Multi-Access Edge Computing

Solving Tomorrow's Problems Today

Christof Wehner, Senior Technical Marketing Architect, Artesyn Embedded Technologies

Multi-access edge computing (MEC), originally conceived for 5G networks, is solving many of the challenges faced by service providers today. It offers opportunities for operators that want to improve their revenue stream, bolstering ARPU while lowering the load on their networks.

This white paper analyzes these challenges and their solutions, as well as the business benefits, of using an MEC architecture. Critically, the paper examines the ways that operators can not only save money, but also make money from the opportunities that MEC enables. Finally, real application examples and system architectures move the discussion from theory to reality, with solutions that can be deployed today.

IOT Solutions Alliance

INTEL® NETWORK BUILDERS PARTNER
ACCELERATING NETWORK TRANSFORMATION

ARTESYN™
EMBEDDED TECHNOLOGIES

Introduction

The recent name change for ETSI's MEC group from Mobile Edge Computing to Multi-Access Edge Computing signifies a change in the world of communications providers, from separate networks with separate requirements to unified networks with unified requirements. MEC, which was originally conceived to support and enable the requirements set for the new 5G networks, is now capable of not only solving those future problems, but it also solving many problems service providers are seeking to solve today, if not yesterday.

What is *Multi-Access Edge Computing*?

MEC is a standard currently under development by ETSI, with all major players in the networking industry participating, from telecom equipment manufacturers (TEMS) to network operators to hardware and software vendors. Having been around in more proprietary formats for a while, MEC is now being standardized to a broader community of developers to allow operators to more rapidly deploy new services and generate new revenue streams.

MEC currently evolves from a software standpoint, ignoring hardware, as it is based on virtualization technology. The goal is to define a set of APIs that enable the creation of virtual network functions (VNFs) that respond to all the needs of a mobile communications network, including security, orchestration and portability, while leaving the actual implementation to the respective provider. There are multiple projects ongoing inside and outside of ETSI, including Open Source MANO, OPNFV and others.

MEC seeks to solve a number of problems that arise from LTE Advanced and 5G advances that place more stress on the network, but the real opportunity is enabling operators to address additional business issues. First of all, like all evolution stages of mobile networks, a major focus is on making more bandwidth available to users, now extending into the Gigabit(s) per second range. This, in turn, results not only in more required bandwidth in the core network, it also enables new applications which can consume such high data rates – which should have a localized counterpart in the network.

Additionally, 5G moves from previous-generations' "speech-acceptable" latencies in the 100 milliseconds range to "machine-acceptable" latencies, which fall into the single-digit millisecond range. This, again, needs more localized processing capacities.

While seeking to solve these new challenges, MEC can also help address other problems present today. We will analyze these problems and their solutions, as well as the business benefits, of using an MEC architecture in the remainder of this white paper.

Problems that need to be addressed

Whenever we talk to network operators today, we find a number of issues that they need to be urgently addressed. There are multiple areas that have a significant business impact on operator revenues and profits:

- Declining ARPU
- Infrastructure cost bottomed out
- Regulatory constraints - net neutrality
- Converging technologies, no more separation between fixed and mobile networks

Declining ARPU

At the core of an operator's revenues is the ARPU (Average Revenue per User). Statistics show that ARPUs in North America topped out some time between mid-2013 and mid-2014, and are now shrinking across the board. In order to sustain revenues, there is a need to either add more users at additional cost, or offer something to users that makes them use more new services, while avoiding the relegation of becoming a utility provider between subscribers and content owners, who are making the money today.

At this very moment, operators are scratching their heads in order to reverse that trend now and create a compelling solution set that not only appeals to subscribers, but also enables them to spend more with their operator rather than an external service. As pure connectivity becomes a commodity, there is a strong need for added, differentiated services for subscribers. This requires new technology in order to work at a reasonable cost, while generating significant additional revenues. Additionally, it requires new ideas for revenue-enabling services.

Infrastructure cost bottomed out

When considering the overall cost for infrastructure, there is little room for further savings. In reality, new services require extended capabilities, new hardware and software, which requires more space and new sites for deployment in order to satisfy not only functional needs, but also new requirements, such as much lower latencies and much higher bandwidth.

Growing usage everywhere, with hotspots in shopping malls, sports stadiums and tourist locations, demands more coverage and more available bandwidth with improved availability. These places are not prepared for large installations, so either more equipment is needed in a brownfield deployment, or equipment needs to be deployed in new locations with very little room to spare. These requirements, paired with flat infrastructure costs as mentioned before, create a need for a new breed of product that enables the network to scale, rethinking today's approach.

Regulatory constraints - net neutrality

Regulation all over the world is kicking in, starting to take a closer look at net neutrality, and ways to control those who control access to the network. Recent cases revolving around renowned companies such as Google and Facebook, with their respective initiatives for generally available network access around the globe, show that authorities in a variety of countries are having a closer look at the practices.

While showing significant opportunities for those cooperating with these companies, authorities also appear concerned with the accumulation of information in the hands of just a few players. Their task at hand is now to balance between companies and end users, to create a level playing field for operators, information owners and customers.

Converging technologies, no more separation between fixed and mobile networks

In the past, there were separate and distinct infrastructures deployed for fixed and mobile networks. Additionally, these were governed by different protocols, different (sometimes local) codecs (speech encoding/decoding software) and different ways of routing traffic. As time moves on, these infrastructure technologies converge, creating new opportunities and challenges both for incumbents, who need to transform their now redundant, application-specific networks, as well as newcomers, who need to build a new future-proof infrastructure from the ground up.

Multi-access edge computing helps solve many problems

MEC as a technology has shown that it may not only be a ground-breaking new, successful technology as well as a stepping stone towards 5G networks, but also it has already started to present itself as a technology that will solve many of the listed problems much earlier than the 5G rollout.

New services appearing

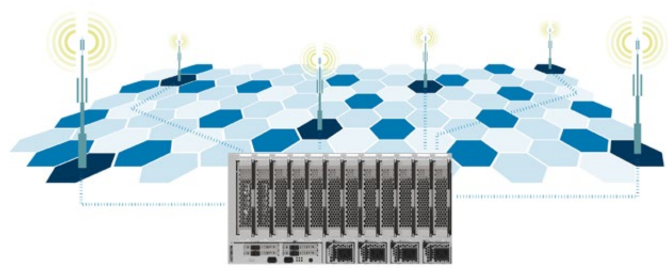
Being close to the edge of the operator's network, MEC enables services that were unthinkable in the past with centralized networks. For example, MEC enables an excellent service provider offering for venue owners, where typical distributed antenna system (DAS) and Wi-Fi implementations have failed to support peak coverage demands at concerts, major sporting events and shopping malls. Today the owners of these facilities bear the CapEx costs of serving these areas with Wi-Fi coverage through DAS equipment and use old models for provisioning for peak capacity. However, these days peak capacity means enough network bandwidth for data coming down and going back up, as in videos uploaded to Instagram during concerts.

MEC, through its use of virtualization, has the potential to scale capacity and set up and tear down different and distinct services that monetize that network dynamically. These new MEC services can also be deployed by the operator and contracted for on an as-needed basis by the venue owner, or as an ongoing service they can license, in lieu of paying for DAS equipment and maintenance. It changes the model by removing the CapEx and overhead costs for the venue owner and allows them to buy it as a service with incremental revenue derived by a more powerful edge network, and thus becomes a valuable enterprise service offering for carriers.

Media independent infrastructure-as-a-service

MEC devices can act as local gateways to provide infrastructure-as-a-service (IaaS) services, thus bringing the "entrance" to these services directly to the user. This enables shortened setup and remote times providing even better end-customer experiences.

Also, functions that need to be run near the customer for improved latencies, may now be split into a non-critical part running in the network data center where they are most cost-optimized, while the parts that affect customer experience can be deployed in a virtual machine close to the customer. Due to the unified access, regardless of whether the connection is fiber, a DSL line or cable, or via the mobile network, those functions may remain where they are while the customer is in the same area. This saves the effort of moving applications and re-configuring the SDN network, as would normally be the case for non-MEC enabled applications.



Micro-localized applications

Today, lots of events with massive attendance are happening in a single place, such as sports events or trade shows. MEC devices now enable service providers to provide very localized additional services at a fee for the respective environment, which can deliver applications that derive new revenue streams from consumers for operators or from attendees for venue owners.

Examples include exclusive footage of certain scenes of a sports game, or location-specific content as well as local search, additional information or augmented reality guided tours, tailored to the specific person's needs. These are just some examples of ways that MEC enables providers to stand out from the crowd and offer valuable services that customers want. There will be more and more arriving in the coming months and years, as technology develops and new business models appear.

All this brings us to a very central question…

Where is the money?

The reason for creating all these business models, as well as solving the ARPU problems outlined earlier, is to find a way to bring additional value and capabilities to customers, creating a reason why they want to pay more. Additionally, the major question remains: How can I, as an operator, save money on the other side, if at all, to survive and satisfy my stakeholders?

Savings enabled by increased density & disaggregation

Looking at the cost of existing infrastructure architectures, the overall cost has bottomed out, restricting direct savings by replacing equipment. However, new technologies and new developments enable increased density as well as disaggregated functions, which allows the use of equipment configurations that deliver a lower total cost per function. OpEx and CapEx cost reduction is further enabled by new CPUs such as the new Intel® Xeon® D processor family as well as new system types such as the Artesyn MaxCore™ family of platforms. These platforms allow MEC users to choose the right platform for a given application profile (for instance high availability or NEBS required or not) and location in the network (3U type for edge or hyperscale for core) to hit the right cost and density point, while at the same time ensuring an application deployment and administration model that remains consistently the same across platforms.

Depending on the requirements profile, add-in cards can be used to support the exact configurations in a single system. This capability was shown in a recent Intel® Network Builders winning submission presented at Verizon Innovation Labs, which showcases an augmented reality (AR) application in a system, with more slots and functions remaining available for other applications.

Monetize through location aware augmented reality

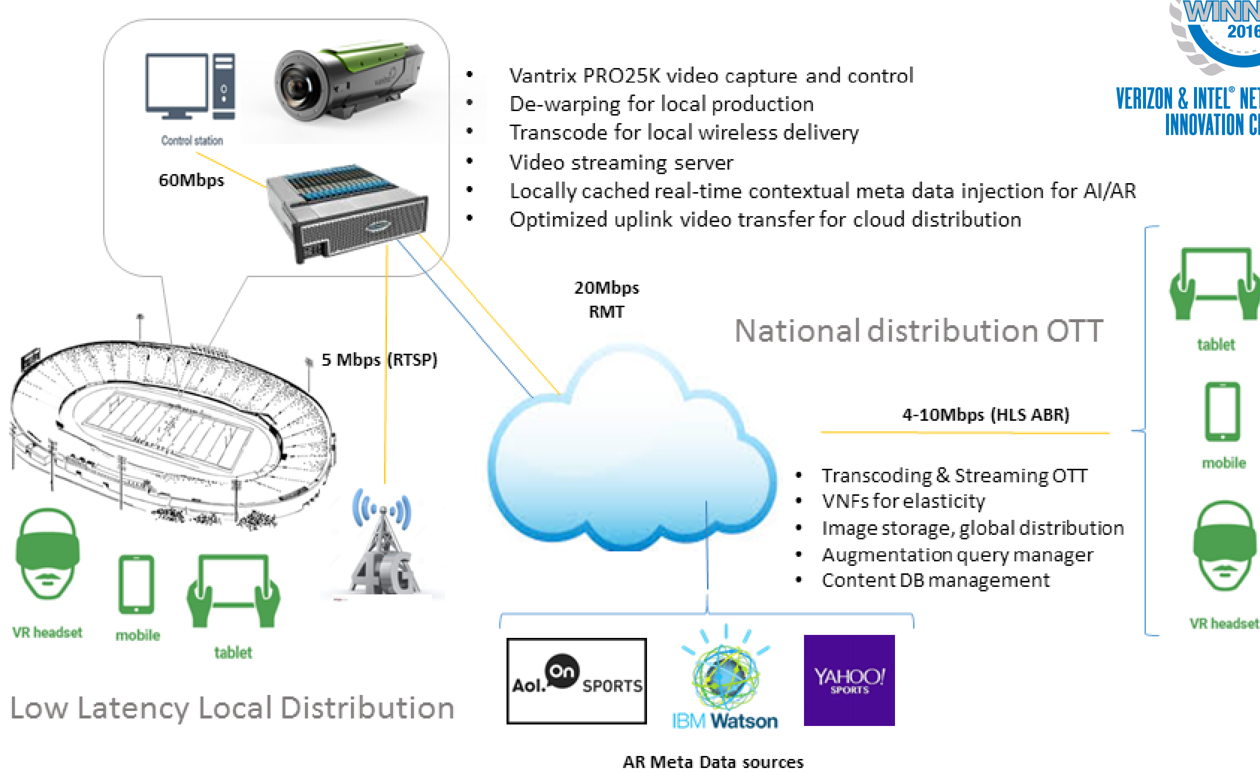
The application submitted was a network edge 360 VR & AR broadcast and delivery infrastructure platform. The entry category was entitled 'Low Latency' and is an ideal fit for the platform. The MaxCore hardware latency is hundreds of nanoseconds, making it an excellent choice for real time or latency-sensitive applications. However, in the case of AR video delivery. the true metric is whether the user sees the delay between what exists and what appears on the screen.

Sitting on top of the hardware platform was an optimized glass-to-glass video processing system from Vantrix. Vantrix understands the need for real-time delivery in AR applications and has constructed a flexible system that takes advantage of the Artesyn MEC platform to insure low latency delivery to the user. The Artesyn-Vantrix system was chosen because the demonstrated latency was just under one (1) second from camera lens to display panel (phone screen). This contrasts with a delay of approximately eight (8) seconds when processed through the cloud, also demonstrated. Further development now has the total latency at around 400 milliseconds.

In this system, quality was not sacrificed for speed; the camera delivers 5K x 5K resolution, enabling broadcast quality. By using an MEC platform to process local content and a feed to the cloud for global distribution, the system demonstrates how a service provider can combine local caching and content to deliver new services. And with the application offered as a VNF, capacity can be added seamlessly or replaced with new services without the need for new hardware.

Deployment System Architecture Event Use Case

GRAND PRIZE WINNER 2016
VERIZON & INTEL® NETWORK BUILDERS INNOVATION CHALLENGE



This AR application is just an example of what can be done in the near future. Example applications can include intelligent city guides that point out important streets, buildings and places, and offer additional information on the item in view. Similar applications would be using the same technology on a fairground in a controlled environment, creating interest-specific tours of the fair. Probably the largest monetary application, however, has been shown by Nintendo's Pokémon Go game. This game can be considered the breakthrough for AR use in mobile gaming.

Pokémon Go as a prime example – and some thoughts on making it even better with MEC

Let's have a look at the numbers for Pokémon Go, the first example of an AR based game, which was and still is a huge success. Latest numbers show it created $950M USD in revenues in the first six months from its release alone, being downloaded 500 million times. Imagine what would have happened without server overloads, with even more localized content, and local compute resources rather than huge backend server farms that have to cope with simultaneous users from all over the world!

Local touch is easily possible as well as adding service-provider-specific functions to the game. This creates the opportunity to generate a shared additional revenue stream through operator-marketed (and exclusive!) local advertising and game extensions for both game maker and local operator. Imagine, say, a 20% share of such revenues in your region!

How Artesyn and ecosystem partners can help

Artesyn's MaxCore™ family of platforms, with its unique ExpressFabric PCI Express implementation, enables new models of cooperation of different servers. Not only can multiple server modules be inserted in a single chassis, they can also share PCI Express resources, or associate PCI Express resources with selected modules. This enables scaling of applications at run time across multiple servers, segregating applications by physically running them on a separate server, or even sharing storage or networking between multiple servers through an intelligent network interface card (NIC) that also allows preprocessing on the NIC, only forwarding relevant data to the respective related server.

As the MaxCore platform features standard PCI Express x16 slots throughout, third-party content such as cards that carry application specific silicon, FPGAs, other NICs, or graphics cards with up to 150W per slot (300W for dual-slot-wide cards) can be easily added and qualified.

Density

The MaxCore platform is designed around a very intelligent PCI Express switch that can be reconfigured at boot time. It can operate in multiple modes, the most flexible of which is the fabric mode. In this mode, any slot can serve as a root complex (the source of a PCI Express device tree), or as a node slot (so a device like a network adaptor, USB controller or others). This allows a seamless upgrade of a system by inserting additional servers, or a reconfiguration that is enabled through a reboot of the system. All this functionality is administrated through a management CPU that presents a graphical user interface to the user of the system. By enabling sharing of cards that allow virtual slicing into multiple components (so a single network adapter can pretend to be ten devices that can now be shared among multiple servers), the system minimizes the needs for generic I/O, allowing more space for server complexes or additional I/O.

With up to 15 slots, and microserver cards configured to carry up to two server complexes with up to 12 cores each, a MaxCore system can house any configuration from a single server with 14 slots holding I/O cards up to 15 microserver cards with 2 servers each, totaling up to 30 servers running 12 cores each!

There are other members to the MaxCore family that combine fewer slots with other attributes such as support for high availability through hot swap, a system configuration without single points of failure or fit into rack mount standard such as Dell DSS9000, open compute (OCP) or CG-OpenRack-19. MaxCore is a flexible architecture that can not only be adapted to support additional use cases, but also offers multiple implementations that are fit for a variety of applications.

Resource allocation for shared use cases

As a MaxCore system allows the dynamic assignment of resources to application CPUs, which is enacted through a reboot of the system, it allows developers to share server complexes among different applications. Depending on the resources needed at a given point in time, servers can be assigned different workloads or execute a family of VNFs according to current load requirements. These can change over time and the system can respond by moving, starting or shutting down the related VMs, or by assigning a complete server complex to a new application. As all of this happens in a single enclosure with a shared external interface, all these reconfigurations can be implemented without the need for network reconfiguration or even manually patching cables to a different adapter card.

This flexibility, together with the support of the intelligent network adaptor and the accelerated vSwitch by Artesyn, enables the flexible application deployment, removal and move that is needed in use cases such as MEC, especially as the small size of the MaxCore system solution fits edge-style uses cases particularly well.

Example application configurations

How can you solve these challenges using a MaxCore system? We have picked some example applications, and will now discuss how we would suggest solving them in a standardized system configuration. Contact Artesyn to help you find your configuration.

Localized cloud gaming

Cloud gaming is a new field of high interest with astonishing growth rates. What used to be simple multiplayer games has now evolved into a massive market, where the compute capacity is no longer provided in peoples' homes but rather provided in the cloud, including the video rendering. In return, this means that much higher complexity games can be created together with far higher resolution and picture quality, while the playback device basically is streaming the movie generated on the cloud gaming platform. Additionally, this enables cloud gaming providers to easily roll out updates to games as well as new products and games.

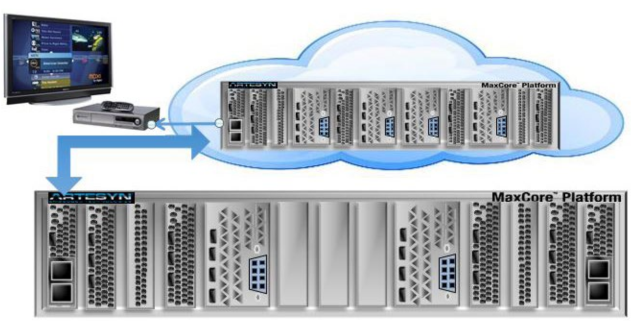
Split application architecture

A major factor in this cloud gaming market, however, is low latency, as players expect an immediate reaction to their action, be it turning, acting or moving items. From a physics perspective, in turn, this means that the front end of the platform that does the rendering on behalf of the player cannot be placed somewhere in a datacenter in a remote network, in fact it needs to be pretty close. MEC is a perfect platform for these applications, as it enables developers to bring the front end portion of the application to the edge of the network, as an accelerated VM very close to the actual player – and only when the player is actually playing.

This function, in return, needs enough compute capacity coupled with a virtualization-aware graphics card. As a result of these requirements, the following components are needed in a compatible MEC system:

- Network adaptor that supports a high-speed connection to the cloud gaming data center
- Compute capacity with virtualization support
- Graphics adaptor that supports sharing between multiple VMs

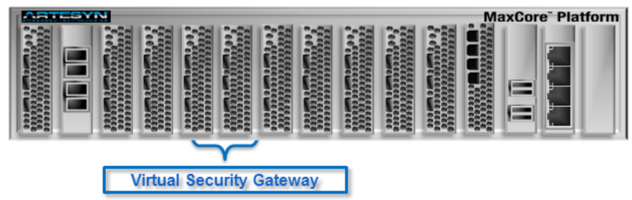
On the backend application, there is a need for high capacity compute to run the connectivity snippets connected to the front-end rendering applications, the game logic and related action and physics software modules.



Localized security

Security is important to customers. It is also important to operators, from a different perspective. As all customer traffic coming from and going to the edge is packaged in personal VPNs for forwarding into the core network, this also opens additional opportunities, such as an operator-provided firewall that guards the edge device from threats coming from the internet. Another interesting business case might be a user-configurable firewall that removes the need to run a firewall in the edge device, freeing cycles, or enabling/disabling parental control filters.

Security Gateway Use Case

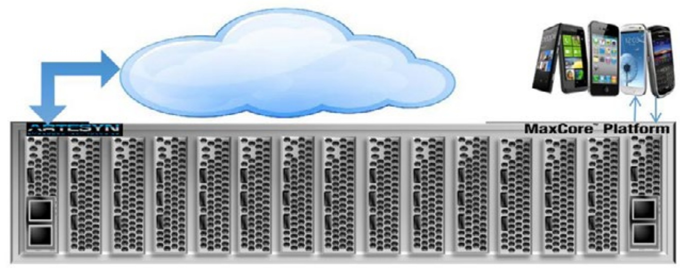


- Clavister Residential Security
  - Multi-tenancy use case
- Clavister Security Software running on Intel Xeon-D microServer cards moves security to Edge
- Enables service delivery with no Network Core action

Test Results with 100 vSGs		
Packet Size (Bytes)	Total throughput (Mbps)	Throughput per VSG (Mbps)
1448	1800	18
512	700	7
64	340	3,4
Note: CPU usage 60%		

The firewall will be implemented as a personal VM running on an MEC device, close to the edge, avoiding unnecessary traffic going back and forth in the network. The actual firewall uses deep packet inspection, possibly using a deep packet inspection development kit, or DPDK inside the virtual machine. To implement this configuration, which requires a large number of x86 cores running in a networked chassis with high bandwidth connected to both the edge and the core network, the following configuration is suggested:

- 2x network adaptors supporting high speed (100G) interfaces and enough virtual functions to enable lots of virtual machines running firewalls in the system.
- Compute nodes with enough cores to process a high number of virtual machines executing DPI. These are preferably in multiple complexes to better protect VMs from each other, while still taking advantage of high density.
- A system that enables easy and convenient use of the virtual functions to avoid waste of CPU cycles distributing the packet streams.



Distributed video delivery, analytics and optimization

Video is a major strain on today's networks. Especially if people attempt to watch full HD or higher resolutions on devices that are not fit for this, this results in many unhappy faces across the value-chain:

- Service provider networks get swamped with data unnecessarily
- Air interfaces are over-utilized, preventing other customers from successfully connecting to the internet and creating paid-for revenues for the operators
- High data usage, depending on the contract, results in early expiry of data packets, creating customer dissatisfaction, resulting in churn
- Re-framing and resizing the video results in high CPU loads on the handheld device, draining the battery early. Again, this results in dissatisfied customers and more churn.

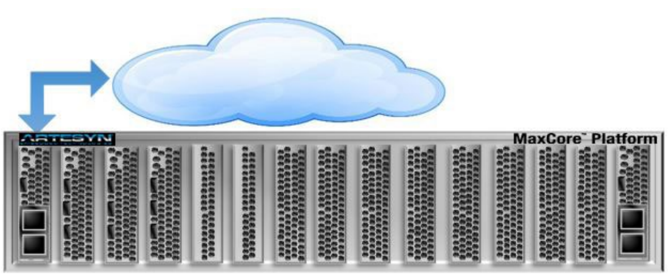
On-edge video optimization enables operators to remove these causes of concern, improving both service provider and customer satisfaction in parallel.

Additionally, analytics run on the videos watched enable opportunities for better targeted advertising, possibly even inserting ads into the video stream as it is delivered.

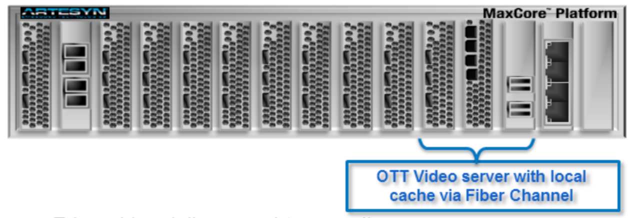
An additional option would be localized storage for highly demanded content, such as brand new videos, local video due to special local events such as concerts or sports events, or other content.

To build a system that delivers all this functionality in a single enclosure, a MaxCore platform should be populated as follows:

- A high bandwidth network adaptor card
- A number of processing units to support the database searches and playouts from local storage cards, one per processing complex
- 2-4 transcoding cards per processing card for video analysis and optimization, depending on the focus of the appliance



OTT Video Use Case



- Edge video delivery and transcoding
- Mix of Intel Xeon-D server and E3 w/ GPU or i7 w/GPU accelerator cards
- Fiber Channel array provides massive, scalable local storage
- Moves high volume video to Edge – key for Enterprise Video Surveillance
- Reduces strain on network due to video traffic
- Enables local caching of local preferred content

Conclusion

Multi-access edge computing (MEC) offers many opportunities today for operators that want to improve their revenue stream, bolstering ARPU while lowering the load on their networks.

Artesyn's MaxCore™ family of platforms and additional cards together with the growing ecosystem delivering application-specific VNF offerings that are guaranteed to run on the platform is ideally suited to give service providers a running start into MEC deployments today, talking full advantage of tomorrow's technology to solve today's problems now.

Please contact Artesyn now to find the perfect product combination for your challenge.

About Artesyn Embedded Technologies

Artesyn Embedded Technologies is a global leader in the design and manufacture of highly reliable embedded computing solutions for a wide range of industries including communications, military, aerospace and industrial automation.

Building on the acquired heritage of industry leaders such as Motorola Computer Group and Force Computers, Artesyn is a recognized leading provider of advanced network computing solutions ranging from application-ready NFV platforms, server acceleration platforms, and add-in acceleration cards to enabling software and professional services.

For more than 40 years, customers have trusted Artesyn to help them accelerate time-to-market, reduce risk and shift development efforts to the deployment of new, value-add features and services that build market share.

Artesyn has over 20,000 employees worldwide across ten engineering centers of excellence, four world-class manufacturing facilities, and global sales and support offices.

www.artesyn.com **computingsales@artesyn.com** **+1 888 412 7832**