

A Two-layer Mobility Load Balancing in LTE Self-Organization Networks

Lin Zhang, Yu Liu, Mengru Zhang, Shucong Jia, and Xiaoyu Duan
Key Lab of Universal Wireless Communications, Ministry of Education
Beijing University of Posts and Telecommunications, Beijing, China
zhanglin@bupt.edu.cn

Abstract—Self-organization and self-optimization of mobile wireless systems is of utmost importance to operators and vendors and investigated for 3GPP Long Term Evolution (LTE). As one of the most important self-optimization issues, Mobility Load Balancing (MLB) has received much attention until now. In this paper, we present a novel method named as TL-MLB (Two-Layer Mobility Load Balancing) in which the over-load cell can choose target cell according to its neighboring cells' load and the surrounding environments, and trigger handover behaviors of users by automatically adjusting the cell-specific offsets. The proposed TL-MLB is subject to a system-level simulation which witnesses an improvement in load distribution index, number of unsatisfied users, total handover times and average throughput.

Keywords- self-organization networks; load balancing; handover; Long Term Evolution

I. INTRODUCTION

To face the ever-growing demand for packet-based mobile broadband systems, the Third-Generation Partnership Project (3GPP) has introduced Long Term Evolution (LTE) specifications [1], in which an enhanced access network and an evolved core network have been defined. The self-organization capability of a mobile network mainly includes three aspects: self-configuration, self-optimization, and self-healing. The self-configuration capability enables fast installation and deployment of future evolved NodeBs (eNBs). Moreover, the newly added eNBs can be integrated in a plug-and-play approach. Comparably, self-optimization techniques enable a mobile network to automatically select and adjust proper algorithms and system parameters, to achieve optimal system capacity and service coverage. Finally, self-healing assists operators in recovering a network when it collapses due to some unexpected reason.

Load balancing among radio access points is very important for user-experience improving by providing users with a more flexible and convenience service environment. The inefficient resource utilization motivates the research and development of more efficient network architecture and new technologies, such as mobility load balancing (MLB) [2]. MLB is defined as the overloading conditions of cells can be detected without manual involvement, and balance actions can be taken in an automatic way to resolve the overloading by shifting traffic towards the

lightly-load cells, consequently, make the use of the radio resource more efficient across the whole network.

The rest of this paper proceeds as follows. We first describe mobility load balancing in LTE in section II. Based on the analysis of MLB and the related works, Section III proposes the new algorithm named as Two-Layer Mobility Load Balancing (TL-MLB), in which the over-load cell can choose target cell according to its neighboring cells' load and the surrounding environments. In section IV, a system-level simulation model is presented and simulation results are analyzed. The paper is brought to a conclusion in section V.

II. MOBILITY LOAD BALANCING

An Evolved Packet System is consisted of an Evolved Packet Core (EPC), Evolved Node B (eNB) and user equipment (UE). In existing load balancing methods, when the load ratio of a served cell exceeds a preset threshold, it would first select the cell with the smallest load ratio from its neighbor cells as the target to execute load balancing. Then select the appropriate users in the overload cell to switch. Finally the users are adjusted (switching, reselection and directed reselection, etc.) to the target cell to achieve load balancing between cells.

There has been a lot of research done to equalize load among cells. Ref [3] presented a traditional handover approach to achieve load balancing, which chooses the highest physical resource block (PRB)-utilization cell as the source cell, and the lowest PRB-utilization adjacent cell as the target cell. In [4], the cell-specific offset is adjusted automatically based on the source cell load and its neighboring cell condition. In [5], a method for load estimation after handover would occur is proposed, which is based on SINR prediction and user measurements. In [6], a mathematical perspective of self-optimizing wireless networks is introduced, but the authors just gave a simple simulation about load balancing in the style of "cell-wise", not "cell-pair-wise" style. In [7], a Autonomic Flowing Water Balancing Method (AFWBM) to achieve ALB for LTE RAN is presented. When over-load conditions detected by the AFWBM modules in eNBs, the HOM (handover hysteresis margin) will be adjusted and handover actions will be triggered to balance load.

However, these kind of methods only consider the load of neighboring cells of the overload cell (we call it 'source cell'), without taking the load of the neighboring cell' adjacent cells into account, which risks the possibility that the overload

This work is partially supported by the 111 Project (No.B08004), the major project of Ministry of Industry and Informationization of China (2010ZX03002-006), and the Fundamental Research Funds for the Central Universities.

conditions of source cell alleviated, but the target cell starts suffering high load conditions. In some cases, a lightly-load cell may be chosen by several adjacent over-load cells as target cell, which may result in a new over-load cell while there are some other lightly-load cells could be chosen. Moreover, it is very likely that there is a cell around the source cell without a lowest load ratio itself but has relatively low load conditions in its surrounding area (except source cell). Then it is more suitable to be the target cell. Therefore, the conventional method which relies just on the target cell's load ratio regardless of its environment is one-sided; more comprehensive considerations are needed in order to make better judgments for the entire system.

III. THE TWO-LAYER MOBILITY LOAD BALANCING ALGORITHM

In this paper, a novel method to achieve load balancing for LTE RAN is presented, which is named Two-Layer Mobility Load Balancing (TL-MLB). The proposed method takes into account of the environment of the candidate lightly-load cells and automatically copes with unequal traffic load by tuning cell-specific offset in order to improve the system capacity and resource utility and minimize the number of handovers and redirections. In TL-MLB, the target cell is selected by the following procedures:

- 1) When the load of a cell (or its load ratio) exceeds a certain threshold, it is identified as the source cell. It will send request to all neighboring cells (denoted as the first layer cells), the request information includes load state and the environment state. The environment state is the load of the first layer cell's adjacent cells excluding the one to be adjusted (denoted as the second layer cells).
- 2) After each first layer cell receives the request, it checks its load status information, and sends the information of "cannot be the target cell" directly back to the source cell if itself is or will soon be overloaded.
- 3) Otherwise, it checks its environment state, which is the average load state of the second layer cells.
- 4) The second layer cells will query and return their load state information whether they are overload or not.
- 5) After receiving the load state information from the second layer cells, the first layer cell obtains its environment state, and sends it back with its load status to the source cell.
- 6) After receiving the load state and the environment state information of first layer cells, the source cell will compute the overall state of each first layer cell, thus chooses the target cell with smallest overall state value.

This TL-MLB method expands the scope of inquiry in the process of finding optimum target cell, aiming to further improve the load balancing performance among adjacent cells. First the parameters of load state and environment state are given in detail as follows.

We denote the load state of each cell i as LS_i which is determined in formula (1) by terminal information U_i , PRB information R_i , and traffic information B_i .

$$LS_i = (\omega_1 * U_i + \omega_2 * R_i) + \omega_3 * B_i \quad (1)$$

in which U_i is the proportion of number of UEs to the cell i maximum UE capacity; R_i is the ratio of the used resource block to the total resource block; B_i is the ratio of real-time and GBR traffic to the total traffic. Parameter ω_1 , ω_2 is determined by the weight of U_i and R_i . Parameter ω_3 is decided by the extent to lower non-real-time traffic or non-GBR traffic resources to accept more users. Now we can obtain the environment state of the first layer cell i :

$$ES_i = (LS_{1i} + LS_{2i} + \dots + LS_{ni}) / n \quad (2)$$

where ES_i (the environmental state of the first layer cell i) is the average load state of its all second layer cells; LS_{ji} is the load state of every second layer cell of the first layer cell j . Its specialty lies in that to calculate environmental state information in most cases is applicable by the fore-mentioned Procedure 5), and then multiply it by an environmental facto. This algorithm practically reflects the average load of the first layer cells, compared to other methods.

In order to evaluate the overall state of the first layer cell we have to combine the load state and environment state into one figure. Thus, the following weighting function has been defined:

$$OS_i = \mu LS_i + (1 - \mu) ES_i \quad (3)$$

OS_i is the resulting overall state of the first layer cell i , in which the μ reflects the influence degree of the LS and the ES to the OS , LS_i is the load state of cell i computed by Formula (1), while ES_i is the environment of cell i brought by Formula (2). The OS reflects the comprehensive information of the first layer cell, thus decides whether the first layer cell can be a target cell. The parameter μ is decided by the weight of LS_i , OS_i .

In our simulation, we simplify Forum (3) by having $LOAD$ stand for LS , and we set μ to 0.2, through several simulation tests we find by setting μ to 0.2 there will be a good effect, so Formula (3) becomes to:

$$OS_i = 0.2 \times LOAD_i + 0.8 \times (\sum_{j=1}^n LOAD_j) / n \quad (4)$$

where OS_i is the overall state of cell i , j is the neighboring cell of cell i , n is the number of neighbor cells.

After finding the source cell and target cell, the problem is how to cause cell-edge users in the source cell to migrate to the target cell. According to [10], a handover event is initiated when UE detects that a neighboring cell offers a better signal quality than its currently serving cell. This condition is referred to as event A3, which has been formulated as (5).

$$M_t + O_{ft} + O_{ct} - Hys > M_s + O_{fs} + O_{cs} + Off \quad (5)$$

where M_t is the measurement result of the target cell, not taking into account any offsets. O_{ft} is the frequency specific offset of the frequency of the target cell. O_{ct} is the cell specific offset of the target cell, and set to zero if not configured for the target cell. M_s is the measurement result of the source cell, not taking into account any offsets. O_{fs} is the frequency specific offset of

the source frequency. O_{cs} is the cell specific offset of the source cell, and is set to zero if not configured for the serving cell. Hys is the hysteresis parameter for this event. Off is the offset parameter for this event. M_t , M_s are expressed in dBm in case of RSRP, or in dB in case of RSRQ. In this paper we simplify the Inequality (5) by just concerning with intra-cell/intra-frequency handover:

$$M_t - M_s > O_{cs} - O_{ct} + Off + Hys \quad (6)$$

We can see from Inequality (6) that when O_{ct} is larger, it will be easier for UEs camping on the source cell to migrate to the target cell. We perform load balancing by automatically adjusting the offsets O_{ct} based on cell load measurements. For the lightly-load cell (target cell) and the overloading cell (the sour cell) have been chosen already, we have a simple way as follows.

$$O_{ct} \leftarrow \min\left(O_{ct} + \Delta * \frac{OS_s}{OS_t}, O_{ct}^{max}\right), \quad OS_s - OS_t \geq OS_{th} \quad (7)$$

where Δ is the offset step-size, OS_s and OS_t are the load of the source cell and the target cell respectively, and OS_{th} is a predefined threshold for triggering load balancing. O_{ct} is initiated to zero and is updated by formula (7) in each load balancing loop, in response to new load measurements reports of each cell. Formula (7) indicates that larger difference between OS_s and OS_t will increase O_{ct} and make load shift from source cell to target cell easier.

In order to find the target cell and source cell pair, a cell load is calculated. Before that, the signal to noise and interference ratio for every user u $SINR_u$ is obtained.

$$SINR_u = \frac{P_{X(u)} * L_{X(u),u}}{N + \sum_{c \neq X(u)} LOAD_c * P_c * L_{c,u}} \quad (8)$$

Where N is the thermal noise, P_c is the transmit power for a cell (assumed to be same for all cell), $L_{c,u}$ is the overall attenuation summed over distance dependent pathloss, shadow fading and antenna gain.

The $SINR_u$ then is mapped to a modulation and coding scheme (MCS) based on look-up-tables obtained from link-level simulation, to obtain an instantaneous achievable data rate $R(SINR_u)$.

We assume a constant-bit-rate (CBR) traffic of 1Mbps that is the UE u required data rate D_u , so the amount of resources required by user u can be expressed as

$$N_u = \frac{D_u}{R(SINR_u)} \quad (9)$$

Now we use a simple method of measuring load of cell c , which can be expressed as the sum of required resources of all

users u connected to cell c to the total number of resources N_{tot} , i.e.

$$LOAD_c = \min\left(\frac{\sum_{u: X(u)=c} N_u}{N_{tot}}, 1\right) \quad (10)$$

where $X(u)=c$ is the connection function between u and c .

For it does not make sense to look at throughputs if we focus on CBR traffic (the UEs either get exactly the CBR or they are totally unsatisfied), so we will use the number of unsatisfied users as an assessment and simulation metric. Equation (10) has already given an expression for the cell load which shall not exceed 1. Based on this we define a virtual load \overline{LOAD}_c which gives us a clear indication how overload a cell is:

$$\overline{LOAD}_c = \frac{\sum_{u: X(u)=c} N_u}{N_{tot}} \quad (11)$$

$\overline{LOAD}_c \leq 1$ means all users in the cell are satisfied, and $\overline{LOAD}_c = 3$ means 1/3 of the users are satisfied. The number of unsatisfied users in the whole network (number of users in cell c is M_c) can be expressed as:

$$Z = \sum_c \max\left(\sum_{u|X(u)=c} M_c * \left(1 - \frac{1}{\overline{LOAD}_c}\right)\right) \quad (12)$$

For performance analysis in our model, we define a load distribution index measuring the degree of load balancing of the entire network, as follows:

$$\nabla(t) = \frac{\left(\sum_c \rho_c(t)\right)^2}{|N| \sum_c (\rho_c(t))^2} \quad (13)$$

where $|N|$ is the number of cells in the simulation scenarios, t is the simulation time. The load distribution index is 1 when load is completely balanced among cells. The target of load balancing is to maximize $\nabla(t)$. The smaller the value of $\nabla(t)$, the worse the unbalanced load distribution among cells is.

IV. SIMULATIONS

System-level simulations for a LTE cellular network are made to evaluate the performance of the proposed TL-MLB in terms of load distribution index, number of unsatisfied users, total handover times, average throughput. In the simulation results we use three different curves respectively to represent performance indicators of traditional load balance, mobility load balance and TL-MLB that we proposed. So we can compare the efficiency of our novel mobility load balance method with other methods. Thus we can find our strengths and possible weaknesses.

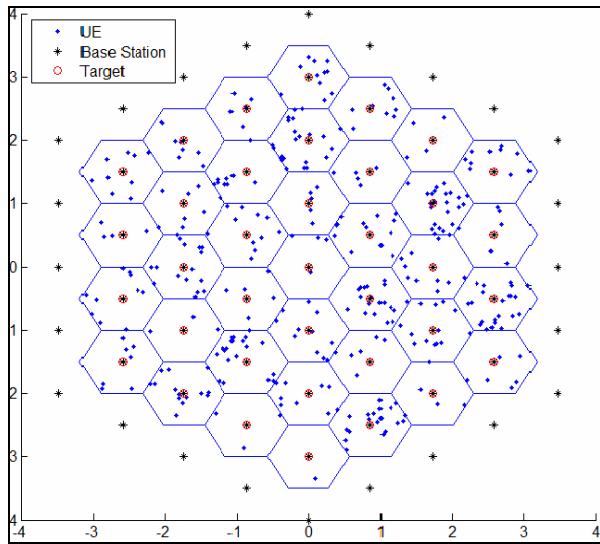


Figure 1. Network model with wrap-around technique

A. Layout, scenarios, parameters

A regular hexagonal 37 cell layout with an inter site distance of 1000m and wrap around technique is used to avoid boundary effects (cf. Figure1). Each cell has one base station situated in the center with no sectors divided. We assume that every user requests a constant bit rate of 1Mbps. The LTE capacity in this case will be 15 UEs per cell (assuming 10MHz bandwidth). We set the simulation time to 2 hours and simulation step time to 30 second. The relatively slow action of load balancing means that long simulation durations are needed to study its behavior. The main simulation parameters are given in Table 1.

TABLE I Simulation Parameters

System bandwidth	10MHz
Cell layout	Hexagonal grid, 37 cell sites, with wrap-around technique
Inter-site distance (ISD)	1000m
Passloss	$-38.4-35.0\log_{10}(R)$ (R is the distance between UE and)
Shadowfading	Log-normal with standard deviation 8dB
Shadowfading correlation distance	50m
Antenna gain	-7dB
eNB Tx antennas	1 per cell
UE Rx antennas	2
eNB Tx power	46dBm
Rate	30 seconds
O_{ct}^{\max}	5dB
Δ	1dB
OS_{th}	0.30
Traffic model	CBR 1Mbps full buffer traffic
Hys	3dB

B. User position

In order to show the necessity of MLB, we will artificially create heavy load concentration in our 37 cell network. At the beginning, a total number of 365 UEs are uniformly dropped according to the setting in Table 2. Each UE engages a random walk in the areas and changes direction every 2.5 second.

However, not all UE can move casually. We assume 180 UEs would not move out of their original cell. Thus the heavy load concentration won't be broken by UEs' random walk.

TABLE II UE Distribution

Number of UE unit	365
Distribution in each cell	Uniform distribution
No. of cells(20 UE units/cell) (15 UEs won't move out)	7
No. of cells(10 UE units/cell) (5 UEs won't move out)	15
No. of cells(5 UE units/cell)	15
Cell capacity	15UE
UE speed	5km/h

C. Results

Simulations are made in terms of load balance index, number of unsatisfied users, total handover times and average throughput of cell to examine the performance of the proposed TL-MLB algorithm. There are 2 reference scenarios, conventional load balancing presented in [3] and a MLB method presented in [8].

Figure 2 shows load distribution index varying with load balancing times. The load distribution index reflects the degree of similarity between cells. When the load values are very similar, this value is accordingly more close to 1. When the load is unbalanced seriously, for example, almost all the load is in one cell, this value is equal to the reciprocal of the total number of the cells. Therefore, this value will increase during load balancing. We can see the ultimate load distribution index of TL-MLB is the highest, MLB inferior, while the index of traditional load balance method is the lowest. As above we can conclude that TL-MLB is better. It is because that TL-MLB executes load balancing via changing the HO threshold, which takes a period of times. This disadvantage can be ignored compared to the advantage of higher load distribution index TL-MLB could reach. Figure 3 shows the number of unsatisfied users versus load balancing times. Using MLB increases user satisfaction significantly. We can see the ultimate number of unsatisfied users of TL-MLB is the smallest, MLB inferior, while the number of traditional load balance method is the biggest.

Figure 4 shows the cumulative handover times versus load balancing times. Some users of high load cell would be switched to light load cell because of load balancing, moreover, as the users are moving randomly, changes of positions will trigger the switch of the serving cell. The value is supposed to be the lower the better, because handover times affects service quality and user experience. It is clearly there are less handover times using TL-MLB than the other two methods, showing that the adverse effect of TL-MLB is less. Figure 5 shows the average throughput versus load balancing times. Throughput is the sum of all users speed in the cell, units Mbits/s. The average throughput is the average of each cell's throughput. After load balancing, due to the reducing of the number of satisfied users, some of the previous unsatisfied user's data can be transmitted normally, so the whole system throughput will increase. It can be seen from the figure that TL-MLB performs better in throughput to finally reach a higher value than the other two methods.

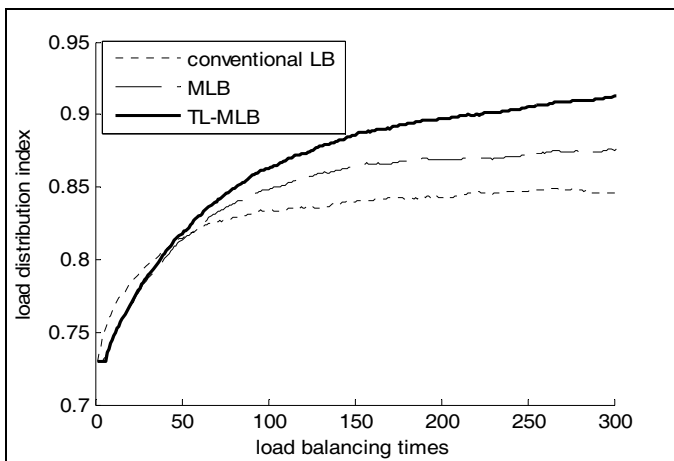


Figure 2. the load distribution index varying with load balancing times

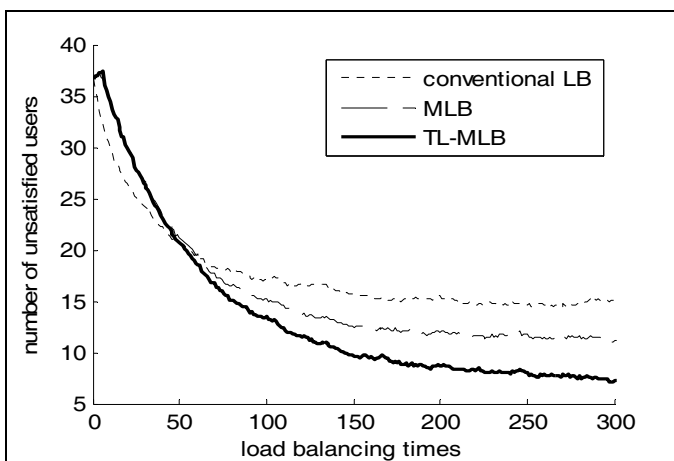


Figure 3 The number of unsatisfied users versus load balancing times

V. CONCLUSIONS

In this paper, a new TL-MLB to achieve MLB for LTE RAN is presented. The new method considers target cell itself and its environment comprehensively, thus avoiding the possibility that all of the over-load cells switch to one lightly-load cell nearby, causing the target cell overloaded. The number of unsatisfied users is reduced with fewer handovers and the higher average throughput compared with conventional solution and current MLB method.

REFERENCES

- [1] 3GPP, Tech. Specif. Group Radio Access Network Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN), 3GPP TS 25.913.
- [2] NEC, "Self Organizing Network: NEC's Proposals for Next-Generation Radio Network Management," Feb. 2009;
- [3] M. Döttling and I. Viering, "Challenges in Mobile Network Operation: Towards Self-Optimizing Networks," IEEE Int'l. Conf. Acoustics, Speech, and Sig. Processing, 2009.

- [4] Raymond Kwan, Rob Arnott, Robert Paterson, Riccardo Trivisonno, Mitsuhiro Kubota, "On Mobility Load Balancing for LTE Systems," IEEE Vehicular Technology Conference Fall (VTC), 2010.
- [5] Andreas Lobinger, Szymon Stafanski, Thomas Jansen, and Irina Balan, "Load Balancing in Downlink LTE Self-Optimizing Network," IEEE Vehicular Technology Conference (VTC), 2010.
- [6] Ingo Viering, Martin Döttling, and Andreas Lobinger, "A mathematical perspective of self-optimizing wireless networks," IEEE International Conference on Communication June 14-18, 2009.
- [7] Heng Zhang, Xuesong Qiu, Luoming Meng, and Xidong Zhang, "Design of distributed and autonomic load balancing for self-organization LTE," IEEE Vehicular Technology Conference (VTC), 2010.
- [8] Weihao Lv, Wenjing Li, Heng Zhang, and Yanguang Liu, "Distribution Mobility Load Balancing With RRM In LTE", IEEE International Conference on Broadband Network and Multimedia Technology (IC-BNMT), 2010.
- [9] Honglin, Jian Zhang, Xiaoying Zheng, Yang Yang, and Ping Wu, "Self-Configuration and Self-Optimization for LTE Networks," IEEE Communications Magazine, Feb., 2010.
- [10] TS 36.331, 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification (Release 8).

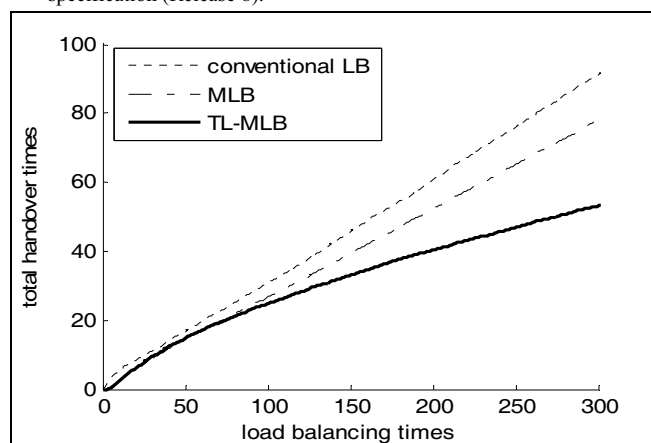


Figure 4 the cumulative handover times versus load balancing times

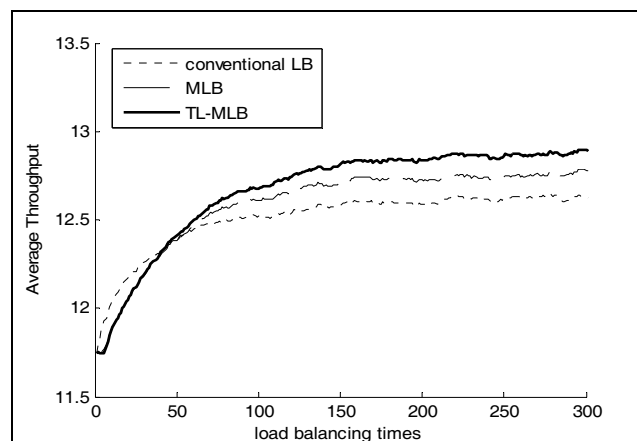


Figure 5 the average throughput versus load balancing times