Traffic-Variation-Aware Connection Admission Control Mechanism for Polling Services in IEEE 802.16 Systems

Jenhui Chen and Chiang-Wei Chang
Department of Computer Science and Information Engineering,
Chang Gung University, Kweishan, Taoyuan, Taiwan
Email: jhchen@mail.cgu.edu.tw

Abstract— **The bandwidth of the broadband wireless access is managed by a centralized base station (BS), which is achieved by a connection admission control (CAC) mechanism and a scheduler dynamically managing hundreds or thousands of ongoing connections persistently. The quality-of-service (QoS) services are also provided in the system for supporting more and more important multimedia services in wireless communications. There is a thorny problem, among these QoS types, of real-time polling service (rtPS) management when deciding the admission of new coming or scheduling ongoing connections due to their indeterminate characteristics of bandwidth requirement, e.g., variable bit rate (VBR) and delay constraint. This paper uses the characteristics of traffic variations as a parameter to construct the CAC mechanism for managing the real-time VBR services. Simulation results show that the proposed CAC can low down the call blocking probability of both rtPS and nrtPS traffics compared with the results of IEEE 802.16 standard. It not only decreases the number of dropped serving PDUs but also increases the system utilization.**

I. Introduction

The broadband wireless access (BWA) technology has been receiving much more attentions recently due to its high bandwidth access [2], [4], [5]. This technology enables more and more multimedia applications such as real-time interactive conference and digital audio/video broadcasting (AVB) applied in the wireless environment. These multimedia applications all have the properties, e.g., VBR, time sensitivity, and delay constraint, etc, and needs an efficient CAC and scheduling mechanism to manage these data flows.

In order to serve multimedia applications, the IEEE 802.16 standard [1] has been proposed to provide QoS capacity in BWA networks. These QoS types include the serving of constant bit rate (CBR) connection as the unsolicited grant service (UGS), the real-time VBR connection as the real-time polling service (rtPS), the non-real-time VBR connection as the non-real-time polling service (nrtPS), and the best effort service (BE). Although IEEE 802.16 constructs a platform for the use of BWA, it is still lack of the parts of bandwidth allocation and CAC polices to manage the radio resource.

Many solutions are investigated in literature to make up for these shortcomings [6], [7], [8], [9]. In [6], the authors propose a QoS packet scheduling scheme for different types of services in 802.16 BWA networks. The paper [7] presents a resource allocation strategy called enhanced staggered resource allocation (ESRA) to maximize the number of concurrent transmissions so that the throughput can be maximized. However, this work does not consider the properties of channel conditions, which will impact on the size of queueing buffer, to adjust the amount of bandwidth allocation to each connection. The work [8] takes the viewpoint of controlling the admission for the bandwidth asymmetry in uplink and downlink to maximize the network resource utilization. However, the real-timer service flows are not considered in this work and will lead to QoS violation of real-time services. In [9], the authors present a queueing-based model for dynamic resource allocation and CAC mechanism, which considers both connection-level and packet-level QoS to overcome the mentioned drawback. Nevertheless, all of these works do not consider the property of bandwidth variation in real-time services, which will influence the QoS.

In this paper, we propose a traffic-variation-aware connection admission control (TCAC) mechanism for real-time services to decrease the connection blocking probability as well as the packet dropping probability due to its delay constraint. This mechanism is achieved by considering the characteristics of rtPS and nrtPS, e.g., minimum reserved traffic rate, maximum sustained traffic rate, and maximum latency, etc. A Markov queueing-based model is also presented to analyze the packet dropping probability for dynamic bandwidth allocation. Under the proposed scheme, the amount of allocated bandwidth to the PS class can be adjusted dynamically according to the variations in traffic load and/or channel quality so that the QoS performances such as packet delay and packet dropping probability can be maintained.

The rest of this paper is organized as follows. In Section II, the system model and assumptions are described; moreover some notations are defined. In Section III, we proposed a traffic-variation-aware connection admission control scheme, and the analytical model for the proposed scheme is also presented. Section IV presents the performance evaluation results. Conclusions are stated in Section V.

II. System Assumptions and Definitions

A. User Satisfaction Degree

The user satisfaction is assessed by taking the utility functions described in [9], which represents the level of satisfaction on the perceived QoS for different service types. The utility of each connection i counts on on the amount of allocated bandwidth, delay degree, throughput for the UGS, rtPS, nrtPS, and BE connections, respectively. In other words, the higher value of the utility for connection i is, the more satisfied the

1-4244-1005-3/07/$25.00 ©2007 IEEE.

connection is. The utility functions represented for UGS and BE connections are defined as

$$U_{\mathrm{UGS}}(b_i) = \begin{cases} 1, & \text{if } b_i \geq b_{\mathrm{UGS}}^{(req)} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

and

$$U_{\mathrm{BE}}(b_i) = \begin{cases} 1, & \text{if } b_i \geq 1 \\ 0, & \text{otherwise} \end{cases}, \quad (2)$$

where the amount of allocated bandwidth b_i for connection i is higher than or equal to the required bandwidth $b_{\mathrm{UGS}}^{(req)}$, the utility for a UGS connection is the one, i.e., highest. Similarly, when the BE connection is admitted into the network, the utility for a BE connection is one. For rtPS and nrtPS connections, the modified sigmoid function is used to obtain utility as a function of the packet-level performance measures. Both in uplink and downlink the utility for rtPS and nrtPS connections can be represented as functions of the allocated bandwidth as follows:

$$U_{\mathrm{rtPS}}(b_i) = 1 - \frac{1}{1 + e^{-g_{rt}\cdot(d(\bar{\gamma},\lambda,b_i) - d_i^{req} - h_{rt})}} \quad (3)$$

and

$$U_{\mathrm{nrtPS}}(b_i) = \frac{1}{1 + e^{-g_{nrt}\cdot(\tau(\bar{\gamma},\lambda,b_i) - \tau_i^{req} - h_{nrt})}}, \quad (4)$$

where $d(\bar{\gamma}, \lambda, b_i)$ and $\tau(\bar{\gamma}, \lambda, b_i)$ imply the average delay and the transmission rate of connection i with the PDU arrival rate λ and the average SNR value $\bar{\gamma}$ when the amount of allocated bandwidth is b. In addition, g_{rt}, g_{nrt}, h_{rt}, and h_{nrt} are the parameters of sigmoid function. The g_{rt} and g_{nrt} determine the sensitivity of the utility function to delay or throughput requirement. The h_{rt} and h_{nrt} represent the center of the utility function.

B. Standard Deviation Estimation for Polling Service

In general, most of strategies of bandwidth management are partitioning the bandwidth into four parts. The size of each portion of bandwidth is not equal. Usually, the portion of UGS is bigger than PS and the portion of PS is bigger than BE. Every part of bandwidth has a quota determined by BS in advance. Because the BE and UGS is CBR, thus the effect of UGS and BE on CAC and bandwidth management can be neglected here.

Now, let us consider more about the features of PS. If the BS admits fewer PS connections, the QoS-guarantee of each connections can be ensured. However, the remainder available bandwidth for PS are totally wasted. On the contrary, it will result in QoS degradation that user cannot tolerate if the BS admits too many PS connections unconditionally. This is because the characteristic of the VBR of PS. In other words, it may happen that the BS admits too many or deficient connections if the CAC is not well designed.

Consequently, in order to ensure the reliability of QoS-guarantee, we propose a standard deviation estimation for PS scheme. First, we assume that polling service arrival process which has its own maximum sustained traffic rate $r_{\max}$ and minimum reserved traffic rate $r_{\min}$ follows a Modulated Markov Poisson Process (MMPP) model. MMPP is a n-state Markov chain that has different rates in each state and with more parameters compared with Poisson distribution. Therefore, it is able to capture diverse characteristics of the connection calls. In other words, MMPP model is more general than a traditional poisson model and is able to capture burstiness in traffic arrival process. The state space is given as follows

$$S = \{1, 2, ..., n\}. \quad (5)$$

Each of the states denotes an unique value which is subtracting $r_{\min}$ from $r_{\max}$. Note that S is system view variation of all connections. The transition probability matrix $\mathbb{T}$ and Poisson arrival rate matrix can be expressed as follows, respectively,

$$\mathbb{T} = \begin{bmatrix} t_{1,1} & \cdots & t_{1,n} \\ t_{2,1} & \cdots & t_{2,n} \\ \vdots & \ddots & \vdots \\ t_{n,1} & \cdots & t_{n,n} \end{bmatrix} \text{ and } \Lambda = \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{bmatrix}. \quad (6)$$

Each diagonal element of Λ is given by,

$$P_x(\alpha_s) = \frac{e^{-\alpha_s T}(\alpha_s T)^x}{x!}, \; s \in S, \; \alpha_s = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}, \quad (7)$$

where $P_x(\alpha_s)$ is the probability that x Poisson events occur during the time interval T, i.e. frame length with the mean rate α_s. The average system deviation of PS is calculated by

$$\bar{d} = \sum_{s=1}^{n} s \times \pi(s), \quad (8)$$

where $\pi(s)$ is the steady state probability of Markov chain.

Thus, the standard deviation estimation function is obtained as follows

$$\sigma = \sqrt{\sum_i (c_i - \bar{d})^2 P_C(c_i)}, \quad (9)$$

which C is the set of all PS connections and c_i denote the difference of $r_{\max}$ and $r_{\min}$ for the i-th connection.

III. Traffic-Variation-Aware Connection Admission Control

The purpose of CAC is to limit the number of connections in the network if all of the bandwidth has been used up to provide the highest QoS to existing users. Thus, a resource reservation with QoS constraints is considered, suppose that there are $\mathbb{B}$ bandwidth in a frame time, and we set four service types UGS, rtPS, nrtPS, and BE with different quotas such that

$$B_{\mathrm{UGS}} + B_{\mathrm{rtPS}} + B_{\mathrm{nrtPS}} + B_{\mathrm{BE}} = \mathbb{B}. \quad (10)$$

In order to obtain the optimal setting values of these four quotas for system, i.e. B_x, $x \in \{\mathrm{UGS}, \mathrm{rtPS}, \mathrm{nrtPS}, \mathrm{BE}\}$. Let the average allocated bandwidth per connection is $\overline{b_x(t)}$ in both uplink and downlink such that

$$\overline{b_x(t)} = \frac{b_1(t) + b_2(t) + \cdots + b_k(t)}{k} = \frac{\sum_{i=1}^{k} b_i(t)}{k}, \quad (11)$$

which the $b_1(t), b_2(t), \ldots, b_k(t)$ is the allocated bandwidth for each connections obtained from history statistics recorded every t interval by BS. Also let connection arrival rate μ, and connection holding time $1/\nu$ for particular service type. Then, we model these four service types with Marcov chain to estimate the connection-level performances such as the connection blocking probability, the average number of connections, and the connection utility.

In addition, the maximum number of ongoing connections which system can tolerate under a certain bandwidth B_x is defined as follows,

$$N_x(t) = \begin{cases} \left\lfloor \frac{B_x}{b_x(t)} \right\rfloor, & \text{when } x \in \{\text{UGS}, \text{BE}\}, \\ \left\lfloor \frac{B_x}{b_x(t) - \sigma} \right\rfloor, & \text{when } x \in \{\text{rtPS}, \text{nrtPS}\}. \end{cases} \tag{12}$$

Furthermore, the state space of the amount of connections is $\mathbb{C} = \{\ i \mid 0 \le i \le N_x(t)\ \}$, where i denote the number of ongoing connections for a specific service type. Also, the transition matrix of Markov chain is obtained as follows,

$$\mathbb{P} = \begin{bmatrix} -\mu & \mu & & & & \\ \nu & -\nu-\mu & \mu & & & \\ & \ddots & \ddots & \ddots & & \\ & & i\nu & -i\nu-\mu & \mu & \\ & & & \ddots & \ddots & \ddots \\ & & & & N_x(t)\nu & -N_x(t)\nu \end{bmatrix}. \tag{13}$$

The steady state probability π_c of Markov is calculated by solving $\pi_c \mathbb{P} = \mathbf{0}$ and $\pi_c \mathbf{1} = 0$, where $\mathbf{0}$ and $\mathbf{1}$ are row marix of zero and column matrix of one. Besides, the steady state space is given by,

$$\pi_c = [\pi_c(0), \pi_c(1), \ldots, \pi_c(i), \ldots, \pi_c(N_x(t))], \tag{14}$$

where π_i denotes the steady state probability that there are total i ongoing connections.

From (8) to (12), the average number of connections obtained as follows,

$$\overline{C} = \sum_{i=0}^{N_x(t)} i \times \pi_c(i). \tag{15}$$

Similarly, the connection block probability is computed as follows,

$$P_{block} = \pi_c(N_x(t)). \tag{16}$$

Once these four quotas B_{UGS}, B_{rtPS}, B_{nrtPS}, and B_{BE} are determined by administrator, the CAC could be implemented in two-phase algorithm mentioned in Fig. 1 and Fig. 2. Note that the two-phase CAC algorithm should be invoked frame by frame. Phase-1 is for admission of new connections, and phase-2 is used to provide QoS for ongoing connections. The two-phase CAC algorithm is described in detail as follows.

PHASE ONE - THE PSEUDOCODE OF CONNECTION ADMISSION CONTROL

```
tmp_u = tmp_rtPS = tmp_nrtPS = tmp_BE = 0;
for( i ∈ C_UGS){
  if(tmp_u + b_UGS^(up,req) + b_UGS^(do,req) ≧ B_UGS)
      continue;      // reject this connection.
  else{  b_i^up = b_UGS^(up,req);   b_i^do = b_UGS^(do,req);   tmp_u = tmp_u + b_i^up + b_i^do; }
}
for( i ∈ C_rtPS){
  if((min_b(d(γ̄_i,λ_i^up,b))≤d_i^(up,req) & min_b(d(γ̄_i,λ_i^do,b))≤d_i^(do,req))==true){
      if((tmp_rt+min_b(d(γ̄_i,λ_i^up,b))+(min_b(d(γ̄_i,λ_i^do,b))) ≧ B_rtPS)
          continue;      // reject this connection.
      else{
          if (σ(c_i, σ) ≧ T_σ)     continue;   // reject this connection.
          else{  b_i^up = min_b(d(γ̄_i, λ_i^up, b));
                 b_i^do = min_b(d(γ̄_i, λ_i^do, b));
                 tmp_rt = tmp_rt + b_i^do + b_i^up;
          }
      }
  }
  else    continue;      // reject this connection.
}
for( i ∈ C_nrtPS){
  if((min_b(τ(γ̄_i,λ_i^up,b))≥τ_i^(up,req) & min_b(τ(γ̄_i,λ_i^do,b))≥τ_i^(do,req))==true){
      if(tmp_nrt + (min_b(τ(γ̄_i,λ_i^up,b))+min_b(τ(γ̄_i,λ_i^do,b)) ≥ B_nrtPS)
          continue;      // reject this connection.
      else{  b_i^up = min_b(r(γ̄_i, λ_i^up, b));
             b_i^do = min_b(r(γ̄_i, λ_i^do, b));
             tmp_nrt = tmp_nrt + b_i^do + b_i^up;
      }
  }
  else    continue;      // reject this connection.
}
for( i ∈ C_BE){
   if(tmp_BE + 1 + 1 ≥ B_BE)    continue;
   else{  b_i^up = 1;   b_i^do = 1;   tmp_BE = tmp_BE + b_i^do + b_i^up; }
}
```

Fig. 1. Admitting new connections.

A. Admission of New Connections

The first phase is a initial state of CAC and it is also for CAC to judge whether a new connection can be admitted or not. Note that each of new connections has its own service type, and each of service types has the different quotas constraint defined by BS in advance. For UGS connections, once the B_{UGS} is exhausted by ongoing connections, the CAC will reject new UGS connections and progress to the step of rtPS condition. Considering about rtPS connections, because rtPS is real-time variable bit rate, the rtPS connections will have delay constraint and min-max traffic rate constraint. Therefore, the BS has to satisfy SSs with the delay condition first. Then, BS is going to decide which and how many connections can be admitted according to the deviation of bandwidth variation caused by rtPS. This behavior not only warrant the QoS level of connection but also ensure the connection blocking probability being certain value, that is to say, the connection blocking probability can be maintained below specified threshold and BS still can provide superior service to SSs. If connections drained of the B_{rtPS}, CAC will decline a new connection likewise. For nrtPS connections, the blocking situation is the same when B_{nrtPS} is consumed by ongoing connections. The BE connections are treated alike. In other words, once BE connections spend out the B_{BE}, the new

PHASE TWO - THE PSEUDOCODE OF CONNECTION ADMISSION CONTROL

for($i \in \mathbb{C}_{\text{UGS}}$){
    $b_i^{up} = b_{\text{UGS}}^{(up,req)}$; $b_i^{do} = b_{\text{UGS}}^{(do,req)}$;
}
for($i \in \mathbb{C}_{\text{rtPS}}$){
    $b_i^{up} = min_b(d(\bar{\gamma}_i, \lambda_i^{up}, b)) \leq d_i^{(up,req)}$;
    $b_i^{do} = min_b(d(\bar{\gamma}_i, \lambda_i^{do}, b)) \leq d_i^{(do,req)}$;
}
for($i \in \mathbb{C}_{\text{nrtPS}}$){
    $b_i^{up} = min_b(\tau(\bar{\gamma}_i, \lambda_i^{up}, b)) \geq d_i^{(up,req)}$;
    $b_i^{do} = min_b(\tau(\bar{\gamma}_i, \lambda_i^{do}, b)) \geq d_i^{(do,req)}$;
}
for($i \in \mathbb{C}_{\text{BE}}$){
    $b_i^{up} = 1$; $b_i^{do} = 1$;
}
$\mathbb{C}_{allo} = \mathbb{C}_{\text{UGS}} \cup \mathbb{C}_{\text{rtPS}} \cup \mathbb{C}_{\text{nrtPS}} \cup \mathbb{C}_{\text{BE}}$
while($(\sum_i b_i \leq \mathbb{B})\&(b_i < b_{\max})\&(\mathbb{C}_{allo}! = \emptyset)$){
    $j = \arg\min_i(U(b_i))$;
    $b_j = b_j + 1$;
    if($b_j == b_{\max}$) $\mathbb{C}_{allo} = \mathbb{C}_{allo} - j$;
    if($\sum_{i \in \mathbb{C}_{\text{UGS}}} b_i == B_{\text{UGS}}$) $\mathbb{C}_{allo} = \mathbb{C}_{allo} - \mathbb{C}_{UGS}$;
    if($\sum_{i \in \mathbb{C}_{rtPS}} b_i == B_{\text{rtPS}}$) $\mathbb{C}_{allo} = \mathbb{C}_{allo} - \mathbb{C}_{\text{rtPS}}$;
    if($\sum_{i \in \mathbb{C}_{nrtPS}} b_i == B_{\text{nrtPS}}$) $\mathbb{C}_{allo} = \mathbb{C}_{allo} - \mathbb{C}_{\text{nrtPS}}$;
    if($\sum_{i \in \mathbb{C}_{BE}} b_i == B_{\text{BE}}$) $\mathbb{C}_{allo} = \mathbb{C}_{allo} - \mathbb{C}_{\text{BE}}$;
}
return Phase-1;   // accept new connections.

Fig. 2. Providing the highest QoS to existing users.

BE connection will also be blocked.

B. Supply of Ongoing Connections

After determining which and how many connections to be served, the second phase CAC will first calculate the allocated bandwidth with the satisfied QoS requirement, i.e. delay, minimum bandwidth, and transmission rate, for each ongoing connection and obtain the remaining available bandwidth of $\mathbb{B}$. Once each connection's minimum requirement is reached, the CAC will search for the connection with lowest utility (i.e. the satisfaction of a certain connection is smallest) and take the advantage of remaining bandwidth $\mathbb{B}$ to improve the satisfaction of a connection which has lowest utility. The two-phase CAC is invoked frame by frame and finishes while either all connections approach to their maximum bandwidth requirement or all available bandwidth $\mathbb{B}$ has exhausted.

C. Connection Admission Control Measure

As mentioned in previous section, we can obtain some CAC metrics to help us whether a connection admitted or not.

Traffic variation: This refers the variation of allocated bandwidth. While the η is getting smaller, the effect causing the QoS violation is getting larger.

$$\eta = \frac{\mathbb{B}}{\sigma} \tag{17}$$

Admitted connections: This denotes that the number of connections can be admitted per interval.

$$\mu(1 - P_{block}) \tag{18}$$

IV. NUMERICAL AND SIMULATION RESULTS

A. Parameter Setting

The simulation environment is set one serving BS and multiple SSs. The frame length is 5 ms and the number of available bandwidth which can be allocated in one frame is 200 units per frame. In other words, there are 200 units per frame used to transmit PDUs. The length of MAC PDU is set fixed 100 bits. And the transmission bandwidth is 25 MHz. Moreover, the size of the queue is 100 PDUs. And we use the ns-2.29.3 with WiMAX module [3] to be our simulation tool.

The parameters g_{rt}, g_{nrt}, h_{rt}, and h_{nrt} of the modified sigmoid utility function are set 2, 2, 0, and 0, respectively. The $b_{\text{UGS}}^{(up,req)}$ and $b_{\text{UGS}}^{(do,req)}$ are set 2 unit per frame. The delay constraints $d_i^{(up,req)}$ and $d_i^{(do,req)}$ are fixed 5 frames. The throughput constraints $\tau_i^{(up,req)}$ and $\tau_i^{(do,req)}$ are defined 15 PDUs per frame. In other words, the throughput constraint is 3000 PDUs per second. All connections request are 5 PDUs per frame at least and 150 PDUs per frame at most in addition. Furthermore, the unit of allocated bandwidth for these connections depends on the modulation and coding rate scheme. Generally, the minimum amount of bandwidth ($b_{\min}$) and maximum amount of bandwidth ($b_{\max}$) allocated per connection is 1 and 10 units, respectively.

The traffic source we used for polling service is the two state MMPP, i.e. n=2. Besides, the PDU arrival process of each polling service connections is the same. Each of connections randomly has different $r_{\min}$ and $r_{\max}$, and the range from $r_{\min}$ to $r_{\max}$ per frame is set from 1 to 150 PDUs.

$$\mathbb{T} = \begin{bmatrix} 0.1 & 0.9 \\ 0.2 & 0.8 \end{bmatrix}, \Lambda = \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix}. \tag{19}$$

The PDU arrival process for UGS connections follows ON-OFF state source model. In addition, the connection arrival time $1/\mu$ and connection holding time $1/\nu$ are assumed to be exponentially distributed. We specify the average holding time for UGS, rtPS, nrtPS, BE to be 10, 15, 20, and 25 minutes, respectively. And we adjust the connection arrival rate to see the connection blocking probability and system Utility under different traffic load scenarios. And, we set $B_\sigma = 20$, $B_{\text{UGS}} = 25$, $B_{\text{rtPS}} = 40$, $B_{\text{nrtPS}} = 25$, and $B_{\text{BE}} = 10$.

B. Numerical and Simulation Results

These results we proposed are comparison with [9]. Fig. 3 shows the connection blocking probability under different traffic load. The blocking probability grows up when connection arrival rate is getting larger. Besides, the blocking probability of TCAC is minor improved.

Moreover, the number of average connections is shown in Fig. 4 and it demonstrates that our scheme accepts more connections than Iterative approach.

The QoS degradation of polling service is shown in Fig. 5 and the connection arrival rate is set as 1. The discontent of delay constraint is the main factor that cause the degradation of rtPS connection. We can investigate that the QoS degradation varies with simulation time. The TCAC approach record the

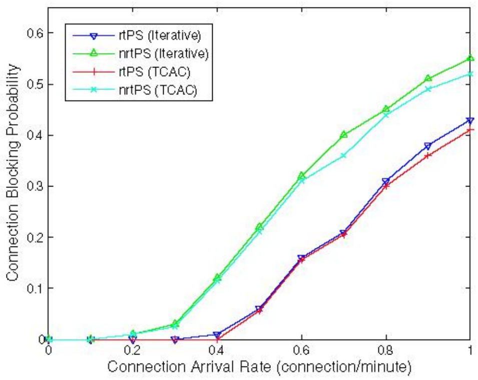

Fig. 3. Connection blocking probability v.s. Connection arrival rate.

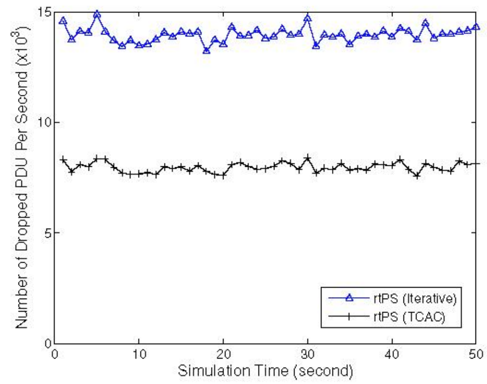

Fig. 5. The simulation time v.s. QoS-violation.

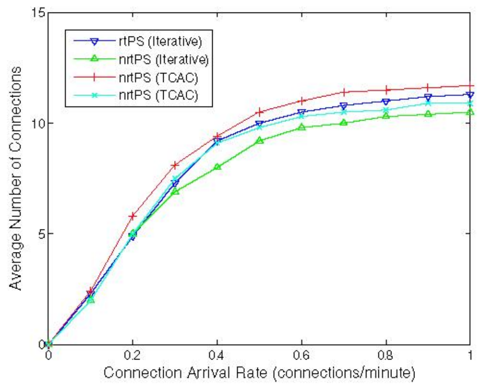

Fig. 4. Number of Connections v.s. Connection arrival rate.

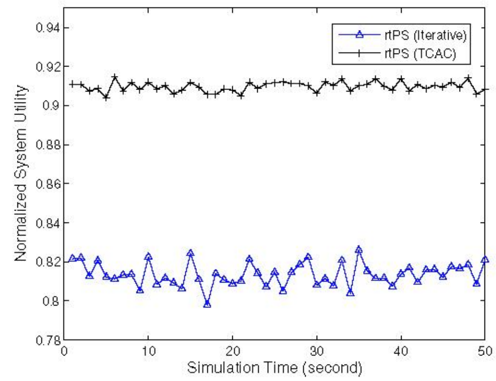

Fig. 6. System utility v.s. Simulation Time.

variation of each traffic rate so that the CAC will not admit connections whose deviation between $r_{\min}$ and $r_{\max}$ is too large when system deviation, i.e. ongoing connections, is huge. Therefore, the QoS degradation we proposed is less than Iterative approach. Also, the dropped PDUs we proposed are more stable than Iterative one.

In Fig. 6, we see the differential in system utility between Iterative approach and TCAC method. It is obvious that the performance we proposed is better than Iterative approach. Because we take the consideration of traffic variation of polling service connections.

V. Conclusion

In this paper we proposed a TCAC mechanism in IEEE 802.16 broadband systems. The TCAC mechanism is more outstanding than previous studies because we take a consideration of deviation of traffic rate of rtPS service type. Moreover, the QoS can be maintain at target level and the computational complexity of our algorithm is only $O(connections)$. It is also possible that binds the CAC and bandwidth allocation in the future.

Acknowledgment

This work was supported by the National Science Council, Taiwan, R.O.C., under Contract NSC95-2221-E-182-023.

References

[1] IEEE 802.16 Working Group, "IEEE Standard for Local and Metropolitan Area Networks–Part 16: Air Interface for Fixed Broadband Wireless Access Systems," IEEE Std. 802.16-2004, Oct. 2004.

[2] B. Fong, N. Ansari, A.C.M. Fong, G.Y. Hong, and P.B. Rapajic, "On the Scalability of Fixed Broadband Wireless Access Network Deployment," *IEEE Commun. Mag.*, vol. 42, no. 9, pp. 12–21, Sept. 2004.

[3] J. Chen, C.C. Wang, F.C. Tsai, C.W. Chang, S.S. Liu, J. Guo, W.J. Lien, J.H. Sum, and C.H. Hung, "The Design and Implementation of WiMAX Module for ns–2 Simulator," *in Proc. ACM VALUETOOLS*, article 5, Pisa, Italy, Oct. 2006.

[4] J. Chen and W.K. Tan, "Predictive Dynamic Channel Allocation Scheme for Improving Power Saving and Mobility in BWA Networks," *ACM/Springer MONET*, vol. 12, no. 1, pp. 15–30, Feb. 2007.

[5] I. Koffman and V. Roman, "Broadband Wireless Access Solutions Based on OFDM Access in IEEE 802.16," *IEEE Commun. Mag.*, vol. 40, no. 4, pp. 96–103, April 2002.

[6] K. Wongthavarawat and A. Ganz, "Packet Scheduling for QoS Support in IEEE 802.16 Broadband Wireless Access Systems," *Int. J. Commun. Syst.*, vol. 16, no. 1, pp. 81–96, Feb. 2003.

[7] K.K. Leung and A. Srivastava, "Dynamic Allocation of Downlink and Uplink Resource for Broadband Services in Fixed Wireless Neworks," *IEEE J. Select. Areas Commun.*, vol. 17, no. 5, pp. 990–1006, May 1999.

[8] X. Yang, G. Feng, and D. Kheong, "Call Admission Control for Multiservice Wireless Networks with Bandwidth Asymmetry Between Uplink and Downlink," *IEEE Trans. Veh. Tech.*, vol. 55, no. 1, pp. 360–368, Jan. 2006

[9] D. Niyato, and E. Hossain, "A Queuing-Theoretic and Optimization-Based Model for Radio Resource Management in IEEE 802.16 Broadband Wireless Networks," *IEEE Trans. Comput.*, vol. 55, no. 11, pp. 1473–1488, Nov. 2006