International Journal of Software Engineering and Its Applications
Vol.7, No.6 (2013), pp.127-136
http://dx.doi.org/10.14257/ijseia.2013.7.6.11

Performance Analysis of Adaptive QoS Handoff Mechanism Using Service Degradation and Compensation

Jihoon Lee[1] and Hyun-chul Kim[2]
[1]*Information and Telecommunication Engineering, Sangmyung University, Korea*
[2]*Computer Software Engineering, Sangmyung University, Korea*
vincent@smu.ac.kr, hkim@smu.ac.kr (corresponding author)

Abstract

User mobility management is one of the important components of mobile multimedia environments. That is, mobile users should obtain network resources after handoff for seamless service supports. So, an adaptive handoff scheme has been proposed to minimize the connection dropping probability for handoff connections. This paper shows the comparative performance analysis for the adaptive QoS handoff mechanism with existing solutions such as non-priority scheme (NPS) and guard-channel scheme (GCS). It has been shown from the performance evaluations that the adaptive QoS mechanism provides lower dropping and blocking probability than existing schemes.

Keywords: *Resource allocation, Adaptive QoS, Performance analysis*

1. Introduction

A channel in the wireless network is a fixed block of communication medium such as a 2-tuple <time slot, carrier frequency> in the time-division multiple access (TDMA) systems, or simply a fixed radio frequency as in the frequency-division multiple access (FDMA) systems [1]. Multiple channels can be allocated to a single user to satisfy higher bandwidth requirements. Otherwise, slots larger than a certain minimum duration may be defined to support higher bandwidth traffic. For example, in the Future Radio Wideband Multiple Access System (FRAMES) [2] multiple access proposal for Universal Mobile Telecommunications System (UMTS), the GSM evolution for 3Gwireless systems, three kinds of multi-slot structure are defined. These are $1/8$th, $1/16$th rate slots, and $1/64$th rate slots. IS-136, North American TDMA (NA-TDMA), is another cellular radio interface standard that specifies allocation of multiple time slots to high bandwidth users.

Compared to wired networks, the fluctuation in resource availability in wireless networks is much more than severe and results from inherent features such as fading and mobility. The adaptive framework only takes mobility characteristics into consideration. That is, an adaptive multimedia call changes its bandwidth level only when there is a new call arrival, a call completion, or a handoff.

The adaptive resource allocation scheme (AREAS) [1] is introduced that manages the bandwidth allocation of every call in each cell. Here adaptation means the bandwidth allocation of incoming calls (new or handoff call) and the change of bandwidth level of the existing calls in a cell depending on network conditions. This paper presents the comparative analysis for the QoS-based handoff schemes and shows how better AREAS scheme is than other schemes.

ISSN: 1738-9984 IJSEIA

2. Adaptive QoS Mechanism

Many applications are adaptive in nature and can therefore generate variable QoS requirements. The adaptive multimedia paradigm can play an important role to mitigate the highly-varying resource availability in wireless/mobile networks. For example, in non-adaptive multimedia framework where the bandwidth of a call is fixed, the handoff call is forced to terminate if there is no available bandwidth in the forward cell. In contrast, in an adaptive multimedia framework, it is possible to overcome the link overhead condition by reducing the bandwidth of individual calls, which is called as bandwidth adaptation, thereby accepting the handoff call [9-14].

Originally, the concept of adaptive multimedia service was introduced in wired networks to cope with network congestion. Broadly, two approaches in adaptive multimedia have been proposed in the literature. In the first approach, the source adjusts the rate (or bandwidth) of a multimedia stream depending on the conditions of the network. The value of the rate is usually continuous. In the second approach, on the other hand, the multimedia stream is compressed in the form of layered (or hierarchical) coding to support heterogeneous receivers. Thus, each receiver can selectively choose the subset of layered coding depending on both its capability and bandwidth availability. For example, most video compression standards, like MPEG, JPEG and JBIG, have a notion of 'progressive mode' or 'hierarchical mode', *e.g.*, a lossy compression using MPEG encoding can yield a data rate from 1.5Mbps to 6.0Mbps for a digital NTSC signal, a lossy compression to a CD quality audio can yield data rate from 384Kbps to 1.41Mbps depending on the quality desired [3]. In the wireless communication environment, recently developed hardware for video coding can adaptively deliver digital video at rates between 60Kbps and 600Kbps [3]. In fact, it is believed that with adequate runtime support, future mobile computing applications can use QoS bounds in order to adapt effectively to dynamic network conditions. Another scheme in the second approach is transcoding where one multimedia coding is changed into another (*e.g.*, MPEG-1, 2 into H.263) [3].

In this paper, the layered coding approach is adopted where the bandwidth of a call can take a set of discrete values, and a sender transmits the layered coding of a multimedia stream to a mobile host (MH). If a cell is 'underloaded', the MHs (with an ongoing adaptive multimedia call) in the cell receive the full multimedia stream, *i.e.*, the whole set of layered coding. However, if congestion takes place in the cell, the layered coding is adapted at the base station (BS). In other words, the subset of the layered coding is filtered or transcoded at the BS to adapt to the situation of the 'overloaded' cell [20].

3. Adaptive Resource Allocation Scheme (AREAS)

An adaptive resource allocation scheme [1] manages the bandwidth allocation of every call (including on-going calls and new/handoff call) in each cell depending on network conditions. During the call setup period, the higher layers are required to provide the following parameters to the QoS framework: (i) maximum bandwidth required, (ii) average bandwidth required, and (iii) minimum bandwidth required. We call this input parameter as a service profile. The higher layers can obtain the service profile information by mandating a reservation setup signaling, before accepting a data flow into the network. The objectives of the adaptive resource allocation scheme are to keep dropping probability of handoff calls low and to reduce blocking probability of new calls, while improving resource utilization.

AREAS tries to minimize the number of calls with lower than average bandwidth with each call. AREAS is arised under two conditions: (a) surplus resources need to be distributed among competing MHs, and (b) an incoming call can be admitted through resource re-allocation, though the currently available resources are insufficient to admit the call. The former case involves increasing the resources for degraded calls (compensation part), while the latter case involves reducing the resources for on-going calls in order to accommodate the incoming call (degradation part).

AREAS for compensation is based on the maxmin optimality criterion [4], which is both fair and efficient, in the sense that all degraded calls get an equal share of this surplus capacity. In addition, AREAS for degradation applies to the case where an incoming call arrives in the given cell and the currently available bandwidth is insufficient. Depending on network conditions, AREAS allocates suitable bandwidths to the incoming call (new or handoff call) and re-allocates the bandwidth of the on-going calls, if necessary. There are six steps in AREAS for degradation (Figure 1). Table 1 shows the components of AREAS.

Table 1. The Components for AREAS

Symbol	Descriptions
BW_{avg}	average bandwidth
BW_{min}	minimum bandwidth
BW_{max}	maximum bandwidth
BW_{A}	available bandwidth in the given cell
BW_{D1}	amount of bandwidth accumulated by changing all calls with more than BW_{min} into calls with BW_{min}
BW_{D2}	amount of bandwidth accumulated by changing all calls with more than BW_{avg} into calls with BW_{avg}
BW_{cur}	the currently allocated bandwidth
BW_{diff}	the required bandwidth for compensation
BW_{thresh}	the threshold bandwidth for bandwidth compensation of degraded calls
BW_{ret}	amount of bandwidth returned by call completion and handoff

When a call leaves the cell or is completed, the BWA will increase. The change in BWA may enable one or more calls to upgrade their bandwidth. A call is referred to be compensated if the bandwidth of the call is changed from lower than BW_{avg} to at least BW_{avg}.

Bandwidth compensation to the average bandwidth implies “re-tuning” the resources. The adaptive resource allocation scheme for compensation will sequentially shuffle the existing bandwidths towards the average bandwidth with each call, requiring users to "tune" to new channels almost instantaneously. AREAS for compensation can sometimes be very expensive and time consuming as a large number of channel re-assignments need to be done. Hence, a partial bandwidth compensation algorithm will be required. The partial bandwidth compensation scheme checks the mobility pattern and starts when the amount of surplus resources (namely, $BW_A + BW_{ret}$) exceeds the threshold level (*i.e.*, fixed percentage of total resources). This scheme is less frequent in resource re-allocation and sustains lower delays.

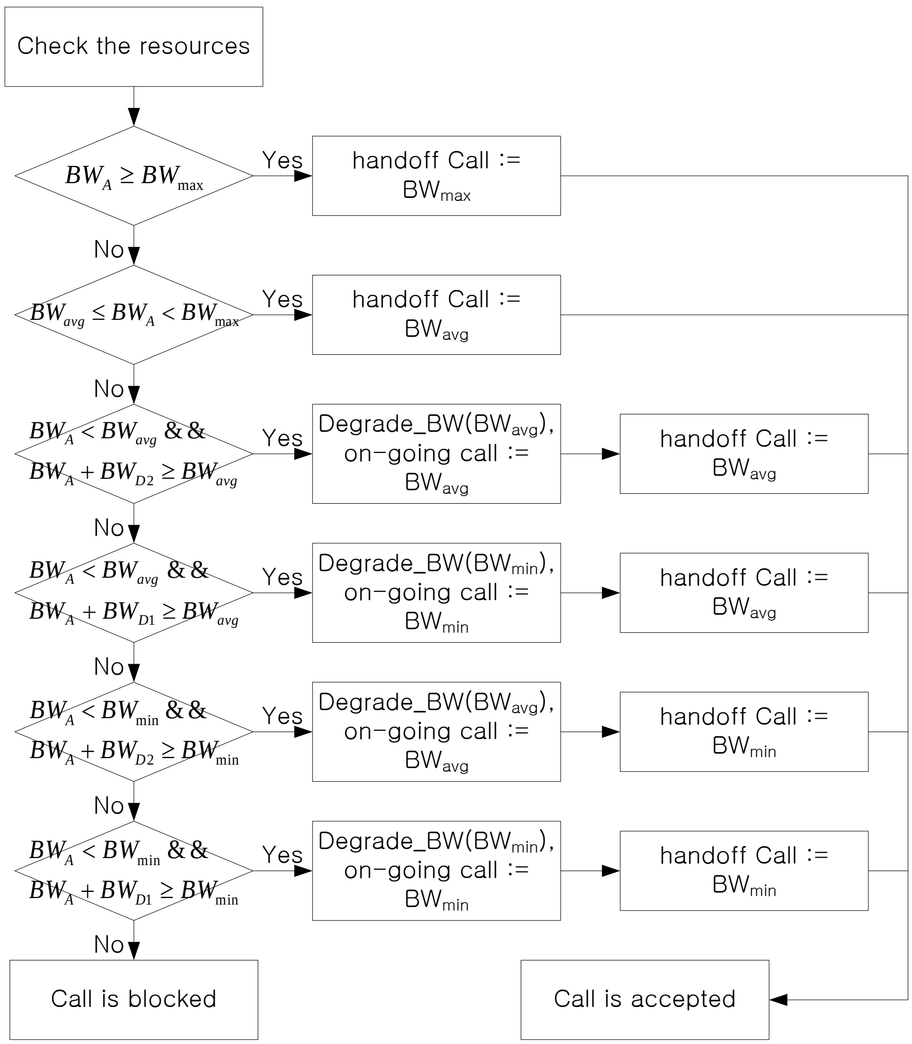

Figure 1. Flow Diagram for AREAS in Bandwidth Degradation Part

4. Numerical Analysis

The system model used assumes a fixed channel assignment homogeneous network in which a number of channels (which can be time slots, frequencies, spreading codes, etc.) are assigned to each cell. The call carries a variable bit rate adaptive service which normally requires the use of two channels ($m_w = 2$), but can tolerate a lower QoS level with the use of only one channel ($m_{wd} = 1$). Note that it is not necessary to assume fixed values for m_w and m_{wd}, however, to simplify the numerical notations later on, we do assume them constant.

The traffic model makes the commonly used assumptions that new calls are generated according to independent Poisson processes with mean rates λ_n for adaptive calls. Furthermore, the mean call holding times are exponentially distributed with means $1/\mu$ for adaptive calls. Finally, the mean dwelling time in a cell, that is, the time spent in one cell before handing off to a neighboring one, are also exponentially distributed with the mean $1/\mu$.

In AREAS model, since the network is assumed to be homogeneous, the system performance can be deduced from the performance of single cell. The system at single cell can be modeled by one dimensional Markov process. Figure 2 shows an example of state space, where each state defined by *i* is the number of on-going adaptive calls. For all *i*, the balance equations are

$$P_i = \begin{cases} \dfrac{(\lambda_n+\lambda_h)^i}{i!(\mu+\eta)^i} \cdot P_0, & 0 \le i < m \\ \dfrac{\lambda_h^{\,i-m}(\lambda_n+\lambda_h)^m}{i!(\mu+\eta)^{i-m}} \cdot P_0, & m \le i \le K \end{cases} \tag{1}$$

where P_0 is given by

$$P_0 = \left[\sum_{i=0}^{m} \frac{(\lambda_n+\lambda_h)^i}{i!(\mu+\eta)^i} + \sum_{j=m+1}^{K} \frac{\lambda_h^{\,j-m}(\lambda_n+\lambda_h)^m}{i!(\mu+\eta)^{j-m}} \right]^{-1} \tag{2}$$

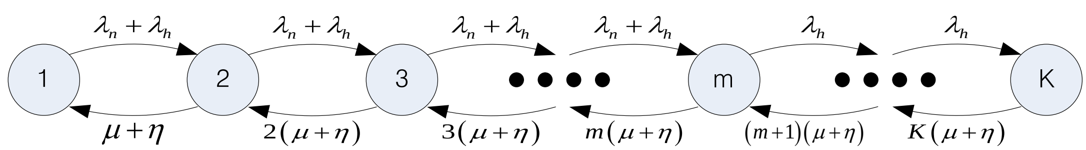

Figure 2. Rate diagram of AREAS Mechanism

The above equilibrium state probabilities can be used to calculate the call blocking probability and dropping probability using the following:

$$\begin{aligned} p_n &= \sum_{m_w j > (N-m_w)} P_j \\ p_h &= \sum_{m_{wd} j > (N-m_{wd})} P_j \end{aligned} \tag{3}$$

With the non-priority scheme (NPS) [5]-[7], $p_n = p_h$. The channel occupancy time is the minimum of the call holding time and the remaining call residual time. Because these are exponentially distributed for adaptive calls, the mean channel occupancy time for adaptive calls is $1/(\mu+\eta)$. Let $\lambda = \lambda_n + \lambda_h$ be the total arrival rates of adaptive calls to a cell. Then a traffic load is given by $\rho = \lambda/(\mu+\eta)$. Figure 3 shows the Markov process for the number of adaptive calls in NPS model. For all i, the balance equations are

$$P_i = \frac{(\lambda_n+\lambda_h)^i}{i!(\mu+\eta)^i} \cdot \left[\sum_{j=0}^{m} \frac{(\lambda_n+\lambda_h)^j}{j!(\mu+\eta)^j} \right]^{-1} \tag{4}$$

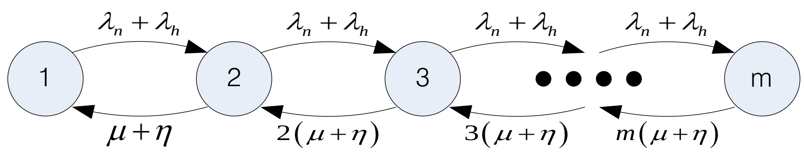

Figure 3. Rate Diagram of NPS Mechanism

Using the above balance equation, the probabilities $p_n = p_h$ are given by $p_n = p_h = \sum_{i>\lceil N/m_w \rceil} P_i$, where $\lceil x \rceil$ denotes the maximum integer value less than or equal to x.

The analysis of the guard channel scheme (GCS) has been considered in [7, 8]. Essentially, the system at single cell can be modeled again using one dimensional Markov process. Figure 4 shows an example state space. The call blocking and dropping probability for adaptive calls can be determined from the following:

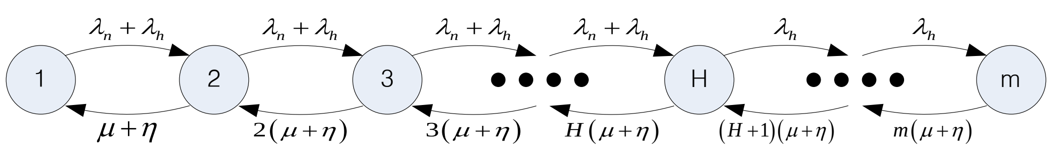

Figure 4. Rate Diagram of GCS Mechanism

$$P_i = \begin{cases} \dfrac{(\lambda_n+\lambda_h)^i}{i!(\mu+\eta)^i} \cdot P_0, & 0 \le i < m \\[2ex] \dfrac{\lambda_h^{\,i-m}(\lambda_n+\lambda_h)^m}{i!(\mu+\eta)^{i-m}} \cdot P_0, & m \le i \le K \end{cases} \tag{5}$$

$$p_n = \sum_{m_w j > (N-N_g-m_w)} P_j \qquad p_h = \sum_{m_w j > (N-m_w)} P_j \tag{6}$$

5. Numerical Results

In this section, to assess the performance of the adaptive resource allocation scheme, we make comparisons with the performance of both the non-priority scheme and guard-channel schemes using numerical analysis.

A single cell with N=20 channels is considered. Also, we choose the call completion rates equal $\mu = 0.3\, calls/\min$ and the dwelling rates $\eta = 0.6\, calls/\min$. For the guard channel scheme, we consider the number of channels reserved for exclusive access by handoff calls to equal four.

Figure 5 shows the connection blocking probability of adaptive calls against various connection arrival rates. As shown in Figure 5, the blocking probabilities of the AREAS are between those of NPS and GPS. By allowing the connections to adapt to using fewer channels, the AREAS clearly increases the utilization of the available resources in the cell, and this causes new connections to be blocked more often compared to NPS where both new and handoff connections are not differentiated.

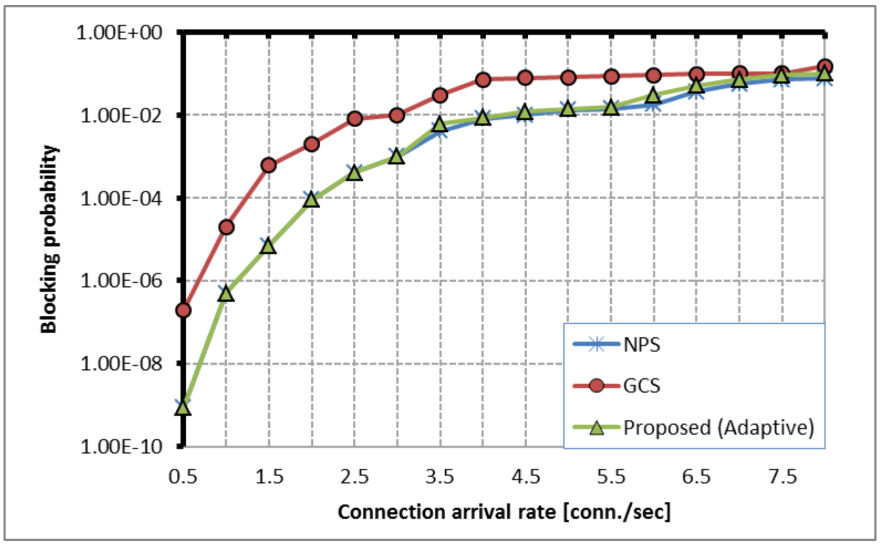

Figure 5. Blocking Probability of New Connection

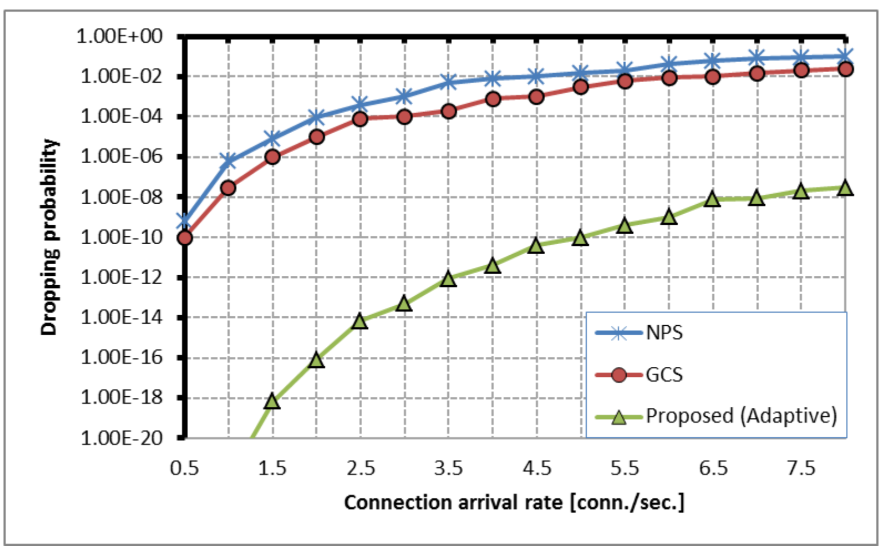

Figure 6. Dropping Probability of Handoff Connections

Figure 6 shows the connection dropping probabilities against the connection arrival rate. It can be seen that the AREAS performs much better than the other schemes such as GCS and NPS. It is due to the fact that with AREAS, there is no resource reservation for handoff connections and the probability of connection dropping probability is reduced by degrading QoS levels of connections that carry adaptive traffic.

To provide another means of comparing the performance of three handoff schemes, Figure 7 depicts the probability of call non-completion (*i.e.*, the probability that either a new call attempts is blocked or handoff call is forced to terminate) against the total call arrival rate. For both adaptive and non-adaptive calls, the AREAS performs better than the other schemes, indicating that the AREAS results in a better utilization of the available bandwidth in the cell (less bandwidth is wasted to support calls that are subsequently not successful).

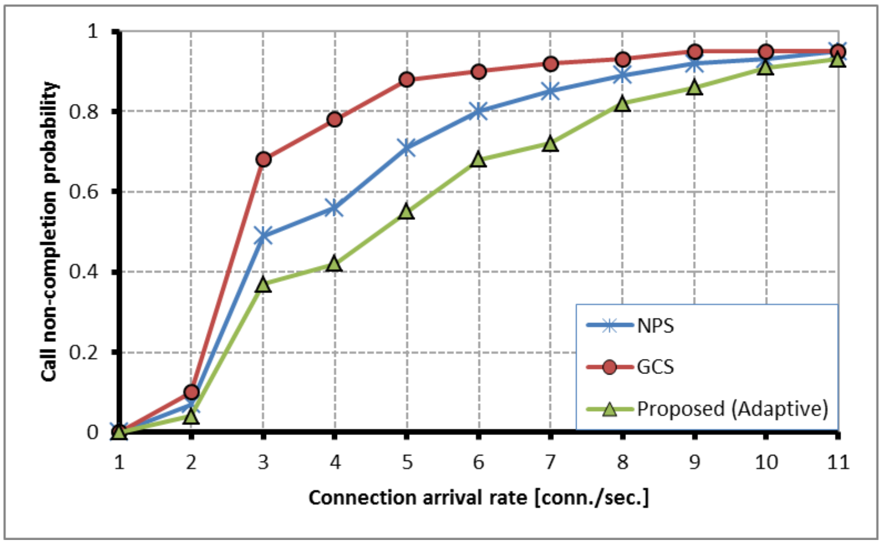

Figure 7. Call Non-completion Probability

6. Conclusion

This paper shows the analysis results for the adaptive resource allocation mechanism in terms of the blocking and dropping probability of new and handoff connections. It is shown from performance results that the adaptive scheme provides lower dropping probability of handoff connections by degrading the QoS levels of on-going connections. It is important that from a user's perspective, a connection terminated in the middle of a call due to a handoff failure is more annoying than having a new call attempt blocked occasionally.

Acknowledgements

This work was supported by Research Funds of Sangmyung University in 2013.

References

[1] J. Lee, T. Jung, S. Yoon, S. Youm and C. Kang, "An adaptive resource allocation mechanism including fast and reliable handoff in IP-based 3G wireless networks", IEEE Personal communications Magazine, DOI: http://dx.doi.org/10.1109/98.892258, vol. 7, no. 6, **(2000)** December, pp. 42-47.

[2] E. Nikula, A. Toskala, E. Dahlman, L. Girard and A. Klein, "FRAMES multiple access for UMTS and IMT-2000", IEEE Personal Communications Magazine, DOI: http://dx.doi.org/10.1109/98.667941, vol. 15, no. 2, **(1998)** April, pp. 16-24.

[3] M. Naghshineh and M. W. LeMair, "End-to-End QoS Provisioning in Multimedia Wireless/Mobile Networks Using an Adaptive Framework", IEEE Communications Magazine, DOI: http://dx.doi.org/10.1109/35.634764, vol. 35, no. 11, **(1997)** November, pp. 72-81.
[4] R. D. Caytiles, Y. E. Gelogo and B. Park, "An Integrated Security Handover Scheme for Seamless Convergence Services over IP-based Mobile Networks", Int'l Journal of Control and Automation, URL: http://www.sersc.org/journals/IJCA/vol4_no4/4.pdf, vol. 4, no. 4, **(2011)** December, pp. 55-62.
[5] R. Cheung, "An Adaptation Control Model to Support Mobile Web Access", Int'l Journal of Control and Automation, URL: http://www.sersc.org/journals/IJCA/vol1_no1/papers/02.pdf, vol. 1, no. 1, pp. 9-16, **(2008)** December.
[6] W. Ni, W. Li and M. Alam, "Determination of optimal call admission control policy in wireless networks", IEEE Trans. On Wireless Communications, DOI: http://dx.doi.org/10.1109/TWC.2009.080349, vol. 8, no. 2, **(2009)** February, pp. 1038-1044.
[7] Z. Zheng, "QoS Based Dynamic Channel Reservation for Handoff Prioritization in Mobile Networks", Journal of Computational Information Systems, URL: http://www.jofcis.com/publishedpapers/2012_8_9_3831_3837.pdf, vol. 8, no. 9, **(2012)** September, pp. 3831-3837.
[8] Z. H. Zheng and W. H. Lam, "Performance Analysis of Dynamic Channel Assignment with Queuing and Guard Channel Combined Scheme for Handoff Prioritization", IEE Electronics Letters, DOI: http://dx.doi.org/10.1049/el:20021177, vol. 38, no. 25, **(2002)** December, pp. 1728-1729.
[9] Y. E. Gelogo, R. D. Caytiles and B. Park, "A Robust Secured Mobile IPv6 Mechanism for Multimedia Convergence Services", Int'l Journal of Multimedia and Ubiquitous Engineering, URL: http://www.sersc.org/journals/IJMUE/vol6_no4_2011/6.pdf, vol. 6, no. 4, **(2011)** October, pp. 61-66.
[10] Y. E. Gelogo and B. Park, "Reducing Packet Loss for Mobile IPv6 Fast Handover (FMIPv6)", Int'l Journal of Software Engineering and Its Applications, URL: http://www.sersc.org/journals/IJSEIA/vol6_no1_2012/7.pdf, vol. 6, no. 1, **(2012)** January, pp. 87-92.
[11] G. Lee and J. Lee, "A Monitoring Method for Supporting QoS in Next Generation Mobile Communication Network", Journal of the Korea Academia-Industrial cooperation Society, URL:http://www.kais99.org, vol. 13, no. 8, **(2012)** August, pp. 3680-3686.
[12] G. Lee, "Resource Allocation and Handoff for Mobile Multimedia Service", Journal of the Korea Academia-Industrial cooperation Society, URL:http://www.kais99.org, vol. 11, no. 2, **(2010)** February, pp. 5029-5035.
[13] K. Ro, A Study on a Mobile Terminal Platform for a High Speed Mobile Multimedia System, Journal of the Korea Academia-Industrial cooperation Society, URL:http://www.kais99.org, vol. 10, no. 1, **(2009)** January, pp. 6-103.
[14] R. D. Caytiles and B. Park, "Mobile IP-Based Architecture for Smart Homes", Int'l Journal of Smart Home, URL: http://www.sersc.org/journals/IJSH/vol6_no1_2012/3.pdf, vol. 6, no. 1, **(2012)** January, pp. 29-36.

Authors

Jihoon Lee, he received the B.S., M.S, and Ph.D. degrees in Electronics engineering from Korea University, Korea in 1996, 1998, and 2001, respectively. From 2002 to 2011, he worked at Samsung Electronics as a senior research member. He is currently an assistant professor in Department of Information and Telecommunication Engineering, Sangmyung University, Korea. His research interests include information centric networking, secure M2M, and network security.

Hyun-chul Kim, he is an Assistant Professor at Sangmyung University. He had worked at Seoul National University as a BK (Brain Korea) assistant professor in 2008~2012, and CAIDA as a visiting scholar in 2006~2007. He received his Ph.D. in Computer Science from KAIST in 2005. His research interests include measurement, analysis and modeling of the Internet traffic and topology.

