the flows (e.g., periodicity and phase alignment, latency and jitter requirements) and a means to differentiate the flows in the PON system (e.g., entry and exit ports on the OLT and ONUs). Regarding the PON TML in the upstream direction, the dynamic bandwidth assignment (DBA) can be made more deterministic by taking the per-TSN flow (per-transmission container (T-CONT) information into account in its choice of burst allocations (timing, rate and size of allocations). This can be considered as a variant of cooperative DBA (CO DBA) as described in [ITU-T G.Sup.71].

A common time of day reference needs to be distributed to the OLT for actions that require accurate timing. TSN uses precision time protocol (PTP) and SyncE to distribute time and frequency synchronization in the industrial network, they can also be used to bring synchronization to the OLT, and the PON system can further distribute synchronization (via the ONUs) to the end-side networks or devices.

An informative example of adding timing awareness to DBA can be found in [Christodoulopoulos 2023].

7 Mechanisms for latency control and deterministic improvement

7.1 Access segment

The latency (and variations on latency) experienced by packets during their end-to-end data transport is determined by the treatment of the packets in the different segments of the network. The sections in clause 7 concentrate on the treatment in the PON-based access segment and more specifically on the PON system segment, corresponding to resp. "access delivery" and "PON delivery" in Figures 5-1, 5-2a and 5-2b. The access segment consists of packet devices (OLT and optional GW, ONU and optional GW), PON layer-specific functions in OLT and ONU, and a fibre-based ODN. The **access segment latency** per packet then equals the PON system latency (PON transmission medium layer in OLT and ONU + fibre propagation over the ODN) + internal latencies of devices performing packet switching and processing. Both upstream and downstream directions are considered.

The **PON ODN** is the most straightforward source of latency as it gives the same propagation delay for all packets in function of the distance, irrespective of the actual load in the fibre. There are two dependencies:

– The first and obvious dependency is the total fibre distance per ONU (patch cabling, feeder fibre, drop fibre).
  - The distance between OLT and ONU is not necessarily fixed, e.g., protection switching from a working path to a protecting path can lead to a decrease or increase in the total fibre length. But in all cases, the maximum distance of the PON class must be respected, so there is an upper limit.
  - During ranging all the ONUs will be time equalized in the upstream direction to the distance of the furthest possible ONU. The effect of time equalization of ONUs on the upstream latency is described in clause 7.2.2.

– The second dependency is the group velocity of the wavelength used, but those variations are of the order of less than 1 ns per km within a wavelength band and hence irrelevant for end-end latencies of tens or hundreds of µs for data traffic. However, note that the effect of such variations on the synchronization accuracy is relevant and is described in the PON TC recommendations. This is not elaborated further in this Supplement.

The **packet devices** in the segment are the optional ingress gateway, the ONU, the OLT and the optional egress gateway. The internal architecture of the packet devices consists of interfaces, (de)multiplexers, buffers and forwarders. Each packet switching device adds latency due to internal forwarding, buffering, traffic management processes, and serialization/deserialization of ethernet frames at interfaces. In congestion points there are mechanisms to manage the order of transmission between packets of the same or different traffic classes, influencing the latency. The internal mechanisms of packet devices are out of the scope of the ITU PON recommendations.

Additionally, there are **PON-specific** contributions to latency (due to the operation of the TML), which are handled in more detail in clauses 7.2 and 7.3.

7.2 PON-specific parameters influencing latency

Several parameters and mechanisms of the TML have an impact on latency. Some parameters are defined in the PON standards as being configurable, others are implementation-specific (e.g., depending on the processing performance).

7.2.1 Influence on downstream PON system latency

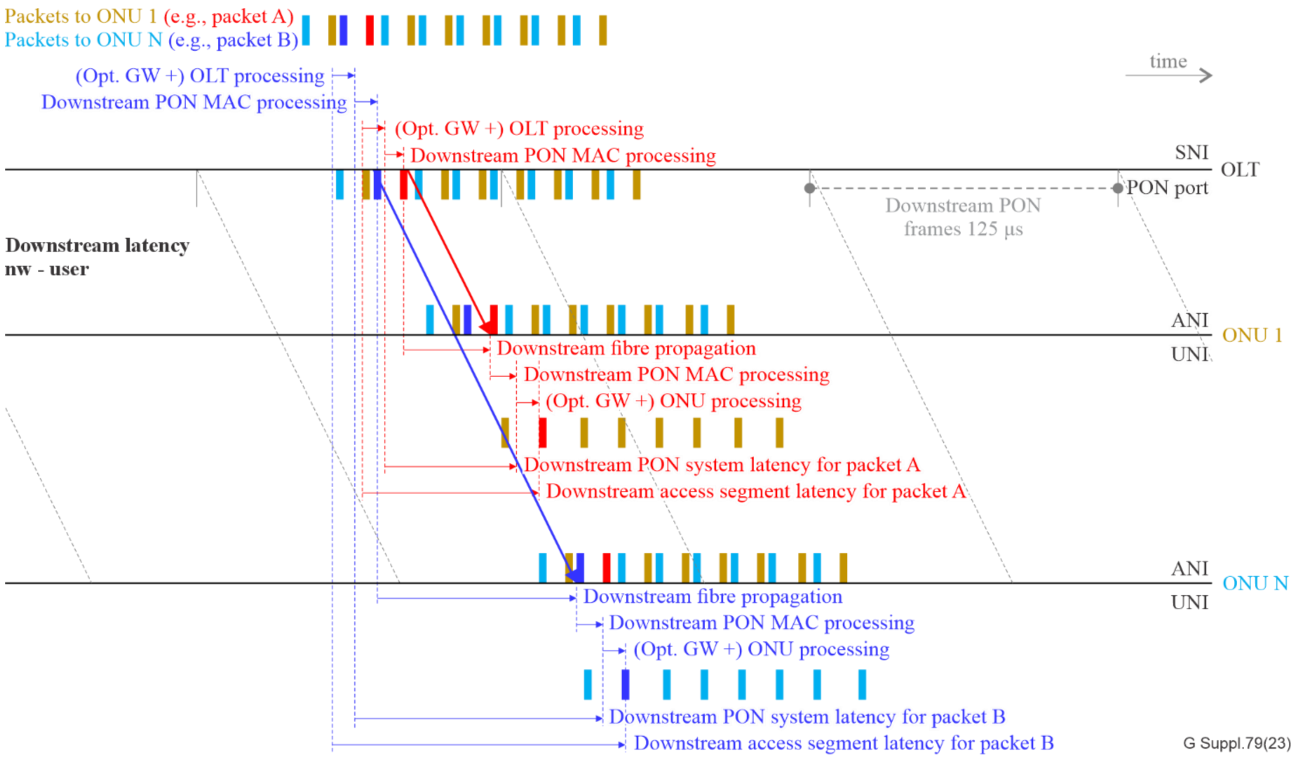

Figure 7-1 – Break-down of per-packet downstream latency

- **OLT**: PON TC layer actions include encapsulation, forward error correction (FEC) encoding, and frame structure (insertion of downstream physical synchronization block (PSBd), multiplexing of ONU management and control interface (OMCI) payload and the physical layer operations, administration and maintenance (PLOAM) payload for ONUs). They will directly impact the service latency but are minor (order of µs) compared to the whole PON system latency.
- **ONU**: PON TC layer actions include FEC decoding and decapsulation, and packet defragmentation. They will directly impact the service latency but are minor (order of µs) compared to the whole PON system latency.

7.2.2 Influence on upstream PON system latency

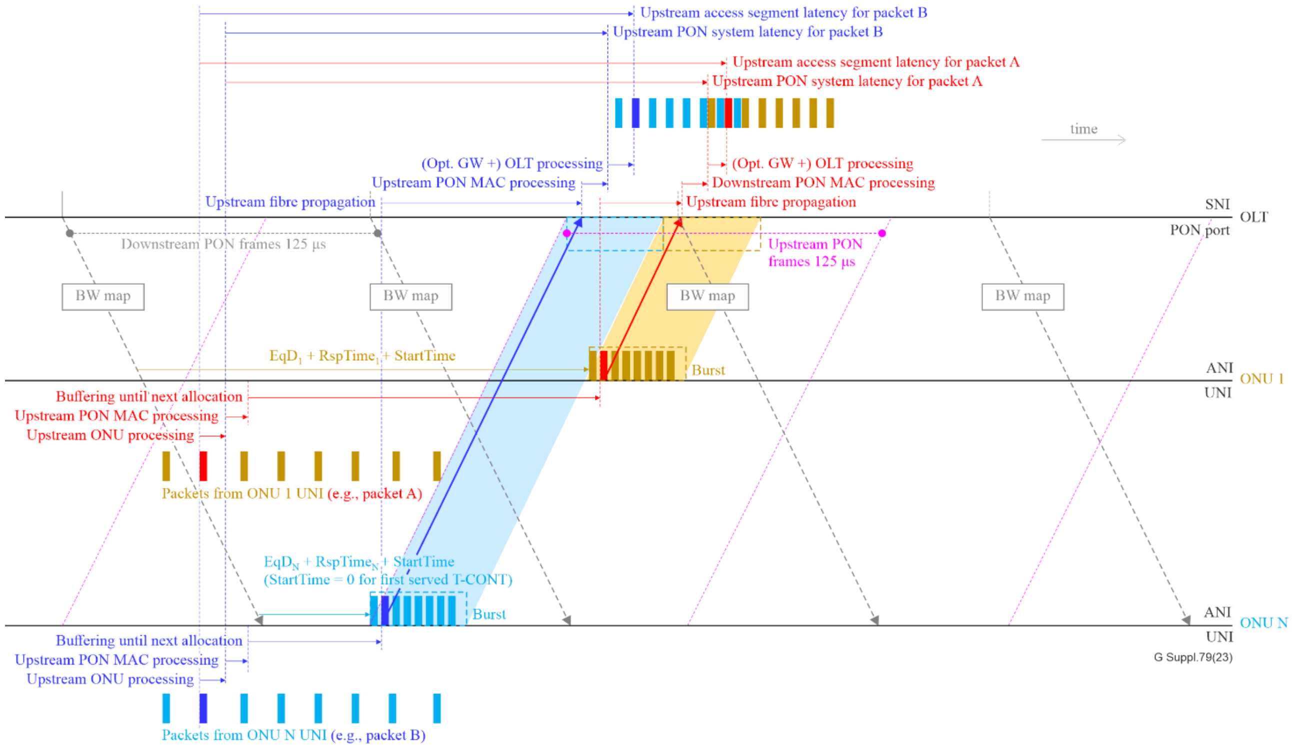

Figure 7-2 – Break-down of per-packet upstream latency

- **ONU**: The ONU must interpret grants for its T-CONTs in the bandwidth map (BWmaps) and generate the corresponding bursts. This includes reception and parsing of the received BWmaps, and the PON TC layer processing actions for generating the bursts which include encapsulation, FEC encoding, construction of the burst structure (insertion of upstream physical synchronization block (PSBu), of OMCI bytes or bursts, of dynamic bandwidth report, upstream (DBRu) bytes, of PLOAM bytes).

  The performance of this processing is stipulated in the PON recommendations by the ONU response time, namely ONUs must be able to generate a burst between 34 and 36 µs after the arrival time of the downstream frame boundary (which is followed by the bandwidth map) at the ONU. This processing latency is fixed and is accounted for when generating a burst at the correct time grant allocated by the BWmap. Therefore, this latency is already part of the time equalisation of the ONUs.

- **OLT and ONU**: The TDMA mechanisms are the main determining factor for the upstream PON latency, namely the combination of packet classification into T-CONTs and DBA allocations per T-CONT:

  – Mapping of traffic to T-CONTs (e.g., separating traffic in dedicated queues with dedicated T-CONT BW profiles). The latency is not due to the (configurable) mapping action itself but is the result of the DBA treatment that the packets of the corresponding T-CONT will receive.

  – The packets in a given ONU queue require burst allocations by the OLT DBA algorithm based on the BW profile of the corresponding T-CONT. The burst allocations (size, rate, and gaps between allocations) relative to the load pattern will determine the upstream buffer filling in the ONU and hence the latency and PDV.

  The configurable parameters defined for all PON flavours are R_F, R_A, R_M, χ_{AB}, P, ω. Additionally, since [ITU-T G.989.3] (optional) time-related parameters have been defined; T_{JT}, T_{DBT} and T_{PST}.

- The effect of T_{JT} is the main parameter as it directly influences the gap between two consecutive allocations to the same T-CONT.
- T_{DBT} determines the waiting time at ramp-up from an inactive (= absence of load) period.
- T_{PST} in the context of protection switching on TDWM PON indicates the time between the downstream synchronization on the protection channel pair and the first allocation.

– Due to time equalization by ranging, a near ONU is forced to wait as long as the furthest possible ONU for an update in the upstream allocations, except if the allocations are fixed in which case there is no update as such. Depending on the arrival time of packets at the ONU UNI with respect to the time of the first updated grant, and depending on the buffer filling with respect to the grant size, this can lead to packets of the ONU experiencing extra latency due to the maximum distance of ONUs on the ODN (e.g., in Figure 7-2, ONU 1 is time equalized to ONU N that represents the ONU at maximum distance).

– The quiet windows opened during ranging interrupt the upstream transmission. The amount of successive quiet windows needed for a ranging event can depend on the amount of ONUs to be discovered and ranged at that event. The duration of quiet windows depends on the maximum allowed differential distance of ONUs on the PON. Note that there are methods (e.g., DAW [ITU-T G.9804.2]) that allow to reduce or remove the impact of ranging windows for a given PON flavour on the ODN.

- **OLT**: PON TC layer processing actions include decapsulation, FEC decoding, and frame defragmentation. These processing steps will directly impact the service latency but can be considered to be minor (order of µs) at the level of the whole upstream PON system latency.

7.2.3 Influence on user-to-user PON system latency

When two PON users communicate with each other, the latencies of both PON-based access segments need to be taken into account. Each user-to-user direction is composed of two legs: an upstream leg from source ONU to OLT and a downstream leg from OLT to destination ONU. The latency in each user-to-user direction is then determined by the sum of the latencies in both legs, with the same dependencies as described in clauses 7.2.1 and 7.2.2.

Figure 7-3 shows the example of a particular case of direct communication between two users with ONUs connected to the same ODN on the same access node, with local interconnection inside the access node like in Figure 5-2b. The diagram shows the latencies for packet A sent from ONU 1 to ONU N (in red) and for packet B sent from ONU N to ONU 1 (in blue).

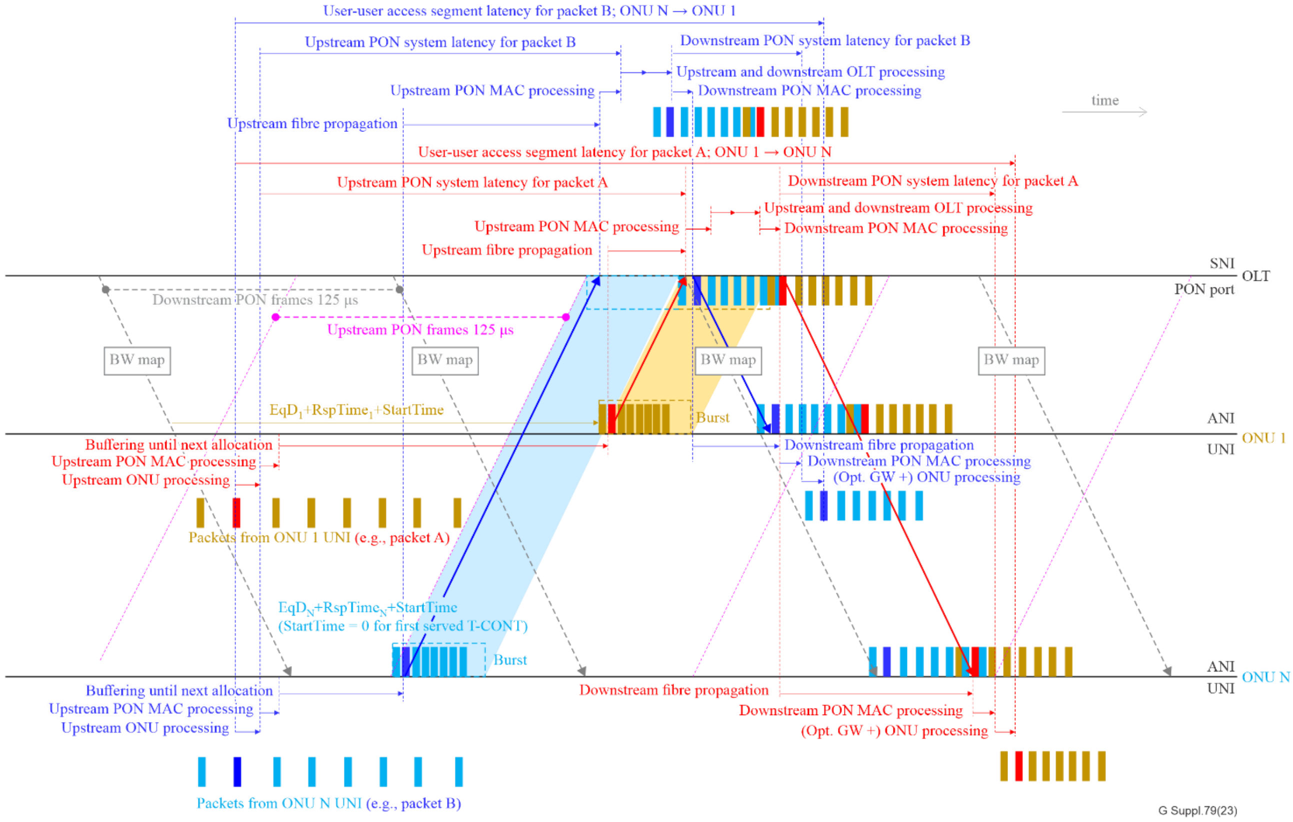

Figure 7-3 – Break-down of per-packet user-to-user latency

7.3 Impact of configuration of PON-specific parameters (TML)

The parameters identified in clause 7.2 that are configurable can be used to control latency and jitter.

7.3.1 Control of worst-case downstream latency and jitter

There are no TML level configurations that can influence the downstream latency and jitter.

Note that for most PON technologies the use of downstream FEC encoding/decoding is mandatory as it is required to ensure that all ONUs can be reached with the optical loss specification of the considered PON class. The exceptions are GPON classes B, B+ and C which do not need FEC. However, the FEC latency is minor compared to the overall downstream PON system latency, so the use of FEC will not make a significant difference in latency.

7.3.2 Minimizing the worst-case (maximum) upstream latency for low-latency traffic

DBA: MAPPING of traffic to T-CONTs

The best result is achieved when separating low-latency traffic from non-low-latency traffic into separate T-CONTs in the ONU in order to follow the DiffServ principles (relative priorities and adequate queueing per traffic class).

DBA: Per-T-CONT BW PROFILES

Each T-CONT has dynamic bandwidth assignment (DBA)-related traffic descriptors.

From a bandwidth perspective, it is straightforward that the higher the bandwidth gets allocated to a T-CONT, the less buffering will be needed at the ONU for a given traffic pattern. The considered **bandwidth-related descriptors** of a T-CONT are R_F, R_A, R_M (the descriptors χ_{AB}, P, ω relate to the priority and fairness of bandwidth sharing between T-CONTs, for simplicity they are not elaborated here). Note that the DBA does not differentiate between packets inside a T-CONT, so these values are not for different traffic classes but any packet in the T-CONT queue could either be served by an allocation driven by R_F, by R_A or by R_M. Selecting values for R_F, R_A, and R_M will determine how much of the traffic load will be treated as guaranteed bandwidth (fixed or assured), or as non-guaranteed bandwidth (non-assured or best effort). Ultimately the packets with the maximum latency will determine the maximum latency of the whole traffic class associated with the T-CONT.

- Assigning an R_F value > 0 guarantees allocation of a fixed bit rate. As there is no need for a dynamic reaction time to change the allocations for the bit rate, using fixed bandwidth provides lower maximum upstream latency for a T-CONT compared to assured and non-assured / best-effort (all other traffic descriptors and traffic loads being equal for the comparison).
  - It is naturally suited for constant bit rate (CBR) traffic.
  - If the traffic is a variable bit rate (VBR) between some minimal and maximal value, there are several possibilities for R_F, in decreasing order for the latency and packet drop probability and increasing order of the bandwidth consumption:
    - Setting R_F to zero means all packets will depend on the reactivity of the DBA.
    - Setting non-zero R_F below the minimal load value leads to (temporary) ONU buffer filling (which can lead to packet drop when overfilled).
    - R_F can be set to the minimal load value, reducing the ONU buffering.
    - Depending on the abundance or scarcity of capacity on the PON, R_F can be increased above the minimal load value, further lowering the ONU buffering but trading off bandwidth efficiency for lower maximum latency.
    - Only when setting R_F to the maximal load value, will the maximum latency be as low as possible.
- Assigning an R_A value > 0 also provides a bandwidth guarantee, but there is a DBA reaction time involved as the assured capacity is only allocated when there is an actual demand. Such demand needs to be detected and then the allocation needs to be actuated, which represents some reaction time and hence a higher maximum latency than with pure fixed allocations.
  - When the traffic is CBR, assured bandwidth is sometimes preferred over fixed bandwidth for cases of known but intermittent loads in order to reassign unused bandwidth to other T-CONTs.
  - For VBR traffic between some minimal and maximal value, the same possibilities exist as for R_F, albeit with some extra latency due to the DBA response time.
- Assigning a non-guaranteed bandwidth ($R_M > R_F + R_A$) is the most bandwidth-efficient but also the most influenced by the DBA responsivity and hence most prone to latency and jitter. The amount of latency and jitter depends on the rate and amplitude of changes in the load versus the responsivity of the DBA to follow load changes.
- As a general remark the use of DBRu reporting benefits the DBA responsivity (except for purely fixed T-CONTs set at or above the peak load).

The **time-related descriptors** are T_{JT}, T_{DBT}, T_{PST}

- The jitter tolerance T_{JT} determines the inter-burst-gap, hence directly influences the maximum latency during steady states and transients of the traffic. The same bandwidth can be allocated by different values of the jitter tolerance. The value can be set to a high frequency of bursts, but there is a trade-off between the latency and available capacity on the PON (more bursts per time unit represent more burst mode overheads).
- T_{DBT} determines how quickly the change of a traffic load from inactive to active is detected, hence directly influencing the maximum latency for transients of the traffic from zero (or low load) to some load. It has to be considered together with T_{JT} for the overall maximum latency of the traffic during operation. It is also subject to the same trade-off between low latency and burst overheads.

- T_{PST} on its own does not guarantee a maximum latency, it always has to be considered together with the other descriptors. If a protection switching event is not allowed to temporarily increase the maximum latency, this parameter must be set sufficiently low compared to the target maximum latency.

Note that these timing values do not apply during quiet windows due to the interruption of the upstream traffic.

CO DBA [ITU-T G.Sup.71]

With cooperative DBA (CO DBA) there is external information provided to the OLT about the expected traffic in the immediate future. The information is provided by an external node over a common interface. The external node must have knowledge of the application-level traffic per time interval. An example is a mobile fronthaul traffic whereby DUs manage the upstream scheduling of UEs and communicate the corresponding RU traffic to OLTs over the cooperative transport interface (CTI).

The OLT maps the traffic report information to the corresponding T-CONT and applies appropriate burst allocations in size, repetition, and timing in order to follow the changing demands per T-CONT. The result is to keep the advantage of statistical multiplexing between T-CONTs of a reactive DBA while keeping the latency sufficiently low for the application. Note that the resulting latency is much improved compared to the classic DBA but is not lower than the approach of the fixed bandwidth allocations (when dimensioned at peak level).

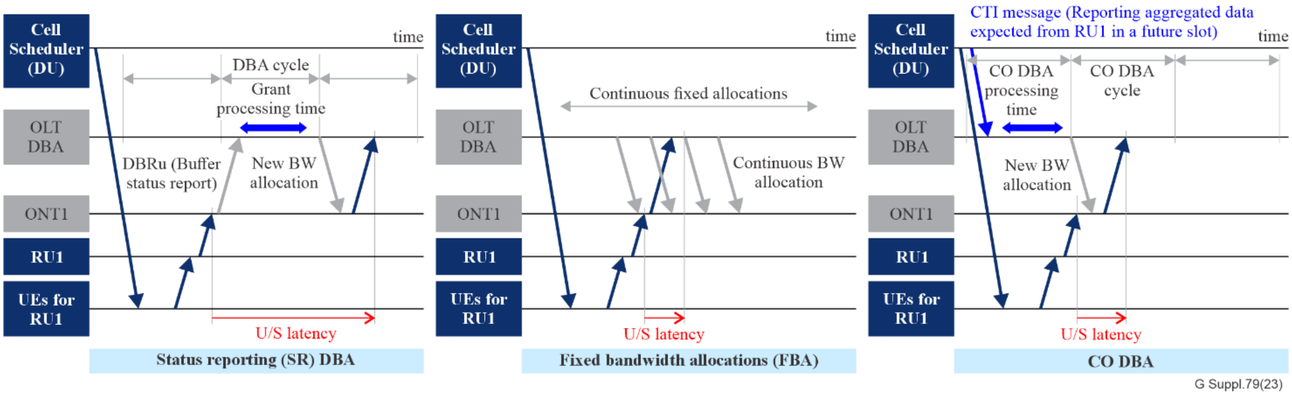

Figure 7-4 – Comparison between SR/TM DBA, fixed bandwidth allocation and CO DBA

Use of FEC

Note that for most PON technologies the upstream FEC encoding/decoding is required to ensure that all ONUs can operate within the optical loss specification of the considered PON class. The exceptions are GPON classes B, B+ and C which do not rely on FEC. For the other PON types, the upstream FEC can only be disabled for ONUs that have enough optical margin to reach the required BER without FEC. However, the FEC latency is minor compared to the overall upstream PON system latency, so the use of FEC will not make a significant difference in latency.

ONU activation and RANGING

As quiet windows are an abrupt interruption of upstream transmission there is an obvious relationship between the duration and frequency of quiet windows and the maximum latency measured over periods of time in which ranging occurs. Reducing or avoiding the use of quiet windows directly benefits latency and jitter.

7.3.3 Minimizing worst-case upstream jitter = maximum latency – minimum latency

The minimum latency is not controllable at the PON level as it is not possible to artificially increase the minimal latency of a given packet at the PON TML level.

All aspects of clause 7.2.3 that reduce the maximum latency will hence also limit the jitter.

7.3.4 Impact of a quiet window on regular US transmission

As every ONU is forced to wait a period of time during each quiet window period, the temporarily opened quiet windows interrupt the regular US transmission of operating ONUs. Assuming the standard differential fibre reach is 20 km, the maximum duration of the quiet window is normally about 250 µs, including RTD, ONU response time, and additional random delay for SN acquisition.

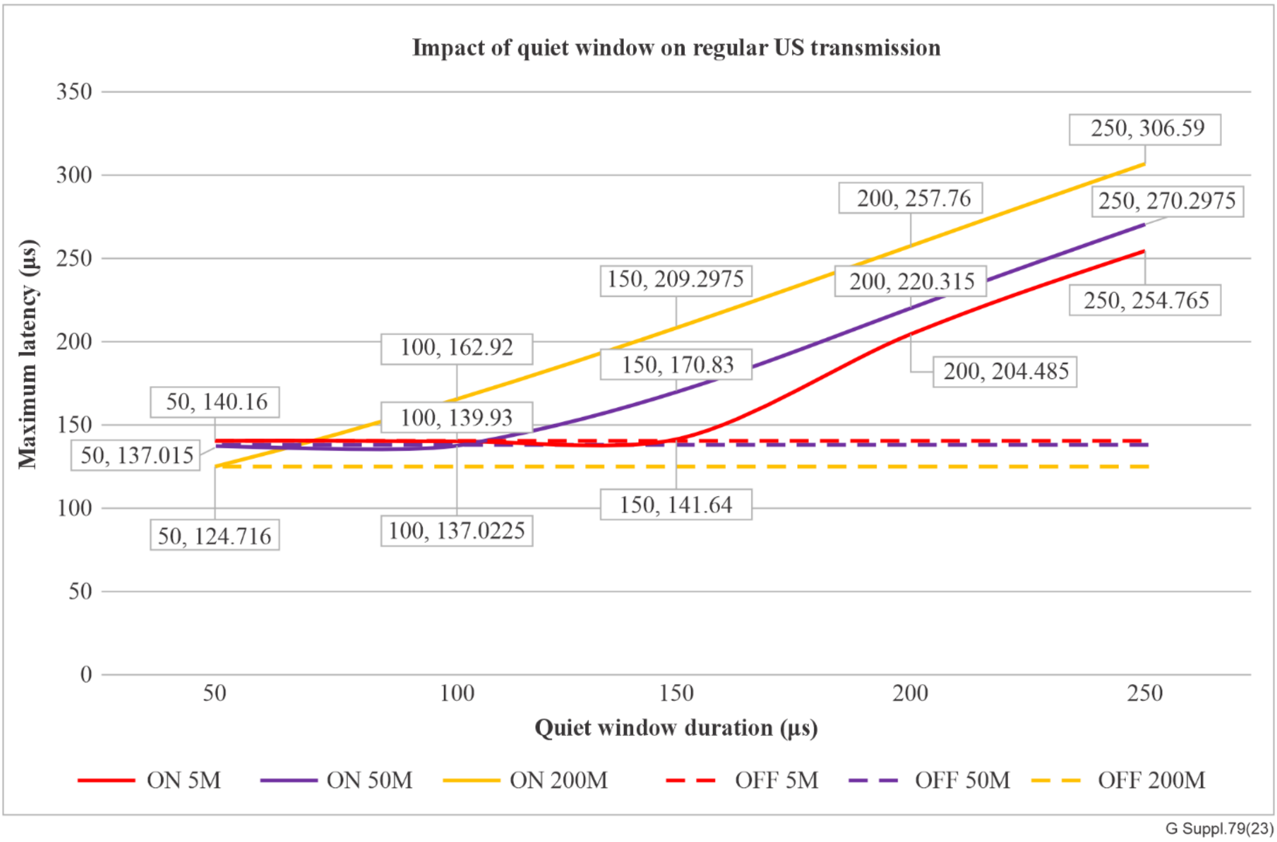

Figure 7-5 – Impact of a quiet window on regular US transmission

The test results of a lab test based on a GPON system with 4 ONUs are shown in Figure 7-5. Each ONU was granted the same US traffic. Three US traffic cases of 5M, 50M and 200M per ONU were tested. The solid lines indicate the maximum latency under different durations of quiet window, while the dotted lines indicate the case of quiet window disabled. Note that only the common single burst allocation was tested.

The measurements show the buffering effects due to the quiet window interruptions under stable burst allocations. For 50 µs quiet window, when the total bandwidth is enough for the US traffic following such a short quiet window, there seems no obvious difference between the quiet window enabled or disabled. The maximum latency levels are gathered around 125~140 µs. However, for 150 µs quiet window, it brings ONUs with 50M and 200M US traffic to the maximum latency levels of around 170 µs and 209 µs respectively. The 250 µs quiet window is relatively long and leads to maximum latency levels of around 255 µs, 270 µs and 307 µs respectively for ONUs with 5M, 50M and 200M US traffic.

According to the test results, the longer the quiet window duration (longer transmission distance), the greater the impact on regular US transmission. Moreover, the latency introduced by the quiet window increases with the US traffic rate, especially when the quiet window duration is relatively long.

7.4 Practical experience of latency control mechanisms

7.4.1 Frame-based dense burst allocation mechanism

For upstream transmission in TDM PON systems, OLT is in charge of the transmission time slot allocations, the belonging ONUs are only to transmit in the permitted time slot. In practice, the time slot allocation also requires processing time in the circuit. Consequent data awaiting to be transmitted in the upstream direction may need to wait for a complete upstream frame period. Moreover, if the number of belonging ONUs increases, the OLT may not be able to process allocations within one upstream frame period. Therefore, for the general PON system, the OLT may be configured to do time slot allocation in multiple 125 μs periods, leading to a worst-case delay of multiple 125 μs.

In industry PON scenarios, applications such as closed-loop fieldbus control and high frequency periodic signal collecting are required to minimize the upstream transmission latency and jitter level. A frame-based dense burst allocation mechanism is introduced to reduce the latency and jitter level in the upstream transmission by increasing the frequency of the time slot allocation.

As shown in Figure 7-6, in the general case, ONU1 transmits the arrived data 1 in the first permitted time slot, and the consequent data 2 to 4 need to wait a long delay for other time slot allocations. If the upstream frame period is split into 4 sub-periods equally, and, to execute the time slot allocation in each sub-period, the consequent data 2, even data 3 and 4 may have the opportunity to be granted a respective time slot within a single 125 μs upstream frame period, which can significantly reduce the time slot allocation delay for time sensitive upstream transmission. On the other side, due to the implementation complexity, stacking of burst mode overheads (i.e., Tplo), and reducing of time slot period, the number of time slot granted sub-periods has a practical limit.

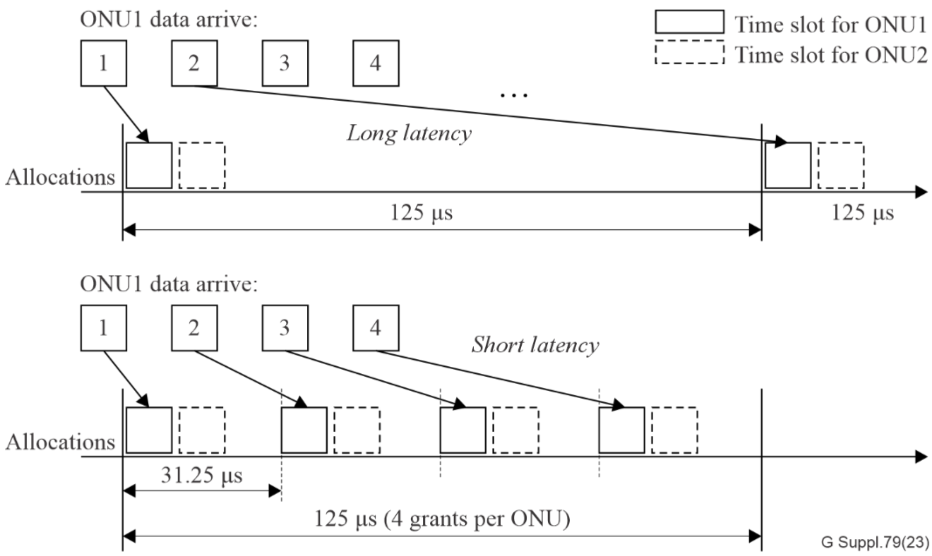

Figure 7-6 – Illustration of general burst allocation versus frame-based dense burst allocation mechanism

Theoretically, this mechanism introduces a maximum delay of 125 μs/N plus other fractions, where N indicates the number of sub-periods in a single upstream frame. Figure 7-7 illustrates the results from a lab test of a typical industrial control scenario with four belonging ONUs. The two sub-periods and four sub-periods in a single upstream frame are analysed.

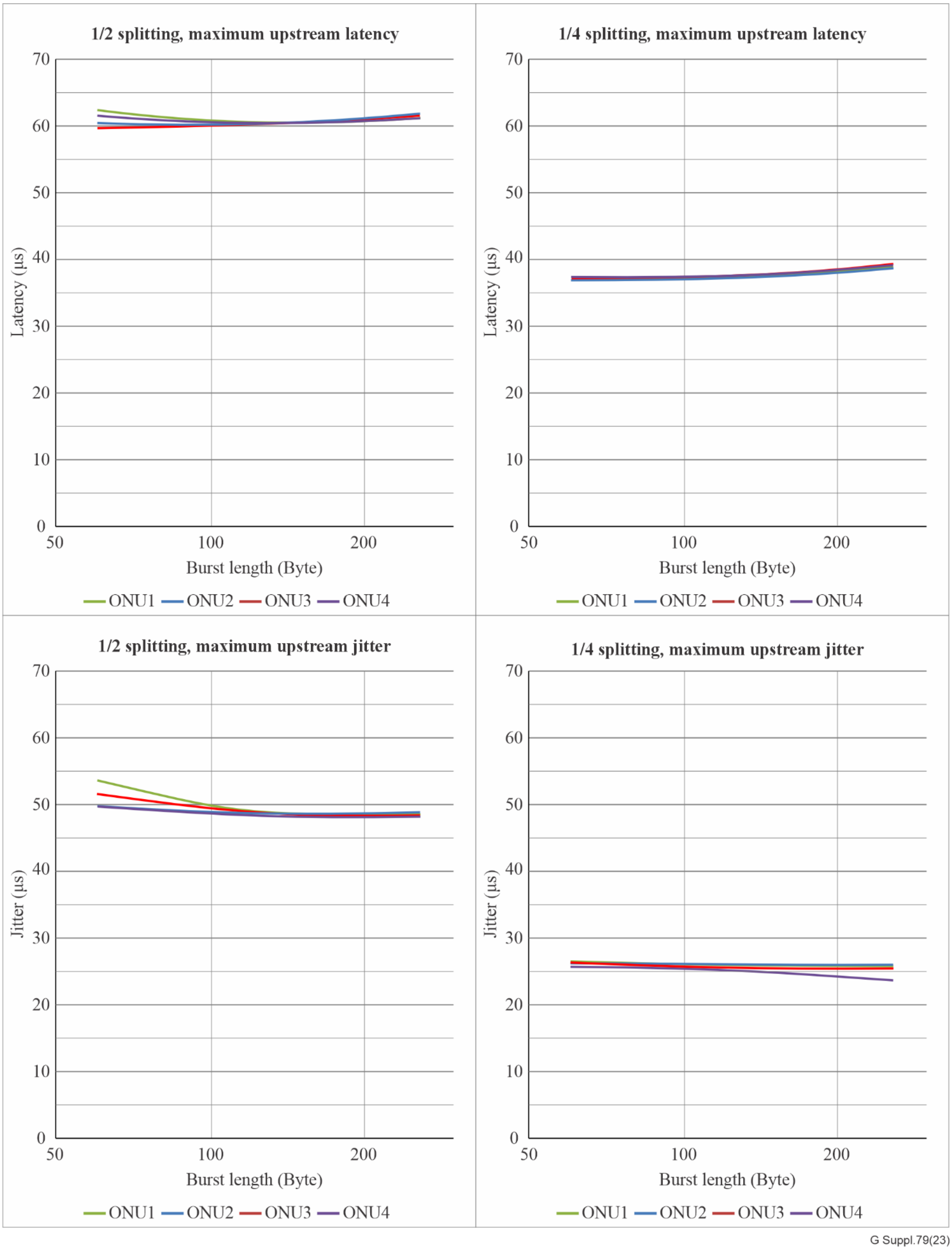

Figure 7-7 – Illustration of frame-based dense burst allocation mechanism

For industrial requirements, analysis on the overall maximum level is more valuable. Accordingly, the upper boundaries of the maximum upstream latency for two and four sub-periods are roughly 64 µs and 39 µs respectively, while the upper boundaries of the maximum upstream jitter for two and four sub-periods are roughly 55 µs and 27 µs respectively.

The test results reflect the gradient reduction on the maximum latency and jitter level with different numbers of sub-periods, which also show the performance degradation (i.e., 39 µs larger than half of 64 µs) when increasing the time slot allocation frequency in a single upstream frame.

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

Published in Switzerland
Geneva, 2024