The road to 5G - Densification and Aggregation at the Mobile Edge

5G will bring several groundbreaking new technologies into play with its new air interface and Radio Access Network (RAN) architecture extensions. It will also need an emerging software defined infrastructure, one that provides greater topology flexibility, essential to deliver on the promises of high availability, high coverage, low latency and high bandwidth connections that will open up new parallel industry opportunities. 5G unlocks many new doors and one of the keys to its enablement lies in the elasticity and flexibility of the underlying infrastructure described in this whitepaper.

The road to 5G
Densification and Aggregation at the Mobile Edge

Introduction

5G will bring several groundbreaking new technologies into play in its new air interface, offering massive multiple-input, multiple-output (MIMO) that groups antennas at the transmitter and receiver to provide better throughput and spectrum efficiency, beam forming for performance optimization, and flexible band sizing allowing previously unavailable bands to be used.
Its Radio Access Network (RAN) architecture extensions bring further innovation with technologies like CoMP, connecting a user to multiple cells to provide greater reliability, small cell support for greater indoor coverage and cell density, 5G-NR for expansion in unlicensed bands and D2D for device-device connectivity as some examples.

But 5G will also need an emerging software define infrastructure, one that provides greater topology flexibility, essential to deliver on the promises of high availability, high coverage, low latency and high bandwidth connections that will open up new parallel industry opportunities. 5G unlocks many new doors and one of the keys to its enablement lies in the elasticity and flexibility of the underlying infrastructure.

Software Defined Networking(SDN) and Network Function Virtualization (NFV) will be two of the essential conduits to 5G infrastructure agility making networks easier to install, secure, automate and manage in a cloud-based infrastructure comprised of:

- Virtual Enterprise Customer-Premise Equipment (vE-CPE), providing a new generation of intelligent gateways into company buildings and enterprise offices connecting multiple business and Internet of Things (IoT) solutions.

- Multiple-access Edge Computing (MEC) which extends computation, storage, and networking out from the hyper-scale datacenters, closer to consumers of applications providing access to RAN subscriber information and offering the low-latency round trip times between user's and applications. MEC is recognized by Europe's 5G-PPP as one of the key enabling technologies for 5G

- Centralized Cloud services, either in-house, hybrid or public consolidating non-latency critical applications and services

Open Commercial-off-the-Shelf (COTS) Networking Equipment based on Intel Architecture from companies like Advantech Networks & Communications Group (NCG) coupled with enabling software from network builder ecosystem partners, will play an important role in the virtual infrastructure that will precede and underpin 5G. This whitepaper focuses on how those foundation elements for 5G are built.

Edge, Cloud or Edge Cloud?

According to the Gartner Hype Cycle for Emerging Technologies, 2016, Virtual Reality (VR) was the only technology on its slope of enlightenment last year, followed close behind by Augmented Reality (AR). Gartner considers the Hype Cycle for Emerging Technologies to be unique among most Hype Cycles because it condenses insights from more than 2,000 technologies into a concise set of must-know emerging technologies and trends that will have the single greatest impact on an organization's strategic planning. Both VR and AR technologies are expected to introduce increasing transparency between people, businesses and things as the technology becomes more adaptive, contextual and fluid.

With VR, AR and assisted reality on the horizon, the next wave of applications are likely to be mobile and latency-sensitive if they are to be immersive everywhere. There is also an expectation that VR, AR and gaming will require user equipment (UE) to offload certain computational algorithms to the cloud even though the latest smart phone devices are increasing in CPU performance. On one hand UEs may still lack in performance when heavy computational loads are needed in short bursts of time while on the other hand, battery consumption remains a concern for a prolonged user experience when all CPU cores are loaded and draining precious power.

To be fully immersive, mobile apps will need to deliver on Quality of Experience (QoE). However, latency in traditional centralized cloud topologies is far too long and it is obvious that slow round-trip response times will negatively impact user experience expectations. Distance means delay and applications requiring near real-time response will need compute and storage resources closer to the users in order to achieve the Quality of Service levels expected by subscribers. Cloud services will need to be placed at the edge of the network in proximity of user equipment (UE) in order to meet latency requirements.

A decentralized architecture with compute and storage capability at the edge of the network provides an intermediary processing stage to reduce the amount of data shipped back up to the cloud by executing algorithms for applications such as face recognition, local building and landmark labelling for augmented reality and cognitive assistance or even crowd-sourced video analytics. Based on the Pokemon-GO AR experience, uptake of such applications can reach viral proportions, causing sudden upswings in demand. With further increases in UE traffic placing even greater demand on the network, Communications Service Providers (CSPs) will need to respond fast to capacity increases while maintaining latency.

AR tests have actually been carried out on MEC testbeds and they demonstrate that latency can be reduced by up to 88% and UE energy consumption by up to 93% by computational offload to MEC servers. [1]

The benefits of MEC on consumers is above all better battery life and a better QoE through low-latency response, adaptive QoE based on real-time network analytics , and content caching at the network edge. CSPs and third parties can also reap benefits from new edge services such as Big Data, IoT and connected vehicles. For example user and IoT data gathering at a MEC aggregation point can be used to offer localized services. IoT Gateway services running on the MEC server can be used to aggregate data from different RF technologies and offered as a service. Connected car services can provide low-latency messaging for car-to-car and car-to-infrastructure communications as demonstrated by MEC ecosystem tests in 2016[2]

In addition, MEC enables better network performance and QoE optimization by providing real-time information to the backhaul for traffic shaping and re-routing. It can also help offload congested backhaul though local content caching at the MEC server. Finally the MEC can gather more UE-related information at the edge and process it in real-time without going back to the central cloud.

There are an increasing number of processes and use cases that can benefit from edge computing, but the realization that remote hyper scale data centers will not be able to deliver the latency response times required is forcing CSPs to take a closer look at where they will locate their edge services. In a recent 2017 webinar poll [3], attendees were asked how they expected to use edge locations for hosting cloud services over the next 3 years. The response may differ depending in which part of the world and which cities the respondent is based, but there is a clear view that strategic edge locations will need to be upgraded.

- Radically expand use of edge locations for cloud services 19.4%
- Select a number of strategic edge locations for upgrade 41.7%
- Upgrade a few edge facilities but not many 19.4%
- Don’t expect to increase use of edge DCs 19.4%

The Importance of Location

The two fundamental motivations for edge computing are access to radio network information and low-latency response time. As such the importance of location is paramount. New deployment models are evolving to satisfy services that need lower latency, real time features, data reduction, local caching and greater security. And although the virtual RAN (vRAN) is probably one of the most demanding areas for deploying edge services, it is quintessential for edge computing.

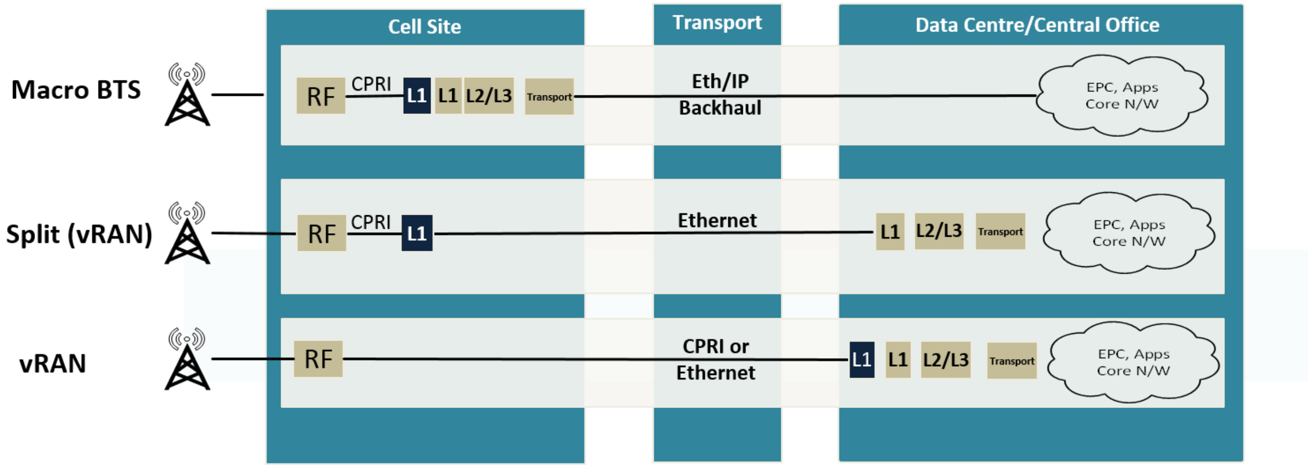


Figure 1 Deployment Scenarios

In a traditional macro cellular network, remote radio heads (RRH) are mounted on the base station tower along with the antennas. They are typically connected to a baseband unit (BBU) located at the base of the tower over a fiber link. The BBU performs the baseband signal processing from the RRH.

In this approach all Layer1 to Layer 3 processing takes place at the cell site close to the tower and RRHs. This is shown at the top of Fig. 1 above. In contrast, in a fully centralized vRAN approach, shown at the bottom of Fig. 1, this processing is moved into a centralized location such as a telco data center or a central office. While this centralization may provide obvious benefits for CAPEX and OPEX, the latency and bandwidth on the transport network becomes a real challenge. In the middle slice of Fig. 1, a split implementation addresses this issue by performing the most critical processing at the cell site. Other tasks which do not require very low latency or consume very high network bandwidth can be moved out to a centralized location.

Virtual RAN

In an ideal world without constraints, the Centralized RAN would run entirely on newly built, highly optimized Telco Cloud Data Centers which can host low cost hyper scale cloud hardware such as Open Compute Project (OCP) racks that are 1200mm deep and rated at 25kW/rack.

However, establishing these data centers, especially in densely populated areas in order to keep the latency to the cell sites down, is a real challenge. Actually there is not really much debate that a part of the implementation needs to be hosted in legacy locations such as central offices. The only debate is how big the percentage split between green field and brown field sites is.

Central offices usually top out at 10kW/rack cooling and power. Moreover, central offices normally use shorter 600mm deep telco racks. Substantial remodeling is required to turn these sites into data centers that would host cloud style hardware. This essentially means, a Carrier Grade platform is needed to implement a Centralized

RAN in a central office environment. Depending on whether the implementation is based on a split or completely centralized architecture, and depending which transport type is used, CPRI to Ethernet termination devices may be needed which would normally be implemented as standalone, rackmount units.

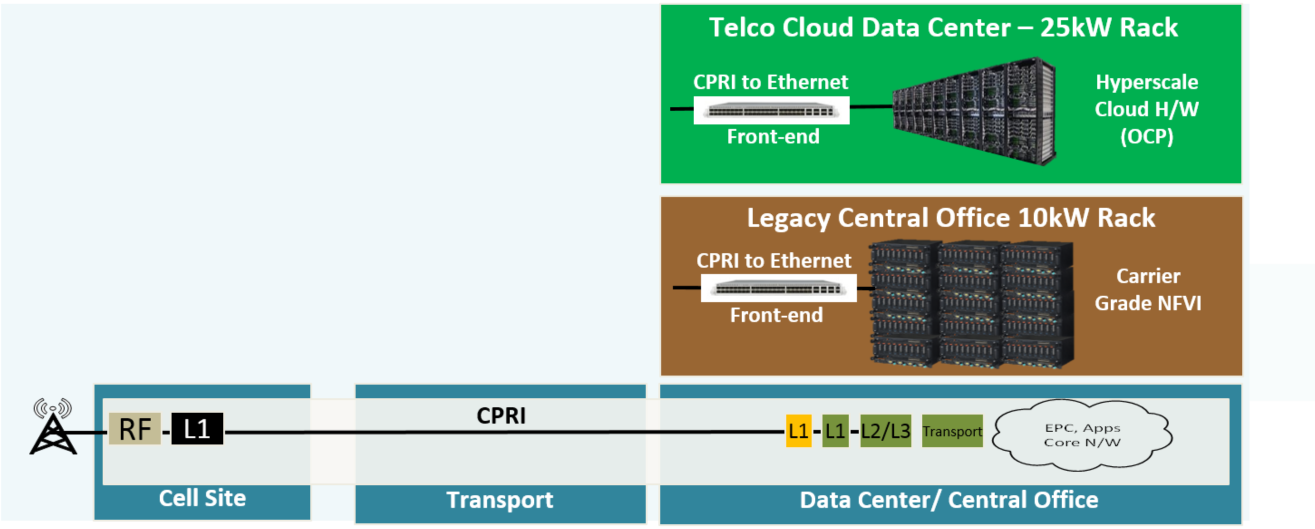


Figure 2 Centralized RAN

The main issue with this implementation is that there is no possibility to add edge computing. Edge computing sits above the L2/L3 RAN stack. If that stack is hosted in a centralized location, then just by the terminology, the services sitting on top cannot be at the edge. In practice, this means that for 5G, IoT and all of the related uses cases mentioned before, a centralized RAN approach will face huge challenges as transport latency will become extremely critical.

Edge Architecture Approaches

A distributed virtual RAN approach puts compute where it should be; at the edge. In a distributed RAN, the processing is implemented at edge sites, aggregation points and other CSP points of presence. The benefits of implementing this approach are that latency requirements are easier to manage. It is also easier to integrate heterogeneous infrastructure in the RAN and connect that into a unified mobile core.

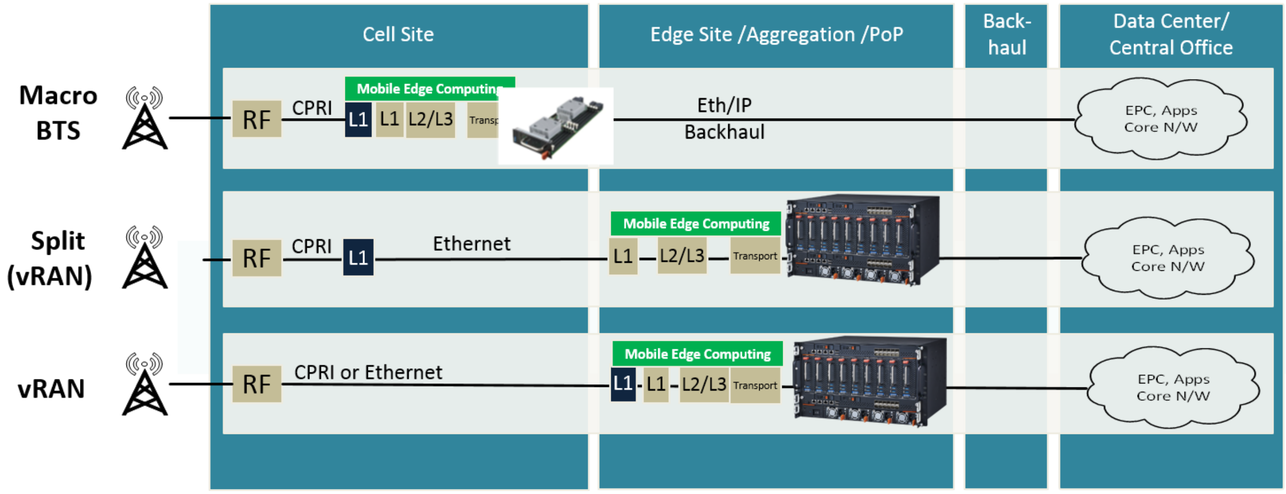


Figure 3 Edge Architecture approaches – vRAN

By locating compute at the edge, not only can latency issues inherent in the radio processing be resolved, but also latency critical new services which are key aspects on the path to 5G can be deployed. The ability to guarantee low latency response times to an Enterprise slice in the future 5G network will be a key differentiator for CSPs and essential to offering competitive service level agreements.

Moving Compute to the Edge

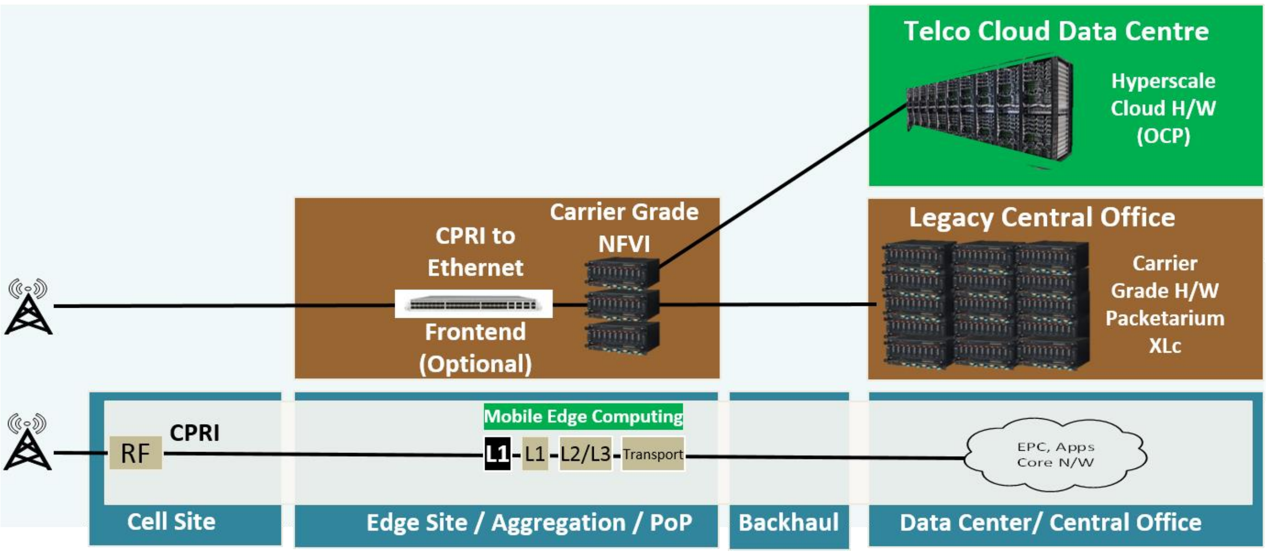


Figure 4 Moving Compute to the Edge

Implementation options for a distributed RAN enabling low latency edge services poses a problem not just of location but also of equipment type. Hyper scale cloud hardware is unsuitable for edge sites which are traditionally unmanned and also space, power and cooling constrained. Even if OCP could be installed and powered at the edge, operations management would become a major obstacle as OCP has been fully optimized for operation in large data centers where operators maintain the equipment and failover of one node or a complete rack is of less consequence based on the number of backup server racks present.

However, in an Edge Computing application there may only be a few racks full of compute at each edge location. If a compute element at the edge fails, given that there are no aisles full of spare backup compute servers an availability challenge needs to be addressed. Failure in a network slice causing many autonomous cars to revert to safe mode will have a huge impact on service quality. Seen from this angle, 5G and edge computing will not be able to rely on equipment that has been architected for completely different use cases as the CAPEX savings would soon be offset by increased Operating Expenses.

Meeting Brownfield Edge Data Center Requirements

One technology that addresses the gap for deployment in Brownfield Edge Data Centers is Advantech's Packetarium XLc, a carrier grade blade server optimized for the network edge and designed for the brownfield installation constraints mentioned above.

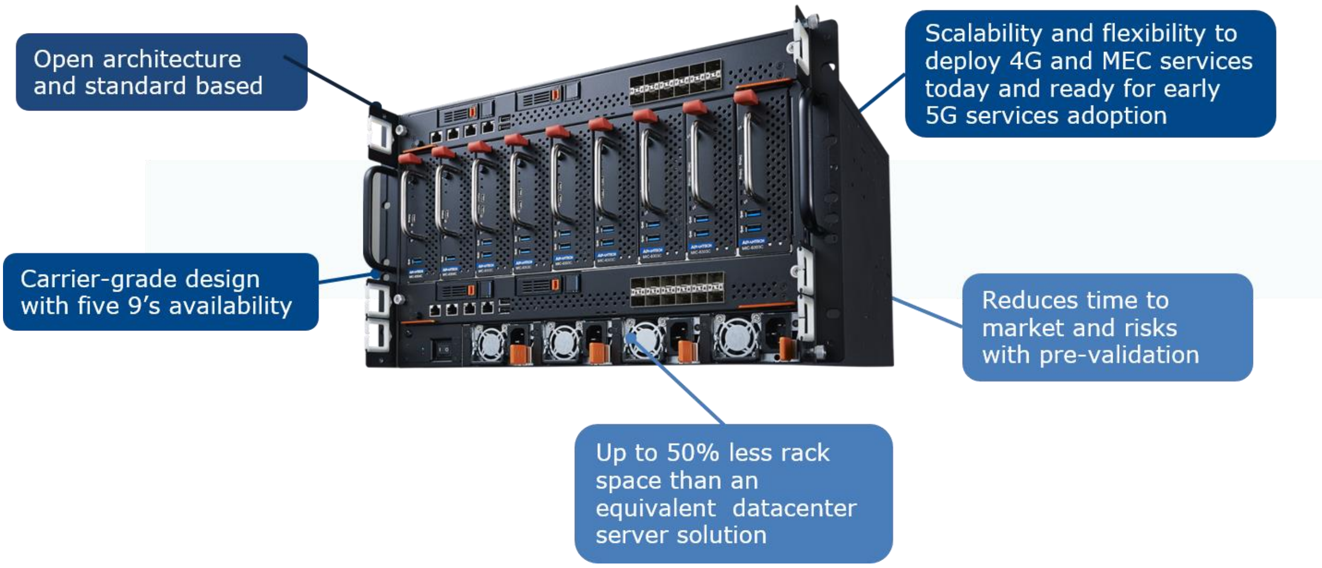
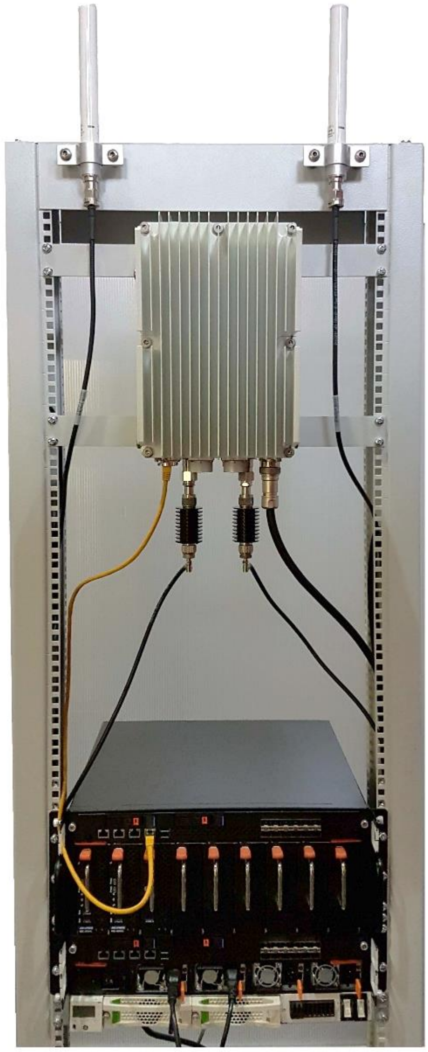


Figure 5 Benefits of Advantech Packetarum XLc Carrier Grade Blade Server

vRAN Proof-of-Concept

On the Amarisoft booth at Mobile World Congress 2017, Advantech and AW2S demonstrated a software based vRAN, with Amarisoft eNodeB and virtual evolved packet core (vEPC) software running on an Advantech Packetarium XLc connected to an AW2S Ethernet RRH. The eNodeB is based on the 3GPP release 13 standard including features for IoT (NB-IoT and Category M1), VoLTE, and Carrier Aggregation. Moving the LTE stack from the lab to a virtual macro base station demonstrates new opportunities and offers a cost-optimized densification solution to CSPs. The Packetarium XLc is capable of handle 108 sectors, equivalent to 36 Macro Cells with all software executing on Intel® Xeon® Processor D-1587 based blades with sufficient processing headroom available for MEC apps.



- Radio over Ethernet on RJ45 cat 5 cable
- Compliant ETSI TS 136 104 Release 13
- AW2S Ethernet RRH
- MIMO 2x2 20MHz TDD Band 40
- Advantech Packetarium XLc
- 2 x Intel® Xeon® Processor D-1587 per blade (2 x 16 cores)
- Amarisoft LTE 100 (3GPP rel 13 ready)
- Fully functional LTE RAN in a box

.

Summary

The boundary of the Network Edge is being redefined as very diverse range of radio protocols need to be supported and co-exist with new revenue generating applications. While new green field sites will need to be deployed, existing brown field sites will have to be retrofitted – this is where Advantech's Packetarium XLc fits in. The system was designed to meet increasing RAN densification trends by bringing higher aggregate compute performance out to the edge of the network. Easily connected to existing CPRI fronthaul, to additional small cells or new Ethernet-ready Remote Radio Heads (RRH) from companies like AW2S, and coupled with software defined radio solutions from Amarisoft, the system accelerates the process of building solutions for vRAN, MEC and NB-IOT. It lays the foundation for 5G enabling software solutions to make use of precise RAN subscriber information through a variety of mechanisms from ETSI MEC, NFV and Openstack to more carrier-grade solution such as Titanium Cloud from Wind River or latency specific implementations such as SlapOS from Nexedi.

References

P. Mach, Z. Becvar, "Mobile Edge Computing: A Survey on Architecture and Computation Offloading"

[1] J. Dolezal, Z. Becvar, and T. Zeman, "Performance Evaluation of Computation Offloading from Mobile Device to the Edge of Mobile Network", IEEE Conference on Standards for Communications and Networking (CSCN), 1-7, 2016.

[2] https://networks.nokia.com/solutions/multi-access-edge-computing.

[3] http://www.lightreading.com/webinar.asp?webinar_id=828

Advantech Contact Information

Hotline Europe: 00-800-248-080 | Hotline USA: 1-800-866-6008

Email: NCG@advantech.com

Regional phone numbers can be found on our website at **http://www.advantech.com/contact/**

www.advantech.com/nc