Adaptive Resource Allocation Mechanism for Broadband Mobile Network

Moo Wan Kim
Tokyo University of Information Sciences
Chiba, Japan
mwkim@rsch.tuis.ac.jp

Shintaro Uno
Aichi University of Technology
Gamagoori, Japan
Uno-shin@aut.ac.jp

***Abstract*—Mobile networks have been continuously developed from 3.5G to 3.9G/4G with high speed wireless technologies (i.e., broadband mobile network). So, mobile networks need to provide more sufficient QoS mechanism to provide enhanced user's satisfaction. In this paper, we propose a new adaptive resource allocation mechanism based on utility function borrowed from the field of microeconomics. Through the simulation and calculation evaluation, we show that adaptive resource allocation based on user preferences is effective.**

Keywords- QoS; utility function; mobile network;

I. INTRODUCTION

Currently mobile networks are being evolved from 3.5G to 3.9G/4G, which is broadband mobile network, integrated into IP core network based on IMS (IP Multimedia Sub-system). In this environment, it is essential to provide sufficient QoS (Quality of Service) mechanism which enables different services to provide enhanced user satisfaction. QoS for mobile networks is being studied mainly in 3GPP [1], but still deal with only class oriented mechanism. For example, QoS class is defined according to the service types, and mapping between QCI (QoS Class Identifier) and DSCP (DiffServ Code Point) is studied [2] [3]. So, it is required that QoS mechanism has to provide more flexible resource allocation mechanism based on each user's preferences.

In this paper, we propose a new adaptive resource allocation mechanism based on each user's preferences by using the utility function borrowed from the field of microeconomics. The utility function qualifies the value that a user perceives for all possible amount of resources allocated. The field of microeconomics addresses the issue of resource allocation when many users compete for a limited amount of resource. Microeconomics has been used to address problems in networks by several authors. Some authors use it to address pricing issues [4] [5] [6] [7], and others to the problem of resource allocation [8] [9].Utility functions have been used by other authors for resource allocation in networks [7] [10]. These papers develop distributed mechanisms for resource allocation assuming that the network does not have knowledge of the utility functions of users. Our work is different in that we assume that the network has knowledge of the utility functions of all users.

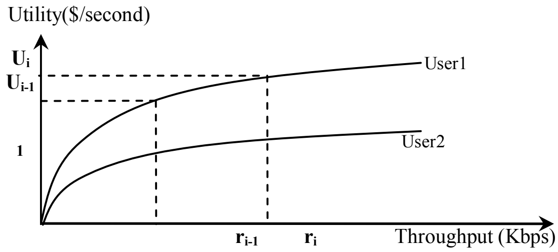

Figure 1. Example of Utility Function

II. BASIC TECHNOLOGIES

A. Utility Function

Utility is defined in the field of microeconomics that the level of satisfaction acquired from the consumption of properties such as services or commodities [11]. The user's total utility obtained from a network service will depend on several QoS metrics, such as throughput, delay, and jitter. In the throughput perspective, the user's utility depends on the bandwidth availability in the network to satisfy the resource requirement of service. Fig. 1 shows the example of utility function regarding the throughput allocated to user. User should define or select multiple utility functions for each service because the utility functions will be different from the kinds of services (e.g., streaming service).

Let us consider the network situation in which we have M users in the system. We let $U_i(r_i)$ denote the utility derived by user's flow i for a bandwidth allocation r_i and C is the total link capacity. User i is allocated $\underline{r}_i$ units of resource that is the i-th component of the solution $r=[r_1, r_2, \ldots, r_M]$ of the following optimization problem: [12]

$$\max_{[r1,r2,\ldots r_M]} \sum_{i=1}^{M} U_i(r_i) \qquad (1)$$

$$\text{subject to } \sum_{i=1}^{M} r_i \leq C$$

$$r_i \geq 0 \text{ for } i = 1,2, .., M$$

In this network situation, in order to adopt utility function to the network QoS management, the following requirements should be satisfied.

 ISBN: 978-1-61208-193-9

- Continuous utility functions should be represented in discrete functions in order to allocate the resource unit where the discrete segmental value of utility function is compared each other.
- Resource allocation should be done based on discrete utility function satisfying above (1)

Fig. 2 shows the example of discrete utility functions, which is the case that the two lines are differentiated by users. This value can be thought as the price that a user would be willing to pay to obtain a specific amount of resource. For example, in the upper line to obtain the throughput $r_{i-1,}$ a user will pay U_{i-1} ($/second) and to obtain the throughput $r_{i,}$ a user will pay U_i($/second). In Fig. 2, the slope of utility (ΔU) is more important since it means a unit price for a unit bandwidth. If a flow is a high priority flow, it has a higher ΔU than other flows.

Following, there are the properties of discrete utility functions:

- Non-negativity: $U(r) \geq 0$ for all $r \geq 0$. Obviously the users cannot associate a negative utility with a positive resource allocation.
- Non-decreasing nature: $U(r)$ has to be a non-decreasing function. Clearly also is the fact that users cannot associate a higher utility with a smaller allocation that with a higher allocation.

B. MSS Resouce Allocation Algorithm

To satisfy (1), the MSS (Maximum Segmental Slope) resource allocation algorithm [12] has been developed by our colleagues based on the standard optimization problem solving method [15]. The main characteristic of this algorithm is that it allows resource allocation maximizing the user's satisfaction by allocating a unit of resource firstly to the flow that has the highest segmental slope. But, this algorithm is rather straightforward and it should be updated to enhance the performance. For example, some heuristic sorting algorithm should be considered because the major portion of the algorithm is a sorting process.

.

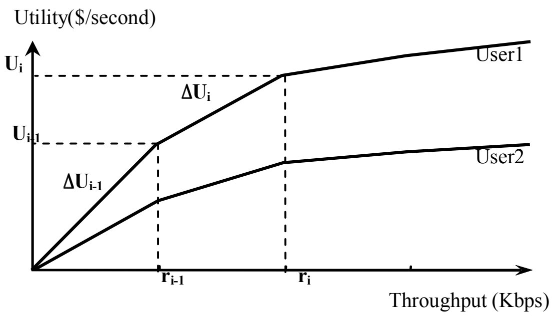

Figure 2. Example of Discrete Utility Function

III. ADAPTIVE RESOURCE ALLOCATION MECHANISM

A. Network Architecture Model

We assume the conceptual network architectural model as depicted in Fig. 3, based on 3.9G network (i.e., LTE) which comprises RAN(Radio Access Network) with MNs (Mobile Nodes), and CN (Core Network) with two main players: the QS (QoS Server) and the MM(Mobility Manager) [7]. The ARs (Access Router) in CN are key control points in the network. They are IP routers that are one IP hop distant from the mobile node via BS (Base Station, i.e., e-Node B in case of LTE). All data packets to and from the mobile node, and signaling messages between the mobile node and various servers in the network pass through the ARs. The CRs (Core Router) are high speed routers that lie in the core network. The main function of QS is an admission control based on the adaptive resource allocation mechanism proposed in the next section. The MM manages MN's location information where the MN is located in the mobile network, and performs handover based on the handover policies provided by the network operator. The MM interacts with QS during the call setup and termination.

B. Adaptive Resource Allocation Mechanism

There are already some studies in which the utility function is adopted to QoS [13] [14]. But, these studies do not deal with the detailed adaptation algorithm considered the flow (i.e., consider caller and callee side together) proposed in this paper. Under the assumption that network has knowledge of utility function for all users, we adopt MSS algorithm from caller to callee side together. Fig. 4 shows the proposed a new adaptive resource allocation mechanism and details of the mechanism are as follows.

- AR1: caller AR, AR2: callee AR
- fi: new flow's entry
- fi-AR1-re-flow-list: a list of existing flows in the AR1 to be re-allocated due to fi's joining

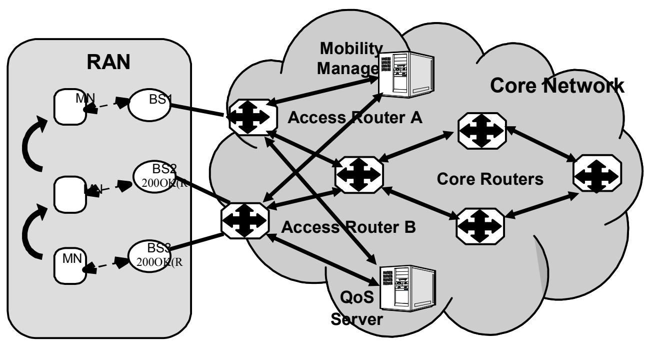

Figure 3. Architectual Model for Mobile Network

- fi-AR2-re-flow-list: a list of existing flows in the AR2 to be re-allocated due to fi's joining
- fi-AR1-alloc: An obtained temporary allocation to fi in AR1 by using the utility function based on MSS algorism
- fi-AR1-re-flow-list-temp-alloc: a list of obtained flows in the AR1 by using the utility function based on MSS algorism
- fi-AR2-alloc: an obtained temporary allocation to fi in AR2 by using the utility function based on MSS algorism.
- fi-AR1-re-flow-list-temp-alloc: a list of obtained flows in the AR2 by using the utility function based on MSS algorism
- fi-mid-alloc: minimum of [fi-AR1-alloc, fi-AR2-alloc]
- bw: bandwidth
- fi-AR1-re-flow-list-last-alloc: the last allocation for the existing flows in AR1, resulting in last-alloc-a, last-alloc-b, last-alloc-c,.. by re-calculating the bandwidth of the existing flows in fi-AR1-re-flow-list if there is remaining un-utilized bandwidth in AR1.
- fi-AR2-re-flow-list-last-alloc: the last allocation for the existing flows in AR2, resulting in last-alloc-x, last-alloc-y, last-alloc-z,.. by re-calculating the bandwidth of the existing flows in fi-AR2-re-flow-list if there is remaining un-utilized bandwidth in AR2.
- fi-AR1-remain-bw: the remained bandwidth in AR1
- fi-AR1-remain-bw: the remained bandwidth in AR2

a) Initial phase(1-4),

QS provides AR's flow list for the caller and callee side.

1). QS receives resource allocation request from caller or caller's AR called AR1.

2). QS generates this new flow's entry called $\boldsymbol{f_i}$ which has the information of its caller's access router ($\boldsymbol{f_i}$ – AR1), and callee's access router ($\boldsymbol{f_i}$ –AR2).

3)&4). From previous step 2, QS generates a list of the existing flows that need to be re-allocated due to $\boldsymbol{f_i}$'s joining. It is called $\boldsymbol{f_i}$-re-flow-list. This list is made of flows that are passing through the caller and callee ARs and that are affected by $\boldsymbol{f_i}$ joining. For example, $\boldsymbol{f_i}$-AR1-re-flow-list and $\boldsymbol{f_i}$-AR1-re-flow-list

b) The second phase(5-7),

Whenever new flow is added, for AR's existed flows for both of caller and callee side, temporary bandwidth for each existing flows and new flow are calculated according to the MSS algorithm. For the new flow, select the minimum value between caller's temporary bandwidth and callee's temporary bandwidth and set it as middle allocation value.

5). By using the utility function based on MSS algorithm, a temporary allocation is made to $\boldsymbol{f_i}$ and the

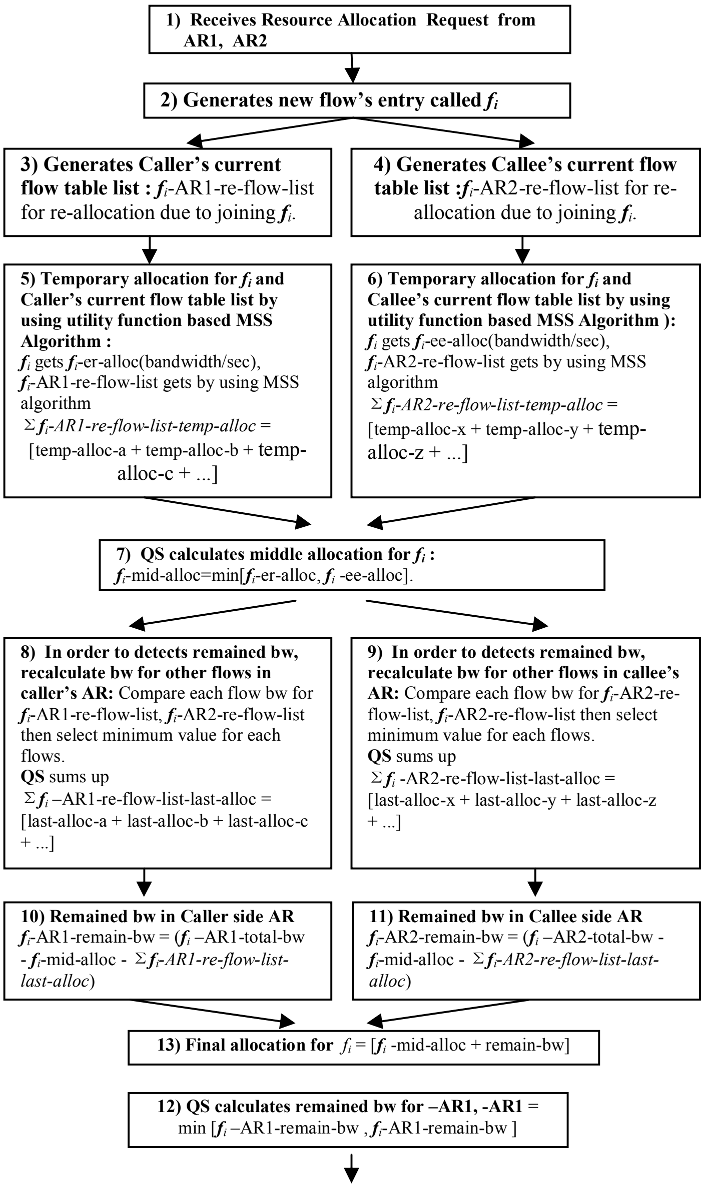

Figure 4. Adaptive Resource Allocation Mechanism

flows in the caller AR list. So, $\boldsymbol{f_i}$ gets $\boldsymbol{f_i}$-er-alloc (bandwidth/sec), and the flows in $\boldsymbol{f_i}$-AR1-re-flow-list gets $\Sigma\boldsymbol{f_i}$-*AR1-re-flow-list-temp-alloc* = [temp-alloc-a + temp-alloc-b + temp-alloc-c + ...]).

6). Same procedures are done for the flow $\boldsymbol{f_i}$ at the callee AR. Flow $\boldsymbol{f_i}$ gets $\boldsymbol{f_i}$-ee-alloc (bandwidth/sec), and the flows in the list for callee AR($\boldsymbol{f_i}$-AR2-re-flow-list) get $\Sigma\boldsymbol{f_i}$-*AR2-re-flow-list-temp-alloc* = [temp-alloc-x + temp- alloc-y + temp-alloc-z + ...] .

7). QS calculates middle allocation for $\boldsymbol{f_i}$, this value is called $\boldsymbol{f_i}$-mid-alloc which is the minimum of [$\boldsymbol{f_i}$-er-alloc, $\boldsymbol{f_i}$ -ee-alloc].

c) The third phase(8-11),

In order to detect remained bandwidth for both caller side AR and callee side AR, compare with the temporary value for each existed flow's in caller side AR and one of callee side, select minimum value and set as a last allocation for each existed flows, then sum up all as the last allocation for each caller side AR and callee side AR.

8). QS will detect if there is remaining un-utilized bandwidth in the ARs. This is done by re-calculating the bandwidth of the other flows in f_i-AR1-re-flow-list similarly to the process above. The calculation is done for the corresponding callee-AR of every other flow in the list. And again a minimum is selected for each flow. This is the last allocation for the other flows in the caller AR, resulting in last-alloc-a, last-alloc-b, last-alloc-c. Finally, QS sums up these values [last-alloc-a + last-alloc-b + last-alloc-c + ...], which can be expressed in Σf_i*-AR1-re-flow-list-last-alloc.*

9). In the similar manner, QS re-allocates each flow in f_i-AR2-re-flow-list and results in Σf_i*–AR2-re-flow-list-last-alloc* which is the sum of [last-alloc-x + last-alloc-y + last-alloc-z + ...].

10). Then the remaining bandwidth in caller side AR is f_i-AR1-remain-bw = (f_i –AR1-total-bw - f_i-mid-alloc – Σf_i*-AR1-re-flow-list-last-alloc*).

11). Similarly, the remaining bandwidth in the callee side AR is calculated in the way as above and results in f_i –AR2-remain-bw = (f_i-AR2-total-bw – f_i-mid-alloc – Σf_i*–AR2-re-flow-list-last-alloc*).

d) Final phase(12-13),

Compares the remained bandwidth for caller side AR with the one for callee side AR, then select minimum value and set as a network remained value. Then, allocate the middle allocation bandwidth and remained bandwidth to the new flow.

12). The remaining bandwidth for AR1 and AR2 for the flow f_i is remain-bw = min [f_i –AR1-remain-bw , f_i-AR2-remain-bw].

13). At last, as the result, the final allocation for f_i = [f_i -mid-alloc + remain-bw].

IV. EVALUATION

A. *Network Topokogy and Conditions*

In order to evaluate the feasibility of our proposal, we have developed the model system by using NS2 simulator with the topology shown in Fig. 5. We have implemented the proposed adaptive resource allocation mechanism. All the entities in the model system are implemented on the NS2. It is assumed that location registration is conducted before call setup. When a MN enters into an area covered by BS, MN sends Registration Request message to its AR. Then the MN's location is recorded in the Location server (i.e.MM). Each AR has 4 units of bandwidth to administrate (1 unit = 40 kbps) and the data flows use RTP packets. In

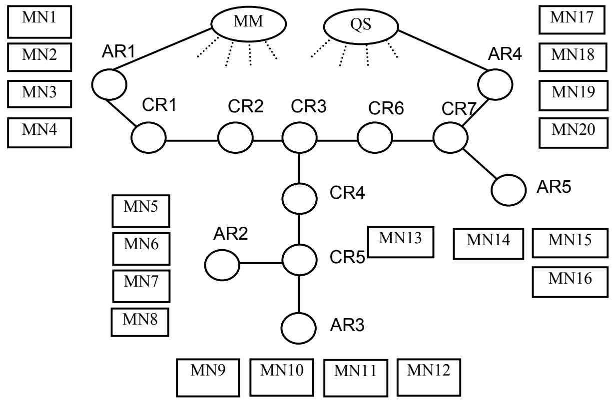

Figure 5. Simulated Network Model System

this topology, MN1, 2, 3 and 4 under AR1 have session with MN5, 9, 13, and 17 respectively. Each node request 4 units of bandwidth (1unit = 40 kbps) and each AR has 4 units of bandwidth to administrate (1 unit = 40 kbps). In detail, the bandwidth between AR and CR, AR and MN, are 160kbps (4units). The data flows are using RTP packet and are called Flow1, 2, 3 and 4. Packet size is 250Byte. In the simulation video phone service is assumed as the application.

B. *Simulation Results*

Fig. 6 shows the bandwidth allocation of each user's flow when a handover takes place. Utility value of Flow B, between MN2 and MN9, Flow S between MN6 and MN10, and Flow H (handover), between MN8 and MN14 are listed in Table I. Flow H starts its call at 30 sec and MN8 starts to move toward to new AR (AR3) at 40 sec. Before Flow H enters to AR3, there are two flows in AR3, our algorithm will therefore allocate 2 units of bandwidth to each flow (Flow H and Flow Sl). Once a new MN having an additional utility function joins to an AR, there is a need to re-allocate the existing resources among the users in the AR according to the utility function that the users have contracted. Hence, when Flow H enters the area of AR3, the allocation becomes Flow B : Flow H : Flow S = 2 : 1 : 1 units. The reason for this allocation is that Flow H has higher utility than Flow H, and therefore it is not affected by Flow H joining the AR. On the other side, Flow S has lower utility than Flow H and therefore one unit is re-allocated to Flow H.

TABLE 1. UTILITY VALUES FOR FLOWS

BW / Flow	U1	U2	U3	U4
Flow B	0.7	0.5	0.14	0.1
Flow S	0.45	0.3	0.14	0.1
Flow H	0.5	0.4	0.14	0.1

BW: allocated bandwidth

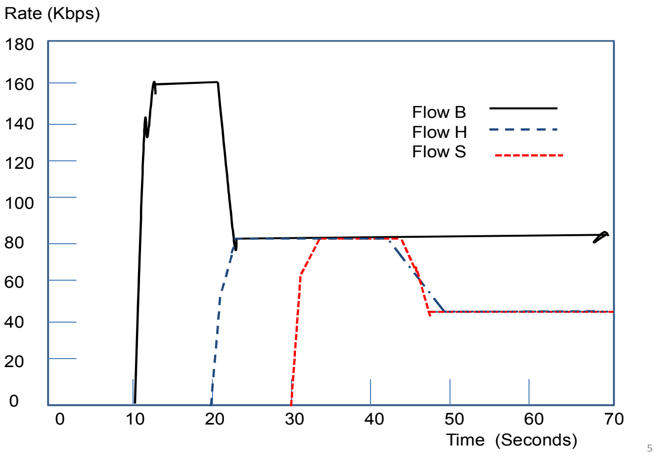

Figure 6. 3 Flows Evaluation Result with Handover

Fig. 7 shows another bandwidth allocation result for 4 flows. The utility value of Flow1 from MN1 to MN5, Flow2 from MN2 to MN9, Flow3 from MN3 to MN13, Flow4 from MN4 to MN17 are listed in Table2. For example, for Flow1, the utility values are 0.9 in U1 (0-40kbps), 0.3 in U2 (40-80bkps),0.15 in U3 (80-120kbps), and 0.1 in U4 (120-160kbps) respectively. Flow1 enters the network at time 10s. As it is the first flow in the system, it is allocated the entire AR1's bandwidth. At time 20s, Flow2 enters the network. Then at time 30s, with Flow3's join, the allocation becomes Flow1:Flow2:Flow3 = 2:1:1. Finally, after Flow4 enters, each flow has same bandwidth and this is equal to the allocation that expected. We should consider both the caller side and callee side. However, in this case the caller side is congested and the callee side is not congested at all in each flow so that it is enough that only the caller side is considered.

TABLE 2. UTILITY VALUES FOR FLOWS

BW / Flow	U1	U2	U3	U4
Flow1	0.9	0.3	0.15	0.1
Flow2	0.8	0.25	0.14	0.1
Flow3	0.7	0.2	0.2	0.1
Flow4	0.4	0.4	0.1	0.1

Based on Fig. 7, the bandwidth allocation can be explained as follows.

1) At time 10s there is only Flow 1 so that 4 units (160kbps) are allocated to Flow1.
2) At the time 20s, Flow2 enters the network. According to calculation based on utility values in Table 1, 2 units (80kbps) are allocated to Flow1 and Flow2 respectively.
3) At the time 30s, Flow3 enters the network. Then 2 units

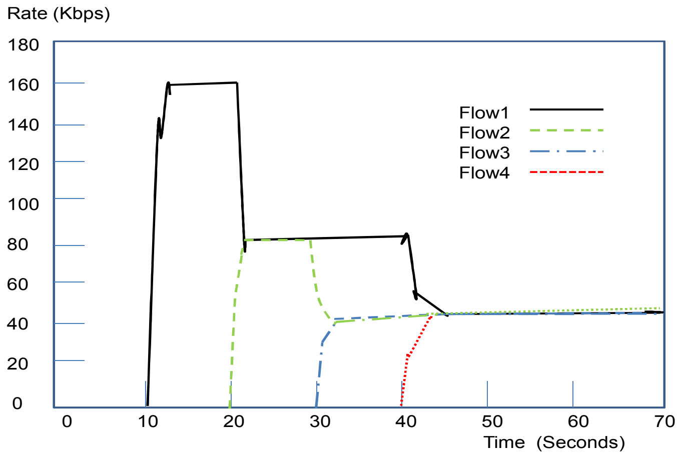

Figure 7. 4 Flows Evaluation Result

(80kbps) are allocated to Flow 1, and 1 unit (40kbps) is allocated to Flow 2 and Flow 3 respectively.
4) At the time 40s, Fkow4 enters the network. Then 1 unit(40kbps) are allocated to Flow1, Flow2, Flow3 and Flow4 respectively.

Fig. 8 shows the total bandwidth allocation result of the current method (i.e. static resource allocation method). The situation in Fig.7 is the same as Fig.6. That is, at time 10s, 20s, 30s and 40s, Flow1, Flow2, Flow3 and Flow4 enters the network via AR1 respectively. The allocated bandwidth is fixed (40Kbps) and always the same for all users. So the total bandwidth at AR1 is increased from 40kbps to 160kbps from 10s to 40s. On the other hand, the total bandwidth of the adaptive allocation method is constantly 160kbps.

We have calculated and compared the total user satisfaction in the case of Fig. 7 and Fig. 8. The calculation result is as follows,

1) Proposed adaptive allocation method (Fig. 7)
Flow1:0.9*10+1.2*10+1.2*10+1.45*10=47.5,
Flow2:1.05*10+0.8*20=26.5
Flow3:0.7*20=14, Flow4:0.4*10=4

- Total user satisfaction: 47.5+26.5+14+4=92

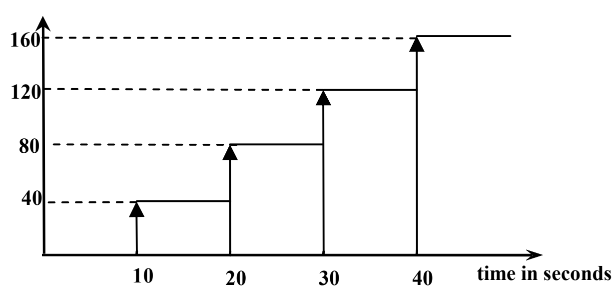

Figure 8. Total Bandwidth Allocation for Current Method

2) Current static resource allocation method (Fig. 8)
Flow1: 0.9*40=36, Flow2: 0.8*30=24,
Flow3: 0.7*20=14, Flow4: 0.4*10=4

- Total user satisfaction: 36+24+14+4=78

C. Considerations

Through the simulation and calculation we have confirmed the following facts.

1) First, we have confirmed that the allocation in the simulation is equal to the allocation calculated theoretically. Actually, the proposed algorithm has been simulated and it is confirmed that allocation in Fig. 6 and in Fig. 7 are equal to the allocation calculated theoretically.

2) Second, from the user's viewpoint, users can obtain the more satisfied service by this mechanism. Normally when the bandwidth is fully utilized (i.e., congested), the new service request is rejected. But by this mechanism, the new service can be prioritized even in the congested situation case.

3) Third, from the operator's viewpoint, the revenue of operators will be increased by using this mechanism. Actually by this mechanism, it is confirmed that the bandwidth can be utilized at maximum. Moreover, if service price is linked to the value of the utility function, the user will pay more payment so that the operators can obtain more revenue and profit.

V. CONCLUSIONS

In this paper, we have proposed a new adaptive resource allocation mechanism based on the utility function. As the next step, we can expand this mechanism on the following items.

1) In case that some of existing user's utility values are lower, and new users having higher utility value are joining, some of existing user's flow may be suddenly suspended based on our proposed mechanism. This is issue to be solved from the service quality viewpoint. However, even in this case, the existing flow can be continued with some minimum bandwidth if we modify MSS algorithm and the procedure described in Fig, 4.

2) In this paper, it is not clearly mentioned who will assign the utility function for each flow. About this we assume two cases; first case is that the mobile operators can define utility functions for flows and save the information in QS, second case is that mobile users can select a utility function for a flow and send this information to QS by using signaling. resources.

3) We have focused on only bandwidth allocation, but other QoS metrics such as the delay or jitter can be studied in the next step. Also, we have to consider the performance aspect of the proposal, especially how to decrease the call setup delay by applying more efficient and heuristic sorting algorithm because the major portion of MSS algorithm is a sorting process.

REFERENCES

[1] 3GPP TS 29.212. : 3rd Generation Partnership Project; Technical Specification Group Core Network and Terminals ; Policy and Charging Control over Gx reference point(Release 8), 2008-09

[2] Ludwig, R., Eksrom, H., Willars, P., Lundin, N., : AN Evolved 3GPP QoS Concept, VTC Spring 2006, May 2006

[3] Y. Kaneda. : "Policy-based End-to-End QoS Guarantee Using On-Path Signaling for Both QoS Requests and Feedback," ICOIN 2008, I-1, Jan.2008

[4] Jerey K. Mackie-Mason and Hal R. Varian, "Pricing the internet," in Public access to the Internet, Brian Kahin and James Keller, Eds. Prentice Hall, Englewood Cliffs, New Jersey, 1995.

[5] Jerey K. Mackie-Mason and Hal R. Varian, "Pricing congestible network resources," IEEE Journal on Selected Areas in Communications, vol. 13, no. 7, pp. 1141-1149, September 1995.

[6] R. J. Gibbens and F. P. Kelly, "Resource pricing and the evolution of congestion control," Automatica, vol. 35, no. 12, pp. 1969-1985, December 1999.

[7] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," Journal of Operations Research Society, vol. 49, no. 3, pp. 237-252, March 1998,

[8] James F. Kurose and Rahul Simha, "A microeconomic approach to optimal resource allocation in distributed computer systems," IEEE Transactions on Computers, vol. 38, no. 5, pp. 705-717, May 1989.

[9] N. Liu, J. Bigham, "Utility-maximization bandwidth adaptation for multi-class traffic QoS provisioning in wireless networks",the 1st ACM international workshop on Quality of service and security in wireless and mobile networks, 2005-10.

[10] S. H. Low and D. E. Lapsley, "Optimization flow control, I: Basic algorithm and convergence," IEEE ACM Transactions on Networking, vol. 7, no. 6, pp. 861-875, December 1999.

[11] H. R. Varian. : Microeconomic Analysis, W. W. Norton and Co., 1992.

[12] O. Gonzalez and Michael Needham, "QoS Provisioning Architecture for Next Generation Mobile Networks," IEICE Trans. Commun., Vol.E87-B, No5, May. 2004.

[13] T. Nomura, K. Yamori, E. Takahashi, T. Miyoshi, Y. Tanaka, : "Waiting time versus utility to download images," APSITT2001, session 5, pp.128-132, Nov.2001

[14] K. Yamori, Y. Tanaka. : "Relation between willingness to pay and guaranteed minimum bandwidth in multiple-priority services," APCC2004, no.MA06-1, pp.113-117, Aug.2004

[15] Cao, Z., Zegura, E. "Utility max-min: An application oriented bandwidth allocation scheme," IEEE INFOCOM`99. (1999) 793-801