EUROCONTROL

Data link Performance Metric Definitions

Data link Performance Monitoring

Edition: 1.0
Edition date: 03-03-2022
Classification: White
Reference nr: NMD/INF/ICNS/DPM/2022-01

SUPPORTING EUROPEAN AVIATION

FOUNDING MEMBER SESAR JOINT UNDERTAKING

NETWORK MANAGER

DOCUMENT CONTROL

Document Title	Data link Performance Metric Definitions
Document Subtitle	Data link Performance Monitoring
Document Reference	NMD/INF/ICNS/DPM/2022-01
Edition Number	1.0
Edition Validity Date	03-03-2022
Classification	White
Accessibility	Extranet
Status	Released Issue
Author(s)	David Isaac (iCNS) Christophe Visée (iCNS)
Contact Person(s)	David Isaac

EDITION HISTORY

Edition No.	Validity Date	Author(s)	Reason
V0.02		David Isaac	Initial draft for internal review
V0.03	24/03/2021	David Isaac	New draft for internal review. Updated to address discussion with ANSPs on Feb 2021. Includes a lot of contextual content from the previous 'Performance Monitoring Requirements PMR' document and identifies a set of 'essential' metrics as requested by EASA.
V0.04	02/04/2021	David Isaac	Second draft for internal review. Moved most of the discussion sections to appendices.
V0.1	09/04/2021	David Isaac / Christophe Visée	AVLC metrics added. Draft for review by stakeholder group.
V0.2	21/05/2021	David Isaac / Christophe Visée	Updated to take account of comments received against V0.1, from ANSP group and ACSP group.

			Some significant changes: PA rate added and Air Initiated Transaction Continuity removed from the 'essential metrics'. DLIC Initiation Logon counts removed (as not considered useful without complete LOF/NAN information). 'Essential' metrics renames as KPIs. A new metric 'CSR Success rate' is added and considered a KPI. Proposal for measuring availability included. Appendix 3 and 4 removed – to be maintained as a separate document.
V0.3	22/06/2021	David Isaac / Christophe Visée	Updated to address comments from the meeting of the 15th June 2021, primarily about the use of UM183 CURRENT ATC UNIT to mark the start of an active CPDLC session, plus some comments received from ARINC on the 'Additional VDL2 metrics'
V0.4	14/07/2021	David Isaac / Christophe Visée	Updated to address comments from the 30th June 2021 meeting. Primarily related to the definitions for availability.
V0.5	05/08/2021	David Isaac / Christophe Visée	Updated to address final comments from stakeholder group. Minor changes to availability definitions.
V0.6	23/11/2021	David Isaac / Christophe Visée	Updated to address comments from review by all ANSPs/CSPs/EASA.
V0.7	03/02/2022	David Isaac / Christophe Visée	Updated to address comments made during meeting of 1st Dec 2021 and added appendix 3 on confidence interval and compliance assessment.
V0.8	22/02/2022	David Isaac / Christophe Visée	Updated to address comments from the 21st Feb 2022 meeting.
V1.0	03/03/2022	David Isaac / Christophe Visée	First definitive version to be published.

TABLE OF CONTENT

TABLE OF FIGURES

1 Introduction

1.1 Purpose

The purpose of this document is to propose a set of metrics for monitoring the performance of the CPDLC system in Europe as required by EC regulation 29/2009. It provides a rationale for the choice of the metrics, a definition of the metrics, a set of target values as well as a means to assess compliance against these targets for those metrics that are considered necessary to ensure compliance against the performance requirements.

Neither the regulation nor the industry standards dictate how data link performance should be monitored. This document is intended to support an agreement between each ANSP and their national supervisory authority (NSA) as to how to monitor the performance of the system and may form the basis for some future additional guidance from EASA. It is the result of a collaborative work among the data link community participating to the Datalink Performance Monitoring Group (DPMG). It is hoped that by starting from a common proposal a high degree of commonality between the different NSAs can be achieved. It is also intended to provide the European data link community with a common definition for a number of other useful metrics.

The document is not intended to define performance requirements, these are defined by the industry standards. Nor is it intended to define how National Supervisory Authorities (NSAs) should monitor performance, nor how to handle any failure to meet the performance requirements; these are issues to be addressed by the NSAs.

The proposed metrics are divided into two categories: i) Key Performance indicators (KPIs) and ii) different groups of 'Additional' Metrics.

The KPIs are the minimum set considered to be necessary to monitor system performance and are expected to form the basis for an agreement between the ANSP and their NSA. The Additional metrics are other parameters that are considered useful for overall system monitoring, grouped according to subject area.

1.2 Scope of the document

This document is addressing performance of datalink defined by ATN Baseline 1. This document will need to be revised in the future to consider Baseline 2.

1.3 Structure of this document

This document is structured as follows:

- **Chapter 1: Introduction**. This section.
- **Chapter 2: Metrics Definitions.** This chapter provides a detailed description of each metric. The rationale for these metrics is contained in Appendix 1 and Appendix 2.
- **Appendix 1: Rationale for Proposed Metrics.** This appendix describes how the proposal for the metrics was derived from consideration of the required performance.

- **Appendix 2: Derivation of Target Values for KPIs.** This appendix describes which metrics are considered essential and why. It also provides a rationale for the proposed target values for the KPIs.
- **Appendix 3: Confidence Intervals and compliance assessment**. This appendix proposes a method to calculate the confidence interval and how that can be used to demonstrate compliance against performance requirements.
- **Appendix 4: Ideas for future metrics.** This appendix lists some candidate future metrics that have yet to be defined.

1.4 Monitoring Objectives

Article 5, paragraph 6 of EC regulation 29/2009 requires the ANPSs to "…monitor the quality of service of communication services and verify their conformance with the level of performance required for the operational environment under their responsibility". The regulation refers to ED120 [1] as the source of the performance requirements, although this document proposes to use the later EUROCAE standard (ED-228A) as the reference for performance requirements

ED-78a [2] defines the guidelines for the provision and use of services supported by data link and has guided the development of the data link services covered by regulation 29/2009. It defines the purpose of in-service monitoring as being to provide "…creditable operational data to determine that requirements for the CNS/ATM system….continue to be met[1]" and further clarifies that the measurement should be transaction based and not separately measure the performance of individual elements of the system. So it is clear that the purpose of monitoring the system performance is not specifically to measure every requirement from the EUROCAE standards but rather to assure that the system is operating smoothly and achieving its overall performance requirements.

1.5 Document Hierarchy

This document is intended to replace the previous description of proposed parameters and their target values as contained in the 'Link 2000+ DLS CRO Performance Monitoring Requirements' document [3].

This document supplements the 'DPMF Report Catalogue' [4] which defines the types of report that the DPMF will create and their frequency.

[1] See para 2.5 of Ed78a.

2 Metric Definitions

This chapter presents a set of metric definitions. The metrics are divided into two separate categories: i) Key Performance Indicators and ii) Additional Metrics.

The KPIs is the set of metrics that are considered to be necessary to demonstrate the system meets the performance requirements defined in the EUROCAE standards referenced by the EC Regulation 29/2009. Most of these metrics have a target value derived directly from the EUROCAE performance requirements.

All other metrics are referred to as 'Additional Metrics' – these metrics are considered useful for measuring some aspect of performance but do not have target values tied to the requirements and are not considered to be strictly needed to demonstrate compliance. They are presented grouped by subject area: Additional system level metrics, VDL2 metrics and Operational metrics.

2.1 Metric Definition General Considerations

This section provides some general comments that are applicable to all the metrics.

2.1.1 Estimating when the aircraft sends messages.

It is not possible to know from the available ground recordings exactly when a message was received by the flight crew. However for CPDLC the time provided in the header of the LACK message sent from the aircraft is considered as giving a reasonable indication of when the associated uplink message has been processed and is available to the pilot.

Similarly the timestamp in the header of the CPDLC request from the aircraft is considered as giving a reasonable indication of when the pilot made the request.

2.1.2 Measuring when an uplink message is sent.

To determine the time at which an uplink message is sent, the timestamp from header of the CPDLC message should be used rather than the time at which the ground system logs the message.

2.1.3 Measuring metrics per frequency or ACSP

It is useful to be able to calculate some application level metrics per VDL2 frequency or, in the case of shared frequencies, per ACSP. In order to achieve this, some association between application level messages and VDL2 exchanges is required.

Ideally the payload of any VDL data recorded would be parsed to identify the associated application level payload. EUROCONTROL intend to use VDL2 recordings to estimate the time at which an aircraft is transferred from one frequency to another as the means to identify on which frequency/ACSP any particular application level transaction occurred and thereby to calculate the metrics per frequency/ACSP. However other methods may be also be employed to achieve the desired association.

2.1.4 The volume of airspace considered

The DLS regulation applies above FL285, but many ANSPs provide the service starting from lower Flight Levels. It is considered appropriate to measure performance in all airspace where the service is supposed to be provided and not just above FL285.

In most cases the service is provided above a defined flight level in the airspace of a particular centre but some ANSPs/Centres may have a more complicated structure. For some metrics, particularly when being used to assess the performance of the ACSPs, it is preferable to only use the data received from aircraft above the flight level from which the service is offered. When required the DPMF will measure performance above a single flight level defined for each ATC centre based on information obtained from the ANSP, typically from the AIP. This flight level is referred to as the 'Local level of implementation' and the currently used value is available from the DPMF wiki.

It is proposed not to use altitude filtering when measuring the KPIs – these metrics should include all messages and events.

Note: If ANSPs attempt to perform data link with aircraft below the local level of implementation, or more generally outside the volume of airspace for which the service is designed, then clearly it will have an adverse effect on the measured performance and would make the metrics rather meaningless from a compliance point of view.

2.1.5 The display of LACKs in the diagrams of this section

In some diagrams in this chapter, LACKs (shown in blue) have been omitted for clarity. They are included where they are used as a measuring point.

2.1.6 Precision of measurement and performance compliance assessment

It is considered appropriate to calculate confidence intervals for KPI metrics where a sample of data is being used as an estimate of a broader domain and assessed against a target value. Guidance for determining confidence intervals and how these can be used to demonstrate compliance is proposed in Appendix 3.

2.2 KPI Metrics

This is the minimum set of metrics considered necessary and sufficient to measure the overall performance of data link against the key performance requirements. If the system meets the defined targets for these metrics then it is considered to perform acceptably from an end-user perspective.

It is recommended to measure the overall performance on at least a monthly basis but shorter periods may be appropriate.

There are two groups within the KPI metrics: i) a set for which a target value can be directly defined from the EUROCAE standards, and ii) a set intended to cover the availability of the service per aircraft (referred to as Availability(Use) in EUROCAE standards) for which it is not considered possible to directly define a target value from the EUROCAE standard.

The following set of KPI metrics have a target value defined directly from the EUROCAE standard:

- Ground Initiated Transaction Continuity
- RCTP Technical Continuity
- ANSP System Availability
- ACSP System Availability

The KPI metrics below are intended to measure the availability of the service to an individual aircraft and do not have a target defined directly from the EUROCAE standard:

- CPDLC Start Request (CSR) Success Rate
- Active Session Provider Abort Rate

An explanation of the rationale behind the selection of these metrics and their target values is contained in Appendix 1 and Appendix 2.

For each metric a general description and a more precise definition is provided in the following subsections and the target values for the KPIs are given separately in a table in section 2.3.

2.2.1 Ground Initiated Transaction Continuity

2.2.1.1 Description

The Ground Initiated Transaction Continuity (GITC) is a measure of overall usability from the controllers' point of view. It measures the probability of the controller getting an expected operational response with a specified time[2]. For example the controller should receive an expected operational response within 2 minutes for 99.9% of uplinks that they send. An ERROR message received as a response is not considered to be an expected operational response as it is not a desired response from the controllers' point of view. Two values for continuity are calculated:

- Continuity at 120 seconds (the Expiration Time, ET).
- Continuity at 60 seconds.

The ANSP end-system should record all ground initiated requests that require a response and associated closing responses.

2.2.1.2 Definition

GITC(120) = number of ground initiated CPDLC messages requiring a response for which a closing response is received within 120 seconds or less / total number of ground initiated CPDLC requests requiring a response.

GITC(60) = number of ground initiated CPDLC messages requiring a response for which a closing response is received within 60 seconds or less / total number of ground initiated CPDLC requests requiring a response.

The following responses are considered closing responses:

- DM0 WILCO
- DM1 UNABLE
- DM2 STANDBY
- DM3 ROGER
- DM4 AFFIRM
- DM5 NEGATIVE

[2] This time is measured end-to-end and includes the time taken by the crew to consider their response.

- DM32 PRESENT LEVEL
- DM81 WE CAN ACCEPT [level] AT [time]
- DM82 WE CANNOT ACCEPT [level]
- DM106 PREFERRED LEVEL [level]
- DM109 TOP OF DESCENT [time]

2.2.2 RCTP Technical Continuity

2.2.2.1 Description

This is the probability that a LACK or an ERROR message is received for an uplink message within a certain delay. It is used as an approximation of the Required Communications Technical Performance (RCTP) defined in the EUROCAE standards.

The requirements are stated in terms of the probability of receiving a response within a specified time. Two values for continuity are calculated:

- Continuity at 32 seconds
- Continuity at 20 seconds

The ANSP end-system should record all ground initiated messages that require a LACK and the associated LACK or ERROR messages received.

2.2.2.2 Definition

RCTP_TC(32) = number of uplink messages requiring a LACK for which a LACK or an ERROR response is received within 32 seconds or less / total number of uplinks requiring a LACK.

RCTP_TC(20) = number of uplink messages requiring a LACK for which a LACK or an ERROR response is received within 20 seconds or less / total number of uplinks requiring a LACK.

2.2.3 ANSP System Availability

2.2.3.1 Description

Availability is interpreted as meaning the probability that the CPDLC system is in service within a planned service area for planned hours of CPDLC operation i.e. it excludes planned outages of the service. Also unplanned outages that last less than 6 minutes (the 'Unplanned outage duration limit') are not considered to have an impact on system availability; the impact of these short outages is considered to be against the continuity requirements.

The date, time and duration of any planned and unplanned service outages lasting more than 6 minutes should be reported on a monthly basis.

Different system elements should be reported separately and should include: ANSP ATN-router, ANSP ground system, availability of the connection to each CSP and the availability of any interconnecting networks used (e.g. leased lines, gateway, New PENS). Outages of the network connecting the ANSP with the CSPs should be reported under the ANSP availability when they impact a single ANSP and under the ACSP when they impact more than one ANSP.

2.2.4 ACSP System Availability

Description

Availability is interpreted as meaning the probability that the CSP system providing the ATN/VDL2 communication network is in service within a planned service area for planned hours of operation i.e. it excludes planned outages of the service.

The date, time and duration of any planned and unplanned service outages lasting more than 6 minutes, as well as the delay between when the outage began and the problem was reported should be reported on a monthly basis.

Different systems elements should be reported separately and should include: Air-Ground Router, Ground-Ground Router, VGSs that results in a loss of coverage[3] in airspace required to be covered and availability of the connection to each ANSP and the availability of any interconnecting networks used (e.g. leased lines, gateway, New PENS).Outages of the network connecting the ANSP with the CSPs should be reported under the ANSP availability when they impact a single ANSP and under the ACSP when they impact more than one ANSP.

2.2.5 CPDLC Start Request (CSR) Success Rate

2.2.5.1 Description

This metric is intended to measure problems with the establishment of an active CPDLC session. It measures the probability of a CPDLC start request resulting in a UM183 CURRENT ATC UNIT message being sent by the ground system. The UM183 CURRENT ATC UNIT message marks the point at which the ground system considers the CPDLC to be enabled and so is considered the point at which the CPDLC session is 'active'.

2.2.5.2 Definition

CSR Success Rate = No. of UM183 CURRENT ATC UNIT messages sent in response to a CPDLC Start Request / No. of CPDLC start-request messages sent.

N.B. In some cases more than a single UM183 CURRENT ATC UNIT message is sent; only the first such message sent after each CPDLC Start request should be counted.

2.2.6 Active Session Provider Abort Rate

2.2.6.1 Description

This measures the rate at which Provider Aborts are experienced for aircraft with an active CPDLC session. A CPDLC session is considered active once the UM183 CURRENT ATC UNIT message has been sent by the ground system. This is the point from which the controller expects to be able to use CPDLC. The rate is expressed at PAs per 100 hours of CPDLC session.

[3] Coverage is only considered to be lost when there is no VDL2 service available on any frequency in the airspace.

2.2.6.2 Definition

The PA rate would be calculated as the number of PAs experienced during an active session per 100 hours of active CPDLC session i.e.

$$\left(\frac{NumberOfPAsWithinActiveSessions}{TotalDurationOfActiveCPDLCSessionsInHours}\right)x100$$

For example, consider two aircraft, one of which has CPDLC sessions with a total duration of 60 hours and the second with a total duration of 70 hours. One of those aircraft suffered 4 PAs before a CPDLC session was established and 1 PA after the session was active and the second aircraft suffered 2 PAs after the session was active, then the PA rate would be:

$$\left(\frac{1+2}{60+70}\right)x100 = 2.3\ PAs\ per\ 100\ Hours\ CPDLC$$

An 'Active CPDLC session' is considered to start when the UM183 CURRENT ATC UNIT message has been sent.

An 'Active CPDLC session' is considered to end when a CPDLC end-response is received by the ground or a User Abort or Provider Abort is declared. If for some reason none of those events are observed then the last message sent or received from the aircraft should be considered as the end of the session; note this should not occur in practice

2.3 KPI Metrics target Values

The following targets have been taken directly from the performance requirements defined in Table E-2 and Table 5-14 of the EUROCAE ED-228A standard, as described in Appendix 2.

Definition Para	Metric	Target Value	Min. Compliance sample size[4]
2.2.1	**Ground Initiated Transaction Continuity**		
	Probability of operational response within 60 seconds	>= 0.95	200
	Probability of operational response within 120 seconds	>= 0.999	10000
2.2.2	**RCTP Technical Continuity**		
	Probability of LACK/ERROR reception within 20 seconds	>= 0.95	200
	Probability of LACK/ERROR reception within 32 seconds	>= 0.999	10000
2.2.3	**ANSP System Availability**	For ANSP system a maximum of 40 outages lasting more than 6 minutes and a maximum outage of 240 minutes per year. The maximum delay between the outage and it being reported to the ATS unit should be less than 5 minutes.	N/A

[4] See Appendix 3 for compliance assessment methodology.

2.2.4	**ACSP System Availability**	For the CSP system a maximum of 40 outages lasting more than 6 minutes and a maximum outage of 240 minutes per year. The maximum delay between the outage and it being reported to the ATS unit should be less than 5 minutes.	N/A

Figure 1 KPI Target Values from ED228A

Targets that cannot be taken directly from the EUROCAE requirements for the metrics are defined in the table below. These figures are proposed targets based on expert judgement as described in Appendix 2.

Metric	Target Value
CSR Success Rate	No target currently defined[5].
Active Session Provider Abort Rate	1 PA per 100 hours

Figure 2 KPI Target Values from expert judgement

2.4 Additional System Level Metrics

This section contains other metrics defined at the end-to-end/system/overall service level for which it is considered beneficial to have a common definition.

- Overall PA rate
- Technical Round Trip Delay
- Technical Uplink Delay
- Technical Downlink Delay
- Message Acknowledgement Rate
- DLIC Contact Continuity
- Ground Initiated CPDLC Transaction Delay
- Downlink Error Rate
- Air Initiated CPDLC Transaction Delay
- Uplink Error Count Rate
- User Abort excluding CT and CDAA Rate
- Command Termination User Aborts

Further ideas for future metrics are captured in Appendix 3.

[5] Current value (June 2021) is around 90%. Reasons should be investigated and an appropriate target set.

2.4.1 Overall PA Rate

2.4.1.1 Description

This measures the rate at which Provider Aborts are experienced for aircraft operating above the local level of data link implementation It is simply a ratio of the total number of PAs over the total number of CPDLC session hours, regardless of whether a CPLDC session was established at the time the PA occurred or not. The rate is expressed as PAs per 100 hours of CPDLC session.

2.4.1.2 Definition

The PA rate would be calculated as the number of PAs experienced per 100 hours of CPDLC session i.e.

$$\left(\frac{NumberOfPAs}{TotalDurationOfCPDLCInHours}\right)x100$$

For example, consider two aircraft, one of which has CPDLC sessions with a total duration of 60 hours and the second with a total duration of 70 hours. One of those aircraft suffered 4 PAs before a CPDLC session was established and 1 PA after the session was active and the second aircraft suffered 2 PAs after the session was active, then the PA rate would be:

$$\left(\frac{5+2}{60+70}\right)x100 = 5.3\ PAs\ per\ 100\ Hours\ CPDLC$$

A CPDLC session is considered to start when the CPDLC start confirmed message is received by the ANSP end system

A CPDLC session is considered to end when a CPDLC end-response is received by the ANSP end system or a User Abort or Provider Abort is declared. If for some reason none of those events are observed then the last message sent or received from the aircraft should be considered as the end of the session; note this should not occur in practice

2.4.2 Technical Round Trip Delay

2.4.2.1 Description

The Technical Round Trip Delay (TRTD) is the time taken by the system to uplink a CPDLC message and receive its application level acknowledgement.

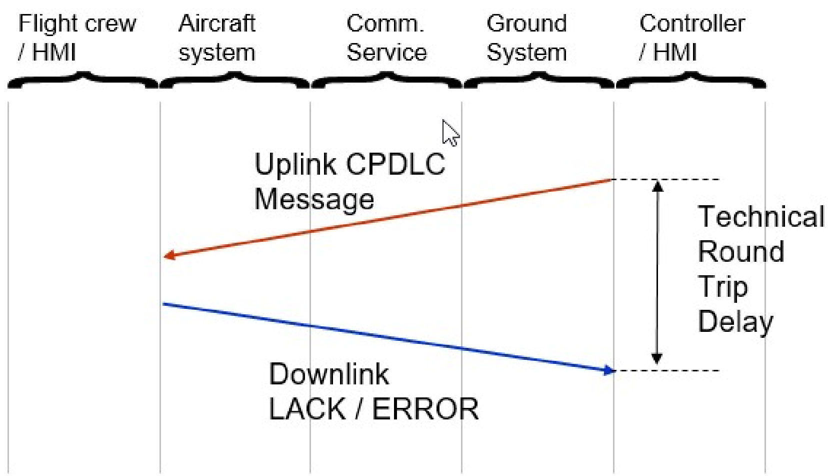


Figure 3: Technical Round-Trip Delay

Technical Round-Trip Delay can be calculated using recordings made by the ANSP end system.

All uplink CPDLC messages requesting a LACK should be logged with the time-stamp of when the message was sent[6] and the downlink application level acknowledgment[7] from the aircraft should be logged with time of receipt at the ANSP end-system.

The uplink messages may be associated with their corresponding LACK/ERROR as a post-process through use of the CPDLC Message Reference Number and the round trip delay calculated.

Note this metric is closely related to the RCTP Technical Continuity KPI metric; the difference is that the TRTD measures the delay for received responses whereas RCTP Technical Continuity metric calculates the probability of receiving a response within a specified time considering uplinks for which no response is received.

2.4.2.2 Definition

TRTD = LACK/ERROR response reception time – Uplink message transmission time.

It is recommended to calculate the 95th and 99th percentile values.

2.4.3 Technical Uplink Delay

2.4.3.1 Description

The uplink element of the Technical Round Trip Delay (TRTD) is the time from when the uplink message is sent and the corresponding LACK/ERROR is sent by the aircraft. This is an approximation of the uplink delay as it also includes some processing of the message by the aircraft to generate the LACK/ERROR.

The timestamp for the time at which the LACK/ERROR message is sent is taken from the header of the LACK/ERROR message and is generated by the aircraft.

[6] As embedded in the uplink message

[7] Note that the application acknowledgement may be a LACK or an ERROR.

Note: This metric will provide inaccurate results if the aircraft time system is not in sync with the ground time system. This is known to occur for some aircraft from time to time.

2.4.3.2 Definition

TUD = LACK/ERROR response transmission sent– Uplink message transmission time.

It is recommended to calculate the 95th and 99th percentile values.

2.4.4 Technical Downlink Delay

2.4.4.1 Description

The downlink element of the Technical Round Trip Delay (TRTD) is the time from when the downlink LACK/ERROR message is sent and the time it is received by the ground system.

The timestamp for the time at which the LACK/ERROR message is sent is taken from the header of the LACK/ERROR message and is generated by the aircraft.

Note: This metric will provide inaccurate results if the aircraft time system is not in sync with the ground time system. This is known to occur for some aircraft from time to time.

2.4.4.2 Definition

TDD = LACK/ERROR response reception time - LACK/ERROR response transmission time.

It is recommended to calculate the 95th and 99th percentile values.

2.4.5 Message Acknowledgment Rate

2.4.5.1 Description

This is the probability that a LACK or an ERROR message is received for an uplink message requiring a LACK within 40 seconds, which is the technical response (tr) timer value from ED110B. [8]

2.4.5.2 Definition

Message Acknowledgment Rate= number of uplink messages requiring a LACK for which a LACK or an ERROR response is received within 40 seconds or less / total number of uplinks requiring a LACK.

[8] This has been called the 'Technical Continuity' in the past but the name is changed to avoid confusion with the RCTP Technical Continuity metric.

2.4.6 DLIC Contact Continuity

2.4.6.1 Description

The DLIC Contact Continuity (DCC) is the probability that the DLIC contact transaction completes before the expiration timer expires (120s).

The ANSP ground system should record for the time at which the Contact Request was sent and should also record the time at which a corresponding contact response was received.

2.4.6.2 Definition

DCC = number of contact responses received within 120 seconds or less / total number of contact requests sent.

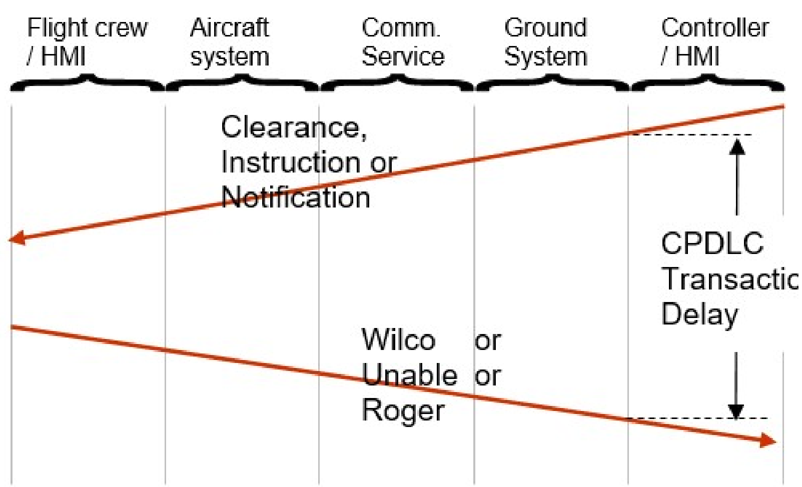
2.4.7 Ground Initiated CPDLC Transaction Delay

2.4.7.1 Description

The Ground Initiated CPDLC Transaction Delay (GICTD) is the delay between the message that initiates a transaction being sent and the corresponding message that closes the transaction being received by the ANSP end-system.

If the initial response is an ERROR message then the transaction should not be included in the statistic (since the transaction will not be closed). Also if the initial response is a LACK but an ERROR message is received subsequently for this transaction (because the flight crew did not respond before the timer expired) then the transaction should also not be included in the statistic[9] but again the error should be counted.

For transactions that are initiated by the ground the times can be derived directly from recordings in the ground system of when the initiating message was sent and the closing message received, as illustrated below.



[9] These are cases where the pilot or controller has not responded for some reason. They are important events that should be counted and will impact the Continuity performance but if included in the CPDLC transaction delay would obscure the normal transaction delay which is what this parameter is trying to measure.

Figure 4 Ground Initiated CPDLC Transaction Delay

An operational exchange that involves a STANDBY message is considered as two separate transactions as illustrated below.

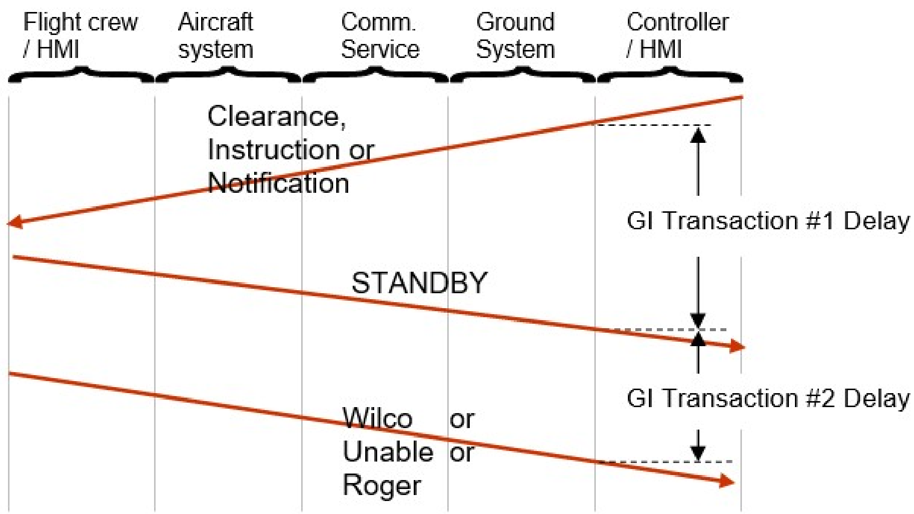


Figure 5 Ground Initiated CPDLC Transaction Delay involving a STANDBY

2.4.7.2 Definition

GICTD = Downlinked response message reception time – uplinked message transmission time

For transactions involving a STANDBY:

$GICTD_1$ = STANDBY reception time – uplinked message transmission time

$GICTD_2$ = Downlinked response message reception time – STANDBY reception time

It is recommended to calculate the 95th and 99th percentile values.

2.4.8 Downlink Error Count Rate

2.4.8.1 Description

This shows the rate of each type of freetext associated with a downlinked error message (DM62) as a percentage of the number of ground initiated transactions.

Examples of downlink error messages include:

- AIR SYSTEM TIMEOUT
- BUSY FLIGHT PHASE. USE VOICE
- TRANSFER IN PROGRESS. REPEAT REQUEST WHEN TRANSFER COMPLETE
- UPLINK DELAYED IN NETWORK AND REJECTED. RESEND OR CONTACT BY VOICE
- UPLINK TIMESTAMP INDICATES FUTURE TIME

- THIS CONCATENATION NOT SUPPORTED BY THIS AIRCRAFT.

2.4.8.2 Definition

$$\text{Downlink Error \%} = \frac{counts\ of\ a\ specific\ error}{number\ of\ ground\ \ initiated\ transactions}$$

2.4.9 Air Initiated CPDLC Transaction Delay

2.4.9.1 Description

The Air Initiated CPDLC Transaction Delay (AICTD) is the delay between the message that initiates a transaction being sent by the flight crew and the corresponding message that closes the transaction being received by the flight crew.

If the initial response is an ERROR message then the transaction should not be included in the statistic (since the transaction will not be closed). Also if the initial response is a LACK but an ERROR message is received subsequently for this transaction (because the controller did not respond before the timer expired) then the transaction should also not be included in the statistic[10] but again the error should be counted.

The time in the header of the downlink message is used as the approximation for the time at which the request was made by the flight crew and the time in the header of the downlink LACK message acknowledging the uplink closure response is used to approximate the time at which the aircraft received an uplink message, as illustrated below.

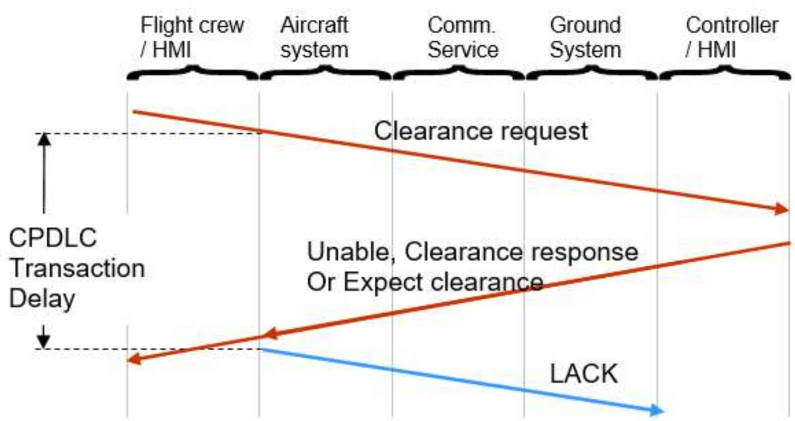


Figure 6 Air Initiated CPDLC Transaction Delay

As for ground initiated transactions an air initiated operational exchange that involves a STANDBY message is considered as two separate transactions as illustrated below.

---

[10] These are cases where the pilot or controller has not responded for some reason. They are important events that should be counted and will impact the Continuity performance but if included in the CPDLC transaction delay would obscure the normal transaction delay which is what this parameter is trying to measure.

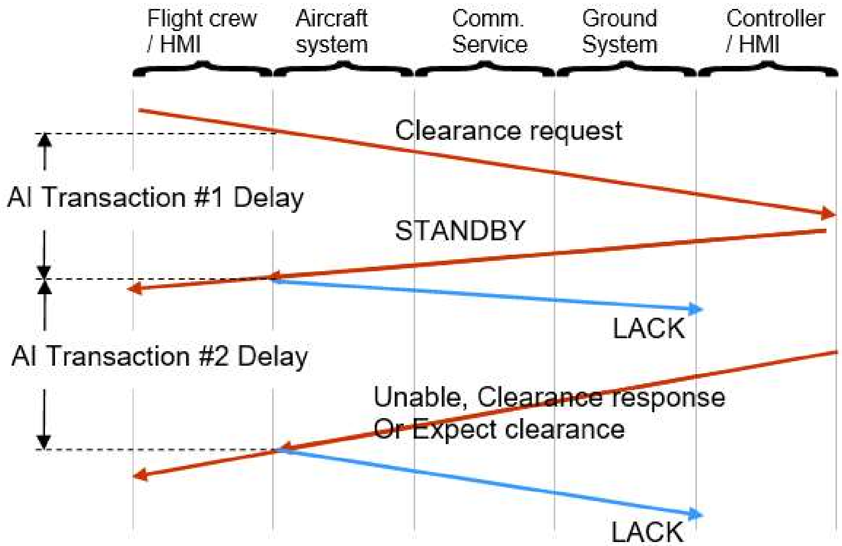


Figure 7 Air Initiated CPDLC Transaction Delay involving a STANDBY

2.4.9.2 Definition

AICTD = Time in the header of the LACK message acknowledging the response - Time in the CPDLC header of the downlinked request message.

For transactions involving a STANDBY:

$AICTD_1$ = Time in the CPDLC header of the downlinked request message - Time in the header of the LACK message acknowledging the STANDBY

$AICTD_2$ = Time in the header of the LACK message acknowledging the response - Time in the header of the LACK message acknowledging the STANDBY.

It is recommended to calculate the 95^{th} and 99^{th} percentile values.

2.4.10 Uplink Error Count Rate

2.4.10.1 Description

This shows the rate of each type of freetext associated with an uplinked error message (UM159) as a percentage of the number of air initiated transactions.

Examples of uplink error messages include:

- ATC TIMEOUT - REPEAT REQUEST
- CONTROLLER TERMINATED CPDLC - USE VOICE
- CPDLC MESSAGE FAILED - USE VOICE
- CPDLC MONITOR MESSAGE FAILED - USE VOICE ON FREQ: 120.935
- CPDLC TRANSFER NOT COMPLETED - REPEAT REQUEST
- DOWNLINK DELAYED - USE VOICE
- DOWNLINK MESSAGE REJECTED - SEND 2 ELEMENTS MAX.

2.4.10.2 Definition

$$\text{Uplink Error \%} = \frac{counts\ of\ a\ specific\ error}{number\ of\ air\ initiated\ transactions}$$

2.4.11 User Aborts excluding CT and CDAA rate

2.4.11.1 Description

The number of User Aborts experienced per 100 hours of CPDLC, excluding user aborts with a reason code of 'commanded-termination' or 'current-data-authority-abort'.

Commanded termination and current-data-authority-aborts are omitted as neither are considered to be of direct interest in monitoring the technical system performance. The flight crew or controllers terminating CPDLC causes user Aborts with a reason code 'commanded termination' so do not really represent any kind of abnormal system behaviour. The 'current-data-authority-abort' reason code is sent to the NDA centre if the aircraft raises a User Abort with the CDA centre, so a particular problem resulting in the aircraft raising a User Abort could result in two Users Aborts being counted if the user abort occurs when the aircraft has a connection with the NDA at the time.

2.4.11.2 Definition

The rate is the number of UAs (excluding CT and CDAA) experienced per 100 hours of Aircraft CPDLC usage i.e.

$$\left(\frac{NumberOfUAs}{TotalDurationOfCPDLCInHours}\right)x100$$

2.4.12 Commanded Termination User Aborts

2.4.12.1 Description

The number of Commanded Termination User Aborts experienced per 100 hours of CPDLC.

The controller or flight crew terminating CPDLC causes user Aborts with a reason code 'commanded termination' so do not necessarily represent any kind of abnormal system behaviour but do indicate the controller or crew terminating CPDLC for some reason, either manually or potentially automatically (e.g. x minutes after doors are open).

2.4.12.2 Definition

The rate is the number of Commanded Termination UAs experienced per 100 hours of Aircraft CPDLC usage i.e.

$$\left(\frac{NumberOfUAs}{TotalDurationOfCPDLCInHours}\right)x100$$

2.5 Additional VDL2 Level Metrics

These metrics are targeting the VDL mode 2 subnetwork performance, but it should be noted that the measurements also include some processing by the aircraft equipment so are not a direct measurement of only the air-ground network performance

These metrics are intended primarily as a means to track the evolution of performance, rather than as a means of compliance against the technical standards, since the measurements do not correspond precisely with how the requirements are stated. They should allow performance to be monitored in different parts of Europe consistently, but are not expected to necessarily form part of any agreement between the ANSPs and the ACSPs.

It is considered appropriate to calculate a confidence interval in accordance with the formula in paragraph A.3.6 below.

2.5.1 Uplink AVLC Round Trip Delay

2.5.1.1 Description

The Uplink AVLC Round Trip Delay is the delay between the time an AVLC frame is received at the VDL Ground Station (VGS) for an uplink transmission and the time at which an acknowledgement is successfully received at the VGS from the aircraft.

The AVLC RTD is expressed through percentiles of the delays distribution. It is recommended to calculate the 95th, 99th and 99.9th percentile values.

2.5.1.2 Definition

AVLC_RTD= Time of an AVLC ACK – time of the first AVLC INFO frame transmission on uplink.

Only acknowledged frames should be counted in the AVLC RTD.

Only uplink AVLC INFO frames conveying ATN packets should be used for the computation.

Valid acknowledgement from the aircraft is considered to be conveyed by one the following AVLC frames:

- RR
- SREJ (for frames up to N(R)-1)
- INFO

2.5.2 Uplink AVLC Reliability

2.5.2.1 Description

The Uplink AVLC Reliability is the probability that an uplink AVLC INFO frame received at the VGS for an uplink transmission is acknowledged within a specified time by the aircraft.

2.5.2.2 Definition

AVLC_R = Number of uplink AVLC INFO frame acknowledged / total number of uplink AVLC INFO frames to be delivered to the aircraft.

Guidance: AVLC_R(18) = Number of uplink AVLC INFO frame acknowledged within 18 seconds or less / total number of uplink AVLC INFO frames to be delivered to the aircraft.

Guidance: AVLC_R(10) = Number of uplink AVLC INFO frame acknowledged within 10 seconds or less / total number of uplink AVLC INFO frames to be delivered to the aircraft.

Only uplink AVLC INFO frames conveying ATN packets should be used for the computation.

Valid acknowledgement from the aircraft is considered to be conveyed by one of the following AVLC frames:

- RR
- SREJ (from frames up to N(R)-1)
- INFO

2.6 Additional Operational Metrics

This section contains some metrics that measure some aspect of the operational service.

2.6.1 R/T Time Saved

2.6.1.1 Description

This metric is intended to provide a measure of one of the benefits of data link; the amount of radio-telephony time saved. It is calculated by associating an average number of seconds against a number of categories of uplink messages and then counting the number of each category of message to work out the overall R/T time saved.

The following categories are used. The assumed R/T time is shown in brackets:

- Level Change (10s)
- Transfer of communications (15s)
- Instructions (12s)
- Route (15s)
- Speed (10s)
- Heading (14s)
- Check Stuck Mic (5s)
- Freetext (10s)
- Others(0s)

2.6.1.2 Definition

The R/T Time saved is the sum of the number of each category of message multiplied by the number of messages sent of that category.

The following uplink messages are assigned to each category:

- Level Change ('6','19','20','23','26','27','28','29','46', '47','48','61','171','172', '173', '174')
- Transfer of communications ('117', '120')
- Instructions ('123', '133','135', '148','179','196', '203','205', '211','213','215', '231', '232', '237')
- Route ('53','64', '73','74', '79','80', '82','92', '190')
- Speed ('55', '61','106', '107','108', '109','116', '222')
- Heading ('72', '94','96','190', '215')
- Check Stuck Mic (157 or 183)
- Freetext (183)
- Others(all other uplink messages)

2.6.2 Average Uplinks per flight per Centre

2.6.2.1 Description

This metric provides an indication of how much CPDLC is being used by controllers. It counts the average number of clearance/instruction messages uplinked per CPDLC flight by the average ATC centre over a specified period.

2.6.2.2 Definition

Average uplinks per flight per centre = Total number of operational uplink messages across all ATC centres / total number of ATC Centre-Flight pairs.

For example if there are two ATC centres of interest (A and B) and a total of 5 flights (a,b,c,d,e) and flights a,b,c and d pass through centre A and B but flight e just passes through centre B. If there are a total of 26 uplinks sent from centre A and 14 sent from centre B then the average uplinks per flight per centre would be (26+14)/(4+5).

The following uplinks are considered as operational messages : 6, 19, 20, 23, 26, 27, 28, 29, 46, 47, 48, 55, 61, 64, 72, 73, 74, 79, 80, 82, 92, 94, 96, 106, 107, 108, 109, 116, 117, 120, 123, 133, 135, 148, 157, 171, 172, 173, 174, 179, 190, 196, 203, 205, 211, 213, 215, 222, 231, 232, and 237.

2.6.3 Check Stuck Mic Count

2.6.3.1 Description

This metric is provides a count of the number of ‘Check Stuck Mic’ messages.

2.6.3.2 Definition

A simple count of the number of check stuck mic messages seen over a given period. Both UM157 messages and UM183 messages containing the text ‘CHECK STUCK MICROPHONE’ are to be counted.

References

[1] ED120. Safety and performance requirements standards for air traffic data link services in continental airspace (continental SPR standard). May 2004. As amended by Change 1, 2, and 3.

[2] ED-78a. Guidelines for approval of the provision and use of air traffic services supported by data communications. December 2000 [1]: ED120. Safety and performance requirements standards for air traffic data link services in continental airspace (continental SPR standard). May 2004.

[3] Link 2000+ DLS CRO Performance Monitoring Requirements. Edition 1.4 dates 19th May 2014.

[4] The DPMF report catalogue. Latest edition available from the DPMF OneSky team website.

[5] ED-228A. Safety and performance requirements standard for Baseline 2 ATS data communications (Baseline 2 SPR standard).March 2016.

Appendix 1 Rationale for Proposed Metrics

This section proposes a set of metrics to monitor and a supporting rationale.

Regulation 29/2009 only applies to traffic operating above FL285 but in practice CPDLC will also be used in airspace below FL285, so it is proposed not to limit the monitoring to flights operating above FL285.

Regulation 29/2009 refers to ED120 as the source for performance requirements and as such it forms the foundation for monitoring the performance of the services required by the regulation. However since ED120 was developed a new EUROCAE Safety and Performance Requirements Document has been developed (ED-228A) which is intended to define the performance requirements for Baseline 2 as well as Baseline 1 so it provides another possible source of performance requirements.

A.1.1 How requirements are stated in ED120

The transaction delay requirements in ED120 are stated in terms of the overall required communication performance (RCP) which is the total time from the initiation to the completion of a transaction. It is made up of a transaction time (TRN) plus the time taken to compose the CPDLC message and display the information to the flight crew or controller (the "Initiator" time[11]). The transaction time (TRN) is divided into two elements: a Required Communications Technical Performance (RCTP) which is the time taken by the technical systems to exchange the data between the air/ground/air and the 'Responder' time which is the time taken by the human to react to the message received.

The diagram below illustrates the ED120 terminology using an exchange initiated by the flight crew, but the terms also apply to transactions initiated by the controller.

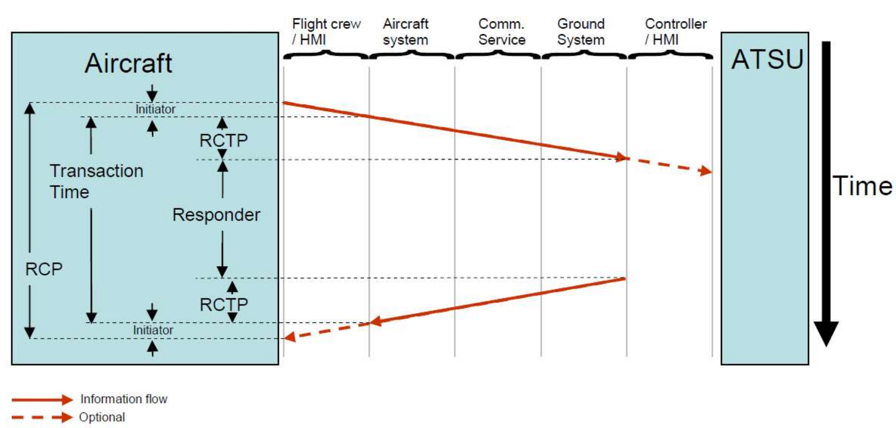


Figure 8: ED120 Breakdown of Transaction Time

Requirements are expressed in ED120 as an expiration time (ET), a transaction time for 95% of all transactions (TT_{95}), a continuity probability (C) , a probability of the availability of the service as a whole ($A_{PROVISION}$), a probability of availability of the service for a particular aircraft (A_{USE}), and an integrity level(I).

[11] Ed-120 does not allocate any specific value to the 'initiator' time.

The table below shows the descriptions of the various parameters as given in ED-78a.

Parameters	Value	Description
Transaction Expiration Time (ET_{RCP})	Time	Maximum time for completion of a transaction after which peer parties should revert to an alternative procedure. The rate at which a transaction expiration time can be exceeded is determined by the continuity parameter
95% Transaction Time (TT_{95})	Time 95%	Time before which 95% of the transactions are completed. This is the time at which controllers and pilots can nominally accept the system performance and represents normal operating performance.
Continuity (C_{RCP})	Probability	That the transaction will be completed before the transaction expiration time, assuming that the communication system is available when the transaction is initiated
Availability (A_{RCP}[12])	Probability	That the communication system between the two parties is in service when it is needed.
Availability ($A_{Provision}$)	Probability	That communication with all aircraft in the area is in service.
Integrity (I_{RCP})	Acceptable Rate	Of transactions completed with undetected error.

Figure 9 : ED78a Definitions of performance parameters

It should also be noted that ED120 states[13] that "…communication transactions that have multiple responses i.e. the STANDBY, followed by the operational response, are treated as two transactions…"

A.1.2 Scope of interest from ED120

There are many requirements in ED120 which are not of interest for the purposes of performance monitoring of the data link services. The regulation 29/2009 only applies to en-route and only to a subset of the services specified in ED120.

The Expiration Timer (ET) is a system parameter after which the transaction is considered to have timed out. It is used in the definition of Continuity (C) such that continuity represents the probability that transactions complete within ET.

The RCP values are often not specified in ED120. In practice it is the transaction (TRN) values that are of interest for monitoring the system performance. The

[12] This corresponds to the A_{USE} defined in ED120.

[13] Para 1.4.3.2.2

difference between the RCP value and the TRN values is the initiator portion and is governed by the design of the system HMI in the aircraft and the ATSU.

So the performance requirements from ED120 that are of interest are the TRN allocation values for the transaction time, continuity, availability and integrity requirements that apply in the en-route environment for the DLIC, ACM, ACL, and AMC services.

The relevant requirements for the individual services are reproduced in Figure 10 below. Continuity, A(Use) and A(Provision) are expressed in this table "per flight hour".

Parameter	Expiration Timer (ET)	Transaction Time (95%)	Continuity	Availability (Use)	Availability (Provision)	Integrity
DLIC initiation	60	30	0.99	0.993	0.999	10^{-5}
DLIC Contact	120	60	0.99	0.993	0.999	10^{-5}
ACM	120	60	0.99	0.993	0.999	10^{-5}
ACL Flight Crew Initiated	270	60	0.99	0.993	0.999	10^{-5}
ACL Controller Initiated	120	60	0.99	0.993	0.999	10^{-5}
AMC					0.999	10^{-3}

Figure 10 : Key performance requirements from ED120

These are the end-to-end performance requirements for transactions that in principle should be monitored. However Integrity is defined as the acceptable rate of transactions having undetectable errors and so by definition (since the errors are undetectable) cannot be regularly monitored in service. Integrity requirements have to be satisfied at the design stage.

ED-120 expresses Continuity as a probability "on a per flight hour basis" that the transaction completes successfully[14] before the expiration time. However, the concept of probability "on a per flight hour basis" is not considered meaningful so a more straightforward probability per transaction is preferred in [3]. In order to transform the requirement from a value per flight hour to a value per transaction it has been assumed that there will be ten transactions per hour, so the continuity requirement is made more stringent by a factor of ten i.e. the Continuity per transaction should be 99.9%. This value is consistent with the definitions given in ED-228A for Baseline 2.

The end-to-end performance requirements in Figure 10 are further broken down in ED120 to differentiate between the time taken by the system ("RCTP") and the time taken by the pilot/controller to respond ("Responder"). The RCTP component is further broken down in ED120 to allocate time to the avionics and the ground system (including ACSP).

[14] ED120 is not very clear about whether the transaction has to complete successfully, but this was the intent and is clearly stated in ED-228A para D.5.2.2

Whether all these requirements need to be monitored and how far the end-to-end requirements should be broken down for monitoring is a matter of judgment and is discussed in the following section.

A.1.3 Appropriate level of monitoring performance from ED120

This section discusses to what level of granularity the performance requirements from ED120 should be monitored.

It is clear that the end-to-end performance requirements should be monitored, but whether each requirement needs to be measured for each service and how far those overall figures should be broken down is open to question. There are several factors to consider:

- Whether to monitor the response time of the system separately from the pilot/controller response time.
- Whether to monitor the technical performance of the system separately.
- Whether to differentiate between the different services.
- How to measure availability.

As mentioned earlier it is not the objective of the ongoing performance monitoring to establish formal compliance with all the ED120 requirements; monitoring is not a form of acceptance testing. The objective is to monitor the performance of CPDLC at a suitable level to ensure that performance problems can be identified for more detailed investigation and also to monitor trends in performance so that action can be taken before the performance becomes unacceptable. So although the ED120 requirements are used as the basis for identifying a set of parameters to monitor and provides guidance for the expected performance, it is not necessary and indeed not practical in some cases to measure performance using the precise definitions given in ED120.

A.1.3.1 System v Human

The overall transaction times (including both the human element and the system element) will be measured as this is a key performance measure and is the scope of the operational monitoring proposed by ED-78a.

In the event of the overall transaction times failing to meet the requirements a more detailed investigation of the causes will be required and this may include looking at different elements of the system or the human performance of the pilots and controllers, but it is not considered necessary to do this on a regular, systematic basis.

A.1.3.2 Monitoring the system

The requirements given in Figure 10 represent the end-to-end performance requirements including both the system and the human. As stated previously these requirements have been broken down and allocated between the human and the system, and the system allocation has been divided between the avionics, the ACSP and the ATSU.

ED-78a does not propose monitoring the performance of individual elements of the system as part of operational monitoring. So rather than attempting to separately monitor the performance of the different elements of the technical systems, it is proposed to include a single measure, the ‘RCTP Technical

Continuity' to provide a good indicator of the performance of the technical system as a whole (the ATSU system, the ACSP and the avionics). It will measure the probability of the ground system receiving an acknowledgement[15] from the aircraft to an uplink. Monitoring the overall technical performance of the system will allow any adverse trends or events to be identified which may then require more detailed investigation to discover the cause.

A.1.3.3 Differentiating between Services

Although the requirements have been stated separately for each type of service, it could be argued that it is not necessary to monitor the performance separately for each service[16]. If the actual implementation treats the different services the same then there is little to be gained by separately monitoring the performance of each service as it would not have any real significance.

The total set of data link services are provided by two different applications (CM and CPDLC) which are implemented differently so it is proposed to monitor the CM application service (DLIC) separately from the CPDLC application services (ACM, ACL and AMC).

In practice there will be very little difference between sending one type of CPDLC message and another; they all use the same systems, it is just the operational meaning of the messages that differs between the services used. So it is not proposed to monitor the performance of the different CPDLC services separately. However the CPDLC transactions initiated by the controller should be monitored separately from the CPDLC transactions initiated by the pilot as they have different expiration timers.

A.1.3.4 Monitoring the CM application

For ATN Baseline 1 there are two elements to DLIC that it may be considered appropriate to monitor: i) Logon and ii) Contact. It would seem sensible to monitor the success rate for both. But to monitor the Logon success rate would require access to all LOF exchanges between ANSP systems and it is not considered to be worthwhile putting the infrastructure in place to provide that data. So the proposal is to monitor the DLIC contact function only as this can be done from the logs of air/ground messages exchanged.

A.1.3.5 Measuring Availability

For measuring the availability(provision) i.e. the availability of the service in the area as a whole it is proposed to report a simple metric based on unplanned outages of the service which affect more than one aircraft.

Gathering data to measure availability(use) accurately is problematic. The most obvious indication of availability(use) i.e. the loss of availability for an individual aircraft is a Provider Abort[17]. One of the most common reasons this occurs is when there is a lack of Air Ground connectivity for 6 minutes after which the system is considered to be unavailable. However it is not simple to determine precisely at what time the system becomes available again. It is possible that

[15] Either a LACK or an ERROR.

[16] This is consistent with the concept of grouping from ED78a (see para F.4.1)

[17] It may be possible to identify other common errors that indicate a loss of availability. Measuring the provider aborts is considered to be a good starting point, but others may be added later.

the pilot or controller may choose not to re-establish CPDLC even though it would be technically possible and so the system should be considered as available.

Provider Aborts are declared by the communication system, but there are also User Aborts that are declared by the end system (either triggered by the system itself or triggered by the controller/flight crew). These also affect whether the service is available to an individual aircraft.

It is proposed therefore to report the number of Provider Aborts and User Aborts rather than to calculate a formal probability of the availability(use).

The rate of Provider Aborts has been used as a proxy measure for both availability and continuity and a target rate of 1 PA per 100 hours of CPDLC has been derived in [3] from the performance requirements defined in ED120. This target has been used for several years and has been subject to a lot of discussion. The three main issues are

- **Geographic scope:** Whether to include all PAs or just those within the declared service volume.
- **Temporal scope:** Whether to include all PAs or just those that occur whilst a CPDLC session is active..
- **Service Scope:** Whether to include just CPDLC PAs or also CM PAs

Some ANSPs have implemented a process that makes several attempts to establish a CPDLC session if the CPDLC start request message does not at first succeed. This can lead to a number of PAs being declared for a single flight before the CPDLC session is active and if these PAs are included in the overall PA rate calculation it may arguably make the performance of the overall service appear worse than how the controller would perceive it (as these PAs would not cause a direct disturbance to the controller since the aircraft is not yet able to use data link with the centre). So it is proposed not to include PAs that occur before the CPDLC session has started in the calculation of PA rate used for the KPI metrics. A CPDLC session would be considered to have truly started only once the UM183 CURRENT ATC UNIT[18] message has been sent by the ground system.

Only counting PAs that occur once a session is established means we need a way to monitor performance before the session is properly established (i.e. before the UM183 ‘ CURRENT ATC UNIT …’ message has been sent). So it is proposed to add a metric to measure the proportion of CPDLC start requests messages that result in a session being properly established. This will highlight any issues with the communication system, or elsewhere that prevents the timely establishment of the CPDLC session.

The following metrics are proposed to address the availability of the service to an individual aircraft (referred to as Availability (use) in ED120):

- **Active Session PA Rate:** which will only include PAs declared for aircraft with an active CPDLC session i.e. the UM183 ‘ CURRENT ATC UNIT …’ message has been sent..

[18] There is a clear requirement to send this message defined in Tables 4-3, 4-4, 4-6, 4-8, 4-9 and 4-11 of ED-110B.

- **CSR Success Rate.** The proportion of CPDLC start requests that result in a UM183 CURRENT ATC UNIT message being sent by the ground system.

No CM PA rate metric is proposed. It is not thought to add much in addition to the CPDLC PA rate (the PA rate is a proxy for service availability which would apply equally to CPDLC and CM).

An **Overall Provider Abort Rate** is kept from the previous metrics definitions [3] but modified to include all PAs declared whilst the aircraft is above the flight level above which the data link service is offered by each ANSP. This metric should also take into account the maximum defined latitude for Sweden/Norway/Finland.

A.1.3.6 The recommended metrics to monitor performance

The above considerations lead to the proposal of the following set of recommended metrics:

- **RCTP Technical Continuity**. The probability of receiving a technical response to an uplink message within a defined time. This will give a good indication of the overall performance of the technical system (i.e. the ACSP, avionics and the ground end-system).
- **DLIC Contact Continuity.** The probability that a contact request results in the reception of a contact response before the expiration timer expires.
- **CPDLC Transaction Delay**. Two separate distributions of the delay for all CPDLC ACL, ACM and AMC transactions; one for air initiated transactions and the second for ground initiated transactions, plus a count of any error responses.
- **CPDLC Continuity**. The probability that CPDLC transactions are closed before the expiration timer expires. Calculated separately for air initiated and ground initiated transactions.
- **Overall Provider Abort Rate.** The number of Provider Aborts experienced per 100 hours of CPDLC usage.
- **Active Session PA Rate.** The number of Provider Aborts experienced during an active CPDLC session expressed per 100 hours of CPDLC usage.
- **CSR Success Rate.** The proportion of CPDLC start requests that result in a CDA message being received by the ground system
- **User Abort Rate.** The number of User Aborts experienced per 100 hours of CPDLC usage.
- **System Availability.** Any unplanned outages of the service as a whole. This reflects the overall system availability.

Appendix 2 KPI Metrics and Target Value Derivation

This appendix identifies the subset of the recommended metrics that are considered to be necessary to measure the overall performance of data link against the key performance requirements. If the system meets the defined targets for these metrics then it is considered to perform acceptably from an end-user perspective. This set is referred to as the KPI metrics.

There are two groups within the KPI metrics: i) those for which a target value can be directly defined from the EUROCAE standards, and ii) a set intended to cover the availability of the service per aircraft (Availability(Use)) for which it is not considered possible to directly define a target value from the EUROCAE standard.

A.2.1 The list of 'KPI Metrics'

The following set of metrics are considered to be necessary to adequately demonstrate compliance against the performance requirements in the EUROCAE standards and are expected to form the basis for a demonstration of compliance against the performance requirements if required by any supervisory authority:

- Ground Initiated Transaction Continuity
- RCTP Technical Continuity
- CDPLC Start Request (CSR) Success Rate
- Active Session Provider Abort Rate
- ANSP System Availability
- ACSP System Availability

When compared to the recommended set of metrics in A.1.3.6 above it should be noted that the following list of metrics are not considered necessary for the reasons given below:

a) The DLIC Contact Continuity metric is not considered essential since Baseline 2 defines no specific performance requirements for DLIC.
b) The CPDLC transaction delay metrics are not considered essential since the requirements are more precisely covered by the transaction Continuity metrics.
c) The Air Initiated Transaction Continuity metric is not considered essential. ED-228A does not define different continuity requirements for Flight Crew initiated and Controller initiated transactions – RCP130 applies to both. However it does state (in Appendix E para E.5.1.1)" …For ground initiated exchange transactions, the maximum acceptable transaction time is stated as an Expiration Time (ET). The need for ET is driven by safety. NOTE: In accordance with the CPDLC-OSA (refer to Appendix B), air initiated transactions do not require an indication for display to the flight crew upon exceeding the maximum acceptable time…." So it is not considered essential to measure the Flight Crew initiated transaction Continuity as it seems to not be related to safety requirement
d) The Overall Provider Abort rate is not considered necessary. The combination of Active Session Provider Abort Rate and CSR Success Rate is considered a better combination to measure availability per aircraft.
e) The User Abort rate is not considered essential since the Active Session Provider Abort Rate and CDA Success Rate metrics are considered to cover the availability requirements adequately.

A.2.2 The choice between ED120 and ED-228A

Since ED-120 was developed EUROCAE have developed a new performance standard for data link; ED-228A. ED-228A is intended to address the requirements for Baseline 2 as well as the current performance requirements defined in ED-120. The baseline 2 requirements represent a more up to date industry view of the required performance and removes some of the inconsistencies and uncertainty in the baseline 1 requirements and so is the preferred reference for determining the target values.

A.2.2.1 The scope of ED-228A

The extract from ED-228A below shows it is intended to include the requirements of ED120

1.2 SCOPE

This document supports provision of data communications in all operational environments e.g., continental, oceanic, and surface in support of provision of air traffic control services.

This document retains the operational, safety and performance requirements that were specified in legacy SPR documents:

- operational, safety, and performance requirements for data link services that support ATM Operations in oceanic and remote airspace as specified in ED-122/DO-306 (including Change 1)
- operational, safety, and performance requirements for data link services that support ATM Operations in domestic airspace as specified in ED-120/DO-290 (including Change 1 and 2)

Figure 11 Extract from ED-228A concerning scope covering ATN B1

A.2.2.2 Primary differences between ED120 and ED-228A

The performance requirements defined in ED120 and ED-228A are very similar. They are discussed below in two sets: i) the transaction continuity requirements and ii) the availability related requirements. In both cases the requirements are similar but ED-228A provides greater clarity.

The main differences that are of interest between ED120 and ED228A are listed below:

- ED-228A provides some clarifications and removes some inconsistencies in the baseline 1 requirements documents (ED110B and ED120).
- ED-228A does not define performance requirements for DLIC (CM Logon etc.)
- ED-228A allocates the overall transaction delay using a different methodology which results in less stringent requirements for the allocation of delay to the different elements of the system.
- ED-228A provides requirements in terms of transactions whereas ED120 defines most requirements "on a per flight hours basis" which then have to be converted into a metric per transaction to be measured practically.

(1) Transaction Continuity Requirements

The relevant transaction performance requirements for CPDLC from ED120 and ED228A are highlighted in the tables below

TABLE 5-32 : ACL CONTROLLER INITIATED TRANSACTIONS PERFORMANCE REQUIREMENTS (CONTINUED)

Service	ACL – controller initiated transactions					
Parameter	ET	TT(95)	C	A(USE)	A(PROVISION)	I
Source	O	O	$QOPL_{UIT}$	$QOPL_{LOCP}$	$QOPL_{LOS}$	$QSPL_{UCT}$
RCP_{ER}	See Note 4	See Note 4	0.99	0.993	0.999	10^{-5}
RCP_{TMA}	See Note 4	See Note 4	0.99	0.993	0.999	10^{-5}
$RCP_{ARR/DEP}$	See Note 4	See Note 4	0.99	0.993	0.999	10^{-5}
$RCP_{Aerodrome}$	See Note 4	See Note 4	0.99	0.993	0.999	10^{-5}
RCP Allocations						
Initiator	See Note 4	See Note 4	See Note 4	See Note 5	See Note 5	See Note 5
TRN_{ER}	120	60	0.99	0.993	See Note 6	10^{-5}
TRN_{TMA}	45 See Note 3	30 See Note 3	0.99	0.993	See Note 6	10^{-5}
$TRN_{ARR/DEP}$	45 See Note 3	30 See Note 3	0.99	0.993	See Note 6	10^{-5}
$TRN_{Aerodrome}$	120 See Note 3	60 See Note 3	0.99	0.993	See Note 6	10^{-5}
TRN Allocations						
$Responder_{ER}$	100	44	0.995	See Note 5	See Note 5	See Note 5
$Responder_{TMA}$	25 See Note 3	14 See Note 3	0.995	See Note 5	See Note 5	See Note 5
$Responder_{ARR/DEP}$	25 See Note 3	14 See Note 3	0.995	See Note 5	See Note 5	See Note 5
$Responder_{Aerodrome}$	100 See Note 3	44 See Note 3	0.995	See Note 5	See Note 5	See Note 5
RCTP	20	16	0.995	0.993	See Note 6	10^{-5}

Figure 12 Extract from ED-120 - ACL performance requirements

TABLE E-2: RCP130/A1 AND RCP400/A2 TIME ALLOCATIONS

RCP communication Transaction Time, 99.9% Continuity and 95% Continuity				
Defined Allocations	RCP130/A1		RCP400/A2	
Transaction Time Parameter	ET (sec) C=99.9%	TT (sec) C=95%	ET (sec) C=99.9%	TT (sec) C=95%
Transaction Time	130	67	400	350
RCP Time Allocations				
Initiator	30	13	30	13
RCMP	120	60	380	174
RCMP Time Allocations				
Responder	100	44	371	161
RCTP	32	20	32	20
RCTP Allocations				
ATSP	23	14	23	14
ATSU	*14*	*6*	*14*	*6*
CSP	*18*	*10*	*18*	*10*
Aircraft	23	10	23	10
Notes				
NOTE: *For the 95th percentile total transaction time, a value of 350 seconds has been taken to maintain consistency with RCP400 specification, used in ENR-2 (refer to section E.5.2.1) .In accordance with the DCL validations, the truly 95th Percentile total transaction time (TT_{95}) should be 180 seconds. This value has been taken for deriving the 95th percentile values of the statistical (lognormal) TIME allocations.*				

Figure 13 RCP130 definition from ED-228A

ED120 expresses Continuity as a probability "on a per flight hours basis", so to convert the 99% requirement from a probability "on a per flight hours basis" to a probability per transaction, we assume 10 transactions per hour[19] to give a requirement of 99.9% of transactions must complete within 120 seconds. This is consistent with the requirements in ED228-A (the baseline 2 SPR document).

After converting the ED120 requirements from "a per flight hours basis" into transactions we see the requirements for CPDLC transaction delays are identical apart from the RCTP.

Requirement	ED120	ED228A
Probability of response to a controller message within 120 seconds	99.9%	99.9%
Probability of response to a controller message within 60 seconds	95%	95%
RCTP	20 seconds	32 seconds

Figure 14 ED120 v ED-228A Transaction performance requirements

There is an apparent inconsistency in the ATN Baseline 1 requirements defined in ED110B (Interop standard) and ED120 (Safety and Performance standard). ED110B has the technical response (tr) timer defined as 40 seconds, whereas the RCTP Expiration Time is defined as 20 seconds. ED110A (the predecessor to ED110B) has tr defined at 20 seconds so would be consistent with the ED120 requirement, so it appears as though ED110 was updated without reflecting the change in ED120.

So from the ATN baseline 1 standards there are two possible maximum delays that could be used when establishing the target RCTP Technical Continuity: i) the 40s seconds from the tr timer in ED110B but without a stated requirement for the minimum percentage that must receive a LACK before the timeout, or ii) the 20 seconds from ED120 for 99.95%[20] of the transactions. It should also be noted that the implementation of the tr timer is not strictly a requirement in ED110B, neither is it a stated as a performance requirement. It is a timer which when it expires results in the notification to the controller and the implementation is stated in a note in paragraph 4.2.1.2 of ED110B as being ' a local matter'.

The baseline 2 documents (ED228-A and ED229-A) have consistent values (32 seconds) defined for the technical response timer (tr) and the RCTP expiration time.

(2) Availability Requirements

ED120 defines two types of availability: Availability (Provision) which relates to the service as a whole in a given area and Availability(Use) which relates to the service between two parties (i.e. the ground system and an individual aircraft), but gives very little guidance on how to measure them.

ED228-A specifies an overall RCP availability for the service as a whole and individual allocations of that availability to the ATSU, CSP and Aircraft.

[19] See C-ENV-11

[20] The requirement in ED120 is expressed per flight hour, so assuming 10 transactions per hour the target needs to be a factor of ten more strict. But note ED228-A uses 99.9%.

ED-228A has a rather complicated proposed method for calculating the availability of the service as a whole, as shown in Figure 15 below.

D.7.4 Availability

Availability is a target for continued compliance demonstration (i.e. post-implementation monitoring) in addition to being a 'requirement' to be formally demonstrated for certification of the aircraft system or ground approval.

- **Measurement of RCP/RSP Availability:**

The (overall) RCP/RSP Availability (A) is measured over 'flight hours' for which the ATSU, CSP, and Aircraft are planning communication or surveillance services. Flight hours are in quotes since service is planned for aircraft at the gate where, technically, no flight hours are being accumulated.

Example:

An ATSU may plan only to provide data communications services in the ENR-1 airspace. For an ENR-1 sector over the measurement period, there may be a total of 6,000 flights, equipped with data communications, each spending an average of 10 minutes in the sector. The total planned number of flight hours in the measurement period is 1000 flight hours. If an outage affects 18 flights for 10 minutes each and 10 other flights for 6 minutes each, the total outage time 4 flight hours. Thus, the overall Availability (A) measurement is calculated as (1000-4)/1000 = 0.996.

Figure 15: Method of measuring overall availability in ED-228A

But ED228-A also specifies allocations of availability to the ATSU, CSP and the aircraft and for the ATSU and CSP it specifies the availability in much simpler terms; the maximum number and duration of outages as shown in Figure 16 and Figure 17 below.

	RCP 130		RCP 240		RCP 400	
Parameter	ET	$TT_{95\%}$	ET	$TT_{95\%}$	ET	$TT_{95\%}$
Transaction Time (Sec)	130	67	240	210	400	350
Continuity (C)	0.999	0.95	0.999	0.95	0.999	0.95
Availability (A)	0.989		0.989 (safety) 0.9899 (efficiency)		0.989	
Integrity (I)	1E-5 per FH		1E-5 per FH		1E-5 per FH	
RCP Monitoring and Alerting Criteria						
MA-1	The system shall be capable of detecting failures and configurations changes that would cause the communication service to no longer meet the RCP specification for the intended use.					
MA-2	When the communication service can no longer meet the RCP specification for the intended function, the flight crew and/or the controller shall take appropriate actions.					

Defined Allocations for RCP Specifications Note 4								
	RCP 130/A1		**RCP 240/A1**		**RCP 400/A1**		**RCP 400/A2**	
Parameter	ET	$TT_{95\%}$	ET	$TT_{95\%}$	ET	$TT_{95\%}$	ET	$TT_{95\%}$
Transaction Time (Sec)								
Initiator	30	13	30	30	30	30	30	13
RCMP	120	60	210	180	370	320	380	174
Responder	100	44	60	60	60	60	371	161
RCTP	32	20	150	120	310	260	32	20
$RCTP_{ATSP}$	23	14	n/a	n/a	n/a	n/a	23	14
$RCTP_{ATSU}$	14 Note1	6 Note1	15	10	15	10	14 Note1	6 Note1
$RCTP_{CSP}$	18 Note1	10 Note1	120	100	280	240	18 Note1	10 Note1
$RCTP_{Aircraft}$	23	10	15	10	15	10	23	10
Continuity (C)								
$C_{ATSU,\ CSP,\ and\ Aircraft}$ (See *NOTE 2*)	0.999	0.95	0.999	0.95	0.999	0.95	0.999	0.95
Availability (A)								
A_{ATSU}	0.9995		n/a		n/a		0.9995	
A_{CSP}	0.9995		0.999 (safety) 0.9999 (efficiency)		0.999		0.9995	

Figure 16: Table 5-14 of ED-228A Showing availability allocations (part 1)

Parameter	RCP 130/A1 ET	RCP 130/A1 TT95%	RCP 240/A1 ET	RCP 240/A1 TT95%	RCP 400/A1 ET	RCP 400/A1 TT95%	RCP 400/A2 ET	RCP 400/A2 TT95%
$A_{Aircraft}$	0.99		0.99		0.99		0.99	
Unplanned outage duration limit$_{ATSU\ \&\ CSP}$ (min)	6		10 (CSP only)		20		6	
Defined Allocations for RCP Specifications Note 4								
	RCP 130/A1		**RCP 240/A1**		**RCP 400/A1**		**RCP 400/A2**	
Parameter	ET	$TT_{95\%}$	ET	$TT_{95\%}$	ET	$TT_{95\%}$	ET	$TT_{95\%}$
Availability (A) (continued)								
Maximum number of unplanned outages$_{ATSU}$	40		n/a		n/a		40	
Maximum number of unplanned outages$_{CSP}$	40		48 (safety) 4 (efficiency)		24		40	
Defined Allocations for RCP Specifications								
	RCP 130/A1		**RCP 240/A1**		**RCP 400/A1**		**RCP 400/A2**	
Parameter	ET	$TT_{95\%}$	ET	$TT_{95\%}$	ET	$TT_{95\%}$	ET	$TT_{95\%}$
Maximum accumulated unplanned outage time$_{ATSU}$ (min/yr)	240		n/a		n/a		240	
Maximum accumulated unplanned outage time$_{CSP}$ (min/yr)	240		520 (safety) 52 (efficiency)		520		240	
Unplanned outage notification delay$_{ATSU\ \&\ CSP}$ (min)	5		5		10		5	

Figure 17: Table 5-14 of ED-228A Showing availability allocations (part 2)

The 'Unplanned outage duration limit' is important as it defines the maximum amount of time an outage may last without being considered against the availability requirement. Appendix D of ED228-A gives definitions for the various term used in the availability requirements as is reproduced below:

Unplanned outage	An outage for which no advance notification was provided to the appropriate parties.
Unplanned outage time	The time from when an unplanned outage begins to when the ATS unit receives notification that the service has been restored.
Unplanned outage duration limit (min)	A value applied to a given airspace that defines the maximum time for the duration of an unplanned outage at which time there is an operational impact. ***NOTE 1:*** *In literature also often called Mean Time to Repair (MTTR)* ***NOTE 2:*** *Unplanned outages that are less than the unplanned outage duration limit are considered against the continuity.*
Maximum number of unplanned outages	A value that defines the acceptable number of unplanned outages that exceed the unplanned outage duration limit in a specified time period ***NOTE:*** *Measured separately for each relevant operational airspace over any 12-month period.*
Maximum accumulated unplanned outage time (min/year)	The time that defines the acceptable accumulated duration of unplanned outages that exceed the unplanned outage duration limit in a specified time period. ***NOTE:*** *Over a 12-month period, it is measured in minutes/year by accumulating **only** the duration times for unplanned service outages greater than the unplanned service outage duration limit during any 12-month period. The accumulation is performed separately for each relevant operational airspace.*
Unplanned outage notification delay (min)	The time from when the unplanned outage begins to when the ATS unit receives notification of the unplanned outage.

Figure 18: Availability terminology definitions from ED-228A

Figure D-7 from ED-228A illustrates the fact that short term outages (less than the unplanned outage duration limit) are not considered against the accumulated unplanned outage time.

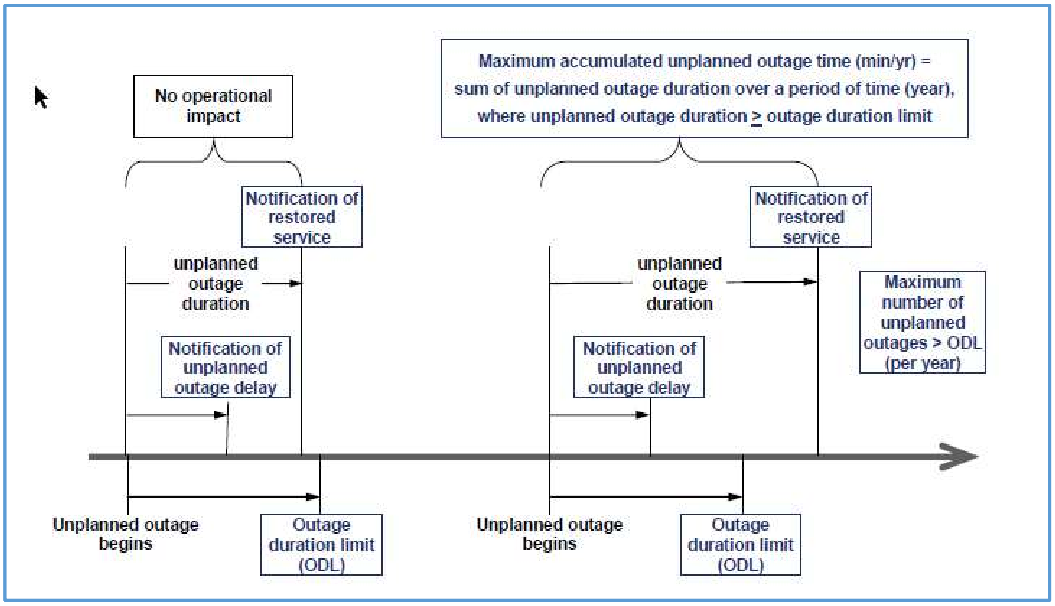


Figure 19: Figure D-7 from ED-228A: Time sequence diagram for unplanned outages

For periodic monitoring of the availability of the Aircraft ED-228A proposes in paragraph D.7.4 to measure the Provider Abort rate, although it is recognised that this is not a precise measure of the actual aircraft availability requirement.

The stated Availability(Aircraft) of 0.99 is considered to be an error. It has been recognised as an error by the ICAO OPDLWG group and the figure to be included in the PBCS manual definition of RCP130 will be 0.999. This is currently due for publication in 2022.

A.2.2.3 Conclusion

In conclusion ED-228A is considered to be a suitable basis for the assessment of performance of datalink in Europe, covering both ATN Baseline 1 operations and Baseline 2 operations in the future[21].

A.2.3 Derivation of Target Values

This section describes how the target values given in section 2.3 for the KPI metrics were derived.

The following set of KPI metrics have a target defined directly from the EUROCAE standard:

- Ground Initiated Transaction Continuity
- RCTP Technical Continuity
- ANSP System Availability
- ACSP System Availability

The KPI metrics below are intended to measure the availability of the service to an individual aircraft and do not have a target defined directly from the EUROCAE standard:

- CSR Success Rate

[21] This is in line with section 1.3.2 of ED-228A.

- Active Session Provider Abort Rate

A.2.3.1 Ground Initiated Transaction Continuity

Two requirements are stated for Ground initiated Transaction Continuity in RCP 130, see Figure 13 above

- 99.9% of ground initiated transactions should complete within 120 seconds
- 95% of ground initiated transactions should complete within 60 seconds.

A.2.3.2 RCTP Technical Continuity

Two requirements are stated for RCTP Continuity in RCP 130, see Figure 13 above

- 99.9% of message exchanges (uplink and downlink elements combined) should complete within 32 seconds
- 95 of message exchanges (uplink and downlink elements combined) should complete within 20 seconds.

A.2.3.3 System Availability

For the sake of simplicity it is proposed not to monitor the overall system availability (as described in Figure 15 above) but to only monitor the system outages of the ATSU and CSP as shown in Figure 17 above. i.e.

- A maximum of 40 outages lasting more than 6 minutes of the CSP system and a maximum total outage of 240 minutes per year.
- A maximum of 40 outages lasting more than 6 minutes of the ANSP system and a maximum total outage of 240 minutes per year.

A.2.3.4 Active Session Provider Abort Rate

A provider abort most often implies a loss of communications of 6 minutes. So if we consider that (with hindsight) the system was not available during those 6 minutes, then one PA in an hour would constitute a 10% loss of availability (i.e. for 6 minutes out of 60 minutes the system was not available) so each PA suffered per 100 hours would constitute a 0.1% loss of availability. On the assumption that the stated aircraft availability requirement in ED228A will be changed to match the PBCS definition of RCP 130 then the corresponding target PA rate of 1 PA per 100 hours would 'match' the revised ED-228A target figure of 99.9% availability for the aircraft.

A.2.3.5 CSR Success Rate

There is no target value for the CPDLC Start Request Success Rate that can be derived from the EUROCAE standards, so no target is explicitly defined at this time.

An initial implementation of the metric shows the average value to be around 90%[22] which seems to indicate a need for improvement. Once the underlying reasons behind this low success rate and the operational implications are understood it may be possible to agree a common target value, but in the

[22] June 2022

meantime it is proposed not to set a specific target but instead to a) monitor the trend and b) investigate why the success rate appears relatively low.

Appendix 3 Confidence Intervals and compliance assessment

The aim of this section is to explain in plain English (with as little maths as possible!) the statistical methodology proposed for compliance assessment. Assessment of Continuity is taken as an example to drive us through the explanation. It reflects the statistical material found in section D.7.5 of ED-228A but the proposed methodology differs slightly.

A.3.1 Introduction

Continuity is defined as the probability that a transaction completes before a specific time (e.g. Expiration Time). It is a proportion: the ratio between the number of transactions completed before the specific time over the total number of transactions (requiring an answer).

The “True Continuity” is the continuity of the system as a whole and cannot be “measured” directly. Monitoring the Continuity then consists in “estimating” the True Continuity with adequate sampling. In a “live” environment, the only samples available are the transactions initiated by the controllers/crews. These transactions are used to estimate Continuity.

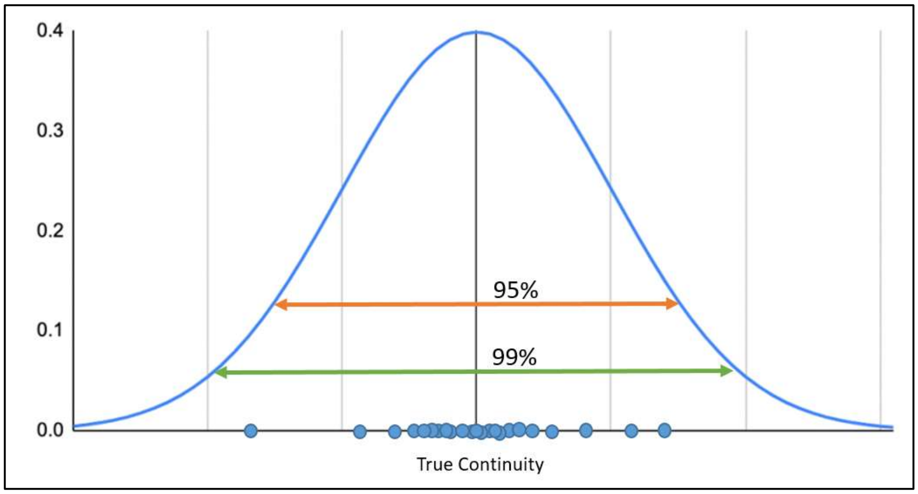
A.3.2 The construction of confidence intervals

If we were to perform several samplings, we would come up with a several Continuity estimates. As Continuity is a proportion, the estimated Continuity values follows a Binomial distribution.

From the Central Limit theorem, as the number of samples used to compute each estimate increases, the Binomial distribution (which is most of the time asymmetric) tends to a Gaussian distribution whose mean is equal to the “True Continuity”. Moreover, the more samples used to estimate Continuity, the more “precise” the estimation is and the smaller the “width” of the distribution.

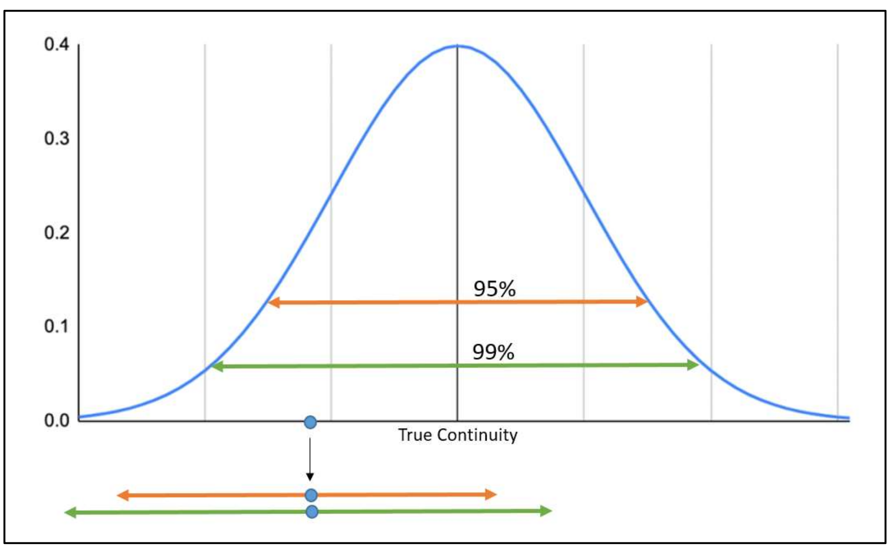
The following graph shows a Gaussian distribution around the True Continuity. The blue dots represent the values for our Continuity estimates if we were to perform several samplings. The blue line represents the proportion of our continuity estimates that give a particular value of continuity. If we were able to perform a sufficiently large number of samplings over a specific time period, each of which with a sufficient number of samples, we would achieve the nice smooth blue line and could get a close estimate of the True Continuity by averaging the estimates. This average is the black vertical line in the centre of the distribution below.

The orange and green arrows represents the intervals containing respectively 95% and 99% of the Continuity estimates around the True Continuity. Assuming a normal distribution of our Continuity estimates, we should expect half of our estimates to have a value greater than the True Continuity and half to have a value less than the True Continuity.



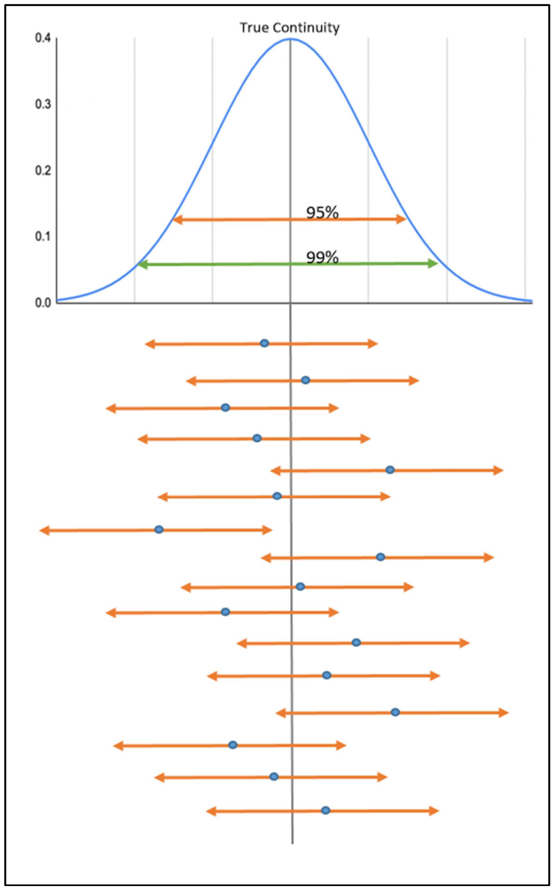
In the scope of our (real life) monitoring, we are generally able to make only a single estimation of Continuity over a specific time-period with a sufficient number of sample (we will come back to the number of sample later). Since we do not know the True Continuity value and cannot in practice take an infinite number of samples to measure it, instead of looking at the probability that our measured estimate is within a range of the True Continuity (orange and green arrows in the graph above), we can reverse the situation by looking at the probability that the true mean is within a range of our single measured estimate.

Assuming that we can estimate the parameters (i.e. the shape) of the distribution of Continuity estimates we can construct the same orange and green arrow around our single measured Continuity estimate as shown in the graph below. Saying that the arrows contains 95% or 99% of the estimates around the True Continuity is equivalent to saying that these arrows, when put around the Continuity estimate, will contain the true Continuity 95% or 99% "of the time". The term "of the time" is of importance and is described hereafter.



The orange and green arrows around our Continuity estimate (the blue dot) are called "confidence intervals" for respectively a 95% and a 99% "confidence level".

If we were to perform several samplings and construct for each of them an orange confidence interval (the 95% confidence level), then 95% of the time, these confidence intervals would contain the true continuity, as illustrated below.



The confidence interval is thus a means to "locate" the true Continuity value using an interval based on a certain level of confidence (confidence level).

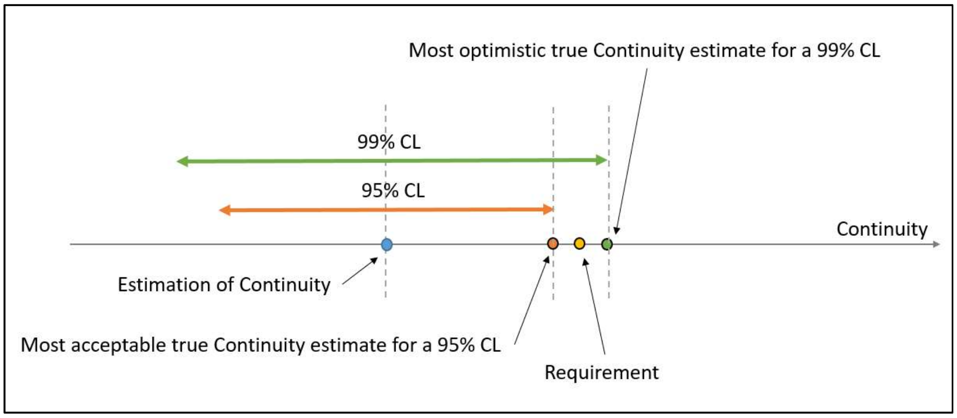
A.3.3 Compliance assessment

At this stage, we need to keep in mind that we need to compare the True Continuity, not the estimated Continuity, against the requirement (the target value).

The system is compliant when the True Continuity is above or equal to the required Continuity target value.

As we cannot compare the True Continuity against the requirement value (because we don't know the True Continuity), we compare an interval in which we are confident to find the True Continuity against that requirement. One should remember that in the scope of our (real life) monitoring it is very unlikely we are able to perform sufficient large samplings over a specific period of time, each of which with a sufficient number of samples, to get a close estimate of the True Continuity by averaging the estimates.

The graph below shows our two confidence intervals (orange and green) around the Continuity estimate (blue dot). The green line represents the range of values within which we are 99% sure the True Continuity lies and the orange line represents the range of values within which we are 95% sure the True Continuity value lies. As the True Continuity could be anywhere within the intervals, we could assume the "system" to be compliant as long as the upper value of the confidence interval is above or equal to the requirement value.



In the example above, the system will be considered compliant in the case of the 99% CL (green dot is above the yellow one) and not compliant in the case of the 95% CL (orange dot is below the yellow one).

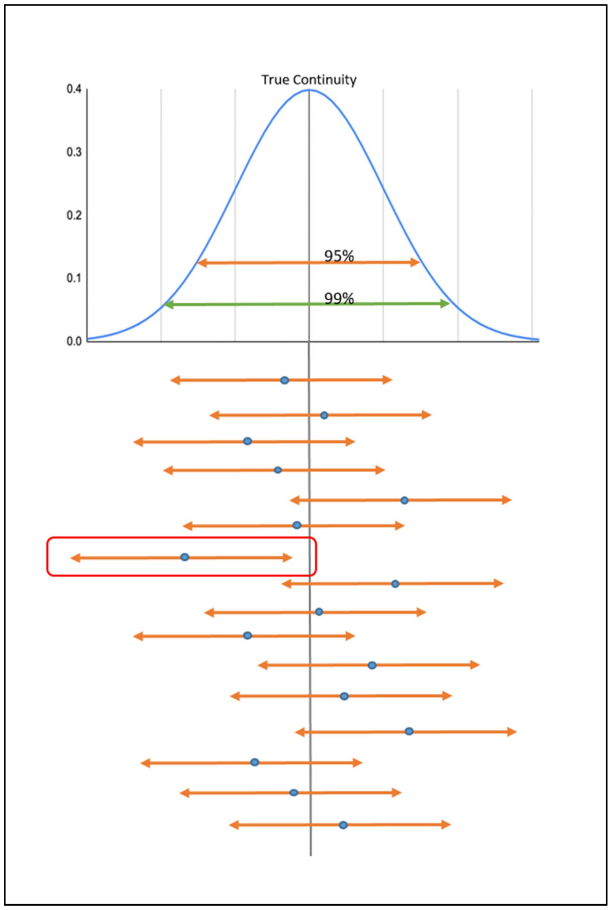
Looking at the above example, one would be tempted to increase the confidence level to a higher value so that we could make the system compliant! This would however lead to a less reliable assessment. This is the scope of the next section.

A.3.4 Types of errors

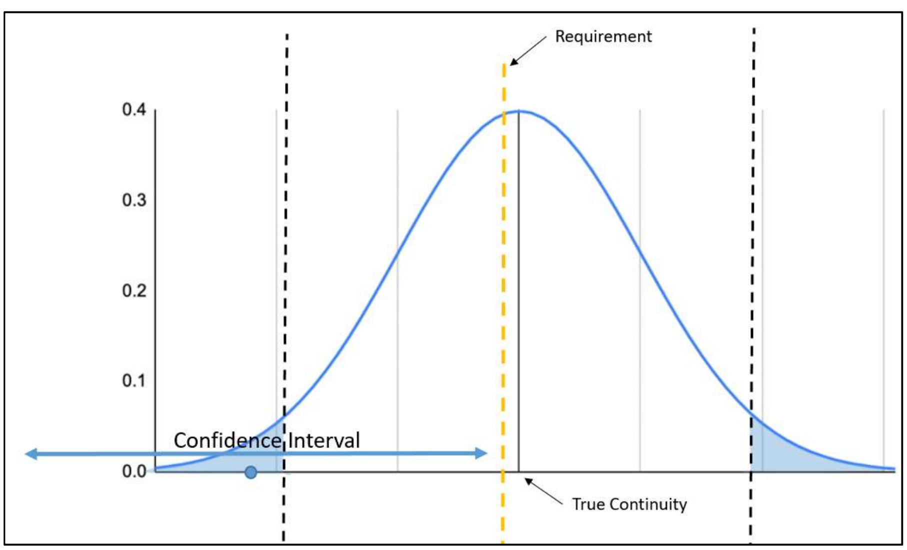
There can be two types of errors: when we state something is false when in fact it is true (known as a Type-I error) or when we state something is true when in fact it is false (a Type-II error).

A.3.4.1 Type I errors

For a 95% confidence level (orange arrow), we expect the True Continuity to lay within the confidence interval 95% of the time. As illustrated in the graph below, in 5% of cases it will not. So if we state the system is not compliant because the Required Continuity is outside the confidence interval, we are making an error of Type-I 5% of the time.



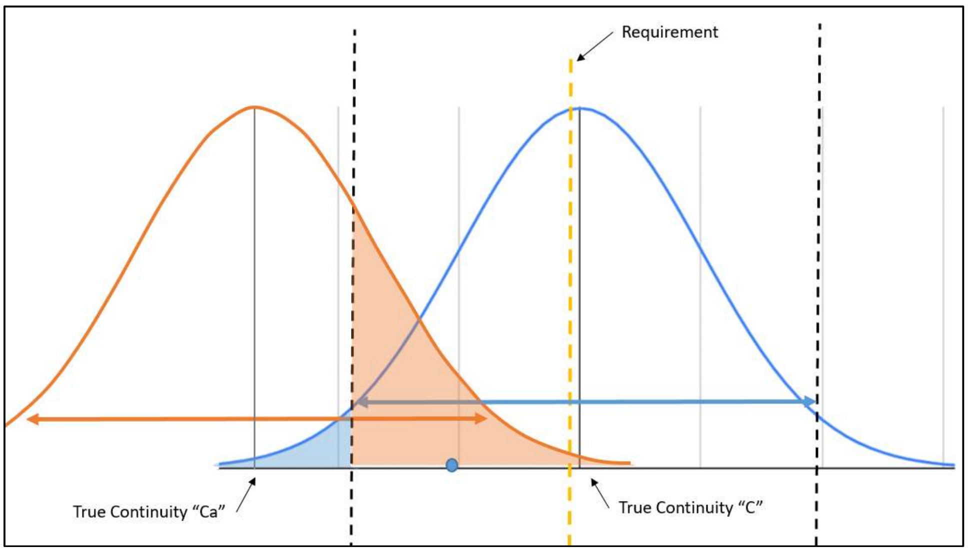
Looking at the distribution of Continuity estimates (graph below), Type-I error are observed when the estimate (blue dot) lies outside of the confidence interval around the true Continuity. The two blue shaded regions at the end tail of the distribution refers to the probability of having a Type-I error.



The probability of making a Type-I error is the complementary value of the confidence interval (1-CL). In the case of a 95% confidence level, the probability of making a Type-I error is 5% - which should be equally split between the two tails of the distribution (2.5% for each blue area). In ED-228A this probability is denoted by the Greek letter “α”.

A.3.4.2 Type II errors

The next graph explains Type-II errors. If the estimate (blue dot) lays within the confidence interval of the blue distribution (letting us think that it represents the True Continuity “C” as above) when in fact it is an estimate of the orange distribution (it is an estimate of what is in fact the real True Continuity “Ca”), then we are making a Type-II error. In this case, we state that the system is compliant when in fact it is not. The orange shaded area is the probability of making a Type-II error. In ED-228A this probability is denoted by the Greek letter “β”.



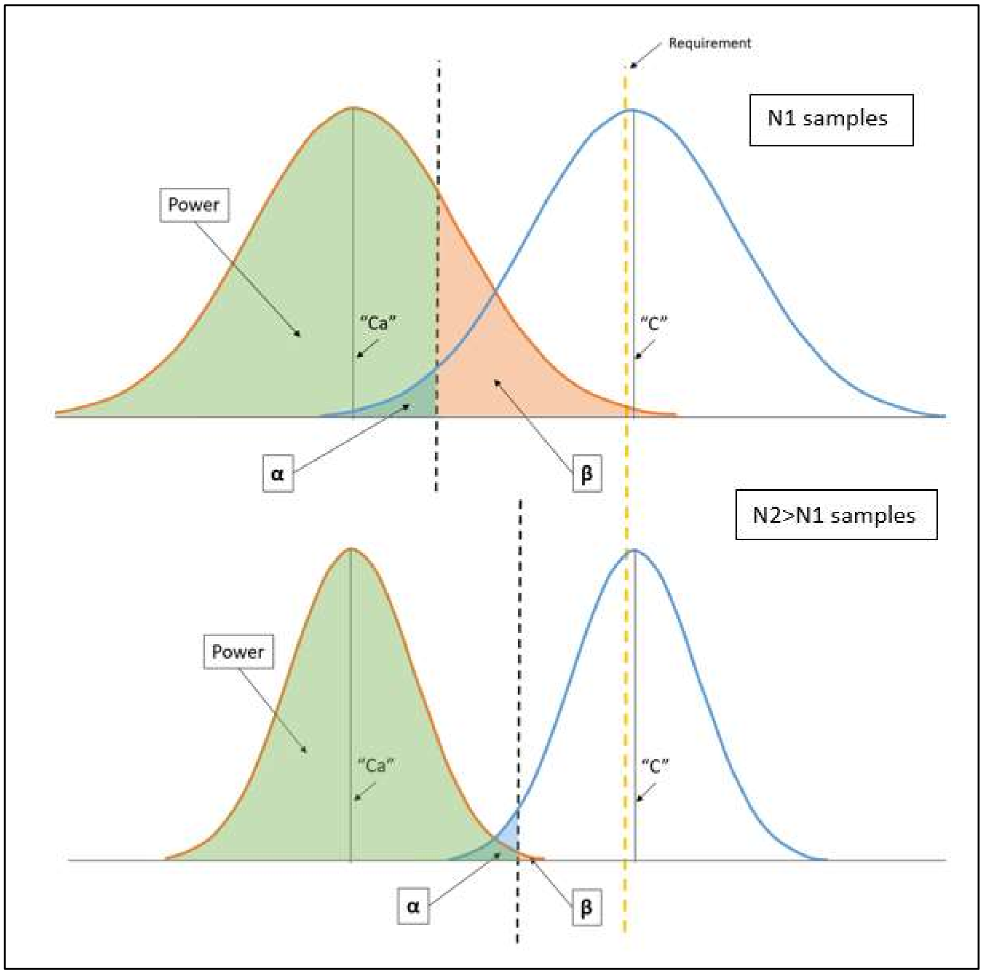
One can easily observe that increasing the confidence level (reducing the blue shaded areas) will decrease Type-I errors but increase Type-II errors (increasing the orange shaded area). On the other hand, decreasing the confidence interval will increase Type-I errors and decrease Type-II errors.

We now understand that we cannot infinitely increase the confidence interval without increasing Type-II errors. The following discussion explains how to deal with this problem.

A.3.5 Power of a test

The “power” of a test is the complementary value of the probability of making a type-II error, thus (1-β) i.e. it is the probability that we are not making a type II error. It is a measure (probability) of how confident we are when we state the system is compliant. The higher the power the more confident we are in our assessment. In the following graphs it is represented by the green shaded area.

Increasing the power can be done by increasing the difference between “C” and “Ca” or by increasing the number of samples for the Continuity estimate. The latter will decrease the “width” of the distribution of the estimates, as shown in the following graphs, and is the preferred method to increase the power.



A.3.6 Discussion

Now that we have introduced the statistical principles needed to understand the compliance assessment from monitoring as proposed in ED-288A, we can now proceed in discussing the proposed method.

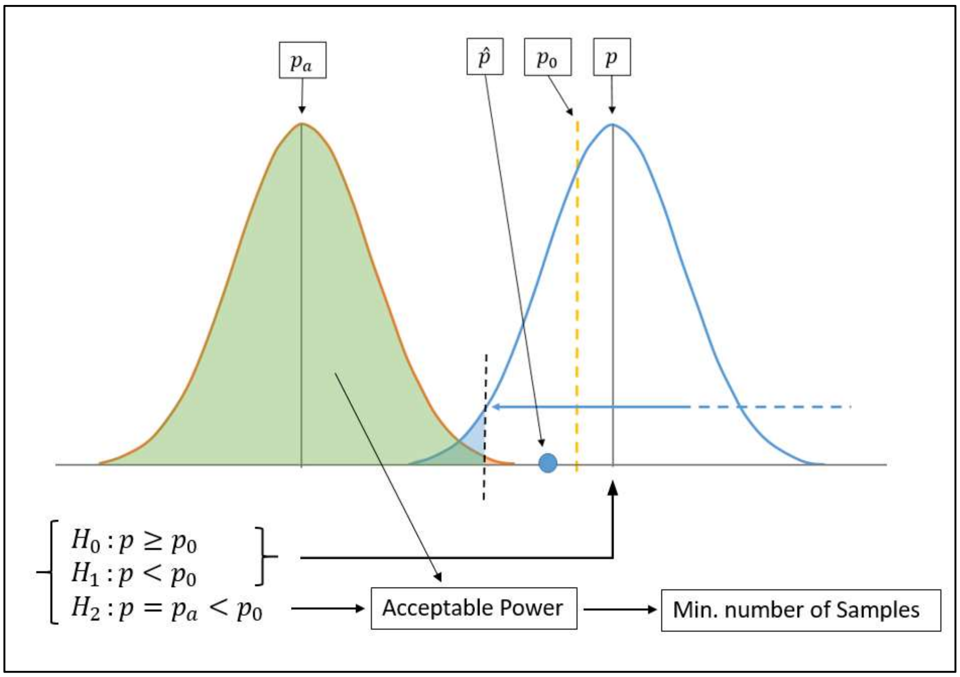
As stated at the beginning of this discussion, all the "maths" can be found in section D.7.5 of ED-228A. We provide here a cross-reference table of the terms used across the discussion and the ones used in ED-228A.

ED-228A	This document
p	True Continuity "C"
$\hat{p}$	Continuity estimate
p_0	Required Continuity value (target)
p_a	True Continuity "Ca" from the "orange" distribution of Continuity estimates
n	The number of samples used to compute a Continuity estimate

As we have been able to understand from the earlier explanation, assessing compliance to performance requirements from monitoring requires an appropriate "tuning" of our statistical parameters.

In ED-228A, compliance is proposed to be assessed using hypothesis testing. A hypothesis test is used to state, based on our estimate and some specific assumptions, if our system is likely to be compliant or not.

The proposed test in ED-228A is composed of 3 hypotheses. As described in the following picture, the first two hypotheses (H_0 and H_1) are used to assess compliance (as in a normal test) while the last one (H_2) is used to assess the "power" of the test. The minimum required power for a test is often used to compute the required number of samples needed to perform the test.



Without entering into too much detail (see [Fleiss][Brown][Wallis] for details), we need to note that looking at a confidence interval or performing a hypothesis test can be considered to be equivalent (assuming the confidence interval is built from a hypothesis test) . So it is proposed to assess compliance using a confidence interval instead of using the hypothesis testing described in section D.7.5.3.1 of ED-228A. It will then be easy to assess compliance by simply checking that the required value lies within the confidence interval.

It is important to understand that the hypothesis test in ED-228A (like the requirement itself) is a "one-sided" test ($H_0 : p \geq p_0$ instead of $H_0 : p = p_0$ for a "two-sided" test). This is equivalent to an "open" confidence interval ranging from -infinity to the upper confidence value ($[-\infty; U.L.]$). Practically, as proportions are bounded between 0 and 1, the interval becomes ($[0; U.L.]$. This has an important implication in the computation of the confidence interval.

Several methods are available to compute confidence intervals. The method presented in ED-228A is called the "standard" method, and although very simple to use, it is generally admitted today that it should no longer be used for proportions [Brown][Wallis] especially outside of the proportion range (0.3 to 0.7). We propose to use the method proposed by Wilson that is generally agreed to be more accurate even for low “n” and for proportions near “0” and “1” [Brown][Wallis].

The main advantage of the Wilson method is that the confidence interval can be derived from “inverting” a hypothesis test. It is guaranteed to obtain the same result as the equivalent “z-test” or “chi-squared” test. The confidence interval can then be directly used to asses compliance by “seeing” if the requirement lies within the confidence interval.

As the hypothesis test is “one-sided”, only the upper level of the confidence interval is needed for the compliance assessment. However, both levels can be computed to show the true confidence interval for example for displaying on a graph.

Wilson confidence interval formula for upper (+ sign) and lower (- sign) levels

$$W_{-,+} = \frac{\left(\hat{p} + \frac{z_\alpha^2}{2n}\right) \pm z_\alpha \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_\alpha^2}{4n^2}}}{\left(1 + \frac{z_\alpha^2}{n}\right)}$$

Where $z_\propto$ is a constant taken from statistical tables (see below).

The above formula takes into account the Yate’s correction for “continuity”[23] and is usually named the “Wilson score interval with continuity correction”. For the sake of simplicity, the “Wilson confidence interval” naming will we kept in this document.

A.3.7 Minimum required number of samples

The minimum required number of samples has been a source of much discussion.

From an estimation point of view (we want to estimate the True Continuity), the more the samples the better your estimate is. As explained at the beginning, as you increase the number of samples used for an estimator you also decrease the “width” of the distribution of estimates. As this “width” is related to the confidence interval, the latter relates to the precision on your estimate; as you continuously increase the number of samples, you tend to a value closer to true size of the population you want to estimate (transaction delay). If you were able to get all the samples (assuming a finite set) of a population (transaction delays), your Continuity “estimate” will become the “True” Continuity.

However, as there is no guidance on how to choose an adequate confidence interval “width” in the literature (which is usually a user’s choice), the minimum required power of a test is often used to compute the minimum required number of samples. Formula (28) of section D.7.5 in ED-228A can be used to calculate the minimum sample size, but it requires that values be chosen for a set of parameters that are not trivial to set and so we propose a much simpler approach.

[23] When the Normal distribution is used as an approximation to the Binomial distribution a small error is introduced into the calculation that needs to be corrected (see [Wallis]).

Building an "exact" confidence interval[24] requires the use of complex mathematics. Even with the increase of computational power available nowadays, there is still a need for simple and reliable approximations. The "standard" method (as in ED-228A) and the Wilson method (as proposed here) are based on a Gaussian approximation to the binomial distribution[25]. The literature [Fleiss][Brown] often assumes this Gaussian approximation to be true if

$\min(n\hat{p}, n(1-\hat{p}))$ is greater than 5 (or 10)[26]

This formula leads to a minimum sample size of 5000 (or 10000) for a 99.9% Continuity estimate and 100 (or 200) for a 95% one[27].

One should note that the minimum number of samples computed using formula (28) of section D.7.5. in ED-228A (based on the power of a test) [28] and the ones computed from the "Gaussian approximation" hypothesis are very close to each other.

Moreover, the proposed confidence interval methodology (Wilson) is an accurate approximate (close to the "exact" one) for low "n" and for proportions near "0" and "1" [Brown][Wallis]. This, combined with the relative simplicity of the method, is the reason why we propose to determine the sample size based on the Gaussian approximation to the Binomial distribution.

A.3.8 Recommended compliance assessment

The following compliance assessment method is valid for the assessment of any proportion.

The method consists of calculating the 95% confidence interval of the measured value and then determining whether the required value lies within the confidence interval. If it does, and the sample size is sufficiently large, the performance is considered compliant with a confidence level of 95%. If the requirement is not contained within the confidence interval, the performance is not considered to have demonstrated compliance. In the case where the sample size is not sufficient, the assessment should be done with care and conclusions might not be derived.

The method is described in more detail below using the RCTP_TC (32) and RCTP_TC (20) continuity metrics as examples.

It is proposed to use a confidence level of 95% (leading to $\propto = 1 - 0.95 = 0.05$) to compute the confidence interval and to use a minimum number of samples of 10000 to assess the RCTP_TC(32) value (as it is required to be 99.9% or better) and 200 samples for the RCTP_TC (20) as it is required to be 95% or better.

---

[24] For a binomial distribution, the "exact" confidence has been derived by Clopper and Pearson, leading to the name "Clopper-Pearson" confidence interval. This confidence interval is always "exact" whatever the number of samples or proportion (even close to '0' and '1').

[25] The correct characterisation of the Gaussian approximation used in the Wilson confidence interval is somehow counter-intuitive and is well explained in [Wallis].

[26] Five and ten are empirical values. We propose to use the more stringent one (10). We then have $\geq \frac{10}{(1-\hat{p})}$, where $\hat{p}$ is 0.95 or 0.999 for respectively the 95% and 99.9% Continuity.

[27] The reader is invited to note that there is no criteria to validate the Gaussian approximation to the Binomial approximation in ED-228A.

[28] using a power of 95% (β=0.05), p_a=0.94 for 95% and p_a=0.998 for 99.9% and z_α=z_β=1.645

The confidence interval is calculated for both RCTP_TC(32) and RCTP_TC(20) using the formula below:

$$W_{-,+} = \frac{\left(\hat{p} + \frac{z_\alpha^2}{2n}\right) \pm z_\alpha \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_\alpha^2}{4n^2}}}{\left(1 + \frac{z_\alpha^2}{n}\right)}$$

Where n is the number of transactions used in the measurement, $\hat{p}$ is the measured value of RCTP_TC(32) or RCTP_TC(20) and the value[29] of z_α is a constant taken from a statistical reference table.

The following tables summarises the recommended parameters.

Parameter	Recommended value
α	0.05 (5%)
N_{min} (99.9%)	10000 samples
N_{min} (95%)	200 samples
z_α	1.644853627

Table 1: Confidence interval parameters

Practical example

The following table provide an example of the above methodology for a RCTP Continuity.

Input Parameters		
Parameter	**Value**	**Comments**
α	0.05	95% Confidence Level
z_α	1.644853627	From table
N_{min}(95%)	200 samples	
N_{min} (99.9%)	10'000 samples	
RCTP Continuity(20s)	0.95	From requirements
RCTP Continuity(32s)	0.999	From requirements
Measurements		
Parameter	**Value**	**Comments**
N TRN	210734 samples	Greater than 10000
N TRN <= 20s	206446 samples	
N TRN <= 32s	207752 samples	
Calculations and Assessment		
Parameter	**Value**	**Comments**
$\hat{p}$ (20s)	0.979652	$N_{\le 20s}/N_{TRN}$
Wilson L.L	0.979140	
Wilson U.L.	0. 980152 **>** 0.95 (**PASS**)	
$\hat{p}$ (32s)	0.985849	$N_{\le 32s}/N_{TRN}$
Wilson L.L.	0.985420	
Wilson U.L.	0.986266 **<** 0.999 (**FAIL**)	

Summary

[29] The value of z is computed using the following formula: $z = \sqrt{2} \cdot erfinv(1 - 2\alpha)$, where "erfinv" is the inverse error function (Gauss error function). These can be taken from statistical look up tables.

Parameter	Value ($\hat{p}$)	Wilson CI (CL=95%)	Assessment
RCTP Continuity(20s)	97.97%	(97.91%, 98.02%)	PASS
RCTP Continuity(32s)	98.585 %	(98.542%, 98.627%)	FAIL

The RCTP Continuity (20s) is estimated 97.97% with an upper limit confidence interval of 98.02%. As the latter is above the requirement of 95%, we can state that "***with a 95% confidence level, the RCTP Continuity (20s) is met***."

The RCTP Continuity (32s) is estimated 98.585% with an upper (time stamp corrected) limit confidence interval of 98.627%. As the latter is below the requirement of 99.9%, we can state that "***with a 95% confidence level, the RCTP Continuity (32s) is not met***."

Notation

Multiple notations are available for confidence intervals. [Wallis] even propose that, in scientific reporting, the estimate value itself should be replaced by the confidence interval. However, we propose to display both information; the estimate itself and the Wilson CI using "parenthesis" notations (see example below).

It is further proposed to round values down to 2 additional digits than the requirement ones (0.123 for 95% and 0.12345 for 99.9%)

A.3.9 Further considerations

As stated before, the methodology above is valid for any proportions. It is then applicable to the following KPI metrics:

- Ground Initiated Transaction Continuity
- RCTP Technical Continuity
- CSR Success Rate

Further discussions are needed on how to apply this methodology to the "Active Session Provider Abort Rate".

The accuracy of the timestamp used in delay measurements could be a source of error in Continuity assessment. However, we assume that the delays are measured from timestamp coming from the same "clock" source so that the error in estimating the delay is negligible.

A.3.10 References

[Fleiss] Statistical Methods for Rates and Proportions, Third Edition, J.L. Fleiss, B. Levin and M.C. Paik, John Wiley & Sons, Inc., 2003.

[Brown] Interval estimation for a binomial proportion, L.D. Brown, T.T. Cai and A. DasGupta and comments from A. Agresti and B.A. Coull, G. Casella, C. Corcoran and C. Mehta, M. Ghosh, T.J. Santner in Statistical Science, 2001, Vol.16, No 2, 101-133.

[Wallis] Binomial confidence intervals and contingency tests: mathematical fundamentals and the evaluation of alternative methods, S. Wallis, Survey of English Usage, University College London.

Appendix 4 Ideas for future metrics

This section provides a brief description of some other ideas for metrics which may be considered useful but for which no definition has yet been agreed.

A.4.1 CPDLC Operational Usage

This metric is intended to provide a measure of how much operational usage is made of CPDLC by the controllers and flight crew. It would measure the number of uplink and downlink operational messages per 100 hours of CPDLC usage, perhaps broken down by flight level band.

A.4.2 Controller workload savings

Some work has been done by DFS to analyse the amount of controller time (or workload) saved when using data link rather than voice, so this would provide a useful benefit metric, but it is thought the work done by DFS may need to be broadened to consider the way other ANSPs operate.

A.4.3 Datalink to voice clearance ratio

A metric to establish what percentage of clearances are issued by voice and by data link would be useful to measure how widely CPDLC is being used, but it is considered difficult to collect the information about the clearances issued by voice.

A.4.4 Ground/Air Initiated Continuity Problem

The Ground/Air Initiated Transaction Continuity metric measures the rate of unsuccessful uplinks/downlinks. This metric would measure the proportion of the different reasons why an uplink/downlink was unsuccessful e.g. response was received too late, an ERROR message was received instead, or no response was received at all.

A.4.5 Avionics support of key protocol features

ANSP metrics on the % of avionics supporting some key protocol features such as: correct ADM_ARS, correct aircraft position (Lat, Lon, Alt and destination airport). It would provide an estimation of the population of aircraft adhering to the Standards and explain some performance differences.

EUROCONTROL

SUPPORTING EUROPEAN AVIATION

www.eurocontrol.int