2019

chetan sharma
CONSULTING

Edge Internet Economy

The Multi-Trillion Dollar Ecosystem Opportunity

Chetan Sharma

Table of Contents

Disclaimer

Chetan Sharma Consulting is one of the most trusted advisory firms in the global mobile industry. This research document presents some in-depth analysis about the future of the mobile industry. However, the author or the company assumes no liability whatsoever.

Any use or reprint of the material discussed in the paper without prior permission is strictly prohibited.

Introduction

In his timeless classic, "The Structure of Scientific Evolutions," preeminent historian of science, Thomas Kuhn defined "paradigm shift" as a *fundamental change in the basic concepts and experimental practices of a scientific discipline*. Computing and Communications have evolved over the many decades shaping the global economy. The Internet gave us a way to connect information and people. Modern mobile communication paved the way that collapsed time and distance. We are yet again at the precipice of a fundamental paradigm shift in Edge Internet that is going to redefine computing and communications to the core by empowering existing applications and enabling a new crop of breakthroughs, business models, and revenue streams. The next generation of applications and services will be "edge native" that are developed with the highest performance in mind, while ensuring lower cost of deployment.

The last ten years of industry growth was powered by LTE broadband, smartphones, and the centralized cloud infrastructure. The centralized cloud helped reduce the cost to launch new services and innovative startups who could scale rapidly without the need of tremendous amounts of capital. LTE broadband moved consumer's interaction with the Internet to the smartphone. Apps and commerce flourished, and we got a very thriving global ecosystem.

We believe that we are at the cusp of a seismic paradigm shift wherein computing and communications will move from the core network and a centralized cloud architecture to the Edge. It won't happen overnight, but it is inevitable. The reasons are manifold but the basic premise is that in order to serve the data, computing, and communications demand of objects, sensors, and people, resources, compute, and intelligence has to move to the edge to not only do it in the most cost-effective way but also to enable new use cases that just can't be supported by the traditional cloud architecture.

The long-wave economists have articulated the emergence of 30-40-year long technology cycles that have changed the face of the global economy over the last 250 years.[1] We are already on the Connected Intelligence cycle[2] that is powered by sensors and software. This cycle won't be possible without the shift to an Edge Internet enabled architecture that optimizes for both performance and cost at the same time.

It is daunting to analyze what the next 10-20 years of the technology evolution will bring to the global marketplace but if we understand the evidence of demand, trajectory of technology trends, insights into the new business models, innovation happening across the value-chain, we can piece together a vision that connects the dots.

To help us understand the Edge Internet roadmap, we must look at how Edge will integrate with legacy and existing infrastructure, applications and services that are already running and how Edge enables new classes of services that have been waiting for this fundamental paradigm shift to occur. This paper on Edge Internet looks at the architecture and investments required to make it happen. We will prove the disruptive power of the Edge with the help of field-data from real use cases and will formulate the thesis on why the mobile ecosystem would adopt the framework in their capex plans.

The use cases will look at the operational performance, costs, and deployment methods to help developers understand the power of an end-to-end Edge architecture. Given the distributed nature of the Edge Internet, enterprises, mobile operators, tower companies, and venue owners will play an important role in the build out. It will require new operating, business, and revenue models. This paper will delve into the business case of rethinking the computing and communications value chains that are Edge-led. Just like the App-Economy[3] turned the tech world upside down, we believe that over the next decade, the Edge-Economy will reshape the global technology ecosystem in fundamental ways.

[1] The three prominent economists have explored long-waves of technology revolution in their influential texts. In 1942, Joseph Schumpeter wrote Capitalism, Socialism, and Democracy. Kondratieff explored the economic cycles in his classic 1925 classic, "The Long Wave Cycle." Carlota Perez provided a more modern analysis of the thesis in her book, "Technological Revolutions and Financial Capitals (2002)"

[2] We have explored the Connected Intelligence cycle in more details here - http://www.chetansharma.com/connected-intelligence/ We believe, the global economy is a new 30-40 cycle powered by sensors and software that will change the trajectory of every industry vertical

[3] Sizing Up the Global Mobile Apps Economy, Chetan Sharma Consulting, 2010

Why Edge? Why Now?

The notion of edge computing has been around for 30-40 years. The dial-up era saw its edge in placing the modem banks closer to the user, the Internet era saw placing content delivery network nodes to reduce the RTT (round trip time) of web object requests. In the mobility/smartphone world, there is a lot of data processing that takes place at the RAN (Radio Access Network) before it is further passed on to the core and the Internet. So, we have been doing some form of "edge computing" for a long time, however, if you consider how the technology environment around us has changed over the last decade and what's to come, it will become clear that the Internet needs an architectural upgrade that is fundamental to efficient operations and improved profitability.

In 2007/8 when iPhone, Android, and the Appstore were just getting started, the cellular data traffic in the US was roughly 86 Petabytes/year.[4] Now, we consume that amount of traffic in less than 18 hours. Looking a decade forward, the same data traffic will be consumed in less than 10 minutes. The number of connected devices has risen 200% over the last decade and the growth is accelerating for the next 10. This is just from the growth in applications and services we already know and understand.

There is a new class of applications, both in consumer and the enterprise spaces, that are emerging which will require tremendous amounts of compute and communication capability that the current centralized cloud will not be able to serve. The last cycle was smartphone and consumer driven. While new consumer applications like VR streaming, cloud gaming, connected car entertainment and safety will demand new technology and business model innovations to scale, it is the enterprise where the demand for agility, high fidelity will be the strongest.

If we just consider the connected car segment, vehicles generating 20 GB/mo will create a 60 PB/mo stream[5] to the cloud which will be impossible for the centralized cloud to handle, never mind the cost. In this case, a multi-tier network architecture is needed wherein the vehicle provides the initial computing capability which interacts with the localized distributed edge that in turn interfaces with the centralized cloud. In this case, the local edge will process the data, the network edge will process the insights, and the centralized cloud will process the intelligence. In some cases, the centralized cloud will not be necessary at all as the orchestration layer will distribute the intelligence to other network nodes without ever touching the cloud. The centralized cloud in such a scenario will be primarily used for the storage of intelligence and wisdom from the transactions for future use.

Similarly, the data generation in hospitals is expected to reach over 1 TB/day and in factories almost 1 EB/day.[6] Obviously, only a small fraction of the data produced needs to be transmitted but it needs to be processed nevertheless, sometimes in real-time, and the current architecture is inadequate for handling the new demands and the new opportunities.

[4] Source: Chetan Sharma Consulting, 2019

[5] Automotive Edge Computing Consortium, General Principle and Vision, Dec, 2018, https://aecc.org/wp-content/uploads/2019/04/AECC_White_Paper_v2.1_003.pdf

[6] Source: Intel, 2018-19

In the enterprise, the move to the edge will be quite dramatic. Gartner estimates that in 2018, only 10% of the enterprise-generated data was created outside of the centralized cloud. This figure will flip to 75% by 2025.[7]

Coming back to the consumer demand and the industries that directly interact with consumers like advertising, retail, media and entertainment, commerce, and others; given the number of sensory touchpoints is accelerating at a feverish pace, we need better ways to capture and process data that creates meaningful outputs for further use in different scenarios.

Additionally, there are new requirements due to data sovereignty laws, competitive dynamics, consumer preferences for data handling, ethical filtering of data, and more that will necessitate a new way for handling data at each node of the transmission process.

All this leads to the creation the "**Edge Internet Economy**," a framework that is not only dependent on the Edge but also the assembling of the platform that will become the foundation for the next generation of "new" trillions of dollars in global economic activity.

Edge Internet is an overlay architectural upgrade to the infrastructure and is independent of the access mode. While technologies such as 5G only make it better, Edge Internet architecture works in-sync with 3G/4G, with WiFi and unlicensed cellular, with fixed and fixed wireless.

Edge Internet is not just an extension of the current cloud architecture but rather a new and necessary programmable computing paradigm to meet the demands of current and future applications and services. There are use cases such as esports arcades or manufacturing facilities or greenhouses where the Edge Internet infrastructure is going to be quite different, perhaps self-contained, and won't necessarily need access to the real-time macro network.

Thus, there are numerous existing and new use cases where the end-to-end architecture will be different based on the requirements of cost and performance. The mission statement for Edge Internet is *autonomous optimization of the cost economics and the network delivery performance at the same time*.

In the next few sections, we will go deeper into the details of the Edge Internet and how it enhances current use cases and enables new ones.

[7] https://www.gartner.com/smarterwithgartner/what-edge-computing-means-for-infrastructure-and-operations-leaders/

Computing Nirvana: Autonomous Optimization of Cost Economics and Network Performance

The factor that determines if a trend will grow at a linear scale or an exponential one often boils down to one factor above anything else and that is *"delivery economics."* Sometimes the optimal balance is delivered by technology while the business models play an important role in others. Let's look at what Jio in India was able to do in the last couple of years. When it came to data services and consumption, India wasn't considered the leading market. In 2014, despite having three times the number of subscribers, the total data traffic in India was one-fourth that of the US.[8]

However, by 2018, the data traffic had twice that of the US. What dramatically changed in the intervening years was the emergence of Jio which built a very economical all-IP LTE network at a fraction of the cost and was able to change the "delivery economics" by offering unlimited data for approximately $2/month. The mobile industry has never ever seen such a meteoric rise of data traffic that has sustained the surge of consumer growth over a long period of time. Now Jio carries more data traffic than all the US mobile operators combined.[9] Even on a per subscriber basis, Jio is ahead of all major large operators in the world.[10]

Through systematic engineering, apt business model transformation, and ruthless execution, Jio was able to redefine the economics of the business. We have seen similar impacts from Facebook, Google, Amazon, Tencent, Paytm, Airbnb, Uber, and others.

Often times, disruption is driven by the change in economics. What if we could change the **"network delivery economics"** by a factor of 10 or 100, then, suddenly, the value-chains will be reorganized for efficiency and innovation. The applications and services that were bogged down by the burden of cost will feel liberated by delivery economics. The cost of sequencing a genome dropped from $100M in 2001 to less than a $500 by 2020[11]. This has had a profound impact on our understanding of medicine and how it is applied to humans. Edge Internet has the potential to have similar effects on different industries ranging from advertising to manufacturing; from entertainment to transportation; and from cybersecurity to agriculture.

Some argue that 5G will be sufficient to address the Edge use cases and one doesn't really require a new architectural design or that for indoors WiFi is more than enough. However, such arguments miss the central premise of Edge Internet which is that it is independent of the access technology. 5G needs Edge more than Edge needs 5G. Edge empowers WiFi to expand to new use cases. The fundamental design change is that through programmable virtualization, computing and communication resources become available closest to the requesting node in real-time while optimizing for delivery economics.

[8] Source: Chetan Sharma Consulting, 2019
[9] Source: Chetan Sharma Consulting, 2019
[10] Source: Chetan Sharma Consulting, 2019
[11] Source: Wetterstrand KA. DNA Sequencing Costs: data from National Human Genome Research Institute, https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data The data is available till July, 2017. Author has extrapolated the trend to end of 2020.

Let's just consider 5G case for the sake of argument. The efficiencies in 5G help reduce the cost per GB to one-tenth but the traffic is expected to grow 10-20 times more as a result of better network and device capabilities.[12] Thus, the same network design topology even with increased number of small cells will not be able to profitably deal with the 5G traffic unless new software techniques are applied to the problem.[13]

As we alluded to before, Edge Internet doesn't mean there is a box sitting in front of every consumer or a server at every base station but rather through the magic of programmable virtualization and orchestration, computing and communications resources can be made available to meet the performance requirement of the application and cost constraints of the business to move the bits from point A to point B. The future of applications is immersive, interactive, and personalization to the n^{th} degree.

By having an architecture that can deliver these resources on demand, we can not only enhance user experience of the current applications but also pave the way for new ones. Sometimes, the edge is the device itself. Google, by using a federated distributed machine learning model instead of a centralized cloud training model was able to reduce the 100 GB voice recognition model down to just 0.5 GB on the device which resulted in serving up the answers from Google Assistant 10 times faster.[14] We have seen similar results from other use cases that we will discuss throughout the paper.

Challenges

In 2018, the top three centralized cloud companies Amazon, Microsoft, and Google spent $68 Billion in Capex. Collectively, these three have spent more than $270 Billion over the last twenty years with 43% of the spend coming in the last two years. At the highest levels, building a centralized cloud model is straight forward, we rely on a few companies to invest heavily to build, operate, and offer cloud services. There are roughly 100-150 datacenters amongst the top three cloud players servicing the entire globe.

Edge by its nature is very distributed and involves many players – tower companies, mobile operators, service providers, cable companies, infrastructure vendors, city, state and national governments, site owners, enterprises, and more. How do we coordinate, develop, and deploy software that manages the data at unprecedented scale? How do we orchestrate the resources, demand, and intelligence between the widely distributed nodes? How do we interface with legacy be it the LTE networks or Azure and AWS? How do we apply the data regulations and the data residency laws in real-time? How do we preserve data, insights, and intelligence and where? Who makes the determination of the data flows and why? Will we need new laws for ephemeral data? Will there be new opportunities for transitory data?

It is inevitable, a disruptive notion like Edge Internet will raise these and many more questions. However, just like the industry has done in the past, we figure out the solutions. Some are

[12] Source: Chetan Sharma Consulting, 2019

[13] Will Operators rise to meet the challenge of 5G? Chetan Sharma, Fierce Wireless, 2019, https://www.fiercewireless.com/wireless/industry-voices-sharma-industry-s-5g-moment-will-operators-rise-to-meet-challenge

[14] Source: https://syncedreview.com/2019/05/07/i-o-2019-your-data-stays-on-your-phone-google-promises-a-better-ai/

technical challenges, while others are regulatory and business issues. New governance models will be needed to manage the integrity of the process. New toolsets and APIs will enable developers to onboard their applications on to the Edge platform.

In the next section, we will delve into the lifecycle of a *bit* and look at the architecture required to address the Edge Internet opportunity.

Connecting the Dots Between Performance and Revenue

From the early days of the cloud, the researchers and the practitioners of the trade understood the relationship between performance and revenue. Noted researcher, Greg Linden wrote that experiments at Amazon showed a 1% sales decrease for an additional 100msec,[15] and that a specific experiment at Google, which increased the time to display search results by 500 msecs reduced revenues by 20% (based on a talk by Marissa Mayer at Web 2.0). If time is not directly part of your overall evaluation criterion, make sure that a new feature that is losing is not losing because it is slower.[16] In 2009, to convince the eCommerce world, Akamai concluded that the threshold to response times acceptability was 2 seconds.[17] Over the last decade, we have clearly understood the relationship between response times to drop off rates to conversions that yields real dollars.

Under the current Internet architecture, we are hitting the limits of how many milliseconds we can shave off from end-to-end latency. Edge Internet architecture by eliminating the need to go to the cloud is one of the best ways we can reduce RTT. The Internet traffic has been steadily going mobile. In markets like India, mobile is over 50% of the overall IP traffic. In such markets, Edge Internet will become a "must-have" to manage costs and performance to just scale the existing demand and applications.

Many new applications that require tremendous amount of compute and low latency, Edge Internet provides an efficient onramp to the existing Internet architecture. To be successful and have broad adoption, Edge Internet must straddle the world of old with the new in a graceful manner.

On the flip side, using edge, new business models can be created that were not possible before in different industries. For example, a tier-1 mobile operator was able to use Edge Computing to convert regular streaming subscription to HD streaming subscriptions because Edge helped manage congestion better and thus increased capacity of the RAN to handle more concurrent streams. There was an uplift of ARPU by 30%.[18]

Similarly, Edge-enabled cloud games could be a new revenue stream for mobile operators. Multiplayer cloud games, AR-games, VR-games, all require consistency in low latency which can be only managed through an edge architecture that manages the session, data processing, and interactivity with the players involved. The market for such applications and services is big.

Later in the paper, we will take a look at the potential market size of the Edge Internet Economy.

[15] Practical Guide to Controlled Experiments on the web, Kohavi, Henne, Sommerfield, 2007, https://ai.stanford.edu/~ronnyk/2007GuideControlledExperiments.pdf
[16] The Cost of a Cloud, Greenberg, Hamilton, Maltz, Patel, 2007, https://mvdirona.com/jrh/TalksAndPapers/CostOfClouds_CCR.pdf
[17] https://www.akamai.com/us/en/about/news/press/2009-press/akamai-reveals-2-seconds-as-the-new-threshold-of-acceptability-for-ecommerce-web-page-response-times.jsp
[18] Source: Chetan Sharma Consulting, 2019

The Edge Internet Architecture: Future-Proofing the Investments

The Edge Internet architecture provides benefits on some key metrics such as: bandwidth and network congestion, compute availability, network resource and intelligence, data localization and regulations, but the two factors are most critical in understanding why Edge Internet architecture is inevitable are **latency and delivery network economics**. Latency is key to service delivery and performance. Furthermore, at the end of the day, the success of any new technology comes down to delivery economics.

Latency

Latency is an important metrics that engineers in the ecosystem watch with great intensity because it impacts performance which in turn changes the revenue profile. As we discussed earlier, even a slight improvement in latency can yield significant revenue gains. We have understood this notion for at least two decades. But the traditional Internet architecture is reaching its limit, especially on the network latency front. Hence, the urgent need of an Edge Internet framework that fundamentally alters the latency equation thus enabling existing applications and services to gain performance boost and unlocking the potential of future services.

Latency is caused by several factors as illustrated in Figure 1. There is an inherent delay on the end-device that processes information due to screen and sensor response times and screen refresh in case of streaming. Then, there is delay due to I/O at the source to pull and package the information to be transmitted. But the bulk of overall latency comes from the network which depends on the transmission medium and its propagation characteristics, routers and nodes and how fast they process and route packets from point A to point B.

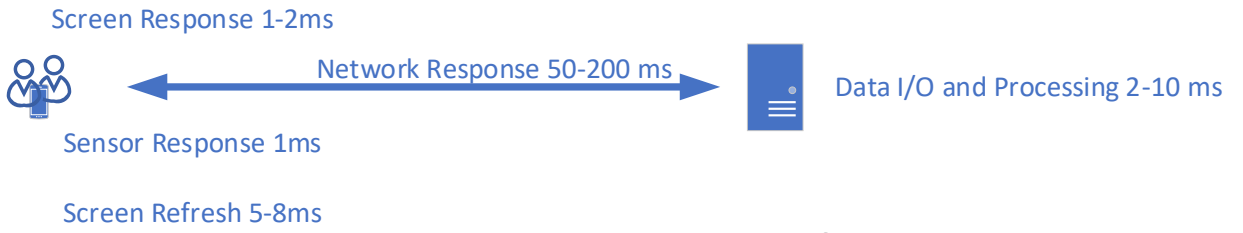

Figure 1. The Anatomy of Latency[19]

It is the network latency we are focused on in this paper and more specifically the wireless network latency. If we further distill down into how the latency is broken up in between different nodes, each node means additional time for RTT. The overall average RTT in a 4G network is roughly 50-100ms, sometimes as high as 200ms. There is delay from the user device to the tower which with 5G comes down significantly to the order of only 1-5 ms but the remaining

[19] Source: Chetan Sharma Consulting, 2019

latency components of an E2E network must be still dealt with. The overall network latency is still 20-30ms in 5G. The Edge Internet architecture helps take out two-thirds of the delay without impacting the traffic whether it is 4G or 5G.

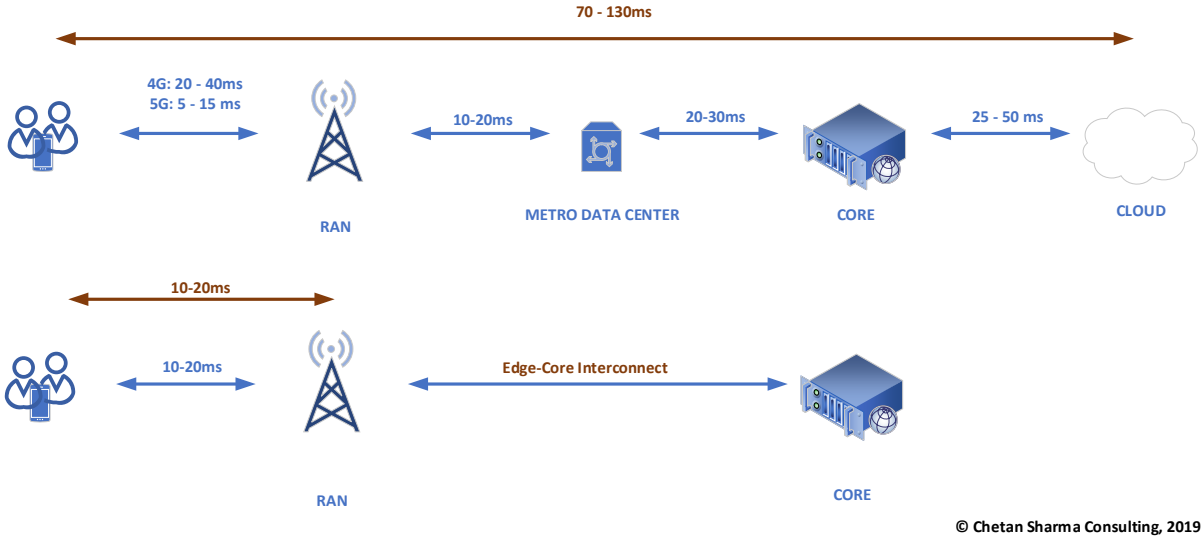

Figure 2. a) Legacy Architecture b) Edge Internet Architecture[20]

Thus, the only way to effectively reduce E2E latency is using an Edge architecture.

Figure 2. further expands on how such an architecture would work in a wireless network. It should be mentioned that there are numerous scenarios wherein a macros cellular network is not needed such as applying computer vision in a manufacturing plant where through wired infrastructure or WiFi, one can attain the desired results as it is a controlled environment. Eventually, it comes down to the use cases and the requirements.

The notion of Edge Internet expands beyond cellular, however. From an architectural perspective, it is independent of the underlying access protocols. For example, if a Netflix show has to be streamed to a mobile device, it can be done from Netflix servers over the cellular network, from servers in the operator RAN, from access point inside the user's home or a hotspot in the neighborhood. The Edge Internet orchestration process plays a key role in figuring out what content needs to play where to optimize for cost and revenue. Similarly, for components of the app, infrastructure could reside near the user/enterprise depending on the needs. The orchestration server can work with any Edge node, even a smartphone that is able to share its resources for the purposes agreed upon by the application and the network.

In cases, where cellular access is needed, the key technology to make it work end-to-end is "Edge-Core-Interconnect" that keeps the existing network working without any modifications (Figure 3). The interconnect is used to update the records at the core with needing to be involved real-time in transactions happening at the Edge. This keeps the resources at the core focused on other traffic and important tasks.

[20] Source: Chetan Sharma Consulting, 2019

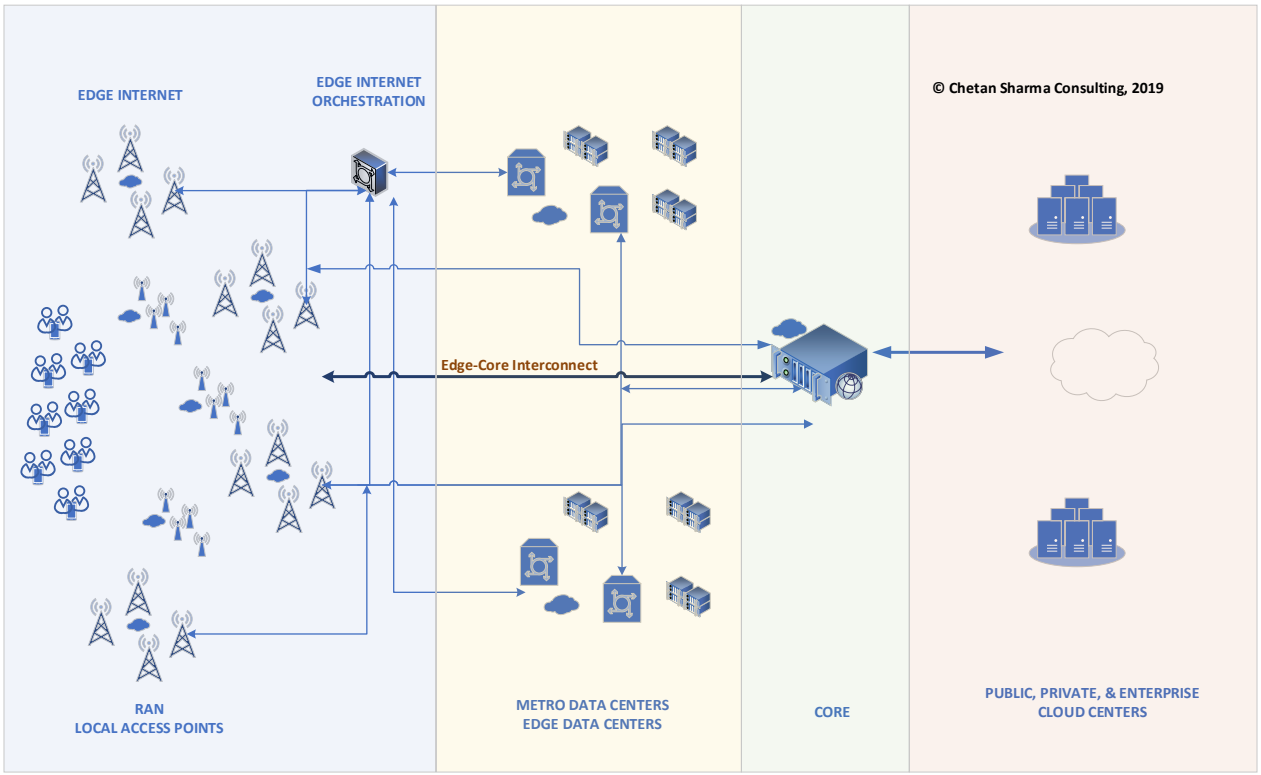

Figure 3. Edge Internet Network Topology[21]

This type of architecture will work across new technologies – upgrades to WiFi and 5G/6G, etc. It is also not intrusive or disruptive to existing operations which is a key factor for such solutions to scale. A greenfield operator like Rakuten Mobile can deploy it with relatively the same ease as an incumbent like KDDI.

Network Delivery Economics

Each network evolution generation is defined by what we choose to optimize for. As we prepare for 5G, the success of operators will be defined by "network delivery economics."[22] How cost effectively can a bit be delivered to the end point will be the single most important metric that a mobile operator CEO should watch and optimize the business for. If they get this metric wrong or if their cost basis is significantly worse than their domestic competition, they are likely to see precipitous decline in ARPU and net revenue. France and India are perfect case studies of such a scenario becoming reality. We think that the network economics thesis will be universally applicable as operators embark on the 5G journey.

[21] Source: Chetan Sharma Consulting, 2019

[22] Source: 5G Economics, An Economics Based Analysis of the Future of Wireless Network, Chetan Sharma Consulting, 2016

Generation	Engineering Focus
1G	Bit/s/Hz
2G	Bit/s/Hz
3G	Bit/s/Hz/m2
4G	Bit/s/Hz/m2/joule
5G	**Bit/s/Hz/m2/joule/$**

Figure 4. Engineering focus of evolving network generations[23]

As Figure 4 indicates, the optimization vector for 5G will be Bit/s/Hz/m^2/joule/$. How should an operator plan their network so that they have enough left to invest in the upper layers of the stack that are generating the most revenue. A great network is obviously table stakes but if all the energy of the organization goes in making the network work and none on the application layer, they will march towards becoming a utility, perhaps a profitable utility, but utility, nonetheless.

Figure 5 plots out the revenue and cost per GB in the US market. As the traffic has increased rapidly over the last decade, the margins have been shrinking. Even with 5G, though it can yield 10x cost reduction over the deployed period, it is not enough to compensate for the growth we are likely to see as a result of faster networks, improved devices, and new form factors and interfaces.

So, how does the operator deal with this juxtaposition?

[23] Source: Chetan Sharma Consulting, 2015, Adapted from "5G Network Capacity: Key Elements and Technologies," IEEE Vehicular Magazine, 2014

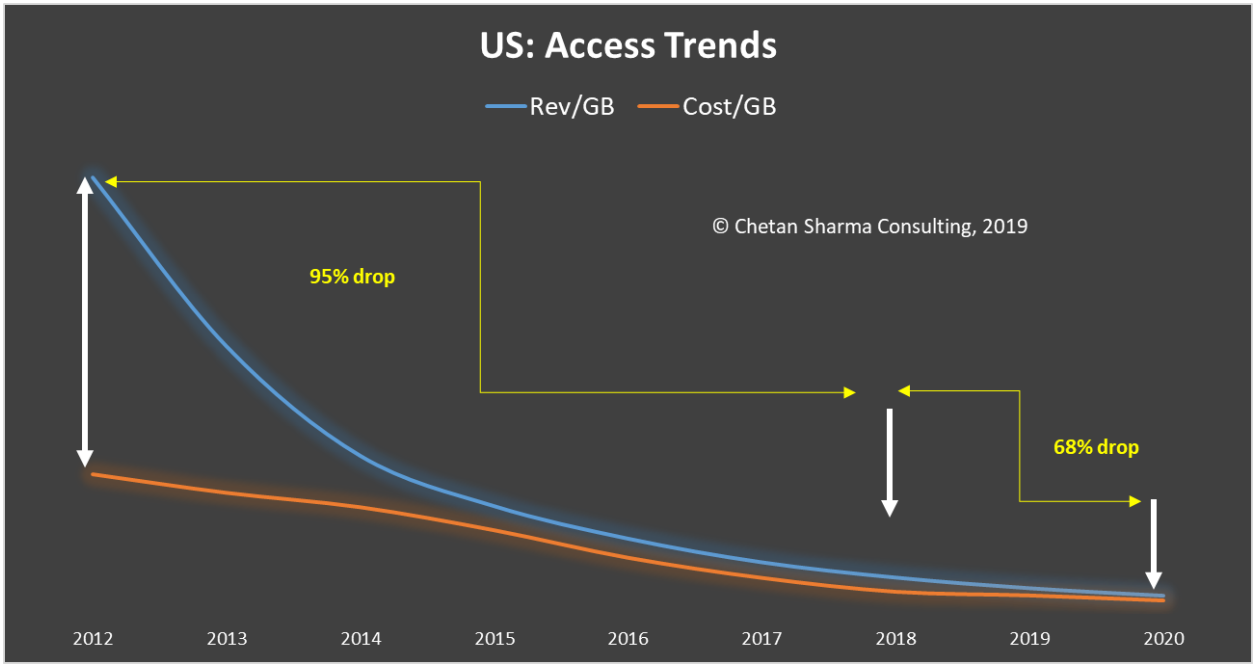

Figure 5. The decline in Revenue and Cost of GB in the US wireless market[24]

We have argued that without a software-led approach to capacity management and network delivery economics, operators are going to find it extremely challenging to keep the costs down.[25] This is where the Edge Internet architecture comes into play. It not only helps in lowering the Capex but also increases the revenue potential. Combined, this can have a multiplier effect on the business and market perception of the value of the investments.

We will prove out this thesis in the next few sections using use cases from the field.

[24] Source: Chetan Sharma Consulting, 2019

[25] Will Operators rise to meet the challenges of 5G? Chetan Sharma, 2019, Fierce Wireless, https://www.fiercewireless.com/wireless/industry-voices-sharma-industry-s-5g-moment-will-operators-rise-to-meet-challenge

Sizing up the New Edge Internet Economy

When the app economy was in its infancy in 2008-9, it was difficult to envision the exact trajectory the ecosystem will take but it was quite clear that developers will latch onto the open development and distribution model of the app store that had been long shackled by the telecom industry. The switch in the revenue model also helped accelerate the growth in the ecosystem.[26] Mobile operators lost their share of the market almost overnight. Thus, a combination of new tools (app store) and the new business models (revenue share and app distribution) helped enable one of the most creative periods of the last two decades.

We think that we are on the verge of something similar with Edge Internet playing a key role in the development of next generation applications and services. To really understand the potential of the total Edge Internet Economy, we must look at the applications and services in two overlapping phases:

- **Existing Use Cases** – Use cases that are already in play but will benefit from the Edge Internet architecture

- **New Use Cases** – Use cases that are not possible without an Edge Internet architecture in any meaningful way

It is important to understand these use cases so we can determine what Edge Value (EV) does Edge Internet bring to the ecosystem. This will help us in understanding the overall Edge Internet Economy (EIE).

Let's briefly look at how Edge can play a role in enhancing existing use cases and empowering new ones.

Existing Use Cases

When many executives in the industry or the press talk about Edge Internet use cases, the discussion immediately turns to AR/VR and autonomous vehicles making it sound like the use of Edge is something for the future. However, as we mentioned earlier, Edge Internet is an architectural upgrade to the mobile Internet which is independent on the access technology. As such, Edge Internet can work very well with existing use cases to solve current performance and business model problems in various industry segments. As such, there should be more emphasis on Edge Internet enhancing current applications and services or solving some thorny problems in the ecosystem that are performance-constrained like the ad-fraud problem in advertising or network congestion in telecom networks or computing/communications constraints in agriculture. We will discuss some of them in the section below.

Advertising

Advertising is more than half a trillion-dollar industry. Digital forms half of that spend and mobile has increased its share to more than 50-60% of that digital spend. However, if we look

[26] Sizing up the Global Mobile Apps Market, Chetan Sharma Consulting, 2010

closely, there is a lot of inefficiency in how the demand is auctioned, how location and profile data is attained and used, and this creates a lot of fraud and inaccurate metrics in the marketplace. Furthermore, as mobile advertising has moved to video-centric delivery of ads and engagement, the delivery of that video becomes critical to the metrics such as startup time, completion rates, engagement, quality of video, etc. In the case of auctions, the exchanges have 10-15ms to fulfill the demand and many times, the ad servers aren't able to fulfill the demand in time. Same thing happens for location data, the servers can't return the API call for location, profile and targeting data within the time allocated. What if the queries are speeded up by 100-200% or what if you could run the whole exchange on the Edge so that latency is cut down by a significant factor such that ad servers are able to compete more effectively for each ad request.

Similarly, by cutting down on the video delivery latency and the congestion, the startup time for the video will decrease, completion rates will see improvements, all of which translates into much better engagement metrics and revenues for the advertising ecosystem. Thus, by applying Edge Internet strategies to existing workflow process in advertising, one can improve the yield and provide better user experience.

We will discuss the advertising use case in more detail with real data to give you a sense of the degree to which an Edge Internet architecture can make a difference.

Mobile Network Operations

Whenever there is an introduction of a new network technology or a new class of devices, there is a step change in network traffic. This is especially problematic for managing networks in dense urban areas where most of the network congestion takes place. In fact, the average 4G network throughput of 30-40 Mbps could crawl down to less than 1 Mbps under these conditions making the user experience terrible. Operators are forced to spend more capex and opex on upgrading sites, splitting cells, and buying more spectrum.

Edge Internet provides a more cost-effective way of managing network traffic wherein some of the network can be processed and offloaded at the RAN without ever touching the normal route through the core. With the help of additional software techniques and strategies, this load can be managed more cost-effectively. So, there is a side benefit of deploying Edge Internet architecture for the operators – in addition to new revenue streams, it can help with the existing operations. As previously mentioned, Edge Internet doesn't really need 5G to be successful (though it can take advantage of it). The layered architecture can be applied to any network access technology such as 3G and 4G.

New Use Cases

Just like Uber or Facebook were sparked by the dual S-curves of smartphone and LTE Broadband growth, we expect the intersection of Edge Computing with new application areas such as AR/VR, Holograms, Autonomous, AI, etc. will have tremendous potential for new services and revenue streams. Entrepreneurs always figure out ways to exploit the performance gains as a result of new technology, architectures, and business models. There are several use cases that are being tested out in the industry, some are more near-medium term, others will take longer-term to scale. We will discuss one of each to give you a sense of what's possible and why Edge plays an important role in these scenarios.

AR/VR and Multiplayer Games

The new interfaces of AR/VR/MR have started to enter the consumer lexicon. We are still a ways away from any mass adoption and scaling the category but we can see that the experience can provide some remarkably new set of experiences for gaming, immersive entertainment, and new workflows in enterprise use cases (Figure 6). The requirements for computation and communications are enormous. Just doing some basic math, for VR, each of the human eyes can see up to 64 million pixels at any given moment, and with 120 fps requirement to generate a real-life view, we are looking at 15.5 billion of pixels/second. By storing each colored pixel in 36 bits and with 1:600 video compression available in H.265 HEVC encoding, we are looking at the consistency of 1 Gbps to guarantee acceptable user experience.[27] Additionally, 5ms latency is required to ensure high resolution and undistorted vision. These kinds of requirements are impossible to fulfill with legacy Internet and Communications architectures. Furthermore, in a gaming scenario, when multiple users might be attached to the same basestation or access point, the infrastructure must handle congestion and resource management.

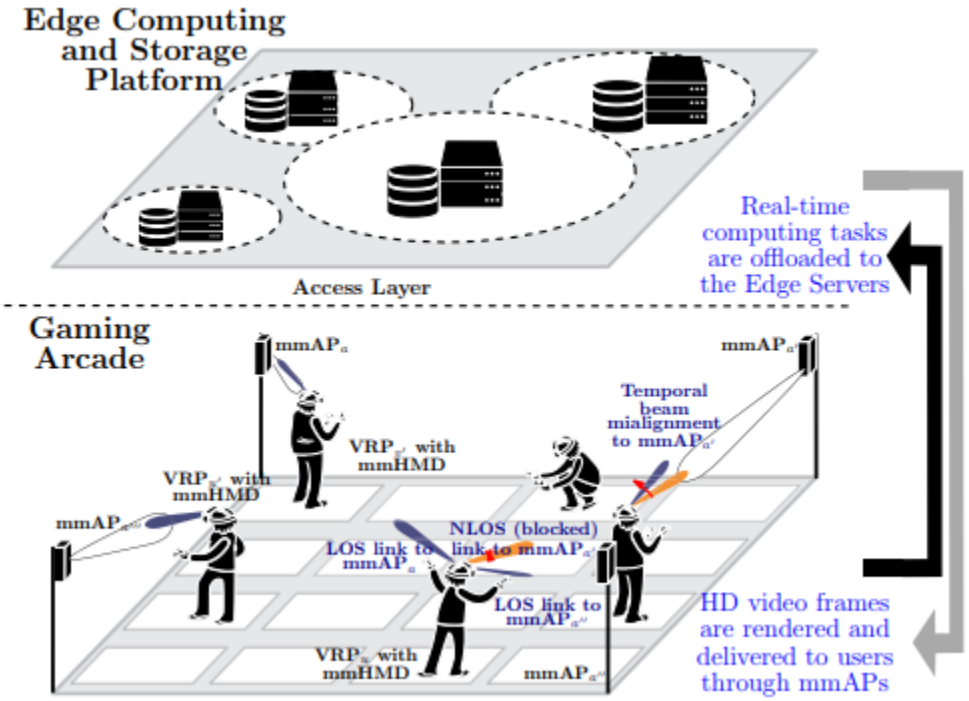

Figure 6. VR and Edge Computing[28]

In multiplayer gaming scenarios, there is even more complexity in synchronizing what's displayed on screens at the same instant or what network resources are used (such as location or profile) within 1-2 ms.[29] In new category such as VR Arcades which will invariably start mushrooming once we have the technology nailed down, we will need to use advances in

[27] Towards Low-Latency and Ultra-Reliable Virtual Reality, Elbamby, Perfecto, Bennis, Doppler, University of Oulu, Bell Labs, 2018

[28] Towards Low-Latency and Ultra-Reliable Virtual Reality, Elbamby, Perfecto, Bennis, Doppler, University of Oulu, Bell Labs, 2018

[29] In VR, the network latency shouldn't be more than 7-8ms to avoid motion sickness

network technology, Edge Internet and computing, storage, data I/O and synchronization, and error correction, congestion management, and latency optimization.

It is inevitable, that these experiences will find their way into outdoor environments as well through the use of normal looking glasses which will be a platform to enable new forms of interfaces and engagement. Further, lowering the latency to less than 5ms will enable new multi-sensory tactile control of objects and information.

In the enterprise environment, the integration of application workflow to enable new gestures to sift through tasks and data will become common place. The amount of information that can be overlaid on the physical objects to help in industries such as manufacturing, medicine, construction, and education is just staggering. This will fundamentally alter how we process information and intelligence. However, all this requires both compute and communication at the Edge and a tight synchronization between the devices and the network.

Figure 7. Mixed Reality[30]

The growth in AR/VR/MR will give rise to new forms of occupations and skillsets such as VR surgery where surgeons wear headsets that are essentially microscope and a supercomputer mixed in one to help them navigate the human anatomy for better outcomes.

Holographic Entertainment

[30] Source: Microsoft, 2019

Music and cinema have always captured our imagination beyond the basic needs of entertainment. These sectors have always adopted technology, the vision of what could be much earlier than it becomes a reality. Star Trek inspired an entire generation of engineers to search for voice communications, interactive screens, artificial intelligence, remote surgery, holographic transmissions, and the list goes on. As the technology has advanced, some of the fictional setups have become real and consumers have helped make them mainstream.

SM Entertainment in Korea is a phenom and they have gained global stardom. It is the largest Korean entertainment company. They created Korean-pop or K-pop which is at the forefront of using technology to create art, characters, and stories. Their work has generated almost 29 billion views on YouTube, brought 17 million fans to their concerts, and work with composers and new artists worldwide. Technology is what enables and empowers them to be different and explosive. They have been using AR/VR and lately experimenting with creating Holographic imagery.[31]

In fact, they are looking at creating new genres using hologram and virtual characters who can create dramatically new entertainment experiences for the audiences worldwide.[32] The fusion of culture and technology requires new cutting-edge innovation and Edge Computing will be at the center of this engagement as fans look to engage with platform using mobile. Even the setup required in the arenas will require significant computation load to be handled at the edge with almost zero latency to make the holographic experience real. In fact, NTT DoCoMo showcased this concept at MWC in Barcelona in 2019. The coming decade will provide a ripe platform for new modes of immersive entertainment, the likes of which we have never seen, all enabled by new Edge Internet architectures and ecosystems.

[31] http://www.koreaherald.com/view.php?ud=20150126000821

[32] Author discussions with SM Entertainment

Use Cases from the Field

Earlier we mentioned that contrary to the popular belief, Edge doesn't need 5G though it will be equally applicable when current networks are upgraded to 5G over the next decade. All the data and use cases discussed below are generated from either existing service operating in the field or simulated on existing networks enhanced by the Edge architecture discussed above. We wanted to prove that the Edge Internet is applicable to all network architectures and that's why it is such a powerful disruptive force.

Advertising

To amplify the notion that Edge Internet enhances existing applications, we will discuss a real case study from the field that is yielding real results and new revenue. Advertising is the foundation of the today's Internet, especially the mobile Internet. By 2020, over 30% of the global advertising expenditure will come from mobile advertising which is twice of desktop advertising.[33] In fact, by 2021, mobile Internet will even surpass Television as the top source of advertising revenue in the global industry.

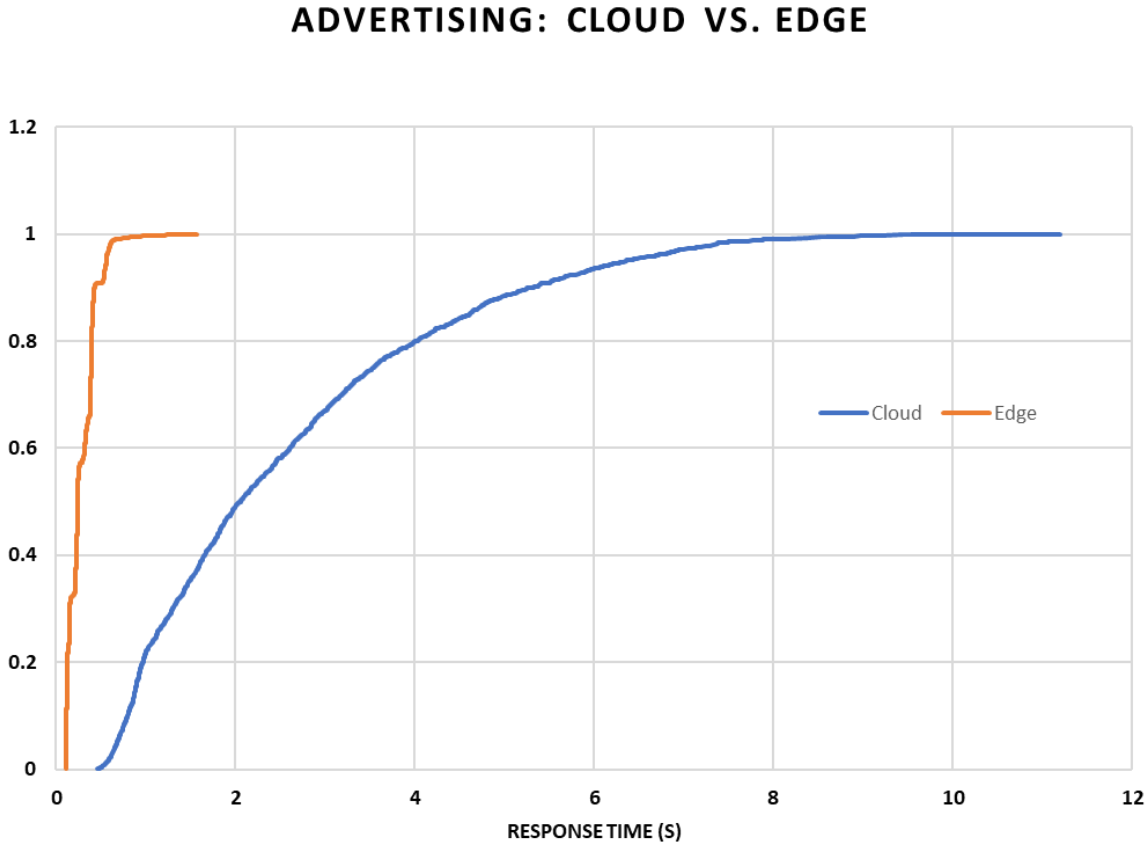

Figure 8. Latency CDF: Edge performance vs. the Cloud[34]

[33] Source: Zenithmedia, https://www.zenithmedia.com/insights/global-intelligence-issue-06-2018/mobile-share-of-advertising-market-to-exceed-30-in-2020/

[34] Source: Analysis based on Mobile Operator data

The tier-1 mobile operator in this case study was looking to solve a specific set of advertising problems. The advertisers wanted to stream video ads but congestion at peak times made the experience a poor one for the consumer which resulted in low customer engagement yielding low advertising revenue which is based on some specific performance metrics.

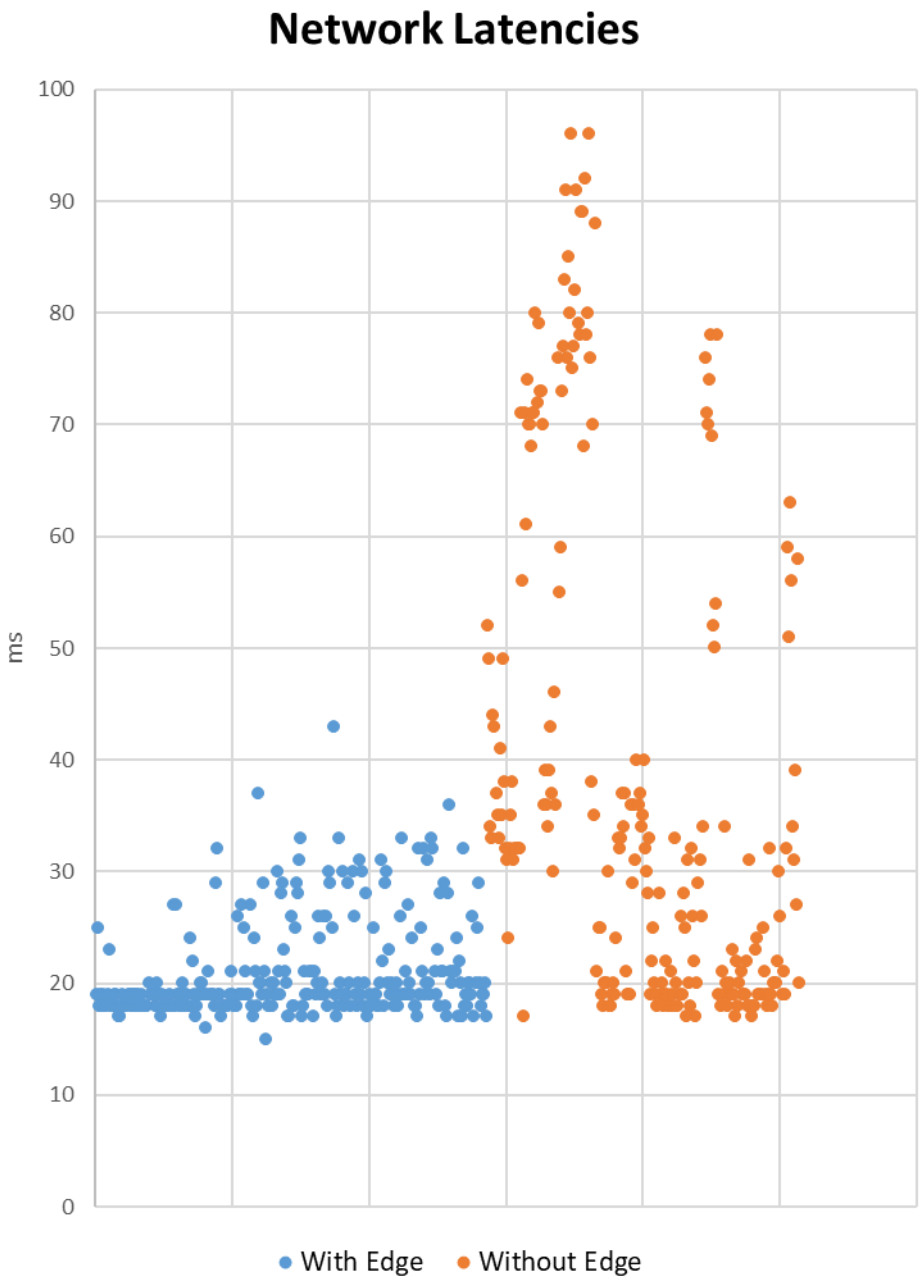

Figure 9. Network Latency: Edge vs. No Edge[35]

[35] Source: Source: Analysis based on Mobile Operator data

By overlaying an Edge Internet architecture and placing the advertising videos closer to the consumer and managing the complexity through Edge and APIs, the solution had a dramatic impact on the performance and revenue metrics.

Performance impact:

- The video resolution went up by 27%
- Data throughput increased by 62%
- Startup time reduced by 33%
- Video ad completion increased by 42%
- Playback time increased by 26%

Revenue impact:

- Revenue increased by 30%
- MAU metric went up by 39%
- Reduction in revenue erosion by 88%

Thus, by applying Edge Internet architecture on top of an existing access network, operator was not only able to manage its network much better but also started a new profitable revenue stream that has already yielded millions of dollars in revenue. This is not a futuristic use case but something that can be applied in the field today with immediate results.

Now multiply this by the number of use cases in different regions and you can see the scale of the opportunity that Edge Internet can unlock. It should also be noted that the reason the architecture played such an important role in this case was the significant reduction in latency (Figures 8 & 9) compared to the traditional cloud architecture. In this case, there was a *10-fold* reduction in latency by using Edge vs. the centralized cloud. The other key aspect worth noting is the consistency in performance is significantly higher for edge vs. the cloud.

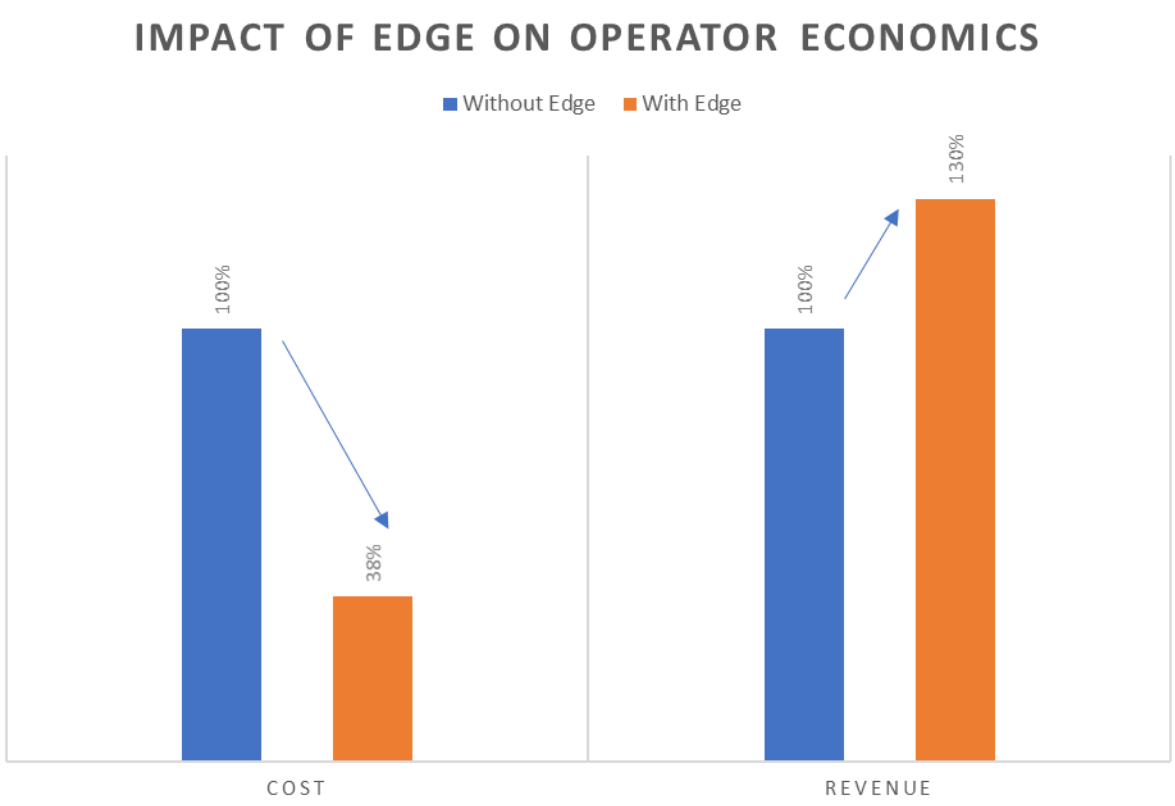

Figure 10. Impact of Edge on Operator Economics[36]

So, let's review how the use of an Edge Internet Architecture impacted operator economics. The cost of delivering the bits decreased by 62% and revenue increased by 30% giving a tremendous boost to net revenue for this case study (Figure 10). Furthermore, by becoming an OTT player on its own network platform, they can expand on their share of value capture.

[36] Source: Analysis based on Mobile Operator data

How Big Is the Edge Internet Economy?

To really understand how big the Edge Internet could become, we must split the analysis into two broad buckets of existing and new use cases and determine the value Edge can provide to each industry segment. We also must analyze where Edge Internet fits in with our understanding of the existing technology ecosystems. We have the experience of determining how cloud economics impacted the global economy, and more specifically the various industry segments.[37] We also understand how various industry verticals have benefitted from the rise of mobile broadband and smartphones.[38]

There have been several studies estimating the size of the impact of various technology cycles such as Internet, Mobile, and Cloud on the larger economy and the global GDP. Some of these studies have focused regionally like the US and OECD/EU where pertinent data for analysis is easier to model.

GSMA estimated the total contribution of the mobile industry to the global economy to be $3.9 Trillion with roughly $ 2.77 Trillion coming from indirect and productivity measures.[39] The Internet Association summarized the contributions of the Internet sector on the economies of the US and EU in the range of 5-6% of the GDP.[40] Even further back in time, researchers have quantified the impact of Internet on the global economy. BCG estimated a $4.2 Trillion Internet economy in 2016 in G-20 economies.[41]

We also have good insights into the long-wave cycles of industry change and disruption through the study of Connected Intelligence.[42] Using these inputs into the model, we have come up with a basic framework to analyze how the Edge Internet Economy is likely to grow as is illustrated in the equations below.

$$Edge\ Internet\ Economy^{43} = \sum_{i=1}^{n} Edge\ Value * Industry\ Segment$$

$$Edge\ Value = TAQ * MR * ER * MA$$

[37] Examining The Economic Contributions of The Cloud To The United States Economy, Internet Association, 2019, https://internetassociation.org/wp-content/uploads/2019/02/IA_Examining-The-Economic-Contributions-Of-The-Cloud-digital.pdf

[38] http://www.chetansharma.com/insights/

[39] Source: The Mobile Economy, 2019, https://www.gsma.com/r/mobileeconomy/

[40] Source: Refreshing Our Understanding of the Internet Economy, 2017, https://internetassociation.org/publications/refreshing-understanding-internet-economy-ia-report/

[41] Source: The Internet Economy in the G-20, 2012, https://www.bcg.com/publications/2012/technology-digital-technology-planning-internet-economy-g20-4-2-trillion-opportunity.aspx

[42] http://www.chetansharma.com/connected-intelligence/

[43] Source: Chetan Sharma Consulting, 2019

TAQ = Technology Adoption Quotient – How fast the industry segment adopts new technology
MR = Mobile Relevancy – How much is mobile relevant to the industry segment[44]
ER = Edge Relevancy – How much is Edge Internet relevant to the industry segment
MA = Market Adoption – How fast can the adoption of Edge Internet in the industry segment grow[45]

Industry segments are split into *existing* and *future* use cases

To calculate the Edge Value to an industry segment, we look at the quotient that determines how fast Edge will be adopted in the segment. Just because a technology is available doesn't mean, the industry is ripe for adoption. For example, healthcare has a notoriously low adoption quotient while entertainment sector has a high degree of probability that a new technology will impact in the short term. The more the use case or the industry segment is mobile dependent, the more Edge can help with some of the thorniest performance issues in the segment.

Just because an application is mobile doesn't mean Edge plays an immediate role in its evolution, for example, in the case of manufacturing, lot of the facilities are hardwired and while mobility and Edge can make an enormous difference on how the plant operates, their short-term needs are probably met by fixed fiber deployment. However, when the layout needs to change, wiring becomes cumbersome and time is lost, it is also hard to change the work-flow process like inserting a machine in between two human processes or vice-versa. Finally, we have to determine how fast the market is likely to grow.

We looked at 25 existing use cases and over 20 future use cases to understand the growth patterns of the Edge Internet Economy.

Some of the use cases are listed in the table below:

Table 1. Example Use Cases considered in the analysis of the Edge Internet Economy

Existing Use Cases	New Use Cases
Advertising	AR Games
Network Operations	Cloud Gaming
Commerce	VR Entertainment
Transportation	Stadiums
Healthcare	Sports Betting
Smart Cities	Enterprise – Drones
Agriculture	Enterprise – Manufacturing
Airlines/Airports	Autonomous Cars
Media and Entertainment	Fleet Management
Finance and Insurance	Holographic Entertainment

[44] Though Mobile has a stronger case of using the Edge architecture, Edge is applicable to fixed scenarios as well and such scenarios are considered as part of the analysis
[45] The rate of growth of each Industry segment is also considered in the analysis

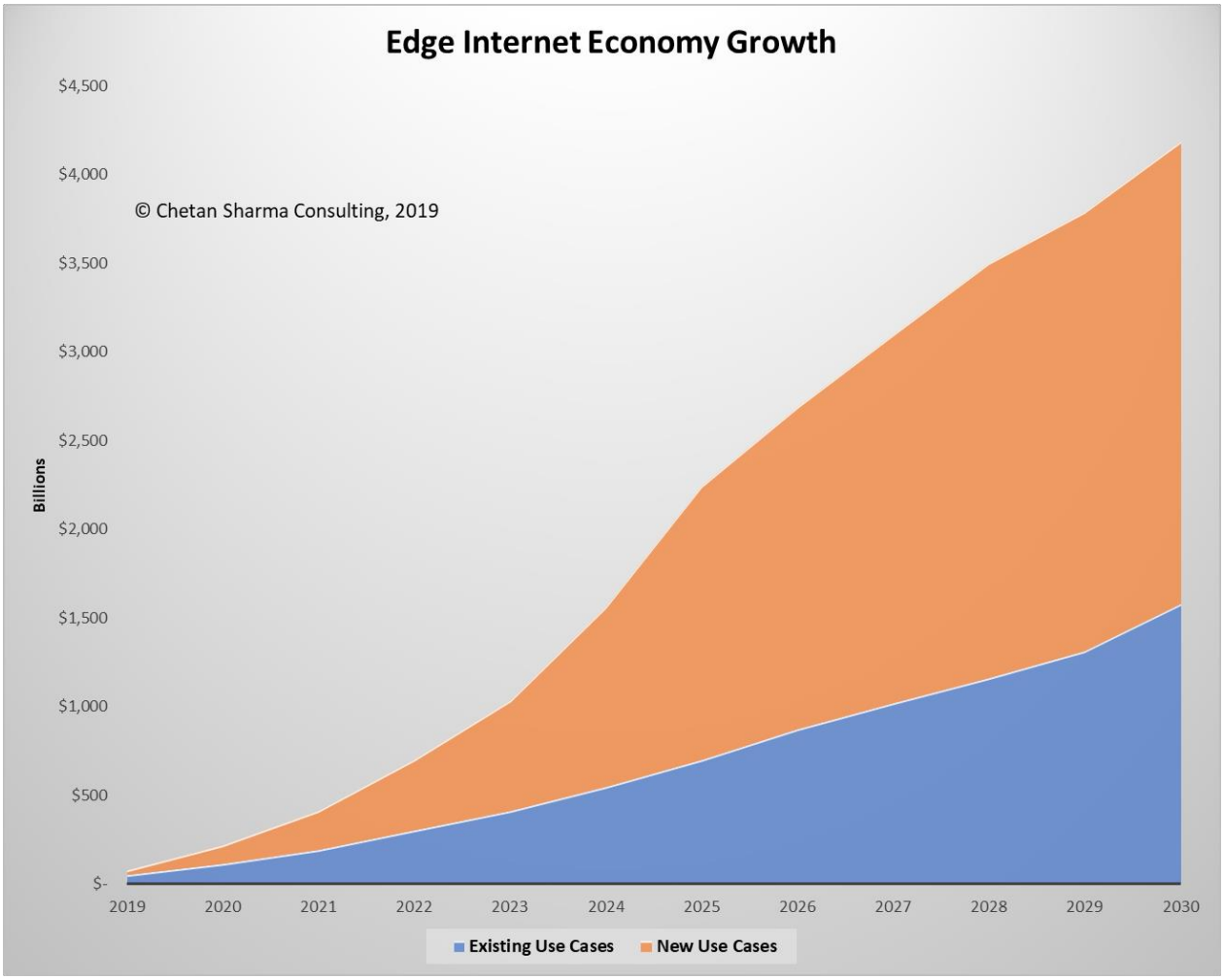

Figure 11. Edge Internet Economy Growth[46]

We estimate that by 2030, the Edge Internet economy will be over **$4.1 Trillion** worldwide (Figure 11). The initial growth will come from Edge serving existing use cases and will gradually be accelerated by the new use cases as the deployment becomes more widespread and developers learn to take advantage of the Edge Internet architecture for applications across industry domains in all major markets.

[46] Source: Chetan Sharma Consulting, 2019

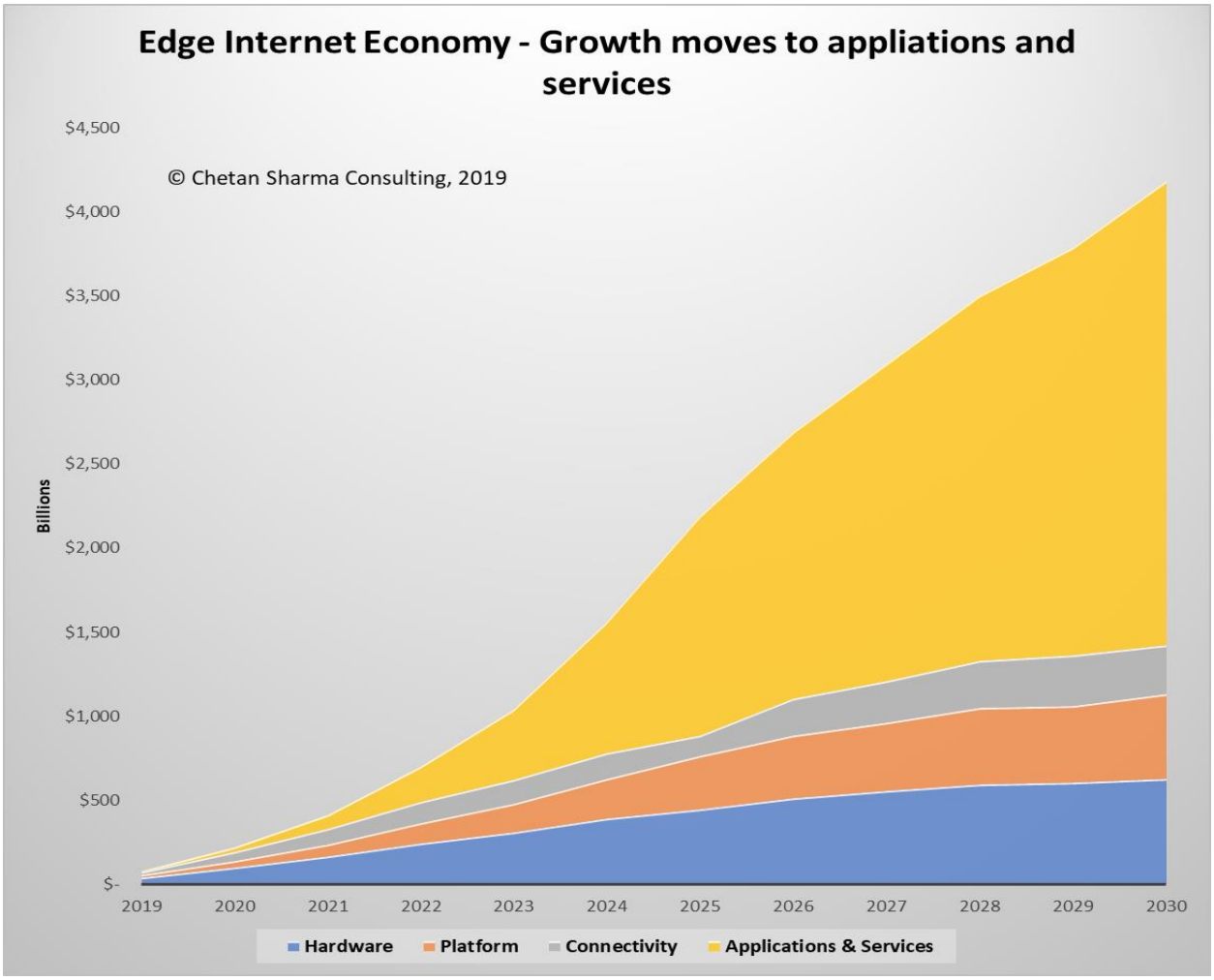

Figure 12. Edge Internet Economy Segmentation[47]

In terms, of Hardware, Connectivity, Software, and Services, the initial growth will be powered by capex spending initially but there is no question that applications and services will take over the bulk of the industry services over time (Figure 12).

[47] Source: Chetan Sharma Consulting, 2019

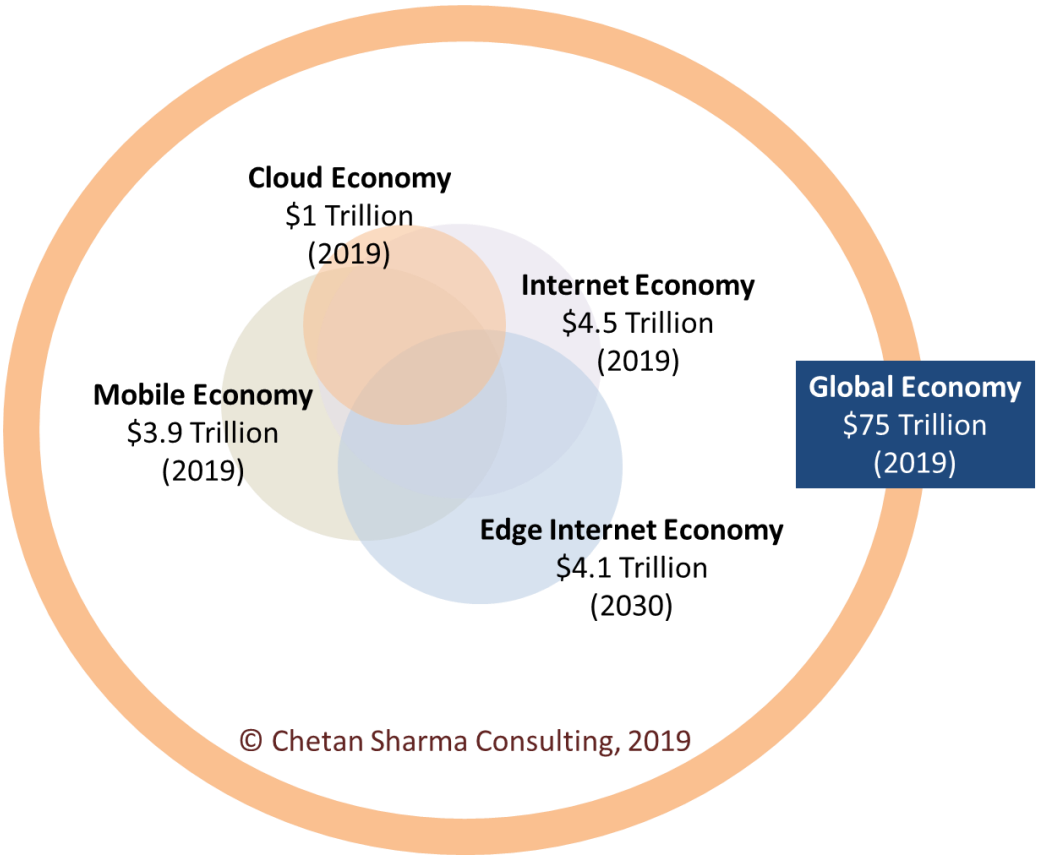

Figure 11. Where does Edge fit in?[48]

Coming back to the question of where Edge Internet fits, one can imagine a series of overlapping circles representing technology ecosystems. Edge Internet is of course in its infancy, but we think the economic impact of the Edge will be bigger than the cloud and similar to that Internet and the mobile technologies have had over the years. It will influence all industry sectors and will for sure, help create new streams of innovation and disruption, new industries and companies, and new jobs and ecosystems.

[48] Source: Chetan Sharma Consulting, 2019

Role of Mobile Operators

The distributed nature of the Edge Internet necessitates having a virtualized Edge environment with computing, storage, and networking resources at locations near the requesting end point. Mobile operators and service providers with their networking resources are a natural place for Edge nodes. What role they play in the business models and revenue streams will depend on the comfort level and ambition of the individual operator. Edge is not only a new revenue opportunity for the mobile operator, but it is also necessary for managing data traffic. In fact, without playing a role in the Edge Internet economy, mobile operators won't be an effective player in the 5G ecosystem.

Existing Operations and Networks

Most of the mature markets constantly face congestion issues at 20-30% of their cellsites depending on the time of the day or traffic surges driven by events. Operators face the dilemma of getting more spectrum, splitting the cells, or find innovative solutions to relieve congestion. Software based Edge Internet solutions will provide an effective way to manage traffic at peak times but more importantly the Edge Internet architecture enables operators to partner with content providers to place streaming content closer to the user thus relieving core and backhaul from content request surges.

New Revenue Streams

The basic model of Edge for operators is simple. For the cellsites they own, they can lease computing and communications capacity to applications and services, existing cloud players and just seek a monthly rent which will be a new source of revenue. Many if not most mobile operators will probably seek this as their going-in Edge strategy. However, as we have discussed in our 4th Wave series papers,[49] the value has moved up from the access layers to the applications and services layers where the bulk of the industry revenues and profits reside now. Some of the progressive operators will seek to go up the value chain with the Edge Internet architecture and formulate partnerships and ecosystems to deliver a complete end-to-end experience to industry segments just like what we saw in the advertising case study earlier. To be able to achieve that requires foresight and ability to create new developer ecosystems. Operators missed out on the cloud over the last decade but by the virtue of their assets, they can play an important role in the Edge Internet economy.

[49] http://www.chetansharma.com/fourth-wave/

Edge Internet Roadmap

Internet wasn't built in a day and similarly, Edge Internet buildout will take time. However, the rollout will be faster as it is going to happen concurrently in different geographies albeit at a different pace. To have a thriving ecosystem, hundreds of players need to be actively participating in doing the plumbing for building an Edge Internet architecture that will serve the economy for years to come. Players across the value chain – chipmakers, component providers, server and hardware manufacturers, facility operators, access and connectivity providers, technology and tools providers, system integrators, infrastructure stewards, platform enablers, traffic accelerators, open source communities, application and service developers will all play a role in building the Edge Internet stack that both works with legacy and existing architectures but also clearly and assertively lays the foundations of the future of Internet.

We think even the funding and business models of infrastructure buildout will change. Mobile operators worldwide are under significant revenue pressure and not everyone can allocate funds for Edge Internet unless they see demand or customers at their doorstep. As such, new players will step-in to leverage the operator physical assets and build the business on top with new revenue sharing models depending on the industry and the problem set they are trying to address. New technologies will be needed to provide breakthrough in routing, messaging, I/O, security, AI, data management, personalization, energy management, etc. which will keep the ecosystem thriving for a long time to come.

Conclusions

The scholars of the Long Wave thesis like Nikolai Kondratieff, Joseph Schumpeter, and Carlota Perez have brilliantly articulated the role of creative destruction in rejuvenating the economic process by "industrial mutation that incessantly revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one."[50] We have seen this basic cyclical structure of the economy repeat itself over the last three centuries. More recently, we have lived through the evolution of the Internet, Mobile Broadband, and Cloud Computing. Internet brought information to our fingertips, provided a common user interface in the form a browser, and upended the development process and consumption patterns in the global economy. Mobile removed time and space from the equation by allowing us to connect and consume from anywhere. It brought Internet to the billions who otherwise wouldn't be connected. The centralized cloud enabled startups to launch new products and services at a feverish pace. They no longer needed substantial capital to bring an idea to the market. With literally, point-and-click, developers could bring services to their consumers in a matter of hours.

Internet, Mobile, and the Cloud brought dynamism to every industry, created new ones, and growth accelerated around the world. We are on the verge of another disruptive cycle that brings resources to the point of need. The Edge Internet will provide an upgrade to the existing Internet, Mobile, and Cloud infrastructure that will change the performance and cost equations dramatically. The Edge economics will lay the foundations for a new breed of applications and services that were just not practical or feasible before. Like every cycle, there will be new winners and losers.

Edge Internet is inevitable. It is already here. We discussed the advertising use case that is generating new revenue for the value chain, raising the bar on performance metrics, and showing what an upgraded network architecture can do to consumer engagement. We will see similar results across all industry verticals. Furthermore, new application categories will be created that are not possible with existing network architectures as they require low latency and high degree of local computing. Edge Internet will enable new forms of data economies, bring in new players to the data supply-chain, and help us invent new business models.

Internet, Mobile, and Cloud enabled new types of developers who were native to the medium. Similarly, we are going to see edge-native developers who will take advantage of the technical superpowers to create new experiences for consumers and workflows for the enterprises. History has shown that companies and entrepreneurs who understand the significance of the introduction of new technology cycles early enough tend to have a higher probability of inventing new capabilities, introducing radical business models that alter the money flows in the ecosystem, and creating new platforms and ecosystem that will serve as the foundation of the next disruptive cycle. Therefore, one must get involved early in understanding how the Edge Internet platform could be used in your own industry.

In this paper, using historical perspective, use cases, and data analysis, we have outlined the business and technology case for Edge Internet.

[50] Capitalism, Socialism, and Democracy, Joseph Schumpeter, 1942.

Acknowledgements

This research paper was supported by AlefEdge. More information can be found at www.alefedge.com.

About Chetan Sharma Consulting

Chetan Sharma Consulting is one of the most respected management consulting and strategic advisory firms in the mobile industry. We are focused on evolving trends, emerging challenges and opportunities, new business models and technology advances that will take our mobile communications industry to the next level. Our expertise is in developing innovation-driven product and IP strategy. Our clients range from small startups with disruptive ideas to multinational conglomerates looking for an edge. We help major brands formulate winning, profitable, and sustainable strategies.

Please visit us at www.chetansharma.com.

About the Author

Chetan Sharma is CEO of Chetan Sharma Consulting and is one of the leading strategists in the mobile industry. Executives from wireless companies around the world seek his accurate predictions, independent insights, and actionable recommendations. He has served as an advisor to senior executive management of several Fortune 100 companies in the wireless space and is probably the only industry strategist who has advised each of the top 9 global mobile data operators. Chetan serves on the advisory boards of Ericsson, Telefonica, Kymeta, NextNav, Zeotap, Opanga Networks, Mast and a number of other startups. Some of his clients include NTT DoCoMo, Disney, KTF, China Mobile, Toyota, Comcast, Motorola, FedEx, Sony, Samsung, Alcatel Lucent, KDDI, Virgin Mobile, Sprint Nextel, Skype, AT&T Wireless, Reuters, Juniper, Qualcomm, Microsoft, Google, Facebook, Mozilla, SAP, Merrill Lynch, American Express, and Hewlett-Packard.

Chetan is the author or co-author of a dozen best-selling books on wireless including *Mobile Advertising: Supercharge your brand in the exploding wireless market* and *Wireless Broadband: Conflict and Convergence*. He is also the editor of the *Mobile Future Forward Book* Series. His books have been adopted in several corporate training programs and university courses at NYU, Stanford, and Tokyo University. His research work is widely quoted in the industry. Chetan is interviewed frequently by leading international media publications such as *Time* magazine, *New York Times, Wall Street Journal, Business Week, Japan Media Review, Mobile Communications International,* and *TechCrunch,* and has appeared on NPR, WBBN, and CNBC as a wireless data technology expert. He is also the chief curator of the mobile thought leadership executive forums – Mobile Future Forward and Mobile Breakfast Series.

Chetan is an advisor to CEOs and CTOs of some of the leading wireless technology companies on product strategy and Intellectual Property (IP) development and serves on the advisory boards of several companies. He is also a sought-after IP strategist and expert witness in the wireless industry and has worked on and testified in some of the most landmark cases in the industry such as Qualcomm vs. Broadcom, Samsung vs. Ericsson, Sprint vs. Verizon, Openwave vs. 724 Solutions, and Upaid vs. Satyam. Chetan is a senior member of IEEE, IEEE Communications Society, and IEEE Computers Society. He has Master of Science degree in Electrical Engineering from Kansas State University and Bachelor of Science degree from the Indian Institute of Technology, Roorkee.