ITUPublications
Recommendations

International Telecommunication Union
Standardization Sector

Recommendation

ITU-T P.1503 (03/2023)

SERIES P: Telephone transmission quality, telephone installations, local line networks

Methods for objective and subjective assessment of quality of services other than speech and video

---

Extended methodology for cross-country and inter-operator digital financial services testing

ITU-T P-SERIES RECOMMENDATIONS

TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	P.10–P.19
Voice terminal characteristics	P.30–P.39
Reference systems	P.40–P.49
Objective measuring apparatus	P.50–P.59
Objective electro-acoustical measurements	P.60–P.69
Measurements related to speech loudness	P.70–P.79
Methods for objective and subjective assessment of speech quality	P.80–P.89
Voice terminal characteristics	P.300–P.399
Objective measuring apparatus	P.500–P.599
Measurements related to speech loudness	P.700–P.709
Methods for objective and subjective assessment of speech and video quality	P.800–P.899
Audiovisual quality in multimedia services	P.900–P.999
Transmission performance and QoS aspects of IP end-points	P.1000–P.1099
Communications involving vehicles	P.1100–P.1199
Models and tools for quality assessment of streamed media	P.1200–P.1299
Telemeeting assessment	P.1300–P.1399
Statistical analysis, evaluation and reporting guidelines of quality measurements	P.1400–P.1499
Methods for objective and subjective assessment of quality of services other than speech and video	**P.1500–P.1599**

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

Recommendation ITU-T P.1503

Extended methodology for cross-country and inter-operator digital financial services testing

Summary

Recommendation ITU-T P.1503 is based on Recommendation ITU-T G.1033, which standardizes a conceptual framework for quality of service and quality of experience (QoE) aspects of digital financial services (DFSs), and on Recommendation ITU-T P.1502, which specifies a methodology for QoE testing of DFSs for the basic person-to-person (P2P) money transfers between two devices using the same network and DFS operator.

Recommendation ITU-T P.1503 has three main parts.

First, the methodological framework and specifications for a generalized P2P money transfer use case are given. In this framework, the DFS operator used to send money, and the operator receiving the same funds (i.e., the A and B side of a money transfer) are parameters in the use case, which integrates all variations (same operator or inter-operators; intra- or international) into the same methodological context.

Second, Recommendation ITU-T P.1503 provides a comprehensive framework for data elements and related processing, including tools and procedures providing operational robustness and a high level of data quality. The data objects specified here support test planning and management as well as provide the input data foundation for efficient processing of data. Guidance is also given on how data processing can be done in a consistent and efficient way.

Third, this Recommendation describes a new tool and the corresponding methodology, designed to assist field test teams in data collection. This "multi-stopwatch" tool is conceptually suggested in Recommendation ITU-T P.1502. It is an electronic time-taking tool similar to a stopwatch but supporting testers to record the events within a DFS test case and upload data entered by testers directly to a central location. This tool eliminates the needs and weaknesses of manual entering of time readings and the process of transferring them to post-processing through multiple transformation stages, e.g., from handwritten notes to entries in a spreadsheet.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.1503	2023-03-01	12	11.1002/1000/15471

Keywords

Digital financial services, QoE, QoS, quality of experience, quality of service.

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

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

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

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

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

NOTE

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

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

INTELLECTUAL PROPERTY RIGHTS

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

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the appropriate ITU-T databases available via the ITU-T website at http://www.itu.int/ITU-T/ipr/.

© ITU 2023

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Table of Contents

		Page
1	Scope	1
2	References	1
3	Definitions	2
	3.1 Terms defined elsewhere	2
	3.2 Terms defined in this Recommendation	2
4	Abbreviations and acronyms	2
5	Conventions	3
6	Test scenario under consideration	3
	6.1 Roles, entities and action or /event flow	3
	6.2 Test parameterization and neutral starting state	4
	6.3 Re-initialization after unsuccessful transactions	5
	6.4 Disappearing money	5
	6.5 Automation of tests	5
7	Transaction model and digital financial service key performance indicators	6
8	Creating the use case model from actual use case examples	8
9	Data sources	8
10	Special procedures in the field	9
	10.1 Operational protocol if party A and B do not communicate directly	9
11	Testing modes	11
	11.1 Basics	11
	11.2 Synchronized and asynchronous testing mode	11
	11.3 Dual-function set-up and operation	13
12	Recording events of DFS test cases	14
	12.1 Basic functionality	14
	12.2 Practical application	14
13	Measurements in the background	15
14	General considerations about errors in measurements	16
15	Data validation and processing	17
	15.1 Overview	17
	15.2 Plausibility and validity checks	18
	15.3 Data processing	19
	15.4 Time profile	20
16	Background testing of mobile networks	20
Appendix I – Working example for creation of a use case model from actual implementations		22
Appendix II – Best-practice example of data structures and procedures used in DFS testing campaigns		23

		Page
II.1	Basic considerations for data collection and processing	23
II.2	Additional considerations for cross-border and inter-network testing	23
II.3	Data structure overview	25
II.4	Naming and formatting conventions	26
II.5	Naming and formatting conventions for data objects	27
II.6	Local log sheets	27
II.7	Team assignment list	28
II.8	Device assignment list	30
Appendix III – Best-practice examples for background testing of mobile networks		33
Bibliography		35

Electronic attachment: Spreadsheet log file templates

Recommendation ITU-T P.1503

Extended methodology for cross-country and inter-operator digital financial services testing

1 Scope

This Recommendation[1] describes the quality of experience (QoE) assessment methodology for the person-to-person (P2P) money transfer use case, i.e., from party A to party B. It covers the full range of possible combinations of operator of digital financial services (DFSs) and country where the parties are located, i.e., for the case where parties A and B use the same or different operators, as well as intra- and international transfers. Therefore, this Recommendation is an extension of [ITU-T P.1502].

This Recommendation also specifies extensions of testing methodology and introduces a new type of tool to assist in testing.

The methodology is designed to be easily extended to other use cases in future editions of this Recommendation.

It is important to understand that this Recommendation only covers the methodology for tests done from an individual user's (end-to-end) perspective, acting within a given DFS ecosystem under current load conditions.

It may be desirable to extend the scope of testing to capacity tests, which would involve creation of specified load scenarios to a DFS ecosystem to determine the robustness of its functionality under these conditions. Such extensions can be easily created from the methodology described in this Recommendation. Their execution is mainly a matter of scale of required resources.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T E.804] Recommendation ITU-T E.804 (2014), *Quality of service aspects for popular services in mobile networks*.

[ITU-T E.840] Recommendation ITU-T E.840 (2018), *Statistical framework for end-to-end network-performance benchmark scoring and ranking*

[ITU-T G.1033] Recommendation ITU-T G.1033 (2019), *Quality of service and quality of experience aspects of digital financial services*.

[ITU-T P.1502] Recommendation ITU-T P.1502 (2020), *Methodology for QoE testing of digital financial services*.

[1] This Recommendation has an electronic attachment containing spreadsheet log file templates for project usage.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 quality of experience (QoE) [b-ITU-T P.10]: The degree of delight or annoyance of the user of an application or service.

3.1.2 quality of service (QoS) [b-ITU-T E.800]: Totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service.

3.2 Terms defined in this Recommendation

This Recommendation defines the following term:

3.2.1 digital financial service (DFS): Methods to electronically store and transfer funds; to make and receive payments; to borrow, save, insure and invest; and to manage personal or business finances; and where these services are accessed via mobile communication devices.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AVDTA	Accumulated Validated Data on Transaction
DAL	Device Assignment List
DFS	Digital Financial Service
FTL	Field Test Lead
GPS	Global Positioning System
ID	Identifier
IMEI	International Mobile Equipment Identity
KPI	Key Performance Indicator
MDR	Mean Data Rate
MSW	Multi-Stopwatch
MTACT	Money Transfer A-side Completion Time
MTCD	Money Transfer Core Duration
MTCR	Money Transfer Completion Rate
MTFCT	Money Transfer Full Completion Time
MTRCT	Money Transfer Raw Completion Time
P2P	Person-to-Person
PIN	Personal Identification Number
QoE	Quality of Experience
QoS	Quality of Service
SMS	Short Message Service
SQL	Structured Query Language
TA	Transaction

TAL	Team Assignment List
USSD	Unstructured Supplementary Service Data
UUID	Universally Unique Identifier
Wi-Fi	Wireless Fidelity

5 Conventions

None.

6 Test scenario under consideration

The basic scenario under consideration is the P2P money transfer between two parties in the following cases:

– in the same country, but using different DFS providers (inter-operator scenario);

– in different countries, using the same DFS providers (e.g., national branches of a multi-national network operator);

– in different countries, using different DFS providers.

It is important to state that the methodology for these variants is the same since in all cases money is transferred between two entities. There may be differences in the details of an operating sequence; these are, however, not greater than differences between same-operator, same-country operating sequences.

NOTE – In some countries, the DFS might be registered under the central bank, and money transfers considered to be bank transfers. This can have an effect on the appearance of such transactions where, in actual implementations, it has been shown that in some cases entities other than the sending party (e.g., agents) are used for international transactions. From an end-to-end perspective, this is still a P2P transfer; technically, money received in this case appears to come from that particular entity rather than from the actual sending entity. If reception notifications are used for data evaluation (which is not the case in the current methodology), this would have to be considered in data processing.

The P2P basic scenario and its modelling is described in detail in [ITU-T G.1033] and [ITU-T P.1502]. For the sake of convenience of reading, the essential parts are explained here. For more details, see [ITU-T G.1033] and [ITU-T P.1502].

6.1 Roles, entities and action or /event flow

In the P2P money transfer scenario, money is transferred from party A (the active party which is sending money) to party B (the receiving party).

In a practical test implementation, each party is represented by one testing team. In practice, money is transferred in a cyclic fashion, so teams switch roles after each transfer, as shown in Figure 1.

NOTE – Figure 1 only shows the basic case. In order to take care of the whole spectrum of possible cases during testing, some additional, derived cases need to be considered.

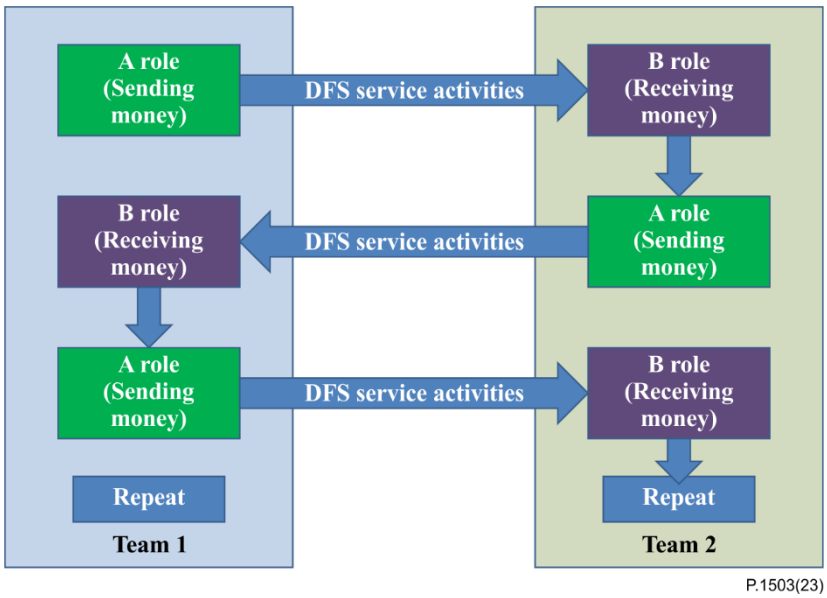


Figure 1 – P2P money transfer roles and team activities

6.2 Test parameterization and neutral starting state

A particular property of systematic service tests is frequency of usage that is significantly higher than that created by a typical end user.

While a high testing frequency leads to a high yield of samples for computation of quality of service (QoS) key performance indicators (KPIs), it is conceivable that the system has a certain dead time after each transaction, where the system does not accept a new transaction or creates unexpected results of a transaction attempted within this time period. It is advisable to be aware of this possibility and obtain relevant information before actual parameters of a test campaign are determined.

The testing frequency can be controlled by a pause between transactions, which also acts as a guard time to allow the service under test to reach its neutral state again. Respective considerations are in full analogy to testing e.g., of telephony.

A testing campaign, therefore, should contain a pre-testing phase with systematic tests to make sure that usage frequencies typical for testing do not affect testing results with respect to the end-user perspective.

NOTE – In actual implementation, DFS providers can apply limitations on the number of transactions per day, or the total amount of transferred money per day or another time period. It is also conceivable that mechanisms exist that limit the frequency of testing. When setting up tests, it is important to check for such conditions. This starts with making testers aware of the effects of such mechanisms and their preparation to adapt data processing accordingly after detection during a testing campaign. Otherwise, misclassification can result, such as ascribing effects of such limitation to functional failures of a DFS.

As the starting point for systematic testing, the guard time should typically be in the range 10 s to 30 s.

When testing is done manually, it is assumed that the system can handle all testing speeds that can be realized by human testers, as even an experienced tester will not work significantly faster than an experienced regular user of DFS. Therefore, no special requirements to slow down testing are applied.

In fully automated testing, it would also be possible to use the high degree of repeatability of such control to determine the appropriate guard time by probing, i.e., by systematically varying the guard time and check for respective effects.

There is a second category of effects that needs to be considered, namely the possibility of a service-specific local memory (analogously to a browser's cache) that stores information related to previous

transactions. The effect would be that in subsequent transactions, such information would be read from local memory instead of obtaining it by an over the air request to the service. This could then impact related measurement values or KPIs.

As long as effects are quantitative rather than qualitative, it may not practicable and is not necessarily required to exclude frequency-dependent effects entirely. However, such effects need to be recorded and documented carefully as part of the reporting in order to understand their impact on the testing conditions.

6.3 Re-initialization after unsuccessful transactions

If a transaction fails, in particular after a time-out condition has occurred, ensure that the service and the device or application are in the typical neutral starting state again, i.e., that no memory of previous error states remains in the system.

6.4 Disappearing money

It is possible that during a transaction, the amount of money deducted is not correct with respect to transferred amount and fees. This includes the case that the amount is correct but sent to a third party by an error in the system. From an end customer perspective, this is either a loss (if too much money is deducted), or an unjustified gain (if money is credited but not deducted on the other side of the transaction. For simplicity, the term "disappear" is used for both variants of this kind of effect.

There may be undelivered transactions where money is deducted from the sender's account along with transfer charges. In such cases, filing of a complaint with the DFS operator is typically required. If this complaint is resolved successfully, money is returned at a later point in time (depending on the process and the terms of service of the DFS operator, transfer charges may not be refunded).

Retrieval of lost money is understood as a second stream of activities outside the scope of this methodology. Functionally, even if money is returned later, it reduces the available credit for further tests. Therefore, in all cases of disappearing money, insertion of fresh funds may be necessary to keep up the necessary level of credit for further testing.

The matter of transaction failures needs special consideration. In that case, it is assumed that a typical user seeks confirmation, e.g., by calling or messaging the recipient (i.e., using an external means of communication). In many cases, the recipient also issues a receipt confirming incoming payments. The sending party might wait for that statement for a certain time period, and enquire about one from the recipient if none is received.

In any case, in particular in testing modes where party A has no direct visibility of events on party B (this issue is also discussed in clauses 11, 13, 15 and 16), reasonable and appropriate measures and conventions, adapted to the actual scope and goals, shall be considered and set-up as part of a testing campaign.

6.5 Automation of tests

The methodology in this Recommendation describes generic testing, i.e., service tests can be done manually as well as automatically. It is understood that automation of tests is desirable to achieve a greater degree of repeatability, and less variation in quantitative data values due to inaccuracy, e.g., of manual time measurements.

Automation can have different forms with respective degrees of automation up to fully automated testing. Using the multi-stopwatch (MSW) concept as described in this methodology is the next step of evolution, significantly improving the robustness of testing with respect to manual event logging.

The next step may be to still use manual operation of transactions but to record low-level activities on the DFS device itself, e.g., from recording of layer 3 messages or Internet protocol-level activities. The ultimate goal is a system that executes all DFS processes automatically. Such implementations

require, at least, an extended level of access to platform devices ("rooting", e.g., having system-level access) and substantial technical efforts, in particular because testing such systems requires actual use of DFSs. Making these systems fit for unsupervised operation or to enable operation on a larger number of mobile device types, would further increase the necessary effort.

Design or implementation of further steps of automation, or related evolution of the methodology lies outside the scope of this Recommendation.

7 Transaction model and digital financial service key performance indicators

The basic model is identical to that described in [ITU-T P.1502]. This clause describes the special variant for inter-operator and cross-border testing where teams do not have direct contact with each other.

In the following, the basic considerations and principal definitions from [ITU-T P.1502], as well as extensions, are described. For a more thorough background reading, as well as for the full set of DFS KPIs, see [ITU-T P.1502].

Figure 2 shows the basic structure and event flow of the DFS implementation; the collection of required details ("Collect transaction data") is shown summarily; the details are different between operators.

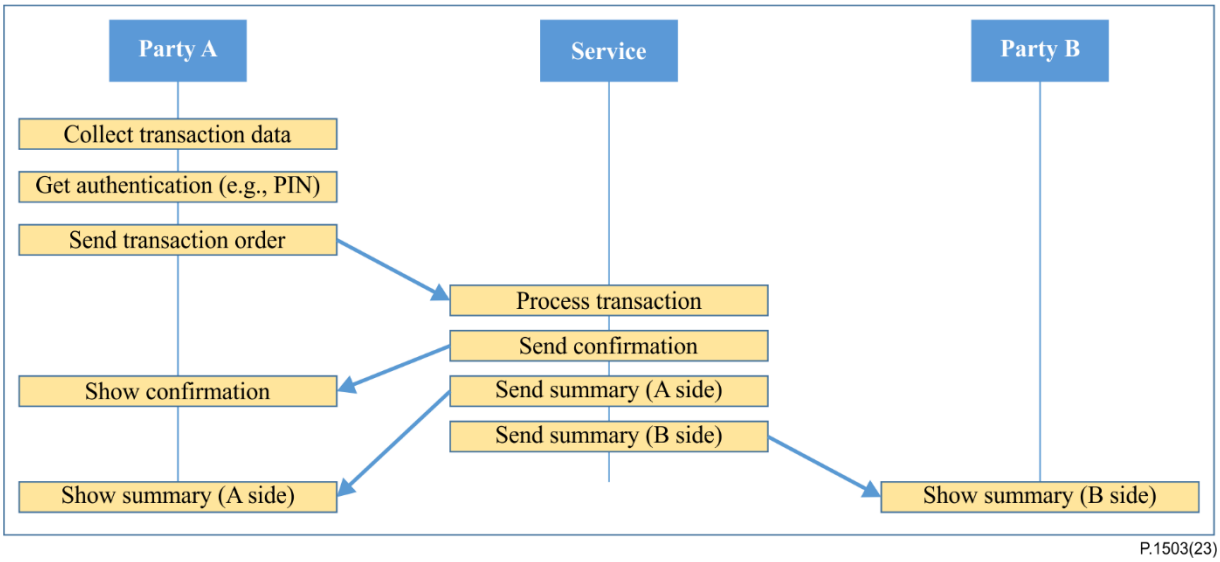


Figure 2 – Basic model of a P2P money transfer implementation

Figure 3 shows how the action and event flow is mapped to the "timer flag" elements that are then used to compute the KPIs. Again, the basic processing is the same as that for the basic P2P case.

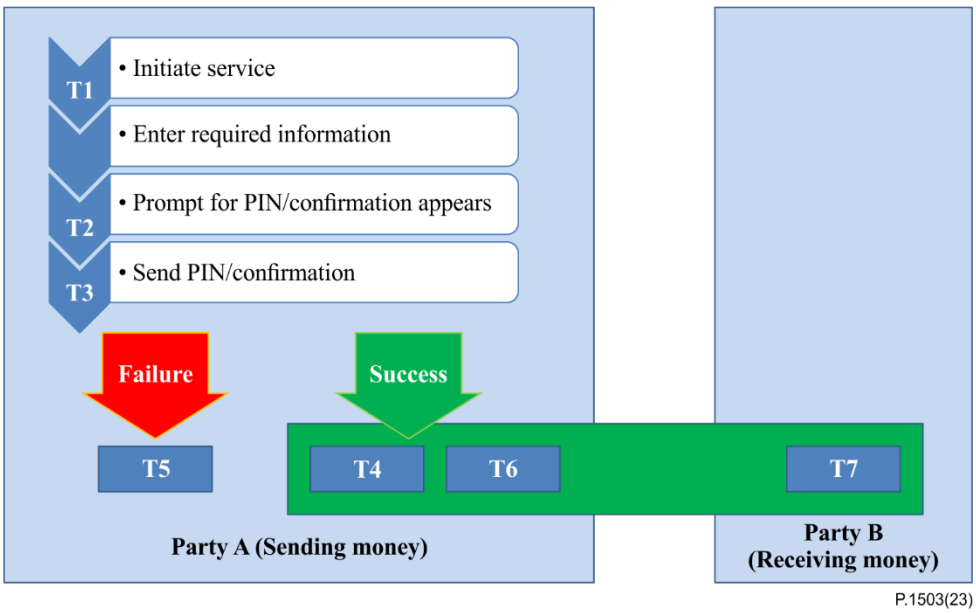


Figure 3 – Connection between events and timer flags (see text for details)

T4 is the primary (in-application) success criterion while T6 refers to the success criterion provided by receipt of a notification short message service (SMS). Please note that in this generalized case, T4 and T6 can appear in any combination and order.

The principal difference to an intra-country, intra-network is that most likely teams are working in different locations.[2] Therefore, not all of timer flags are recorded in the same place. The practical consequence is that data has to be combined from different sources, which is described in more detail in clause 15.

It is important to keep in mind that due to the different paths for events T4, T6 and T7, they can appear in any order.

Due to the different types of implementation, it is possible that on the party A side, either T4 or T6 is missing. For data processing, this means that transaction success is indicated by either criterion.

The KPIs used are a subset of the simplified set defined in clause 10 and Table 5 of [ITU-T P.1502].

NOTE – The event names are those specified in this Recommendation; the set differs from that in [ITU-T P.1502].

[2] [ITU-T P.1502] assumes that both the party A and B devices are in the same place. Of course, even for national, intra-network testing it is possible that operation in different places is desired. Therefore, this Recommendation provides an extension to [ITU-T P.1502] even in this basic case.

Table 1 – Simplified set of key performance indicators used in this Recommendation, based on the set in [ITU-T P.1502]

Indicator	Abbreviation	Definition/Remark
Money transfer core duration [s]	MTCD	T4-T3
Money transfer raw completion time [s]	MTRCT	T4-T1
Money transfer completion rate [%]	MTCR	T1 present, T4 present: success (see remark 1)
Money transfer full completion time [s]	MTFCT	T7-T1: Not used when asynchronous mode is used
Money transfer A-side completion time [s]	MTACT	T6-T1 (see remark 2)

It is assumed that some data cleansing has to be applied here. With manual operation of the test case, input errors may occur. In data recording, a way to mark such transactions as "to be ignored" should be provided. When KPIs are calculated, such transactions need to be excluded from the set because they may cause distortions to other KPIs (e.g., longer response times in the case of non-existent account numbers or personal identification number (PIN) errors).

In addition, it is recommended to align KPI calculation with known general network problems. For instance, if there is a mobile network outage during a test, respective transactions should be ignored in KPI calculation also.

There is also the possibility that a transaction is reported to the party A as successful but actually is not. As part of data cleansing, the actual arrival of money should be checked (e.g., as a bulk action in post-processing) and appropriate corrections applied in case. This can, for example, be done by setting respective transaction attributes, or – in the simplest case – invalidating T4 and or T6. These procedures need to be aligned with the specific implementation of evaluation and KPI processing.

Remark 1: When a test campaign is designed, a decision should be made about whether the MTCR is reported at all, based on an assessment if the test is seen as representative.

Remark 2: If T6 is used as surrogate for T4 and T4 is missing, MTACT is typically not computed as there would be partial overlap with MTCD and the sample basis is smaller in any case. If computation of MTACT is nevertheless required, it needs to be made clear in the accompanying documentation (project report), that this is a secondary or auxiliary KPI.

Special consideration is required for the case of MTCR, as the reported completion rate depends on the number of unsuccessful transactions. There might be cases where either the network or the DFS infrastructure is temporarily down, i.e., a part of the service is systematically unavailable. MTCR is therefore not only a technical element, but also a matter of testing perspective and scope of a testing campaign. It needs to be clearly specified and documented how these cases are treated, i.e., if and which transactions shall be removed from the valid set.

In the case of a campaign having an explorative character or seeking a broad perspective, a solution might be to report different variants of MTCR to show the corridor of values, depending on respective decisions.

8 Creating the use case model from actual use case examples

This Recommendation does not prescribe explicit ways to create use case models. However, Appendix I provides a best-practice example of use case models from existing implementations.

9 Data sources

In order to compute DFS KPIs, relevant input data need to be collected.

Basically, there are three sources of information as follows.

– Recorded events from observation of the DFS use case. In the context of this Recommendation, these are events recorded by the MSW app.

– Results from background measurement of the transport network at each side of the DFS use case. Basically, this can be done by every suitable QoS testing tool on the market.

– Optionally, notification SMS on both sides of the DFS use case. In the current context, these are not used as there is considerable effort to collect these data and their additional value is considered to be small. If their use is desired, see the relevant clauses of [ITU-T P.1502] or [b-ITU-QoS_01].

Data recording and logging needs to ensure a high level of robustness with respect to data quality for the entire processing chain towards DFS KPIs and reproducibility of testing actions. This includes taking into account that testing teams may be in different locations without direct means of communication, multiple teams may be testing at the same time or role assignments between teams may change over time.

This Recommendation does not make explicit prescriptions of procedures, data structures or naming schemes for data sources or other information. For guidance, Appendix II, however, provides a set of working examples covering these aspects.

10 Special procedures in the field

10.1 Operational protocol if party A and B do not communicate directly

If teams are operating in different locations, a robust protocol is required that tells teams how and when to act.

This protocol needs to cover the following situations (based on cyclical role switching in the DFS use case).

– Normal operation: Teams need to start the procedure by agreeing which team has the initial party A role.

– If a team currently has the party B role and T7 occurs; how long should it wait until continuing in the party A role? This issue becomes important when it is possible that T7 precedes T4 or T6.

– If a team is in the party B role, it is possible the transaction was successful, but no notification SMS is received (i.e., no T7 can be set). How should teams act if expected events are overdue?

– How should teams communicate when they want to pause or end testing?

Figures 4 to 6 are graphical representations of the main cases and recommended parameters of the communication protocol.

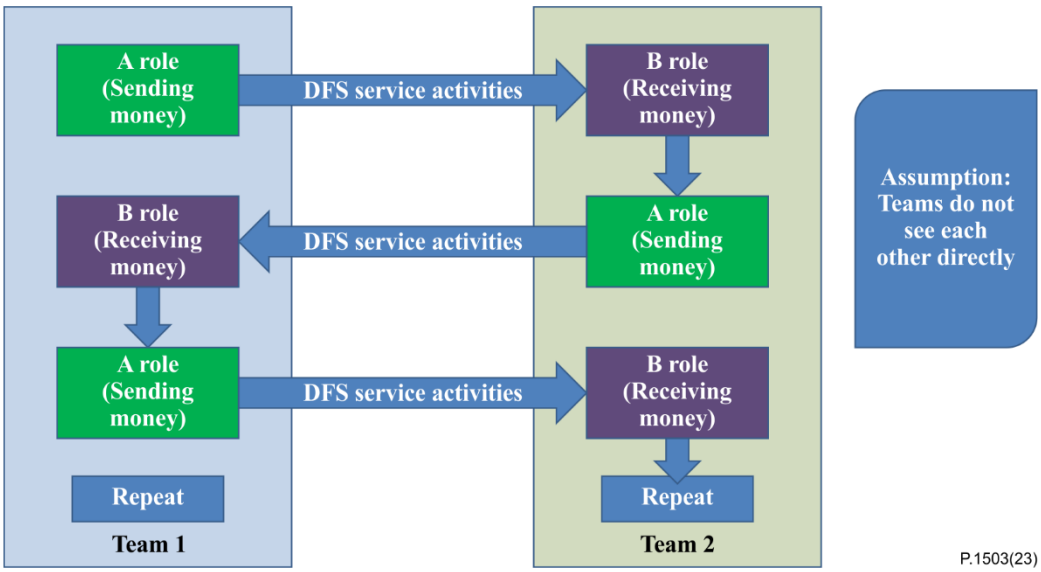


Figure 4 – Team interaction – normal case

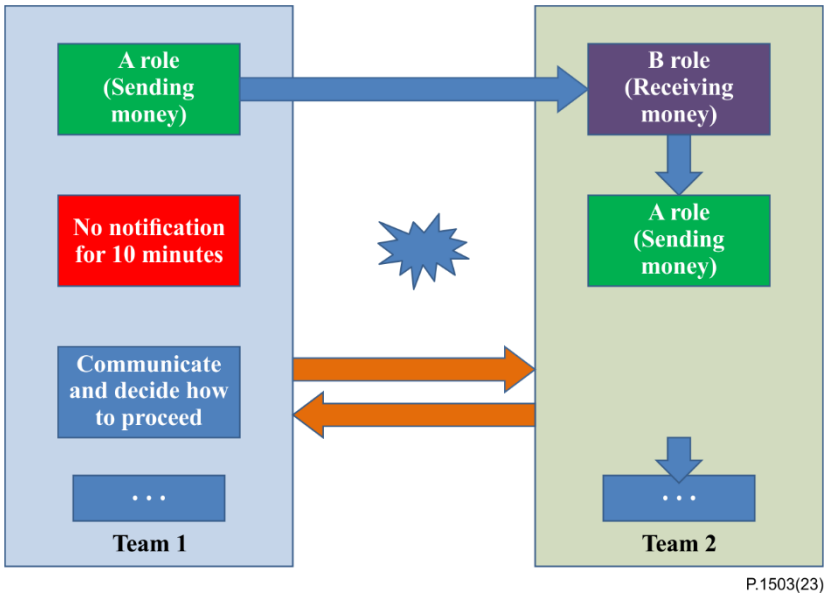


Figure 5 –Team interaction - unplanned interruptions

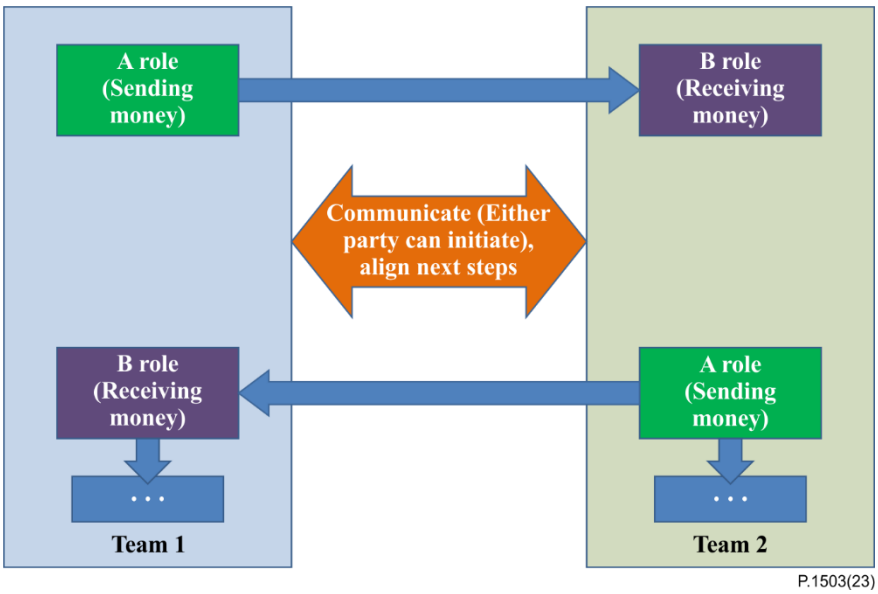


Figure 6 – Team interaction - planned interruptions

11 Testing modes

11.1 Basics

In [ITU-T P.1502], it is assumed that testers have both party A and B roles in a test case configuration under direct control and in full visibility. There are, however, situations where this is not the case, e.g.,

– in cross-border testing where teams are in different locations or countries by nature of the test case;

– in testing situations where special aspects of the service are under test, e.g., network coverage that is good in its location for one party and poor for another;

– in special circumstances, e.g., in a lockdown during a pandemic.

In these situations, further differences are made in the degree to which teams can communicate with each other. A situation where there is full-scale real-time communication, including video feeds on respective devices is practically equal to the standard situation of all elements being in the same room. At the opposite end of the scale, there may be no communication at all between the testers.

11.2 Synchronized and asynchronous testing mode

The following variants of testing and appropriate terminology are specified to take care of these situations. The field test lead (FTL) may select from these variants according to its assessment of the situation and the requirements of the KPIs to be fulfilled.

Both modes still assume paired teams send money to each other.

Synchronized testing: In this mode, party A and B roles are exchanged cyclically. The party B role ends on receipt of a T7 event or when a predetermined time-out is reached.

In situations where teams are working in the same location or have real-time communication, one tester records events, and a data set can contain all events (T1 to T7) specified in the modelling of the use case.

In situations where teams are separated, the team in a party A role records events T1 through T6 in one data set, and the team in the party B role records T7 in another (optionally, T1 can also be recorded in this set, indicating the point in time from which receipt of a notification is expected). Relevant data are then combined during post processing.

Figure 7 shows the principal flow in synchronized testing.

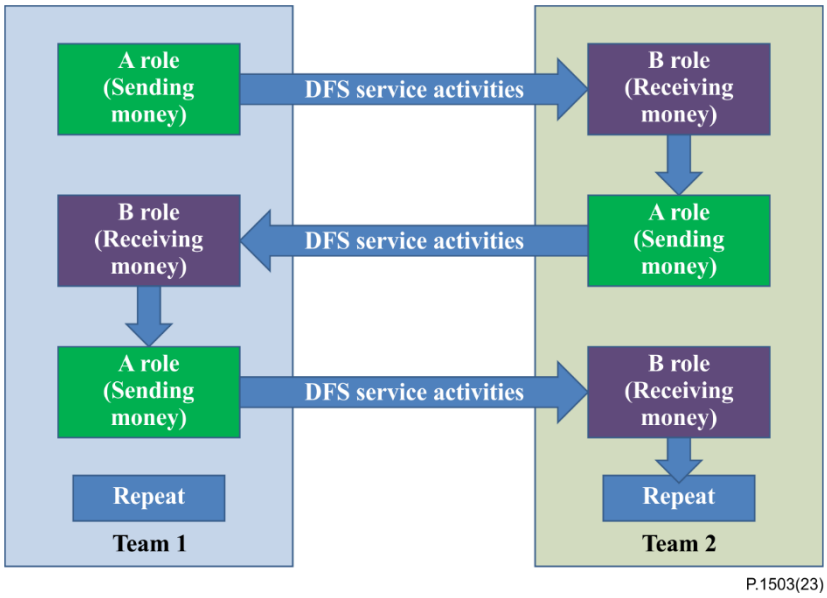


Figure 7 – Principal action flow for synchronized testing

Asynchronous testing: In this mode, testers on both teams act independently in the appropriate party A role. This means that team 1 and team 2 generate transactions independently towards each other. In this mode, T7 is not recorded, and KPIs containing T7 are not computed and reported.

Asynchronous testing makes sense where teams have different working hours, frequency of operation or where KPIs using T7 are not considered to be required.

In comparison to synchronous testing, the amount of money to be held in credit on each device has to be higher, as in the extreme, only one team may send money for an extended time period. See Figure 8.

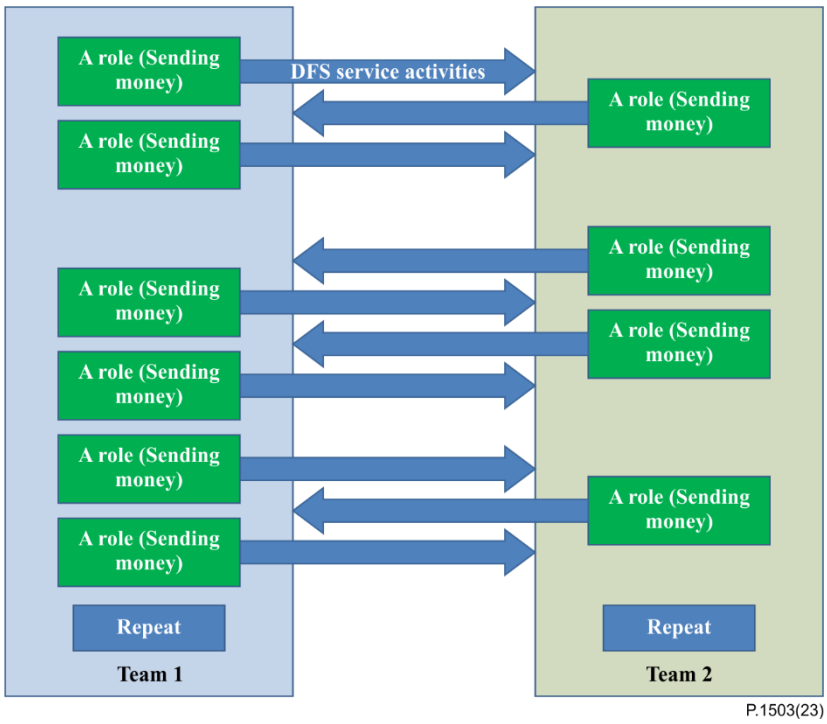


Figure 8 – Principal action flow for asynchronous testing

11.3 Dual-function set-up and operation

There is a special situation when intra-operator tests in the same country are made in the same location, and it is required that the number of devices used in testing be minimized.

In the standard case, devices have fixed functions; the DFS device is used to run the use case while the observer device is used to run the MSW app and the background-testing app. In the generic case, this would require a total of four devices where each device has a dedicated role.

The number of devices can be reduced to two if devices switch functions. In that case, both devices need DFS accounts, MSW and ObsTool apps installed. Assuming device designations X and Y, the sequence of testing would be:

– X is used as the DFS device; MSW and ObsTool are started on device Y.

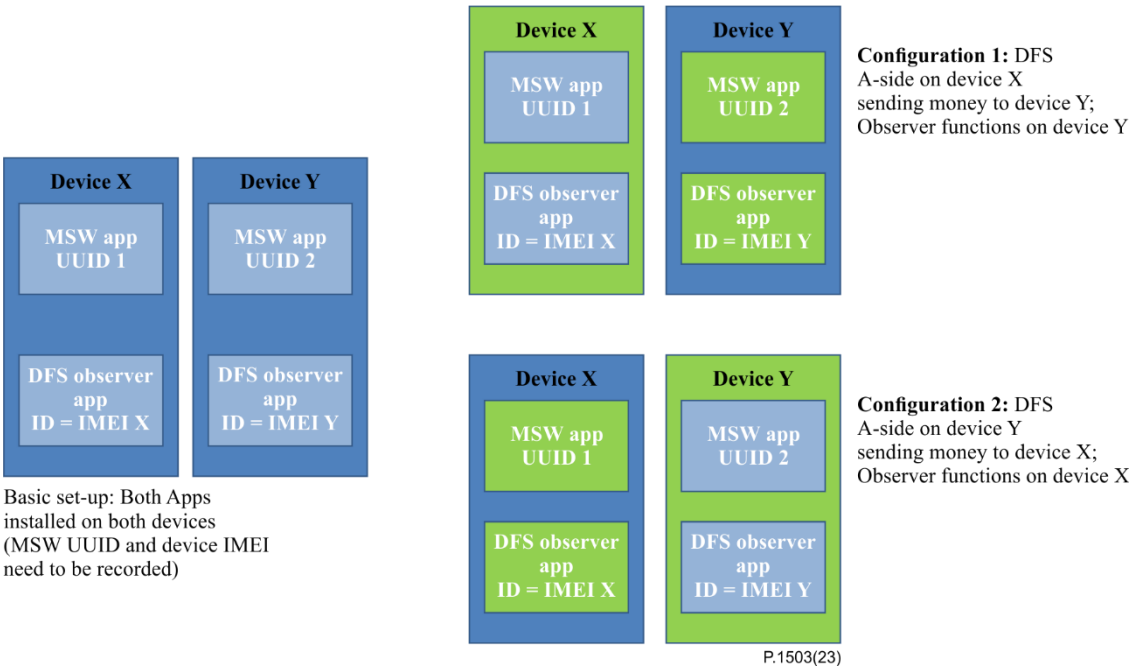
– X is in the party A role, sending money to device Y which is therefore in the party B role. Events are recorded using the MSW app on device Y.

– If synchronous mode is used, device functions are swapped now. In asynchronous testing, first a number of transactions is run, then functions are swapped.

– After swapping, MSW and DFS observer is ended on device Y, and started on device X (see remark (1)). The DFS party A role is now performed on device Y with device X in the party B role.

Remark (1): This is a precaution to exclude cross-effects from background testing into DFS performance. By assessment and decision of the FTL, respective apps can run continuously on both devices.

Figure 9 shows the set-up and operation graphically.



UUID: universally unique identifier

Figure 9 – Set-up and operation for dual-function operation of devices

12 Recording events of DFS test cases

12.1 Basic functionality

As stipulated in [ITU-T P.1502], manual collection of test case data requires considerable effort during the tests, and is also a major source of errors and potential degradation of data quality due to transfer of handwritten notes into electronic media.

Therefore, the methodology recommended here assumes there is an electronic tool that supports direct collection of events and related timestamps.

In order to enable independent implementations, this methodology does not prescribe a particular implementation. It does, however, use terms related to a reference implementation in order to provide a comprehensive picture of functionality, operation and practical considerations related to such an application.

The implementation is based on the following functional requirements and design considerations.

– The app provides buttons for the timer flags defined in the methodology, i.e., T1 to T7.

– The button naming includes the timer flag names plus a descriptor text that is derived from the analysis of use cases obtained as described in clause 10.

– There is an optional text input field "comments" where the user can enter additional information.

– There are explicit submit and discard buttons. Even when the user hits "discard", the current data set (timer flag buttons pushed and content of the comment field) is recorded or uploaded.

There is button logic aimed at preventing unintended actions, as well as providing the required freedom of operation. The following button logic is recommended, but may be chosen differently in different implementations.

– The submit button acts immediately while the discard button produces a dialogue box asking for confirmation. This is to reduce the number of button taps in the case considered as the most common.

– The discard button is permanently enabled. The submit button is enabled after T3 has been pushed. For handling of time-outs in early stages of the transaction, see clause 12.2.

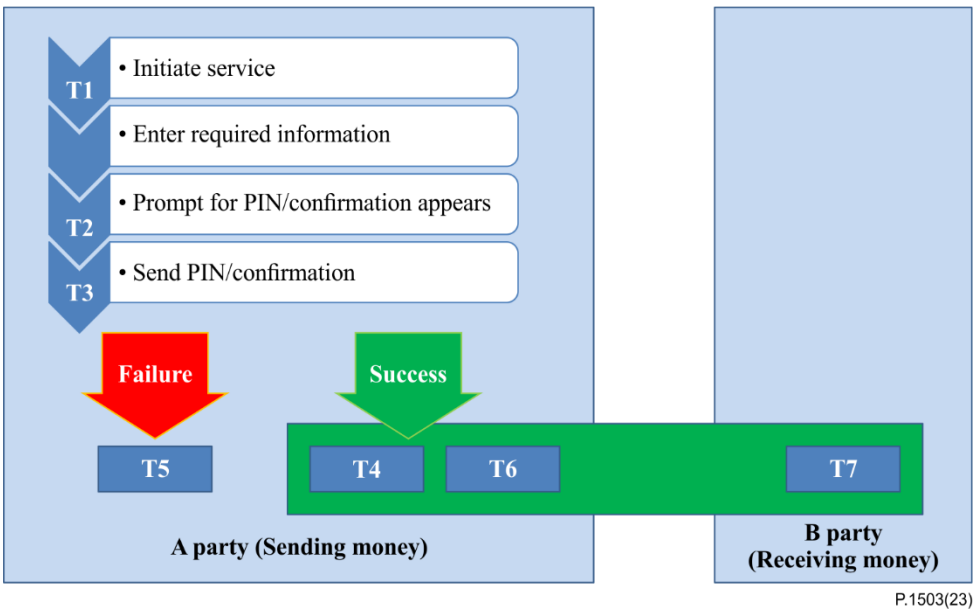
– The T7 button is permanently enabled. Button T1 enables T2, T2 enables T3 and. T3 enables T4 to T6. This follows the sequence prescribed by the use case modelling and shall reduce the risk of submitting erroneous data. As T4, T6 and T7 can appear in any order, sequence forcing after T3 or automatic submission of data cannot be supported.

If a specific implementation uses a different set of button logic, adjustments to the practical hints as given in clause 12.2 are required.

Different implementations may make other choices in actual user interface aspects, such as screen layout and button-locking logic colorization. It is, however, assumed that the functional core is the same, so the descriptions in subsequent clauses of this methodology refer to the generic functional elements as specified in the preceding list.

12.2 Practical application

The following applies for the synchronous testing mode as described in clause 13. Figure 10 symbolizes the action flow. For asynchronous testing, party B activities are omitted.



For the asynchronous testing mode, party B activities are omitted.

Figure 10 – Symbolized flow of action (synchronous testing mode)

– Party A situations, normal testing workflow
  - Normal flow of actions: Successively tap T1, T2, etc. To conclude and record or upload the data set, tap submit.
  - No response (timeout) after activating the service by unstructured supplementary service data (USSD) command (T1 has been recorded): Enter "A timeout" in the comment field, then tap discard.

– Party B situations, normal testing workflow
  - Getting ready to expect incoming notification: tap T1 (optional). When notification comes, tap T7, then submit.
  - No notification after agreed time-out period: Enter "B timeout" in the comment field, then tap discard.

– Handling erroneous actions:
  - If for some reasons a transaction should be ignored (e.g., because during data entry wrong actions have occurred), a short text describing the cause should be entered in the comment field, and discard tapped. The wording can either be freely chosen (avoiding texts used to describe party A or B situations as previously specified) or can be specified to flow a set of terms agreed in the specific campaign. In any case, the text shall be suitable to clearly mark transactions that need to be excluded from post processing.

13 Measurements in the background

While staying within the general conceptual framework of [ITU-T P.1502], the context of multi-network or cross-border testing require some differentiation and careful consideration of respective conditions.

For convenience of reading, Table 6 of [ITU-T P.1502] is reproduced in Table 2 to show in which cases the local mobile network performance has an effect of DFS QoS.

Table 2 – Categorization of impact of mobile network and DFS infrastructure performance on end-to-end DFS QoE

	Well-performing DFS functionality	**Poorly performing DFS functionality**
Well-performing mobile network	High level of overall QoE, only vulnerable to local or temporal impairments of each component	Mobile network performance not relevant/not visible
Poorly performing mobile network	Overall DFS QoE strongly depends on mobile network performance	Low level of overall QoE, no clear dominance of each component

Background testing typically uses a mix of test cases that are considered to be relevant for the DFS-related performance of the mobile network used for the use case, i.e., in the current context, packet-data service performance.

In the case of inter-operator and cross-border operation, based on the general principle that the scenario should be the same for all teams, the following guiding rules apply:

– the network load caused by background testing should be moderate;

– there should be no country-specific elements; it follows that, for example, web browsing should use standardized reference pages only or live pages that can be reasonably assumed to be general enough.

This leads to the following recommended scenario:

– fixed-size HTTP download with moderate content size (e.g., 3 MB), and generous time-out or pause values;

– fixed-size HTTP upload with moderate content size (e.g., 1 MB), and generous time-out or pause values;

– web browsing with a rather lightweight standardized reference page, e.g., ETSI Kepler Smartphone.

USSD and SMS elements can be added or omitted, based on the assessment of the FTL). USSD can be problematic because a uniform set-up across all teams would need USSD codes that fulfil the requirements of [ITU-T P.1502] for a multitude of countries. In the case of SMS, the set-up of devices would be considerably more complex due to requirements by the Android operating system. If USSD or SMS are not primary for the implementation of the DFS service (also in the case of SMS, the use case itself provides information about relative SMS performance), it is recommended to leave them out of the scenario.

14 General considerations about errors in measurements

Before entering into specific considerations, the usage and definition of the term *error* needs clarification as it is used in several conceptual contexts.

A summary of potential meanings of *error* follows.

1) A statistical error in the sense of an error margin. If a quantity is calculated from a limited number of data samples from a system that exhibits, from the user viewpoint, somewhat random behaviour (such as a failure probability), this quantity does not describe the relevant property of the system exactly, but only within a given margin. This margin can be calculated based on statistical formulae; relevant information can be found in [ITU-T E.840] or [ITU-T E.804]. In short, the only way to reduce this error margin is to increase the number of samples taken.

2) Errors caused by incorrect reading or transmission of readings, i.e., "human error" in the data collection process. [ITU-T P.1502] deals extensively with such errors in the context of

measurements on DFS. Avoiding such errors requires careful execution of testing and data-collection steps. The checklists and procedures described in this Recommendation are a tool to provide robustness of the measurement process and to reduce the probability of such errors. However, there is always a trade-off between the effort for such checking procedures and impact of actual undetected errors. In general, single errors decrease the accuracy of measurements. As far as such errors are effectively random in nature, increasing the number of samples is also a means to reduce their impact on output data quality. Applying cross-checks and logical tests is also a way to reduce the probability that such errors take place undetected.

For instance, checking the number of samples against the testing time, by estimating expected sample counts and comparing them to actually received ones, would be a simple first-level way of data quality assurance.

The way the team assignment list (TAL) and device assignment list (DAL) are constructed, and the usage of field test logs, are also expressions of this strategy. It needs to be mentioned, however, that applying a pre-determined recipe alone is not sufficient. Considering the concrete situation in the field, and applying respective checking steps, are equally important parts of an overall error reduction strategy.

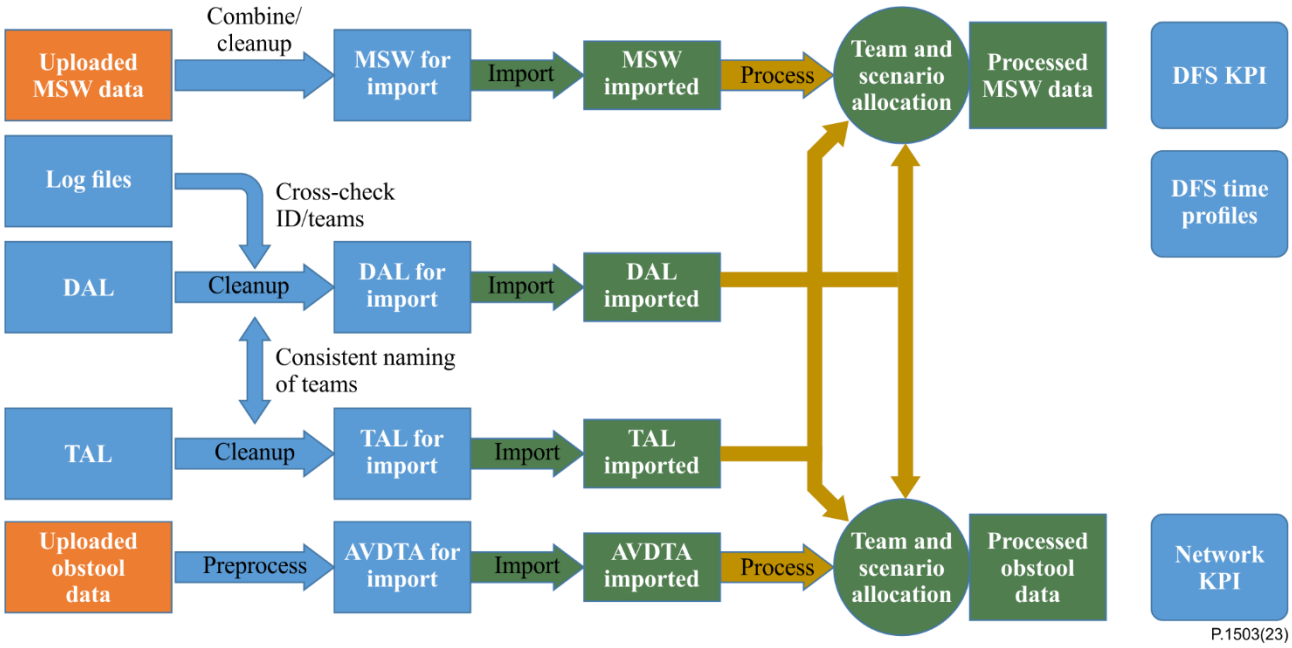
3) Errors arising from operation, i.e., a special type of human error but with an impact on more than one data point. Examples are: insufficient power supply (low-battery condition), which can cause atypical device behaviour; overheating of devices due to insufficient air flow or exposure to heat sources; and forgetting to activate functions on the devices. The log templates and associated regular checking procedures are designed to provide protection against such errors. Again, these measures need to be complemented by assessment of concrete field situations and appropriate judgement and formulation of additional measures based on actual circumstances.

4) Errors in the implementation of data processing. The way to reduce this risk is to run algorithm tests (e.g., structured query language (SQL) queries) with a limited number of data points, and compute reference values manually (typically in a spreadsheet calculation application). Even if pre-defined processing algorithms are provided (e.g., by a set of SQL statements used in previous measurement campaigns) it is advisable to apply such tests, unless the processing environment is assured to be exactly the same.

15 Data validation and processing

15.1 Overview

Figure 11 is a schematic description of the structures and processes for post-processing (see also clause 11.2 for a description of the data entities).



AVDTA: accumulated validated data on transaction level. AVDTA can be understood as a placeholder for any suitable type of pre-processed raw input data for creation of the required KPI.

Figure 11 – Symbolic overview of data objects and processing

The steps for data cleansing and processing are:

– Complete and validate the TAL; make sure team column headers are consistent with data base requirements (see also clause II.7).

– Complete and validate the DAL against field test logs; make sure column headers are consistent with data base requirements (see also clause II.8 for reference).

– Cross-validate DAL and TAL, make sure that team names are consistent.

– Cross-validate MSW and ObsTool data, make sure that data source identifiers (IDs) can be resolved to team names and scenario names.

15.2 Plausibility and validity checks

15.2.1 Basics

The intensity of checks depends on assumptions about outer conditions of testing; if there are factors that can lead to a higher risk of data loss, checks should be run more often and *vice versa*. It is good practice to run these tests on the "data harvest" every day or every other day.

– General yield of data: Check whether the number of data items from the MSW and the network performance test roughly corresponds to the overall testing time.

– Cross-checking with global positioning system (GPS) data: If the location permits, GPS data can be an important source of information for cross-validation. For instance, the GPS data yield (data points per hour) should correspond to the overall testing time. Time information in GPS can also provide information for cross-checking device settings. For that purpose, the background-testing tool should provide the original NMEA (National Marine Electronics Association) sentences or an adequate equivalent.

– Cross-checking the session or location logs with the information in the TAL or DAL.

15.2.2 DAL and TAL validity checking

In order to correctly combine the data from different sources (e.g., T1 to T6 from the party A side with T7 from the party B side), information about the pairing of teams, and temporal assignment to DFS testing scenarios, is required. This information is provided by the DAL (see also clause II.8) and

TAL (see clause II.7). Therefore, DAL and TAL are also imported to the database, and after validity checking, typically processed into respective internal tables with added (constructed) content.

In order to process data, a unique scenario descriptor is required that is used to aggregate (group) data to the respective KPIs. Depending on the actual TAL structure, such a descriptor can be either included in the TAL directly, or – preferably –constructed in the database from basic elements. In any case, validation is required to make sure that this indicator is indeed unique, and data can be assigned without creating gaps or duplicates.

It is also helpful to produce unique scenario IDs as numbers. A scenario name will typically be a rather long string of text that may be impractical for use in dense tables or graphs. A scenario index, in combination with a look-up table, makes labelling easier.

For TAL validation, it is helpful to create a visualization of the TAL in the form of a Gantt diagram as shown in Figure 12. With the help of such a visualization, it can easily be checked if all scenario or time ranges are present and consistent. The source table for this visualization can also be used to check the scenario names for uniqueness.

Figure 12 – Example of a Gantt visualization of a TAL

15.3 Data processing

A good practice for data processing is to import data from MSW and network background testing into a central database and run the final processing there.

In the first step of processing, MSW and network KPIs are combined (joined) with the relevant scenario and team information. In the second step, appropriate grouping of DFS and network KPIs is done based on this information.

The basis of these join operations is information relating to data items to scenarios. This is done in a multi-step operation. In the first step, technical IDs are used to connect to the configurations or owner teams. This can be done by creating a look-up table from the DAL with respective join operations or by assigning these elements directly in respective SQL statements when the processed MSW and network KPI tables are created.

The exact process for join operations also depends on the way the measurements are done. If devices are kept within the same team, and different configurations are used, the device or app ID has to be

linked to the owner team (see TAL and DAL) and the time range information provided. If devices with fixed configurations are exchanged between teams, linkages have to be made using the device configuration information.

The result of such join operations would be extended MSW and network KPI tables that contain the scenario names used in the TAL, as the element required to perform aggregation.

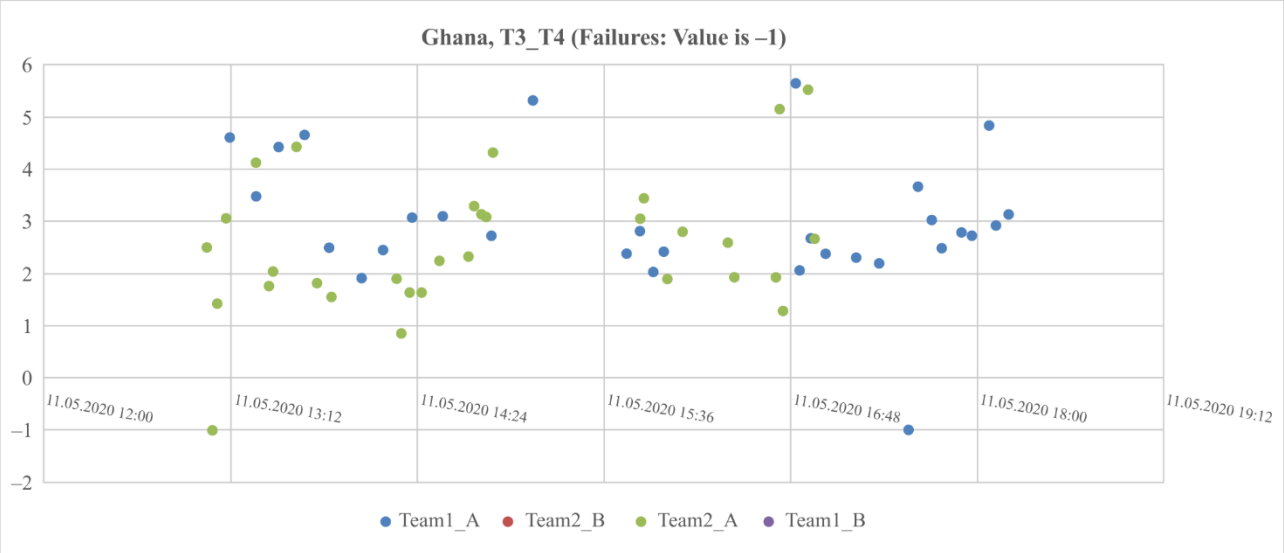
The third and final step of processing is then to create appropriate table and graphical visualizations. Typically, this is done by either creating tables in a spreadsheet application for conversion to graphics or directly using graphical front-end tools.

15.4 Time profile

A time profile is useful for both data validation and reporting.

The profile is an *x*-*y* graph created from MSW. It is basically a scatter plot where, on the *x*-axis, date and time is shown, while on the *y*-axis, a selected indicator is shown.

Figure 13 shows a time profile for a pair of teams.



P.1503(23)

Figure 13 – A time profile for a pair of teams

From such a graph, various information is directly visible:

– the time range of tests;

– the value range (band);

– these tests were done in asynchronous mode, i.e., there are no party B events;

– there are a few failure cases (data points having the value −1); it would have to be clarified if these were real or false positive cases.

16 Background testing of mobile networks

The considerations and recommendations on mobile network background testing provided in [ITU-T P.1502] fully apply as running such tests on the local network does not depend on the test scenario. The following is therefore meant to be as an extra perspective on the matter.

This Recommendation dies not make explicit prescriptions for testing tools or for specific testing scenarios to be used. However, Appendix III provides a working example for a set of testing tools and mobile-network QoS KPIs.

As outlined in [ITU-T P.1502], the effect of mobile network performance on overall DFS QoS depends on the performance and interplay between DFS infrastructure and the network. Only if the DFS infrastructure works well, i.e., the processing times are consistently short, mobile network performance may become the determining or limiting factor in overall DFS QoS. With slow or strongly fluctuating DFS performance, the influence of the mobile network may not be visible in output data at all.

If mobile network performance is also a matter of interest at all depends on the overall scope and goals of a campaign. Therefore, the effort made when testing mobile network performance (e.g., if devices are allocated for testing, or the budgeted cost of mobile data plans) is typically decided case by case. For instance, if tests are done stationary, spot testing may be sufficient to assess the mobile network coverage quality, instead of running data-intensive testing all the time.

Appendix I

Working example for creation of a use case model from actual implementations

(This appendix does not form an integral part of this Recommendation.)

This appendix is based on relevant content in [b-ITU-QoS_03].

It has been shown that for creation of the basic model of a DFS transaction, a well-produced video is best practice.

A well-produced video is a persistent source of information in detail and can be analysed easily.

In order to fulfil its purpose, videos should be produced along the following guidelines.

– Show the device screen in good, uniform lighting and clarity, avoiding light reflections.

– Make sure that while there is the need of manual operation, the screen is visible long enough in each step to allow the flow of events to be followed.

– Have a high-quality audio comment providing explanations of the steps to be taken, and comments on results where necessary. Ideally, these comments already include references to actual event-recording processes, e.g., mentioning the "timer flags" to be recorded.

– If feasible and for completeness, the screen of the party B device should also visible. While this will be impractical if party B is in a different location, it may be provided by another video showing the reaction of the device on an incoming DFS transaction.

Appendix II

Best-practice example of data structures and procedures used in DFS testing campaigns

(This appendix does not form an integral part of this Recommendation.)

II.1 Basic considerations for data collection and processing

This appendix is based on relevant content in [b-ITU-QoS_03].

The data source apps are installed on platform devices. On the Android operating system, running apps concurrently without cross-effects is not guaranteed so the ideal configuration, from a fundamentalist point of view, is to run each app on a separate device. On the other hand, it is desirable to minimize the handling effort, i.e., to combine apps on the same platform. The actual decision is made by the FTL on the basis of a benefit-to-effort consideration. It is recommended also that checks (e.g., on the data from the first days of a measurement campaign) be run to ensure that there are no negative effects or only effects on a tolerable scale.

The DFS application must be always visible in order to record events properly, and the MSW app also needs to be on top of the screen for delay-free recording of events. Therefore, from all possible combinations, only a few remain: There must be at least two devices; the network-testing app can run on either the device running the DFS use case ("DFS device"), or the device running the MSW app. With the further consideration that network testing also uses packet data resources, the second configuration is that of choice, as shown schematically in Figure II.1.

Please note that this is a rather schematic and simplified view. Details given in clauses II.2 to II.8 have precedence.

Figure II.1 – Schematic allocation of data sources to devices

II.2 Additional considerations for cross-border and inter-network testing

In the case of more complex measurement campaigns, such as those involving cross-border transactions, transactions between different networks or multi-team activities, one or several of the following conditions can apply.

– Testing teams may be in different locations without direct communication between them. This means that each team only sees a part of the overall set of events belonging to a transaction.

NOTE – Enabling real-time communication between teams, e.g., via audio or video conferencing solutions), requires adequate data bandwidth, which limits applicability to relevant situations. A powerful means of communication between teams is an automated dashboard, making processed information available in near real time. Technically, such a solution is easily feasible given the amount of budget required. In the actual design of such a system, using as little data bandwidth as possible has to be a major design goal in order to ensure operation under a wide range of on-site conditions.

– There may be multiple teams at work during the same time period.

Consequently, additional information – beyond what is collected by the testing tools itself – is required to produce the desired information, i.e., DFS KPIs.

Basically, it must be known which team or pair of teams is running which scenario in which time period. Only then is it possible to create appropriate KPIs.

For instance, assume team 1 and 2 are running a national test between two different DFS providers, and at the same time, team 3 and 4 are running a cross-border test on another set of DFS providers. For a given transaction for the national test, timer flags T1 to T6 are taken by team 1, while T7 is taken by team 2. For the cross-border test, T1 to T6 is taken by team 3 and T7 is taken by team 4. Likewise, network performance tests are also collected by respective devices in different locations.

Assuming that all data is imported to the same database – which is the usual way to process data for best efficiency – allocation of appropriate device IDs to DFS service test scenarios has to be made, which requires corresponding information about measurement system allocation and testing schedules. The process of data cleansing – removal of data that is considered to be not valid – also requires appropriate information.

In many cases, it may be possible to deduct information that identifies the scenario under test by using information in the primary data source, e.g., GPS locations in the background-testing data. However, as the methodology is supposed to work in a robust way under a wide range of conditions, it cannot be guaranteed that these information sources will always have sufficient information. For instance, if tests are done from within a building, a valid GPS position fix may not be available. Therefore, frame or top-level information should be provided.

Basically, such frame information can be provided centrally or locally. As such tests are typically done in a planned manner with a central management entity, there should be a register of team, device, and scenario allocations versus time. For maximum robustness, it is also recommended that this information again be collected locally, i.e., using log sheet listings of what has actually been done by the teams. As this information is typically high level, i.e., taken only once per location or testing session, load on teams is low and the extra redundancy improves the overall robustness of testing.

This methodology is designed to be robust in the sense that a certain degree of redundancy is maintained for information that is essential to proper data evaluation. Essential information relates to which teams are paired for a given test case and the assignment of electronic IDs of relevant tool installations.

These electronic IDs are essential single elements of a test and measurement set-up, as they are required to assign measurement data to the right context.

Basically, there are several types of electronic ID as follows.

– Fixed IDs, such as international mobile equipment identity (IMEI) or media access control addresses. In most cases, they can be read electronically by apps through the relevant application program interfaces. However, in recent years operating systems put restrictions on this type of ID as they allow identification of devices and therefore – in case where devices are also used for personal purposes – may create privacy issues.
– Dynamic IDs that are created with every new installation, and are not linked to any static attributes or properties of the platform.

In any case, suitable IDs need to be unique, i.e., it must be made sure that no two devices or rather, data sources, have the same ID to prevent data assignment mix-ups.

In the current case, it is assumed that devices are sourced exclusively for testing purposes, so fixed IDs are not problematic. The MSW provides, however, dynamic ID generation and is therefore more versatile.

There are two central lists, maintained by the FTL in co-operation with the entity managing the overall testing activities, as follows.

– The DAL holds information about the assignment of devices and tool app IDs to teams.

– The TAL holds information about the time schedule of scenarios and associated team pairings (under the assumption that each team runs one end of a two-way DFS use case as specified in the transaction model). The TAL is used both for planning activities and for documenting them afterwards.

If team assignments change, it is especially important to keep good record of these assignments, to prevent data loss or artefacts due to unclear or incorrect combination of data from different teams or data sources.

Copies of the DAL and TAL are imported to the post-processing database and serve as the source of assignment operations required to generate KPIs and other report information.

In addition, each team uses log sheets (in paper or electronic form) to document their activities locally. These log sheets provide an additional layer of robustness by providing redundancy of information with respect to central lists.

TAL and DAL are described in detail in clause II.7 and clause II.8, respectively.

II.3 Data structure overview

The overall data structure is shown in Figure II.2. It is assumed that this data structure exists in a central data processing environment, typically an SQL database.

It needs to be pointed out that the methodology does not provide a single, prescribed data structure. Actual data structures can have additional members and shapes. There is also no absolute way to process data; the methodology can therefore be embedded in a wide range of post-processing environments and tool chains. This applies, in particular, to background testing (network KPI) data that can be created in multiple ways.

Data are typically processed in steps that also involve data validation and inspection. The goal is in any case to obtain a robust database for subsequent processing, i.e., any ambiguities, missing assignments or contradictions should be detected and resolved prior to creation of actual deliverable output.

Data preparation and validation is usually a multi-step process starting with coarse data cleansing on input data basis (e.g., visual inspection of data in spreadsheet files and alignment with log data).

As a database is an efficient environment for data inspection and structural checks, data cleansing is typically a cyclical, incremental process.

Examples for data that may need to be cleaned out are:

– data resulting from test runs which have not been done under defined conditions;

– data taken in situations that are deemed to be exceptional and should not be part of statistics;

– data resulting from unintended operation, e.g., an incorrect PIN, or from an operation cancelled due to some other erroneous entry.

Pre-cleaned data is imported to the database, checked, corrected if errors are detected, and reimported until the desired state is reached. During this process, individual data sets may be masked out, i.e.,

tagged as to be ignored during further processing steps. This is necessary if information is incomplete or contradictory due to missing or inconsistent data collected in the field.[3]

Data cleansing may be a cyclical, repetitive process because in order to detect some artefacts, a certain level of cleanliness is required in the first place. When it comes to processing larger amounts of data, data also need to have some formal structure before meaningful checking procedures can be applied efficiently.

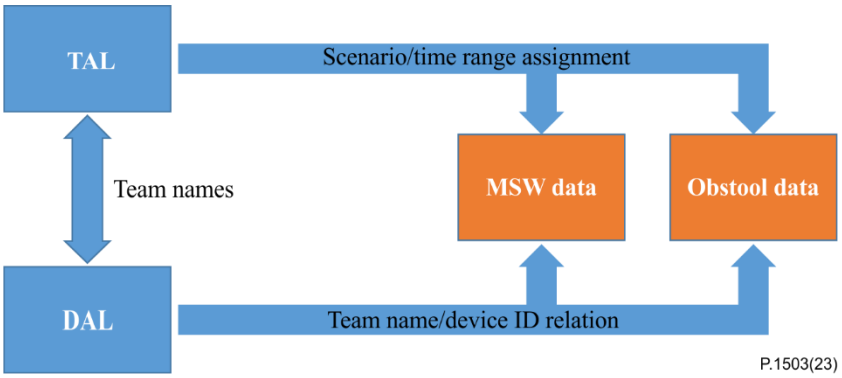


Figure II.2 – Symbolic overview of data structure for post-processing

II.4 Naming and formatting conventions

The conventions in clause II.4.1 are essential to ensure error-free and efficient data processing over the whole chain.

II.4.1 Team naming

For data processing, consistent naming over all data sources is of primary importance as it is used to combine data, and errors will lead to incorrect KPI evaluation.

The following assumptions or prescriptions are made.

– A DFS operator is identified by, in this order, the network name and the country name, e.g., MTN Rwanda; Airtel Uganda; Vodafone Ghana.

– Team names shall be chosen accordingly, i.e., the team name should be the same as the operator name.

– A team operates a logical group of one DFS device and one Observer device (see clause 6).

– There are only two permitted types of device or function allocation as follows.

  a) One device is fixedly assigned to the DFS role (i.e., is set up for a particular DFS operator), and the second device is fixedly set up for the observer role.

  b) Both devices are set up for both roles, but the DFS operator on both devices is the same.

  If configuration b) is used, data taken by the MSW and network performance background testing are treated as equivalent with respect to the generating devices.

The word "team" has historic origins. In consequence, the team name represents, and stands for, a particular DFS operator. In this sense, team is a logical entity rather than meaning a group of people.

Therefore, a team can be represented by a single person. It is possible that a single person or a group of persons operates multiple devices or switches between different set-ups. In that case, the

[3] Data processing and cleansing can compensate for a certain amount of simplification or omissions in the data collection process, e.g., by inferring missing or incorrect information from circumstantial data. However, this cannot be taken for granted, so careful data collection in the field is essential for a high yield of useful data and high quality of resulting information.

documentation described in clause II.7 needs to list each configuration where the DFS operator changes as a separate team with the respective time window information.

This is because there must be absolute clarity (i.e., at all times) in the relation between the observer device or data IDs (MSW and background testing) and the DFS operator under test.

II.5 Naming and formatting conventions for data objects

In order to facilitate efficient and error-free transfer of data between file media and databases, the following conventions must be met.

– Column names in data files (e.g., a spreadsheet) shall only contain alphanumeric characters; shall not begin with a numeric character, and shall not contain whitespace characters.

– Date and time shall be given in one of the standard formats. Formats shall not be mixed within the same column.

– In particular, for time information, data integrity has to be maintained over the whole processing chain. In particular, this applies to the decimal separator and delimiters in general. It is also known that millisecond formats can be troublesome. It is highly recommended to run tests before going into full-size data processing.

– A proven method to enhance conversion robustness is using "decorated" elements. For instance, if a timestamp is required to be given with milliseconds, in can be written with a preceding fixed-size string. In that case, it is imported as a string (e.g., nvarchar) to the database. Once in the database, substring, cast, and parsing operations can be used to convert this element to a datetime element again without the risk of loss of information.

II.6 Local log sheets

There are two variants of log sheets. The FTL assesses the given situation and decides which variant is to be used, and which information is pre-printed, and which has to be entered by the teams. This decision is based on the level of skill and experience of teams and expected frame conditions in a given location.

– A rather explicit location log sheet that has fields for basic scenario and set-up related information and provides a detailed check list for elements to be checked periodically (on an approximately 2 h schedule).

– A short-form session or location log sheet that also collects information essential for measurement data allocation, but has a simplified status check section that just asks for a general confirmation that operating conditions are still within valid parameters.

This Recommendation has an electronic attachment containing spreadsheet templates for project usage. It is good practice to use file names that contain a reference to the project or campaign to which they belong, some working that describes their function and a revision number to support document maintenance and evolution.

This Recommendation is accompanied by two spreadsheet files containing ready-for-use table templates:

DFS_Full_Location_Log_Template_R01.xlsx

DFS_Short Form_Location_Session_Log_Template_R01

These templates can be used electronically, i.e., for data entries into appropriate file copies on a computer or by printing from templates and filling out these copies by hand.

In the case of electronic input, files should be copy protected after completion, and a copy kept in a safe location, as a precaution against data loss or accidental alteration of content in later stages of processing.

The first type of log sheet, providing more guidance but also demanding a higher effort to be filled in, is meant to be used by teams that have limited experience or are meant to work under conditions with a higher level of potential distraction where formalized checklists can provide additional operational robustness. The second type of log sheet is designed for experienced teams or teams that have to operate under conditions where time to fill in details is limited. In practical situations, it would be, for example, the task of the FTL to select the most appropriate type with respect to the actual situation.

The log sheets make provisions for some elements of the testing procedure that are optional, to be decided by the FTL as follows.

– Photographing or scanning log sheets and sending them by e-mail (applicable to cases where printed copies are used): frame information should be available at the point of data evaluation as fast as possible, to allow for further steps in post processing and measurement data quality management. Taking a scan or photo and sending them immediately after a testing session at a given location is a means to that end; it is also a precaution against potential loss of such sheets along the way. If it can be reasonably expected that log sheets reach the point where they are collected and processed further in a short time and with low probability of loss, this step can be omitted. In that case, log sheets with respective fields removed or marked as "not applicable" should be created from the respective templates.
– Entering check marks for intermediate tests of network connectivity or power supply status: if it can be expected that teams have the required skill level and proficiency, log sheets with respective fields removed or marked as "not applicable" should be created from the respective templates.

Another practical simplification can be made by printing log sheet copies with fields already pre-set. Appropriate decisions should be made by the FTL based upon judgement of the actual situation.

II.7 Team assignment list

Remark: This type of list is also used with the name scenario master list.

This list holds information about the assignment between team names, scenarios and time windows associated with scenarios.

In data processing, it is used – in combination with a DAL (see clause II.8) – to link measurement data with relevant scenarios.

The TAL structure is chosen so that it is user-friendly (in the sense that its structure is not more formal than required). However, it must be clearly emphasized that great care is strongly advised when creating and maintaining this list. Undetected errors can cause artefacts and errors in later stages of processing that are hard to detect. In particular, the naming of teams and configurations must be consistent between the DAL and the TAL.

Assuming that the DAL and TAL are imported to a database for final processing, it is highly recommended that appropriate check and validation procedures be included for TAL and DAL structure and content.

As the content of a TAL is quite project-specific, no templates are provided. It is, however, recommended that Table II.1 be used for construction of the appropriate files.

Table II.1 – Structure of a typical team assignment table with example content

Column name	Example	Type	Function	Hints/Further remarks
Test_Scenario	InterNetwork	Text	Scenario description	see further remarks
OwnerTeam	MTN Ghana	Text	Team name	Can be set by spreadsheet formula (default: equal to FromConfigName)
FromCountry	Ghana	Text	Country ("From" role)	Primary input field
FromOperator	MTN	Text	DFS/network operator, "From" role	Primary input field
ToCountry	Ghana	Text	Country ("To" role)	
ToOperator	Vodafone	Text	DFS/network operator, "To" role	
Status	Completed	Text	Status of tests	As the TAL can also serve as planning element, the status may contain different states such as "Ongoing" or "Cancelled"
Remarks	Unlimited transfers	Text	Additional scenario-specific information	
MaxTAPerDay	10	Integer	Indicates if there is a systematic limit for the maximum number of transactions per day	Leave this field empty if there is no such limit. The value is intended to be used to compute mask-out indicators during data evaluation. Enter only values that affect the number of outgoing transactions.
From_Date_Time	11.05.2020 00:00	Date (time optional)	Start of assignment	
To_Date_Time	15.05.2020 23:59	Date/time	end of assignment (inclusive)	Make sure to include time for proper time windowing
FromConfigName	MTN Ghana	Text	Configuration name, "from" role"	Typically constructed from respective operator or country fields by formula
ToConfigName	Vodafone Ghana	Text	Configuration name, "to" role"	Typically constructed from respective operator or country fields by formula
TransactFlow	MTN Ghana to Vodafone Ghana	Text	Auxiliary field for readability	Default: constructed from respective ConfigName fields, can be overwritten

Remarks:

This list in Table II.1 is understood to describe the minimum required for processing. The list can be amended by additional columns as seen fit by campaign management.

The TAL uses a couple of primary input fields (e.g., for operator and country). This facilitates generation of "logical names" needed for processing in an automated way, e.g., by spreadsheet

formulas. Experience has shown that when entering free text, errors that are difficult to detect by eye can easily slip in (e.g., two blanks instead of one), which can cause substantial trouble in automated processing. Therefore, it is recommended that formulae be used by default and only override content when strictly necessary.

Some of the fields (e.g., TransactFlow) are there for convenience of usage. If the TAL is a spreadsheet list, these fields should also be pre-set with respective formulae, to be overwritten when appropriate.

Please note that the TAL uses the configuration names (not the team names) as the purpose of the TAL is to link measurement data to scenarios.

For final processing and computation, a unique scenario descriptor is required (typically, in databases, aggregation of data into respective KPIs is done using GROUP statements). This can be done directly using the scenario description field in the TAL. In that case, ensure that each description text is unique. Another way can be to use a generic scenario descriptor (e.g., for a category such as "cross-border" and to construct the final scenario description use for aggregation from the corresponding fields of the TAL, e.g., from the scenario, the FromConfigName, and the ToConfigName.

Date range information may be crucial for later allocation of data. Typically, JOIN statements with corresponding ON conditions including time-range BETWEEN clauses will be used. Make sure that date and time ranges given in the TAL are consistent, either by using complete date and time information or by using respective DATEADD statements.

Column names should be chosen not to contain blank spaces or other non-alphanumerical characters. This may ease data processing e.g., in SQL databases. If desired, creation of "friendly text" for output can be done in respective SQL statements, e.g., replacing underscores (_) by blanks.

II.8 Device assignment list

A DAL connects measurement data to teams.

In the following, two IDs are assumed.

The MSW ID identifies data from the MSW app.

Data from network background measurement are identified by the IMEI of the device on which the respective testing app is installed.[4]

The DAL structure is chosen so that it is user-friendly (in the sense that its structure is not more formal than required). However, it must be clearly emphasized that great care is strongly advised when creating and maintaining this list. Undetected errors can cause artefacts and errors in later stages of processing which are hard to detect. For instance, a typing error in an IMEI may lead to "ignored" data items with associated consequences for completeness and correctness of results. The same goes for date and time entries.

The basic structure of the DAL mirrors a team, named after the DFS operator and the country it is testing. This team uses a pair of devices according to the basic role assignment and set-up described in clause 11 onwards of this Recommendation. However, the structure also provides for extensions of this basic scheme as follows.

– A team may change the devices during the campaign or re-installation of apps can lead to new IDs being allocated.
– Devices can assume different roles, e.g., devices can switch between the DFS and the observer role.
– Devices can be swapped between teams or teams can be allocated to different testing tasks (e.g., a team testing operator A can start to test operator B after some time).

[4] This also provides the appropriate degree of genericity, as there is no specific application prescribed for these tests.

The DAL has a column structure that is, for reading convenience, shown there transposed with explanations for each column.

As the content of a DAL is quite project-specific, no templates are provided. It is, however, recommended that Tables II.2 and II.3 be used for construction of appropriate files.

Table II.2 – Structure and content example of a device assignment list

Column name	Example	Type	Function	Hints/Further remarks
CheckStatus	Ok	Text	Can be used for validation vs. log files	Free text, can be empty
OwnerTeam	MTN Ghana	Text	Name of the team	By default, same as configuration name (A spreadsheet DAL can use a formula here)
Country	Ghana	Text	Country	
Operator	MTN	Text	Operator (DFS/mobile network)	
ConfigName	MTN Ghana	Text	Logical configuration name (default: Operator<blank>Country)	A spreadsheet DAL can use a formula here
Start_Allocation	30.04.2020 00:00	Date/Time	Date (time is optional)	see remarks
End_Allocation	19.06.2020 00:00	Date/Time	Date/time (see remarks). Can be left empty	see remarks
Config1_DFS_IMEI	354481115999999	Text	IMEI of the device in the primary DFS role	Although an IMEI only contains numeric characters, relevant fields should be treated as text to avoid artefacts by conversion. May also use "text-decorated" variants to ease import to databases
Config1_Obstool_IMEI	354481115888888	Text	IMEI of the device in the primary "observer" role	
Config1_FIMSW_ID	c07eead8-62e8-41dc-b3fb-d5c1b3a7cde6	Text	MSW ID of the device in the primary Observer role	
Config2_DFS_IMEI	354481115888888	Text	IMEI of the device in the secondary DFS role	A spreadsheet DAL can use a formula here to get content from "opposite" fields
Config2_Obstool_IMEI	354481115999999	Text	IMEI of the device in the secondary "observer" role	
Config2_FIMSW_ID	d12c73fe-8961-450a-8b9f-bc431e394c3e	Text	MSW ID of the device in the secondary Observer role	

Table II.2 – Structure and content example of a device assignment list

Column name	Example	Type	Function	Hints/Further remarks
Remarks		Text	Free text	

Remarks:

The DAL structure shown here has the functionally required minimum number of elements. Additional elements may be useful, e.g., the phone number associated with a device.

Typically, the Start_Allocation and End_Allocation fields include pre- and after-campaign ranges, and actual time windowing for KPI reporting is made in appropriate processing steps. In a typical database-based post processing, JOIN operations also using these time ranges are applied. These fields can, however, also be used for tighter time-windowing only considering actual testing-campaign times.

If device allocations do not change over time, both the Start_Allocation and End_Allocation fields can be left empty. In that case, database procedures must make sure proper treatment of content e.g., by setting default dates and times.

If device allocation changes during the campaign, Start_Allocation and End_Allocation must be used to properly describe these allocations. In corresponding JOIN operations, SQL BETWEEN statements are used; if only a date is given without time information, 00:00 is assumed. Ensure that appropriate time regions are complete (i.e., include hour and minute information, e.g., 1.2.2020 23:59 in End_Allocation to cover the whole day of 1.2.2020), or use respective DATEADD functions in BETWEEN statements.

Column names should be chosen not to contain blank spaces or other non-alphanumeric characters. This may ease data processing e.g., in SQL databases. If desired, creation of "friendly text" for output can be done in respective SQL statements, e.g., replacing underscores (_) by blanks.

Appendix III

Best-practice examples for background testing of mobile networks

(This appendix does not form an integral part of this Recommendation.)

This appendix is based on relevant content in [b-ITU-QoS_03].

The following set of use cases and KPIs attempts to provide a good overview with reasonable effort. It does not include SMS or USSD for the following reasons.

– SMS is not a primary transport service for DFS-related information. It is only used to transfer notification SMS to party B (which is irrelevant in asynchronous mode anyway), and for side A, SMS acts only as a secondary indicator.

– As can be seen in relevant reports ([b-ITU-QoS_01], [b-ITU-QoS_02]), USSD performance has been shown not to be highly correlated with DFS performance. In a multi-network, multi-country campaign, it will also be hard to find USSD codes that work for all networks involved. In summary, the effort to include USSD in such campaigns should be carefully considered from a cost to value point of view.

Test case	Key performance indicator
Web Browsing (ETSI Kepler SP reference web site)	**End-to-end session time (E2E ST)** if transactions are successful. In contrast to the ETSI TS 102 250-2/ITU-T E.804 session time, the time window begins with the start of web site download (not with receipt of the first package). **End-to-end success rate** (percentage of transactions successfully completed, from all valid transactions (TAs). A valid TA is a TA run via mobile network (not via wireless fidelity (Wi-Fi)), and not blanked out, e.g., by a user break indication.
HTTP DL with 3 MB; time-out 30 s	**End-to-end session time (ST_E2E)** if transactions are successful. Analogously to web browsing, this session time includes the initial start time. **End-to-end success rate** (percentage of transactions successfully completed, from all valid TA. Evaluation is done in fixed-size mode, i.e., a TA that ran into a time-out is not counted as successful. **Mean data rate end-to-end: (MDR_E2E)**: Effective data rate. This value is also output if the result if the TA is unsuccessful (e.g., dropped or ran into time-out); in that case the transferred data up to the stopping point, and the time expired, is used to compute the MDR.
HTTP UL with 1 MB file in fixed-time mode; time window 30 s (hybrid mode)	In hybrid mode, the TA ends either when the intended data volume is transferred or the time window is expired. In this mode, reaching the end of the time-window does not result in the result "unsuccessful". If desired, a computational "unsuccessful" state can be created by evaluating the TA duration. **End-to-end session time (ST_E2E).** By computation, this value is created only when the end of the time window is not reached (to stay consistent with standardized KPI computation. Analogously to web browsing, this session time includes the initial start time. **End-to-end success rate** (percentage of transactions successfully completed, from all valid TA. Evaluation is done in "computational fixed-size mode", i.e., a TA that ran into a time-out is not counted as successful.

Test case	Key performance indicator
	Mean data rate end-to-end: (MDR_E2E): Effective data rate. This value is also output if the result if the TA is unsuccessful (e.g., dropped or ran into time-out); in that case the transferred data up to the stopping point, and the time expired, is used to compute the MDR.

The definition of valid TA excludes transactions that are taken via Wi-Fi, were interrupted by the user ("user break") or are masked-out otherwise. Through joining with the TAL, there is also an effective time-windowing to exclude TA taken outside the date range of the relevant scenario. Due to the fact that measurements were taken stationary (in the same location), there is, however, no time windowing with respect to MSW time ranges.

MDR values are, different from standard MDR averaging, taken over all TA including unsuccessful ones. This avoids biasing towards higher expected values, which occurs when timed-out transactions are excluded from averaging.

Session time values are calculated over values from successful TA only to avoid inconsistencies by clipping. When interpreting data, success rates need to be considered along with ST values.

When setting up a scenario for network testing, it also needs to be considered where respective content is hosted. The effort to be taken is, again, a matter of scope and purpose of measurement. If network KPIs are only required to have an indicative or secondary character, a simple approach, i.e., by hosting all content on the same server, can be taken (located in one of the participating countries, or elsewhere). If precision measurements are intended, multiple server locations with high supported bandwidth may be required, possibly accompanied by calibration and validation testing.

Bibliography

[b-ITU-T E.800] Recommendation ITU-T E.800 (2008), *Definitions of terms related to quality of service.*

[b-ITU-T P.10] Recommendation ITU-T P.10/G.100 (2017), *Vocabulary for performance, quality of service and quality of experience.*

[b-ITU-QoS_01] Security, Infrastructure and Trust Working Group of the Financial Inclusion Global Initiative (2020). *Methodology for measurement of quality of service (QoS) key performance indicators (KPIs) for digital financial services.* Geneva: International Telecommunication Union. 94 pp. Available [viewed 2023-05-13] at: https://figi.itu.int/wp-content/uploads/2021/04/Methodology-for-measurement-of-Quality-of-Service-QoS-Key-Performance-Indicators-KPIs-for-Digital-Financial-Services_f-1.pdf

[b-ITU-QoS_02] Security, Infrastructure and Trust Working Group of the Financial Inclusion Global Initiative (2020). *Methodology for inter-operator and cross-border P2P money transfers – Report of the quality of service workstream*. Geneva: International Telecommunication Union. 44 pp. Available [viewed 2023-05-13] at: https://figi.itu.int/wp-content/uploads/2021/04/Methodology-for-measurement-of-Quality-of-Service-Key-Performance-Indicators-for-inter-operator-and-crossborder-P2P-money-transfers-1.pdf

[b-ITU-QoS_03] Security, Infrastructure and Trust Working Group of the Financial Inclusion Global Initiative (2020). *Test campaign report: Cross-border and inter-operator testing of digital financial services, P2P money transfer – Report of the quality of service workstream*. Geneva: International Telecommunication Union. 47 pp. Available [viewed 2023-05-13] at: https://www.itu.int/en/ITU-T/extcoop/figisymposium/Documents/Test%20Campaign%20Report%20for%20Cross-border%20and%20inter-operator%20testing%20of%20Digital%20Financial%20Services%2c%20P2P%20money%20transfer.pdf)

[b-ETSI TS 102 250-2] Technical Specification ETSI TS 102 250-2 v2.7.1 (2019), *Speech and multimedia transmission quality (STQ); QoS aspects for popular services in mobile networks; Part 2: Definition of quality of service parameters and their computation*.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	**Telephone transmission quality, telephone installations, local line networks**
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Published in Switzerland
Geneva, 2023