파테토적 개선을 위한 네트워크 운영
Making your network Pareto-improving

김용재

경영정보학과, 경영대학
건국대학교,서울,광진구 화양동 1
Tel: +82-2-450-3608, Fax: +82-2-3436-6610, E-mail: yjaekim@gmail.com

Abstract

새로운 인터넷 응용 프로그램들이 등장하면서 망 중립성에 대한 논의는 계속되고 있다. 백본망 제공자의 입장에서 자신의 망에 부과를 주는 응용프로그램들 때문에 비용을 추가로 지불해야 한다는 사실은 오래된 문제이긴 하지만 새롭게 주목을 받고 있다. 많은 연구가 인터넷혼잡에 대한 문제 해결을 위해 가격체계를 도입해야 한다는 사실을 밝혔지만, 종량제 도입시도는 사회적 저항을 불러일으켰다. 하지만 IPTV나 VoIP와 같은 응용프로그램들이 요구하는 품질을 맞추기 위해서 망 투자는 계속되어야 하지만, 정치적 사회적 반대로 인해 사업자의 선택은 제한되어 있다. 한편 인터넷 버전 6를 위한 구축비용이 엄청나며, 새로운 표준 도입의 기술적 문제 때문에 많은 사업자들이 주저하고 있다. 이 논문은 우선대기열을 사용해서, 망 사업자가 무차별, 무계급(non-priority)에 기반한 인터넷에서 차별화된(prioritized) 인터넷 망 서비스를 제공할 때 일어날 수 있는 문제를 살펴본다. 현재의 서비스 수준보다 더 좋거나 같은 수준의 (즉 Pareto-improving) 서비스를 차별화된 가격체계에서 제공할 수 있을지에 대한 가능성과 문제점들을 살펴 보면서, 망 사업자와 사용자 모두에게 유익할 수 있는 방안을 검토해 본다.

Keywords:

Pareto-improving transition; Priority queue; Revenue Management

1. Introduction

In recent years, the “net neutrality” has surfaced only to rekindle the interest in preferential treatments of heterogeneous Internet traffic. With Internet applications such as IPTV taking up larger chunks of Internet traffic, a few network carriers have implicitly considered or planned differential Internet services to control the unwieldy traffic growth. However, such moves have been met with public protests and regulatory concerns both in the US and in Korea. They fret that prioritized networks would drive out less fortunate users and induce discrimination, directly opposing the free spirit of the original Internet. Joining the camp are Internet portals and service providers that insist that the carriers should remain neutral to the Internet traffic. Although incidents such as carriers refusing VoIP traffic were quickly noticed and resolved by regulating authorities, whether the net neutrality should be maintained or not will be recurring in the near future because the Internet is still rapidly evolving and it is hard to expect what it will take to get this matter settled (Kwak 2006; Laxton 2006; Hahn and Wallsten 2006).

Furthermore upgrading the current Internet on a global basis has been found extremely difficult to achieve for lack of compatible network protocols and exorbitant upgrading costs. Complicating the situation is the fact that the market is driven by network carriers confined in a specific country, regulated by different laws and governmental agencies, and under different market conditions. As a result, carriers may be left with few options to exercise to change their status quo.

The primary goal of this research is to illustrate that a network carrier still has a way to improve its operation with no one worse off. In other words, a carrier can improve its operational efficiency without hurting any customers, thereby making itself immune from net neutrality controversy. To illustrate such potentials, we use a multi-class M/G/1 queuing model approximating a queueing network owned by a network carrier who is a monopolist maximizing social net value defined as the sum of values of finished jobs minus the total delay costs.

Combating network delay has a long history and a substantial body of research suggests that networks better be operated as prioritized (multi-class) rather than non-priority (single-class) systems. For example, empirical studies (Edel, R. and Varaiya, P. 1999; Dovrolis and Ramanathan 1999; Cochi 1993) have demonstrated the needs and benefits of prioritized operation. In the economics literature, it was Pigou (1920) who first studied the queueing delay effect in a congestible resource and Naor (1969) advanced the idea in different contexts. The model used in this paper largely borrows from Mendelson and Whang (1990) who showed that priority- and time-dependent pricing induces individual users to select a correct priority class and that the resulting state is both optimal and incentive-compatible.

In the area of network management, although there have been a few papers that address transition issues similar to ours (for example, see Cochi et al. 1993), none of them have examined welfare aspects of the transition from a non-priority to a prioritized system systematically, which is the primary focus of this paper.

The plan of this presentation is as follows. First, we state fundamental theorems showing that the total delay cost decreases after the transition and, under fixed system traffic among the user classes, the full price (viz., social marginal cost) faced by individual jobs also diminishes. Unfortunately a Pareto-improving transition is not always achieved and thus we consider a theoretical optimization model that incorporates transition cost, an externality cost capturing the impact of individual user behavior after the transition. However, the computational complexity solving the problem motivates us to develop an efficient heuristic. Through a genetic algorithm, the quality of post-transition solutions and the quality of solutions generated by the heuristic are examined. Simulation results demonstrate that the initial post-transition solutions are typically Pareto-improving. For non Pareto-improving solutions, we compare the heuristic with a genetic algorithm that balances an objective (either net social value or revenue) with satisfaction of Pareto-improving and incentive-compatibility constraints. Results of a simulation that tests the quality of solutions generated by the heuristic and the genetic algorithm are followed.

2. Pre- and Post-transition Description

We assume that arrival of jobs to the network is governed by N independent Poisson processes, where class-i jobs arrive at rate λ_i. Following Mendelson and Whang (1990), let $V_i(\lambda_i)$ denote the contribution of class-i jobs when the class's arrival rate to the system is λ_i where $V_i(\lambda_i)$ is monotone increasing, continuously differentiable, and strictly concave. The marginal class-i user's valuation of a completed job will be $\partial V_i(\lambda_i)/\partial\lambda_i$ and the social value function, $V(\underline{\lambda})$, is defined as the sum of individual classes, i.e. $V(\underline{\lambda}) = \sum_{k=1}^{N} V_k(\lambda_k)$ where $\underline{\lambda} = (\lambda_1, .., \lambda_N)$.

Jobs are served on FCFS basis and class-i service time distribution is generally distributed with mean c_i and second moment $c_i^{(2)}$. If the network is run as a non-priority M/G/1, the expected sojourn time of class-i of non-priority system, ST_i^1, is $ST_i^1(\underline{\lambda}) = \Lambda_N / \overline{S}_N + c_i$ where $S_i = \sum_{k=1}^{i} \lambda_k c_k$, $\overline{S}_i = 1 - S_i$, and $\Lambda_i = \sum_{k=1}^{i} \lambda_k c_k^{(2)}/2$ (Kleinrock 1976). If v_i is the delay cost per unit time for a class-i user, the total delay cost is then defined as $\sum_{k=1}^{N} v_k \lambda_k ST_k^1(\underline{\lambda})$ and the net value maximizing problem is to choose $\underline{\lambda} = (\lambda_1, .., \lambda_N)$ to maximize $\sum_{k=1}^{N}\left(V_k(\lambda_k) - v_k\lambda_k(\Lambda_N / \overline{S}_N + c_k)\right)$ so that the optimal arrival rate vector $\underline{\lambda}^+ = (\lambda_1^+, .., \lambda_N^+)$ satisfies the first-order conditions $\partial V_i(\lambda_i)/\partial\lambda_i = v_i ST_i^1(\underline{\lambda}) + \sum_{k=1}^{N} v_k\lambda_k\, \partial ST_i^1(\underline{\lambda})/\partial\lambda_i$.

Assuming that at least one internal solution exists under the demand relation, the class-i externality cost be equated with the optimal price, i.e., $p_i^+(\underline{\lambda}^+) = \sum_{k=1}^{N} v_k\lambda_k\, \partial ST_i^1(\underline{\lambda}^+)/\partial\lambda_i$. However the price is not incentive-compatible and we need a time-dependent pricing $p^1(t) = \left(\frac{t^2}{2\overline{S}_N^+} + \frac{\Lambda_N^+ t}{\overline{S}_N^{+\,2}}\right)\sum_{k=1}^{N} v_k\lambda_k^+$ where $\overline{S}_i^+ = 1 - \sum_{k=1}^{i}\lambda_k^+ c_k$ and $\Lambda_i^+ = \sum_{k=1}^{i}\lambda_k^+ c_k^{(2)}/2$. (Kim and Mannino 2002).

Now suppose that the network carrier decides to transform the non-priority system to a nonpreemptive priority M/G/1 and to apply the v_i/c_i priority assignment rule. The expected sojourn time of class-i, ST_i is $ST_i(\underline{\lambda}) = \Lambda_N / \overline{S}_i\overline{S}_{i-1} + c_i$ (Keinlock 1976) and the total delay cost is given as $TC(\underline{\lambda}) = \sum_{k=1}^{N} v_k\lambda_k\left(\Lambda_N/\overline{S}_k\overline{S}_{k-1} + c_k\right)$. The net value maximizing problem for the prioritized system is to choose $\underline{\lambda}$ to maximize $\sum_{k=1}^{N}\left(V_k(\lambda_k) - v_k\lambda_k ST_i(\underline{\lambda})\right)$.

Assume that the optimal arrival rate vector $\underline{\lambda}^{\circ} = (\lambda_1^{\circ}, .., \lambda_N^{\circ})$, solving the first-order conditions $\partial V(\underline{\lambda})/\partial\lambda_i = v_i ST_i(\underline{\lambda}) +$

$$\left\{\sum_{k=1}^{N}\frac{v_k\lambda_k c_i^{(2)}}{2\overline{S}_{k-1}\overline{S}_k} + \frac{v_i\lambda_i c_i\Lambda_N}{\overline{S}_{i-1}\overline{S}_i^2} + \sum_{k=i+1}^{N}\left(\frac{v_k\lambda_k c_i\Lambda_N}{\overline{S}_{k-1}^2\overline{S}_{kj}} + \frac{v_k\lambda_k c_i\Lambda_N}{\overline{S}_{k-1}\overline{S}_k^2}\right)\right\}$$

, exists. Then the optimal price for class i, $p_i(\underline{\lambda}^{\circ})$, will be set as

$$p_i(\underline{\lambda}^{\circ}) = \begin{array}{l} c_i\left(\frac{v_i\lambda_i^{\circ}\Lambda_N^{\circ}}{\overline{S}_{i-1}^{\circ}\overline{S}_i^{\circ\,2}} + \sum_{k=i+1}^{N}\left(\frac{\Lambda_N^{\circ}v_k\lambda_k^{\circ}}{\overline{S}_k^{\circ}\overline{S}_{k-1}^{\circ\,2}} + \frac{\Lambda_N^{\circ}v_k\lambda_k^{\circ}}{\overline{S}_{k-1}^{\circ}\overline{S}_k^{\circ\,2}}\right)\right) \\ + \frac{c_i^{(2)}}{2}\left(\sum_{k=1}^{N}\frac{v_k\lambda_k^{\circ}}{\overline{S}_{k-1}^{\circ}\overline{S}_k^{\circ}}\right) \end{array}$$

where $\overline{S}_i^{\circ} = 1 - \sum_{k=1}^{i}\lambda_k^{\circ}c_k$ and $\Lambda_i^{\circ} = \sum_{k=1}^{i}\lambda_k^{\circ}c_k^{(2)}/2$ respectively.

As Mendelson and Whang (1990) showed, the above optimal prices are not incentive-compatible and a priority- and time-dependent pricing scheme should be used:

$$p_i(t) = t\left(\frac{v_i \lambda_i^{\square} \Lambda_N^{\square}}{\overline{S}_{i-1}^{\square} \overline{S}_i^{\square 2}} + \sum_{k=i+1}^{N}\left(\frac{\Lambda_N^{\square} v_k \lambda_k^{\square}}{\overline{S}_k^{\square} \overline{S}_{k-1}^{\square\ 2}} + \frac{\Lambda_N^{\square} v_k \lambda_k^{\square}}{\overline{S}_{k-1}^{\square} \overline{S}_k^{\square 2}}\right)\right)..$$

$$+\frac{t^2}{2}\left(\sum_{k=1}^{N}\frac{v_k \lambda_k^{\square}}{\overline{S}_{k-1}^{\square} \overline{S}_k^{\square}}\right)$$

3. Fundamental Transition Theorems

THEOREM 1 (Delay Costs)

Given a fixed arrival rate vector $(\lambda_1,..,\lambda_N)$, *the transition from* non-priority M/G/1 *to* nonpreemptive priority M/G/1 *results in a lower total delay cost. In other words,* $\sum_{k=1}^{N} v_k \lambda_k ST_k(\underline{\lambda}) < \sum_{k=1}^{N} v_k \lambda_k ST_k^1(\underline{\lambda})$ ·

Therefore the following corollary holds.

COROLLARY 1 (Increased Net System Value)

The transition from nonpreemptive non-priority M/G/1 *to nonpreemptive* M/G/1 *will result in the net social value increased.*

THEOREM 2 (Social Marginal Costs)

For a fixed traffic $\underline{\lambda}$, *the sum of price and sojourn time cost of nonpreemptive* M/G/1 *is less than that of non-priority* M/G/1. *In other words,*

$$v_i ST_i(\underline{\lambda}) + \sum_{k=1}^{N}\frac{v_k \lambda_k c_i^{(2)}}{2\overline{S}_{k-1}\overline{S}_k} + \frac{v_i \lambda_i c_i \Lambda_N}{\overline{S}_{i-1}\overline{S}_i^2}$$

$$+c_i \sum_{k=i+1}^{N}\left(\frac{v_k \lambda_k \Lambda_N}{\overline{S}_{k-1}^{\ 2}\overline{S}_k} + \frac{v_k \lambda_k \Lambda_N}{\overline{S}_{k-1}\overline{S}_k^2}\right) <$$

$$v_i ST_i^1(\underline{\lambda}) + \left(\frac{c_i^{(2)}}{2\overline{S}_N} + \frac{\Lambda_N c_i}{\overline{S}_N^2}\right)\sum_{k=1}^{N} v_k \lambda_k \cdot$$

However we cannot assume that the arrival rate vector remains unchanged.

EXAMPLE 1:[1]

Consider a system that has two classes with $V_1(\lambda_1) = 9 - 20\lambda_1$ on $\lambda_1 \in [0,0.45]$, and $V_2(\lambda_2) = 12 - 30\lambda_2$ on $\lambda_2 \in [0,0.4]$. Let $v_1 = 2$, $v_2 = 1$, $c_1 = 0.1$, $c_2 = 2$, $c_1^{(2)} = 2c_1^2$, and $c_2^{(2)} = 2c_2^2$ respectively, indicating that the class-1 users are more sensitive to delays. By solving the two first order conditions $\partial V_i(\lambda_i)/\partial\lambda_i = v_i ST_i^1(\underline{\lambda}) + \sum_{k=1}^{N} v_k \lambda_k \, \partial ST_i^1(\underline{\lambda})/\partial\lambda_i$ and

$$\partial V(\underline{\lambda})/\partial\lambda_i = v_i ST_i(\underline{\lambda})$$

$$+\left\{\sum_{k=1}^{N}\frac{v_k \lambda_k c_i^{(2)}}{2\overline{S}_{k-1}\overline{S}_k} + \frac{v_i \lambda_i c_i \Lambda_N}{\overline{S}_{i-1}\overline{S}_i^2} + \sum_{k=i+1}^{N}\left(\frac{v_k \lambda_k c_i \Lambda_N}{\overline{S}_{k-1}^2 \overline{S}_{kj}} + \frac{v_k \lambda_k c_i \Lambda_N}{\overline{S}_{k-1}\overline{S}_k^2}\right)\right\}$$

respectively, we obtain $\underline{\lambda}^{+} =$ (0.3754, 0.1110) and $\underline{\lambda}^{\square} = (0.3718, 0.1517)$.

A serious problem with the transition in the example is that class-1 users' optimal price increases from 0.082 to 0.096 while that of class-2 users decreases from 6.06 to 4.49. Worse yet, class-1 users are worse off (their full price increases from 1.492 to 1.565) while class-2 users' full price (defined as the sum of access charge and sojourn time cost) reduces from 8.669 to 7.449. The public will notice that the transition favors class-2 users over class-1. In short, the transition is not Pareto-improving.

4. Simulation Results and Implications

The simulations used random problems created by a sampling procedure. Table 1 shows the range of values for service times, waiting costs, arrival rates, and coefficients of value functions used by the sampling procedure. Without loss of generality, the sampling procedure used value functions with derivatives of the form $V'(\lambda_i) = A_i - B_i\lambda_i$ $(i = 1, 2)$ and exponential service time distributions.

Table 1: Parameter Ranges for the Generated Problems

A_i	B_i	c_i	v_i
0 to 65	0 to 265	0 to 2	0 to 3

Table 2 summarizes the generated problems by the number of classes

Table 2: Summary of Generated Problems

No of Class	Pareto Improving	Non Pareto Improving	Infeasible	Non Converging
2	24,298	200	1,129	4,138
3	5,548	200	245	3,093
4	1,976	200	62	2,010

Each simulation compared the genetic algorithm to random search using the generated non Pareto-improving problems as input. The parameters used in the genetic algorithm (Table 3) are consistent with values used in other studies of constrained optimization problems (Goldberg 1989; Michalewicz 1996). In the random search, the heuristic objective function was used along with Δ_i values randomly generated in a range determined by the gap between the two social marginal costs of class-*i* users after and before the transition. The genetic algorithm used these randomly generated solutions as its initial population. The genetic algorithm was executed for each combination of a number of classes and a fitness function.

[1] We use the same example used by Mendelson and Whang (1990) for comparison.

Table 3: Parameters for the Genetic Algorithm

Population Size	Crossover Rate	Mutation Rate	Number of Generations
10	0.6	0.1	40

In the simulation study for two, three, and four user classes, we found that marginal contribution of the genetic algorithm may be marginal given the parameter values used in the simulation study. If random search typically finds some improvement to reduce the transition externality, the genetic algorithm may not have enough flexibility for much additional improvement.

The modest gain that we observed in the study can be still a larger gain to a network operator whose revenue size is in millions of dollars. Given the steep cost of upgrading the current networking gears to cope with the heavy traffic, it could be argued that only a minimal change is necessary to make the revenue prospects better without causing little technical difficulties.

References ←level 1 Heading 12pt, Bold

[1] Balachandran, K. and Radhakrishnan, S. (1994). "Extensions to class dominance characteristics," *Management Science* 40, 10, 353-360.

[2] Cocchi, R., Estrin, D., Shenker, S., and Zhang, L. (1993). "Pricing in computer networks: Motivation, formulation, and example," *IEEE/ACM Transactions on. Networking*, 1, 6, 614-629.

[3] Dovrolis, C. and Ramanathan, P. (1999). "A Case for Relative Differentiated Services and the Proportional Differentiation Model," *IEEE Network* 13, 5, 26-34.

[4] Edel, R. and Varaiya, P. (1999). "Providing Internet Access: What We Learn from INDEX," *IEEE Network* 13, 5, 18-25.

[5] Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, MA.

[6] Hahn R. and Wallsten S. (2006). "The Economics of Net Neutrality," http://www.bepress.com/ev/vol3/iss6/art8/

[7] Kim Y. and Mannino, M. (2003). "Optimal incentive-compatible pricing for M/G/1 Queues," *Operations Research Letters* 31, 6.

[8] Kleinrock, L. (1976). *Queueing Systems, Vol. II: Computer Applications*, John Wiley & Sons, Inc.

[9] Kwak, JH., "On network neutrality issues in the US," http://www.kisdi.re.kr/imagedata/pdf/10/1020061003.pdf

[10] Laxton, Jr. W., "The End of Net Neutrality," www.law.duke.edu/journals/dltr/articles/2006dltr0015.html

[11] Mendelson, H. and Whang, S. (1990). "Optimal Incentive-compatible Priority Pricing for the $M/M/1$ Queue," *Operations Research* 38, 5

[12] Michalewicz, Z. (1990). *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, Third Edition, Berlin, Germany, 1996.

[13] Mitrani I. (1998). *Probabilistic Modelling*, Cambridge University Press,

[14] Naor, P. (1969). "The Regulation of Queueing Size by Levying Tolls," *Econometrica* 37, 1.

[15] Pigou, P. (1920). *The Economics of Welfare*, Macmillan, First Edition, London.