Congestion-Aware Handover in LTE Systems for Load Balancing in Transport Network

Safdar Nawaz Khan Marwat, Sven Meyer, Thushara Weerawardane, and Carmelita Goerg

Long-Term Evolution employs a hard handover procedure. To reduce the interruption of data flow, downlink data is forwarded from the serving eNodeB (eNB) to the target eNB during handover. In cellular networks, unbalanced loads may lead to congestion in both the radio network and the backhaul network, resulting in bad end-to-end performance as well as causing unfairness among the users sharing the bottleneck link. This work focuses on congestion in the transport network. Handovers toward less loaded cells can help redistribute the load of the bottleneck link; such a mechanism is known as load balancing. The results show that the introduction of such a handover mechanism into the simulation environment positively influences the system performance. This is because terminals spend more time in the cell; hence, a better reception is offered. The utilization of load balancing can be used to further improve the performance of cellular systems that are experiencing congestion on a bottleneck link due to an uneven load.

Keywords: LTE, handover, load balancing, congestion, QoS, transport network.

Manuscript received Oct. 14, 2013; revised May 10, 2014; accepted May 27, 2014.

This work was supported by the University of Engineering and Technology of Peshawar, Pakistan and the International Graduate School for Dynamics in Logistics, University of Bremen, Germany.

Safdar Nawaz Khan Marwat (corresponding author, safdar@comnets.uni-bremen.de) and Carmelita Goerg (cg@comnets.uni-bremen.de) are with the Department of Communication Networks, University of Bremen, Germany.

Sven Meyer (sven.meyer@intel.com) is with Intel Mobile Communications Technology GmbH, Dresden, Germany (this work was done during the time Mr. Sven Meyer was writing his diploma thesis at the University of Bremen).

Thushara Weerawardane (tlw@kdu.ac.lk) is with the Department of Electrical and Electronic Engineering, Sir John Kotelawala Defence University, Ratmalana, Sri Lanka.

I. Introduction

With the rising popularity of smartphones as well as flat rates for voice and data services, the requirement for capacity both in mobile and fixed networks has increased rapidly. Additionally, carriers are under pressure to reduce the cost of every transported bit to stay competitive. The demand for higher capacity, reduced costs, and higher cell throughput has led to a fast introduction of Long-Term Evolution (LTE) into the market. LTE can be used to bring fast Internet services to regions that don't have high-speed Internet access; such regions tend to use alternative systems such as digital subscriber line. LTE is designed to reduce system latency and to better exploit the channel quality of its users. Moreover, the system uses beamforming and multiple-input and multiple-output (MIMO) to optimize spectral efficiency and increase peak user throughput.

Mobility in cellular networks is one of its key features. Call drops or long gaps in transmission, which can be perceived by the end user, should be avoided at all costs. This becomes even more critical with LTE since this standard is supposed to support mobile terminals at high velocities. While soft handover and softer handover, both of which are used in the Universal Mobile Telecommunications System (UMTS), are not applicable in LTE, handovers in LTE are performed and are known as "hard handovers." Such hard handovers mean that there is an interruption in reception, and as a result, connection with the network is lost for a short duration. To comply with the strict quality of service (QoS) requirements for Voice-over-Internet Protocol (VoIP), the number of interruptions as well as their durations has to be reduced as much as possible.

The users at the edges of the loaded cells can be transferred to the lowly loaded cells in the neighboring eNodeBs (eNBs)

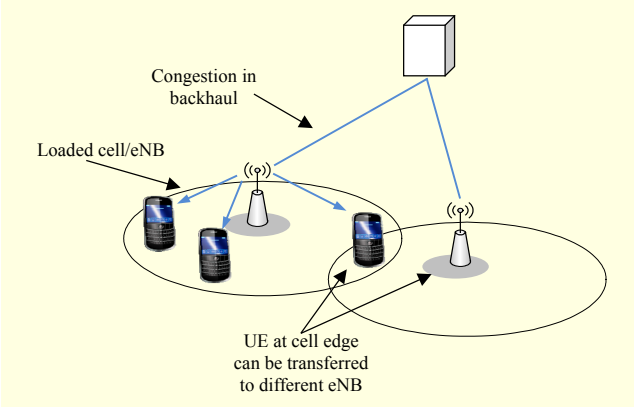

Fig. 1. Backhaul load balancing.

by making inter-eNB and intra-eNB handovers (Fig. 1). Therefore, the offered load through the bottleneck link in the transport network (TN) can be redistributed to other links that are not congested at that time. Moreover, from the point of view of the radio network, cell overloading can be avoided by diverting traffic to the lowly loaded cells. All in all, the TN and radio network can be optimized by applying knowledge-based adaptive handovers; thus, enabling a guaranteed QoS for end users.

Since most ongoing research regarding load balancing concentrates on the detection of high loads within the air interface of cells (for example, [1]–[2]), this paper presents a mechanism for resolving congestion situations in the TN. Therefore, the objective of this paper is to design an LTE handover mechanism using an LTE OPNET simulator. In this work, the OPNET simulator is used to analyze the impact of handover particularly in the case of a congested TN. Moreover, this paper investigates how end-user performance can be enhanced using load balancing in such loaded scenarios.

LTE TN congestion can lead to low end-to-end performance. When offered load over a certain router link is high, the corresponding egress router drops the packets. In such cases, the TCP protocol controls user flow rates over the bottleneck link in the LTE backhaul. This may result in a reduction of the achievable user throughput; consequently, unfairness among the users of the bottleneck links can occur. When cells are unevenly loaded in a particular eNB, it is the last-mile link of that eNB that, more often than not, causes congestion compared to other links in the TN. All these effects result in lower overall LTE performance, which negatively impacts upon both the end users and the network operators.

II. Handover in LTE and Related Work

In contrast to UMTS, LTE does not support soft or softer handovers. Therefore, all handovers are performed as hard handovers. Hence, at some point, the user equipment (UE) will disconnect from the serving eNB (SeNB) and connect to the target eNB (TeNB). Therefore, the UE will be without a connection to a cell for a certain amount of time. The decision as to whether a UE should perform a handover is based on measurements, performed by the UE, that are transmitted to the eNB. Thus, although the decision is made by the eNB, it requires the assistance of the corresponding UE.

The handover from one eNB to another eNB is also referred to as Intra-Mobility Management Entity/Serving Gateway (MME/SGW) handover. This refers to the fact that the Evolved Packet Core (EPC) is not directly involved in the handover and is only informed afterwards. To achieve this, the SeNB has to be able to forward data to the TeNB. After the connection is re-established, the downlink path is switched to the TeNB.

1. Handover Measurements

During the RRC[1)]_Connected mode, the UE continuously performs measurements of the SeNB and neighboring cell signals. A UE can be configured to various measurement options [3]. In general, measurements are based on the reference signals (RS) that are constantly transmitted with every frame. By measuring these signals, the UE determines the reference signal received power (RSRP) and the reference signal received quality (RSRQ). The 3rd Generation Partnership Project (3GPP) specifications [3] describe several events, each of which is triggered under certain conditions, that will cause a UE to generate and transmit a measurement report to the SeNB. The specifications define two main parameters that influence the triggering of the events described in [3]. Time-to-Trigger (TTT) is the timespan required for the entering condition to be fulfilled without triggering the leaving condition, which in turn would trigger the handover. Handover margin (HOM), or hysteresis factor, is a threshold factor that needs to be fulfilled by the measured signal to trigger the entering condition (Fig. 2).

In general, a provider will configure the parameter values so as to have only the minimum necessary number of handovers. This is because each handover consumes valuable network resources that might otherwise be used for the benefit of users. If these settings are not selected carefully, then unnecessary handovers may be triggered. For instance, if the selected TTT value is too small, a handover might be triggered, even if the received RSRP is high for only a very short amount of time. This triggering of handovers might be frequent at cell edges, where the received signal strength of neighboring and SeNB cells alternates a number of times. Frequent handovers result in

1) Radio Resource Control

http://dx.doi.org/10.4218/etrij.14.0113.1034

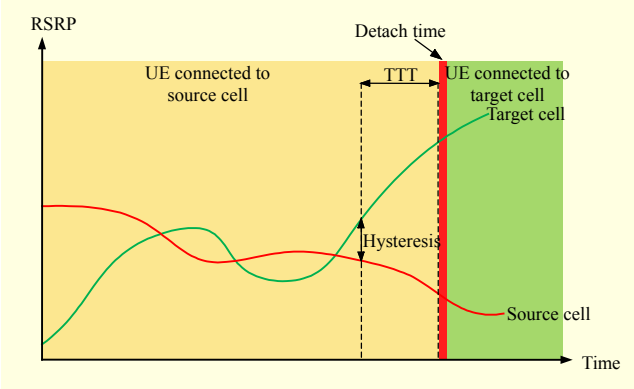

Fig. 2. TTT and HOM parameter selection.

a waste of system resources. This is commonly referred to as the ping-pong effect.

The parameters for the entering and leaving conditions are not specified by the 3GPP. They have to be defined by the provider in such a way that optimal network performance is achieved. Therefore, network providers can set the parameters according to their requirements and network conditions. Some network providers might optimize their network to the point where only a minimum number of call drops are to be tolerated, whereas other providers might favor to optimize their network to reduce the number of unnecessary handovers.

To optimize the selection of TTT values, the 3GPP specifications recommend grouping all UEs of a cell into three groups according to their velocity. The estimation of UE velocity is also a focus of ongoing research. For instance, [4] proposes an algorithm for velocity estimation that is based on the variance of the channel frequency response of two adjacent OFDM symbols. Other papers take a different approach to estimating the speed of a UE; for instance, through measuring the fading slope duration [5] or through Rician fading channels [6]. Reference [7] presents a unified approach to the performance analysis of speed-estimation techniques and presents several such major techniques.

The selection of handover-triggering parameters is a crucial point and is also the focus of current research. Reference [8] compares different methods on how to assign TTT values according to a UE's velocity. Other research papers propose self-optimizing mechanisms that constantly evaluate the experienced rate of call drops or the rate of unnecessary handovers [9]–[10]. Other papers recommend adaptive Layer 3 filtering based on either UE velocity [11] or radial velocity [12].

2. Phases of Handover

The handover process is divided into three phases [13]: preparation, execution, and completion.

A. Handover Preparation

In the handover preparation phase, the SeNB has to acquire information from the TeNB to determine whether it will accept another UE. This is initiated by the SeNB sending the handover request message to the TeNB. Consequently, the TeNB will perform admission control (AC) to decide whether it should accept new Evolved Packet System (EPS) bearers. To do this, the AC in the eNB has to estimate whether the eNB is able to successfully provide the required resources for the EPS bearers of the new user, while still being able to provide enough resources for users that are already in the cell. Thus, the eNB determines the QoS requirements for the new EPS bearers of that user, its priority levels, and the QoS that is provided to the active sessions in the cell. Admission rights for the new user are only granted if the QoS requirements can be fulfilled and all existing connection that have the same or higher priority can be served with an acceptable quality. The result of the AC will be transmitted to the SeNB within the handover-request acknowledge message. After AC has been performed, the SeNB initiates the handover.

B. Handover Execution

The handover execution phase starts with the UE disconnecting from the SeNB and the sending of the sequence number (SN) status transfer message. To establish a forwarding tunnel to the TeNB, all necessary information, such as the Packet Data Convergence Protocol (PDCP) SN and hyper frame number for the UE, is transferred to the TeNB with the "SN Status Transfer" message. The eNB sends the SN Status Transfer message when it considers the transmitter/receiver status to be frozen. Once this has been done, the SeNB will forward all incoming downlink data to the TeNB over the X2 interface. The X2 interface connects the eNBs of the network and allows direct forwarding of user and control plane data without the involvement of the EPC. The X2 interface helps to reduce the time span of the period where no data is being received at the UE; thereby, ensuring the strict delay requirement for a seamless handover. All data that is forwarded is buffered at the TeNB until the UE reconnects. After re-attachment of the UE to the TeNB, the SeNB transmits the buffered data and new incoming data over the "old path," as indicated in Fig. 3. In the uplink direction, data is sent directly from the TeNB to the MME/Gateway over the "new path."

C. Handover Completion

After the UE is attached to the TeNB and a data connection is re-established, the TeNB will instruct the gateway or the MME to switch the downlink path to the TeNB. The SGW will switch the downlink path to the TeNB and will send the end

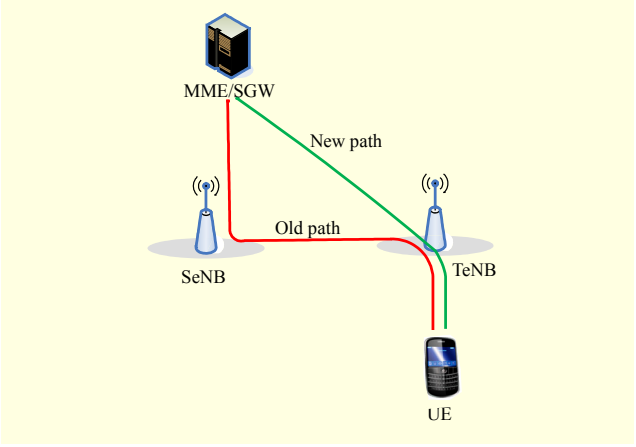

Fig. 3. Switching of data path by MME.

marker over the old path. This is done to inform the SeNB that no further data is to be sent from the SGW to the SeNB. The SGW has to send at least one end marker for every E-UTRAN[2] Radio Access Bearer (E-RAB). After sending the end marker, no more data should be transmitted over the old path. Therefore, the SeNB has to discard every packet that arrives after the end marker. The SeNB has to forward the end marker to the TeNB. During the handover procedure, in particular during the switching of the downlink path, packets might arrive interchanged via the old or new paths. The PDCP layer has to ensure in-sequence delivery of packets.

III. Load Balancing

Load balancing describes the process of passing on users of a highly loaded cell to neighboring cells with low load. By balancing the load among the cells of the network, the overall system performance of the cellular network can be improved. This is due to the efficient use of the unused resources of the less-loaded cells.

In current networks, system parameters are adjusted manually to improve the performance. According to 3GPP, the basic concept behind a self-optimizing network (SON) is that it is able to auto-tune system parameters based on network optimization measurements. The 3GPP introduced LTE, an example of an SON, and proposed SON algorithms for its self-configuration and self-optimization. Thus, LTE supports capacity planning, interference reduction, and automated configuration of newly introduced cells.

Load balancing can be realized in several ways. For example, by modifying the pilot power and reducing the cell size. This is also referred to as load balancing through cell breathing. Another way is to modify the offset values of users that are close to the cell edge. By increasing the offset value for the serving cell and assigning a negative offset value to a target cell, the handover triggering algorithm in the UE is forced to trigger a handover. The offset values ensure that after handover completion, the UE will not return to the highly loaded cell. This method is referred to as "mobility load balancing." Mobility load balancing can also be utilized to reduce or avoid congestion in the TN.

Several approaches have been investigated in the literature. In [14], an analytical framework is used to study the performance of dynamic load balancing schemes for cellular networks. The authors of [15] investigate partial frequency reuse with load balancing to improve network performance. The performance of probability-triggered and utility-based load balancing schemes is provided in [16]. The authors of [17] summarize several load balancing schemes and propose an algorithm based on considerations of system overhead. In [18], a joint mobility load balancing algorithm is proposed, which takes the load situation of the radio network and the TN into account. The use of the aforementioned load balancing mechanisms usually results in a reduced call blocking rate and higher cell edge throughput. In our work, we utilize the One-Way Active Measurement Protocol (OWAMP) to achieve load balancing, which has not been used before for load balancing.

IV. Handover Implementation

In this work, we implemented the handover mechanism using an LTE model in the OPNET simulation environment [19]. OPNET is used as the primary modeling, simulation, and analysis tool of this paper. The designed LTE simulation model consists of several nodes with LTE functionalities and protocols (Fig. 4). The remote server node supports user applications in all the cells and acts as a sink for the uplink data. The remote server and the access gateway (AGW) (Fig. 4) nodes are interconnected via an Ethernet link of 20 ms average delay. The AGW node consists of certain peer-to-peer protocols, such as User Datagram Protocol, Internet Protocol (IP), and Ethernet, toward the TN and other peer-to-peer protocols toward the remote server. The AGW and eNB nodes (eNB1, eNB2 …) are connected through the TN of IP routers (R1, R2), which are configured according to the standard OPNET differentiated services model and routing protocols. The QoS parameters for the traffic differentiation at the TN are configured at the QoS Parameters node. The mobility node consists of the mobility and channel models. The mobility model emulates the user movement in a cell and updates the user location at every sampling interval. The user mobility information is stored in the global user database (Global_UE_List in Fig. 4) and is accessible from every node in the system

2) Evolved UMTS Terrestrial Radio Access Network

ETRI Journal, Volume 36, Number 5, October 2014
http://dx.doi.org/10.4218/etrij.14.0113.1034

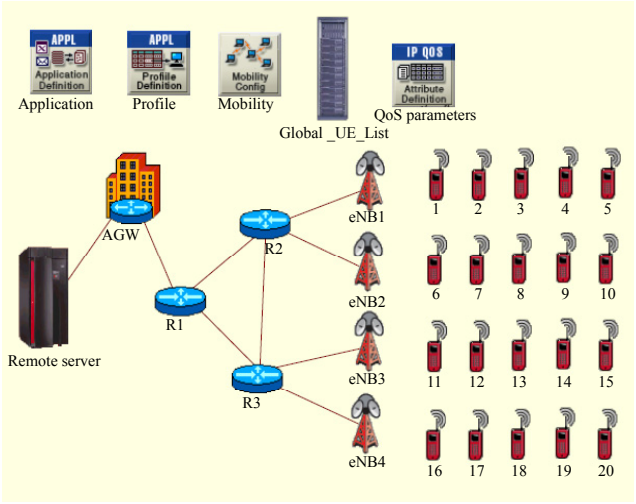

Fig. 4. OPNET project editor.

at any given time. The channel models include path loss as well as slow fading and fast fading models. The mobility of the UEs is modeled with a random waypoint model and a random direction model.

1. Handover Mechanism

Since there are no standard values defined for the handover related parameters, the TTT and HOM values are set in accordance with [9], [10], and [20] in this work (Table 1). As proposed in the specifications, UEs are categorized according to their speed and assigned different TTT values (if specified). In addition to the TTT and HOM values, a ping-pong timer has been implemented, which helps in avoiding unnecessary handovers. The timer is activated upon a handover and prevents the triggering of another handover until this timer is expired (Table 1).

All the sent handover-related RRC messages as well as the AC that is performed during the handover preparation phase are modeled by specifying an appropriate delay (Table 1). The handover execution phase is implemented so that the detachment of the UE is initiated when the state informs the MAC layer to stop scheduling packets for the UE. This is realized by setting a UE-specific state variable, which indicates that the UE is now disconnected. Afterwards, all packets still in the buffer are forwarded to the TeNB. To initiate the same procedure within the UE, an interrupt is also sent to the PDCP buffer of the UE. The PDCP of the UE also has a state called "disconnecting," which resets all buffers as well as SN. The detachment duration is simulated by another timer, which is called "handover execution delay" (Table 1). The timer specifies the time duration, during which the scheduling for the user is paused. Upon expiration of this timer, the MAC layer

Table 1. Handover settings for simulations.

Parameter	Value
TTT	480 ms
HOM	6 dB
Load balancing offset	9 dB
Ping-pong timer	1 s
Handover preparation delay	40 ms–43 ms
Handover execution delay	50 ms–70 ms
Handover completion delay	15 ms–30 ms

automatically starts the scheduling of the buffered and new incoming packets. Upon UE detachment, the SeNB starts forwarding incoming packets to the TeNB. This applies to data that is incoming over the new path and to data that is still in the PDCP buffer but has not yet been transmitted. After the handover execution phase, the UE is reattached with the TeNB, and data in the uplink and downlink directions is transmitted. Since downlink data is still incoming over the old path, the switching of the data path is initiated. Therefore, a delayed interrupt is sent from the MAC layer to the AGW. This interrupt refers to the "path switch" request message (specified in the 3GPP specifications).

2. Load Balancing Mechanism

For load balancing for the TN links, a mechanism able to measure the delay on the network links is implemented. The OWAMP, standardized by the IP Performance Metrics working group of the Internet Engineering Task Force (IETF) [21], is used for detection of delay in the traffic network. The OWAMP server in the AGW transmits test packets that are received by the OWAMP Client in each eNB. The packet delay may vary, depending on the load situation of the TN. It is assumed that the network will experience congestion as soon as a certain delay threshold is reached. The congestion indication interrupt is then sent to the PDCP layer of the same eNB. The PDCP layer then initiates load balancing to resolve the critical load situation of that particular link by assigning an offset of –9 dB to all the UEs connected to the congested eNB, which increases the probability to initiate a handover to a neighboring cell. The load balancing is turned on for a delay of 0.08 s and then turned off for 0.02 s.

3. Intra eNB Handover

For an intra eNB handover, both cells involved in the handover belong to the same eNB. Therefore, path switching by the AGW, as previously described, is not required. In our

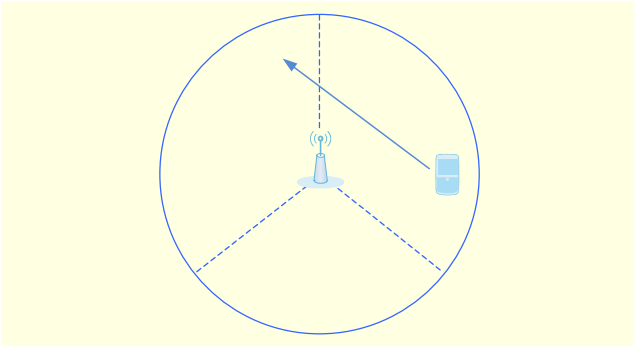

Fig. 5. Intra eNB handover.

simulation model, the simulation area is divided into three sectors, as depicted in Fig. 5. The position of the UE is determined by the mobility model. When a UE crosses the sector line, a cell sector change is initiated. To simulate the synchronization with the target cell, the MAC scheduler for the UE is paused for a certain amount of time, which is similar to the handover execution delay (detach time) in the inter eNB handover. Since the forwarding of data over the X2 connection is not necessary in intra eNB handover, only the detach time of the UE is taken into account. The new target cell is geometrically determined by the mobility model.

V. Simulation Results

The simulations are performed under the settings illustrated in Table 2. The simulations are performed to analyze the impact of handover on system performance during network congestion and the application of handover for load balancing under several load situations.

A varying number of users with different traffic classes are simulated. The settings of the traffic models are listed in Table 3. We demonstrate how the different traffic types are handled and prioritized by the TN. The two main traffic classes are guaranteed bit rate (GBR) and non-GBR.

Services like File Transfer Protocol (FTP) or Hypertext Transfer Protocol (HTTP), which typically use the TCP protocol, are usually categorized as non-GBR services. These kinds of services are assigned with free resources of the LTE system after all GBR bearers have been served.

1. QoS Performance under Different Load Situations

In this subsection, the implemented LTE handover mechanism is tested under different load situations. We compare three scenarios. In scenario A, 27 VoIP, 27 HTTP, and 27 FTP users are simulated. In scenario B, the number of users is raised to 36 for all traffic types. In scenario C, 45 users of all the aforementioned traffic classes are simulated. An increase of active users results in an increased throughput in both the downlink and the uplink directions. The downlink and uplink MAC schedulers are designed as in [19] and [23], respectively.

In Fig. 6, the accumulated throughput for all three cells of eNB1 is depicted for the simulated scenarios. It shows that with

Table 2. Simulation parameters.

Parameter	Value
Simulation length	1000 s
Number of eNBs	4
Cells per eNB	3
eNB coverage radius	350 m
Cell distance	640 m
Min. eNB – UE distance	35 m
Max UE power	23 dBm
User mobility profiles	120 km/h
Mobility model	Random Way Point (RWP)
Frequency reuse factor	1
Transmission bandwidth	5 MHz (for downlink and uplink each)
No. of PRBs	25 (for downlink and uplink each)
MCS	QPSK, 16 QAM, 64 QAM,
Path loss	$128.1 + 37.6 \log_{10}(R)$, R = distance in kilometers
Slow fading	Log-normal shadowing, 8 dB standard deviation, correlation 1
Fast fading	Jakes-like method [22]
UE buffer size	Infinite
Power control, α	0.6
Power control, P_0	–58 dBm

Table 3. Traffic models.

Simulation parameter	Setting
VoIP traffic model	
Silence length	Exponential distribution, 3 s mean
Talk spurt length	Exponential distribution, 3 s mean
Encoder scheme	GSM EFR (Enhanced Full Rate)
Call duration	Till the end of simulation
HTTP traffic model	
Page size	100 kB
Page interarrival time	12 s
FTP traffic model	
File size	20 MB
File inter-request time	Uniform distribution (80 s to 100 s)

ETRI Journal, Volume 36, Number 5, October 2014
http://dx.doi.org/10.4218/etrij.14.0113.1034

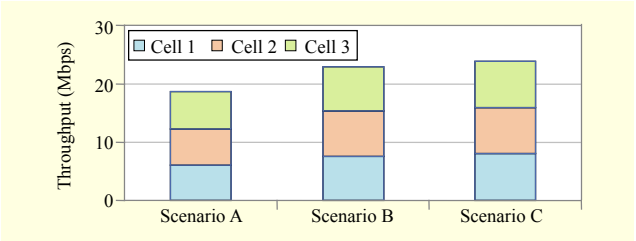

Fig. 6. Total Throughput of eNB1 (Mbps).

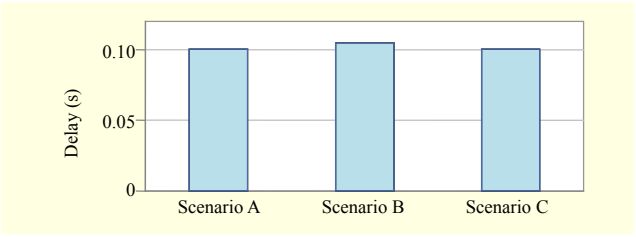

Fig. 7. VoIP average packet end-to-end delay.

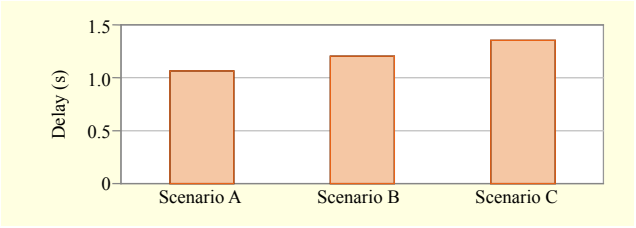

Fig. 8. HTTP average page response time (s).

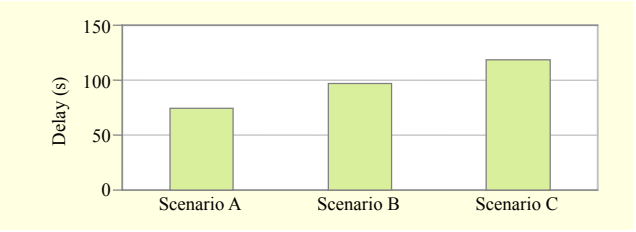

Fig. 9. FTP average download response time.

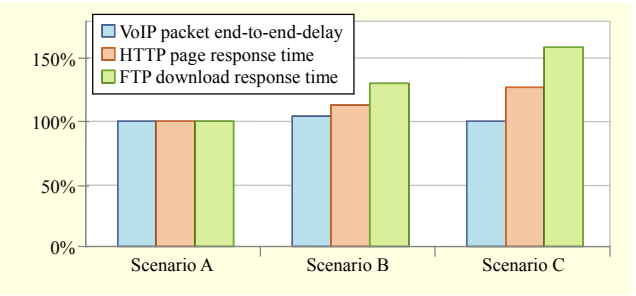

Fig. 10. Change of delay in percent of all three services.

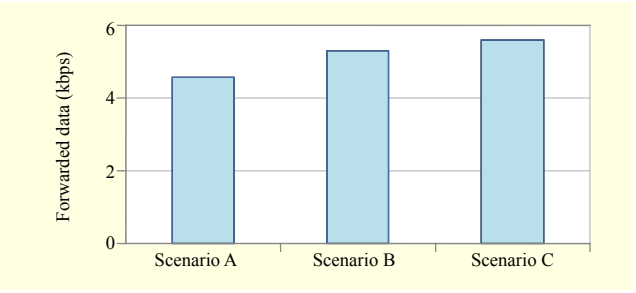

Fig. 11. Average forwarded data over X2 (eNB1 to other eNBs).

an increased number of users, the total throughput for an eNB increases. But, the throughput can only increase until all radio resources have been assigned. If even more users are added to the eNB afterwards, then the users have to share the radio resources. In Fig. 6, the total throughput increase from scenario B to scenario C is not as steep as from scenario A to scenario B. This is an indication that the cells of eNB1 are fully loaded.

The GBR connections, such as VoIP, are scheduled with the highest priority by the MAC scheduler. Hence, increasing the traffic load has no significant impact on the performance of VoIP traffic. The average packet end-to-end delay for VoIP users in eNB1 is depicted in Fig. 7.

On the other hand, non-GBR classes, such as HTTP and FTP, are scheduled with low priority, and the physical resource blocks (PRBs) leftover from GBR classes are allocated. This results in a degradation of performance by increasing the load. The average page response time for HTTP and the average download time for a FTP file are depicted in Figs. 8 and 9, respectively.

Figure 10 depicts the percentage of increase of average delay for each QoS type with increasing traffic load, showing that the FTP download response time is affected the most by a loaded cell. This indicates that the scheduling of FTP data is given low priority by the MAC scheduler, although the order of prioritization of different traffic classes is vendor specific.

Figure 11 depicts the average throughput for data that is forwarded to the target cell over the X2 interface. The X2 interface is used for data forwarding during the handover when the UE is connected to the TeNB. Meanwhile, the downlink data is still being sent to the SeNB and then forwarded to the TeNB. An increased cell load causes the amount of data transferred over the X2 interface to increase.

2. Congestion Detection with OWAMP

The influence of a congested transport link on the system performance is investigated in this subsection. System performance in terms of delay on the TN is measured with the OWAMP algorithm. The congestion in the network is simulated by limiting the maximum possible throughput for the link between router R2 and the eNB1 (Fig. 4). In these simulations, each eNB is set up with 30 UEs comprising 12

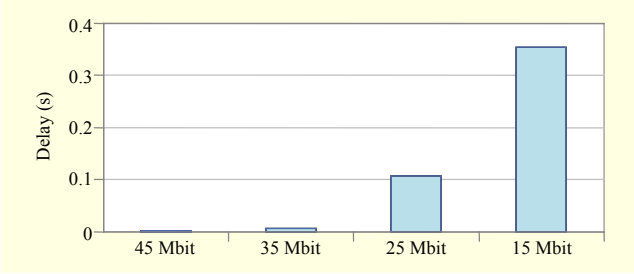

Fig. 12. Packet delay buildup recorded by OWAMP.

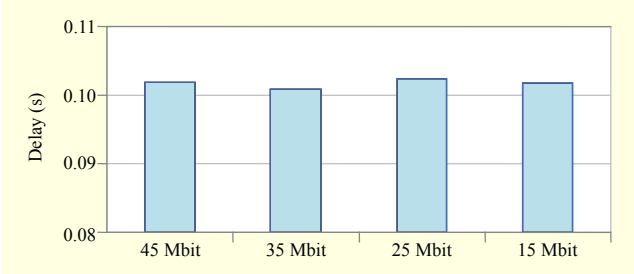

Fig. 13. Average VoIP packet end-to-end delay.

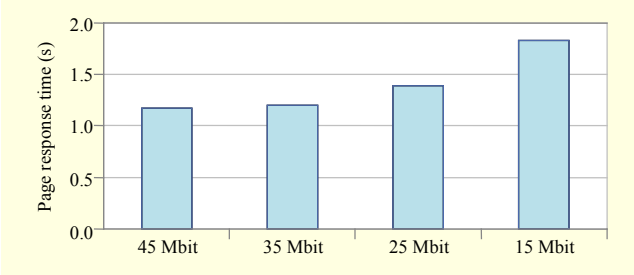

Fig. 14. Average HTTP page response time.

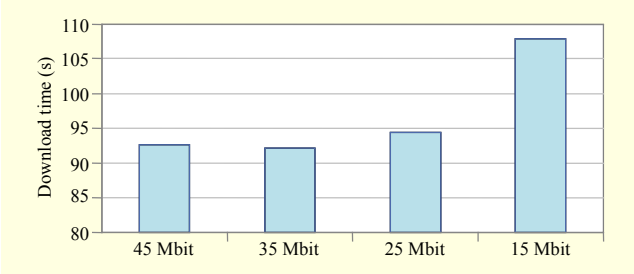

Fig. 15. Average FTP download duration.

VoIP users, 12 HTTP users, and 6 FTP users. The maximum throughput of the chosen link is reduced stepwise, starting with a throughput of 45 Mbps down to 15 Mbps.

Decreasing the maximum link capacity increases the packet delay measured for eNB1 (Fig. 12). It is shown that the measured delay for a throughput of 45 Mbps and 35 Mbps is quite low, but it increases dramatically if the throughput is decreased further to 25 Mbps or 15 Mbps (heavily congested).

Since routers in the TN prioritize GBR packets for services such as VoIP, the delay of those packets remains almost constant for all scenarios. This effect is depicted in Fig. 13, which displays the average end-to-end delay for VoIP packets. Even if the link throughput is reduced to 15 Mbps, the packet end-to-end delay does not change significantly.

However, non-GBR classes, such as HTTP or FTP, experience an increase in average delay, as depicted in Figs. 14 and 15. The increase of the transport delay depends on the prioritization applied within the TN. These figures also demonstrate that the increase of the page response time and the download time becomes severe in the scenario with a maximum throughput of 15 Mbps.

3. Impact of Load Balancing

This subsection demonstrates how knowledge about the load and the delay situation within the TN, gathered by the OWAMP algorithm, can be utilized to apply load balancing. The implementation of the OWAMP algorithm allows the definition of a threshold level to identify the congestion level in the TN, which defines when load balancing is to be activated. Upon this, load balancing is activated by the PDCP layer of the eNB1 to control the number of UEs connected to it.

Figure 16 shows that reducing the throughput causes the load balancing algorithm to further reduce the number of users connected to eNB1. Without load balancing, the number of users remains almost constant for all eNBs within the simulation area. For the link throughputs of 45 Mbps and 35 Mbps, the number of users remains relatively unchanged. This is due to the fact that almost no congestion occurs in these scenarios; therefore, the load balancing algorithm is not

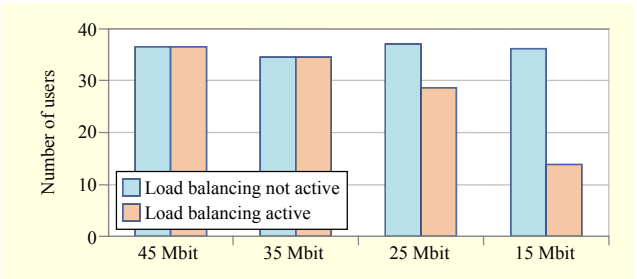

Fig. 16. Average number of users connected to eNB1.

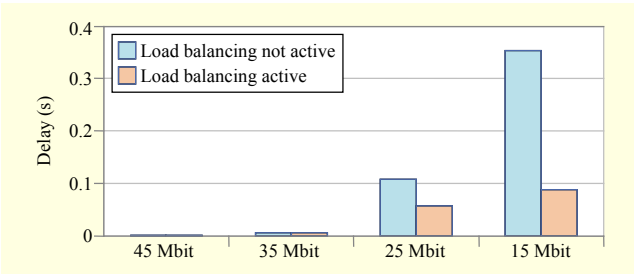

Fig. 17. Average delay measured with OWAMP.

ETRI Journal, Volume 36, Number 5, October 2014
http://dx.doi.org/10.4218/etrij.14.0113.1034

activated. If the throughput is reduced further, then the congestion in the link increases; consequently, the number of times the load balancing algorithm is active also increases. As shown, the number of users is reduced to an average of below 15 UEs connected to eNB1 for a throughput of 15 Mbps.

Figure 17 displays the OWAMP-recorded delay by comparing the scenarios with and without active load balancing. The results are consistent with those in Fig. 16, showing an unchanged delay for a bandwidth of 45 Mbps and 35 Mbps. With an increased delay in the link, the load balancing algorithm is activated frequently; and by reducing the number of users connected to eNB1, the delay is reduced.

The reduction of the delay on the link to eNB1 results in a better system performance for users that are connected to eNB1. In Fig. 18, the VoIP packet end-to-end delay for the complete system is compared for different throughput scenarios with load balancing activated and deactivated. The results depict the average delays for UEs that are connected to the eNBs affected by the congested link and for those that are connected through non-congested links. As in Fig. 18, the VoIP delay is not affected by the congestion significantly. Therefore, there is almost no difference if load balancing is activated or deactivated. As observed in previous simulations, this is due to the high priority given over to VoIP traffic.

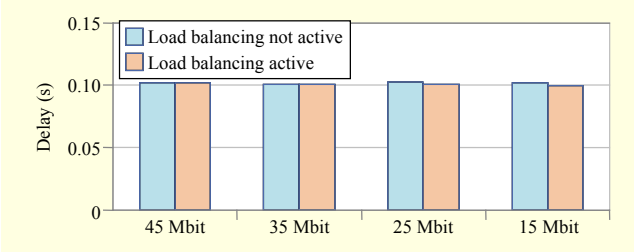

Fig. 18. Average VoIP packet end-to-end delay.

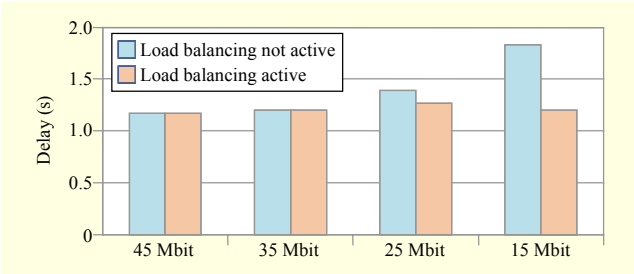

Fig. 19. Average HTTP page response time.

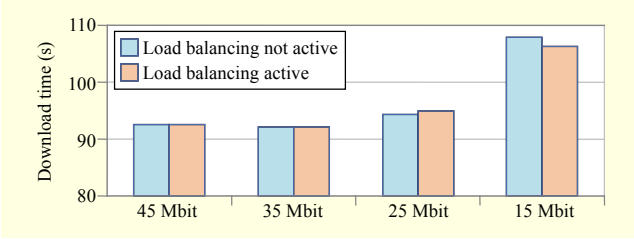

Fig. 20. Average FTP download duration.

Other services used in these simulations suffer from an increased delay. Here, active load balancing can improve the system performance considerably. The page response time for the HTTP service remains almost constant for all scenarios of this section when the load balancing is activated, as illustrated in Fig. 19. Thus, even in situations with heavy congestion, HTTP users experience almost no change in performance.

There is no significant improvement in FTP download duration time (Fig. 20). On the one hand, the UEs shifted from a congested eNB will experience a reduced download time; while on the other hand, UEs that are connected to non-congested eNBs will be joined by even more UEs. Therefore, those UEs need to share the resources that can be assigned to the FTP service. The low priority of FTP traffic undoubtedly means that it fails to benefit from load balancing.

VI. Conclusion

This work presents the design and implementation of a handover and load-balancing mechanism for an LTE system model. The handover performance under different load scenarios is studied by simulating various numbers of users with various traffic classes. The results illustrate the impact of high load on the performance of various traffic types and throughput. The load balancing algorithm utilizing the OWAMP algorithm to detect congestion in the TN is analyzed in the scenarios with congested links within the TN. It is shown that these bottleneck links cause delays and degrade the overall system performance. The load balancing algorithm is able to shift users to less-loaded eNBs, thereby resolving the congestion in the TN. This is done by configuring an offset parameter in the handover decision mechanism, which causes UEs that are at the cell edge to handover to another less-loaded eNB. The results show that the load balancing algorithm can help in balancing the load among the network components.

In future, we intend to study self-optimizing offset values for load balancing. Furthermore, we would also look for ways to resolve congestion in radio interfaces along with TN.

References

[1] L. Yun et al., "Dynamic Optimization of Handover Parameters Adjustment for Conflict Avoidance in Long Term Evolution," *China Commun.*, vol. 10, no. 1, Jan. 2013, pp. 56–71.

[2] N.N. Rapiei et al., "Handover Mechanism in Dynamic Load Balancing for LTE System," *IEEE Symp. Wireless Technol. Appl.*, Bandung, Indonesia, Sept. 23–26, 2012, pp. 43–47.

[3] 3GPP Technical Specification 36.331 V 11.3.0, *Evolved Universal*

Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol Specification, Mar. 2013.

[4] Z. Shu et al., "Fast and Accurate Velocity Estimation for OFDM System Based on Channel Frequency Response," *IEEE Veh. Technol. Conf.*, Yokohama, Japan, May 15–18, 2011, pp. 1–5.

[5] L. Zhao and J.W. Mark, "Mobile Speed Estimation Based on Average Fade Slope Duration," *IEEE Trans. Commun.*, vol. 52, no. 12, Dec. 2004, pp. 2066–2069.

[6] Y.R. Zheng and C. Xiao, "Mobile Speed Estimation for Broadband Wireless Communications over Rician Fading Channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, Jan. 2009, pp. 1–5.

[7] A. Abdi, H. Zhang, and C. Tepedelenlioglu, "A Unified Approach to the Performance Analysis of Speed-Estimation Technique in Mobile Communication," *IEEE Trans. Commun.*, vol. 55, no. 12, Dec. 2007, p. 2381.

[8] Y. Lee et al., "Effects of Time-to-Trigger Parameter on Handover Performance in SON-Based LTE System," *Asia-Pacific Conf. Commun.*, Auckland, New Zealand, Oct. 31–Nov. 3, 2010, pp. 492–496.

[9] J. Alonso-Rubio, "Self-optimization for Handover Oscillation Control in LTE," *IEEE Netw. Operations Manag. Symp.*, Osaka, Japan, Apr. 19–23, 2010, pp. 950–953.

[10] T. Jansen et al., "Handover Parameter Optimization in LTE Self-organizing Networks," *IEEE Veh. Technol. Conf.*, Ottawa, Canada, Sept. 6–9, 2010, pp. 1–5.

[11] H. Zhang et al., "A Novel Self-optimizing Handover Mechanism for Multi-service Provisioning in LTE-Advanced," *Int. Conf. Res. Challenges Comput. Sci.*, Shanghai, China, Dec. 28–29, 2009, pp. 221–224.

[12] J. Gu et al., "Mobility-Based Handover Decision Mechanism to Relieve Ping-Pong Effect in Cellular Networks," *Asia-Pacific Conf. Commun.*, Auckland, New Zealand, Oct. 31–Nov. 3, 2010, pp. 487–491.

[13] 3GPP Technical Specification 36.300 V 11.5.0, *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall Description; Stage 2*, Mar. 2013.

[14] O.K. Tonguz and E. Yanmaz, "The Mathematical Theory of Dynamic Load Balancing in Cellular Networks," *IEEE Trans. Mobile Comput.*, vol. 7, no. 12, Dec. 25, 2008, pp. 1504–1518.

[15] K. Son, S. Chong, and G. Veciana, "Dynamic Association for Load Balancing and Interference Avoidance in Multi-cell Networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, July 2009, pp. 3566–3576.

[16] A.S. Adeyemi and D.U. Ike, "A Review of Load Balancing Technique in 3GPP LTE System," *Int. J. Comput. Sci. Eng.*, vol. 2, no. 4, July 4, 2013, pp. 112–116.

[17] Y. Yang, P. Li, and W. Wang, "Algorithm about Mobility Load Balance Considering System Overhead on LTE System," *Wireless Opt. Commun. Conf.*, Chongqing, China, May 16–18, 2013, pp. 231–235.

[18] L. Zhao et al., "Joint Load Balancing of Radio and Transport Networks in LTE System," *Int. Conf. Ubiquitous Future Netw.*, Dalian, China, June 15–17, 2011, pp. 65–70.

[19] Y. Zaki et al., "Long Term Evolution (LTE) Model Development within OPNET Simulation Environment," *OPNET Workshop*, Washington, DC, USA, Aug. 29–Sept. 1, 2011.

[20] M. Anas et al., "Performance Evaluation of Received Signal Strength Based Hard Handover for UTRAN LTE," *IEEE Veh. Technol. Conf.*, Dublin, Ireland, Apr. 22–25, 2007, pp. 1046–1050.

[21] S. Shalunov and B. Teitelbaum, "One-Way Active Measurement Protocol (OWAMP) Requirements," The Internet Eng. Task Force (IETF), RFC 3763, Apr. 2004.

[22] J.K. Cavers, *Mobile Channel Characteristics*, Dordrecht, Netherlands: Kluwer Academic Publishers, 2002.

[23] S.N.K. Marwat et al., "Performance Evaluation of Bandwidth and QoS Aware LTE Uplink Scheduler," *Int. Conf. Wired/Wireless Internet Commun.*, Santorini, Greece, vol. 7277, June 6–8, 2012, pp. 298–306.

Safdar Nawaz Khan Marwat received his BS degree in computer systems engineering at the University of Engineering and Technology, Peshawar, Pakistan, in 2006. He worked as a lecturer at Peshawar College of Engineering, Peshawar, Pakistan, from 2006 to 2008. He obtained a master-leading-to-PhD scholarship from the University of Engineering and Technology, Peshawar, Pakistan, in 2008. He received his MS degree in communication and information technology at the University of Bremen, Germany, in 2011. Since 2011, he has been a PhD student at Communication Networks Group (ComNets), University of Bremen, Germany. His research focuses on machine-to-machine communications and future mobile networks.

Sven Meyer received his diploma in electrical engineering, with an emphasis on communication and information technology, from the University of Bremen, Germany, in 2012. His diploma thesis on "Congestion-Aware Handover in LTE" was awarded with the FFV Prize for the best thesis 2013 by the ComNets Friends and Promoters Club. Since 2012, he has been working as an LTE firmware engineering specialist at Intel Mobile Communications GmbH, Dresden, Germany. His research interests include radio resource management and firmware for LTE networks.

Thushara Weerawardane received his BS degree in electrical engineering at the University of Moratuwa, Sri Lanka, in 1998 and his MS degree in communication and information technology at the University of Bremen, Germany, in 2004. He worked as an assistant network manager in the Department of Electrical Engineering and also as a system engineer for Lanka Educational and Research Network, Sri Lanka, from 1999 to 2002. From 2004 to 2012, he was a research scientist at ComNets, University of Bremen, Germany, where he was responsible for research projects on UMTS, HSPA, and LTE performance optimization, all of which were funded by Nokia Siemens Networks. Since 2012, he has been working as a senior lecturer in the Department of Electrical, Electronic, and Telecommunication Engineering at General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka. During his career, he has published more than eight journal publications and over thirty conference publications.

Carmelita Goerg received her diploma in computer science from the University of Karlsruhe, Germany, in 1976. She obtained her Doctor rerum naturalium degree in traffic theory and appointment as lecturer from the Faculty of Electrical Engineering, Aachen University of Technology, Germany, in 1983. From 1985 to 1989, she worked as a consultant in the field of communication networks. From 1989 to 1999, she was at first a group leader and then later (1997) an assistant professor at the Communication Networks Institute, Aachen University of Technology, Germany. Since 1999, she has been a professor at the University of Bremen, Germany. Her research interests include performance analysis of communication networks; stochastic simulation; rare-event simulation; wireless networks; mobility support; new services and applications; and network virtualization.