Load Balancing Early Handoff Scheme for Multimedia Data Transmission in Wireless Networks

Cheng-Wei Hu and Yung-Chung Wang
National Taipei University of Technology, Taipei, 10608, Taiwan
Email: {cwhu, ycwang}@ntut.edu.tw

Jenn-Shing Wang
Jinwen University of Science and Technology, New Taipei City, 23154, Taiwan
Email: jswang@just.edu.tw

Abstract—This study developed a load balancing early handoff scheme that enables transfer part of heavy load in a base station to the others using the overlap between base stations in wireless cellular networks. The proposed scheme reduces the new call blocking probability (CBP) and handoff call connection dropping probability (CDP) as well as to increase the resource utilization of wireless networks. Since the scheme is triggered by a connection request received by the base station, it immediately reflects the loading of the base station. The proposed load balancing early handoff policy can be integrated in diverse call admission control schemes. The effectiveness of the load balancing early handoff scheme is demonstrated by integrating it with DMTBR to improve the call admission control function. Additionally, NS2 is used to simulate the load balancing early handoff scheme in a wireless network with wrap-around model. The simulation results show that the proposed scheme can effectively improve the wireless network resource utilization.

Index Terms—Call admission control, handoff, resource allocation, drop probability, block probability.

I. Introduction

As the use and convenience of the Internet increases, the requirements for various network services tend to diversify. New mobile communication services are used not just for conventional voice-based services, but also for multimedia applications [1]. Hence, it is important to utilize resources appropriately to meet various traffic characteristics, such as bandwidth, delay time, and call drop probability.

Cellular networks commonly use handoff when a mobile device is moving from the current serving cell to a new cell in order to provide continuous communication. Handoff can be requested by the mobile device or by the base station. A failed handoff procedure may result from a base station dropping the call for various reasons, e.g., lack of resources; it will force termination of the communication in progress. Two important factors in the QoS of wireless cellular networks are handoff call drop probability (CDP) and new call blocking probability (CBP) [2][3][4].

Mobile service users are highly sensitive to service interruptions. Services should not be stopped when a user moves from a current service base station to another because an abruptly terminated call is more annoying than a blocked call. To provide an uninterruptible communication service, the handoff process should be performed smoothly and without affecting existing services. The destination base station must have sufficient resources for the handoff.

The most common architecture for mobile communications is the cellular network. A land area is divided into regularly shaped cells. The most common shape is the hexagon. Therefore, each hexagon-shaped cell has six adjacent neighbors and overlapping coverage on each side. Overlapping coverage areas can access multiple services from different base stations. Users in the overlapping coverage area can choose which base station to access.

Based on the overlapping features of the base station in wireless cellular networks, this study proposes a load balancing early handoff scheme (LBEHS). Multiple cellular base stations cooperate to balance load in each base station and improve the system capability. The proposed LBEHS is able to integrate with different admission control policies. The DMTBR [5] is used to demonstrate the integration and NS2 [6][7] is used to simulate the network environment and the usage-situations. Simulation results show that the proposed scheme outperforms conventional schemes.

The rest of this paper is organized as follows. Section 2 describes the related works. Section 3 describes the system design of load balancing early handoff scheme. Next, we analysis the performance of multiple-threshold bandwidth reservation scheme in section 4. The simulations are verifying the effectiveness of the proposed scheme and results are presented in Section 5. Finally, Section 6 draws conclusions.

II. Related works

Due to limited wireless network bandwidth, network system resources must be managed effectively, and bandwidth must be distributed properly for different communication requirements. Therefore, effective admission controls and resource allocation schemes are essential. Admission control policies can be categorized as fixed channel allocation (FCA) and dynamic channel allocation (DCA). The FCA is designed for a network with known modeling features that cannot be used in a dynamic networking environment. In contrast, DCA is appropriate for dynamic networks. The related literature is discussed below.

Reference [8] introduced and compared fixed, dynamic, and hybrid allocation schemes, and discussed the effect of handoffs and prioritization schemes; [9][10][11] discussed the effect of handoffs and prioritization schemes. In a series of studies [11][12] authors categorized admission control into guard channel schemes and queuing priority schemes. The prioritized and non-prioritized handoff schemes with FCA are proposed in [13]. Furthermore, [14] compared the effects of FCA communication networks on user load, user mobility and the distribution of users among cells. To improve the efficiency of system resource usage, [15] proposed a queuing algorithm for new call requests to increase the system capacity of an FCA handoff channel reservation scheme. The designed first-in-first-out queue can reserve more channels for handoffs compared to conventional schemes. The performance improvement was studied in [16] by using channel allocation schemes in Time Division Multiple Access (TDMA) type of Digital Mobile Cellular systems. Reference [17] explored dynamic channel allocation combined with reservation and used soft reservation to adjust the number of reserved channels by monitoring network parameters. A QoS adaptation framework was proposed in [18] that stabilizes frequent handoffs in wired high speed ATM networks with different approaches: application layer adaptation and network layer adaptation. Reference [5] proposed a dynamic multiple-threshold bandwidth reservation scheme to adjust reservation bandwidth. In [5][19], they granted differential priorities to different traffic classes, including new and handoff traffic and adjusts bandwidth reservation thresholds. Reference [20] proposed an application-aware scheme and classified traffic as TCP and Simple UDP to enhance network utilization. Reference [21] used local information to predict resource demands and to determine the resources available for handoff calls in multimedia wireless IP networks.

Studies of load balancing issues include changing different Internet gateways in a mesh network [22] and allowing the mobile device to migrate to the existing network when the new network is saturated [23]. In an IEEE 802.16j wireless network environment, a multi-hop relay can extend network coverage [24]. A mobile station can access a base station through multi-hop relay stations. Reference [25] proposed a path selection algorithm for improving overall network performance.

Mobility direction is another important factor in handoff because it enables the target cell to reserve resources for the handoff call. Reference [26] used aggregate history to predict the direction of movement by mobile users and hand-off times in a call. GPS measurements can determine the direction of movement by mobile users and sending channel reservation requests for possible handoff calls to neighboring cells [27]. Reference [28] integrated GPS-based handoff prediction and proposed a dynamic guard bandwidth scheme that considers changes in handoff parameters and changes in terrain and manmade features. In [29], combined stochastic information with topological information in micro-cellular environments is used to predict the next handoff cell.

III. Load Balancing Early Handoff Scheme

Cellular network works networks use groups of cells to extend their coverage area. Multiple base stations can provide full coverage, but they may cause an overlap in the overlap coverage area between each adjacent base station. After accessing a serving base station (SBS), mobile users tend to move to another base station without disconnecting its service(s). Therefore, a service is required to transfer the connection(s) to the target base station. This procedure is called handoff or handover. Fig. 1 shows how a mobile node (MN) in an overlapping area provides access to BS_A or BS_B resources. Generally, the coverage depends on the signal strength.
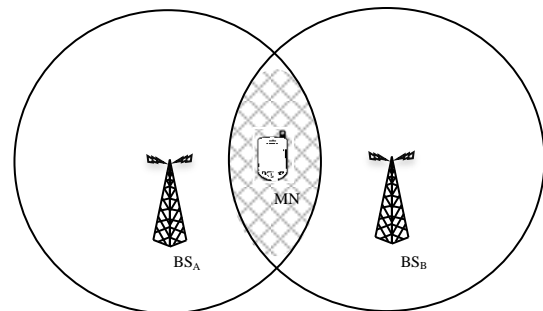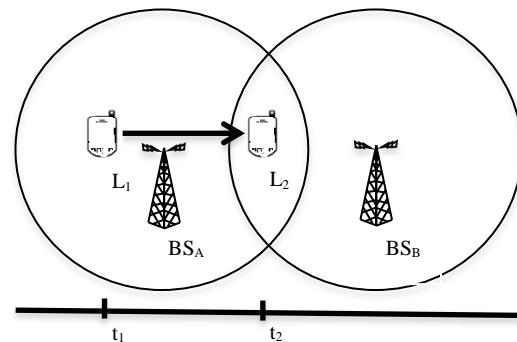

Fig. 1. Overlap area between base stations

Fig. 2. Early handoff scenarios

As the system load of a base station increases, the usable resources decrease. In this situation, the proposed load balance early handoff scheme requests an MN

handoff to the base station with overlapping coverage to release the bandwidth. Fig. 2 illustrates a handoff scenario. At t_1, MN is at L_1 and uses BS_A resources while moving toward BS_B. At t_2, MN is at L_2, which is in the overlapping coverage area between BS_A and BS_B. The MN does not meet the handoff requirement and is still accessing BS_A. When BS_A lakes of resources, the early handoff procedure is triggered.

Fig. 3 shows the early handoff procedure. When the BS_A has a load increase or resource leaks, the early handoff procedure is triggered. The BS_A sends a request to the serving MNs in the overlapping coverage area to start the early handoff. The MNs then request the target base station to give permission to handoff. If the target base station does not have sufficient resources to service the handoff connection, it denies permission to handoff. Conversely, the target base station gives permission to handoff and starts the handoff procedure.
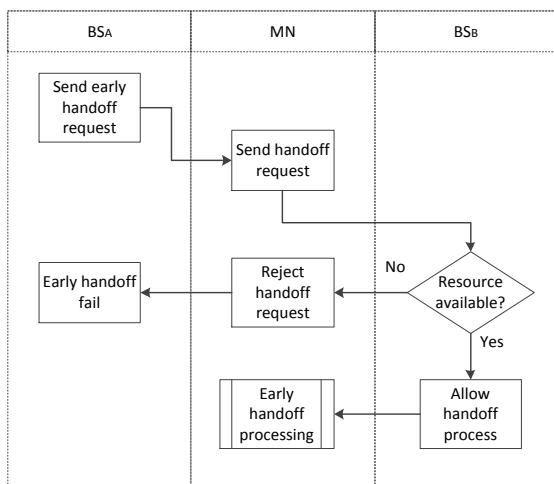

Fig. 3. Early handoff procedures

To improve performance, a bandwidth reservation scheme is implemented on each base station to control the connecting requests and to manage bandwidth resources. However, the proposed LBEHS is suitable for a large network of cells. Moreover, the LBEHS can be integrated into various call admission policies. The LBEHS is triggered when the current base station resources are insufficient. The LBEHS can react to the system load immediately and overcome the bottleneck of usage in individual base stations.

A difficult problem is how to integrate LBEHS into the admission control policy. Since the bandwidth reservation scheme manages bandwidth in a base station and does not affect the handoff procedures, it benefits LBEHS implementation and functions like an extension of the original system.

The performance of LBEHS was evaluated by integrating it with multiple-threshold bandwidth reservation scheme (MTBR), which characterizes different traffic classes into real-time and non-real-time traffic; furthermore, MTBR grants new and handoff traffic to each traffic type. The multi-thresholds bandwidth reservation scheme MTBR reserves resources

for different traffics by setting up three reservation thresholds to reduce call block and drop probabilities. A connection from a neighboring base station is named a handoff call, and handoff calls are categorized as real time and non-real time calls. For example, video and voice conference calls are deemed as real time connections; emails and file transfer are deemed as non-real time connections.

Interrupting a connecting call is assumed to be worse than an uninitiated call. The connection quality for a real time multimedia service is also assumed to be more considerable than a non-real time service; it should not be disconnected even though the user is moving continually. According to the conditions above, MTBR grants priorities for traffics are set as: Handoff call for real time traffic > Handoff call for non-real time traffic > New call for real time traffic > New call for non-real time traffic. Fig. 4 shows the thresholds are defined in MTBR for reserve bandwidth for different traffic and connection types. The thresholds are applied to the call admission control policy, and each size should be less than the total bandwidth.
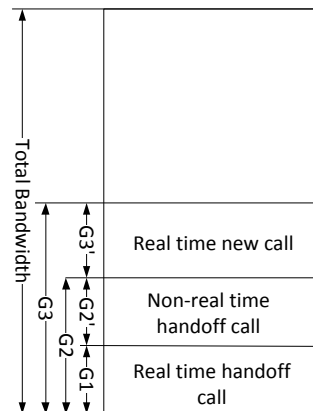

Fig. 4 Thresholds of MTBR

Fig. 5 shows the system block of call admission control (CAC), which includes MTBR and LBEHS. When an MN attempts to access the base station, the first step is to get the permission from the CAC policy. The CAC queries the MTBR to determine whether the base station has enough bandwidth for the coming connection. If the current base station lacks the resources, the CAC starts the LBEHS process. After LBEHS, CAC checks the current resources again. If the base station still cannot provide services for the connection, CAC rejects the accessing request.
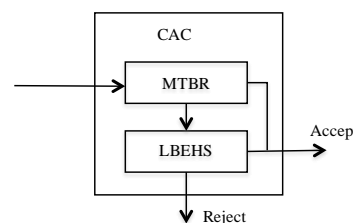

Fig. 5. Call admission control

The CAC is based on MTBR, which can process handoff and new incoming traffic. Fig. 6 shows that MTBR works with LBEHS for handoff connections. A real time handoff connection has priority and is allowed to access the base station when the bandwidth of the base station is higher than the basic real time connection bandwidth requirement, BWrt BU. Otherwise; the base station executes LBEHS then repeats resources check. If the resources available for a connection are insufficient,

the CAC rejects the connection. If the requested connection is non-real time handoff traffic, the available bandwidth should exceed BWnrt + G1 BU. Else, the request is rejected. If a call is initialed in a base station, it follows the process shown in Fig. 7. The admission policy is the same as in MTBR, and LBEHS is performed only when the base station lacks resources for real time connections.
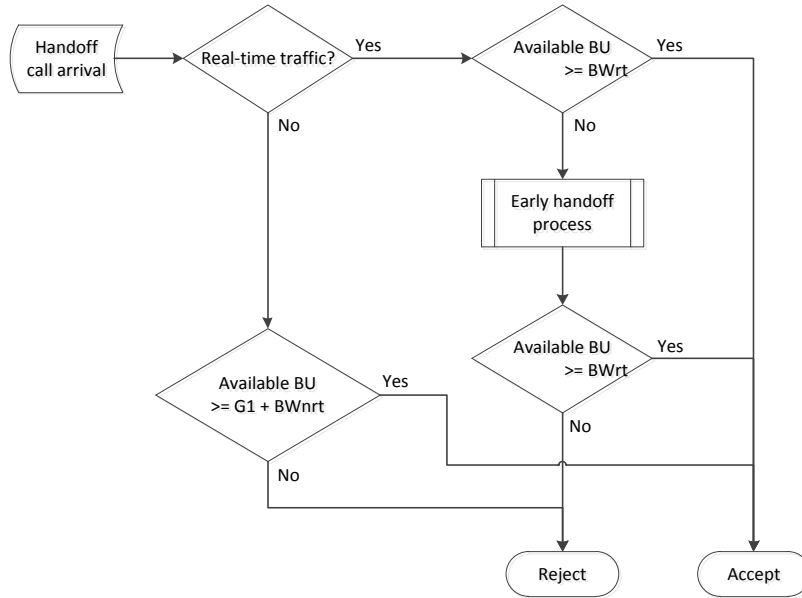

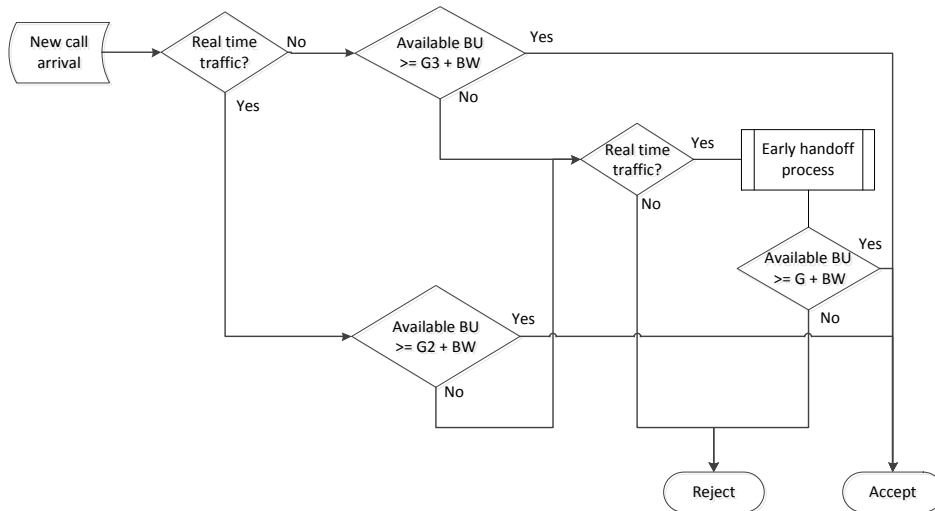Fig. 6. MTBR with LBEHS for handoff connection

Fig. 7. MTBR with LBEHS for new connection

IV. PERFORMANCE EVALUATION OF MULTIPLE-THRESHOLD BANDWIDTH RESERVATION SCHEME

Real time and non-real time new arrival call will be modeled using a Poisson process with parameter $\lambda_{rt,n}$ and $\lambda_{nrt,n}$. Real time and non-real time handoff inter-arrival call will be modeled using a Poisson process with parameter $\lambda_{rt,h}$ and $\lambda_{nrt,h}$. We assume that ongoing call (new or handoff) holding time is exponentially distributed with parameter μ_h. The time spent in a given cell, before

handing off, is called the cell dwell time. We assume this time is also exponentially distributed with parameter μ_d.

Consider the embedded continuous-time Markov chain $\{L(t), t \geq 0\}$ of the queuing system on the state space $\{0,1,\ldots,C\}$, where $L(t)$ denotes the capacity occupancy at time t. The embedded Markov chain now has an infinitesimal generator Q of the following block form

$$\begin{bmatrix} B_0 & A_0 & 0 & 0 \\ D & B_1 & A_1 & 0 \\ 0 & D & B_2 & A_2 \\ 0 & 0 & D & B_3 \end{bmatrix} \qquad (1)$$

Where

$$A_0 = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ \lambda_{rt} + \lambda_{nrt,h} & \cdots & 0 \end{bmatrix}, A_1 = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ \lambda_{rt,h} + \lambda_{nrt,h} & \cdots & 0 \end{bmatrix}, A_1 = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ \lambda_{rt,h} & \cdots & 0 \end{bmatrix} \qquad (2)$$

$$B_0 = \begin{bmatrix} -(\lambda_{nrt} + \lambda_{rt}) & \lambda_{nrt} + \lambda_{rt} & 0 & \cdots & 0 \\ \mu & -(\mu + \lambda_{nrt} + \lambda_{rt}) & \lambda_{nrt} + \lambda_{rt} & \cdots & 0 \\ \vdots & & & \vdots & \vdots \\ 0 & 0 & 0 & \mu & -(\mu + \lambda_{nrt,h} + \lambda_{rt}) \end{bmatrix} \qquad (3)$$

$$B_1 = \begin{bmatrix} -(\lambda_{nrt,h} + \lambda_{rt}) & \lambda_{nrt,h} + \lambda_{rt} & 0 & \cdots & 0 \\ \mu & -(\mu + \lambda_{nrt,h} + \lambda_{rt}) & \lambda_{nrt,h} + \lambda_{rt} & \cdots & 0 \\ \vdots & & & \vdots & \vdots \\ 0 & 0 & 0 & \mu & -(\mu + \lambda_{nrt,h} + \lambda_{rt,h}) \end{bmatrix} \qquad (4)$$

$$B_2 = \begin{bmatrix} -(\lambda_{nrt,h} + \lambda_{rt,h}) & \lambda_{nrt,h} + \lambda_{rt,h} & 0 & \cdots & 0 \\ \mu & -(\mu + \lambda_{nrt,h} + \lambda_{rt,h}) & \lambda_{nrt,h} + \lambda_{rt,h} & \cdots & 0 \\ \vdots & & & \vdots & \vdots \\ 0 & 0 & 0 & \mu & -(\mu + \lambda_{rt,h}) \end{bmatrix} \qquad (5)$$

$$B_3 = \begin{bmatrix} -(\lambda_{rt,h}) & \lambda_{rt,h} & 0 & \cdots & 0 \\ \mu & -(\mu + \lambda_{rt,h}) & \lambda_{rt,h} & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \mu & -(\mu) \end{bmatrix} \qquad (6)$$

$$D = \begin{bmatrix} 0 & \cdots & \mu \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \qquad (7)$$

Where $\mu = \mu_d + \mu_h$, $\lambda_{rt} = \lambda_{rt,n} + \lambda_{rt,h}$, $\lambda_{nr,t} = \lambda_{nrt,n} + \lambda_{nrt,h}$.

Let x=(x0, x1, $\cdots$, xC) be the stationary probability vector of the Markov chain Q, i.e. $\mathbf{xQ} = 0$, $\mathbf{x} \geq 0$ and $\mathbf{xe} = 1$.

Now let $X_{block}^{(nrt,n)}(t)$ be the number of the non-real time new arrival call blockings in the interval [0, t). Then the expected value of $X_{block}^{(nrt,n)}(t)$, denoted by $E[X_{block}^{(nrt,n)}(t)]$, is given by

$$E[X_{block}^{(nrt,n)}(t)] = \sum_{i=C-G_3}^{C} \lambda_{nrt,n} x_i t. \qquad (8)$$

Consequently, the non-real time new arrival call blocking probability, denoted by $P_{block}^{(nrt,n)}$, can be calculated by

$$P_{block}^{(nrt,n)} = \frac{E[X_{block}^{(nrt,n)}(t)]}{\lambda_{rt,n} t} = \sum_{i=C-G_3}^{C} \lambda_{nrt,n} x_i \qquad (9)$$

Now let $X_{block}^{(rt,n)}(t)$ be the number of the real time new arrival call blockings in the interval [0, t). Then the expected value of $X_{block}^{(rt,n)}(t)$, denoted by $E[X_{block}^{(rt,n)}(t)]$, is given by

$$E[X_{block}^{(rt,n)}(t)] = \sum_{i=C-G_2}^{C} \lambda_{rt,n} x_i t. \qquad (10)$$

Consequently, the real time new arrival call blocking probability, denoted by $P_{block}^{(rt,n)}$, can be calculated by

$$P_{block}^{(rt,n)} = \frac{E[X_{block}^{(rt,n)}(t)]}{\lambda_{rt,n} t} = \sum_{i=C-G_2}^{C} \lambda_{rt,n} x_i \qquad (11)$$

Now let $X_{drop}^{(nrt,h)}(t)$ be the number of the non-real time handoff arrival call dropping in the interval [0, t). Then the expected value of $X_{drop}^{(nrt,h)}(t)$, denoted by $E[X_{drop}^{(nrt,h)}(t)]$, is given by

$$E[X_{drop}^{(nrt,h)}(t)] = \sum_{i=C-G_1}^{C} \lambda_{nrt,h} x_i t. \qquad (12)$$

Consequently, the non-real time handoff arrival call dropping probability, denoted by $P_{drop}^{(nrt,h)}$, can be calculated by

$$P_{drop}^{(nrt,h)} = \frac{E[X_{drop}^{(nrt,h)}(t)]}{\lambda_{nrt,h} t} = \sum_{i=C-G_1}^{C} \lambda_{nrt,h} x_i \qquad (13)$$

Now let $X_{drop}^{(rt,h)}(t)$ be the number of the real time handoff arrival call dropping in the interval [0, t). Then the expected value of $X_{drop}^{(rt,h)}(t)$, denoted by $E[X_{drop}^{(rt,h)}(t)]$, is given by

$$E[X_{drop}^{(rt,h)}(t)] = \lambda_{rt,h} x_C t. \qquad (14)$$

Consequently, the real time handoff arrival call dropping probability, denoted by $P_{drop}^{(rt,n)}$, can be calculated by

$$P_{drop}^{(rt,n)} = \frac{E[X_{drop}^{(rt,h)}(t)]}{\lambda_{rt,h} t} = \lambda_{rt,h} x_C \qquad (15)$$

V. SIMULATION RESULTS

The performance of LBEHS was evaluated by integrating it with dynamic multiple-threshold bandwidth reservation scheme (DMTBR) [5], which characterizes different traffic classes into real-time and non-real-time traffic; furthermore, DMTBR grants new and handoff traffic to each traffic type. The DMTBR reserves resources for different traffics by setting up three reservation thresholds to reduce call block and drop probabilities. According to the analysis results in section 4, we dynamically adjust the three reservation thresholds in this simulation work. A connection from a neighboring base station is named a handoff call, and handoff calls are categorized as real time and non-real time calls. For example, video and voice conference calls are deemed as

real time connections; emails and file transfer are deemed as non-real time connections.

A cell wrap-around model [30] was used to simulate the wireless signal coverage environment as Fig. 8. The coverage area in this example has 37 cell base stations. Each cell has the same bandwidth capabilities and adjoins six neighbors. We use Poisson process is used to control the connection arrives, and the assumptions are 25% real time traffic and 75% non-real time traffic. The connection duration time, cell residence time, and handoff occasion follow have an exponential distribution.
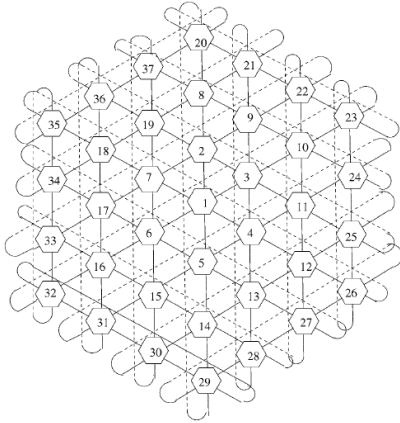

Fig. 8. Cell warp around model [30]

The NS2 is used to describe the above environment; call drop probability (CDP) and call block probability (CBP) are observed in different new connection arrival rates. The LBEHS and DMTBR are also compared under different conditions. Additionally, the assumed parameters are shown in Table 1.

Table 1. Parameters of simulation

Parameter	Amount
System simulation time	3 hours
Numbers of cell	37
Bandwidth units in a cell, BU_{total}	50 BU
Bandwidth units for real time connection, BU_{rt}	1 BU
Bandwidth units for non-real time connection, BU_{nrt}	4 BU
Average call duration time for real time connection, CDT_{rt}	300 sec
Average call duration time for non-real time connection, CDT_{nrt}	1500 sec
Average call residence time for real time connection, CRT_{rt}	150 sec
Average call residence time for non-real time connection, CRT_{nrt}	750 sec

A. The occasion of early handoff

In Fig. 9, an MN in $Cell_A$ service area at t_0 is moving towards $Cell_B$ direction and is expected to leave $Cell_A$ at t_2. At t_1, MN is in the overlapping coverage area between $Cell_A$ and $Cell_B$; MN can process an early handoff to $Cell_B$ when $Cell_A$ lacks sufficient bandwidth resources. The T_1 is the average duration of the connection before the early handoff. The T_2 is the average cell residence time in $Cell_A$. The T_1 is related to the overlapping coverage area between both $Cell_A$ and $Cell_B$. As the

overlapping coverage area increases, T_1 decreases. In contrast, as the overlapping coverage area decreases, T_1 increases. As the number of MNs in the overlapping coverage area increases, opportunities to handoff to neighbor base stations and share the loading of the current base station increase. Table 2 compares handoff times for three situations to observe the effects on the early handoff scheme in different size of overlap coverage area.
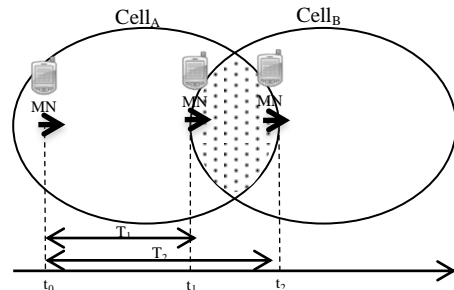

Fig. 9. Early handoff occasion

Table 2. Average early handoff time

Handoff time (T_1, T_2)	Situation (1)	Situation (2)	Situation (3)
Real time traffic	(50, 150)	(100,150)	(150,150)
Non-real time traffic	(250,750)	(500,750)	(750,750)
			Unit: Sec

Fig. 10 shows the CBP results for the three different handoff situations in Table 2. The real time and non-real time traffic in situation 1 are denoted by EH_DMTBR RT1 and EH_DMTBR NRT1 and so forth in Fig. 10-11. The CBP in situation 1 is lower than that in situation 2 and 3 in real time traffic; it means the size of overlap coverage zone will positively affect the new call accept rate for real time traffic. However, the CBP is not affected in the lower priority traffic class, non-real time traffic.

The priority of real time traffic for handoff calls is in the first place, as the result the CDP in real time traffic is close to zero in Fig. 11. The non-real time traffics in different overlap coverage zones are similar due to the priority of the handoff call. However, we still can observe the average of the CDP in larger overlap coverage area is lesser than the others. Therefore, the size of the overlap in coverage affects the performance of the LBEHS.
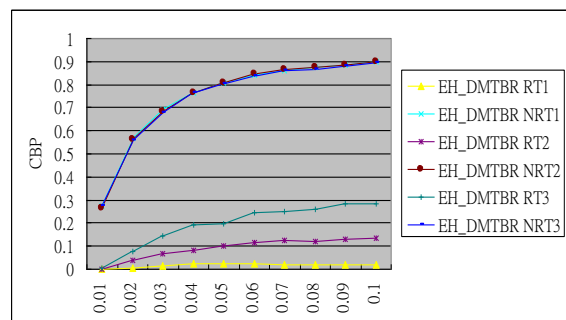

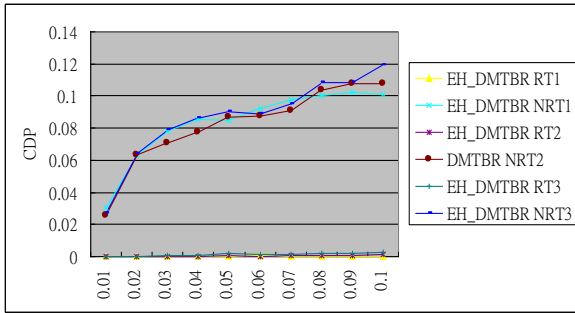Fig. 10. CBP effected on different sizes of overlap coverage area

Fig. 11. CDP effected on different sizes of overlap coverage area

B. DMTBR with LBEHS vs. without LBEHS

The effectiveness of the LBEHS is evaluated by comparing CBP and CDP of DMTBR with and without LBEHS in real time and non-real time traffic. Changes in CBP and CDP are observed for different arrival rates of connections from 0.01 to 0.1. Fig. 12-13 show the comparison results.

DMTBR RT (NRT) denotes the DMTBR (non-) real time traffic, and EH_DMTBR RT (NRT) denotes DMTBR with LBEHS (non-) real time traffic. Fig. 12 shows that CBP in EH_DMTBR RT is substantially reduced. Fig. 13 shows the CDP comparison in both real time and non-real time traffic; the result shows that EH_DMTBR outperforms DMTBR.
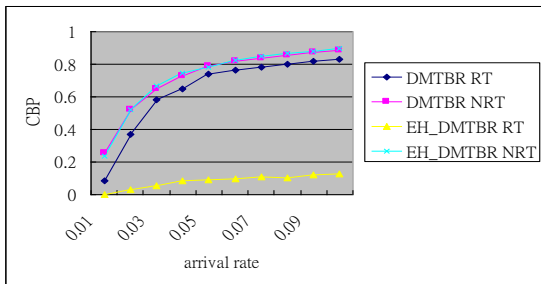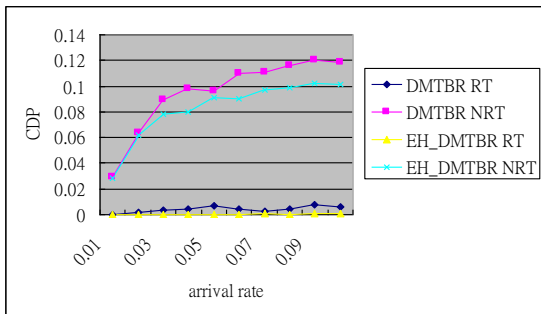

Fig. 12. CBP of DMTBR vs. EH_DMTBR

Fig. 13. CDP of DMTBR vs. EH_DMTBR

C. Load imbalance effect

To determine the effect of LBEHS on the overall system. We compare the both real time traffic and non-real time traffic in DMTBR with and without LBEHS in a system load imbalance situation. Such as the raise call arrival rate was increased in partial base stations. In this simulation, we assume the six stations: 7, 9, 11, 13, 15, and 17 perform higher loading than others in Fig. 8.

Fig. 14-15 show the results of system reaction when the partial cell loading is two times higher than average loading. The EH_DMTBR still transcends DMTBR.

Fig. 16-17 are the comparison of the partial cells perform four times higher than average loading. Although both CBP and CDP upper then perform in double loading systems in Fig. 14-15, the LBEHS keeps the system performance and utilization.
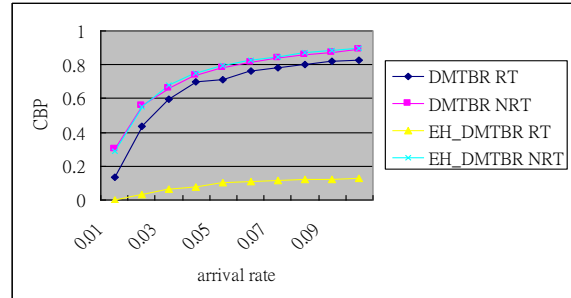

Fig. 14. CBP effected on two times of loading in partial cells
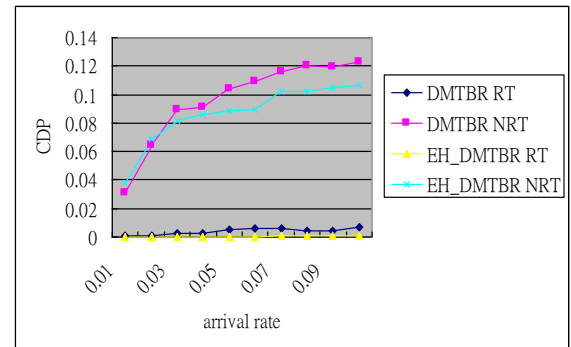

Fig. 15. CDP effected on two times of loading in partial cells
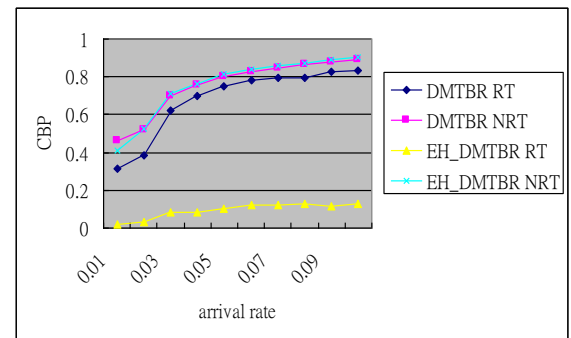

Fig. 16. CBP effected on four times of loading in partial cells
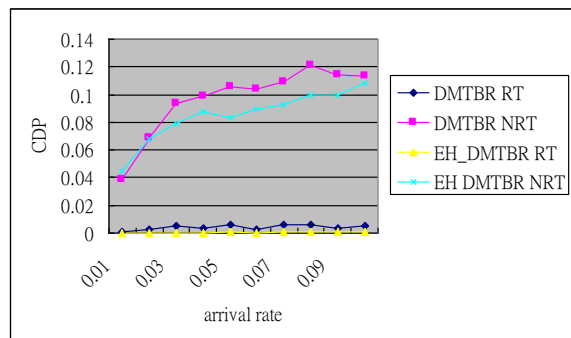

Fig. 17. CDP effected on four times of loading in partial cells

VI. Conclusion

In this study, a load balancing early handoff scheme is proposed and is able to integrate with various call admission control schemes. We combine the load balancing early handoff feature with the dynamic multiple threshold bandwidth reservation policy as the demonstration. Hence, the system can maintain priorities for different services and balance the system loading. Load balancing early handoff scheme increases usage in individual base stations, especially when the system loading is imbalance. The simulation results confirm that the proposed scheme effectively improves wireless network resource utilization. In the future work, the load balancing early handoff scheme can be applied to numerous physical wireless communication systems, such as WiMAX, LTE or 5G networks as a part of the radio resource management feature.

References

[1] J. F. Huber, "Mobile next-generation networks," *MultiMedia, IEEE*, vol. 11, no. 1, pp. 72–83, 2004.

[2] J. Cai, "A Probabilistic Priority Adaptive Time-Threshold-based scheme for cellular multimedia networks," presented at the *Advanced Computer Control (ICACC), 2010 2nd International Conference on*, 2010, vol. 5, pp. 267–272.

[3] Y. Iraqi and R. Baoutaba, "Handoff and call dropping probabilities in wireless cellular networks," *Proceedings of International Conference on Wireless Networks, Communications and Mobile Computing*, 2005, pp. 209-213.

[4] L. Song, L. Wu, and X. Yang, "Call admission control based on degradation and queues in wireless mobile networks," presented at the *Network Infrastructure and Digital Content, 2009. IC-NIDC 2009. IEEE International Conference on*, 2009, pp. 159–163.

[5] X. Chen, Bin Li, and Y. Fang, "A dynamic multiple-threshold bandwidth reservation (DMTBR) scheme for QoS provisioning in multimedia wireless networks," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 583–592, Mar. 2005.

[6] T. Issariyakul, *Introduction to Network Simulator Ns2*. Springer Science+Business Media, 2012.

[7] "The Network Simulator - ns-2." [Online]. Available: http://www.isi.edu/nsnam/ns/. [Accessed: 26-Dec-2014].

[8] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems: a comprehensive survey," *Personal Communications, IEEE*, vol. 3, no. 3, pp. 10–31, 1996.

[9] W. Li and A. S. Alfa, "Channel reservation for handoff calls in a PCS network," *Vehicular Technology, IEEE Transactions on*, vol. 49, no. 1, pp. 95–104, 2000.

[10] R. Fantacci, "Performance evaluation of prioritized handoff schemes in mobile cellular networks," *Vehicular Technology, IEEE Transactions on*, vol. 49, no. 2, pp. 485–493, 2000.

[11] W. Zhuang, B. Bensaou, and K. C. Chua, "Adaptive quality of service handoff priority scheme for mobile multimedia networks," *Vehicular Technology, IEEE Transactions on*, vol. 49, no. 2, pp. 494–505, 2000.

[12] J. Li, N. B. Shroff, and E. K. P. Chong, "Channel carrying: a novel handoff scheme for mobile cellular networks," *Networking, IEEE/ACM Transactions on*, vol. 7, no. 1, pp. 38–50, 1999.

[13] D. Hong and S. Rappaport Stephen, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures," *Vehicular Technology, IEEE Transactions on*, vol. 35, no. 3, pp. 77–92, 1986.

[14] P. Camarda, G. Schiraldi, F. Talucci, and R. Valla, "Mobility and performance modeling in cellular communication networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 1, no. 4, pp. 25–32, Oct. 1997.

[15] V. K. N. Lau and S. V. Maric, "Mobility of queued call requests of a new call-queueing technique for cellular systems," *Vehicular Technology, IEEE Transactions on*, vol. 47, no. 2, pp. 480–488, 1998.

[16] P. L. Hiew and M. Zukerman, "Teletraffic issues related to channel allocation in digital mobile cellular networks," *Proceedings of INFOCOM '98. Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies*, IEEE, 1998, vol. 1, pp. 43–50.

[17] J. Jiang and T.-H. Lai, "An efficient approach to support QoS and bandwidth efficiency in high-speed mobile networks," presented at the *Communications, 2000. ICC 2000. 2000 IEEE International Conference on*, 2000, vol. 2, pp. 980–984.

[18] C. K. Toh and W. K. Tsai, "Intelligent control of QoS adaptation in next generation wireless broadband networks," *Journal of High Speed Networks*, vol. 10, no. 1, Jan. 2001.

[19] L. M. Far and H. Beigy, "A new fractional call admission control scheme in integrated cellular network," presented at the *Electro/Information Technology, 2009. eit '09. IEEE International Conference on*, 2009, pp. 332–337.

[20] Y. Zhu, "A New Class-Based Traffic Queue Management Algorithm in the Internet," *KSII TIIS*, vol. 3, no. 6, pp. 575–596, 2009.

[21] T. Zhang, E. van den Berg, J. Chennikara, P. Agrawal, J.-C. Chen, and T. Kodama, "Local predictive resource reservation for handoff in multimedia wireless IP networks," *Selected Areas in Communications, IEEE Journal on*, vol. 19, no. 10, pp. 1931–1941, 2001.

[22] M. T. Tran, Y. H. Kim, and J. Le, "Load Balancing and Mobility Management in Multi-homed Wireless Mesh Networks.," *KSII TIIS*, vol. 5, no. 5, pp. 959–975, 2011.

[23] S. Prapojjanaporn, K. Kittiwaytang, and C. Aswakul, "Performance Evaluation of Heterogeneous Wireless Networks with Location Distribution and Migration of User Equipments," presented at the *ISMS '13: Proceedings of the 2013 4th International Conference on Intelligent Systems, Modeling and Simulation*, 2013, pp. 584–588.

[24] S. Ann, K. G. Lee, and H. S. Kim, "A Path Selection Method in IEEE 802.16j Mobile Multi-hop Relay Networks," presented at the *Sensor Technologies and Applications, 2008. SENSORCOMM '08. Second International Conference on*, 2008, pp. 808–812.

[25] S. H. Lee and Y. B. Ko, "A Bandwidth Adaptive Path Selection Scheme in IEEE 802.16 Relay Networks," *KSII TIIS*, vol. 5, no.3, pp. 477–493, 2011.

[26] S. Choi and K. G. Shin, "Predictive and adaptive bandwidth reservation for hand-offs in QoS-sensitive cellular networks," *Proceedings of the ACM SIGCOMM'98 conference on Applications, technologies, architectures, and protocols for computer communication*, New York, New York, USA, 1998, pp. 155–166.

[27] Z. Xu, Z. Ye, S. V. Krishnamurthy, and S. K. Tripathi, "A new adaptive channel reservation scheme for handoff calls in wireless cellular networks," *Proceedings of*

NETWORKING 2002: Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications Lecture Notes in Computer Science, 2002, pp 672–684

[28] W.-S. Soh and H. S. Kim, "Dynamic guard bandwidth scheme for wireless broadband networks," presented at the *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, 2001, vol. 1, pp. 572–581.

[29] B. Tolg and W. Bziuk, "A new topology-based algorithm for next cell prediction in cellular networks," presented at the *Global Telecommunications Conference 2003, GLOBECOM '03. IEEE*, 2003, vol. 2, pp. 919–923.

[30] H. Zeng, Y. Fang, and I. Chlamtac, "Call Blocking Performance Study for PCS Networks under More Realistic Mobility Assumptions," *Telecommunication Systems*, vol. 19, no. 2, pp. 125–146, 2002.

Authors' Profiles

Cheng-Wei Hu, Ph. D. Candidate at National Taipei University of Technology from Taiwan. From 2007 to 2011, he was a Software/Firmware Engineer at the Institute for Information Industry, Taiwan, where he dedicated to WiMAX and LTE related projects. His research interests lie in the areas of wireless networking, software-defined network, mobile systems and embedded systems.

Yung-Chung Wang received the M.S. and Ph.D. degrees in electrical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 1990 and 2000, respectively. From 1990 to 2001, he was a Research Engineer with the Chung-Hwa Telecommunication Laboratory, where he was engaged in research on the development of ATM switching systems and IP switch router systems. Since 2001, he has been with the Department of Electrical Engineering, National Taipei University of Technology (NTUT), Taipei, Taiwan, where he is a Full Professor. His research interests include cloud computing, wireless networks, optical networks, software-defined network, queuing theory and performance evaluation of communication networks.

Jenn-Shing Wang was born in Taiwan, Rep. of China, on August 4, 1953. He graduated from Taipei Institute of Technology, Taipei, Taiwan, in 1974, and received the MS and Ph.D. degrees from Stevens Institute of Technology and the University of Oklahoma, in 1980 and 1989, respectively, all in electrical engineering.

In 1981, he was a Research and Development Engineer at Nuclear Research Corporation, Denver, New Jersey. In 1982 he joined the faculty of Taipei University of Technology, Taipei, Taiwan, where he retired from the position of the Director of Computer Center. In 2006 he joined the faculty of Jinwen University of Technology, Taipei, Taiwan, where he is an Associated Professor of the department of computer science and information engineering. He is interested in computing machinery, digital systems design, and network communications.