About Mobile Edge Cloud Computing – An Introduction

Version 1.0

Guenter I. Klas
Nov 08, 2015

The 1 minute takeaway

We are all using applications and apps, from mobile devices like smartphones and tablets. A great deal of application code is hosted in the so-called Cloud. There tends to be physical distance between me as the user of the application and the Cloud where various kinds of computations are happening. Long distance is the enemy of low latency, and thus the enemy of applications which require super-fast response time, e.g. those from the emerging Tactile Internet. How can the problem be solved? Via distributing the Cloud, moving some application code and data from the main Cloud to the fringes of networks, to the *Edge* of networks, into closer proximity to consumers. For wireless networks, this leads to the concept of Mobile Edge Cloud Computing. The industry has taken several first steps in this direction. Some more work remains to be done which is outlined here as well.

Tags: 5G, 5G radio, cloud computing, Cloudlet, Edge Computing, Fog Computing, haptic feedback, low latency, MEC, Mobile Edge Cloud Computing, Mobile Edge Computing, Mobile Edge Computing Congress, Robotics, Tactile Internet, Victor Bahl

Mobile Edge Cloud Computing

Everybody surely knows that most of our on-line applications and services are served out of the "Cloud". Cloud computing has become more and more important, with ever increasing size of datacenters and degree of automation. Where next? What comes after the hyper-scale datacenter and ever-growing consolidation? Might it be distributed computing? Edge cloud computing? Micro datacenters?

Well, the evolution of cloud computing hasn't stopped of course. A new trend to highlight here is what's called Edge Cloud Computing, or short Edge Computing. This is all about computation, storage, and networking outside the main hyper-scale datacenters, closer to where the consumers of applications are. As many applications are today consumed over wireless/cellular networks, closer to the consumer means closer to the edge of ISPs and cellular networks. Example: Instead of having the application code for a game partitioned between smartphone and Cloud, split it into three parts: The client part executed on the smartphone, the back end part in the main Cloud (e.g. Amazon, HP, IBM, Google), and a middle part deployed to the edge of a mobile network, right next to its 4G base station (or a step behind it)!

Why edge cloud computing?

There are two main motivations for this:

1) Access to radio network information. When execution of server-side application code is located at the edge of a mobile network, the application can make use of real-time information from the radio access network (E.g. how congested is the radio cell of user Susan just now? How much bandwidth is she being granted over the radio connection?) Through this, applications can become smarter and more intelligent in how they deliver a service to the consumer (e.g. an on-line video stream).

2) Super-short application response time. This is important e.g. for applications of the Tactile Internet which require very low application-level latency, say from 100ms down to a challenging 1ms in the extreme. Mind: This latency is measured on application level, the round-trip-time from

sending the first query from the mobile device to receiving the first server response at the same device (e.g. on http level). Do you know how many milliseconds it takes for a photon to ride from my app on a London smartphone to the Amazon datacenter say in California and back again? Well, flying distance between London Heathrow and LAX is only 5,437 miles. A photon travels at the speed of light: 6.706e+8 miles per hour. So that's 186.3 miles per millisecond. Alas, one way it takes 29ms. Until the photon emitted from my smartphone comes back, 2 * 29ms = 58ms will have passed. That's latency. LATENCY. L A T E N C Y. Too much time for many apps of the Tactile Internet. And mind: there was no buffering, no queuing, not even any processing in the Cloud involved.

To explore what needs to be done to enable the industry to progress based on above motivations, a number of companies launched an international standardisation project in ETSI, called MEC (Mobile Edge Computing) in Sept 2014.

Sept 2014, the **prime white paper** which outlines the vision and goals was published. It's downloadable at [1].

Sept 2015, a **follow-up white paper** was published which is downloadable at [2].

So far so good. However, edge cloud computing doesn't stop with the ETSI MEC project. In contrast, edge cloud computing has a number of aspects, and more than one organisation has been conducting research in this space. I want to highlight the following…

Main areas of edge cloud computing

Area 1: **Interfaces and APIs for (Cloud) applications** which get deployed at the edge of mobile networks. This is the prime focus of the ETSI MEC standardisation project.

Area 2: **Application code provisioning from main cloud to edge cloud**: Research about moving server-side application code, which e.g. is usually hosted in a hyper-scale datacenter, to an edge cloud infrastructure in close proximity to the user, in order to cut down on application-level latency. This is the next hot topic being progressed these days. Carnegie Mellon University's School of Computer Science is the place to look at closer (and I do hope no poaching happens as with CMU's robotics lab back in Jan 2015). See [9] and the power of Cloudlets.

Area 3: Dynamic application code offload from device to network: Research about off-loading application code dynamically, in real-time, from devices like smartphones to more capable "cloud" server infrastructure in proximity of users, naturally at the "edge" of networks. Mind: I'm not going to offload some Java code from my smartphone in London to a Google DC in California; instead I would love to boost my app performance by offloading some compute-heavy code from my burning-hot phone to a micro datacenter actually in London. A topic that has been hot for a good while, however, that is rather challenging.

What about developers?

Why would any ISV or software developer become interested in edge cloud computing? Two answer:

1) Beat your competitors through outstanding quality of experience for consumers: make use of the intimate knowledge the radio access network has about the user's technical circumstances: perfect conditions? Soon out of coverage? Radio cell getting overcrowded? What is an optimal bit rate I should select to deliver content over a radio interface?

2) Making money with applications of the Tactile Internet and Internet of Things (short IoT, including automotive, V2X, driver-less cars and real-time analytics for IoT) and with apps designed to benefit from super-fast future 5G radio access networks in 2020. The use cases and applications of the Tactile Internet are only being discovered now. It's about using the sense of touch, about great

system responsiveness to touch, about haptic feedback to remotely conducting tasks like performing a surgery, controlling robots or drones, teaching students over the web how to do this all, with an immersive, highly effective experience.

So then, how far-fetched is this all?

Well, several players are trying to crack the nut. For now, it's all about researching different aspects of cloud computing at the edge. Sept 2014 saw also the first Mobile Edge Computing Congress happening in London [3]. Those presentations were highly interesting. Unfortunately I haven't found them yet publicly available.

What the future holds is down to crystal ball reading. Comes 2025 (i.e. 10 short years from now into the future), we will look back and may be able to connect the dots. Which are the dots I see today?

Aug 2014: ITU-T Technology Watch Report, published about "The Tactile Internet". It demonstrates what the future may hold and in which cases super-short application response times (and therefore very low system latency) are necessary [4].

Sept 2014: A few companies establish ETSI MEC, the project to standardise some key interfaces for mobile edge computing [5].

May 2015: There is a magnificent presentation authored by Victor Bahl from Microsoft Research. Why giants like Microsoft should look to complement hyper-scale datacenters with micro datacenters at the edge of networks [6].

Sept 2015: Several companies share their view on why mobile edge computing might well be on the path to future 5G networks (at Mobile Edge Computing Congress London). You may want to read the speaker interview with Professor Mahadev Satyanarayanan from Carnegie Mellon University [10].

Nov 2015: Toyota announces it is going to invest one billion dollars in a Silicon Valley research company to develop robotics, not only in the context of driver-less cars, but as they say also for everyday life, including healthcare (a robot helps you to pick up your stick once you have dropped it at the age of 85…. I may add: The robot might soon be remotely controlled by your most loved relative, of course with haptic feedback to the person holding the joystick, adding a bit of human intelligence to the process.) [7]

Dec 2015 at Middleware 2015: Victor Bahl is supposed to present on Micro Datacenter Middleware for Mobile Computing [8]. I would love to be there!

Cloud computing at the edge is mostly a vision today, however, as far as technology is concerned, many features have already been shown to be feasible and have been prototyped. However, there are a few miles still to go.

A number of aspects need to be further looked at:

1. Types of interfaces and APIs necessary to make edge cloud fly.
2. How to cater for mobility of users (Once you move away from a micro datacentre at the edge of a network, your experience with your latency-sensitive application will deteriorate. So then, how to get the application in the distributed cloud to catch up with your current geographical location? Time for a new kind of virtual machine migration…).
3. How to provision and deploy application code and data at the edge of networks? What about deployment on demand, just in time?
4. When it comes to code and data provisioning on IaaS, what about synergy with network function virtualization, the hot topic currently worked on by the telecoms industry?
5. What about the *deployment* APIs for software developers (in contrast to the application *programming* APIs)?
6. And many others.

For now, good enough to get the interested reader going. I hope to find the time to follow up on some of above aspects in future blogs.

References

[1] First white paper on Mobile Edge Computing from ETSI, Sept 2014.
https://portal.etsi.org/Portals/0/TBpages/MEC/Docs/Mobile-edge_Computing_-_Introductory_Technical_White_Paper_V1%2018-09-14.pdf

[2] Second white paper on Mobile Edge Computing from ETSI, Sept 2015.
http://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp11_mec_a_key_technology_towards_5g.pdf

[3] First Mobile Edge Computing Congress, London, Sept 2015. http://meccongress.com

[4] ITU-T report about Tactile Internet, Aug 2014.
https://www.itu.int/dms_pub/itu-t/oth/23/01/T23010000230001PDFE.pdf

[5] ETSI Mobile Edge Computing Portal:
http://www.etsi.org/technologies-clusters/technologies/mobile-edge-computing

[6] Victor Bahl, Microsoft, 13 May 2015.
http://research.microsoft.com/en-us/um/people/bahl/Present/Bahl_mDC_Keynote_May.pdf

[7] Toyota announces massive investment in research into robotics, Nov 2015.
https://uk.news.yahoo.com/toyota-invests-silicon-valley-robotics-programme-062410422.html#JXq7Sn0

[8] Keynote speech, Victor Bahl, Dec 2015.
http://2015.middleware-conference.org/keynote-talk-victor-bahl/

[9] Cloudlets: all about cloudlet-enabled mobile computing. http://elijah.cs.cmu.edu/

[10] Interview with Prof. Satya: http://meccongress.com/professor-mahadev-satyanarayanan-carnegie-mellon-university/