

QoS-aware Active Queue Management Scheme for Multimedia Services

Hyun Jong Kim*, Pyung-Koo Park**, Ho Sun Yoon**, Seong Gon Choi*

*College of Electrical & Computer Engineering, Chungbuk National University,
410 Seongbong-ro, Heungdeok-gu, Cheongju Chungbuk, Korea

**Electronics and Telecommunications Research Institute, 138 Gajeongno, Yuseong-Gu, Daejeon, Korea

*{hjkim78, sgchoi}@cbnu.ac.kr, **{parkpk, yhs}@etri.re.kr

Abstract—In this paper, we propose QoS-aware AQM(Active Queue Management) method considering video quality level that service users request in DiffServ network environment. QoS provision is essential in order to provide multimedia services because real-time multimedia services such as IPTV and VoIP are sensitive to delay and packet loss. However, existing AQM methods can't guarantee QoS for a multimedia service because they control queue threshold just considering queueing delay or accept the service considering only packet loss rate. So, we propose the QoS-aware AQM method guaranteeing end-to-end network delay and packet loss rate which can derive the QoS lever range for multimedia service requested in each hop according to the number of hops in IPTV CDN(Contents Delivery Network). As another contribution of the proposed scheme, its compatibility in DiffServ is also considered by cooperating with the provider's quality policy.

Keywords— AQM, DiffServ, QoS, Delay, Loss, IPTV

I. INTRODUCTION

The growing relative proportion of video, voice and gaming applications such as IPTV (Internet Protocol TV), VoD (Video on Demand) and VoIP (Voice over IP) based on user datagram protocol (UDP) and alongside the legacy transmission control protocol (TCP) applications emphasizes the importance of quality of service (QoS) provisioning. Currently, QoS is addressed by the active queue management (AQM) and the differentiated services (DiffServ) architecture [1].

AQM in routers has been proposed as a solution to some of the scalability issues associated with end-to-end approach to congestion control and QoS provision. Beyond congestion control, controlling queues in routers is important because unstable router queues can cause poor application quality. Although, existing AQM schemes[2-8] explicitly try to control router queues by probabilistically dropping packets, they don't realize the requested level of quality at application layer but provide only the congestion avoidance depending on the network state.

DiffServ arose as a simpler, more scalable, manageable, and easily deployed solution for service differentiation in IP networks to provide QoS to service users. Its premise is that individual flows with similar QoS requirements can be

aggregated in larger traffic groups that use a certain set of forwarding rules at the core routers, their per-hop-behavior (PHB). So, DiffServ can exploit the statistical multiplexing gains of aggregation, it does not require any signaling protocol, and it is a distributed architecture, because the PHBs have strictly local semantics. The basic drawback is that the network cannot enforce a required level of QoS to individual flows within an aggregate[12][13].

Existing AQM schemes and DiffServ architecture do not guarantee the requested service quality through the control of maximum end-to-end delay, maximum packet loss rate and jitter. They don't consider the minimum network performance that must be guaranteed according to the required QoS level at each hop in the service delivery path[14-17].

Therefore, we propose the EST(Either Side Threshold) scheme as QoS-aware AQM scheme, which can consider the level of quality that each hop can support in DiffServ environment. The proposed scheme can deliver the information of required quality level at application layer and the number of end-to-end hops to network layer by using the modified DSCP(DiffServ Code Point) field. Using the proposed scheme, we allow the network equipments to recognize the quality of service information, and can provide the reliable multimedia service through the QoS-aware AQM and congestion control.

II. RELATED WORKS

In this section, the various AQM (Active Queue Management) schemes are introduced and we look around the performance analysis model of these schemes.

A. Congestion avoidance mechanism

Lots of proposed queue management schemes [3-6] are RED[2] oriented. The difference of these schemes is mainly in the parameter-adjusting method. Their mechanisms for adaptation of dropping probability in response to network condition are different. In this section, brief descriptions of these queue management schemes are given. These representative queue management schemes include Drop-tail, RED[2], DRED[5, 6], DSRED[7], WRED[8].

For Drop-tail, arrival packets are dropped when queue overflow occurs. Drop-tail incurs large queue length and high packet loss rate at congested links. Especially, Drop-tail

results in a phenomenon, called global synchronization, when a lot of TCP flows compete in a bottleneck. However, Drop-tail is the most widespread queue management scheme due to its simplicity.

Random early detection (RED) [2] is a queue management scheme that is intended to remedy the shortcomings of Drop-tail. The dropping probability of RED is decided by the queue length. An arriving packet may be dropped before the queue is full. It is an early congestion notification. The dropping probability increases in order to provide enough early congestion notifications.

[5] proposed the Dynamic Random Early Detection (DRED) model, which uses a newly introduced parameter i.e. warning line. DRED scheme responds early enough to the increased number of packets at the gateway. Also, the maximum drop probability of packets show improved performance over the original RED. It was concluded that our scheme demonstrated superiority by avoiding global synchronization and there is great reduction in the fluctuations of the actual queue size.

[7] proposed a new active queue management scheme (Double Slope Random Early Detection: DSRED) which achieves a higher throughput than RED and retains all the advantages of RED. The principle of DSRED is that the gateway buffer segment between K_l and K_h is divided into two sub-segments separated by K_m . The overall drop function from K_l to K_h are described by two linear segments with slope α and β respectively. The slopes for these two linear segments are complementary and are adjusted by the mode selector γ . Here, the K_m is set as $0.5(K_l + K_h)$, which can be configured by gateway administrator.

DSRED is similar to RED in two respects. First, both of them use linear drop junctions to give smoothly increasing drop action based on average queue length. Secondly, they calculate the average queue length using the same definition to account for the effect of long term congestion. Therefore, DSRED inherits the advantages of RED. However, DSRED's two-segment drop function provides much more flexible drop operation than RED.

However, RED, DRED and DSRED are insufficient in guaranteeing QoS for the high priority service because the drop probability which is applied according to the network condition is flexible.

So, WRED[8] (Weighted RED) discards incoming packets with dropping probabilities for multi-class traffic. The flexibility of the WRED parameters will be illustrated with respect to performance parameters and traffic characteristics. However, in the network congestion situation, this scheme is unable to actively guarantee the QoS of the high priority service.

B. AQM in a DiffServ network

The diverse service requirements of emerging Internet applications increase the need for flexible and scalable IP QoS schemes. The Differentiated Services (DiffServ) architecture has a scalable QoS mechanism providing different service levels in the backbone networks. So, in the Internet environment, many research for supporting the various

services (e.g. VoIP, IPTV, and Videoconference) through the service differentiation (DiffServ) has been in progress.

[12] proposed MRED (Multi-level RED) to deal with three drop preference levels. It is an extension of RED and has three parameter sets for three different drop preferences. Multiple queues can be assigned to each class in order to improve the fairness among flows in the same class. The number of queues in a class is a design choice and depends on the queue management capability of the router.

On the other hand, [15] introduced another attractive property for an AQM scheme and propose a modification of Hybrid RED (HRED), which overcomes weakness of the pure RED, for the property. The modified AQM is designed to assured drop probability in a properly provisioned network as well as to offer sheltering and load tolerance. Also, [13] proposed a new framework for modeling AQM queues which is applied to a network of RED queues with DiffServ capabilities termed DiffRED which is similar to the WRED algorithms.

In this way, the many schemes for guaranteeing QoS for the various services in the Internet environment are proposed. However, it is difficult actively to guarantee the QoS of the high priority service through these schemes.

III. THE PROPOSED EST(EITHER SIDE THRESHOLD) ACTIVE QUEUE MANAGEMENT IN DIFFSERV

Existing AQM schemes just allocate a queue threshold by considering only queueing delay or determine whether the service is admitted in access point by considering the packet loss rate. However, existing schemes require a method which can actively packet processing according to QoS level of multimedia service because legacy network equipments don't recognize the quality information in application layer.

So, we propose the QoS-aware AQM scheme which can recognize the network quality parameters such as delay, jitter and packet loss rate, and the number of end-to-end hops. Using the proposed scheme, network equipment such as routers recognizes network delay boundary and loss rate boundary to be guaranteed in each hop, and can provide reliable service quality to end users.

It is most reasonable to mark AF classes in such a way that the traffic type is homogeneous within a particular AF class. For example, one might allocate AF classes as follows.

- AF1: IPTV (HD and SD), VoIP
- AF2: VoD, Video conference
- AF3: Bulk data transfers
- AF4: Transactions (instant messaging, database access, interactive applications)

After that, the priorities of the different traffic flows (and packets within the flows) can be marked using the three AF drop precedence within each class.

However, there are limitations to reflect the service quality information using only drop precedence of AF class because multimedia service is sensitive to network performance. Network equipment such as switches and routers don't

consider queue threshold, and carry out best effort packet processing because they do not gauge the level of quality using DSCP information. Therefore, we modify existing DSCP field to delivery multimedia service characteristics, required QoS level and the number of end-to-end hops.

For example, IPTV service and VoIP service is considered as a killer application service. Here, multimedia service quality indicators and standards of ITU-T and DSL-Forum are referred to transmit quality information. IPTV and VoIP service are classified as AF1 service class because they are very sensitive to the delay, packet loss rate. Table 1 shows end-to-end delay, packet loss rate and jitter of real-time multimedia.

As shown in table 1, two bits of existing drop precedence is used to transport QoS information because AF1 service class delay and loss rate can be classified broadly into two types. Also, the number of end-to-end hops should be limited within 20 hops in order to provide reliable HD IPTV service. Therefore, the information of end-to-end hop counts is delivered to the network using unused field of DSCP.

Figure 1 shows the functional architecture of QoS-aware AQM scheme and EST queue management algorithm. First, as shown in figure 1, service class of incoming packets is classified according to the DSCP field in IP header. To determine the threshold of the queue, packet arrival rate and the modified DSCP field reflecting QoS level for delay and loss rate are confirmed. Next, queue threshold(TD) for queueing delay in one hop to reliably provide the multimedia

services and another queue threshold(TL) for packet loss in one hop are calculated.

TABLE 1. MULTIMEDIA SERVICE QoS PARAMETERS AND STANDARD IN RECOMMENDATIONS[9][10]

Application Service	Delay Requirement	Loss rate Requirement	Jitter Requirement
HD IPTV	Less than 200ms	Less than 10^{-6}	Less than 50ms
SD IPTV	Less than 200ms	Less than 10^{-5}	Less than 50ms
VoIP	Less than 100ms	Less than 10^{-5}	Less than 50ms

After two threshold(TD, TL) to meet QoS quality parameters such as delay and loss rate is calculated, max threshold and min threshold are allocated as figure 2. Depending on network conditions, while ensuring the required quality level, queue threshold can be actively controlled within the threshold margin. For example, the service can be stably provided within threshold margin when TD is greater than the TL. In contrast, the multimedia service can't be provided due to delay and packet loss rate increase when TD is less than the TL. At this time, EST notifies path re-search message to media server and QoS management system.

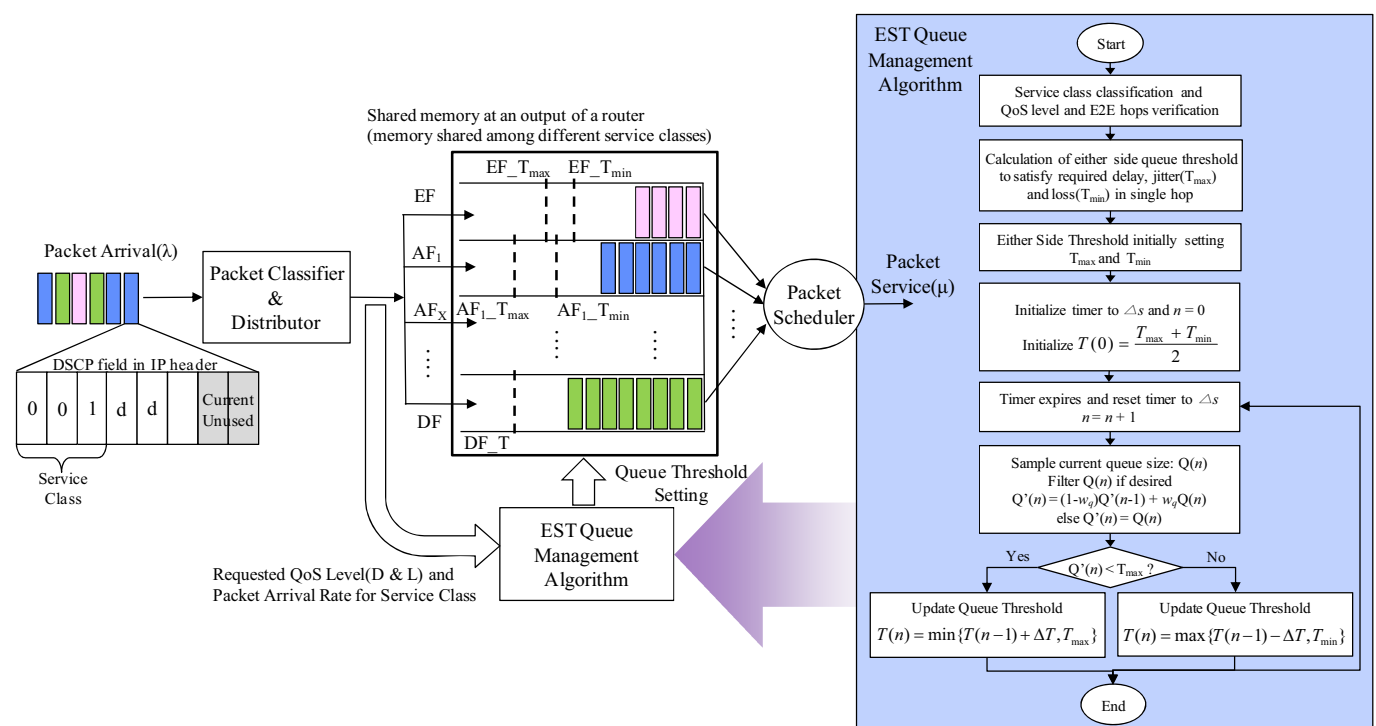


Figure 1. Proposed QoS-aware AQM functional architecture and EST queue management algorithm

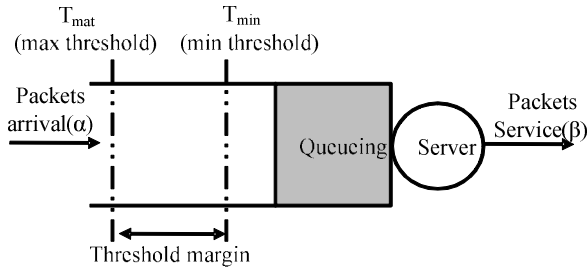


Figure 2. Dynamic threshold control between required QoS parameter(delay, loss rate and jitter) and threshold margin

IV. PERFORMANCE EVALUATIONS AND ANALYSIS BETWEEN QoS AND END-TO-END HOP COUNTS

This section presents a discrete-time queueing model for the performance analysis of proposed QoS-aware AQM method, which includes the arrival model, system model description and performance measures[11].

In the discrete-time queueing system for the proposed QoS-aware AQM, we will assume that departures always take place before arrivals in any unit time (slot). There is a finite waiting room of K packets, including any in service, and a queue threshold is at position TD or TL . The queueing discipline is first-come first-served. If the number of packets in the queue reaches the threshold (TD), then the overflowed packets are discarded. The mean queue length and mean arrival rate are measured over each time window k and the information is used to compute the value of the threshold for the next time window $k+1$ to bound the delay at the required value. Adjusting the threshold, in turn, guarantees the requested QoS level.

We assume that in each state of the source model, the probability that a time-slot contains an arrival is a Poisson process with a parameter that varies according to a Markov process. So we use $M/M/1/K$ queue model to analyze our AQM performance.

In $M/M/1/K$ queueing model, the delay between end hops and packet loss rate can be calculated by using following formulas. In order to perform the steady state analysis of the system, we use following formulas (1) ~ (8). In steady state, the queue state probability is calculated through equation (1).

$$p_k = \begin{cases} p_0 \prod_{i=0}^{k-1} \frac{\lambda}{\mu} = \frac{1 - \lambda/\mu}{1 - (\lambda/\mu)^{K+1}} \left(\frac{\lambda}{\mu} \right)^k & \text{for } 0 \leq k < K \\ 0 & \text{for } k \geq K \end{cases} \quad (1)$$

where p_k is the stationary state probabilities in queue, λ is packet arrival rate, μ is packet service rate.

The aggregate mean queue length (MQL) can be expressed from the equilibrium state probabilities p_k :

$$MQL = \sum_{i=0}^N i P_i \quad (2)$$

The overall mean delay(queueing delay) can be obtained from Little's result for this finite capacity queue as:

$$W = \frac{L}{S} \quad (3)$$

where S is the mean throughput of the discrete-time finite capacity queue given by the fraction of time the server is busy:

$$S = (1 - P_0) \times \beta \quad (4)$$

The blocking probability for an $M/M/1/K$ system with $\rho < 1$ is well-known

$$p_k = \frac{(1 - \rho) \rho^K}{1 - \rho^{K+1}} \quad (5)$$

where $\rho (= \lambda/\mu)$ is utilization.

If packets arrive on the proposed system, ETS algorithm calculates the either side queue thresholds to process service packet based on packet arrival rate, required QoS level in DSCP, and the number of end-to-end hops. First, max queue threshold(TD) is calculated by formula (6) based on required delay in single hop, and min queue threshold(TL) is also calculated by formula (7) based on required packet loss rate in single hop.

$$\frac{\text{Delay}_{e2e}}{\text{hop_counts}} \geq \frac{\sum_{i=0}^{T_D} i P_i}{\mu} \quad (6)$$

$$(1 - L)^h \geq (1 - \frac{1 - \rho}{1 - \rho^{T_L+1}} \rho^{T_L}) \quad (7)$$

$$T_{jitter} \geq T_{\max} - \frac{\text{jitter} \times \mu}{\text{hop_counts}} \quad (8)$$

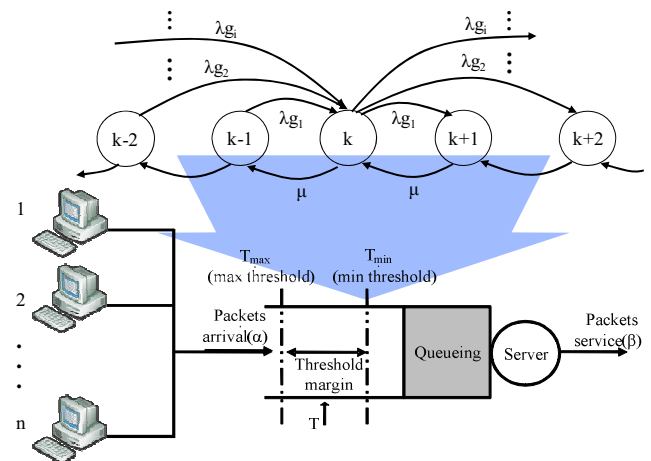


Figure 3. Throughput according to arrival rate and end-to-end hop counts

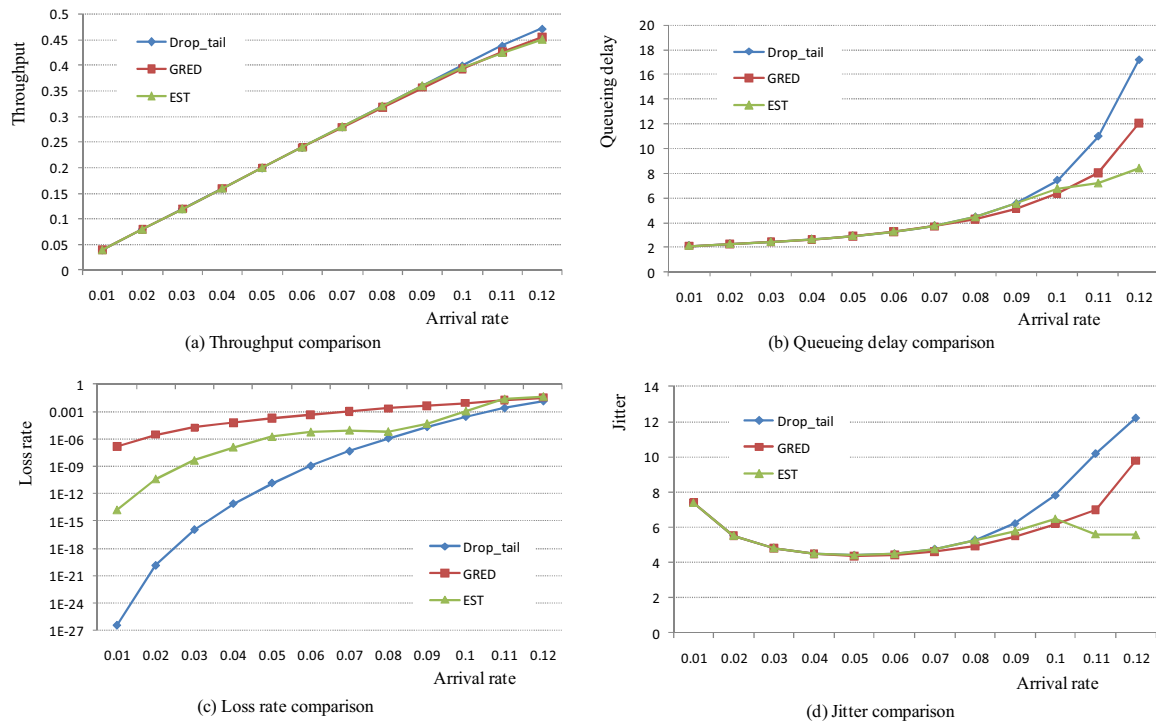


Figure 4. Performance analysis and comparison with legacy AQM(Drop-tail, GRED)

Figure 4 shows the results of performance analysis and comparison with legacy AQM(Drop-tail, GRED). Four types queue is allocated to classify service classes, and to analyze the performance of the proposed algorithm we use the fixed service rate($\mu = 0.5$) and variance arrival rate in queueing model. Figure 4-(a) shows the variation of the throughput about the fixed service rate and variable arrival rate, we know that its results are similar. This shows that the performance is not bad when the proposed algorithm is compared with existing schemes. Figure 4-(b) shows changes of queueing delay in each hop. The nearly same queueing delay occurs in the low arrival rate, while the queueing delay is limited by TD in the high arrival rate. As figure 4-(c), we know that packet loss rate is increased by TD. However, as figure 4-(d), we know that packet processing delay variation can be stabilized by threshold margin between TD and TL in single hop. This means that it is possible to provide multimedia service with stable QoS level.

Figure 5 shows the queue throughput according to arrival rate and the number of end-to-end hops. As mentioned earlier, EST reduces the queue threshold to guarantee QoS in single hop when service delivery distance increases. This increases the packet loss rate and eventually leads to lower throughput. Thus we can arrange the foundation for IPTV CDN(Content Delivery Network) to stably provide multimedia services when network provider configures the network for IPTV service. Therefore, when the proposed EST algorithm is used for streaming multimedia services, the required quality level such as end-to-end delay, jitter and packet loss rate must be maintained.

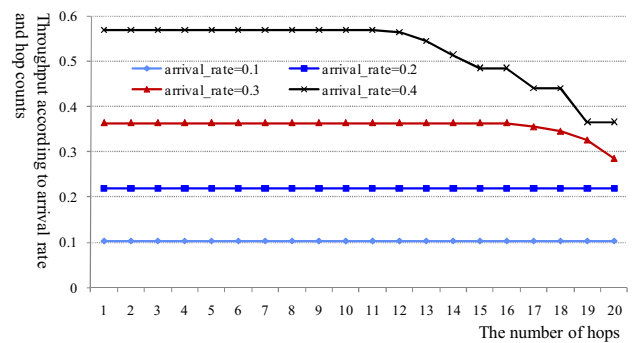


Figure 5. Throughput according to arrival rate and end-to-end hop counts

V. CONCLUSIONS

In this paper, we propose QoS-aware AQM(Active Queue Management) method considering video quality level which service users request in DiffServ network environment. In order to provide multimedia services, QoS provision is essential because real-time multimedia services such as IPTV and VoIP are sensitive to delay and packet loss. Thus we can arrange the foundation for IPTV CDN(Content Delivery Network) to stably provide multimedia services when network provider configures the network for IPTV service. Therefore, when the proposed EST algorithm is used for streaming multimedia services, the required quality level such as end-to-end delay, jitter and packet loss rate must be maintained. In the future, the analysis of the relationship between QoS level

and the number of end-to-end hops is required in terms of IPTV CDN configuration because the long service path increases end-to-end delay and packet loss rate.

ACKNOWLEDGMENT

This work was supported by the project of Integrated Signaling Technology Development for All IP Converged Networks of ETRI(Electronics and Telecommunications Research Institute).

* Correspondence author: Seong Gon Choi (sgchoi@cbnu.ac.kr)

REFERENCES

- [1] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, "An architecture for differentiated services," IETF Network Working Group RFC 2475, Dec. 1998.
- [2] Dong Lin, Robert Morris, "Dynamics of Random Early Detection," *ACM, Computer Communication Review*, Vol. 27, No. 4, October 1997.
- [3] Yang Xiao, Moon Ho Lee, "Nonlinear Control of Active Queue Management for Multiple Bottleneck Network," *IEICE Trans. Commun.*, Vol.E89-B, No.11, pp.3108-3113, Nov. 2006.
- [4] Y. Zheng, M. Lu, Z. Feng, "Performance Evaluation of Adaptive AQM Algorithms in a Variable Bandwidth Network," *IEICE Trans. Commun.*, Vol.E86-B, No.6, pp.2060-2067, Jun. 2003.
- [5] A. A. Akintola, G. A. Aderounmu, L. A. Akanbi, and M. O. Adigun, "Modeling and Performance Analysis of Dynamic Random Early Detection (DRED) Gateway for Congestion Avoidance," *InSITE2005*, June 2005.
- [6] Hussein Abdel-Jaber, Mike Woodward, Fadi Thabtah, Amer Abu-Ali, "Performance evaluation for DRED Discrete-time queueing network analytical model," *Journal of Network and Computer Applications*, September 2007.
- [7] Bing Zheng, Mohammed Atiquzzaman, "DSRED: An Active Queue Management Scheme for Next Generation Networks," *LCN 2000*, November 2000, pp. 242-251.
- [8] Mark Wurtzler, "Analysis and Simulation of Weighted Random Early Detection (WRED) Queues," *EECS 891 Project*, 2002.
- [9] ITU-T Std. Recommendation Y.1541, "Network performance objectives for IP-based services," Feb. 2006.
- [10] DSL Forum Technical Report TR-126, "Triple-play Services Quality of Experience (QoE) Requirements," Dec. 2006.
- [11] L. Guan, M.E. Woodward, I.U. Awan, "Control of queueing delay in a buffer with time-varying arrival rate," *Journal of Computer and System Sciences*, Vol.72, Issue 7, pp.1238-1248, Nov. 2006.
- [12] Pablo J. Argibay-Losada, Andrés Suárez-González, Cándido López-García, Manuel Fernández-Veiga, "A new design for end-to-end proportional loss differentiation in IP networks," *Computer Networks*, Available online, Dec. 2009.
- [13] R. Makkar, I. Lambadaris, J.H. Salim, N. Seddigh, B. Nandy, J. Babiarez, "Empirical Study of Buffer Management Scheme for DiffServ Assured Forwarding PHB," *ICCCN 2000*, October 2000, pp. 632-637.
- [14] Hideyuki Shimonishi, Ichinoshin Maki, Tutomu Murase and Masayuki Murata, "Dynamic Fair Bandwidth Allocation for DiffServ Classes," *ICC 2002*, Vol. 4, April 2002, pp. 2348-2352.
- [15] Changhee Joo, Jaesung Hong and Saewoong Bahk, "Assuring drop probability for delay-insensitive traffic in a differentiated service network," *CCNC 2005*, January 2005, pp. 515-520.
- [16] Xiaolong Jin, Geyong Min, "Performance analysis of priority scheduling mechanisms under heterogeneous network traffic," *Journal of Computer and System Sciences*, Vol. 73, Issue 8, December 2007, pp. 1207-1220.
- [17] Alhussein A. Abouzeid, Sumit Roy, "Modeling random early detection in a differentiated services network," *Computer Networks*, Vol. 40, Issue 4, November 2002, pp. 537-556.