Digicel

Digicel (PNG) Limited
Section 494, Lot 1&2,
Kennedy Road, Gordons
PO Box 1618
Port Moresby
Papua New Guinea
www.digicelgroup.com
(675) 7222 2222

15 October 2019

Mr. Charles Punaha
Chief Executive Officer
NICTA
Frangipani Street, Hohola
PO Box 8444
Boroko
National Capital District]

By Email & By Hand

Digicel's Response to NICTA's Consultation Paper on "Draft Rule on Telecommunications Quality of Service Performance Monitoring"

Thank you for the opportunity to provide comments in relation to the consultation described above. Digicel's submission is enclosed with this letter.

We trust that you find Digicel's comments helpful and would welcome any further opportunities to discuss our views.

Yours sincerely
Digicel (PNG) Limited

Michael Henao
Head of Legal & Regulatory

DIGICEL (PNG) LIMITED

Submission to NICTA

Consultation Paper on "Draft Rule on Telecommunications Quality of Service Performance Monitoring"

Tuesday, 15th October 2019

Contents

Executive Summary

Digicel welcomes this opportunity to respond to the National Information and Communications Technology Authority's ("NICTA") public consultation on "Draft Rule on Telecommunications Quality of Service Performance Monitoring". As the leading provider of telecommunication services in Papua New Guinea, Digicel welcomes this opportunity to share its global experience of system performance with NICTA.

Digicel has provided in this document comprehensive comments and feedback on the NICTA proposals contained in the QOS consultation and can summarise these as follows:

Availability:

Availability is a very important parameter and Digicel fully supports the collation and reporting of this parameter. However, the unique and challenging environment in Papua New Guinea in which networks operate must be accommodated, and Digicel respectfully requests that Availability performance is reported separately for Urban, Semi Urban and Rural areas in Papua New Guinea as the environment and challenges listed can seriously affect availability performance.

Call Set-Up time:

Call set up time is a very important mobile network parameter and in a mobile network this can be determined by a number of factors, some of which are outside the control of the operator. In order to capture these aspects, Digicel recommends that call set up time data is separately collected for On-Net M2M Calls, Off-Net MTF calls, Off-Net M2M calls and Off-Net Mobile to international calls. In addition, fixed networks should also report CSSR performance.

Speech Quality:

There does not appear to be any further information in the consultation document on this topic and any recording or reporting requirements. Based on experience elsewhere, Digicel would guide NICTA towards the ITU work on speech quality assessment and in particular the topics of Perceptual Objective Listening Quality Analysis ("POLQA") and associated Mean Opinion Score ("MOS") work.

Reliability:

Reliability - often confused with network quality - is a very important measure of quality of service on a mobile network and Digicel welcomes NICTA's approach to collecting information on unsuccessful calls (CSSR) and dropped calls. Digicel has reviewed the proposed details in attachment B and is of the view that this approach is over-complicated with measurements of successive and non-successive calls. Instead, a simple call failure rate, as described in the text in the table on page 9 under 2 should be adopted and based on network stats. Digicel would also suggest that the 5% threshold is discontinued with a reporting of the actual CSSR being made. The proposed methodology for recording dropped call information in the table on page 10 is satisfactory. However, the actual dropped call rate should be reported and not just in the case where it is in excess of 3%.

Fixed Telephony:

Fixed communication networks in Papua New Guinea, in particular the Telikom PSTN network, should be included and availability, call set up times, voice quality and CSSR captured and reported.

Broadband Internet Access:

Digicel notes that broadband internet access is divided into mobile and fixed. In some cases, the same assessment criteria are used for both networks whereas different criteria are used in other cases. Given the broadband scenarios in Papua New Guinea, Digicel believes that the fixed and mobile distinctions should be removed and, instead, all networks assessed in a similar manner.

Several assessment criteria are listed in Figure 1 but there does not appear to be any further detail in the consultation document as to how to assess or report these criteria. Attachment C contains details on how to report on issues connected with the delivery of broadband services, but apart from availability these do not cross link to the contents of Figure 1. Having assessed the proposed criteria in Figure 1 and based on experience in other markets, Digicel proposes a possible broadband assessment regime be adopted in Papua New Guinea. It is based on similar schemes used by operators and regulators elsewhere.

This broadband testing regime will provide an objective measure for the following KPIs:

- FTP download success rate.
- Average downlink throughput.
- Average Uplink throughput.
- FTP upload success rate.
- Download success rate (DLSR).
- Average download time of web pages.
- Network Latency.

Details on the definitions of these criteria, collection mechanisms, reporting criteria are contained in this Digicel Submission.

1. Introduction

Digicel welcomes this opportunity to respond to the NICTA public consultation on "Draft Rule on Telecommunications Quality of Service Performance Monitoring". As the leading provider of telecommunication services in Papua New Guinea, Digicel welcomes this opportunity to share its global experience of system performance with NICTA.

In the early stages of the deployment of a telecommunications network the emphasis on performance monitoring will very much concentrate on roll-out, speed of deployment, signal strength on the ground as detected by mobiles and other light touch resilience and availability requirements. However as networks mature, as is the case with the mobile phone networks in Papua New Guinea, the emphasis on performance monitoring shifts to service performance including speed of services, setup and dropped call performance as well as tighter requirements on resilience and performance.

It is important that any new measures are well-defined, measurable in a cost-effective way, can be used to inform customers and can also support NICTA's requirements to understand network performance.

2. Comments on NICTA's Proposed QOS Focus

Digicel welcomes NICTA's proposals to collect performance information on both Mobile networks as well as Broadband networks. While there is considerable current discussion and debate around the continued relevance of fixed or legacy PSTN networks, Digicel is of the view that certainly for the next 5-10 years in Papua New Guinea, NICTA should also collect performance data on the Telikom Fixed or PSTN network. The proposed headings for Mobile telephony could also be applied in most cases to a fixed network.

NICTA proposes to adopt the QOS criteria as set out in Figure 1. Digicel would like to comment on these proposed criteria as follows:

2.1 Mobile Telephony

(1) ***Availability***

There does not appear to be any further information on the definition of availability in the consultation document. Given this, it is difficult to understand what NICTA intends with regards availability.

This could be availability as perceived by the end user or subscriber so it cannot be collected by the network. Usually such data is collected via a rigorous drive trial program and it is relatively easy then to collect data on when the network was available and mobiles could select it. In the absence of drive trials which are not mentioned in the consultation document, any network-based availability data would be confined to availability of network nodes. This would record any outages of switching platforms, billing platforms, HLR/HSS, BSC, RNC and so forth.

Digicel requests further details from NICTA on availability assessment and performance recoding and collection, perhaps as an addendum to the consultation, so that it can comment further.

Similar issues determining availability would apply to fixed networks where as far as customers are concerned, failure to obtain a dial tone would be considered as unavailable but this is not collectable by the network.

In devising and implementing a network availability recording/reporting regime, NICTA should make provision and allowance for the unique and challenging environment in which networks operate in Papua New Guinea.

The operational environment in Papua New Guinea is a very adverse environment which can and will impact availability performance due to factors such as:

- With the majority of sites being off-grid and supplied by generator, power supply issues can arise;
- With requirements to operate a network where the majority of infrastructure is off-grid, this places extreme strain on logistic channels such as refuelling requirements and maintenance visits. Roads are often washed out or otherwise interrupted which can mean sites stop working or have to operate on reduced capacity. Indeed, Digicel supports some key parts of its network with airborne refuelling and these can be easily affected by weather which is a common occurrence in Papua New Guinea;
- As NICTA will understand, Digicel experiences levels of vandalism to its infrastructure, more so in rural and remote locations, and this includes the theft of equipment such as generators, damage to towers, theft of fuel and the like. These all impact network availability; and
- 3rd party suppliers, many supporting Digicel operations in rural and remote locations, can fail to deliver or maintain services with excessive outages while awaiting repair and as a consequence, Digicel's availability in these remote and rural locations can be impacted.

Given the real potential for the unique and challenging local conditions and environment to impact on network availability in Papua New Guinea, Digicel believes that network availability performance should be collated on three separate basis: Urban, Semi Urban and Rural.

This will allow easy recognition, justification and understanding of the impact of outside and adverse issues such as those mentioned above on network availability.

Indeed, page 1 of the consultation document outlines the current requirements and obligations re QOS. The separate requirements under C and D for mobile operators for main centres, mid-sized centres etc demonstrates that NICTA is very aware of the challenging issues facing operators in Papua New Guinea and operates relaxed performance criteria in more challenging environments. This recognition of the challenges in providing services across the whole of Papua New Guinea should be continued and enhanced in any new QOS reporting schemes.

(2) ***Call Set-Up Time***

Call set up time is a very important mobile network parameter and in a mobile network this can be determined by a number of factors, some of which are outside the control of the

operator. The mobile operator is and should be responsible for all on-net calls. These are calls where both the A and B parties are subscribers on that mobile network and are receiving service from their home network at the time of the call. However, mobile networks also support calls to off-net subscribers and these can be mobile to national mobile calls, mobile to national fixed calls as well as mobile to international calls. In the case of these third party connected or interconnected networks, the performance of these networks in supporting call setups is not under the control of Digicel. Indeed, with international calls, there can be many diverse routes with calls, for many reasons, not necessarily taking the quickest or fastest route.

Digicel believes that NICTA should collect call setup time performance data but in addition to overall or total call performance data, this data should be collected on the following sub basis too.

- On-Net Digicel to Digicel Calls
- Off-Net Digicel to Telikom fixed calls
- Off-Net Digicel to Bmobile calls
- Off-Net Digicel to international calls

It is vitally important, given NICTA's statement on page 2 that in the event of poor performance it may introduce mandatory performance standards, to have this breakdown so that networks can be properly compared and it can very quickly be identified where any issues with call set up may be.

Digicel believes that this data should also be collected, with similar subdivisions, for the Telikom Fixed network.

Apart from this enhanced collection and reporting granularity, Digicel has no issues with the proposed collection methodology and recording process.

(3) ***Speech Quality***

There does not appear to be any further information in the consultation document on this topic and any recording or reporting requirements. Simple reference is made to using a sample basis but is this achieved through human assessment of Speech quality or some automated measurement approach?

Speech quality and recording its performance is very difficult from the network side as ultimately its down to what the user or caller perceives. Instead, such performance is normally measured and performance determined using drive trials or network trials.

Based on experience elsewhere, Digicel would guide NICTA towards the ITU work on speech quality assessment and in particular the topics of Perceptual Objective Listening Quality Analysis ("POLQA") and associated Mean Opinion Score ("MOS") work. The derivation of this speech quality data is covered in ITU T recommendations P862 and the later version P863. Many automatic field-based survey tools used for automatically recording and assessing speech quality use these approaches to providing speech quality assessment and in particular comparing competing networks.

Digicel has attached the relevant ITU recommendations to this response.

(4) ***Reliability***

Reliability - often confused with network quality - is a very important measure of quality of service on a mobile network and Digicel welcomes NICTA's approach to collecting information on unsuccessful calls and dropped calls.

Unsuccessful calls, often referred to as Call Setup Success Rate or CSSR, shows how successful a caller is in establishing a call. There are many reasons for call failures ranging from incorrect dialling, air interface congestion, core network congestion, Interconnect route congestion, unavailability of called parties etc. Some of these events will be detectable within the network while others such as air interface congestion can be missed as the network may not have detected the mobile. For this reason, CSSR is often obtained from a mix of network statistics and drive trial results.

Digicel has reviewed the proposed details in attachment B and sugests that this approach is over complicated with measurements of successive and non-successive calls. Instead, a simple call failure rate, as described in the text in the table on page 9 under 2 should be adopted and based on network stats. This could then perhaps be expanded at a later time to include call failures detected from drive trials. While these drive trials can be conducted by individual operators, ideally such drive trials would be conducted by NICTA or an appointed independent contractor on a multi-operator basis, across all networks. NICTA's CSSR data could be compared to the results from individual operators to determine any missed data.

Digicel would also suggest that the 5% threshold is discontinued with a reporting of the actual CSSR being made.

Dropped calls are another important indicator of quality and can be the cause of considerable annoyance and frustration with subscribers. There is nothing worse than being cut off mid-call with a friend, colleague or relative and having to call back. These dropped calls can happen for a number of reasons including network congestion, loss of signal, failed call handovers, base station settings and so forth.

The proposed methodology in the table on page 10 is satisfactory. However, the actual dropped call rate should be reported and not just in the case where it is in excess of 3%.

2.2 Fixed Telephony

Digicel has already highlighted the omission of fixed telephony from these proposed Quality of Service reporting requirements and believes this is a serious omission by NICTA.

Fixed communication networks in Papua New Guinea, in particular the Telikom PSTN network, should be included and availability, call set up times, voice quality and CSSR captured, and reported.

2.3 Broadband Internet Access

Digicel notes that broadband internet access is divided into mobile and fixed. In some cases, the same assessment criteria are used for both networks whereas different criteria are used in other cases.

Given the broadband scenarios in Papua New Guinea, Digicel believes that the fixed and mobile distinctions should be removed and instead, all networks assessed in a similar manner. Mobile broadband is far more widespread in Papua New Guinea while the performance of

fixed broadband is oftentimes hindered by legacy equipment and networks. The contents of attachment C also make no fixed-mobile distinction.

Several assessment criteria are mentioned/listed in Figure 1 but there does not appear to be any further detail in the document as to how to assess or report these criteria. Attachment C does contain details on how to report on issues connected with the delivery of broadband services but apart from availability, these do not cross link to the contents of Figure 1.

Having assessed the proposed criteria in Figure 1 and based on experience in other markets, Digicel proposes the following broadband assessment regime as a possible broadband assessment regime to be adopted in Papua New Guinea. It is based on similar schemes used by operators and regulators elsewhere.

(1) ***Possible Broadband Assessment Regime***

Based on extensive world-wide experience, Digicel believes that the majority of data use is stationary when people are sitting down in restaurants, food courts, receptions, lounges, at home or in the office and so forth. While there is some data use on the move, this is by far in the minority. As a result, it would make better sense for these measurements and assessments to be made predominately in static locations with some mobile use.

This broadband testing will provide an objective measure for the following KPIs:

- **FTP download success rate.**

  The probability of completing an FTP download. All samples are used to calculate. The KPI is calculated using the inverse of the ETSI FTP Session Failure Ratio KPI.

  Reference [1]

- **Average downlink throughput.**

  This KPI is the arithmetic mean of the ETSI Mean File Throughput (kbps) KPI. The average throughput achieved whilst downloading a predefined file using FTP.

  Reference [2]

- **Average Uplink throughput.**

  This KPI is the arithmetic mean of the ETSI Mean File Throughput (kbps) KPI. The average throughput achieved whilst uploading a predefined file using FTP.

  Reference [2]

- **FTP upload success rate.**

  The probability of completing an FTP upload. All samples are used to calculate. The KPI is calculated using the inverse of the ETSI FTP Session Failure Ratio KPI.

  Reference [1]

- **Download success rate (DLSR).**

  The probability of completing an http (web page) download. All samples are used to calculate. This KPI is the inverse of the ETSI HTTP Session Failure Ratio [%] KPI.

  Reference [3]

- **Average download time of web pages.**

  How long does it take to down load a number of web pages. Calculated from successful samples only. This is the ETSI HTTP Session Times [s] KPI.

  Reference [4]

- **Network Latency.**

  This will measure the delays in the network and will be achieved by pinging a given in-country IP address. Each operator can report on the average delay for all ping attempts. (total ping time/# pings)

  These should be delivered with sufficient accuracy to enable all the network operators to be compared.

(2) ***Data Collection***

Testing should be:

- Performed statically either on foot or in the car.
- Take place between typical network hours in Papua New Guinea. Testing should also take place on both weekdays and weekends.
- The latest or fastest current USB modems/handsets should be used for each operator.
- The TCP window size for each operator should be as their standard modem settings.
- Automated test equipment could be used to ensure that logging capability is delivered for validation purposes.
- Testing over a reasonable time window should take place at each location. Times such as 30-60 minutes at each location should be considered to mimic typical user behaviour.
- GPS positioning and or Cell information is required to validate test location. These can be written into the log file.
- The Test server for FTP throughput tests should be an in-country hosted server that requires internet access to carry out data sessions. Perhaps this could be NICTA hosted and available to the Operators for testing purposes.
- Ideally these tests are carried out on All operators simultaneously but can be Operator conducted too.

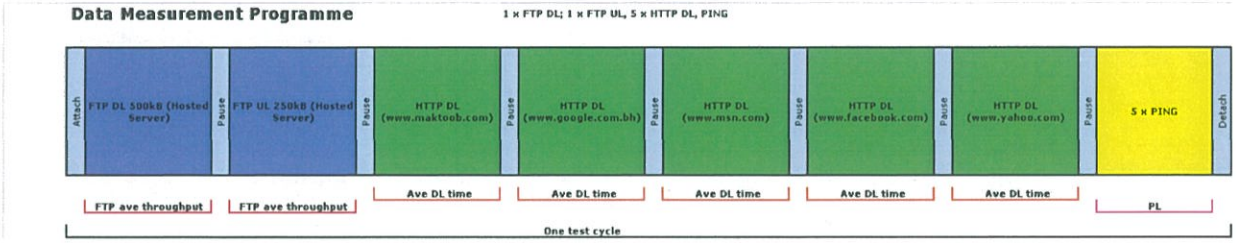
(3) ***Measurement locations***

As previously mentioned, a number of test and measurement locations would be chosen. These would be in locations such as retail locations, malls, airports, cinemas, stadiums etc. It is also important that the locations are all over Papua New Guinea including regional and provincial centres.

Perhaps some thought could be given by NICTA to drawing-up a list of approved or accepted measurement locations for all operators to use. In this way, NICTA could then compare network performance at the same location for each operator.

(4) *Measurement Programme*

The following cycle could be tested by operators or for all operators. This is repeated until 30-60 minutes testing has been completed. This test cycle aims to mirror a user's behaviour so there is a single network attach and PDP activation before the data sessions. There is no requirement to deactivate and detach from the network between test locations.



A 500kB DL FTP file size and a 250kB UL FTP file size gives an acceptable trade-off between ensuring that a true indication of likely throughput can be achieved while minimising the overall impact of the TCP/IP flow control operations. This is regardless of how operators have configured their radio resource management features and ensuring a suitable number of samples can be achieved. Although it is known that file size impacts the overall maximum achievable FTP throughput rate, actual file sizes are not considered to significantly impact relative operator performance. However, if very large file sizes are used then sample size could be significantly reduced due to a single congestion event that may disproportionately impact the overall comparison.

In order to properly test HTTP performance a number of websites should be chosen to support the HTTP testing and these sites should represent typical current web browsing behaviours in Papua New Guinea. Examples might include:

- www.google.com
- www.msn.com
- www.facebook.com
- www.yahoo.com

For the ping and FTP testing, it is important to limit the effects of any outside influences on the testing. Such influences would include any delays accessing overseas FTP servers. As a result, best practice would be to establish/adopt an in-country server presence to ensure that all in-country testing is to an in-country server.

(5) *Reporting Dashboard*

A typical dashboard, such as that below could be used for all reports. Dashboards would be prepared for each test location, and a dashboard showing the overall result for all the test locations combined together.

		All web browsing									Downlink FTP Throughput								
Location	Operator	Samples	Successful samples	Average DL Time (mS)	Operator Ranking (DL Time)	DLSR (%)	Operator Ranking (DLSR%)	DLSR % Upper	DLSR % Lower	DLSR Sample Error %	Samples	Successful	Ave Mean Throughtput (kbps)	Operator Ranking Speed	Success (%)	Operator Ranking Success	FTP Success % Upper	FTP Success % Lower	FTP Success Sample Error %
Area 51	Operator 1	5276	5107	2566	2	96.80	3	97.28	96.32	0.48271	3238	3210	1784	1	99.14	1	99.45	98.82	0.31891
Area 51	Operator 2	4639	4503	2225	1	97.07	2	97.56	96.58	0.49259	2908	2826	1401	2	97.18	2	97.78	96.58	0.60167
Area 51	Operator 3	3201	3167	3044	3	98.94	1	99.30	98.58	0.35667	2009	1938	601	3	96.47	3	97.27	95.66	0.80741

Uplink FTP Throughput									Ping			www.google.com.				www.msn.com				www.facebook.com				www.yahoo.com			
Samples	Successful	Ave Mean Throughtput (kbps)	Operator Ranking Speed	Success (%)	Operator Ranking Success	FTP Success % Upper	FTP Success % Lower	FTP Success Sample Error %	Packet Loss % (URL)	Packet Loss % (IP)	Latency 90th Percentile mS	Samples	Successful samples	Average DL Time (mS)	Success (%)	Samples	Successful samples	Average DL Time (mS)	Success (%)	Samples	Successful samples	Average DL Time (mS)	Success (%)	Samples	Successful samples	Average DL Time (mS)	Success (%)
3238	3210	1784	1	99.14	1	99.45	98.82	0.31891	14.9	11.0	675	1081	1051	2281	97.2	1044	1017	2617	97.4	1035	986	3358	95.3	1045	1016	2438	97.2
2908	2826	1401	2	97.18	2	97.78	96.58	0.60167	0.5	0.4	453	948	927	1133	97.8	918	884	2826	96.3	919	887	2497	96.5	914	896	2206	98.0
2009	1938	601	3	96.47	3	97.27	95.66	0.80741	0.0	0.0	387	654	653	1316	99.8	631	627	4424	99.4	626	625	2759	99.8	620	614	2730	99.0

(6) *Metric Definitions*

1. Average download time (mS)
    a. This is for the web page download only i.e. the time taken from a user hitting the enter button to the point where the web page has completed downloading following inserting a URL.
    b. Web pages consist of a number of files; a successful download is defined as the entire page loading i.e. all files are successfully downloaded.
2. Downlink success rate (DLSR %)
    a. Successful downloads divided by attempted downloads.
    b. Web pages consist of a number of files; a successful download is defined as the entire page loading i.e. all files are successfully downloaded.
3. Average Mean Throughput (kbps)
    a. This is calculated for successful uplink and downlink throughputs.
    b. Success is where the entire file has been up or downloaded.
    c. The average throughput for each sample is provided by the test tool.
4. Success (%)
    Total successful uploads or downloads divided by total attempts.

5. Network Latency (mS)
    Based on simple ping test to determine round trip time to various IP addresses.

(7) *References*

[1] ETSI TS 102 250-2 v1.6.2 (2008-09). Section 6.1.5 FTP {Download/Upload} Session Failure Ratio [%]. Page 40.

[2] ETSI TS 102 250-2 v1.6.2 (2008-09). Section 6.1.7 FTP {Download/Upload} Mean Data Rate [kbit/s]. Page 42

[3] ETSI TS 102 250-2 v1.6.2 (2008-09). Section 6.8.5 HTTP Session Failure Ratio [%]. Page 120.

[4] ETSI TS 102 250-2 v1.6.2 (2008-09). Section 6.8.6 HTTP Session Times [s]. Page 118.

International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.862

Amendment 2
(11/2005)

SERIES P: TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Methods for objective and subjective assessment of quality

Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs

Amendment 2: Revised Annex A – Reference implementations and conformance testing for ITU-T Recs P.862, P.862.1 and P.862.2

ITU-T Recommendation P.862 (2001) – Amendment 2

ITU-T

International Telecommunication Union

ITU-T P-SERIES RECOMMENDATIONS

TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Subscribers' lines and sets	Series	P.30
		P.300
Transmission standards	Series	P.40
Objective measuring apparatus	Series	P.50
		P.500
Objective electro-acoustical measurements	Series	P.60
Measurements related to speech loudness	Series	P.70
Methods for objective and subjective assessment of quality	**Series**	**P.80**
		P.800
Audiovisual quality in multimedia services	Series	P.900
Transmission performance and QoS aspects of IP end-points	Series	P.1000

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

ITU-T Recommendation P.862

Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs

Amendment 2

Revised Annex A – Reference implementations and conformance testing for ITU-T Recs P.862, P.862.1 and P.862.2

Summary

This revision to Annex A/P.862 describes the revised ANSI C reference implementation of PESQ, which includes modes of operation for Recommendation P.862.1 (narrowband MOS-LQO mapping) and Recommendation P.862.2 (the wideband extension). It also describes the conformance testing procedures for P.862 and P.862.2. This revision replaces Annex A to P.862 (February 2001) and Amendment 1 to P.862 (March 2003).

Source

Amendment 2 to ITU-T Recommendation P.862 (2001) was approved on 29 November 2005 by ITU-T Study Group 12 (2005-2008) under the ITU-T Recommendation A.8 procedure.

FOREWORD

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

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

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

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

NOTE

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

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

INTELLECTUAL PROPERTY RIGHTS

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

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

CONTENTS

ITU-T Recommendation P.862

Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs

Amendment 2

Revised Annex A – Reference implementations and conformance testing for ITU-T Recs P.862, P.862.1 and P.862.2

A.1 Files provided

A.1.1 List of files provided for the ANSI-C reference implementation

The ANSI-C reference implementation of ITU-T Recs P.862, P.862.1 and P.862.2 is contained in the following text files which are provided in the `source` sub-directory of the CD-ROM distribution:

- dsp.c *Basic DSP routines*
- dsp.h *Header file for dsp.c*
- pesq.h *General header file*
- pesqdsp.c *PESQ DSP routines*
- pesqio.c *File input/output*
- pesqmain.c *Main program*
- pesqmod.c *PESQ high-level model*
- pesqpar.h *PESQ perceptual model definitions*

The ANSI-C reference implementation is provided in separate files and forms an integral part of this Recommendation and ITU-T Recs P.862.1 and P.862.2. The ANSI-C reference implementation shall take precedence in case of conflicts between the high-level description and the ANSI-C reference implementation.

A.1.2 List of files provided for conformance validation

The conformance validation process described below makes reference to the following files, which are provided in the `conform` sub-directory of the CD-ROM distribution:

- supp23_16k.txt *File pairs and PESQ scores for test 1(a)*
- supp23_16k.bat *Batch script to assist with test 1(a)*
- supp23_8k.txt *File pairs and PESQ scores for test 1(b)*
- supp23_8k.bat *Batch script to assist with test 1(b)*
- voipref_16k.txt *File pairs and PESQ scores for test 2(a)*
- voipref_16k.bat *Batch script to assist with test 2(a)*
- voipref_8k.txt *File pairs and PESQ scores for test 2(b)*
- voipref_8k.bat *Batch script to assist with test 2(b)*
- process.bat *Sample batch script to assist with preparing material for tests 1(b) and 2(a)*
- supp23_wb.txt *File pairs and P.862.2 scores for conformance validation to Supplement 23 (wideband operation)*

- supp23_wb.bat *Batch script to assist with validation of P.862.2 scores*
- *Speech files provided for validation of P.862 with variable delay:*

or105.wav	or109.wav	or114.wav	or129.wav	or134.wav	or137.wav
or145.wav	or149.wav	or152.wav	or154.wav	or155.wav	or161.wav
or164.wav	or166.wav	or170.wav	or179.wav	or221.wav	or229.wav
or246.wav	or272.wav	dg105.wav	dg109.wav	dg114.wav	dg129.wav
dg134.wav	dg137.wav	dg145.wav	dg149.wav	dg152.wav	dg154.wav
dg155.wav	dg161.wav	dg164.wav	dg166.wav	dg170.wav	dg179.wav
dg221.wav	dg229.wav	dg246.wav	dg272.wav		

u_am1s01.wav	u_am1s02.wav	u_am1s03.wav	
u_am1s01b1c1.wav	u_am1s01b1c7.wav	u_am1s01b1c15.wav	u_am1s02b1c9.wav
u_am1s03b1c16.wav	u_am1s03b1c18.wav	u_am1s01b2c1.wav	u_am1s01b2c8.wav
u_am1s02b2c4.wav	u_am1s02b2c5.wav	u_am1s02b2c14.wav	u_am1s03b2c5.wav
u_am1s03b2c6.wav	u_am1s03b2c7.wav	u_am1s03b2c11.wav	u_am1s03b2c18.wav
u_af1s01.wav	u_af1s02.wav	u_af1s03.wav	
u_af1s01b2c16.wav	u_af1s03b2c16.wav	u_af1s02b2c17.wav	u_af1s03b2c17.wav

The variable delay speech files are in Wave format (16-bit linear PCM, Intel byte ordering, 44 byte header), at 8 kHz sample rate.

These files form an integral part of this annex.

A.2 Sampling rate

An implementation of ITU-T Rec. P.862 may, at the implementer's discretion, operate at 8 kHz sampling rate, 16 kHz sampling rate, or both. However, the implementation must pass the conformance tests set for all sampling rates that are offered by the implementation.

Implementations of the P.862.2 wideband extension must operate at 16kHz sampling rate.

A.3 Conformance tests

A.3.1 Conformance data sets

The data sets for the conformance tests are as follows.

Test	Number of file pairs	(a) 16 kHz data set	(b) 8 kHz data set	Type of test
1	1736	ITU-T P-series Supplement 23	Downsampled from ITU-T P-series Supplement 23 using ITU-T Software Tool Library (version 2000, release 3[1]) and `process.bat`.	Mandatory
2	40	Upsampled from P.862 VoIP variable delay data using Software Tool Library (version 2000, release 3) and `process.bat`.	ITU-T Rec. P.862 Annex A VoIP variable delay data.	Mandatory
3	No data set defined. This test is open-ended, based on general, unknown data.			Mandatory
4	1736	ITU-T P-series Supplement 23	Not applicable	Mandatory for P.862.2

[1] ITU-T Rec. G.191 (2005), *Software tools for speech and audio coding standardization.*

A.3.2 Conformance requirements

The test requirements are summarized in the following table and are set out in detail below. The requirements are based on the absolute difference in PESQ score between the implementation under test and the ANSI-C reference implementation, calculated for each reference and degraded file pair. For the conformance tests defined in Annex A/P.862 (February 2001), there is no change.

Test	Number of file pairs	Lower threshold	Upper threshold	Type of test
1(a)	1736	Difference may not exceed 0.05 in any situation.	Not applicable	Mandatory
1(b)	1736	Difference may exceed 0.05 in not more than 2 file pairs (approx. 0.1% of cases).	Difference may not exceed 0.1 in any case.	Mandatory
2(a)	40	Difference may exceed 0.05 in not more than 1 file pair (2.5% of cases).	Difference may not exceed 0.5 in any case.	Mandatory
2(b)	40	Difference may exceed 0.05 in not more than 1 file pair (2.5% of cases).	Difference may not exceed 0.5 in any case.	Mandatory
3	No data set defined	Difference may exceed 0.05 in not more than 0.5% of cases.	Difference may exceed 0.05 in not more than 5% of cases.	Lower threshold is advisory. Upper threshold is mandatory.
4	1736	Difference may not exceed 0.05 in any situation.	Not applicable	Mandatory for P.862.2 operation

A.3.2.1(a) Conformance test 1(a) (16 kHz sampling rate)

In this test, all files from all ten experiments as released with ITU-T P-series Supplement 23 are used, on a file-by-file basis. The Supplement 23 data is all at 16 kHz sampling rate. The original and degraded file names, and the PESQ score given by the reference implementation, are provided in the files listed above.

An implementation passes this test when the absolute difference in the raw PESQ score compared to the reference implementation is not greater than 0.05 in all cases.

This conformance test is mandatory for all implementations of PESQ at 16 kHz sampling rate.

ITU-T P-series Suppplement 23 can be obtained separately from the ITU.

A.3.2.1(b) Conformance test 1(b) (8 kHz sampling rate)

In this test, 8 kHz resampled versions of the Supplement 23 files are used, on a file-by-file basis. The original and degraded files must be downsampled using the ITU-T Software Tool Library 2000 release 3, program `filter`, using the following command:

```
filter -down HQ2 inputfile.raw outputfile.raw
```

This assumes that the 16 kHz input speech file is called `inputfile.raw` and the 8 kHz output file is called `outputfile.raw`.

A batch script to assist with this, and the original and degraded file names, and the raw PESQ score given by the reference implementation, are provided in the files listed above.

An implementation passes this test when the absolute difference in the raw PESQ score compared to the reference implementation is not greater than 0.05 in more than 2 file pairs (these may be any two of the file pairs), and not greater than 0.1 in all cases.

This conformance test is mandatory for all implementations of PESQ at 8 kHz sampling rate.

ITU-T P-series Suppplement 23 can be obtained separately from the ITU.

A.3.2.2(a) Conformance test 2(a) (16 kHz sampling rate)

This test is based on data provided with PESQ and described in this annex. In this test, 16 kHz resampled versions of the Annex A/P.862 VoIP test files are used on a file-by-file basis. The original and degraded files must be upsampled using the ITU-T Software Tool Library 2000 release 3, program `filter`, using the following command:

```
filter -up HQ2 inputfile.raw outputfile.raw
```

This assumes that the 8 kHz input speech file is called `inputfile.raw` and the 16 kHz output file is called `outputfile.raw`.

A batch script to assist with this, and the original and degraded file names, and the raw PESQ score given by the reference implementation, are provided in the files listed above.

An implementation passes this test when the absolute difference in the raw PESQ score compared to the reference implementation is not greater than 0.05 in more than 1 file pair (this may be any one of the file pairs), and not greater than 0.5 in all cases.

This conformance test is mandatory for all implementations of PESQ at 16 kHz sampling rate.

A.3.2.2(b) Conformance test 2(b) (8 kHz sampling rate)

A composite database was constructed for Annex A/P.862 from 40 conditions (file pairs) from two subjective tests covering real and simulated VoIP connections that exhibit time-varying delay. Many of these file pairs also trigger the bad interval realignment process. This data is provided at 8 kHz sampling rate as the Annex A/P.862 VoIP test files, and these are used on a file-by-file basis.

The original and degraded file names, and the raw PESQ score given by the reference implementation, are provided in the files listed above.

An implementation passes this test when the absolute difference in the raw PESQ score compared to the reference implementation is not greater than 0.05 in more than 1 file pair (this may be any one of the file pairs), and not greater than 0.5 in all cases.

This conformance test is mandatory for all implementations of PESQ at 8 kHz sampling rate.

A.3.2.3 Conformance test 3 (8 kHz or 16 kHz sampling rate) – Additional comparisons

To prevent implementers from specifically tailoring an algorithm to conform to requirements for the files described above, a further test is available. An implementation of PESQ that conforms to ITU-T Rec. P.862 must, in at least 95% of cases, give an output score that is within 0.05 of the raw PESQ score given by the ANSI-C reference implementation. These cases must be based on speech files covering a representative sample of reasonable telephone network conditions, and must lie within the scope of ITU-T Rec. P.862.

In practice it has been found that this is a much wider margin than required for most implementations. Users should expect that, in at least 99.5% of cases, an implementation should give an output score that is within 0.05 of the raw PESQ score given by the ANSI-C reference implementation. This should be considered to be a desirable level of accuracy, but it is not mandatory.

A.3.2.4 Conformance test 4 (16 kHz sampling rate) – P.862.2 operation

This test applies to implementations of the wideband extension to P.862 defined in ITU-T Rec. P.862.2. Wideband operation of the ANSI-C reference code is enabled by using the `+wb` command line option.

In this test, all files from all ten experiments as released with ITU-T P-series Supplement 23 are used, on a file-by-file basis. The Supplement 23 data is all at 16 kHz sampling rate. The original and degraded file names, and the wideband PESQ score given by the reference implementation, are provided in the files listed above.

An implementation passes this test when the absolute difference in the wideband PESQ score compared to the reference implementation is not greater than 0.05 in all cases.

This conformance test is mandatory for all implementations of P.862.2.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	General tariff principles
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	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
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 and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2006