# Location Management and Resource Allocation Using Load Balancing in Wireless Heterogeneous Networks

G. Adiline Macriga[1] and V.S. Surya[2]

[1] Department of Information Technology,
Sri Sai Ram Engineering College, Chennai
[2] Department of Computer and Communication,
Sri Sai Ram Engineering College, Chennai

**Abstract.** In the next generation of wireless networks, the entire network operating on different radio frequencies under wireless mode will be available for communication. The rapidly growing demand for "any service – anywhere - anytime" high-speed access to IP-based services is becoming one of the major challenges for mobile networks. However, in several situations, mobile terminals tend to associate with networks guaranteeing the best performances to stay "Always Best Connected" which leads to overload the most attractive technology while keeping the others technologies underutilized. As a recent research focus, load balancing is one of the key technologies in the convergence of heterogeneous wireless networks. Load balancing is a significant method to achieve the resource sharing over heterogeneous wireless networks, and it can improve resource utilization, enlarge system capacity, as well as provide better services for users. When dealing with high speed traffic over heterogeneous, we propose a load balancing algorithm using novel approach. For wireless packet network, this novel load metric is based on the packet scheduling and the radio link quality information. The solution can be used in on-line system because it requires less computation time and it operates in a distributed way instead of a usual centralized way. The two main targets of the proposed algorithm are the admission control and the network initiated handover. Further we have to investigate the joint load balancing and resource allocation optimization in heterogeneous networks.

**Keywords:** Heterogeneous, load balancing, admission control, handover.

## 1 Introduction

In the heterogeneous networks, generally the available networks in the transmission path (e.g., WLAN, 3G, 3GPP, UTRAN, WiMax) operating on the Radio frequencies under wireless mode available for communication are brought under a single head, based on their common operating procedures and functions such as request – acknowledge - response. As the radio access network is considered the parameters such as - signal strength, maximum coverage, power level, network load, flow rate, traffic rate are taken as the major factor. Along with the increase of multimedia and data-intensive applications, future networks will experience an extremely high load.

N. Meghanathan et al. (Eds.): CCSIT 2012, Part I, LNICST 84, pp. 383–393, 2012.

This load metric definition takes into account not only the user's required resource but also the radio link quality between the user and the BS. An operator can deploy different technologies or interwork with other technologies owned by other operators to enable the global roaming capability through a coordinated heterogeneous access network environment. Load balancing plays an important role in the Common Radio Resource Management (CRRM). It is defined as a platform to gather information from the Base Stations (BS) of different Radio Access Technology (RAT), and to control the resource allocation of all BSs to optimize the overall system performance.

An advanced Common RRM (CRRM) is a motivation for interworking among these networks, and also a challenge to overcome the existing load balancing techniques. The load balancing algorithm consists of accepting or denying a new incoming network request and forcing users connected to a heavily loaded BS to hand over to a lightly loaded one considering the loosely coupled architecture as shown in fig 1. The stronger the coupling is, the more efficient the resource can be utilized. In other words the balancing scheme is based on the load values of different access nodes regardless of underlying technologies and underlying scheduling schemes.

In the proposed model, load balancing is divided into two parts: network architecture and load balancing algorithm. The former is the foundation of load balancing, and a good network architecture can improve the efficiency of load balancing. In the perspective of control mode, load balancing mechanisms can be classified as centralized, distributed and semi-centralized and semi-distributed. A few problems are faced in the first two mechanisms: the centralized one is relatively low reliability, while the distributed one has a huge overhead. This solution can be used to rectify the above mentioned problems and can be used in on-line system because of its less computation time and the way it operates in a distributed way instead of a usual centralized way.

The remainder of this paper is organized as follows. In Section 2 we briefly discuss the related researches into overhead and reliability of network architectures. Section 3 introduces the system model we proposed. In Section 4, the proposed model is presented followed in Section 5 by our simulation and the results. Finally in Section 6 we conclude this paper with discussion of our work.

## 2 Related Work

The authors of [14] developed a mathematical framework that can be used to compactly represent and analyze heterogeneous networks that combine multiple entity and link types. They generalized Bonacich centrality, which measures connectivity between nodes by the number of paths between them, to heterogeneous networks and used this measure to study network structure. The authors of [9] proposed a semi-centralized and semi-distributed architecture (SCSDA), in which a BS just exchanges load information with several neighboring BSs. Although the architecture can reduce the overhead of control signaling, the authors neither expressed the overhead in mathematical formula, nor proved it by simulation. Reference [15] designed hybrid wireless network architecture, [16] proposed a multiple mobile routers based network architecture to support seamless mobility across heterogeneous networks, and they both tested the overhead by NS2 simulator. However, the model of the overhead was derived in neither [4] nor [12]. Route overhead was analyzed in theory in [8], by calculating the number of control messages generated in a BS/AP service area due to maintaining route. Nevertheless, the simulation for overhead was not given. The communication overhead of the scheme presented in [6] was calculated, and an algorithm for minimizing the communication overhead was given, which was proved to be effective through simulation. The authors of [15] considered a general heterogeneous network architecture with two basic entities in the system: mobile nodes (MNs) and access points (APs). They formulated the overhead of AP discovery which is divided into hello messages and RREQ messages, and gave the simulation results. Reference [2] proposed a hierarchical and distributed (HD) architecture with three hierarchical levels of mobility management being distinguished: end terminal remains connected to the same radio access network but it changes its point of attachments, end terminal changes its radio access network but it remains associated to the same operator and end terminal changes its operator network. And it also studied signaling cost generated by QoS negotiation during handover process in both theory and simulation. The research on reliability of telecommunication network starts at the study on switched telecommunication network by Li [10]. Li defined call blocking as the link failure, and measured reliability taking connectivity [13] as standard. Reference [12] mentioned the concept of integrated reliability, which took call loss as the evaluation indicator of network reliability, and proved that the integrated reliability can reflect the practical situation much better than taking connectivity as standard. The authors of [5] analyzed the reliability aspects of some access network topologies to insure a certain level of quality of service at the lowest cost for the end users.

Most of the previous work mainly focused on identifying the functionalities of the CRRM architectural components, and designing the protocols for control exchanges between these components. Besides, the resource allocation scheme which aims at quantifying the amount of resources allocated to each user in such a way to maximize the operator's revenue or the user's satisfaction has also been increasingly studied. However, the load balancing between different BSs and different RATs has not been sufficiently considered. Although the load balancing is much related to the resource allocation, they are two separable aspects. The load balancing can be considered on the one hand as an objective of the resource allocation scheme and on the other hand as a constraint for the resource allocation optimization. In this work, we only focus on the load balancing issue.

## 3 System Model

The load can be computed in different manners for different systems. As a result, the same load value for two different systems does not mean the same load situation. As such a comparison is the basis of any cross-system load balancing solution, having a same semantic of the load metric is mandatory. The existing load computation methods, which are based on the interference or the throughput, do not allow the load variation anticipation prior to the situation where a user moves into/out of a cell. The estimation of future interference or throughput values is really challenging. Accordingly, we will not be able to make the right decision to achieve an efficient resource balancing.

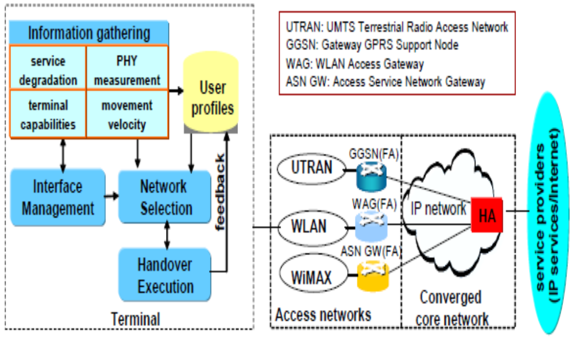



**Fig. 1.** Loosely Coupled Architecture

In order to solve the general load-balancing problem different algorithms have been proposed. Examples, mainly from the area of distributed computing, are diffusion algorithms, random matching algorithms, pre-computation based load distribution algorithms, algorithms based on microeconomics, force models (also called particles approach) and simple transmitter receiver-based algorithms. The force-based load-balancing algorithm introduced is a descriptive and easily extendable CRRM algorithm for combined UMTS/GSM networks.

A. DISTRIBUTED Vs CENTRALIZED ARCHITETURE

The semi-centralized and semi-distributed architecture (SCSDA), in which a BS just exchanges load information with several neighboring BSs. Although the architecture can reduce the overhead of control signaling, it does not consider the load history of the neighboring nodes. The hybrid wireless network architecture (figure 2) comparatively distribute the load on multi tier mode.

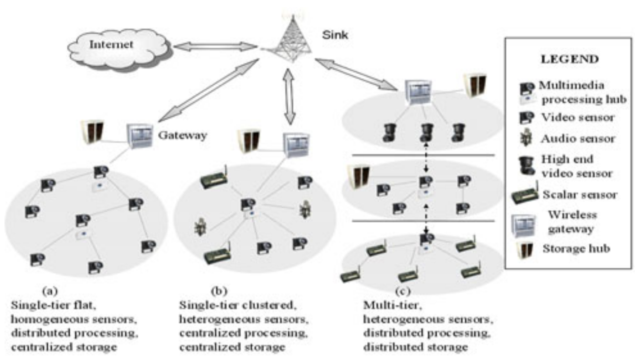



**Fig. 2.** Hybrid wireless network architecture

B. HIERARCHICAL SEMI-CENTRALIZED ARCHITECTURE

The hierarchical semi-centralized architecture based on basic grids, which takes three different types of access networks (UMTS, WLAN and WiMax) for example. A basic grid is made up of several adjacent cells. IS (Information Server), RA (Resource Allotter) and RS (Resource Statistics) are collectively referred to as Resource Management Unit (RMU), which are responsible for managing the resources of basic grids. Installed in the access point, a RS is used to calculate resources of the cell administered by the RS. A RA collects load information from RSs, and balances the load according to the load and resources of the basic grid. Normally an IS allocates resources for the borders of basic grids and stores information of cell identification, location and load states. However, it can take over the broken RA immediately. In order to improve the system reliability, a main IS and a standby IS are set up. Once the main IS stops running, the standby one will take over it.

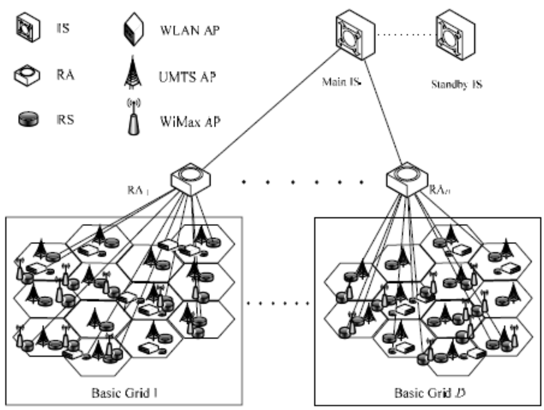



**Fig. 3.** Hierarchical semi-centralized architecture

## 4 Proposed Load Balancing Algorithm

The aim of the proposed work is to design a feasible and suboptimal solution for load balancing while minimizing the resource rearrangement and the computation effort. When a user initiates a connection, the end-user device selects a suitable access network among available ones using the network selection mechanism. The load value of each access node may be used in the network selection evaluation if the user has access to this information.

The user will not be allowed to select the heavily loaded access node. Besides, the access node may refuse the user's connection request based on its admission control policy if it is heavily loaded. Despite the use of an admission control, the overload of an access node still happens due to the transmission channel fluctuation, the mobility or the application data rate changes. To handle the load balancing, on-going calls will be transferred from an access network to another. The two main targets of our proposed algorithm are the admission control and the network-initiated handover.

A. ADMISSION CONTROL

The admission control is employed to admit or reject a new originating or handing over communication in order to avoid overload situations. A connection request to a specific BS will be accepted if the BS's load, including the contribution of the incoming communication, is below an admission threshold $d_{th}$. Otherwise, the new incoming communication will be redirected to the least loaded overlapped access network. If all BSs in the coverage area could not accommodate the new communication, the connection request is rejected. If the incoming communication is a handing over one, the admission threshold is greater than the one used for a new originating communication. It is generally preferable to refuse the new calls rather than to drop the on-going calls. That explains also why we choose a load balancing threshold $d < 1$. In our solution, we choose to always accept the handing over users.

However, the admission control is just a first step in the load balancing process as it only deals with incoming communications and it does not treat the load fluctuation of ongoing ones. Moreover, trying to redirect an originating communication to a less loaded access system (redirect from one technology to another) may not be possible if the communication is initiated from a single-mode terminal. In this case, it may be better to accommodate the originating single-mode user and to force a multimode user to make a vertical handover to a coordinated access system. That motivates the need to use handover enforcement to effectively distribute the load over the heterogeneous systems.

B. HANDOVER ENFORCEMENT

In addition to the admission control, it is essential to have a mechanism to detect and handle imminent overload situations. Such mechanism is known as the handover enforcement since its main role is to select suitable users in a heavily loaded access network and force them to handover. Instead of balancing the resources of the overall system as described in the optimal algorithm, our proposed solution aims at redistributing locally the load of a heavily loaded BS around its neighboring overlapped BSs. In turn, the neighboring BS will redistribute its load to its own neighboring BSs and so on. By doing so, the load of the overall system will be then balanced. In fact, the handover enforcement will be triggered when the load of a specific BS is greater than $d$. The algorithm execution is continued until $x2 = 0$ or we cannot find a handover to improve index $x2$. In our proposition we only consider one-move and two-move operations during the handover enforcement since considering more than two consecutive moves is not realistic in on-line system due to its computation time.

## 5 Simulation and Results

In this section, we derive an analytical model for reliability of handover procedure between UMTS and WLAN/WiMax network. The following notations are used in our analysis.

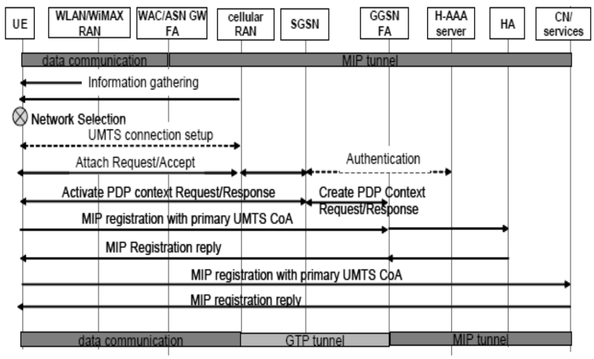



**Fig. 4.** Handover procedure from WLAN/WiMAX RAN to UMTS RAN

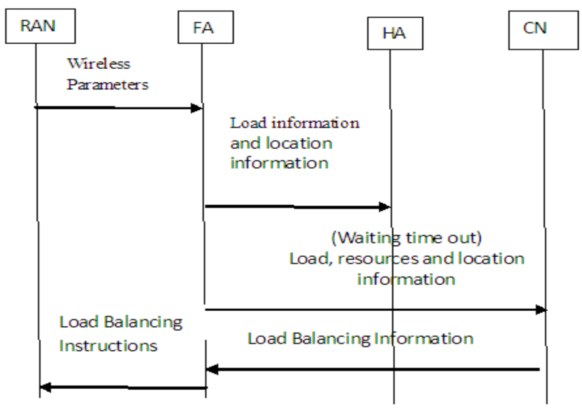



**Fig. 5.** Signaling Flow Chart

$i$: System type. $i$=1 denotes UMTS system; $i$=2 denotes WLAN system; $i$=3 denotes WiMax system.

$b_{ij}$: The traffic intensity between $FA_i$ and $HA_j$.

$c_j$: The traffic intensity between $HA_j$ and CN.
$R_{CN}$: The reliability of CN.
$R_{HA}$: The reliability of HA.
$Rc$: The reliability of junction line between CN and HA.
$a_i$: The signaling overhead of transferring load information once between one FA located in system type $i$ and HA.
$d$: The signaling overhead of transferring load information once between one HA and CN.
$e$: The signaling overhead of transferring load information once between one main CN and standby CN.
$A_i$: The number of RANs in system type $i$.
$D$: The number of HAs.

$A_{ij}$: The number of RANs for system type *i* in the basic grid *j*.
$\lambda_i$: The traffic arrival rate of system type *i*.
$\mu_i$: The service rate of system type *i*.
$m_i$: The cell capacity of system type *i*.
$k_{i1}$: The light threshold of system type *i*.
$k_{i2}$: The heavy threshold of system type *i*.
*T*: The period of transferring load information among RMU.
To facilitate the analysis, we assume that there are only one main CN and one standby CN.

*A. Simulation Scenario*

The scenario is a medium urban area, where both UMTS system and WiMax system cover the whole area while WLAN system covers the hot spots only. In order to reduce the complex of simulation, we assume that there are all the three types of APs in each basic grid, and the number of APs for the same system is equal in every basic grid.

The values of parameters used in simulation are as follows. $T$=0.1s, $A_1$=600, $A_2$=900, $A_3$=600; $R_{IS}$=0.99,

$R_{HA}$=0.98, $R_c$=0.97; $b_i$=1$e_{rl}$; $K_1$=1, $K_2$=1, $K_3$=1; $a_1$=1, $d$=1, $e$=1; $m1$=60, $m2$=20, $m3$=80; $\eta_{iTH1}$=0.7, $\eta_{iTH2}$=0.9; $\mu_i$=1/180s. Where $i$=1, 2, 3.

*A. Simulation Results*

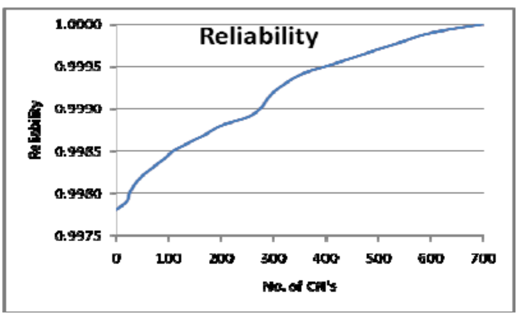
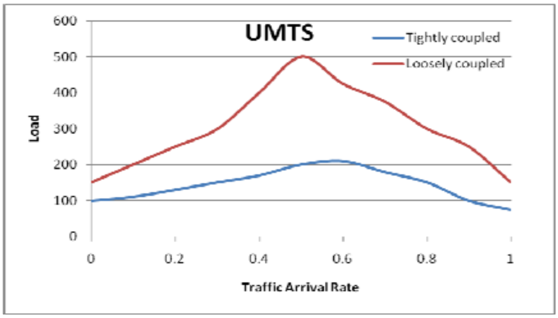
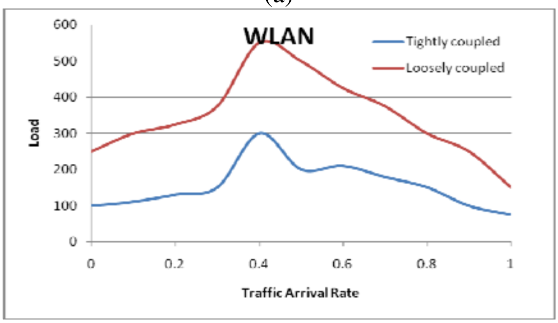



**Fig. 6.** Network Reliability in a Heterogeneous environment

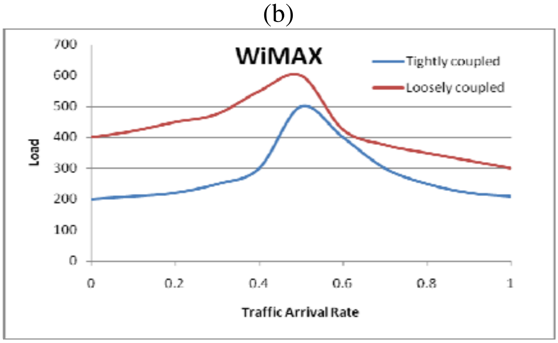
Figure 6 gives the reliability while choosing the network. In the heterogeneous environment the reliability is high even when the number of nodes is high. The higher transmission latency experienced in the heterogeneous network can be observed in the graph provided (Figure. 6). On the transmit side, the transmission is performed with no silence period. On the receiver side, handing over to the cellular network introduces more latency, results in a silence period the order of magnitude of which is equal to the latency difference between both networks. The use of an adaptive buffer at the receiver side makes it transparent to the user which is reflected as a smooth seamless flow in the heterogeneous Networks. When considering the 3G/ Wimax cellular network, the number of users is high compared with the other networks and also had a wider coverage but there is a pitfall at the end, the bandwidth fluctuates



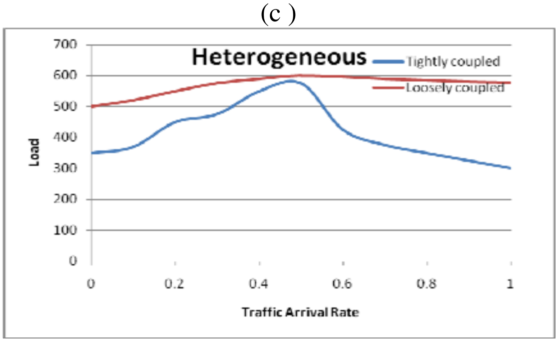

(a)




(b)




(c )




(d)

**Fig. 7.** (a, b, c, d). Load sharing in different networks

beyond 80%. At the time of mobility the network coverage is limited as shown in figure 7 (a,b,c,d). Load sharing in the Heterogeneous Network Environment, by considering the positive measures of the above mentioned networks and by having a thorough understanding between the available networks the heterogeneous network is designed. The heterogeneous network provides maximum throughput, minimum number of handoffs and maximum coverage at mobile. By designing a proper QoS standard and having proper understanding between the network the desires which are explained at the initial paragraph can be achieved. By improving the performance measures by deploying and allocating the code spectrum for the 3GPP network and by having proper power management in the 3G network and by making use of antennas with wider coverage in WLAN environment, the available bandwidth can be maximum utilized and also the number of handoffs can be reduced as the nature of the network present in the graphical architecture between the source and the destination is studied in advance, a maximum throughput can be achieved with minimum tolerable delay or no delay based on the nature of the information that is taken for transmission. The data rate of the heterogeneous network is very close to the available rate as shown in fig. 7.

## 6 Conclusion

In this paper, we have proposed a new load metric which makes it possible to formulate the load balancing as a classic optimization problem. This novel load metric for wireless packet networks is based on the packet scheduling and the radio link quality information. Thank to this new metric, the heterogeneity of different access technologies can be removed. It also facilitates the load balancing operations since it allows load variation anticipation. We introduced a new load balancing index to measure the overload degree of a system. This balancing index leads to minimize the overload degree of a system instead of equalizing the load among the access nodes within a system. We designed a load balancing scheme which consists of admission control and handover enforcement. The proposed handover enforcement based on one-move and two-move iterative search is one of the feasible suboptimal solutions to the problem. The solution can be used in on-line system because it does not require much computation time and because it operates in a distributed way instead of a usual centralized way. It was shown that our proposed approach outperforms the existing approaches. In the future work, we plan to investigate the joint load balancing and resource allocation optimization in heterogeneous networks.

## References

[1] Yu, F., Krishnamurthy, V.: Optimal Joint Session Admission Control in integrated WLAN and CDMA Cellular Networks with Vertical Handoff. IEEE Transaction on Mobile Computing 6(1), 126–139 (2007)

[2] Holis, J., Pechac, P.: Elevation Dependent Shadowing Model for Mobile Communications via High Altitude Platforms in Built- Up Areas. IEEE Transactions on Antennas and Propagation 56(4) (2008)

[3] Imre, S.: Dynamic Call Admission Control for Uplink in 3G/4G CDMA-Based Systems. IEEE Transactions on Vehicular Technology 56(5) (2007)
[4] Zhang, Y., Fujise, M.: Location Management Congestion Problem in Wireless Networks. IEEE Transactions on Vehicular Technology 56(2) (2007)
[5] Morris, D., Hamid Aghvami, A.: Location Management Strategies for Cellular Overlay Networks—A Signaling Cost Analysis. IEEE Transactions on Broadcasting 53(2) (2007)
[6] Chen, H., Wu, H., Kumar, S., Tzeng, N.-F.: Minimum-Cost Data Delivery in Heterogeneous Wireless Networks. IEEE Transactions on vehicular Technology 56(6) (2007)
[7] Chen, Y.-S., Chuang, M.-C., Chen, C.-K.: DeuceScan: Deuce-Based Fast Handoff Scheme in IEEE 802.11 Wireless Networks. IEEE Transactions on Vehicular Technology 57(2) (2008)
[8] Olama, M.M., Djouadi, S.M., Papageorgiou, I.G., Charalambous, C.D.: Position and Velocity Tracking in Mobile Networks Using Particle and Kalman Filtering With Comparison. IEEE Transactions on Vehicular Technology 57(2) (2008)
[9] Misra, A., Roy, A., Das, S.K.: Information-Theory Based Optimal Location Management Schemes for Integrated Multi-System Wireless Networks. IEEE/ACM Transactions on Networking 16(3) (2008)
[10] Xiao, Y., Pan, Y., Li, J.: Design and Analysis of Location Management for 3G Cellular Networks. IEEE Transactions on Parallel and Distributed Systems 15(4) (2004)
[11] Huang, D.-W., Lin, P., Gan, C.-H.: Design and Performance Study for a Mobility Management Mechanism (WMM) Using Location Cache for Wireless Mesh Networks. IEEE Transactions on Mobile Computing 7(5) (2008)
[12] Zhu, Y.-H., Leung, V.C.M.: Optimization of Sequential Paging in Movement-Based Location Management Based on Movement Statistics. IEEE Transactions on Vehicular Technology 56(2) (2007)
[13] Rodríguez-Dagnino, R.M., Takagi, H.: Movement-Based Location Management for General Cell Residence Times in Wireless Networks. IEEE Transactions on Vehicular Technology 56(5) (2007)
[14] Ma, W., Fang, Y., Lin, P.: Mobility Management Strategy Based on User Mobility Patterns in Wireless Networks. IEEE Transactions on Vehicular Technology 56(1) (2007)
[15] Niyato, D., Hossain, E.: A Noncooperative Game-Theoretic Framework for Radio Resource Management in 4G Heterogeneous Wireless Access Networks. IEEE Transactions on Mobile Computing 7(3) (2008)
[16] Roberts, M.L., Temple, M.A., Raines, R.A., Mills, R.F., Oxley, M.E.: Communication Waveform Design Using an Adaptive Spectrally Modulated, Spectrally Encoded (SMSE) Framework. IEEE Journal of Selected Topics in Signal Processing 1(1) (2007)