A Case Study of Policy-based QoS Management in 3G Networks

Mi Kyoung Kim, Kiho Nam, Jaeki Lee and Hwang-Soo Lee
Department of EECS, Division of Electrical Engineering, KAIST
373-1, Guseong-dong, Yuseong-gu, Daejeon, 305-701, Republic of Korea
miki@mcl.kaist.ac.kr

***Abstract*—The proposed core network for the third-generation (3G) mobile telecommunication has the all IP architecture that IP protocol is used for transport of all user data and signaling traffic within the network. While this architecture can bring many benefits (e.g. providing various IP-based multi-media services), it is also noted that the QoS problems still exist since it provides the multi-media services using ordinary IP packets and routing. Especially, for the operator, it is necessary to manage the network resources and capabilities according to their business requirements. As a solution for this problem, the policy-based network management (PBNM) would be suggested by 3GPP. In this paper, we obtain the comparative performance results of specific cases of the network with the policy-based QoS management and with the existing QoS mechanism through ns-2 simulation. Simulation results show that the proposed PBNM architecture can guarantee the required quality of service to meet the business requirements better.**

***Index Terms*—Policy-based Network Management, Third Generation, All IP Network, QoS, COPS, SIP.**

I. Introduction

THE third-generation (3G) mobile telecommunication network which is being standardized by the Third Generation Partnership Project (3GPP) is proposed to have an all IP architecture that IP protocol is used for transport of all user data and signaling traffic within the network. This architecture can bring a lot of benefits such as being efficient to provide various IP-based multi-media services, being able to offer seamless services through the use of IP, being independent of the access technologies, etc.

However, this all IP architecture that supports multi-media services using ordinary IP packets and routing requires proper mechanisms by which the Quality of Services (QoS) should be guaranteed according to their specific traffic characteristics. For example, voice service is sensitive to delay, while data service is sensitive to transmission errors, but both are to be handled in the same IP network. In addition to these different aspects of services, there are also different classes of subscribers. Some subscribers want higher data rate services despite of expensive cost and others want lower data rate services with a cheaper cost. This QoS management issue has not yet been solved completely even in the existing wired IP networks. Wireless telecommunication networks basically have different characteristics compared to the wired networks, so we should pay a particular attention to the QoS issues for the next generation all IP network in the wireless environment.

For this problem, 3GPP has been making progress in developing standards that state the overall network architecture and procedures to guarantee the end-to-end QoS in the all IP network. The main idea for these QoS standards is based on the policy-based network management (PBNM) [1]. It ultimately guarantees required communication qualities through managing the network resources and capabilities by the policies defined by network operators. In this architecture, whenever network operators need new requirements, they only have to create proper policies and apply them to networks.

3GPP has been revising the end-to-end QoS-related standards [1] ~ [5] in order to resolve the several issues like structural problems and scalability. However, there have been few experimental researches to obtain the performance of the policy-based QoS management methods by applying them to an arbitrary virtual network like the 3G network. Therefore, in this paper, we focus on the performance analysis of them through simulation. We obtain the comparative performance results of specific cases of the network with the policy-based QoS management and with the existing QoS mechanism through ns-2 simulation. The simulation results show that the proposed PBNM architecture can guarantee the required quality of service to meet the business requirements better.

The paper is organized as follows. In section II, we present the overall network architecture and procedures as a 3GPP solution to guarantee end-to-end QoS in the all IP network. In section III, we describe the general architecture and components of the PBNM method. In section IV, we suggest specific policies and analyze the experimental results obtained by applying them to a virtual network that is similar to the 3G network. Finally, we make conclusions.

II. End-to-End QoS in 3G Networks

3GPP has proposed the following overall network architecture using PBNM technology to guarantee end-to-end QoS in the all IP network [1] as shown in Figure 1. In this figure, we omit some components for the radio interface and access network between UE and GGSN: i.e. Universal Terrestrial Radio Access Network (UTRAN), Serving GPRS Support Node (SGSN), etc. It is the architecture that the QoS management functions for controlling the external IP bearer services are added on top of the Universal Mobile Telecommunications System (UMTS) bearer service QoS management functions already existing in the control plane. IP bearer service manager (IP BS Manager) in Gateway General Packet Radio Service Support Node (GGSN) and policy control function (PCF) in Proxy-Call Server Control Function (P-CSCF) are the QoS management functions for controlling the external IP bearer services. According to the service-based local policy, PCF maps the Session Description Protocol (SDP) parameters received from P-CSCF into the authorized IP QoS parameters and sends them to the IP BS Manager in GGSN. Inter-working between the mechanisms and parameters used within the UMTS bearer service and those used within the IP bearer service is provided by the translation/mapping function in the GGSN. It maps the authorized IP QoS parameters received from PCF into the authorized UMTS QoS parameters. The translation/mapping function in the user equipment (UE) maps the SDP parameters into some UMTS QoS parameters [1].

The IP BS Manager and PCF are also the components to handle PBNM. PBNM manages the network resources and capabilities by the policies defined by network operators. It is implemented by communications between a policy decision point (PDP) and a policy enforcement point (PEP). We would present a detailed description of the technology in the next section. In the all IP network of 3GPP, the IP BS Manager in GGSN is equivalent to PEP and the PCF in P-CSCF is equivalent to PDP. 3GPP also define the interface between GGSN and PCF as the Go interface using the Common Open Policy Service (COPS) protocol. The Go interface shall conform to the Internet Engineering Task Force (IETF) COPS framework as a client/server interface between GGSN and PCF. The messages used for the COPS protocol are REQ, DEC, RPT, etc [7].

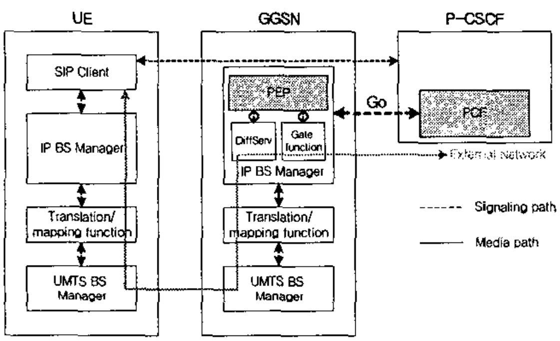

Figure 1. The PBNM architecture suggested by 3GPP.

We have to look into the end-to-end session flows in order to know how to achieve the policy-based QoS management in the all IP network. Especially, we will be able to understand more exactly by examining the messages exchanged through the Go interface. Those procedures are described in the dotted lines in Figure 2. They are 'Authorize QoS Resources', 'Resource Reservation', and 'Approval of QoS Commit' procedures. We would explain in detail those procedures but omit the description of the end-to-end session flows because they are described minutely in [3]. In the 'Authorize QoS Resources' procedure, PCF authorizes the IP QoS resources, and maps the SDP parameters into the authorized IP QoS parameters. In the 'Resource Reservation' procedure, PCF makes an authorization decision of the requested IP flows, and returns a COPS DEC message including the policy information to be enforced by GGSN in order to perform the policy-based admission control according to the decision. After UE sends a 200 OK message to P-CSCF in order to complete the session setup, PCF receiving the 200 OK message approves the QoS commit, and sends a COPS DEC messages to GGSN to open the gate. Therefore, the end-to-end QoS in the 3GPP all IP network can be managed and guaranteed by using both the Session Initiation Protocol (SIP) [6] and the COPS [7] mechanisms [3] [5].

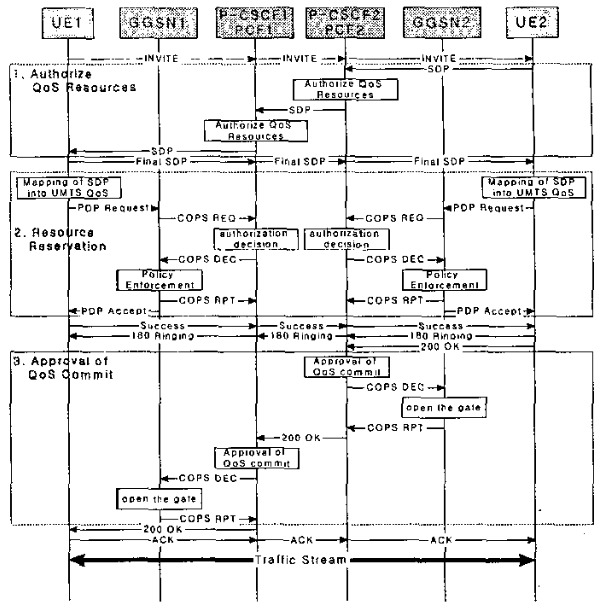

Figure 2. End-to-end session flows and QoS interaction procedures.

Then, how can a network operator apply the policies when a new business requirement comes up and he wants to manage the network resources based on the policies? To find out the solution for this question, first, we have to know the configuration of PBNM and the procedures of applying a policy to the network presented in the next section.

III. Policy-Based Network Management

Policy-based Network Management (PBNM) ultimately guarantees required communication qualities through managing the network resources and capabilities by the policies defined by network operators. In general, the PBNM technology needs four components. They are policy management tool, policy repository, PDP, and PEP. The interaction and protocols between these components are shown in Figure 3 [8]

Policy is a rule invented to manage the network resources based on business requirements. It is usually defined by network operators for accomplishing their business requirements. For example, there are various policies: to give a higher priority for a specific application or to provide different subscribers with differentiated services, etc. In a standard policy-based network, policy rules consist of two components: *conditions* and *actions* [11]. *Conditions* are kinds of situations for applying policy. They might include parameters such as user names, addresses, protocols, application types, and time of the day. *Actions* are behaviors taken when *conditions* are met. They are such as bandwidth guarantee and access control. These policies are created by a policy management tool and stored in a policy repository. The policy management tool allows the network operator to define policies in a human readable fashion and then translates each policy to a computer executable command. A PDP makes policy decision through interpretation, and PEP enforces the policy decision. PEP might have the specific QoS guarantee functions such as DiffServ and IntServ to enforce the policy decision. According to the 3GPP standard, as the capability of IP BS Manager in GGSN, the DiffServ edge function is mandatory and the IntServ function is a network operator choice [1]. Hence, in this paper, we just consider DiffServ as a solution of policy control of PEP.

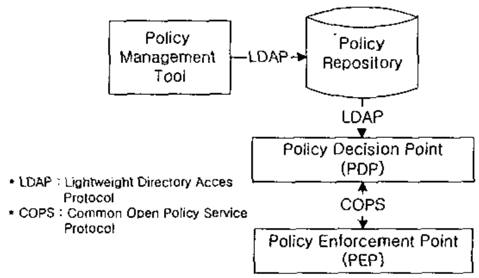

Figure 3. Policy-based network management components [8].

On the other hand, the architecture of the all IP network currently specified in the 3GPP does not include functions of the policy management tool and policy repository. It is because the policy which is currently being specified is based on specific services and implemented in the visited network rather than the service level agreement per individual subscriber or the strategy of the home network operator. This policy is called the service-based local policy. In this architecture, whenever a new policy requirement is needed, we have to inform all PDPs of the new policy and update them. This is not scalable. Due to these architectural limits, 3GPP has been making constant efforts such as modifying architecture and defining new signaling to apply dynamic policies efficiently.

Until now, we considered the end-to-end QoS management mechanisms of the 3GPP standards and a general PBNM technology. 3GPP has been making progress in developing standards of the overall network architecture and procedures that the PBNM technology is applicable. However, their standards are based on applying the existing IP QoS guarantee mechanism to the 3GPP standard core network rather than concrete experimental results. The experiments for QoS guarantee technologies just performed so far are by applying them to the existing core network like Internet and its specific application traffic. In reality, there exist far different characteristics between the existing application traffic and the 3G mobile telecommunication traffic. However, there have been few researches for quantitative analysis of PBNM in the all IP network architecture using the 3G mobile telecommunication traffic. Therefore, we compare the performances of the network in some specific cases with the policy-based QoS management and with the existing QoS mechanism through simulation. More specifically, we consider the DiffServ architecture with and without PBNM.

IV. Simulation Scenarios and Results

In this section, we obtain the comparative performance results of specific cases of the network with the policy-based QoS management and with the existing QoS mechanism (here, DiffServ) through ns-2 simulation. The version of ns-2 simulator used is ns-2.1b9a.

The UMTS specifications define four QoS classes according to delay sensitivity: conversational, streaming, interactive, and background. Conversational class is the most delay sensitive traffic class while background class is the least delay sensitive. Conversational and streaming classes are intended for carrying real-time traffic flows like voice/video telephony and streaming audio. Interactive and background classes are mainly used for traditional Internet applications like WWW, e-mail, telnet, and ftp. They do not have severe delay requirements but have to preserve payload content. Interactive class has three traffic handling priorities and those priorities are higher than that of background class. So, background class has the least priority like the best-effort service of Internet [2]. Considering the relation between the UMTS QoS classes and the DiffServ Per Hop Behaviors (PHBs), each class can be mapped as follows: conversational into EF, streaming into AF4, each interactive with traffic handling priorities 1, 2, and 3 into AF3, AF2, and AF1, and background into BE [10]. Even though mobile telecommunication network will evolve into 3G network, unless a new killer application comes up with a far different characteristic, there will be only traffic similar to the UMTS traffic. Hence, we consider the UMTS traffic as simulation traffic.

The specific policies applied are 'preference of the conversational class' and 'preference of the real-time classes'. From now on, we would call 'preference of the conversational class' as 'policy 1' and 'preference of the real-time classes' as 'policy 2'. Both policies have the same condition. If the link utilization monitored by PEP is over a specific rate (it can be changed by a network operator choice) of the output link capacity, PEP reports it to PDP. The results of this report are divided into two cases as the applied policies. In the case of the policy 1, PEP changes each DiffServ Code Point (DSCP) of the others except for conversational class (EF) into BE DSCP. While, in the case of the policy 2, PEP changes each DSCP of the others except real-time classes (EF and AF4) into BE DSCP. These policies seem to be very extreme. However, we can understand more clearly the influences of them through applying them. They are also feasible policies because network operators can apply very various policies according to their business requirements. The pseudo coding of the suggested policies is shown in Figure 4.

```
IF ((link_throughput >= link_bandwidth*0.95) && (out-of policy)) THEN {
    IF (policy_type ==1) {  // preference of the conversational class
        AF4_DSCP = BE_DSCP;
        AF3_DSCP = BE_DSCP;
        AF2_DSCP = BE_DSCP;
        AF1_DSCP = BE_DSCP;
    } ELSEIF (policy_type==2) {  // preference of the real-time classes
        AF3_DSCP = BE_DSCP;
        AF2_DSCP = BE_DSCP;
        AF1_DSCP = BE_DSCP;
    }
} ELSEIF ((link_throughput < link_bandwidth*0.95) && (in-policy)) THEN {
    return to original queue and codepoint
} ELSE THEN { no action }
```

Figure 4. Pseudo coding of the specific policies 1 and 2.

The procedures that the policies are applied in the real 3G core network are as follows. Each GGSN is monitoring periodically the output link utilization. When the link throughput becomes above a specific threshold over the output link capacity, it sends a COPS REQ message. Then PCF receiving the COPS REQ message makes a policy decision, and return a COPS DEC message to the corresponding GGSN. Now GGSN enforces those policy decisions. After GGSN enforces those decisions, it sends a COPS RPT message to PCF. On the contrary, even when the link throughput becomes below a specific threshold under the output link capacity, GGSN sends a COPS REQ message to PCF and waits the decision of PCF. The following procedures are the same as the former case. These procedures can be executed in the middle of both call processing explained in section II and traffic transmission after call processing. The signaling used at this time is based on the 3GPP specifications.

The network for the simulation study is shown in Figure 5. We omit some components for the radio interface and access network between UE and GGSN. This is reasonable according to the concept of the 3GPP all IP architecture, which specifies that the access and core networks are independent [10]. GGSN is a regular router supporting the DiffServ edge function. P-CSCF is a policy server and SIP server including PCF. There is also one core router having a bottleneck link because the effect of the bottleneck router is the greatest even if traffics are passed through several routers. The bandwidth of 5 Mbps is assigned to the bottleneck link for fast convergence to a congestion condition. Each UE generates a specific class traffic with a specific DiffServ PHB. UE1 generates voice traffic for EF PHB, UE2 generates video traffic for AF4 PHB, and each UE3 ~ UE6 generates data traffic for AF3 ~ AF1 and BE PHB, respectively. UE1 and UE2 generate UDP traffic for real-time services. Voice flows are generated by AMR codecs at 12.2 kbps, while video flows are generated by H.263 codecs at the average rate of 28 kbps and the peak rate of 40 kbps. These traffics generated are exponential distributed. UE3, UE4, and UE5 generate TCP traffic for WWW. UE6 also generates TCP traffic, while it is for ftp service. Each traffic of UE3 ~ UE5 has the average rate of 64 kbps, 144 kbps and 384 kbps, respectively. UE6 is a merged traffic of 64 kbps, 144 kbps and 384 kbps. Those four traffics are generated by using the pareto distribution.

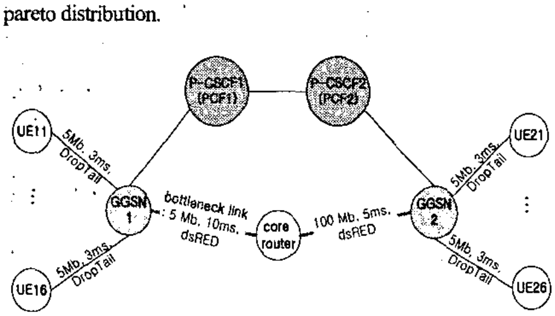

Figure 5. A network environment for simulation.

We use the weighted round robin (WRR) packet scheduler for scheduling of the output link in GGSN. The weights are given as EF : AF4 : others = 0.1 : 0.15 : 0.75 [10]. So, conversational class traffics are guaranteed to have 10 percent, streaming class traffics are guaranteed to have 15 percent and interactive and background class traffics are guaranteed to have 75 percent of the output link capacity.

In this experiment, we generate voice and video flows so that the sum of voice flows yields 10 percent and that of video flows yields 15 percent of the output link capacity. We generate only one flow for each interactive class because we are just interested in making a congestion situation by background traffic. A 592 kbps flow is added every 10 sec simulation time for background flows, i.e. 592 kbps, 1184 kbps, 1776 kbps, and so on. Subsequently, the rate of background flows is raised gradually to result in a congestion situation. The output link capacity of this experiment is 5 Mbps, so the suggested policies will be applied during 5 background flows are generated. That is, the experiment considers voice and video flows while background traffic is increased gradually.

The end-to-end delay and packet loss rate for real-time traffic are measured. These metrics are selected because the major quality factor among UMTS QoS classes is delay sensitivity. The simulation results are provided in Figure 6. a), b).

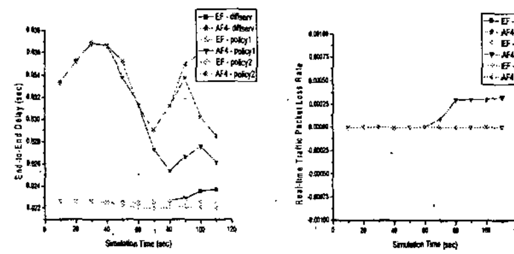

Figure 6. a) End-to-end delay; b) packet loss rate

In Figure 6, we show that there are some variations even before applying the policies. This is because generated traffic is based on probability distribution. When the policies 1 and 2 are applied, the end-to-end delay of voice flows is decreased even though the network is congested. After the policies are applied, the average end-to-end delay of voice flows is changed from 23 msec to 22.03 msec under the policy 1 and to 22.63 msec under the policy 2 in comparison with 'DiffServ without PBNM'. In case of the policy 1, voice flows are not interrupted by the other traffics because it gives higher priority only to voice flows and the same priority as background traffic to the other traffic. While in case of the policy 2, video flows are added as the obstacles of voice flows. So the average end-to-end delay of the policy 1 is lower than that of the policy 2. But these differences are very small because voice and video flows are generated less than or equal to the amount of the guaranteed load, so the effects of the policies are small. For this reason, on the contrary, the average end-to-end delay of the video flows under the policy 2 is a little longer than that with 'DiffServ without PBNM'. But the variation of the video flows is the least under the policy 2. This shows that applying the policy 2 stabilizes the video flows. The packet loss rate of the video flows under the policy 2 is decreased to zero in contrast to that with the policy 1. It is worst under the policy 1 because it does not guarantee video flows. Therefore, as a result, we show that the policy 1 improves the characteristics of voice flows and the policy 2 improves the characteristics of voice and video flows.

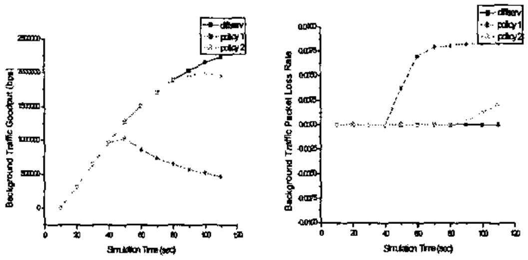

Figure 7. a) Goodput (bps); b) packet loss rate

In Figure 7, we show the goodput and packet loss rate of the background traffic. We do not present the results of the interactive class traffic because the characteristics of them change little by the relatively small traffic generated in the experiment. The impact of background traffic is the biggest because DSCPs of both the interactive and the background classes are changed into BE DSCP when the policies are applied. The goodput of the background traffic decreases in order 'DiffServ without PBNM', 'DiffServ with the policy 2', and 'DiffServ with the policy 1' when the policies are applied. This is because much more traffic crowd to the queue for BE. Therefore, the packet loss rate of the background traffic is getting worse during applying the policies. By the way, the packet loss rate of the background traffic under the policy 2 is lower than that under the policy 1. These results are explained in that the policy 2 does not impose duty for video flows to the queue for BE traffic. Therefore, we note that both the policies 1 and 2 obtain the better performance for real-time traffic at the sacrifice of the other traffics.

In the experiments of Figures 6 and 7, the clear effects of the real-time flows do not appear because they occupy small portion of the output link capacity. To show the performance variations of the real time flows, we perform the experiment with more generated traffic flow. The voice flows are increased up to 30 percent of the output link capacity and so are the interactive class traffics. The packet loss rate of the real-time traffic in this scenario is given in Figure 8. The packet loss rate of voice flows are decreased greatly when the specific policies are applied, while those of video flows are decreased rather small by relatively generating small traffic. As a result, we note that the real-time flows are guaranteed when the policies are applied.

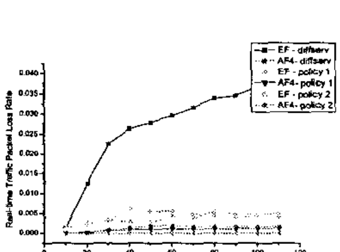

Figure 8. Real-time traffic packet loss rate when voice flows increase up to 30% and other traffic increase.

Those experimental results are also similar in various environments such as increasing voice flows. That is, when the 'DiffServ with PBNM' is applied, the behaviors of the real-time traffic are better while those of the other traffics are worse. This shows that the PBNM architecture can reflect the business requirements, 'preference of the conversational class' and 'preference of the real-time classes', well and also guarantee the qualities of the traffic.

Applying these policies is different from the existing QoS mechanisms where real-time traffic is prioritized. The existing QoS mechanisms only provide the static service differentiation, while applying policies can provide it and can also provide the functions that change from existing policy to other policy at any time. For example, when the policy 1 is applied, if a network operator wants to apply the policy 2 after office hours, the existing QoS mechanisms cannot do that but PBNM can do. That is, the existing QoS mechanisms cannot reflect dynamic business requirements but the PBNM can do that. Therefore, we can apply easily any business requirements by introducing the policy-based network management technology.

V. Conclusions

In this paper, we compared the performance of various traffic flows in the all IP network with the proposed policies and that with the existing QoS mechanism through ns-2 simulation. The simulation results show that the proposed PBNM is effective to meet the QoS requirement of the specific traffic classes. Therefore, whenever network operators need new requirements, they only have to create proper policies and apply them to the network. This means that the network operators are not affected by the limits of the existing QoS mechanisms once they apply the PBNM technology. Especially, this is more effective when a new killer application comes up.

In this experimental study, we consider just a few scenarios. There exist many policies applicable to the real networks. Therefore, it is necessary to simulate and analyze more cases and architectures for other traffic classes in order to obtain effective PBNM architectures and mechanisms.

References

[1] 3GPP TS23.207 v5.3.0, "End-to-End QoS Concept and Architecture (R5)," Mar. 2002

[2] 3GPP TS23.107 v5.4.0, "QoS Concept and Architecture (R5)," Mar. 2002

[3] 3GPP TS23.228 v5.4.1, "IP Multimedia (IM) Subsystem; Stage 2 (R5)," Apr. 2002

[4] 3GPP TS29.207 v1.5.0, "Policy control over Go interface (R5)," May, 2002

[5] 3GPP TS 29.208 v1.4.0, "End-to-End signaling flows (R5)," May, 2002

[6] M. Handley, "SIP: Session Initiation Protocol," rfc2543, Mar. 1999

[7] D. Durham, et al., "The COPS (Common Open Policy Service) Protocol," rfc2748, Jan. 2000

[8] A. Mehra, et al., "Policy-Based Diffserv on Internet Servers: The AIX Approach," *IEEE Internet Computing*, vol: 4, Issue: 5, pp. 75~80, Sept.-Oct. 2000

[9] Salsano S., et al., "QoS control by means of COPS to support SIP-based applications," *IEEE Network*, V.16, N.2, pp. 27~33, March/April, 2002

[10] Sotiris I. Maniatis, et al., "QoS Issues in the Converged 3G Wireless and Wired Networks," *IEEE Communications Magazine*, Vol. 40, No. 8, pp. 44~53, Aug. 2002

[11] Roberto Mameli, "The Strong Need for Policy in the QoS Environment," *Fitce2000 conference*, Limerick (Ireland), Aug. 2000