A New Approach for Pricing the Internet Service

Hoon Lee ***Regular Member***

ABSTRACT

In this paper, we propose a method of determining the price for the elastic traffic in the current or future Internet services. First, we investigate the behavior in the consumption of bandwidth of elastic traffic in IP network. Next, we propose a new method to relate the bandwidth usage with the pricing for the elastic traffic, which is based partially or fully on the usage rate of the network bandwidth. Next, we propose an optimal charging function for elastic traffic, which is applicable to any Internet services. Finally, we will illustrate the implication of the work via simple numerical experiments.

Key Words : NGN, Internet Pricing, Usage-based Charging

I. Introduction

At present almost all the Internet customers in Korea pay a fixed amount of charge for the Internet services such as web browsing, file transfer as well as the exchange of E-mails irrespective of the amount of data generated and transferred over the commercial high speed Internet. The leased-line and xDSL services adopt this charging scheme. The customers pay neither transfer charges nor content charges. They only pay the access charge to the Internet service provider irrespective of the usage of the network resources. Thus, this is called a flat charging, and it is a kind of a subscription charge.

The flat charging scheme has reasons in a shared network with best effort service architecture, because there exists no classes or priorities in services. So, there exists a high probability that greedy users can use up the network resources, especially the bandwidth, so that lazy users experience high delays upon their visit to the network when the network is congested with packets generated from the heavy users.

Recently, we could find new applications which require timely delivery of data such as the Internet phone or applications which favor guarantee of minimum amount of bandwidth during data transfer such as web browsing or Intra/Extranet via VPN (Virtual Private Network). To cope with these differences in the requirements for the network performance, differentiated and quality-based service policies have been proposed in the world of network service providers, system manufacturers as well as the standardization organizations such as IETF (Internet Engineering Task Forces). In line with these approaches, the concepts of charging in the current Internet services are undergoing changes toward the usage-based charging [8,12,15,18].

We could find a lot of literature in this field. To name a few for the usage-based pricing, Firdman [6] advocated the necessity of usage-based pricing based on revenue from usage. He gave alternatives such as the usage-sensitive pricing, priority-based pricing, value-added pricing, etc. McKnight [15] presented an overview on pricing the next generation Internet services after flat rate scheme. Blot et al. [3] reported a functional framework called *NetCounter* on charging the individual connection in IP commercial network. Karsten [7] proposed a scheme for a linear price calculation in IETF integrated service

Changwon National University, Changwon, Korea, 641-773 (hoony@changwon.ac.kr)
논문번호 : 030007-0103, 접수일자 : 2003년 1월 6일

architecture advocating that the internal price calculation should be linear, based on resource usage in the network. However, he assumed a reservation-based service differentiation scheme. Pras [18] gave a general discussion on the current state of the art in Internet accounting such as the objectives, protocols and methods for Internet accounting, the trend on standardization and the architecture for implementation.

Lee [10] gave a formal discussion on bandwidth sharing and its impact on user utility and pricing for IP network. In [10], they argued that the network service provider has to levy charge based on the usage of the network bandwidth illustrating the quantitative numerical results for elastic traffic with best effort service architecture. In [11,12], the authors extended the concept of usage-based charging to more specific applications, the VPN services. There, the usage rate charging is advocated by showing some numerical results. Recently, Park [17] advocated the use of two-part tariff scheme from the economic point of view. Lee also advocated the two-part scheme even for the current best effort Internet services in [13].

This paper is an extension of the works of [11,12,13]. We argue that the usage-based charging has to be tuned to the objectives of the network operator's purpose as well as the provisioning of the bandwidth to the users. That is, when the bandwidth is in the form of reservation the charging has the form of fixed charging, whereas if the bandwidth is completely shared the charging scheme has the complete usage-based charging. However, neither scheme is suited to the elastic Internet services. So, this paper proposes an intermediate approach between the flat charging and the complete usage-based charging. We discuss this approach in more systematic way, and we try to give a formal framework and quantitative discussion for charging the elastic traffic.

Before entering into the discussion for the charging the customers, we have to differentiate the concept between pricing, charging and billing: Pricing is the process of determining a cost per unit bandwidth the connection uses, whereas charging is the process of translating the customer's bandwidth usage information into an amount of money the customer has to pay. Finally, billing is the procedure of issuing the bill to the customer [2]. This paper discusses methods to determine the first two ones: pricing and charging.

This paper is composed as follows: In Section II we describe the attributes of elastic traffic from the bandwidth usage. In Section III we propose a new method to determine the price curve under the flat & residual pricing scheme for elastic traffic. Section IV gives the results for numerical experiments, where the implication of the proposed methods is shown with graphs. In Section V we summarize the paper and give some comments on further research areas.

II. Elastic Traffic and Implication to Pricing

A. Elastic Traffic

It is well known that the elastic traffic (ET) is named from the property of the elastic services such as files of data, text, picture, WWW pages and other documents that is transferred from one point to the other point in the network in that it can cope with a non-guaranteed, variable throughput. Thus, ET can tolerate packet delays gracefully and it would rather wait for reception of traffic in the correct order, without losses. So, the traffic in elastic services needs a large buffer and an elastic bandwidth allocation mechanism like TCP in IP network. Examples of elastic service include traditional data services such as remote terminal, file transfer, name services and electronic mail. Note that the attribute of elastic traffic is very similar to that of ABR or UBR (Unspecified Bit Rate) traffic in ATM network. As such, the Internet user and IP network can negotiate the transfer of packets via two different methods: minimum bandwidth with plus-alpha for ABR-like-ET (for simplicity, we call it ABR-ET) and no bandwidth contraction for UBR-like-ET (UBR-ET).

As to ABR-ET, the specification for QoS (Quality of Service) is expressed in terms of minimum throughput, which is represented by the file size divided by transfer time [14]. Minimum throughput of elastic traffic is synonymous with the minimum bandwidth the network has to provide to the traffic, which is contracted with the customer before traffic is transferred. The contracted minimum bandwidth (CMB), can be allocated (in the form of reservation) a priori or statistically guaranteed, and an additional bandwidth (we had describe it as "plus-alpha") is provided by network (especially, in Next Generation Network) if there is any available bandwidth unused by other connections in the network.

For UBR-ET, no contraction with respect to the bandwidth usage is needed in order to transfer the data, and so no bandwidth is reserved to the connection with UBR-ET, and packet transfer occurs only if there exists available bandwidth not used by other high priority traffic, which is called the best effort traffic of the current Internet.

At any instant, a customer may generate traffic less than or greater than the CMB. If there is sufficient bandwidth in the link, the network can carry out all the traffic in excess of the contracted value, otherwise some packets are forced to wait in the queue for later transmission. Fig.1 illustrates a rough graph of the behavior of elastic traffic as a function of time. The solid flat line denotes the maximum link capacity and the dotted line denotes the agreed bandwidth such as CMB, so M is the maximum rate the users can use the bandwidth, whereas μ is the CMB. Because the variation of the traffic volume generated from a connection for Internet access is very harsh [4], there may happen a case in which a connection can or can't use the agreed bandwidth. So, the traffic curve goes up and down across the CMB. Of course, the traffic rate of a connection should not exceed the maximum link capacity M at any time.

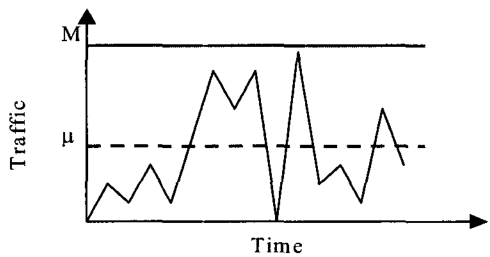


Fig.1. Behavior of elastic traffic

B. Flat Pricing and Its Drawbacks

Flat pricing levies a fixed charge to a connection irrespective of the connection's actual usage. For example, a network can provide a connection with at most M amount of bandwidth, and it levies a fixed amount of Π to a connection whether a connection uses it or not. Fig.2 illustrates the flat charge.

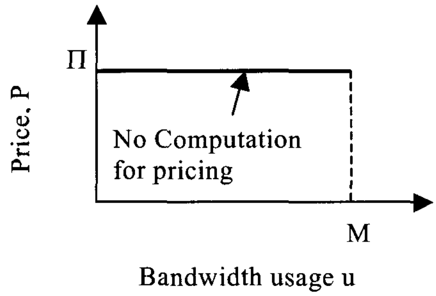


Fig.2. Concept of flat pricing

So, from traffic point of view, the flat pricing is not suited to ABR-ET and UBR-ET from the following reasons. First, FP (Flat pricing) results in unfairness between heavy users and light users. Second, moral hazard exists between users, because the charge is not sensitive to the actual usage. Moral hazard results in an excessive use of network bandwidth. Under the network congestion, heavy users do damage to light users in terms of delayed access and/or delayed packet transfer, which results in a negative network externality [17]. Finally, FP gives no incentive to ISP's effort for the differentiation or upgrade of the quality of services. The flat pricing is best suited to the VPN type services in which a customer can use the bandwidth up to full volume at any time. From this discussion, we find that a new pricing scheme that is suited to the elastic traffic has to be developed.

Ⅲ. Pricing and Charging the Elastic Traffic

The price for a packet transfer over the current IP network in Korea is based on the fixed rate and the only metric for charging is the speed of an access link either the user is accessed via xDSL or Leased-line [12]. As we may have a lot of tables for the unit charge with respect to the use of the bandwidth for IP network depending on the network service operators, we use the current charge for ADSL services, for example, X-Premium(We used an anonymous name X: Maximum speed of down link: 8Mbps) and X-Lite (Maximum speed of down link: 1.5Mbps) charge a fixed monthly price of Internet access at 33$ and 25$, respectively [9].

Recently, MIC (Ministry of Information & Communication) of Korea announced a recommendation for the contraction of SLA (Service level agreement) for the high-speed Internet services of Korea [5]. Table 1 summarized an example of SLA in terms of the minimum bandwidth. The SLA can be interpreted as a means for the enforcement of the minimum bandwidth, which corresponds to CMB in our work, that has to be guaranteed to a user. This is similar to the concept of the AF (assured forwarding) service of DiffServ (Differentiated service) architecture in the future Internet services. Therefore, we assume that an ISP is obliged to guarantee a minimum bandwidth contracted by SLA to a user under the proposed pricing architecture. Let us call this kind of service architecture as an ABR-like service in the current Internet service. There may exist a number of methods to guarantee a minimum bandwidth under the current Internet service. An alternative would be over-provisioning of the network. However, this is beyond the scope of the present work.

Table 1. Required bandwidth for Internet access services

Access service	Maximum bandwidth (Mbps)	Minimum Bandwidth (Mbps)
X-Premium	8	1
X-Lite	1.5	0.5

From the movement toward the enforcement of SLA as well as the discussion in Section II we can infer that a new pricing scheme that takes into account the additional usage exceeding the CMB has to be devised in any manner. To that purpose, we have to devise a method to translate the bandwidth used by the elastic traffic into price. The pricing can be divided into two cases: price for ABR-like elastic traffic (For simplicity, we called it ABR-ET) or price for UBR-like elastic traffic (UBR-ET). The former discriminates the value of the minimum bandwidth and additional bandwidth used by availability of the network at any instant, whereas the latter cares only the usage of bandwidth. In this work, we focus on the former one.

A. Flat & Residual Price for Elastic Traffic

First, let us remind the attribute of ABR-ET. For ABR-ET, the minimum bandwidth, denoted by μ in the previous section, can be looked upon as a kind of reserved bandwidth within which a customer can use at his/her own will, whereas the additional bandwidth, called as *plus-alpha*, is the excess bandwidth that can be used by a customer if there exists additional bandwidth the network operator can provide to the customer. For ABR-ET, let us levy a charge in a fixed and residual way such that a fixed amount of charge is levied unless the usage exceeds a minimum bandwidth. When the usage exceeds a minimum bandwidth, additional charge is levied in proportion to the usage. Let us call it an FRP (flat & residual pricing) scheme. The basic assumption behind the concept of flat & residual pricing is that the customer pays a fixed amount of charge for the minimum bandwidth whether he/she uses it or not, whereas the implicit agreement in the provision of additional bandwidth is that the customer is ready to pay additional price for the additional bandwidth provided by the network. Let us call the price of residual bandwidth to be the residual price. From these discussions, we can find that the concept of flat & residual pricing is well suited to the purpose of bandwidth provision of ABR-ET.

We can easily find the similarity in levying the price for ABR-ET in an IP world and the concept of residual price in an ATM world. CANCAN [16] announced a recommendation for residual price for

ABR traffic based on the Committed Information Rate (CIR) because the network operator reserves a minimum bandwidth relevant to CIR to an ABR connection. However, to the best of author's knowledge, we do not know the scheme is implemented in the real field. The concept of flat & residual pricing is illustrated in Fig.3. As we can find from Fig.3, the customer pays a fixed amount of charge irrespective of the usage of the network bandwidth so far as the usage rate does not exceed the predefined minimum bandwidth μ. So, a minute computation for pricing is not carried out by the network operator. However, the customer has to pay additional charge for the usage of the bandwidth in an amount he/she used in addition to μ when the bandwidth usage is greater than μ, where an intensive price computation is carried out.

There may exist various ways for levying prices differently to the usage of the bandwidth (see [10,11] and references therein). The typical curve for the residual price is a linear function connecting the two points of μ and M.

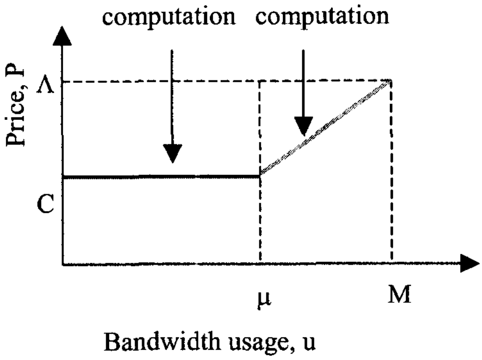


Fig.3. Concept of fixed & residual pricing

The slope of residual price, which is denoted by σ in eq.(1), is determined by the pricing policy of the network company. We assume that maximum price that can be levied is for the usage of maximum allowed bandwidth of M. Then, σ is given by

$$\sigma = \frac{\Delta P}{\Delta u} = \frac{\Lambda - C}{M - \mu}. \qquad (1)$$

The remained problem is to determine the values Λ and C. To the best of author's knowledge, there does not exist a well-defined way to determine the values Λ and C. They may be determined by the network company's own pricing policy or determined purely from the market forces. Let us propose a new approach for determining the values Λ and C in this work. Our proposition is to compute Λ and C by adopting the price and usage graph as shown in Fig.4. The proposed price and usage graph is introduced from the idea of the supply and demand graph in standard Economics, which is qualitatively described before in [17].

Let us assume that, in an ideal case, a customer can use a maximum amount of network bandwidth M and he/she pays a charge Π under the flat charging scheme. Note that this is just one of several assumptions on the pricing scheme for the flat rate charging (Some may levy a flat charge based on the peak rate, while others may levy based on the average rate. Let us assume the former one in this work). Under the flat charging scheme, the amount of flat charge one has to pay is equal to an area determined by the rectangular oΠEM whether he/she uses it fully or not.

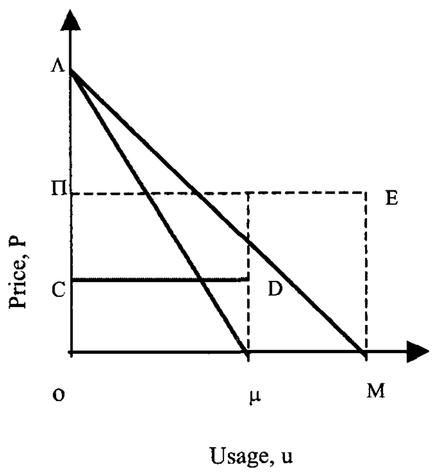


Fig.4. Price and usage graph

On the other hand, the price and usage graph implies that a user uses no bandwidth if the price is Λ, whereas a user uses M amount of bandwidth if the price is zero. In between the prices $[0, \Lambda]$ a user uses the bandwidth in reverse-proportion to the price shown in Fig.4. The area of a triangle formed by a triangle oΛM is the utility, where utility is defined to be the value of a service that a customer can obtain for an arbitrary price .

The ideal situation is the case when the utility, which is the value of service, is equal to the charge a customer pays. Optimal utility is realized at the point obtained by equalizing the two areas made from triangle oΛM and the rectangle oΠEM. From this fact, we obtain the following result:

$$\Lambda = 2\Pi \cdot \tag{2}$$

Now it remains to determine the price C. How much do we have to levy a price for a customer that uses a bandwidth less than or equal to μ? To that purpose, let us remind the current flat charging scheme. That is, the amount of charge levied to a customer is C irrespective of the actual usage unless the usage does not exceed the threshold μ, the minimum bandwidth guaranteed anytime. Therefore, we if compute the proportional charge the current flat charging scheme, we obtain the following formula for C.

$$C = \frac{\mu}{M}\Pi \; . \tag{3}$$

B. Charging Function

Let us define some variables and parameters for the calculation of the charge levied to a connection. Let v(t) be the traffic volume (unit: bits) which is generated by the customer at time t. Let T be the time interval of the measurement of the network usage. Then, the usage rate of the network bandwidth (in a unit of bit per second) in an i-th measurement period, u_i, i=1,2,...,N, is defined by the amount of bits transmitted in the network during a certain time period, and it is given by

$$u_i = \frac{1}{T}\int_0^T v(t)dt. \tag{4}$$

The duration of measurement T indicates the specification of the monitoring frequency and it is closely related to the speed of the network link, which is also related to delay characteristics expected from the network and the amount of bandwidth allowed by the network. The more sensitive the application is, the higher the monitoring frequency should be. T is also related to the accuracy of the measurement. In [1] a discussion on this value is given in a qualitative manner for three traffic classes: very frequent, frequent, and unspecified duration of measurement. The elastic traffic is categorized as an unspecified duration for the measurement. Even for the elastic traffic with unspecified duration, the period of monitoring has a close relationship with the accuracy of the charging. However, we assume that the monitoring period is much shorter than the connection duration, from which we can accumulate enough data for the estimation of usage rate of the connection. Finally, let us take the average u of u_i, i=1,2,...,N, which is given as follows:

$$u = \frac{1}{N}\sum_{i=1}^{N} u_i \; . \tag{5}$$

Note that eq.(5) is the average usage rate of a customer. One point we have to know is that the above result considers only one-way traffic. The usage rate of both-way traffic for the asymmetric link can be obtained in the same way [12]. If we have a usage rate u of network bandwidth, we can relate the usage rate into a charging function F(u), and it is given as follows:

$$\begin{aligned} F(u) &= C, if u \langle \mu, \\ F(u) &= C + \sigma \times (u - \mu), else \end{aligned} \tag{6}$$

Ⅳ. Numerical Results

From Table1 we can find that the minimum (or maximum) bandwidth corresponds to our minimum (or maximum) bandwidth. For the purpose of easy illustration, let us rewrite the relationship between M and μ as follows:

$$\mu = 0.33 \times M, X - Lite$$
$$\mu = 0.125 \times M, X - Premium \qquad (7)$$

From the equations (6)-(7) and the price for each access service given in Table 1, we obtain the results shown in Table 2.

In order to represent the result in the form of Fig.3, let us draw a graph for each case. Fig.5 illustrates the charge (unit:$) as function of bandwidth usage (unit:Mbps) levied to a customer subscribed to a X-Lite service, whereas Fig.6 illustrates the result for the charge (unit:$) as function of bandwidth usage (unit:Mbps) levied to a customer subscribed to a X-Premium service.

From Fig.5 we can find that customers of X-Lite service with bandwidth usage less than 0.5Mbps can be benefited in an amount of 16.67$ per month, whereas customers with bandwidth usage greater than 0.9Mbps has to pay more charge than the current fixed charge of 25$, which is proportional to an amount of their actual usage. On the other hand, customers of X-Premium service with bandwidth usage less than 1Mbps can be benefited in an amount of 28.87$ per month, whereas customers with bandwidth usage greater than 4.27Mbps has to pay more charge than the current fixed charge of 33$, which is also proportional to an amount of their actual usage.

Table 2. Maximum and minimum price for each case

Access service	Max. Charge ($)	Min. Charge ($)
X-Premium	66	4.13
X-Lite	50	8.33

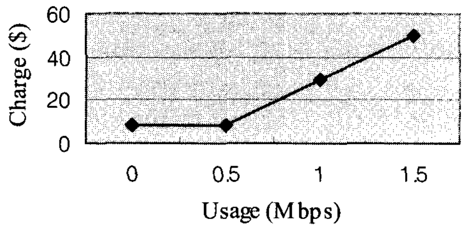


Fig.5. Charge for FRP (X-Lite)

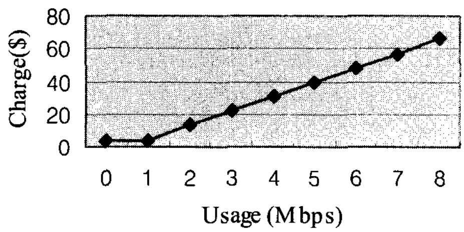


Fig.6. Charge for FRP (X-Premium)

From Fig.6 one can find that customers under the flat charging scheme of the present pay the minimum charge of 33$ irrespective of the actual usage. With the newly proposed charging scheme the network company can collect more charge (up to 66$!).

One point we have to point out at this stage is the following finding: There exists a contradiction in the charge when the usage rate is lower than 0.75Mbps. That is, a user of X-Lite pays more charge than a user of X-Premium if the usage is the same and it is lower than 0.75Mbps. However, this is due to the SLA that has been set up between the customer and the ISP, and therefore it is beyond the scope of our study. This problem has to be revisited in the future study.

Ⅴ. Conclusions

In this paper we proposed a method for determining the charges for the elastic traffic in the IP network. We assumed services with elastic traffic for the Internet services which corresponds to the typical AF services in DiffServ service architecture in the future Internet.

We argued that the flat charging has to be replaced with the proposed charging with FRP (flat & residual pricing) scheme due to several reasons described in the paper by proposing a systematic method for determination of the upper limit (threshold value) for flat charging and derived formulas for the maximum charge and the minimum charge by using the concept of the price and usage graph. Via simple numerical experiments we could provide an intuition for the users to consider price for transferring the elastic traffic over IP network.

This work is just a first step in the quantitative research for the charging the Internet based on the usage of the bandwidth. Thus, there remain lots of problems, as such areas for the further study would be wealthy: the determination of the optimal price for the use of links with different speeds, sophistication in the method for monitoring the traffic usage, time granularity for metering, etc. Our next research will be concentrated on these areas.

References

[1] O. Aboul-Magd and B. Jamoussi, "QoS and service interworking using CR-LDP", IEEE Communications Magazine May 2001.

[2] A. Pras et al., "Internet accounting", IEEE Communications Magazine May 2001.

[3] S.M. Blot, et al., "User-level billing and accounting in IP networks", Bell Labs Technical Journal, Oct-Dec. 1999.

[4] J. Charzinski, "Problems of elastic traffic admission control in an HTTP scenario", Proceedings of IWQoS2001.

[5] "On the quality regulation of high-speed Internet services of Korea", http://www/chosun.com/

[6] E. Firdman, "Rx for the Internet: usage-based pricing", Data Communications on the web, http://www.data.com/business_case/rx_internet.html

[7] M. Karsten et al., "Provider-oriented linear price calculation for Integrated services", Proceedings of IWQoS'99, 1999.

[8] F.P. Kelly, "Charging and accounting for bursty connections", Internet Economics, ed. L.W. McKnight and J.P. Bailey (eds) MIT Press, 1996.

[9] The price for Megapass services for the KT, http://www.kt.co.kr, as of July 10, 2002.

[10] Hoon Lee, "Bandwidth sharing, utility and pricing for the best effort Internet services", Proceedings of the 3rd Int'l conference on advanced communication technology (ICACT2001), February 8-10, 2001, Muju, Korea.

[11] Hoon Lee, "A usage based charging in IP-VPN", NCS2000, December 4-6, 2000.

[12] Hoon Lee and J.-H. Eom, "A usage-rate based charging for QoS-free traffic in IP-VPN", Proceedings of the ICC2001, June 11-15, 2001, Helsinki, Finland.

[13] Hoon Lee et al., "Differentiated charging for elastic traffic", Journal of KICS, Vol.26, No.12, December 2001.

[14] K. Lindberger, " Balancing Quality of Service, pricing and utilization in multi-service networks with Stream and Elastic Traffic", ITC 16, Edinburgh, June 1999.

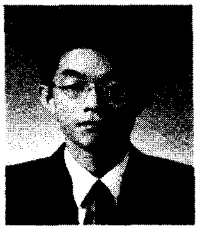
[15] Lee W. McKnight and J. Boroumand, " Pricing Internet services: after flat rate", Telecommunications Policy 24 (2000).

[16] D. Morris and V. Pronk Ed., "ATM charging Recommendations II, Final Recommendations, CANCAN Deliverables, August 1998.

[17] J.S. Park and J.H. Lee, "A study on problems of the high-speed Internet pricing in Korea and its optimal solution", Summer Conference of KICS 2002, July 11-13, 2002, Muju, Korea.

[18] A. Pras, et al., "Internet accounting", IEEE Communications Magazine, May 2001.

Hoon Lee regular member



February 1984: B.E. degree in Electronics at Kyungbook National University

February 1986: M.E. degree in Communications at Kyungbook National University

March 1996: Ph.D. degree in Electrical & Communication Engineering from Tohoku University, Sendai, Japan

February 1986 - February 2001: KT R&D Group

March 2001 ~ : Assistant Professor of Changwon National University

Research interests: Teletraffic engineering, Network design, performance analysis and provision of QoS, and pricing for high speed telecommunication networks.

Dr. Lee is a member of IEEE, KICS and IEEK.