NICC ND 1704 V1.1.2 (2008-09)

NICC Document

End-to-End Network Performance Rules & Objectives for the Interconnection of NGNs

NICC

Ofcom,

2a Southwark Bridge Road,

London SE1 9HA.

If you have any comments concerning the accuracy of the contents of this document, please write to:

The Technical Secretary, Network Interoperability Consultative Committee,

Ofcom,

2a Southwark Bridge Road,

London SE1 9HA.

Contents

Intellectual Property Rights 4

Foreword 4

Introduction 4

2 References 5
2.1 Normative references 5
2.2 Informative references 5

3 Key words and abbreviations 5
3.1 Key words 5
3.2 Abbreviations 5

4 General 6
4.1 Relationship with ND1701 6
4.2 Assumptions Regarding the Maximum Number of Interconnected NGNs 6
4.3 Enterprise Networks 6

5 Network Performance Rules and Objectives 7
5.1 General 7
5.2 Packet Delay and Delay Variation 7
5.2.1 General Guidance 7
5.2.2 Detailed Guidance 8
5.2.3 Quantitative Objectives 8
5.2.3.1 Connections Using Adaptive Dejitter Buffering 9
5.2.3.2 Connections Using Static Dejitter Buffering 9
5.3 Codecs 10
5.4 Packet Loss 11
5.5 Echo Control 11
5.6 Post Dial Delay 12
5.6.1 Definition of Post Dial Delay 12
5.6.2 Assumptions 12
5.6.3 Apportionment Model 12
5.6.4 Targets 13

History 15

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to NICC.

Pursuant to the NICC IPR Policy, no investigation, including IPR searches, has been carried out by NICC. No guarantee can be given as to the existence of other IPRs which are, or may be, or may become, essential to the present document.

Foreword

This NICC Document (ND) has been produced by NICC E2E QoS WG.

Introduction

This document describes rules governing the end-to-end network performance of interconnected fixed and mobile Next Generation Networks providing PSTN/ISDN services. Per CP quantitative objectives are associated with some of the performance parameters.

The network performance planning rules described in this document are intended to ensure that end-users experience high quality of service when using PSTN/ISDN provided over interconnected NGNs. Careful consideration needs to be given to the potential effect on user-perceived QoS (notably delay) when planning the deployment of IP-based technologies in Next Generation Networks. The rules given in this document **shall** be taken into account in order to minimise the impact on QoS of the introduction of NGNs and any other new technologies.

Future versions of this document may be expanded to include other services (such as multimedia services and non-PSTN voice) and access networks.

This document only addresses aspects of the performance of public NGNs, both fixed and mobile. It does not address the network performance or QoS implications of enterprise networks or, in the current version, of non-PSTN VoIP.

1 Scope

The present document describes rules governing the end-to-end network performance of interconnected fixed and mobile Next Generation Networks providing PSTN/ISDN services. Per CP quantitative objectives are associated with some of the performance parameters.

Future versions of this document may be expanded to include other services (such as multimedia services and non-PSTN voice) and access networks.

2 References

2.1 Normative references

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

[1] ETSI SR 001 262 ETSI drafting rules Verbal Forms For The Expression Of Provisions Version 2004-07

[2] ND1701 Recommended Standard for the UK National Transmission Plan for Public Networks Issue 5 2006/03

[3] ITU-T Rec. G.711 Pulse code modulation (PCM) of voice frequencies

[4] ND1612 Generic IP Connectivity for PSTN/ISDN Services between UK Next Generation Networks Issue 2 2006/06

[7] ITU-T Rec. Y.1541 Network performance objectives for IP-based services

[8] ITU-T Rec. G.168 Digital network echo cancellers

[9] ITU-T Rec. Q.115.1 Logic for the control of echo control devices and functions

[11] ND1423 Guidelines for Usage of Enbloc/ Overlap Signalling in UK Networks Issue 1 2007/06

2.2 Informative references

[5] ITU-T Rec. G.113 Transmission impairments due to speech processing

[6] ITU-T Rec. G.107 The E-model, a computational model for use in transmission planning

[10] ITU-T Rec. G.108.2 Transmission planning aspects of echo cancellers

3 Key words and abbreviations

3.1 Key words

The key words “**shall**”, “**shall not**”, “**must**”, “**must not**”, “**should**”, “**should not**”, “**may**”, “**need not**”, “**can**” and “**cannot**” in this document are to be interpreted as defined in the ETSI Drafting Rules [1].

3.2 Abbreviations

A/D	Analogue to Digital
ATM	Asynchronous Transfer Mode
CP	Communications Provider
CPE	Customer Premises Equipment
D/A	Digital to Analogue
E2E QoS WG	End-to-End Quality of Service Working Group
ETSI	European Telecommunications Standards Institute
GoS	Grade of Service
IAM	Initial Address Message

IETF	Internet Engineering Task Force
IP	Internet Protocol
IPDV	IP Packet Delay Variation
ISC	Interconnect Standards Committee (replaced by TSG WP)
ISDN	Integrated Services Digital Network
ISUP	Integrated Services User Part
ITU-T	International Telecommunications Union - Telecommunications Standardization Sector
NGN	Next Generation Network
NICC	Network Interoperability Consultative Committee
NTP	Network Termination Point
PLC	Packet Loss Concealment
PNO-IG	Public Network Operators' – Interest Group (replaced by TSG)
PNO-ISC	Public Network Operators' – Interconnect Standards Committee (replaced by TSG)
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RFC	Request for Comments
RTP	Real-time Transport Protocol
SIP	Session Initiation Protocol
SIP-I	SIP ISUP mapping
TCP	Transmission Control Protocol
TDM	Time Division Multiplexing
TSG	Technical Steering Group
UK	United Kingdom of Great Britain and Northern Ireland
VLC	Voice Line Control
VoIP	Voice over Internet Protocol

4 General

4.1 Relationship with ND1701

This document describes rules governing the end-to-end network performance of interconnected fixed and mobile UK Next Generation Networks providing PSTN/ISDN services. NGNs are based on IP technology and the interconnect between them is also IP-based.

ND1701 [2] describes rules governing the end-to-end network performance of the UK PSTN and of public networks evolving towards NGN. The scope of ND1701 covers networks based on TDM, ATM and IP technology but, importantly, it only deals with TDM-based interconnect between such networks.

4.2 Assumptions Regarding the Maximum Number of Interconnected NGNs

Certain aspects of network performance (e.g. packet delay and error ratios, dejitter buffer dimensioning) are a function of the number of CP networks in a call path. It is not technically possible to ascertain the number of CPs involved in a path on a call-by-call basis. In order that recommendations can be made regarding these aspects of performance this document assumes that a maximum of six CP networks will be involved in any call between interconnected NGNs in the UK.

4.3 Enterprise Networks

Enterprise networks are outside of the scope of this document but the existence of enterprise networks **cannot** be ignored in network performance planning because they will have an effect on end to end impairments. Historically the NTP-NTP delay performance of the PSTN was below 15ms and enterprise network operators have been able to use this low delay to allow the implementation of low bit rate codecs and IP technology and still achieve “acceptable” quality for calls from the enterprise network to the PSTN/ISDN. As IP technology is deployed in the PSTN and NGNs, enterprise network operators are encouraged to bear in mind the recommendations of this document concerning delay in the public network.

5 Network Performance Rules and Objectives

5.1 General

Rule G1. In order that the number of "packetisations", and therefore delay, can be kept to a minimum, NGNs in the UK **should** be interconnected using IP. *Note: The availability of TDM and IP or other broadband interconnects within the UK are driven by performance and other non-performance related considerations. Timescales for the implementation or withdrawal of interconnect products are outside the scope of this document.*

Rule G2. Care **shall** be taken to ensure that the performance of a service is not adversely affected by other traffic types. NGNs **shall** preserve grade of service (GoS) even under high or rapidly fluctuating load and/or mix of other traffic types.

5.2 Packet Delay and Delay Variation

5.2.1 General Guidance[1]

Rule PD1. Transmission delay **shall** be minimised. See section 5.2.2 and 5.2.3 for detailed guidance.

Rule PD2. IP packet delay variation **shall** be minimised.

Rule PD3. IP dejitter buffer delay **shall** be minimised. *Note: There is a trade-off between dejitter buffer size (and therefore end-to-end delay) and packet loss, with shorter dejitter buffer delays potentially resulting in higher packet loss. This should be borne in mind when optimising these two parameters.*

Rule PD4. A dejitter function **shall** be provided at the point of transition from a packetised voice bit stream to a synchronous voice bit stream. For calls within the UK NGN in the scope of this document, dejitter functions **should not** be applied on a per network element basis or at the boundary between two IP-based networks interconnected with IP. *Note: There may also be a need to use a dejitter buffer at points where the codec type or speech sample size are different to those specified in Section 5.3 Rule C1 and Rule C2 (see Section 5.3 Rule C5).*

Rule PD5. In addition to the dejitter function to be provided in accordance with Rule PD4, additional dejitter buffers **may** be required on calls entering the PSTN/ISDN from enterprise networks, from outside the UK or from other non- PSTN/ISDN services. Such additional buffers **shall** be provided to ensure that the CP receiving such a call does not pass on more delay variation to the next network than that allowed by section 5.2.3.

Rule PD6. In order to help minimise delay, adaptive dejitter buffers **shall** be used. Exceptionally, when the network detects the presence of a call involving fax, dial-up modem or ISDN clearmode the dejitter buffer **shall** be fixed at the value specified in section 5.2.2.2. Static dejitter buffers are used in these cases to avoid the impact of buffer adaptation on data transmission. For ISDN clearmode the static dejitter buffer **shall** be established on call set-up (i.e. before any user data is sent). For fax and other modem calls the call **may** start as a voice call with an adaptive dejitter buffer; in this case the dejitter buffer **should** then be switched to a static behaviour. See Section 5.2.3 for further details on the sizing of dejitter buffers. *Note: When in-band data is detected and the dejitter buffer switches to a static setting it may have an impact on modem training or data transmission.*

[1] CPs should be aware of the contribution that delay variation in packet streams entering the public network from an enterprise network makes to the performance of the overall end-to-end connection.

5.2.2 Detailed Guidance

CPs **shall** endeavour to keep delay to a minimum. Using IP interconnects for IP networks reduces the possibility of having multiple delay-intensive packetisation functions. The IP interconnect also reduces the number of dejitter buffer instances. Packet delay variation is treated as additive in this document, hence a single static dejitter buffer instance will have a nominal delay equal to the sum of the assumed per-network delay variations. However future versions of this document may propose alternative (lower) values for the nominal delay based on a better understanding of jitter accumulation. The use of a single dejitter buffer instance results in lower end-to-end delays than the use of multiple instances.

Except in the cases indicated below, adaptive dejitter buffers **shall** be used as they result in the lowest practicable end-to-end delay. The dejitter buffer will contribute less to the end to end delay of calls with low jitter than calls with high jitter.

The exception to this is when the networks detects the presence of in-band data calls (such as fax and dial-up modems) and ISDN clearmode, in which case static dejitter buffers **shall** be used.

The delay of a connection is the sum of

- packetisation delay
- transmit delay (scheduling and coding)
- receive delay (scheduling and decoding)
- packet network fixed delay
- total jitter-related delay
- the delay associated with signal propagation time, A/D & D/A conversion, packet loss concealment and echo cancellation

5.2.3 Quantitative Objectives

The delay objectives given below do not include delays associated with the access network or the following functions which introduce a small amount of delay but which **should** occur no more than once in the call path:

- Analogue-Digital conversion and Digital-Analogue conversion
- Echo cancellation
- Packet loss concealment (PLC)

Each network in a call path **shall** introduce no more than 1ms of propagation delay for each 100km the call is carried across their infrastructure. CPs **shall** design and operate their network in such a way as to minimise total propagation delay (e.g. by appropriate choice of network architecture, interconnects and routing).

The overall effect across all networks of propagation delay, A/D and D/A conversion and echo cancellation is expected be less than 10ms delay for all UK calls.

Delay and IPDV objectives for access networks are for further study. Different objectives may be defined for specific access technologies. Until such objectives are agreed CPs **shall** apply the general guidance in 5.1 and 5.2.1 where applicable to access networks in order to minimise the effects of delay and IPDV.

Each network in the IP part of the call path is allocated a delay allowance associated with its fixed delay F and its variable delay V. N.B. The variable delay allocation includes the contribution from the VLAN across the interconnect in the originating speech direction.

- Fixed delay F=4 ms
- Variable delay V=7 ms

The source network (the one that packetises the speech) is additionally allocated a delay allowance associated with speech packetisation P and packet transmission T.

- Packetisation P=10 ms (defined by the packetisation interval, which is 10ms by default. See section 5.3.)
- Packet transmission T=5 ms

The sink network (the one that depacketises the speech) is additionally allocated a delay allowance associated with packet receiving R.

- Packet receiving *R*=5 ms

The sink network contains the dejitter buffer which **shall** be dimensioned to handle the delay variation introduced to the packet flow by itself and all preceding networks (up to a maximum of six ND1704 compliant networks in total). Guidance on dimensioning the dejitter buffer is given in section 5.2.3.1 for adaptive dejitter buffers and 5.2.3.2 for static dejitter buffers.

It should be noted that for a given two-way call path a network will be a *source* for one direction of the call and a *sink* for the other. For "on net" calls the source and sink will be on the same network (this scenario is outside of the scope of this interconnect document).

In order to ensure good voice quality in the UK NGN, CPs are encouraged to procure and deploy equipment which meets or performs better than the above values for delay and delay variation.

5.2.3.1 Connections Using Adaptive Dejitter Buffering

For connections using adaptive dejitter buffers, the buffer will train itself to an optimum size based on the observed IPDV of the packet flow concerned. Adaptive dejitter buffer algorithms vary and might sometimes add a constant delay above the best-case minimum, or might add delay proportional to the spread of the delay variation they are handling. Such a buffer might adapt to apply a minimal additional delay to the maximally delayed packet. Rather than assume "perfect" adaptive dejitter buffer behaviour it is assumed in this document that an adaptive dejitter buffer might typically size itself at 150% of the observed IPDV or less. A future version of this document may make a recommendation regarding the maximum sizing of an adaptive dejitter buffer.

Table 1 shows the delay contributions for a series of connections involving two to six interconnected CPs. Per CP allocations of the fixed components of delay are shown, along with the maximum expected dejitter buffer delay and the end-to-end delay.

Table 1 One-way delay components for connections using adaptive dejitter buffers

No. of CPs in call path *n*	Source network max. fixed delay	Transit network 1 max. fixed delay	Transit network 2 max. fixed delay	Transit network 3 max. fixed delay	Transit network 4 max. fixed delay	Sink network max. fixed delay	Total of network fixed components of delay	Max. expected delay due to dejitter buffer (Note 1)	Total contribution to end-to-end delay (maximum) (Note 2)
	=*P*+*T*+*F*	=*F*	=*F*	=*F*	=*F*	=*R*+*F*	=*P*+*T*+*R*+*nF*	=(*V* x *n*)x1.5	
2	19ms	-	-	-	-	9ms	28ms	21.0ms	49.0ms
3	19ms	4ms	-	-	-	9ms	32ms	31.5ms	63.5ms
4	19ms	4ms	4ms	-	-	9ms	36ms	42.0ms	78.0ms
5	19ms	4ms	4ms	4ms	-	9ms	40ms	52.5ms	92.5ms
6	19ms	4ms	4ms	4ms	4ms	9ms	44ms	63.0ms	107.0ms

Note 1: In addition to its fixed allocation for the fixed component of delay (*F*) each CP is allocated a variable component (*V*) of up to 7ms, which is handled by the dejitter buffer in the sink network. These values assume that dejitter buffer sizes itself at 150% of the observed IPDV.

Note 2: These values assume that a) IPDV accumulates linearly, b) each CP uses its maximum allocation for *F* and *V*, and c) the dejitter buffer sizes itself at 150% of the observed IPDV. In practice many connections are likely to observe total end-to-end delays of less that the illustrated maximum.

5.2.3.2 Connections Using Static Dejitter Buffering

Static dejitter buffers are required for certain types of connection, including those supporting ISDN clearmode and in-band data calls (such as fax and dial-up modems). Except where the call routing is known to involve fewer than six CPs, static dejitter buffers **shall** be dimensioned to allow for the maximum six network IPDV contributions of 7 milliseconds i.e. 42ms. The dejitter buffer really only needs to apply a delay of 42ms to an undelayed packet of the flow, but it has to fix the delay on receiving the first packet, and it cannot determine whether this is an undelayed packet, a maximally delayed packet, or something in between. The best case happens only when the dejitter buffer happens to initialise on an undelayed packet. In the worst case the first packet will be maximally delayed and the static dejitter buffer will still add a further 42ms playout delay.

Table 2 shows the maximum permitted delay contributions for a series of connections involving two to six interconnected CPs. Per CP allocations of the fixed components of delay are shown, along with the maximum permitted dejitter buffer delay and the maximum end-to-end delay to be expected.

Table 2 One-way delay components for connections using static dejitter buffers

No. of CPs in call path *n*	Source network max. fixed delay	Transit network 1 max. fixed delay	Transit network 2 max. fixed delay	Transit network 3 max. fixed delay	Transit network 4 max. fixed delay	Sink network max. fixed delay	Total of network fixed components of delay	Max. expected delay due to dejitter buffer (Note 1)	Total contribution to end-to-end delay (maximum) (Note 2)
	$=P+T+F$	$=F$	$=F$	$=F$	$=F$	$=R+F$	$=P+T+R+nF$	$=(V \times n)+42$	
2	19ms	-	-	-	-	9ms	28ms	56ms	84ms
3	19ms	4ms	-	-	-	9ms	32ms	63ms	95ms
4	19ms	4ms	4ms	-	-	9ms	36ms	70ms	106ms
5	19ms	4ms	4ms	4ms	-	9ms	40ms	77ms	117ms
6	19ms	4ms	4ms	4ms	4ms	9ms	44ms	84ms	128ms

Note 1: In addition to its fixed allocation for the fixed component of delay (F) each CP is allocated a variable component (V) of up to 7ms, which is handled by the dejitter buffer in the sink network. These values assume that initial packet which is seen by dejitter buffer is maximally delayed.

Note 2: These values assume that a) IPDV accumulates linearly, b) each CP uses its maximum allocation for F and V, and c) that initial packet which is seen by dejitter buffer is maximally delayed. In practice many connections are likely to observe total end-to-end delays of less that the illustrated maximum.

5.3 Codecs

Rule C1. The G.711 A-law codec [3], without silence suppression, **shall** be used for interconnecting NGNs providing PSTN/ISDN services [4]. This **shall** also be used as the default interoperability option for other voice services. *Note: Non-PSTN/ISDN connections can be established using other codec pairs, either by bilateral interconnect agreement or end-to-end per-call negotiation. Similarly other codec pairs can be employed within one CP's network or in access networks. CPs need to be aware of the potential impact that codecs other than G.711 can have on perceived speech quality.*

Rule C2. In order that the effects of packetisation on delay can be kept to a minimum, for PSTN/ISDN voice carried over interconnected NGNs a speech frame size of 10ms for G.711 [3] **shall** be used with one speech frame per packet [4].

Rule C3. Silence suppression may adversely affect the perceived quality of a call and **should not** be used except in situations where there is certainty over the routing of calls.

Rule C4. The following add delay and/or coding impairment and **should** be avoided where possible:

- multiple packetisations *Note: The benefit of avoiding multiple packetisations is only gained if networks use the same speech frame size and number of speech frames per packet. See Rule C2 above.*
- transcoding

- asynchronous tandeming[2]

Rule C5. Codec translation within the call media path is undesirable unless it is to enable a call connection that would otherwise be impossible due to incompatible codecs. The Codec Translation Function introduces delay and potential conversion impairment. The recommended national delay objectives for connections as per section 7.3.1 of [2] are applicable to a call subject to codec translation if a section of it has PSTN/ISDN emulation status. For such calls it is recommended that the routing of a call is as determinate as possible to ensure that end to end call delay maxima are not exceeded. A CP **may** introduce a Codec Translation Function in their own fixed or mobile network, or direct calls to a transit network which provides such a function, based on known routing. Where provision of the Codec Translation Function in an applicable fixed network exceeds the delay allocation in section 5.2.3 above, it is recommended that the total delay introduced by the CP's network **should** not exceed a delay objective of 35ms. Exceptionally, this figure **may** be exceeded where the call routing is so determinate that the call will meet the appropriate recommended national end to end delay objective given in [2]. When choosing to employ a Codec Translation Function the speech quality resulting from the delays and choice of codec needs to be considered in order that end-to-end impairments are minimised.

5.4 Packet Loss

Rule PL1. NGN packet loss for PSTN/ISDN traffic **should** essentially be zero. Objectives for end-to-end packet loss for support of PSTN/ISDN traffic will be added to a future issue of this document, but initial studies suggest that a packet loss ratio of around $1x10^{-6}$ per CP will be required for voice-band data. The requirement for voice only traffic will be less stringent.
Note 1: There is a trade-off between packet loss and dejitter buffer delay, with shorter dejitter buffer delays potentially resulting in higher packet loss. This should be borne in mind when optimising these two parameters. Appendix I of ITU-T Recommendation G.113 [5] *gives some guidance on the effect of packet loss on speech quality as estimated by the G.107 E-model* [6].

Note 2: Excessive packet wander/jitter and/or synchronisation differences between the source clock, network clocks and the sink clock can lead to packet loss due to dejitter buffer overruns or under-runs. The contribution of such losses to overall packet loss will be negligible if G.811 clocks are employed in accordance with ND1612. Packet loss due to dejitter buffer overruns or under-runs caused by source or sink CPE clocks that are not the responsibility of the CP should not be counted in the CP's per network limit.

Rule PL2. The support of ISDN clearmode in NGNs requires more stringent packet loss objectives than the support of voice traffic. Objectives for end-to-end packet loss for support of clearmode will be added to a future issue of this document, but initial studies suggest that a packet loss ratio of around $1x10^{-6}$ or $1x10^{-7}$ per CP will be required. The allocation of this objective to individual CP domains is also for further study. *Note: The most stringent packet loss requirements in Y.1541 [7] are for the QoS Classes 6 and 7, which have an end-to-end packet loss objective of* $1x10^{-5}$.

Rule PL3. CPs **should** take account of the fact that they will be unaware of packet loss in preceding networks and therefore packet loss concealment **should** be applied when decoding PSTN/ISDN calls. In NGN the CP terminating the multi-CP IP connection segment **shall** be responsible for PLC. It is possible that another CP's preceding network might introduce a level of packet loss which makes PLC necessary in a very low loss network which terminates a multi-CP NGN connection segment. Any PLC algorithm employed **must** be transparent, i.e. not reliant on any extra data carried in the RTP stream. The use of PLC will slightly increase the fixed delay of the sink network.

5.5 Echo Control

Rule EC1. Echo cancellation **shall** be employed in NGNs as end-to-end delay will be above 25ms. Echo cancellers **should** be collocated with the codec functions.

Rule EC2. Echo cancellers compliant with the requirements of ITU-T Recommendation G.168 [8] **shall** be employed.

Rule EC3. The presence of an echo canceller in a call path **shall** be signalled to adjacent networks in accordance with ITU-T Recommendation Q.115.1 [9]. Ideally all echo cancellers **should** be disabled with the exception of the

[2] The term asynchronous tandeming originally referred to a series connection of speech coders that requires digital to analog conversion followed by re-sampling and re-encoding. Today it also refers to cases where the speech samples must be reconstructed and then re-encoded by the next codec.

two closest to the two potential sources of echo. *Note: Some signalling systems do not enable the presence of echo cancellers in private networks to be signalled to public networks. If G.168 [8] compliant echo cancellers are used in private networks then the potential presence of two echo cancellers working in the same direction of a call path will not degrade speech quality.*

Rule EC4. In general, the tail capacity **shall** be at least 64ms to prevent any potential failure to cancel echo in some routeing scenarios. As an exception to this, the tail capacity may be smaller when the echo path is known to be less than 64ms.

Note that much new detail regarding the use of echo cancellers is provided in ITU-T Recommendation G.108.2 [10].

5.6 Post Dial Delay

NGN calls are based on different technology to that used in the TDM-based PSTN. From a service perspective, it is important that customer perception of connection processing performance parameters such as post dial delay (PDD)[3] in NGN **should** be comparable to that in the PSTN.

5.6.1 Definition of Post Dial Delay

Post Dial Delay (PDD) is experienced by the originating customer as the time from the sending of the final dialled digit to the point at which they hear ring tone or other in-band information. Where the originating network is required to play an announcement before completing the call then this definition of PDD excludes the duration of such announcements.

Each network that forms part of the call path contributes to the overall experience. Although more complex routings are potentially possible, the defined worst-case call set-up involves six network 'stages', of which two involve location-related services (for example: 'Find me anywhere' or mobile telephony). Under this assumption, all calls will naturally involve both an Originating and Terminating stage.

5.6.2 Assumptions

The PDD apportionment model makes the following assumptions:

1. Overlap signalling is not used. In general terms SIP-I 'overlap signalling' is strongly discouraged and to be used only in accordance with NICC overlap signalling best practice guide [11] due to the risks of imposing additional processing load on signalling devices arising from increased messages per call.

2. Normal PSTN operation is assumed, where the return media path is established based upon the forward signalling and ring-tone is generated at the far end. Local ('near end') ring-tone generation is of course technically possible, and it is noted that for example SIP defaults to local application of ring-tone; it is important that operators of NGN consider their SIP configuration, particularly when inter-working with legacy networks, to ensure that callers do in fact receive ring-tone (ideally from only one source).

3. In general a particular network has no knowledge or control over the performance of the other networks involved in delivering a call.

It is appropriate to consider the behaviour of interconnected networks at the network boundaries. It is recognised that there are many possible approaches to the implementation of signalling, however it is expected that the following simplified model will cover the majority of cases (and the impact of other cases is noted for information).

5.6.3 Apportionment Model

The apportionment model is that, as call set up progresses, there are clearly defined handovers of call setup control in each of the forward and backwards directions at each network boundary.

The Forward PDD element of a network **must** be viewed as the sum of delays for processing all messaging up to the point the far end is able to alert the calling party. The Backward PDD element **must** be viewed as the sum of delays for

[3] Prior to this document there were no UK standards for PDD. This section contains a considerable amount of detail as it is necessary to define a model for the measurement of PDD in NGNs before per CP network objectives can be set.

processing all messages up to the point that the calling party receives the network response. The total PDD contribution of a network is the sum of all Forward and Backward elements.

The target forward and backward contributions for each network are based upon the functions that each network is expected to perform in a given call path and are identified as the time difference between Input and Output signalling events. It is assumed that optimal routing is used. The recognised functions are shown in Tables 3 and 4.

Table 3 PDD functions associated with the Forward path

Network	Input Event	Output Event	Network Functions
Originating	End of customer dialling	IAM/Invite to next network	Authentication; Routeing
Transit	IAM/Invite from preceding network	IAM/Invite to next network	Routeing (simple) Note: 'simple' means that the destination was known when the call was passed to the transit network
Terminating	IAM/Invite from preceding network	Alerting destination	Identify customer; internal routeing; alerting.
Mobility	Service trigger	Instruction to serving network	Locate mobile terminal
Number Translation Service	Service trigger	True Destination	Number translation services

Table 4 PDD functions associated with the Backwards path

Network	Input	Output	Network Functions
Terminating	Alerting destination	Return ring-tone to preceding network	Connect media path; generate ring-tone
Transit	Ring-tone	Ring-tone	Connect media path
Originating	Ring-tone	Ring-tone	Connect media path

5.6.4 Targets

The PDD target is in two parts: the 95th percentile of the PDD distribution (i.e. representing the situation that when meeting the target no more than 5% calls will experience longer than the specified PDD); and the target maximum value that any call **should** experience. This approach recognises that the PDD will depend upon the (variable) instantaneous load.

CPs **may** be expected to demonstrate that they can meet the target 95th percentile across the mix of calls under normal operation; the maximum figure is considered a purely design target and in the event that a particular call exceeds this target there is no requirement to abort the establishment of that call. A CP that exceeds their target allowance will be expected to bring their contribution back to the specified target by appropriate engineering methods (e.g. equipment improvements or architectural changes).

Network designers **shall** take account of the fact that under adverse load conditions (overload) it is vital to manage the network in order to constrain PDD and prevent worsening of the overload due to high customer abandonment rates. Clearly PDD will increase in the face of overload but failure to constrain PDD risks high levels of customer abandonment and repeat attempts, which can actually lead to worsening of the overload by increasing the level of ineffective processing.

Overload is effectively the situation in which the processing load on elements of the network is high enough that it is likely that normal load PDD targets (identified in the Table 5) will be breached, in which case it is reasonable that higher targets (for further study) **should** apply.

Table 5 Normal load PDD targets

Network 'stage'	PDD (95th Percentile) (Sum of forward and backward path contributions)	PDD (Target maximum all calls) (Sum of forward and backward path contributions)
Originating	250ms	500ms
Transit	100ms	200ms
Mobility Service		10s (See note 2)
Number Translation Service		1s
Termination (fixed line)	200ms	400ms

Note 1: Some complex services **may** be out of scope of these targets. Such services will be characterised by being of interest to "sophisticated" end users who will recognise that the service complexity will result in comparatively long set-up times, might be subscription only and might not be "mass market".

Note 2: Whilst it is recognised that mobility services will in general require more time for PDD than Number Translation Services, the additional time in the target is intended only to allow the CP the necessary allowance to locate the destination customer/terminal and is not intended as a general consent to longer PDD without good cause.

History

Document history		
V1.1.1	March 2008	Approved by NICC E2E QoS WG
V1.1.2	September 2008	Editorial corrections approved by NICC TSG