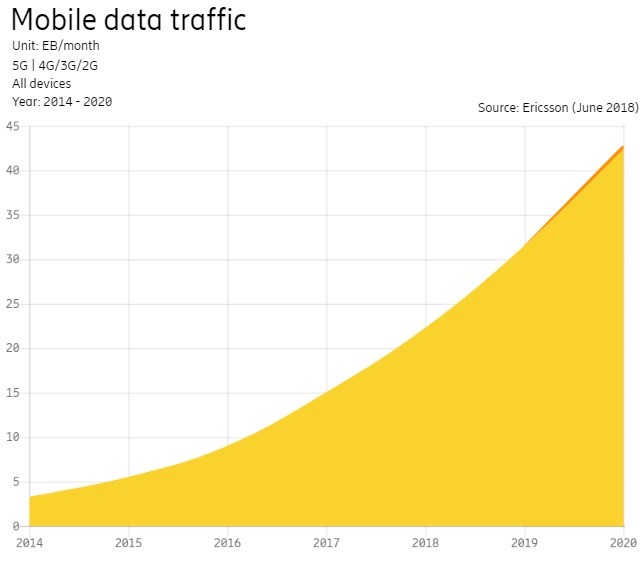
Embedded CRAN Analytics

Introduction

The world is moving towards Internet of Things (IoT). It consists of a wide range of use cases, some of which are complicated. Recently, the industry has coined the term Massive IoT, referring to the large number of devices and machines connected over the Internet. IoT applications require high availability, coverage and low latency and this is being done using 5G.

According to a study by McKinsey, IoT has the potential to become a $11 Trillion Industry by 2025. IoT covers various business verticals such as Connected Wearables, Connected Cars, Connected Homes, Connected Cities and Industrial IoT.

Forecasts show that the IoT Market will grow by several orders of magnitude in the next few years. The predicted number of Cellular IoT Devices will be 28× in 2020 compared to the number of 2014. In addition to this, we are seeing a non-linear rise in data traffic over the Internet due to the rise of IoT Devices. According to Ericsson’s Mobility Report, the increase in data traffic will be from approximately 4 EB/month in 2014 to over 40 EB/month in 2020.

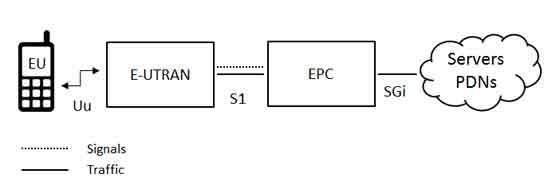


Problem Definition

As seen above, there will be an explosion in the data traffic over the Internet due to the increase in the number of connected devices. These devices will collectively form the Massive Internet of Things, a term coined by 3GPP in their 5G White Paper. The variety of these devices leading to Massive IoT will give rise to different challenges in the field of communication. Solving these challenges would need a new approach. The communication system needs to address the diversity of the devices, security, scalability along with the quality of service.

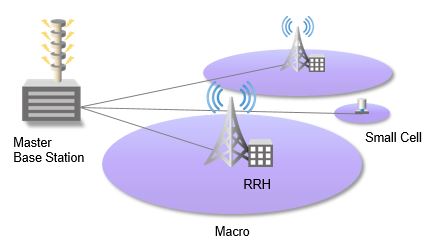
As shown in the figure below, the current LTE System consists of the following functional units:

* User Entity (UE): This is the end-user device. Any cell phone or device connected to the network is known as the User Entity.
* Evolved UMTS Terrestrial Radio Access Network (E-UTRAN): The E-UTRAN handles the communication between the User Entity and the Core. It has only one component, that is the evolved base station. Each base station controls one or more User Entities. Each base station is connected to the Core through an S1 Interface and to other base stations through an X2 Interface. The X2 Interface facilitates signal and packet forwarding during handover.
* Evolved Packet Core (EPC): EPC is a framework for providing converged voice and data over 4G LTE Network. This is connected to the Servers for Corporate Access. The insights for the data can usually be obtained through the Core.



With many User Entities connected to one base station, the load on one base station would increase. Adding more base stations would increase the cost since each base station would come with its own processing power requirements. As we saw earlier, we will have multiple times the devices that we have today within the next few years which will lead to an exponential rise in data traffic over the Internet. Too many base stations in E-UTRAN can lead to interference.

As the amount of traffic on the mobile networks rises, more radios would be required to connect the devices to the Core. More radios give rise to more basebands which in turn leads to increase in cost. In order to solve this issue, we use the concept of Cloud/Centralized Radio Access Network. In CRAN, we separate the base station in two parts – Remote Radio Head (RRH) and the Baseband Unit (BBU). This allows network operators to maintain or increase the number of network access points (RRHs), while centralizing the baseband processing functions into a Master Base Station. The Master Base Station contains all the BBUs at one location instead of having them all over the place like we had in E-UTRAN. Using the Master Base Station simplifies radio resource management in complex operating environments such as Carrier Aggregation. We can add more RRH to the network without significantly adding to the price since the processing takes place at the Master base station and we do not need to add processing nodes at the RRH. This is seen to be cost efficient and reduces interference. The data from this is sent to the EPC which is connected to the Servers.



Now, as the number of UEs increases, the data received by all these UEs would increase. While the CRAN is equipped to tackle this data using Master Base Station, it would be difficult for EPC to keep up with this. The computational power of the Servers will not rise proportionately with the rise of data and the UEs. This would create problems with Latency and Interference at the EPC.

Sample Scenario

Today, most of the Analytics done on the Network is performed on the Server connected to the EPC. Nokia is using Machine Learning on the EPC. In a MIMO (multiple-input multiple-output) network, cellular base stations send and receive radio frequency signals in parallel through many more antennas than are normally used on a base station. This means the base station can transmit and receive more data, but these signals also interfere with one another.

Beamforming is a signal processing technology that lets base stations send targeted beams of data to users, reducing interference and making more efficient use of the radio-frequency spectrum.

One of the challenges in building these systems is figuring out how to schedule the beams. Nokia, for example, has a system with 128 antennas all working together to form 32 beams and wants to schedule up to four beams in a specified amount of time. The company also wants to schedule those beams in a sequence that will provide the highest spectral efficiency, which is a measure of how many bits per second a base station can send to a set of users.

The number of possible ways to schedule four of 32 beams mathematically adds up to more than 30,000 options. There’s simply not enough processing power on a base station to quickly find the best schedule for that many combinations.

Nokia says it was able to train neural networks how to find the best schedule offline, and then later quickly predict the best schedules on demand, although the company did not provide data to back up their performance or allow comparisons to other possible heuristics.

Proposed Solution

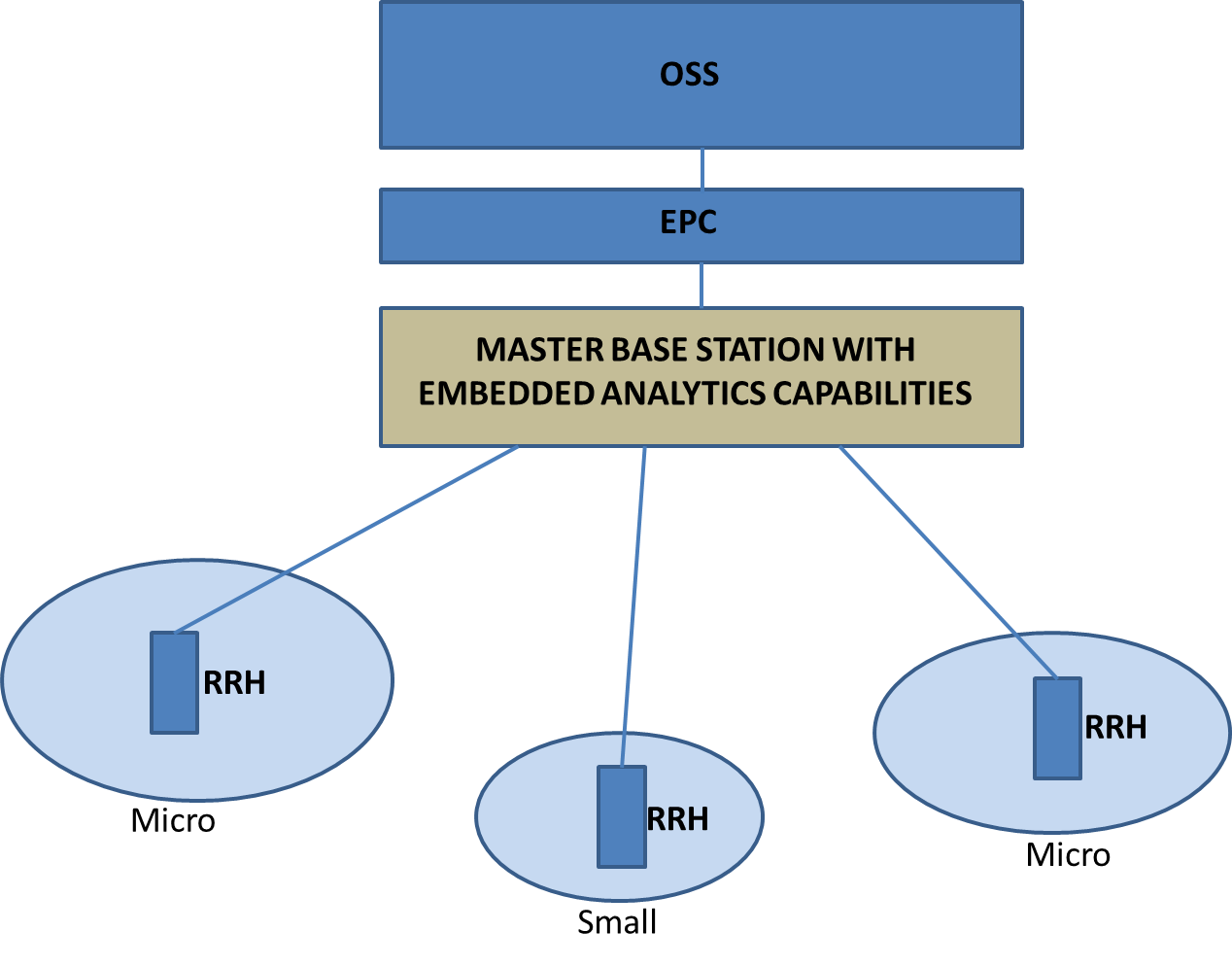
The solution relies on optimizing the operations and the process with the technology advancements. As the technology evolves, the process to use that technology needs to evolve as well. For example, it took nearly 50 years for Electricity to be utilized properly in the manufacturing industry to make the operations more efficient. Similarly, we need to look from the process point of view to address the challenges posed by the data explosion on the Internet. The way to handle the 5G technology, which is a giant leap from 4G should be from this perspective. Hence, one of the techniques could be to make use of Edge Processing for faster operations, reduce latency and avoid bottlenecks at the Core.

Nokia is using Neural Networks, but, they do not have enough processing power on the base station to solve this issue in real time. For this, we need low power Machine Learning Algorithms that will work on the base station utilizing its power. Many low power algorithms utilizing the Kernel Trick exist which reduce the computational complexity of the problem and enables us to run Machine Learning on the Edge.

As the Master Base Station utilizes Distributed processing, we can make use of low power algorithms over a distributed network such as Distributed SVM Algorithm developed at the Embedded and Codesign Lab at Texas A&M University. This algorithm would fit on to the Master Base Station and would provide insights on the data received from the RRUs. This would reduce the load on the EPC by solving most of the problems at the Edge itself.

The New Network Architecture would not be very different from the CRAN Architecture. The only change would be the flow of operation. The User Entity would send data to the Core through the RRH, in the form of Macro Cells and Small Cells, and the Master Base Station. Instead of processing the entire data at the core which is computationally expensive, there is Embedded Processing at every Master Base Station. These are distributed algorithms that do not require high processing power. Most of the data is processed here and some of the data is sent to the core.

Following diagram describes the modified architecture



Proposed Network Architecture

References

1. LTE Progress Leading to the 5G Massive Internet of Things – 3GPP Enhancements up to Release 14 – December 2017
2. <http://www.3gpp.org/NEWS-EVENTS/3GPP-NEWS/1930-SYS_ARCHITECTURE>
3. <https://www.ericsson.com/en/mobility-report/mobility-calculator>
4. https://www.tutorialspoint.com/lte/lte\_network\_architecture.htm
5. Cloud RAN Architecture for 5G – Telefonica, In collaboration with Ericsson
6. Elastic RAN Guidelines – Ericsson
7. <http://www.artizanetworks.com/resources/tutorials/cran.html>
8. <https://spectrum.ieee.org/tech-talk/telecom/wireless/3-ways-nokia-is-using-machine-learning-in-5g-networks>