Making 5G Proactive and Predictive for the Automotive Industry

5GAA Automotive Association
Presents a B2B Industry White Paper

5GAA Automotive Association

CONTACT INFORMATION:
Lead Coordinator – Thomas Linget
Email: thomas.linget@5gaa.org

MAILING ADDRESS:
5GAA c/o MCI Munich
Neumarkter Str. 21
81673 München, Germany

www.5gaa.org

VERSION:	**1.0**
DATE OF PUBLICATION:	**9.12.2019**
DOCUMENT TYPE:	**White Paper**
CONFIDENTIALITY CLASS:	**P (Public use)**
REFERENCE 5GAA WORKING GROUP:	**Working Group 2**
DATE OF APPROVAL BY 5GAA BOARD:	**13.11.2019**

Contents

Executive Summary

5G Automotive Association (5GAA) has developed the concept of predictive Quality of Service (QoS), which is a mechanism enabling mobile networks to provide advance notifications about predicted QoS changes to interested consumers. This makes it possible to adjust application behaviour before the predicted QoS change takes effect, which is important to certain automotive use cases, such as remote and autonomous driving.

The concept of predictive QoS is spreading in the industry and among standards- developing organizations (SDO). It is, therefore, of interest to achieve a common understanding of what predictive QoS is – what problems it addresses, as well as how they are addressed.

This White Paper describes predictive QoS in order to reach a common understanding.

1.0 Introduction

We are at the dawn of a new era of mobility. Future intelligent transportation systems will provide tremendous benefits to people and society in terms of lives saved, increased time and money efficiencies, reduced environmental impact, as well as better utilization of the global road and highway networks. Automated and connected driving, intelligent driver assistance and data-driven transportation network optimisation are examples of what tomorrow's transportation will mean for people and governments. A key enabler of this development is that mobile networks fulfil the requirements of future transportation.

1.1 What Are the Needs of the Automotive Industry?

The automotive industry is evolving towards connected, cooperative and automated driving. With this development follows a large number of connectivity-related use cases to be supported: for example, in the areas of safety, traffic efficiency, convenience and autonomous driving.

Cellular Vehicle-to-Everything (C-V2X) technology – currently based on 3GPP R14 LTE [1] [2] evolving into 5G R16 [3] [4] – provides good support for most of today's use cases. C-V2X defines two transmission modes: long-range communications via the mobile network, using Uu interface, and direct short-range communications using PC5 interface, also known as sidelink.

Automotive applications relying on network support for connectivity have specific Quality of Service (QoS) requirements.[3] These requirements may be expressed in terms of ubiquitous coverage, minimum required uplink, downlink and sidelink data rates, acceptable packet loss ratio, maximum allowed packet delay, inter-packet reception time on sidelink, and other related Key Performance Indicators (KPIs).

As for long-range, network-based communications, which is the main focus of this White Paper, the 5G System (5GS) [5] [6] [7] already provides mechanisms to fulfil the majority of the performance requirements. V2X applications in the vehicle can request one or more communication links with a V2X application server through the mobile network, setting specific QoS KPIs for such links depending on their usage – for example, sending or receiving video streams, downloading maps, receiving software updates.

However, due to varying network conditions, there could still be circumstances in which the 5GS is unable to deliver the QoS required by the application. In such scenarios, the application will experience an unanticipated QoS degradation. Such unanticipated degradation of QoS is not ideal for certain automotive applications (tele-operated driving, autonomous driving, etc.) because of the nature of the adaptation of application required when there is a QoS change. Such adaptations may include slowing down the vehicle, changing lanes, aborting an overtaking operation, or bringing the vehicle to a location where it can be stopped in a controlled way. These adaptations may involve mechanical reactions, as well as possible changes of driving patterns.

The 5GS R15 [8] supports mechanisms allowing applications to **react** to QoS changes, but it does not provide mechanisms to support applications to adapt their behaviour in case of possible QoS changes in the near future. Mechanisms such as these could be considered **in-advance notifications** about upcoming QoS changes. In order to proactively provide such notifications, there may be a need to predict upcoming non-fulfilment of QoS requirements.

With such functionality, the network would be able to send a message to an automotive application, informing it about a potential future change of a QoS KPI that is relevant for the application, and possibly including an estimate of how much the QoS is likely to change and how long it is estimated to remain in the changed state. Being aware in advance of an estimated QoS change permits the automotive application to use that time to adapt its behaviour before the change takes place.

Consequently, there is a need for the automotive industry to have mechanisms allowing applications to **proactively** take action in response to estimated QoS changes. In other words, there is a need for **predictive QoS.**

1.2 What is Predictive QoS?

Predictive QoS is a mechanism that enables the mobile network to provide notifications about predicted QoS changes to interested consumers in order to adjust the application behaviour in advance. Such prior notifications, whenever predictions are made with sufficient confidence, should be delivered with a **notice period** before the new predicted QoS is experienced. The **notice period** depends on the specific application and use case, and should be long enough to give the application sufficient time to adapt to the new predicted QoS.

The message carrying such information is called **In-advance QoS Notification (IQN).** In Figure 1.2-1 an IQN is received by the vehicle in Step 2, thus enabling the V2X application to take appropriate action **prior** to the predicted QoS change taking effect in Step 4.

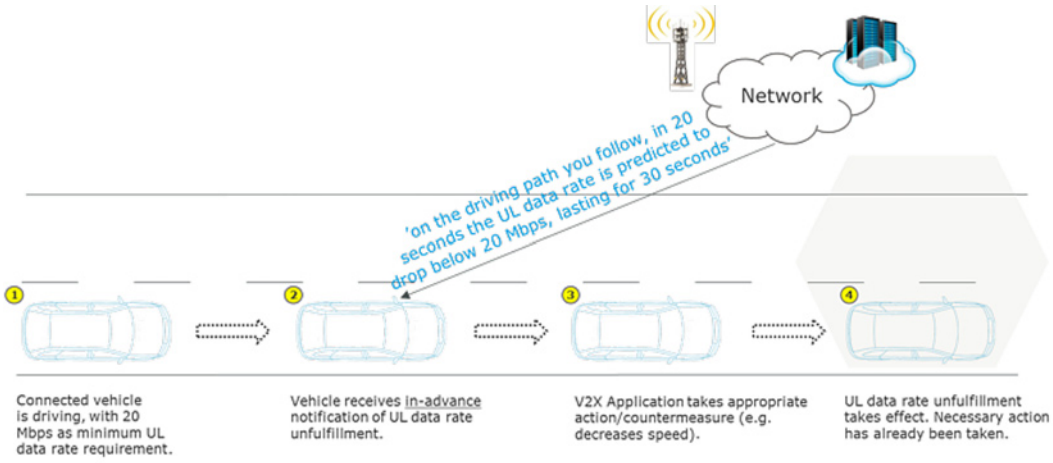

Figure 1.2-1: C-V2X application supported by predictive QoS.

1.3 Examples of Use Cases Demanding Predictive QoS

An example requiring predictive QoS is when a connected, automated vehicle is operated remotely by a tele-operated driving (ToD) application.

The remote operation at a command centre can be performed either by a human or by a software application. In order for the remote operation to be successful when performed by a human, the remote driver must receive video data of sufficient quality from on-board cameras of the vehicle, as well as vehicle status data, such as speed and direction. At the same time, the remote control commands must be delivered to the V2X application on board the vehicle with a latency in the order of 20 ms or less because a higher delay may cause lack of driving responsiveness.

In such scenarios, the network must provide a specific QoS to the ToD application in terms of the minimum (uplink) data rate and maximum (downlink) latency, as well as other potential relevant KPIs. However, situations when the network is temporarily not able to fulfil the agreed QoS can still be properly handled as long as the V2X application in the car is notified in advance. This allows the vehicle to slow down and adapt to a speed permitting tele-operation to continue under the upcoming QoS conditions. In case the upcoming QoS is not sufficient for tele-operation at all, the vehicle reaction could – in the worst case scenario – still bring the vehicle to a controlled, safe stop.

Additional examples of use cases that benefit from predictive QoS are given in Table 1.3-1, together with relevant QoS KPIs for each use case and examples of potential application reactions after receiving an IQN message. More information on QoS KPIs is given in Section 3.2.

Table 1.3-1: Use Case Analysis

Use Case [9] [10]	QoS KPIs To Be Predicted	Examples of Potential Application Reactions
Tele-operated driving	Data rate, latency, reliability	Change route, park vehicle, hand over to nearby driver, change sensor set/properties, change teleoperation mode (e.g. from manoeuvring to trajectory provision).
High-density platooning	Latency, reliability	Change inter-vehicle distance, hand over to driver, change platoon speed or length, terminate platoon.
Hazardous location warning	Reliability	Inform user about availability of warning service, change speed, change route.
Lane merge	Latency, reliability	Change speed of merging attempt, abort lane merge.
Software update	Data rate	Reschedule, stop or resume download.
Infotainment	Data rate	Change video quality.

2.0 Considerations for Predictive QoS

In order to further explore the benefits of predictive QoS, this section contains an analysis of how the needs of the automotive industry can be addressed, as well as a comparison between different approaches.

2.1 Addressing the Automotive Industry's Needs

While a vehicle is moving, the network performance experienced over time and space varies due to changes in network load, radio link quality, etc. As a result, automotive applications have to be designed with consideration of the capabilities of adapting application behaviour – for example, encoding, flow priority, packet inter-arrival time – to the changes in network performance.

One example of application adaptation is video streaming with variable resolutions to accommodate bit-rate variations. This can be achieved by implementing specific, built-in features integrated with rate adaptation algorithms,[11] [12] providing target bit-rate suggestions based on queue length, estimated latency, etc.

The adaptation can also be supported from the network by enforcing specific QoS treatments in relation to requirements on the lowest acceptable and highest needed bit rates. For instance, this is considered for live uplink streaming applications,[13] where the network tries to fulfil the minimum requirement so the video can run without interruption. The application then manages adaptations between the minimum and maximum bit rates based on video needs and network capabilities, also considering the impact the adaptation has on the service. For a video – streaming application, it might be better to reduce the video resolution when the data rate of the network connection is degraded. The impact is a degradation of video quality, but it avoids video interruption.

A key aspect of some automotive applications, such as those related to information-sharing among vehicles or remote controls, is the mutual relationship between the behaviours of the application and of the vehicle. For instance, considering the video of a remote-driving application, a left turn may generate a higher video bit rate due to changes of background images and change of camera focus while turning. Thus, a change of vehicle behaviour (left turn) will have an immediate impact on application behaviour (increase of bit rate).

On the other hand, a reduction of video resolution by the application to accommodate a reduced network capacity could result in speed reduction because the video quality might not be enough to safely keep the current speed. The change of application behaviour (resolution reduction) will have an immediate impact on vehicle behaviour (speed reduction), but in this case the vehicle reaction will take a certain amount of time to be fully and safely completed.

It should, therefore, be highlighted that adaptation capabilities for automotive applications should be designed with consideration of their impact on vehicle behaviour, especially the interaction and time lag between the application adaptation and the corresponding change of vehicle behaviour. From this point of view, predictive information on expected changes of network performance can (i) enable application reactions with reduced impact on associated vehicle behaviour changes and (ii) support completion of vehicle behaviour change before the predicted network change takes place.

2.2 Reactive vs Proactive Network Behaviour

A proactive network behaviour for providing information about QoS unfulfillment, such as sending information in advance of predicted QoS changes, has clear advantages over reactive reporting of such changes because it gives applications sufficient time to react and make adaptations accordingly.

It must, however, be noted that introducing a proactive behaviour in the 5G system through delivery of IQN messages does not necessarily replace the ability of an application to react to unexpected changes in network behaviour. Applications must always be capable of adjusting to sudden QoS changes by an appropriate reaction – as it happens today – without any notification in advance. The proactive network behaviour is thus to be seen as an addition – and a complement – to the already-existing mechanisms for reactive network behaviour.

Compared with today's reactive behaviour, applications enabled by notifications in advance may provide a better user experience because they are capable of anticipating the reactions and adapting to the new QoS situations in more graceful and controlled ways.

2.3 Predictive QoS vs Over-Dimensioning

In an environment with no QoS variation, when the resources offered by the network can be considered unlimited compared with the demand from the various applications, there would be no need for predictive QoS.

However, resource deployment in mobile networks requires careful consideration because of cost and because licensed spectrum is, by nature, a finite resource. It is also widely agreed that pre-allocation of resources to a specific industry vertical – for example, automotive, manufacturing, utilities – even if technically possible, may not be business viable.[14]

There are multiple tools and mechanisms for providing a reliable service: over-dimensioning, dual connectivity/multi-connectivity, etc. However, a 100% guaranteed QoS, at every point in time and in every location, is something that is not feasible.

In such context, and together with the ability to enable an adapted application behavior, predictive QoS becomes a fundamental enabler to minimise disruption caused by the inherently random characteristics of mobile networks.

2.4 MNO-Based vs OTT-Based Predictive QoS

An Over-The-Top (OTT) solution is provided by an entity other than an MNO. In the case of OTT-based predictive QoS, this entity could be a car manufacturer, automotive supplier, or a third party. For example, a car manufacturer could gather data from its connected vehicles ('probe data') about the radio network (coverage, performance, etc.) and use this to provide QoS predictions.

A car OEM could create coverage maps for dedicated MNOs, and with modern machine-earning technologies, quality could be further enhanced. Poor coverage regions, most probable handover regions and some call drops could be located and identified as a trend over a longer period of data collection.

However, there are some limitations of an OTT-based approach:

- Network topology changes/upgrades and traffic pattern changes could only be detected slowly due to the required amount of 'probe data'.
- The UE cannot force certain radio bands to be used. This implies that a survey will be dependent on what the UE is being instructed to use from the network in terms of bands and technologies.
- Detailed radio level information will depend greatly on available APIs provided by the UE modem.
- There will be a lack of awareness concerning: capacity information and resource utilisation of the serving cells; specific traffic prioritisation enforced by the MNO (for example, subscriber category and policy information); mobility management optimisation for vehicular UEs; root causes such as outages of any network components.

On the other hand, an MNO-based solution for predictive QoS could show a clear added value. For operational aspects, such as network monitoring and customer experience management, MNOs operate several systems and continuously analyse negative impacts on the mobile network. Such information is derived from cell-specific performance metrics; subscriber signalling information and events; dedicated network measurement protocols; legacy drive tests or minimisation of drive tests (MDTs); alarming and monitoring systems, as well as probes. Infrastructure information combined with geographical information is already used for radio coverage simulations during the network design, deployment and optimisation process.

A combination of all such data sources with real-time processing can not only create highly accurate coverage maps, but also time-sensitive data rates and latency predictions which, once enabled by an MNO, can be provided to all subscribed OEMs.

Furthermore, an MNO could efficiently coordinate testing sessions in order to measure network metrics at different points without impacting the service delivered to other users sharing the same network. A traditional speed test with file download/upload is a simple way to get an E2E achievable data rate, but this is not a scalable solution. It does not allow multiple users to share the same network at the same time to train the AI models because all users performing a speed test at the same time will share their bandwidth, thus leading to an incorrect result as the measurements will affect everyone else.

A summary of both approaches is given in Table 2.4-1:

Table 2.4-1: Comparison between OTT and MNO Deployment

	OTT Deployment	MNO Deployment
Advantage	• No geo-graphical limitation beyond what the road/vehicle network allows	• Data availability with deep network insights • UE measurements benefit from MNO-wide, UE footprint • Combination with network slices, QoS management and network analytics function
Limitations	• Limited to UE measurements, lacking information on network infrastructure KPIs • Partial network view due to limited number of clients compared to MNO UE footprint	• Limited to MNO footprint*
Challenges	• Sufficient market penetration	• Interoperability* • Longer time to market

Finally, it is worth highlighting that both approaches have their own benefits and the major gains will be achieved if all available data could at some point be combined for an efficient QoS prediction. A recent study [15] analysed the impact of different training data sets on the prediction accuracy and clearly showed that a combination of all information sources collected from the UE and MNO for data rate measurements achieve the highest prediction accuracy.

*Challenge could be overcome with a 3GPP standardized solution.

3.0 Key Concepts of Predictive QoS

5GAA has defined a high-level procedure required for predictive QoS support between the V2X application and the network. This procedure is based on request/response and subscription/notification mechanisms.

In the 5GS, the connectivity service provided by the mobile network to the vehicle is represented by the PDU session.** A PDU session contains one or more QoS flows, as depicted in Figure 3-1. The QoS flow is the finest granularity of QoS differentiation in the PDU session. 5GS has already defined methods and procedures to manage QoS for QoS flows. Therefore in 5GS, the IQN should carry QoS predictions concerning either one or a subset of the QoS flows within a PDU session, or the whole PDU session in cases where the carried information is relevant for all the QoS flows within the PDU session.

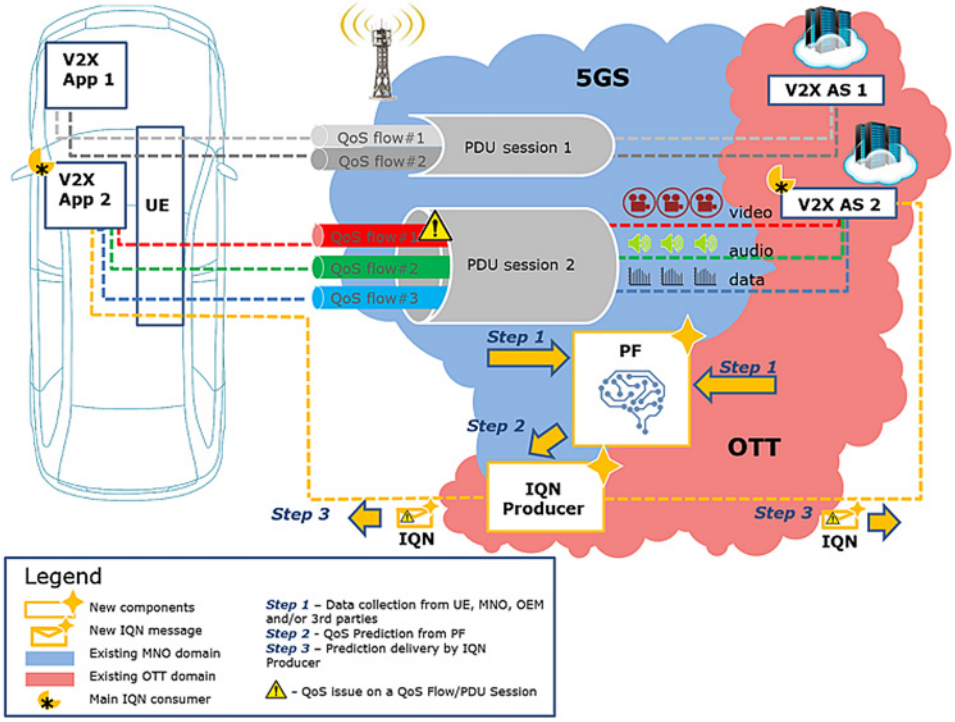

Figure 3-1: Predictive QoS in the 5GS

**3GPP TS 23.50 [5] defines the PDU session as an association between the UE and a data network that provides a PDU connectivity service.

The **IQN consumer** is the entity that requests IQNs and uses the information within the IQN to adapt the behaviour of the vehicle and the application accordingly – for example, V2X application, and/or V2X application server. The **IQN producer** is the network entity that receives the consumer request and delivers the IQN to the consumer. There is also a need for a **prediction function (PF)** that collects necessary information – such as statistics including network / vehicle / third party data to generate predictions through, for example, machine-learning algorithms. Figure 3-1 shows an example of a vehicle running two applications – one of them, V2X App 2, using predictive QoS. In this example, both V2X App 2 and the application server to which the application is connected (V2X AS 2) are requesting and receiving an IQN for a predicted QoS change on QoS flow 1 of PDU session 2. More information on the IQN content is described in Section 4.3.

From a network perspective, supporting predictive QoS involves the realization of three logical steps or functionalities:

Step 1: Collecting Data.
QoS prediction is assisted by data collection from different sources. For example, PF collects data from vehicles, networks, and third parties (e.g. weather information).

Step 2: Making Predictions.
The computation is performed by the PF to produce content to be placed inside the IQN (i.e. the QoS prediction), using collected prediction-supporting data.

Step 3: Delivering Predictions.
These procedures are required in order for the IQN producer to deliver the IQN containing the QoS prediction to the intended consumers over the specified interfaces.

These steps are described in the following sections.

3.1 Collecting Data

5GAA has identified a list of information categories that can be collected and used to assist the prediction function. These categories are grouped according to relevant actors supplying such information. For each Information category, Table 3.1-1 lists related sources and examples of included information components.

Table 3.1-1: Prediction-Supporting Information Exchanged for QoS Predictions

Information Group	Information Category	Information Source	Example of Information Components in this Information Category	Potential Exploitation of the Information Category in Making the Prediction	
				5G-Based	OTT-Based
OEM	Vehicle information	V2X application on UE	Potential routes (e.g. number of routes, set of locations and time-stamps per potential route), application class (e.g. 3GPP 5QI as defined in [3]), QoS KPI requirements, UE capabilities (e.g. radio capabilities, software capabilities, etc.)	Yes***	Yes
	Client performance measurement	UE	Network performance KPIs measured in the UE or in the application (e.g. data rate, latency, packet loss rate, etc.)	Yes	Yes
Mobile network Operator/ vendor	Radio access network information	Network (RAN)	Radio network load available vs free resources, connection setup success rate, current number of users and other relevant information, such as users data rate, signal strength, latency, etc.	Yes	No
	Core network information	Network (CN)	Core network information, such as network load, resources, PDU session information and relevant QoS	Yes	No
			Analytics information (e.g. historical data on QoS collected by OAM and other NFs)	Yes	Yes***
Third party player	Weather information	AF	Rainfall, fog/visibility levels, etc.	Yes***	Yes***
	Coverage information	AF	Coverage map with additional information (e.g. average data rates per geographic area)	Yes***	Yes***
	Road traffic information	AF	Traffic congestion level, etc.	Yes***	Yes***
	Road infrastructure information	AF	Road topology, tunnel locations and lengths, road conditions, roadwork plans, traffic signs, traffic-light status, etc.	Yes***	Yes***
	Event-based information	AF	e.g. sports event in a certain location at a certain time	Yes***	Yes***

For example, the PF may obtain information on potential routes from the V2X applications, weather information from third party AFs, radio network load from RANs and historical data on QoS from CNs. The data identified for each Information category can be used by the PF to generate QoS predictions.

***It cannot be assumed by default that data is available, but under agreement it could be shared.

3.2 Making Predictions

This generation of QoS prediction is performed by the prediction function once it has collected enough information. Which procedures and algorithms will be used for generating QoS predictions depends on the specific use case, the KPIs of interest to be predicted for that specific use case, as well as the relevant timing and accuracy constraints related to that KPI.

The timing constraints depend on the potential application reaction time, where the reaction is the set of actions that may be performed in the vehicle by the application when a QoS prediction is received. As there could be several potential reactions to a QoS prediction – for example, reduce speed, change lanes, change routes – it is assumed that for every use case there is a worst-case, application-reaction completion time that will set a lower boundary for how long in advance the QoS prediction must be delivered to the consumer. The application-reaction, completion time for most use cases is expected to be in the order of a few seconds to tens of seconds.

The following QoS KPIs (defined in [3] [5]) are of interest as the bases of predictions:

- **Latency** (ms)
- **Reliability** (%)
- **Packet delivery ratio** (%)
- **Data rate** (Mbps)
- **UP connection** (active/inactive)

Additional KPIs may be considered, such as **network utilisation, delay variation** (jitter) or **availability.**

3.3 Delivering Predictions

Figure 3.3-1 illustrates the four steps in the delivery of an In-advance QoS Notification (IQN) to a connected vehicle running a ToD application.

In Step 1 the vehicle is driving under normal conditions, and the application is utilizing QoS flow 1 for video streaming from on-board cameras and QoS flow 2 for driving instructions.

In Step 2 the vehicle receives an IQN containing a prediction of a QoS degradation that is relevant to QoS flow 1: for example, a downgraded UL data rate.

In Step 3 the application adapts to the upcoming conditions before they take effect. Adaptation may be done in the client, in the server or in both. The adaptation may include, for example, reduction of vehicle driving speed and reduction of video quality.

In Step 4 the predicted QoS degradation takes effect – the UL data rate actually drops to lower (yellow) performance. The vehicle avoids potential service disruption despite the new network QoS conditions because the necessary adaptations have already been done prior to the degradation taking effect.

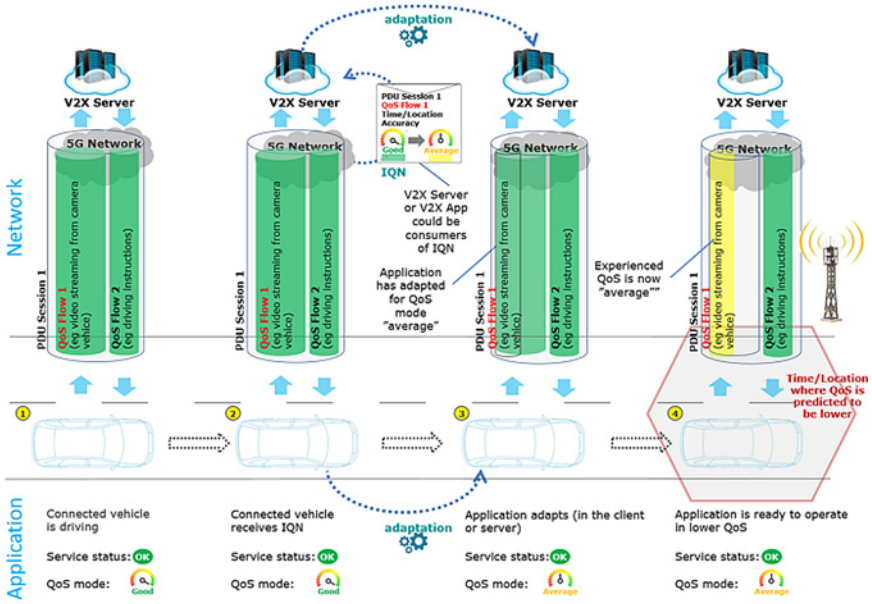

Figure 3.3-1: Application Adaptation after Receiving IQN

The potential IQN consumers are the **V2X application** running in the vehicle and/or the **V2X application server (V2X AS)** deployed, for example, in the cloud. The specific setting of consumers depends on which entity should take the decision about how to react to the reception of an IQN. The actual solution-specific process of IQN delivery may also involve other entities such as **application function (AF)**, third party external AFs, as well as other network functions (NF) or nodes, such as UE.

In the vehicle, the V2X application is the ultimate intended receiver of the IQN, and also the entity that is supposed to react (at application layer) to QoS changes (improvement, degradation, or loss of connectivity) that are predicted for the near future. As mentioned above, the delivery towards the V2X application and V2X AS/AF is solution-specific.

Delivery towards the V2X AS (and to a generic AF) may consider direct delivery from the IQN producer or via another NF. Delivery towards the V2X application may consider different approaches, such as direct delivery from the IQN producer, or via the V2X AS, or by using 5GS network procedures: for example, via the in-vehicle UE. QoS predictions may also be exposed to the V2X application through a MEC service API.[16] Depending on the solution and on the needs of the use case, network functions/nodes may also trigger certain reactions. For example, the UE may initiate a PDU session management procedure, or the AF may initiate specific procedures: for example, application function influence on traffic routing to request a PDU session modification.

4.0 Key Questions of In-Advance QoS Notifications

In order to summarize and further clarify the predictive QoS concept within an automotive scope, two key questions are looked at in sections 4.1 and 4.2. Here it is assumed that In-advance QoS Notifications (IQNs) are triggered by a prediction function (PF) in relation to a PDU session, or to a QoS flow within a PDU session. The IQN, containing the QoS prediction, is delivered to the IQN consumer according to the consumer's specific IQN request or IQN subscription. The QoS prediction is related to one or more KPIs, such as latency and uplink/downlink data rate.

4.1 What Does an IQN Consist of?

In a general case, an IQN contains the following information, as depicted in Figure 4.1-1:

- Information regarding which QoS flow/PDU session may be affected.
- Predicted KPI and its predicted value with respect to the threshold: 'UL data rate predicted to cross the threshold A by 1 Mbps downwards'
- Time when the prediction takes effect: 'in 25 seconds', or alternatively, the distance where the prediction takes effect: for example, 'in 10 km'
- Predicted change duration: 'for 15 seconds' or alternatively 'for 5 km'
- Prediction accuracy 'with 98% accuracy'

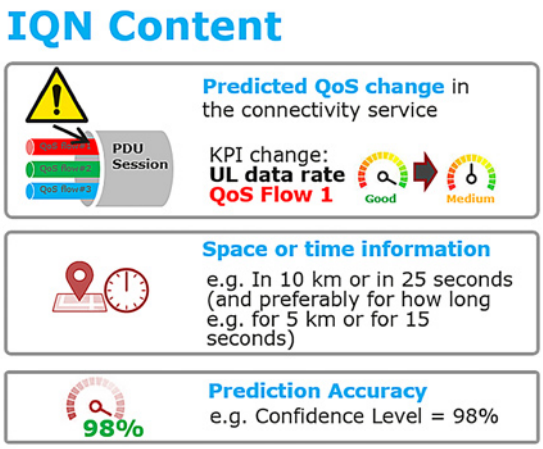

Figure 4.1-1: Content of an IQN

There are different granularity levels possible, and the actual content of an IQN depends on what the IQN consumer configures when sending the predictive QoS request to the IQN producer, which may vary between use cases. In its light implementation, the IQN may simply inform whether there may be a change in the QoS KPI when compared with the QoS currently associated with the QoS flow: for example, at the location of a tunnel or a coverage hole. The message to the IQN consumer would then be: 'On the driving path you are following, in **20 seconds/1 km** the QoS KPI **UL data rate** is predicted to **change**.'

Although such a simple IQN can be helpful to a V2X application, for complex use cases it may not be enough for the V2X application to decide on an appropriate reaction. The V2X application may also need to know how much the QoS may change and for how long because its countermeasure depends on those factors. For example, a predicted data rate drop from 120 Mbps to 80 Mbps may result in one type of application adaptation, while a predicted data rate drop from 120 Mbps to 60 Mbps may result in a different type of adaptation. This is illustrated in Fig 4.1-2, which shows different threshold levels associated with a tele-operated driving use case, with video streaming of four HD cameras from the vehicle, each with a data rate from 15 to 29 Mbps, plus 4 Mbps object data.

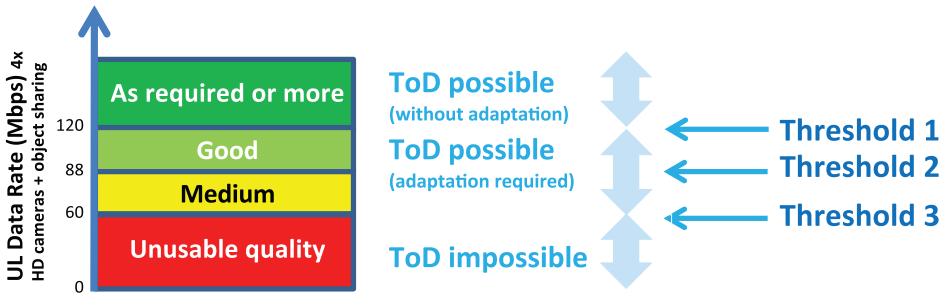

Figure 4.1-2: UL Data Rate Threshold Levels, Tele-operated Driving (ToD)

In this use case, IQN information needs to be given with a finer granularity level in order to enable an appropriate application action. Examples of different granularity levels of QoS information that can be part of an IQN are as follows, from coarser to finer levels.

- 'On the driving path you follow, in **20 seconds / 1 km**, the QoS KPI **UL data rate** is predicted to:
- '**Become Good** (for example, 88–120 Mbps) / **Medium** (for example, 60 – 88 Mbps) / **Bad** (< 60 Mbps)
- '**Drop below** the **60 Mbps** threshold
- **'Drop below** the **60 Mbps** threshold with 95% probability'

In summary, an IQN consists of all the information that is required by the IQN consumer to implement an application adaptation to the upcoming QoS changes. Such information can include predictive information of QoS KPIs, prediction accuracy, and information on when or where the QoS change is expected to happen. The exact content depends on the specific application reaction that needs to be performed in a specific use case.

4.2 When Should an IQN Be Triggered?

The IQN consumer (for example, V2X application or V2X AS) determines which QoS KPIs are to be predicted by the PF, depending on the use case. For example, one IQN consumer may be interested in data rate predictions, whereas another may be interested in predictions for latency and reliability.

In order to know when to trigger an IQN, the PF needs to know the thresholds associated with each QoS KPI of interest. These thresholds are defined by the application. The event that triggers an IQN is when a threshold of a QoS KPI is predicted to be passed. However, since the application has to complete its reaction before the predicted QoS becomes effective, the IQN has to be delivered well in advance. The IQN notice period defines how long in advance the IQN shall be received by the consumer, as described in Figure 4.2-1. The **IQN notice period** is use-case specific and is set by the application.

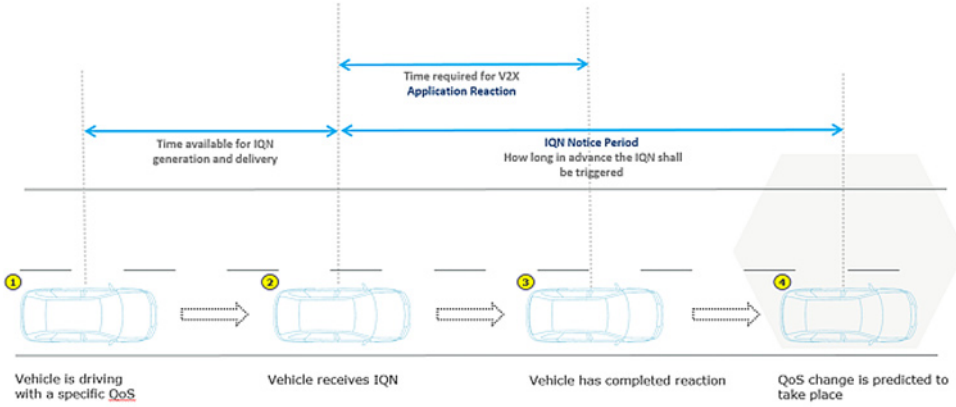

Figure 4.2-1: IQN Notice Period Giving Time for Application Reaction

The time scale of the reaction that the prediction involves, including mechanical actions, is not on the atomic time scale (for example, µs or ms), but rather is measured in seconds or longer.

Figure 4.2-2 indicates that the stopping distance for a car driving at 50 km/h is around 30 m.[17] In the tele-operated driving use case, assuming that identified application reaction for a specific scenario is to stop the vehicle, this means that a car has to start taking corrective action at least 2–3 seconds in advance. In other words, when a car is driven at 50 km/h, an in-advance notification about a potential QoS degradation has to be delivered at least 2–3 seconds or 30–40 metres ahead of a point where the potential QoS degradation may occur. However, the worst-case distance or time in advance may be higher and that depends on capabilities and conditions of the car, local weather, road traffic conditions, supported V2X application, etc.

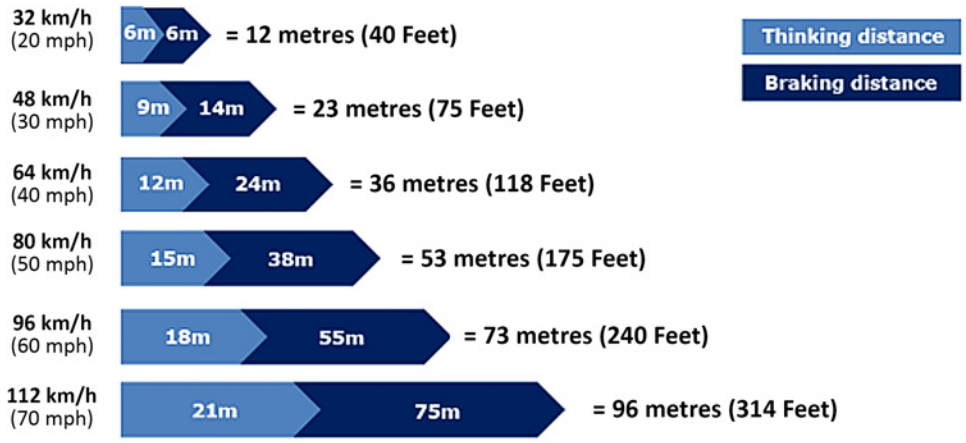

Figure 4.2-2: Typical Stopping Distances for Different Speeds [17]

For tele-operated driving or hazardous location warning use cases, consider a worst-case scenario where a vehicle driven at 100 km/h loses a connection and has to be stopped safely and parked in a safe place. This may take a few seconds up to minutes. This means that in order to ensure safe operation of a connected vehicle, a driver must be notified sufficiently in advance before a connection is predicted to be impaired.

In summary:

An IQN is expected to be triggered in a time scale of seconds to minutes in advance, rather than µs or ms in advance.

The IQN notice period is use-case, as well as application-state specific.****

****Application may have several potential reactions in a specific application state and in a specific use case. Considering all the potential reaction times, the IQN notice period shall be larger than the maximum of all the potential application reaction times under a particular use case and state of an application.

5.0 Towards a Standard Mechanism for Predictive QoS

Predictive QoS has attracted interest among several industry groups, and work is ongoing within different organizations to define standard mechanisms. This section briefly mentions a few of the most important activities and studies currently taking place.

5.1 5GAA

5G Automotive Association (5GAA) was created in September 2016 to connect the telecom industry and vehicle manufacturers in developing end-to-end solutions for future mobility and transportation services. Currently 5GAA has more than 120 member companies.

5GAA WG1 (use cases and technical requirements) and WG2 (system architecture and solution development) are performing studies related to predictive QoS.

A specific WG2 study on predictive QoS and network slicing,[18] finalised in February 2019, resulted in a list of predictive QoS requirements derived for 3GPP that were discussed and adopted in 3GPP Release 16. During 2019 a continued study has been ongoing in WG2, aiming at a deeper understanding of the concept, as well as helping 3GPP derive new requirements for future releases.

5.2 5G-ACIA

5G Alliance for Connected Industries and Automation (5G-ACIA) was created in April 2018 to ensure the best possible applicability of 5G technology for connected industries, in particular the process and manufacturing industries. Currently 5G-ACIA has more than 50 member companies. A specific study on In-Advance/Predicted QoS Notifications in 5G (ADAPT) is ongoing in 5G-ACIA WG1 (use cases and requirements).

5.3 3GPP

3rd Generation Partnership Project (3GPP) is performing work related to predictive QoS mainly in working groups SA1 (services) and SA2 (architecture), targeting Release 16 and Release 17. Input requirements from 5GAA have been received and adopted.

In 3GPP SA1, predictive QoS-related requirements are studied and captured in 'Enhancement of 3GPP Support for V2X Scenarios'.[3]

In 3GPP SA2, predictive QoS-related studies are captured in 'Architecture Enhancements for 5G System to Support Vehicle to-Everything (V2X) Services' [4] and 'Architecture Enhancements for 5G System to Support Network Data Analytics Service'.[19]

Further, in SA2, the R17-related study on Enablers for Network Automation for 5G – Phase 2 [20] includes a key issue 'NWDA-Assisted Predictable Network Performance', which addresses predictive QoS topics.

5.4 ETSI

The European Telecommunications Standards Institute (ETSI) has studied predictive QoS for V2X in ISG MEC (multi-access edge computing) as part of the completed study item on MEC support for V2X use cases. The results of the study have been captured in the ETSI GR MEC 022.[21] The resulting normative work on V2X information service API is ongoing and being documented in the ETSI GS MEC 030.[16] At the time of publication of this White Paper, the GS MEC 030 was still in draft state, a work in progress.

5.5 GSMA

GSM Association (GSMA) has taken predictive QoS into account when defining generic network slice template (GST),[22] by including performance prediction as an attribute, allowing the mobile system to predict the network and service status.

5.6 5G-PPP Research Projects

5G Infrastructure Public Private Partnership (5G-PPP) research projects feed their results as inputs to standards-developing organisations (SDO): for example, 3GPP and ETSI. The project 5G Cross-Border Control (5GCroCo) within EU research program H2020 has the objective to validate advanced 5G features, including predictive QoS.[23]

6.0 Conclusions

Predictive QoS enables a new proactive behaviour in 5GS through In-advance QoS Notifications of predicted QoS changes. This is a complement to the procedures for QoS management, policy and control, which are already included in the 5GS. Without predictive QoS, an application can only react to QoS changes after they have occurred: for example, without any prior notice. With predictive QoS, on the other hand, an application can take action before a potential QoS change takes effect, thus enabling a more graceful adaptation. Such in-advance adaptation is especially important for mission-critical applications. By taking necessary countermeasures before a potential QoS drop, the applications in the vehicle may continue their operations, potentially with reduced functionality or speed, rather than stopping their operations, which could be very costly.

The concept of predictive QoS has been developed in close cooperation between the telecom and automotive industries. Examples of automotive use cases that are foreseen to benefit from this new concept are: tele-operated driving, high-density platooning and hazardous location warning. However, predictive QoS is applicable not only to the automotive industry: other industry verticals, such as industry automation, are foreseen to benefit from it.

Work has been done in a number of SDOs and industry groups in order to further detail and standardise predictive QoS. Some examples are 5GAA, 3GPP, 5G-ACIA, ETSI, and GSMA. In 3GPP, basic support for predictive QoS is provided in R16, and further development is ongoing for R17.

Predictive QoS is being widely investigated by multiple parties for adoption in the automotive industry, and the foundation for standardisation has been laid. More work is needed to understand and describe additional use cases and derive requirements from them. However, the first steps have now been taken to make 5G proactive and predictive.

List of Abbreviations

5G-PPP	5G Infrastructure Public Private Partnership
5GS	5G System
5QI	5G QoS Indicator
AF	Application Function
API	Application Programming Interface
AS	Application Server
C-V2X	Cellular Vehicle-to-Everything
E2E	End-to-End
IQN	In-Advance QoS Notification
ISG	Industry Specification Group
KPI	Key Performance Indicator
MDT	Minimisation of Drive Tests
MEC	Multi-Access Edge Computing
MNO	Mobile Network Operator
NF	Network Function
OEM	Original Equipment Manufacturer
OTT	Over-The-Top
PF	Prediction Function
QoS	Quality of Service
SDO	Standards Developing Organization
ToD	Tele-operated Driving
UPF	User Plane Function

References

[1] 3GPP TS 23.285, V14.7.0 (2018-06): Architecture Enhancements for V2X Services

[2] 3GPP TS 22.185, V14.4.0 (2018-06): Service Requirements for V2X Services, Stage 1 (R14)

[3] 3GPP TS 22.186, V16.2.0 (2019-06): Enhancement of 3GPP Support for V2X Scenarios

[4] 3GPP TS 23.287, V2.0.0 (2019-08): Architecture Enhancements for 5G System to Support Vehicle to-Everything (V2X) Services

[5] 3GPP TS 23.501 V16.0.2 (2019-04): System Architecture for the 5G System; Stage 2 (Release 16)

[6] 3GPP TS 23.502 V16.0.2 (2019-04): Procedures for the 5G System; Stage 2 (Release 16)

[7] 3GPP TS 23.503 V16.0.0 (2019-03): Policy and Charging Control Framework for the 5G System (5GS); Stage 2

[8] 3GPP TS 23.501 V15.7.0 (2019-09): System Architecture for the 5G System; Stage 2 (Release 15)

[9] 5GAA TR T-190028: 5G Use Cases and Requirements – Wave 2.1

[10] 5GAA TR T-190099: "G Use Cases and Requirements – Wave 2.2

[11] IETF RFC 8298: Self-Clocked Rate Adaptation for Multimedia

[12] I. Johansson, S. Dadhich, U. Bodin, T. Jönsson: Adaptive Video with SCReAM over LTE for Remote-Operated Working Machines, Wireless Communications and Mobile Computing, August 2018.

[13] 3GPP TR 26.939: Guidelines on the Framework for Live Uplink Streaming (FLUS), Release 15

[14] GSMA White Paper (2017-11): An Introduction to Network Slicing

[15] Samba, A., Busnel, Y., Blanc, A., Dooze, P., & Simon, G. (2018): Predicting File Downloading Time in Cellular Network: Large-Scale Analysis of Machine Learning Approaches. Computer Networks, 145, 243-254.

[16] ETSI GS MEC 030 (Draft): "V2X Information Service API"

[17] https://driving-test-success.myshopify.com/pages/stopping-distances-and-the-theory-test (retrieved 19 July 2019)

[18] 5GAA TR A-190176: Architectural Enhancements for Providing QoS Predictability in C-V2X

[19] 3GPP TS 23.288, V16.0.0 (2019-06): Architecture Enhancements for 5G System to Support Network Data Analytics Service

[20] 3GPP SP-190557: Study on Enablers for Network Automation for 5G – Phase 2

[21] ETSI GR MEC 022 V2.1.1 (2018-09): Study on MEC support for V2X Use Cases

[22] GSMA NG.116 V1.0 (2019-05): Generic Network Slice Template version 1.0

[23] 5GCroCo D1.2 (2019-09): 5GCroCo First Intermediate Project Report version 1.0