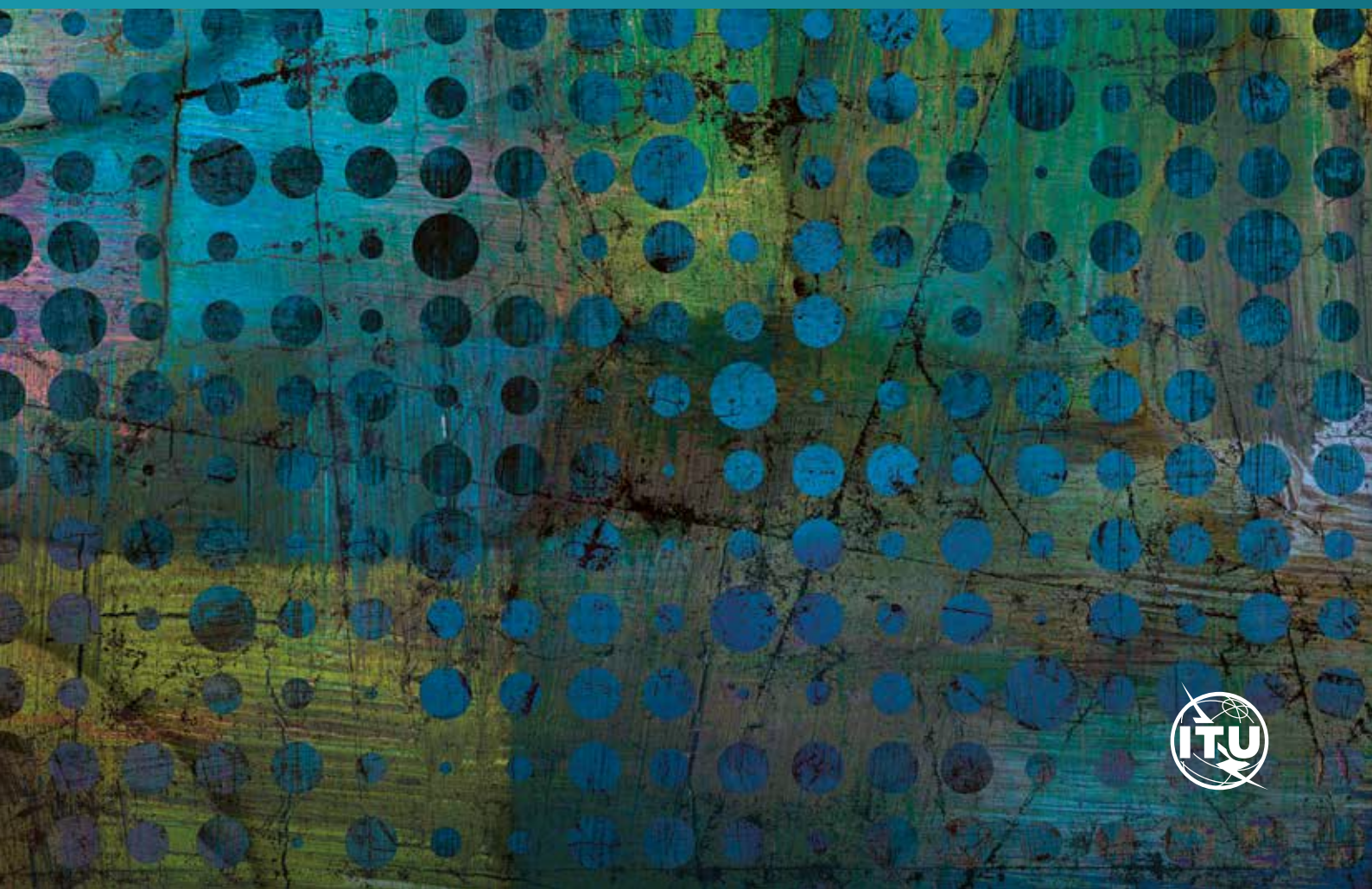
FIGI

FINANCIAL INCLUSION
GLOBAL INITIATIVE

SECURITY, INFRASTRUCTURE AND TRUST WORKING GROUP

Methodology for measurement of Quality of Service (QoS) Key Performance Indicators (KPIs) for Digital Financial Services

REPORT OF QUALITY OF SERVICE WORKSTREAM



SECURITY, INFRASTRUCTURE AND TRUST WORKING GROUP

Methodology for measurement of Quality of Service (QoS) Key Performance Indicators (KPIs) for Digital Financial Services

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.

A new global program to advance research in digital finance and accelerate digital financial inclusion in developing countries, the Financial Inclusion Global Initiative (FIGI), was launched by the World Bank Group, the International Telecommunication Union (ITU) and the Committee on Payments and Market Infrastructures (CPMI), with support from the Bill & Melinda Gates Foundation.

The Security, Infrastructure and Trust Working Group is one of the three working groups which has been established under FIGI and is led by the ITU. The other two working groups are the Digital Identity and Electronic Payments Acceptance Working Groups and are led by the World Bank Group.

© ITU 2019

This work is licensed to the public through a Creative Commons Attribution-Non-Commercial-Share Alike 4.0 International license (CC BY-NC-SA 4.0).

For more information visit https://creativecommons.org/licenses/by-nc-sa/4.0/

About this report

This report was written by Wolfgang Balzer and Joachim Pomy. This report describes a methodology to measure Quality of Service (QoS) Key Performance Indicators (KPIs) for digital financial services. The QoS KPIs for DFS were based on the ITU-T Focus Group Digital Financial Services report on *QoS and QoE aspects of Digital Financial Services* developed by the ITU-T Focus Group Digital Financial Services (FG DFS).

The methodology described in this report was validated on the basis of field measurements carried out in Ghana. The QoS KPIs for DFS and the methodology described in this report are also being considered as an international standard in ITU-T Study Group 12.

Special thanks to Focus Infocom GmbH, Kwame Baah-Acheamfuor, ITU-T SG12 Chairman, National Communication Authority of Ghana and the members of the Security, Infrastructure and Trust Working Group for their comments and feedback.

For queries regarding the report, please contact Vijay Mauree at ITU (email: tsbfigisit@itu.int)

Contents

Executive Summary

The report describes a methodology for measurement of Quality of Service Key Performance Indicators (QoS KPIs) for Digital Financial Services (DFS), based on the Technical Report *QoS and QoE aspects of Digital Financial Services* developed by the ITU-T Focus Group Digital Financial Services (FG DFS).

The purpose of this methodology is to:

- Define processes for designing and performing DFS QoS field tests and computation of respective KPI.
- Define procedures to assure data quality and integrity of the KPI results.
- Establish a data driven basis for guidance of regulators and enables definition of requirements, which assure good service quality for DFS.

The methodology is beneficial to both telecom regulators and DFS operators in the following ways.

- Regulators are enabled to identify parameters that relate to quality of mobile money transactions that can be tested. Therefore, means of measurement of end-to-end quality of mobile money transactions can be established. Results can be used to benchmark service providers, and to set meaningful targets for service quality. Finally yet importantly, regulators have a tool that allows understanding the state of the mobile money market on a solid data-driven basis.
- For network operators and, likewise, service providers, the methodology and related quality metrics guide the planning of service provisioning to satisfy targets set for each parameter. Service quality parameters can be tested and monitored to ensure regulatory compliance and consumer satisfaction, and results can be used to improve service performance and user experience. The methodology can be applied across country markets and therefore improves overall efficiency of operations.

The methodology combines two approaches that complement each other. The first approach is to define computation for DFS KPI, which cover the whole service from an end-to-end, user related perspective. The second approach is a combination of direct tests of DFS related use cases with background measurements of standard basic network services such as USSD, SMS, and packet data services.

A DFS ecosystem consists of two components, the mobile network(s) over which the services are delivered, and the infrastructure, which handles the actual money transfer. Different parties may operate these components. The methodology allows testing the end-to-end performance of the entire service and actually provides KPIs, which implicitly address the money transfer infrastructure. The methodology focuses on the mobile network part as this is the component where location and usage dependent mechanisms affect DFS performance and therefore field tests are required.

DFS are delivered using existing and well-defined mobile network services as transport medium. Therefore, there is a solid basis for using respective service tests as proxies for DFS QoS. This can help to simplify testing and makes it easier to define meaningful regulatory requirements for mobile networks over which DFS are offered; this is done by linking basic KPI to user expected DFS performance.

The methodology describes money transfers in a generic way that is independent of the DFS implementation. It uses a set of events to define a set of KPIs which describe the performance of DFS and which can be linked to observable events in a specific implementation.

The set of KPIs provided by the methodology covers basic end-to-end metrics, this includes:

- KPIs such as the success rate and completion time for money transfers; focused on person-to-person money transfers but expandable to other DFS use cases.
- KPIs that can be used to express potential weaknesses or failures of the underlying money transfer management process, such as the False Positive and False Negative Rate, where the system is reporting the outcome of transfers incorrectly.
- KPIs for loss and duplication of money
- KPIs for final account stabilization.

A supplementary set of KPIs is provided which is specifically designed for tests executed manually on normal end-user devices, i.e. where only a subset of events can be observed; the accuracy of time-taking is limited by the manual execution, and where components of manual activities can only to some extent be separated from results obtained by measurements.

Testing DFS, like field quality assessment in general does not only require well-defined KPIs, but careful design of testing procedures. In particular, data quality assurance is essential to ensure validity of results and optimum use of resources. The methodology therefore covers the entire range of manual tests up to automation of tests, to provide best-practice information for implementers of testing campaigns.

1 ABBREVIATIONS AND ACRONYMS

QoS	Quality of Service
QoE	Quality of Experience
DFS	Digital Financial Services
KPI	Key Performance Indicator
SMS	Short Message Service (also used for a single text message transmitted by SMS)
NSMS	Notification SMS
DID	Device Identifier
XML	Extensible Markup Language
PIN	Personal Identification Number

Wording

The following table shows abbreviations used in the presented document, as well as working names in cases where different commonly used names exist. In such cases, one of these names has been chosen as the working name throughout the present document, and alternative names are understood as aliases.

WORKING NAME OR DEFINITION	TERM/ALIAS
DFS (Digital Financial Services)	MoMo (Mobile Money)
A or B Party, Account (actually the representation of an user's account on a mobile device or another type of TE)	Digital wallet, Wallet
PFT (Pilot field test)	Only internal use to designate the pilot test campaign in Ghana
TA	Transaction
ObsTool	Observer Tool: User Equipment running software for active and passive network testing

Extensions

For the time being, the PFT uses only the P2P use case. There are requests to extend to other use cases, such as

- Person to Government payments
- Variants involving interoperability (national and international, between different service providers)
- Transfers between a person's MoMo and bank account

These extensions will required extended and/or modified modelling and related procedures for measurement and data evaluation. In the course of the current project, it will be decided how to extend the scope, and to integrate these or other use cases, into the methodology and potential additional field tests.

Methodology for measurement of QoS KPIs for DFS

2 INTRODUCTION

Regulators of both the financial and the telecommunication sectors are encouraged to collaborate in using the present report as an initial toolbox for the assessment of DFS related aspects of QoS and as far as possible as QoE.

DFS applications preferred by customers will swiftly change in functionality, structure and thus also in complexity. These swift changes will differ by country or region and international interoperability will add even more complexity.

There is not and there will not be a particular QoS and QoE test suite that could be applied to all DFS applications. Therefore, the challenge for regulators of both sectors is to standardize QoS and QoE test suites tailored to the needs in their country or region such that customers can rely on smoothly flowing DFS services can be trusted alike the many other utilities keeping an economy up and running.

It is however important to know for regulators that they are not left alone with this challenge. International SDO like the ITU have started work items planning to come up with new standards in the field of QoS and QoE for DFS. As this work is contribution driven, regulators are encouraged to actively participate in the work in this sector to make sure that these standards are optimally suited to their needs.

The present document is based on the recommendations for end-to-end QoS KPI definitions first published in the Focus Group Technical Report QoS and QoE aspects of Digital Financial Services by the ITU-T Focus Group Digital Financial Services. It details the methodology and connects to a field test using this methodology, which has been conducted in Ghana in the first half of 2018.

Money transfer from end user's devices to other devices or to other entities has become an important element of everyday life in many countries. This service, however, relies on the functionality of mobile networks. Therefore, a connection exists between the functioning and the QoE of money transfer services, and the QoS and proper functioning of those mobile networks, and respective quality metrics and testing methodologies need to be defined.

The main part of the present document describes the testing methodology and concludes with a summary of lessons learned from conducting the pilot field trial and a proposed way forward.

Annex A provides examples for basic tests on a target service prior to setting up a testing campaign. Annex B describes an example set of procedures used to ensure proper set-up and functioning of systems used for testing. Annex C describes details on the set-up and specifics of devices when using the circular testing

scheme described in the present document. In Annex D, the basics for design and definition of QoS and QoE metrics are described for readers which are not already familiar with the topic. Annex E gives examples for consistent naming of files and other data items in a campaign. Annex F provides an overview table of KPIs and related trigger points.

Furthermore, appendices provide specific information on the pilot testing campaign itself, which has been performed in Ghana in the first part of 2018. Appendix I shows the log sheets used which can serve as a template for similar campaigns. Appendix II, provides an overview of the testing campaign in Ghana itself, while Appendix III shows how a component of data sourcing, a specific tool for backing up notification SMS, is set up.

Please note that the current document 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 defined load to a DFS ecosystem in order to determine the robustness of DFS functionality under these conditions. Such extensions can be easily created from the given methodology. Their execution is mainly a matter of scale of required resources.

3 TEST SCENARIO UNDER CONSIDERATION

In the following, the use case "Person-to-Person" (P2P) money transfer is described. The methodology is designed to be easily extended to other use cases in future projects.

3.1 Definitions

3.1.1 Roles, entities and abbreviations

TABLE 3-1: Terms and abbreviations specific to descriptions of DFS test cases

A Party and B Party	Formal roles for transfer, e.g. A (active role) transfers money to B (passive role).
SPx and FPx	Designation of device types used for a transfer, Smartphone x, Feature phone x.
Dx	More general indexed device description (e.g. D1, D2...)
OPx	Observer Phone x
Px	Person x, designation of a tester/ operator (independent of role)

- NOTE 1: The testing methodology comprises typically of a round-trip transfer in definite N steps. As a result, thereof, roles between devices and operators are switched after every transaction.
- NOTE 2: In order to optimize testing efficiency and to minimize the risk of errors during the test preparation, the assignment of devices to accounts should be fixed. Consequently, the assignment of roles is being switched between the devices in a cyclic manner (with Smart Phone to the left and Feature Phone to the right of a Person during manual tests). This is to ensure that the manual cycle of tests is uniform to the transaction cycle as illustrated in Table 3 1.

3.1.2 Definition of action flows

A team, which conducts a test, typically consists of two persons, named P1 and P2. Alternative team sizes (e.g. five persons assigned to one of the four testing phones and the observer phone) or the option to use more than one team per location are for further study. Based on experience made so far, it appears that any such solution should be accompanied by increased tool support (such as partly automated time-taking as described in clause 10.2) in order to manage the increased level of work intensity and to maintain high data quality.

This team, acting in the roles of A and B party, respectively will do a single transfer.

In parallel to the actual transfer action, the person designated as P2 also operates the observer phone (since P1, in the A party role, is engaged with performing the transfer while P2 in the B party role is mostly idle with respect to the money transfer).

A cycle of transfers consists of four (4) transactions, using all combinations of smartphones and feature phones assigned to the A and B party roles. After this cycle, the money transferred (less operators' charges) is again available on SP1 and FP1, respectively.

3.2 Neutral starting state

A particular property of systematic service tests is a frequency of service usages which is significantly higher than the usage frequency created by a typical end user. A testing campaign, therefore, should contain systematic tests to make sure that usage frequencies typical for testing do not affect testing results with respect to the end-user perspective. A possible way to realize this can be systematic tests, which are run prior to the actual campaign and where the testing frequency is varied.

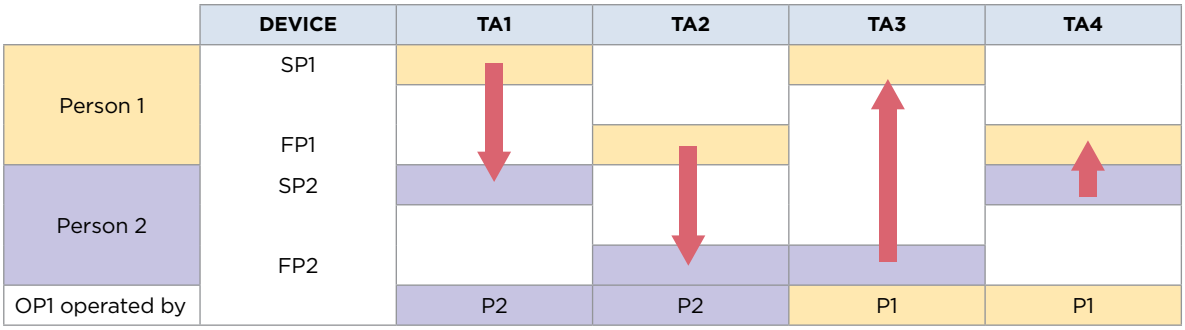
There are two categories of effects. Firstly, as for services other than DFS, after each service usage a certain relaxation or guard time is expected to be required. This is well understood, e.g. for guard times after telephone connections have ended — in order to enable the service to reach its neutral state again. This first category is considered to be uncritical as such guard times are typically in the range of 10 to 30 seconds. Depending on the respective implementation, a second category of effects may, however, exist which concerns longer periods of time.

For example, some kind of local memory (e.g. like a web browser's cache) where content already downloaded will be kept for a longer period of time or even for an indefinite time may be involved.

The request for the same content at a later point in time would then access the local cache, directly instead of triggering an actual data transfer, and therefore would massively impact the correctness of the results of the test campaign (by showing quite short transfer delays).

According to findings of respective pre-tests, appropriate steps should be taken (such as clearing local memories). As long as effects are quantitative rather

TABLE 3-2: Role and activity assignments during a 4-transaction cycle



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

With respect to guard times, it is conceivable that the system has a certain "dead time" after each transaction, where the system would not accept a new transaction or create unexpected results of a transaction attempted within this period. It is advisable to be aware of this possibility and obtain respective information before actual parameters of a test campaign are determined. Technically, it would also be possible to probe for such effects. This would, however, require a sufficiently controlled execution of testing, i.e. automated test control with the ability to systematically reduce inter-transaction time and check for related effects.

3.3 Re-initialization after unsuccessful transactions

If a transaction fails, in particular after a time-out condition has occurred, it shall be ensured 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.

3.4 Disappeared 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, we use the term "disappear" for both variants of this kind of effect.

- NOTE 1: In cases of disappeared money, insertion of fresh money will be necessary.
- NOTE 2: Retrieval of lost money should be treated as a second stream of activities.

3.5 Automation of tests

The methodology in the present document describes testing in a generic way, i.e. service tests can be done manually as well as in an automated way. 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 of e.g. manual time measurements. It is likewise understood that such automation requires a higher initial effort to ensure reliability of operation under unsupervised conditions or to cover a wider range of end-user devices.

4 TRANSACTION MODEL

4.1 Person to Person (P2P) Mobile Money (MoMo) transfer

4.1.1 Transaction description

Abstract: Transfer of a known amount of M units of money from account A to account B.

Success definition: The correct amount plus applicable operators' fees have been deducted from the A party account and the correct amount (net) has been credited to the B party account within the defined time window.

Examples for unsuccessful execution are cases

- Where the system sends—at any stage of the transfer—an explicit response indicating failure of the transfer.
- Where the transfer has been done but the amount is wrong.
- Timed-out and still pending TA.

- NOTE 1: The description does not explicitly refer to assignment of roles to devices or operators. For instance, if a particular device is assigned to represent a given account, the device may be operated as A Party or B Party. Related events occur, and related activities are performed, on the respective device.

- NOTE 2: Some service implementations may also offer a "tokenized" transfer which is in effect also a P2P transfer. In this case, the transfer done by the A party would create a token which can be transferred to a B party. This type of transfer is considered to be a special case and is not considered here.

4.1.2 Event and action flow

The core of a P2P MoMo transfer consists of instructing the DFS to transfer money from the A party's account to the B party's account.

In order to do so, the service requires information items such as the respective account ID's, information text for the transaction and the amount to be transferred. Also, the transfer will be authenticated by providing a respective token such as a PIN.

There are many conceivable ways the user interface may be designed. Most details are not relevant for modelling of a generic use case—such as the order in which required information items are gathered.

4.1.2.1 Involvement of the mobile network in the MoMo process

There is, however, an important exception which is highly relevant. This is the degree to which the mobile network is involved in the MoMo process. There are two general options:

a) All information is collected locally, and afterwards a single data block is sent to trigger the actual money transfer. This will be referred to as type A.

b) The information is collected item-wise, with exchange of data over the network after each step. This will be referred to as type B.

These options define the extremes of a *network involvement type scale* where an actual implementation is described by a value between those extremes (assigning them, eventually, type identifiers for easier reference). For instance, the local (A-party side) application may collect type and recipient of a payment, then validate the user exists; then it may request the amount to be transferred to check if it is within the limits of the A party's balance and contract, and finally request the remaining elements, including the A party's authorization, to validate the transfer.

- NOTE: The differences belong, from a generic modelling perspective of a MoMo transaction, to the 'service set-up phase'. Collecting the information is prerequisite to conduct the transaction, but these steps do not provide any customer value by themselves. The customer value materializes in the actual performance of the money transfer which is the subsequent step.

Figures 4-1 and 4-2 illustrate the perspectives graphically.

Figure 4-1 shows a MoMo implementation where all information is collected locally in the A side DFS agent (e.g. an app, or a function implemented in the SIM of the device) and is then transferred to the DFS. In this example, the DFS sends three data items in response:

- The primary confirmation is sent to the A side's local agent
- A secondary confirmation may be sent also to the A party through another channel, such as SMS
- A confirmation that money has been transferred is sent to the B side. As this is an unsolicited message (the B party is not actively participating in the transfer), a respective channel (such as SMS) is used.

In Figure 4-2, a MoMo implementation is shown where the information required for a DFS transaction is collected successively by prompts from the server (intermediate variants are also possible, where some information is requested as a group).

The figures also show a common element that is important for both modelling and methodology. There is an event "Show TA completion" on the A side. It represents a message from the service indicating that the transaction has been completed. It is therefore called the *primary completion indicator.* Completion is used here as the most general case for a distinctive message from the system which by itself only marks a defined end of the transaction, which can be a successful operation or an unsuccessful one. If the transaction has been performed successfully, this event is also called *primary success indicator.*

In real MoMo implementations, there are additional messages generated by the MoMo implementations, which contain a summary of the transaction (including, for the A side, information about fees charged). These messages are typically sent by store and forward service such as SMS.

From a functional point of view, they can be considered as additional information which is, at least for the B side, important from of the customer perspective but not critical or indicative for the DFS core transaction; debiting and crediting money has taken place already. Consequently, these events and information elements are considered to be secondary indicators; they are not crucial for the following considerations of type-variant dependent dynamics.

In the context of the current methodology, it is assumed that the SMS containing summary information represent the final and correct information about the A and B side account balance. Technically, it is possible that these SMS may contain erroneous content with respect to actual book-keeping. It is, for state-of-the-art systems, unlikely that such an essential element of a DFS implementation is faulty.

FIGURE 4-1: Entities and event flow for a DFS implementation where required information is collected locally, and then transmitted to the service (Type A).

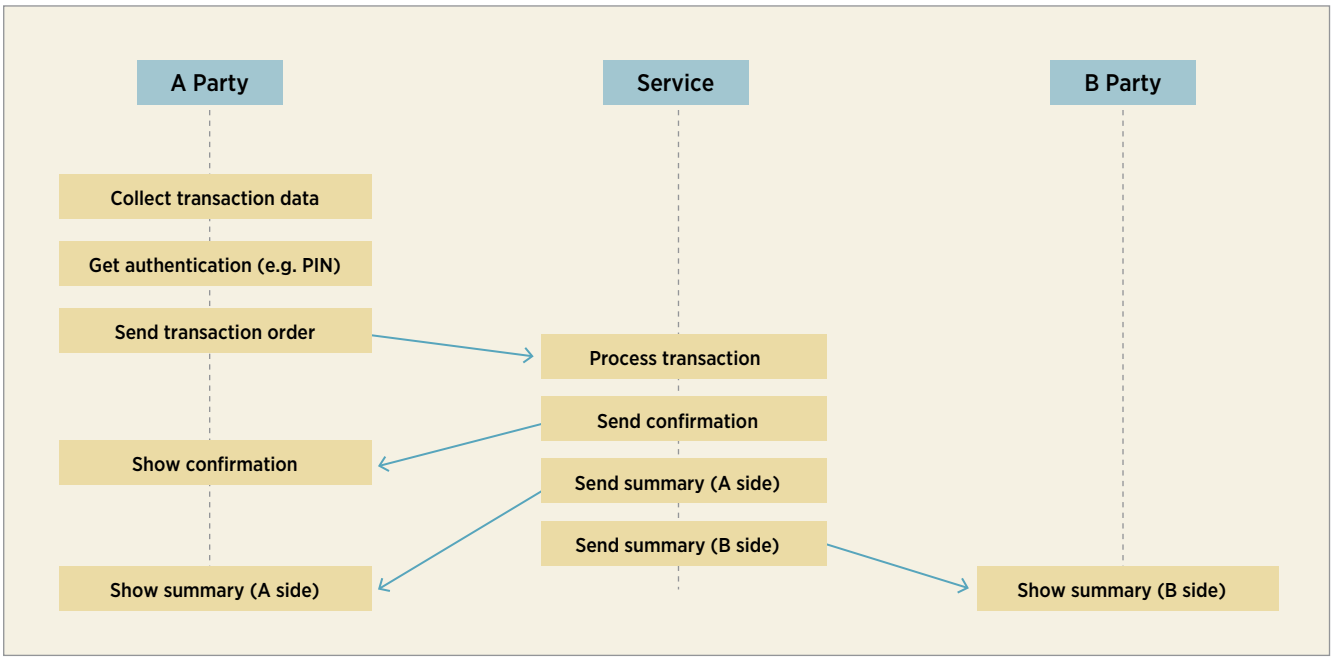
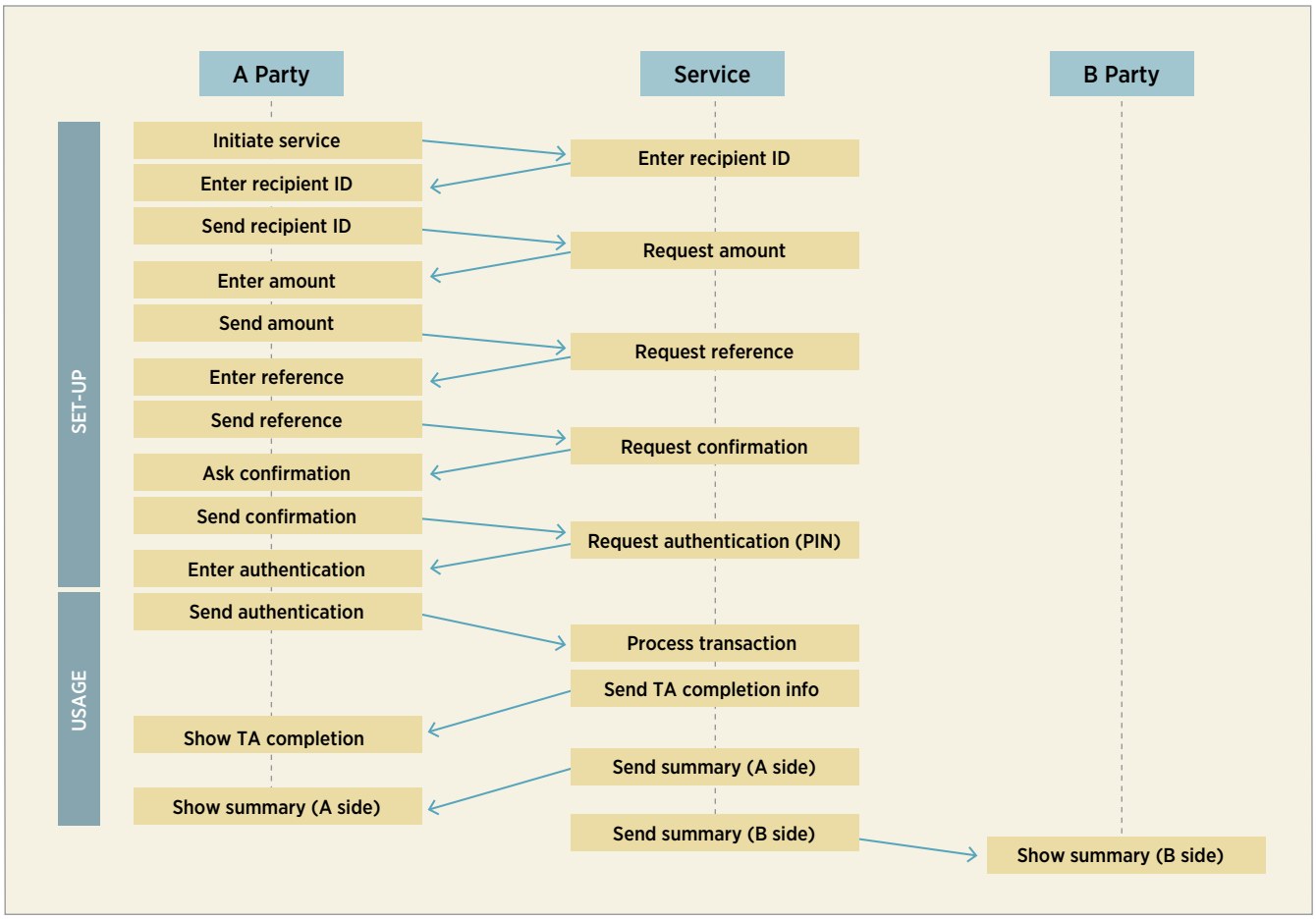


FIGURE 4-2: Entities and event flow for a MoMo implementation where required information is collected element by element by the service (Type B).



From a QoS perspective, and therefore also for network testing, the degree of network involvement should be considered to be mission-critical.

The number of data exchanges through the network is much higher in type B than in type A. As the overall success of the MoMo transaction depends on the success of each of those steps, the MoMo success rate has a stronger dependency on network performance than in type A. Secondly, the collection of information items involves human interaction, i.e. typing. This extends the time during which the network needs to perform well considerably which plays a role in mobile scenarios.

On the other hand, Type B implementations enable collection of more information about the network's performance as each step in the information-gathering phase provides an information source for respective indicators. This topic will be discussed in detail in subsequent clauses. Briefly, the question is if it makes sense to define KPIs for every possible combination of events—which is technically possible but may obscure things rather than provide insights.

In this context, not only network quality should be considered but also mobile device related effects, like running out of battery power. Annex B provides an overview of elements for checking.

Of course, this discussion does not change the necessity of using the actual implementation of the service. From a QoE point of view, there is no choice—the whole transaction has to be taken into account if the test results are assumed to describe the customer perspective properly.

4.1.3 Phase definition

4.1.3.1 Top-level phases

Set-up: Preparations for the actual transfer:

- Activation of service.
- Input of required information (destination account, amount of money to be transferred, reference, credentials to enable the transfer e.g. password or PIN).

Usage:

- Performance of actual money transfer (including service-related transfer of information on A and B side).

■ NOTE: The set-up phase may or may not include access to functions within the service. Typically, the information required for a money transfer consist of several items of information. These items can be collected on the A party side and sent in one block of data, or can be sent one after another. From a diagnostic point of view, these variants will have different appearance and relation to transport network properties. However, from an end-to-end related functional view the actual mode is not relevant.

4.1.4 Failure information in top-level views

Depending on DFS implementation, collection of information needed to perform a DFS money transfer may involve data transfers through the network. In the hierarchical phase model, such steps are described by respective sub-phases of the set-up phase.

While it is formally possible to define respective KPI from these sub-phases, this may not be the best choice. It would increase the number of KPI vastly. This may weaken the value of each KPI and obscure the function of a KPI as indicator of quality from the user's perspective. When benchmarking services, each contender can be the "test winner" in some category if there are enough KPI in the portfolio. In the end, this decreases transparency instead of creating it. Therefore, the set of KPI should be as small as possible, with each KPI carrying a strong meaning with a clear relation to user perception.

Moreover, a KPI is essentially an isolated quantity. A phase consists of individual steps or sub-phases which occur in a given sequential order. With KPI for each sub-phase, information about this sequential order is not visible anymore. Therefore, a single KPI describing the overall success (or failure) rate of that phase, plus detail information about unsuccessful cases is more useful. Such detail information would then consist of information at which step of the sequence the failures have occurred. If required, statistics on such causes can be created or further processed to KPI-like indicators, i.e. this way is still open if required. The advantage as compared to the primary use of KPI to convey this information is that the information about failure causes is preserved on the transaction level and can be used to create additional diagnostic insight.

In the set of DFS KPI, the Money Transfer completion rate is a very good example for this approach. With the abstract model described in Event and action flow, and the practical example shown in clause 4.2, this approach is demonstrated as follows.

The information required to perform a DFS transaction is prompted sequentially. After the user has entered a value, it is transmitted to the service, which in effect triggers the prompt for the next item of information. In order to make this happen, two data transfers are required. As seen from the A party's mobile device, this is:

- Sending an item of information, via the transport network, to the service, and
- Receiving the next item from the service.

As seen from the A Party device, there are two ways this sequence can be interrupted.

a) Sending an information item can fail, with a failure information; this can be a temporary failure when a retry takes place, or a permanent failure when e.g. the maximum number of retries or a time-out condition is reached.

b) The expected response may not occur. This essentially a matter of time-out condition. Without additional information, the A Party cannot determine if the request — the data sent to the service — or the response of the service has been lost.

If there is, in a particular implementation of a test or the DF service, no sending failure information on the A side, case a) cannot be technically distinguished and all interruptions appear to be of type b).

In any case, the A side has information of the last successful step, and the next one attempted. In case of failure, this information can be output together with the failure information and used in subsequent processing.

4.1.5 Time corrections for human interaction

If interactions require human input, time measurements will need adjustments. The top-level phase for set-up (see Top-level phases) consists, as shown in Event and action flow, of a series of prompts for information items, and respective input by the user. Therefore, a time measurement for the whole set-up phase will contain elements which depend on the user's typing speed which is clearly not useful for an objective measurement.

If time measurements are sufficiently fine-grained, it is possible to separate human interaction-related time spans from time spans caused by network or service response. For instance, if a prompt to enter data appears, the user needs some time to read the prompt, enter the requested information, and send it to the service. The service then responds with the next prompt until all required steps are made.

When the DFS event flow is monitored and recorded manually, the granularity of time measurements, and their accuracy, is limited. Therefore, it may be difficult to separate service response times. Time measurements for larger groups of activity — such as the entire set-up phase as shown in Figure 4-2 — will inevitably contain human-interaction times. It can be expected that after some initial training, the time to enter data will be quite constant from transaction to transaction. However, time measurements should be expected to be of limited accuracy.

It is plausible to assume that service response times in the set-up phase are of interest nevertheless. A possible way to create respective data — at least on an averaged basis — is to record a number of interactions by e.g. video and determine a typical "typing time".

For a practical example, see the extended table in clause 4.2, and the definitions provided there.

4.2 Trigger point IDs

4.2.1 Trigger point ID basics

A trigger point ID is a short-form notation describing a specific action or event. The difference between action and event is somewhat arbitrary and also depends on the point of observation (POO). For a POO on the A Party side of a use case implementation, action refers to an activity performed on the A side (by human action or some programmed activity) while event refers to something incoming (e.g. a message received via a mobile network).

- NOTE: In older literature, the term PCO was used (Point of Control and Observation). The newer term POO reflects the fact that in most cases, respective data comes from sources which do not allow control anyway (e.g. IP layer traces); also in general it is better to not intermix control and data layers.

Trigger Point ID (TPID) = <Service and use case code> _<Type>_<Index>

where

<Service and use case code>: in the present document, always DFSP2P

<Type> is either

- AE event observable on the A side.
- AA action to be performed by the user on the A side.
- BE event observable on the B side.
- BA (not used) Action to be performed on the B side.

<Index> is a continuous index, three digits, leading zeroes. Please note that numbering is not necessarily consecutive, i.e. choice of index does not carry meaning by itself.

For practical purposes in cases where the use case context is clearly defined, there is also a short-form TPID being used that omits the service and use case code and the related delimiter.

4.2.2 Trigger point IDs used

The following list of events has been derived from video analysis of an actual DFS P2P money transfer, for two variants:

a) App based. This category also includes browser based web applications. Typically such applications use https or other secure protocols..

b) USSD based (typically used on feature phone based).

The actual platform was, however, a smartphone in both cases.

For further reference see also Event and action flow.

Table 4-1 shows the trigger point ID for the MoMo P2P transaction model used.

With respect to the considerations discussed in Time corrections for human interaction, and Special considerations for manual testing and time-taking the table also contains color-coding describing the nature of the phase between respective trigger points.

The fields marked in blue identify parts of the event flow that relate to user activity. They are to be read in the following way: Beginning of the user activity is marked by the TPID preceding this element; the end of user activity is marked by the TPID assigned to the respective element.

TABLE 4-1: Trigger point IDs for the MoMo P2P model case[2]

TRIGER POINT ID	SHORT TPID	DESCRIPTION (APP)	DESCRIPTION (USSD)
DFS_P2P_AA_100	AA_100	Start DFS app	enter start USSD command
DFS_P2P_AE_104	AE_104	Prompt to select TA type	Prompt to select TA type
DFS_P2P_AA_108	AA_108	Select: Transfer	enter 1 to select “Transfer Money”
DFS_P2P_AE_112	AE_112	Prompt to select recipient type	Prompt to select recipient type
DFS_P2P_AA_116	AA_116	Select: To mobile user	enter 1 to select “to Mobile Money user”
DFS_P2P_AE_120	AE_120	N/A	Prompt to select category of recipient
DFS_P2P_AA_124	AA_124	N/A	enter 1 to select “to subscriber”
DFS_P2P_AE_128	AE_128	Prompt to select recipient ID	Prompt to select recipient ID
DFS_P2P_AA_132	AA_132	Enter B number and continue	Enter B number and continue
DFS_P2P_AE_136	AE_136	Prompt to select recipient ID again	Prompt to select recipient ID again
DFS_P2P_AA_140	AA_140	Enter B number again and continue	Enter B number again and continue
DFS_P2P_AE_144	AE_144	Prompt to enter amount	Prompt to enter amount
DFS_P2P_AA_148	AA_148	Enter amount and continue	Enter amount and continue
DFS_P2P_AE_152	AE_152	Prompt to enter reference	Prompt to enter reference
DFS_P2P_AA_156	AA_156	Enter reference and continue	Enter reference and continue
DFS_P2P_AE_160	AE_160	Request to confirm transaction appears	N/A
DFS_P2P_AA_164	AA_164	Confirm	N/A
DFS_P2P_AE_168	AE_168	Request for PIN appears	Request for PIN appears
DFS_P2P_AA_200	AA_200	Enter PIN and confirm	Enter PIN and confirm
DFS_P2P_AE_210	AE_210	Display TA in progress info	Display TA in progress info
DFS_P2P_AE_300	AE_300	Display payment confirmation	Display payment confirmation
DFS_P2P_AE_310	AE_310	Receive A side payment info	Receive A side payment info
DFS_P2P_BE_320	BE_320	Receive B side payment info	Receive B side payment info

Example: For the TPID AA_148 (Enter amount and continue), the user-activity starts with TPID AE-144 (Prompt to enter amount). At this point in time, the respective prompt appears on the user interface. The duration of this sub-phase is the time difference between these two trigger points, T(AA_144, AE_148) in the notation defined in Expressions. It includes the time the user needs to read and understand the prompt, to perform the action asked for (in this case, to type the amount, and to tap or press a button to confirm/send).

5 END TO END DFS KPIS

5.1 KPI abbreviations and reference

The Table 5-1 below is a quick-reference index between KPI abbreviations, basic types, and the respective KPI definitions.

The abbreviation is given for easier reference; it also provides a way to add the actual test case type description in a similar way as in other KPI definitions. For ease of reading—because the present document only deals with the P2P MoMo case—the core abbreviation is used.

Full abbreviation: DFS-<Test case type>-<KPI abbreviation>

Example: DFS-P2P MoMo-MTCR

All definitions are using the trigger point codes defined in clause 4.2.

5.2 Money Transfer Completion Rate (MTCR)

5.2.1 Functional description

Probability that a money transfer can be completed successfully.

5.2.2 Formal definition

MTCR = *ratio between the number of successful instances of the use case, and all valid attempts to perform the use case.*

With AA_100 as indicator for a valid attempt (successful activation of the DFS function) and AE_300 as success indicator, the expression becomes (see Expressions)

MTCR = R (AE_300, AA_100)

5.2.3 Specific definition

Using the primary success definition, i.e. the summarizing SMS are not considered.

5.3 Money Transfer Completion Time (MTCT)

5.3.1 Functional description

Time to complete a money transfer.

5.3.2 Formal definition

Using the primary success definition, i.e. the summarizing SMS are not considered.

This value is determined from the time between the activation of the used case until the completion of the transfer as indicated by the primary success indicator; it is therefore only valid for a successful transaction.

As the overall time contains human interaction, the technical definition excludes such times, but adds a typical time assumed to express the respective portion of the use case.

MTCT = T(AE_104, AE_300) – MTHI + TTHI

MTHI stands for the measured and TTHI for the (assumed) typical time for all human interaction in this use case.

The meaning of this expression is "take the measured overall duration of the transaction, eliminate times caused by actual human interaction (which can vary from instance to instance) and replace them by a generalized (typical) value).

The special case TTHI=0 stands for the ideal (practically unreachable) case where data is entered so fast that the duration becomes negligible.

5.3.3 Specific definition

MHTI can be expressed in terms of trigger point timestamps as follows:

MHTI = T(AE_t04, AA_108) – T(AE_112, AE_116) – AE(120, AA_124) – T(AE_128,AA,_132) – T(AE_136,AA_140) – T(AE_144,AA_148)-T(AE_152,AA_156)-T(AE_160,AA_164)

TABLE 5-1: KPI abbreviations and full names

ABBREVIATION	TYPE	REFERENCE
MTCR	Rate/Probability	Money Transfer completion rate
MTCT	Time	Money Transfer completion time
MTFPR	Rate/Probability	Money Transfer False Positive Rate
MTFNR	Rate/Probability	Money Transfer False Negative Rate
MTFTRR	Rate/Probability	Money Transfer Failed Transaction Resolution Rate
MTASSR	Rate/Probability	Money Transfer Account Stabilization Success Rate
MTAST	Time	Money Transfer Account Stabilization Time
MTLR	Rate/Probability	Money Transfer Loss Rate
MTDR	Rate/Probability	Money Transfer Duplication Rate

By reference to Trigger Point IDs used, the terms of this equation are the sub-phases related to entering required information elements for the DFS transaction.

If a specific DFS implementation does not use and request a specific item, respective events and actions are not present, and associated T(x,y) are likewise not valid and are not used in the computation.

5.4 Money Transfer False Positive Rate (MTFPR)

5.4.1 Functional description

Probability that a transaction is reported as successfully completed but has not actually been performed.

5.4.2 Formal definition

Using the event flow, receiving a primary or secondary success event without a corresponding attempt.

5.4.3 Specific definition

For further study. In order to determine the actual account balance, either secondary information (e.g. A/B side summary information SMS), or an evaluation of an account record could be used.

5.5 Money Transfer False Negative Rate (MTFNR)

5.5.1 Functional description

Probability that a money transfer is reported as unsuccessful but in fact has taken place (i.e. money has been transferred)

5.5.2 Formal definition

Reception of a negative response from the system (an event other than a primary success criterion) which ends the transaction. There are different variants of this type of result depending if secondary success criteria (SMS) or SMS with respective content are received or not (or if SMS content contradicts other system responses). To detect this type of outcome, two checks are possible:

- Account balance check on A side (as reported by SMS)
- Account balance check on B side (as reported by SMS)

It is also possible that account information by SMS is not given by a SMS related to the current TA but in a later SMS from a subsequent TA. This can be checked (optionally) in post processing.

5.5.3 Specific definition

For further study. In order to determine the actual account balance, either secondary information (e.g. A/B side summary information SMS), or an evaluation of an account record could be used.

5.6 Money Transfer Failed Transaction Resolution Rate (MTFTRR)

5.6.1 Functional description

Probability that a failed transaction (by time-out through inaction or loss of network coverage) leads to a correct account balance

- NOTE: This will be treated as out of context for the current project but should be subject of further study. Respective cases from the project can be used as input for failure assessment.

This is a secondary KPI which implies an error resolution process outside the scope of actual testing. It involves cases where initially money is lost (with respect to reported account balance), and where this lost money is retrieved though

a) An active process, e.g. by filing a claim to retrieve lost money, or
b) Some automated process in the realm of the DFS operator that restores lost money automatically.

This KPI is not subject to testing or measurement within the scope of the present pilot testing campaign.

5.6.2 Formal definition

Subject to further study.

5.6.3 Specific definition

Subject to further study.

5.7 Money Transfer Account Stabilization Success Rate (MTASSR)

5.7.1 Functional description

Probability that a DFS transfer leads to a consistent account on both sides when all information is considered (i.e. primary status information on the A side, and summary information on A and B-side.

For the current project it is assumed that the content of A and B side summary messages is correct. This KPI can then be computed as soon as both of these messages (e.g. SMS) have arrived.

5.7.2 Formal definition

Subject to further study. It needs to be defined how missing A or B side summary messages should be treated (e.g. ignoring them for KPI computation or not).

Furthermore, the computation for a consistently negative case needs to be defined (i.e. when a transaction fails, the expected result would be that the account balance is not changed). If this is not desired, respective definition of valid transactions is needed.

Preliminary definition:

$$\text{MTASSR [\%]} = 100 * \frac{\text{Number of transactions where information in summary messages is correct}}{\text{Total number of successful transactions (: AA\_200 valid, AE\_300 valid)}}$$

MTASSR = Ratio of transactions where the information is correct, to all valid and successful transactions (i.e. where AA_200 and AE_300 are valid).

5.7.3 Specific definition

Start/valid try when the MT is actually triggered, i.e. with last user confirmation. End after both the A and B side summary SMS (or equivalent data elements of a particular DFS implementation) have been received. Evaluation is made based on content of these elements.

See also the considerations in Event and action flow.

If the actual DFS implementation does not provide respective information, this KPI cannot be computed.

5.8 Money Transfer Account Stabilization Time (MTAST)

5.8.1 Functional description

Time (after the DFS money transfer has been triggered) until all status and account information is correct and consistent.

Start event: when the MT is actually triggered, i.e. with last user confirmation

With reference to Money Transfer Account Stabilization Success Rate (MTASSR), the stop time is the time when the last of the A and B side summary messages e.g.by SMS have been received.

For the current project, it is assumed that the content of these messages is correct.

- NOTE: In order to validate the content of confirmation SMS against primary account reports may be a subject of further study.

5.8.2 Formal definition

$$MTAST = \max(T(AA\_200, AE\_310), T(AA\_200, BE\_320))$$

This definition takes into account that the A and B side confirmations (e.g. by SMS) do not necessarily have a fixed order.

5.8.3 Specific definition

Start time taken when the MT is actually triggered, i.e. with last user confirmation

5.9 Money Transfer Loss Rate (MTLR)

5.9.1 Functional description

Probability that a money transfer ends in a loss, i.e. money is deducted on the A side but not credited on the B side.

For the current project it is assumed that the content of A and B side summary messages is correct. This KPI can then be computed as soon as both of these messages (e.g. SMS) have arrived.

5.9.2 Formal definition

Computation of this KPI needs further study to determine how unsuccessful transfers should be treated.

Preliminary definition:

$$MTLR\ [\%] = 100 * \frac{\text{Number of transactions where money is deducted on the A side but not credited on the B side}}{\text{Total number of successful transactions}}$$

5.9.3 Specific definition

This KPI requires a timeout which determines the time after it is assumed unlikely that the money sent by the A party will appear in the B party account. The timeout value should be determined based on the specific implementation of the Service under test (see also Annex A for respective considerations).

5.10 Money Transfer Duplication Rate (MTDR)

5.10.1 Functional description

Probability that a money transfer is credited to the B side but is not deducted from the A side account.

For the current project it is assumed that the content of A and B side summary messages is correct. This KPI can then be computed as soon as both of these messages (e.g. SMS) have arrived.

5.10.2 Formal definition

Computation of this KPI needs further study to determine how unsuccessful transfers should be treated.

Preliminary definition

Number of transactions where money is credited on the B side but not deducted on the A side)/(Total number of successful transactions):

$$MTDR\ [\%] = 100 * \frac{\text{Number of transactions where money is credited on the B side but not deducted on the B side}}{\text{Total number of successful transactions}}$$

5.10.3 Specific definition

There are two possible cases to differentiate:

a) TA is reported as unsuccessful, but money actually appears on B side (but is not debited to A side; the other case is treated in the MT false Negative Rate).

b) TA is reported as successful, money is credited to B but not debited from A.

6 ACQUISITION OF DATA ON DFS TRANSACTIONS

6.1 Overview

In order to compute DFS KPI, respective input data need to be collected.

The method used should be robust and provide a high level of data quality. Robustness means that the system should ensure security against loss of data. Data quality refers to aspects such as reproducibility and plausibility tests to detect wrong data.

Figure 6-1 is a graphical representation of measurement data flow and handling. Please note that this is a rather schematic and simplified view. Details given in the following sub clauses have precedence.

In the present methodology a manual method will be used to collect the primary information, i.e. timestamp data for events needed to compute KPI will be entered manually by a member of the measurement team.

There is in addition, secondary information in the form of summary SMS sent by the system at the end of the transaction. These SMS will be read from the devices in a bulk fashion, and also transmitted to the data processing system.

For primary data collection on DFS transactions, there are two possible approaches:

a) Collection on paper and subsequent transfer into electronic forms (e.g. Excel®).

b) Direct entry to electronic forms (e.g. Excel® tables).

Both methods have their respective merits and will therefore be described subsequently. See also (Description of the Ghana Pilot Campaign) where respective considerations have been used.

6.2 Primary DFS data collection modes

6.2.1 General remarks

The procedures in the following are defined to provide operational robustness. They include steps which are intended to provide some redundancy and elements of data backup.

The term "uploading" is used in a functional way. Where smartphones are the platform (e.g. when taking a photo of a completed data log), it is assumed, unless otherwise mentioned, that this means sending respective data by e-mail.

As far as PCs are the platform, it is assumed that FTP or http upload will be used. It is further assumed that for this upload, the ITU IFA server will be used.

6.2.2 Collection on paper, later transfer

Paper printouts of respective tables are created. These printouts are called 'data capture sheets' (DCS) from here on.

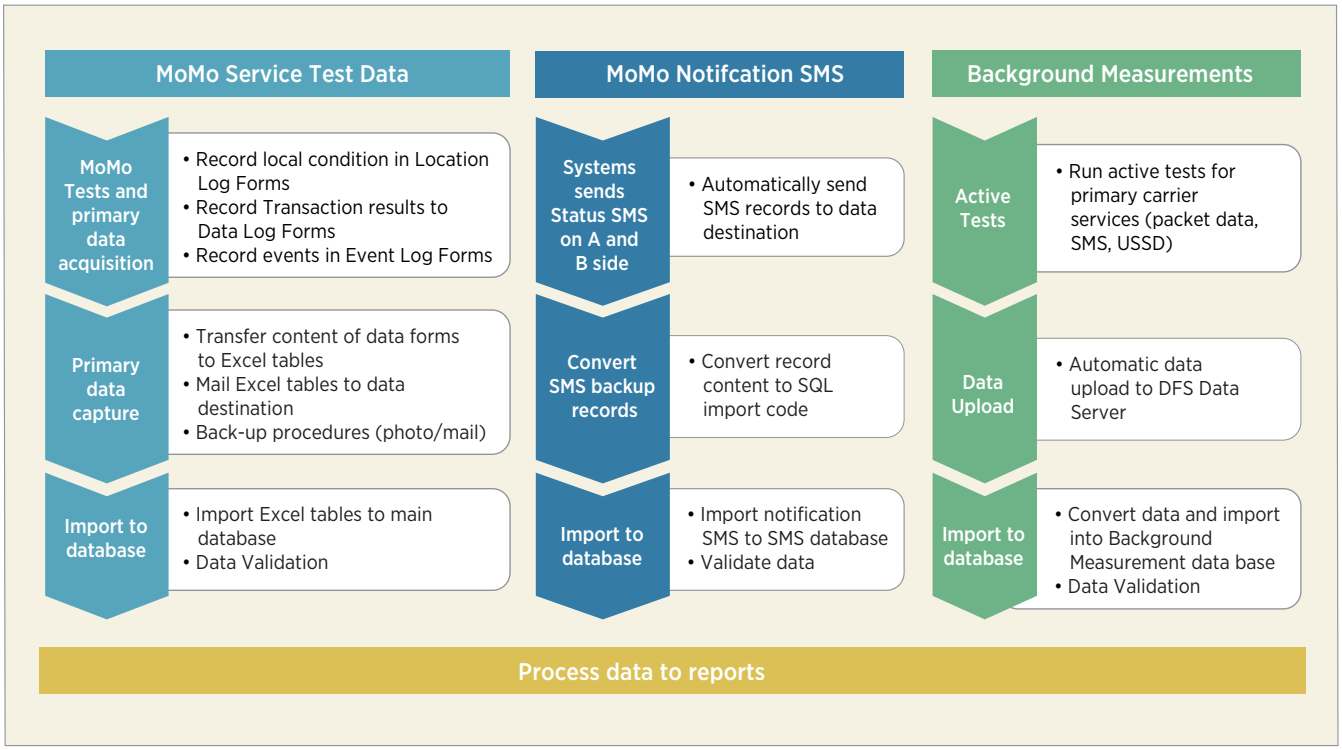
Each DCS shall carry some information to allow data consistency and completeness checking:

- Identification of the team.
- Date.
- Location of test.
- Running number of test in this specific location.

When a new location is used, a new DCS is used.

During testing, the team member enters data manually into the DCS.

FIGURE 6-1: Schematic overview of measurement data flow and handling



The exact means of time-taking are not prescribed provided the required time resolution is given. However, the overall procedure must make sure that time and date settings are correct.

When a DCS is completed (all rows filled in), it is photographed and uploaded. Each such upload is logged in the general event log. Likewise, if the location of the test is changed, and at the end of a measurement day, the last DCS used is photographed and uploaded.

After the end of a measurement day, the data sheets of that day are entered into an electronic file (e.g. Excel spreadsheet) by a member of the team.

For the name of data files, see Data file naming.

The data file is then uploaded. In addition, a copy of the file is made to a suitable data medium (CD or USB stick, to be kept in a safe place). The file is also kept on the PC.

If an upload is not possible (if no connectivity for upload is available), attempts to upload the file shall be repeated in a reasonable time pattern, at latest at the following day.

All DCS originals are collected and kept in a safe place.

6.2.3 Direct entry into electronic form

During testing, the team member allocated for this task enters data directly into a data file. Respective procedures are the same as described in the previous clause.

Upload attempts for data files shall be made on the following occasions:

a) The team changes the location, and

b) At the end of a measurement day, and .

c) When a time of 4 hours after the last upload has expired.

6.3 Data file naming

6.3.1 General file naming

These generic file naming rules apply to files not specifically listed in sub clause 6.3.2.

Each electronic document (data table) is named in a consistent and unique way.

This information is also duplicated in the document itself. The information shall contain:

- A common text identifier (to be defined).
- Identifier of the team.
- Date and time of creation (time resolution: minutes, e.g. hh:mm).

The following table contains file/content types used, and their respective file naming rules.

6.3.2 Specific file names

The naming of electronic log files is tentative and users of this specification are encouraged to reasonably adapt naming conventions to local circumstances.

■ NOTE: Data, location and event log files may contain information for different locations and therefore have no location name in the file name. Instead, they carry hhmm in case there are multiple files per day.

6.4 Campaign logs

Each team maintains a campaign log (paper or electronic form) where all relevant events are logged with date/time. Such events are:

a) Entering and leaving a given location.

b) Start, end and possible interrupts of background measurements.

c) Start and end of test activities.

d) Data logging and transfer related activities (depending on the mode selected).

e) Unusual events which occurred during measurement (e.g. power outages, planned or unplanned pauses).

The forms being used should at least include the following:

TABLE 6-1: Specific files, naming conventions

FILE TYPE	NAMING DEFINITION
Scanned/photographed log files (per location)	TeamName_YYMMDD_LocationName.pdf Example: Team2_180618_Bubuashie.pdf YYMMDD should indicate the day to which the log file set refers (this implies that each file should only contain log files for one and the same day)
Electronic version of Data Log	DataLog_TeamName_YYMMDD_hhmm.xlsx ***YYMMDD*** should indicate the date of entries (implying that each log file should only contain data for one day). ***hhmm*** should indicate the earliest timestamp of content. With respect to the paper versions, this would be the "sheet started" time. If no paper version is used, the time should be the time of the first item of content.
Electronic version of Location Log	LocationLog_TeamName_YYMMDD_hhmm.xlsx For YYMMDD and hhmm, see above.
Electronic version of Event Log	EventLog_TeamName_YYMMDD_hhmm.xlsx For YYMMDD and hhmm, see above.

- Location Log Sheet: Initial, intermediate and final checks on device set-up and status.
- Data Log Sheet (P2P Transfer): Acquisition of results for service tests.
- Event Log Sheet: capture of unusual conditions or events during tests.

An example of an actual campaign log used in the pilot campaign done in Ghana is shown in Appendix I

6.5 Handling of confirmation/information SMS (secondary information)

This data shall be retrieved from the device at least once per day, and transmitted/uploaded to a target destination (typically by email).

For retrieval, several tools are available. For the pilot project, the app "SMS Backup & Restore" has been selected. The app copies local SMS to a XML file. This file can be stored locally and be sent to a remote destination via e-mail.

The following sub clause describes the model set-up of the app.

After the data has been successfully uploaded, it can be deleted on the device. The data file just uploaded can be moved to a backup storage location. Until then, the data shall be kept on the device as a back-up copy.

■ NOTE: If the devices are restricted in functionality (e.g. to act as "Feature phones", transfer via e-mail requires that these restrictions (e.g. "no mobile data") are removed for the transfer. It is important to re-establish the correct settings for DFS testing afterwards, or prior to a new set of tests.

7 SPECIAL CONSIDERATIONS FOR MANUALLY OPERATED TESTING AND TIME-TAKING

The considerations described so far assume that time-taking provides a precision of time measurement which is sufficiently higher than typical times for respective phases of the event flow.

In case of fully automated data acquisition, typical time resolution is 1msec while typical phase durations are at least a couple of 100msec or longer.

The other extreme is entirely manual time-taking where time resolutions are much longer, typically 1 s or even more considering that times have to be read from a display which by itself may contain additional delay. Even in the case of semi-automatic data acquisition where some kind of stopwatch with high resolution is used, human reaction time and its jitter will result in an effective time resolution in the order of some 100 msec.

This means that a fine-grained time recording as indicated by Figure 4-2 will not be possible and use case modelling will have to be restricted to main phases. From a practical point of view, this will be the overall transaction time from invocation of the MoMo service to its completion (end to end duration), and the core transaction time, i.e. the time between triggering the transfer after all input information has been provided, and its completion.

Data acquisition may deliberately be done fully manual, or points of observation to obtain tripper point events may be limited. . In such a situation, some of the generic KPI, as described in clause 5, are not applicable due to the reasons described. The following set of practical KPI can be used:

In all cases, a valid sample requires that all trigger points used in computation are valid, i.e. present. Indicators of type 'time' are therefore computed from transactions where respective phases have been completed successfully.

For the overall completion times, the E2E version using T1 was selected although it includes times for manual activity. Reasoning is as follows: A KPI, as an indicator expressing the end-user perspective, should provide a realistic estimate of a service's behavior. Manual activity is an integral part of service usage and therefore it makes sense to include respective times into an indicator. Assuming that a testing team can be compared to an experienced user, times taken by such a team can be viewed as a valid estimation of manual components of service usage.

TABLE 7-1: Timestamps (trigger points) used for the practical KPI

SYMBOL	DESCRIPTION
T1	Start of transaction (activation of the DFS function/application on the device)
T2	All necessary input data has been entered and the actual money transfer is triggered.
One of T3	Reception of the primary success s criterion (information about the successful completion of the transaction), or
T4	Reception of an information stating that the transaction has failed
T5	Time-out limit reached without a positive or negative reaction from the service
T6	Reception of the summary SMS in the A-side mobile device
T7	Reception of the summary SMS in the B-side mobile device

TABLE 7-2: Simplified set of DFS KPI

INDICATOR	ABBREVIATION	COMPUTATION	REFERENCE TO FORMAL KPI
Money Transfer Core Duration	MTCD	T3-T2	New KPI
Money Transfer Raw Completion Time	MTRCT	T3-T1	MTCT
Money Transfer completion rate	MTCR	T1 present, T3 present: success	MTCR
Money Transfer Full Completion Time	MTFCT	T7-T1	New KPI
Money Transfer A-side Completion Time	MTACT	T6-T1	New KPI

8 MEASUREMENTS IN THE BACKGROUND

8.1 Overview and basic assumptions

The performance of Digital Financial Services over mobile networks is related to the properties of the network over which these services are provided.

It is important to keep in mind that the actual DFS is usually provided by some distinct ecosystem or functionality domain. A good mobile network alone does not guarantee a well-performing DFS as other components of such services also need to function well. A poorly performing mobile network can, however, degrade DFS performance massively.

Table 8 1 shows the categorization of relative impacts of mobile network and DFS infrastructure performance, and conclusions with respect to field testing of DFS. With a poorly performing DFS functionality, effects of mobile networks are not or only weakly visible. In that case, field tests in different locations will most probably not be efficient, as the same results could be obtained by testing in fixed locations. If, on the other hand, if it is possible to ensure that mobile network performance is high, no field tests are required either. In the remaining cases, field tests will be needed to get a correct picture of overall DFS performance and QoE.

One of the goals of the methodology described in the present document is to provide guidance to regulators with respect to service performance levels of mobile networks in order to secure well-working digital financial services. While the present document describes KPI to express DFS QoE, it is desirable to provide insights about the connection between basic transport network QoS and their relation to DFS quality. Basic service KPI can then be used as proxies to create assessments on expected DFS quality. The methodology therefore also provides ways to link these KPI.

DFS can be implemented in various ways. Many implementations are based on the SIM Application Toolkit (STK) and access transport network services through functions provided by the STK.

With unmodified mobile devices it is not possible to access such services through STK, but this is considered to be not essential as these services can be accessed directly.

- REMARK: STK offers encryption of traffic which is not an intrinsic property of the generic services such as SMS or USSD. In the current context, this is considered to make no difference. Encryption may lead to additional delay and/or increased size of data content. It can however be assumed that this will not qualitatively affect the sensitivity to factors impairing service quality.

Using basic service as proxies to create assessments on expected DFS performance, and to provide guidance for e.g. regulators to set meaningful targets for network performance, has potential benefits; it is however also important to understand the limitations. A benefit is that the measurement of basic network services is technically easier than a full end.to-end measurement of DFS, not the least because the actual transfer of money is involved. It should be kept in mind, however, that the full DFS ecosphere also includes actors or parties beyond the mobile network infrastructure.

The figure below shows a generic model of the elements involved in the interaction between the A party (left side) and DFS system it uses.

Each component has a certain influence on the overall result, i.e. on the QoE of the DFS as perceived by the user of the service. If mobile network performance is the dominating element, there will be a distinctive correlation between the KPI of transport services used by the DFS implementation, and these service KPI can be seen as good proxies for actual DFS performance. If other elements dominate, e.g. the infrastructure which handles the money transfer or elements between the mobile network and this infrastructure, respective correlation will be weak and transport service KPI will not be good proxies for DFS performance assessment or formulation of target value corridors.

The goal of the field trial was therefore to run a wide spectrum of basic services tests. This allows to evaluate the correlation between DFS and transport service KPI and therefore identify the most useful proxies for DFS quality assessment.

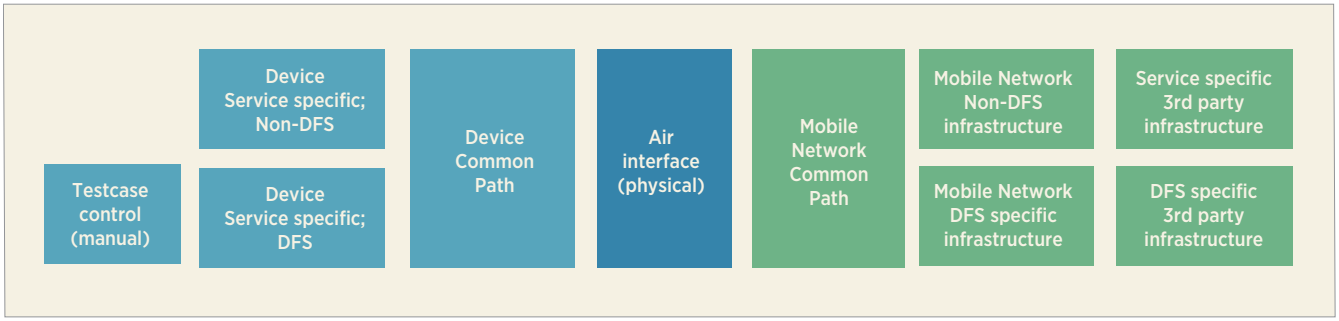
8.2 Acquired data

During execution of the DFS use cases, the transport network is tested actively in parallel with a repeated se-

TABLE 8-1: 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

FIGURE 8-1: Generic model of the elements involved in the interaction between the A party and the DFS system



quence of test cases for different services. The purpose is to evaluate the general condition of the network. The intensity of these tests is however moderate in order to not stress the network too much.

Also, some basic network parameters as well as GPS information are taken continuously. However, the extent of these passive data is limited. On purpose, in this methodology only unmodified ("out of the box") mobile devices are used.

The following parameters are recorded:

- Signal strength.
- Type of network (Radio network technology, RAT).
- Cell identity (as far as the device supports this).
- GPS position and speed.

If more information is desired, modifications to the phones are unavoidable. Such an extension of the methodology is for further study.

In the following clauses, considerations about the design of this sequence and the accompanying methodological considerations are described.

8.3 Test cases for transport network background testing

Scenarios for testing the transport network in the background have to be selected and defined on a country-by-country basis.

As an example, the following test cases can be used:

- SMS.
- USSD.
- Web browsing (to a live and a reference page).
- http download and upload.

These test cases—with respective guard times and additional pauses to achieve a desired density of tests—are repeated cyclically.

Most of these use cases have parameters such as the amount of data transferred. Choice of parameters is made in a way to avoid overloading the transport network. This relates to pauses between test cases as well as use case specific parameters, e.g. data volume transferred in upload or download, and selection of the web sites used for testing.

8.4 Monitoring

Some baseline data should be collected for assessment of packet data performance. It is recommended to also run a monitoring device under good radio conditions (or via Wi-Fi connected to a fixed-line connection) which accesses the same server (for UL/DL) or web site respectively.

By analyzing the performance, times where the server or web site itself is down (or its performance is degraded) can be easily identified.

9 DATA VALIDATION AND PROCESSING

9.1 Plausibility and validity checks

The tables in the following subclauses are meant to be checklist templates, e.g. validated items would receive respective check marks.

9.1.1 Tests on DFS data

❏ Are backup records (photos of filled-in sheets) complete?

❏ Check time spans for electronic data (Excel tables from primary data) vs. backup copies (range checks, i.e. first and last transaction on each data log sheet).

❏ Check timestamps of DFS data against respective location logs

Does the timestamp range match the time window recorded for that location?

❏ Check timestamps of background measurement data against respective location logs

Does the timestamp range match the time window recorded for that location?

❏ Decide on necessity to exclude time ranges.

Does the location log indicate special events and conditions, which set the need to exclude data from the set?

❏ Visualize timestamps of transactions: Are there any gaps or unusually dense transactions during a period of time? If yes, validate reasons.

❏ *(further check items to be added)*

9.1.2 Tests on background test data

❏ If GPS data are available, does the location indicated, and the GPS location match?

❏ Visualize timestamps of transactions: Are there any gaps or unusually dense transactions during a period of time? If yes, validate reasons.

❏ *(further check items to be added)*

9.1.3 Cross tests between data (after import)

❏ Validate time stamps of DFS and background data for consistency.

❏ Validate consistency between network unavailability in DFS and background data. A possible consistency problem exists if background data indicate network unavailability but DFS transactions work during a given timespan. If such periods of time exist, mark them in the database and seek further clarification.

❏ *(further check items to be added)*

9.2 Additional processing

With respect to some KPI definitions, additional check procedures may be done.

Examples are:

1. Check consistency of accounts throughout a sequence of information SMS.
2. Check for "false negatives" (ref. Money Transfer False Negative Rate MTFNR) by comparing account balance against transaction results.

10 LESSONS LEARNED

10.1 Overview

The manual capture of DFS TA has shown to be a major weak point in the campaign.

This was not entirely unexpected; manual mode was deliberately chosen to provide the widest possible angle of view and maximum transparency of the data acquisition process.

The weaknesses of manual operation are manifest in some main fields:

Time-taking of the transaction, with typical time scales in the order of a few seconds, introduces quantization errors at best where times have to be read from displays with a typical resolution of one second. Other effects come on top, such as potential errors due to time offsets when using different devices. Time-synchronization of such devices can only provide limited protection as this compensation is itself quantized to one second steps (unless modified devices are being used).

Transferring readings to paper logs open up additional sources of error due to handwriting.

Further transferring of paper logs to electronic means—which is a prerequisite of data processing—is again prone to reading errors.

All process steps are essentially dull and repetitive, and are therefore vulnerable to human errors.
Typical error patterns are:

a) Reading from the display: subsequent timestamps 10:49:58 10:50:02 logged as 10:49:02

b) Transferring from paper logs: number switches such as 1<>5, 2<->3, 2<->5, 3<->5, 1<-7, 4<->9 depending on handwriting.

c) Transferring from paper logs: eye-jumping to the line above or below the actual one

d) Transferring from paper logs: Number-switching, e.g. 12:30:14 ->13:20:14

Transferring from paper logs: simple typing errors.

Of course, it is possible to extend the manual established data quality assurance procedures, which can prevent or eliminate errors. However, this is a significant cost driver and therefore needs to be considered against automation or partially automation of the data acquisition process. Some suggestions how this could be done are given subsequently.

10.2 Recommended measures

Fully automated DFS transactions would eliminate all of the above mentioned sources of error. If budgetary conditions allow, this would be the method of choice. It should however kept in mind that careful—and periodically repeated — validation of automated solutions is part of the design of such a solution and therefore needs to be considered in a cost assessment.

The next best solution — in case budgetary or other considerations lead to the decision to not use full automation—is tool-assisted time-taking. A respective tool would have the following basic functionality:

- Android app with a simple and user-friendly user interface, e.g. showing a group of buttons with one button per timer flag.

- The app should have a built-in time synchronization with network or GPS time (which may need modification of the phone, i.e. "rooting" to get required access rights), or at least captures GPS (combined with procedures to make sure that there is at least a minimum level of GPS data capture) in order to have a high-precision time source on board.
- Captured time stamps, along with information on the measurement team and other respective data, is stored locally as well as automatically updated to a server location.

If for any reason a solution involving paper logs and manual transfer has to be used, the following improvements are recommended:

- Use Excel® templates which contain a set of built-in initial checks and create visual warnings, e.g. if timestamps are inconsistent (based on expected ranges or relations between entries.
- If budget allows, prescribe a four-eyes method for data transfer.
- Further improve paper logs by visual elements which reduce the risk of "eye slips".

11 CONCLUSIONS AND WAY FORWARD

The methodology described in the present document provides all means to conduct and evaluate QoS measurements on Mobile Money services. The current focus is on person-to-person money transfers, but the overall framework has been designed with extension to other use cases in mind.

With respect to the actual execution of tests, all-manual testing and data acquisition has been deliberately chosen to provide maximum transparency on the procedures, despite restrictions in accuracy. As expected, the manual processes exposed various ways data can be compromised, in particular where information is transferred between different media. Respective consistency checking procedures have been designed and tested, and a broad range of experience and ways to handle such errors has been created.

From the robustness and data quality point of view, automated systems are encouraged for testing, similar to common practice in most field quality assessments. Where this cannot be done for practical reasons (i.e. budget restrictions), technically supported time-taking should be used. A practical way would be to have a multi-step time-taking tool with automatic upload of acquired data.

The result of this pilot testing and evaluation is a set of procedural insights and recommendations as well as guidance useful for design and performance of future testing. The basic expectations, with respect to QoE assessment of MoMo services and correlation with carrier network performance, have been field-tested and found to be valid. These results provide a functional methodology as well as a clear path to further extension and refinement.

The existing use case P2P needs to be extended, e.g. for the following topics:

- Areas with non-optimal radio coverage.
- Mobility aspects.
- Better statistical relevance of the data base.

Additional use cases should be elaborated (besides P2P). However, care has to be taken to correctly isolate the payment process from the application in which it may be embedded. Also with any new use case taken into consideration, it has to be analyzed which events or trigger points are accessible. This again may vary depending on who is planning to conduct the testing. Possible ways can be to do a "friendly" testing with all stakeholders involved may enable access to internal trigger points; it could also be testing by a third party (e.g. the regulator) which however may turn out to be significantly more difficult.

An important type of use case is G2P, i.e. payments of governmental bodies to induvial. From a testing perspective, this would also provide a good basis as real money flows, which are a necessity in such kind of tests, can rather easily be controlled in order to create a mostly circular type of transfer with respectively moderate need with respect to used capital.

Studies are underway which seem to indicate that users under certain circumstances prefer dedicated hardware solutions over an app on the smartphone. With the advent of Internet-of-Things (IoT) and Low-Power-Networks (LPN) being rolled-out a new class of DFS solutions may appear on the market which are using dedicated hardware in the context of LPN enabled IoT devices. Dedicated hardware DFS solutions would have the potential to reduce human errors on the users' side of DFS. The methodology laid out in the present report, while in principle sufficiently wide in scope, will require a thorough review in order to explicitly include this class of solutions.

ANNEX A

One-time tests

This clause deals with tests which should be performed once per campaign to determine basic properties of the DF service under test.

A.1 DETERMINE TIME-OUTS

Determine the time-outs for each step of a DFS use case (e.g. entering destination ID, amount, and reference). Make sure the time-outs do not cause failures with typical typing speed/time for entering values. Consider also typical reading times for information presented by the service, e.g. prompt texts.

ANNEX B

Check lists to be used in testing campaigns

This clause contains elements of check lists for usage in measurement campaigns. The lists describe the points to be checked; the way to do so will have to be defined case by case.

Figure B-1 illustrates the use of the check list during a particular day of a test campaign.

B.1 DAILY, PRIOR TO BEGINNING OF TESTS

- Make sure the time-taking device has correct time and date settings.
- Make sure the device is set up to use network date/time (in case the network is providing this feature and information is assessed to be reliable).
- Make sure the devices have sufficient airtime/data volume credit to perform their respective actions (e.g. sufficient prepaid credit, or remaining data volume). Query and record respective information.

B.2 AT EACH NEW TESTING LOCATION

B.3 DAILY, AFTER COMPLETION OF TESTS

- Make sure the device is set up to use network date/time (in case the network is providing this feature and information is assessed to be reliable).
- Check that airtime/data volume credit is sufficient to perform their respective actions (e.g. sufficient prepaid credit, or remaining data volume). Query and record respective information. Re-charge if necessary.

Remark: When respective action should be taken will depend on the actual testing situation (i.e. if it is better to do it in the evening for the following day, or in the morning of the next day). Choice should be made to give the best overall test team productivity under given circumstances.

FIGURE B-1: Measurement related checking procedures

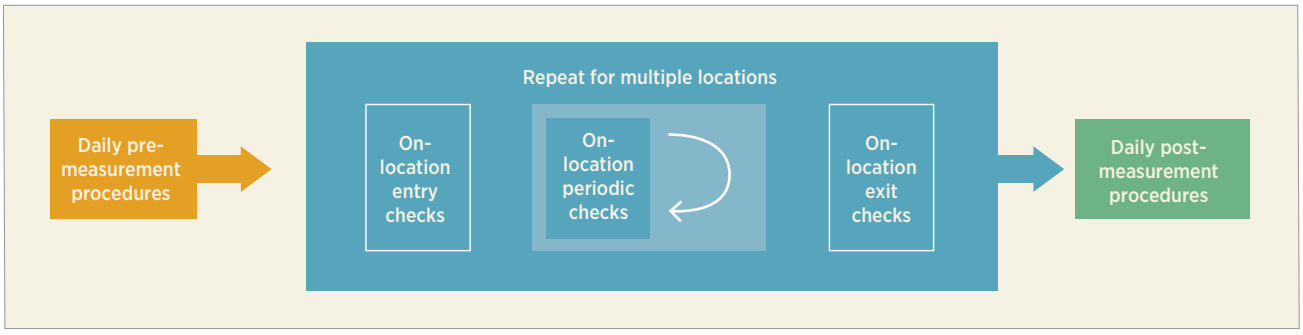


TABLE B-1: Checking actions to be taken at each new testing location

ACTION	FREQUENCY
Make sure the ObsTool UE is in the same cell as the DFS UE	Initially and periodically every ~ 2 hours
Make sure the UE used have sufficient battery charging level	Initially and periodically every ~ 2 hours
Make sure UE used for DFS testing do not run extensive background activities (e.g. download of new OS versions or apps requiring substantial system resources	Initially and periodically every ~2 hours

ANNEX C

Device set-up for the Ghana pilot

C.1 GENERAL

Figure C 1 shows the device set-up schematically. Please note that this diagram is shown for convenience and overview. Explicit textual descriptions have precedence.

All settings and selections made during the set-up process shall be recorded and stored electronically (e.g. in an Excel® table file) to facilitate overview and reproduction in case of need.

C.2 BASIC DEVICE SET-UP

All devices are set-up following the usual procedure for Android.

In particular, the Google user account and associated mail address shall be recorded to be able to identify mails sent from this device, and facilitate emergency remote access to this device over respective Google services.

Set-up of optional services and features for the MoMo test phones shall be made in a way assuming a typical user (i.e. accepting default settings suggested by the set-up process).

In case of the observer phone, set-up shall be made in a way which results in minimally possible background data traffic.

All devices shall be set up to use network date/time to ensure time stamp consistency. This set-up shall also be verified periodically, at least once per day at the beginning of measurements.

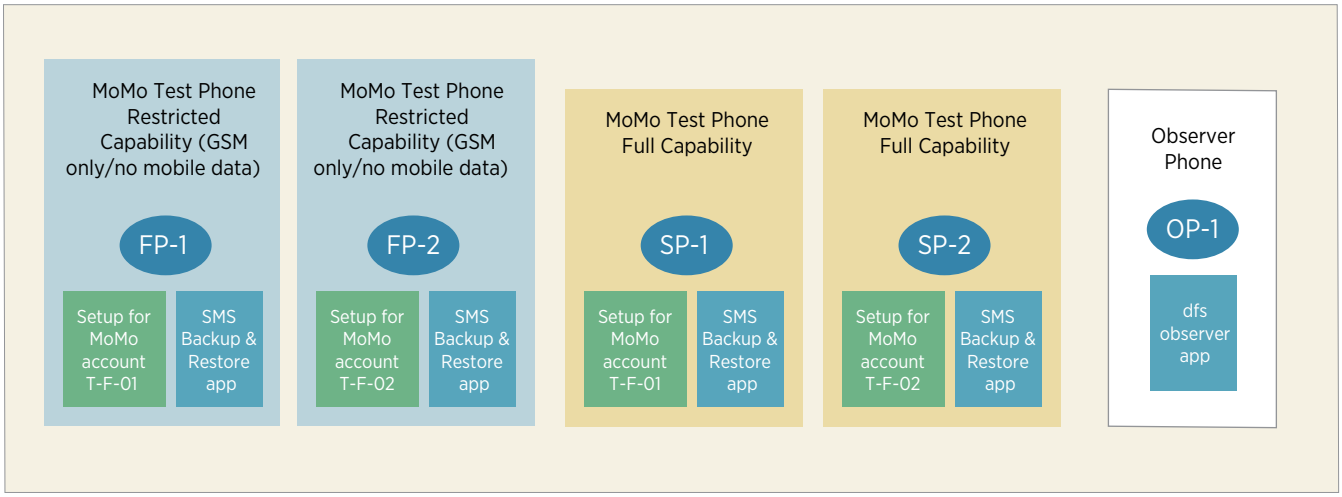
C.3 SETUP FOR MOMO ACCOUNT

The accounts on each MoMo test phones shall be set up in the way deemed typical for a subscriber of the respective service.

C.4 SMS BACKUP & RESTORE APP

The app shall be installed and parameterized as described in detail in clause Installation and set-up of the SMS retrieval tool.

FIGURE C-1. Device set-up for the Ghana pilot (per team)



C.5 DFS OBSERVER APP

C.5.1 General

This app (installation name: dfs_observer) shall be installed using the standard procedure for Android apps, and, eventually, with additional support from Focus Infocom (detailed instructions given in a separate document).

Please note that the SMS test case needs customization of the scenario for each individual device to use the correct destination phone number. Respective instructions will eventually be given by Focus Infocom Support.

C.5.2 Scenario used for the pilot

The scenario used combines various data tests, an SMS test, and two different USSD tests (order of testing may differ)

- Google start page.
- ETSI Kepler SP reference page on 2 different servers (fixed-time mode).
- ETSI Kepler Full reference page in fixed-time mode, on 2 different servers.
- Download 100 kBytes, fixed-time mode, on 2 different servers.
- Upload 100 kByte, fixed-time mode.
- SMS to self.
- USSD: *156# (show own number).
- USSD: *151# (unknown code, see remark below).

■ NOTE: Deliberately using an invalid USSD code is a means to get a kind of “ping” to the USSD subsystem. There is, however, the risk that the network negatively reacts to repeated sending of invalid codes after some time. The data shall be monitored in order to detect indications of such reactions and the scenario may be changed respectively.

C.6 ADDITIONAL SOFTWARE

In order to make remote support easier, it is recommended to install the TeamViewer app (or other remote support apps) on each device.

■ NOTE: It is assumed that the terms of use for this app allow the intended usage. Respective terms need to be monitored and checked against the mode of usage. In case of conflicts, respective resolution by e.g. purchase of required license or selection of another app needs to be considered.

ANNEX D

KPI basics

D.1 OVERVIEW

The KPI defined in this clause have been introduced in the Focus Group Technical Report QoS and QoE aspects of Digital Financial Services (05/2016) by the ITU-T Focus Group Digital Financial Services.

D.2 TERMINOLOGY

The definitions for computation of KPI are based on the formal event codes defined in clause 5.

The following is, for the purpose of easy reading, a condensed description of the model developed and explained in detail in ETSI TS 102 250-2[3] (or Rec. ITU-T E.804[4]).

A transaction is defined to be a single instance of a particular use case. Each transaction produces a sequence of events. From these events, the properties for the respective transaction are derived. The minimum set of information includes the result of the transaction (in the simplest case, success or failure; other result types can also occur).

Typically in case of success, the elapsed time (duration of the transaction) can also be computed; Depending on the nature and structure of a transaction, such computations may be complex and involve a multitude of events.

A transaction can consist of several phases. Typically, a transaction has a layered or hierarchical structure, with a single-phase description at the top representing the 'end to end' view.

Each such layer of description is called a view on the transaction, with the degree of detail, i.e. the number of phases, usually increasing with lower-level views. In a lower-level view, phases may be partitions of respective higher-level phases, or phase boundaries may have a different phase composition.

In the QoS context, a phase is usually related to a part of an overall transaction which has meaning in a user-perception way and is connected to events which are observable by a typical user. In a more diagnostics-related context, however, phases can also be related to protocol messages or other events which are not typically user-visible.

Each phase is marked by exactly two events. The starting event (E1) stands for an attempt to reach the functional goal associated with this phase.

A successful completion of a phase is indicated by an event E2 (positive success). If the phase has not been completed successfully, this is either indicated by an explicit event, or by reaching a given time-out condition. Every phase must have such a time-out condition as parameter. This ensures that no undefined state occurs,

- NOTE 1: Without such a time-out condition, an explicit list of events indicating failure would be required. In case the system under test yields a response previously unknown or not caught by a definition, the test case state would be undefined. With the requirement of a positive success indicator and a catch-all time-out condition, this case is prevented.
- NOTE 2: Instead of an explicit phase-wise time-out condition, there may be a 'global' time-out, with the clock started at the very beginning of the transaction.

Each event is associated with a time stamp. Therefore, an event has two basic properties: Its presence (i.e. if a particular event is present within a given transaction), and, if it is present, its time stamp.

- NOTE: Events are understood to be logical events, which are assumed to be unique by definition. These logical events are set by actual technical input, such as occurrence of particular protocol messages or data items at some point of observation. Such technical events are not necessarily unique. In such cases, a state logic is assumed to exist which produces respective logical events from technical input.

Example: Assume two events E1 and E2, where E1 stands for an attempt to achieve a given functional goal ('try'), and E2 indicates that this goal has been reached ('success'):

D.3 EXPRESSIONS

For KPI computation, the following expressions are defined. For actual KPI definition, respective Ex are replaced by their trigger point ID's (see also Event Codes).

TABLE D-1: Expressions used in definitions of KPI computation

N(EI)	NUMBER OF EVENTS WITH INDEX I.
T(Ei, Ej)	Time elapsed between the events with indices I and j. This quantity applies to one particular transaction and is only valid if both events are present. A KPI of 'time' type is the average (usually arithmetic mean) of respective transaction-wise T values.
R(Ei, Ej)	Rate (percentage) of events with index I with respect to events j. Typically, Event index j represents a "try" and index I the related success indicator for a given phase. In that case, R stands for the success rate of the given phase. This expression is only valid if N(Ej) > 0 which means that for a valid rate indicator there must be at least one 'try' occurrence of the respective phase. Technically, the condition N(Ei) <= N(Ej) is also met. This is however a technical cross-checking condition which is assumed to be fulfilled always if the underlying measurement and processing mechanism is properly defined and functioning.

D.4 UNDERSTANDING OF KPI

Reported KPI represent the results of respective measurements. Under the assumption that a statistically significant number of samples has been taken, they also represent a prediction on the outcome of tests done with the same set of testing conditions, i.e. parameters of a test.

- NOTE: In statistics, sample usually refers to a set of measurements, i.e. the entirety of all data from a given test.[5] For the purpose of this document, sample (singular) denotes a single data point for a KPI computation, i.e. information related to one particular transaction. This is equivalent to the term "sample point" or "observation" in the statistics context.

Therefore, the functional descriptions use the term 'probability' for KPI which have the type of a rate, in accordance with the wording in ETSI TS 102 250-2 and Rec. ITU-T E.804.

The term 'time' is used in two ways. If the context is individual transactions, it means respective single values for that particular transaction. In a KPI context, it designates an aggregated value. If no other definition is made, this shall mean the average of transaction-wise values.

To avoid duplications of text, validity rules are assumed to be generic, i.e. relate to the formal validity definitions outline in Terminology.

D.5 SPECIFIC PROBLEMS OF DFS TRANSACTIONS

DFS is, at least from a technical testing perspective, a store and forward service. For pragmatic purpose, a time-out condition for a test case is necessary; otherwise, a 'hanging' transaction would effectively block a test.

Considering that a field test for DFS is transferring real money, there is the basic question of clean-up. In case a transaction is unsuccessful, the money involved in this transaction would have to be assumed, and new money would have to be inserted into the loop.

Even though DFS appears primarily to be a direct, interactive type of service, it has some store-and-forward properties. This relates to the matter of using reasonable values for time-outs.

Here, several aspects have to be considered carefully. If time-out values are too short, this would not only represent customer perspective by painting a too-negative picture of the service. It would also increase the amount of money needed for insertion after assumed failure.

From an operational point of view, it would also create additional complexity. If time-out is declared due to a missing response of the system, the next transaction will be started. This would then either need a dedicated cancellation of the ongoing transaction, or the test would be in a kind of undefined state.

Moreover, due to the secondary response (summary SMS), there is actually a double time frame. The primary confirmation may have arrived, but the summary SMS are still under way. It is assumed for the time being—with a note that this should be validated—that these SMS are actually decoupled from the DFS process. If the waiting time for these SMS has expired, and the next transaction is started, they can still appear. The procedure also needs to cover this possibility in order not to introduce confusion in case it occurs.

Using long time-outs—to reduce this risk, understood as hoping a transaction without a clear response may turn out to be successful after all—will however reduce the yield of a measurement campaign in case of a high actual loss rate.

Also, some care needs to be taken in definition of a clean-up process. A clean-up process should not pose the risk of messing up the test; i.e. an attempt to roll back a transaction may not only cause time delay but also create additional disturbance in the system and endanger data integrity. At present it appears to be the most sensible decision to refrain from any situational roll-back attempts and assume that some final tidying-up is made. For test design this means that sufficient reserves—and a good monitoring—have to be allocated to keep the testing process going.

ANNEX E

Naming rules, data structures and related processes used in the pilot project

E.1 NAMING

E.1.1 General

Element names shown in bold face are functional names which shall be used consistently throughout all relevant documents. They may also have abbreviations used for brevity but only in the current clause of this document.

E.1.2 Teams

Each team shall be given a unique Team ID (TID), made of alphanumeric characters. The TN can be freely chosen but it shall not be changed over time.

E.1.3 Devices

The Device ID (DID) is composed of the device's role and index (e.g. SP1, FP2, and OP), underscore ('_') and the last 6 digits of the device's IMEI.

On log sheets, an abbreviated name using only the device's role is used. The full DID can be looked up by the respective entry in the device/team assignment data (see below).

The IMEI is the identifier displayed if the code *#06# is entered in the Phone Dial window.

In case of dual-SIM devices, they may have 2 IMEI. In that case, the IMEI for the first SIM position is used for the DID. Usually this is the first IMEI displayed for *#06# (to be verified).

Example for a full DID: SP1_123456

On log sheets, Device Role Aliases can be used instead of the short role names. The following aliases are defined:

SP	Full Capability
FP	Low Capability

E.2 TEAM/DEVICE ASSIGNMENT LIST

A list is kept which records the assignment of devices to teams. As this assignment may change over time, the respective time window is also recorded.

The list has the following elements:

TABLE E-1: Data base format definition for team/device assignment list

ELEMENT	TYPE
Team ID	Varchar(128)
Device ID	Varchar(64)
Start time and date of assignment	datetime
End time and date of assignment	datetime

The end time can be NULL indicating that the assignment is ongoing.

E.3 NOTIFICATION SMS

E.3.1 Transfer and data handling process

The **Notification SMS** (NSMS) (sent to the A and B party) contain information about the DFS transaction. This information is used to complement the overall information.

The steps of this process are:

- NSMS arrive at respective devices
- The SMS backup process (see SMS Backup & restore app) sends, when invoked, an e-mail with an XML file attachment to a specific location. This XML file contains a copy of all SMS which were stored on the device at the time of invocation.
- The attachment is processed by importing it into the project data base.

The NSMS do not contain information about the devices involved. This information must therefore be added during the overall process of NSMS collection.

This is done using the following definition and process:

- The set-up of SMS backup allows to configure the Subject. This Subject shall contain the DID of the respective device.
- For import, the DID shall be added to the respective data items.
- As each backup file is a snapshot of all SMS on the device, subsequent executions will produce duplicates of NSMS. The data structure/import process must have provisions to handle these duplicates.

E.3.2 Notification SMS data table structure

TABLE E-2: Notification SMS data table structure

ELEMENT	TYPE
Device ID	Varchar(64)
Import date and time	datetime
SMS content	Mirroring of XML structure

E.3.3 Assignment of primary test data and SMS

Each successful DFS transaction is assumed to produce a set of primary data (timestamp information according to the definitions elsewhere in the present document) and two confirmation SMS on the A and B device, respectively.

By processing SMS back-up copies uploaded to the data base, these SMS are assigned. There are two basic types (A side and B side SMS). There can be other SMS on the device. Therefore, the classification and assignment process has the following stages:

i) Identify if a SMS is of type A-side, B-side or other.

ii) If A-side, attempt to find the matching B side SMS from another device.

iii) If B-side, attempt to find the matching A-side SMS (actually steps 2 and 3 are symmetrical).

iv) Attempt to find the matching primary transaction for the A-side and B-side SMS, respectively, using device/team allocation and timestamp

Ideally, the process assigns all A-side and B-side SMS. It is expected that "orphans" exist which do not have a counterpart. For such orphans, the first step is to check if SMS exist on devices which have not been covered by the backup process. If this check finds previously missing SMS, they shall be processed.

The remaining orphans are again sorted into categories:

- A or B side SMS which have matching DFS transactions. This indicates transactions where such SMS are missing and shall be notified along accordingly.
- A or B side SMS which have no matching DFS transaction. An investigation shall be conducted to clarify the circumstances.

E.3.4 Storage and deletion aspects of SMS on devices

The process of SMS back-up is based on periodic copies of all SMS on a particular device.

In the course of the test campaign, locally stored SMSD will accumulate unless they are deleted. Deletion procedures carry the risk of unwanted deletion of meaningful data. A hard cause to delete SMS would be capacity issue. Unless this is given, it is assumed to be better to handle SMS duplicates — which is technically quite simple in data processing — than to run a deletion process.

In case a deletion process is required after all, it is performed along the following process:

i) There are regular device maintenance cycles (e.g. once per week) where all devices used in the test campaign are participating.

ii) From processing of previously uploaded data, a reference point in time for DFS confirmation SMS is calculated (SMS-RP, type time/date). It is assumed that up to this RT, all uploaded SMS are checked and assigned (see Assignment of primary test data and SMS) and that any hints on missing SMS (detected as missing in uploaded data, to be checked for on devices) are clarified.

iii) In the maintenance process, all locally stored SMS older than the SMS-RP are deleted.

TABLE E-3: KPI/Trigger point reference

			MTCR	MTCT	MTFPR	MTFNR	MTFTRR	MTASSR	MTAST	MTLR	MTDR
			Rate/ Probability	Time	Rate/ Probability	Rate/ Probability	Rate/ Probability	Rate/ Probability	Time	Rate/ Probability	Rate/ Probability
			Money Transfer completion rate	Money Transfer completion time	Money Transfer False Positive Rate	Money Transfer False Negative Rate	Money Transfer Failed Transaction Resolution Rate	Money Transfer Account Stabilization Success Rate	Money Transfer Account Stabilization Time	Money Transfer Loss Rate	Money Transfer Duplication Rate
DFS_P2P_AA_100	Start DFS app	enter start USSD command									
DFS_P2P_AE_104	Prompt to select TA type	Prompt to select TA type	Start	Start	Start		Start				
DFS_P2P_AA_108	Select: Transfer	enter 1 to select “Transfer Money”	X (2)	X (3)							
DFS_P2P_AE_112	Prompt to select recipient type	Prompt to select recipient type	X (2)	X (3)							
DFS_P2P_AA_116	Select: To mobile user	enter 1 to select “to Mobile Money user”	X (2)	X (3)							
DFS_P2P_AE_120	0	Prompt to select category of recipient	X (2)	X (3)							
DFS_P2P_AA_124	0	enter 1 to select “to subscriber”	X (2)	X (3)							
DFS_P2P_AE_128	Prompt to select recipient ID	Prompt to select recipient ID	X (2)	X (3)							
DFS_P2P_AA_132	Enter B number and continue	Enter B number and continue	X (2)	X (3)							
DFS_P2P_AE_136	Prompt to select recipient ID again	Prompt to select recipient ID again	X (2)	X (3)							
DFS_P2P_AA_140	Enter B number again and continue	Enter B number again and continue	X (2)	X (3)							
DFS_P2P_AE_144	Prompt to enter amount	Prompt to enter amount	X (2)	X (3)							
DFS_P2P_AA_148	Enter amount and continue	Enter amount and continue	X (2)	X (3)							
DFS_P2P_AE_152	Prompt to enter reference	Prompt to enter reference	X (2)	X (3)							
DFS_P2P_AA_156	Enter reference and continue	Enter reference and continue	X (2)	X (3)							
DFS_P2P_AE_160	Request to confirm transaction appears	0	X (2)	X (3)							

TABLE E-3, *continued*

			MTCR	MTCT	MTFPR	MTFNR	MTFTRR	MTASSR	MTAST	MTLR	MTDR
			Rate/ Probability	Time	Rate/ Probability	Rate/ Probability	Rate/ Probability	Rate/ Probability	Time	Rate/ Probability	Rate/ Probability
			Money Transfer completion rate	Money Transfer completion time	Money Transfer False Positive Rate	Money Transfer False Negative Rate	Money Transfer Failed Transaction Resolution Rate	Money Transfer Account Stabilization Success Rate	Money Transfer Account Stabilization Time	Money Transfer Loss Rate	Money Transfer Duplication Rate
DFS_P2P_AA_164	Confirm	0	X (2)	X (3)							
DFS_P2P_AE_168	Request for PIN appears	Request for PIN appears	X (2)	X (3)							
DFS_P2P_AA_200	Enter PIN and confirm	Enter PIN and confirm	X (2)	X (3)		Start		Start	Start	Start	Start
DFS_P2P_AE_210	Display TA in progress info	Display TA in progress info	X (2)	X (3)							
DFS_P2P_AE_300	Display payment confirmation	Display payment confirmation	Success	End (1)	Success (use for validity check)	Success (use for validity check)	Success (use for validity check)	Success (use for validity check)	Success (use for validity check)	Success (use for validity check)	Success (use for validity check)
DFS_P2P_AE_310	Receive A side payment info	Receive A side payment info			use content for validation	use content for validation	use content for validation	use content for validation	use content for validation	use content for validation	use content for validation
DFS_P2P_BE_320	Receive B side payment info	Receive B side payment info			use content for validation	use content for validation	use content for validation	use content for validation	use content for validation	use content for validation	use content for validation

- NOTE 1: For a time value, all sub-phases where human interaction is involved need to be eliminated, and, eventually, a normalized/typical time value has to be used instead.
- NOTE 2: Used to create detail information in case of failure (identify sub-phase where failure occurred).
- NOTE 3: use all available elements to calculate eligible time intervals (only use the times which do not contain human action, e.g. time from confirmation of an info item to appearance of the next prompt).

APPENDIX I

Campaign log examples

Location Log Sheet

Team ID	
Team Leader	
Location name	
Date	

(Date/Time)

Sheet started	*Point in time when the sheet is started to be used*
Sheet completed	*Point in time when the sheet is full/completed*
Sheet photographed	*After completion, the sheet shall be photographed and the photo uploaded (e-mailed) to a given location*
Sheet Photo uploaded	*Checked after the sheet has been successfully e-mailed*

INITIAL CHECK OF CONDITIONS AND MEASUREMENT SET-UP // *After entering the location*

	Low Capability 1	Low Capability 2	Full Capability 1	Full Capability 2	Observer Phone
Phone ID					
Time: (24h format)					
Operator					
RF level					
Battery level					
Charger connected?					
Date/time correct					
Network mode	2G only ☐	2G only ☐	LTE/3G/2G auto connect ☐	LTE/3G/2G auto connect ☐	LTE/3G/2G auto connect ☐
Mobile data	disabled ☐	disabled ☐	enabled ☐	enabled ☐	enabled ☐

REGULAR CHECKS // *Remark: Regular checks should be scheduled approx. every 2 hours*

	Low Capability 1	Low Capability 2	Full Capability 1	Full Capability 2	Observer Phone
Phone ID					
Time: (24h format)					
RF level (no network=0)					
Battery level					
Date/time correct					
Network mode	2G only ☐	2G only ☐	LTE/3G/2G auto connect ☐	LTE/3G/2G auto connect ☐	LTE/3G/2G auto connect ☐
Mobile data	disabled ☐	disabled ☐	enabled ☐	enabled ☐	enabled ☐
Time:					
RF level (no network=0)					
Battery level					
Date/time correct					
Network mode	2G only ☐	2G only ☐	LTE/3G/2G auto connect ☐	LTE/3G/2G auto connect ☐	LTE/3G/2G auto connect ☐
Mobile data	disabled ☐	disabled ☐	enabled ☐	enabled ☐	enabled ☐

FINAL CHECK OF CONDITIONS AND MEASUREMENT SET-UP // *Before leaving the location*

	Low Capability 1	Low Capability 2	Full Capability 1	Full Capability 2	Observer Phone
Phone ID					
Time: (24h format)					
RF level (no network=0)					
Battery level					
Date/time correct					
Network mode	2G only ☐	2G only ☐	LTE/3G/2G auto connect ☐	LTE/3G/2G auto connect ☐	LTE/3G/2G auto connect ☐
Mobile data	disabled ☐	disabled ☐	enabled ☐	enabled ☐	enabled ☐
SMS Backup executed					

Data Log Sheet P2P transfer

Team ID	
Team Leader	
Location name	
Date	

(Date/Time)

Sheet started	*Point in time when the sheet is started to be used*
Sheet completed	*Point in time when the sheet is full/completed*
Sheet photographed	*After completion, the sheet shall be photographed and the photo uploaded (e-mailed) to a given location*
Sheet Photo uploaded	*Checked after the sheet has been successfully e-mailed*

AMOUNT OF MONEY ON ACCOUNTS, BY TEST DEVICE

Device ID	at sheet start	at sheet end

Field description

Device ID	Unique ID of the device
Sender ID	Device ID of the device used to send money (A party)
Receiver ID	Device ID of the device used to receive money (B party)

Please enter times in hh:mm:ss format

TEST DATA LOG

	Amount	Sender ID	Receiver ID	T1	T2	T3	T4	T5	T6	T7
				Start of TA	Trigger transfer	Success	Failure	Time-out	A side SMS	B side SMS
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										

Event Log Sheet

Team ID	
Team Leader	
Location name	
Date	

(Date/Time)

Sheet started	*Point in time when the sheet is started to be used*
Sheet completed	*Point in time when the sheet is full/completed*
Sheet photographed	*After completion, the sheet shall be photographed and the photo uploaded (e-mailed) to a given location*
Sheet Photo uploaded	*Checked after the sheet has been successfully e-mailed*

	Time (24h format)	Description of event
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		

Examples of events to be logged:

- Changes of power supply (battery/charger operation, power outage,...)
- Changes of network coverage (loss of coverage/return of coverage)
- Stopping or resuming the test measurement (pause/end of pause, external events,...)
- Any unusual events which occured during the test

APPENDIX II

Description of the Ghana pilot campaign

II.1 DATA COLLECTION METHOD

For the Ghana pilot project, full manual acquisition of data, i.e. method a) as defined in clause 6.2, has been selected, as it is the most generic one.

II.2 EVENT DEFINITION

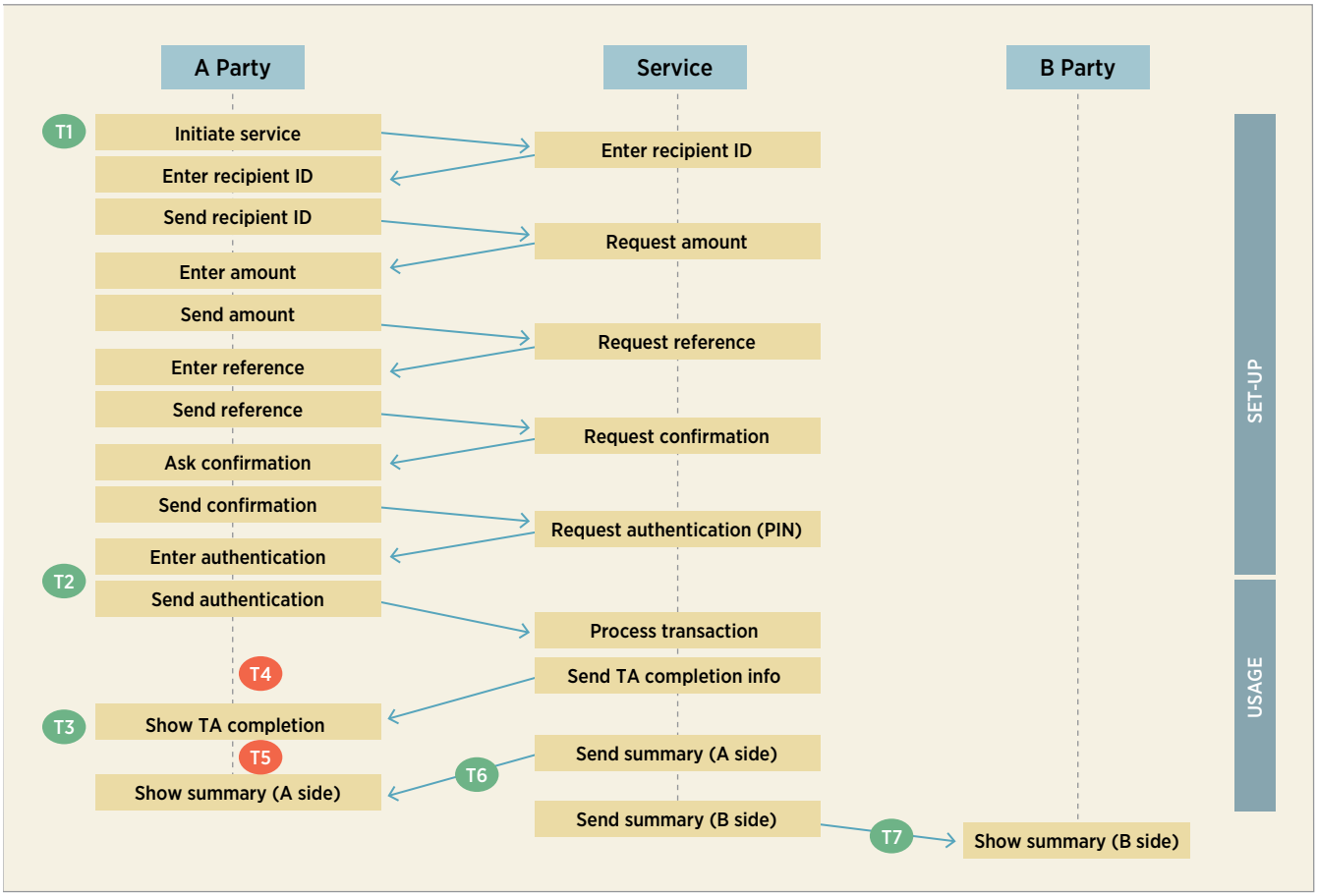
Events have to be recorded with their respective time-stamps. Manual recording of those events requires a certain amount of time. This should not delay the DFS process under test. This sets practical limits to the granularity or number of events per DFS use case. Therefore, the extent of data will be limited to the set of practical KPI as defined in clause 7.

The decision about the time-out condition needs actually to be made by a member of the observer team. This requires special element in the toolset used, e.g. an alarm timer started with T1.

- NOTE: It is assumed that T6 and T7 can also be derived from captured SMS on respective phones later. It is however desirable to record these events in the data logs too.

Figure II-1 (based on Figure 4-2) shows the event flow with the recording points T1 to T7 for manual time measurement. Events which belong to the positive result case are shown with green background color; negative events (indicating failure or time-out) have red background color.

FIGURE II-1: DFS event flow with recording points for manual time measurement



II.3 MAPPING OF ACQUIRED DATA TO FORMAL TRIGGER POINTS

By comparison to the full trigger point list shown in in Table 4-2, the timestamps used in the Ghana pilot campaign, as shown in Table II 1 are a subset (see the full discussion of consequences of manual execution of tests in clause 7). Consequently, a mapping of timer flags to formal trigger points needs to be done and is shown in Table II-1.

Please note that there are no format trigger points for T4 and T5, as they are not linked to events from the activity flow in a DFS implementation. In case of T4, it will be set from a failure indication given by the DFS implementation which cannot be provoked directly from A or B side, but needs to be interpreted as part of human or automated monitoring of the test. In case of T5, it is set by a time-out condition determined by some external time-keeping process.

II.4 BACKGROUND TESTING OF THE TRANSPORT NETWORK

For SMS testing, sending SMS to the same device is used to simplify data capture.

For USSD testing, a code (or multiple codes) should not make permanent changes to the state of the subscription, or to the mobile device. For the tests, the USSD code *135# has been chosen which queries the own telephone number. In addition, using a code, which relates directly to DFS, should not be used, as this may bring the DFS system into undesired states. Suitable USSD codes would serve as suitable proxies for the functioning of the USSD subsystem in the network under test, without having undesired side effects.

For the choice of web sites, small sites were selected, i.e. the Google search engine start page, and the ETSI Kepler for Smartphones page[6] hosted on a reference server.

Even though the DFS implementation in Ghana uses USSD and SMS as its principal carrier services, packet data related test cases have been added to collect some potentially useful additional information.

After some validation tests, it has been determined that using a data server in Germany (hosted at Focus Infocom) provides the best operational value also with respect to maintenance. During the pre-pilot phase, a second server (at Strato, a large German web hosting system) has been tested and verified to work. This was done to make sure a fallback solution is available in case of server problems during the campaign.

Table II-1: Reference table: Ghana campaign timestamps to formal trigger points

TIMESTAMP	FORMAL TRIGGER POINT	REMARKS
T1	AA_100	Start of test case execution
T2	AA_200	Start of core transaction
T3	AE_300	Successful completion of transaction
T4		Used as failure indicator
T5		Used as timeout indicator
T6	AE_310	Reception of information SMS on A side
T7	BE_320	Reception of information SMS on B side

APPENDIX III

Example for the set-up of a SMS Backup Tool

Installation is made on all mobile devices used for a specific DFS testing campaign.

- NOTE 1: It is assumed that the terms of use for this app allow the intended usage. Respective terms need to be monitored and checked against the mode of usage. In case of conflicts, respective resolution by e.g. purchase of required license or selection of another app needs to be considered.

The following description has been created from a test installation as of 24 February 2018. If a set-up is made later, the user interface may be different. In this case, please inform one of the editors of the present document.

- NOTE 2: The app asks for a number of permissions, which are quite far going. It is assumed that this uncritical if the device is used only for test and measurement purposes. In case of installation on a device which is also used for other purposes (business and/or private), it is highly recommended to carefully consider if this mode of operation is actually desirable.
  - ❏ Activate the Play Store and select for "SMS Backup & Restore" (SyncTech Pty Ltd).
  - ❏ Install the app.
  - ❏ Run the app and follow the instructions
  - ❏ A series of requests to allow access is displayed. Allow all of them:
    - Access contacts.
    - Access photos, media and files on the device.
    - Send and view SMS messages.
    - Make and manage phone calls.
  - ❏ Next screen offers to set up a backup; accept by tapping "Set up a backup".
  - ❏ The app asks what should be backed up; select "messages", un-select "Phone calls" and tap "Next"
  - ❏ The app asks where backups should be stored. Select "Your phone". In the dialog box presented subsequently, select Local backup, Default app folder" (the default setting) and tap "OK".
  - ❏ Tap "Next". A dialog box appears reminding that backup files may be lost if stored locally. Select "Do not remind me again" and tap "Yes".
  - ❏ A screen for setting up a schedule for recurring backups appears. Select "Daily". The default time is 0:00.
  - ❏ Tap "back up now" to create the first backup. The backup process runs and a summary screen is shown. There is a section labelled "Backing up locally isn't safe.
  - ❏ Tap "Change location". A dialog for additional backup locations appears. Select "Email". Another screen appears. Open the drop-down list at "Email Service type". Select Gmail.
  - ❏ Enter the recipient email address (to be provided by the campaign management).
  - ❏ In the Subject field, enter "SMS Backup" followed by the IID of the device
  - ❏ Tap Log In. A dialog box asks for the mail account to be used. Select the account, which has been created at initial configuration of the device (a Gmail account), and tap "OK". The app checks access and should indicate success.
  - ❏ Tap "Test". Check the destination mail account if the test email has been received and tap "OK" in the dialog box (transfer of the email may take some time, wait for a reasonable amount of time. If no mail is received, check and repeat the email set-up process.
  - ❏ Tap "Save" to revert to the overview screen.
  - ❏ Open the app's option menu (symbol in the upper-left corner of the screen), select the gear symbol (settings), and select "Backup Settings". Scroll down to "Email backup" and make sure it is activated. If it has been previously inactive, activation will show the email setup dialog again. Check the settings and confirm by tapping "Save".
  - ❏ Revert to the main screen (Home). To finalize the set-up, tap "Back up Now". Tap "Back Up" to start the back-up process.

After completion of the set-up, the app should run an automated back-up at the scheduled time. To create an ad-hoc back up (after finishing the measurements at a given location), use the "Back up now" function as previously described.

Endnotes

1. Where required information is collected element by element by the service (Type B).
2. Derived from a practical implementation. Blue fields in column Short TPID mark user-activity sub-phases.
3. https://www.etsi.org/deliver/etsi_ts/102200_102299/10225002/02.06.01_60/ts_10225002v020601p.pdf
4. https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-E.804-201402-I!!PDF-E&type=items
5. https://en.wikipedia.org/wiki/Sample_(statistics)
6. http://docbox.etsi.org/STQ/Open/Kepler/Kepler_for_Smartphones.zip

Telecommunication Union,
Place des Nations, CH-1211
Geneva 20, Switzerland

September 2019

E-mail: mailto:tsbfigisit@itu.int