WHITE PAPER

Celona MicroSlicing™

Essentials for Enterprise Network Admins

OCTOBER 2021

celona

CONTENTS

Introduction

Traditional Quality of Service (QoS) within corporate networks, such as the use of VLAN assignments and DSCP markings used to determine traffic classification for network data, have long been used as the de facto method of controlling and prioritizing the convergence of voice, video and data traffic flows. However, modern 5G technologies have taken portions of what's great about QoS and built-in additional improvements to codify what's generally referred to as 5G network slicing. This is a set of standards formulated by the 3GPP -- the primary standards organization responsible for advancing the architecture 5G.

Taking 5G network slicing one step further, Celona has developed what's known as MicroSlicing – a set of network functions that allows for highly-granular control over data flows at an application level. On the Celona radio access network, a QoS scheduler accounts for all the pre-defined QoS requirements for a given application or device group. These QoS metrics are applied on both the uplink and downlink for each cellular connection over the air. Concurrently, dedicated network edge resources, such as real-time compute, memory and data path enforcement are allocated for each application or device group to create unique traffic isolation for each MicroSlice that can then be tagged or translated to the requisite VLAN or network policy that are in place within the existing enterprise network. This enables granular end-to-end service levels applied on a per application or subscriber identity basis that span the RAN and the LAN.

This document provides an educational overview, design considerations and guidelines for implementing Celona MicroSlicing™ within an enterprise environment. It assumes that readers are already familiar with traditional QoS terms, concepts and integration techniques that are commonly deployed on enterprise-grade routers, switches and endpoints.

MicroSlicing Overview

This section answers the following questions:

What is MicroSlicing?

MicroSlicing is a unique and patented technology developed by Celona specifically for private 4G LTE/5G NR network infrastructure including the EPC / 5GC core network platforms. MicroSlicing is defined as a set of network functions within a private 5G LAN that allows various quality of service metrics and thresholds to be automatically defined and enforced on discrete application traffic flows or device groups to form a cohesive, end-to-end logical network that maps quality of service from the ingress cellular RAN through the entire existing enterprise L2/L3 network. A single MicroSlice can be configured to meet an application's network requirements from a service level agreement (SLA) perspective. An SLA within a MicroSlice includes a range of network performance requirements as it traverses the 5G RAN. Configurable options include:

- Flow priority
- Packet delay budget (PDB)
- Packet error rate (PER)

The control plane portion of the Celona 5G LAN platform continuously monitors and adjusts traffic transmissions if network congestion occurs. Unlike traditional network QoS that throttles traffic that's already been transmitted, 5G technologies more efficiently control the transmission of data prior to it being sent. These processes and functions provide strict service level agreements (SLAs) that guarantee that higher-priority traffic is given preference and special treatment over lower-priority flows in real-time as defined by network administrators

MicroSlices also add an extra layer of security to corporate networks. For one, data flowing within each MicroSlice is logically isolated from all other slices. Data encryption can be used within each MicroSlice on a per-flow basis.

The other aspect of a 5G LAN in general is that MicroSlice data can be architected in such a way so that as traffic exits the RAN and enters the LAN, it can be directed to a secure network segment such as a firewall DMZ. This provides further network separation for instances where 5G traffic should not be allowed to flow across trusted portions of the corporate LAN.

.

Different MicroSlice types

MicroSlicing uses the telecommunications concept of bearer channels. Bearer channels are logical paths across the network that are solely responsible for user and device data transport. A MicroSlice can be configured as one of three bearer types depending on the performance needs of the application data being transported across the bearer channel. The three types of MicroSlice bearer channels are:

- **Guaranteed bit rate (GBR)** – used to set traffic policy for device groups and applications that are bit rate sensitive. GBR defines the minimum bit rate value allocated to a bearer. Bit rates higher than the GBR are possible if sufficient network resources are available.
- **Non-guaranteed bit rate (Non-GBR)** - used to set traffic policy for device groups and applications that are not bit rate sensitive.

Defining, prioritizing and enforcing MicroSlices

While choosing a bearer channel sets the framework for a MicroSlice, it must still be defined and prioritized using network-based SLA parameters. This is accomplished using what's known as QoS class options. Celona MicroSlices use the 4G LTE QoS Class Identifier (QCI) and 5G QoS Identifier (5QI) as defined by 3GPP TS 23.203 standard. This is done to simplify the process of setting the appropriate network quality parameters used in packet scheduling across the 5G RAN. For example, for 4G LTE operation, most Celona RAN administrators will only need to concern themselves with QCI's 1 through 9 for the vast majority of traffic flows (see below Table):

QCI	Bearer Type	Priority Level	Packet Delay Budget (PDB)	Packet Error Rate (PER)	Example Services
1	GBR	2	100 ms	10-2	Conversational Voice
2	GBR	4	150 ms	10-3	Conversational Video (Live Streaming)
3	GBR	3	50 ms	10-3	Real Time Gaming
4	GBR	5	300 ms	10-6	Non-Conversational Video (Buffered Streaming)
5	Non-GBR	1	100 ms	10-6	IMS Signalling
6	Non-GBR	6	300 ms	10-6	Video (Buffered Streaming)
7	Non-GBR	7	100 ms	10-3	Voice, video, gaming
8	Non-GBR	8	300 ms	10-6	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file
9	Non-GBR	9	300 ms	10-6	file sharing, progressive video, etc.

3GPP QCI 1-9 for 4G LTE. (5G NR has larger set of priority classes: 5QIs that enable much lower Packet Delay Budget values)

The QCI diagram above indicates several 3GPP standardized settings for classes 1-9. This includes the bearer type (either GBR or Non-GBR), the defined QoS priority level, packet delay budget (PDR), packet error rate (PER) and some examples of the types of common enterprise-grade applications/services that are commonly applied to the class. Also note that QCI is a term used in conjunction with 4G LTE networks. 5G networks use a slightly different class identification system known as 5G QoS Identifier (5QI). However, in the context of this document, QCI and 5QI can be used interchangeably. As you can see, GBR bearer channels are assigned to QCI values 1 through 4 while values 5 to 9 are for non-GBR bearer channels. Also, it's important to point out that QCI values to not always align with how data flows are prioritized. The higher the priority number is, the more likely it will be discarded if congestion occurs. Thus, for example, QCIs 1 to 4 will be discarded before QCI 5 traffic as it has a preferred priority value of 1.

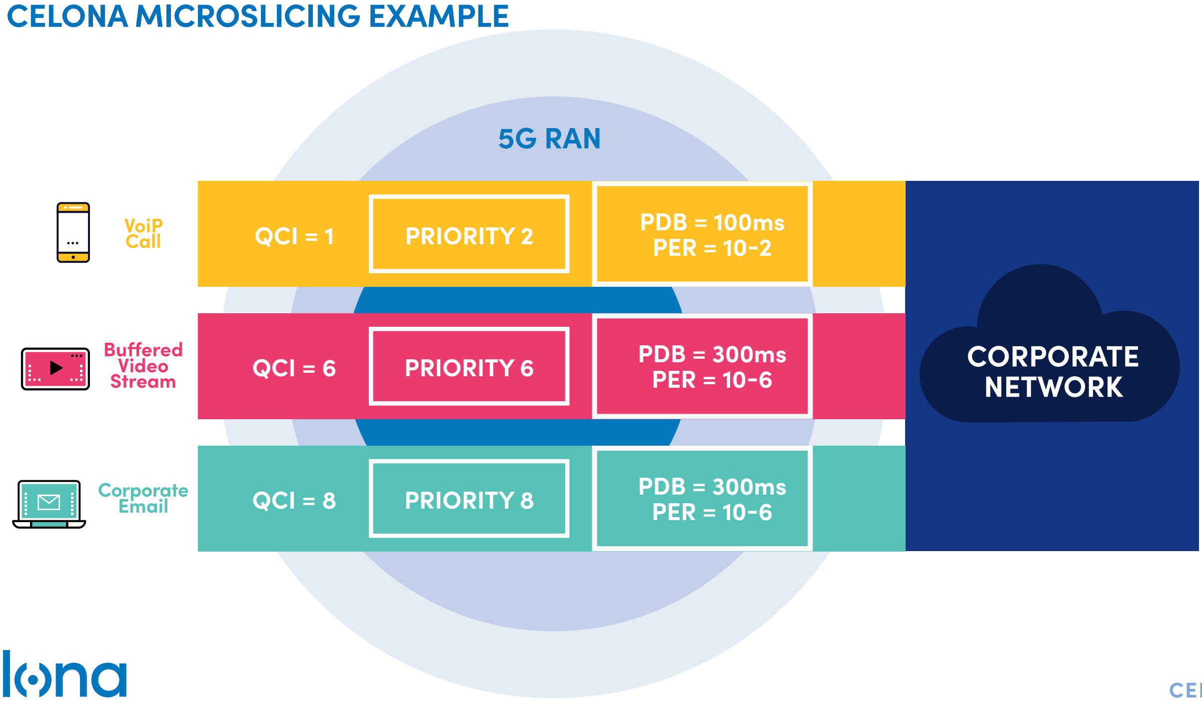
The other two QCI parameters are defined as follows:

Packet delay budget (PDB) - used to set the maximum amount of time (in milliseconds) that a packet can be delayed. For example, a 5G RAN with packets assigned to QCI of 1 and a priority of 2 can delay traffic up to 100ms before having to resort to discarding traffic with lowest priority value at that time.

Packet error rate (PER) - the ratio of the number of packets not successfully received compared to the overall number of packets transmitted. For example, a packet with a QCI value of 1 and a priority of 2 can drop a maximum of 10-2 (10^{-2}) packets. In other words, 1 packet out of every 100 can be in error. Once this error rate exceeds 1 error out of every 100, the 5G platform will begin discarding the lowest priority traffic to attempt to bring the higher priority flow back within SLA guidelines.

The following diagram illustrates how a Celona 5G LAN with three MicroSlices defined operates. The thre Microslices directly align with three distinct services : 1) a VoIP call, 2) a video streaming and 3) corporate email. The priority number assigned to each MicroSlice dictates its likelihood of transmitted data being throttled at the source device in the event that network congestion occurs on the RAN. For example, if network congestion reached the point where either PDB or PER exceeds the set QCI values for the VoIP call, the first MicroSlice to be impacted would be corporate email flows as this MicroSlice is set with a less preferred priority value:

CELONA MICROSLICING EXAMPLE



The difference between Microslicing and 5G Network Slicing

5G LAN Microslicing and 5G network slicing, a standard feature of the 5G technology standard widely used by mobile operators are distinctly different, but these differences are often misunderstood.

Network slicing is an architecture that allows multiple virtual networks to be created on top of a common public or shared physical infrastructure using overlay techniques. From a 5G operator's perspective, a network slice is an independent, end-to-end logical network that runs on a shared physical infrastructure that is capable of delivering an agreed upon network service quality.

The customizable network slicing service options include transport speed, quality, latency, security and reliability. These service levels are delivered based on service quality requirements negotiated between the mobile operator and the business customer. The 5G operator is solely responsible for managing and orchestrating network resources to meet the customer SLA. From the carrier perspective, network slicing is a way to gain additional revenue opportunities for customers that wish to pay more for improved 5G service experiences.

In most cases, 5G operators create and manage network slices on a customer-by-customer basis. This means that all customer traffic traversing a public 5G network slice will receive identical QoS service levels regardless of the type of application or service that is being transmitted across the RAN.

Why 5G network slicing does not meet the QoS needs of enterprise networks

While network slicing at a customer-level may be sufficient for certain wireless use cases, most enterprise organizations are seeking ways to better optimize and guarantee traffic flows and specific rates for mission critical business applications and services. This is something that network slicing simply cannot achieve as it would become too complex and cumbersome to manage this number of slices across a nationwide 5G infrastructure.

MicroSlicing introduces flow-level network slices

Operating a private 5G LAN gives administrators the opportunity to create virtual MicroSlices at a highly-granular level when compared to carrier 5G network slicing options. From a per-application perspective, MicroSlicing is one of the only ways to provide strict, SLA-backed traffic flow guarantees across a wireless medium that can be directly integrated with the existing enterprise QoS framework already in place, such as VLANs, DSCP values, security policies, etc.

Celona creates a more granular approach by allowing administrators to organize their RAN by application and device group:

- **Device groups** – Administrators can place 5G endpoints that connect to the RAN via SIM card or eSIM into groups to logically segment them by device type. In many cases, the type of device often dictates what apps/services are being operated on a regular basis. This helps to coordinate which devices are assigned to a specific MicroSlice.
- **Applications** – Applications can be configured within the Celona platform so that the control plane can identify the data flow being transmitted across the RAN. Once the application is identified by the Celona platform, the flow can be assigned to the appropriate MicroSlice and associated network performance SLAs.

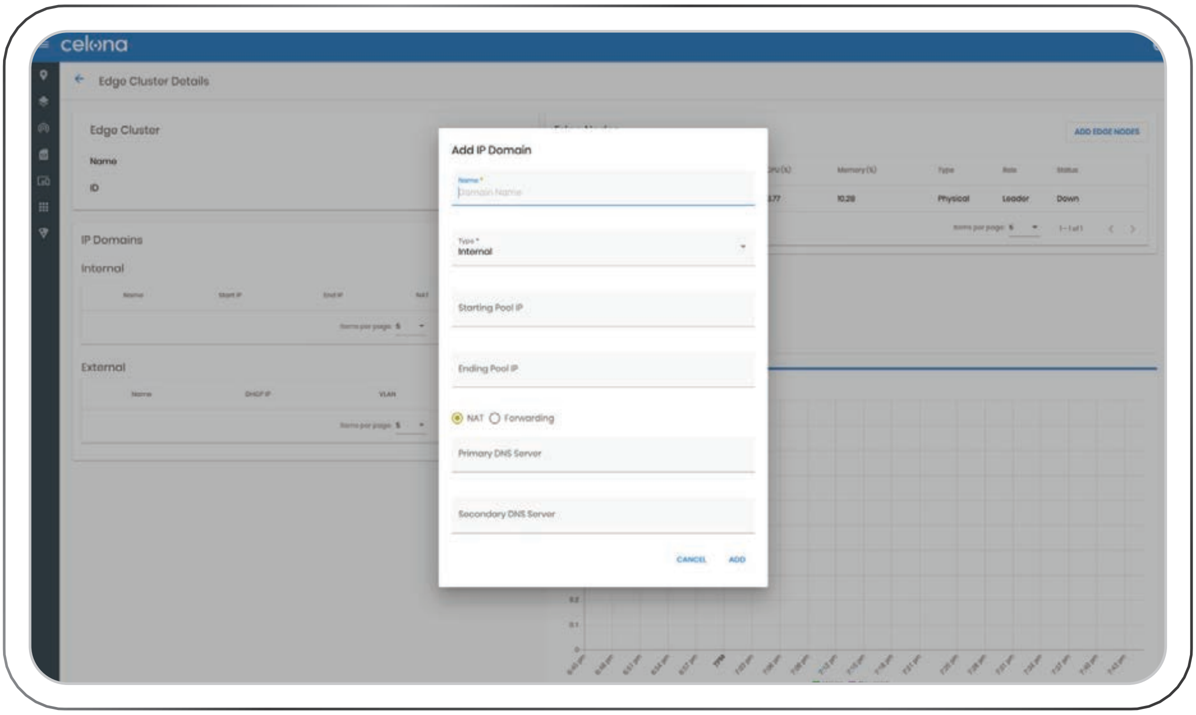
Configuring MicroSlicing on the Celona Orchestration Platform

A successful MicroSlicing implementation requires an understanding of the following concepts and configuration options:

- **Understanding and configuring IP Domains**
- **Celona MicroSlice default configuration settings**
- **Organizing 5G LAN endpoints into device groups**
- **Identifying applications**
- **Configuring a new MicroSlice**

Understanding and configuring IP Domains

In order for the Celona 5G LAN to be integrated and used as an extension of the corporate LAN, the proper IP Domain setup must be setup as shown here:



Depending on the needs of the business and network architecture, Celona provides the following three LAN integration types:

Internal IP Domain

- **NAT Mode (default)** – Configuring your IP Domain for in this internal domain mode means that the Celona Edge platform acts as a network address translation (NAT) gateway in a similar fashion that a traditional router or layer 3 switch translates internal IP addresses to one or more external IP addresses. All devices connected to the Celona 5G LAN receive their IP addresses through a DHCP service and will share the same DNS settings as the Celona Edge. In the event that the Celona Edge appliance is not configured with specific DNS settings, 5G LAN clients will default to using dns.google.com. Keep in mind, however, that in NAT mode, 5G LAN endpoints sitting behind the NAT gateway will not be accessible from devices on the outside of the NAT gateway. If traffic must be initiated from the LAN into the RAN, you must use Routed or VLAN mode.
- **Routing Mode (Internal DHCP)** – In this internal IP Domain mode, the Celona Edge platform operates as a standard layer 3 routed gateway for transport in and out of the 5G LAN. Unlike NAT mode, this allows devices on the traditional LAN to reach out and communicate with 5G endpoints.

External IP Domain

- **VLAN Mode** – This is considered an external IP Domain mode as the gateway attaches itself to an existing VLAN/subnet already configured on the corporate LAN. Thus, the VLAN number must be assigned to the IP Domain along with the IP address of the LAN's DHCP server, if needed.

For more detailed IP Domain configuration steps, please **click here.**

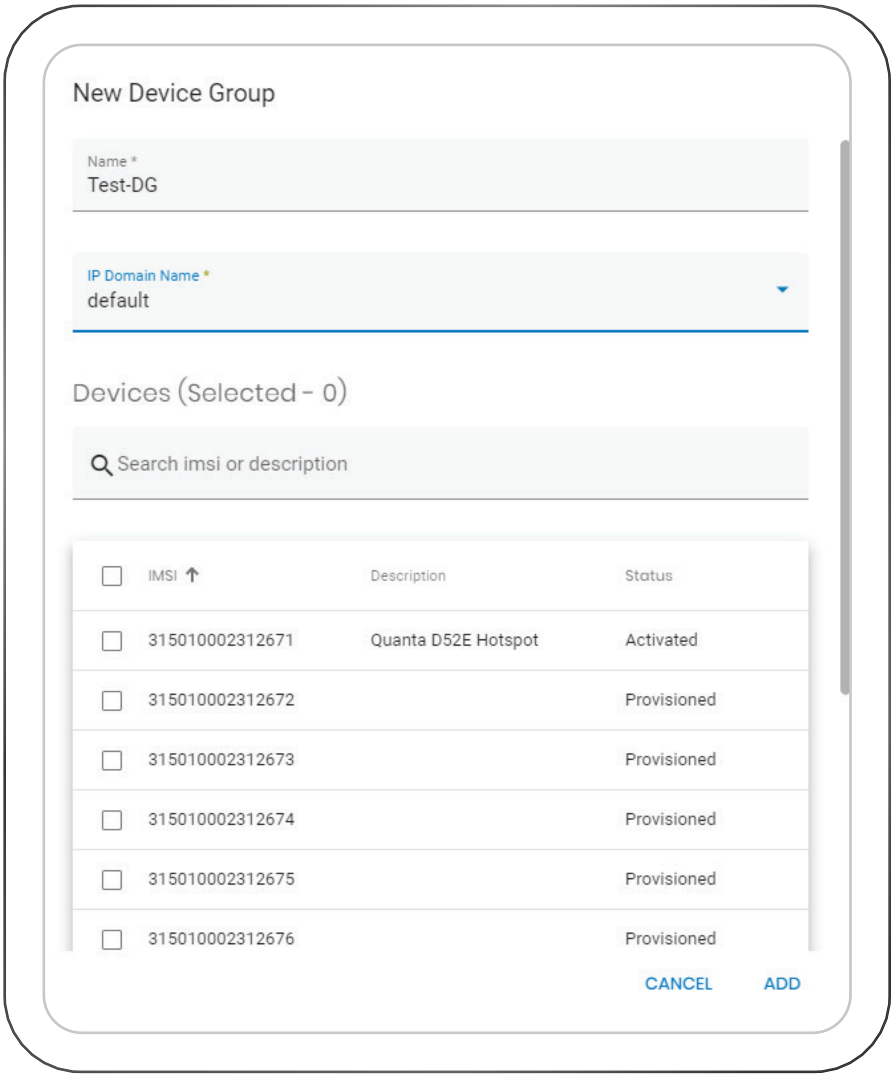
Celona MicroSlice default configuration settings

By default, all traffic that traverses the Celona 5G LAN is placed into a single, pre-configured MicroSlice. For most deployments, non-essential and non-latency sensitive data flows can remain in the default MicroSlice while administrators create and assign new MicroSlices for mission-critical and latency-sensitive traffic. The default MicroSlice is assigned with a QCI of 9. This means that the following QoS parameters are applied:

- Priority Level = 9
- BDP = 300 ms
- PER = 10-6

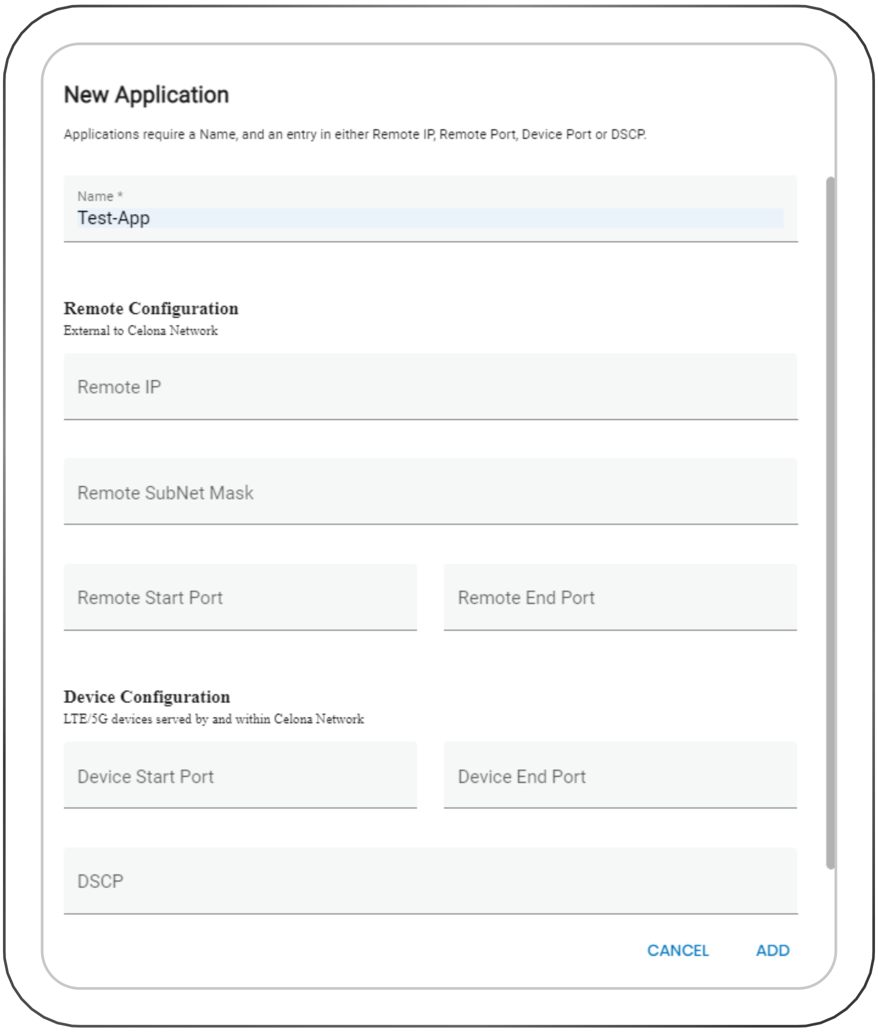
Organizing 5G LAN endpoints into device groups

For MicroSlices to provide unique network performance service levels for different types of traffic flows, administrators must identify the devices and users of those devices for the purpose of adding them to the slice. As you can see in the following Celona orchestrator screen capture, the creation of a device group requires that administrators name the group, assign the group to the correct IP Domain and identify/add devices using either the unique SIM card International Mobile Subscriber Identity (IMSI) number or the device description, if one is available:



Identifying applications

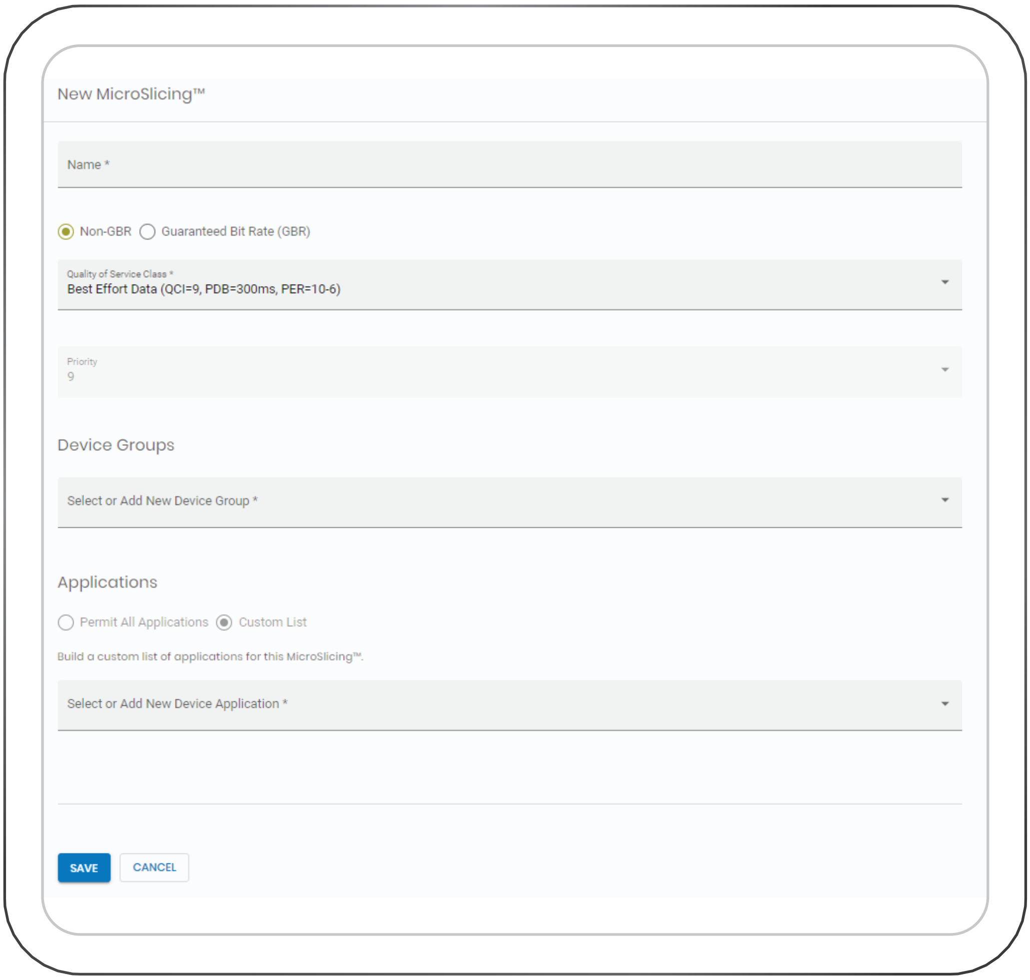
A MicroSlice must also be able to identify which traffic flows should be assigned to a specific MicroSlice. This is accomplished by defining the requisite applications on the Celona orchestrator that have specific traffic handling requirements. Creating a new MicroSlice application (see below screen capture) simply requires that an administrator assign a unique name for the app and then identify it using either remote or device (local) configuration identification methods. Remote configuration requires that the administrator input the IP address and subnet mask of the application server as well as the server's starting and ending port numbers that are used in the communication. Alternatively, an application can be defined on the device side which includes the device starting/ending ports as well as what Differentiated Services Code Point (DSCP) value the device will tag for each packet in the application flow:



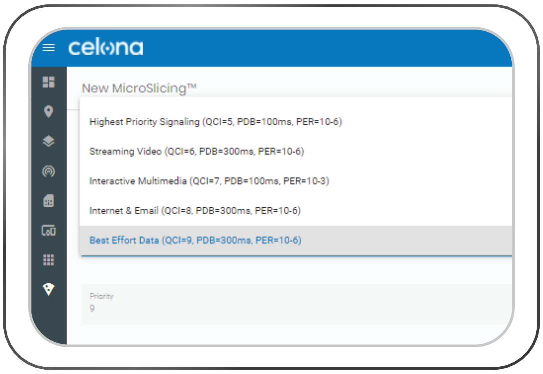
Configuring a new MicroSlice

Once the devices and applications are identified and organized into their respective groups, it's time to create a new MicroSlice. Within the MicroSlicing section, simply click on the "CREATE MICROSLICING™" button. The Celona orchestrator will then bring up the following configuration screen:

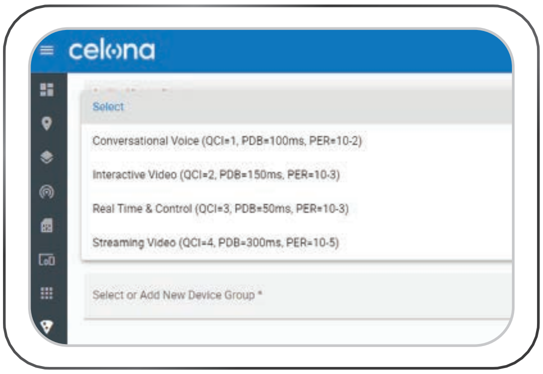
The first step is to type in a unique name that identifies the purpose of the MicroSlice. Next, administrators must choose between a Guaranteed Bit Rate (GBR) or a Non-GBR slice. Based on that selection, the following Quality of Service Class parameters will be available to choose from:



Guaranteed Bit Rate (GBR) QoS Class Options:

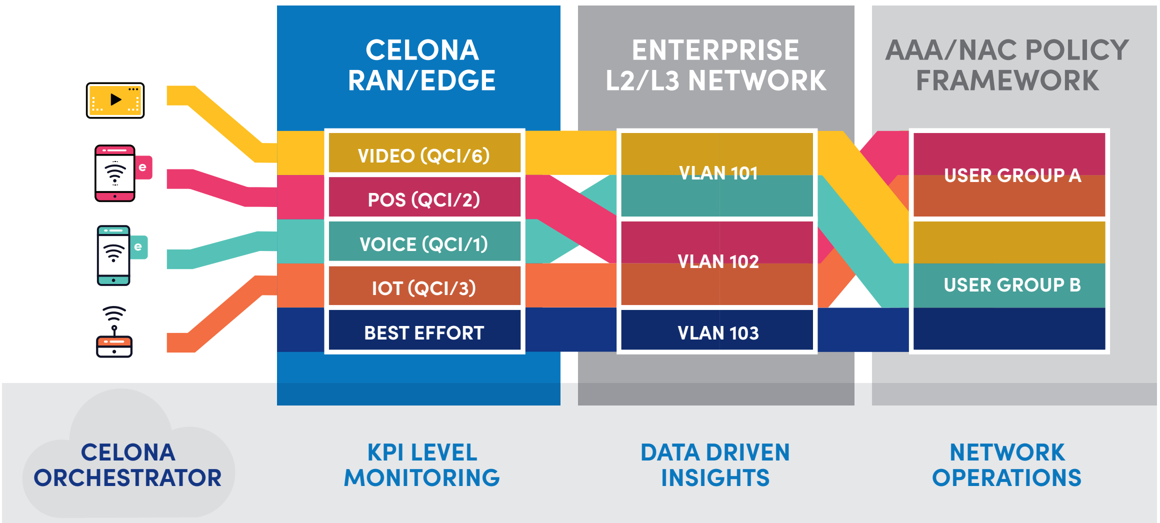


Non-GBR QoS Class Options:



Once a bearer type and QoS class option is selected, the final step is to assign the appropriate Application(s) and custom device group(s) to the MicroSlice.

To view a video demonstration that covers the initial Celona platform setup and example MicroSlicing configuration, please [click here.]

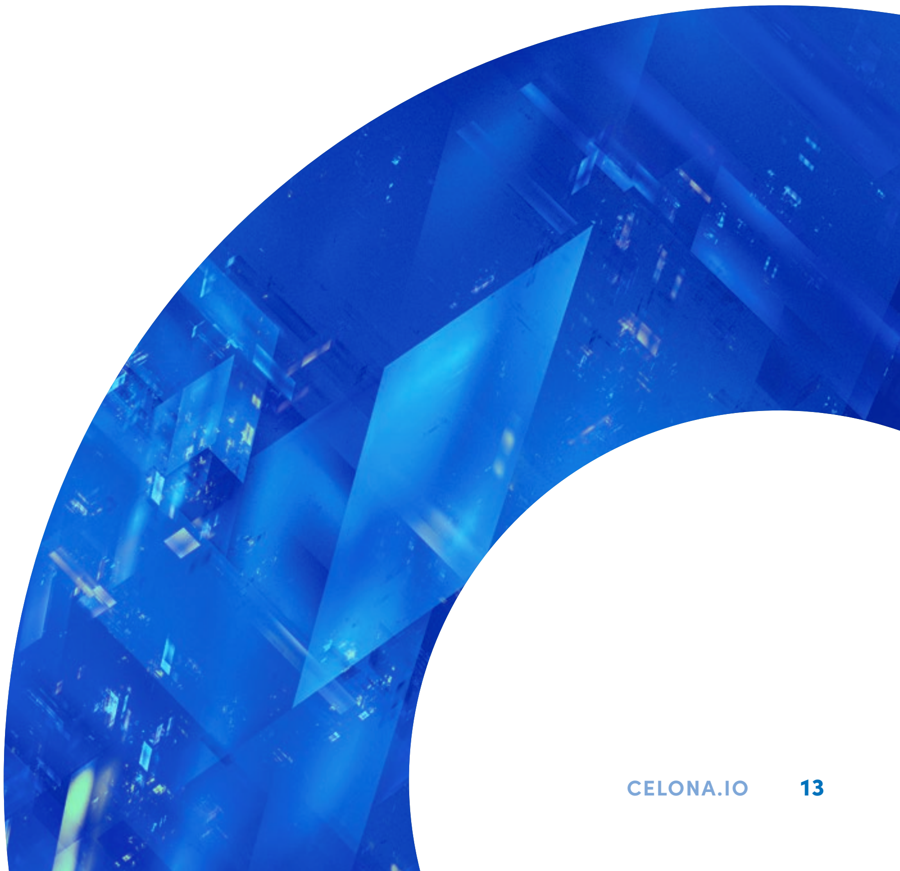


Translating MicroSlice SLA Policy into Traditional QoS

MicroSlice SLAs operate from one end of the 5G LAN to the other. However, once data leaves the radio network and flows into the corporate LAN, it's important to be able to translate MicroSlice settings into traditional QoS policy that enterprise-grade routers and switches understand.

Fortunately, this is an easy task as the Celona platform uses the same DSCP tagging mechanic found in traditional QoS to identify application traffic. QoS policy can then be created around these DSCP tags to provide preferential treatment to mission-critical application flows using queuing, forwarding and discarding mechanisms such as traffic shaping or traffic policing.

The following diagram depicts how 5G devices are segregated into individual MicroSlices such as video, POS, voice and IoT. Then once these packets hit the LAN, the same DSCP tags can be used to organize data into specific VLANs and apply QoS enforcement depending on the level of network performance required.



SEE THE CELONA TECHNOLOGY IN ACTION

Request a free trial and custom product demonstrations by visiting us at celona.io/journey.

hello@celona.io | celona.io

celona