ETSI TR 102 720 V1.1.1 (2009-10)

Technical Report

Speech and multimedia Transmission Quality (STQ);
Delay variation on unshared access lines

Reference
DTR/STQ-00150

Keywords
quality, voice

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://webapp.etsi.org/IPR/home.asp).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Speech and multimedia Transmission Quality (STQ).

Introduction

Limited bandwidth is a major source for delay variation in packet networks. This is especially the case for residential customers with a low physical bandwidth connection to their ISP, and it is of concern for VoIP services.

Why this is the case, even with a perfect implemented prioritisation, is shown in the present document.

There are also some numbers shown to get a feeling of the impact a low bandwidth connection has in respect of delay variation.

The content of the present document is valid for unshared access lines.

1 Scope

The intention of the present document is to provide guidance for VoIP transmission planners and SDOs in the area of delay and delay variation, especially on access lines.

The present document provides an introduction on the effect of different IP services on lines with limited bandwidth (e.g. DSL). It explains the mechanism of serialisation delay, gives a (very general) overview over prioritisation and shows how the maximum delay variation due to concurrent traffic can be calculated. The calculations shown in the present document are valid for unshared lines only, shared lines are excluded.

2 References

References are either specific (identified by date of publication and/or edition number or version number) or non-specific.

- For a specific reference, subsequent revisions do not apply.
- Non-specific reference may be made only to a complete document or a part thereof and only in the following cases:
  - if it is accepted that it will be possible to use all future changes of the referenced document for the purposes of the referring document;
  - for informative references.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication ETSI cannot guarantee their long term validity.

2.1 Normative references

The following referenced documents are indispensable for the application of the present document. For dated references, only the edition cited applies. For non-specific references, the latest edition of the referenced document (including any amendments) applies.

Not applicable.

2.2 Informative references

The following referenced documents are not essential to the use of the present document but they assist the user with regard to a particular subject area. For non-specific references, the latest version of the referenced document (including any amendments) applies.

[i.1] ITU-T Recommendation Y.1540: "Internet protocol data communication service - IP packet transfer and availability performance parameters".

3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

ATM	Asynchronous Transfer Mode
DSL	Digital Subscriber Line
IP	Internet Protcol
IPDV	IP Packet Delay Variation
ISP	Internet Service Provider
PVC	Permanent Virtual Connection
SDO	Standards Development Organization
VoIP	Voice over Internet Protocol

4 Introduction

For some time now, the coverage of broadband accesses for customers is getting higher and higher. These broadband accesses are used for different services, starting with Internet, nowadays more and more also TV and voice services. Packet based networks offer a high flexibility to deliver all of these services over the same network. If more than one service is used at the same time or if one service uses more than one session at a time, there is the possibility of interference. One very real effect will be the influence of other services/sessions on VoIP-media traffic.

5 Serialisation delay

Serialisation delay of a packet is the time it takes to clock every bit of a packet onto the line. A packet ready to be sent will be normally put in a play out buffer, from where it will be clocked onto the line with the physical line speed (see figure 1).

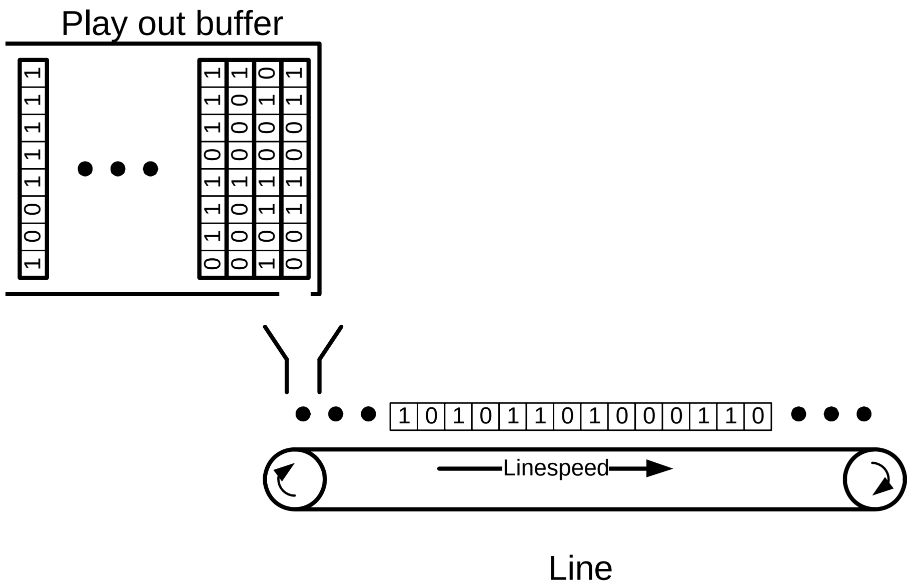

Figure 1: Serialisation delay

The formula to calculate the serialisation delay $t_{serialisation}$ is as follows:

$$t_{Serialisation}[s] = \frac{Packetsize[Bit]}{Linespeed[Bit/s]}$$

With:

Packetsize = size of a packet on the physical layer.

Linespeed = linespeed on the physical layer.

Table 1 shows some example calculations.

NOTE: It is recommended to do this calculation on the physical layer even if it could be done on any other layer, as long as the packet size and the line speed are calculated for the same layer. It has to be taken into account that the packet size need to represent the size of a packet, including all headers and trailers (for further calculations it may also be necessary to include the minimal distance between two packets).

Table 1: Example of serialisation delays with different linespeed and packetsizes

Linespeed	Packetsize	Serialisation delay
1 Gbit/s	1 500 Bytes	0,012 ms
1 GBit/s	200 Bytes	0,0016 ms
100 MBit/s	1 500 Bytes	0,12 ms
100 MBit/s	200 Bytes	0,016 ms
10 MBit/s	1 500 Bytes	1,2 ms
10 MBit/s	200 Bytes	0,16 ms
1 MBit/s	1 500 Bytes	12 ms
1 MBit/s	200 Bytes	1,6 ms
100 kbit/s	1 500 Bytes	120 ms
100 kbit/s	200 Bytes	16 ms

6 Prioritisation

The general concept of prioritisation is that traffic with higher priority is favoured against traffic with lower priority on the same line. Prioritisation is important in the case where a bandwidth limitation exists (more input capacity than output capacity) or in case of congestion (these two effects can be related). The prioritisation is a strong way of queue management (strict priority), which means that this traffic is prioritised in any case, or it can be used a weaker bandwidth allocation (like a weighted fair queuing) to allow also lower "prioritised" traffic to pass even in the case of congestion.

Normally VoIP media traffic is in the highest priority class which will use strict priority queues (as a matter of fact, internal network control traffic will be even higher prioritised). For the reminder of the traffic there are several priority classes possible, which normally will use a fair queuing.

Figure 2 shows the functional diagram of a typical prioritisation algorithm in network equipment.

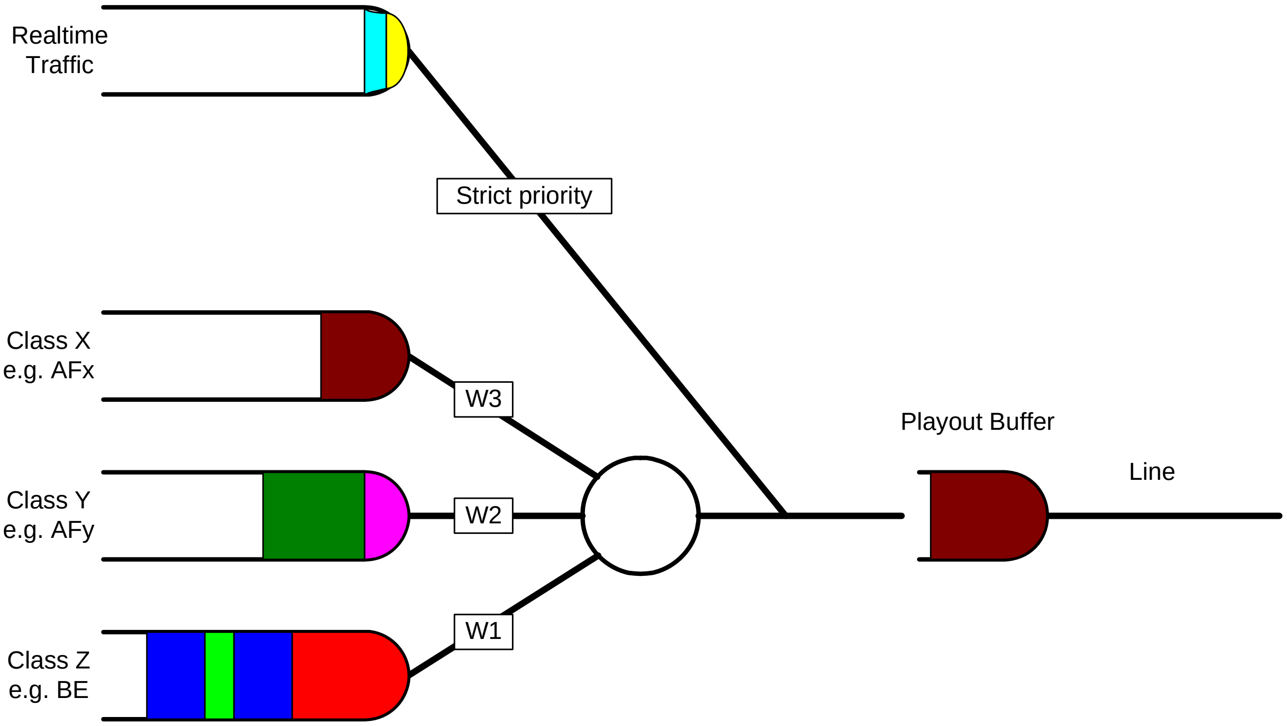

Figure 2: Prioritisation principle

Classes X, Y and Z are three differently prioritised classes (in the picture, classes X and Y are two different Assured Forwarding classes, while class Z is a Best Effort class). In this case the prioritisation is according to different weights (W1, W2 and W3) of the classes, but every class will get its time slot to send (based on the weighting). This means, that it is possible that a lowest prioritised packet from the queue of class Z can be sent to the playout buffer, even if there are packets in the higher prioritised queues of class X and Y.

Since the realtime class is allocated strict priority, no packet out of the queues X, Y and Z can be sent to the playout buffer, if there is a packet in the realtime queue (this implicitly means that there has to be some sort of bandwidth control for the realtime class, to avoid a blockage of all other traffic).

In the playout buffer, there is no prioritisation. Packets in the playout buffer will be sent to the line, in the order they arrived (FIFO, first in, first out).

If the playout buffer is full, not even a packet from the realtime queue can be sent to it. The prioritised packet from the realtime queue has to wait until the packet in the playout buffer is sent to the line.

With the formula for the serialisation delay, it can be calculated, how much time this takes. This effect leads to delay variation for realtime traffic.

In a usual priority implementation the playout buffer has the capability to hold two packets; this means that in the worst case, two low prioritised packets have to be sent, before a high priority packet can be sent.

The formula for the maximum delay variation $t_{delayvariation}$ due to this effect will be the one for the serialisation delay multiplied with the number of packets, which the playout buffer can hold.

NOTE 1: In difference to the formula for the serialisation delay, where the actual packet size of the packet under interest has to be taken, in the formula for the maximum delay variation, the maximum packet size possible on the link has to be taken.

$$t_{delay\,variation}[s] = \frac{NbrofPackets \times MaxPacketsize[Bit]}{Linespeed[Bit/s]}$$

With:

NbrofPackets = max. number of packets in the playout buffer.

MaxPacketsize = maximum size of a packet on the link (physical layer).

Linespeed = linespeed on the physical layer.

NOTE 2: A typical maximum packet size for IP networks is around 1 500 Bytes at the IP-layer.

7 Measurement examples

Figures 3 to 5 show real IP packet delay variation measurements for a VoIP call between two DSL-Accesses, customer A with 6 400 / 640 kbit/s access speed, customer B with 4 608 / 576 kbit/s access speed.

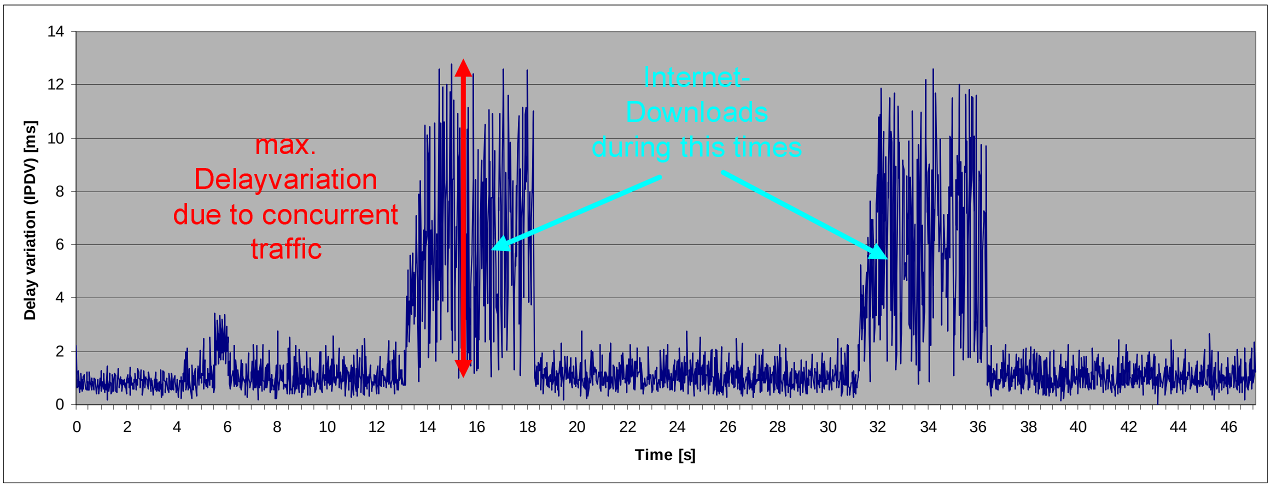

Figure 3: Delayvariation (IPDV) of a call A to B, with parallel download at B

Figure 3 shows the delay variation (IPDV according to [i.1]) over time measured at customer B side, with two intermittent Internet downloads also on the B side. According to the formula for the serialization delay, one Internet packet (1 500 Bytes at IP layer -> app. 1 696 Bytes on the physical layer in this case) will have a serialization delay of 2,94 ms (downstream bitrate B: 4 608 kbit/s). Since the maximal delay variation measured is much higher (nearly 12 ms, taken the difference of the maximum IPDV and the IPDV without Internet download), it can be assumed, that the playout buffer of the network equipment involved in the prioritisation towards the DSL-line holds up to 4 IP-packets. With a better implementation of the prioritisation algorithm, IPDV could be reduced by 9 ms.

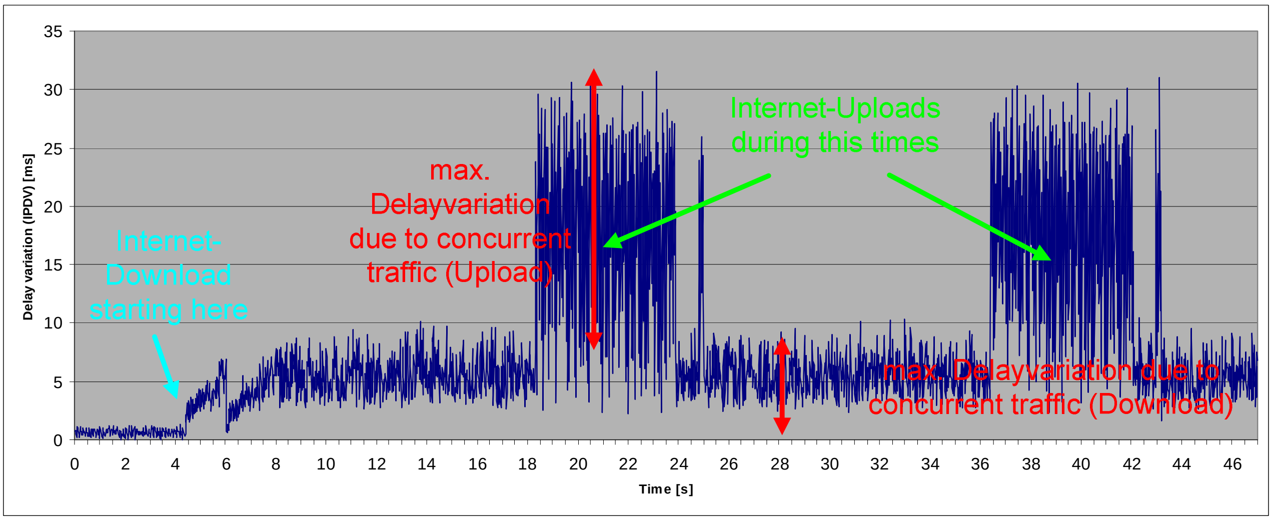

Figure 4: Delayvariation (IPDV) of a call B to A with parallel download at A and upload at B

This graph shows the delay variation (IPDV according to [i1]) over time measured at customer A side, with a continuous download at the A side (starting at 4 s) and two intermittent uploads at the B side.
The serialisation delay of one internet packet (1 500 Bytes at IP layer -> app. 1 696 Bytes on the physical layer in this case) on the A side will be 2,12 ms (downstream bitrate A: 6 400 kbit/s). Since the maximal delay variation measured is much higher (about 8 ms, taken the difference of the maximum IPDV without upload and the IPDV without up- and download), the conclusion from the previous measurement is confirmed: The playout buffer of the network equipment involved in the prioritisation towards the DSL-line holds up to 4 IP-packets.
The serialisation delay of one internet packet (1 500 Bytes at IP layer –> app. 1 696 Bytes on the physical layer in this case) on the B side will be 23,5 ms (upstream bitrate B: 576 kbit/s). This is the delay variation measured during the two intermittent periods of upstream traffic (taken the difference of maximum IPDV during up- and download and maximum IPDV during download). This means that the playout buffer of the customer router holds only 1 IP-packet, which is the optimum.

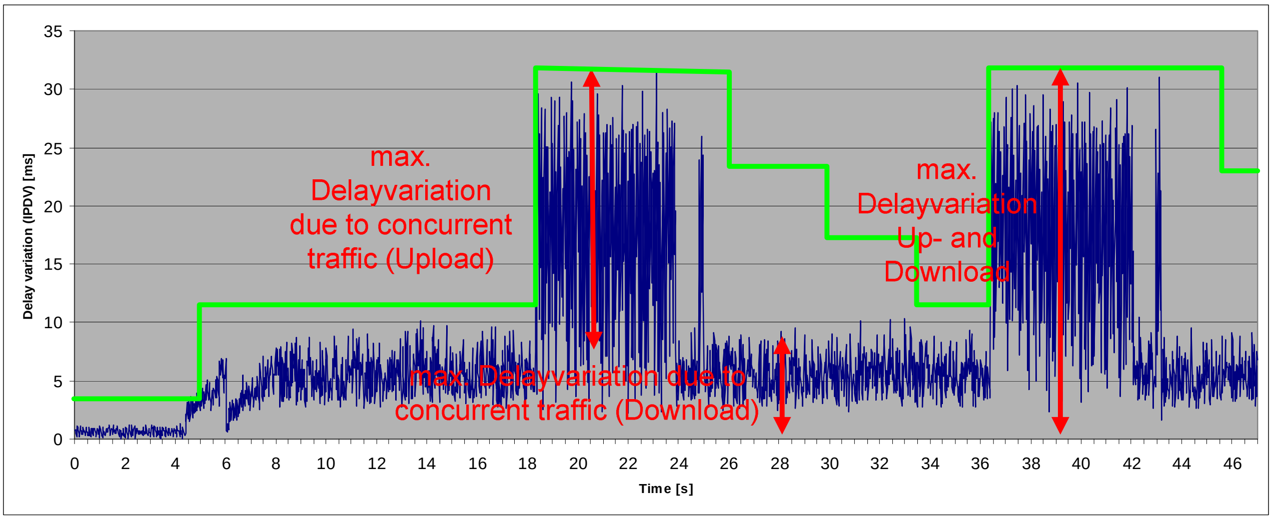

Figure 5: Delayvariation (IPDV) of a call B to A with parallel download at A and upload at B

Figure 5 shows a possible End-to-End (audio) delay variation. This delay variation depends on the dejitter buffer behaviour, so the green line is an example. Figure 5 also shows that if there is delay variation on both accesses of a connection, the resulting delay variation will be the sum of the two delay variations.

8 Theoretical considerations

Unfortunately, there are not many options to overcome the problem of exceeding delay variation on access lines.

- Using ATM with different PVCs per prioritisation class (and a playout buffer per PVC). With this solution, the maximal packetsize on the link will be 53 bytes (ATM frame size) instead of over 1 500 bytes. Unfortunately ATM will no longer be an option as it will be phased out more and more.

- Fragmentation of lower prioritised packets: If the maximum packet size for lower prioritised traffic is set to one half (e.g. from 1 500 Bytes to 750 Bytes), the maximum delay variation will also be only half of the delay variation without fragmentation. This approach has two major disadvantages: There will be more overhead, leading to less net bandwidth, and there will be additional load on the network equipment which has to do the fragmentation/defragmentation.

- If the physical bandwidth is enhanced, the maximum delay variation will decrease. Obviously the enhancement of the physical bandwidth is often not an appropriate solution due to economical constraints.

- Tell the customer, that he should not use the Internet or any other service, while using VoIP. For one point this would not help much, because there is still parallel signalling traffic used for the VoIP connection also causing delay variation, for another point, no operator will tell his customer, that he cannot use all IP services in parallel.

9 Conclusions

The mechanism showed in the present document leads to the conclusion, that for VoIP-services the access parts of a connection lead to substantial delay variation, which has to be taken into account for network planning purposes.

Furthermore if there is concurrent traffic on both sides of a VoIP connection (upstream on one side, downstream on the other side), the resulting delay variation is the sum of both delay variations and consequently a normal jitter buffer on the receiver side will produce an additional delay of at least the sum of both delay variations.

The possible theoretical solutions are often not applicable in practice.

For an operator it is necessary to find a balance between coverage (which is given with the minimum access bandwidth needed for VoIP) and voice quality (which is the maximal delay variation allowed on a connection).

If delay variation limits in planning guidelines are set too low, a lot of customers will not be able to ever get VoIP-services.

Therefore it is important to find a solution for planning guidelines, which allows lower bandwidth accesses (-> higher delay variation limits) without relaxing the delay variation limits for high bandwidth accesses.

History

Document history		
V1.1.1	October 2009	Publication