A Review of Load Balancing Techniques in 3GPP LTE System

Alatishe S. Adeyemi[1], Dike U. Ike[2]
[1,2]Department of Electrical and Information Engineering,
Covenant University, Ota, Nigeria.

Abstract- Load imbalance that reduces network performance is a serious problem in 3GPP Long Term Evolution Networks. Load imbalance occurs in communication networks due to non-uniform user deployment distribution. Several load balancing techniques have been proposed, this paper reviews only two techniques such as heaviest-first load balancing.

Keywords-Load Imbalance, Handoff Algorithm, 3GPP LTE

I. INTRODUCTION

With the ever increasing rise in demand for mobile broadband services, content on-the-move, it is evident that customers are always looking for more bandwidth, better quality, and services tuned to their personal profile at little costs. These user demands require not only faster networks and radio interfaces but also higher cost-efficiency networks.

LTE, whose radio access is called Evolved UMTS Terrestrial Radio Access Network (E-UTRAN), is expected to substantially improve end-user throughputs, sector capacity and reduce user plane latency, bringing significantly improved user experience with full mobility [1]. The main objectives of LTE are to minimize the system and User Equipment (UE) complexities, allow flexible spectrum deployment in existing or new frequency spectrum and to enable co-existence with other 3GPP Radio Access Technologies (RATs) [2, 3], The result includes a flexible and spectrally efficient radio link protocol design with low overhead, which meets the challenging targets [7] that were set to ensure good service performance in varying deployments. The data rate can vary from more than 300 Mb/s in the downlink and 75 Mb/s in the uplink. The LTE architecture also contributes to reducing the cost of network deployment, as discussed in the following section.

LTE has performance requirements that rely on physical layer technologies, such as, Orthogonal Frequency Division Multiplexing (OFDM) and Multiple-Input Multiple-Output (MIMO) systems, Smart Antennas to achieve these targets. LTE uses Orthogonal Frequency Division Multiple Access (OFDMA) as its access technology in the downlink (DL) [4] while it uses Single-carrier frequency division multiple access (SC-FDMA) as its uplink multiple access technology [5]. In OFDMA, the entire bandwidth is divided into many sub-carriers. 3GPP LTE system is a self-organizing network (SON) focused on self-configuration, self-healing and self-optimization techniques [6]. Self optimization handles functions like load balancing, handover parameter optimization (HPO), and capacity and coverage optimization.

II. OVERVIEW OF 3GPP LTE

The objective of LTE is to develop a framework for the evolution of the 3GPP radio access technology towards a *high-data-rate*, *low-latency*, and *packet-optimized* radio access technology [8]. The key issue is to increase spectral efficiency and coverage while maintaining acceptable complexity and cost. From the radio interface point of view, the current 3GPP Release 5 and 6 solutions can achieve up to 14.4 Mbps downlink and 5.7 Mbps uplink peak data rates (without channel coding) using HSPA. Thus for the long term evolution, clearly more ambitious goals were set up. Among others, the main targets are presented below [8]:

- Significantly increased peak data rate e.g. 100 Mbps (downlink) and 50Mbps (uplink).
- Significantly improved spectrum efficiency (e.g. 2-4 times that of Release 6 HSPA system)
- Radio Network user plane latency below 10 ms *Round Trip Time* (RTT) with 5 MHz or higher spectrum allocation.
- Scalable bandwidth up to 20 MHz (lowest possible bandwidth: 1.25 MHz).
- Support for inter-working with existing 3G systems and non-3GPP specified systems.
- Support for packet switched domain only (voice is carried by e.g. VoIP).
- Optimized for low mobile speed, but also with support for high mobile speed.

The *Transmission Time Interval* (TTI) is made as short as 1 ms in order to improve the *Round Trip Time* (RTT) performance. The downlink transmission scheme is based on conventional OFDM using a cyclic prefix [9]. The system has a scalable bandwidth up to 20 MHz, with smaller bandwidths covering 1.25 MHz, 2.5 MHz, 5 MHz, 10 MHz and 15 MHz to allow for operation in differently sized spectrum allocations. The transmission bandwidth is varied by varying the number of OFDM sub-carriers while keeping the sub-carrier spacing constant. The throughput enhancing mechanisms in LTE downlink are explained in details as follows, namely the Orthogonal Frequency Division Multiplexing (OFDM), Link Adaptation (LA), fast Frequency Domain Packet Scheduling (FDPS), Hybrid Automatic Repeat Request (HARQ) and Multiple-Input Multiple-Output (MIMO).

In LTE, downlink and uplink transmissions are organized into radio frames with 10 ms duration. Two radio frame structures FDD (Frequency Division Duplex) and TDD (Time Division Duplex) are both supported. The time-domain structure is shown in Fig. 1. Each radio frame consists of 20 slots of length 0.5 ms, numbered from 0 to 19. A sub-frame is defined as two consecutive slots. In FDD, uplink and downlink are simultaneously transmitted within the 10 ms interval every frame and separated in frequency domain.

Figure 1. LTE time-domain structure

LTE downlink transmission scheme is based on OFDM (Orthogonal Frequency-Davison Multiplexing). This technique breaks the available bandwidth into many narrow subcarriers and transmits the data in parallel streams. The basic LTE downlink physical resource can be seen as a time-frequency resource grids which are grouped into Resource Blocks (RB) [10] (see Figure 2). Each resource block consists of 12 subcarriers with spacing of 15KHz. A downlink carrier can include from 6 RBs up to more than 100 RBs which corresponds to the downlink transmission bandwidth ranging from 1.25 MHz to 20 MHz Since LTE is designed to work on a wide bandwidth up to 20 MHz, the channel will experience severe frequency selective fading. By splitting it into several narrowband channels at different frequencies, OFDM-based transmission can cope with severe channel conditions, inter-symbol interference (ISI) and contribute to spectral diversity gain.

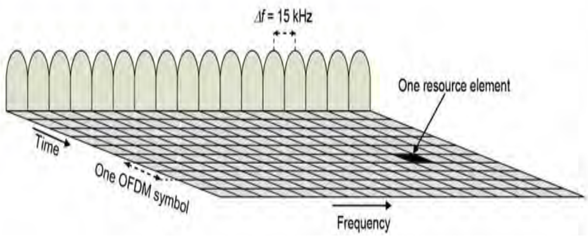

Figure 2. The LTE downlink physical resource

LTE uplink transmission scheme is based on SC-FDMA. The basic transmitter and receiver architecture is very similar to OFDM which can be seen as a conventional OFDM transmission combined with DFT-based (Discrete Fourier Transform) pre-coding. Therefore, similar to OFDM, the SC-FDMA physical resource can be seen as a time-frequency grid with the additional constraint that the overall time-frequency resource assigned to a mobile terminal must always consist of consecutive subcarriers. The principle advantage of SC-FDMA over conventional OFDM is a lower peak-to-average power ratio (PAPR) because the underlying waveform is essentially single-carrier.

III. SYSTEM ARCHITECTURE OF THE 3GPP LTE

The evolved packet system (EPS) that consists of the core network part, the evolved packet core (EPC) and the radio network evolution part, the evolved UTRAN (E-UTRAN), also known as LTE. The EPC also can be connected to other 3GPP and non-3GPP radio-access networks. As illustrated in Fig. 3, the EPC consists of one control- plane node, called a mobility management entity (MME), and two user-plane nodes, called serving gateway (S-GW) and packet-data network gateway (P-GW). The LTE radio-access network consists of the base stations, denoted as enhanced NodeB (eNB), that are connected to each other through the X2 interface and to the EPC through the S1 interface. The mobile terminal is denoted as user equipment (UE). The architecture in EPC/LTE, with only two user-plane nodes (eNB and S/P-GW),1 is simpler than in UTRAN Release 6 with four nodes (NodeB, radio network controller [RNC], serving general packet radio service [GPRS] support node [SGSN], and gateway GPRS support node [GGSN]) and reduces the user-plane latency. One consequence is that some functionality performed by the RNC in UTRAN, such as ciphering and header compression, is performed by the eNBs in LTE. Further, handovers between eNBs are handled through packet forwarding over the X2 interface rather than by means of a central automatic repeat reQuest (ARQ) entity in the RNC as in UTRAN [11].

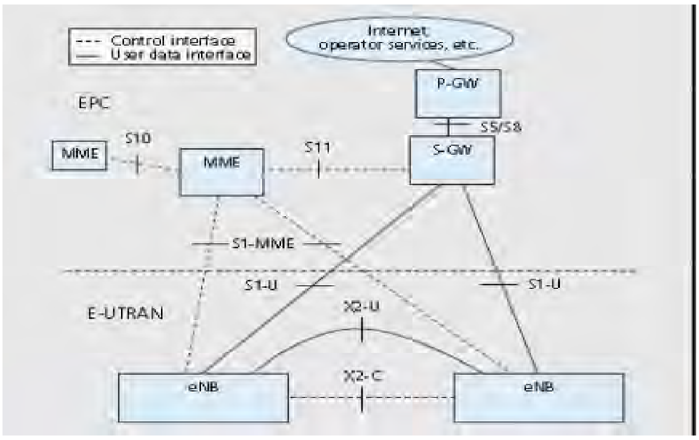

Figure 3. LTE Architecture

IV. LOAD BALANCING

In existing networks, parameters are manually adjusted to obtain a high level of network operational performance. In LTE the concept of self-optimizing networks (SON) is introduced, where the parameter tuning is done automatically based on measurements. A challenge is to deliver additional performance gain further improving network efficiency. The use of load-balancing (LB), which belongs to the group of suggested SON functions for LTE network operations, is meant to deliver this extra gain in terms of network performance. For LB this is achieved by adjusting the network control parameters in such a way that overloaded cells can offload the excess traffic to low-loaded adjacent cells, whenever available. In a live network high load fluctuations occurs and they are usually accounted for by over-dimensioning the network during planning phase. A SON enabled network, where the proposed SON algorithm monitors the network and reacts to these peaks in load, can achieve better performance by distributing the load among neighbouring cells [12].

When the loads among cells are not balanced, the block probabilities of heavily loaded cells may be higher, while their neighbouring cells may have resources not fully utilized. In this case load balancing can be conducted to alleviate and even avoid this problem. There has been a lot of research done on load balancing, which can be classified into two categories: block probability-triggered load balancing [12-14], and utility based load balancing [15-17].

In the first category, the overhead is low because the load balancing is triggered only when the block probability is larger than a certain threshold. However, the block probability is not minimized, since load balancing can be done before block happening to reduce it. For the second category, i.e., utility-based load balancing schemes, the performance is better because the load balance and throughput are considered in both cell selection and handover phases. However, their overheads are heavy, because the load of each cell has to be exchanged instantaneously.

In [18], they deal with load imbalance in LTE networks, and propose a load balancing algorithm intended to balance the load among cells and keep network throughput with reasonable overhead but still in the line of the designing strategies of practical cellular system. They first formulate the problem as an optimization problem, which considers the trade-off between network throughput and load balancing and employs user-cell matching indicator as parameters. Then analyzed its complexity and propose a suboptimal algorithm, called Heaviest-First Load Balancing (HFLB). In the algorithm, new users access the cell with the maximum signal strength and load balancing is triggered when the load of the busiest cell in the network exceeds a threshold in each load balancing cycle. Appropriate users are switched out under the metric considering throughput and load jointly. Since only the heaviest loaded cell is chosen to do load balancing, the overhead of HFLB is low. Simulation results show that the HFLB algorithm can significantly decrease the load balance index (i.e., enhance load balancing) while keeping network throughput as high as possible at a small price of only a bit more handovers.

Then the load of cell *i* at time slot *t* is:

$$\boldsymbol{\rho_i(t) = b_i^u / b_i(t)} \tag{1}$$

In LTE networks, all the cells are often allocated the same number of PRBs so that we use *b* instead of $b_i(t)$.

For performance analysis in the heaviest-first model, they define a load balance index measuring the degree of load balancing of the entire network, as follows:

$$\boldsymbol{\xi(t) = \sum_{i \epsilon N} (\rho_i(t) - \overline{\rho}\,(t))^2} \tag{2}$$

Where N is the set of cells (or eNodeBs) in the network, and $\overline{\rho}(t) = \sum_{i \in N} \rho_i(t)/|N|$ is the average load of the network at time slot *t*, where $|N|$ is the number of cells in the network. The load balance index is 0 when load is completely balanced among cells. The bigger the value of $\xi(t)$, the severer the unbalanced load distribution among cells. The target of load balancing is to minimize $\xi(t)$.

In Simulating the network model for OFDMA based LTE downlink, partial frequency reuse (PFR). PFR can reduce inter-cell interference for most users. They only consider signal to noise ratio (SNR) for simplicity. The set of users is denoted by K. Then the SNR of the signal received by user $k \in K$ from eNodeB $i \in N$ at time slot t can be written as:

$$SNR_{i,k}(t) = S_{i,k}(t)/N_0, \qquad (3)$$

Where $S_{i,k}(t)$ represents the power of received signal by user k from eNodeB i on the allocated PRB at time slot t, and N_0 is average power of the additive white Gaussian noise (AWGN) on the same PRB.

Given $SNR_{i,k}(t)$, the achievable Shannon rate at time slot t for user k from cell i is:

$$r_{i,k}(t) = W_{i,k} \log_2(1 + SNR_{i,k}(t)) \qquad (4)$$

where $W_{i,k}$ is the bandwidth of the PRB allocated by cell i. Considering that adaptive coding and modulation is used in LTE networks, we will use Shannon rate in equation (4) as the throughput of a user. The throughput of the entire network $R(t)$ is the sum of all the users' throughput at time slot t. In [18], they dealt with the optimization issue on dynamic load balancing and network throughput in 3GPP LTE networks. they formulate the issue as a joint integer optimization problem and propose a suboptimal method to solve it in a distributed manner. The algorithm for the method is developed. The generality and validity of the algorithm has been evaluated in various cases.

In [19], Load Balancing is expressed as a multi-objective optimization problem whose constraints are physical resource limits (PRB) and QoS demands. An algorithm was developed which includes QoS guaranteed hybrid scheduling, Handoff of users with and without QoS requirements, and call admission control. This algorithm reduces call block probability.

In their Network Model, they considered users with two kinds of QoS requirements, Constant bit rate (CBR) and Best Effort.

SINR for user k K from cell i N at a subframe τ is given by

$$SINR_{i,k}(\tau) = \frac{g_{i,k}P_i(\tau)}{N + \sum_{j\epsilon N, j\neq i} g_{j,k}P_j(\tau)} \qquad (5)$$

Where N is AWGN, g is the gain between eNodeB i and user k, and p is the transmit power of eNodeB i.

Hence bandwidth efficiency is

$$e_{i,k}(t) = \log_2[1 + E(SINR_{i,k}(t))] \text{ bps/hz.} \qquad (6)$$

The load of a cell i at a given time t is

$$\rho_i(t) = \frac{S_i^c(t)}{S_i(t)} \qquad (7)$$

Where $S_i(t)$ is the total resources, $S_i^c(t)$ is resources occurred by CBR users, and $S_i^b(t)$ is resources occurred by BE users.

In this algorithm LB is done by enforced Handover. Users with higher QoS requirements are guaranteed. It uses QoS aware Handover.

V. CONCLUSION

In this paper, LTE was discussed; Its Architecture, frame structure, downlink and uplink transmission modalities were explored. Load Balancing, which is an issue in every telecommunication network, was explored. Two Algorithms were reviewed, the heaviest-first LB and the QoS-aware LB. Both techniques aid LB in a unique way, but i recommend the QoS-aware LB, because it considers the heterogeneous nature evident in LTE networks and reduces call blocking probabilities.

REFERENCES

[1] R. Divya and A. Huseyin, “3GPP – Long Term Evolution – A technical Study,” Spring 2009’

[2] A Technical White paper,"Long Term Evolution (LTE): A Technical Overview". Motorola

[3] D. Astely, E. Dahlman, A. Furuskar, Y. Jading, M. Lindstrom, and S. Parkvall, “LTE: The Evolution of Mobile Broadband”, LTE PART11: 3GPP Release, Ericsson Research.

[4] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, “LTE-advanced: next-generation wireless broadband technology [invited paper,” Wireless Communications, IEEE, vol. 17, pp. 10-22, 2010.

[5] A. Omri, R. Bouallegue, R. Hamila, and M. Hasna, “Channel Estimation for LTE Uplink System by Perceptron Neural Network,” International Journal of Wireless & Mobile Networks (IJWMN), vol. 2, pp. 155–165, 2010.

[6] Doo-Won Lee, Gye-Tae Gil, and Dong-Hoi Kim, “A Cost-Based Adaptive Handover Hysteris Scheme to Minimize the Handover Failure Rate in 3GPP LTE,” EURASIP Journal on Wireless Communications and Networking, Article ID 750173, vol. 2010, Hindawi Publishing Corporation, pp. 1-7.

[7] M. Meyer et al., “ARQ Concept for the UMTS Long-Term Evolution,” VTC Fall 2006, Montreal, Canada,Sept. 2006.

[8] 3GPP.Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN). Technical Specification TR 25.913 (V 7.3.0), Technical Specification Group Radio Access Network, Mar. 2006.

[9] 3GPP.Physical Layer Aspects for Evolved UTRA. Technical Specification TR 25.814 (V 7.1.0), Technical Specification Group Radio Access Network, Oct. 2006.

[10] E. Dahlman, S. Parkvall, J. Sköld, P. Beming, 3G Evolution: HSPA and LTE for Mobile Broadband, Academic Press, 2007.

[11] A. Larmo, M. Lindstrom, M. Meyer, G. Pelletier, J. Torsner, H. Wiemann, "The LTE Link Layer Design," LTE PartII: 3GPP Release 8, IEEE Communications Magazine, April 2009.

[12] O. K. Tonguz, and E. Yanmaz, "The mathematical theory of dynamic load balancing in cellular networks," IEEE Trans. on Mobile Computing, vol. 7, no. 12, pp. 1504-1518, Dec 2008.

[13] B. Eklundh, "Channel utilization and blocking probability in a cellular mobile telephone system with directed retry," IEEE Trans. Comm., vol. 34, no. 3, pp. 329-337, Apr. 1986.

[14] H. Jiang, and S. S. Rappaport, "CBWL: A new channel assignment and sharing method for cellular communication systems," IEEE Trans. Vehicular Technology, vol. 43, no. 4, pp. 313-322, May 1994.

[15] S. Das, H. Viswanathan, and G. Rittenhouse, "Dynamic load balancing through coordinated scheduling in packet data systems," IEEE Proc. INFOCOM, 2003.

[16] T. Bu, L. Li, and R. Ramjee, "Generalized proportional fair scheduling in third generation wireless data networks," in IEEE Proc. INFOCOM, Apr. 2006.

[17] K. Son, S. Chong, and G. Veciana, "Dynamic association for load balancing and interference avoidance in multi-cell networks," IEEE Trans. on Wireless Communications, vol. 8, no. 7, pp. 3566-3576, Jul. 2009.

[18] H. Wang, L. Ding, P. Wu, Z. Pan, N. Liu, and X. You, "Dynamic Load Balancing and Throughput Optimization in 3GPP LTE Networks," IWCMC, June 28-July 2, 2010, Caen, France.

[19] H. Wang, L. Ding, P. Wu, Z. Pan, N. Liu, and X. You, "Dynamic Load Balancing in 3GPP LTE Multi-cell Networks with Heterogeneous services," ICST Conference, 2010, Beijing, China.