Low latency communication Monitoring, Measurement and estimation

Njakarison Menja Randriamasinoro (njakarison@gmail.com)
Benoit Gendron (bglatence@gmail.com)
Kim Khoa Nguyen (kim-khoa.nguyen@etsmtl.ca)
Milad Amiriyan (amiriyan.milad@gmail.com)

February 18, 2021

Contents

Abstract

Providing low latency communication is the deal of all telco provider actually. Previous generation of communication has difficulty to assume such requirement, especially while the number of the customer increase. Adopting 5G communication helps them to overcome that difficulty and will provide 10 times bandwidth, a very high reliability and a minimal time response. Hence, such network allow strict network applications which requires a very low latency communication to work effectively. However, some event in the network will change its efficiency, such as network overload or congestion as well as devices breakdown. Monitoring the network is then important to avoid network crash and disturbance in application functionality. Here we are presenting a tool which helps providers and customer to measure and monitor in real time network activities then will predict future issues. This tools measure generally the network latency and based on historical data obtained from long term measurement, it will predict future event. Multiple methods are used in this tools according to the use case of the customer, in order to keep measurement precision. Important requirement are considered to adapt it with 5G network communication as well as WiFi and LTE networks. In this document, we present several measurement obtained with the tool using low latency 5G network communication.

Keywords : *5G Network Communication, Network Latency, Latency Components, Jitter, Network Reliability, Round Trip Time, O-RAN, V-RAN, SLA*

1 Introduction

1.1 5G Network advantages

5G network bring multiple advantages to its users. In term of bandwidth, this technology will provide more than 10 times bandwidth compared with previous generation of network. It will ensure a very high reliability, more than 99.999 percent, between end points int network, . Response time will be divided by 5 compared with the latest 4G.

1.2 Importance of 5G latency

The new generation of network known as 5G provides a very high communication quality compared with previous generation. It will enable new services and deeply transform the value chain of several industries. Align with the concept of “Industry 4.0”:

- Remote control of robots
- Real-time accurate location of assets
- Track connected vehicles on the move
- Smart automation of processes

Latency calculation and monitoring must be precise and continuously performed. It is therefore critical to understand latency ”behavior” in different environments and scenarios... and identify ways to improve it.

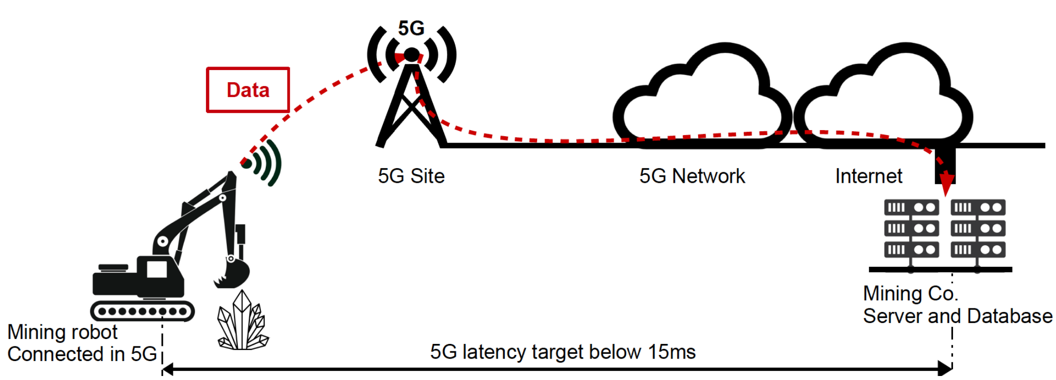


Figure 1: 5G Network scenario

1.3 5G mobile operator challenge

Mobile operator acting in 5G network in now faced in challenge of monetizing ultra low latency and high reliability. The SLA [1] indicate for example the measured latency lower than 15ms with a reliability more than 99.992 % per month. Mobile operator needs tools to track in real time and continuously their client latency and reliability according to this SLA. Latence Technologies will offer this tools with more options and multiple measurement and tracking and methods. This B2B [2] cooperation will monetize the network low latency, generate additional revenue on both side of the B2B, and increase client trust.

1.4 Big question on 5G latency

The problematic of mobile operators is focused on defining the key indicators explaining the end-to-end latency of my 5G network. How ensure the commitments for 5G Ultra Reliable and Low Latency Communication (URLLC) and selling it to new B2B customers. They will question also on what is the value or price premium compared with other connectivity technology. On the other side, Industry 4.0 find out why how explain the latency variance of their private 5G network and how to improve it. The tools we propose addresses both the technical issues which are latency reduction and optimization as well as business imperatives which consist on valuation and monetization of low latency.

2 Measurement and metrics

Managing network consist on monitoring its activity, preventing future problem and maintain the better Quality of Service (QoS). Several QoS metrics should be controlled continuously to make sure that no user will lose connection and facing bad experience, caused by some service interruption. The tools we propose in this work help generally its user to measure, monitor continuously and store network status with several metrics. It include also a module which helps doing AI/ML with stored measurement data in order to forecast future network status and helping network administrator on taking decision on what to do. Table 1 summarize the used metrics and their definition.

[1] Service level agreement

[2] Business-To-Business

Table 1: Definition of some metrics

Metrics	Definition
Network Latency (ms)	Network latency is defined as the time duration between the generation of a packet and its correct reception at the destination [1]. The tool we propose measure the round trip time (RTT) of the packet trough path in the network. The network latency is generally measured in millisecond
Jitter (ms)	Is a variance in delay between packets sent over the network, usually measured in milliseconds (ms). Network jitter happens due to network congestion, interference, route changes, etc. There is no network free of jitter. Network Jitter is expressed in millisecond
Packet loss (%)	The packet loss indicate the quantity of packet fail to be transmitted across the network due to multiple factor such as network congestion. The packet loss expressed then the percentage of lost packet according to the total number of packet sent.
Reliability (%)	The network reliability at time t, is the probability that all the nodes are operational and can communicate with each other over entire time interval. It will be expressed by the ratio between the packet delivered successfully with the total number of packet sent to the network according to the time required by a service to be executed.
Availability	Is the ratio between the time where connection is available in an interval of observation and the total time in the interval.

3 Latency characteristics

The principal metrics which interest us is the Latency. As it definition presented in table 1, latency is the time that a packet takes to cross the network path from the origin to the destination of the packet.

3.1 Network latency effects

3.2 Origin of latency

This latency will be caused by multiple factor such as Physical distance between the 2 end points. The capacity of each link which form the path in the network will also induce latency in the network. The characteristics of each network devices (speed, memory size, what firmware is used by the devices, and so one) will generate latency in each packet. That will be repeated according to the number of node that packet should cross inside the path.

3.3 Latency composition

Theoretically, network latency may be decomposed according to its physical deployment and all used devices. According to [1], there is generally 5 latency components

✧*propagation latency* : it's the time that signal take to cross the path from the start point to the end point

✧*Transmission latency* : the time to transmit all bits in a packet

✧ *Processing time* : time that device takes to process data as packets

✧*Routing time* : time to route packet

✧*Queuing time* : time waiting by packet to be transmit

3.4 Latency components evaluation

As described in article [1, 2], these latency components are due to the influence of factor described in section bellow. General formulation which can estimate the values of these components are proposed in [2].

3.4.1 propagation latency ($\mathcal{L}_{Prop}$)

The propagation latency is the latency induced by the propagation speed of the electrical signals across the network path. It depends on the distance between the two end points of the path, and may be expressed as follows :

$$\mathcal{L}_{Prop} = \frac{D}{V} \tag{1}$$

D is the distance (km) and V is the speed of light ($\approx 2.1 * 10^8 - 3 * 10^8 m/s$)

3.5 Transmission latency ($\mathcal{L}_{Tran}$)

Transmission latency is the time that take a packet to be transmitted through each link in the end to end path. This latency component depends on the capacity of each links and the size of a packet. It will be evaluate as follows:

$$\mathcal{L}_{Tran} = \frac{P_S}{L_{Bw}} \tag{2}$$

P_S indicate the Packet size ($bits$) and L_{Bw} is the link rate in bits per second (bps)

3.5.1 Processing latency ($\mathcal{L}_{Proc}$)

To be transported through the network, information and data are divided in packets which contains some information about sources, destination, the data, ...

The time that takes devices to convert original data into packets constitute the processing latency component. This component depend then generally on the devices characteristics (hardware and software such as, hardware speed, memory size, firmware version,...). It may be evaluated with the following formulation [2] :

$$\mathcal{L}_{Proc} = \frac{B_S}{S_P} \tag{3}$$

with B_S indicate the device buffer size ($bits$) and S_P is the device rate processing *(bit per second)*.

3.5.2 Queuing latency($\mathcal{L}_{Queu}$)

Queuing latency is the time that a packet spent in the device buffer waiting to be transmitted. This time depends on the number of waiting packet and the size of the buffer. The queuing delay for a packet depends on the average of packet arrival rate and the service rate. It will be formulated as follows :

$$\mathcal{L}_{Queu} = \frac{1}{\alpha - \lambda} \tag{4}$$

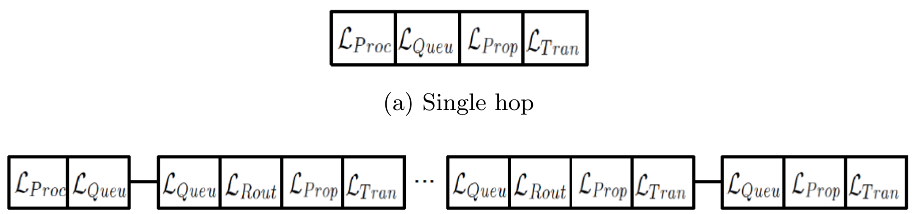
Where, α is the packet output rate and λ the arrival rate *(packets per second)*.

3.5.3 Routing latency($\mathcal{L}_{Rout}$)

The routing latency is the time that device take to decide the route for the packet. It depends then on how network device compute routing decision. It is a metric that estimates the average waiting time for each potential next hop [3], and this component appears only in multi-hops network. For high speed network device, this routing time is slightly negligible compared to other components.

3.6 End to end Latency

An end to end latency will be composed by a single or multiple hops. The number of hops that the packets have to cross in a path define the appearance of each components. For a single hop network, the data need to be processed as packet, buffered in the device, transmitted to the path and then cross the physical network path. In this case, the routing latency component is not required.



(a) Single hop

(b) Multiple hops

Figure 2: Latency components in a path

4 URLLC use cases

New categories of applications requires high network reliability and very low latency communication to be able to achieve tasks in network. According to [4], the emerging 5G is designed to support enhanced Mobile BroadBand (eMMB), Ultra-Reliable Low Latency Communication (URLLC) and massive Machine Type Communication (mMTC). URLLC communication ensure at the same time a very high network reliability and low latency communication, that claims most of network applications. Using the 5G communication, mobile provider network will offer high performance networks to their customer, and will meets applications requirements. Table 2 shows some URLLC use cases with their requirements in terms of latency and reliability [1].

Maintaining the QoS in order fulfill customer SLA is a puzzle game for the provider. The most efficient way to ensure that QoS to be maintained is the real time network monitoring.

Table 2: Low latency 5G use cases [1]

Use case	**Latency**	**Reliability (%)**	**Bandwidth**
Factory automation	1-10 ms	99.9999999	Low data rate
Process automation	100 ms-1s	99.9999999	Low data rate
Virtual reality	1ms	high	high data rate
Automated guided vehicle	few ms	99.99999	high data rate
Tele-surgery	1-10 ms	98	high data rate
Exoskeletons and Prosthetic hands	few ms	high	Low data rate
Protection traffic in smart grid	1-10 ms	high	Low data rate
Control traffic in smart grid	100 ms	high	Low data rate
Financial market	few ms	high	Low data rate
Connected 5G Glasses	5-10 ms	high	high data rate

It consist on measuring continuously sensitive metrics which will affect user experience while using network. Network latency and reliability are metrics that should be monitored. The type of use case also plays a role in this monitoring task. We are analyzing 3 use cases which seems to have a very high potential in the market. It will help us identifying the suitable measurement method and tools according to each use case.

4.1 Factory automation use case

The factory automation use case describe the industry 4.0 paradigm which enable interconnection and communication between machines, devices, sensors and people. High reliability and a guaranty of lo latency is required in an industrial manufacturing [1]. Figure 3 shows single example scenario of a factory automation use case, with all metrics requirements. In this example, the robotic hand receive command from remote server control in order to automatize its actions.

4.2 AR - VR Use case

Augmented and Virtual reality use case consist on the use of virtual environment to execute remote task. Figure 4 illustrate for example this use case, by controlling remotely a robotic hand. The robotic side site environment is captured using the camera. Output video/sound

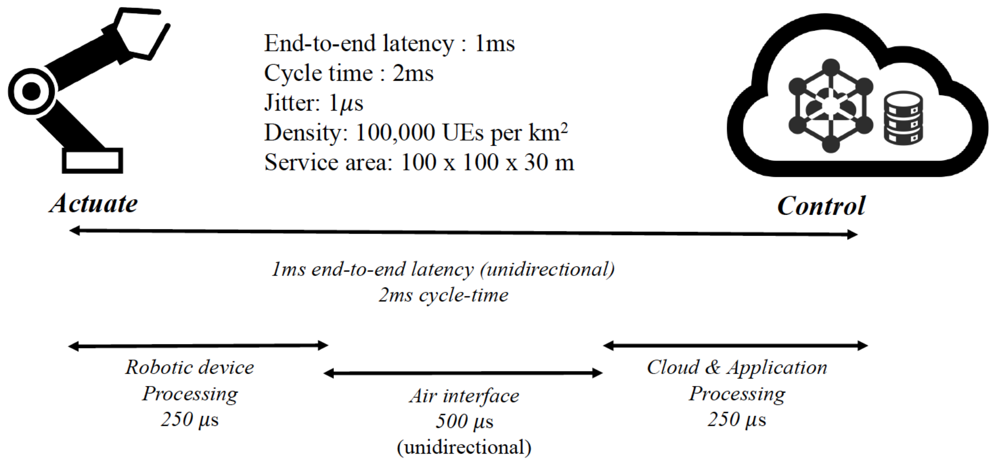


Figure 3: Factory automation Use case [5]

are sent to the operator side through the 5G network. The VR glass reproduce the robotic site environment using the captured video, in real time. The operator will perform a command to send to the robotic hand, using the joystick.

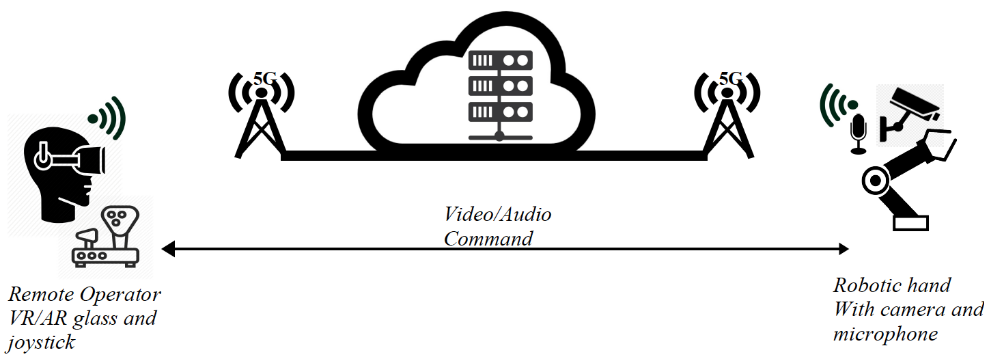


Figure 4: AR/VR use case illustration

4.3 Connected glass use case

According to [6], connected glass is a multiple-purpose wearable device initially designed for health and road safety applications. This kind of device is has multiple sensors such as camera/microphone, infrared proximity,....

It is supposed to be able continuously monitor and detect the different human activities, then send all collected data to the network while staying non intrusive. Another use of connected glass consist of directing a technician who perform a remote work. It has similarity with the AR/VR use case but the human execute the action instead of a robot hand.

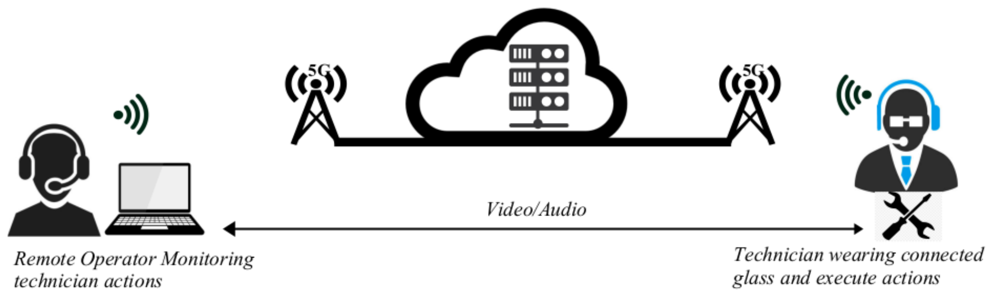


Figure 5: connected glass use case illustration

And instead of command, the remote operator provide vocal instructions to the technician, remotely.

5 Latency measurement

Multiple method are used to measure latency in a network but it can be generally measured as one way or two ways delay. Some tools evaluate directly the two ways delay, know as round trip time (RTT), of the packet an some of them measure only the one way delay between the two end points. Both end points of the network (source and destination) are supposed to be accessible, synchronized and should cooperate for one way delay measurement [7]. In one way delay measurement, the receiver side act only as a reflector of the sent packet.

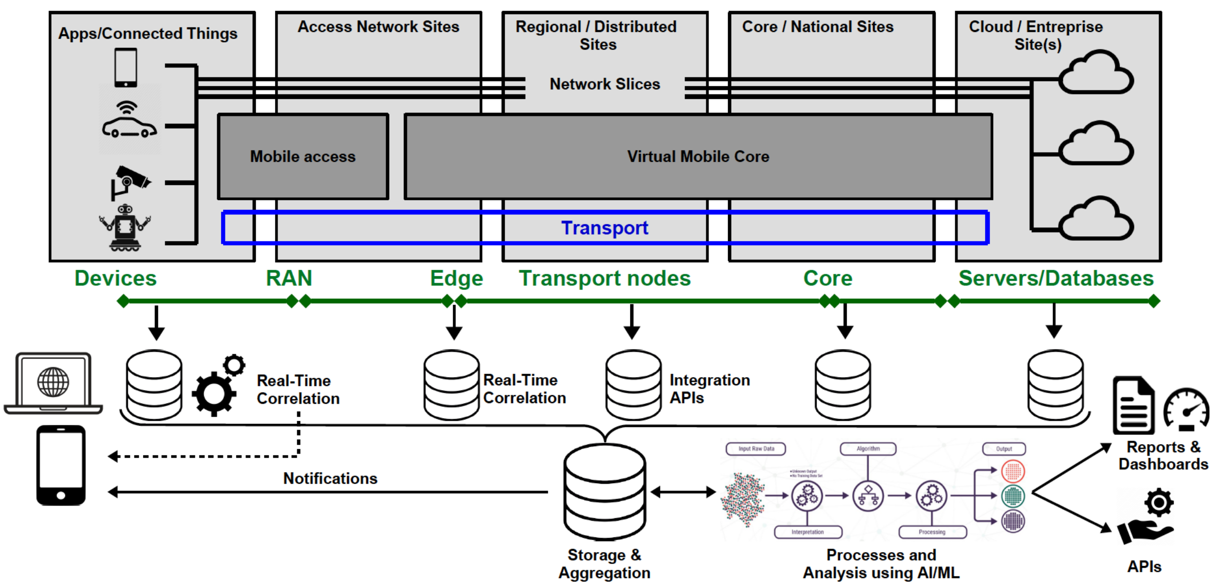


Figure 6: Precise and continuous 5G latency measurements with detailed analysis

5.1 Measurement concept and approach

Generally, measurement to evaluate network metrics is based on 3 approaches [8] :

☆ **Passive measurements :** this measurement approach consist on measuring networks metrics without modification in the network traffics. The passive measurement will deliver a detailed information about the measured end point.

☆ **Active measurements :** Active measurement consist on generating probe packets which is then sent trough the network, from the sender host to the destination. The probe packets are observed in the sender or the receiver host to get information about the network performance. This approach evaluate several metrics such as packet delay time, average of packet loss as well as connection bandwidth. Active measurement is the primary measurement method as it is flexible, provide accurate result and easy to deploy in the network [9]. The most popular active measurement tools are ping and traceroute which are generally used by most of operating system[7].

☆ **Hybrid measurements :** this approach consist on combining both active and passive measurements. More clearly, specific probe packets are injected to the network using active approach. These packets are tracked and monitored with passive way in network devices where they cross through the end to end path.

For the case we are investigating here, Figure 15 show the architecture of the measurement approach. To process that, we have adopted both measurement approach.

5.2 Measurement tools

A multitude of network measurement tools already exists according to their measurement methods. Table 3 shows some of them. We are actually investigating more deeply on some of these tool, such as

✧ The tools which based on ICMP[3], TCP[4] and UDP[5] protocols (ping and traceroute)

✧ Tools based on Precision Time Protocols (PTP) such as PTPD.

✧ Tools based on One and Two Way Active Measurements Protocols (Owamp and Twamp).

[3]Internet Control Message Protocol

[4]Transmission Control Protocol

[5]User Datagram Protocol

Table 3: Latency measurement methods and tools [10]

Tools	Method	Probe type	Availability	Storage/analysis
Ping	RTT; Packet loss ratio	ICMP	Single Measurement	Locally; analyze independently
Traceroute	RTT	ICMP ECHO/TCP	Single Measurement	Locally; analyze independently
Cisco IP SLA	RTT(average); one-way delay; packet loss	ICMP/UDP/ TCP/HTTP/DNS	Always-on	Locally; analyze independently
Pingmesh	RTT; Packet loss ratio	TCP/HTTP	Always-on	Cosmos and SCOPE
NetROAD	RTT; Packet loss ratio	UDP	Always-on	Scribe and Scuba
Everflow	link RTT	Packet marked with debug bit	Single Measurement	Custom analyzer and SCOPE
SLAM	Network path latency distribution	Crafted probe	Single Measurement	Controller
INT	End to end latency	Crafted probe	Single Measurement	Last switch on path; analyze independently
LossRadar	Packet losses at switches	No probe	Always-on	Custom collector and analyzer
TIMELY	RTT	TCP	Always-on	Locally; analyze independently
PTPmesh	One-way delay (average); Packet loss ratio	UDP	Always-on	Locally; analyze independently
OWAMP	One-way delay average	TCP/UDP	Always-on	Locally; analyze independently
TWAMP	Two-way delay average	TCP/UDP	Always-on	Locally; analyze independently

5.2.1 ICMP based tools based tools

Ping and Traceroute tools use ICMP protocol echo request and echo reply [11]. It provide the round trip latency between the sender and the target end point. Measurement done with these tools are active measurement, since the agent send probe packet as a request to the server and the latter reflect this packet to the sender.

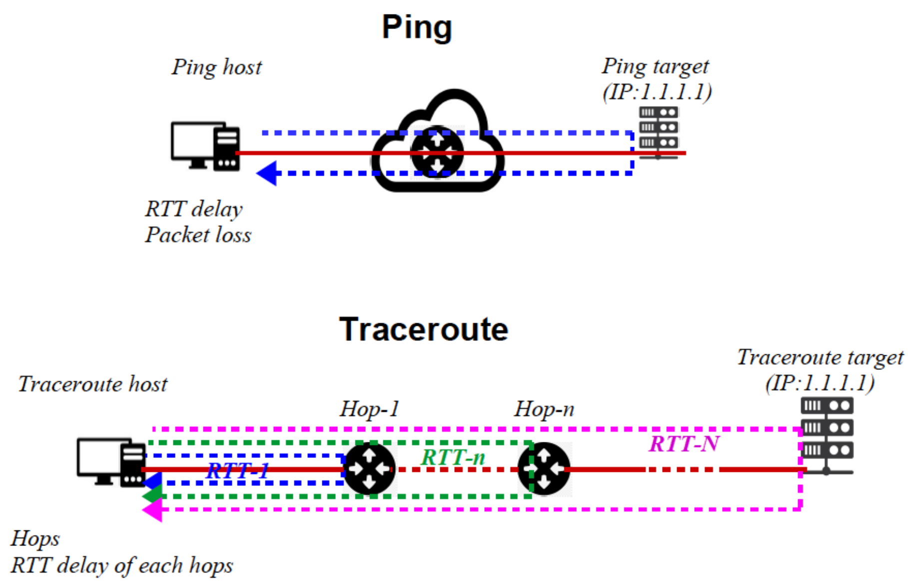


Figure 7: Ping and Traceroute principle

These tools (ping/traceroute) proviede in their output the time that the probe packet takes to cross the network from the sender to the server and the return. The output time also called as RTT of the packet is considered as the measured end to end latency.

5.2.2 Ptp protocol based tools

The Precision Time Protocol known as PTP is a time protocol used to synchronized time in the network. This protocol employ the Master/Slave architecture to apply time synchronization based on the timestamp from the master server. PTP protocol my be also used to infer network latency and packet loss in data-center according to [10]. Figure 8 show how ptpd master and ptpd slave send messages to each other. First, at time T1, the master send a sync message to the slave. which record locally the arrival time of this message as T'1. If the sending time (T1) in not sent with this sync message, the server send a followup message containing the time T1. The difference T'1-T1 represent the master to slave delay. At T2, the

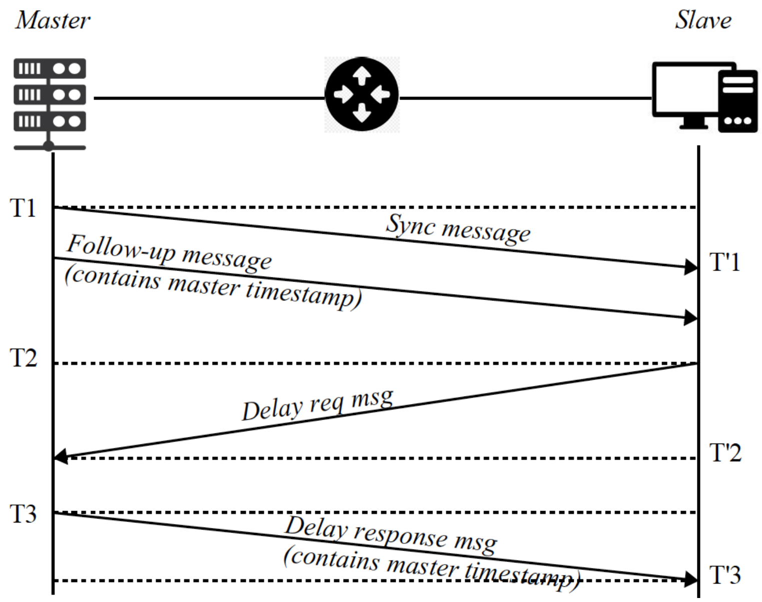


Figure 8: PTP messaging protocol

slave send a delay request to the master and record this time T2 locally. The server receive this request at time T'2, and send a delay response message to the slave containing the time T'2. The difference T'2-T2 represent the slave to master delay. According to [10], the one way delay is the mean of the master to slave and slave to master delay.

5.2.3 OWAMP/TWAMP protocol based tools

With roundtrip-based measurements, it is hard to isolate the direction in which congestion is experienced. One-way measurements solve this problem and make the direction of congestion immediately apparent [12]. Owamp provide a common protocol for measuring one way metrics between network devices [13].

Owamp/twamp architecture usually comprised of two hosts with specific roles (Fig. 9)

This tool can be use bidirectionally (one way metrics from both direction in the network) way, but not suitable for round trip active measurement. Owamp/twamp tools measure accurate and precise value of QoS parameters such as network latency, jitter and packet loss.

This tools work as a client/server application. The client request to the daemon server a test session to the server. This request contains an indicator and parameters according to the packet characteristics. The server will accept or deny this client request. Figure 10 describe the detailed message between the server and client.

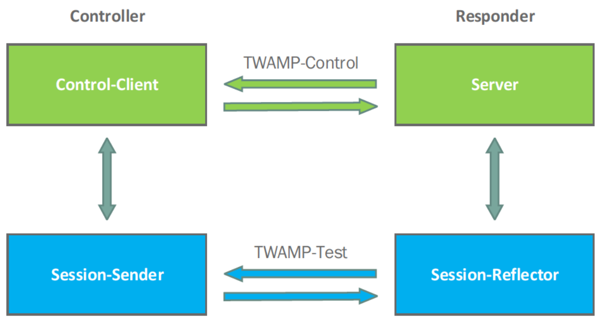


Figure 9: Twamp based architecture[14]

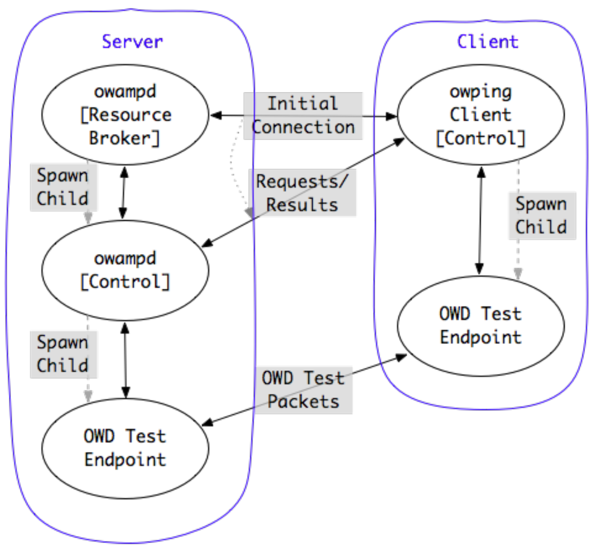


Figure 10: OWAMP detailed structure[12]

6 Application presentation

In this section, we present the implemented architecture that we have used to calculate, measure and visualize the latency data between the 5G collecting device and the beacon server. For this purpose we are using the open source applications which will be explained in the following sections.

6.1 Architecture

The proposed architecture follows several specifications in order to assure both the network operators and also the customers a reliable and scalable real-time latency data collection,

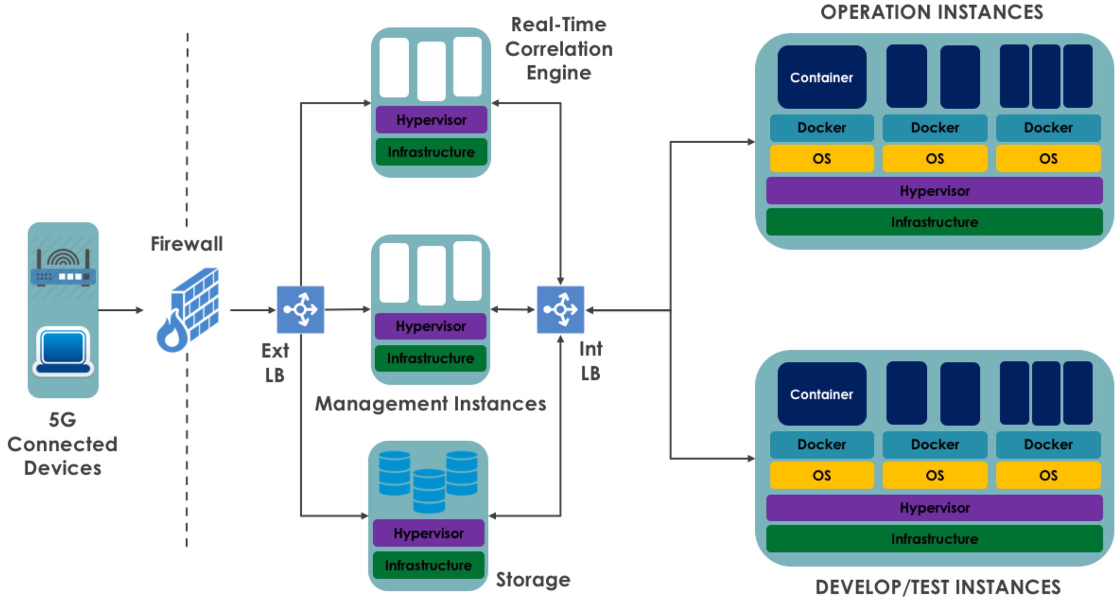


Figure 11: Cloud-Based Proposed Architecture

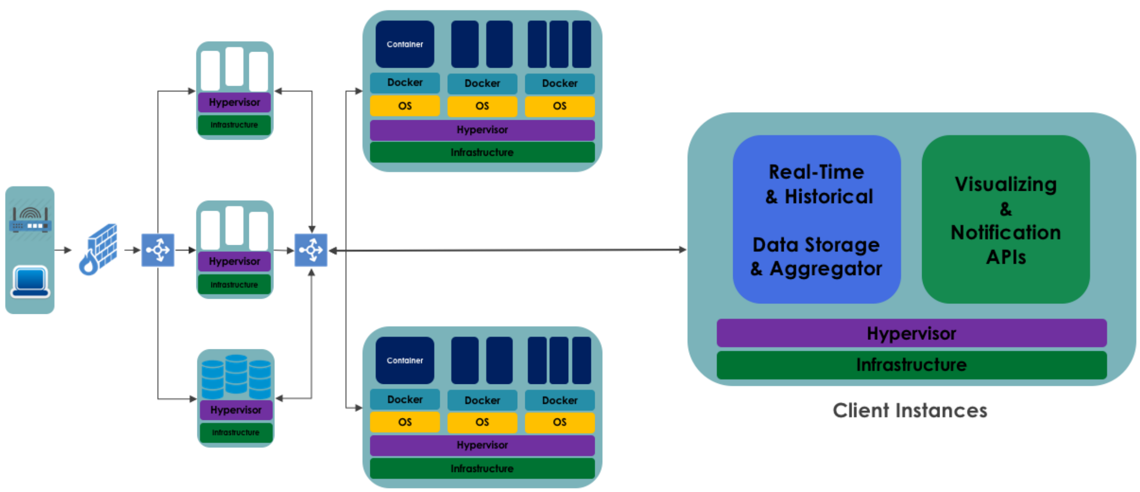


Figure 12: Small Customer Architecture

monetization and analysis within a secure private network. Some of these target specifications are:

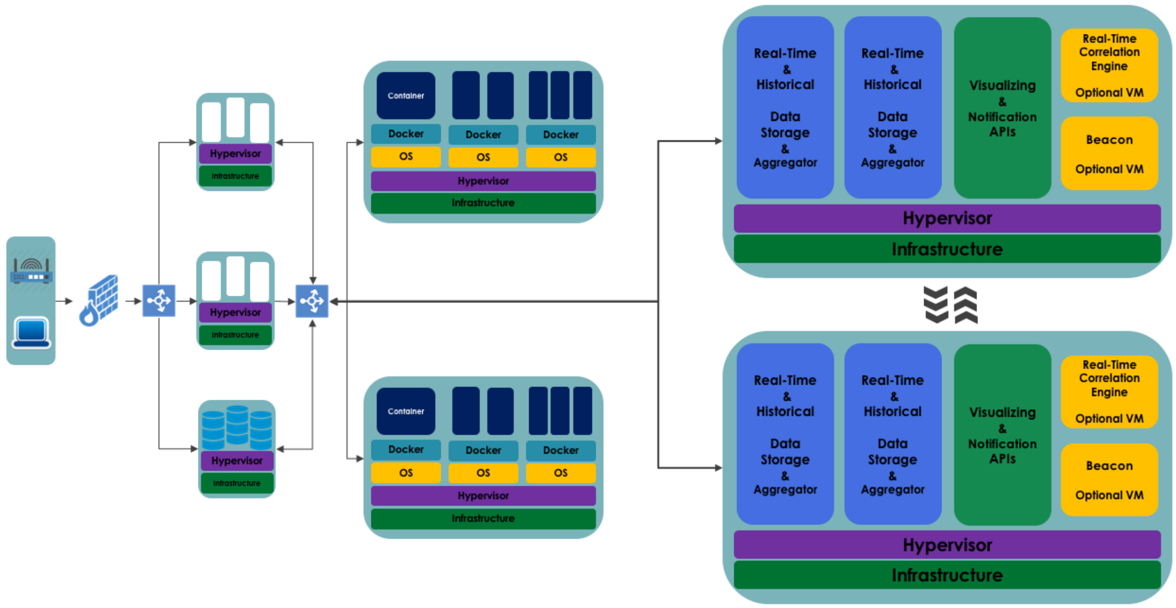


Figure 13: Large HA Customer Architecture

- Collection software agent shall be platform/OS agnostic
- Storage/Aggregation components shall be SAAS/Vendor agnostic (i.e. capable to run on Azure, AWS or GCP)
- Private instances per customer
- Manage own instance and customer instances (without seeing customer private data)
- Ability to aggregate and store latency data for historical purposes and reporting (e.g. 10Gbs per customer, 1 year retention)
- Capacity to extract dataset for AI-based analysis
- Capability to export data or invoke function using API
- Ability to collect environmental data together with latency data (location, frequencies, devide type, weather status, etc.)
- DevOps approach to quickly iterate between coding-testing-production
- Ability to present data using customized dashboards

- Capacity to generate alerts and notifications when thresholds are reached
- Ease of scaling to ten's of customers
- Architecture shall not use proprietary hardware
- Secure File Sharing for Businesses

6.2 Data Collection

The key element in every study and analysis is data. So our first step is to collect data from different sources through different destinations or beacons. For this purpose, different devices for different solutions have been proposed. In this chapter we will introduce three different type of our collecting devices and their purpose.

(a) Type I

(b) Type II

(c) Type III

Figure 14: Collecting devices

- Type I: The collecting device type I is our standard device with 1.5GHz 64-bit quad-core ARMv8 CPU and 4Gb RAM which supports 5G(NR,SA), 4G/LTE, 3G and GPRS and also could be connected through Ethernet, WiFi and USB. (Fig.14a)
- Type II: This device has the same system architecture as Type I but it has also a screen, so the collected metrics could be shown an analysed at the first step in the collecting device which is the most near real-time monitoring and analyzing point in data collection and analysis. (Fig.14b)
- Type III: This device is a very advanced device with lot of resource power specially for ML/AI solutions at the collecting point. It uses Quad-core ARM A57 @ 1.43 GHz CPU

and also 128-core Maxwell GPU and 4 GB 64-bit LPDDR4 25.6 GB/s RAM to ensure the AI/ML performance and supports all the RF bands and connectivity interfaces. (Fig. 14c)

6.3 Data storage and Analysis

This section represents the structure of the collected data and the open source applications that we use to store it. We use InfluxDB time series database which is designed to handle high write and query loads. It is using as a backing store for the use cases with large amounts of timestamped data. Here are some of its features that makes it a great choice for working with [15]:

- Custom high performance datastore written specifically for time series data. The TSM engine allows for high ingest speed and data compression
- Written entirely in Go. It compiles into a single binary with no external dependencies.
- Simple, high performing write and query HTTP APIs.
- Plugins support for other data ingestion protocols such as Graphite, collectd, and OpenTSDB.
- Expressive SQL-like query language tailored to easily query aggregated data.
- Tags allow series to be indexed for fast and efficient queries.
- Retention policies efficiently auto-expire stale data.
- Continuous queries automatically compute aggregate data to make frequent queries more efficient.

6.4 Data presentation and Reporting

The last important part of all data collection and analysis is to show and monitor the network latency in near real-time and also have the capability to investigate the historical data if needed. By monitoring the latency and its dependant metrics in real-time, the system has the ability to send alerts and notifications for differnet scenarios according to the use case and demands of the network and latency. For this purpose, we use Grafan which is a completely well-known open source monitoring API. Some of its powerful features are:

- Visualize: Fast and flexible visualizations with a multitude of options allow you to visualize your data any way you want
- Dynamic Dashboards: Create dynamic and reusable dashboards with template variables that appear as dropdowns at the top of the dashboard
- Explore Metrics: Explore your data through ad-hoc queries and dynamic drilldown. Split view and compare different time ranges, queries and data sources side by side
- Explore Logs: Quickly search through all your logs or streaming them live
- Alerting: Visually define alert rules for your most important metrics. Grafana will continuously evaluate and send notifications by mail or to systems like Slack, PagerDuty, VictorOps, OpsGenie

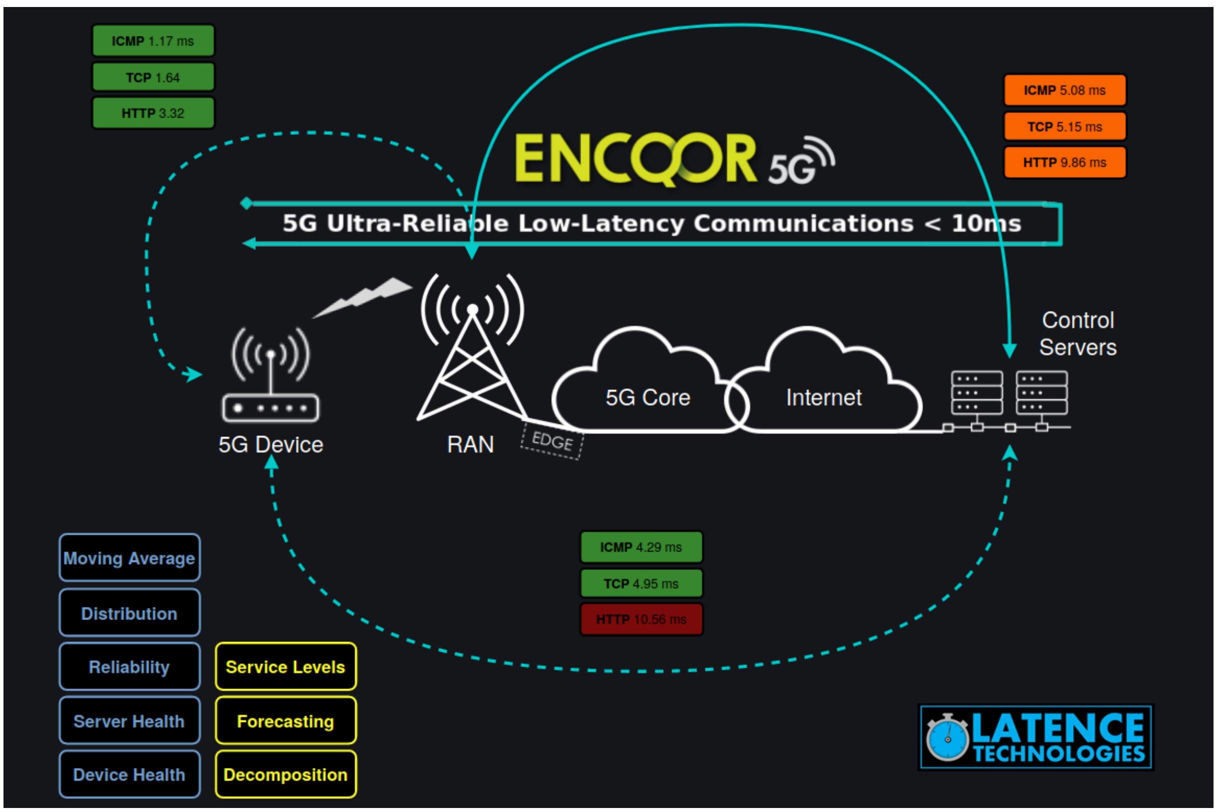


Figure 15: Real-Time Presenting Dashboard

6.5 Data output structure

The output data generated by the network agent install in the collecting device shows the measurement values and several tags that can identify it. These values are inserted in the influx database in real time.

Table 4: Historical data structure

Structure of historical data				
General		**Key**	***Value type***	***Comment***
	Global Tags			
		Site	*VarChar*	Measurement site name/description
		host	*VarChar*	Hostname where agent is launched
		Interface	*VarChar*	Network interface used by agent (Ethernet, Wifi,...)
		SessionID	*VarChar*	Identification of the session(Composed by session Date-time, host,...)
		5GDevice	*VarChar*	Used devices for measurement (5G devices)
		5GBand	*VarChar*	Band used by 5G network
		Beginning_date	*VarChar*	Beginning date-time of the collection
		End_date	*VarChar*	Ending date-time of the colllection
		Periodicity	*VarChar*	Aggregation Period (hourly, daily, weekly, mounthly, year,...)
Network				
	Http			
		http_response_time	*float*	Average response time received bye the http agent during a period (in second)
		http_server	*VarChar*	Web adress of the beacon server where request is sent (eg. : http://52.229.88.107)
	Ping			
		ping_average_response	*float*	Average value of response time average received bye the icmp agent (in millisecond)
		png_packet_loss	*float*	Average value of percentage of packet loss (%)
		reliability	*float*	Network reliability
		ping_url	*VarChar*	Ip adress of the beacon server (eg.: 52.229.88.107)
	Tcp			
		tcp_response_time	*VarChar*	Average response time received bye the tcp agent (in second)
		tcp_server	*VarChar*	IP adress of the beacon server where request is sent (eg. : 52.229.88.107)
	Aggregate			
		agg_latency	*float*	aggregate latency calculated

The configuration of the agent define the collection time periodicity.

Following is a portion of data output in Json format. This Sample result shows a single measurement of Ping response and percentage of packet at defined timestamp.

```
...
{
    "fields":{
        "average_response_ms":26.358,
        "percent_packet_loss":0
    },
    "name":"ping",
    "tags":{
        "Interface":"Ethernet & Wifi",
        "Site":"encqor",
        "host":"whitepi4",
        "url":"10.10.10.2"
    },
    "timestamp":16039974560
}
...
```

In short, the structure of the output data shows all measured data and metrics. The measurement time stamp will be considered as an index of each measured value. All tags in the measurement store information about each metrics, and some information about the measurement.

7 Conclusion

Bibliography

[1] X. Jiang, H. Shokri-Ghadikolaei, G. Fodor, E. Modiano, Z. Pang, M. Zorzi, and C. Fischione. Low-latency networking: Where latency lurks and how to tame it. *Proceedings of the IEEE*, 107(2):280–306, 2019.

[2] A. A. Bisu, A. Purvis, K. Brigham, and H. Sun. A framework for end-to-end latency measurements in a satellite network environment. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–6, 2018.

[3] E. P. C. Jones, L. Li, J. K. Schmidtke, and P. A. S. Ward. Practical routing in delay-tolerant networks. *IEEE Transactions on Mobile Computing*, 6(8):943–959, 2007.

[4] T. Fehrenbach, R. Datta, B. Göktepe, T. Wirth, and C. Hellge. Urllc services in 5g low latency enhancements for lte. In *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, pages 1–6, 2018.

[5] GABRIEL Brown. Ultra-reliable low-latency 5g for industrial automation. *Technol. Rep. Qualcomm*, 2:52065394, 2018.

[6] A. Arcaya-Jordan, A. Pegatoquet, and A. Castagnetti. Smart connected glasses for drowsiness detection: a system-level modeling approach. In *2019 IEEE Sensors Applications Symposium (SAS)*, pages 1–6, 2019.

[7] Kim Ervasti. A survey on network measurement: Concepts, techniques, and tools.

[8] Venkat Mohan, YR Janardhan Reddy, and K Kalpana. Active and passive network measurements: a survey. *International Journal of Computer Science and Information Technologies*, 2(4):1372–1385, 2011.

[9] Li Wenwei, Zhang Dafang, Yang Jinmin, and Xie Gaogang. On evaluating the differences of tcp and icmp in network measurement. *Comput. Commun.*, 30(2):428–439, January 2007.

[10] D. A. Popescu and A. W. Moore. Ptpmesh: Data center network latency measurements using ptp. In *2017 IEEE 25th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS)*, pages 73–79, 2017.

[11] G. Jackson, P. Keleher, and A. Sussman. A ping too far: Real world network latency measurement. In *2015 IEEE 11th International Conference on e-Science*, pages 580–588, 2015.

[12] Aaron Brown. One-way ping (owamp), December 2020.

[13] Kaynam Hedayat, R Krzanowski, Al Morton, Kiho Yum, and Jozef Babiarz. A two-way active measurement protocol (twamp). Technical report, RFC 5357, October, 2008.

[14] R. Zhu, B. Liu, D. Niu, Z. Li, and H. V. Zhao. Network latency estimation for personal devices: A matrix completion approach. *IEEE/ACM Transactions on Networking*, 25(2):724–737, 2017.

[15] Inc InfluxData. Influxdata documentation. *https://docs.influxdata.com/*, 2020.