This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/FGCT.2017.8103399, 2017 Sixth International Conference on Future Generation Communication Technologies (FGCT)

Small Cell Deployment for Data Only Transmission Assisted by Mobile Edge Computing Functionality

Raed Sadoon
Dept of Electronic and Comp Eng
Brunel University London
United Kingdom
Email: Raed.Sadoon@brunel.ac.uk

Raid Sakat
Dept of Electronic and Comp Eng
Brunel University London
United Kingdom
Email: Raed.Sakat@brunel.ac.uk

Maysam Abbod
Dept of Electronic and Comp Eng
Brunel University London
United Kingdom
Email:maysam.abbod@brunel.ac.uk

Abstract—The continuous evolution of mobile networks is demanding and promising. The increase in mobile population, the introduction of new applications and use cases which represents a major shift in the paradigm from voice to data especially with the adoption of the IP network from the internet in the mobile networks. This has impose challenges in current architecture of mobile networks in terms of Backhaul capacity and efficiency which can hardly satisfy such user demands due to heavy burden on the backhaul links and long latency. Deployment of the next generation mobile networks require an ecosystem of changeable components that represent an integration of IT and Telecommunications Networking in order to bring new possibilities and capabilities that is a key source to the Average Revenue Per User for operators. A key transformation is the ability to provide network functionality and content delivery at the edge of the network based on Mobile Edge Computing (MEC) borrowed from the IT network while providing the service by deploying a heterogeneous Radio Access network.

Keywords—next generation mobile network, MEC, latency, content delivery, IT.

I. INTRODUCTION

After the introduction of the Short Message Service (SMS) that represents the first usage of data in mobile network presented in the second generation (2G) of the system. We continue to see development in the subsequent generations till the adoption of the Lte-Advance of Long Term Evolution (LTE) system to be the fourth generation (4G) of the mobile network. LTE has been designed to support only Packet-switched services with all IP connectivity. This shift towards data usage with the continuous evolution of mobile networks, leads to an explosion in the usage of mobile applications supported by existing network capabilities. The rapid increase in smart devices with the new use cases that are presented as a result of need or as industries evolvement resulting in a huge amount of data being generated which represent additional burden of the already loaded network backhaul which in turn result in more delay and affect the user experience.

In 2016, the Global mobile data traffic grew 63 percent, and (4G) traffic accounted for 69% of this growth although 4G connections represented only 26 percent of mobile connections in that year. Almost half a billion mobile devices and connections were added where smartphones accounted for most of that growth. Between the overall mobile data traffic, 60 percent of that traffic is a Mobile video traffic. Within the next 5 years, Global mobile data traffic will increase sevenfold, and by 2021, 4G will be 53 percent of connections, but 79 percent of total traffic. There will be 8.3 billion handheld or personal mobile-ready devices and 3.3 billion M2M connections, among that smartphones will surpass four-fifths (86 percent) of mobile data traffic, and Over three-fourths (78 percent) of the world's mobile data traffic will be video [1].

As the demand for mobile broadband is increasing in a rapid way with the proliferation of mobile and portable devices added to the networks and as new applications is presented every day that are in direct or indirect touch to the people's lives. Newer advanced mobile services as wwww (worldwide wireless web), Machine to Machine (M2M), Internet of things (IoT), intelligent or smart society, augmented reality (AR) and Virtual reality (VR) with the development and wide spread use of sensor network. It can be clearly seen that there are vast ranges of technology developments on the horizon and that most of the development is based on the higher demand of data and low latency connected to the term of IoT in one way or another and the mobile/Wireless technologies and due to many reasons such as the dependency on the mobile devices, remote applications, radio capabilities and the integration of different and many networks are becoming the foundation for other industries like telemedicine and online gaming which require higher demands such as 100% network availability in addition to the very high bandwidth and almost zero latency [2].

The structure of the current mobile system will not be able to fulfil the requirements of the new presented technologies. Therefore, new structure is required. While 4G connectivity is predicted to have the majority share of the market, there are field trails currently underway for 5G in some countries like UK, Korea, Japan and USA.

Many of the features that are seen as 5G solution could be in fact be implemented as 4G extensions. And many studies and researches have been done in this field. Different Content delivery (CD) techniques implemented in the core and radio access network (RAN) to allow efficient content distribution and delivery as in [3, 4, 5, 6, 12&13]. In the same aspect of content delivery the authors in [7] studied the concept of

Quality of Experience (QoE) for video traffic in LTE systems and how video users are more affected by unpredictable disruptions. In [8, & 9] the authors make use of the Software-Defined Network concepts by splitting the control plane and user plan, the first paper proposed an approach that enables the incorporation of Mobile Edge Computing (MEC) services by bringing the content and the decision functions to the network edge in order to achieve network capacity saving. While the second paper present a design for content delivery over software defined mobile social networks (SDMSNs). Study for Network Aware and RAN-Aware Content delivery in [10, &11]. Traffic offloading and energy saving in Heterogeneous Networks (HetNets) has been covered in [14, 15, 16 &17]. Many papers were presented in the field of Coordinated Multipoint (CoMP) as in [18].

All the above studies and proposed techniques aimed to enhance the current 4G network in terms of reduce transmission delay, minimizing packet loss, backhaul traffic savings, for efficient contents delivery to end devices and to propose new scenarios for next generation mobile system that requires greater service portability and interoperability.

II. Caching Analysis

In this section we calculate the cost structure of caching data at eNodeBs. We see the problem from two standpoints: The Delay and the Bandwidth (hence the throughput). In the second scenario we will show that caching at eNodeB can lead to many benefits for both mobile operators and end users. We also show that in-network caches can effectively improve the eNodeB's caching performance.

1.1. eNodeB Caching (Scenario 1) Delivery cost

Selecting N to be the number of eNodeBs in the network, n_j be the number of data requests received by the j_{th} eNodeB and P be the mean size (bytes) of a requested object. We also denote the cost components to be as follows:

1. U is the cost per byte from UE to eNodeB,
2. C is the backhaul link cost per byte from eNodeB to PGW.
3. W is the transit cost per byte between core network and the content provider.

If no caching is provided in the cellular network, each request will incur the costs U+C+W. And the total costs for complying all the requests in the network can be calculated as follows:

$$M_{nocache} = \sum_{j=1} (n_j) \times (U+C+W) \times P \qquad (1)$$

In the case of adding cache to the eNodeB we will consider the following parameters for the equation

(i) n_{jc} be the number of request for objects that are cached at the j_{th} eNodeB.

(ii) K is the additional cost per byte of caching objects at the eNodeBs (Server, storage cost, etc.).

With the existence of eNodeB caching, the cost for the mobile operator to serve the requests will be the sum of the following parameters:

$$M_1 = \sum_{j=1} (n_j - n_{jc}) \times (U+C+W) \times P \qquad (2)$$

is the cost when requested objects need to be fetched from the origin server

$$M_2 = \sum_{j=1} n_{jc} \times U \times P \qquad (3)$$

Is the cost incurred on the network path from the UEs to the caching eNodeB.

$$M_3 = \sum_{j=1} n_{jc} \times K \times P \qquad (4)$$

Is the cost when adding cache to the eNodeB

This will result in :

$$M_{cache} = M_1 + M_2 + M_3$$

$$M_{cache} = \sum_{j=1} n_{jc} \times U \times P + \sum_{j=1} (n_j - n_{jc}) * (U+C+W) \times P + \sum_{j=1} n_{jc} \times K \times P \qquad (5)$$

By subtracting (5) from (1) this will result in the benefit we get from adding cache to the network and as follows:

$$M_{Benefit} = \sum_{j=1} n_{jc} \times (C+W-K) \times P \qquad (6)$$

1.2. In-network cache (Scenario 2)

We observe that the new trend of in-network caching can achieve additional delay reduction for end users. Keeping in mind that the advantage of in-network cache might be a disadvantage if the in-network cache cannot be efficiently utilized by the main eNodeB. An example of that is the situation when the cached content is duplicated in both the eNodeB and the in-network caches, which most likely occur in the case of full load [(Scenario 4) in our simulation].

In the next section, we will illustrate a framework presenting small cells solution where eNodeBs can use in-network cache information to improve performance. We simulate the model in order to find an optimal eNodeB caching solution to reduce end to end latency and increase the throughput in the framework under several network load conditions.

III. System Model Description

MEC provides IT and cloud-computing capabilities within the RAN in close proximity to mobile subscribers. In this work, we consider the concept of adding computing capacity and to deploy services; such as content delivery; at the main (e-Node B) that used to provide coverage and handle most of the control signalling and controls the small cells that are operating in different frequency band to deliver the required data services. The main e-Node B is connected to all SCs within its coverage area with fibre links as illustrated in Fig. 1. The Vendors and operators start looking for more cost effective options such as deployment services at an aggregation points at the IP transport network. Recently this concept (MEC/CDN) has evolved to draw on network functions virtualization (NVF) technology to allow virtual network functions (VNFs) to be run on this platform making the concept a strategic option in the 5G network architecture. On the other hand the separation of control plane (C-Plane) and the user Plane (U-Plane) using the software defined feature in the platform will improve the adaptability and manageability of the Network.

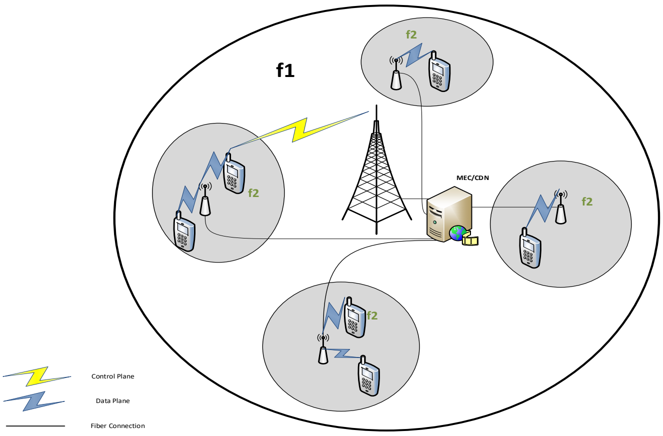

Figure 1: Network Architecture

A. Mobile Edge computing

Mobile Edge Computing (MEC) is a network architecture concept that enables cloud computing capabilities and an IT service environment at the edge of the cellular network.[20] The concept originated with the idea to develop mobile CDNs for mobile networks specifically after the wide spread and popularity of smart phones. Some of the interactive aspect of MEC is its influence on the architecture of future network and to meet the needs of low latency and balance mobility 5G services. This concept has also been proposed by the European Telecommunications Standards Institute (ETSI) in recent years for the same purpose.

Correct deployment of MEC nodes in the RAN, will reduce excessive application layer handovers and can meet the performance targets needed for latency-critical applications. As the service environment at the RAN edge can offer low latency and rich bandwidth as well as direct access to real time radio network information (like subscriber location, cell load, etc.) which can be used to deploy applications that are capable of differentiating the mobile broadband experience [22]. This is important to operators who start to distribute PGW's for reasons of performance and economics. MEC presence can reduce the need for extra PGW's to be distributed, thus reducing inter-PGW handoffs [21].

MEC can be seen as a cloud server operating at the edge of a mobile network and can carry out tasks that could not be achieved with conventional network infrastructure, in a way that can improve user experience, and minimize congestion in the other parts of the network.

MEC will create an eco-system where new services are developed in and around the base station benefiting from real time network data that can be used by the applications to generate more useful data such as cell-ID, location of the subscriber, cell load and throughput guidance. Information regarding the user equipment (UE) activation parameters can be provided by Radio Network Information Services (RNIS) based on 3GPP radio network-layer signalling messages (such as S1 Application Protocol (S1-AP), X2 Application Protocol (X2-AP) and Radio Resource Control (RRC)), The parameters on the UE context and the established E-UTRAN Radio Access Bearer (E-RAB), such as identity of the UE-associated logical signalling connection, QoS, Cell ID for the E-RAB, and when a modification to an E-RAB or to a UE context occurs, or when an E-RAB is released, in addition to measurement and statistics information related to the user plane based on the 3GPP signalling messages and Performance Measurements (PM) defined by 3GPP and gathered by the mobile network element.

Deployment of the MEC server that provides computing resources, storage capacity, connectivity, and access to user traffic and radio and network information must comply with existing 3GPP specifications and need to satisfy 3GPP-related security requirements [22].

The deployment options within LTE macro base station (eNB) site, or 3G Radio network Controller (RNC) site, or at multi-technology (3G/LTE) cell aggregation site, or an operator small cells (SC). Are presented in the Fig. 2 that includes the MEC server platform overview. The MEC platform provides capabilities to perform access control and integrate user/device authorization in a heterogeneous network (HetNet) with the implemented access control system; that classify different levels of services per end-user within.

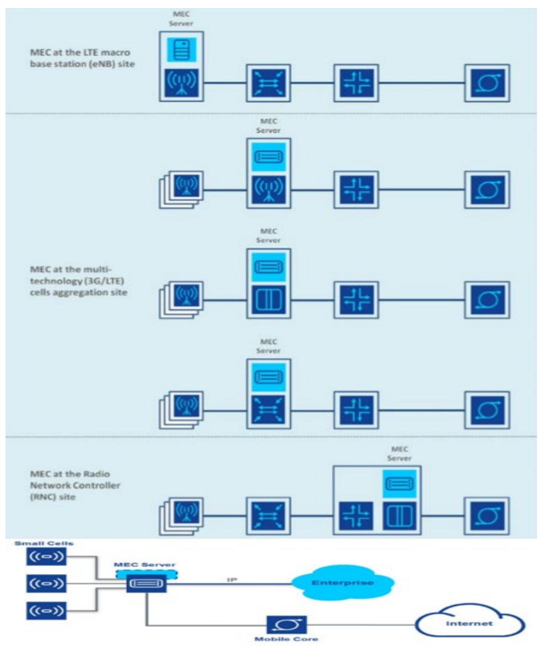

Figure 2: MEC Deployment scenarios [22]

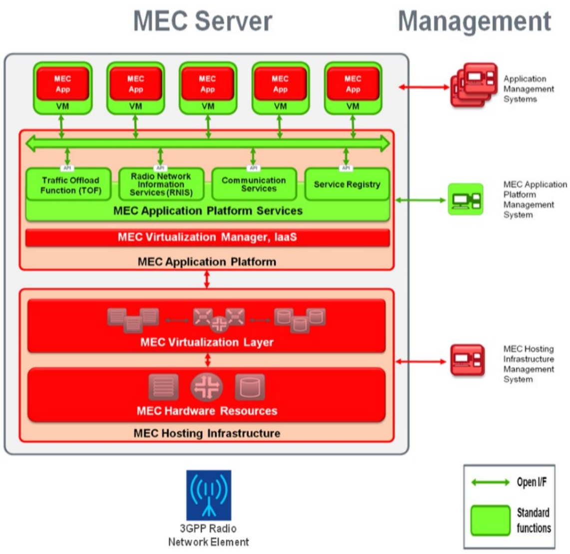

Figure 3: MEC server platform overview [22]

B. Small cells

Low power small cells (SC) are considered promising technology to cope with mobile traffic explosion, especially for hotspot deployments in indoor and outdoor scenarios. According to 3GPP TR [23] a low power node means a node with transmitting (Tx) power lower than macro and BS classes nodes, for example Pico and Femto eNB are both applicable.

Based on the specification requirements of 3GPP in [23], and as in Fig.1, our deployment scenario consider small cells (SC) layer to be deployed under the main macro (MeNB) cell layer, where the UE could be under coverage of both the macro cell (MeNB) and the small cell (SC) simultaneously or only under the coverage of the macro cell only. The small cells are connected to the main macro cell using fibre cable that terminate the connection to a controller which is suggested to be implemented to handle the MEC/CDN functionality like caching, and other specific operations such as synchronization, baseband processing, interference coordination, carrier aggregation, inter-eNB COMP and dual connectivity.

Enhancement in the network such as increasing throughput, reducing signalling overhead and lower delay can be achieved through the implantation of the small cells (SC) under the coverage of the main macro cell (MeNB) using dual connectivity mechanism [24]. In dual connectivity the control plane and user plane are separated and every node has control of its own resources and coordination is still required between the nodes to enable efficient resource management between the nodes. MeNB is primarily responsible for handling UE's RRC state and generates the final RRC message and send it to the UE after radio resource management coordination with the SC. In this way the small cells can be designed to handle pure data downlink, as the macro cell retain all the public signals and channels, while most of the C-plane public signals and channels of the small cells removed. The UE attain access to the MeNB using the normal attached procedure and maintain RRC connection with the MeNB as long as it is under its coverage. After that, when a service is required by the UE, the macro cell can direct the UE to the best SC (i.e. through SIB) based on the specific measurements such as location of the UE and distance from the SC as well as the on-off status of the SC.

IV. Simulation Setup

We chose the Riverbed Modeller to implement and study the practicability of the proposed network. The Riverbed (formerly known as OPNET) library offers more than 400 out-of-the-box protocols and vendor device models including TCP/UDP, IPv6, VoIP/Video/FTP/HTTP/Email, WiMAX, LTE, UMTS, WLAN (a/b/g/n) etc, to support accurate event-driven simulation scenarios.

The LTE HetNet contains 1 main eNodeB of equal size with 4 Small cells belonging to the central cell. The interference and multi-path are modelled. IP traffic is established between the UEs (50 in the main eNodeB cell , 25 in the small cells) and HTTP server is connected to the LTE network through internet backbone. The UEs are randomly distributed in a uniform manner. The typical parameter settings, according to [19], are tabulated in Table I. Adaptive modulation and coding was enabled in order to enable the UE to communicate with the eNodeB in variable channel conditions.

Table I: Simulation parameters

Parameter	**Value**
FTP File Size	2 Mbytes
File inter-arrival distribution	Exponential
Average File inter-arrival Time	16s, 20s, 24s
Total MeNB/SeNB Tx Power	46/30 dBm
Noise figure	9 dB in UE, 5 dB in eNBs
UE Tx Power	23 dBm
Carrier frequency	f1 at 2 GHz (MeNB); f2 at 2.6 GHz (SeNB)
LTE Bandwidth/Duplexing	10 MHz/FDD
Sub-carrier spacing	15 KHz
Sub-frame length (TTI)	1 ms
Symbols per TTI	14
Data/control symbols per TTI	11/3
Path loss	Macro cell (MeNB): 140.7+36.7 log10 (R[km]) Small cell (SeNB): 128.1+37.6 log10 (R[km])
Shadow fading	Macro cell (MeNB): Lognormal, std = 8 dB Small cell (SeNB): Lognormal, std = 10 dB

V. RESULTS AND DISCUSSION

In this section we analyse the proposed scheme based on the scenarios and settings discussed in the previous section . The sent and received IP data packets all over the LTE network are examined.

The throughput and packet end-to-end (E2E) delay are chosen as the key performance factors. In the first scenario, the network is configured with low load traffic to decrease the probability of packet loss due to either the buffer overflow or repeated retransmissions due to the traffic congestion.

Three main cases are considered;
1) No content is cached
2) Content cached in the M-eNodeB
3) The UE is connected to the M-eNodeB in the UL and to the SC-eNodeB in the DL, with Content cached in the SC-eNodeB.

Fig. 4 below shows the response of the network in terms of end-to-end delay for the three scenarios. It can be observed that the delay is very high when the UE is connected to the M-eNodeB provided that no data is cached in the network while it is acceptable when the contents are cached in the M-eNodeB and has dropped significantly when the UE is connected to the SC-eNodeB and using the proposed scheme.

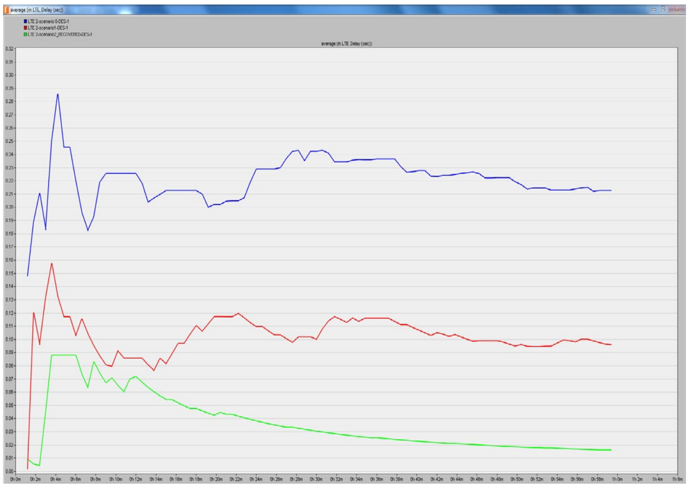

Figure 4: Overall end-to-end delay

The explanation of drop in the delay is the fact that the distance to the M-eNodeB is quite larger than the distance to the SCeNodeB, provided that the DL frequency differs from the UL frequency which reduces the interference in the network in order to help decreasing the losses as proposed.

Considering the same network simulation we examine the results in terms of the throughput. Fig. 5 illustrates the throughput delivered in bits/seconds. It can be observed that throughput is increasing when the content server is getting closer to the UE, achieving its best when the UE downlink is connected to the SC-eNodeB and using the proposed scheme.

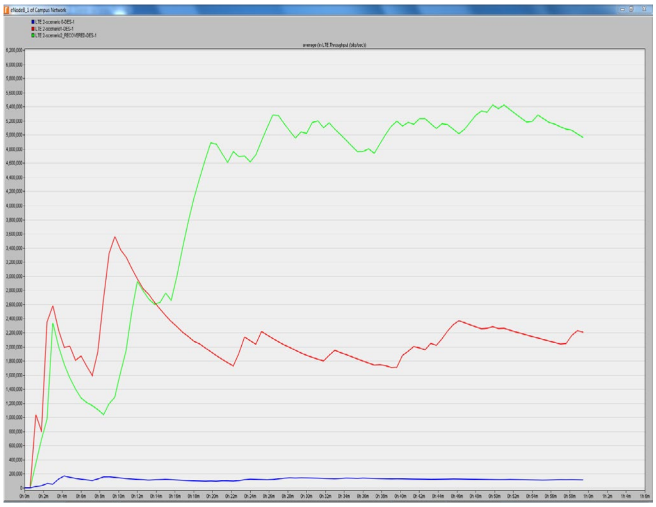

Figure 5: Throughput in bits/sec

Logically, the throughput (bits/sec) increases in 2 cases i) when the data traffic increase, ii) when the elapsed time decrease, hence in our model, when the content are cached in the M-eNodeB the network delivers and performs best at the begging of the simulation, whilst it start to perform even better in the third scenario when the DL is connected to the SC-eNodeB after 20% of the simulation time, this is due to SC-eNodeB initialization and time spent fetching the content from the main sever.

Finally we experienced the case when the same network runs and routes full load in its data plane. Fig. 6 below shows the response of the network in terms of end-to-end delay for the three scenarios. It can be observed that the delay is very high when the UE is connected to the M-eNodeB provided that no data is cached in the network while it is acceptable when the contents are cached in the M-eNodeB and has dropped remarkably when the UE is connected to the SC-eNodeB and using the proposed scheme.

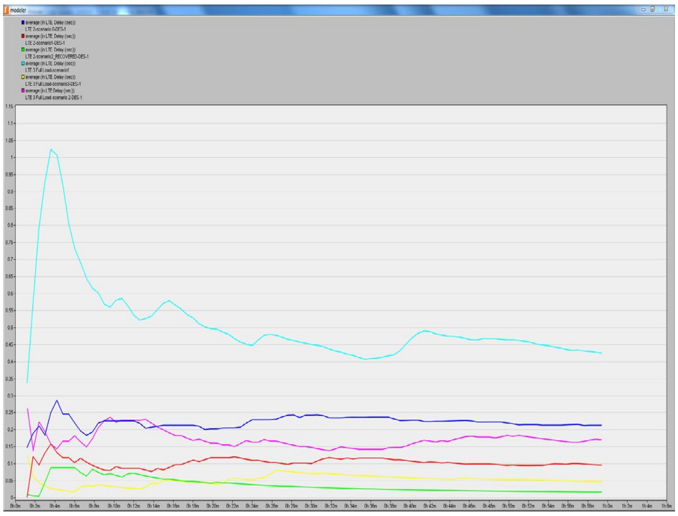

Figure 6: Overall end-to-end delay with full load

In the same way Fig. 7 illustrates the throughput delivered in bits/seconds. It can be observed that throughput is increasing when the content server is getting closer to the UE, achieving its best when the UE downlink is connected to the SC-eNodeB and using the proposed scheme. Noting that when the content server is placed in the M-eNodeB the throughput drops dramatically when the network is running full load, this is due to high traffic that is being processed and requests from the UEs to be fulfilled by one content server.

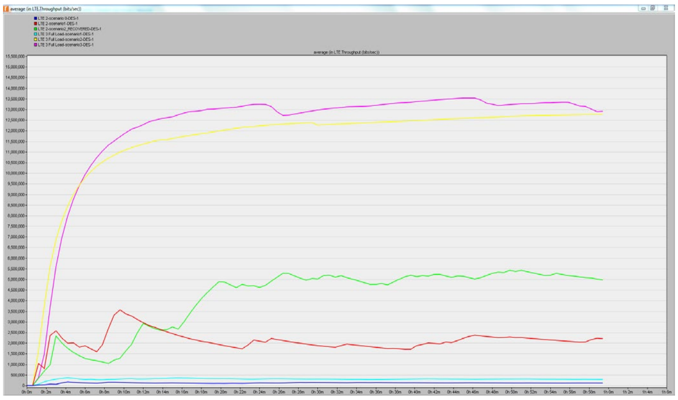

Figure 7: Throughput (bits/sec) with full load

VI. Conclusion

In this paper we presented a new transmission/reception scheme based on the CoMP concepts when UL and DL links are physically separated into M-eNodeB and SC-eNodeB. When comparing the data traffic received in the DL direction under many network circumstances, it can be observed that the proposed method had successfully enhanced the performance of the network and neutralized any performance degrading that may occur due to a sudden high network load or conditions. The results show that the proposed scheme can reduce the packet loss and latency when routing full load in the network.

References

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021

[2] Raed Sahib Sadoon, "Explosion of Data (BIGDATA)", Chapter "3" in the "Internet of Things and Big Data Analysis: Recent Trends and Challenges", Nov, 2016, ISBN-10:0692809929.

[3] Rim Haw and Choong Seon Hong, "A seamless content delivery scheme for flow mobility in Content Centric Network," *2012 14th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, Seoul, 2012, pp. 1-5.

[4] N. Katanekwa and N. Ventura, "Mobile content distribution and selective traffic offload in the 3GPP evolved packet system (EPS)," *The International Conference on Information Networking 2013 (ICOIN)*, Bangkok, 2013, pp. 119-124.

[5] C. Schwartz, J. Eisl, A. Halimi, A. Rafetseder and K. Tutschku, "Interconnected content distribution in LTE networks," *2010 IEEE Globecom Workshops*, Miami, FL, 2010, pp. 2028-2033.

[6] Z. Su and Q. Xu, "Content distribution over content centric mobile social networks in 5G," in *IEEE Communications Magazine*, vol. 53, no. 6, pp. 66-72, June 2015.

[7] D. J. Vergados, A. Sgora, A. Michalas, D. D. Vergados, J. P. Laulajainen and Y. Jiang, "A QoE-Driven Adaptation Scheme for Video Content Delivery in LTE Networks," *ISWCS 2013; The Tenth International Symposium on Wireless Communication Systems*, Ilmenau, Germany, 2013, pp. 1-5.

[8] V. G. Vassilakis, I. D. Moscholios, B. A. Alzahrani and M. D. Logothetis, "A software-defined architecture for next-generation cellular networks," *2016 IEEE International Conference on Communications (ICC)*, Kuala Lumpur, 2016, pp. 1-6.

[9] Z. Su, Q. Xu, H. Zhu and Y. Wang, "A novel design for content delivery over software defined mobile social networks," in *IEEE Network*, vol. 29, no. 4, pp. 62-67, July-August 2015.

[10] R. B. Miller, J. P. Hearn, C. Purzynski, F. Cuervo and M. Scheutzow, "Mobile video delivery using network aware transcoding in an LTE network," in *Bell Labs Technical Journal*, vol. 16, no. 4, pp. 43-61, March 2012.

[11] S. K. Kovvali and C. W. Boyle, "RAN-aware content delivery, proxy optimizations, and offload in cellular wireless networks," *2014 IEEE Latin-America Conference on Communications (LATINCOM)*, Cartagena de Indias, 2014, pp. 1-6.

[12] P. Suthar and M. Stolic, "Deploying information centric networking in LTE mobile networks," *2016 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob)*, Bandung, 2016, pp. 130-136.

[13] A. Gomes and T. Braun, "Load balancing in LTE mobile networks with Information-Centric Networking," *2015 IEEE International Conference on Communication Workshop (ICCW)*, London, 2015, pp. 1845-1851.

[14] S. Chen, Z. Yuan and G. M. Muntean, "A traffic burstiness-based offload scheme for energy efficiency deliveries in heterogeneous wireless networks," *2013 IEEE Globecom Workshops (GC Wkshps)*, Atlanta, GA, 2013, pp. 538-543.

[15] J. Zhang, X. Zhang and W. Wang, "Cache-Enabled Software Defined Heterogeneous Networks for Green and Flexible 5G Networks," in *IEEE Access*, vol. 4, no. , pp. 3591-3604, 2016.

[16] P. Sermpezis and T. Spyropoulos, "Offloading on the edge: Performance and cost analysis of local data storage and offloading in HetNets," *2017 13th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*, Jackson, WY, 2017, pp. 49-56.

[17] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch and G. Caire, "FemtoCaching: Wireless video content delivery through distributed caching helpers," *2012 Proceedings IEEE INFOCOM*, Orlando, FL, 2012, pp. 1107-1115.

[18] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa and M. Tanno, "Coordinated multipoint transmission/reception techniques for LTE-advanced [Coordinated and Distributed MIMO]," in *IEEE Wireless Communications*, vol. 17, no. 3, pp. 26-34, June 2010.

[19] A. Ahmed and E. Ahmed, "A survey on mobile edge computing," *2016 10th International Conference on Intelligent Systems and Control (ISCO)*, Coimbatore, 2016, pp. 1-8.

[20] Mobile Edge computing use cases and & deployment options , Juniper white paper

[21] Milan Patel, et al, Mobile-Edge Computing – Introductory Technical White Paper, September 2014.

[22] 3GPP TR 36.932 V13.0.0 (2015-12)

[23] 3GPP TR 36.842 V12.0.0 (2013-12)

[24] ITU-R M.2135-1, "Guidelines for evaluation of radio interface technologies for IMT-Advanced," December 2009.