HetNet OpenSchema QoS/QoE Score

Enabling Intelligent User-Oriented and Data Driven Traffic Steering

Wi-Fi Solution Project Group

June 2021

Document version: v1

Confidentiality Level: **Green [Public Access]**

Authors

Magnus Olden
Domos
magnus@domos.no

Behrad Analui
Shoelace Wireless
behrad@shoelacewireless.com

Contributors

Many ecosystem partners contributed and continue contributing to this effort by generously providing their methods and experiences or participating in reviews, integrations and trials:

Organization	Individual / Team
Deutsche Telekom (DT)	Sascha Dech Seamless SDK Team
Shoelace Wireless	Behrad Analui Jim Mains Mervin Mathew
Domos	Magnus Olden Bjørn Ivar Teigen Olav Nedrelid
Vodafone	Gavin Young
Plume	Bill McFarland Atieh Madani
NetExperience	Marcel Chenier
Boingo Wireless	Peter Barany
Celona	Srinivasan Balasubramanian
Parallel Wireless	Kaitki Agarwal
Taoglas	Carlo Terminiello
Facebook	Evgeniy Makeev Jacky Tian Brian Barrit Shah Rahman Rajesh Rasalkar
Aspire	Declan Friel Padraig O'Seighin

TIP Document License

By using and/or copying this document, or the TIP document from which this statement is linked, you (the licensee) agree that you have read, understood, and will comply with the following terms and conditions:

Permission to copy, display and distribute the contents of this document, or the TIP document from which this statement is linked, in any medium for any purpose and without fee or royalty is hereby granted under the copyrights of TIP and its Contributors, provided that you include the following on ALL copies of the document, or portions thereof, that you use:

1. A link or URL to the original TIP document.

2. The pre-existing copyright notice of the original author, or if it doesn't exist, a notice (hypertext is preferred, but a textual representation is permitted) of the form: "Copyright © <<year>>, TIP and its Contributors. All rights Reserved"

3. When space permits, inclusion of the full text of this License should be provided. We request that authorship attribution be provided in any software, documents, or other items or products that you create pursuant to the implementation of the contents of this document, or any portion thereof.

No right to create modifications or derivatives of TIP documents is granted pursuant to this License. except as follows: To facilitate implementation of software or specifications that may be the subject of this document, anyone may prepare and distribute derivative works and portions of this document in such implementations, in supporting materials accompanying the implementations, PROVIDED that all such materials include the copyright notice above and this License. HOWEVER, the publication of derivative works of this document for any other purpose is expressly prohibited.

For the avoidance of doubt, Software and Specifications, as those terms are defined in TIP's Organizational Documents (which may be accessed at https://telecominfraproject.com/organizational-documents/), and components thereof incorporated into the Document are licensed in accordance with the applicable Organizational Document(s).

Disclaimers

THIS DOCUMENT IS PROVIDED "AS IS," AND TIP MAKES NO REPRESENTATIONS OR WARRANTIES, EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, NON-INFRINGEMENT, OR TITLE; THAT THE CONTENTS OF THE DOCUMENT ARE SUITABLE FOR ANY PURPOSE; NOR THAT THE IMPLEMENTATION OF SUCH CONTENTS WILL NOT INFRINGE ANY THIRD PARTY PATENTS, COPYRIGHTS, TRADEMARKS OR OTHER RIGHTS.

TIP WILL NOT BE LIABLE FOR ANY DIRECT, INDIRECT, SPECIAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF ANY USE OF THE DOCUMENT OR THE PERFORMANCE OR IMPLEMENTATION OF THE CONTENTS THEREOF.

The name or trademarks of TIP may NOT be used in advertising or publicity pertaining to this document or its contents without specific, written prior permission. Title to copyright in this document will at all times remain with TIP and its Contributors.
This TIP Document License is based, with permission from the W3C, on the W3C Document License which may be found at https://www.w3.org/Consortium/Legal/2015/doc-license.html.

Change Tracking

Date	Revision	Author(s)	Comment
10/2020	v0.1	Magnus Olden, Domos	1st Draft version
04/2021	v0.2	Behrad Analui	2nd Draft version
06/29/2021	v1.0		Final Version

Table of Contents

1

Executive Summary

1. Executive Summary

Heterogeneity is a core characteristic of current and, more importantly, future networks to support the ever accelerating demand in usage and everywhere, always connected, access. With each new network generation, new wireless access technologies are being introduced while the previous ones mature and become more widespread. At the same time, new user equipment (UE) technologies have become increasingly capable of utilizing multiple radio access technologies, thus leveraging all available access networks, whether licensed or unlicensed spectrum based, for expanded and affordable network coverage. Heterogeneous Network (HetNet)[1] environments provide a tremendous opportunity for the wireless communication ecosystem community to innovate by deploying new services that can drive additional revenue streams and can also reduce Operator CAPEX/OPEX spend by optimizing spectrum efficiency and utilization, both while improving customers and end users' quality of experience.

To realize the HetNet opportunity, UEs need smarter decision making capabilities, in general, but especially in the last mile of connectivity for the radio access network. Enabling smarter decision making on UEs as well as the access side HetNet elements (e.g., APs, Gateways, Small Cells, and etc.) requires a unified and consistent framework for performance related metrics measurement, collection, reporting, and analysis. This unified framework facilitates building online and offline learning models and applications using machine learning and AI. For example a standard QoS/QoE score for HetNets can make it easier for UEs to select the best and most cost effective network for the right time, place, and function.

Current approaches to network metrics collection have some disadvantages, for example:

- They are a partial representation of the state of the network since these metrics only focus on one point of view (e.g., from network out)
- They often require vendor specific integration which makes the cycle of build and innovation complex and long or
- They are based on diverse and inconsistent measurement methods and tools (e.g., some prefer latency over bandwidth, or prefer using one tool over the other to measure certain metric)

Magma Unified HetNet OpenSchema Data Lake is an open source effort to create a standard framework by defining a unified schema for metrics collection across all elements of the access network (e.g., UE and radio equipment). Magma converged core solution is very well positioned to facilitate HetNet metrics collection. We are currently developing a scalable unified data lake pipeline to process and store the metrics collected using HetNet OpenSchema. This unified data lake pipeline will be integrated as a Service Management and Orchestration(SMO) plugin component.

[1] HetNet in this document is used as a generic concept, where multiple wireless network technologies are available, and not to be confused with specific use of HetNet in coexistence of LTE/NR macro and small cells. For the latter please refer to: https://www.3gpp.org/technologies/keywords-acronyms/1576-hetnet

This white paper will review multiple methods and approaches presented in academia and industry that address the problem of measuring network performance and improving QoE and QoS. The paper focuses and tests several core HetNet use cases such as:

- Seamless HetNet transition[2] and UX improvement
- HetNet aggregation[3]
- Traffic/Application Predictive Steering
- Analytics for network asset positioning
- Handovers and Mobile Data Offload decision

In the future, HetNet QoS/QoE Services will be implemented as part of Magma Unified Service Management and Orchestration and synergies with other QoE focused efforts, such as O-RAN and NR-RIC's(near real time and non-real time radio intelligent controller) UE and Cell metrics collection will be explored with the goal of creating a unified schema to facilitate a 360 degrees view of the access network.

[2] HetNet transition refers to UEs moving in and out of a HetNet environment. An example of this condition is when users walk into a mall with one or more public Wi-Fi available.
[3] HetNet aggregation refers to a situation where two or more wireless technologies are combined to provide a better user experience.

2

Project Overview

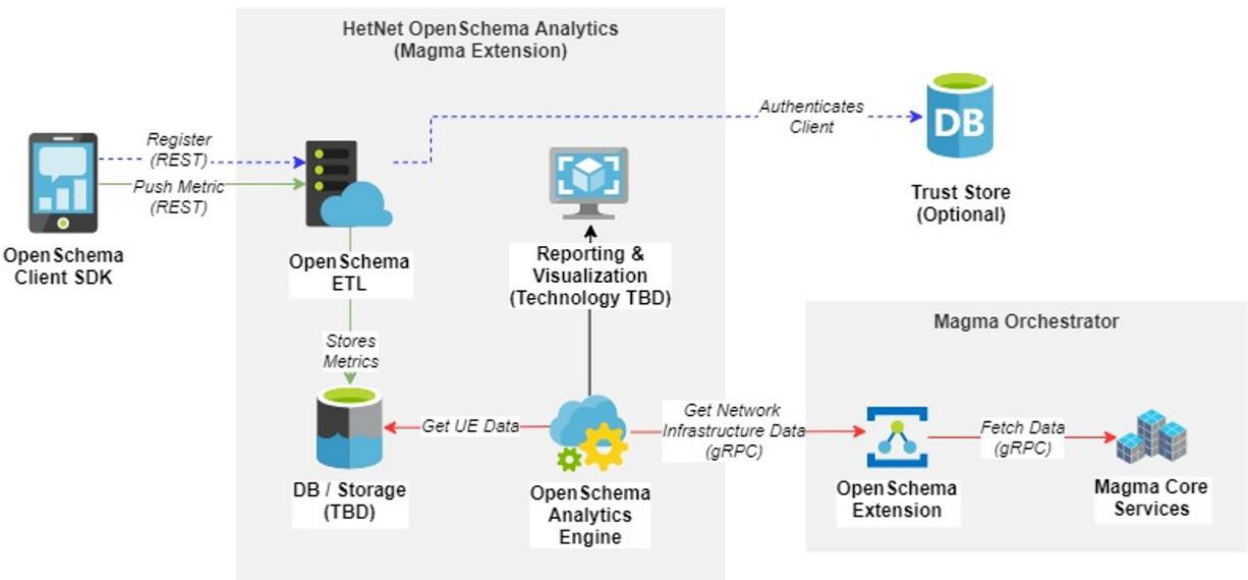
2. Project Overview

2.1 HetNet OpenSchema

Heterogeneity and coexistence of multiple radio access wireless networks is key in the current and future generation networks. This means there are multiple wireless networks available for the User Equipment (UE) to choose from, or UE can aggregate multiple wireless networks to provide a better quality of experience (QoE).

Independent and non-standardized capture and storage methods of network and device related metrics data increase integration complexity and slow innovation. OpenSchema is a unified, plug and play model to identify, collect, store and retrieve metrics and statistics from user equipment (UEs), CPEs and different elements of Access Networks in a Heterogeneous Networks (HetNet) environment. Adopting OpenSchema as a common data store for network and device related metrics can provide faster and easier integration for eco-system innovation and collaboration for MDO as well as enable new services for monetization and user experience enhancements.

Picture below shows the end to end vision of OpenSchema UE data collection as a plug and play extension in Magma unified Service Management and Orchestration (SMO).



The OpenSchema architecture as a Magma Service Management and Orchestration (SMO) Plug-In, facilitates separation of scalability discussion from other SMO[4] components. Additionally a unified data lake

---

[4] Please refer to https://www.magmacore.org for more and update info about Unified SMO

facilitates ingestions of metrics from other HetNet Access sources such as APs, WACs, GWs and etc. With this architecture, building and deploying data intensive applications such as HetNet QoE will be much easier.

project.

3

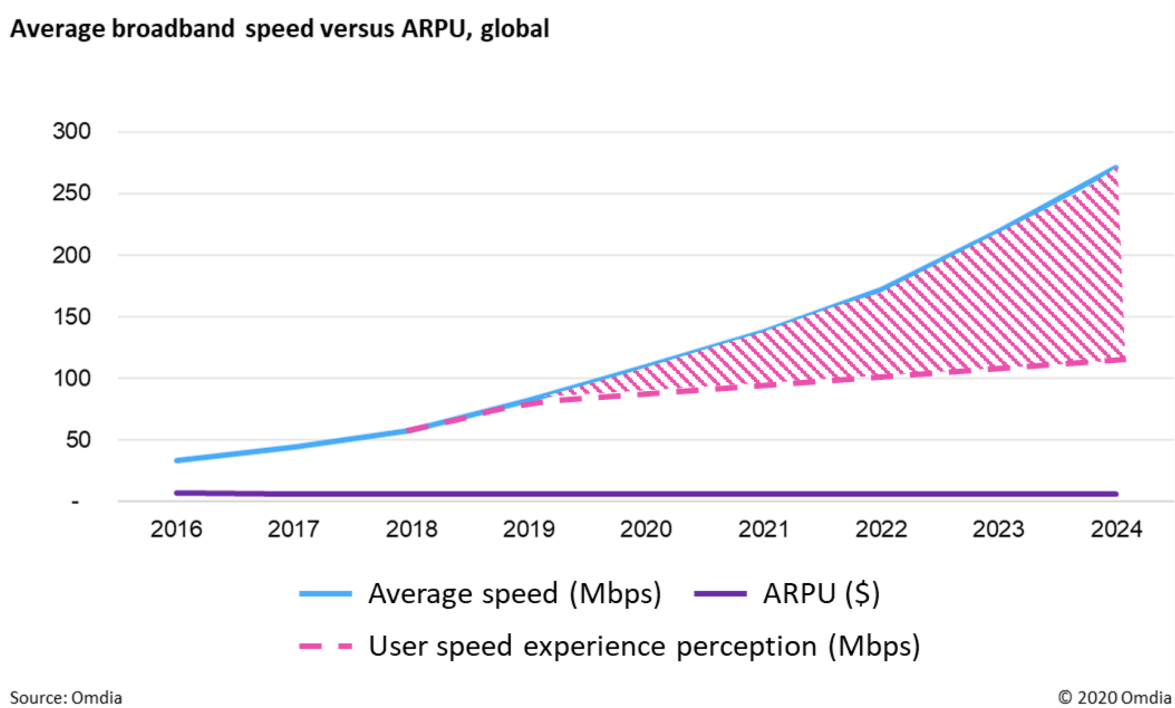
Network Quality and Quality of Experience

3. Network Quality and Quality of Experience

Understanding and measuring network quality is harder than one might think. For starters, network quality is not a single agreed upon thing. It is measured in many different ways [1]...[12], and differently across network technologies. Comparing these measures is often impossible.

But that is just the start of the difficulties. Network Quality has many dimensions (bandwidth, latency, packet loss, variations) and what "*better* network quality" means is difficult to reason about.

The returns on increasing Mbps for user experience are diminishing [see graph below]. The vast majority of users will not have any better experience upgrading for 100 Mbps to 1Gbps. There are no applications, except large downloads, commonly used today that would see a difference going from 1Gbps to 2Gbps.



Calculating what a megabit per second is worth is a complex, but not impossible, issue. Putting a perfect price on a Mbps in a proprietary technology is likely impossible (or requires vast amounts of reverse engineering). Doing so with proper techniques on Open Networks is possible, and opens a new world of end-to-end optimization and understanding of user experience.

First, we need to answer the question, what is so difficult about network quality? Turning the question around, what are the questions we would like to be able to answer about network quality:

From an end user perspective:

- Will this application work on this network?
- Why does the application work sometimes, but other times fail?
- Which network should I use?

From an operator perspective:

- Why are most networks over dimensioned?
- How to dimension a network for the greatest ROI?
- What investment in network equipment has the highest ROI?
- How are the network quality and Quality of Experience coupled?
- How can networks be efficiently priced?
- What are the optimization criteria and how can we optimize a network?
- If there are multiple paths, what comparative metrics can be used to determine a preferred path to use for packet routing?
- What are the implications of choosing a potentially suboptimal route?
- Which services require seamless connectivity and how can we ensure seamless connectivity for mobile devices?
- Why is root cause analysis so difficult?

Many of these questions require a good understanding of the deployed networks along with their interworking and needs sufficient intuition to be developed for optimal operation of the network. Let us postulate the goals of better measuring network quality:

Better end-user experiences, at the lowest possible cost

Although this seems obvious, even with QoS guarantees, the current networks are optimized for system capacity and peak individual user throughput.

3.1 Difficulties in measuring network quality – A top level view

- There are many different ways of measuring network quality and are dependent on:
    - Different technologies (4G,5G, Wi-Fi, fiber, cable etc. and their generations and associated bands of operation)
    - OSI layers and the associated nodes hosting the functions
    - The perspective end user Equipment and Network Equipment

- The metrics that need to be collected and the determination of the network quality tends to be non-trivial.
    - The required metric collection can span nodes and aggregating the information accounting for the timing of the information being collected can be complex.
    - Do counterfactuals. What if we had upgraded this network equipment

- The end user experience is knowing the user / user equipment and require indirect measurements are required to determine the performance experienced at the network
- Quality of Service guarantees based on assumed performance expectations may not directly map to Quality of Experience
- There are multiple dimensions of network quality, making it difficult to understand what “better” really means. More bandwidth does not always translate to better end-user experience.
- Applications have complex contextual behaviors and the network tuning for better performance is very much dependent on a given specific application.

Research shows that there is a link between perceived network quality and ROI. But perceived network quality depends on the application, which means there is no trivial way of pricing network usage or network quality because the link between network quality and Mbps is broken. What a Mbps is worth is guesswork, often driven by the idea that more Mbps will create better user experiences . That worked great from the kilobit era to the 100 Mbps era, but is now failing.

3.2 Speedtest and Megabits per second are misleading

A speed test does not contain enough information to answer several of the questions previously posted. For example, it does not answer whether an application (other than the speedtest) will work on the network. Nor does a comparison between two speedtests on two different networks necessarily tell you which one is better for you. Let's get into the details of why that is.

Complication

You have to look at Bandwidth, Latency and Packet Loss to understand End-User Experience

3.2.1 Speedtests are momentary

Most network equipment is in one way or another shared between multiple users. That means the number of resources you have available changes, which is why two speedtest results on the same network are rarely identical. Variations in network quality is an issue in itself, which we will come back to later. The larger point is that while there is information in a speedtest, a network can be too volatile to make that information useful in many contexts

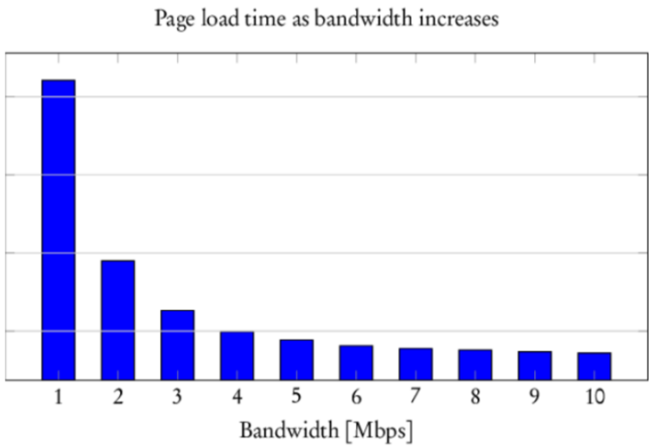
Complication

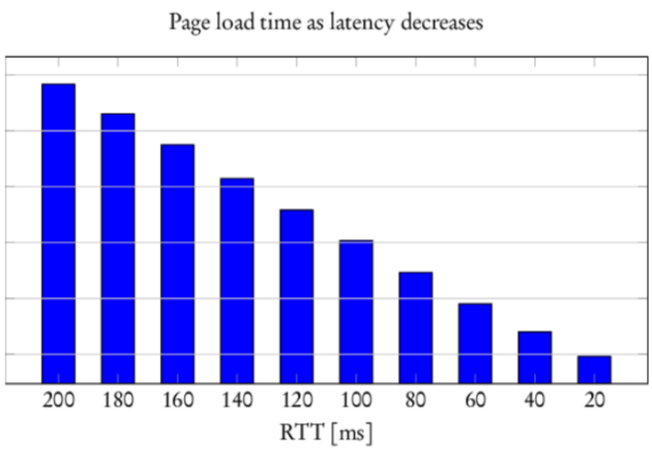
Network Traffic loads from other users are always impacting you. Your network quality is constantly changing

3.2.2 Mbps does not translate to user experience

Megabits per second is the dominant network quality metric, but unfortunately it does not translate well to end user experience. There are other dimensions of network quality, namely packet loss and latency.

When you order a Pizza, what do you care about? The amount of time it takes to get the Pizza? Or the size of the delivery van? When you browse the web, what do you care about? The amount of time it takes to load a page? Or the amount of Mbps? Exclusively looking at Mbps is analogous to exclusively looking at the size of the pizza delivery van. Its *"Pizza delivered in 20 minutes"* vs *"Pizza delivered with semitrailer" and "Page Loaded in 1s"* vs *"Page Loaded with 1Gbps"*





Source:
https://www.researchgate.net/publication/299368164_Towards_a_Low_Latency_Internet_Understanding_and_Solutions

Page Load Time (PLT) of a website as a function of bandwidth (top graph) and as a function latency (bottom graph) from High Performance Browser Networking. As we can see on these graphs, there are next to no returns from increasing bandwidth above 8 Mbps. Reducing latency on the other hand linearly reduces PLT.

Video Conferences are another example. They can rarely handle more than 100-300 ms latency, but require no more than 4 mbps. increasing from 4 to 1000 mbps has no impact on the Video Conference. As is cloud gaming and any collaboration activity.

Application	Mbps requirement	Mbps point of diminishing returns
Video conferencing HD	4 (up/down)	8 (up/down)
Video conferencing 4K	25 (up/down)	50 (up/down)
Video Stream HD	4 (down)	8 (down)
Video streaming 4k	25 (down)	50 (down)
Online Gaming	2 (up/down)	5 (up/down)
Cloud Gaming	5 (up/down)	25 (up/down)
Web Browsing	5(down)	10 (down)

Many in the telecom industry are discussing what to do beyond the gigabit (1000 Mbps), but the question remains, why are there still poor user experiences when we supposedly have many times the requirements?

Complication

There are situations where you could have identical speedtest results, but vastly different end-user experience.

You can easily optimize a network in such a way that you increase throughput (Mbps) but create poorer user experience due to increased latency or packet loss. In fact, the bufferbloat project [**13**] exists for the sole reason that someone has optimized for Mbps by increasing queue sizes, causing more efficient bandwidth use. But it also vastly increases latency, by making packets wait for long times in queues. This has degraded user experiences. More about the dimensions of network quality and how they are coupled later in this paper.

Complication

You can increase Mbps and create a worse end-user experience

Megabits per second is also just an average over a second. Averages are often misleading and packets are sent on a milli- or micro-second scale. Plenty of important information is lost in this average, for example, sending 50% of packets on 10 Mbps and 50% on 20 Mbps can create a worse user experience than sending everything on 15 Mbps.

Complication

Are you looking at network quality per second? millisecond? minute? hour? day? Are you averaging? Sampling? Statistical analysis? Two identical averages for longer time periods can have vastly different quality of experience

3.2.3 Not all bandwidth is created equal

Driving on an empty road is different from driving on the same road when congested. In the same way, 50 mbps on an empty network is not the same as 50 mbps on a loaded network, due to queueing and buffering. Getting onto a 2-lane road with a single busy lane, is different to getting onto a 10-lane road with a single busy lane. In the same way, 50 Mbps on a 100 Mbps port is not the same as 50 Mbps on a 10Gbps port, due to serialization delays.

There is no physical layer consistency. 100 Mbps on Wi-Fi is not the same as 100 Mbps on DOCSIS or LTE or PON. Let's use Wi-Fi as an example. The Wi-Fi bandwidth commonly reported in sales is the highest achievable data rate, under ideal conditions. However, a Wi-Fi router does not have access to send data for an entire second, meaning it can't achieve that data rate for a full second. Let us rephrase that, a Wi-Fi router with a 5.4 Gbps max speed cannot send 5.4 Gigabits per second, because it has to wait some milliseconds between transmissions to assess if the channel is clear. It can however send data at a 5.4 Gbps rate while it has access to the channel. For fiber, this is different.

Complication

Different network technologies measure bandwidth differently

Insight

End-to-End Network Slicing with QoS Requirements will not work well when different technologies interpret the QoS Requirements differently.

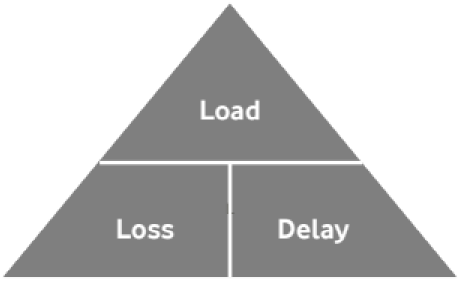
3.3 The Coupled dimensions of network quality

Packet loss, latency and bandwidth (aka load/throughput) are the three main dimensions of network quality. But how are they coupled?

What happens when you have insufficient bandwidth? Your packets get put into queues. Waiting in these queues causes a great variation in latency. That is the coupling mechanism between bandwidth and latency. There are of course some parts of latency that are not caused by queues, but by physics, like the speed of light and by processing. So, one part of latency is caused by things like distance and packet size, another part of the latency is caused by momentary insufficient bandwidth at one or more network links.

What about packet loss? What happens if the queues are full? The network equipment may drop the packet. Getting into the higher levels of the network stack, you can also get packet loss by TCP timing out the packet because it has taken too long. So, packet loss is coupled with latency in two ways: The queues get so long that the Network Equipment drops the packet. The second way: The overall round trip time gets so long that TCP considers the packet lost. But there are other forms of packet loss. Collisions, bit errors, faulty routing, etc. Some technologies, like Wi-Fi, retries certain types of packet loss. Some other network technologies do not. Since TCP retransmits packets if it deems them lost, the application that is sending the data will perceive the packet loss as higher latency (as it does not realize the packet was ever lost, it just took a lot of time to get back).

As a short summary: Insufficient bandwidth causes queues, which causes latency, which may cause packet loss. But not all packet loss is caused by latency and not all latency is caused by insufficient bandwidth.



"Insufficient bandwidth" really means: "At the offered load, the resulting **packet loss/delay** exceeds the acceptable performance bounds of the user's application"

3.3.1 Network Quality constantly changes

Almost all modern network equipment is in some way shared by multiple users. That means that which share of the network a single user has access to is constantly changing. Bandwidth, latency and the chance of Packet Loss is constantly changing. These changes complicate network quality a lot, for two main reasons:

1. Many Applications react to their perceived network quality, if they perceive the network as great, they may try to consume more by increasing resolution and other qualities. Changes can often be worse than plain low quality.
2. What the network looked like a second ago may tell us very little about what it will look like in a second.

> Complication
>
> Knowing what a network will look like at any point in the future is guesswork. There is no certainty. We're dealing with probabilities

3.3.2 Jitter, Stability, and Reliability

As throughput varies, queues materialize and queues create latency and packet loss. There is a language developed for the variation of latency and packet loss. For Packet Loss there are two axes of variation.

Reliability: The chance this packet will make it to its destination. Example: a network where 50% of packets are reach their destination have a low reliability (Most networks perform at 99+%)

Stability: The change of reliability over time. From the previous example: If the network always has a 50% chance of packets reaching the destination, it is stable (we can call it stably unreliable). An example of low stability is a network where one second 99% of packets reach the destination while the next second 80, and the next second 100%, and the next second 50%.

Throwing a dice to get a 6 is unreliable as you will on get 6 about ⅙ of the times, but it is highly stable, keep doing it and you will get a 6 roughly a ⅙ of the times no matter how long you keep at it. Sports are often more unstable, a team may reliably win most games through a season, but a few seasons down the line they can reliably lose games (there are of course exceptions, but you get the point).

A wise reader will notice that we once again run into a problem of at what time scale should we measure. If I look at the stability example across the 4 seconds one may conclude it is stable at 82.25% (99+80+100+50/400) reliability. But by measuring the network at the second mark it looks unstable. If I were to measure it at the millisecond it may look even more unstable, or more stable. It depends on the statistical characteristics.

Insight

Our choice of timescale for a network quality metric changes the perceived variations.

For latency the variations are often referred to as Jitter. Jitter is more well documented and there are plenty of resources where you can read about jitter.

3.3.3 Increasing bandwidth does not, necessarily, decrease latency, and packet loss

Unless you increase bandwidth at every point at once, from end to end, you are in no way guaranteed to decrease latency by adding more bandwidth. In fact you may increase latency or jitter by increasing bandwidth at a network node. How does this work? Network traffic goes through many networks' nodes on its way to its destination. Just like adding more lanes onto a ramp to a highway can increase the amount of congestion(and thereby "latency"), the same thing will happen in networks.

Insight

Increasing bandwidth is guaranteed to reduce latency if, and only if, either:
a) Bandwidth is increased end to end
b) The network node where bandwidth is increased was **always** the bottleneck
As available network resources are always changing, it is not very often a single link is always the bottleneck.

3.3.4 Optimizing for bandwidth can increase latency and packet loss

As previously mentioned, optimizing for bandwidth can increase latency, for example by increasing queue sizes blindly, you will get more efficient bandwidth usage, but without considering the increased latency you will often create worse user experiences. This phenomenon is often called bufferbloat. Decreasing the queue sizes too much on the other hand will cause more packet loss. Sending data at rates close to the SNR limit will have a very high bandwidth, but also a lot of packet loss.

3.4 Applications

3.4.1 Applications are complicated

If our goal is to improve the end-user experience at the lowest possible cost, it means we have to understand how different applications (YouTube, Netflix, Teams, Meets, Zoom, etc.,) and different classes of applications (Streaming, Web-Browsing, Video Conferencing, Voice, Download, Backups, Online Gaming, Cloud Gaming) respond to different levels of network quality.

3.4.2 Application types

Below is a table of application types, examples and what their primary network quality dimension is that causes them to fail. That does not mean that they don't rely on other dimensions, only what most often causes them to fail.

Application Type	Application examples:	Primary Network Quality Dimension(s) of failure
Interactive	Video Conferencing, Online Gaming, Cloud Gaming, voice, collaboration tools (google docs), Remote Robot Surgery	Latency (Absolute, Jitter), packet-loss (Reliability)
Interactive with hidden buffering	Video Conferencing	Latency (Jitter), Packet-Loss (Stability)
Download driven	Video Streaming (Netflix, YouTube), downloads	Bandwidth (down)
Page Load Time / sequential driven	Websites	Latency, packet-loss (Reliability)
Upload driven	Data Backups, Broadcasting (Twitch)	Bandwidth (up)

Insight

Network 1: Ranging from 100 to 300 mbps and latency ranging from 1-100.
Network 2: Consistently 10 mbps and 100 ms latency.

Which one creates the best Video Conferencing?

Network 2, which by looking at averages would look a lot worse: 200 mbps and 50 ms latency vs 10 mbps and 100 ms latency

Complication

Skype, Zoom, Microsoft Teams, Google Meets, WebEx, etc. have proprietary techniques for handling packet loss and latency. The same network quality can give different results for different video conferencing solutions

3.4.3 What does better network quality mean?

We can only be sure the network is "better", if it improves all dimensions of network quality, also the statistical components are improved. You can for example, improve throughput, packet loss and average latency and create poorer QoE (by increasing jitter). In general, "better" network quality depends on the application being used. In this context "better" means better application outcomes, i.e. web pages that load faster, removing lag from video conferences, higher definition streaming. There is a subjective component of user-experience which may be captured by metrics such as NPS, but is out of scope for this document.

3.4.4 Choosing a networking layer for the measurement

Networks are divided into different independent layers. At what layer should you measure network quality? There is a fundamental issue regarding choosing a layer to measure. At the highest layer, the TCP/UDP IP layer you can measure something that is possible to map to end-user experience. At lower layers, it is easier to do root cause analysis. However, at the lower layers, we are becoming dependent on technology specifics, which ruins many use cases. The best solution would be to measure at TCP/UDP IP layer.

3.4.5 Where should we be able to take the measurement

At the end-user device? At any given Network Equipment? To answer all the stated questions, we should be able to measure network quality in all of these points. There is useful information at every point. At the end-user equipment we are the closest to the end-user experience, and there are things that are only possible to know from this perspective. But there are plenty of things you cannot know from an UE perspective. There is a need for a single network quality measurement that works across technologies.

To summarize, network quality is complicated. But not impossible. We have to remember that you can decrease user experience when increasing Mbps, which means you can spend money upgrading a network equipment that creates poorer user experience. That seems like bad business.

Today, you can't look at a network measurement and say anything certain about the end-user experience. Which means you can't map network equipment to end-user experience, NPS, Support calls or churn in a meaningful way.

Closed networks have the disadvantage that you either have to rely on proprietary network quality metrics or measure the closed part of the network in a larger context, there is always information lost that way. Open Networks can measure everything in the same way, at every point. And by that get a huge advantage for better investment and optimization. Telecom can become data driven.

By understanding what parts of the networks the network quality bottleneck are, and importantly which dimensions of the network quality is the bottleneck, we can make much wiser investment decisions. It is highly likely that a better investment than new equipment can be better latency management at a certain node, and also a lot cheaper.

If the network quality is measured correctly, we can do much better:

1. Dimensioning
2. Testing
3. Investing
4. Mapping network quality to user experience
5. Mapping from user experience to NPS or similar
6. Counterfactual
7. Pricing
8. Finding the best paths
9. Optimizing
10. Create better user experiences

3.5 Quick look at the State of HetNet QoS/QoE measurement methods

3.5.1 QoE/QoS in Academic Research

Vertical handover in the HetNet environment has been the subject of numerous academic research ever since the introduction of UEs with multiple radio access technologies (multi-RAT). Depending on the higher level policies, for example mobile data offload, handover decisions should be based on measurements from available wireless networks. Earlier studies focus mostly on the metrics from the network side. These metrics are often collected on the Access Point or Cell level or collected by UE and sent over specific 3gpp interfaces. More recent studies are more user oriented and focus on QoE. We did an evaluation of selected research on user oriented vertical handover methods[14]. Table below is summary of academic HetNet decision making algorithms, from: “State of Art: Vertical Handover Decision Schemes in Next-Generation Wireless Network”:

3.5.2 HetNet QoE/QoS in Standards Organizations

	existing vertical handover decision schemes	advantage(s)	disadvantage(s)
single metric	RSS-based schemes	simple design	increased unnecessary handovers increased ping-pong effect
	bandwidth-based schemes	good throughput performance good network selection	inefficient bandwidth computation
	cost function	less call drop probability reduced ping-pong effect	increased system overload
multiple metric	context-aware schemes	very less call-drop blocking good context collection	complex design leads to implementation issues
	media-independent handover	good network selection reduced latency	high-signaling overheads increased resources consumption
	MADM schemes	better decision on dynamic parameters	performance dependence on traffic class
computation	FL	user-satisfied handovers reduced handover delay	heavy complex design higher processing time
	NNs	successful handovers reduced handover delay	heavy complex design centralized control
multi-attribute + computation	MADM-AI	reduced handover decision delay precise data for handover decision	heavy complex design terminal-based decision huge training process
	MADM-context-aware schemes	improved QoS for user	unreliable handover decision at high speed terminal-based decision

In this section we will have a quick overview of how some of the Network and Wireless standard organizations look at HetNet QoE/QoS measurements or use cases. Please refer to [14] for more structured review of these methods using methodology in Appendix B.

3GPP ATSSS

Access traffic steering, switching and splitting(ATSSS) is the 3gpp rel.16 standard for 5G systems to facilitate utilization of non 3gpp access networks. UEs route traffic to access networks using route selection policy(URSP). Since 5G promises a high level of QoE, it is imperative that proper measurement of all access networks is required for an efficient URSP delivery.

Wireless Broadband Alliance OpenRoaming™

OpenRoaming™ provides federation among Identity Providers and Access Providers to facilitate Wi-Fi integration, Seamless Roaming and high quality Mobile Data Offload. OpenRoaming™ defines QoS profiles with some minimum bandwidth requirements, but the exact QoE requirements depend on the implementation.

Broadband Forum's ΔQ

The idea of quality attenuation or ΔQ, is to go beyond speed test, and use multiple metrics(bandwidth, latency, jitter, loss and etc.) of different applications, to create a statistical model that can later be used to get more network insights, Root Cause Analysis, Wi-Fi slicing equipment selection and others.

O-RAN Alliance

Traffic Steering and QoE Optimization are two important use cases of O-RAN phase1 class. O-RAN provides detailed specification for interfaces to collect UE based metrics and deliver policy and intent to UEs. Although UE metrics are collected on the network side using E2 interface, there is potential synergy where unified metrics collection(UE application based metrics in addition to UE network side metrics) can be utilized for RAN intelligent controller (RIC) application development. Additionally integration of non 3GPP access (for example managed and unmanaged Wi-Fi) metrics could improve HetNet use cases implementations.

4

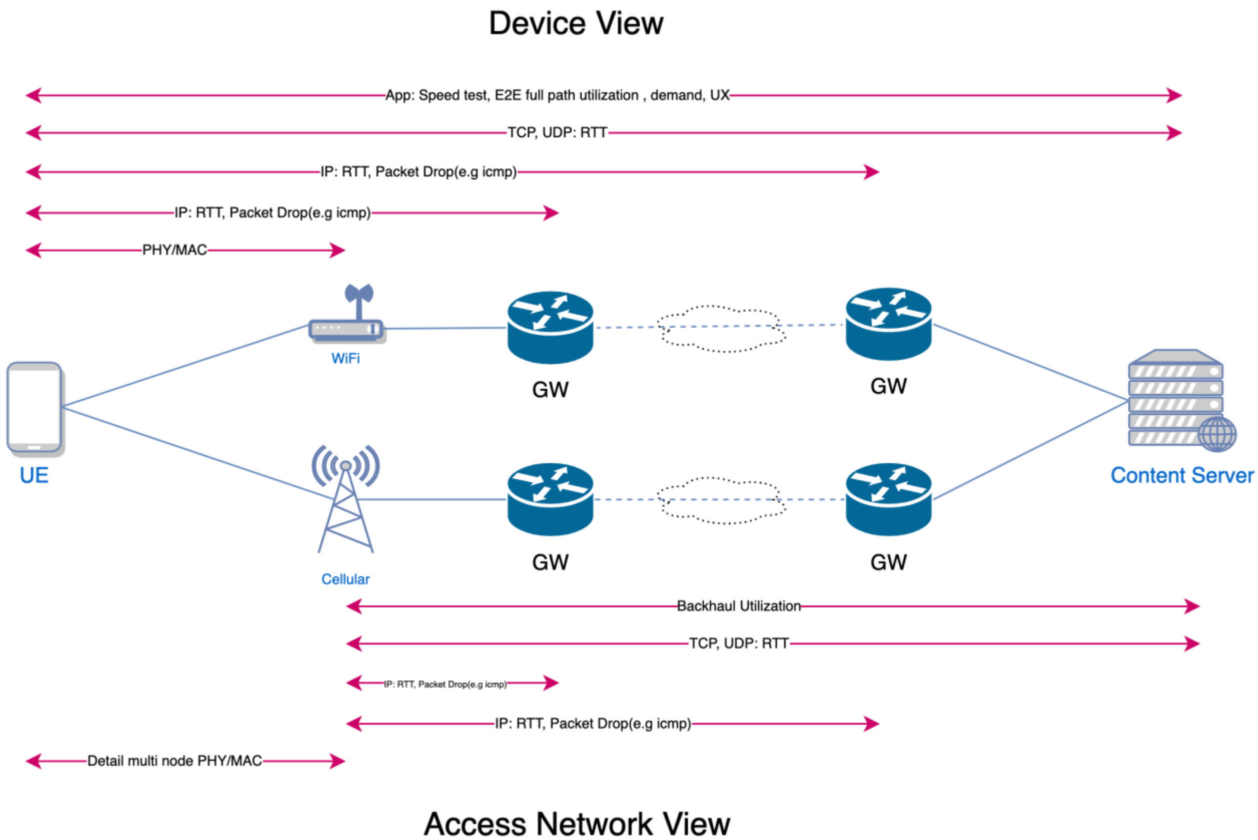
HetNet QoE Score Use Case

4. HetNet QoE Score Use Case

4.1 Deeper Look at HetNet Challenges and Opportunities

HetNet is the environment where multiple wireless access networks are available and there are UEs with the ability to connect to these different network access technologies. A familiar example is mobile phones that can choose between Wi-Fi and 4G/5G. HetNets face a challenge very similar to you on your vacation in that weird foreign country whenever deciding which network path to choose. In a scenario where you need a Video Conference to work, should you choose 4G/5G or Wi-Fi? It depends on which is best. Ideally, the phone would know which is best. But it is surprisingly tricky.

The point of view for the phone and you as a bewildered tourist is similar. You don't know what paths are congested. You don't know if a path is blocked. Also, there is this weird thing where roads report speed differently. The theoretical limit of semi-trailers per hour (achieved with perfect driving in simulation) has very little to do with how fast you can get to your destination. Something analogous to "the theoretical number of semi-trailers per hour" is the most common speed report in networks. How do you use it to calculate the best path?



HetNets are going to be part of all the current and future generations of networks and as said before introduce many opportunities as well as challenges. Obviously more access technologies means more capacity. But there is the challenge of licensed and unlicensed spectrum. On the other hand, despite the fact that many of the UEs are capable of transmitting using different wireless access technologies, coexistence of these access technologies introduces challenges such as interference and excessive energy consumption and could have a direct negative QoE effect.

Overcoming these challenges of HetNet can unlock a huge opportunity for QoE improvement for users and CAPEX/OPEX optimization for Service Providers. Additionally with the HetNet environment there is potential for new revenue streams and business models.

The first step towards proper HetNet utilization proper measurement of key network metrics and KPIs. As can be seen in the picture above there are two views when it comes to metrics and KPI collections - the 'View from Device' and 'View from Access Network'.

Both directions provide valuable data points that can be used in QoE measurements and HetNet decision making. Key point is that in an environment with millions of UEs, real time centralized decision making might not provide the optimal result. Certainly there are use cases where network side KPIs can help with predictive steering. For example, knowing an AP or Cell is congested in advance can improve QoE by steering UE away from that access network. On the other hand, UE has real time access to network, device, application and user behavior measurements, which leads to an opportunistic traffic steering that can improve QoE and spectrum utilization at the same time.

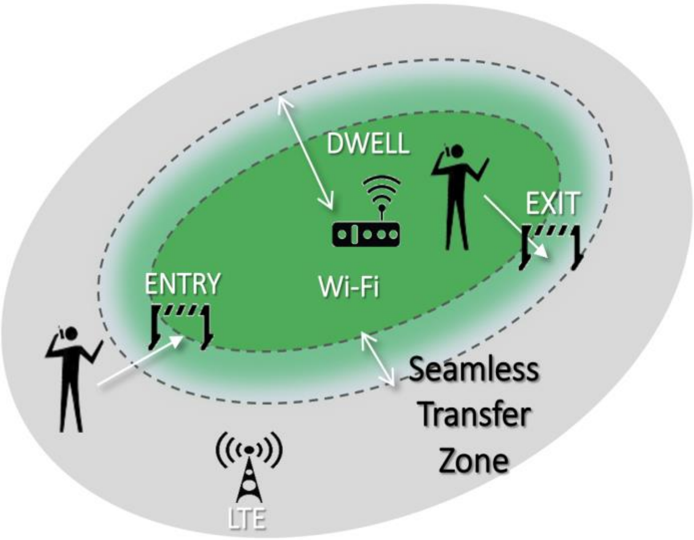
4.2 Use Case 1: HetNet Network Performance Measurement

With billions of multi-RAT supporting UEs (from Smartphones to IoT devices) and several different wireless access technologies(3G/4G/5G, mmWave, Wi-Fi, LPWAN and etc.) available, it is indispensable for service provider to have a 360 degrees view of the network using all available tools and technologies, in order to provide the best user experience while minimizing CAPEX/OPEX.

Service Providers have access to metrics collected from their managed networks over standard interfaces and vendor specific technologies. But as soon as UE is off their network, they have no intelligence. The view from the UE, using SDK based metrics collection, provides this additional piece of information and completes the 360 view of the network performance. For example UEs can record every time handover happens, vertical or horizontal, and provide additional insights such as effect on applications, energy consumption, QoE and etc. Combining and correlating network side metrics and UE side metrics enables Service Providers to make intelligent data driven decisions on where to position new assets, especially in the super densified current and future networks.

4.3 Use Case 2: Seamless Connectivity and Improved UX

For an average smartphone user, having a continuous and uninterrupted internet connection when they need it, regardless of the technology is the most important concern. This is called zero touch or worry free connectivity. With the increase in video consumption and 2 way video communication, slow and congested networks are much more noticeable to end users. The average user might not notice hundreds of milliseconds of delay in loading a web page or delivering a message, but they notice a buffered video instantly. To avoid these kinds of bad user experiences and create seamless transitions(vertical handovers) both at the device level and application level, UEs need to have a clear view of access network changes.



4.4 Use Case 3: HetNet Aggregation

To provide a better user experience while multiple networks are available, UEs can aggregate bandwidth from all access networks to improve speed, reliability and security of mobile connectivity. For example when two wireless links are combined, theoretical throughput available to the UE equals to the sum of instantaneous throughput of each wireless link minus the overhead of additional encapsulation. To achieve this theoretical throughput, UEs need to overcome two barriers. First the constant changing nature of wireless links due to interference, congestion and other factors and second the asymmetrical nature of different wireless technologies, for example high bandwidth high latency Cell versus low bandwidth and low delay Wi-Fi. To overcome these and properly distribute traffic over multiple HetNet channels, UEs need to have a real time view of different metrics(bandwidth, packet loss, latency, jitter, RSS and etc.) on each channel.

4.5 Use Case 4: Mobile Data Offload (and Onload) and Network Augmentation

Mobile Network Operators have been offloading cellular traffic to other managed or unmanaged networks for several users. However with the surge in mobile traffic, especially video consumption which leads to higher QoE expectation, on one side and more unlicensed bands becoming available and deployed on the other side, Mobile Data Offload decisions have become more and more important and critical.

Wi-Fi offloading is the most common way of network augmentation. A large amount of Wi-Fi offload happens in Home and Office, unmanaged Wi-Fi networks. For MNO to benefit from this offload opportunity, users should keep their Wi-Fi radio on all the time. Imagine a scenario where a user is offloaded to a congested Wi-Fi network. Users would immediately shutdown Wi-Fi radio to avoid a bad experience. For MNOs this means that they will lose any future offload opportunity.

CBRS and unlicensed midband spectrums provide another offload opportunity. On one hand more and more Private LTE and Private 5G networks are being deployed and on the other hand more UEs and devices with dual sim and eSIM capabilities are coming to market. Private cellular requires more capex compared to Wi-Fi but provides better indoor coverage and potentially other use cases. One big challenge with private/public cellular coexistence is the users QoE during handover(most devices support DSDS or dual sim dual standby where only one sim card can be active at a time).

In all MDO scenarios having a clear HetNet QoE score from all available networks can improve MDO in two ways. First the offload decision will be focused on user experience, thus providing worry free or zero touch connectivity(no user interaction with radio to lose future offload potentials). Second if the offloaded network is congested then the user can be onloaded to a better network (Always Best Connected).

4.6 Use Case 5: Fixed Mobile Convergence (FMC) or Hybrid Access

More Service providers are moving toward providing Fixed line and Mobile converged access. These services are provided by converged home/residential gateway devices with dual backhaul availability. The gateways are on the data path and can provide dual connectivity to all the local devices. Realtime QoE metrics of all the WAN connectivity is required in order to provide better user experience(per user, device, application on device) to all the connected devices.

5

Lab Testing Summary

5. Lab Testing Summary

5.1 Methods Selected for Lab Testing

In this phase of the project 2 network side methods (Plume QoE and Domos) and two UE side methods were selected for further evaluation. Below is a summary of each method.

DT Connection Classification

- Thin speedtest + passive measurement
- On mobile device ML pipeline
- Widely tested and deployed
- Well defined Quality classes

Domos/QED

- QED published by BB and focuses on QoE and defines QTA
- Measure quality attenuation at TCP/IP with G,S,V characterization
- Domos' Wi-Fi Slicing model estimates QED in real-time on CPE

Plume QoE

- Measures real-time available throughput to devices and mesh AP using base metrics collected
- Predefined Device class and categories with active and Idle need
- Provides simple computation for Device Happiness Score and QoE score
- Integrated to OpenSchema

Shoelace NMA

- On device real-time network assessment: per Network and Application
- Uses both passive and active measurements. Active to measure last mile throughput, latency and loss.
- On device custom ML to reduce drops, failures, and timeouts
- Implements several use cases and integrated to OpenSchema

6

Next Steps and Future Roadmap

6. Next Steps and Future Roadmap

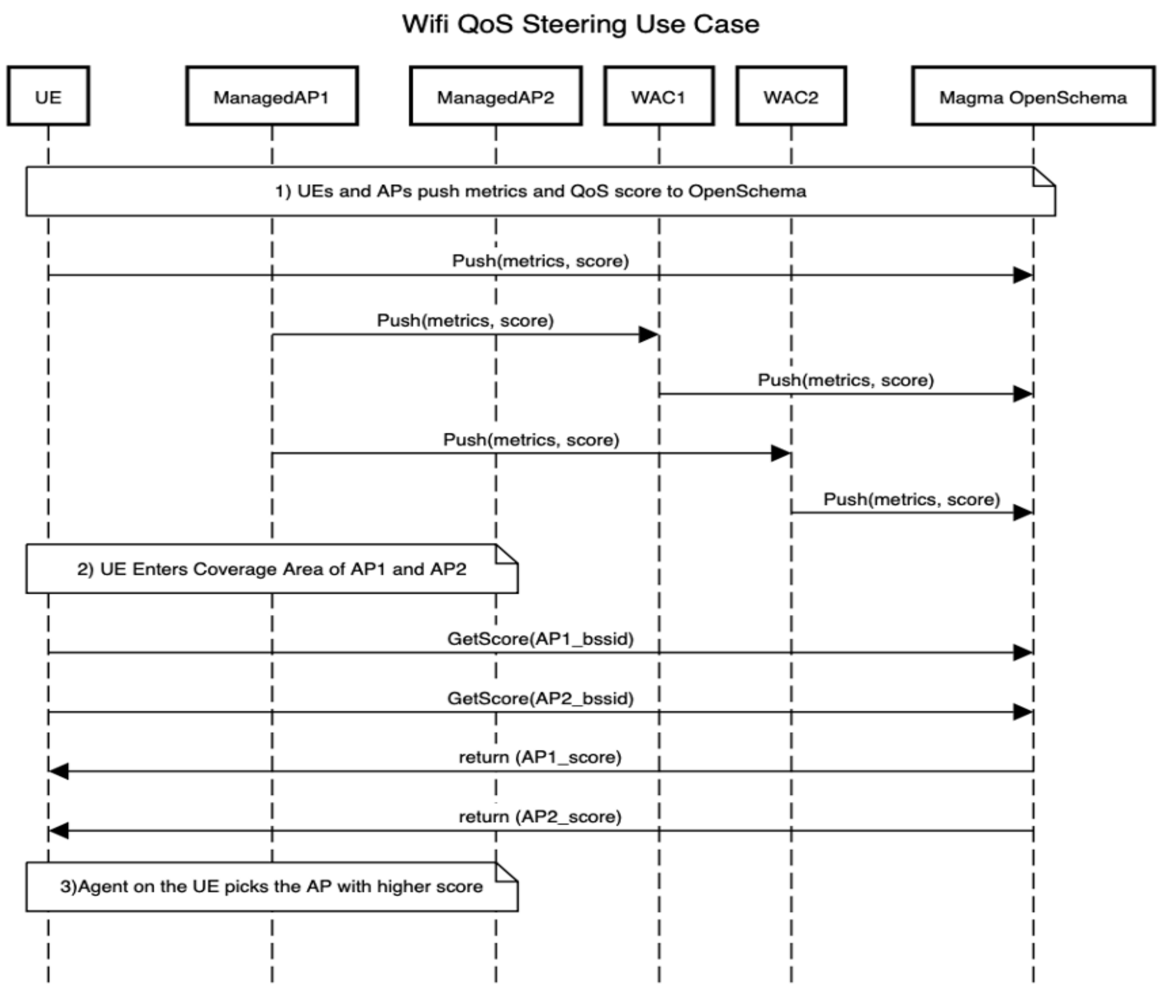
Here are a few proposed next steps:

- Continue testing UE side and Network Side methods
- Design and develop PoC of Unified OpenSchema QoE scoring app/service as part of Magma Unified SMO
- Develop the 360 metrics view by exploring synergies with other groups such as TIP Open Converged GW, Open Ran and RIC, MDT
- Field Tests and Trials and Publishing Results

References

1. IETF RFC 5337: TWAMP (Two-Way Active Measurement Protocol)
2. ITU-T Y.1564: Ethernet testing
3. IETF RFC 2544: Benchmarking Methodology for Network Interconnect Devices
4. Quick Test
5. betterspeedtest.sh.
6. netperfrunner.sh
7. Flent
8. netperf
9. http://fast.com
10. http://speedtest.net
11. http://testmy.net
12. http://speedof.me
13. https://www.bufferbloat.net/
14. https://bit.ly/3phwYgF

Appendix A: Sample Magma Wi-Fi QoS Steering Call Flow




```
//Sample Schema to Store AP Scores message AP {
      String bassid = 1;
      Location lcoation = 2;
      Optional int32 score = 3;
}

//Sample Service to query score service ApScore {
      rpc GetScore(AP) returns (AP) {}
}
```

Appendix B: Review Criteria

Criteria	Definition/Description
Type/Developer/Year	Academic/industry, organization developed or standardized, year published.
Single or Multi Metrics	Does the method use single metric or multi metrics?
Metrics	What are the metrics used? (PHY, MAC, IP, Transport, Application, Device Related, User Prefs, and etc.)
Measurement method	How are the metrics measured? Passive or Active? Tools and Techs used.
Single or Multi Node	Are the metrics used retrieved from single node or multiple nodes (aka crowdsourced).
Evaluation Result	Is the method evaluated? What are the results?
Math or Computation method	Is there certain computation or math used? (ML, NN and etc.)
Open Source	Is the method or implementation open source?
UE side or Network Side	Is the method on UE side or Network side?
Traffic/Application Specific	Is the method application traffic specific? (e.g., video, voice and etc.)
Wireless Technology	Can be implemented on other technologies: (LTE, 5G, NR and etc.)
Scoring scale	Is there a scoring scale define?
System and Architecture	What is the architecture or system component required.
OpenSchema Integration	Is the method integrated with OpenSchema or can be easily integrated?