Performance measurements according to ITU-T Q.3930 and considerations in benchmarking

Why are we here today?

We usually end our presentations with this statement by Lord Kelvin

"If you can't measure it,
you can't improve it"
- Lord Kelvin

"If you can't measure it,
you can't compare it"
- Michael Mild

Why is Benchmarking more and more requested?

Why is benchmarking more and more requested?

Benchmarking at Telecoms (Singtel today), Singapore 1984

Benchmarking of IT systems has come far during the 35 years since this happened!

Why is benchmarking more and more requested?

In the connected world excellent performance is a necessity!

Performance is the most visible asset of a product or service, since it is experienced in every interaction!

Products are judged by their performance in the connected world.

Products with best performance are always challenged by competitors, i.e. a product cannot rest on old performance merits.

This is where benchmarking begins!

Background to ITU-T Recommendation Q.3930

Background to ITU-T recommendation Q.3930

It all started with the need for a generally accepted terminology!

If you ask ten people what performance testing is about you are likely to get at least five different answers.

In many cases terminology in performance testing reflect *how test tools are used* rather than *what system characteristics are actually measured*.

An example:

Is it *Load tests*, or is it *Capacity measurements / Responsiveness measurements?*

The purpose of ITU-T recommendation Q.3930 is to create a standard for terminology and concepts in performance testing where we all talk the same language.

The terminology in Q.3930 is used in this presentation.

Structure of
ITU-T Recommendation Q.3930

The structure of ITU-T recommendation Q.3930

The document covers most aspects of performance measurements

The document contains a set of terminology and description of concepts in performance testing to establish a common base for discussions about performance and performance tests in the following sections:

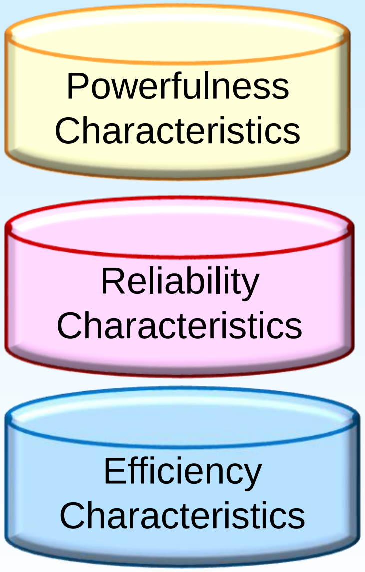
- *Section 6: Categories of performance characteristics.*
- *Section 7: Measured objects and service characteristics.*
- *Section 8: Requirements on metrics and collected data.*
- *Section 9: Abstract performance metrics*
- *Section 10: Performance data processing*
- *Section 11: General performance test concepts*
- *Section 12: Performance test environment*
- *Section 13: Performance test specifications*
- *Section 14: Workload definitions*

Categories of Performance Characteristics

Categories of Performance Characteristics

The standard Q.3930 introduces three categories of performance metrics

A system's performance can be described by an endless number of metrics.
To enable more focused performance characteristics we introduced three categories:


Powerfulness Characteristics
Reliability Characteristics
Efficiency Characteristics

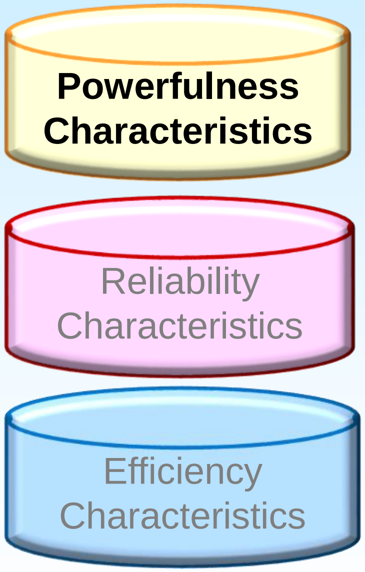
The performance categories are:

1. Powerfulness characteristics
2. Reliability characteristics
3. Efficiency characteristics

Categories of Performance Characteristics

Powerfulness characteristics

The Powerfulness category contains metrics describing the *delivery limits* of a measured system, i.e. metrics for how much services can be handled and/or how fast produced services can be delivered.



Powerfulness Characteristics
Reliability Characteristics
Efficiency Characteristics

Powerfulness characteristics has three groups of metrics:

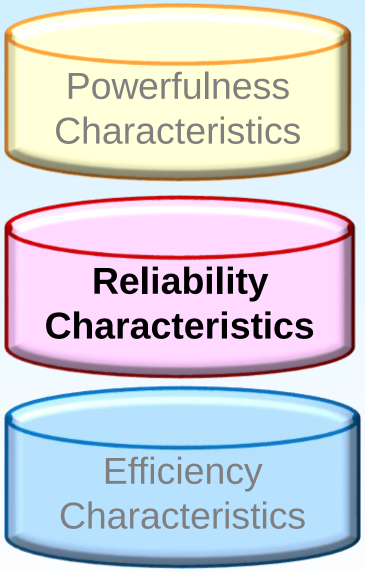
1. Capacity metrics
2. Responsiveness metrics
3. Scalability metrics

***Note!** Metrics in the powerfulness category are always relative to the used platform!*

Categories of Performance Characteristics

Reliability characteristics

The Reliability category contains metrics describing the *conditional limits* of a measured system, i.e. metrics for maintained service levels.



Powerfulness Characteristics

Reliability Characteristics

Efficiency Characteristics

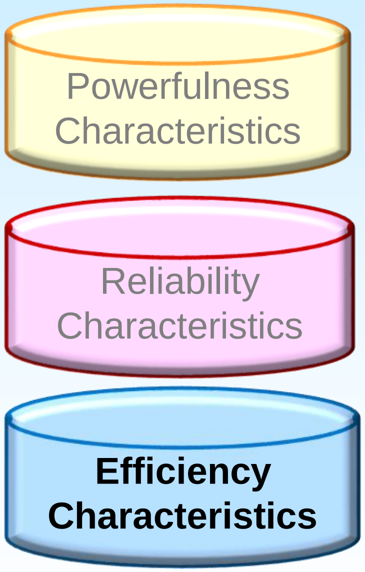
Reliability characteristics has five groups of metrics:

1. Stability metrics
2. Availability metrics
3. Robustness metrics
4. Recovery metrics
5. Correctness metrics

Categories of Performance Characteristics

Efficiency characteristics

The Efficiency category contains metrics describing the *productivity limits* of a measured system, i.e. metrics for required efforts of produced services.



Powerfulness Characteristics

Reliability Characteristics

Efficiency Characteristics

Efficiency characteristics has seven groups of metrics:

1. Service resource usage
2. Service resource linearity
3. Service resource scalability
4. Service resource bottlenecks
5. Platform resource utilization
6. Platform resource distribution
7. Platform resource scalability

Requirements on Metrics and Collected Data

Requirements on Metrics and Collected Data

General objectives for performance metrics

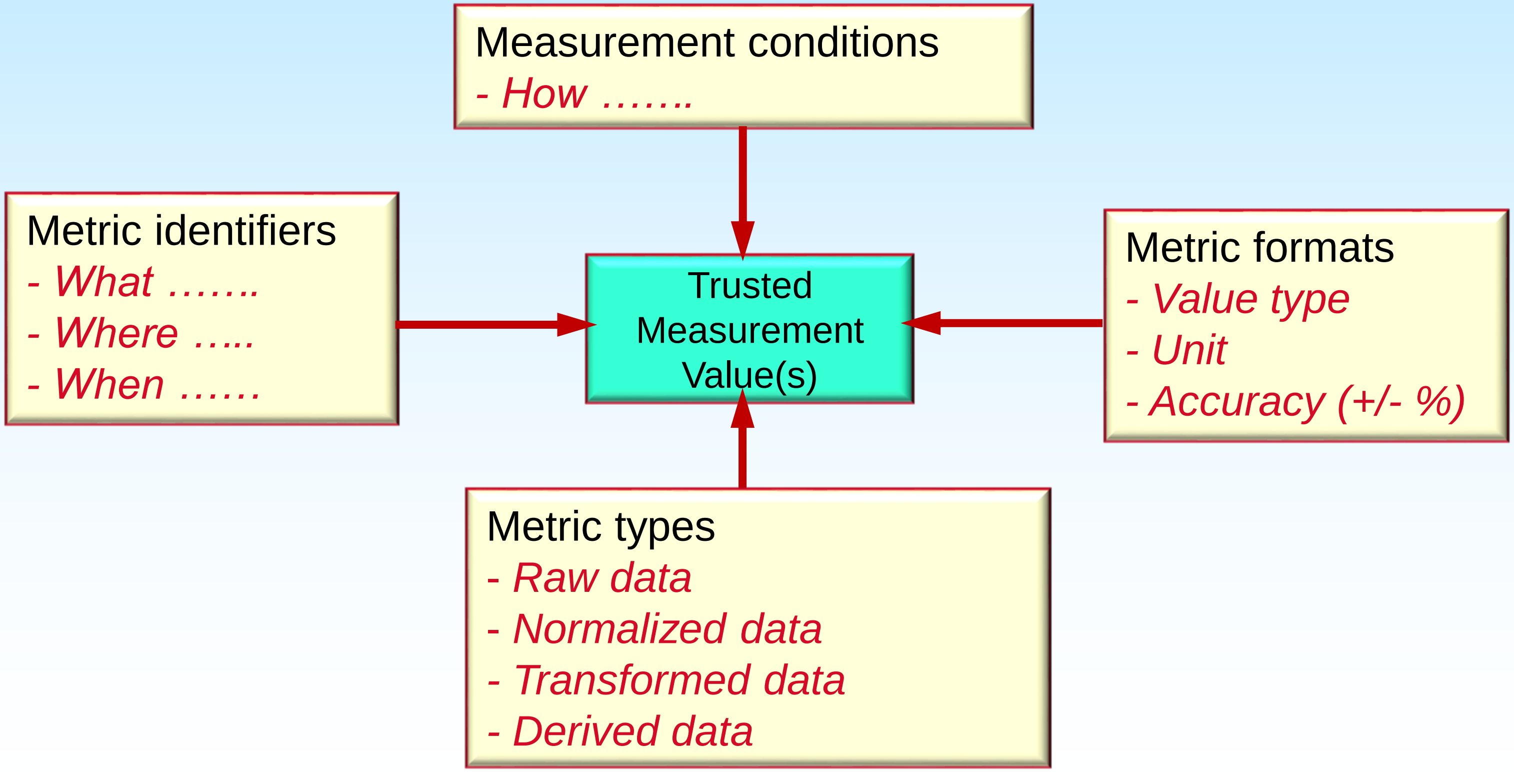
Good performance metrics must comply to a set of requirements:

1. *Understandable*, i.e. are the metrics easy to interpret?
2. *Reliable*, i.e. are produce values always in accordance with real values?
3. *Accurate*, i.e. are produced values precise or very close to real values?
4. *Repeatable*, i.e. are produced measurements values possible to repeat?
5. *Linear*, i.e. are produced figures proportional to the changes in the tested object?
6. *Consistent*, i.e. are metric units and their definition the same on tested platforms?
7. *Computable*, i.e. are metric units and definitions, measurement accuracy, and measurement methods precise enough to enable computations?

Requirements on Performance Data

What makes performance figures trusted and comparable?

Four groups of attributes describing collected performance data are required to make performance characteristics trusted and comparable.



Measurement conditions
- *How …….*

Metric identifiers
- *What …….*
- *Where …..*
- *When ……*

Trusted Measurement Value(s)

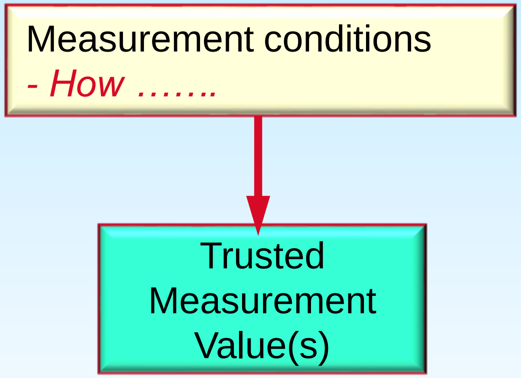
Metric formats
- *Value type*
- *Unit*
- *Accuracy (+/- %)*

Metric types
- *Raw data*
- *Normalized data*
- *Transformed data*
- *Derived data*

Requirements on Performance Data

1. Measurement conditions

Measurement conditions describe what must apply when data are captured.


Measurement conditions

- How …….

Trusted Measurement Value(s)

There are two kinds of captured performance data:

1. Data that is, or is part of the requested performance characteristics
2. Data that shows actual status of required conditions when requested performance data were collected!

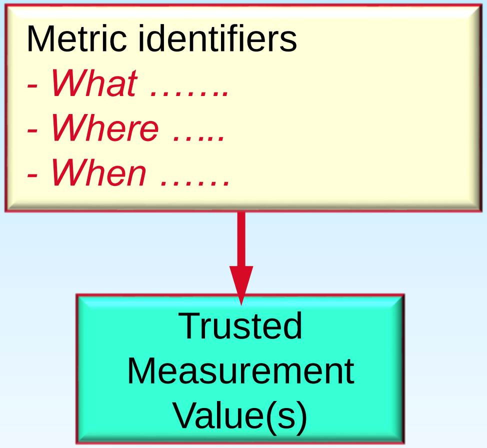
Recorded data about actual conditions during the performance measurements are extremely important for the validity of captured performance data.

If you can't verify that presented performance data were collected under requested conditions they can't be trusted and are consequently worthless!

Requirements on Performance Data

2. Metric identifiers

Metric identifiers provide basic information about collected performance data.



Collected performance data must have identifiers for :

1. What they represent
2. Where they were captured (logical or physical locations).
3. When they were captured (during the performance measurement).

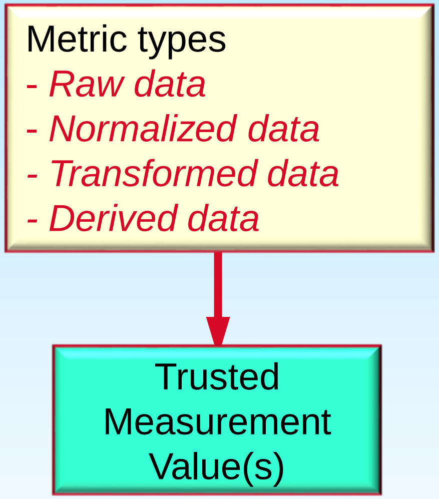
The need for metric identifiers applies both for requested performance data and required condition measurement data.

Without metric identifiers performance data don't represent anything!

Requirements on Performance Data

3. Metric types

Metric types provide information about the kind of performance data.



Metric types
- *Raw data*
- *Normalized data*
- *Transformed data*
- *Derived data*

Trusted Measurement Value(s)

Performance data belong to one of four type:

1. Raw data, such as response time.
2. Normalized data, such as transactions *per second*.
3. Transformed data, such as memory usage in Mbyte transformed to memory usage as percentage of total memory.
4. Derived data, performance characteristics resulting from a well defined computation.

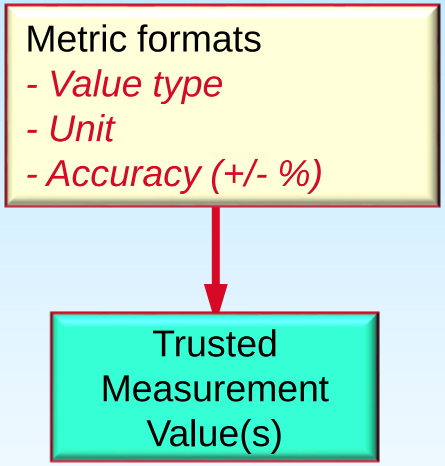
Metric types are required to make performance figures comparable.

Without metrics types performance characteristics can’t be compared!

Requirements on Performance Data

4. Metric formats

Metric formats provide information about value type, scale, and precision.



Metric formats

- Value type

- Unit

- Accuracy (+/- %)

Trusted Measurement Value(s)

Performance data has three format attributes:

1. Value type data (time, counters, size, etc.)
2. Unit data (For example time: Hour, sec, mSec)
3. Accuracy data (the precision or quality of data)

Metric formats are required to make performance figures comparable, such as miles per hour and kilometers per hour.

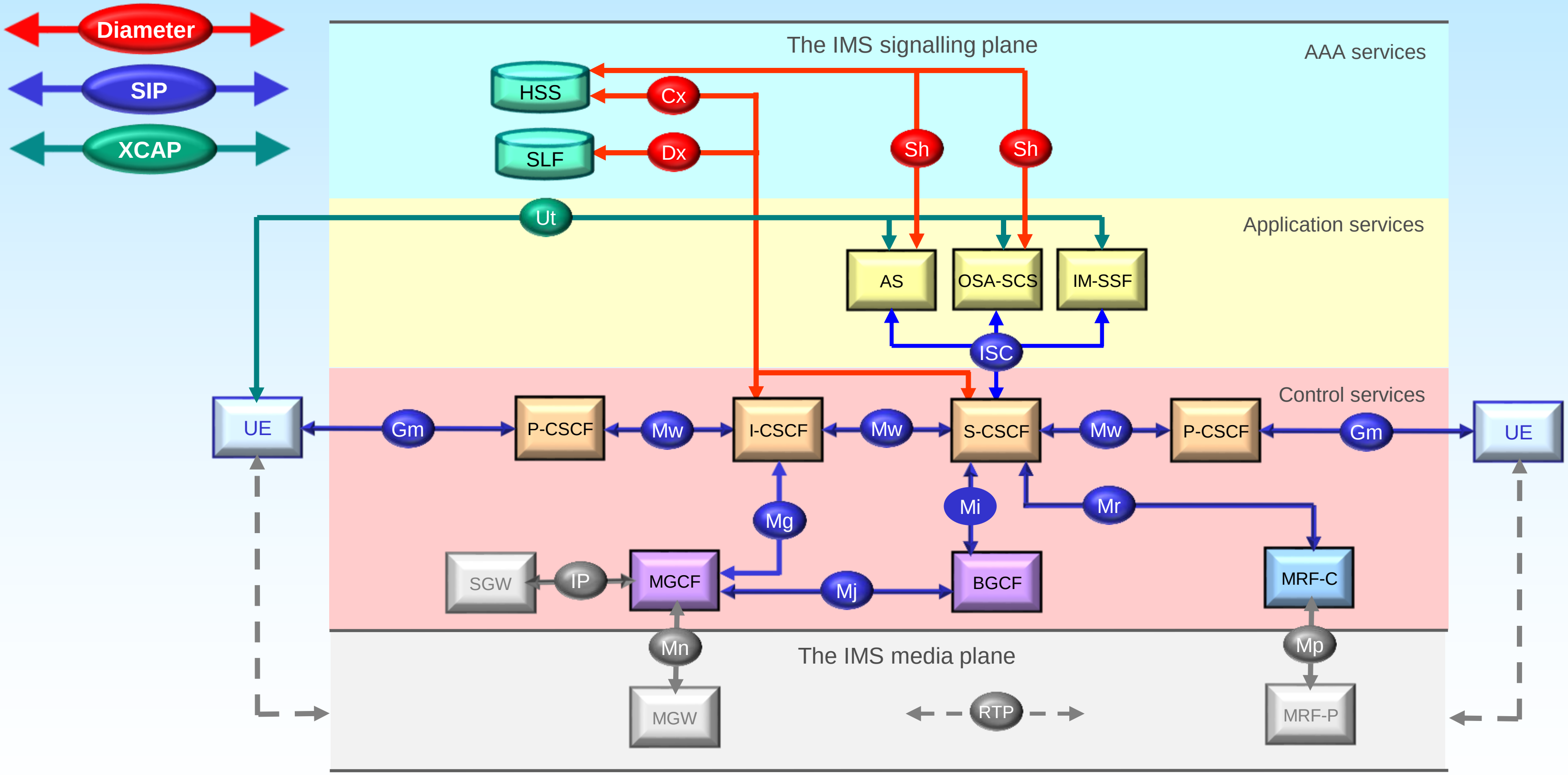
Without metrics formats performance characteristics can’t be compared !

Measured Objects

Measured objects

IMS is a good example of a distributed service

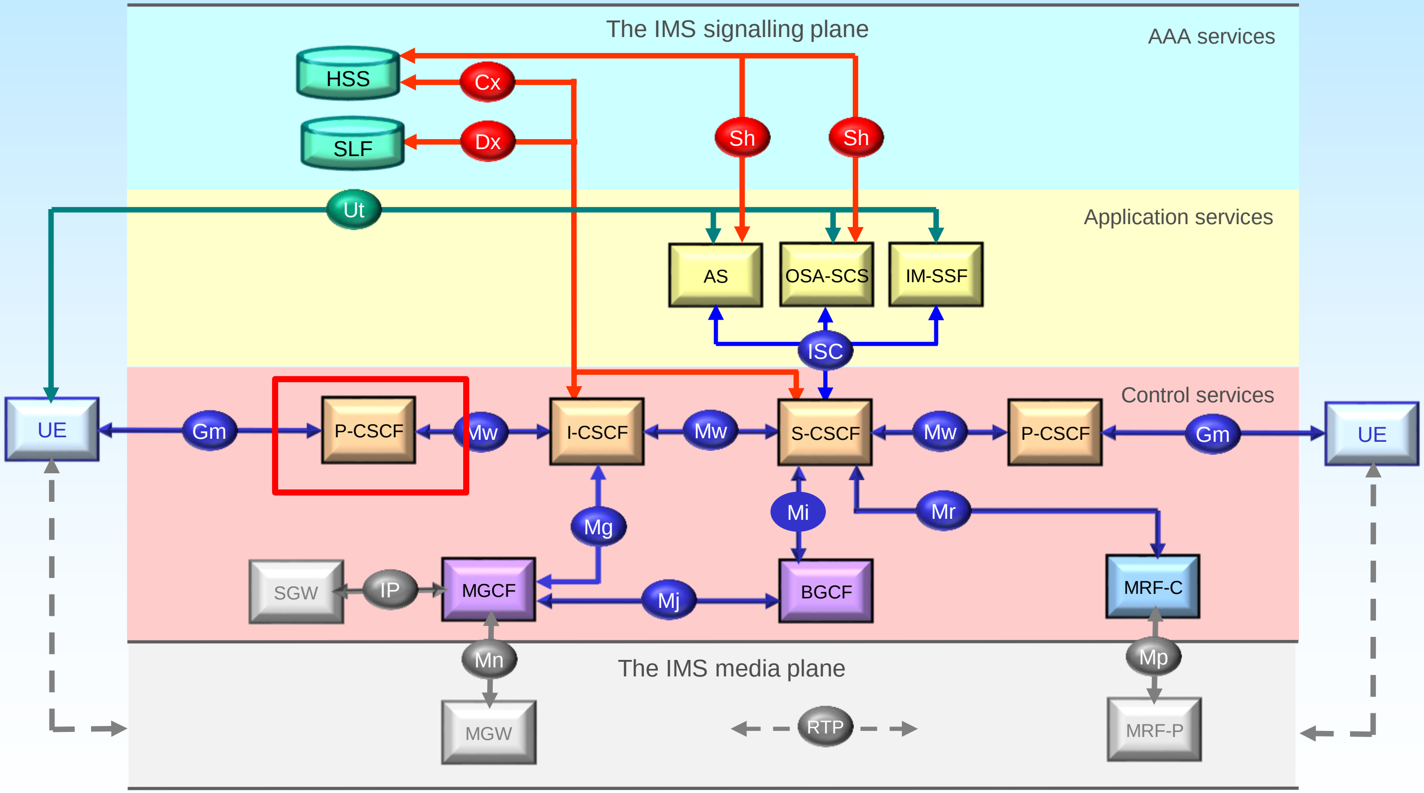
IMS has a number of defined services interconnected by a set of protocols.



Measured objects

What is actually measured?

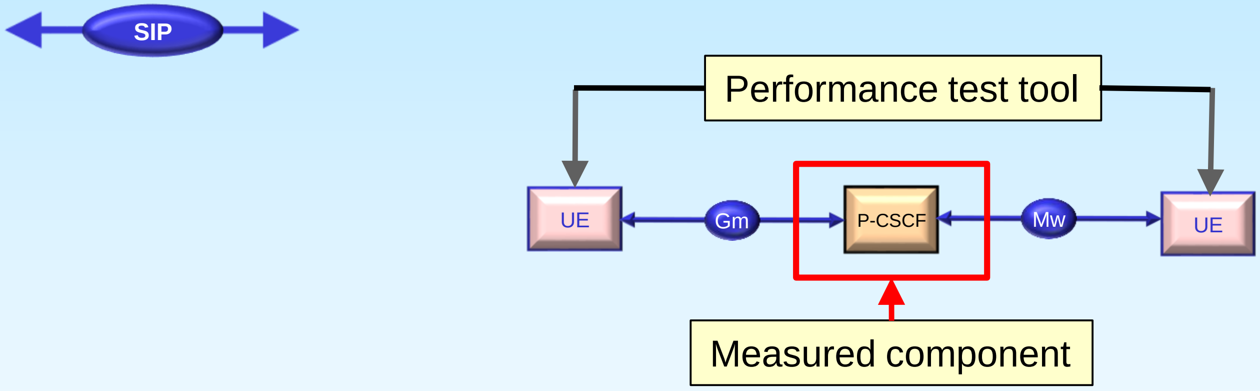
From performance characteristics of single IMS components



Measured objects

What is actually measured?

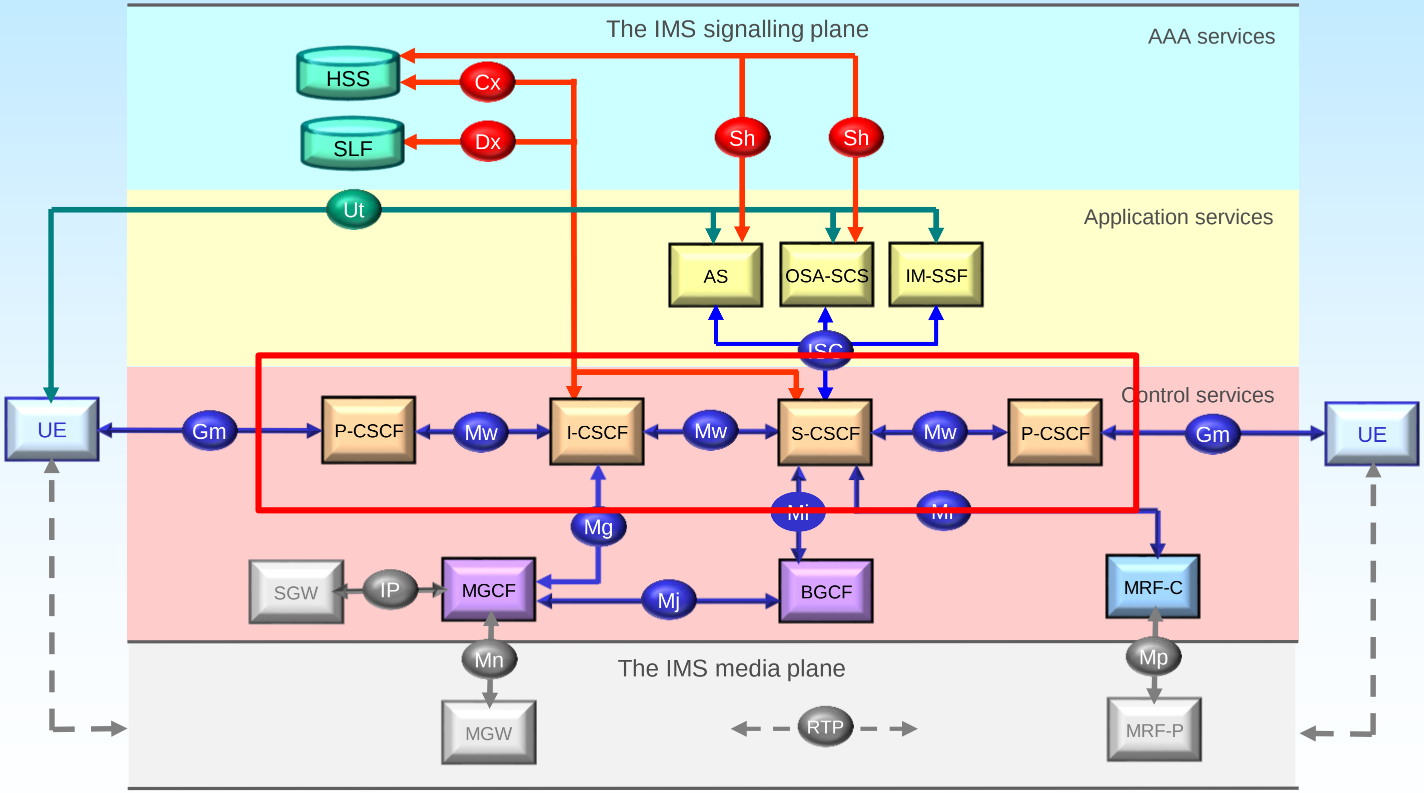
From performance characteristics of single IMS components



Measured objects

What is actually measured?

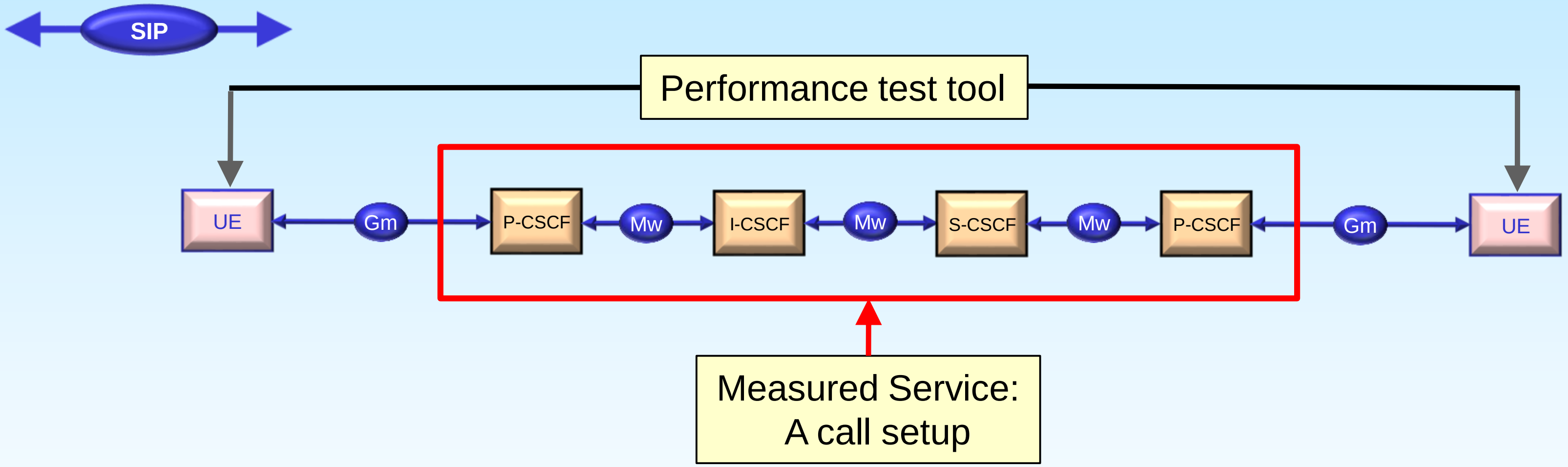
To performance characteristics of IMS services



Measured objects

What is actually measured?

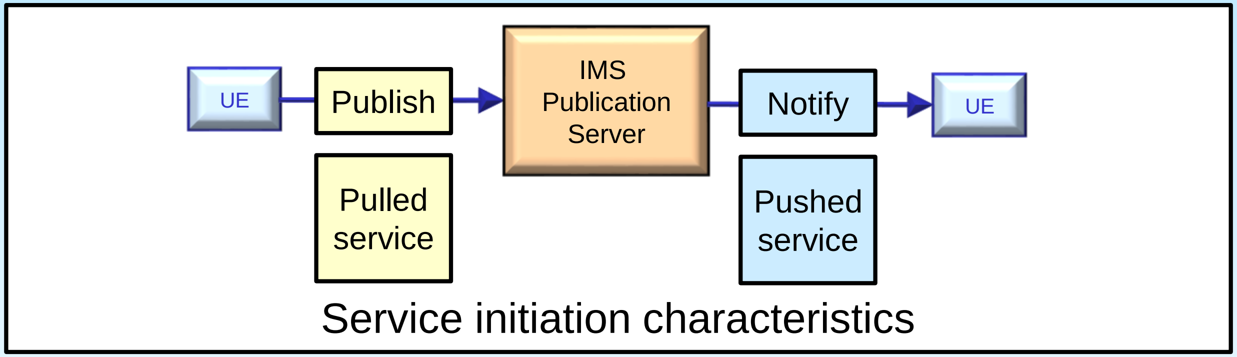
To performance characteristics of IMS services



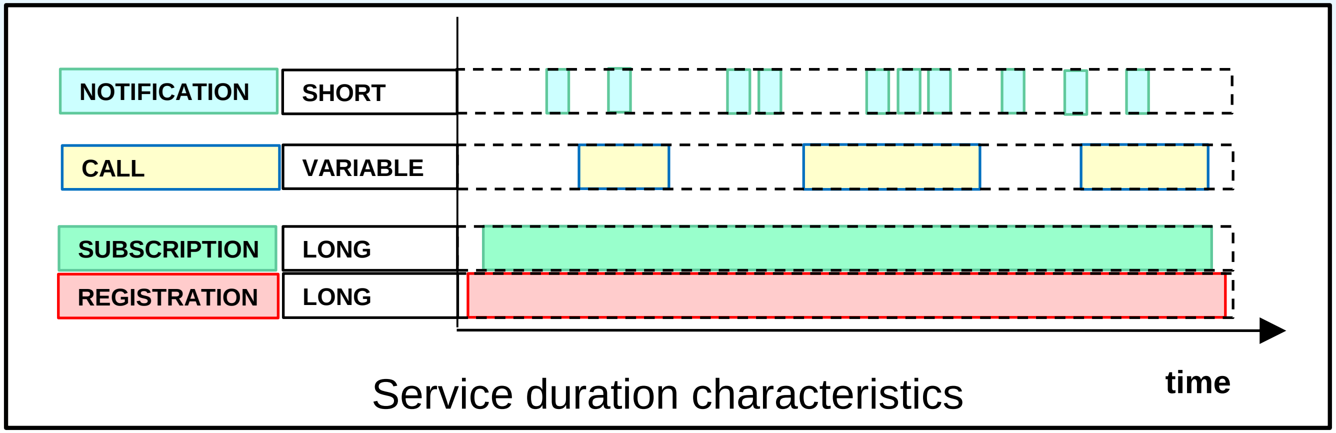
Measured objects

Service characteristics

Provided services can be very different, here are two examples:



Service initiation characteristics

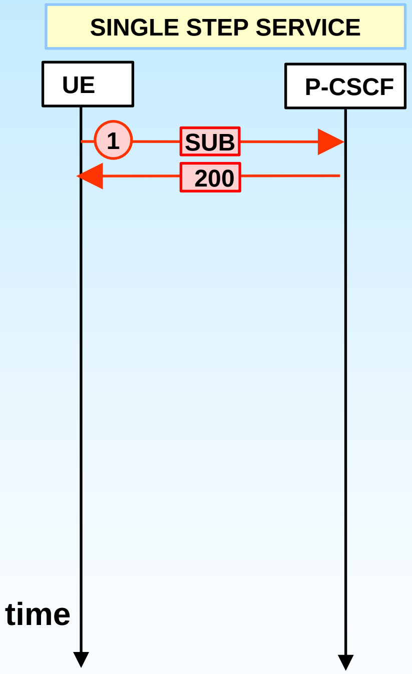


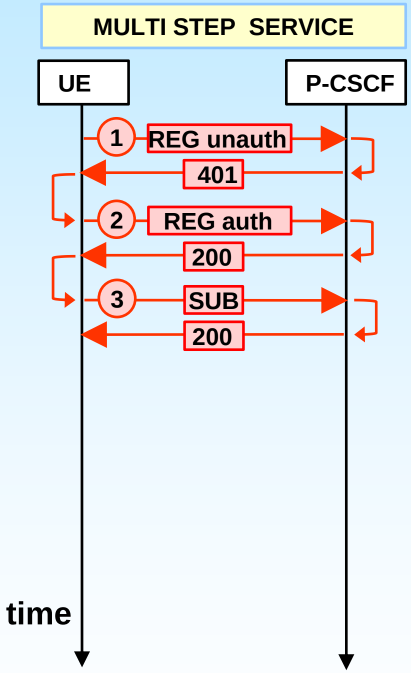
Service duration characteristics

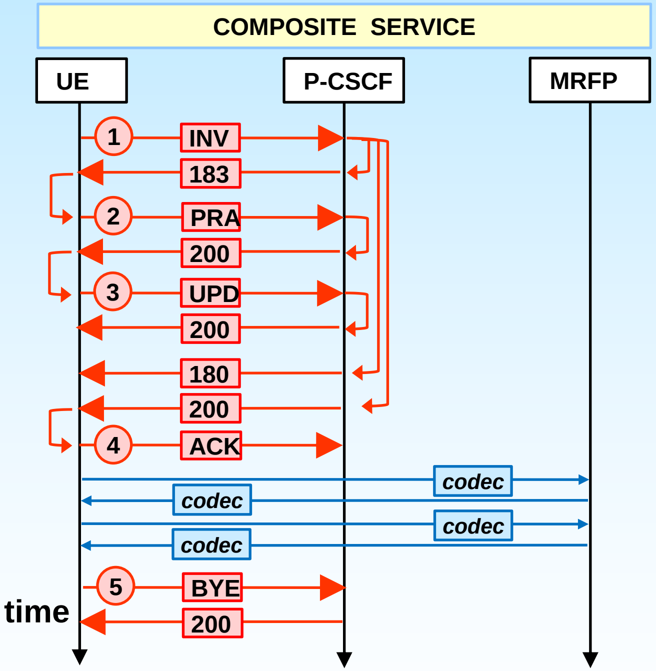
Measured objects

Service design

Provided services can have different designs, here are three examples:



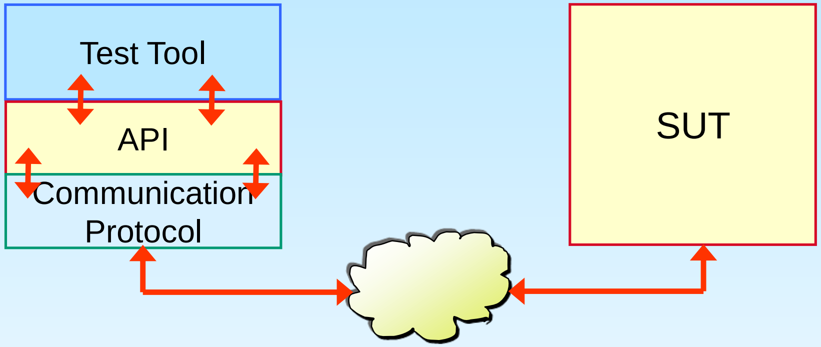


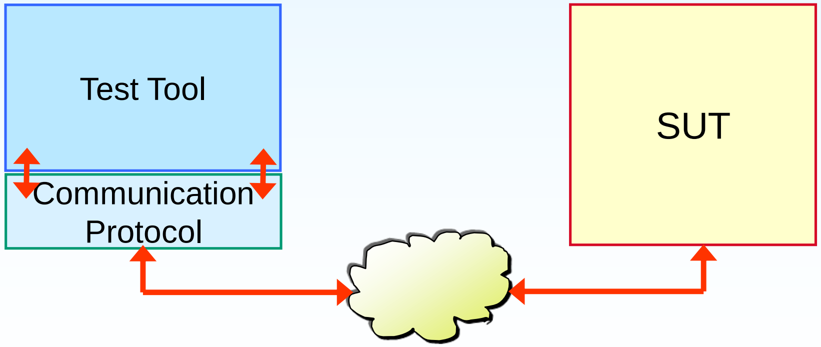


Measured objects

Service interfaces

Provided services can be provided with different interfaces, here are two examples:



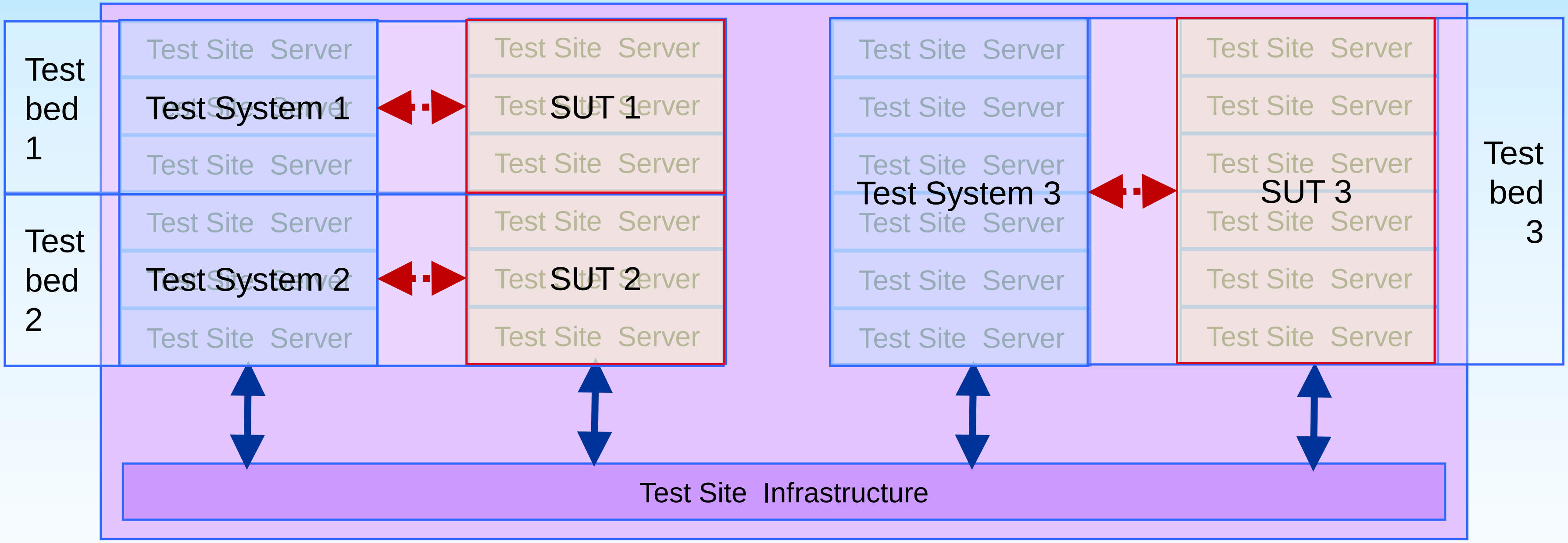


Performance Test Environment

Performance Test Environment

Test site and test beds

A test site is a collection of hardware installed for performance measurements.

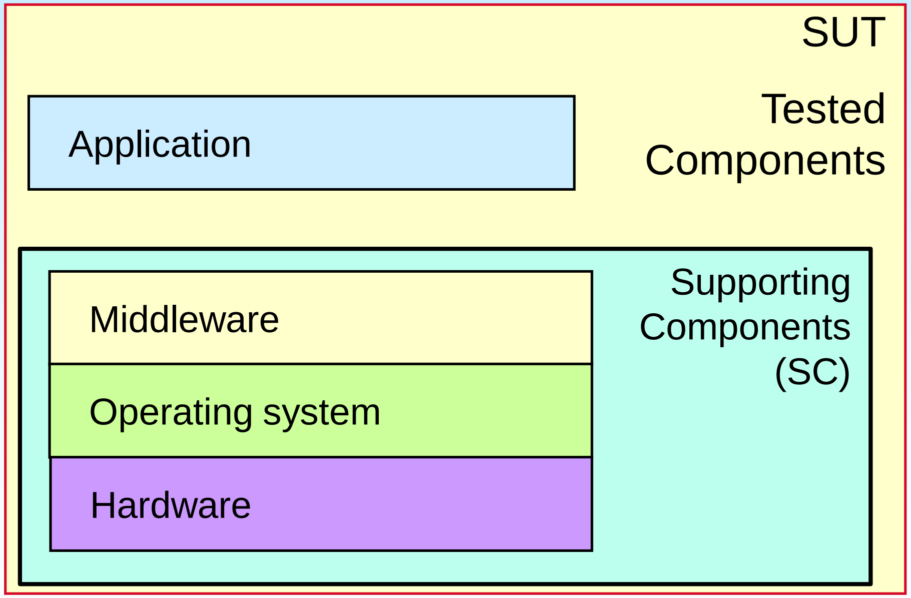


Test beds are how different SUT:s utilize the test site.

Performance Test Environment

Layers in a System Under Test (SUT)

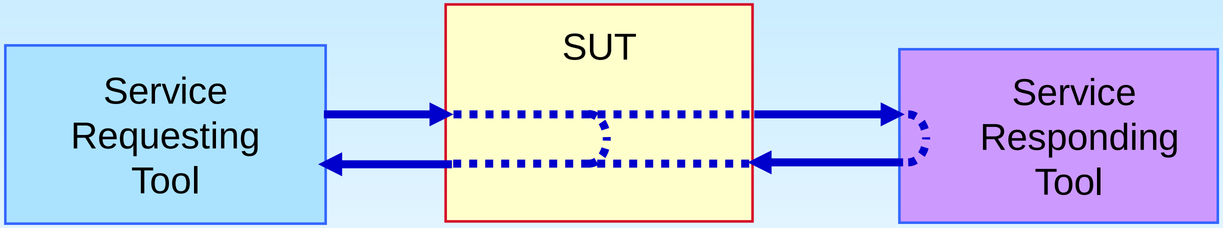
An application is the target for performance measurements.



Performance Test Environment

A test bed with service requesting and service responding tools

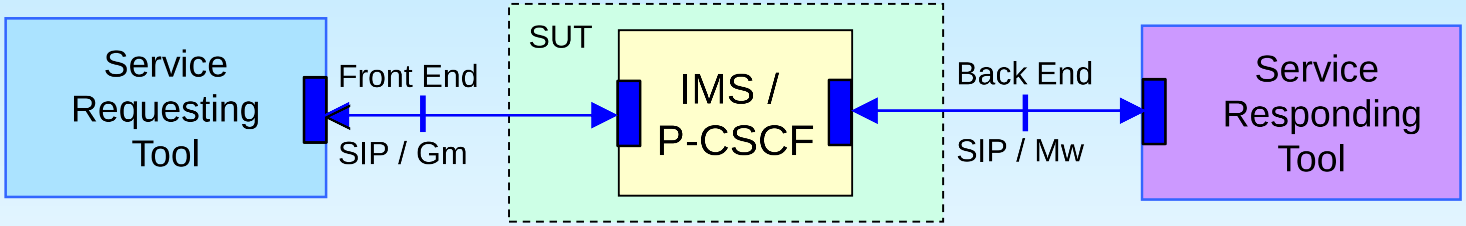
A SUT can be surrounded by test tools playing different roles.



Performance Test Environment

Example of front-end border and back-end border

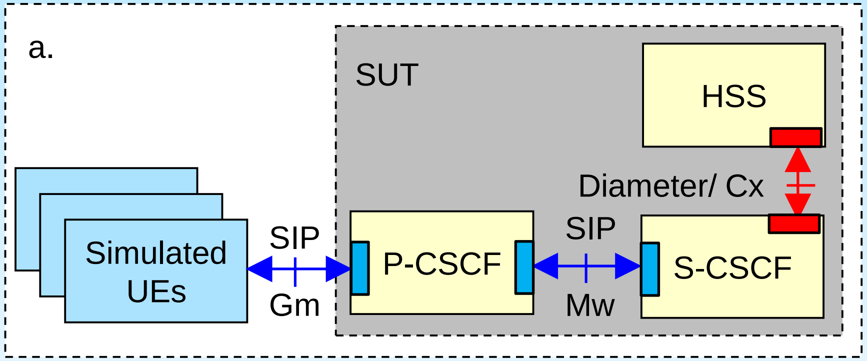
A test tool has one or more borders to a System Under Test.



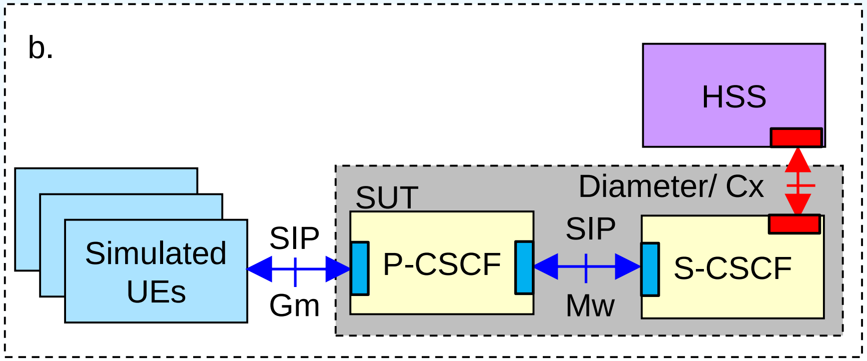
Performance Test Environment

Example of a SUT with simulated components (HSS)

Components of a System Under Test can be replaced by test tools.



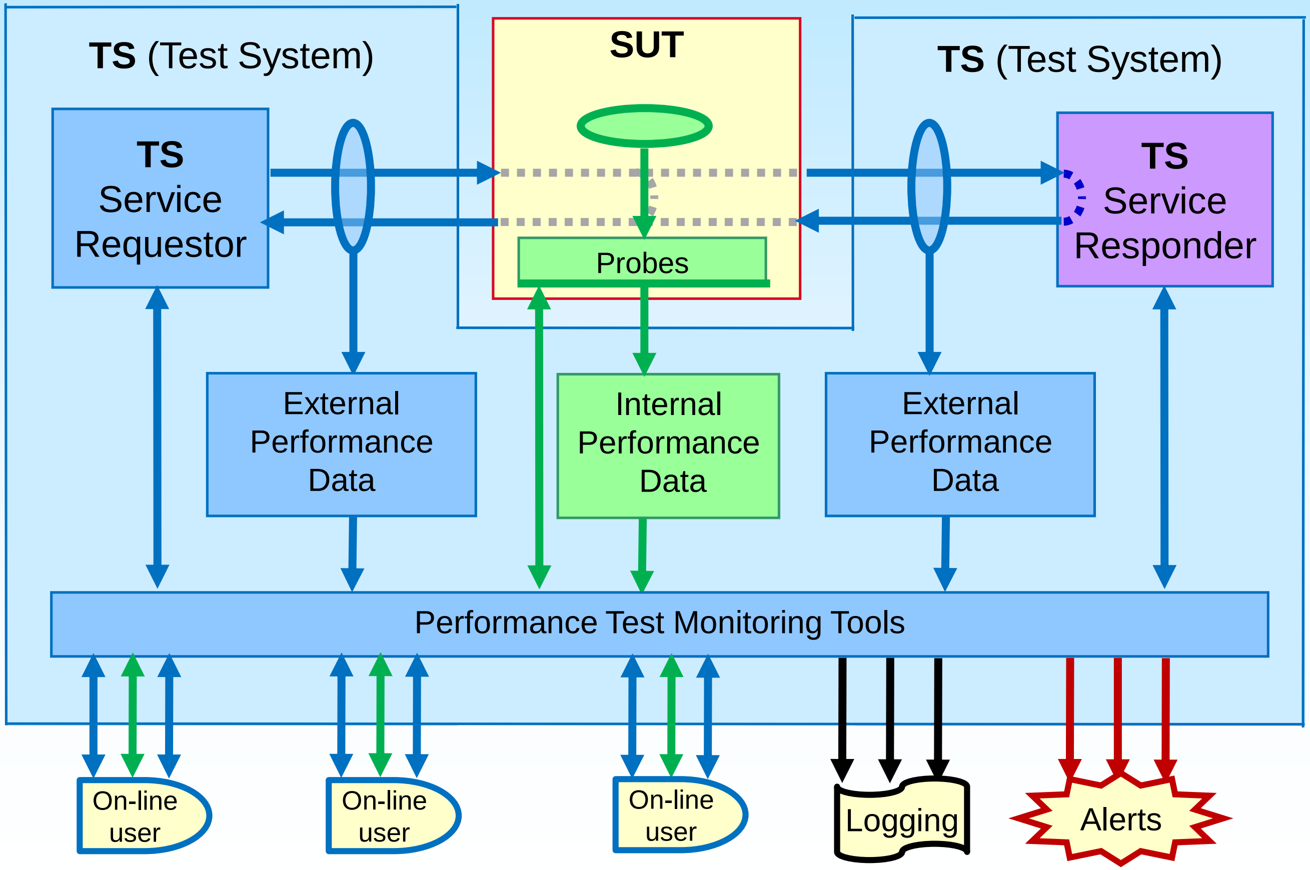
In this case an HSS.



Performance Test Environment

Monitoring performance tests

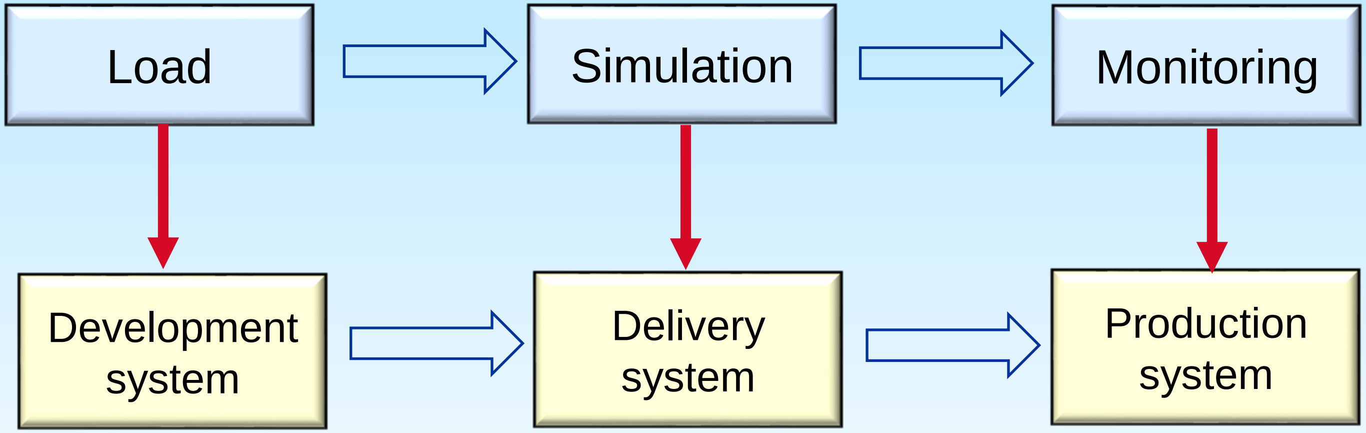
A performance test tool shall support monitoring of an ongoing test.



General Performance Test Concepts

General Performance Test Concepts

Performance measurements are required during the whole life cycle



Load

Simulation

Monitoring

Development system

Delivery system

Production system

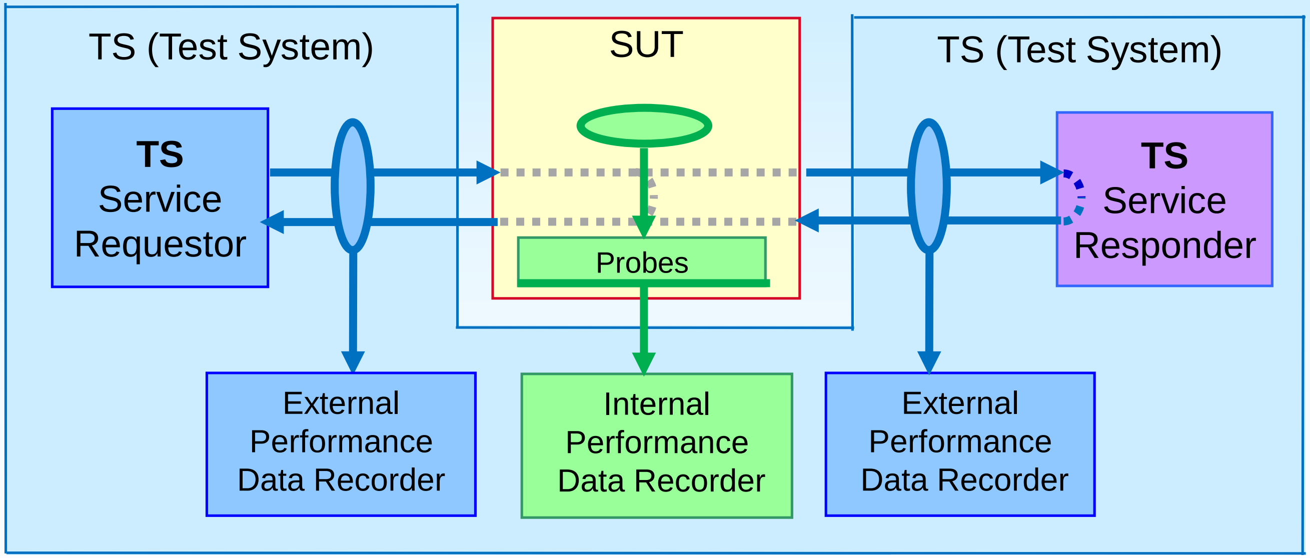
When:	Development	Pre Deployment	Post Deployment
Why:	Measure limits	Verify requirements	Confirm requirements
What:	Isolated services .	Complex traffic Critical services	Proactive monitoring .

General Performance Test Concepts

Recording of external and internal performance data

There are two types of performance data:

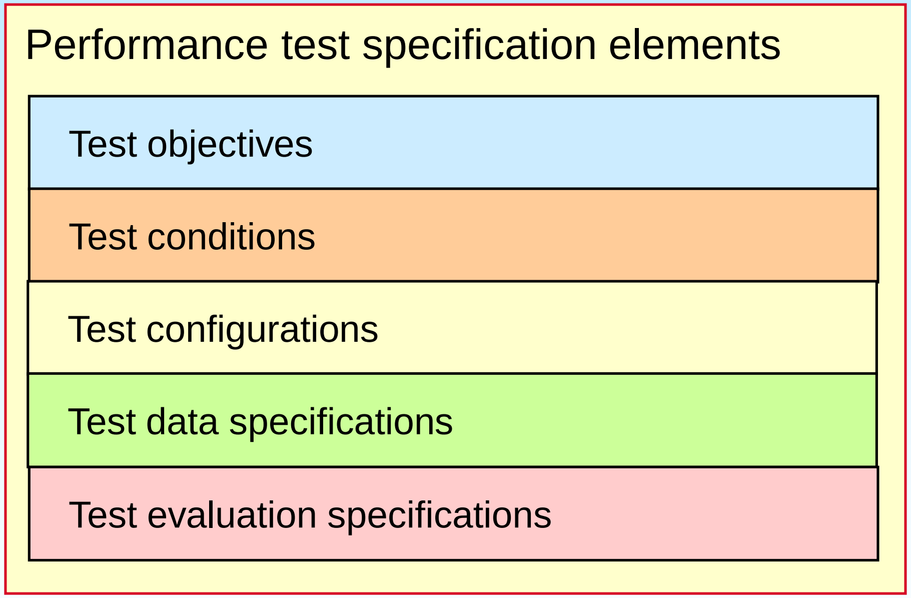
1. What is captured externally or outside the System Under Test (SUT).
2. What is captured internally or inside the SUT (with probes or profiling tools).



Requirements on Performance Test Specifications

Requirements on Performance Test Specifications

A performance test specification has many elements



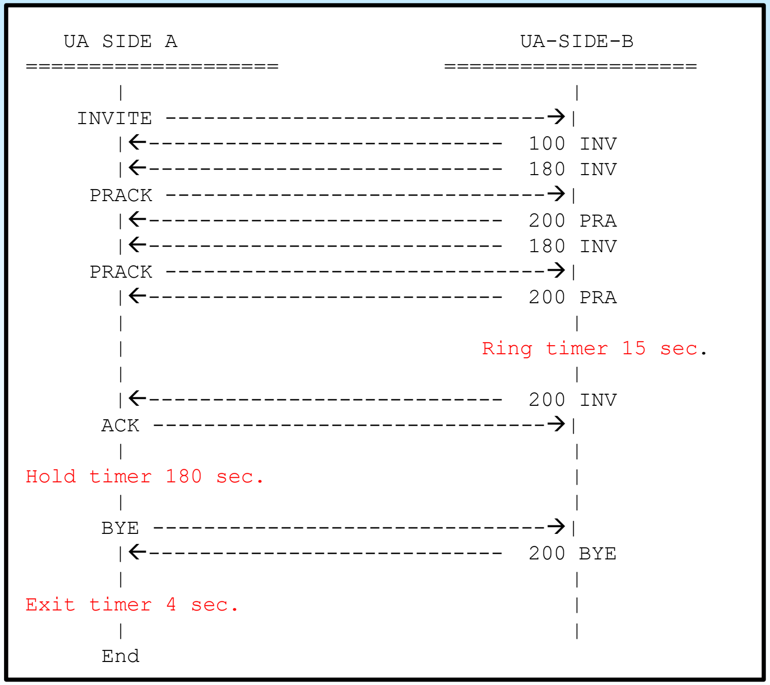
Performance test specification elements

- Test objectives
- Test conditions
- Test configurations
- Test data specifications
- Test evaluation specifications

Requirements on Performance Test Specifications

The core of a test case is the service scenarios to be used

A performance test case requires working services to be tested.



```
   UA SIDE A                          UA-SIDE-B
====================              ====================
       |                                   |
    INVITE ------------------------------→|
       |←---------------------------  100 INV
       |←---------------------------  180 INV
     PRACK ------------------------------→|
       |←---------------------------  200 PRA
       |←---------------------------  180 INV
     PRACK ------------------------------→|
       |←---------------------------  200 PRA
       |                                   |
       |                         Ring timer 15 sec.
       |                                   |
       |←---------------------------  200 INV
      ACK --------------------------------→|
       |                                   |
Hold timer 180 sec.                        |
       |                                   |
      BYE --------------------------------→|
       |←---------------------------  200 BYE
       |                                   |
Exit timer 4 sec.                          |
       |                                   |
      End
```

Workload Concepts

Workload Concepts

The description of what the measured system shall produce

A workload has three parts, each with many variations:

1. The mix of requested services and the weight of each service.
2. The amount of each requested service (can be described in several ways).
3. The time distribution of requested services, or the service request load.

The service request load can be defined in two ways:

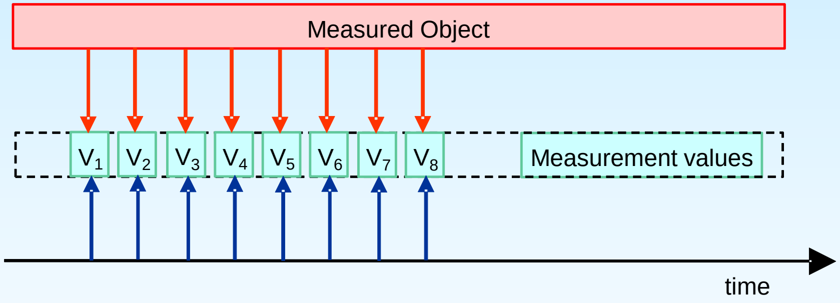
1. User session driven load (suited for web performance measurements).
2. Traffic rate driven load (suited for telecom performance measurements).

Performance Data Processing

Performance Data Processing

Performance data processing has many steps

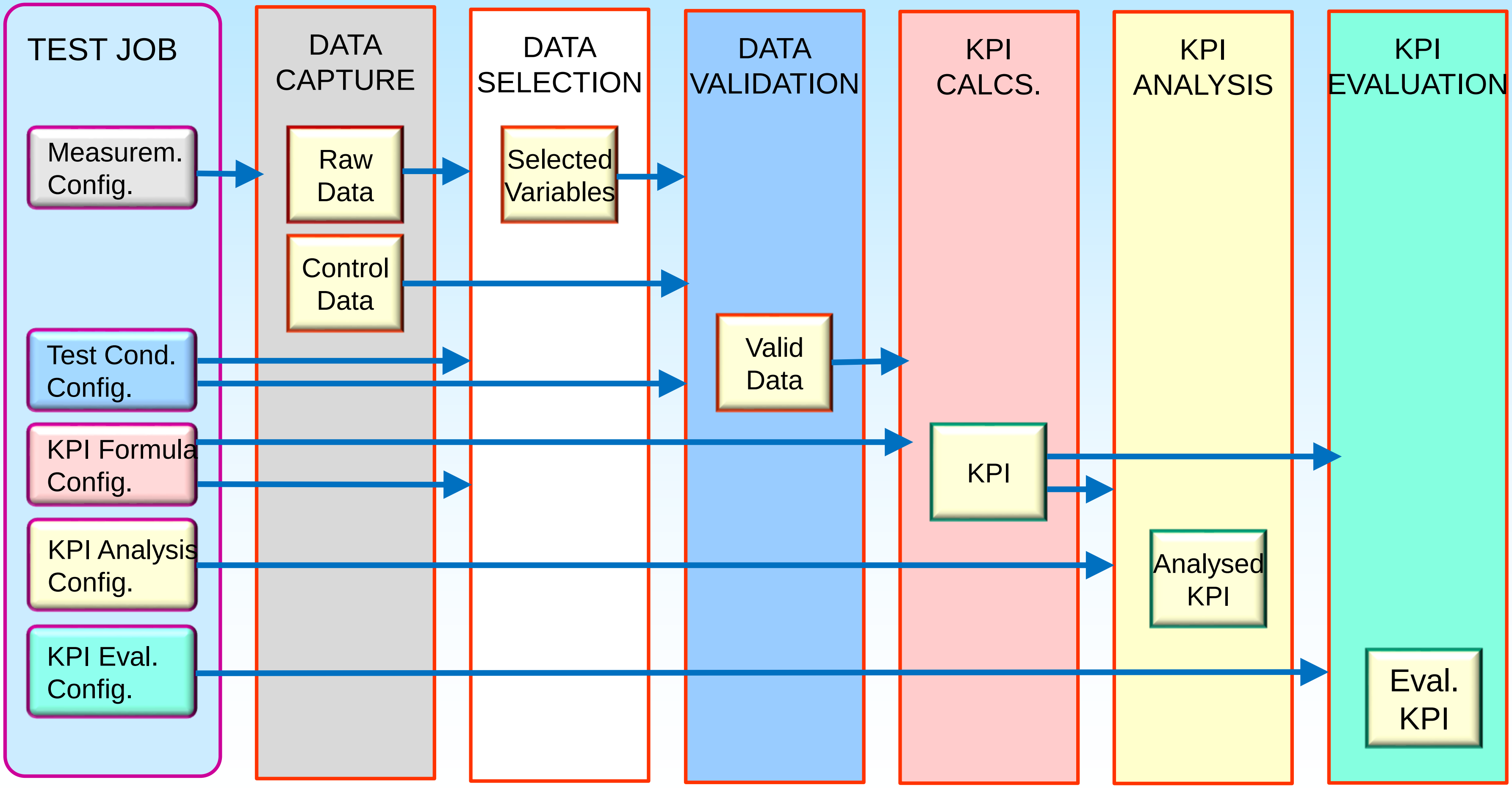
Collected measurement data are usually time series of measurement values.



Different measurement values are synchronized by timestamps.

Performance Measurement data flow

Performance data processing has many steps: Here for KPI production



TEST JOB

- Measurem. Config.
- Test Cond. Config.
- KPI Formula Config.
- KPI Analysis Config.
- KPI Eval. Config.

DATA CAPTURE

- Raw Data
- Control Data

DATA SELECTION

- Selected Variables

DATA VALIDATION

- Valid Data

KPI CALCS.

- KPI

KPI ANALYSIS

- Analysed KPI

KPI EVALUATION

- Eval. KPI

I end this presentation with another statement by Lord Kelvin

"To measure is to know"

- Lord Kelvin

"To benchmark is to know"

- Michael Mild