

Adaptive Resource Allocation Method for Mobile Network

Xi LIANG*, Moo Wan Kim*, Shintaro Uno*, Souichiro MIURA**

*Department of Information Systems, Tokyo University of Information Sciences

** Department of Electronic and Communication Engineering, Chuo University

Ykim5jp@ybb.ne.jp

Abstract— Mobile networks are being evolved from 3.5G to 3.9G/4G, 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. In this paper, we propose a new QoS mechanism providing adaptive resource allocation based on each user's preferences by using the utility function borrowed from the field of microeconomics.

Keywords— QoS, Utility Function, Resource Allocation, Signaling, Mobility Management

I. INTRODUCTION

Mobile networks are being evolved from 3.5G to 3.9G/4G, integrated into IP core network based on IMS (IP Multimedia Sub-system). Figure 1 shows the general trend of mobile network developments. In Japan LTE (Long Term Evolution), or 3.9G, commercial service has been started from December of 2010 by NTT DoCoMO as the latest mobile network service.

In this environment, it is essential to provide sufficient QoS (Quality of Service) mechanism which enables different services to provide enhanced user satisfaction. Currently QoS is being studied in 3GPP, 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. Also, even in the operator's perspectives, the needs of user differentiation by their rates are being increased to make the best use of the limited resources. Considering such requirements, in the user's perspective, it is required that QoS mechanism has to provide adaptive resource allocation based on each user's preferences.

In this paper, we propose a new QoS mechanism providing adaptive resource allocation 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. In the proposed mechanism, we have developed a resource adaptation algorithm based on utility function which considers the caller and callee side together. Also, we have developed signaling mechanism to integrate the QoS and mobility management.

The rest of the paper is organized as follows: Chapter 2 presents the basic technologies description. Chapter 3 presents a description of the adaptive resource allocation method with the mobility management and describes the associated signaling mechanism. Chapter 4 presents the simulation result for call setup and handovers with some consideration. Finally conclusion is described in Chapter 5.

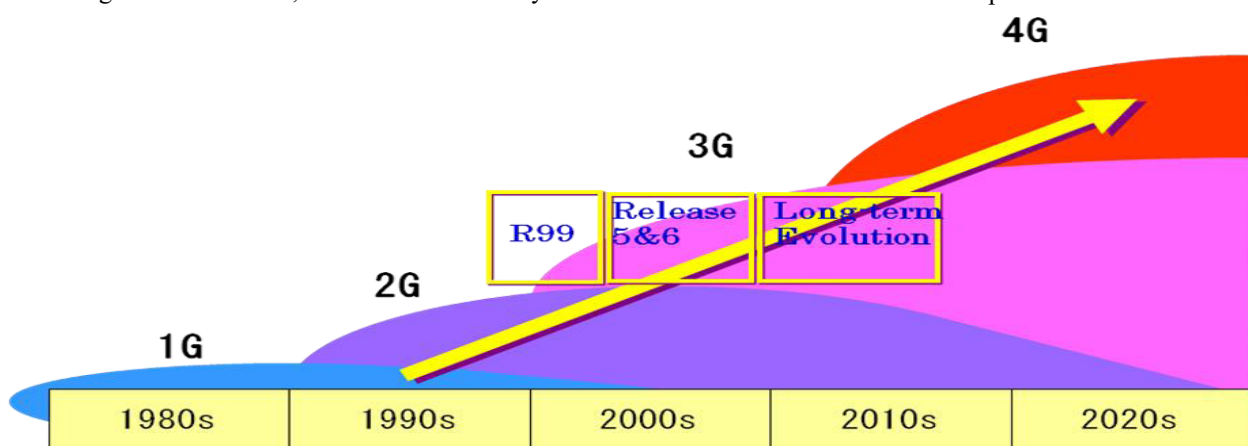


Figure 1. Trend of Mobile Network Development

II. BASIC TECHNOLOGIES DESCRIPTION

There are a number of mobile network architecture, but we assume a simple network architectural model presented in Figure 2 for the sake of simplicity. The network architectural model comprises RAN (Radio Access Network) with MNs (Mobile Nodes) as the wireless domain and CN (Core Network) as the wire line domain with two main nodes: the QS (QoS Server) and the MM (Mobility Manager).

A. Utility Function

Utility function is a notation that we have borrowed from the field of microeconomics to refer to a user's benefit from all possible resource allocation amounts. This work assumes that the network knows the users utility functions by means of the signaling mechanism. Utility functions quantify the value that an application or users perceives for all possible amount of resource allocated. Utility can also be thought as the price that a user would be willing to pay to obtain a specific amount of resource. Therefore the users can tell the system that for obtaining χ units of resource, they are willing to pay b\$/second.

We let $U(\chi)$ denote the utility functions where χ is the amount of resource and $U(\chi)$ is the utility obtained from that allocation. The user's total utility will depend on several QoS metrics, such as bandwidth, delay, and jitter. As a first step we are going to concentrate on the case that the users only are concerned with the bandwidth allocation they receive. This is a step towards a more general solution that will take into account the other factors as well.

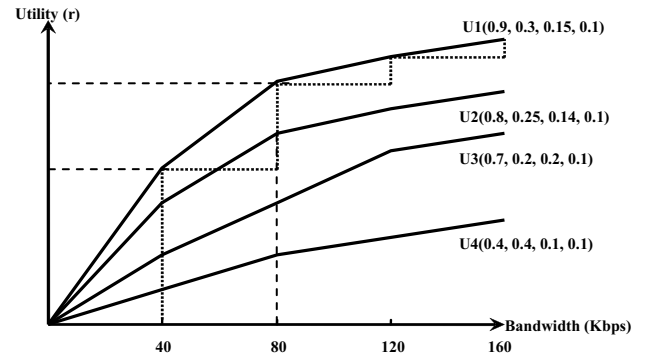


Figure 3. Examples of Utility Function

Utility functions need to satisfy the following properties:

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

Figure 3 shows some examples of utility functions. From there we can observe another property of the utility functions which is concavity. $U(\chi)$ is concave for all $\chi \geq 0$. The meaning of this is that a user attaches a lot more of value to obtaining a unit of resource when it does not have any allocation than when it already has a certain amount allocated.

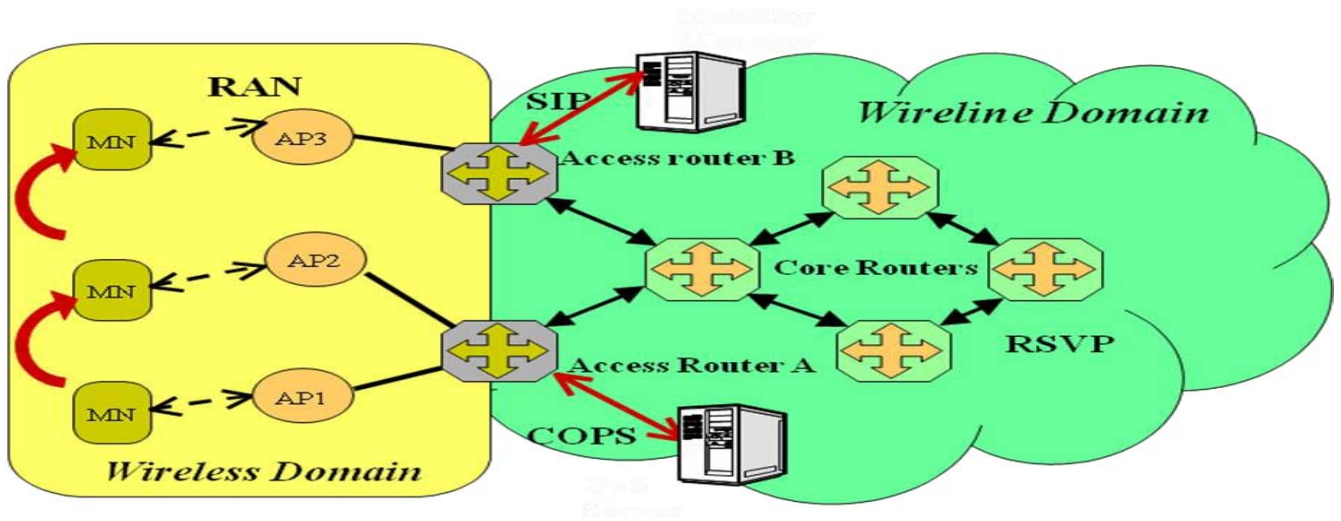


Figure 2 Target Network Architecture

B. MSS Resource Allocation Algorithm

Let us consider where we have M users. 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 r_i^* units of resource that is the i -th component of the solution $r^* = [r_1^*, r_2^*, \dots, r_M^*]$ of the following optimization problem:

$$\begin{aligned} & \text{Max} \quad \sum_{i=1}^M U_i(r_i) \quad (\text{Equation 1}) \\ & [r_1, r_2, \dots, r_M] \\ & \text{subject to} \quad \sum_{i=1}^M r_i \leq C \\ & \quad \quad \quad r_i \geq 0 \text{ for } i = 1, 2, \dots, M \end{aligned}$$

We have developed MSS(Maximum Segmental Slope) resource allocation algorithm to satisfy the Equation 1. The algorithm depicted in Figure 4 is devised to manage the bandwidth resource efficiently among competing users. 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.

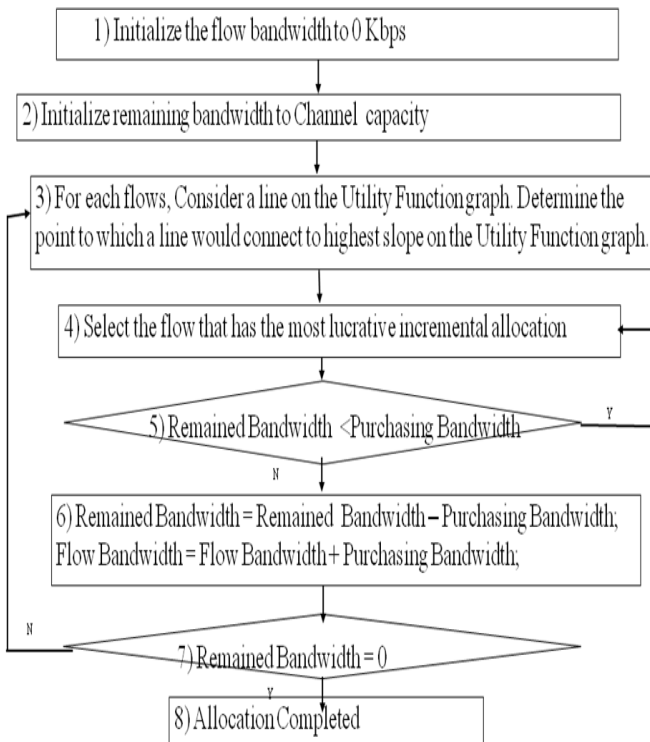


Figure 4 Procedure of the MSS Algorithm

III. ADAPTIVE RESOURCE ALLOCATION METHOD

A. Adaptive Resource Allocation Algorithm

The followings are the proposed adaptive resource allocation algorithm based on MSS:

- 1) QS receives resource allocation request from caller or caller's AR called erAR.
- 2) QS generates this new flow's entry called f_i which has the information of its caller's access router (f_i -erAR), and callee's access router (f_i -eeAR).
- 3) From 2, QS generates a list of the existing flows that need to be re-allocated due to f_i 's joining. It is called 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 f_i joining. For example, f_i -erAR-re-flow-list (for example, it includes flow-a, flow-b, flow-c, ...) and f_i -eeAR-re-flow-list (for example, it includes flow-x, flow-y, flow-z, ...).
- 4) By using the utility function based on MSS algorithm, a temporary allocation is made to f_i and the flows in the caller AR list. So, f_i gets f_i -er-alloc, and the flows in f_i -erAR-re-flow-list get $\sum f_i$ -erAR-re-flow-list-temp-alloc = [temp-alloc-a + temp-alloc-b + temp-alloc-c + ...].
- 5) Same procedures are done for the flow f_i at the callee AR. Flow f_i gets f_i -ee-alloc, and the flows in the list for callee AR (f_i -eeAR-re-flow-list) get $\sum f_i$ -eeAR-re-flow-list-temp-alloc = [temp-alloc-x + temp-alloc-y + temp-alloc-z + ...].
- 6) QS calculates middle allocation for f_i , this value is called f_i -mid-alloc which is the minimum of [f_i -er-alloc, f_i -ee-alloc].
- 7) Now 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 -erAR-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 $\sum f_i$ -erAR-re-flow-list-last-alloc.
- 8) In the similar manner, QS re-allocates each flow in f_i -eeAR-re-flow-list and results in $\sum f_i$ -eeAR-re-flow-list-last-alloc which is the sum of [last-alloc-x + last-alloc-y + last-alloc-z + ...].
- 9) Then the remaining bandwidth in caller side AR is f_i -erAR-remain-bw = (f_i -erAR-total-bw - f_i -mid-alloc - $\sum f_i$ -erAR-re-flow-list-last-alloc).
- 10) Similar, the remaining bandwidth in the callee side AR is calculated in the way as above and results in f_i -eeAR-remain-bw = (f_i -eeAR-total-bw - f_i -mid-alloc - $\sum f_i$ -eeAR-re-flow-list-last-alloc).
- 11) Then the remaining bandwidth for caller AR and callee AR for the flow f_i is remain-bw = min [f_i -erAR-remain-bw, f_i -eeAR-remain-bw].
- 12) At last, as the result, the final allocation for f_i is = [f_i -mid-alloc + remain-bw].

B. Call Setup and HO Signaling

In this paper, we classify 4 classes of signaling messages that used to the assumed network architectural model.

1) Path Signaling : These messages traverse the data path of the flows, will be processed by routers in the path from source to destination. We use RSVP signalling.

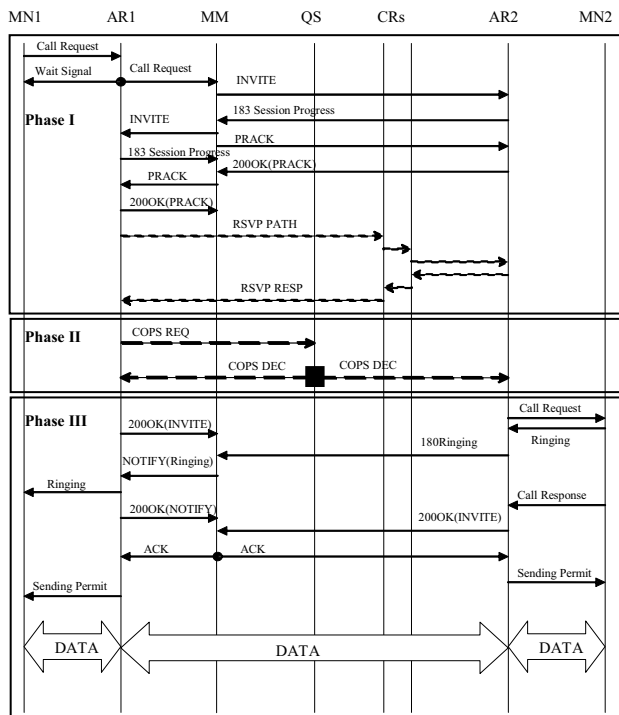


Figure 5. Call Setup Signaling

2) Access Signaling : These message travels between the MN and the AR. As these signals use last hop air-link, the frequency and size of messages must be kept to a minimum. Specific signaling depends on wireless access technologies.

3) Server Signaling : Many entities in a network may need to communicate with each other. MM and QS should communicate to enable handover. In this paper, we use COPS(Common Open Policy Service) signaling.

4) Control Signaling : This signaling refers to Layer 3 messages support call control and mobility management. Proxy MIP(Mobile IP), SIP signaling may be considered as candidate signaling. In this paper, we use SIP in order to verify the feasibility of integration of QoS and mobility management. We don't extend it to MN as usual SIP operation does; this is done to allow the use of non-SIP capable legacy MN.

Figure 5 shows the call setup signaling procedure integrating mobility and QoS management among the entities in the mobile network. In the call procedure, the notification of AP, between MN and AR is abbreviated to simplify the simulation. Also, all the algorithms in the adaptive QoS mechanism are implemented in QS entity.

This procedure comprises 3 phase as follows.

1) The first phase : Call initiation procedure is performed between caller's and callee's AR by the initiative of MM. After MM receives Call Request from MN1 via AR1, MM

confirms the location of MN2 in AR2, then, proceeds call initiation procedure by exchanging INVITE and PRACK message between AR1 and AR2. Then, in order to check the availability of the path between AR1 and AR2 in the core network, AR1 send RSVP PATH message to AR2, then AR2 reply RSVP RESP message if path is available.

2) The second phase : Adaptive QoS mechanism is performed at the QS when it receives COPS REQUEST message from AR1. At this point, the QS calculates the final bandwidth allocation information for both caller's and callee's side according to the algorithm, then send it to both AR1 and AR2 via COPS DEC message.

3) The third phase : MM continues call proceeding procedure by exchanging the OK and NOTIFY messages and send ACK message between AR1 and AR2, then data path is open.

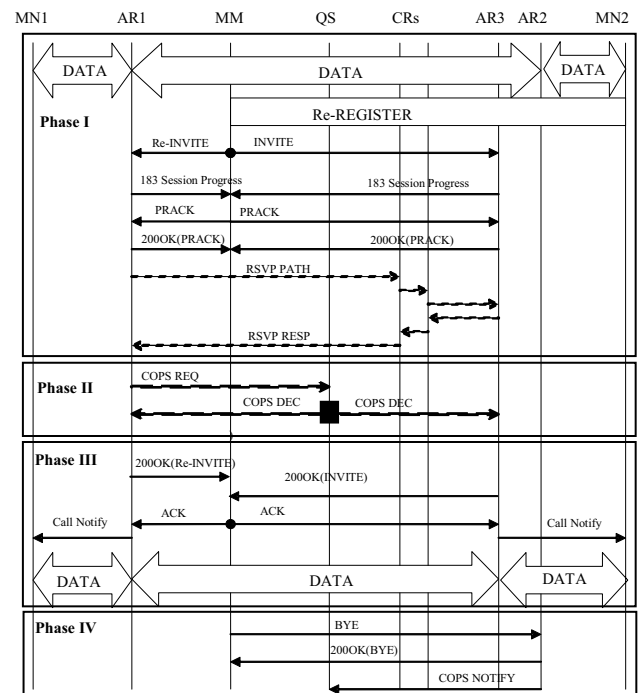


Figure 6. Handover Call Signaling

Figure 6 shows the handover signaling procedure and the entities involved integrating mobility and QoS management.

When a MN moves from one AR to another, resources must be reserved for the MN in the new AR and resources that were reserved in the old AR should be released. When a handover is about to occur, the MM initiates handover QoS signaling based on inputs from the MN.

In details, through the first phase to the third phase, MM proceeds to setup another session with the moving MN's new AR using SIP INVITE method and modify the established session with the corresponding MN's AR using SIP Re-INVITE method. At the last fourth phase, after the new session open, MM should close the old session using SIP BYE. Then, the sessions' setup and modification should be notified to QS using COPS NOTIFY. In detail of the last phase in the Figure 6, we can observe that if necessary, other

MNs in the target AR3 will have re-allocated their resources; however this will depend on the utility function of the handover call. If the handover call has lower utility than other calls in the target AR, then probably depending on operator policies, there will be re-allocations only. If the handover session can't be setup successfully in the new AR (due to the traffic congestion), the call should be sustained until the current radio access is disconnected. When either caller or callee's AR detects the termination, AR should notify MM by using SIP BYE. Then the MM uses the same method to notify the AR of the other calling party. Whenever the AR receives SIP BYE, AR should

IV. EVALUATION

A. Simulation Results

We have used the NS2 simulator with the topology shown in Figure 6. We have implemented call setup and hand over signaling with the proposed algorithm. Each access router has 4 units of bandwidth to administrate (1 unit = 40 kbps). In this simulation, the utility threshold (U_{th}) of core routers is set to be zero where all packets are passed through.

MN1, 2, 3 and 4 under AR1 have session with MN5, 9, 13, and 17 respectively. Each node request 4 units of bandwidth and each AR has 4 units of bandwidth. The data flows are called flow1, 2, 3 and 4. Flow1, 2, 3 and 4 have utility functions U_1 , U_2 , U_3 and U_4 (shown in Table 1) respectively. In the non handover call simulation, flow 1 enters the network at time 10s. As it is the only flow in the system, it is allocated the entire AR1's bandwidth. At time 20s, flow2 enters the network. From MSS algorithm, flow1 and flow2 should be allocated by 2 units and this is shown in Figure 7. Then at time 30s, with flow3's join, the allocation becomes $f1:f2:f3 = 2:1:1$. Finally after $f4$ enters, each flow has same bandwidth at receiver and this is the allocation theoretically.

Figure 8 shows the result when a handover takes place. Utility value of Flow-big, between MN2 and MN9, Flow-small between MN6 and MN10, and Flow-h (handover), between MN8 and MN14 are listed in Table-2. 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-big and Flow-small). And when Flow-h enters the area of AR3, the allocation becomes $Flow-big:Flow-h:Flow-small = 2 : 1 : 1$ units.

B. Consideration

Through the simulation we have confirmed the following items.

- 1) We have confirmed that the allocation in the simulation is equal to the allocation calculated theoretically.
- 2) Users can obtain the more satisfied service. Normally when the bandwidth is fully utilized, the new service request is rejected and in case of handover, the flow is

suspended. The proposed method can solve these issues in some degree.

- 3) By this method 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.

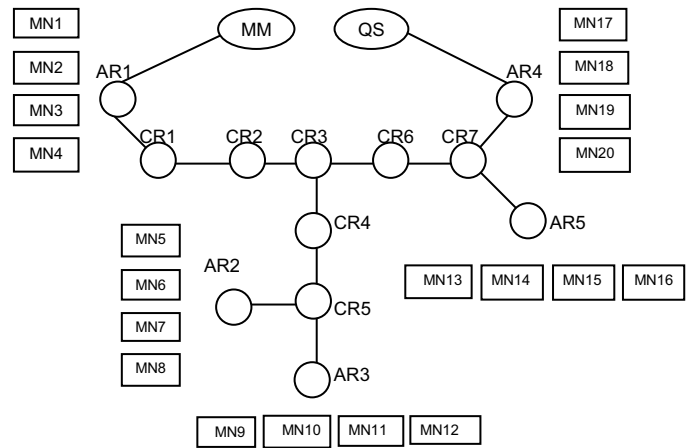


Figure 6. Simulated Network

V. CONCLUSION

We have proposed a new adaptive resource allocation method. As the next step, we can expand this method on the following items.

- 1) In the case that the incremental value of user's utility function is not high and a new flow, where it's incremental value of utility function is higher, is generated, it may be happened that the existing service will be suspended. However, the service can be continued if we modify the method that existing services should be guaranteed by some minimum bandwidth.
- 2) About the assignment of the utility function, we assume two cases; (a) mobile operators can define utility functions and save the information in QoS server, (b) mobile users can select a utility function for a flow and send this information to QoS server by signalling.

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] 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.
- [3] H.R. Varian, Microeconomic Analysis, W. W. Norton and Co., 1992..
- [4] O. Gonzalez, F. Gotoh, and S. Gilbert, "Maximum Network Operator Revenue Resource Allocation, 2002 IEICE Society Conference, Sept. 2002.

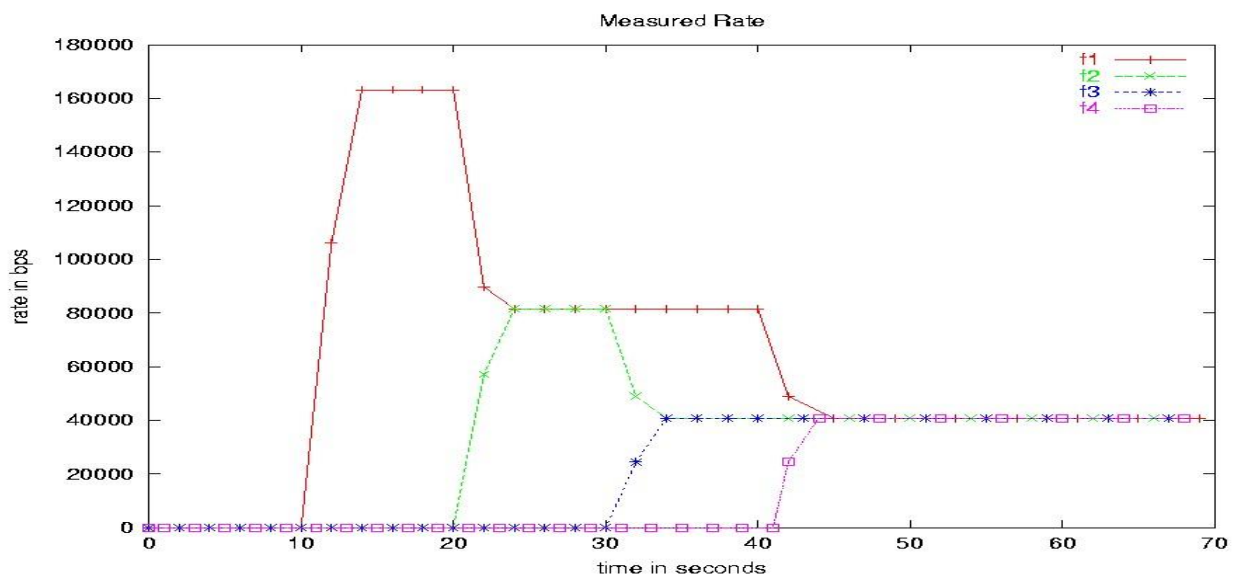


Figure 7. Simple Flows Evaluation : Non-Hand Over

Table1. Utility Values for Flows

Flow	1u	2u	3u	4u
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

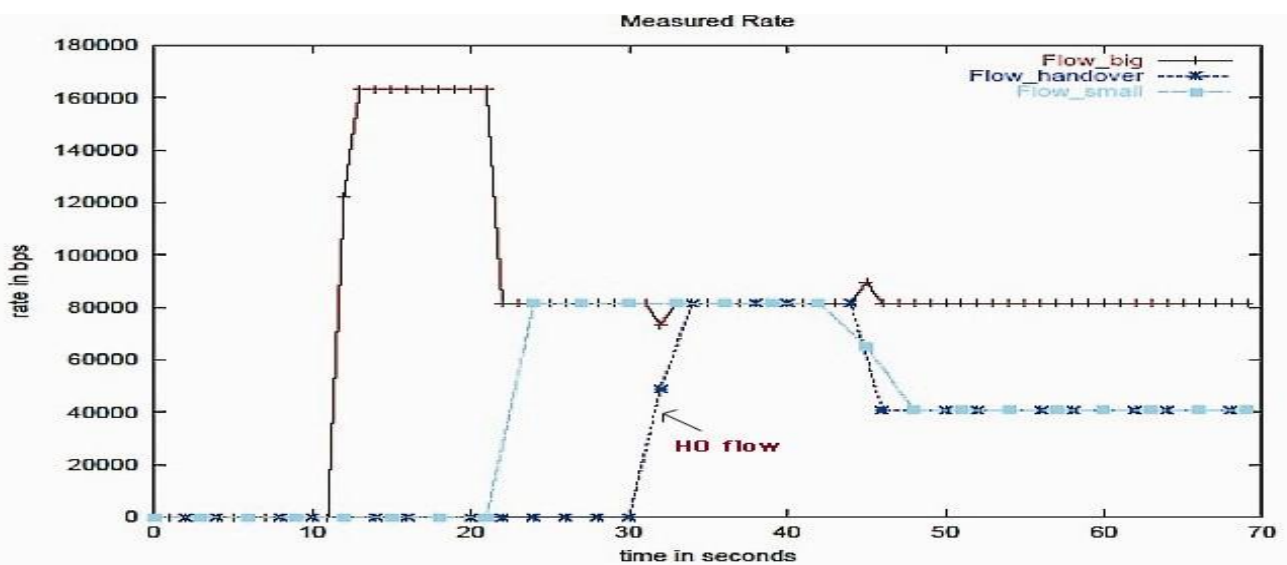


Figure 8. Simple Flows Evaluation including Hand Over

Table 2. Utility Values for Flows

Flow	1u	2u	3u	4u
Flow-big	0.9	0.3	0.15	0.1
Flow-h	0.8	0.25	0.14	0.1
Flow-small	0.7	0.2	0.2	0.1