Use Case #5

Wearable, mobile, point-of-view, wireless video service delivery

Overview and Objectives

This UC deals with auto scaling video routers in 5G Edge and aims to evaluate the performance of a 5G network in various use cases related to video and audio streaming quality, system scalability, and core performance metrics. The KPIs measured (outlined in Table 1) include video resolution, packet loss, audio quality (Mean Opinion Scorescore), system scalability, CPU, RAM, and network loads, and video flow capacity.

Use Case Description

For this use case, RedZinc provides wearable video for mobile telemedicine applications. BlueEye Handsfree is a wearable video solution for paramedics and nurses. A video camera is worn on an ergonomically designed headset. This camera sends live point-of-view video to a remote doctor. The doctor could be at a helpdesk, different part of the hospital, or at home at night. Real-time video of emergencies can benefit both patients and first responders. When the video is relayed to the emergency doctor at the hospital, the doctor can help with diagnosis, treatment and oversight before the patient reaches a hospital. For a stroke patient 'time is brain tissue' and for a heart attack patient 'time is heart muscle'. Using point-of-view video from the paramedic to 'immerse' the hospital-bound doctor in the remote scene allows for quicker delivery of, for example, clot busting drugs, which can benefit patient outcome.

RedZinc has also come up with wearable video for educational and training set-up. The medical professional gives a live demonstration of a medical procedure to students in remote locations such as students at home. The professor or medical professional wears the hands-free headset through which the live video and audio feed are transmitted to the students who are in remote locations. The students log in to the BlueEye portal using a laptop or smartphone and access the real-time video and audio feed.
This UC aims to evaluate the performance of a 5G network in various use cases related to video and audio streaming quality, system scalability, and core performance metrics. The KPIs measured (outlined in Table 1) include video resolution, packet loss, audio quality (MOS score), system scalability, CPU, RAM, and network loads, and video flow capacity.

Table 1 : UC5 KPIs

KPIs	Results expected	Experimentation results
UC 5.1	Achieve better video quality than 640x480.	The reported video quality was consistently 1280x720, which substantially exceeds the baseline KPI of 640x480.
UC 5.2	Achieve proper video feed quality by having less than 5% of data packet loss.	The video feed quality maintained less than 5% packet loss throughout all test scenarios, ensuring a stable and continuous video stream.
UC 5.3	Have no impact on audio feed from paramedic/remote doctor with audio MOS>4.	Audio feeds maintained a Mean Opinion Score (MOS) of over 4, indicating high clarity and minimal disruption.
UC 5.4	Achieve video and audio MOS>4 at the functional testing of the video.	During functional testing, both audio and video streams achieved a MOS greater than 4.

This project has received funding from the European Union's Horizon 2020
Innovation Action programme under Grant Agreement No 101016521.

UC 5.5	Achieve a "nice" video quality, perceived by the end user, with a video MOS>4.	End-user feedback consistently rated the video quality with a MOS greater than 4, reflecting satisfaction with the visual clarity and detail
UC 5.6	Achieve a "nice" audio quality, perceived by the end user with an audio MOS>4.	End-user feedback consistently rated the video quality with a MOS greater than 4, reflecting satisfaction with the visual clarity and detail.
UC 5.7	Ability to auto scale and deploy in under 60 seconds.	The system demonstrated exceptional scalability, with auto-scaling and deployment executed in under 60 seconds in 98% of cases. This rapid capability highlights the system's agility and readiness to adapt to sudden changes in demand without sacrificing performance.
UC5.8	Core performance: • CPU loads <50% • RAM loads <50% • Disk Loads <50% • Ingress/Egress Network loads < 50% of line card capacity • Frame packet drops < 3% Video flow capacity > 50 flows per CPU per virtual.	Core performance metrics were well within optimal ranges: CPU, RAM, and Disk loads were maintained below 50%, ensuring efficient resource usage. Network loads were kept under 50% of line card capacity, preventing any bottleneck. Frame packet drops were consistently below 3%, facilitating a smooth data transmission. Video flow capacity exceeded 50 flows per CPU per virtual, showcasing the system's ability to handle high traffic volumes effectively.

Thought to be used anywhere at any time, the Mobitrust platform is now leveraging the microservice architecture, which allows the split of its internal components into multiple geographical locations. Hence, certain components can be instantiated near the scenarios where operations are taking place. Such proximity will allow a decrease in latency for communications and also the reliability of the whole system by being tolerant to failures in backhaul connectivity, which are known to occur in certain catastrophe scenarios.

Experiment Setup/Methodology/Deployment

RZ collected the set-up time of new servers after 1000 runs. This data collection is part of a comprehensive effort to evaluate and optimize our system's performance across several critical parameters. The choice of these specific parameters is rooted in their direct impact on the efficiency and reliability of server deployment, which is vital for maintaining high-quality service delivery in dynamic environments.

KPI 5.7 specifies that autoscale and deployment need to be completed in less than a minute. This is crucial for scenarios that require rapid scaling and deployment to handle varying loads efficiently. However, our evaluation was not limited to KPI 5.7. We also measured other key performance indicators to ensure a comprehensive assessment of the system's capabilities.

Parameters monitored during server set-up were:

- **Start Timestamp:** Indicates when the test application begins its operation.
- **Registered Timestamp:** Marks the moment when the test application receives an acknowledgement from the orchestrator.
- **Available Timestamp:** The point at which the video router becomes responsive to commands.
- **Provision Workflow:** Describes the process of the orchestrator contacting Kubernetes and Domain Name System (DNS) for setting up the video router.
  Steps:
    PROVISION – Starting provision (1): The initial step of initialization of the provision workflow.

PROVISION - Added Server Definition (2): Logs on the orchestrator side documenting the addition of the new VNF to the system.
PROVISION - Created Pod and Service (3): Logs confirming the successful receipt of an 'OK' from Kubernetes for the VNF.
PROVISION - Added DNS Entry (4): Final step in the provisioning process, involves registering the network address of the new VNF.

- **Register Workflow:** Involves the orchestrator configuring the newly set-up video router.
  Steps:
  REGISTER - VNF Requested Registration (1): VNF side log indicating to the orchestrator that it is operational.
  REGISTER - Got Service Definitions (2): Logs documenting the VNF's receipt of configuration data from the orchestrator.
  REGISTER - Loaded Certificates (3): Pertains to the VNF loading HTTPS certificates required for secure operations.
  REGISTER - Sent Configurations to Service (4): Final configuration adjustments reported to the orchestrator, completing the setup.

The results we looked for were:

- **Response Time:** This is the duration between the start and registered timestamps, which shows how quickly the system responds to initialization commands.
- **Setup Time:** The time from the start timestamp until the video router becomes available. This provides insights into how long it takes to fully set up and make the video router operational.
- **Workflow Efficiency:** Comparing the provision and registration workflows' durations to assess efficiency and identify any potential bottlenecks.

The deployment was carried out using Kubernetes for orchestration, with necessary DNS configurations for network management. The system required quick and efficient setup capabilities to meet the stringent KPI of deployment within 60 seconds. This was achieved in 98% of the cases, demonstrating the system's robust deployment mechanism.

Experiment Execution and Results

We obtained the following results for the three different parts (external provider, BlueEye Cloud and Altice Labs Cloud) of the setup process:Table 2: Different response time in the set-up process

Process	Average time (ms)
user server response time	6680,44
Provision workflow	
1→2[1]	9,25
2→3	882,20
3→4	5183,42
Register workflow	
1→2	8,24
2→3	1
3→4	8,71

[1] 1→2 means from process 1 to 2, inside the provision workflow.

The processes are as follow:

- Provision workflow:
  1. Start Provision.
  2. Add Server definition.
  3. Create pod and service.
  4. Add DNS entry.
- Register Workflow:
  1. VNF Requested registration.
  2. Got service definitions.
  3. Load certificates.
  4. Sent configuration to service.

The *user server response time* suggests a moderate response time when considering server communications. It indicates the typical time taken for a server to process a request. The *Register* workflow is consistently faster and shows less variability compared to the *Provision* workflow. This suggests that once the initial provisioning of resources is completed, subsequent configurations and registrations are handled with high efficiency. In both workflows, the presence of outliers (especially in the *Provision* workflow) suggests that while typical operations are efficient, occasional delays or issues can significantly impact performance. This could point to specific areas where resource allocation, or process optimization, could be further improved. Figure 1, Figure 2 and

Figure 5 present the graphs that depict the three processes.

As a general outcome, we can conclude the system benefits from private DNS with "slice" to reduce time to availability and enhance privacy and security. Among the 100 runs, only two were failures. We can also conclude that the system has a high reliability and success rate.

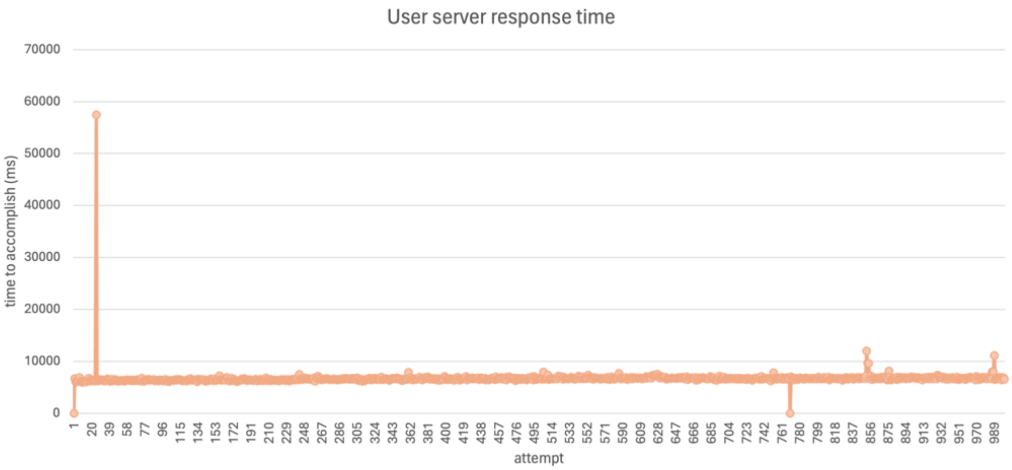

Figure 1: User server response time

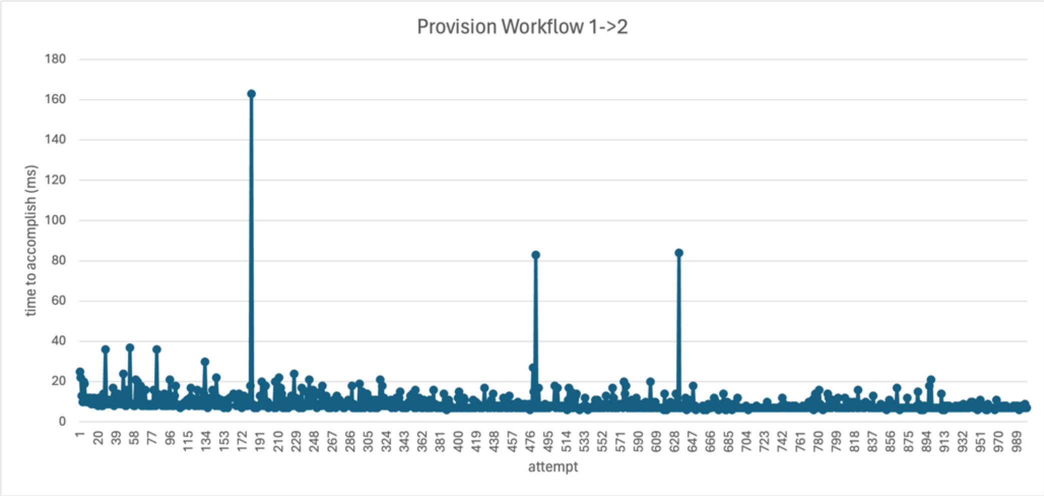

Figure 2: Provision workflow first step

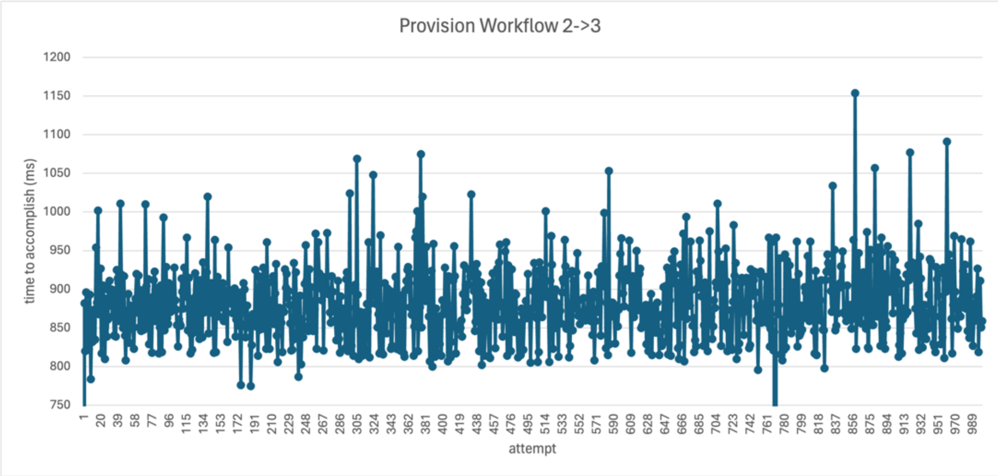

Figure 3: Provision flow second step

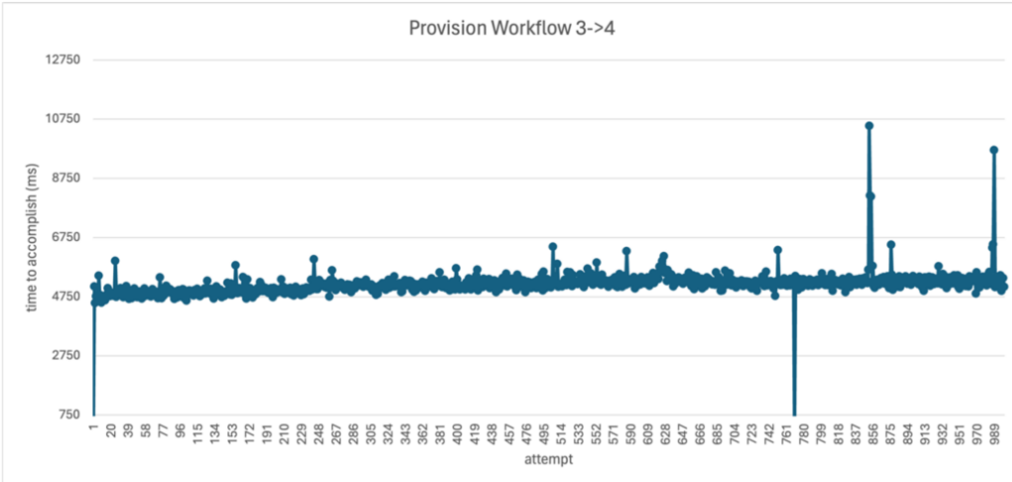

Figure 4: Provision flow third step

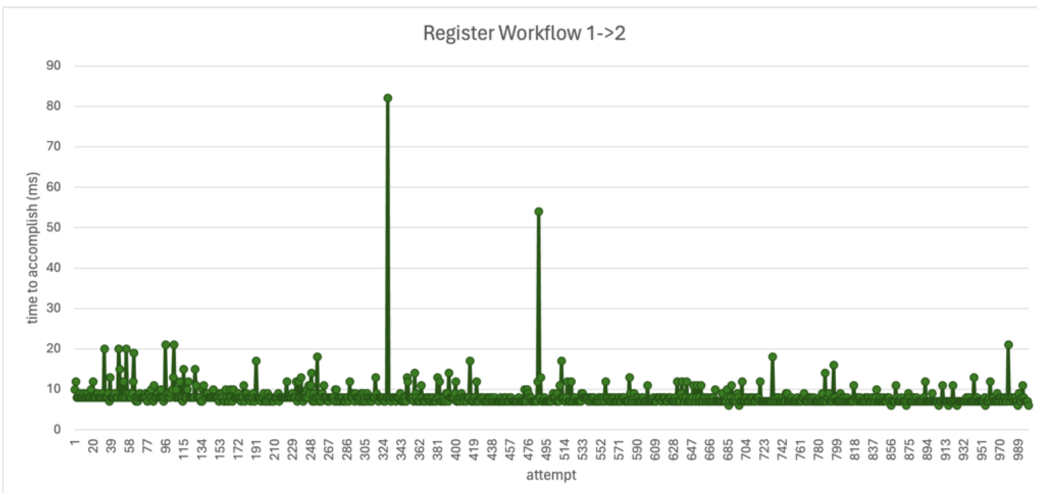

Figure 5: Register workflow first step

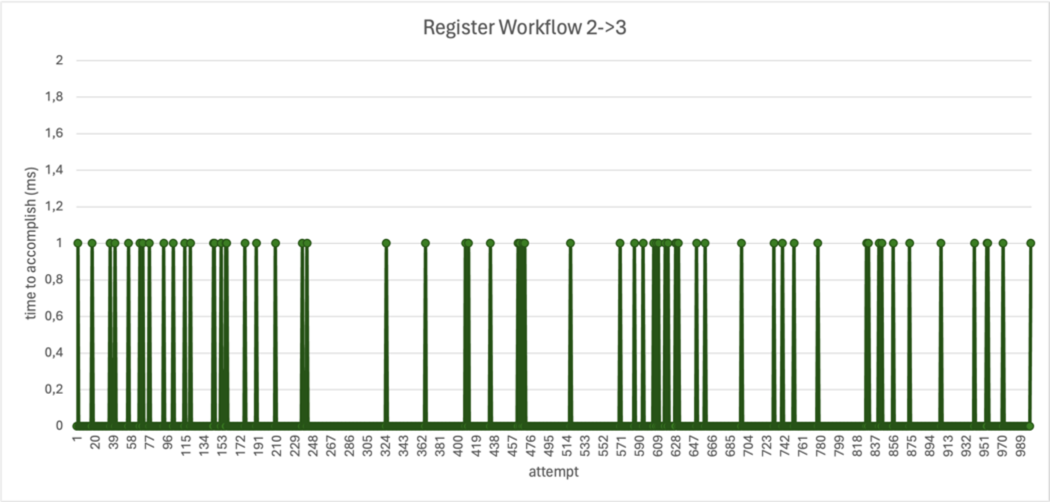

Figure 6: Register flow second step

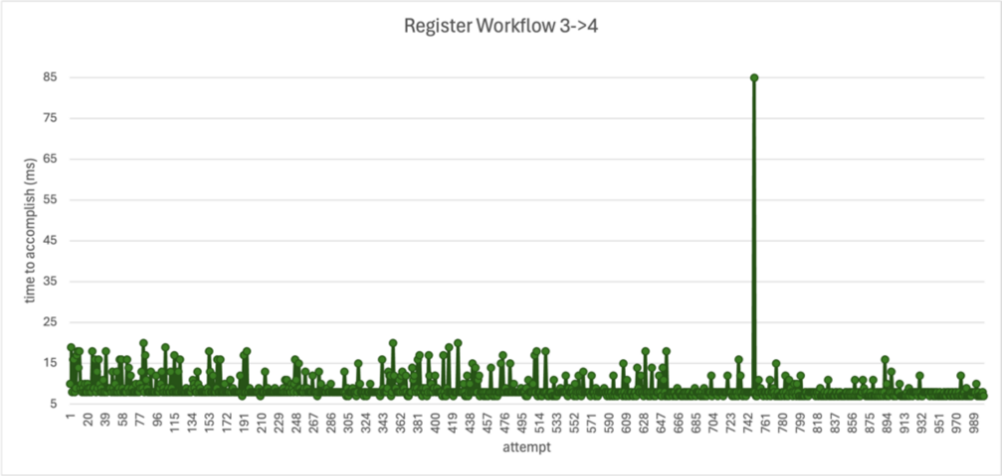

Figure 7: Register flow third step

In Table 3, the results concerning the KPIs set can be found.

Table 3 : KPI results

KPI	Results
UC 5.1: Video Quality	Achieved 1280x720 resolution, surpassing the 640x480 baseline.
UC 5.2: Packet loss	Maintained below 5%, ensuring stable video feed.
UC 5.3: Audio MOS	Maintained above 4, indicating high clarity.
UC 5.4;5.5;5.6: Video and Audio MOS	Consistently above 4, reflecting end-user satisfaction.
UC 5.7: Audio scaling and deployment	Achieved under 60 seconds in 98% of cases.

UC 5.8: Core performance	CPU, RAM, Disk, and Network loads maintained below 50%, with frame packet drops below 3%.

Conclusions

The numerical results indicate that the system performs exceptionally well within the defined KPIs. The video and audio quality met or exceeded expectations, providing clear and stable streams. The auto-scaling and deployment times demonstrated the system's agility. Core performance metrics were maintained within optimal ranges, ensuring efficient resource usage and smooth data transmission. The results underscore the significant impact of 5G networks on enhancing application performance, particularly in Public Protection and Disaster Relief (PPDR) scenarios, highlighting improved reliability, scalability, and quality of service.

ALB platform conclusion

The ALB testbed used in 5G-EPICENTRE experimentation activities is an evolution of the 5G experimental platform built and expanded in the framework of previous internal and collaborative 5G-related projects, targeting multiple 5G use cases and vertical areas such as energy and transportation. In the context of 5G-EPICENTRE, the platform was integrated in the 5G-EPICENTRE ecosystem and evolved to fulfil the requirements of the supported PPDR use cases. Overall, the results are in line with the expectations and demonstrate the value of 5G technologies to support both use cases, in relevant PPDR application scenarios. In particular, throughput and latency performance enabled the stable execution of the use cases without significant issues. During the time frame of 5G-EPICENTRE, the available bandwidth for ALB experimental activities was decreased from 100 to 20 MHz, as a result of the beginning of 5G commercial exploitation. However, this did not represent a limiting factor against the accomplishment of relevant KPIs, particularly throughput, latency and reliability.

ALB testbed follows the Open RAN architecture, which means that the gNB is disaggregated into 3 separate components: the Radio Unit, the Distributed Unit and the Centralized Unit, which can be deployed separately in multiple combinations. In Altice Labs testbed, different solutions have been tested and deployed for the Radio Unit component, as well as antenna, and the variations of the performance results have been considerable, particularly in outdoor environments.

UC5 was deployed and evaluated on Altice Labs testbed infrastructure and the results are globally in line with the expectations. The main issue was the interoperability between Samsung A52/A53 phones and SA 5G core, which was overcome with an alternative terminal equipment (Samsung S23+). Again, high throughput and low latency enabled by 5G were key to accomplish the expected performance required by the use case.