The T-Mobile UK measurement strategy for meter and bill integrity of public electronic communication services.

Scope

This document defines how T-Mobile will measure meter and bill integrity. The principles described are general, but are also intended to satisfy obligations under the Meter and Billing Direction, formerly known as OTR003: 2001.

This version has been made suitable for public distribution through the removal of all information sensitive to T-Mobile UK.

Purpose

By defining the basis of measurement, this document

- provides a theoretical basis for calculating performance even when systems and processes change; and
- satisfies an explicit obligation of the Meter and Billing Direction.

It is intended that this document also be used as a general guide to the issues of measuring meter and bill integrity.

The public version of this document is offered to

- increase transparency, enabling customers and inter-working partners to better scrutinise and understand how T-Mobile UK seeks to monitor the accuracy of customer charging; and
- provide a template approach that other communications providers may wish to emulate.

T · ·Mobile· · ·

Table of contents

Table of tables

Table of figures

1. Introduction

1.1 What is meant by metering and billing integrity in this document

All businesses need to have a means of recording sales and then charging customers for those sales. In some cases, the two are simultaneous. In other cases, the exchange is asynchronous; the customer may execute a financial transfer before or after the desired goods or services are provided. Whatever the convention applied, any reputable business will want to:

- correctly record what was supplied to customers; and
- correctly charge customers for what was recorded as supplied.

The first bullet relates well to what is called here by the term “metering”. The second bullet more roughly relates to what is called here “billing”. However, these terms can be confusing so they are precisely defined later on. The main point of this text is to describe how to measure the integrity of metering and billing. Integrity just means that the supply is correctly recorded, and the consequent charge is correctly rendered.

1.2 The benefits of measuring integrity

1.2.1 Consumer protection

Any customer deserves to receive an accurate charge for what they actually received. That charge should be calculated according to the basis agreed between supplier and customer. There are many cases where customers are dependent on the integrity of systems and processes they cannot readily examine. In these cases, the onus is on the business to establish the integrity of their systems and processes.

This particular document is an output from T-Mobile UK’s programme for compliance with Oftel’s regulations regarding the integrity of metering and billing. The integrity expectations are stated in the Meter and Billing Direction, formerly the OTR003: 2001 standard. This document meets an explicit requirement of the Direction. However, the principles discussed remain true irrespective of the legal or regulatory environment, and hence should be good practice even in the absence of any external driver to demonstrate integrity.

Measuring integrity is a means of demonstrating that the consumer’s rights have been protected.

1.2.2 Commercial benefits

There are commercial benefits that follow from avoiding inaccuracy in processes and systems:

- errors in the customer’s favour will mean less revenue is earned than should have been;
- errors in the business’ favour may lead to increased cost in handling complaints, loss of customer goodwill or churn.

Measuring integrity is a means of calculating how much the business is undercharging or overcharging. In either case, it enables a basis for appropriately prioritised investigation and remedy of any issues leading to incorrect charges. In the absence of systematic regular measurement, integrity issues may go unidentified for an indefinite period of time.

1.3 How this document is structured

1.3.1 Main body

The main body of this document discusses the principles and theory for how to measure integrity in any relevant business. It is based on the experience of T-Mobile UK but not specific to it. Wherever practical, real-life issues likely to occur in other businesses are discussed in detail.

The main body is to be made public in the interests of transparency as to how T-Mobile UK meets its consumer protection obligations.

It is not expected that the main body will need regular revision.

1.3.2 Confidential appendices

The appendices will be made available solely to authorised T-Mobile staff and its Approval Body. They include commercially sensitive information regarding

- what processes and systems in T-Mobile UK are subject to measurement and hence details of how they operate; and
- how the principles discussed in the main body were applied in practice.

The appendices may need revision on a regular basis, to reflect changes in systems, processes and such. There is likely to be a regular need to change these to reflect changes in how T-Mobile UK records sales and charges for them. Changes may also occur as T-Mobile UK seeks to augment or enhance its measurement approach.

The confidential appendices have been written according to T-Mobile UK's best guess of what an Approval Body might require. The total extent of all detail that might be provided would be unwieldy to include in one document, so only key matters of concern that might be of interest to the Approval Body are included here. No advice or guidance, whether general or specific, has been received regarding an Approval Body's expectations for what the content should be.

The confidential appendices are split in a manner to aid the Approval Body's audit team to mimic a split of services based on the naïve understanding of the "man on the Clapham Omnibus". This split has nothing to do with how technology or processes work in practice, but closely represents how the end customer *believes* they are segmented. This enables a closer approximation of decisions on the status of performance to the judgement that would be reached by a layman with little or no understanding of how metering and billing is performed in practice.

2. Concepts and definitions

2.1 What is measurement

The verb “to measure” means to

- ascertain
- the extent or quantity [of a thing]
- by comparison to a known standard or fixed unit.

This also means that measurement excludes the following:

- it is not guesswork or a subjective judgement;
- it cannot be expressed in qualitative terms, as the aim is to represent in numbers the relationship between the thing measured and the unit of measurement; and
- it is not subject to reinterpretation, as the standard or unit does not change.

2.1.1 The thing that is measured

Measuring integrity is the same as measuring its opposite, error. What complicates things is that the number of errors may be uninteresting, because we really want to understand how serious the errors are. Some errors may have no impact, but others may have enormous consequences, and it can be difficult to extrapolate from the kind of error to the likely impact. The purpose of this measurement approach is to quantify the impact of all errors that affect the final charges levied. These errors are determined by comparison to what *should* have been the charged.

It may be thought that it is necessary to know what the errors are before they can be quantified. That is one approach, but knowing an error has taken place need not mean the error itself has been identified. Imagine a situation where a customer in a shop receives the wrong change. The person serving them may have incorrectly keyed the sales into the till, or added up the change wrongly, or the till may have been in error, or the bar code on one item may have been associated with a price different to that stated on the product’s labelling. Whatever the reason, the important thing is that the customer can tell a mistake has occurred without knowing exactly why. There is hence a difference between

- detecting and measuring error in general; and
- investigating and analysing the causes of error.

This document is about the former, not the latter. The emphasis here is to correctly quantify error. Once quantified, appropriate resources can be applied to the investigation and resolution of the causes of errors. As a result, the techniques described do not *assume* that certain errors are likely or unlikely. Any measurement technique that presupposes the outcome is by definition *biased*. The techniques described are intended to correctly quantify the extent of error without prejudging the outcome.

2.1.2 Numbers and units

A measure is ultimately always expressed as a number and a unit. Clarity is needed as to what are the relevant units for measuring the integrity of metering and billing. The two types of units that are primarily used in measuring meter and bill integrity are:

- units of monetary value; and
- counting discrete events and charges.

Given the purpose of measurement is to assure accuracy of charging, it follows that the extent of error in determining the value of charges presented is the key single numerical property to be determined. Customers and businesses are likely to be most concerned about the monetary value of errors that take place more than they are the number of items that are in error. However, it is also worth considering how many discrete charges are in error. This is because the cost of correction or the impact on customer perception may be more closely related to the number of individual items in error than the value of each error. This is most obviously true when comparing a large number of errors, each of small value, to a small number of errors of large value.

Depending on how a charge is calculated, the business will need to record properties expressed in other kinds of units, such as

- time; and
- data volumes.

Quantifying meter and billing integrity may hence mean that the accuracy of recording properties in these units must also be measured. However, the purpose is subordinate to determining the impact in terms of value or numbers of incorrectly rendered charges. Hence, the appropriate units for measuring error in terms of time, data volumes and the like depends on the way charges are calculated and the rates applied.

2.1.3 Defining the measure

"It is after all beyond doubt that a measurement with one yardstick can be used to predict the results of measurement with others. And, further, that if it could not – our whole system of measuring would collapse."[1]

Measures should be clearly defined. Without definition, it is impossible to practically implement consistent measures. Definition also clarifies what is *not* being measured. By defining a measure, it is also possible to state what level of confidence can be obtained from any given measure as compared to any given target. To avoid the need to revise a measure, it is a valuable discipline to state a clear definition at the outset and stick to it. It is

[1] Ludwig Wittegenstein, Remarks on the Foundation of Mathematics

undesirable to have a situation where the measurement is in need of revision, as this undermines the original measure's credibility. Measurement is based on fixed standards, so measurement methods should not change on a regular basis. Changing the basis of measurement also impedes comparisons between results drawn before and after the change.

There is not much purpose to measurement if the results are subject to perpetual revision. If there is reason to doubt the validity of the measure, the method, not the output, needs to be changed. It is inappropriate to revise a single result in isolation, because this introduces a risk of bias. If results indicate a problem with the method of measurement, then the method should be corrected and the measure collected afresh and consistently from then on. Where a measure has been correctly obtained but is flawed by chance, say because a sample is not representative, it should still be necessary to *demonstrate* the flaw instead of just assuming it. This might require the collection of further and larger samples, or analysis of the cause of the error in order to show that the real likelihood of the error is significantly different to the results obtained from the sample.

Any basis of measurement that permits results to be disregarded if they do not fit preconceptions is flawed. Any basis of measurement that ignores errors on the basis that they can be treated as "resolved" because of subsequent actions is also flawed. Either of these so-called "corrections" of measures should be properly considered as forms of *bias*. The measurement approach stated in this document provides for no post-measurement revision of results, with the sole exception of methods to address sampling error.

A measure works well if it can be applied consistently in different places and at different times. It serves no purpose if each measure is itself a "one-off". As the purpose is to establish the integrity of metering and billing, there is no value to a measurement approach that permits each new measure to be determined on a one-off basis without consistency.

It is reasonable to expect a degree of trial and error when establishing measures of meter and bill integrity. However, that period of trial and error should not last indefinitely. A flaw in the method of measurement will typically imply the results obtained using that method are also flawed and should be disregarded. Hence, the objective is to design and implement measures that are free from any flaw.

2.1.4 Setting appropriate expectations for precision in measurement

It is reasonably obvious that in order to measure to a certain level of precision, it is necessary to implement devices, processes and the like capable of measuring to *at least* that level of precision. The precision of the measuring device depends on how it works. For equipment, attainable levels of precision should be documented in a reputable manufacturer's specifications. These may be independently tested to show that the actual performance is consistent with specification.

Although discussions of precision usually focus on automation, it is also important to consider human elements that may affect the reliability of results. For example, a person required to manually inspect an invoice is unlikely to find all the errors if given insufficient time. This is not a reason to object to a human element to measurement, but human factors must be allowed for. There are advantages to using people rather than automation for measurement, especially where automated checking is too expensive or inflexible. Just as it is necessary to test and monitor automated checks to ensure they are correct, similar techniques can be applied to manual checks. Such techniques include varying the people who perform checks, performing cross-checks or reviews, and training people to be aware of changes before they take place.

There may be other factors affecting the precision of measurement unrelated to how the measurement device itself works. These factors may be described as environmental. They relate to a fundamental (though usually very small) inconsistency between what is being measured and how it is being measured. A simple real-life example would be to measure the time at which a signal occurs from the point of view of the source and the destination. However quickly the signal travels the distance between the two points, there is some delay. If the measurement device is located at the signal's origin, but is used to measure when the signal is received (or vice versa) then the delay may affect the level of precision in measurements taken with that device. If the delay can be precisely determined, and is consistent, then it may be allowed for without a reduction in precision. If the delay is negligible, then it may be ignored. However, if the delay cannot be precisely determined, is variable, and cannot be considered negligible, then it will reduce the level of precision in measurement. Just because a delay (or similar property) may be determined *in theory*, this does not mean that the relevant scenario in the real world has been shown to be consistent with the theory. For example, although it is possible to determine how long it takes a signal to travel between two specific points, it may be impractical to determine or even estimate the delay where signals may travel a variety of paths between a large number of combinations of source and destination. Although the delay can be precisely measured for any two points, it is more likely that the *range* between the shortest possible and longest possible be used if it is impractical to relate the exact delay to each specific event. The wider the range, the greater the imprecision that results from this environmental factor.

When measuring performance by counting discrete numbers of events, the more events that are counted, the more precise the measurement. For example, suppose a sample of 10 events are observed, and 9 are found to be correctly charged. The error rate in this sample is 10%. Ignoring the risk that the sample is unrepresentative, the most precise conclusion possible is that the population error rate lies between 5% and 15%. Probability tells us that 9 out of a sample of 10 is the most likely outcome for a population error rate between 5% and 15%. To be more precise in estimating the population error rate requires a bigger sample. If the sample had been of 100 items, with 90 correctly charged, the population error rate would have been measured to between 9.5% and 10.5%. Ignoring the risk of unrepresentative samples, it is true that the more data there is, the more precise the ability to measure the rate of error. So, it is wrong to try to measure errors to a higher level of precision than permitted from the volume of data obtained. This is most pertinent where the *expected* level of error is very low. If an error is expected to occur only once in 100,000 events, and it is not possible to separately monitor the causes of error in that one case, then *at least* 100,000 events need to be inspected in order to conclude whether 1: 100,000 is the error rate[2]. Allowing for sampling error further increases the amount of data needed to obtain a conclusion to the desired level of confidence.

When formulating a measurement method, it serves little purpose to attempt to measure to a level of precision not permitted by:

- the accuracy of measurement equipment and processes;
- the quantity of data available; and
- environmental factors.

It may be difficult to assess these without trial and error, but once assessed the expectations of measurement must be set at a realistic level. If not, results will not be meaningful.

A good analogy to the above is that of a marksman shooting at a target. The closer he is to the target, the smaller the bull's eye he may realistically intend to hit. If he stands further away, the same deviation in aim will mean a miss instead of a hit. Also, standing further away means he must pay more attention to environmental factors like crosswinds. The variability of these reduce the accuracy he can achieve. This means his realistic target must be larger at a greater distance. When devising a measurement approach, it is worth considering how "close" the measure is to the property being measured, and set the measurement expectations accordingly.

2.2 Ratios

It follows that a key consideration in the measurement of error rates is how that error rate itself is to be expressed. It is likely that some form of ratio will be used, which of course requires a numerator and a denominator. This ratio of error rates will reflect the number of errors identified in a given sample. For example, there may be 1 event in error per 1000 events, or €1 aggregate error per €1000 charged.

The numerator in such a ratio is straightforward. This will be the extent of error identified. However, an element of choice exists around the selection of the denominator for an error ratio. The denominator might be either the total including errors, or excluding errors. In other words, it might be the total prior to identifying errors (as might have been originally recorded or presented), or the total that should have been presented if error-free. In essence, there is a choice whether to present error ratios as

$$\frac{\text{Errors}}{\text{Total including errors}} \quad \text{or} \quad \frac{\text{Errors}}{\text{Total excluding errors}}$$

On a philosophical level, it may appear preferable to calculate the proportion of errors against values excluding error. If errors are included in the numerator and denominator, errors effectively bias the

[2] Note that discussion relates to establishing the expected error rate, not to checking for deviations from the expected error rate once an expected error rate is established. In the former, the concern is to establish a reliable estimate of real-world performance without bias or supposition. In the latter, statistical models can be used to calculate the likelihood that results indicate a significant deviation from the expected error rate, but only if the expected error rate has already been reliably established. In this context, what is considered a significant deviation will depend on the target, or tolerable, error rate.

ratio when comparing it to any target. However, where the intention is to calculate based on total excluding errors, it must be recognised that this figure cannot be determined in practice with absolute certainty because of the risk that real errors were not identified.

Take the example of a test for over-metering. 1000 events form the test sample, 1 of which has been duplicated (and thus over-metered). Therefore, 1001 events have been recorded, although only 1000 events actually took place. This is the only error that has been identified within the sample. The error rate can be expressed in one of two ways: either 1 error in 1000 events, or 1 error in 1001 events. The difference in the error rates based on totals including and excluding error is approximately equal to the square of either error rate. For example the difference between 1/1000 and 1/1001 $\approx (1/1000)^2 \approx (1/1001)^2$. This means the importance of using a consistent denominator increases relative to the expected rate itself. As noted, the error rate as calculated using either denominator need only be squared to determine the total numerical impact of the decision to included or exclude errors in the denominator.

2.3 Metering and billing

The terminology for describing how a charge is processed for presentation to a customer is confusing. Many processing steps may occur between the point where a supply is made and the point where a charge is presented. Terminology tends to be based on the names of the automated systems that comprise the chain of processing from supply to bill. Often there is no good terminology to describe the manual activities and workarounds that are also vital to processing. The term "metering and billing" is vaguely used to encompass everything in the processing chain from supply to charge. In that sense, it is misleading to believe that there are two activities, one called "metering" and one called "billing" where every activity can be categorised in under one or other heading. Some of the automated and human activities encompassed by the term "metering and billing" are: recording, logging, mediation, provisioning, rating, real-time account queries and updates and suspense recycling. Although it would be possible to try to redefine the terms "metering" and "billing" to try to aid the categorisation of each processing step, the exercise would satisfy no real purpose. Hence, we define "metering and billing" as follows

"Metering and billing" means all cumulative activities that enable the sale of a supply to be recorded and then rendered as a charge to the customer.

Presented below is a model that breaks this down into its logical components. Note that the logical components need not readily correspond to different discrete systems or processes, because it may be that several logical activities take place in parallel. In turn, a single logical activity may involve the operation of several discrete systems or processes acting in sequence.

3. A generic model for measurement: The MCC Model

Processing data for "metering and billing" involves three logical activities:

- the recording of the supply that is made;
- the calculation of the charge due for the supplies made; and
- the transmission of data from the point of recording the supply to the point of presentation to the customer.

For simplicity, these might best be respectively termed "metering", "calculation", and "conveyance". This document will discuss a model for measuring developed by T-Mobile UK based on the logical activities of metering, conveyance and calculation. This will be referred to as the Metering Conveyance Calculation model, or MCC model for short.

3.1 MCC Model: metering

The term "metering" is used here to cover activities commonly known as metering, logging or recording. The analogy with an old-fashioned meter is helpful, to a point. The basic purpose of metering is to generate sufficient data when a sale is made to correctly charge for the sale. Here are some simple examples:

- a voice call is made, and the switch produces a Call Detail Record which contains relevant parameters such as the origin and destination, the time it took place and the duration;
- a customer telephones a helpdesk to request a new service which incurs a one-off activation charge, and this is typed into the sales order system by the helpdesk assistant;
- a customer clicks on a webpage banner to download an information service, causing the customer details, type of information requested and time of the event to be recorded in a database that will later be queried in order to charge for the content downloaded.

Metering is different to calculation and conveyance in its nature because it is time-critical. The data must be recorded when the relevant events take place. It is not possible to retrospectively record a sale if no data was produced at the time of the sale. Unlike conveyance or calculation, there is no possibility of reworking if there are indications of error. Even with real-time charging, it will still be possible to rework for any errors in conveyance or calculation, so long as the original meter data is retained and is reliable. If there are reasons to believe that meter data is incomplete or invalid, it will be impossible to systematically correct these afterwards. The only way to do that would be, in effect, to implement a second, separate meter to serve as a potential substitute. However, if the two meters disagree, that only begs the question of which one is right.

One source of confusion is that the data output by the meter may not be retained "at" the meter for any significant length of time; it is often produced by the meter but soon afterwards conveyed so it can be stored in a data repository elsewhere. This may complicate the investigation of meter errors if testing is based on the repository's data instead of the raw meter output; errors identified in the repository's data may be actually be the result of failures in the process to convey and write data to the repository. As it is possible to construct a measure based on the direct meter output, it remains true to regard metering as a separate logical activity from conveyance. As with many aspects of measurement, we can understand a distinction between a theoretical goal of measurement and real-life compromises to get a reasonable approximation. It is important to distinguish between the theoretical ideal and practical compromises in order to recognise the implications of compromises.

Because metering is uniquely real-time, and cannot be subsequently reviewed, it means that any measurement of meter error must be based on collecting real-time data from a comparable system. To check the meter output, it is necessary to use an independent system that is provided with identical inputs and is designed to emulate its outputs. In effect, an alternative meter is needed to provide the test data. Any disagreement between the meter and test system begs the question of which (if either) is deemed to be in error, though the practical intention must be to generate a test approach where any discrepancy can be safely ascribed to error by the meter.

"And when I put the ruler up against the table, am I always measuring the table; am I not sometimes checking the ruler? And in what does the distinction between the one procedure and the other consist?"[3].

3.2 MCC Model: conveyance

The word "conveyance" is used here because it is meant to represent the fact that the data output from the meter must be conveyed through a number of intermediary systems before being conveyed to the customer on a bill. There is also an analogy to a manufacturing conveyor belt. In this metering and billing production line, data that records a sale is placed on the conveyor belt at one end, a series of processes take place in sequence to reformat, normalise and add to that data. This ultimately leads to the production of data presented on a bill or similar that states a monetary charge. The idea of conveyance may appear trivial and obviously error-free. Because it exists "between" systems and its complexity is often underestimated, it can be a poorly-controlled source of very significant errors that go unidentified for long periods. Here are some examples of conveyance:

- CDRs are polled from an exchange and transmitted to a mediator;
- CDRs pass "through" a mediator, being input to it, normalised and then output to the destination system;
- The billing system generates a print file which is sent electronically to another business to print the bills;
- A provisioning system flags the completion of a previously outstanding order, and this flag is identified by a batch process causing it to update a record in the billing system;
- CDRs fall into suspense at the billing system, then are successfully recycled;
- A salesman writes down an order on a paper form which is later returned to another department to key in the details into the sales order system;
- Printed bills are sent through the post to the customer's address.

[3] Ludwig Wittegenstein, Remarks on the Foundation of Mathematics

Some of these examples suggest an important question: what are the "ends" to this conveyor belt? Is it from the customer to the customer, or is it enough just to consider the inputs and outputs of a single business, or even a single system? This question becomes even more pertinent when the activities of more than one business contribute to the same "conveyor belt" for a given customer. For example, operator A may record a call using its meter, pass the CDR through a regular electronic file transfer to service provider B to calculate the billable value, which creates a print file and sends it to billing bureau C so it can print and send out the bill. Errors may occur at A, B or C, but all will affect the charge the customer receives. A suggested practical definition of the end points is given below.

- The output of the meter is the starting point for conveyance.
- The end of conveyance is the point where data is output in a form where any further processing is for *presentation* purposes only.

By definition, the meter produces data intended to be used to determine the charges to be rendered. This definition is somewhat arbitrary, but the point is to focus attention on the actual system or systems used to make a record of a sale where that record is the basis for subsequent charging. Other systems may also make records that *could* have been used as the basis of charging, or *may* even be reconciled to the record that is used, but for the purpose of integrity they are irrelevant. What matters is the correctness of the output from the device that *does* produce records used as the basis for charging. This is hence the most sensible starting point for measuring correctness of data conveyance.

A reasonable end point would be where any other processing to be performed should have no impact on the final value of the charges to be presented to the customer. Although it is conceivable that an error may take place during formatting that affects the charges presented, this would normally be a negligible risk. Of course, other conveyance errors may still take place, such as the bill being lost in the post. These kinds of "errors" are of a very different nature to the issues this kind of measurement is intended to address, but there is nothing to prevent measurement of these kinds of errors as well.

Note that some of the examples stated above would be excluded from measurement if the suggested end points are used. For example, losing a bill in the post falls outside the suggested end points. Whichever end points are selected, it is most pertinent that they are clearly identified and consistently used. The suggested end-points are those that T-Mobile UK will use as the theoretical basis for measurement and are consistent with the requirements of OTR003: 2001. However, they are also practical suggestions and may help any operator better understand what processes it intends to monitor and why.

Conveyance is not something that necessarily occurs before or after calculation. It is best considered as something that occurs in parallel with calculation. In the example above using suspense, a calculation could not be performed, hence the CDR could not be conveyed to the next step. Instead, it is redirected to a "holding" data repository until it can be resubmitted (or "recycled" to use the normal terminology). This can be thought of as a small loop that comes off our conveyor belt and returns items back to an earlier point. The point here is that performing the calculation, so that we get the calculated output from the pre-calculation input, is actually part of the conveyor belt.

3.3 MCC Model: calculation

The term calculation is not used here to mean *any* manipulation of data. Data is often manipulated in many ways that do not in any way change the nature of the information relayed by that data. A change in the format of data is not a change in the information it conveys. Changing a number represented in, say, binary to the same number represented in base ten could be said to involve calculations of a certain type, but that is not how the term "calculation" is used here. With respect to metering and billing integrity, calculation includes any operation necessary to determine a monetary value from the conveyed meter output. These operations may be performed at once or in a series of discrete parts. However, their cumulative output is the final monetary value of the charge due.

Some examples of calculation are as follows:

- a CDR for a voice call is rated according to the event type, the A and B party, the tariff plan associated with the A party, the time the event took place and its duration;
- all calls that a customer makes in a certain month are aggregated by the billing system and a discount applied that varies with the total duration of calls made in that month;
- a leased line is implemented, causing an update in the billing system and from that point the line is charged at a standard recurring rate;
- outages of a leased line are recorded in a database in order to determine if service levels have been met, and if service levels have not been met then a flag is set causing the billing system to calculate credits to net against the recurring charge;
- a customer phones a helpdesk and requests a new service, so the helpdesk assistant flags a new order in the billing systems causing a one-off flat fee to be applied to that customer's account;
- a customer sends an SMS to a number which is recognised as charged at a premium rate as it is a request for an automated SMS response of weather information;
- a customer sends an SMS which causes an automated SMS response of weather information, the receipt of which gets recorded as a chargeable event at a premium rate;
- a certain number of events are allocated against an allowance each month, the consequence being that these are provided at zero charge.

Just as conveyance occurs between many systems, more than one system may be involved in calculating the value of a charge. It may even be the case that more than one business is involved in calculating the charge. For instance, operator A meters a call and applies a charge, then passes it to operator B. Operator B has the billing relationship with the end customer. Operator B adds a margin to the call and then charges it to the customer. In this case, both operators perform part of the total calculation of the charge presented to the customer, and either or both might commit a calculation error.

It is not always the case that there is a simple relationship between each event and the charge incurred. This is most evident where there is a discount applied to the total value of a bill. There may be no means of calculating how much each individual event has been effectively discounted. We might think of the discount being netted against the standing charge, instead of the individual calls.

Calculation implies augmenting the data associated with the metered events. Metering produces the original raw data recording a sale, but calculation produces further data that represent charges (or at least intermediate steps in their calculation). Calculation is use of an algorithm based on data relating to the metered events and the appropriate rate of charge for the customer. In practice, this means that customer data needs to be looked up to perform the calculation. The way data typically is organised, the customer will be associated with tariff plan data. For a calculation to be correct, both the algorithm and the standing data must be correct. These are correct if they are consistent with the basis for charging agreed with the customer. Any measurement of the correctness of calculation must at some stage link back to the price plan communicated between supplier and customer.

In one of the examples given above, the concept of a zero charge is described. This must be distinguished from the case where no event is presented on the bill because that kind of event is not chargeable. A zero charge implies a charge has been calculated. Where no charge will ever be levied for a certain kind of event, it may be easier to avoid metering the event, or else fail to convey the event, instead of calculating a zero value for it. The case where a zero charge is calculated and conveyed to the bill needs to be clearly distinguished from the case where the event is simply not metered or conveyed to the bill. This is because there may be integrity reasons to present zero charge events on the bill. Though the event itself may not be charged, it may alter the final charge. The most obvious example would be where a certain volume of events is subject to an allowance. Not conveying a record of an event to the bill is only acceptable if there could never be an affect on the end value of the bill *and* the customer can be safely assumed to have no need for this information.

3.4 A conceptual model for the Total "Metering and Billing" System

Here is a diagram representing the concepts described above.

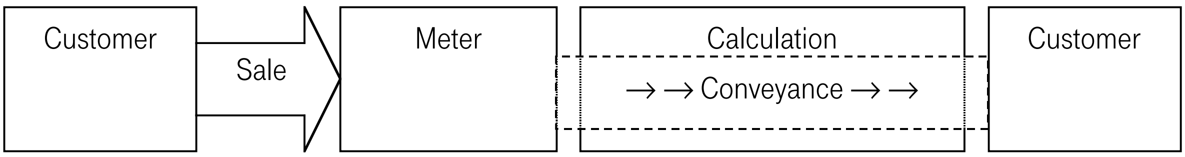

Figure 1: Concept diagram of metering and billing

Previous models have made the mistake of failing to present the *concepts*. Instead, they tend to be very abstract system diagrams. A very simplified view of how real-life systems interact is unlikely to give an adequate and consistent basis for forming a measurement model. Real-life system architectures will always deviate from any attempt to describe typical systems operations. Instead, it is preferable to start with a clear view of the principles, then map these to real-life systems. That way, the concepts always remain true and a consistent basis for a measurement approach.

The table below gives a simple practical analysis of the different implications for each kind of error.

	Metering Error	Conveyance Error	Calculation Error
If sufficient raw data is retained, will it be possible to reprocess it and hence correct the errors identified?	No	Yes	Yes
Could this kind of error be identified by customers checking itemised bills against agreed tariffs?	No	No	Yes

Table 1: The implications of errors using the MCC classification

The impact on the customer of each kind of error is discussed further below.

3.5 Understanding consumer protection issues with the MCC Model

The following diagram illustrates how telecommunications metering and billing differs from many other industries and how this changes the nature of the billing relationship with the customer. In telecommunications, a number of processing stages exist between the meter's output and its "display" in the form of a bill. This can be compared to a gasoline pump, a gas meter or a grocer's scale, where the meter output is displayed directly in front of the customer. In many industries, the meter's output is exactly what is displayed to the customer. In telecommunications the meter output is but a data stream that needs further processing in order to be "displayed" in the form of an itemised bill.

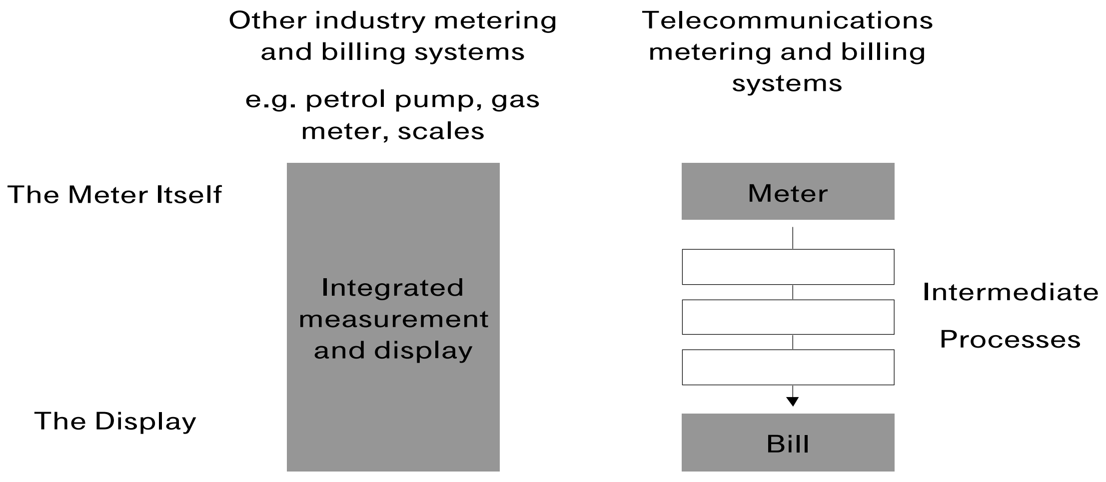

Figure 2: The link between meter and display in telecommunications and other industries

The following diagram illustrates how the customer's knowledge of the meter output is dependent upon the correctness of processes that occur after the meter. This is different to industries where there are no intermediate processing stages or where intermediate processing is trivial.

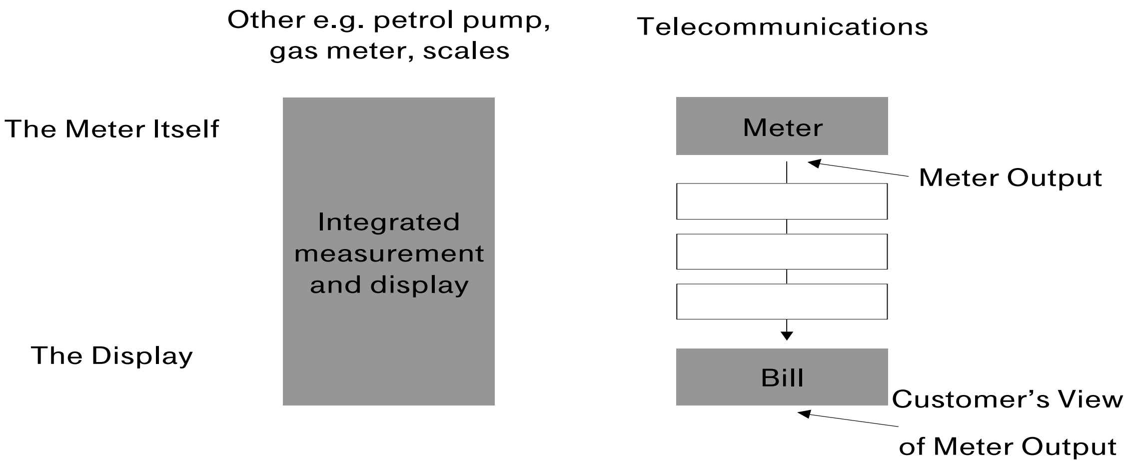

Figure 3: The difference between the meter output and display

The following diagram illustrates how conveyance and calculation can be linked in the mind of the customer, and in measurement. However, measuring each separately is still good practice. The causes of conveyance errors and calculation errors are likely to be distinct, and have a different impact on the customer. Whilst a customer will generally be able to identify a calculation error if they receive an adequate itemisation, they are unlikely to be able to identify a conveyance error. Furthermore, they will not be able to distinguish a conveyance error from a metering error. This last point is important as conveyance errors may be corrected in a way that metering errors cannot.

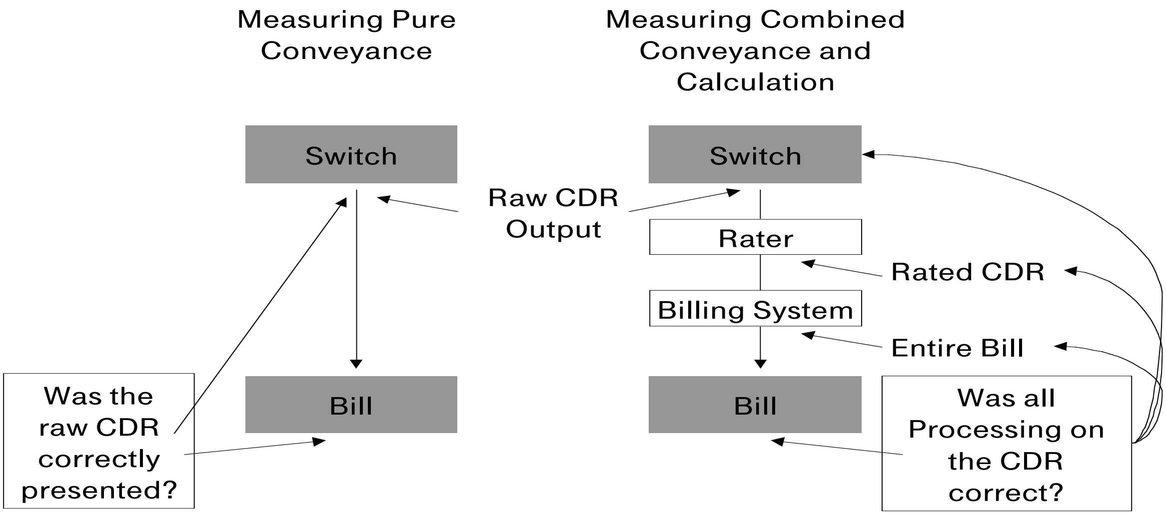

Figure 4: Measurement of conveyance on its own and with calculation

Conveyance is measured by determining whether any corruption occurs to the data produced by the meter as it passes through intermediate systems. With calculation, the issue is that the algorithms for calculating the chargeable value are correct, and that the customer and tariff data that support these algorithms are correct. Separating the two concepts ensures they are not confused and hence that they are both properly measured.

One complication occurs when the itemisation is not presented on the customer's bill. In this case, the operator can still perform the equivalent of a check of the bill itemisation. Instead of checking the itemisation itself, the data repository used to hold the meter data that would have been presented on an itemised bill is checked. It is important that some repository of this type is installed and maintained properly, because otherwise the business will be unable to respond to any queries regarding the charges it has presented. Such a repository is a precondition of analysing conveyance separately from calculation. It should also be considered an obligation in its own right, as the business will be unable to justify its charges if it did not possess even this minimal form of record keeping.

Consumer protection issues are intimately related to what information is presented to the customer, and when. If the customer can "see" the meter for themselves, they are able to double-check the calculations of any charges and only rely on the accuracy of the meter itself. To some extent they may even be able to observe obvious meter errors (perhaps the petrol pump reading continues to increase though nothing is being pumped). If the customer sees the meter output, but only at a later date, their task is made more complicated because they need to remember the actual events to compare this to the data presented to them. If the customer never sees the meter data then the onus is on the business to fully ensure that the charges are correct in every instance. In telecommunications, customers who do not receive a full itemisation are, in effect, unable to see the meter. To summarise, the customer's ability to protect themselves is related to:

- the level of detailed information they receive confirming the sales they are being charged for; and
- the timeliness of presenting that information.

The following table illustrates the point with some simple examples.

	Time of meter presentation	**Time of bill presentation**
Petrol pump	At time of sale	At time of sale
Gas meter	At time of sale	Later
Itemised postpay telco bill	Later	Later
Telco bill without itemisation	Never	At time of sale (prepay/real-time) Later (typical postpay)

Table 2: Timing of meter and bill presentation in different industries

The reason for not providing customers with itemised bills is usually cost or convenience. But note that there is good reason to balance the saving in terms of cost against the increased obligation to protect customers from error. Furthermore, because customers lack the data to check their own bills, there is a case for saying that the business has eliminated one very useful source of data regarding the correctness of its bills.

4. Black Box and White Box measures

The concepts of black box and white box testing helped with the application of scientific principles to software verification. However, the ideas of the black box and white box can also be applied to measuring at a more general level that covers metering and billing integrity.

4.1 A Black Box view of metering and billing

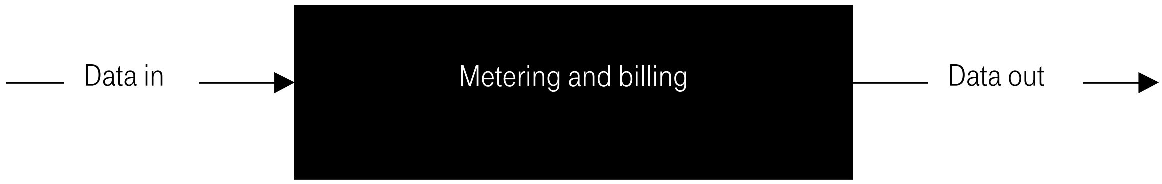

Figure 5: A Black Box view of things

A black box view is so-called because no knowledge is assumed about how the systems and processes being assessed actually work. It is as if they are contained in a black box we cannot see into, and all we know is what goes in and what comes out the other end. Hence all we can do to monitor is to compare what came out to what went in and our expectations of how the two should be related.

This means our definition of measurement comes solely from our expectations of how things *should* work, without knowing anything about how they *do* work. We defined metering and billing as

all cumulative activities that enable the sale of a supply to be recorded and then rendered as a charge to the customer

so if we know nothing about those activities, all we can do is base our measurement on whether the data output (the charges and any itemisation presented to the customer) is correct given the input (what was sold).

4.2 A White Box view of metering and billing

A white box view assumes a knowledge of not just of the ins and outs but also the details of workings within the systems and processes that lie between. Hence, the box is “white” because we see inside it. Because we can see inside, we can devise an approach based on the details of how things work in practice as well as end-to-end expectations.

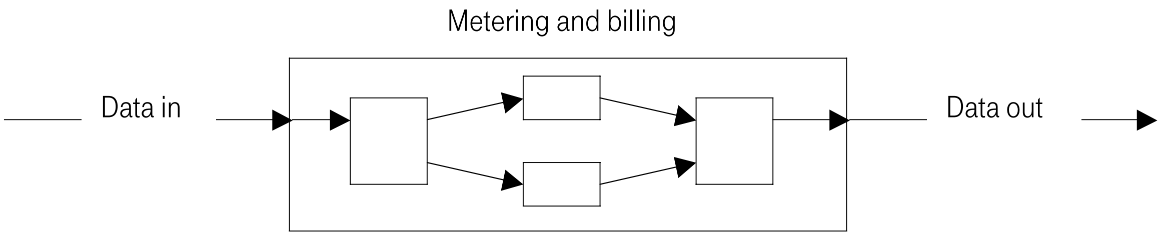

Figure 6: A White Box view of things

Using a white box approach, the definition of measurement can come from both how things *should* work, and how they *do* work. Instead of just measuring between the end points, it is also possible to measure between intermediate points and "add up" the answers. Measures can be devised to place more emphasis on likely errors given the way that things actually work.

The diagram above is obviously a gross simplification of the complexity in using a white box approach at the lowest level. The following diagram gives some feel for how many systems and processes are connected in a typical telecommunications business. Labels have been blanked in the interests of commercial confidentiality, but it should be obvious that it requires a significant level of resource to just to identify the relevant measurement points, and even more to implement measures across all of them.

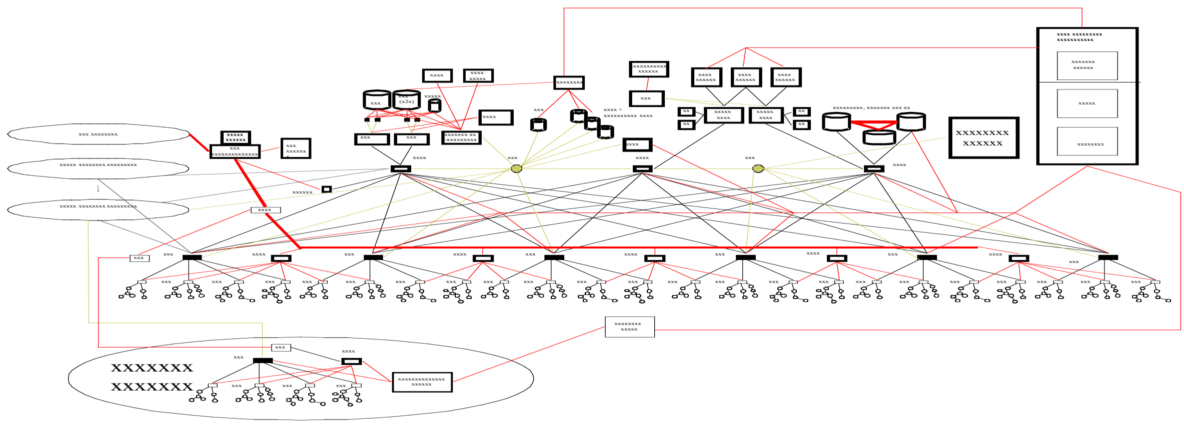

Figure 7: Diagrammatic representation of White Box complexity

4.3 Measurement in a Black Box or White Box context

In the diagram below, we see that a black box measure is based on just the main points "in" and "out". In a white box measure, we can measure performance between each of the points within the various systems and processes, then combine these to get a cumulative score. Both measures should give the same results. However, there are numerous practical reasons why they may give different results, which can be summarised as follows:

- an incomplete white box approach may miss some of the steps necessary to get a truly complete measure;
- a black box approach may be overly simple and hence exclude certain real kinds of error from measurement results; and

- a white box approach may not correctly weight each of the results before they are combined, exaggerating the importance of some results but undervaluing the importance of others.

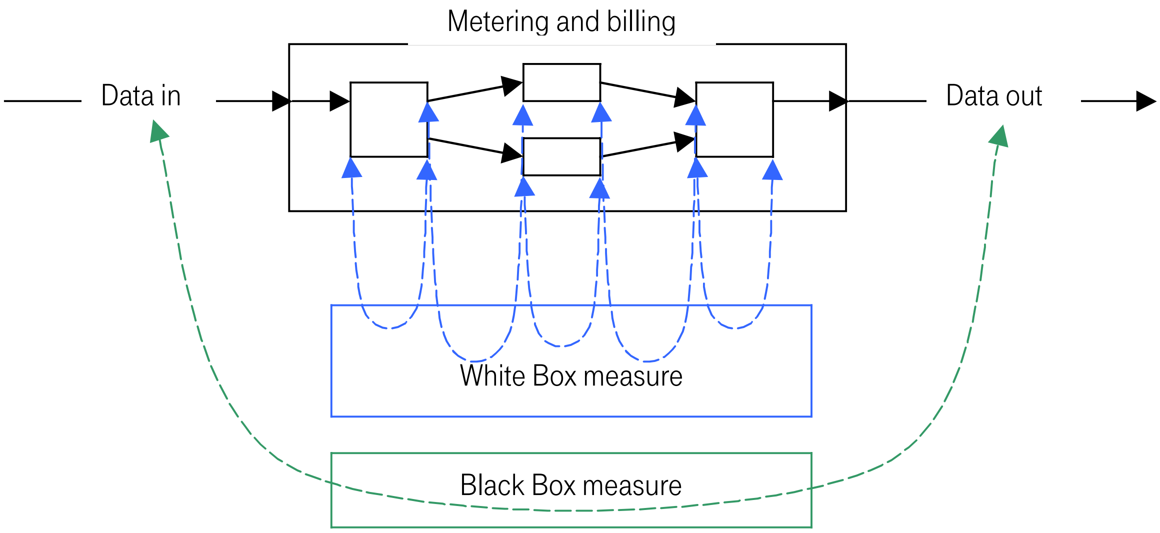

Figure 8: Black Box and White Box measurement points

The causes of discrepancies between the two approaches are discussed further below.

4.4 Relative merits of Black Box and White Box measurement

4.4.1 Advantages of Black Box measurement

A black box approach is:

- easy and quick to understand and audit because it requires no detailed knowledge of real-life systems or of relative risks;
- easier to maintain because it can remain valid even when the measured real-life metering and billing systems change;
- less likely to omit relevant data than a white box approach;
- better at providing high-level metrics without the need to appropriately weight and combine data from many sources;
- easier to reproduce in different contexts because it is not related to the underlying systems;
- a useful high-level “barometer” especially if the results are produced on a timely basis;
- not capable of giving a misleading “partial” measure because of gaps in the data;
- not subject to some of the errors that may occur with a white box method because it does not require a comprehensive understanding of systems and processes;

- easier to adjust for changes in tariff policy, service offerings, and customer behaviour; and

- intuitively easier for a customer or auditor to understand, because the measurement is based on what *should* be charged given what *was* supplied, irrespective of *how* this is done.

4.4.2 Advantages of White Box measurement

A white box approach:

- gives useful information even when incomplete;

- can be assembled by combining existing sources of data and filling gaps rather than implementing a completely new approach;

- permits effort to be focused on likely causes of errors instead of just providing a lot of generally uninteresting and repetitive results;

- puts less onus on devising a comprehensive understanding of tariffs, customer behaviour and other factors that alter the relative importance of errors across the whole metering and billing chain;

- is less subject to macroscopic errors due to inappropriate or unrepresentative samples;

- should encourage a sufficiently complex view of how metering and billing actually works that there is less risk of errors due to oversimplification;

- is more useful in diagnosing the cause of problems because it works at a lower level;

- forces a thorough understanding of how systems work and are integrated;

- forces a thorough understanding of where the greatest risks lie;

- will provide data of more immediate use and relevance to specific parts of the business.

4.4.3 Conclusions on which to use when

In general, a black box approach is very well suited to high-level measurement across a whole business or chain of related platforms. The simpler the tariff plans, customer options, customer behaviour and the like, the easier it will be to use a black box approach to assess the relative impact of error in a simple and intuitive fashion. Black box approaches become less useful when the nature of metering and billing is so complicated that it is no longer possible to devise simple end-to-end measures. For example, the more complicated the variety of sales recorded, and means of recording them, or the more complicated the charges and means of rendering them to bills of different formats, the harder it will be to perform checks that cover a broad range of outputs.

A white box approach is better suited to a business that already has many of the relevant sources of data but needs to tie them together. It will be much more costly and difficult to implement, however, if there are many gaps in existing knowledge. If there are gaps in the knowledge, the white box approach is nothing more than an incomplete inventory of issues. Like any incomplete inventory, the total result will be understated, and no conclusion can be formed about the significance of items missing from the inventory. A real advantage, however, is that the data is more relevant to a given system or process, meaning it can be used in the absence of end-to-end measures and is better suited to diagnosing the cause of problems.

In reality, white box and black box are two ends of a spectrum. White-box tends to be low-level, giving better data at that level but making it harder and costlier to generate a comprehensive view of performance. Black box is high-level, which increases the risk of over-simplification, but also gives a more intuitive overall understanding of the relative issues in performance. The one caveat for black box approaches is that it will often be necessary to compromise to get working measurement technology and human processes because of the difficulty of putting a genuinely "end to end" measure in place.

4.5 Combining Black Box and White Box measurement

T-Mobile UK has adopted both a black box and white box approach to measurement.

- The black box approach is high-level and designed for end-to-end measurement on a continuous basis. As such, it is the definitive source for compliance measurement.

- The white-box approach is low-level and is designed to record each and every detailed issue identified from low-level monitoring, investigation, problem management and such. It is a reflection of a primarily diagnostic approach to integrity and the individual elements are subject to revision and update on different bases. It is recognised that this inventory will never be complete because of the diminishing returns in addressing what are *judged* to be increasingly improbable risks of error. As such, it is not a suitable source for reliable and recurring end-to-end measurement of use to an external auditor (who might reasonably question subjective judgements of risk in the absence of empirical data either way), but it is highly indicative of end-to-end performance.

The two measures are independently generated and then cross-checked for greater confidence. Although the white-box measure is not designed to be as precise or reliable a source for end-to-end measurement as the black-box measure, cross-checking has the following virtues:

- any oversimplifications in the black box model can be identified from the detailed line item analysis in the white box model;

- any gaps in the white box model may be made apparent by comparison to end-to-end totals in the black box model;

- one or other model may be better suited to rapidly cope with any given system or process change, so even where changes take place the time for which there is inadequate measurement is minimised; and

- major discrepancies may indicate flaws in either source of data or the means of calculating their impact in measurement terms, so comparison is a broad-span control over quality of data and appropriateness of measurement calculation.

The disadvantages of using both approaches independently relate to the increased cost and effort to the business of establishing essentially the same information. The black-box model is based on a very high-level view of the business with the fewest possible points of data analysis to minimise the effort required. The white-box model, on the other hand, involves a lot more detailed work to compile, though the actual figures are a secondary output from the detailed monitoring and investigative work that would be necessary anyway to maintain a well-controlled business.

For T-Mobile UK, the black box model is represented by the OTR003 compliance measures. These are produced on a monthly basis and issued to auditors and senior management for their review of overall business performance. The measures are in effect the output of a series of equations with variables populated from raw data obtained from high-level metering, conveyance and calculation checks. The equations enable the probabilities and significance of each kind of error to be appropriately weighted to give a representative total error. The white-box model is reported via what is known as the Revenue Loss and Revenue Gain Reports. These are also produced on a monthly basis and issued to senior management, most specifically to highlight critical and major issues and their impact on the previous month. These reports are made available to auditors for review as they find appropriate. The compliance measures and Revenue Loss/Gain reports are exchanged between the separate teams that produce them for cross-checking. A regular meeting is held to discuss any apparent discrepancies and their implications.

5. Applying the concepts to real life

5.1 Identifying the meter

As stated above, the meter produces the record of the sale that will be used as the basis of charging. To test metering, it is necessary to identify the meter for each chargeable sale. The following discussion covers some of the issues with correctly identifying the meter.

5.1.1 Only one meter for each event

Not everything that may be the meter is *the* meter. This is because more than one device may produce equivalent data that could be used as the basis of charging. The important point is that, where there is a choice, it still remains true that only one is actually intended to be used as the source of data for charging. Data from other systems, not intended to be used as the basis of charging, are irrelevant. However, such data may be used as a control over the quality of the meter, for example by reconciling the two outputs. For each event reviewed, it is necessary to identify which device produced the relevant data and check that output. The correctness or otherwise of data output from other devices is ultimately irrelevant to meter and bill integrity.

A simple example would be where a customer leaves a voicemail on an automated platform. The switch that manages the call, which is automatically redirected to voicemail, may produce a record of the call just like any other call. The voicemail platform may also produce a record of the delivery of the message. Either record may be used as the basis of charging, but just because they could be used does not mean they are being used. It is possible to imagine a case where the customer receives a charge for making a call (like any other) and for depositing a voicemail where the former is charged per the switch record and the latter per the voicemail record. This may even be presented on the bill as one charge for one event. In this case, it is best to treat this as two separate events, each with an independent and different meter. This still enables meter errors to be distinguish from conveyance and calculation (if the events are subsequently combined to create one charge, that should be considered an element of conveyance and calculation).

5.1.2 Only one meter but maybe more than one device

A system-based view may lead to the impression that one meter = one device. This is not precisely true. In the conceptual model, *the* meter produces *the* data. Though we talk about *the* meter it may be that more than one device produces data where all the cumulative data is needed to calculate the charge. This is not the same as the example above, because each device is needed for metering because each produces different data.

A simple example would be where a call takes place and the duration and start time are recorded by two separate devices. Call data is passed via an intelligent network device to a real-time billing solution. The start time for the call is recorded by the intelligent network device at the start of the call. That datum is conveyed to and stored by the real-time billing solution. The duration for the call is recorded by the switch at the end of the call and passed via the intelligent network to the billing solution. In this case, the device that records the start time and the device that records the duration are different, though both have their data output conveyed to a third device, which uses all the data to calculate the final charge. One source of confusion here is that the metered start time is produced by the intelligent network, which also conveys the metered duration. On the other hand, the switch may have been able to record both start time and duration, but in this instance the start time recorded by the switch is *not* that used for charging the customer. An attention to detail is needed to correctly identify which device produces which datum as part of metering. However, it remains true that for each datum needed to charge an event, only one device produces *the* datum actually used as the basis of charging.

5.1.3 Contingency meters

It is possible that, where a choice of meters exists, that one may be used as a substitute should the usual meter fail, for whatever reason. This is the same as saying that another device becomes *the* meter on a temporary basis. Any measurement must reflect even a temporary change of meters. If test devices have been deployed, it may be tempting to use these as contingency meters as they produce equivalent data for the same events. However, where they are being used as *the* meter, it is no longer true that they also produce a measure of meter correctness. The measure of meter correctness requires use of a second device independent of the meter. If a test device is temporarily used as *the* meter then it cannot also be considered to produce measurement data about its own correctness.

5.1.4 Timing of metering

Metering must take place at the time when a sale is made. However, this somewhat simplifies what might occur in real life, where the decision as to what was "sold" might not occur just at one point of time. It might be that a sale is recorded, then the details of the sale are altered or updated to take into account subsequent events. In this case, clearly all data needs to be recorded at the relevant point in time. This also means that output of several devices may be needed.

One simple example relates to service levels. Suppose a fixed line is leased on the basis of a guaranteed service level. Failure to meet service levels implies credits will be given to the customer. A sale is recorded when the line is procured. However, performance against the service levels is recorded on an ongoing basis. In other words, outages are recorded as they occur and then at some relevant point in time compared to service level commitments for determining whether a credit is due. So metering of what was in practice sold may occur at more than one point of time, because the original record of sale must be supplemented by data on subsequent service levels achieved.

A more complicated example involves a situation where a sale is recorded where it is known that the sale may not be completed satisfactorily and hence will need to be completely reversed. Although no sale takes place as such, two accurate pieces of data need to be produced at different points in time to correctly record this "non-sale". For example, consider a situation where a text message is recorded when the customer sends it. However, the customer is not charged per attempted send but per successful receipt. Some time later, a message is received by the network indicating the message was not successfully received. Metering has taken place at two points of time, by two separate devices. Furthermore, the devices may not have been operated by the same business. However, their correct cumulative operation is necessary for accurate metering.

A fairly academic question may be raised here: is the customer being charged for one "thing" in these cases? Or is that there is a charge and a netting refund, or two "things" that balance each other. The question may be thought academic as the net effect, in terms of value, is the same either way, and all devices that meter, at all points of time, need to work correctly to get the correct record of events for charging the customer. However, there are two reasons why this question might not be considered academic.

- The customer's perception of how many events took place may not be the same as the business'. In particular, even if netting figures have been correctly presented to the customer, it still is likely to lead to confusion if the customer does not expect to see netting figures. This is exacerbated if the netting charges are not presented at the same time.

- Where a target is stated in terms of number of correctly charged events, it clearly matters whether cases like these are counted as one or two events. It matters not because the question is interesting in itself, but because whatever rule is applied it must be applied consistently to get consistent measures. This is particularly relevant when dealing with expectations of third parties like regulators or commercial partners. A flippant view that sets targets per event, but fails to properly define what is one event, is only likely to lead to unnecessary dispute at a later date.

T-Mobile, for the sake of a consistent convention, works on the following basis.

- Where the intention is to present a single figure on a bill, then one event has taken place.

- Where a charge and a credit relating to the same service is presented separately, the charge based on certain meter output and the credit based on other meter output, then two events have taken place. An example would be a charge for a message and refund because delivery was unsuccessful as presumably these are based on different, unconnected data.

- Where a charge and a credit relating to the same service is presented separately, the charge and the credit based on the same meter output, then one event has taken place. An example would be a charge for an event and a pre-determined discount for that same event as this is just a presentational nicety and both elements are calculated from the same raw data.

5.1.5 A practical approach

The meter is the start of the processing chain. The paradoxical consequence of this is that to determine the meter, we need to first have an understanding of the other "end" of the processing chain, the final output in terms of the value. What needs to be recorded at the start is defined by how the charges are calculated. The best approach is hence to trace back chargeable events from the final presentation through the various calculations and intermediate systems back to the original sources for the raw meter data. Only by tracing backward from end to start is it possible to prove that all relevant meter outputs have been identified. Of course, this approach also ensures that no effort is wasted on irrelevant data outputs.

5.2 Identifying conveyance

The process described above for identifying the meter also ensures that the route by which records are conveyed is identified. The key points here for identifying the path for conveyance are as follows.

- There are many potential "end points" for a conveyance process, so to be sure of covering the right one, it is necessary to start with the end point of interest, e.g. the bill, and trace backwards methodically to the meter.

- Some processes involve the deliberate creation of duplicates or near-duplicates of records so they can be sent to different downstream systems (i.e. to meet different objectives). Effort needs to be methodically focused only on those versions of a record that are being covered by the intended measure. Whereas it may be practical from a diagnostic or controls point of view to cover all outputs from a system at the same time, this is not helpful for end-to-end measuring if it means effort is wasted on outputs that will not be measured any further.

- Conveyance is the transmission of relevant information, not of any and all data in any particular format. This means that attention must be paid to which transformations of data involve a change in the information (i.e. calculation) and those which just give the same information in another way (i.e. reformatting). New information added, where relevant to the final value, must be correctly conveyed from the point it is generated. However, some information may be added for reasons that are irrelevant to the final value of the charge. The correctness or corruption of this information is not part of measuring meter and bill integrity.

- Conveyance may be thought of at a "one record in, one record out" type level. This analogy is helpful but not always reliable. Individual records may be combined to make one, or one record may be split into several, all of which may need to reach their correct destination for the final charge to be correct.

- Though measurement of conveyance should be thought of at a "per event" level, data is normally conveyed in files which batch together large numbers of events. Many controls will hence focus on whether a file gets from one point to the next. However, this may obscure the underlying issues of conveyance. A file may be successfully transmitted without each record being successfully transmitted. Records may be omitted or corrupted, phantom records added to the file on its creation, the file may itself be corrupted or the file may not be properly "read" by the next system that receives it. Ultimately any measurement of conveyance must work at the level of individual records and not of files. Otherwise, certain kinds of real errors may go unmeasured.

- Loss of data is a typical conveyance error. Sometimes loss is real, whereas sometimes the data is not lost, but "caught" in a suspense or error log. Data is lost if it simply is not recorded anywhere, as may happen if memory buffers are exceeded, or where records fail a process but are not written to any suspense file or error log. No error has taken place in conveyance *if* records can be resubmitted and processed successfully. Per the conveyor belt analogy, the items came off the belt and fell into a bucket to be returned back to an earlier point on the same belt. If, however, the items are not resubmitted, then they are effectively lost and a conveyance error has taken place. This is still an error even if the result of a conscious decision, because if the records represented real sales, then the business has failed to charge for all its sales correctly. In effect, the commercial decision is to commit an error because that is more cost-effective than being accurate and because the customer's interests are thought to be unharmed.

- Some loss of data may be deliberate, and not all loss is an error. A device may output records that are known to fall into error or are where there is no intention to pass them to any downstream process. The important point is whether these *should* have been passed on, which is why working back from the end is a helpful practical way of viewing conveyance. Another form of loss would involve getting rid of irrelevant fields in a record. Again, there is no integrity issue if the lost field would not have affected the final value. The last form of loss is loss of precision, most notably in the form of rounding. Rounding is a perfectly acceptable form of loss of precision in data, so long as it is correctly understood. From a consumer protection standpoint, the customer should understand all of the ways rounding may affect the final value of what is presented to them. From a measuring standpoint, rounding needs to be allowed for as part of the conveyance model. For example, if a system is expected to round a certain way, and does not in a given situation, then the failure to round correctly is itself a form of data corruption.

5.3 Identifying calculation

As above, calculation is best identified by tracing from the end to the start. Calculation should be easier to pinpoint because it is unlikely to occur at more than one or two places. Because calculation usually involves looking up standing data, it should be fairly apparent where this occurs. It should also be evident where new data is being written to represent a monetary value. The complexity in measuring calculation relates to ascertaining that the standing data, as well as the algorithms are correct. See below for more.

5.4 Reperforming calculation

In order to reperform a calculation, it is necessary to:

- emulate all the algorithms used in calculation;
- base calculations on equivalent customer data; and
- base calculations on equivalent tariff data.

All three pose different problems.

5.4.1 Precise emulation of algorithms

Care must be taken that calculations are properly understood and are correctly replicated. Otherwise, apparent errors identified by reperformance will in fact represent limitations of the algorithms used in reperformance. A seemingly mundane question that follows from this is the decision as to what, exactly, is being replicated. There is in fact a choice:

- the algorithms as designed and implemented in real life; or
- the calculation as described to customers.

These should, of course, be the same. In practice, differences can occur. Choosing to emulate the calculation as presented to the customer means that the measure also covers any discrepancies between what is communicated and implemented. The additional difficulty with this approach is that any vagueness, inconsistency or ambiguity in what is presented to customers, or any subsequent misreading by those responsible for implementing the check, may lead to the implementation of an unreliable algorithm for checking. However, the identification and resolution of issues of this type is a useful exercise in its own right. If the alternative approach is taken, the algorithms used for measurement should not be a direct copy of the code used in real-life, as the results will always be identical. All that would then be tested is the data, not the algorithms themselves. It is still possible to cover consistency between the wording presented to customers and the algorithms implemented without basing the comparative algorithms directly on the wording. However, no “measurement” as such will be performed of the difference between what is stated to customers and implemented in real-life. Instead, a control will be needed ensure to prevent discrepancies from occurring, rather than to measure discrepancies as they occur. This is a sensible deviation from a purely measurement-based view of integrity. If consistency can be enforced using a preventative control more efficiently than errors can be detected, it is obvious that efforts should be focused on prevention. Consistency is enforced in T-Mobile UK by using a single formal procedure for the Pricing function to communicate tariff requirements to the departments responsible for their implementation and testing. These are all recorded in a single database. The requirements, known as Tariff Change Requests, are stated in a format to avoid any possible ambiguity. Consistency is established because:

- the departments responsible for implementation have an unambiguous specification to work with that accurately represents Pricing's view of the requirements; and
- the Pricing department works with the assistance of the Legal department to ensure that the wording presented to customers is an unambiguous and consistent statement of what was formally stated in the Tariff Change Request.

The chief drawback of measuring by reperformance is the effort needed to implement a second mechanism to reproduce calculations. Given that the activity is intended to find flaws in the usual algorithms, it will always be questionable whether it is efficient to try to reproduce them rather than to apply additional resources to improving the original. Though the algorithms need to be reproduced to enable measurement, it should not be assumed that they can only be reproduced using automation. A manual check of calculations is also a kind of reproduction. The best approach will vary depending on the relative cost and difficulty of implementing an automated or manual check. It will also vary depending on how flexible or varied the reperformance needs to be. This relates to how much variation there is in real life calculations and how often the tariffs change. The cost of implementing and updating an automated system needs to be compared to the cost of employing and training people. Though there will always be some degree of human error, errors will also occur with an automated approach if insufficient resources are applied to implementation or testing of the automated test device. A sensible compromise is to reproduce calculations on an automated level where it is easy to do so, and rely upon the flexibility of manual checking for other aspects of calculations. For example, all calculations may be reperformed using an automated, but limited, tool. Where an exception is reported, this may be manually inspected to determine if the difference reflects an error in the real life calculation or a limitation in the measurement device. T-Mobile UK has adopted a slightly different compromise. Basic usage rate calculations are checked on a sample basis using an automated device able to apply rates to events on a simple per event basis. The device is designed to be consistent with the specifications of tariffs in the Tariff Change Request database. Any discrepancies are examined manually to determine whether they are genuine or due to a limitation in the tool. Where calculations involve some form of allowance or discount in addition to simple rerating, these are checked manually on a sample basis.

5.4.2 Customer data for reperformance

In order to calculate a charge for an event performed by a customer, it is of course necessary to know something about how that customer is charged. To reperform, it is hence necessary to either:

- look up the relevant customer data, or
- somehow *know* in advance the relevant facts about that customer.

Knowledge in advance is possible where checking known test accounts. There is a risk where looking-up customer data that the data being looked-up is already incorrect, hence meaning the error in the actual calculation is being replicated. The risk can be mitigated by taking data from a different source to that used for the real calculation. This source should not be populated downstream from that used for the real calculation, or the same problem of replicating error may occur. An example of a suitable solution would be to interrogate any system that records the customer's order (including changes of tariffs) instead of the main customer database.

5.4.3 Tariff data for reperformance

Though tariff data and the code of any rating or billing system algorithms are separable, they are both necessary to completely describe how to calculate the charge for any given event or customer. The separation is somewhat artificial; tariff data enables flexibility, but potentially at the cost of processing efficiency. As such, there is no definitive split between how much a complete calculation is described in the code of a system and how much in the tariff data. On this same basis, there is no definitive split for any automated system doing reperformance. However, a good principle is to try to get a similar balance in the reperformance system as the actual. Of course, there is no such logical distinction where calculations are being reperformed manually.

6. Principles of calculating suitable samples

6.1 Calculating the sample size

In calculating sample size, it is always necessary to decide what is the acceptable risk that the wrong conclusion is drawn. That risk is also related to the degree of precision expected in measurement. Where there is more than one source of data being used to form a conclusion, then the cumulative quality of evidence needs to be sufficient to mitigate the risks of error. Samples are taken where there are checks of detail, so the size of a sample can be reduced where other controls are in place to prevent or detect errors.

6.2 Picking a representative sample

6.2.1 Which meters to test

It is not necessary to test each individual meter to form an opinion about meter performance in general. The key question is to determine a representative sample. This means, for the properties being tested, the sampled meters behave the same as others in the population. For example, if two devices are known to be identical in all regards relevant for their function, it is possible to draw inferences about the operation of both from the results of testing one. However, if they are fundamentally different, it would be inappropriate to base a conclusion on one using data obtained from testing the other.

If you start with entire population of meters, the question can be described as to whether:

- the entire population is homogeneous; which means that each meter can be considered identical with respect to the operations being measured; or
- the population needs to be segmented into a number of different subsets, each of them homogeneous.

It is only valid to draw inferences on meters homogeneous to any meters that have had their performance measured. So if the aim is to measure performance across a range of different types or builds of meters, it is necessary to measure performance of at least one meter in each sub-population. However, it is not necessary to do more.

Populations can be shown to be homogeneous through controls over consistency. For most meters, the principles to be established, and the controls to show that they are homogeneous would be:

- consistent hardware, shown through the same specification;
- consistent software, shown through controls over the installation of software to the production environment; and
- consistent configuration and operational settings, including any data input as part of configuration, shown through controls over the implementation and testing of new configurations and both automated and manual consistency checks between configurations.

It is of course easier and more beneficial to implement controls to ensure consistency rather than independently measure each meter as if inconsistency was expected.

6.3 Handling unrepresentative samples

In any situation where sampling is performed, there is always a risk that the wrong inference is drawn about the integrity of metering and billing because the sample did not fairly represent the population. This can be mitigated by increasing sample sizes or calculating the impact of known bias or unrepresentativeness of a sample. There is no all-purpose approach to mitigating these kinds of risk, which are intimately related to the details of what is being sampled, how the sample is taken and the purpose of sampling. T-Mobile UK has defined its own detailed procedures to manage instances where sampling error is believed to have occurred in samples used for measuring integrity. This is based on two general techniques to confirm and rectify sampling error:

- increasing sample sizes; and

- extrapolation by forming mathematical connections between known quantified properties for the population and the property that is being tested by the sample.

Increasing the sample size is the simple and conclusive means of determining if the original sample suffered from sampling error. However, it may sometimes be quicker to infer sampling error by observing that other properties, related to the property being measured, are not correctly represented. An example of the latter would be concluding that a certain error in calculating discounts is over-represented because the sample taken over-represents the number of customers entitled to that discount.

7. Determining a test and measure plan

7.1 Metering

7.1.1 Completeness

This is the error of failing to record a sale. It is measured by checking whether a known sale resulted in a record of that sale.

7.1.2 Validity

This is the error of recording a sale when no sale took place. It is measured by checking whether a recorded sale can be substantiated by evidence of a real sale.

7.1.3 Accuracy

Accuracy is the concept that the record generated of an event correctly describes that event. What needs to be recorded, and hence the accuracy expectations, depends on how the charge will be calculated. Here are some common examples:

- Originating party
- Receiving party
- Event type
- Start Time
- Duration

Below is a discussion of issues in measuring duration. Much of it is practical in nature, and it is likely that similar practical analysis would be needed for the measurement of other properties where the final charge is proportional to that property. This is because the necessary accuracy of measuring the property prescribes the accuracy of the eventual charge, and so is intimately related to both price and precision of technology.

7.1.3.1 Defining duration

The duration of a call is the difference in time between the time the call commences and the time the call terminates. Beyond this, there are many possible ways to define the duration of an event because of the choice of what is said to be the start and end of a call. For the purposes of metering and billing integrity, there are two generic extremes.

Input to the meter

A call commences when the answer signal has been acknowledged at the meter.

A call terminates when the clear signal from the first customer to clear has been acknowledged at the meter or when the meter recognises that the call is to be cleared for whatever reason.

User perception

The period of time that the actual service is available to the user.

Implications of the choice

There may be a difference between the call duration and the actual time for which conversation is possible. Such a difference could arise, for example, due to delays between when a signal is sent and received.

An estimate of the average transmission latency and its variability may be necessary to reconcile the user's perception and what is recorded by the meter.

Certain types of delay may be introduced by deliberate design. Commonly these will be intended to ensure that a call is not terminated prematurely. It would not be legitimate business practice to implement delays for the purpose of increasing revenue.

The choice used for compliance measurement purposes

T-Mobile UK has chosen a definition of duration based on input to the meter as the basis for measuring its compliance to OTR003: 2001. It was chosen for the following reasons:

- Environmental variances may cause a difference between measured duration recorded at the meter and at the customer. Customers must be protected from deliberate abuse of environmental factors, such as introducing transmission delays merely to extend recorded duration. However, it was not intended that the scheme be designed to set expectations for an acceptable range of transmission delay that is inherently prior to the meter. The scheme sets clear targets for the actual metering device on the assumption that appropriate diligence is applied in ensuring signals are transmitted promptly. No separate target has been stipulated for pre-meter variances due to signal transmission delays. Such a target could be set (at least one national regulator has explicitly stated maximum variances permitted due to delays in transmissions). However, the current OTR003: 2001 regulations were not written in this way.

- It follows from the above that an operator still has a general duty of care to promptly transmit signals. However, it is not possible to set numerical targets which would be both general enough to apply to all cases, stringent enough to challenge operators to do their best in cases where delays would be short but loose enough to allow for real-life limitations of technology and distances. As such, the OTR003: 2001 standard is read as setting targets explicitly for the meter itself rather than targets intended to cover metering error and pre-meter signal variances. This decision is based on the observations that not only is there explicit mention of pre-meter delays in the rules, but that no analysis was performed as part of setting the targets that would have related to what might be the different signal delay expectations in different circumstances. Technical information in the public domain suggests that these delays are non-trivial at the levels of precision implied by the OTR003: 2001 standard, and that they vary depending on the call scenarios, the technology involved and whether the call involves more than one network. This means that if the intention had been to cover pre-meter signal delays, it would have been appropriate to set different targets for different scenarios in order to set an equitable compliance burden.

- An alternative reading of the OTR003: 2001 obligations is that the stated meter tolerances do cover pre-meter signal delays, but only *after* the impact of that delay has been allowed for in the calculations. In other words, a certain leeway is given for the pre-meter delays. This leeway may be determined from rigorous testing in real-life. This could be used to calculate a range that the "actual" duration is known to be within *after* taking the anticipated range of delay into account, and this calculated range of "actual" duration is compared to the targets set for the meter. We believe this approach would be flawed for two reasons. First, the approach is essentially the same as ignoring the pre-meter signal delay, as it first subtracts an additional margin for error (in effect an additional tolerance) and then compares the calculated result to the meter accuracy expectations. Second, because the approach is based on calculating the margin for error in practice, it still lacks a clear prescription of what is an acceptable margin for error and is based on current levels of actual performance instead of an assessment based on a stipulated target for performance.

- It is reasonable to observe that it is highly unlikely that a typical telecommunications business would design its core networks to produce small but definite delays in signals. If the intended net effect is to increase the duration, the delays cannot be consistent. To systematically increase duration, the delay on the clear signal must be *maximised* relative to the delay of signals to establish the call. Prescriptive regulation and audit of signal delay should hence be considered unlikely to be of benefit to customers. At most, a very occasional ad hoc practical test of signal delays would meet the same objective without the need for systematic and repeated measurement. In other words, it would be inconsistent and onerous to simultaneously argue for a regulatory regime to systematically and repeatedly test signal delay variances whilst acknowledging that the real-life risk to customers stemming from these variances is small or negligible.

- Because of all the above, it is rational to choose a definition of duration that most closely represents the issues relating to the actual meter's accuracy. It also is appropriate to choose a definition that supports precise measurement using automated devices by minimising the need to reconcile the duration per the meter and the duration per the measurement device.

T-Mobile has hence specifically identified that the definition of duration used for measurement is based on the actual time when relevant signals are received by or sent from the actual meter. See below for more on how the choice of measurement devices corresponds more or less precisely to each different definition of duration.

7.1.3.2 Measuring duration error

Note that there may be a difference between each of the following three values:

- the call duration, as defined;
- the call duration as recorded by the meter; and
- the call duration as measured by an independent test or measurement device.

The purpose of measurement is to gain assurance over the accuracy of the call duration, though there is no empirical data that directly states what this is. The call duration can only be inferred from one or other of the two sources of actual data i.e. the meter or from an independent test or measurement device.

The aim of a measurement strategy that addresses duration error must be to compare the meter recording to an independent device. To avoid the impact of environmental factors, and errors due to the comparison device, the actual device should be designed to give the closest possible correspondence to the meter. This is necessary to maximise the assurance derived from practical, empirical observation of systems working in practice and minimise the reliance on theoretical explanations of the differences between the results.

7.1.4 Measurement devices

There are a variety of different kinds of measurement or test devices that may be employed to similar ends. Rather than listing all the kinds of devices, it is more productive to contrast the main differences in design.

7.1.4.1 Invasive and passive

Invasive devices create an event to check that meters work correctly. They record data about the event by virtue of creating it. Passive devices monitor real events, giving an independent output for comparison to that of the meter. They record data about events that they do not create.

7.1.4.2 Near meter and near customer

Near-meter devices are located and designed to give similar output to the meter. Near-customer devices are located and designed to accurately record what the customer perceives. Devices that are near to the meter are to be preferred for measuring accuracy of duration.

Different devices may be regarded as “near” or “far” from the meter. These terms are used to reflect not just geographical proximity, but similarity of functioning. The following devices should be considered near to the switch:

- passive SS7 probes situated near the meter and used to monitor signals; and
- invasive test generators which directly interface with meters.

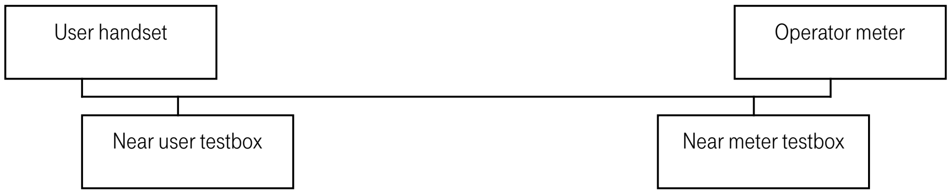

Figure 9: Near user and near meter measurement

7.1.4.3 Hybrid measurement architecture

T-Mobile UK has implemented both invasive near-customer test probes, and passive near-meter probes to analyse performance from both perspectives. Because of the relative imprecision introduced by environmental factors in using near-customer test probes, the near-meter probes are used for measurement. This measurement is then corroborated using the results from the near-customer probes.

7.1.5 Demonstration that devices are fit for purpose

No measurement approach will give satisfactory results, if the devices used introduce errors because they are

- flawed (the results they give are wrong even per the device’s own specification);
- imprecise (the results given are correct to a certain level, but not precise enough in the exact context they are used); or
- poorly aligned to measuring the exact property for which they are used (the results are correct and precise but the device’s implementation and specification gives inherently different results to the properties being measured).

As such, sufficient care must be taken to address all three principles. In other words,

- the device must be tested to determine if it works as specified;
- the precision of the device must be checked both by checking the specification is adequate and by testing that it is met in practice;
- the device must be chosen on the basis that it is well aligned to the properties being tested, or at least that any inherent difference is quantifiable and manageable.

Though appropriate steps should be taken to address these objectives, the nature of measuring meter performance lends itself to an additional, and generally more powerful form of analysis on how well the measurement device is functioning. The basis of measurement of the meter is comparison to the independent device. As such, measurement is nothing more than repeated comparison of the output of the meter to the output of the measurement device. The measurement device is, somewhat arbitrarily, considered to give the "definitive" result for the purpose of measurement, and differences are wholly attributed to the meter. This is a sound approach for the purpose of measurement, as it assumes that all differences are due to the meter, so tends to report the "worst case" for the level of metering error. This is not quite precise, because both the meter and the measurement device are producing an output that represents an ideal e.g. the actual time a signal is sent or received (as opposed to the time recorded by either device). There are hence two separate real variances between the ideal or "actual" property and each device:

- there is a variance between the actual and the meter; and
- there is a variance between the actual and the measurement device.

As stated above, the cumulative difference is, for measurement purposes, assumed to be wholly attributable to the former variance i.e. to the difference between actual and meter. In other words, the latter variance is taken to always be nil. Of course, the real variance is a combination of the two and the latter variance may in reality be a non-trivial component.

The two variances may be correlated or not. If correlated, then:

- the specification of the operation of the meter is more similar to the specification of the measurement device than it is to the definition of the actual property reported by both; or
- the devices are not independent.

Independence should be readily demonstrable from the specifications of the devices. It is not in general likely that devices may be accidentally inter-dependent without that being understood as part of the implementation of one or other. So the likelier risk is that the specifications are more similar to each other than either is to the property being measured. For this reason, care is needed to define that property in a realistic way that can be used as the basis for assessing real performance. In effect, the similarity of the two specifications is a desirable quality, and if the property measured is different to the two outputs, then the difference between outputs and property is intrinsically environmental, as both devices are performing as intended. Clearly, when setting an objective it is pointless to set an ideal property that cannot be precisely measured either by the meter or a test device. On the other hand, if quantifying the idealised property is an aspect of protecting the customer, then the tolerances for performance must reflect any known environmental factors that would cause variances between meter output and the idealised property. It should also allow for environmental factors between measurement device and idealised property. This is because the only real comparison taking place is between the test device and meter. The idealised actual property is not itself a datum in any real analysis, and so consumer protection based on this property can only be hypothetical.

All of this discussion is, to a large extent, an over-complication of the real problem. The theoretical problem is to monitor the variance between meter and actual by using a measurement device whilst allowing for the difference between actual and measurement device. However, from the above it should be clear whether the meter and measurement device is independent. If independent, then the appropriate course of action is:

- compare the two outputs;
- if different, analyse the reason why as far as possible;
- if not different, take that as evidence that the meter is accurate.

Both the meter and measurement device may be different to the actual property in the same way and to the same extent. This is not a limitation on testing the meter, but rather a limitation on testing the idealised "actual" property. The meter can be shown to work correctly per its specification by the measurement device without needing recourse to this idealised third point. If it cannot be shown to work correctly per the idealised third point, and no data can represent this point, then the problem is one of *definition*. In other words, the measurement can never state performance versus the expectations as *defined* because of the way they are defined. The practical solution to this problem is to alter the definition to fit practical reality, not to take practical steps to meet an unattainable ideal.

7.2 Conveyance

Conveyance is checked by ensuring that all meter output is correctly presented to a bill. The check is, in effect, a simple reconciliation between the two. This reconciliation may be performed by

- reconciling a selection of records 2 ways from meter to bill and bill to meter; or
- checking a selection of records per the meter are represented on the bill, and checking a *different* selection of records per the bill were recorded as at the meter.

Performing two separate 1-way checks may attain the same goal as performing a single 2-way check. The choice of whether to implement a 2-way check or two 1-way checks depends on convenience as they are theoretically equivalent. Implementing two 1-way checks does avoid any risk of bias in selecting the records to be checked. Bias may occur in a 2-way check if the records are exclusively selected from one of the two points being reconciled.

7.3 Calculation

Calculation is checked by reperformance. The issues in reperforming calculations are described above.

8. Real-life issues and obstacles

8.1 Producing top-level metrics

It is vital that top-level metrics can be traced back to the various kinds of raw data used to produce them. It is also very likely that to obtain such a top-level metric it will be necessary to establish a series of equations to correctly combine the data. This must also appropriately "weight" data where it combines data sources relating to different populations, and representing tests with different sample sizes, into "overall" measures.

The means of compiling and calculating the top-level metric should be designed to be reviewed easily. This is because a relatively simple error may make the metric useless or misleading. Simplifying the process of review makes it easier for the calculations to be checked by a number of individuals, hence reducing the risk of errors going unnoticed. It also should make it simpler to understand the impact of any necessary changes in the calculation, such as would be required if it needs to incorporate a new source of measurement data.

8.2 Handover of calls between base stations

We have not identified any issues relating to the handover of calls between base stations that would affect meter performance.

8.3 Impact of rounding meter duration and start time output

Rounding the meter duration and start time records means that there is a step-change in the degree of precision in the meter output. This also means that any measurement of meter accuracy has an absolute limit in its precision. Below this limit, it will always be unclear whether the meter was accurate or inaccurate "before" rounding.

Rounding rules need to be properly built in to any process to measure accuracy of recording. The simplest approach may be to round test output in the same way as meter output, but it must be recognised that rounding may accentuate small differences.

8.4 Practical factors when devising a measurement approach

In implementing measures, the following should be considered:

- the cost of implementing a measure should not outweigh the likely benefits to the business and customers;
- existing operational measures should be utilised where available before seeking to implement new measures;
- existing expertise should be utilised when interpreting results to get the best quality of analysis;
- the gathering, interpretation and communication of results should be of a timely nature; and
- the measurement strategy should be reviewed on a regular basis (at least annually) to account for any developments and the ongoing analysis of results from the measures.

Confidential appendices

Deleted from this version.

Document history

Issue	Date	Details
1.0		First issue
1.1	4 April 2003	Amended in response to Martin Gainey first review
1.2	12 May 2003	Edited ready for public distribution by removing sensitive information
1.3	2 June 2003	Miscellaneous typo's corrected, figure 9 added
1.4	4 August 2003	Reworded for change in regulation name from OTR003: 2001 to M&B Direction
1.5	5 November 2003	Additions made to incorporate SMS measurement and combined reporting
1.6	8 January 2004	Additions made to outline rationale behind ratio selection
1.7	21 January 2004	Clarification of document through amended appendices and title to avoid confusion when read by auditors

---

Title - The T-Mobile UK measurement strategy for meter and bill integrity of public electronic communications services

Owner/author (signed)

Name (block capitals): Eric Priezkalns

Approved by (signed) Manager

Name (block capitals): Anthony Sani

Ownership & confidentiality

T- Mobile is the trading name of T-Mobile (UK). No part of this document may be disclosed orally or in writing, including by reproduction, to any third party without the prior written consent of T- Mobile (UK). This document, its associated appendices and any attachments remain the property of T - Mobile (UK) and shall be returned upon request.