Quality assessment of over-the-top (OTT) services in mobile networks

ITU-T SG12-RG-AFR Workshop, Freetown, Sierra Leone

1st – 2nd July 2025

Mohamed Lahouij
Solution Engineering Director | MEA

Agenda

01 New emerging 5G testing trends & challenges

02 What are generic testing techniques?

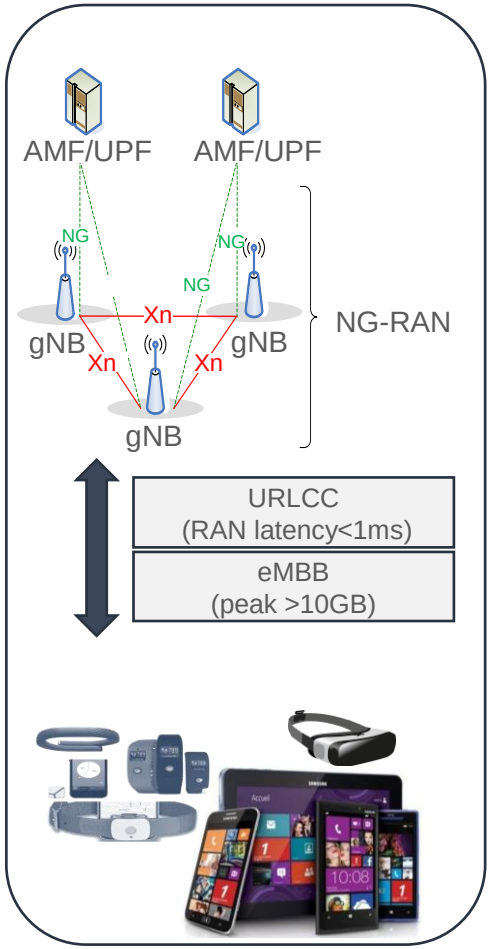
03 Infovista's generic testing solutions within ITU SG12?

- Generic OTT voice testing
- Generic OTT media testing
- Generic OTT interactivity testing

New emerging 5G testing trends & related challenges

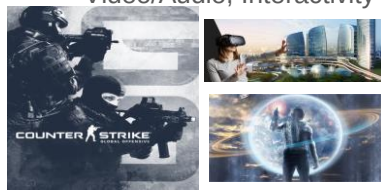
infovista

5G networks enable user centric service evolution…



Gaming/VR/AR/XR services

Video/Audio, Interactivity



IoT verticals

OTT media services

(voice/video, video streaming)

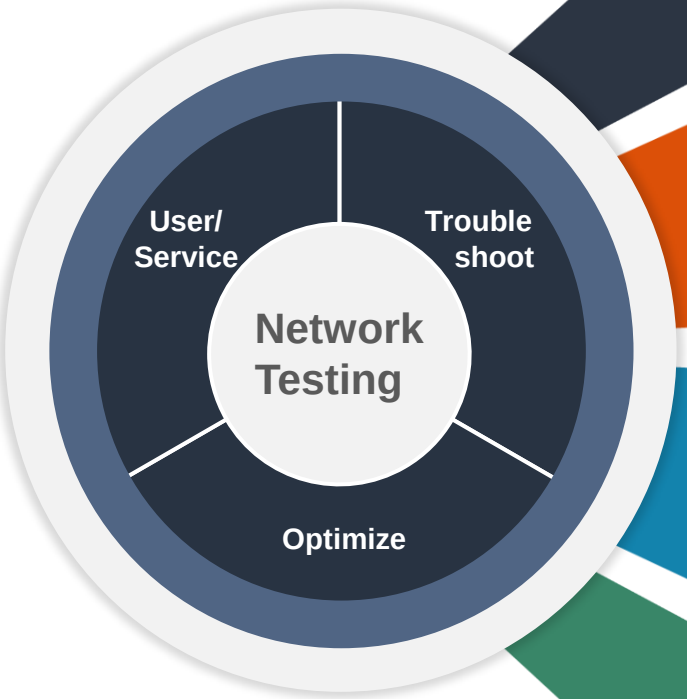
YouTube amazonPrime NETFLIX hulu

5G enablers

5G services/applications

Context aware
user and machine QoE

…with significant impact on network testing



Network complexity/intelligence (technology disruptors)

- mmWave, mMIMO/3D beamforming, high order modulation & coding
- Dynamic ML/AI based network slicing
- Network embedded ML/AI

Devices variety

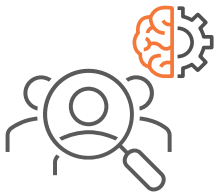
- Characteristics, capabilities, performance,
- Proprietary ML/AI based RF modules (e.g. beam management)
- Services/applications clients/codec

Services/applications/verticals variety & diversity

- Diverse, ML/AI based delivery platforms
- Different server/client performances
- Dynamic OTT versions/releases
- Dynamic proprietary encryption/security

The "new" QoE

- Context aware (network/service/user preferences) >> ML based QoE
- New Types: User, machine(s)
- New Dimensions: interactivity
- Models to cope with non-deterministic network metrics interdependency



What are generic testing techniques?

infovista

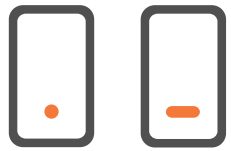
Why take a generic testing approach?

Challenges

Network statistics provide **very little insight** into a customer's QoE across apps and services

App and service **performance is critical** to overall satisfaction with the network

It is **impossible to test** the performance of **all the apps** and service available



Generic Testing Approach

Benefits

Practical and cost-effective approach which closely mimics real apps and services

Delivers **trustworthy results** which are highly correlated to real-world testing

Confidence the network will **deliver the expected user experience** across all apps

Three generic user experience testing techniques examples

OTT voice testing

OTT voice quality testing with a generic client approach

OTT media testing

OTT application testing with a generic framework approach

Interactivity Scoring

User interactivity testing with generic OTT service/application traffic patterns

Infovista’s TEMS generic testing solutions within ITU SG12?

infovista

What are Infovista's TEMS generic testing solutions?

****Generic framework for NATIVE OTT media apps**

WHY: test automation for time efficiency, consistency with same set of ETSI defined KPIs across OTT apps

WHICH apps: various native OTT media apps (e.g. conversational, streaming, posting, etc.)

***GENERIC OTT voice with sQLEAR** MOS scoring (**ITU-T P.565.1**)

WHY: large variety of OTT voice apps with proprietary codec/clients to cope with

WHICH apps: generic OTT voice app emulating WhatsApp

Validation: similar performance behavior for generic and native OTT app (ETSI defined)

Generic OTT Testing

*****GENERIC OTT interactive apps** emulated with traffic patterns for determining UX interactive score (**ITU-T G.1051**)

WHY: need for a feasible solution for drive testing scenarios

WHICH apps: commonly used native OTT interactive app replaced by packet traffic patterns using modified/adaptive TWAMP (ITU-T G.1051)

Validation: traffic patterns to trigger same network resources as a native app would do

*TEMS gOTT voice AQM
**TEMS gOTT media apps
***TEMS gOTT interactive

Generic OTT voice testing

infovista

Generic OTT voice testing

OTT voice quality testing using a generic client approach

Challenge

- Testing mobile OTT voice services/applications is important but practically impossible due to encryption, proprietary codecs, error concealment schemes etc.

Solution

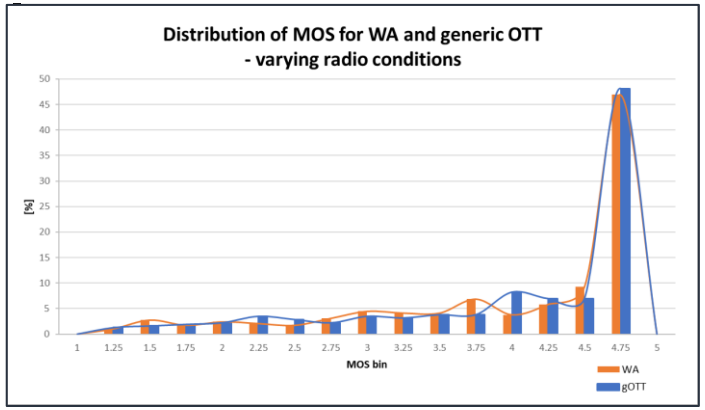
- Infovista's generic OTT voice client accurately mimics the behavior of OTT voice clients (e.g. WhatsApp audio call)

Benefits

- The generic client provides the ability to test only one OTT application, one version and one set of fully accessible KPIs (free of encryption)
- The result is a reference of network performance for OTT voice applications

TEMS generic OTT voice testing explained

- Trained with one of the most common OTT apps (WhatsApp)
- Improved level of accuracy with comparative tests between PJSIP and WhatsApp on the same device at the same time
- Packet-level diagnostic information for detailed analysis and troubleshooting
- RTP data-based KPIs - packet loss, jitter and delay KPIs



Codec

Utilizes the most common codec used within OTT apps

opus

Client

Based on the most common open-source SIP client

PJSIP

VOIP Jitter Buffer Size Increase	Rate of frames being stalled from played out due to increasing the buffer
VOIP Jitter Buffer Overruns	Rate of frames being discarded due to buffer being full or too large
VOIP Decoding Errors	Rate of frames being error concealed due to discards
VOIP Jitter Buffer Underruns	Rate of frames being error concealed due to empty buffer
RTP Lost Packets Rate Audio	Combined rate of underruns and lost packets. Related to frame loss directly due to network issues
VOIP FER Combined Packet Loss	Combined rate of all types of frame loss and discards. Shows very good correlation to the voice quality by including both frame loss due to jitter buffer handling and frame loss directly due to network issues
RTP Jitter Audio RFC3550	Jitter on the down link RTP stream
VOIP Jitter Buffer Playout Delay Average	The audio delay caused by the jitter buffer. Average during one second
VOIP Jitter Buffer Playout Delay Min	The audio delay caused by the jitter buffer. Minimum during one second
VOIP Jitter Buffer Playout Delay Max	The audio delay caused by the jitter buffer. Maximum during one second

MOS Bench test - good radio quality

RAT	WA avg	OTT avg		WA std	OTT std
4G	3.81	3.90		0.43	0.47
3G	3.83	3.99		0.42	0.35
2G	3.23	3.17		0.88	0.90
All	3.62	3.69		0.58	0.57
Difference	**0.07**			**0.00**	

MOS Damper test - bad and varying radio quality

RAT	WA avg	OTT avg		WA std	OTT std
4G	3.17	3.23		1.11	1.16
3G	3.12	3.01		1.14	1.26
2G	2.52	2.47		1.12	1.15
All	2.94	2.90		1.12	1.19
Difference	**0.04**			**0.07**	

All scenarios	3.28	3.30		0.85	0.88
Difference	**0.015**			**0.03**	

infovista

Generic OTT media testing

infovista

Generic OTT media testing

OTT media application testing using a generic framework approach

Challenge

- OTT apps are constantly changing and can differ between devices, countries and even networks – not feasible to test them all

Solution

- Native Python UI automation scripting for setting up the tests solves the changing application challenge
- Generic test methodology and KPIs across all OTT media applications, aligned with ETSI specifications

Benefits

- Generic framework approach allows operators to quickly test any OTT media application with consistency and confidence

infovista

TEMS generic OTT media testing explained

Evaluate QoS of any native OTT application using Python scripts and commercial devices

Video streaming

- Test OTT streaming services in conformance with ETSI QoS standard.
- Get ‘Figure of Merit’ KPI – app agnostic quality trend indicator metric

File sharing

- Share audio, video, picture or document using the native applications.
- Reported events and KPIs include file sharing attempt, success and throughput values

Messaging

- Send text messages using real OTT applications during drive testing.
- Reported events and KPIs include message send start and end to end delivery duration

Remote meetings

- Test remote meeting applications
- Measure retainability and state transition events including success and failure, time and throughput values

Actively test and keep up with the tremendous pace of OTT applications evolution

infovista

Generic OTT interactivity testing

infovista

Generic OTT Interactivity testing

TWAMP based interactivity testing solution

- Interactivity Score is specifically designed to **test latency-sensitive applications**, critical for 5G
- Generic traffic patterns emulate traffic behavior and its adaptability to network conditions in the same way as a real application would

Mobile eGaming

Remote meetings

Video Chat

infovista

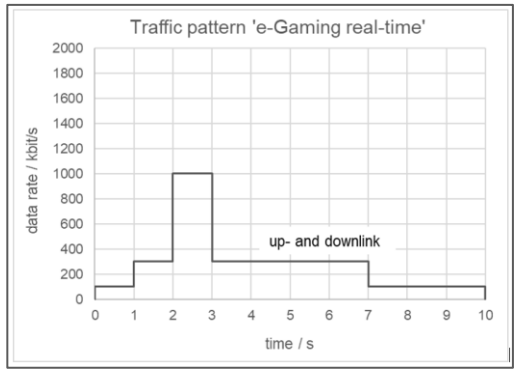
TEMS interactivity scoring explained

Pre-defined traffic profiles (ITU draft proposal) simulate different test scenarios, using TEMS TWAMP server, which include:

- Set of three pre-defined traffic patterns – Mobile eGaming, Remote Meeting and Video Chat
- Interactivity score using TWAMP algorithm, with weighted KPI values
- Chunk-based log reporting for more granular frequent reporting

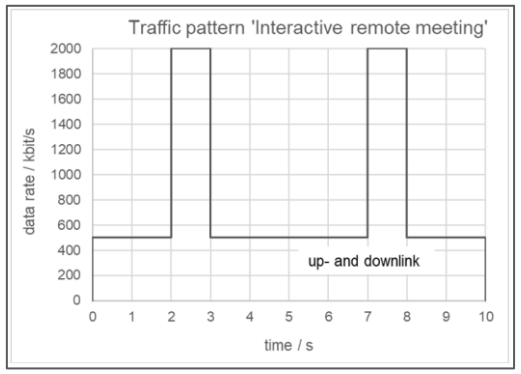
Mobile eGaming

parameter	f_{max}	*a*	*b*	*d*	*e*
value	100	22	6	50	20



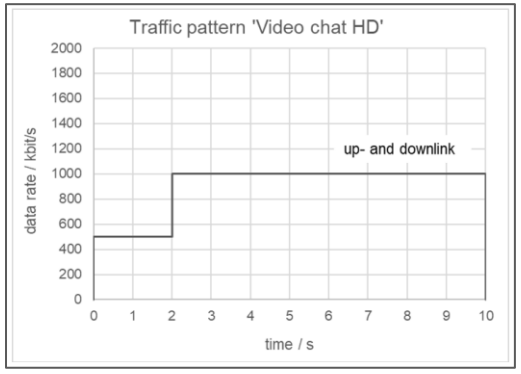
Remote meeting

parameter	f_{max}	*a*	*b*	*d*	*e*
value	100	215	50	500	30



Video chat

parameter	f_{max}	*a*	*b*	*d*	*e*
value	100	215	50	500	30



Sample log report

TWAMP (Quality Index) Packet Train Log Report

Time : 18:17:01.489
ID : 7488
TimeStamp (UtcTime) : 01/24/2022 10:17:00
Packet Train Number : 0
Client Transmission Bitrate : 174.7181
Server Reception Bitrate : 172.4051
Server Transmission Bitrate : 172.1198
Client Reception Bitrate : 171.3617
Inter Train RTT : 23740.2550
Minimum RTT : 2.1550
Maximum RTT : 99.8620
Average RTT : 8.6463
Median RTT : 6.5140
Standard Deviation RTT : 9.6131
Variance RTT : 92.4111
Minimum Jitter : -10.1150
Maximum Jitter : 91.9760
Average Jitter : 0.0004
Median Jitter : -0.8640
Standard Deviation Jitter : 4.6164
Variance Jitter : 21.3115
Jitter Inter Train RTT : 0.0000
PacketLoss : 2.5655
Median Packet Delay Variation : 4.3590
Interactivity Score : 86.5692

Key take aways

infovista

Key take aways

Key factors for generic testing

- OTT app encryption
- OTT variability and diversity extensively based on proprietary ML/AI based media processing/delivery techniques
- Multitude of “no size fits all” QoE models
- OPEX constraints

Generic testing techniques

- Generic OTT voice clients: TEMS gOTT voice
- Generic OTT CG traffic patterns: TEMS adaptive CG traffic patterns and CG Ux interactivity scoring
- Generic framework for testing any OTT app: one drive, one score suitable to drive testing and many/any OTT app

Standardization of generic testing

- Infovista is an active contributor to:
  - ITU-T Study Group 12: Performance, QoS and QoE
  - ETSI STQ Mobile

Questions?

www.infovista.com

Learn more by reading our user experience testing whitepapers

[Assessing OTT voice user experience]

[Validating our generic OTT voice testing approach]

[Assessing mobile gaming user experience]

[Assessing OTT video user experience]