Growing enterprise revenues

Kindly supported by:

Executive Briefing

WHAT EDGE DEVELOPERS WANT FROM TELCOS NOW

The use of edge computing is accelerating rapidly in applications such as live streaming, drone management, and AR/VR. Developers, the key catalyst of growth, need infrastructure fast and certainty on pricing. Telcos can play here but must act now or lose out. We illustrate how and why with seven real-life developer case studies and a guide to the driving factors for edge adoption.

Tilly Gilbert, Consultant | tilly.gilbert@stlpartners.com | October 2019
Dalia Adib, Senior Consultant dalia.adib@stlpartners.com

Executive Summary

Edge computing is at a turning point: from hypothetical use cases to real applications

- Industry discussions are now centring around the "how" and the "when" rather than the "if" and the "why" when it comes to edge computing. They are moving on from exploring use cases at a hypothetical level to seeing the results from developers experimenting with (telecom) edge computing in hackathons, testing environments (edge nodes in networks) and incorporating edge computing into their application.

- Although developers acknowledge the benefits edge computing can bring to their applications, there are still challenges they face in leveraging edge. They need clarity on how they will access the edge and where, which capabilities will be available to them, how workloads can be managed at the edge and how they will be charged for its use.

- In order to drive these conversations forward and ensure adoption, telcos need to listen and learn from these developers as they are likely to be key customers and partners.

- We interviewed seven start-ups from four key domains about their experience exploring edge computing, detailed in the below list. Each application developer was at a different stage of exploring edge computing and had been working with different types of edges. Our focus was primarily on telecom edge (compute capabilities on a telco's wide area network) and on-premises edge (servers running at the customer site.)

Company	Description	Domain	Experience and type of edge used
1000 Realities	Software studio for VR and AR applications across consumer and enterprise	AR/VR	Design, build and testing (telecom edge)
Arvizio	Enterprise mixed reality for collaboration, design and presentation	AR/VR	Commercial offering (on-premise and telecom edge)
Atrius	Infrastructure to enable autonomous drones and real-time data analysis	Drones	Proof-of-concept (telecom edge)
Holo-Light	Industrial augmented reality solutions for enterprises	AR/VR	Proof-of-concept (telecom edge)

Radius Networks	Real-time location-based services for retail and hospitality industries	Location-based services	Scaled offering (on-premise edge)
Section	Content delivery platform running containerised workloads across locations	Video / app optimisation	Commercial offering (cloud / CDN edge) Ecosystem engagement (telecom edge)
Smart Mobile Labs	Video optimisation and delivery solutions for live streaming	Video / app optimisation	Commercial offering (telecom edge)

Application developers have different drivers for adopting edge computing, but latency and bandwidth dominate

- While edge computing can deliver several advantages to running applications in the cloud, reducing latency and the amount of data travelling back to the centralised cloud were the two key benefits application developers valued the most.

- For most, this was because edge computing enables them to strike a balance between the performance they need and the ability to scale their solution across customers. For example, in AR/VR applications, telecom edge computing reduces latency between the server rendering the image remotely (at the edge) and the headset to create an enhanced user experience. Plus, by avoiding the need to install dedicated servers at each user's site, the developer can scale its application more easily.

	Retail & Wholesale – Location Based Services	Augmented Reality / Virtual Reality	Video/Media (Upstream & Downstream)	Drones: Unmanned Traffic Management	Drones: Data Processing & Analytics
Why edge?	Local real-time analysis of location data	Offload processing power and maintain real-time interaction	Live streaming and bursting (high capacity situations)	Autonomous navigation	Avoid cost of moving data to central cloud
Latency Critical Compute	■■	■■■	■■□	■■■	■□□
Heavy I/O	■■■	■■■	■■□	■■□	■■■
Geo-Spatial Knowledge	□□□	□□□	□□□	■■□	□□□
Hyper-Local Grouping	□□□	□□□	■□□	■■■	■□□
Data Residency	■■□	■■□	□□□	■■□	■■□

- For several developers, concerns over data sovereignty and a desire to be "compliant by design" also make edge computing an attractive prospect. In location-based services, only anonymised and aggregated data sets would be sent to the cloud for data analysis while customer location information would always remain on-premise. In AR/VR, edge computing can ensure that proprietary information such as CAD [1]models of buildings are stored on a device (which could be lost or stolen) but instead at a secure edge server.

But each domain is different – telcos must engage with developers to understand and address their needs

- In order for telcos to monetise the edge computing opportunity, they will need to identify areas where they can bring value. This stems from addressing challenges developers face currently and providing solutions to these problems, as well as enabling access to edge infrastructure in a way that is easy and scalable for developers across different domains.

- **AR/VR applications:** AR/VR application developers highlighted the need for telcos to deliver more than just standard hardware at the edge. Their key concern was the availability of GPUs (for image processing) and persistent, fast SSD storage (to create local caches of very large 3D models).

- **Drones:** Those looking to develop solutions for autonomous drones emphasised the network requirements that are needed in order to deliver sub-10ms latency targets; an obvious area where telcos can provide specific network services. They also suggested that telcos may be valuable partners to try and accelerate adoption by working with local governments to shape the regulatory environment.

- **Application and video optimisation:** For application and video optimisation solutions, using standard APIs to interface with and the ability to flexibly spin up and spin down workloads depending on demand are crucial. They too wanted specialised hardware including GPUs and guaranteed speed of processing at the edge to ensure that application latency, as well as network latency, is minimised.

- **Location-based services:** When the telecom edge is used for location-based services, ubiquity of service is crucial. This means that regardless of the network provider a consumer is with, the location information must be available in all scenarios.

The role of the telco: commercialisation models for edge computing

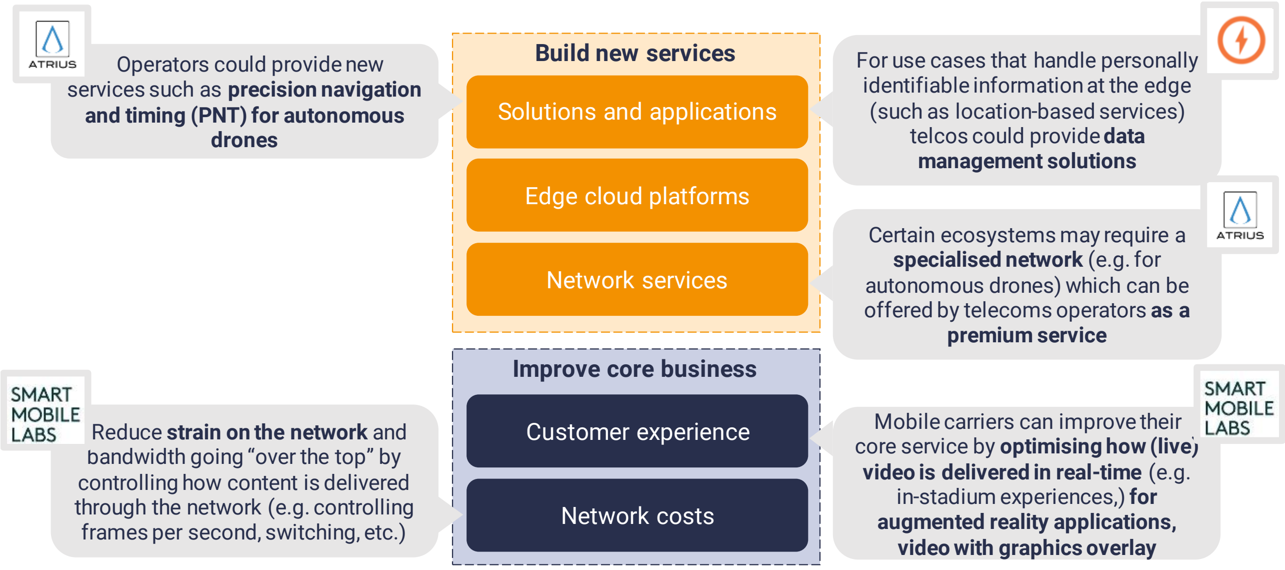
- The key message from developers to telcos was to stop waiting for a myriad of edge computing use cases to be proved before rolling out the infrastructure. Application developers need more sites in order to test their solutions and commercialise them for their users, wherever they are in the world.

---

[1] Computer-aided design

- Not only this, but telcos must be clear about how they choose to charge developers for the edge – this is partly about the absolute amount that they will be charged, but also about having a level of certainty as they develop applications. Without knowing this, they are unable to make a cost/benefit assessment to determine when edge should be used.

- Telcos can use edge computing both to improve their core connectivity business, e.g. reducing latency for consumers when they stream video over their network, and to build new services, e.g. data management solutions at the edge that ensure data sovereignty and sharing requirements are met and enable data analysis. The below diagram summarises some of the key opportunities for telcos, explored further in the final section of this report, **Monetisation opportunities for telcos**.



This document was researched and written independently by STL Partners, supported by MobiledgeX. STL's conclusions are entirely independent and built on ongoing research into the future of telecoms.

Table of Contents

Table of Figures

Introduction

STL Partners has been writing about edge computing since 2015. We've published reports including Edge computing: Five viable telco business models and Telco edge computing: Turning vision into practice. Although this is relatively nascent in the telecoms industry, the domain is maturing rapidly. Discussions are now centring around the "how" and the "when" rather than the "if" and the "why".

In order to drive these conversations forward, telcos need to listen and learn from developers who will, eventually, be making use of their edge computing capabilities. There are developers who are deeply engaged with the issue of edge computing, seeing it as a game-changing capability for their own solution. But, they also have strong messages they want the telecommunications industry to hear. They have their own requirements and expectations for how edge computing should work. They want clarity around what capabilities it will have, how their application will work on the edge and how they will be charged for its usage. **This paper looks to give several application developers at the forefront of edge computing development a platform to address the telecoms industry.**

For our interview programme we have focused on four key industries:

- AR/VR applications
- Drones
- Location based services
- Video and application optimisation.

The focus for this paper is on application developers who primarily serve enterprise markets. However, there is real opportunity and applicability for applications running at the edge in the consumer market as well. In particular, some of the AR/VR applications discussed are currently industry focused but could and will eventually be used by consumers as well.

Our hope with this paper is that it will stimulate discussions within the edge computing community as a whole, including all key stakeholders. We also pull out the key practical implications for telcos in terms of business models, the technology they should look to be developing and the partnerships they may wish to establish.

AR/VR for industry

Application introduction (AR/VR for industry)

The promise of industry 4.0 is being discussed broadly, and has been for several years. Much of the promise of increased productivity and reduced waste comes from the automation of processes that have typically required routine, often physical, human intervention. STL Partners has evaluated some of these use cases at length, as well as forecasting the value they can bring to the industry, in an upcoming report focused on the manufacturing industry.

However, there is also much promise in applications that, rather than replacing humans, look to increase their safety, efficiency and productivity. And this kind of use case can span outside of manufacturing, into industries such as mining, utilities, construction, architecture and beyond. One of these use cases is using AR/VR/MR[2] technology to overlay information for workers. This can span from simpler applications such as improving people management through applications that provide information on the order of tasks that should be performed to more complex applications like using augmented reality to visualise 3D CAD models. Benefits of these kinds of solutions include:

1. **Increased productivity of workers.** For example, instead of needing to refer to manuals or instructions before returning to the task at hand, instructions can be overlaid on smart glasses so they can be referred to as the task is being completed. This will save time and increase the likelihood of the task being completed correctly first time. In another example, information about what task to complete next can save time through maximising the number of jobs a worker can do before they need to return to a base to collect new materials.

2. **Increase productivity of experts.** AR/VR applications can optimise the time of workers out in the field. But this resource is often not the most expensive. These applications also have the ability to save on the time and expense of experts and trainers, who can hold up production or progress on a project for several weeks because of stretched availabilities and the need to travel to site from global locations. VR/AR applications can essentially upskill cheaper labour either through the additional information they can receive through the application or through the ability to more closely collaborate with experts who are not physically in the same place as them. This concept is explored in greater detail later (see Arvizio: edge for dynamic collaboration between remote parties) but can be envisioned as either:

- The ability for a remote expert to see from the point of viewer of the worker in the field what it is they are looking at and provide real-time annotations on their device or instructions via headphones

[2]Augmented reality (AR) consists of digital elements being displayed over a live view of the physical, real world. Virtual reality (VR) consists of complete immersion into a digital world in which there is no view of the physical world. Mixed reality (MR) combines elements of both AR and VR where digital and real-world objects can interact by making digital objects trackable and interactable in the user's real-world environment.

- Or, the ability for two teams in different locations to both view simultaneously a 3D model from individual points of view in order to collaborate on modifying the design.

3. **Tasks performed with more accuracy.** The knock-on effect of both of the above efficiency savings feeds through to various other industry benefits. If workers can be upskilled through the use of overlaid information, then they are less likely to need to redo tasks because mistakes have been made. This feeds into reduced costs on elements like training (because workers can be guided through more challenging task with the support of experienced experts).

4. **Better health, safety and compliance.** Overlays on the smart glasses can warn workers of hazards and enable them to more safely handle challenging situations. Where video is stored, compliance to health and safety standards can be proven.

These use cases can occur on a variety of end devices. Most current applications work on a combination of smartphone and smart glasses (the majority optimised for Microsoft's HoloLens). In the case of a smartphone, the user must hold it out in front of them for the images to be overlaid – it cannot easily be used for virtual reality applications because the real physical world is not blocked out. It also limits the ability for the application to be used hands-free, meaning complex tasks that require two hands cannot be done while using applications on smart phones. The benefit of using smartphones, however, comes in their ubiquity. As BYOD (bring your own device) becomes increasingly common in the workplace, using the application on a smartphone removes the need for expensive dedicated hardware entirely. For the optimal experience, however, the application developers we spoke to all believed that eventually smart glasses will be the key end device.

Some AR/VR applications are beginning to mature in an already fast-growing market

There are already many existing applications in this space. Several are already commercially successful, such as Fieldbit who are working with Israel's national water supplier Mekorot to improve their first-time fix rates.[3] In many cases, the application is more mature when using smartphones as the end device, despite ambitions to move further into the smart glass space.

For this research, we interviewed three start-ups in particular.

- **1000 Realities** is a Poland-based start-up focused on building a platform for other developers to create AR and VR experiences. Their focus has been on applications for industries including manufacturing, logistics and construction, but they are moving to become industry agnostic.

- **Holo-Light** is a German company founded in 2016 which looks to provide industrial augmented reality solutions with a focus on engineering, manufacturing and automotive verticals.

- **Arvizio** provides enterprise software solutions for 3D visualisation, including multi-user shared experiences, in augmented and mixed reality. They focus on architecture, engineering and

[3] For more information: https://www.fieldbit.net/mekorot-success-story-guaranteeing-the-safe-and-reliable-delivery-of-drinking-water/

construction verticals. Arvizio have run pilots with several major US telcos and their hybrid rendering solution is commercially available.

Key difficulties that AR/VR application developers are grappling with include latency and device limitations

While they faced unique challenges too, there were several key points all three companies we spoke to highlighted as challenges or pain points that were preventing them from developing their solution.

- **Stringent latency demands.** The CEO of Arvizio stated that "currently it is often only possible to achieve a low motion to photon latency" which means it takes too long when you move your head for the new scene to be rendered. Solutions hosted in the cloud can often only provide below-optimal solutions where users experience short delays in images being rendered and stream back to their device.

- **Device limitations.** For most applications, the optimal experience would be via a smart glass so that the user can be hands-free to complete tasks while using the application. However, smart glasses today both lack an LTE-enabled SIM card and do not have the necessary compute power on the device to create a high-resolution experience for the user.

- **Slow speed of integration into existing enterprise solutions.** In order to implement applications, application developers today (or their system integrator partners) often have to spend significant time integrating with existing enterprises' Wi-Fi systems and deploying servers on client premises.

> *"For a start-up, it's mandatory to offer something scalable. And going to set up on the client premise every time takes a lot of resources".*
>
> **Arvizio**

Several of these limitations could be solved through the use of edge computing technology

Edge computing is one way in which application developers may be able to overcome some of these challenges. For example:

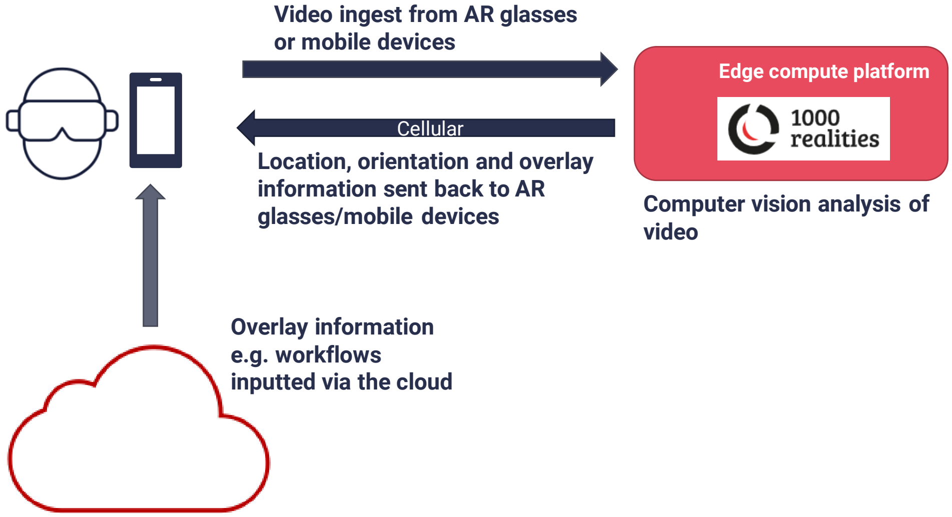
- Edge computing can improve latency through processing information in a location physically closer to the end-user. This will enable mobile use cases (either with LTE or eventually 5G) which previously can only be achieved with wired connectivity.

- Edge computing can enable complex solutions on dumber devices by offloading compute to the edge node.

- Edge computing can improve data security by ensuring that proprietary information remains in-country (rather than being streamed to a cloud data centre anywhere in the world).

1000 Realities: Edge computing for remote AR assistance

Figure 1: Use case diagram of how edge computing can enable 1000 Realities' solution



Source: STL Partners, interview with 1000 Realities

Edge computing will enable the computer vision analysis of the video (where low quality video is analysed and the extracted location and orientation information enables the correct information to be overlaid) to occur at an edge compute location rather than on the device or in the cloud. Edge computing helps to balance out two competing requirements that 1000 Realities are juggling. The first is they require very low latency which could be achieved through running the workloads on the device, and the second being that device limitations and the need for flexibility means doing compute on the device is less than ideal.

1000 Realities have identified 33ms latency as the upper limit of what can be tolerated for their solution to provide a good user experience. Offloading the computer vision analysis means they are reliant on a very fast network to ensure the user does not experience any lag or disorientation. They stated that "LTE with edge computing delivers 20-25ms latency and that if you combine it with private LTE or a campus network it can be as fast as 15ms". But on the other hand, 1000 Realities also highlighted that what customers really want, when it comes to AR services, is being able to experience it on any hardware. To provide this flexibility, this makes putting the compute on the device very difficult. On top of this, there are other undesirable side effects to having the compute on the device:

- **Heavier devices.** Increased compute power needed requires heavier and larger devices. This becomes particularly important when the end device is a smart glass that workers may be expected to wear for extended time periods. Heavy and unwieldy devices may become uncomfortable and used less frequently than their potential.
- **Shorter battery life.** Compute intensive tasks like rendering models or computer vision to analyse user's orientation will drain the battery life of end devices. This may result in the solution

being used less often in order to preserve battery or, in some cases, customers may have 8-10 hours of continuous use as a requirement that, if it cannot be met, will prevent them from using the solution at all.

- **More expensive devices.** Devices that have the needed specialised hardware for computer processing will be markedly more expensive than more generic smart glasses or smartphones. Particularly in industry situations, where workers may be in rugged environments, the cost of replacing devices when they break, are stolen, or mislaid, could become prohibitively expensive, particularly if you wish the vast majority of your workforce to be able to use the devices.

- **Confidential information must be stored on the device.** If all the compute is happening on the device, the information which is being overlaid on the video (such as step-by-step instructions of how to perform maintenance checks on a specific machine) must also be stored on the device. If this is then misplaced, lost or stolen companies are risking potentially proprietary information from being compromised.

Figure 2: Latency and data security were the two most critical factors that edge compute can deliver for 1000 Realities

Edge factor	Level of importance to use case
Latency Critical Compute The necessity that the delay between communications over a network is minimal	■■■
Heavy I/O The transfer of significant amounts of data between devices or networks	■□□
Geo-Spatial Knowledge The dependency of the device, application, content, or industry to maintain geo-location or network information in real time	□□□
Hyper-Local Grouping The dependency that the device, application, content, or industry is able to coordinate sets of geo-location and/or network data	□□□
Data Residency The benefit arising out of the added security benefits of edge security	■■□

Source: STL Partners, interview with 1000 Realities

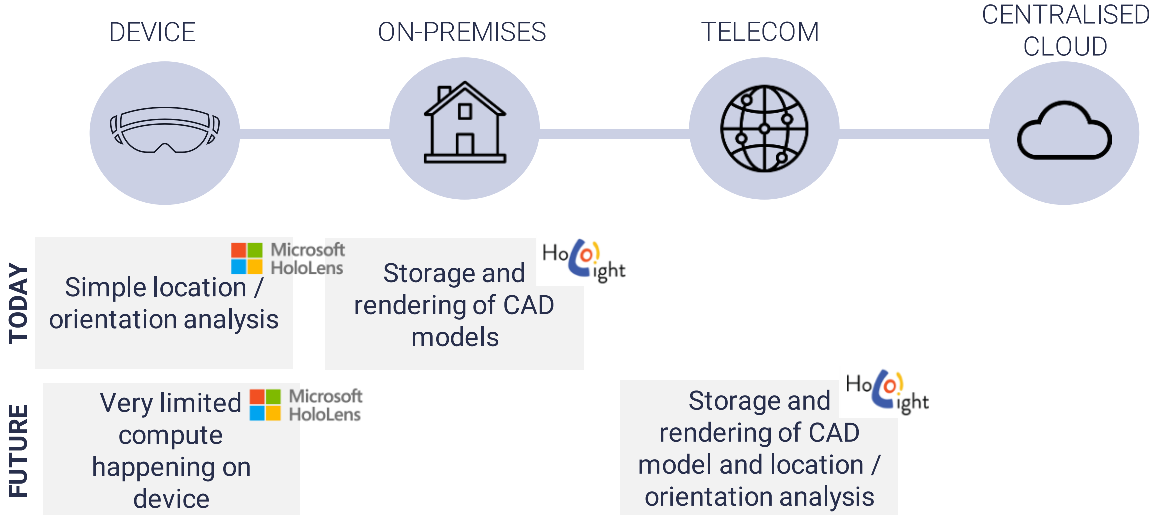
Holo-Light: edge for heavy duty computing with CAD models

Holo-Light had additional requirements when it comes to the amount of computer power needed to deliver their solution in real-time. This is primarily because their solution overlays more than just annotations or written information for the field force worker, but also extremely high definition 3D CAD models. The rendering process for such models is much more onerous – amplifying their need to offload some of this compute power to an edge computing location.

"AR/VR must run at 60 frames per second otherwise there's a kind of a lag and that feels strange, or can give people a headache. With cloud, you can reach 30-40 frames per second. With edge computing, you can reach 50-60 frames per second"

Holo-Light

Figure 3: With edge computing, workloads can be distributed across the compute spectrum

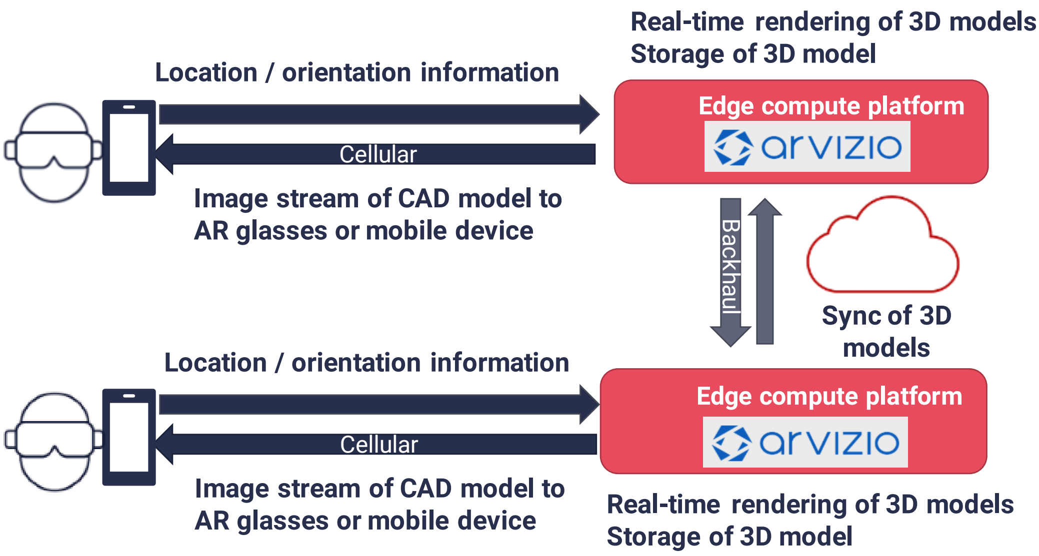


Source: STL Partners

Arvizio: edge for dynamic collaboration between remote parties

For the previous solutions, one party is using the VR/AR application. However, as the number of edge compute locations increases, these solutions should be able to also support situations where multiple parties in multiple locations wish to collaborate. This will enable real-time collaboration where, for example, two teams can visualise a 3D model of a building plan and discuss changes that need to happen, or could be use in a sales scenario to walk a client through how an end product will look. Figure 4 highlights how Arvizio state this will work.

Figure 4: When remote parties want to collaborate, edge computing can ensure there is no lag in the parties' experiences



Source: STL Partners, interview with Arvizio

Each party requires connectivity to their own local edge node where the rendering of the model occurs. The model itself will need to have been uploaded to the edge node before the meeting occurs to minimise loading time during the meeting. Orientation information is sent via the cloud to ensure both edge locations are synced and the users can accurately see where the other party is, while retaining their own unique perspective.

Challenges and implications for telcos

Speaking to developers in the AR/VR world, it is clear that they have specific technical requirements for edge compute locations. To start, all these solutions require GPU at the edge in order to process and render the models. As Arvizio stated "edge compute nodes can't just be bog standard off the shelf x86 servers, without additional hardware resources". Developers, based on how they have already architected their solution, also can have preferences on which operating system runs at the edge too.

Arvizio also have specific requirements around storage capabilities at the edge – "you need persistent, fast SSD storage – enough to build a local cache of the models". Telecoms operators will need to guarantee availability of both storage and GPUs without having dedicated resources for each application. Arvizio indicated that the way their solution was architected means they would need to run it in a virtual machine rather than a container. To complicate matters further, when a meeting or project has been completed, the applications developers require assurance that the model has then been removed from the edge cloud. This is important for matters of data security, particularly as the model may have proprietary information that the enterprise will want to know has been removed.

There are also challenges around hardware which must be resolved. Currently, smart glasses like Holo-lens and Magic Leap do not support mobile technology as they do not have SIM cards embedded.

This must happen before the eventual goal of using 5G and edge computing for non-wired, low latency, high bandwidth connectivity. There are also challenges that come if multiple parties are collaborating together on different hardware. Each device has a different way of doing commands such as zooming in or rotating an object. Some come with a controller, others use only hand gestures. Currently the processing of these commands happens locally on the device but to share that experience the user IO instructions need to be streamed to the edge node, propagated with other users and sent back. Applications will need to be able to handle different ways of doing this depending on the device used. This will allow everyone else to also zoom in to the same bit of the model as the user is, for example, discussing a detail in the design.

AR/VR application developers also highlighted the need for standardisation and automation of edge computing. They want reassurance that workloads will be orchestrated in a way that ensures that their latency requirements will never be missed and that storage will be available at the nearest edge nodes when they require it. Telcos also need to be able to provide simple and effective automated processes so they can run their applications at the edge. Manually building different scripts to interface with different operators' edge computing systems will not be sustainable. As Arvizio put it "we need to come up with standards that will define how you can as an end user application, like ours, request or demand resources in the edge compute servers based on our application needs".

UAV/drones

Commercial drones are struggling to achieve wide scale adoption

Forecasts for the drone market have been optimistic in predicting take-up of the technology across different industries. There are proven cases of how drones can deliver benefits across different sectors, for example:

- Delivering packages, such as Amazon's Prime Air
- Monitoring critical infrastructure, such as bridges and utility lines
- Surveying land and the condition of crop in agricultural settings

Outside of delivery, most drone use cases centre on the ability to capture data that has historically been costly, time consuming or dangerous to do so and make sense of it by creating meaningful maps or interpret the data to identify anomalies. For example, France-based start-up Donecle is enabling automated aircraft inspections through drones to improve efficiencies and reduce the time planes spend in the hangar. Software companies such as Pix4D, DroneDeploy and Bentley are the market leaders for providing photogrammetry tools to translate imagery from drones into practical models.

However, adoption is slower than expected. This is partly due to the nascency of the technology; most drones are limited to 30 minutes of flight time, which restricts the amount of data that can be collected in a single session. Regulation for commercial use is inhibiting use, by putting constraints on how large the drone is, when it can fly and how high, as well as mandating the need for pilot qualifications to fly drones.

Ultimately, the challenge is that, until there is a way to continuously collect data and monitor assets/infrastructure, industries and governments will not be able to access the true benefit of using drones. To make a real economic difference, drones must enable a significant volume of data that is not currently accessible. The current model relies on an individual to manually programme the drone to fly and collect the data, then connect it to a PC, to transfer the data and finally upload it to the photogrammetry software to extract insights. Atrius, a start-up we interviewed who is developing data centre units to enable autonomous drones, likened this to using a bucket to collect oil from an oil field and driving back to the refinery to process it into fuel rather than using a pipeline. Instead of using manual processes, data collection and transformation from drones needs to be autonomous – from the drone knowing when to set off and where to go, to interpreting the data and distributing it to the relevant recipients and systems.

> *"There's a lot of use cases where if you're able to deliver into the command centre a view of an accident scene, that's taking place on one of their interstates, you can save them in that one incident $20,000."*
>
> **Atrius**

Enter edge computing: enabling autonomous drones

In order to create this autonomous pipeline of information captured by drones, there are two things that need to happen:

1. **Autonomous navigation**: Drones should be able to conduct a flight without any human intervention, for example scanning the accident scene in the event of an emergency. This means the drone is aware of where it is in relation to other objects, within six centimetres, and anything that may affect its journey, including using real-time weather information.

   – Problem: The data that determines where the drone is in relation to other drones and the environment around it needs to be accessed in real-time and low latency is critical to eliminate any potential risks, therefore this cannot be done via the cloud.

2. **Real-time data analysis**: The rich images captured by drones, translating into petabytes of data at scale, will need to be turned into useful models and maps, particularly in situations where they are being used to identify anomalies. For example, in an emergency, they may be used to identify whether there is anyone still at the accident site. This requires the data to be analysed and processed immediately to take action, such as direct emergency services to rescue the injured people.

   – Problem: It is too costly to process the data in a remote data centre, given the amount of data that would need to be streamed which results in high bandwidth costs. Ultra-low latency (sub-50 milliseconds) is not necessary for this use case as the actions could still be triggered even if the data collected from the drone was analysed in a matter of seconds. However, as some of these use cases are for mission critical situations, e.g. emergency crises, the delivery of the data needs to be reliable.

Edge computing helps to overcome these challenges by bringing the compute closer to the drone, thus reducing latency. Rather than traffic traversing through the entirety of the network to get to the cloud, it would go to an edge node on the network for processing. This also reduces the cost of accessing data from the drones and allows fleets of drones to be coordinated. The additional benefit, particularly for the public sector, is that critical data recording the status of public infrastructure stays within the local municipality and avoids problems of data leaving a certain jurisdiction.

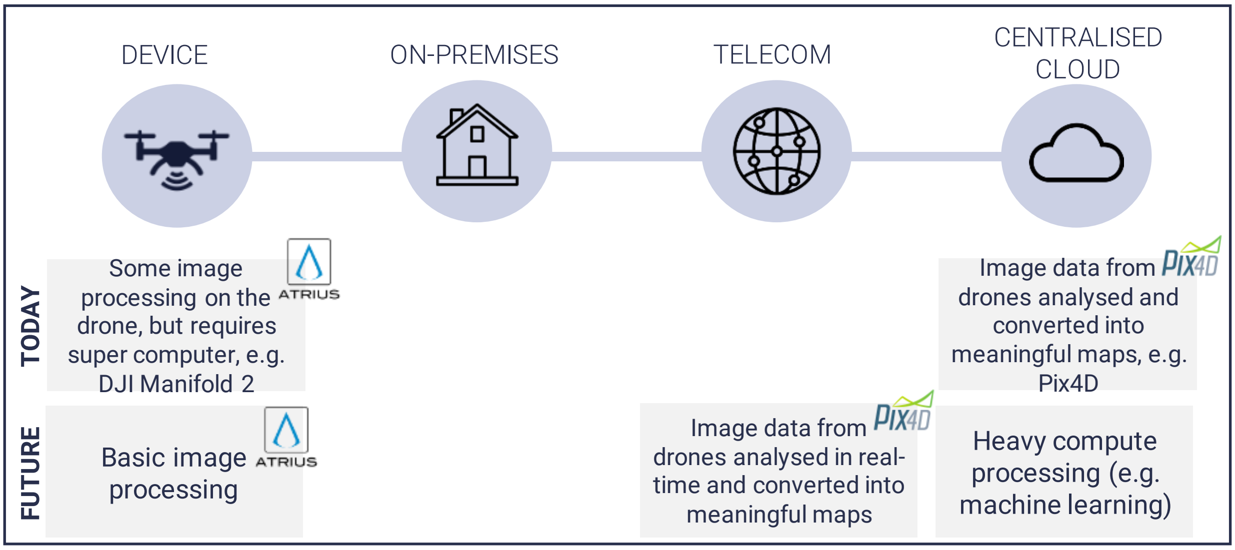
Figure 5: Latency is the most important factor for autonomous navigation, whereas heavy data load is needed for real-time data analysis

Edge factor	Level of importance to use case	
	Autonomous navigation	Real-time data analysis
Latency Critical Compute The necessity that the delay between communications over a network is minimal	■■■	■□□
Heavy I/O The transfer of significant amounts of data between devices or networks	■■□	■■■
Geo-Spatial Knowledge The dependency of the device, application, content, or industry to maintain geo-location or network information in real time	■■□	□□□
Hyper-Local Grouping The dependency that the device, application, content, or industry is able to coordinate sets of geo-location and/or network data	■■■	■□□
Data Residency The benefit arising out of the added security benefits of edge security	■■□	■■□

Source: STL Partners, interview with Atrius

The telecom edge is the most suitable edge in this case, as it provides local processing that is close enough to provide edge computing benefits, but also acts as an aggregation point to collate multiple data points and support drones' mobile nature.

Figure 6: Processing data from drones moving to the telecom edge



Source: STL Partners, interview with Atrius

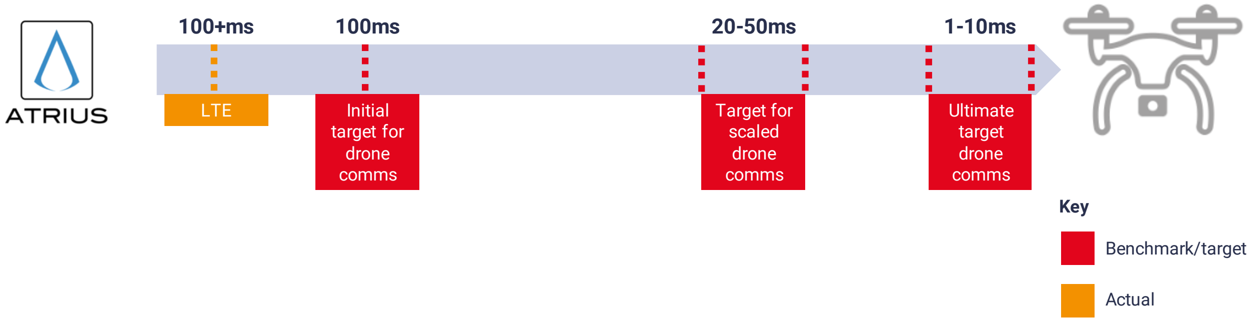
Atrius' experience: edge is necessary, and the network is key

As part of our research, we interviewed Atrius, a start-up developing hardware – nano data centre units and depots - for edge computing-based autonomous drone systems. They are specifically targeting inspection use cases, such as emergency response, traffic management and infrastructure monitoring (bridges, railways, etc.) Similar systems that require low latency sharing of data to enable mobility could be used for autonomous vehicles, although the current focus is on drones. They have been exploring edge computing through partnerships with Vapor IO in the U.S. to run proofs-of-concept on the network and demonstrate latency levels at a testing ground in Austin, Texas.

According to Atrius, application latency (the time it takes to process software requests) is below 5ms and most latency occurs in the network. The initial target was to achieve 100ms from drone to RAN, however even this is difficult to guarantee without "edgier" computing using current (LTE) network architectures. The tests in Austin showed latency is approximately 200ms, as most of the mobile operators hand off at Dallas or Houston and there are too many hops through the network to get to the cloud to process the data.

Beyond initial testing, the hope is to achieve 20-50ms latency in the field, between the drone and the RAN, to be able to scale autonomous drone communications. In addition, any information from the cloud that is required to ensure safe navigation of drones should be delivered in under 50ms for it to feed into the navigation process. Eventually, to stream video from the drones in real-time and augment information on top of these feeds, RAN latency would need to be at 1-10ms.

Figure 7: Drone-RAN latency levels can only be met with edge computing



Source: STL Partners, interview with Atrius

The challenge with this is not only achieving the stringent latency targets but guaranteeing consistent levels of ultra-low latency to persuade regulatory bodies, such as the Federal Aviation Administration[4] in the U.S., that the system is safe and secure. Given this, Atrius emphasises a need for an isolated low latency RAN for drone navigation, which they are current exploring by experimenting with CBRS.

Challenges and implications for operators

The maturity of autonomous drone navigation and real-time data processing is nascent - we are unlikely to have scale deployments for at least five years. Atrius believes that we will start to see requirements for autonomous mobility corridors in 2020, at least in the U.S., although at this stage it would be limited to small groups of cell towers. By 2021, drones could start to be deployed in a larger region.

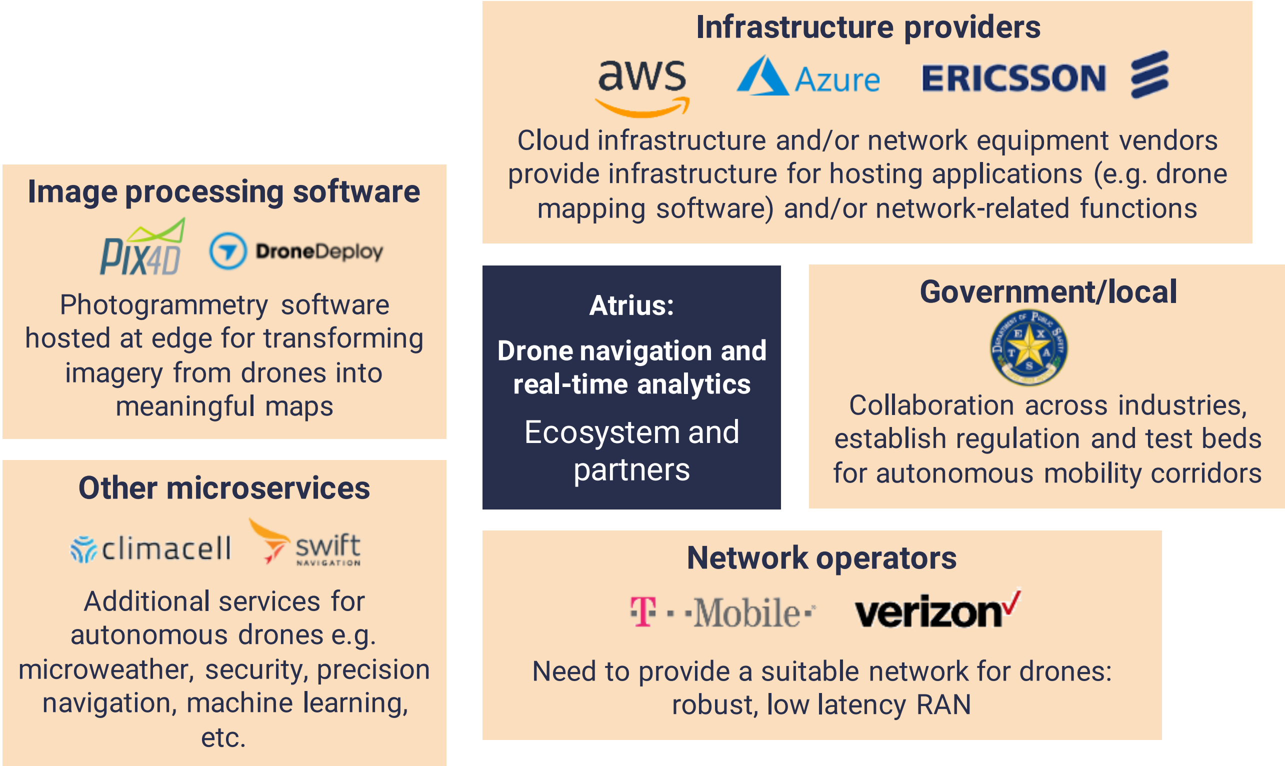
However, there are a number of challenges that must be resolved before such a system becomes available commercially. Not least are the network and low latency needs already discussed, which is where edge computing provides key benefits. The telecoms industry can play a leading role in providing a network that is able to meet the specific communications requirements, combined with infrastructure and capabilities for edge computing. Atrius did point out that this infrastructure needs to be universal and easily accessible, which has implications on how operators provide edge computing to developers and other industries. Data from drones and other sources (e.g. precision weather information) needs to be shared in real-time and should not be inhibited based on network siloes, i.e. whether the drone is on network A but closer operator B's edge computing facility at operator B's cell tower.

Outside of technology, there is a complex ecosystem that will work together to deliver these autonomous mobility corridors and allow data from drones to be used effectively. Atrius is focusing on public sector use cases, such as emergency management, environmental monitoring and public safety and are working with 100 city, state and federal agencies to develop appropriate solutions. The other objective behind working with these local authorities is to gain access to test bed environments and help move regulation forward to open the market. Given telecoms operators' local nature and

[4] FAA: Regulatory body in the USA responsible for unmanned vehicles (e.g. drones) and manned vehicles (i.e. aeroplanes)

existing relationships with public entities, they could play a role in accelerating engagement, bringing the right stakeholders together, as well as demonstrating the technology.

Figure 8: Ecosystem of stakeholders to enable autonomous drone systems



Source: STL Partners, interview with Atrius

Other players in the ecosystem include those who will help provide software services at the edge sites, on top of the edge compute infrastructure provided by Atrius and other hardware vendors. End software applications will include those required to enable autonomous navigation, such as Swift Navigation who provide positioning software for autonomous vehicles, or ClimaCell for micro-weather forecasting. Separately, image processing software such as Pix4D and DroneDeploy would transform image data coming from the drones into models. These applications may rest on (edge) cloud infrastructure platforms that could be provided by new companies or the traditional cloud providers, like AWS or Azure, particularly given their developments in the edge computing space whether it be through Greengrass (AWS) or Azure stack. Cloud providers and network equipment vendors are converging: NEPs[5] such as Ericsson, Nokia and Huawei are developing edge platforms for hosting applications and the cloud providers are targeting operators as customers of their infrastructure for network functions. These are summarised in Figure 8, although it is not an exhaustive list, for example there are other software providers who will support end-to-end solutions with other capabilities and microservices.

[5] Network equipment providers

Video and application optimisation

The changing nature of video and application optimisation

The way in which content, video and applications are optimised to improve performance, scalability and security has evolved. This is due to a number of reasons:

- **Application and web page content is increasingly personalised and dynamic** - caching static content at the edge is not sufficient.
- **Real-time video streaming is growing in entertainment, as well as enterprise/government applications** (e.g. police body cameras) – performance here cannot be improved by moving the content closer to the end-viewer, video has to be optimised as it is captured.
- **Content is being enriched with augmented reality** – for example overlaying live statistics on players when streaming a basketball game.

This is driving a need for edge computing and the ability to run workloads closer to the end user, rather than simply cache content or applications in a CDN. Two of our case studies come from this domain, although have very different propositions: the start-up Section provides a platform deploying workloads for developers at the edge and Smart Mobile Labs' solutions optimise real-time video streaming.

Section

Section provides a DevOps-centric, container-based edge compute platform that is hardware-agnostic and can run on any server across different "edges" to optimise performance, security, and scalability. Its core value proposition is to "run any workload, anywhere," which means it can deploy a workload for a customer in a point of presence on a public cloud, e.g. Google Cloud Platform, on a bare metal server, e.g. RackCorp in Australia, a telecom edge point, or even at a customer site.

> *"We're a software company, everything we do is about making it easy to stand up a Point of Presence very quickly."*
>
> **Section**

The types of workloads they offer to improve application performance, security and scalability include, but are not confined to:

- Dynamic content delivery
- Image optimisation
- Bot mitigation

- Language translation on the fly
- Ecommerce rendering

Smart Mobile Labs

Smart Mobile Labs is a video CDN start-up that uses new technology, edge computing and private LTE/5G to improve real-time, video-based applications: live video broadcasts, in-stadium / event experiences, cloud gaming and public safety networks. Its Edge Video Orchestration (EVO) technology enhances the way video data is streamed across networks to reduce latency to ultra-low levels of up to 1 millisecond by processing the video closer to the source.

> *""It's all about low latency. The edge computing servers need to have a perfect connection to the base station."*
>
> **Smart Mobile Labs**

Benefits of the telecom edge

Section sees edge computing as a continuum from device to centralised cloud – all of which are potential locations for its customers to place workloads. They are actively involved in the Kinetic Edge Alliance, a group of edge computing companies headed up by Vapor IO, and currently piloting deploying workloads at Vapor IO cell tower sites. The telecom edge is an area the company is exploring at the moment to evolve new capabilities and offer customers enhanced scalability. This would be relevant in areas where network capacity may be constrained and there are a large number of users (e.g. during events), therefore workloads need to be moved even deeper in the network to avoid increased levels of latency that degrades application performance. This could be done for short periods using on-demand location scheduling, which enables workloads to 'burst' to an optimal location that meets performance, scalability and cost requirements.

Smart Mobile Labs' core focus is on reducing latency for live video streaming and the driver for using the edge anchors on this. The original idea for the company stemmed from trying to resolve the problems occurring when watching Formula 1: spectators found it difficult to follow the cars as they moved quickly through the race track. Smart Mobile Labs sought to provide a mechanism for tracking the cars from the spectator's seat by providing a constant view of each car in real-time. The solution was to put video optimising capabilities at the telecom edge, closer to the camera that would be connected by fixed or wireless. They are already rolling this out to customers, namely event venues, to enable real-time streaming to the attendees' devices and have been working with MobiledgeX to test telecom edge use cases.

Figure 9: Benefits of edge computing according to Smart Mobile Labs

Edge factor	Level of importance
Latency Critical Compute The necessity that the delay between communications over a network is minimal	■■■
Heavy I/O The transfer of significant amounts of data between devices or networks	■■□
Geo-Spatial Knowledge The dependency of the device, application, content, or industry to maintain geo-location or network information in real time	□□□
Hyper-Local Grouping The dependency that the device, application, content, or industry is able to coordinate sets of geo-location and/or network data	■□□
Data Residency The benefit arising out of the added security benefits of edge security	□□□

Source: STL Partners, Interview with Smart Mobile Labs

Edge use cases in video / application optimisation

Both Section and Smart Mobile Labs are exploring a range of use cases across entertainment, security and web applications. We have selected two to detail below: live event experiences and gaming.

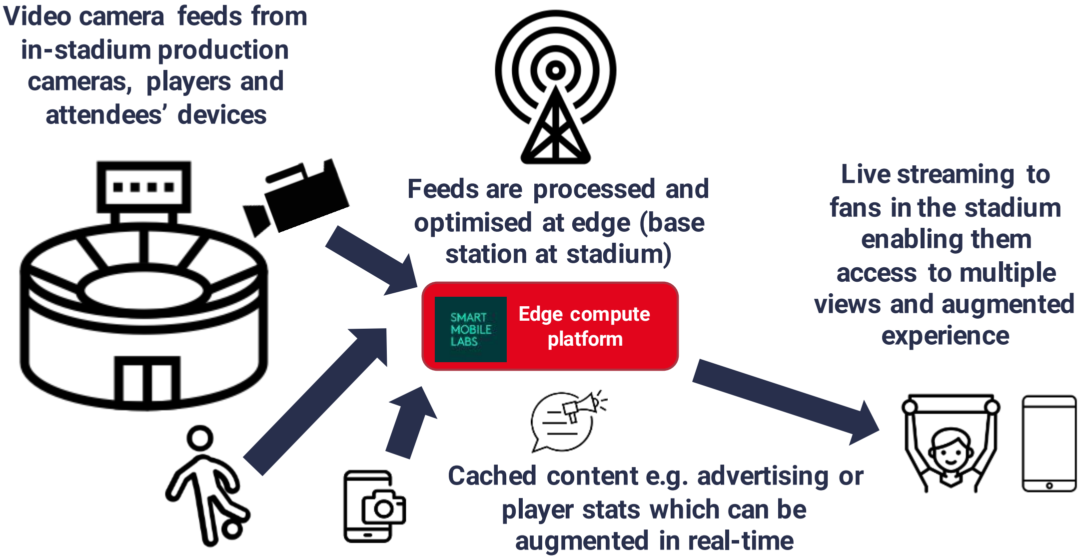
Live event experiences

Smart Mobile Labs is improving the in-stadium experience by using edge computing. As mentioned above, processing video at the edge and reducing latency allows the event venue to provide new experience for fans, such as seeing multiple views of the sport (for example) which would be accessed via an app on a smart phone.

Alternative streaming mechanisms, e.g. YouTube or traditional CDNs like Akamai, use different technology (HTTP as opposed to web RTC) and do not take advantage of the telecom edge. This results in it taking seconds to stream a file – even for YouTube's live streaming service, the minimum latency is five seconds. On the other hand, Smart Mobile Labs can achieve millisecond latency.

There are other potential services that can further enrich the core streaming application in the stadium or live event. Customised information, such as the players' statistics, can be augmented in real-time as the fan is watching their favourite play live. Advertising and other notifications can be cached at the edge and therefore overlaid in real-time.

Figure 10: Smart Mobile Labs using edge computing for live event experiences



Source: STL Partners, Interview with Smart Mobile Labs

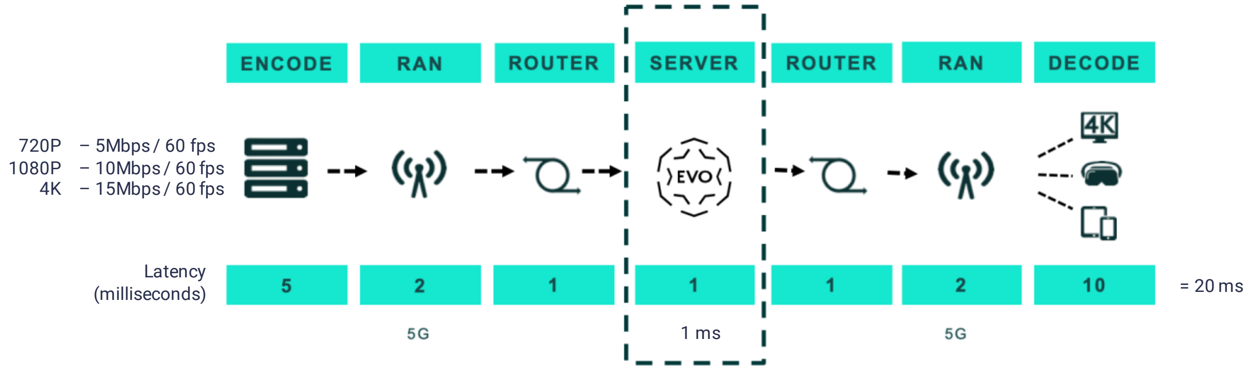
Gaming

Both Section and Smart Mobile Labs offer edge computing services to improve gaming by reducing latency. This is beneficial in a variety of sub-contexts within gaming:

- **Cloud gaming**: ensuring there is limited latency as gamers interact with a cloud-hosted game by placing custom gaming workloads closer to the end-user. This will become more mainstream in the next few years, as many companies launch cloud gaming offerings, following on from EA Games (Project Atlas), Google Stadia and Microsoft xCloud.
- **Multiplayer gaming**: improve latency between multiple gamers by hosting the server in a location that is optimal for those taking part in the game.
- **Augmented reality / immersive experiences:** as seen in the AR/VR section, edge computing can help reduce latency between the end-device and the cloud to allow for rendering off the device and make devices more lightweight.

Smart Mobile Labs has experimented with the impact of adaptive streaming for cloud gaming at the edge in a 5G test environment with Deutsche Telekom and proven the benefit on latency, having achieved 20 milliseconds end-to-end. The test used GPUs to keep processing latency to a minimum and used a MobiledgeX server that was situated next to the base station to significantly reduce network latency.

Figure 11: Smart Mobile Labs' results for cloud gaming latency in 5G test environment



Source: Smart Mobile Labs

Challenges and implications for operators

In terms of the challenges for telecom edge computing that Section and Smart Mobile Labs highlighted, there were three core areas:

Custom infrastructure and hardware requirements

Use cases have unique requirements that means a generic architecture for edge computing and the network to support it does not necessarily always apply. Smart Mobile Labs is able to reduce latency for live video streaming in stadiums by using the telecom edge and connecting cameras via the mobile network. However, one of the challenges with this is that it means the already strained network capacity at the stadium will now be used for high bandwidth video capture. To ensure there is sufficient capacity for this application, operators will need to provide a way to isolate this traffic.

The other complication is regarding hardware requirements at the edge nodes. Given that for many of these use cases latency is critical, the time it takes to process the application at the server must be minimised, as well as the network latency. Use cases that require high processing power, including gaming and augmented reality, will need GPUs at the edge.

Developer interfaces and tools for edge infrastructure

Developers need to be able to build applications that can scale, therefore the first obstacle for them is being able to access (telecom) edges. There is a chicken-and-egg dilemma where, until developers are certain the infrastructure will exist when they expand to new customers, they are less willing to experiment with the edge. However, operators need evidence that there is demand and a monetisation model before they invest in this infrastructure.

Assuming edge capabilities are available, the other point brought up by interviewees was the need to ensure this is accessible in a way that was appropriate for the developer. This includes using standard interfaces, operating systems and APIs for the application providers to host their services on top. The tools that may be more difficult to provide include those for managing and orchestrating workloads across a distributed edge computing portfolio. Section emphasised that this is important, particularly if a customer wants to use the edge in a dynamic way, for a limited period of time. This will demand more sophisticated controls that can quickly spin up or down workloads depending on customer needs.

> *"How do you dynamically move and not have wasted capacity?"*
>
> **Section**

Pricing

Given the early stage we are in for edge computing, there is still no indication for how much the telecom edge will cost. In order to provide telecom edges as options on its edge computing platform, pricing is a key factor for Section. Otherwise, it is difficult to support its customers to make decisions on where workloads should run, as these are often based on price and performance. It will be interesting to see how price points reduce over time as telecom edges become less finite and more are deployed.

Location-based services

Location-based services leverage information about a user's location in order to provide targeted information, advertising or offers. Radius Networks provides these types of solutions for the retail and fast food industry. Specifically, they enable solutions such as:

- **Table service.** Often used in fast food restaurants, when a customer has ordered they are given a beacon and can go and sit at a table. Staff are able to track the customer and bring their food to them when it has been prepared.

- **Curbside pickup of groceries.** When a customer orders groceries in advance and drives to the store to pick it up, their location can be tracked in order for staff to be ready to hand them their order as soon as they arrive in the carpark. This ensures minimum wait time while also minimising the amount of time food is taken out of optimal storage conditions such as a fridge or a freezer.

- **Asset tracking.** Assets such as products or machinery can be tracked throughout a store. This can ensure expensive stock or items are not lost and can help with logistical difficulties such as locating a specific package or item in a large warehouse.

There are current technical limitations that come with location-based services – and edge can help solve them

Radius Networks have been making use of an on-premise edge compute solution since 2015. They define the edge as "where our installed, dependable infrastructure meets the mobile, transient infrastructure that we're interacting with".

Edge computing enables three key elements of their solution:

1. Reduced backhaul for continuous streaming.

> *"In order to track the location of a device, and give real time awareness of it, you have to have a constant stream of signals coming from the devices… our solution would not be viable if we had to do all of our location calculations centrally".*
>
> **Radius Networks**

2. Low latency for real-time location information. In order to accurately track if a customer or asset moves location, latency must be low.

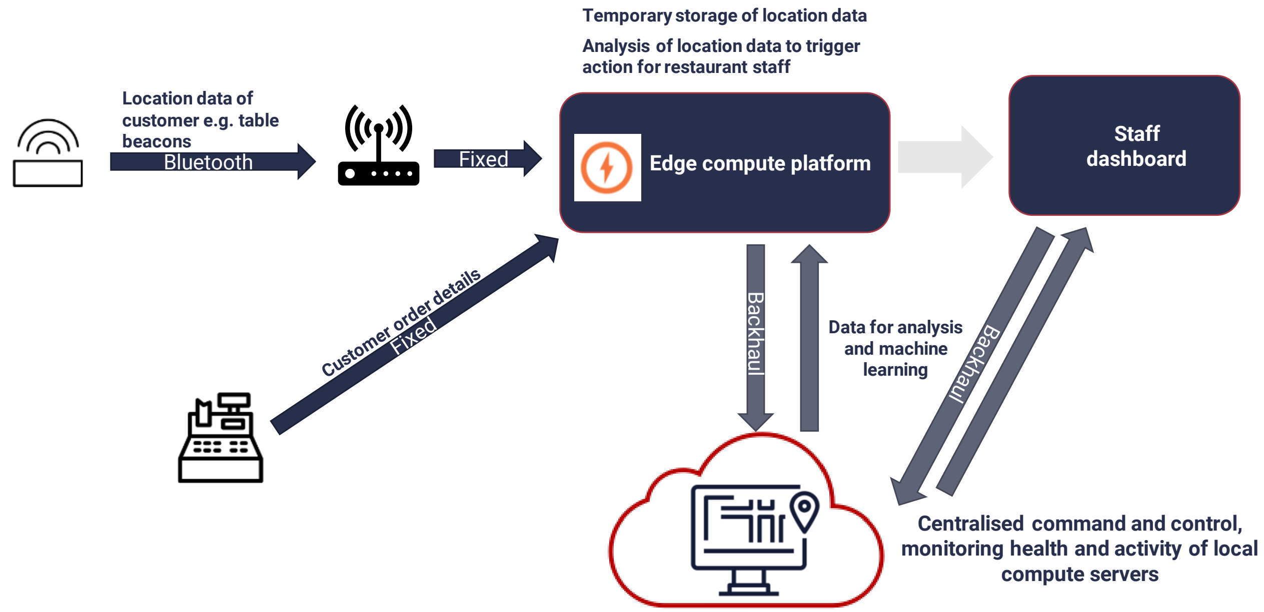
3. Data security for customer information. Information about where individual customers are clearly needs to be carefully stored and managed. For Radius Networks, this means this information being entirely collected, used and destroyed at the edge. The only information that they send to a centralised cloud is aggregated data (e.g. to analyse when was busiest in the

restaurant and therefore where staff should be reallocated) and metadata to determine the success of the solution. This ensures they are always PII (personally identifiable information) compliant.

Edge computing and location-based services: how it works

Radius Network's solution enables information from Bluetooth beacons to be aggregated with customer order details at an edge compute platform. Here analysis of the location data ensures monitors the real-time location of the customer. When food has been prepared staff can consult a dashboard for this information in order to deliver it to the customer. Backhaul to the cloud enables analysis of aggregated information (likely not to be real-time). For example, information from multiple chain locations of one restaurant can be aggregated and insights are which stores are most at which times can be drawn.

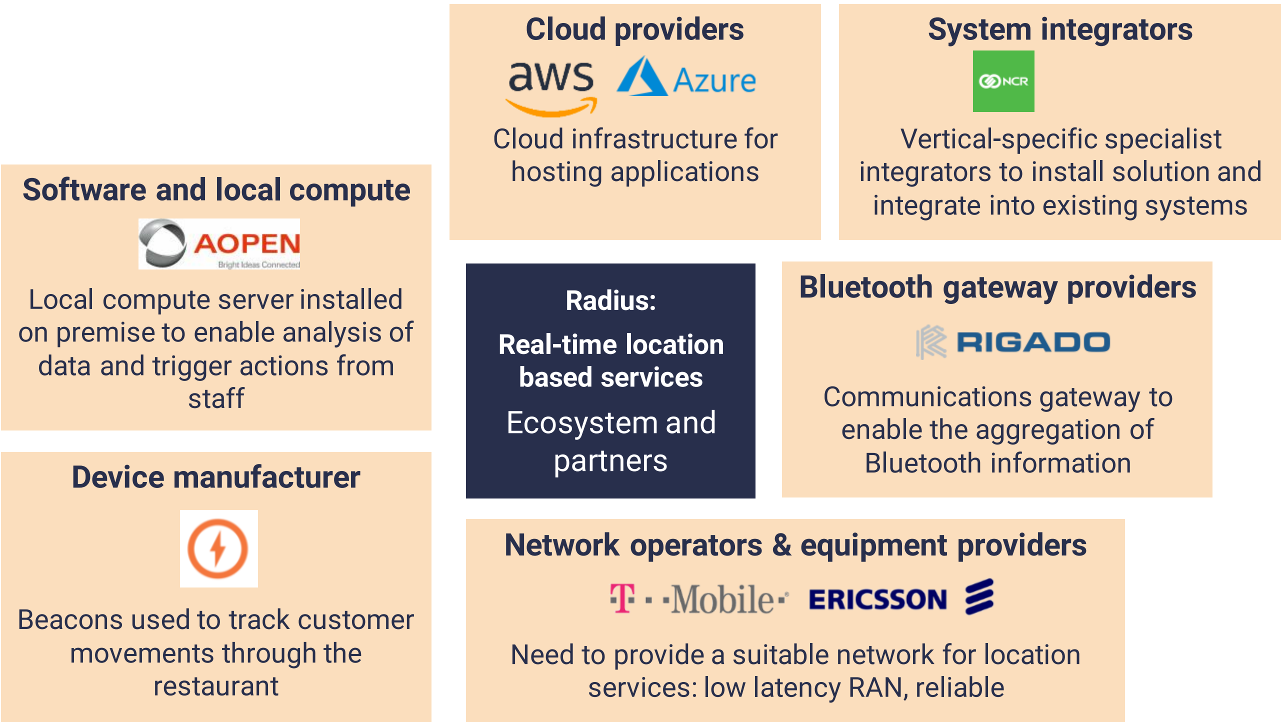
Figure 12: On premise edge computing enables real-time location monitoring



Source: STL Partners, interview with Radius Networks

Figure 13 highlights a non-exhaustive overview of the edge computing ecosystem that Radius Networks is part of. In this solution, Bluetooth, fixed and mobile connectivity are all used, while Radius' aim is for installation to be simple enough to be outsourced to non-specialised systems integrators who have existing relationships with the retail and fast food industry.

Figure 13: Radius has described the ecosystem their solution relies upon



Source: STL Partners, interview with Radius Networks

Challenges and implications for operators

Radius Networks currently make use of an on-premise edge compute solution. However, as telecom edge locations begin to grow in number, they would consider a move onto telecom edge, particularly for certain applications. For example, mobile edge computing could be used for location-based services outside of a fixed premises, such as for timing delivery of groceries to customer or in adjacent industries like tracking the real-time location of a VIP guest on the way from the airport to the hotel in the hospitality industry. In order for these types of solutions to be possible in the future, operators will need to guarantee ubiquity of service. This means whatever device the customer is using, and whatever operator they are with, the nearest edge compute node must be available for use.

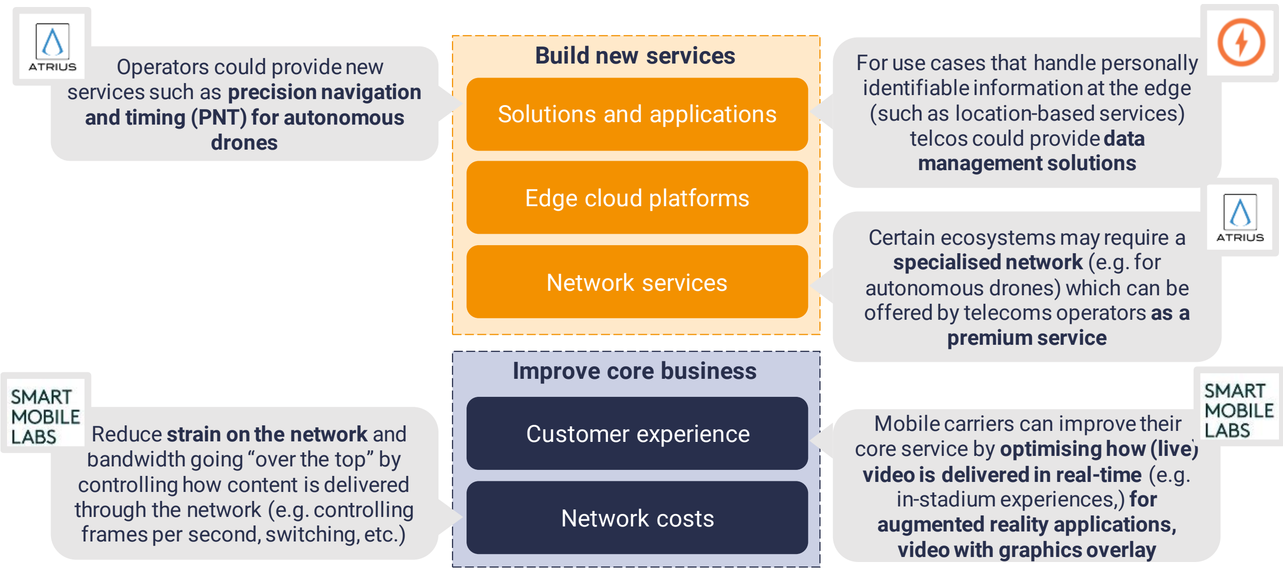
Monetisation opportunities for telcos

The case studies we have explored in the research programme demonstrate that there is real demand for edge computing, whether it be for unlocking new industries, such as drones and mixed reality, or improving customer experience and bringing in new services, such as application and video optimisation.

The question for telcos is how should they monetise the edge? In other words, where will they see the value that will drive a business case for deploying edge computing on their network or at customer sites?

There are two broad ways of thinking about this: first, improving the competitiveness of a telco's core business by reducing the costs of running the network and improving end-user experience. Second, telcos can seek to build and offer new services to enterprises and consumers that either leverage the edge compute infrastructure or improve other edge services.

Figure 14: Edge monetisation opportunities for telcos



Source: STL Partners

1. Improve core business

   – **Network costs**: Processing and filtering large data loads at the edge for video streaming, IoT sensor data, etc. will reduce the amount of data that has to go through the network and help operators better optimise how the networks are used and avoid overloading backhaul. This will become more prevalent as 5G adoption grows and we see a surge in data use driven by high bandwidth applications – more mobile video, mixed reality, etc.

   – **Customer experience**: Network operators may choose to open their telecom edges and onboard as many applications as possible to reduce latency overall for customers. This could significantly improve end-user experience and differentiate their core service from competitors, reduce churn and grow the customer base. For example, Smart Mobile Labs emphasised that using adaptive streaming on the network for cloud gaming would demonstrate a significant improvement for the gamer compared to playing the game over the top. Alternatively, they would be able to provide a superior service in stadiums – a spectator would be able to access applications during the event to enrich their experience, e.g. low latency replays with graphics overlays.

4. Build new services

– **Network services**: Some of the case studies highlighted the need for a reliable, robust network to ensure a consistently low level of latency. Atrius emphasised the requirement for a separate RAN for autonomous drone mobility, which would potentially use new spectrum to reduce interference. Similarly, Smart Mobile Labs is using private LTE or 5G test beds to overcome the issue of insufficient capacity at event venues for broadcasting video streams. Operators should view this as an opportunity to provide additional network services. In some cases, this may require a form of a private network, which could eventually be a virtual network slice on the 5G network, in other cases (e.g. in stadiums) it could require increasing capacity for the enterprise customer in a specified location.

– **Edge cloud platforms**: Developers need to be able to access edge clouds easily, namely the telecom edge. Operators who build out edge compute facilities and hardware at sites on their network will need to consider how best to offer this to enterprise customers. We have previously explored potential business models in our report '*Edge computing: Five viable telco business models*.'

  The interviewees highlighted that there needs to be some form of "universal infrastructure" that can easily be plugged into and that they need to be able to manage workloads effectively across different edges and potentially different operators. Developers do not want to use new tools for each deployment, therefore it is critical that either the telecoms industry develops standards and a unified platform, or a third party provides platforms that can be used as a mediator to access different infrastructure. It could also be that the hyperscale cloud players, like AWS or Azure, extend their capabilities to the edge, as we are already starting to see for on-premises edge through Greengrass / AWS Outposts and Azure Stack.

– **Solutions and applications**: Telcos can also seek to monetise edge by providing new applications and solutions to the market that leverage the infrastructure and unlock innovative new services for customers. For example, in the drone domain, Atrius suggested telcos could offer microservices that enable autonomous communication and navigation such as precision positioning, navigation and timing services that help to ensure drone positioning is accurate. Telecoms operators already have these capabilities, since they are used to synchronise timeslots and manage handover between base stations. Alternatively, Radius Networks proposed operators could offer data management solutions at the edge, given that many use cases anchor on the ability to pre-process and filter data close to the source of the data.

Conclusions: practical next steps for operators

To push adoption of edge computing forward and establish their own role in the ecosystem, telcos should:

- **Engage with the edge community.** There are more lessons to be learned from developers and the broader community engaged with edge computing. There is a growing ecosystem of companies involved in edge computing, as demonstrated by STL Partners' edge ecosystem tool. Telcos should play an active role within these forums, such as the Seamster community and LF Edge, to ensure the collaboration needed to accelerate edge adoption.

- **Explore further edge computing use cases.** For this research we spoke to seven developers focussed on four different key domains. However, we have engaged developers and ecosystems for many other use cases, including in domains such as gaming, smart cities and healthcare. We at STL Partners are supporting telecoms operators to identify and prioritise use cases that they should use to build propositions for their market.

- **Start to invest in edge infrastructure.** Telcos have an opportunity to be key players in the edge ecosystem, but only if they act now. They must overcome the chicken-and-egg dilemma of needing evidence of demand before they invest– because developers will only start to pay for edge when they are certain of scale – that the infrastructure will exist for all their customers. It will be important to ensure the hardware at the edge meets developer needs, for example if GPUs or accelerators are required for certain use cases such as augmented reality. The balance of catering to specific needs and providing an edge that scales will impact operator's deployment and roll out strategy, as well as on logistical constraints such as space and power requirements at the edge.

- **Provide developers easy access to the edge environment for testing.** Developers must be able to test their applications at edge sites in order to establish the type of user experience it will be able to deliver. For most developers, they evaluate metrics like latency and jitter within the context of overall application performance and, most importantly, customer experience. This will also help answer questions around whether their application will need to be re-architected to make best use of the edge in the most economical way and to what extent they need to change how they manage (edge) cloud workloads.

STL PARTNERS

Research

Consulting

Events