Project: H2020-ICT-688712

Project Name:

5G Applications and Devices Benchmarking (TRIANGLE)

D2.2. Appendix 1

TRIANGLE Mark scoring

Date of delivery:	30/06/2017	Version:	1.0
Start date of Project:	01/01/2016	Duration:	18 months

Contents

List of Tables

1 Introduction

1.1 Purpose

The purpose of this Appendix is to describe the process to obtain the TRIANGLE mark scoring from the measurements (raw test results data) when performing the test cases defined in TRIANGLE Test Specifications.

1.2 Acronyms

Table 1 – Acronyms

State	Description
AEC	Application Under Test Energy Consumption
AUE	Application Under Test User Experience
CDF	Cumulative Distribution Function
DRA	Device Under Test user Experience with Reference Apps
IDP	IoT Devices Under Test Data Performance
IDR	IoT Devices Under Test Reliability
IEC	IoT Devices Under Test Energy Consumption
KPI	Key Performance Indicator
RES	Application Under Test Device Resource Usage
SL	Short Lasting
AEC	Application Under Test Energy Consumption
AUE	Application Under Test User Experience
CDF	Cumulative Distribution Function
DRA	Device Under Test user Experience with Reference Apps
IDP	IoT Devices Under Test Data Performance
IEC	IoT Devices Under Test Energy Consumption
KPI	Key Performance Indicator

1.3 References

[1] TRIANGLE, D2.1: Initial report on the testing scenarios, requirements and use cases, 2016.

[2] ITU-T, "G.1030 Estimating end-to-end performance in IP networks for data applications," 02/2014.

[3] ITU-T, "G.1030 Estimating end-to-end performance in IP networks for data applications," 02/2014.

2 Triangle Mark Scoring Framework

Before going in detail with the scoring process the following concepts introduced in D2.1 [1] are refreshed for a better understanding:

- Measurement is a value discovered by measuring, that corresponds to a property of something. The measurements come from the logs that can be collected by the Triangle Testing Framework. For example, the instantaneous current in Amperes measured by a Power Meter is an example of measurement in TRIANGLE.
- Key Performance Indicator is a quantitative evaluation of criteria that a product (i.e., device or App) must meet before release. A KPI is computed from a set of measurements. For example, energy usage while an App is in active state is a KPI computed from current consumption measurement.
- Domain is a categorization of the KPIs from a user satisfaction perspective. In D2.1 [1] the following are proposed:
    - Mobile Devices: Energy Consumption, Data Performance, Radio Performance, User Experience (with Reference Apps)
    - IoT Devices: Networks Adaptation, Reliability, data Performance, Energy Consumption
    - Applications: Reliability, Network resource Usage, User Experience, Devices Resources Usage, Network Adaptation, Energy Consumption.
- Use Case: Showcase applications in the context of 5G services. In D2.1 [1] the following are proposed: Virtual Reality, Gaming, Content Distribution. Live Streaming, High speed Internet, Smart Metering, Smart Grids, Connected Vehicles, Patient Monitoring and Emergency Services.
- Network Scenarios: 5G usage scenarios according to their commonalities in the network deployment and channel conditions. In D2.1 [1] the following are proposed:
    - Urban-Office, Urban-Pedestrian, Urban-driving-normal, Urban-driving-traffic jam, Urban-Internet Cafè-Busy Hours, Urban-Internet Cafè-Off Peak,
    - Suburban-Festival, Suburban-Stadium, Suburban-Shopping Mall-Busy Hours, Suburban-Shopping Mall-Off Peak,
    - High Speed Train-Relay, High Speed Train-Direct connection,
    - IoT-Warehouse, IoT-Outdoor sensors, IoT-Home sensors.

Let

- m, the Measurements (raw test result data)
- N, the number of test iterations of a test case
- KPI, the Key Performance Indicators
- KPI', the normalized KPIs
- P, the number of KPI which are defined for a given Use Case
- Q, the number of Use Cases
- R, the number of applicable Scenarios

- S, the number of Domains
- Δ, aggregation function
- f, reference case evaluation function.

The Triangle Mark is a value within the range 1.0 to 5.0. It is represented by a Spider Web chart with S axis where each axis holds a domain score.

The aggregation functions Δ^N which summarize measurements into KPIs are specified in the test specifications documents. For example, in User Experience domain, the KPI “App Access Time” is defined by the Average, Deviation and CDF from the measured login time throughout the test iterations.

A first set of the reference case evaluation functions “f” are introduced in section 3 and, for the domains inside the scope of this first version of this framework (i.e., User Experience, Device Resource Usage, Energy Consumption), the threshold values are in Annex 1.

The aggregation function Δ^P, which aggregates the normalized KPIs into one single score for each Use Case, is an average function in this version of the framework.

The aggregation function Δ^Q, which summarizes Use Case scores into one single per each Domain, is also an average function.

More sophisticated functions can be defined in further releases of this framework supported by actual testing results. The same rationale applies to Δ^R and Δ^S.

The complete process is illustrated with one example in Annex 2.

The scoring framework is defined by the following algorithm:

For each **Domain** "i"

{

For each applicable **Network Scenario** "j"

{

For each **Use Case** "k"

{

For each **KPI** "l"

{

$$KPI_l = \Delta^N m$$

$$KPI_l' = f\ (KPI_l)$$

}

$$Domain_i\ Scenario_j Use\ Case_k\ \ Score = \ \Delta_l^P KPI_l'$$

}

}

For each **Use Case** "k"

{

$$Domain_i\ Use\ Case_k\ \ Score = \ \Delta_l^R Domain_i\ Scenario_l Use\ Case_k$$

}

$$Domain_i\ Score = \ \Delta_j^Q\ Domain_i\ Use\ Case_j\ Score$$

}

For each **Use Case** "k"

{

$$Use\ Case_k\ \ Score = \ \Delta_l^S Domain_l\ Use\ Case_k\ Score$$

}

$$Triangle\ Mark = \ \Delta_i^Q Use\ Case_i\ Score$$

3 Reference Case Evaluation Functions

The TRIANGLE mark scoring requires the merging of KPI from multiple measurements. This data needs to be transformed into forms appropriate for scoring. The data transformation proposed in this framework involves normalization, where the KPI are scaled so as to fall within small specified range from 1.0 to 5.0.

In this document some normalization functions are proposed as an initial approach to exercise this framework. As actual testing is performed, these functions will be refined and others could be developed.

3.1 Type I

This function performs a linear interpolation on the original data. Suppose that min_{KPI} and max_{KPI} are the worst and best values of a KPI from a reference case. This function maps a value, *v*, of a KPI, to *v'* in the range [1.0, 5.0] by computing

$$v' = \frac{v - min_{KPI}}{max_{KPI} - min_{KPI}}(5.0 - 1.0) + 1.0$$

This function is to be used for KPIs which will scored by a simple linear interpolation from the worst and best expected values from a reference case.

If a future input case falls outside the data range of the KPI, it will be set to the extreme value min_{KPI} (if it is worse) or max_{KPI} (if it is better).

3.2 Type II

This function performs a logarithmic interpolation and is inspired in the opinion model recommended by the ITU-T in [2] for a simple web search task. This function maps a value, *v*, of a KPI, to *v'* in the range [1.0, 5.0] by computing

$$v' = \frac{5.0 - 1.0}{\ln((0.003min_{KPI} + 0.12)/min_{KPI}} \cdot (\ln(v) - \ln(0.003min_{KPI} + 0.12)) + 5$$

This function is to be used in the TRIANGLE framework scoring for KPIs which reflect single time events such as search time or time to load first frame. Therefore, for such a single interaction, max_{KPI} (the "best" value for the scoring) equals to 0.12 seconds, corresponding to an instantaneous perception threshold [3].

If a future input case falls outside the data range of the KPI, it will be set to the extreme value min_{KPI} (if it is worse) or max_{KPI} (if it is better).

4 Annex 1: Reference Case Evaluation Values

This annex shows the values for the reference case evaluation functions parameters for the test specifications covered in this release of the deliverable.

4.1 AUT User Experience

Table 2 – AUE Reference Case Evaluation Values

KPI	**Target**	**Function**		
			min_{KPI}	max_{KPI}
App Access Time (s)	Average	Type II	10	0.1
App Accessibility (%)	Ratio	Type I	50	100
App Availability (%)	Ratio	Type I	50	100
Content Load Time (s)	Average	Type II	10	0.1
Response Time (s)	Average	Type II	10	0.1
Feature Availability (%)	Ratio	Type I	50	100
Content Stall (%)	Index	Type I	5	0
Content Search Time (s)	Average	Type II	10	0.1
Content Download Throughput (Mbit/s)	Average	Type I	1	1000
Content Upload Throughput (Mbit/s)	Average	Type I	1	1000
Content Resolution	Mode	Type I	Lowest	Highest
Broadcast Content Resolution	Mode	Type I	Lowest	Highest

4.2 AUT Device Resources Usage

Table 3 – RES Reference Case Evaluation Values

KPI	**Target**	**Function**	**U-P**	
			min_{KPI}	max_{KPI}
Use of memory (%)	Average	Type I	100	0
Use of CPU (%)	Average	Type I	100	0
Use of GPU (%)	Average	Type I	100	0

4.3 DUT User Experience with Reference Apps

Table 4 – DRA Reference Case Evaluation Values

KPI	**Target**	**Function**		
			min_{KPI}	max_{KPI}
App Access Time (s)	Average	Type II	10	0.1
App Accessibility (%)	Ratio	Type I	50	100
App Availability (%)	Ratio	Type I	50	100
Content Load Time (s)	Average	Type II	10	0.1
Response Time (s)	Average	Type II	10	0.1
Feature Availability (%)	Ratio	Type I	50	100
Content Stall (%)	Index	Type I	5	0
Content Search Time (s)	Average	Type II	10	0.1
Content Download Throughput (Mbit/s)	Average	Type I	1	1000
Content Upload Throughput (Mbit/s)	Average	Type I	1	1000
Content Resolution	Mode	Type I	Lowest	Highest

4.4 AUT Energy Consumption

Table 5 – AEC Reference Case Evaluation Values

KPI	**Target**	**Function**		
			min_{KPI}	max_{KPI}
App Current Consumption	Average	Type I	0.70	0.10

The value in Table 5 corresponds to maximum brightness level in a 3.7 V powered host (mobile device) and is derived from the aggregation of the measurements obtained from the test cases in the deliverable D2.2 Appendix 7.

4.5 IoT DUT Energy Consumption

Table 6 – IEC Reference Case Evaluation Values

KPI	**Target**	**Function**		
			min_{KPI}	max_{KPI}
Current Consumption (A)	Average	Type I	1.00	0.10

The value in Table 6 corresponds to the following usage scenario workload in 5V powered IoT device: Image/Video Capture and Streaming. Some reference values have been found in [4].

4.6 IoT DUT Data Performance

Table 7 – IDP Reference Case Evaluation Values

KPI	**Target**	**Function**		
			min_{KPI}	max_{KPI}
OTA DL U-plane throughput	Average	Type I		
OTA DL C-plane throughput	Average	Type I		
OTA UL U-plane throughput	Average	Type I		
OTA UL C-plane throughput	Average	Type I		
PDCP-SAP goodput UL/DL	Average	Type I		
Number of bearers	Average	Type I		
Number of transport connections	Average	Type I		
Burst inter-generation time at transport level	Average	Type I		

4.7 IoT DUT Reliability

Table 8 – IDR Reference Case Evaluation Values

KPI	**Target**	**Function**		
			min_{KPI}	max_{KPI}
Availability (%)	Average	Type I	50	100
Content Stall (%)	Average	Type I	5	0
Frame Loss Rate (%)	Average	Type I	2	0
Content Resolution	Mode	Type I	Lowest	Highest
Recovery after fail (%)	Average	Type I	50	100
Recovery Time (s)	Average	Type II	60	1

5 Annex 2: Triangle Mark Scoring Example

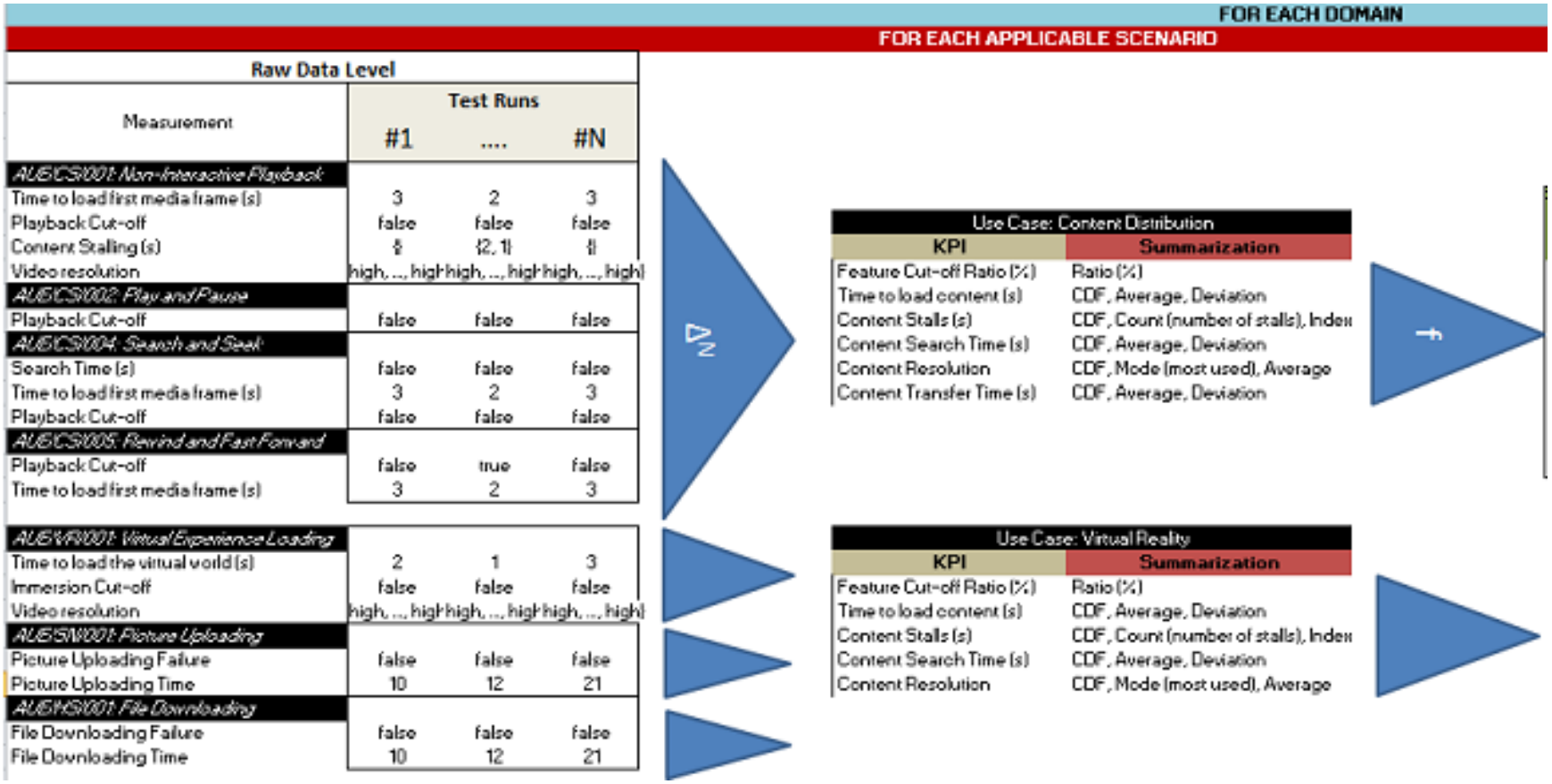

Figure 1 – Example Obtaining KPIs

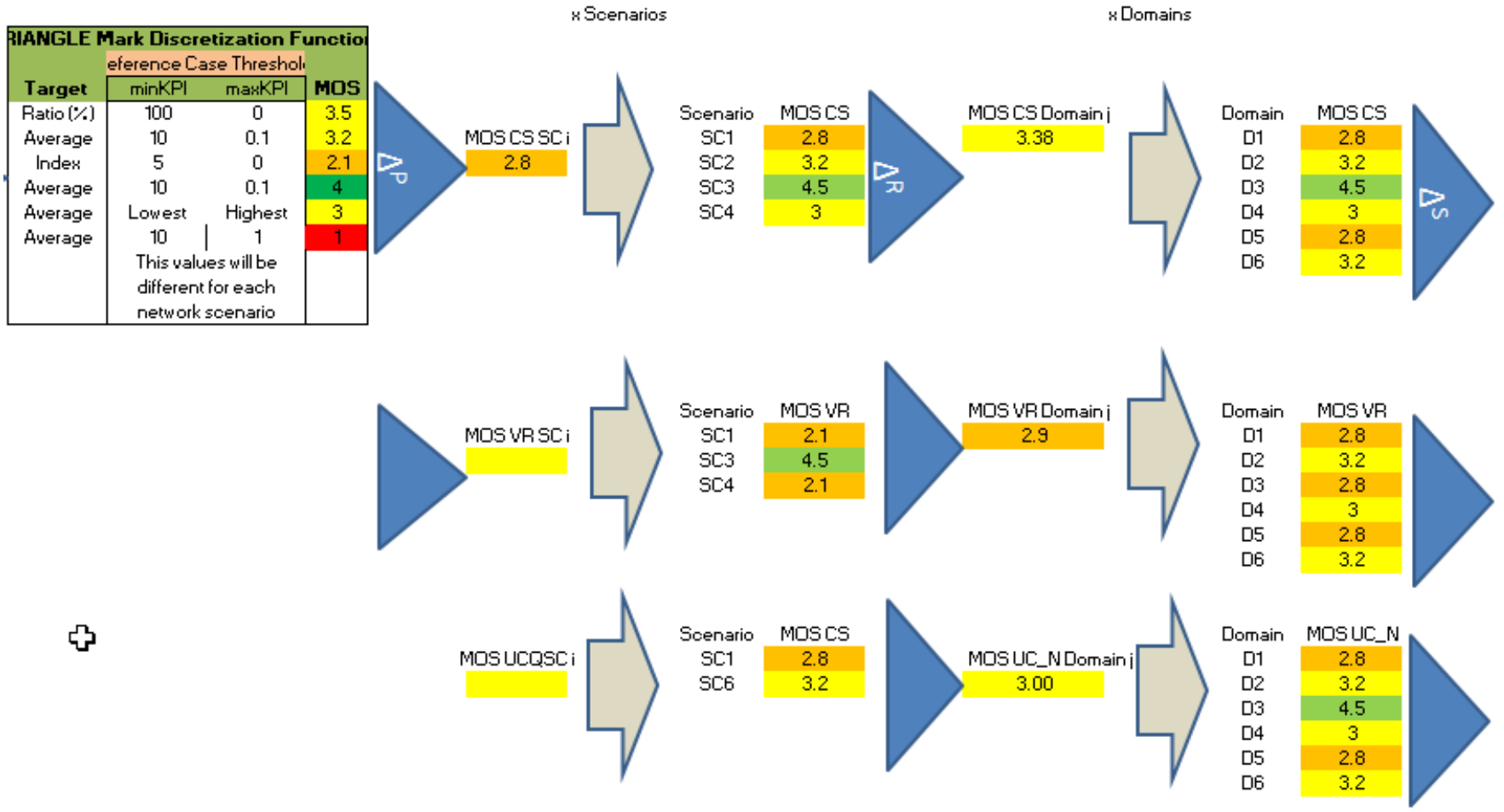

Figure 2 – Example for MOS KPI

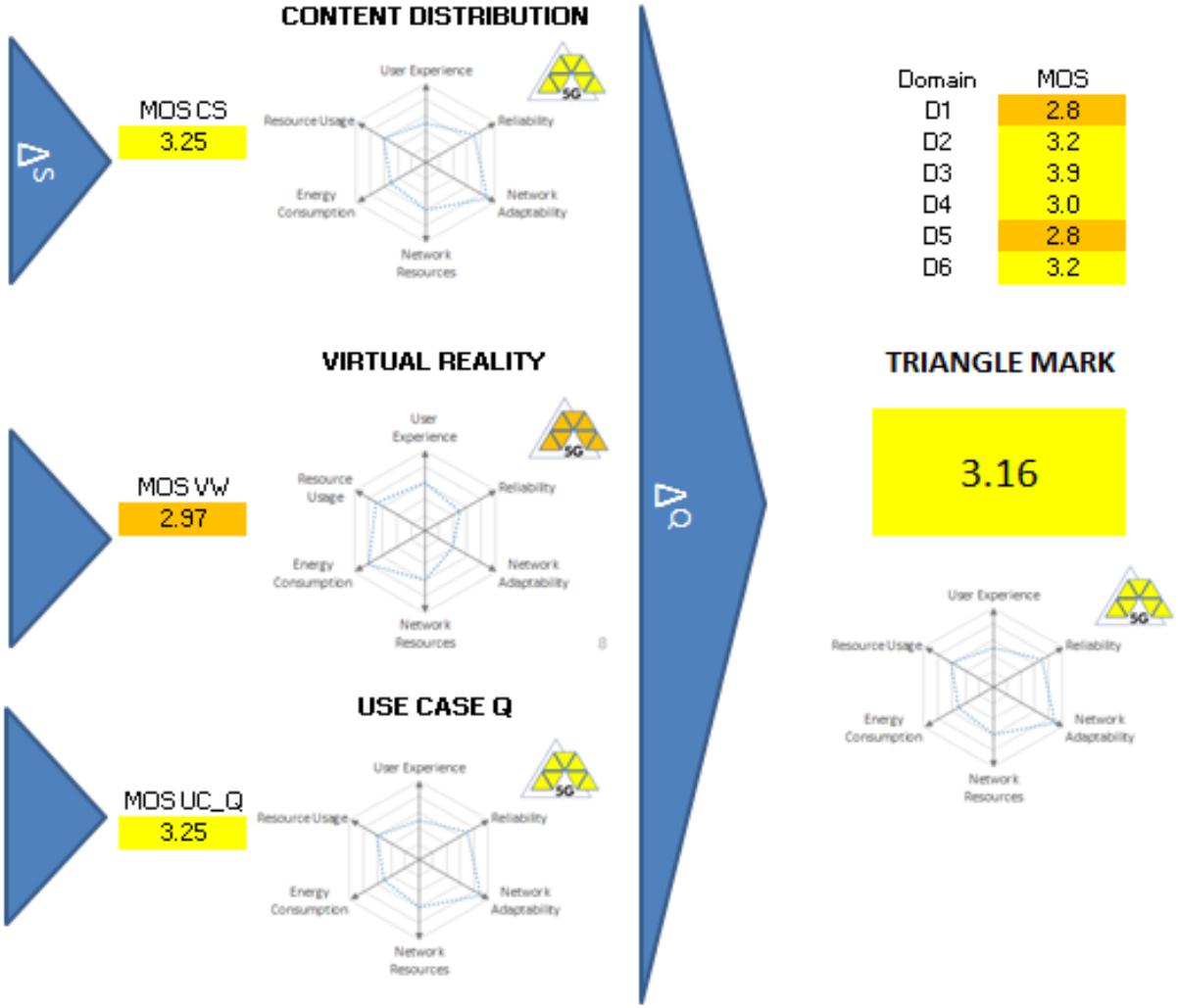

Figure 3 – TRIANGLE Mark example