4

The Fog over the Meadow and the Cloud in the Blue Sky

Rajarshi Sanyal

Belgacom International Carrier Services, Belgium

4.1 Introduction

It is said that 'Change Is Tough But Constant Evolution Is Invigorating'. Last few years, we have witnessed various nuances of telecom network architecture during its continuous evolution trail. The network moved from pure premise based deployments towards the implementation of virtualised instances on the cloud. Technologies like SDN and NFV were introduced which fuelled this network metamorphosis. The evolution to a virtualised network framework had benefited the network operators by reducing the capital expenditure, streamlining the operational processes, reducing the time to deploy new functionalities and rapidly scale up the network capacity. Typically in such scenarios, the network operations are spawned in the cloud environment, the physical location of the clouds are apparently not important anymore. This is indeed one of the selling points of the cloud based technologies. However from a purist's standpoint, this may not be spot-on. The Cloud that we are talking of is not certainly thin mist. The network functions require physical processing resources to operate and the cloud has to fuel its needs. These virtual entities will have to be hosted in real physical machines located in some data centers. The distance between the user devices and the virtual machines have an impact on the data latency and the spatial dimension of the backhaul loop, which determines the overall quality rendered for the given service. The regulatory requirements of certain countries can restrict the user data traffic flow within its boundaries. This makes it overtly important for us to comprehend where the virtual environment is actually hosted, within the country or outside its peripheries. In spite of these constrains, a cloud based mobile network meant for human users appears to be tenable. In fact, it has many 'pros' compared to the COTS (Commercial Off-The-Shelve) installations. However for the machine to machine devices, those simple choices cannot be made. Some machines may need brisker data processing. Some machine may invoke large volumes of queries which can contribute to an avalanche at the backbone network. The centralised network environment in the cloud may be less responsive to such massive computing requests. It is imperative that a part of the processing functions needs to be brought closer to these machines. The telecom researchers thus proposed to lug the cloud from the sky to the ground to mitigate some of these concerns. The cloud morphed into the fog, thus the name ‘fog networks’.

The Fog network concept was coined by Cisco [1] and is well within the scope of the 5G network paradigm. Recently a new industry group, The OpenFog Consortium, was formed to define and promote fog computing. The consortium, founded by ARM, Cisco, Dell, Intel, Microsoft and Princeton University in November 2015, seeks to create an architecture and approach to fog, edge, and distributed computing.

FogNets connect a plethora of devices like wearable devices, connected vehicles, sensors, actuators etc. It provides access to internet via aggregate router and also empowers the devices to communicate amongst themselves in a collaborative fashion. These devices form many “local clouds” at the edge of the network. Contrary to the notion of cloud computing, the fog computing renders capabilities at edge entities with an endeavour to localise computing and relieve the cloud core from the prosaic tasks of repetitive data processing related to control signalling, monitoring, optimisation or even social networking. The edge can also have storage capabilities which is utilised to store chunks of data which can be consolidated to compile the big data. The delta tasks that cannot be accomplished by edge entities are handled by the cloud.

This chapter is organized in 5 sections including introduction (4.1). Section 4.2 provides some background on Fog network with some practical examples. Section 4.3 discusses the overall architecture consolidating fog and the cloud entities Section 4.4 sets out the attributes of the fog. Section 5.5 concludes.

4.2 Background and Examples

Imagine a scenario where a group of college students are travelling together. We see them bantering, sharing pictures and other media files amongst themselves, as well with friends outside their group via social networking platform. The physical environment hosting the social media application may be thousands of kilometres away. The data from the individual smartphones traverses thousands of kilometres every time to reach the central application server which determines the identity and location of the target device. If the originator and the target are in vicinity as in our case, then the trajectory of the data packets becomes a big loop. The traffic trombones via the central server and arrives back at the target device which may be just a meter away. When we have such millions of users exchanging peer to peer information, the backbone traffic and the computing resources expended at the core can be substantial. The pertinent question here is, why cannot the devices communicate with each other at the edges without involving the core? We cannot do it today because most of the device clients are not designed presently to take up a share of the processing tasks done at the cloud core. They are also not fabricated to actuate collaborative communication amongst them. The core applications at present are also not modelled to dispel some processing tasks to the edges, implying the mobile devices.

With the advancement of semiconductor technologies, the edge devices are becoming more intelligent. The network ecosystem is slowly adapting to empower the edge devices to take up some processing load to alleviate the burden of the core, and to make the application more intuitive, real time and responsive. Here are some practical illustrations.

4.2.1 Uber Fog Network

As per the traditional approach the Uber smartphone application of the Uber driver communicates with the Uber backend to continuously update the trip information. The primary server collates the data and replicates across the hot backup servers in other data centres. In case the primary server breaks, the secondary server takes over. As the primary server continuously communicates with the Uber mobile devices and simultaneously needs to update the secondary servers, we envisage substantial data communication between all these elements. Apart from the backhaul and front haul load, we also need to understand that the computing power required in the core to choreograph these processes is substantial. The system is also subject to latency and we need to keep a watchful eye there to avoid timeouts of data transactions.

To circumvent this situation and optimise the ecosystem, Uber has come up with a unique proposition, i.e., to empower the devices to trigger the computing at the edge and diminish the processing burden of the core network (Figure 4.1). The device application has been redesigned. It consumes additional computing resources and memory of the smartphones of the cab drivers. The trip data is locally stored in the device itself which takes decisions unilaterally without invoking any triggers in real time to the core network. The device is also used as a storage device where moving vector information and network mutation information can be stored temporarily. Uber could eventually demonstrate that even if the whole core network is out of service, or isolated from the cluster of end devices, the network operations can still run as normal and unperturbed, thanks to the collaborative edge computing realised by the mobile device processor and its memory. We must acknowledge that this practical demonstration is a giant step towards establishing the fact that the fog concept can indeed work.

4.2.2 IFTTT and Google OnHub

Another novel approach of fog computing is the IFTTT (If This Then That) technology which is presently supported by Google. With this technology, users can orchestrate processes within the mobile device across various applications residing on the handset to trigger a user defined automation task. Before IFTTT was born, this was done at the core network by a mobile service delivery platform which used to house a process choreographer to actuate network driven automation tasks achieving the alike results. With IFTTT, this automation process has been pushed towards the edge. Not only can IFTTT work within a device, but can also help to create a fog network at the edge via an intelligent router with some computing capabilities. Once such example is

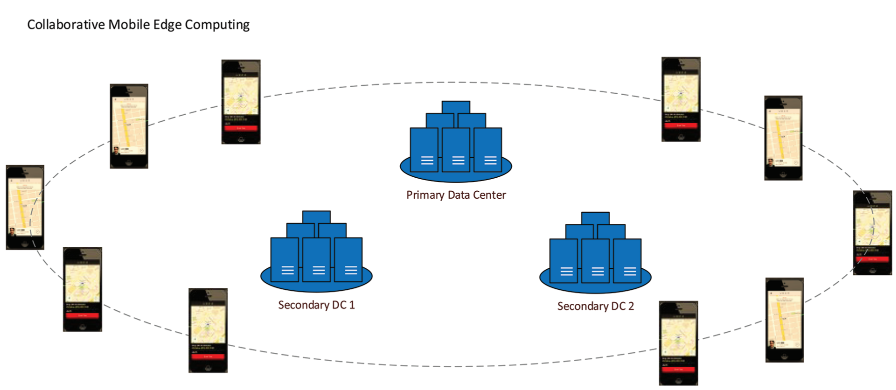

Figure 4.1 Collaborative computing at the edge in Uber network.

Google's Onhub which supports IFTTT and help in connecting a plethora of devices at the edge like smartphones, Internet of Things.

IFTTT can be triggered when devices connect and disconnect from OnHub (Figure 4.2). The user can program any custom built logic, termed as recipe, over a smart interface in their device. It lets users manage and prioritize Wi-Fi to connected devices through an app. Some examples are provided by Onhub on what we can do with it:

- Build a recipe that turns your WiFi Bulb wirelessly. The lightbulb disconnects automatically when you leave your room, actually when your smartphone disconnects from WiFi network.

Figure 4.2 High level representation of the Google OnHub.

- Trigger IFTTT to send you an intimation by email or SMS when your child gets home from school and his/her phone connects to the home's OnHub network.
- Automatically prioritize Wi-Fi to your Chromecast over other existing wifi device connections, when the Chromecast connects to your OnHub network after you plug it in.

4.2.3 Smartgrid

The Power Smart Grid is not a new concept. It aims to pilot the control information flow between the power plants more efficiently, optimizes power distributions, diagonizes and isolates circuit failures, scale up, scale down power level dynamically, actuate smart power distribution etc. The 'As is' power grid is modelled to fetch information from the edge actuators and sensors towards the central cloud to pivot the processing tasks. With the surge in the power requirements, the smart grid network architecture is becoming more intricate. Issues like data privacy, scalability, high availability and latency are more potent than ever. The existing smartgrid model needs to adapt to fit in the new environment. The smartgrid designers are looking towards the fog. In South America, SkyWave IDP satellite terminals control and monitor Smart Grid applications. The SkyWave M2M devices have inbuilt capability to actuate analytics at the edges. The thresholds and filters are implemented at these edge devices so that they can preprocess and filter out the relevant information to be sent to the cloud core for group processing. Hence only the critical messages, like changes in current, voltage and power factor information are conveyed to the core network.

4.2.4 Edge Analytics

An American petroleum company had to abstract essential information like valve pressure, volume of extracted gas, general health of the machines, etc. Earlier, the actuators were disseminating information towards the central application server in the cloud, where the big data was post-processed subsequently for analytics. Latter on, the analytical capabilities were added at the edge, and only vital data like instantaneous high and low pressure and flow, incremental/decremental gas volume, out of range values were being sent wirelessly to the cloud. The company could increase the production by 30 per cent.

Apart from the above examples there are many other uses cases. Few which comes to my mind are smart cities, rail safety systems connected cars, connected wearables, smart traffic lights, sensors for military applications, aircraft sensor network, Virtual Radio Access Networks etc.

4.3 Fog Network Architecture and Its Attributes

Fog network comprises of dense computational virtual structures that provide computing, storage, and networking services at the network's edge. Some of the characteristics of such platforms include low latency, location awareness and operability in wireless hetnet access environment. We shall provide more details on the attributes of the Fog network later in this chapter.

Cisco's impression of Fog Network architecture is in Figure 4.3.

As evident from Figure 4.3, the IoTs and M2M devices will be attached to the fog nodes. The application which governs the fog data services can be run on a separate platform at the edge or can as well run within the fog nodes in a distributed fashion. The data is analysed and sieved by the fog application, and only essential data is transferred to the cloud. The temporary data is retained at the fog layer else which can be used for daily operational work without communicating with the cloud.

The latency drop off is the key to Fog Networks. Figure 4.4 provides a landscape of the latency requirements of Business Intelligence and analytics pertaining to the fog and the cloud network in purview of a typical operation business, say where data is collected from actuators of a petroleum factory.

The fog network directly gathers data from the sensors and the actuators and executes application specific analytical data in real time. The typical time to process the M2M (only) data is in milliseconds. As explained earlier, fog network has memory and retains the generic data. Only critical data is conveyed towards the cloud network for centralised processing and book keeping.

As we go up in the analytics chain (as in the diagram), we observe that the time to process the analytical data increases exponentially. It is also interesting to see that the more we move towards the cloud zone, the use cases are no more only confined to M2M, but it find its use in HMI (Human to Machine Interface).

4.3.1 Fog Network in the Context of 5G

The scope for 5G network does not confine itself in just providing superfast connectivity to users. The scope extrapolates to broader domains. Some of them (though not exhaustive) are as follows:

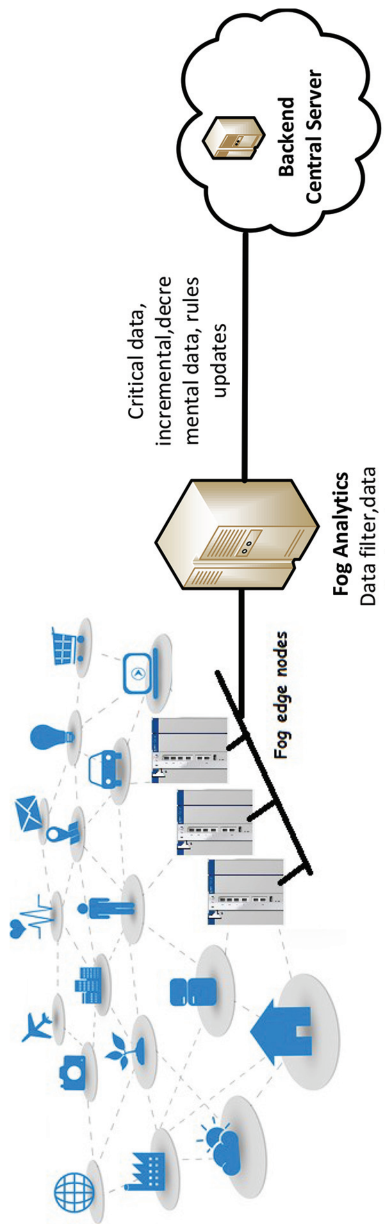

Figure 4.3 Fog network Architecture.

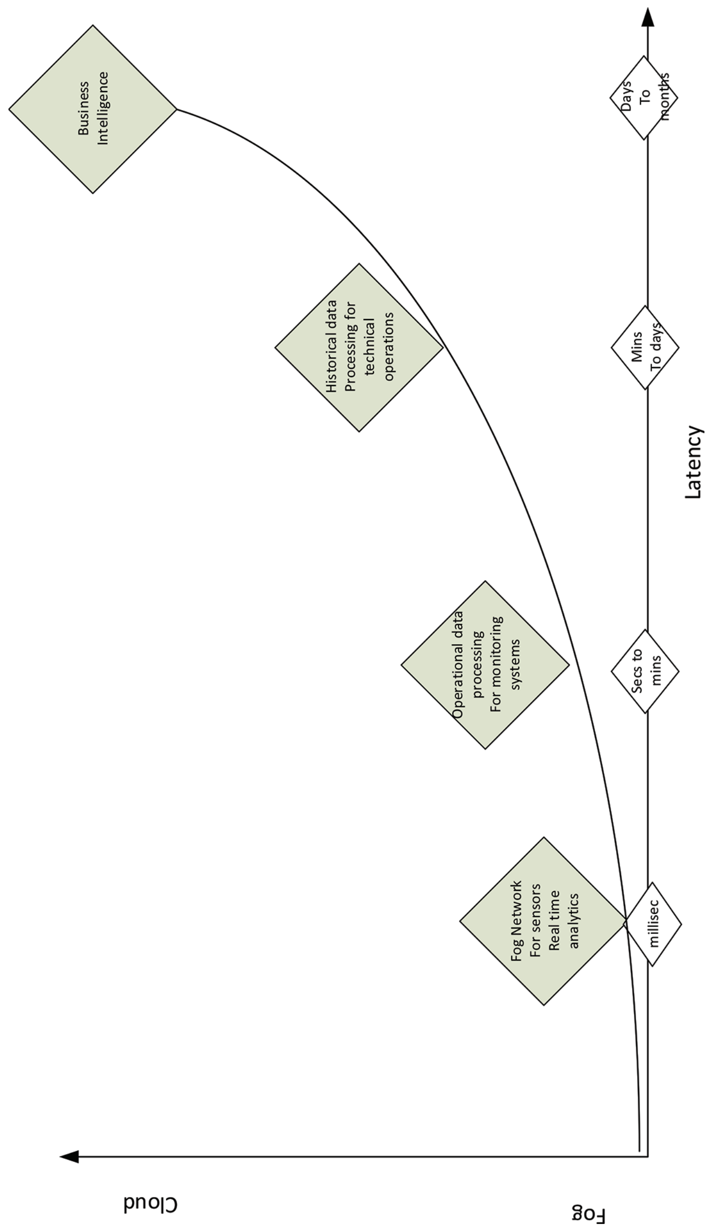

Figure 4.4 The latency w.r.t the analytics chain.

- Integration of the service logic and containers in the network that processes the big data.
- Aggregation of the data feed from the actuators that are linked to varied industrial implementations.
- Cater traffic from human users.
- Comply with the 5G latency requirements.

5G vision for ultra-low latency as envisioned by Ericsson is "To support such latency-critical applications, 5G should allow for an application end-to-end latency of 1 ms or less". Nokia 5G research group has spoken about "Zero latency gigabit experience" for its 5G networks.
Fog network will facilitate this by proposing edge computing in three network domains.

- **Fog Computing**: Assign computing and storage resources at the edge, filter and parse critical data to cloud to reduce backhaul traffic and processing.
- **Fog Mesh Network**: Utilize a swarm of Collaborative device clients at the edges to perform some operations that presently happen in cloud.
- **Fog Radio Access Network (F-RAN)**: Realise real-time Collaborative Radio Signal Processing, Collaborative Radio Resource Management at the edge devices, flexibility to scale up and scale down the access network capacity based on traffic conditions, cognitive radio capabilities that renders the ability to adjust the radio parameters based on traffic conditions and radio environment, and diminish fronthaul payload and processing requirements at the LTE Core.

The fog network and the cloud network landscape in purview of 5G paradigm is captured in Figure 4.5. The 5G LTE-A mobile network core in conjunction with the virtual RAN (comprising of virtual BS pool and RRUs) shapes up the 5 network [2]. The virtual BS pool conveys are user traffic to the core network via backhaul links. The voice and messaging network service are rendered by the IMS core hosted on the cloud platform. This network is designed to serve human users.

Mobile network was initially not designed to serve the machines. Following the evolution trail from 2G to 5G, we witness that the basic network processes, for example location management, addressing, session setup, almost remain the same. The Fog network is typically designed to cater M2M devices. The network is physically isolated from the mobile core network just because the network characteristics are very different due to the machine type usage. The network design and modelling challenges are different here. The traffic patterns we see in some M2M scenario (non-exhaustive) are as follows.

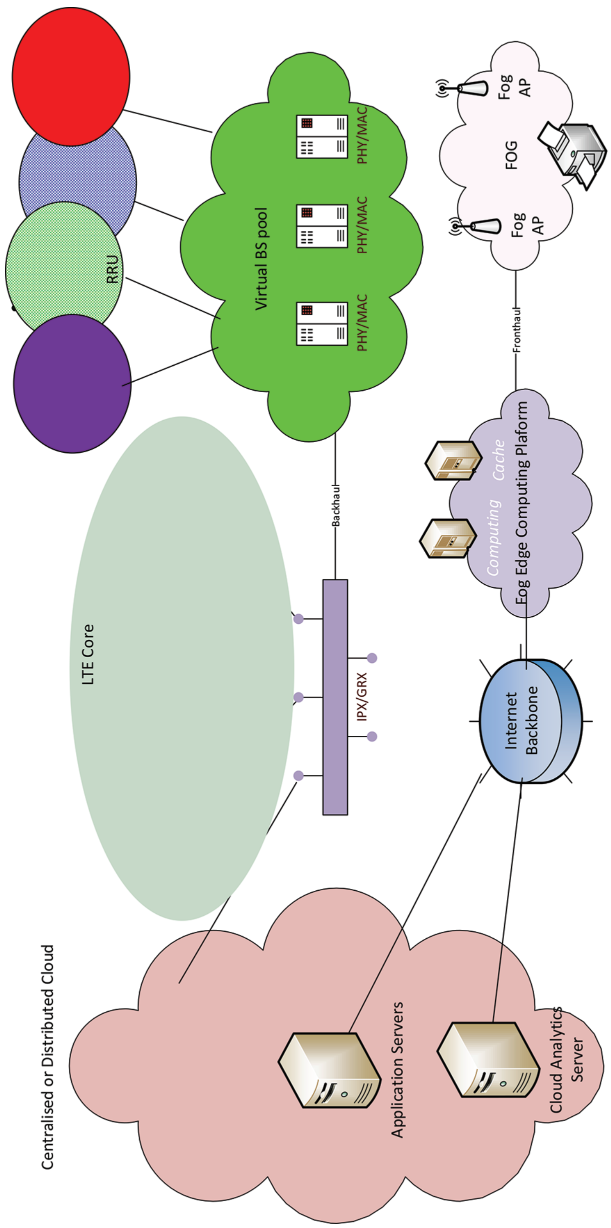

Figure 4.5 Fog and cloud in 5G network paradigm.

- **Electric meters**: Small amount of data, few times a year, no mobility, large number of devices.
- **Health**: Small amount of data, limited mobility, high quality.
- **Medical imaging**: Limited usage but high data volume, needs reliable data channel for quality.
- **SOS**: Once in lifetime communication but needs very reliable channel.
- **ATMs/POS**: Large numbers, small volumes, bursty traffic secure channel.
- **Vehicular communication**: High number of devices, high mobility, low but constant data rate.

Some of the challenges that we face today in sharing the existing mobile network for both human users and M2M are as follows.

1. Lot of devices need to remain attached to network, but remains mostly in idle mode.
2. Many devices can initiate network attach exactly as same time creating congestion at HLR/HSS.
3. Many devices can initiate GTP context activation at same instance creating congestion at GGSN/PGW/PCRF/OCS.
4. Network congestion due to M2M affects human users.
5. Hung GTP sessions: GTP sessions not gracefully terminated by M2Ms while new GTP sessions are initiated $\rightarrow$ leads to capacity loss at GGSN/PGW.
6. Not properly configured periodic location update timer in M2M device which leads to increase in Location updates.
7. Current device addressing scheme may not be able to fuel growth.
8. Support of BYOD (Bring Your Own Device): Device behaviour can vary widely.
9. Security threats pertaining to M2M can be different from human users.

Any surge in M2M traffic can overwhelm the whole mobile network and impact the human users, had we not realised a standalone network for M2M traffic. The access method implemented for M2M can be very different from a 5G cellular networks. Technologies like LoRa (Low power high Range) [2] have been brought in to address the requirements of M2M traffic. Another new technology based on physical layer addressing and mobility management SMNAT (Smart Mobile Network Access Topology) [3] is ideal for the M2M type implementation. The Fog computing platform is connected to the Fog Edge network serving the edge actuators through the fronthaul pipe. LTE-M (LTE for machines) is also a good candidate for M2M access.

The man and machine also need to communicate and interoperate. A central network orchestrator can play the role to choreograph the processes across the two kinds of environments.

4.3.2 Fog Network Attributes

The key attributes of Fog network are as follows:

Cognizance of Location: The Fog network brings in location awareness, and holds the logic for processing the meta data at the edges in accordance to the location of the diminutive cloud. For example, in gaming sort of applications, the latency requirements achieved depends on the physical location of the fog network which is in vicinity of the gaming equipment.

Heterogeneous operational environment: The fog devices can come in various shapes, sizes and behaviour. There is not yet a common framework that binds them together. So interoperability, federation and backward compatibility are some of the areas where the fog researchers have focus. The few companies which are in the fog consortium are attempting to bridge these differences.

Hetergeneous Wireless access: The device can acquire any kind of network, be it cellular, LoRa, WiFi, Zigbee etc. The fog network should harmonise the traffic before feeding the data to the fog computing platform by the fronthaul channel.

Device Mobility: The devices can be dynamic in nature, cross countries or continents. The technologies like LoRa support mobility in a limited way. To attempt ubiquitous mobility, more futuristic technology like SMNAT can be adopted.

Real Time Analytics: The Fog Edge network should be able to perform analytics in real time so as to impart only pertinent information to the cloud in an attempt to diminish the backhaul load and processing weight at the core network.

4.4 Summary and Conclusions

Cloud computing and fog computing play a complementary role in propelling the evolution of the network ecosystem catering the needs of smartphones and M2Ms. The norms widely vary across the two worlds, i.e., man and the machine, and so is their traffic pattern and data distribution. So while we aim to achieve a degree of seclusion between these two network kinds, it is also imperative that interoperability needs to be attained between them. The Fog edge network initiates local data analysis and picks up critical data bearing the changes to the rules of the game. This delta information is transferred towards the central cloud to reduce traffic and processing there. New generation of interfaces are coming up, new architectural choices needs to be addressed. This includes billing framework, security models, operational and analytics processes, session management etc. It is time to think more about our smart watches and the connected coffee machine rather than confining our thought processes only to cloud, NFV, SDN and virtualization. Fog network bears the potential to bring closer the three worlds, the man, the machines and the cloud network.

References

[1] Fog Computing and the Internet of Things: Extend the Cloud to Where the Things Are: CISCO Whitepaper, 2015.

[2] Shao-Chou Hung; Hsiang Hsu; Shao-Yu Lien; Kwang-Cheng Chen, Architecture Harmonization Between Cloud Radio Access Networks and Fog Networks, Access IEEE, 2015.

[3] A technical overview of LoRa® and LoRaWAN™, Technical Marketing Workgroup 1.0, LoRa Alliance.

[4] Sanyal Rajarshi, Prasad Ramjee, Enabling Cellular Device to Device Data Exchange on WISDOM 5G by Actuating Cooperative Communication Based on SMNAT, Journal of International Journal of Interdisciplinary Telecommunications and Networking, Volume 6 Issue 3, July 2014 Pages 37–59, IGI Publishing Hershey.

About the Author

Rajarshi Sanyal is telecommunication engineer with 19 years of industry experience in the field of mobile communication. He started his career as a telecom engineer in Hutchison in the mobile switching domain in India and later on worked on R&D projects on SS7 and CTI applications. He was involved in R&D in a French Telecom startup in the role of a Domain Specialist – Telecom focusing on the development of new solutions related to CAMEL, SMSCs, Mobile Service Nodes and Mobile Payment solutions etc. He was associated with Reliance Communications as Engineering Manager on Mobile Core Networks and later on as Managing Consultant at IBM responsible for designing the Service Delivery Platform Solutions for Mobile Networks. He is presently working as a Network Architect (Voice and Mobile Data core network engineering) at BICS (Belgacom International Carrier Services) at Brussels, Belgium. BICS is one of the largest wholesale voice, signaling and data carriers worldwide. For the last 8 years at BICS, he had been spearheading tier one projects based on roaming hubs, GTP solutions, LTE instant roaming, Voice over LTE, WebRTC, M2M and IoT technologies.

He holds several EU, US and Indian patents in the field of Mobile and Computer communications. He has published 24 papers in international telecom journals and conference proceedings in US, Europe and Far East. He holds a Ph.D. from Aalborg University, Denmark. In his thesis he proposed a new generation multiple access technology based on intelligent physical layer to conceive a smart mobile network and realize device to device communication within the framework of 5G mobile network architecture.