Integrated QoS Signaling with Adaptive Resource Allocation Strategy for Multi-service Networks

Bor-Jiunn Hwang[1]
bjhwang@mcu.edu.tw

Fang-Chang Kuo [2]
kfc@niu.edu.tw

Jung-Shyr Wu [3]
jswu@wireless.ee.ncu.edu.tw

Jiun-Yu Tu [3]
945403002@cc.ncu.edu.tw

Yu-Shun Liu [4]
rescue@vnu.edu.tw

Abstract

Triple play (voice, video and high speed data) over IP is likely to be a popular service in the Next Generation Network (NGN). The major challenges to this kind of service are due to the different qualities of service (QoS) required by different types of applications and the corresponding resource allocation works. In this paper, we analyze the functional architecture of session and resource management with SIP, RSVP, COPS and Megaco/H.248, and propose an integrated QoS signaling architecture to guarantee QoS. In addition, a variation based on adaptive resource allocation strategy, called Variance-based Adaptive Resource Regulation (VARR), is proposed to improve the performance. To provide multi-service with the required QoS, the total bandwidth is divided into three parts for different types of traffic. These parts are regulated according to blocking rate, dropping rate, and variation of variable bandwidth. The effect of mobility on the system performance was evaluated and compared by simulation. The results reveal that the proposed VARR does improve the performance in supporting multi-services heterogeneous networks.

Keywords

QoS, Adaptive resource allocation strategy, integrated QoS signaling, Multi-services

1. INTRODUCTION

With progressive development in the techniques of broadband mobile communication technology (e.g. Beyond 3G cellular, etc.), the B3G/4G network has recently gained significant attention. Offering triple play (voice, video and high speed data) over IP is likely to be a popular service in the Next Generation Network (NGN). However, there are a number of technical challenges with regards to the rollout of triple play services due to different characteristics of these service and different burdens on the network that provides access to these services. The major challenges are mainly come from aspects of heterogeneous wireless access environments, qualities of service (QoS) needed for multiple types of applications with different requirements, adaptive resource allocation, etc. Thus, the NGN networks must use comprehensive technologies to achieve a range of services broader than the traditional services provided in current systems.

We have considered two basic paradigms for QoS signaling, referring to as path-coupled and path-decoupled. In the former case, signaling messages are routed only through nodes that are on the data path, while in the latter case, nodes are not assumed to be on the data path. There are potentially significant differences in the way that the two signaling paradigms should be analyzed. Moreover, interoperation between these two signaling paradigms is an important issue when mutual operation is essential to maintain guaranteed QoS.

To reduce the probability of handoff failure due to lack of resources in adjacent cells, a basic approach is to reserve resources for handoff calls. The best-known reservation scheme is guard channel (GC) scheme and its numerous variations. One of the challenges in moving to a multi-service system is that the limited bandwidth has to be shared among multiple traffic classes. To solve this problem, an effective and efficient bandwidth allocation strategy is necessary, especially for mobile networks with multi-service.

---

[1] Department of Computer and Communication Engineering, Ming-Chuan University, Tao-Yuan, Taiwan, 33348

[2] Department of Electronic Engineering, National Ilan University, Ilan, Taiwan, 260

[3] Department of Communication Engineering,National Central University, Tao-Yuan, Taiwan, 320

[4] Department of Computer Science and Information Engineering, Vanung University, Tao-Yuan, Taiwan, 32061

In this paper, we analyze the functional architecture of a session and resource management with SIP, RSVP, COPS and Megaco/H.248, and propose an integrated QoS signaling architecture exclusively for QoS guarantee. The media transmission path will be changed since the user is mobile. Therefore, how to achieve the QoS of a multimedia session will be studied by investigating the effect on the action of user mobility. Finally, adaptive resource management mechanism based on variation is proposed which is adjusted the allocated resource between different traffic types to improve system performance in terms of reduced blocking probability and dropping probability.

The remainder of the paper is organized as follows. In the next section, we describe related work. The system model is described in section 3. In section 4, the simulation results are presented to demonstrate the system performance. Finally, conclusions are drawn in section 5.

2. Related work

Several methods have been suggested to apply RSVP for mobile Internet, such as Mobile RSVP (MRSVP) and Hierarchical Mobile RSVP (HMRSVP) [4, 5]. Kim and Jeon proposed a resource reservation scheme with PAR-SIP soft handoff to achieve QoS for real-time multimedia communications [6]. Ban, et al. proposed a hierarchical mobile IP with a paging (P-HMIP) scheme to reduce both of the registration and reservation costs [7]. However, these approaches are only considered SIP and RSVP, and they lack COPS and Megaco/H.248. Politis proposed a hybrid scheme to handle macro-mobility for All-IP networks [8], though resource reservation and policy-based management were not taken into account. Sargento, et al. suggested the different phases of a multimedia Internet access session, when using SIP, COPS, Diameter and RSVP in the IP-based access networks without considering user mobile case [9]. Tang and Li, presented an adaptive bandwidth allocation scheme, called Complete Sharing with Preemptive Priority (CSPP) scheme, for integrated voice/data mobile networks, analyzing the model of the CSPP scheme by a two dimensional Markov process [10]. The limitation of that scheme is that the analyzed model will to be complex while multiple types of applications with different requirements are considered. Niyato and Hossain proposed the architecture of a two-tier CAC scheme for a differentiated services cellular wireless network based on call-level and packet-level QoS considerations, though only two types of application were considered in terms of voice and video [11].

A number of articles investigated improving handoff performance by Fractional Guard Channel (FGC) and its numerous variations [12-16]. In [12], they derived recursive formulas for the new call blocking and handoff failure probabilities for FGC policies in cellular networks. That study also compared the effect of user mobility on the maximum system capacity of the GC, Limited Fractional Guard Channel (LFGC), and Uniform Fractional Guard Channel (UFGC) strategies. In [13], a two-level fractional guard channels (TLFGC) scheme to efficiently provide priority access for handoff calls over new calls in cellular systems was proposed. A later study examines the performance of three handover priority schemes in terms of pure guard channel method (GCM), GCM with first-in-first-out (GCM-FIFO), and dynamic priority queuing (DPQ) and compared them for different scenarios [14]. In [15], a handoff technique was proposed by combining the mobile assisted hand off (MAHO) and GC techniques that the MT reports back the received signal strength indicator (RSSI), the BER and the number of free channels available for the handoff traffic. The limitation of these approaches described above is that the analyzed model will be complicated when it considers multiple types of services with different requirements. The adaptive multi-guard channel scheme (AMGCS) for multi-class traffic, an adaptive channel reservation scheme, has been proposed to ensure the quality of service for multimedia wireless cellular networks [16]. However, it will degrade the system performance when various traffic loadings due to the fixed capacity are adopted for each class guard channel.

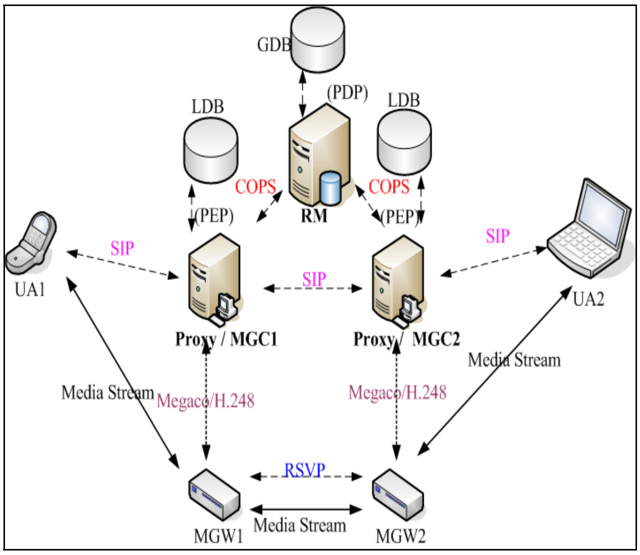


Figure 1: System Architecture

3. System Description

Hierarchical architectures have demonstrated that they will be of a great benefit to scalability and resilience. Therefore, this study proposes hierarchical system architecture by analyzing the functional architecture of IMS, as shown in Figure 1 [1-3], which is composed of a Domain Resource Manager (DRM), Proxy/MGCs and MGWs. Additionally, the media transmission path, signaling path and communication protocol for the interface of the pair components is presented. The wireless access networks can be either WiFi or WiMAX, though they are not shown in Figure 1.

3.1 Integrated QoS Signaling Architecture and Mobility Scenarios

An integrated QoS signaling architecture exclusively for QoS guarantee by analyzing the functional architecture of session and resource management with SIP, RSVP, COPS and Megaco/H.248 is shown in Figure 2. The notations 1.x, 2.x, 3.x, 4.x and 5.x denote SIP, COPS, Megaco/H.248, RSVP signaling and RTP session, respectively. The mobility scenario is shown in Figure 3, in which four cases of mobility scenario are presented, in terms of Intra Media Gateway Mobility, Inter Media Gateway Mobility, Inter Media Gateway Controller Mobility and Inter Domain Mobility, while UA1 (User Agent) moves into its new attached base station, BS2, BS3, BS4 or BS5, respectively. The detail of these scenarios is described as following.

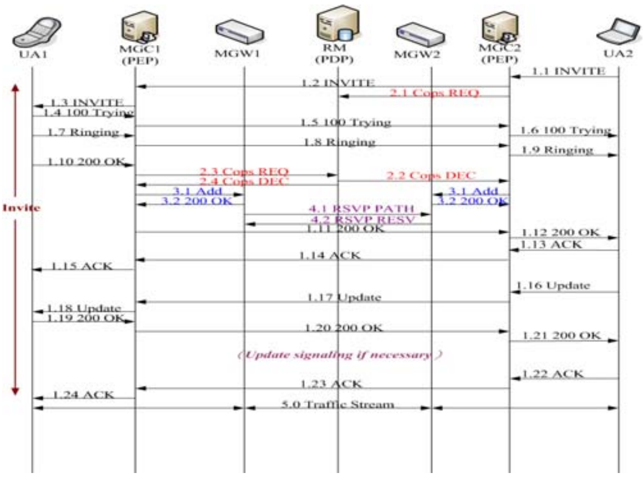


Figure 2: Integrated QoS signaling for Invite

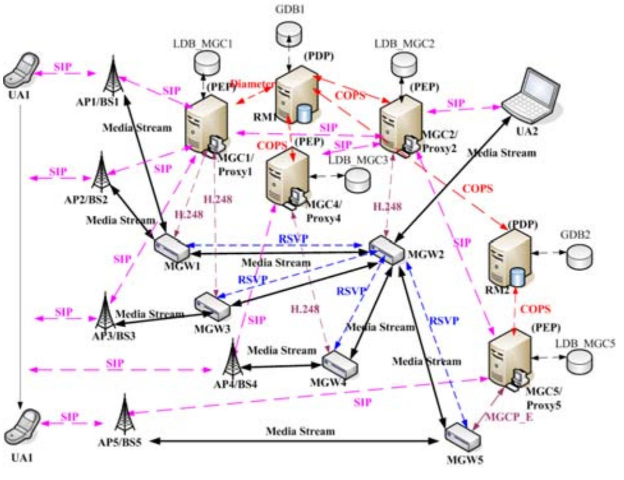


Figure 3: Mobile Scenario

Scenario 1:

UA1 moves into its AP2/BS2 service area, which is connected to the same MGW (i.e. MGW1), denoted as Intra Media Gateway Mobility. In this situation, the Re-invite procedure is executed for UA2 in order to get the new profile of UA1 such as new IP address. Additionally, Update procedure is executed if the media transmission configure was modified.

Scenario 2:

UA1 moves into the AP3/BS3 service area, which is connected to MGW3 controlled by the same MGC (i.e. MGC1), denoted as Inter Media Gateway Mobility. In this situation, it is necessary to perform the resource reservation procedure for new transmission path (i.e. between AP3/BS3 and MGW3) and release the unused resource.

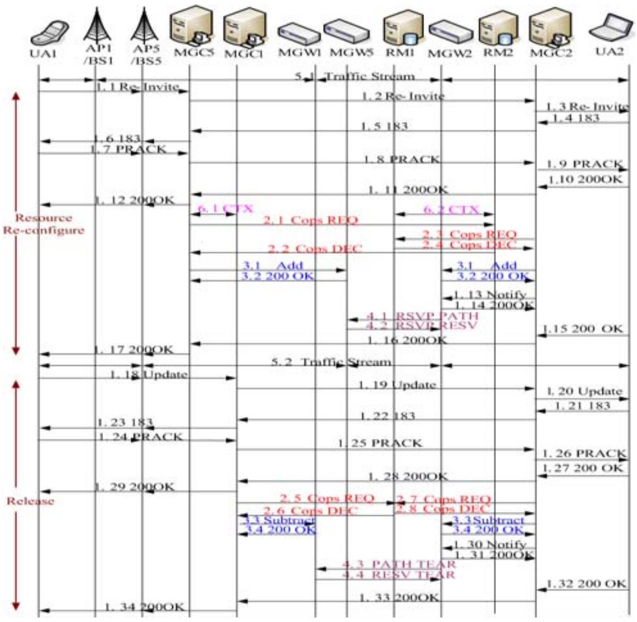


Figure 4: Signaling for Inter Domain Mobility Handoff

Scenario 3:

UA1 moves into the AP4/BS4 service area, which is connected to MGW4 controlled by MGC4, denoted as Inter Media Gateway Controller Mobility. In this situation it is necessary to execute policy procedure, which is similar to the step 2.x in Figure 4, except that resource reservation and release procedures are performed.

Scenario 4:

UA1 moves into the AP5/BS5 service area, which is connected to MGW5 controlled by MGC5, denoted as Inter domain Mobility. This situation is different from others since the MGW5 belongs to another Domain network, dominated by RM2. In this scenario, it is necessary to execute context transfer procedure according from step 6.1 to 6.2 in Figure 4, although policy, resource reservation and release procedures are also performed.

The integrated QoS signaling concerned with the newly added reserved transmission path and release procedures for Inter Domain Mobility is shown in Figure 4.

3.2 Adaptive Resource Regulation

Generally, the GC-based approach is adopted to reduce handoff failure due to insufficient resources in adjacent cells. In response, this paper proposes a novel scheme called Variance-based Adaptive Resource Regulation (VARR) to achieve guaranteeing QoS requirement and efficient resource utilization. VARR is superior to FGC and multi-guard channel with fixed capacity, since VARR is an adaptive FGC with a multi-service scheme, while resources are regulated adaptively between different priority blocks and different service classes according to the variance of bandwidth. Moreover, the handoff rate considered when no dropping occurs. For simplification, only WiMAX service types (including UGS, rtPS, ErtPS, nrtPS and BE) are adopted in the VARR.

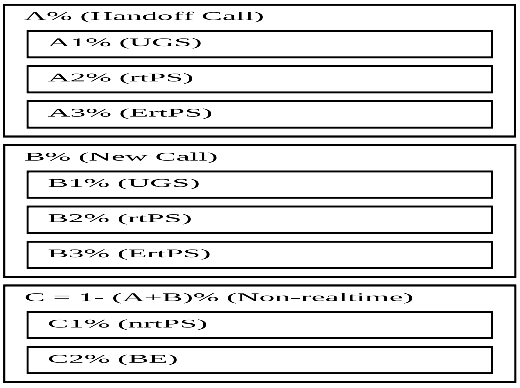


Figure 5: VARR scheme

First, the total bandwidth is divided into three blocks, A, B and C for handoffs, new calls and Non-realtime, respectively. Additionally, only the handoff calls and the new calls for UGS, rtPS and ErtPS types can use the A block and B block resources, respectively. For the Non-realtime calls, including handoff, new and existing calls, only the C block resource can be used. To provide multi-service with required QoS in the A block (resp. B block), each block is divided into three sub-blocks, A1, A2 and A3 (resp. B1, B2 and B3) sub-blocks, for UGS, rtPS and ErtPS, as shown in Figure 5. There are four cases for resource regulation between different blocks and sub-blocks, as described below.

Case 1:

If blocking and dropping do not occurred in the regulated period, it is demonstrable that the allocated resource is in excess for A and B blocks. Thus, distributing some of resource from A and B blocks to C block will improve system performance by reducing blocking and dropping probabilities, and increasing bandwidth of individue session for Non-realtime services.

Case 2:

In this case, blocking occurs while there is no dropping in the regulating period. The bandwidth is regulated in the B block first in the order of B1, B3 and B2. The bandwidth of A and C blocks are regulated in turn while the condition is still not satisfied.

Case 3:

In this case, dropping occurs while there is no blocking in the regulated period. The bandwidth is regulated in the A block first according to the A1, A3 and A2 order. The bandwidth of B and C Blocks are regulated in turn while the condition is still not satisfied.

Case 4:

If both blocking and dropping occur in the regulated period, the bandwidth of A and B blocks proceed with regulating in turn while the condition is still not satisfied.

4. Simulation Model Description and Numerical Results

Table 1 Simulation Parameters

Service Type	Connection hold Duration (sec)	Transmission rate (Kbps)	Arrival rate (1/sec)	Class	Initial Bandwidth
UGS	$\mu_{UGS} = 210$	64	$\lambda_{n\_UGS} = 10$	ClassA1	10%
				ClassB1	20%
ErtPS	$\mu_{ErtPS} = 300$	32	$\lambda_{n\_ErtPS} = 3.3$	ClassA2	10%
				ClassB2	10%
rtPS	$\mu_{rtPS} = 120$	248	$\lambda_{n\_rtPS} = 5$	ClassA3	20%
				ClassB3	10%
nrtPS	$\mu_{nrtPS} = 180$	512	$\lambda_{n\_nrtPS} = 3.3$	ClassC1	10%
BE	$\mu_{BE} = 30$	128	$\lambda_{n\_BE} = 3.3$	ClassC2	10%

The simulation model shown in Figure 3 is implemented in the ns2 simulator [17] and some assumptions involved in this model are stated below.

1. Both the new and handoff connections contain 5 services (i.e. UGS, rtPS, ErtPS, nrtPS, and BE) [18], traffics are generated with the following Poisson distribution with average arrival rates of λ_{n_UGS}, λ_{n_rtPS}, λ_{n_ErtPS}, λ_{n_nrtPS} and λ_{n_BE}, respectively[19].
2. The connection hold duration for UGS, rtPS, ErtPS, nrtPS, and BE traffics are exponentially distributed with average duration μ_{UGS}, μ_{rtPS}, μ_{ErtPS}, μ_{nrtPS} and μ_{BE}, respectively.
3. There are three speed types of mobile hosts, *V_fast*, *V_middle* and *V_slow* following Poisson distributions with average speeds of 25 km/hr, 15 km/hr and 5 km/hr, respectively.
4. The regulation waits are set as $\alpha_{a1} = \beta_{b1} = 1$, $\alpha_{a2} = \beta_{b2} = 2$, $\alpha_{a3} = \beta_{b3} = 3$; the parameters of C1 and C2 are set as C1=2*C2; the regulation thresholds are set as $Dr_{th} = Bl_{th} = 0.01$; the total bandwidth is 50 Mbps.

The simulation parameters are summarized in Table 1.

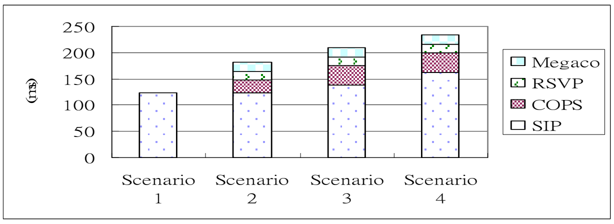
4.1 Performance Evaluation

A. Delay Evaluation

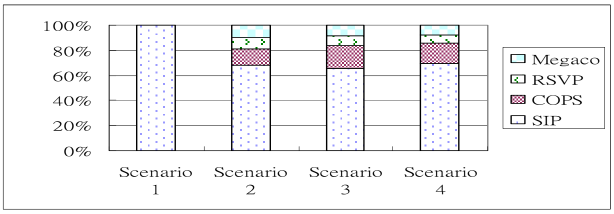
The handoff delay time of integrated QoS signalling will be evaluated as four mobility scenarios in terms of Intra Media Gateway Mobility, Inter Media Gateway Mobility, Inter Media Gateway Controller Mobility and Inter Domain Mobility. Table 2 and Figure 6 present the handoff delay time of integrated QoS signalling for four mobility scenarios. To compare Intra Media Gateway Mobility with Inter Domain Mobility, the amount of difference delay time is 110 ms, which is comprised of 38ms, due to SIP increase and 72 ms due to other signalling. According to Figure 10, the handoff delay time due to SIP dominates since it contains about 66%-69%. Therefore, it is effective to reduce the handoff delay time of SIP by decreasing the total handoff delay time of integrated QoS signaling.

Table 2 Total Handoff signaling delay time

	Intra Media Gateway Mobility	Inter Media Gateway Mobility	Inter Media Gateway Controller Mobility	Inter-domain Mobility
Scenario	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Avg. Handoff delay time	124 ms	182 ms	210 ms	234 ms



(a) Delay time due to different signaling



(b) Ration of delay time for different protocols

Figure 6: Handoff delay time of integrated QoS signaling

B. VARR Evaluation

The performance in terms of blocking and dropping probabilities, and throughput is evaluated by simulating as speed type of V_middle, moreover, it is estimated by comparing with fixed bandwidth regulation method. Figure 7 illustrate the blocking and dropping rates. The total blocking rate and dropping rate for VARR are 4.7% and 3.3%, respectively; where as according to the fixed bandwidth regulation method, they are 14% and 12%, respectively, as shown in Figure 8. The improvements for blocking and dropping rates are 9.3% and 8.7%, respectively, due to adopting variation-based regulating for VARR. In addition, the blocking and dropping rates of UGS and ErtPS for both VARR and fixed bandwidth regulation method are less than 0.1% due to higher priority. The blocking and dropping rates of non-realtime for VARR is less than the fixed bandwidth regulation method caused by dynamically regulating bandwidth for VARR, so, more resource is allocated for non-realtime traffic since the QoS requirement of real-time is met.

4.2 Effects of Mobility

Figure 9 and Figure 10 illustrate the blocking and dropping rates of VARR, respectively for three cases of mobility velocity, indicating that the rates for both UGS and ErtPS are less than 0.1% for all mobility velocity cases due to higher priority. According to Figure 9, the blocking rates for rtPS, nrtPS and BE have a downward trend as the velocity increases, except the nrtPS in the V_fast case. Furthermore, more bandwidth acquired from UGS and ErtPS is allocated for these kinds of traffics. The blocking rate for nrtPS in case of V_fast is higher than the V_middle case since more bandwidth is distributed to higher priorities and lower transmissions rate for BE. According to Figure 10, the dropping rates for rtPS, nrtPS and BE is decrease where comparing V_middle case with V_slow case due to higher variation. In the case of V_fast, the dropping rates for both nrtPS and BE are increased since more bandwidth is occupied by higher priority traffic and the kind of reserved.

5. Conclusion

It is an interesting issue to study integrated QoS signaling supporting multi-services in the heterogeneous networks. In this paper, we propose an integrated QoS signaling architecture and estimated handoff delay time for different mobility scenarios. In addition, an adaptive resource allocation strategy based variation called VARR is proposed to improve system performance. Simulation results demonstrate the proposed system performance and in comparison with a fixed bandwidth regulation method. Furthermore, the effect on system performance due to mobility was evaluated and compared. The simulation results show that VARR clearly improves the system performance in supporting multi-service heterogeneous networks.

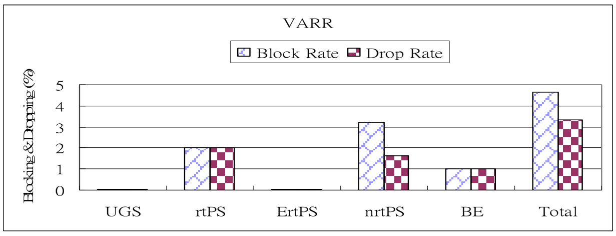


Figure 7: Blocking and dropping rates for VARR

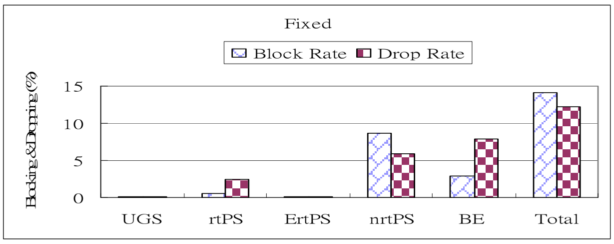


Figure 8: Blocking and Dropping rates for fixed bandwidth regulation method

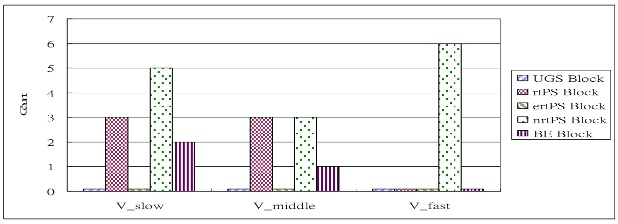


Figure 9: Blocking for different velocity

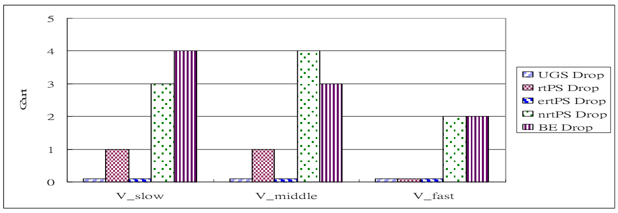


Figure 10: Dropping for different velocity

REFERENCES

[1] 3GPP TS 22.340, IP Multimedia System (IMS) messaging; Stage 1.

[2] 3GPP TS 23.228, IP Multimedia Subsystem (IMS); stage 2.

[3] 3GPP TS 24.229, IP multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3.

[4] A. K. Talukdar, B. R. Badrinath, and A. Acharya, "MRSVP - A resource reservation protocol for an integrated service network with mobile hosts," Wireless Networks, Volume 7, Number 1, pp. 5-19, January 2001.

[5] Chien-Chao Tseng1, et al., "HMRSVP – A Hierarchical Mobile RSVP Protocol," Wireless Networks, Volume 9, Number 2, pp. 95-102, March 2003.

[6] Wooseong Kim and Hongseok Jeon, "Advanced Resource Reservation using PAR-SIP for Real-time Multimedia Communication," IEEE VTC 2006-Spring, Volume 1, pp. 324 - 328, 2006.

[7] Yun-ji Ban, et al., "Network Configuration for QoS in MIP System," IEEE ICACT, Volume 2, pp. 1007 - 1012, Feb. 2007.

[8] C. Politis, et al., "Hybrid multilayer mobility management with AAA context transfer capabilities for all-IP networks," IEEE Wireless Communications, Vol.11, Issue 4, pp. 76-88, August 2004.

[9] S. Sargento, R. Valadas, J. Goncalves, and H. Sousa, " IP-based access networks for broadband multimedia services," IEEE Communications Magazine, Volume 41, Issue 2, pp. 146 - 154, Feb. 2003.

[10] Shen sheng Tang and Wei Li, "An adaptive bandwidth allocation scheme with preemptive priority for integrated voice/data mobile networks," IEEE Transactions on Wireless Communications, Volume 5, issue 10, pp. 2874 - 2886, Oct. 2006.

[11] D. Niyato and E. Hossain, "Call admission control for QoS provisioning in 4G wireless networks: issues and approaches," IEEE Network, Volume 19, issue 5, pp. 5 - 11, Sept.-Oct. 2005.

[12] J. Vazquez-Avila, F.A.C. Cruz-Perez and L. Ortigoza-Guerrero, "Performance analysis of fractional guard channel policies in mobile cellular networks," IEEE Transactions on Wireless Communications, Volume 5, issue 2, pp. 301 - 305, Feb. 2006.

[13] David Tung Chong Wong, Jon W. Mark and Kee Chaing Chua, "Two-Level Fractional Guard Channels for Priority Access in Cellular Systems," IEEE VTC 2006-Spring, Volume 1, 2006, pp. 383 – 387.

[14] Ariton E. Xhafa and Ozan K. Tonguz, "Handover Performance of Priority Schemes in Cellular Networks," IEEE Transactions on Vehicular Technology, Volume pp, issue 99, 2007. Accepted for future publication.

[15] B.B. Madan, S Dharmaraja and K.S. Trivedi, "Combined Guard Channel and Mobile Assisted handoff for Cellular Networks," IEEE Transactions on Vehicular Technology, Volume pp, issue 99, 2007. Accepted for future publication.

[16] Muhammed Salamah, "An Adaptive Multi-Guard Channel Scheme for Multi-Class Traffic in Cellular Networks," IEEE ICCSA, March 8, 2006, pp. 716 – 723.

[17] The Network Simulator - ns-2 http://www.isi.edu/nsnam/ns/

[18] WiMAX Forum, Mobile WiMAX Part I:A Technical Overview and Performance Evaluation August, 2006

[19] Tarhini, C.and Chahed, T., "AMC-Aware QoS Proposal for OFDMA-Based IEEE802.16 WiMAX Systems", IEEE GLOBLECOM'07, November 26,2007, pp. 4780-4784