QoS-Revenue Tradeoff with Time-Constrained ISP Pricing

Yuan Wu[1], Prashanth H. Hande[2], Hongseok Kim[2], Mung Chiang[2], Danny H.K. Tsang[1]
[1]Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology
[2]Department of Electrical Engineering, Princeton University

Abstract*—Usage-based pricing has been recognized as a network congestion management tool. Internet Service Providers (ISPs), however, have limited ability to set time-adaptive usage-price to manage congestion arising from time-varying consumer utility for data. To achieve the maximum revenue, ISP can set its time-invariant usage-price low enough to aggressively encourage consumer's traffic demand. The downside is that ISP has to drop consumer's excessive traffic demand through congestion management (i.e., packet dropping), which may degrade Quality of Service (QoS) of consumer's traffic. Alternatively, to protect consumer's QoS, ISP can set its time-invariant usage-price high enough to reduce consumer's traffic demand, thus minimizing the need for congestion management through packet dropping. The downside is that ISP suffers a revenue loss due to the inefficient usage of its network. The tradeoff between ISP's revenue maximization and consumer's QoS protection motivates us to study ISP's revenue maximization subject to QoS constraint in terms of the number of packets dropped. We investigate two different QoS measures: short-term per-slot packet dropping constraint and long-term packet dropping constraint. The short-term constraint can be interpreted as a more transparent congestion management practice compared to the long-term constraint. We analyze ISP's optimal time-invariant pricing for both constraints, and develop an upper bound for the optimal revenue by considering the specified packet dropping threshold. We quantify the impact of consumer's price elasticity on ISP's optimal revenue and show that ISP should carry out a differentiated QoS protection strategy based on consumer's price elasticity in order to mitigate the revenue loss[1].

I. INTRODUCTION

The rapid growth in Internet traffic requires Internet Service Providers (ISPs) to carry out some forms of network congestion management. Besides some technical forms of congestion management, usage-based pricing, serving as an alternative form of congestion management practice, can be more attractive to ISPs because it is more transparent to public and thus more acceptable by regulators. In fact, the Canadian telecom regulator recognized that "economic practices are the most transparent Internet traffic management practices" [1]. Consumer utility for data exhibits time variation [2], which results in a time-varying aggregate traffic demand faced by ISP. Ideally, ISP should adapt its usage-based pricing according to this time-varying traffic characteristic. For example, the usage-price should take the form of time-adaptive "congestion price" to guarantee a full utilization of ISP's network capacity, thus maximizing its revenue. The associated billing cost and technical complexity, however, restrict price variation over time, which means that ISP's practical pricing is actually "time-constrained".

Recent work [3] studied ISP's revenue maximization with time-constrained pricing. Two different time-invariant pricing strategies were analyzed. **Strategy (i)**: ISP can set its time-invariant usage-price low enough to aggressively encourage consumers' traffic demand, and then drop packets of consumers' excessive traffic demand that exceeds ISP's network capacity[2]. In this case, ISP's revenue is maximized and is equal to that with the time-adaptive congestion pricing described in section II.B. However, the downside is that ISP has to use the congestion management to drop consumers' traffic demand, which may degrade QoS of consumers' traffic [3]. **Strategy (ii):** ISP can conservatively set its time-invariant usage-price high enough so that the peak demand from consumers' aggregate traffic is within its network capacity. Thus, ISP does not have to invoke any congestion management. However, the downside is that ISP may suffer revenue loss because its network capacity is under-utilized during the off-peak periods.

A tradeoff exists between these two extreme strategies. In fact, recent report indicates ISPs are slashing prices to gain traffic and running into the tradeoff between overall revenue maximization and per-user QoS protection [8]. A survey on impact of QoS on ISP revenue is also being carried out by US National Exchange of Carrier Association [19]. Based on these motivations, in this paper we study ISP's time-constrained pricing with QoS measure in terms of the number of packets dropped. A brief overview of this paper is as follows.

(1) We first formulate ISP's revenue maximization problem with the time-constrained price and a packet dropping constraint. We consider two schemes: (a) short-term per-slot packet dropping constraint and (b) long-term packet dropping constraint. We then determine the optimal pricing for revenue maximization in both cases. Under certain conditions, we can develop the upper bounds for ISP's optimal revenues by considering the specified packet dropping thresholds.

(2) We analyze the impact of consumer's price elasticity, a measure of the consumer's change in demand with

[1]This work has been in part supported by NSF CNS-0720570 and CNS-0905086 at the Princeton EDGE Lab.

[2] [5] provides a comprehensive survey on different packet dropping policies used in IP networks and their effects.

[3]Dropping too many packets impairs the performance of TCP connection significantly, e.g., resulting in increase of delay and response time, and the decrease of link throughput and link utilization [6]. Quality of real-time application (which usually is built on UDP connection) is also severely deteriorated if many packets are dropped [7].

change in price, on ISP's optimal revenue. We show that in order to mitigate the revenue loss ISP should carry out a differentiated QoS protection policy based on consumer's price elasticity. Specifically, ISP can minimize revenue loss by providing a weak QoS protection (e.g., long-term packet dropping constraint or a loose packet dropping threshold) if consumer price elasticity is high. In contrast, ISP can afford to provide a strong QoS protection (e.g., short-term per-slot packet dropping constraint or a stringent packet dropping threshold) if consumer price elasticity is low. We also find that the flat-price is a dominant component in ISP's revenue when the packet dropping constraint is loose.

(3) We investigate the dependence of revenue loss on traffic type, each with its own price elasticity, and show that the congestion management is more suitable for best-effort traffic than real-time video traffic. Our analysis captures that the revenue loss can be marginal for best-effort traffic, but can be substantial for real-time video traffic. This result is consistent with the intuition that best-effort traffic has a larger tolerance to packet dropping (and so is more suitable for the congestion management) compared to real-time video traffic.

Pricing in communication networks has attracted a lot of research interests in recent years [11]. In general, two threads of works exist. The first thread of works focus on efficient network resource allocation. Specifically, pricing serves as a control mechanism which can convey the scarcity of resource [12]. [13] provided a comprehensive study on how the economic-driven behaviors of users and network operators influenced the network performance. The second thread of works focus on revenue maximization. Specifically, pricing serves as a revenue reaping mechanism when the network resource is limited [14] [15]. [16] [17] studied pricing strategy with multiple ISPs. This paper provides a new angle to study revenue maximization considering the impact of resource constraint, and explicitly characterizes the impact of pricing on the tradeoff between QoS and revenue.

II. Preliminary Formulation

A. Each Flow's Net-Surplus Maximization

We assume that there exists a set $\mathcal{F} = \{1, 2, ..., F\}$ of F flows requiring bandwidth from a monopolistic access ISP through a link with capacity C. We treat flow and consumer interchangeable in the rest of the paper. Given ISP's pricing decision, each flow f aims to maximize its net-surplus utility over a time horizon $\mathcal{T} = \{1, 2, ..., T\}$

$$\max_{\{z_f^t\}_t} \sum_t (\sigma_f^t u_f(z_f^t) - (h_f^t z_f^t + g_f^t)), \tag{1}$$

where z_f^t denotes flow f's traffic rate demand at time slot t and σ_f^t denotes flow f's utility level at time t. Specifically, σ_f^t represents f's happiness in using the network at time slot t, which can be measured by [9]. Meanwhile, h_f^t, g_f^t denote ISP's usage-price and flat-price for flow f at time slot t, respectively. In this work, ISP is assumed to be monopolistic and has the pricing power, which means (i) ISP can adopt the linear combination of usage-price and flat-price to transfer consumer's entire net-surplus into its revenue [10], and (ii) ISP can carry out price differentiation cross different flows. These assumptions are the most favorable conditions to ISP, and the outcome represents the maximum revenue ISP can expect. For clear presentation, in the following we presume that flow index $f \in \mathcal{F}$ and time index $t \in \mathcal{T}$ unless an additional specification is used. Followed by [3], the utility function $u_f(z_f^t) = \frac{1}{1-\alpha_f}(z_f^t)^{1-\alpha_f}$ when $0 \leq \alpha_f < 1$(such $u_f(z_f^t)$ is similar to the *Cobb-Douglas* utility function which has been widely used in the classic demand theory [20]). Flow f's net-surplus maximization problem (1) can be separated into individual time slot, and the corresponding traffic rate demand function can be expressed as: $D_f^t(h_f^t) = (\frac{\sigma_f^t}{h_f^t})^{\frac{1}{\alpha_f}}, \forall t$. Moreover, flow f's price elasticity [20] can be calculated as $\xi_f = \frac{dD_f^t(h_f^t)}{dh_f^t}\frac{h_f^t}{D_f^t(h_f^t)} = \frac{1}{\alpha_f}$, which is a normalized metric representing how much the demand changes as the price changes.

The price elasticity can be intuitively connected to traffic type. Specifically, when flow f is price inelastic (i.e., α_f is close to 1), then its utility function approaches to a log function, which captures the utility of best-effort traffic. In comparison, when flow f is price elastic (i.e., α_f is close to 0), then its utility function approaches to a linear function, which captures the utility of real-time video traffic [18]. This is because the operating bitrate for networked video is usually low considering the available bandwidth and the network congestion. Hence, the concave utility can be approximated by a roughly linear function in this regime. In reality, scalable video codec and adaptive-rate servers are used to achieve various operating points depending on network status.

B. ISP Revenue Maximization with Time-Adaptive Pricing

ISP's Revenue Maximization Problem (RMP) with time-adaptive pricing can be expressed as follows:

$$\textbf{(RMP):} \quad \max_{\{x_f^t, h_f^t, g_f^t\}_{f,t}} \sum_f \sum_t (h_f^t x_f^t + g_f^t) \tag{2}$$

$$\text{subject to: } x_f^t \leq D_f^t(h_f^t) = (\frac{\sigma_f^t}{h_f^t})^{\frac{1}{\alpha_f}}, \forall t, f \tag{3}$$

$$\sum_f x_f^t \leq C, \forall t \tag{4}$$

$$\sigma_f^t u_f(x_f^t) \geq x_f^t h_f^t + g_f^t, \forall t, f \tag{5}$$

Specifically, x_f^t denotes ISP's rate allocation for flow f at time period t. Constraint (3) represents that ISP's rate allocation for flow f at time period t cannot exceed flow f's demand according to its traffic rate demand function. Constraint (4) represents that ISP's total rate allocation cannot exceed its access link capacity C. Constraint (5) represents that ISP's rate allocation and price decision should guarantee a nonnegative surplus for flow f.

The above presentation of problem (RMP) makes it difficult to solve because it is a nonconvex problem. However, by putting the self-incentive constraint (5) into the objective function[4], problem (RMP) can be transformed into the following ISP Utility Maximization Problem (UMP):

$$\textbf{(UMP):} \max_{\{x_f^t\}_{f,t}} \sum_f \sum_t \sigma_f^t u_f(x_f^t) \text{ subject to: constraint(4).}$$

Let $\{x_f^{t*}\}_{f,t}$ denote the optimal rate allocation for problem (UMP). Meanwhile, let $\{\lambda^{t*}\}_t$ denote the set of optimal dual prices for constraints (4). There exists $x_f^{t*} = (\frac{\sigma_f^t}{\lambda^{t*}})^{\frac{1}{\alpha_f}}, \forall t, f$ and λ^{t*} is chosen so that $\sum_f x_f^{t*} = C, \forall t$. Thus, to solve the original problem (RMP), ISP's optimal pricing $\{h_f^{t*}, g_f^{t*}\}_{f,t}$ can be given by (1) $h_f^{t*} = \lambda^{t*}, \forall t, f$ and (2) $g_f^{t*} = \sigma_f^t u_f(x_f^{t*}) - h_f^{t*} x_f^{t*}, \forall t, f$. With $\{h_f^{t*}, g_f^{t*}\}_{f,t}$, ISP can transfer the optimal network utility $\sum_f \sum_t \sigma_f^t u_f(x_f^{t*})$ into its revenue. Meanwhile, since constraint (3) is strictly binding, ISP does not have to drop any flow's traffic demand.

C. ISP Revenue Maximization with Time-Constrained Pricing

Considering the practical billing system, [3] studied ISP's Revenue Maximization Problem with Time-Constrained (RMP-TC) pricing, i.e., ISP cannot change its price within the time horizon $\mathcal{T}$. The problem is as follows:

$$\textbf{(RMP-TC):} \max_{\{h_f, g_f\}_f, \{x_f^t\}_{f,t}} \sum_f ((\sum_t x_f^t) h_f + g_f) \quad (6)$$

$$\text{subject to:} x_f^t \leq D_f^t(h_f) = (\frac{\sigma_f^t}{h_f})^{\frac{1}{\alpha_f}}, \forall t, f \quad (7)$$

$$\sum_f x_f^t \leq C, \forall t \quad (8)$$

$$\sum_t \sigma_f^t u_f(x_f^t) \geq (\sum_t x_f^t) h_f + g_f, \forall f \quad (9)$$

Problem (RMP-TC) can be solved with the similar method as that for problem (RMP). Specifically, by putting constraint (9) into the objective function, problem (RMP-TC) can be transformed into the exactly same utility maximization problem as problem (UMP) (described in section II.B). Therefore, let $\{\widetilde{x}_f^t\}_{f,t}$ denote the optimal rate allocation profile for problem (RMP-TC), there exists $\widetilde{x}_f^t = x_f^*, \forall f, t$. Meanwhile, let $\{\widetilde{h}_f, \widetilde{g}_f\}_f$ denote the optimal pricing for problem (RMP-TC), then any pricing decision is optimal for problem (RMP-TC) if the following two conditions are met:

$$\widetilde{h}_f \leq \min_t \{h_f^{t*}\} = \min_t \{\frac{\sigma_f^t}{(x_f^{t*})^{\alpha_f}}\}, \forall f, \quad (10)$$

$$\widetilde{g}_f = \sum_t \sigma_f^t u_f(x_f^{t*}) - \widetilde{h}_f \sum_t x_f^{t*}, \forall f. \quad (11)$$

Condition (10) guarantees constraint (7), and condition (11) guarantees the binding of constraint (9). Thus conditions (10) and (11) together can guarantee the feasibility of $\{\widetilde{x}_f^t\}_{f,t}$.

Although in problem (RMP-TC) ISP's pricing is time-constrained, ISP can still achieve the same maximum revenue as that of problem (RMP) (because of $\sum_{f,t} \sigma_f^t u(\widetilde{x}_f^t) = \sum_{f,t} \sigma_f^t u(x_f^{t*})$) by using the optimal pricing decision satisfying conditions (10) and (11).

The downside of using the price strategy according to (10) and (11), however, is that ISP has to drop flow's excessive traffic demand. It is because constraint (7) is not always strictly binding at the optimum. For example, assume that ISP uses the optimal usage-price $\widetilde{h}_f = \min_t \{h_f^{t*}\}, \forall f$, then the dropped traffic rate can be expressed as:

$$\Delta_f^t = \frac{(\sigma_f^t)^{\frac{1}{\alpha_f}}}{(\widetilde{h}_f)^{\frac{1}{\alpha_f}}} - x_f^{t*} = \frac{(\sigma_f^t)^{\frac{1}{\alpha_f}}}{\min_t \{\frac{(\sigma_f^t)^{\frac{1}{\alpha_f}}}{x_f^{t*}}\}} - x_f^{t*} \geq 0, \forall f, t. \quad (12)$$

The rationale behind the above strategy is clear, i.e., ISP lowers down its usage-price as much as possible to encourage consumer's traffic demand, and then drop consumer's excessive traffic demand that exceeds its capacity afterwards.

Although with strategy (10) and (11), ISP does not suffer any revenue loss even its pricing is time-constrained, dropping consumer's packets causes severe QoS deterioration [6] [7].

Special Case: Assume that $\alpha_f = \alpha$ and ISP uses the optimal usage-price $\widetilde{h}_f = \min_t \{h_f^{t*}\}, \forall f$, then the dropped traffic rate (12) can be further expressed as:

$$\Delta_f^t = (\frac{(\sigma_f^t)^{\frac{1}{\alpha}}}{\min_t \{\sum_f (\sigma_f^t)^{\frac{1}{\alpha}}\}} - \frac{(\sigma_f^t)^{\frac{1}{\alpha}}}{\sum_f (\sigma_f^t)^{\frac{1}{\alpha}}}) C, \forall f, t. \quad (13)$$

Therefore, if consumer f has a large fluctuation in its utility-level profile $\{\sigma_f^t\}_t$, then ISP has to drop a large traffic demand to avoid the revenue loss. Our numerical result in Figure 7 verifies this point.

In the following we consider the optimal usage-price for problem (RMP-TC) as $\widetilde{h}_f = \min_t \{h_f^{t*}\}, \forall f$, which serves as a benchmark strategy because ISP aggressively aims to maximize its revenue regardless of how many packets are dropped.

III. Time-Constrained Pricing and Packet Dropping Constraint

A. Time-Constrained Pricing with Short-Term Per-Slot Packet Dropping Constraint

We consider ISP time-constrained pricing with short-term per-slot packet dropping constraint in this subsection. The problem is as follows (where "PS" stands for per-slot):

$$\textbf{(RMP-PS):} \max_{\{h_f, g_f\}_f, \{x_f^t\}_{f,t}} \sum_f ((\sum_t x_f^t) h_f + g_f) \quad (14)$$

$$\text{subject to: } x_f^t \leq D_f^t(h_f) = (\frac{\sigma_f^t}{h_f})^{\frac{1}{\alpha_f}}, \forall t, f \quad (15)$$

$$(\frac{\sigma_f^t}{h_f})^{\frac{1}{\alpha_f}} - x_f^t \leq \Gamma_f^t, \forall t, f \quad (16)$$

$$\sum_f x_f^t \leq C, \forall t \quad (17)$$

$$\sum_t \sigma_f^t u_f(x_f^t) \geq (\sum_t x_f^t) h_f + g_f, \forall f \quad (18)$$

[4]Notice that by using the combination of usage-price and flat-price, ISP can completely transfer each flow's utility into its revenue [10]. Thus constraint (5) is always binding at the optimum.

Specifically, constraint (16) guarantees at time slot t, the number of dropped packets for flow f (i.e., the gap between flow f's instantaneous traffic rate demand $(\frac{\sigma_f^t}{h_f})^{\frac{1}{\alpha_f}}$ and ISP's rate allocation x_f^t) cannot exceed the threshold Γ_f^t.

Problem (RMP-PS) is a nonconvex problem. By putting constraint (18) into the objective function, and then combining constraints (15) and (16) together, problem (RMP-PS) can be transformed into the following ISP utility maximization problem[5]:

$$\textbf{(UMP-PS):} \max_{\{x_f^t\}_{f,t}} \sum_f \sum_t \sigma_f^t u_f(x_f^t) \tag{19}$$

$$\text{subject to:} \frac{(\sigma_f^{t'})^{\frac{1}{\alpha_f}}}{x_f^{t'}+\Gamma_f^{t'}} \leq \frac{(\sigma_f^t)^{\frac{1}{\alpha_f}}}{x_f^t}, \forall t,t',f \tag{20}$$

$$\sum_f x_f^t \leq C, \forall t \tag{21}$$

Given the packet dropping thresholds $\Gamma_f^t, \forall t$, constraint (20) imposes the restriction on the rate allocation profile (i.e., $\{x_f^t\}_t$) for flow f so that at least a feasible h_f exists.

Remark 1: One of the sufficient conditions for problem (RMP-PS) to be equivalent to problem (RMP-TC) can be given as: $\Gamma_f^t \geq \Delta_f^t, \forall f, t$ (Δ_f^t is given in (12)). As described in section II.C, Δ_f^t denotes the minimum traffic rate that ISP has to drop in order to avoid any revenue loss (when its pricing is time-constrained).

Because of the concavity of the objective function and linearity of the constraints, problem (UMP-PS) is a convex problem, and thus KKT condition is applicable. Let $\mu_f^{tt'}$ denote the dual price for constraint (20). Let λ^t denote the dual price for constraint (21). The optimal solution for problem (UMP-PS) can thus be expressed as: $\widetilde{x}_f^t = (\frac{\sigma_f^t}{-\sum_{t'\neq t}(\widetilde{\mu}_f^{t't}-\widetilde{\mu}_f^{tt'})(\sigma_f^{t'})^{\frac{1}{\alpha_f}}+\widetilde{\lambda}^t})^{\frac{1}{\alpha_f}}$, where $\widetilde{\mu}_f^{tt'}$ denotes the optimal dual price for constraint (20). Notice that, different from $x_f^{t*} = (\frac{\sigma_f^t}{\lambda^{t*}})^{\frac{1}{\alpha_f}}$, in problem (UMP-PS) each flow's optimal rate allocation also depends on the corresponding dual price for the packet dropping constraint. For notational simplicity we still use $\{\widetilde{x}_f^t\}_{f,t}, \{\widetilde{h}_f, \widetilde{g}_f\}_f$ to denote the optimal solution for problem (RMP-PS). Notice that for each flow f and any pair of t, t', there exists $\widetilde{\mu}_f^{tt'}\widetilde{\mu}_f^{t't} = 0$ if $\Gamma_f^t\Gamma_f^{t'} > 0$.

Based on the optimal rate allocation $\{\widetilde{x}_f^t\}_{f,t}$, the maximum sum-utility for problem (UMP-PS) can be transferred to ISP's revenue. Specifically, any value of usage-price satisfying the following condition can be used by ISP to extract all flows' utilities:

$$\max_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t+\Gamma_f^t)^{\alpha_f}}\} \leq \widetilde{h}_f \leq \min_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}\}, \forall f. \tag{22}$$

[5]Notice that constraint (15) actually requires that $h_f \leq \min_t\{\frac{\sigma_f^t}{(x_f^t)^{\alpha_f}}\}$ and constraint (16) requires that $h_f \geq \max_t\{\frac{\sigma_f^t}{(x_f^t+\Gamma_f^t)^{\alpha_f}}\}$. Thus, by combining them together we can get constraint (20), which guarantees the existence of a feasible h_f for flow f.

Meanwhile, the corresponding flat-price can be set as: $\widetilde{g}_f = \sum_t \sigma_f^t u_f(\widetilde{x}_f^t) - (\sum_t \widetilde{x}_f^t)\widetilde{h}_f, \forall f$.

Remark 2: For each flow f, if the packet dropping threshold $\Gamma_f^t, \forall t$, is large enough so that the optimal dual price $\widetilde{\mu}_f^{tt'} = 0, \forall t, t'$, i.e., constraint (20) is slack (we provide a sufficient condition for it in **Remark 1**), then $\max_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t+\Gamma_f^t)^{\alpha_f}}\} < \min_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}\}$ (i.e., the value of optimal usage-price $\widetilde{h}_f$ can be chosen from an interval). Otherwise, $\max_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t+\Gamma_f^t)^{\alpha_f}}\} = \min_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}\}$ (i.e., the value of optimal usage-price $\widetilde{h}_f$ can only be chosen from a single value). The reason is as follows. If flow f has the optimal dual price $\widetilde{\mu}_f^{tt'} = 0, \forall t, t'$, then $\frac{\sigma_f^{t'}}{(\widetilde{x}_f^{t'}+\Gamma_f^{t'})^{\alpha_f}} < \frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}, \forall t, t'$, which means that $\max_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t+\Gamma_f^t)^{\alpha_f}}\} < \min_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}\}$. On the contrary, assume that there exists a particular pair of slots t, t' for flow f, which has $\widetilde{\mu}_f^{tt'} \neq 0$, i.e., $\frac{\sigma_f^{t'}}{(\widetilde{x}_f^{t'}+\Gamma_f^{t'})^{\alpha_f}} = \frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}$, then $\max_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t+\Gamma_f^t)^{\alpha_f}}\} \geq \frac{\sigma_f^{t'}}{(\widetilde{x}_f^{t'}+\Gamma_f^{t'})^{\alpha_f}} = \frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}} \geq \min_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}\}$. Followed by the feasibility of constraint (20), there exists $\max_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t+\Gamma_f^t)^{\alpha_f}}\} = \min_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}\}$, which means that $\widetilde{h}_f$ can only be chosen from a single value.

Let $V = \sum_f \sum_t \sigma_f^t u_f(\widetilde{x}_f^t)$ denote ISP's optimal revenue for problem (RMP-TC). Let $V_{PS} = \sum_f \sum_t \sigma_f^t u_f(\widetilde{x}_f^t)$ denote ISP's optimal revenue for problem (RMP-PS)[6]. We provide a bound of revenue loss for per-slot packet dropping constraint as follows (with proof in Appendix I).

Proposition 1: Assume that the homogeneous rate dropping constraint $\Gamma_f^t = \gamma, \forall f, t$ is used. Further assume that both flow's price elasticity and utility level are flow independent, i.e., $\xi_f = \frac{1}{\alpha_f} = \frac{1}{\alpha}, \forall f$, and $\sigma_f^t = \sigma^t, \forall f, t$, then $1 \geq \frac{V_{PS}}{V} \geq \frac{\sum_t(\frac{\sigma^t}{\max_t\{\sigma^t\}})^{\frac{1}{\alpha}}}{\sum_t(\frac{\sigma^t}{\max_t\{\sigma^t\}})}$.

However, the value of $\frac{V_{PS}}{V}$ is difficult to derive for arbitrary γ. We can only provide a upper bound for it as follows.

Proposition 2: Assume that the homogeneous packet dropping constraint $\Gamma_f^t = \gamma, \forall f, t$ is used in problem (RMP-PS). If both flow's price elasticity and utility level are flow independent, i.e., $\xi_f = \frac{1}{\alpha_f} = \frac{1}{\alpha}, \forall f$, and $\sigma_f^t = \sigma^t, \forall f, t$, then

$$\frac{V_{PS}}{V} \leq \frac{\sum_{t\in\Omega_1}\sigma^t}{\sum_t \sigma^t} + \frac{(\frac{C}{F}+\gamma)^{1-\alpha}}{(\frac{C}{F})^{1-\alpha}} \frac{\sum_{t\in\Omega_2}\frac{(\sigma^t)^{\frac{1}{\alpha}}}{\max_t\{(\sigma^t)^{\frac{1}{\alpha}}\}}}{\sum_t \frac{\sigma^t}{\max_t\{\sigma^t\}}}, \tag{23}$$

where Ω_1 and Ω_2 are two complementary sets of time slots defined in the proof in Appendix II.

Remark 3: According to (23), the upper bound of $\frac{V_{PS}}{V}$ is nondecreasing with respect to the value of γ until constraint (16) becomes slack[7]. This is consistent with the intuition that the larger the value of γ, the larger the feasible region, thus ISP's optimal revenue cannot decrease.

[6]Notice that the optimal rate allocation $\{\widetilde{x}_f^t\}_{f,t}$ for problem (RMP-PS) is different from the optimal rate allocation for (RMP-TC).

[7]For example, γ satisfies the sufficient condition provided in **Remark 1**.

Remark 4: According to (23), the upper bound of $\frac{V_{PS}}{V}$ is decreasing with respect to each flow's price elasticity ξ (i.e., increasing with respect to the value of α). The proof is given in Appendix III. The decreasing property of the right hand side of (23) with ξ is consistent with the intuition. Specifically, if each flow has a high price elasticity, then a low usage-price should be used to maximize the revenue [20]. However, the packet dropping constraint restricts ISP from using the low usage-price, resulting in that ISP suffers a revenue loss. Moreover, the higher the price elasticity, the larger the revenue loss.

Figure 1 shows the performance of ISP's revenue upper bound (23) for problem (RMP-PS). For numerical illustration, we consider a network scenario with $F = 3$ flows and $T = 20$. The access link capacity is $C = 2$. Meanwhile, we set the optimal usage-price as $\widetilde{h}_f = \min_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}\}, \forall f$ according to (22), which corresponds to the least packets dropped.

Figure 1 shows that our derived upper (23) is tight. The tightness can be explained by showing that $x^{t_0} = \frac{C}{F}$ (notice that $t_0 = \arg\max_t\{(\sigma^t)^{\frac{1}{\alpha}}\}$) for any value of $\gamma \geq 0$. It can be intuitively explained as follows. Constraint (20) requires that $\frac{x^t}{x^{t_0}+\gamma} \leq \frac{(\sigma^t)^{\frac{1}{\alpha}}}{(\max_t\{\sigma^t\})^{\frac{1}{\alpha}}}$. As described before, when $\gamma = 0$, $x^{t_0} = \frac{C}{F}$. Meanwhile, when γ increases (i.e., $\gamma > 0$), $x^{t_0} = \frac{C}{F}$ can still hold and all the other $x^t, t \neq t_0$ can increase accordingly until they are bounded by $\frac{C}{F}$ (it is clear that in order to maximize the objective function $\sum_f \sum_t \sigma_f^t u_f(x_f^t)$, $\{x^t\}_{t\in\mathcal{T}, f\in\mathcal{F}}$ needs to be set as large as possible).

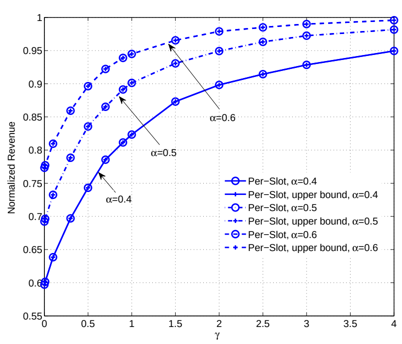

Fig. 1. Performance of the revenue upper bound (23) for problem (RMP-PS). We change $\alpha_f = \alpha = 0.4, 0.5, 0.6$, and set $\sigma_f^t = \sigma^t, \forall f$, which follows the same uniform distribution $u(0.2, 1)$.

Remark 4 leads to a policy implication for ISP. Specifically, in order to mitigate the revenue loss due to its time-constrained pricing, ISP should provide different levels of QoS protection according to different price elasticities. Specifically, ISP should use a large packet dropping threshold (i.e., weak QoS protection) if consumer's price elasticity is high. However, if consumer's price elasticity is low, then ISP can use a small packet dropping threshold (i.e., strong QoS protection) without suffering a large revenue loss. The results shown in Figure 2 and Figure 6 also imply this policy.

Figure 1 also shows that the revenue loss depends on the traffic type. Recall that when α is close to 1, then the utility function roughly captures the utility of best-effort traffic. In comparison when α is close to 0, then the utility function roughly captures the utility of real-time video traffic. As shown in Figure 1, given the same packet dropping threshold, real-time video traffic incurs a larger revenue loss than best-effort traffic does. In practice, real-time video traffic can only tolerate a small number of packets dropped, thus the revenue loss could be substantial. By contrast, best-effort traffic can tolerate a large number of packets dropped, thus the revenue loss could be even marginal. These results confirm our intuition that the congestion management via packet dropping is more suitable for best-effort traffic than real-time video traffic.

B. *Time Constrained Pricing with Long-Term Packet Dropping Constraint*

We also consider ISP's time-constrained pricing with long-term packet dropping constraint as follows (where "LT" stands for long-term):

$$\textbf{(RMP-LT):} \max_{\{h_f, g_f\}_f, \{x_f^t\}_{f,t}} \sum_f ((\sum_t x_f^t) h_f + g_f) \tag{24}$$

$$\text{subject to:} x_f^t \leq (\frac{\sigma_f^t}{h_f})^{\frac{1}{\alpha_f}}, \forall t, f \tag{25}$$

$$\sum_t \max\{(\frac{\sigma_f^t}{h_f})^{\frac{1}{\alpha_f}} - x_f^t, 0\} \leq \Gamma_f, \forall f \tag{26}$$

$$\sum_f x_f^t \leq C, \forall t \tag{27}$$

$$\sum_t \sigma_f^t u_f(x_f^t) \geq (\sum_t x_f^t) h_f + g_f, \forall f \tag{28}$$

Constraint (26) guarantees that in long-term the number of dropped packets for flow f (i.e., $\sum_t \max\{(\frac{\sigma_f^t}{h_f})^{\frac{1}{\alpha_f}} - x_f^t, 0\}$) cannot exceed the threshold Γ_f. Compared to problem (RMP-PS), problem (RMP-LT) has a larger flexibility in dropping each flow's traffic rate demand (i.e., ISP has the freedom to allocate its packet dropping budget over time). Thus, if the threshold $\Gamma_f = \sum_t \Gamma_f^t, \forall f$ (Γ_f^t is the packet dropping threshold in problem (RMP-PS)), then there exists $V_{LT} \geq V_{PS}$, where V_{LT} denotes the optimal revenue for problem (RMP-LT). For fair comparison, we set $\Gamma_f = \sum_t \Gamma_f^t$ in the following numerical experiments.

Problem (RMP-LT) is a nonconvex problem. By putting constraint (28) into the objective function, and then combining constraints (25) and (26) together, problem (RMP-LT) can be transformed into the following ISP utility maximization problem:

$$\textbf{(UMP-LT):} \max_{\{x_f^t\}_{f,t}} \sum_f \sum_t \sigma_f^t u_f(x_f^t) \tag{29}$$

$$\text{subject to:} \frac{\sum_t (\sigma_f^t)^{\frac{1}{\alpha_f}}}{\Gamma_f + \sum_t x_f^t} \leq \frac{(\sigma_f^t)^{\frac{1}{\alpha_f}}}{x_f^t}, \forall t, f \tag{30}$$

$$\sum_f x_f^t \leq C, \forall t \tag{31}$$

Given the packet dropping thresholds $\{\Gamma_f\}_f$, constraint (30) imposes the restriction on the rate allocation profile for flow f so that a feasible h_f exists for constraints (25) and (26).

Remark 6: One of the sufficient conditions for problem (RMP-LT) to be equivalent to problem (RMP-TC) can be given as: $\Gamma_f \geq \sum_t \Delta_f^t, \forall f$.

Problem (UMP-LT) is a convex problem and KKT condition is applicable. Let ψ_f^t denote the dual price for constraint (30). Let λ^t denote the dual price for constraint (31). The optimal solution for problem (UMP-LT) can thus be expressed as: $\widetilde{x}_f^t = \left(\frac{\sigma_f^t}{-\sum_{t'\neq t}(\widetilde{\psi}_f^{t'} - \widetilde{\psi}_f^t)(\sigma_f^{t'})^{\frac{1}{\alpha_f}} + \widetilde{\lambda}^t}\right)^{\frac{1}{\alpha_f}}, \forall f, t$. Based on the optimal rate allocation $\{\widetilde{x}_f^t\}_{f,t}$, the maximum sum-utility for problem (UMP-LT) can be transferred to ISP's revenue. Specifically, any value of usage-price satisfying the following condition can be used by ISP to extract all flows' utilities:

$$\left(\frac{\sum_t (\sigma_f^t)^{\frac{1}{\alpha_f}}}{\Gamma_f + \sum_t \widetilde{x}_f^t}\right)^{\alpha_f} \leq \widetilde{h}_f \leq \min_t \left\{\frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}\right\}, \forall f. \quad (32)$$

Meanwhile, the corresponding value of flat-price can be set as: $\widetilde{g}_f = \sum_t \sigma_f^t u_f(\widetilde{x}_f^t) - (\sum_t \widetilde{x}_f^t)\widetilde{h}_f, \forall f$.

Remark 7: For flow f, if the packet dropping threshold Γ_f is large enough so that $\widetilde{\psi}_f^t = 0$, i.e., constraint (30) is slack (we provide a sufficient condition for it in **Remark 6**), then $\left(\frac{\sum_t (\sigma_f^t)^{\frac{1}{\alpha_f}}}{\Gamma_f + \sum_t \widetilde{x}_f^t}\right)^{\alpha_f} < \min_t\left\{\frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}\right\}$ (i.e., the value of $\widetilde{h}_f$ can be chosen from an interval). Otherwise, $\left(\frac{\sum_t (\sigma_f^t)^{\frac{1}{\alpha_f}}}{\Gamma_f + \sum_t \widetilde{x}_f^t}\right)^{\alpha_f} = \min_t\left\{\frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}\right\}$ (i.e., the value of $\widetilde{h}_f$ can only be chosen from a single value). The explanation is similar to that in Remark 2.

Let $V_{LT} = \sum_f \sum_t \sigma_f^t u_f(\widetilde{x}_f^t)$ denote ISP's optimal revenue for problem (RMP-LT) [8]. We provide a revenue loss bound as follows (the proof is similar to that for **Proposition 1**).

Proposition 3: Assume that the homogeneous rate dropping constraint $\Gamma_f = \eta, \forall f$ is used. If both flow's price elasticity and utility level are flow independent, i.e., $\xi_f = \frac{1}{\alpha_f} = \frac{1}{\alpha}, \forall f$, and $\sigma_f^t = \sigma^t, \forall f, t$, then $1 \geq \frac{V_{LT}}{V} \geq \frac{\sum_t \left(\frac{\sigma^t}{\max_t \sigma^t}\right)^{\frac{1}{\alpha}}}{\sum_t \left(\frac{\sigma^t}{\max_t \sigma^t}\right)}$.

However, the exact value of $\frac{V_{LT}}{V}$ is difficult to derive for arbitrary η. We only provide a upper bound for it as follows.

Proposition 4: Assume the homogenous packet dropping constraint $\Gamma_f = \eta, \forall f$, is used in problem (RMP-LT). If both flow's price elasticity and utility level are flow independent, i.e., $\xi_f = \frac{1}{\alpha_f} = \frac{1}{\alpha}, \forall f$, and $\sigma_f^t = \sigma^t, \forall f, t$, then

$$\frac{V_{LT}}{V} \leq \frac{\sum_{t\in\Pi_1} \sigma^t}{\sum_t \sigma^t} + \frac{(\frac{C}{F}T + \eta)^{1-\alpha}}{(\frac{C}{F})^{1-\alpha}} \frac{\sum_{t\in\Pi_2} \frac{(\sigma^t)^{\frac{1}{\alpha}}}{\sum_t (\sigma^t)^{\frac{1}{\alpha}}}}{\sum_t \frac{\sigma^t}{(\sum_t (\sigma^t)^{\frac{1}{\alpha}})^{\alpha}}}, \quad (33)$$

where Π_1 and Π_2 are two complementary sets of time slots defined in the proof in Appendix IV.

[8]Notice that the optimal rate allocation $\{\widetilde{x}_f^t\}_{f,t}$ for problem (RMP-LT) is different from the optimal rate allocation for problem (RMP-TC).

Remark 8: According to (33), the upper bound of $\frac{V_{LT}}{V}$ is nondecreasing with respect to η until constraint (26) becomes always slack[9].

Remark 9: According to (33), the upper bound of $\frac{V_{LT}}{V}$ is decreasing with respect to each flow's price elasticity (i.e., increasing with respect to the value of α). The proof is given in Appendix IV. The decreasing property of $\frac{V_{LT}}{V}$ with respect to ξ can also be intuitively explained with the similar reason as that in Remark 4.

Figure 2 shows the performance ISP's revenue upper bound (33) for problem (RMP-LT). We also set the optimal usage-price as $\widetilde{h}_f = \min_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}\}, \forall f$ according to (32), which corresponds to the least packets dropped. Figure 2 shows that the derived upper bound for $\frac{V_{LT}}{V}$ is tight when $\Gamma_f = \eta$ is large. Figure 3 further shows the relative error of the provided revenue upper bound. It shows that the upper bound (33) is more accurate when (i) T is small, (ii) η is large, and (iii) ξ is small. The looseness of the upper bound is because we make some aggressive assumption in derivation as described in Appendix IV.

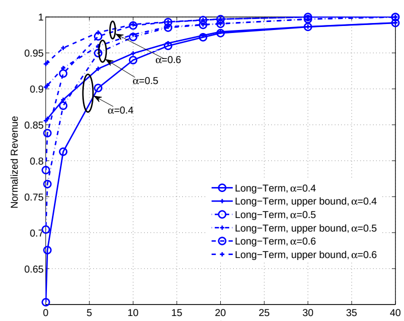

Fig. 2. Performance of the revenue upper bound (33) for problem (RMP-LT). We change $\alpha_f = \alpha = 0.4, 0.5, 0.6$, and set $\sigma_f^t = \sigma^t, \forall f$, which follows the same uniform distribution $u(0.2, 1)$.

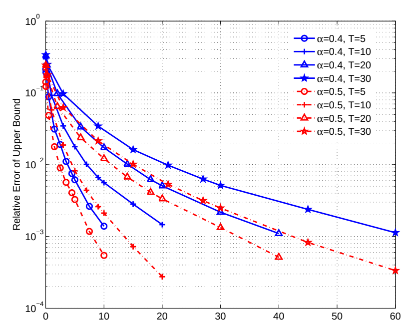

Fig. 3. Relative error of the upper bound (33). $\alpha_f = \alpha = 0.4, 0.5$ and $\sigma_f^t = \sigma^t, \forall f$, follows the same uniform distribution $u(0.2, 1)$. Every point in the figure is averaged over 100 random realizations of $\{\sigma^t\}_t$.

Remark 9 also suggests the differentiated QoS protection policy for ISP. Specifically, to mitigate the revenue loss due

[9]For example, η satisfies the sufficient condition provided in **Remark 6**.

to the time-constrained price strategy, ISP should only provide a weak QoS protection if consumer's price elasticity is high. Considering the traffic type, Figure 2 also suggests that given the same packet dropping constraint, best-effort traffic incurs a smaller revenue loss than real-time video traffic does.

C. Impact of Packet Dropping Constraint on Usage-Based Revenue

Both short-term per-slot and long-term packet dropping constraints impose restrictions on ISP's usage-price, thus influencing ISP's usage-based revenue.

Proposition 5: For problem (RMP-PS), let $V_{PS}^{usage} = \sum_f (\sum_t \widetilde{x}_f^t) \widetilde{h}_f$ denote ISP's optimal usage-based revenue. If all flows have the same price elasticity (i.e., $\xi_f = \frac{1}{\alpha_f} = \frac{1}{\alpha}, \forall f$), then $\frac{V_{PS}^{usage}}{V_{PS}} \leq 1 - \alpha$, and the equal sign is obtained when ISP cannot drop any flow's rate demand, i.e., $\Gamma_f^t = 0, \forall f, t$. (The proof is in Appendix V.)

Similarly, we also have the following proposition.

Proposition 6: For problem (RMP-LT), let $V_{LT}^{usage} = \sum_f (\sum_t \widetilde{x}_f^t) \widetilde{h}_f$ denote ISP's optimal usage-based revenue. If all flows have the same price elasticity (i.e., $\xi_f = \frac{1}{\alpha_f} = \frac{1}{\alpha}, \forall f$), then $\frac{V_{LT}^{usage}}{V_{LT}} \leq 1 - \alpha$, and the equal sign is obtained when ISP cannot drop any flow's rate demand, i.e., $\Gamma_f = 0, \forall f$.

Notice that the upper bound of $\frac{V_{PS}^{usage}}{V_{PS}}$ and $\frac{V_{LT}^{usage}}{V_{LT}}$ are both increasing with respect to each flow's price elasticity. This is consistent with the intuition that usage-based revenue is dominant in the entire revenue if consumer has a high price elasticity. Similar result also appeared in [3]. The exact values of $\frac{V_{PS}^{usage}}{V_{PS}}$ and $\frac{V_{LT}^{usage}}{V_{LT}}$ with arbitrary packet dropping thresholds, however, are difficult to derive, and they depend on the detailed choices of $\widetilde{h}_f$ (according to (22) and (32)).

Figure 4 shows the comparison between $\frac{V_{PS}^{usage}}{V_{PS}}$ and $\frac{V_{LT}^{usage}}{V_{LT}}$. For both problems, we set the optimal usage-price as $\widetilde{h}_f = \min_t \{\frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}\}, \forall f$ according to (22) and (32). Figure 4 shows that $\frac{V_{PS}^{usage}}{V_{PS}}$ decreases when $\Gamma_f^t = \gamma, \forall f, t$ increases. Meanwhile, $\frac{V_{PS}^{usage}}{V_{PS}}$ is lower bounded as $\Gamma_f^t = \gamma \to \max_{f,t}\{\Delta_f^t\}$ (12). Specifically, $\frac{V_{PS}^{usage}}{V_{PS}}$ is lower bounded by the corresponding value of $\frac{V^{usage}}{V}$ for problem (RMP-TC), where V^{usage} denotes the optimal usage-based revenue for problem (RMP-TC). The intuitive explanation is as follows. If $\Gamma_f^t = \gamma$ is so small that $\widetilde{h}_f = \min_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}\} = \max_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t + \Gamma_f^t)^{\alpha_f}}\}$, then $\widetilde{h}_f$ decreases when Γ_f^t increases, which implies that ISP can lower down its usage-price more aggressively. As a result, the value of $\frac{V_{PS}^{usage}}{V_{PS}}$ decreases. However, if $\Gamma_f^t = \gamma$ is so large that $\widetilde{h}_f = \min_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}\} > \max_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t + \Gamma_f^t)^{\alpha_f}}\}$, then according to our previous description in section III.A, problem (RMP-PS) has already become equivalent to problem (RMP-TC). Therefore, there exists $\frac{V_{PS}^{usage}}{V_{PS}} = \frac{V^{usage}}{V}$. We can observe a similar property of $\frac{V_{LT}^{usage}}{V_{LT}}$. The value of $\frac{V_{LT}^{usage}}{V_{LT}}$ is lower bounded by $\frac{V^{usage}}{V}$ as $\Gamma_f = \eta \to \max_f \sum_t \{\Delta_f^t\}$. Figure 4 indicates that the flat-part revenue (i.e., flat-price) is important to ISP's revenue retention. Specifically, the more loose the packet dropping constraint, the more heavily ISP has to rely on its flat-price to achieve the maximum revenue.

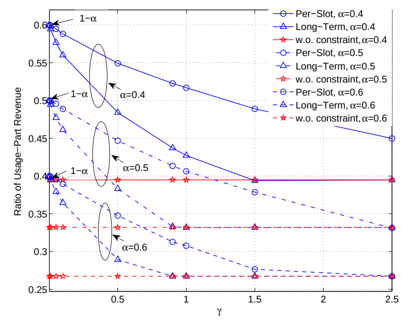

Fig. 4. Comparison of the usage-based revenue ratio between problem (RMP-PS) and problem (RMP-LT)

D. Tradeoff between Per-Slot and Long-Term Constraints

Figure 5 shows the comparison of the optimal revenues and average number of packets dropped among problems (RMP-TC), (RMP-PS), (RMP-LT). Specifically, the horizontal axis denotes the packet dropping threshold $\Gamma_f^t = \gamma, \forall f, t$ for problem (RMP-PS). Meanwhile, for fair comparison the packet dropping threshold for problem (RMP-LT) is set as $\Gamma_f = \eta = T\gamma$. The top subfigure shows the comparison between $\frac{V_{PS}}{V}$ and $\frac{V_{LT}}{V}$. It shows that ISP can always achieve a smaller revenue loss with long-term packet dropping constraint than with short-term per-slot constraint (until both problems (RMP-PS) and (RMP-LT) become equivalent). This result is consistent with the intuition. Because ISP has a larger flexibility in dropping flows' traffic rate with long-term constraint, it can obtain a no smaller optimal revenue.

The downside of long-term packet dropping constraint, however, is that it can only provide a weak QoS protection. The bottom subfigure in Figure 5 verifies this point by showing the average number of packets dropped for the three problems. Figure 5 presents the tradeoff faced by ISP, i.e., between using short-term per-slot packet dropping constraint to provide a strong QoS protection but suffering a large revenue loss and using long-term constraint to provide a weak QoS protection but suffering a small revenue loss.

E. Impact of Price Elasticity and Utility Level Fluctuation

Figure 6 shows the impact of flow's price elasticity on ISP's optimal revenue. Figure 6 shows that ISP will suffer a large revenue loss if each flow has a high price elasticity (for both short-term per-slot packet dropping constraint and long-term packet dropping constraint). These results are consistent with our previous **Remark 4** and **Remark 9**.

Figure 7 shows the impact of the fluctuation of flow's utility level $\{\sigma_f^t\}_{f,t}$ on ISP's optimal revenue. Figure 7 shows that if each flow's utility level has a small fluctuation, then ISP will suffer a small revenue loss. Intuitively this is right since

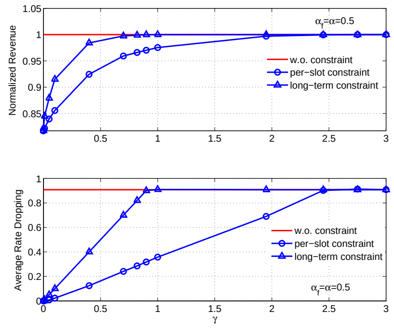

Fig. 5. Comparison of the optimal revenue and the average number of packets dropped for problems (RMP-TC), (RMP-PS) (RMP-LT). Top subfigure: the comparison between $\frac{V_{PS}}{V}$ and $\frac{V_{LT}}{V}$. Bottom subfigure: the comparison of the average number of packets dropped $E[(\frac{\sigma_f^t}{h_f})^{\frac{1}{\alpha_f}} - \tilde{x}_f^t]$.

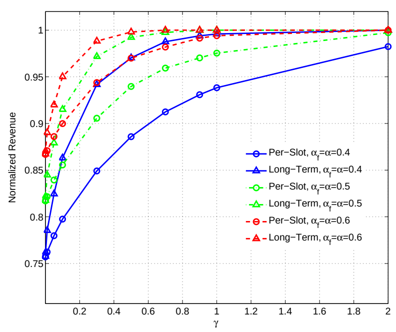

Fig. 6. Impact of flow's price elasticity on ISP's optimal revenue. We change $\alpha_f = \alpha = 0.4, 0.5, 0.6, \forall f$.

a small fluctuation of utility level implies that ISP can more aggressively lower down its usage-price, thus attracting more consumer's traffic rate demand. For example, according to (13), if $\sum_f(\sigma_f^t)^{\frac{1}{\alpha}}$ is a constant within the time horizon $\mathcal{T}$, then ISP can avoid the revenue loss without dropping any flow's packets.

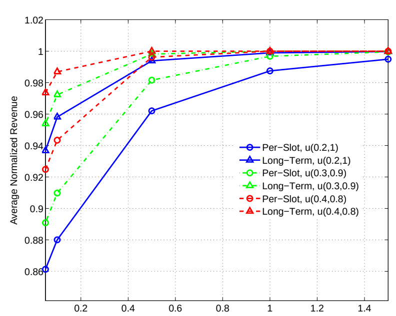

Fig. 7. Impact of flow's utility level fluctuation on ISP's optimal revenue. $\alpha_f = \alpha = 0.6, \forall f$. Every point in the figure is averaged over 100 random realizations of $\{\sigma_f^t\}_{f,t}$. We test three cases, i.e., $\{\sigma_f^t\}_{t,f}$ follows the uniform distributions $u(0.2, 1)$, $u(0.3, 0.9)$ and $u(0.4, 0.8)$, respectively.

IV. Conclusion

This paper studies ISP's revenue maximization trading off with QoS measure (in terms of the number of packets dropped). We consider two QoS time horizons (i.e., short-term per-slot constraint and long-term packet dropping constraint), and quantify the tradeoff between QoS protection and revenue maximization faced by ISP when its pricing has to be "time-constrained". In particular, we demonstrate the impact of consumer's price elasticity on ISP's optimal revenue, and show that in order to mitigate the revenue loss ISP should carry out a differentiated QoS protection strategy based on consumer's price elasticity. We analyze the optimal time-constrained pricing for both cases (short-term per-slot constraint and long-term constraint) and identify the importance of ISP's flat-price in reaping revenue as the QoS protection constraint becomes loose.

Appendix I: Proof of Proposition 1

Proof: The optimal revenue V_{PS} for problem (RMP-PS) is nondecreasing with respect to the value of γ (because of the increase of the feasible region) until constraint (16) becomes always slack (e.g., satisfying the sufficient condition provided in **Remark 1**) and thus $V = V_{PS}$.

On the other side, the lower bound of $\frac{V_{PS}}{V}$ is obtained when $\gamma = 0$. Specifically, constraint (20) becomes that $\frac{(\sigma_f^{t'})^{\frac{1}{\alpha_f}}}{x_f^{t'}} = \frac{(\sigma_f^{t})^{\frac{1}{\alpha_f}}}{x_f^{t}}, \forall f, t, t'$ when $\gamma = 0$. Let $t^0 = \arg\max_t\{\sigma^t\}$ (we omit the flow index here for clear presentation). With the assumption that $\alpha_f = \alpha, \forall t$ and $\sigma_f^t = \sigma^t, \forall f, t$, there exists $\tilde{x}^{t^0} = \frac{C}{F}$. Thus, ISP's optimal rate allocation can be expressed as: $\tilde{x}^t = (\frac{\sigma^t}{\max_t\{\sigma^t\}})^{\frac{1}{\alpha}}\frac{C}{F}, \forall t$ when $\gamma = 0$. Therefore, ISP's optimal revenue can be expressed as:

$$\begin{aligned} V_{PS} &= \sum_f \sum_t \sigma_f^t u_f(\tilde{x}_f^t) \\ &= F^{\alpha} C^{1-\alpha} \frac{1}{1-\alpha} (\max_t\{\sigma^t\})^{1-\frac{1}{\alpha}} \sum_t (\sigma^t)^{\frac{1}{\alpha}} \\ &= V \frac{\sum_t (\frac{\sigma^t}{\max_t\{\sigma^t\}})^{\frac{1}{\alpha}}}{\sum_t (\frac{\sigma^t}{\max_t\{\sigma^t\}})}. \end{aligned} \quad (34)$$

Appendix II: Proof of Proposition 2

Proof: We drop the flow index here for clear presentation. Constraint (16) requires $(\frac{\sigma^t}{h})^{\frac{1}{\alpha}} \leq x^t + \gamma, \forall t$. Therefore, there exists $h^{\frac{1}{\alpha}} \geq \max_t\{\frac{(\sigma^t)^{\frac{1}{\alpha}}}{x^t+\gamma}\} \geq \frac{\max_t\{(\sigma^t)^{\frac{1}{\alpha}}\}}{x^{t_0}+\gamma} \geq \frac{\max_t\{(\sigma^t)^{\frac{1}{\alpha}}\}}{\frac{C}{F}+\gamma}$, where $t_0 = \arg\max_t\{(\sigma^t)^{\frac{1}{\alpha}}\}$ (notice that $x^{t_0} \leq \frac{C}{F}$ because all flows are symmetric). Thus, according to constraint (15), $x^t \leq \min\{\frac{C}{F}, \frac{(\sigma^t)^{\frac{1}{\alpha}}}{\max_t\{(\sigma^t)^{\frac{1}{\alpha}}\}}(\frac{C}{F}+\gamma)\}, \forall t$. Therefore, by substituting this expression into ISP's revenue function (which is equal to all flows' sum-utility), we can get:

$$\frac{V_{PS}}{V} \leq \frac{\sum_{t\in\Omega_1}\sigma^t}{\sum_t \sigma^t} + \frac{(\frac{C}{F}+\gamma)^{1-\alpha}}{(\frac{C}{F})^{1-\alpha}} \frac{\sum_{t\in\Omega_2}\frac{(\sigma^t)^{\frac{1}{\alpha}}}{\max_t\{(\sigma^t)^{\frac{1}{\alpha}}\}}}{\sum_t \frac{\sigma^t}{\max_t\{\sigma^t\}}}, \quad (35)$$

where $\Omega_1 = \{t|\frac{(\sigma^t)^{\frac{1}{\alpha}}}{\max_t\{(\sigma^t)^{\frac{1}{\alpha}}\}}(\frac{C}{F}+\gamma) \geq \frac{C}{F}\}$, and $\Omega_2 = \{t|\frac{(\sigma^t)^{\frac{1}{\alpha}}}{\max_t\{(\sigma^t)^{\frac{1}{\alpha}}\}}(\frac{C}{F}+\gamma) < \frac{C}{F}\}$, and $\Omega_1 \cup \Omega_2 = \mathcal{T}$.□

APPENDIX III: PROOF OF DECREASING OF (23) WITH ξ

Proof: Let $\Phi^t(\alpha) = (\frac{\frac{C}{F}+\gamma}{\frac{C}{F}})^{1-\alpha}(\frac{\sigma^t}{\max_t\{\sigma^t\}})^{\frac{1}{\alpha}}, \forall t \in \Omega_2$. To show the upper bound (23) is increasing with α, it is sufficient to show that $\Phi^t(\alpha)$ is increasing with α.

$$\frac{d\Phi^t(\alpha)}{d\alpha} = (\frac{\frac{C}{F}+\gamma}{\frac{C}{F}})^{1-\alpha}(\frac{\sigma^t}{\max_t\{\sigma^t\}})^{\frac{1}{\alpha}}(-\ln(\frac{\frac{C}{F}+\gamma}{\frac{c}{F}}) - \frac{1}{\alpha^2}\ln(\frac{\sigma^t}{\max_t\{\sigma^t\}})).$$

Thus, to show $\frac{d\Phi^t(\alpha)}{d\alpha} \geq 0, \forall t \in \Omega_2$, it is equivalent to show $(\frac{\frac{C}{F}}{\frac{C}{F}+\gamma})^{\alpha^2}/(\frac{\sigma^t}{\max_t\{\sigma^t\}}) \geq 1$. This always holds for $t \in \Omega_2$ because $(\frac{\frac{C}{F}}{\frac{C}{F}+\gamma})^{\alpha^2} \geq (\frac{(\sigma^t)^{\frac{1}{\alpha}}}{(\max_t\{\sigma^t\})^{\frac{1}{\alpha}}})^{\alpha^2} = \frac{(\sigma^t)^{\alpha}}{(\max_t\{\sigma^t\})^{\alpha}} \geq \frac{(\sigma^t)}{(\max_t\{\sigma^t\})}$. The first inequality comes from the definition for the set Ω_2, which guarantees that $\frac{(\sigma^t)^{\frac{1}{\alpha}}}{(\max_t\{\sigma^t\})^{\frac{1}{\alpha}}}(\frac{C}{F}+\gamma) < \frac{C}{F}$. The last inequality holds because $\alpha < 1$. □

APPENDIX IV: PROOF OF PROPOSITION 4

Proof: We drop the flow index here for clear presentation. Constraint (26) requires $\sum_t \frac{(\sigma^t)^{\frac{1}{\alpha}}}{(h)^{\frac{1}{\alpha}}} \leq \sum_t x^t + \eta$. Therefore, there exists $(h)^{\frac{1}{\alpha}} \geq \frac{\sum_t(\sigma^t)^{\frac{1}{\alpha}}}{\sum_t x^t+\eta} \geq \frac{\sum_t(\sigma^t)^{\frac{1}{\alpha}}}{T\frac{C}{F}+\eta}$ (because all flows are symmetric, there exists $x^t \leq \frac{C}{F}, \forall t$). T denotes the length of the entire time horizon [10]. Thus, according to constraint (25), $x^t \leq \min\{\frac{C}{F}, \frac{(\sigma^t)^{\frac{1}{\alpha}}}{\sum_t(\sigma^t)^{\frac{1}{\alpha}}}(\frac{C}{F}T+\eta)\}, \forall t$. Therefore, by substituting this expression into ISP's revenue function (which is equal to all flows' sum-utility), we can derive the upper bound of $\frac{V_{LT}}{V}$ as follows:

$$\frac{V_{LT}}{V} \leq \frac{\sum_{t\in\Pi_1}\sigma^t}{\sum_t\sigma^t} + \frac{(\frac{C}{F}T+\eta)^{1-\alpha}}{(\frac{C}{F})^{1-\alpha}}\frac{\sum_{t\in\Pi_2}\frac{(\sigma^t)^{\frac{1}{\alpha}}}{\sum_t(\sigma^t)^{\frac{1}{\alpha}}}}{\sum_t\frac{\sigma^t}{(\sum_t(\sigma^t)^{\frac{1}{\alpha}})^{\alpha}}}, \quad (36)$$

where $\Pi_1 = \{t|\frac{(\sigma^t)^{\frac{1}{\alpha}}}{\sum_t(\sigma^t)^{\frac{1}{\alpha}}}(\frac{C}{F}T+\eta) \geq \frac{C}{F}\}$, and $\Pi_2 = \{t|\frac{(\sigma^t)^{\frac{1}{\alpha}}}{\sum_t(\sigma^t)^{\frac{1}{\alpha}}}(\frac{C}{F}T+\eta) < \frac{C}{F}\}$, and $\Pi_1 \cup \Pi_2 = \mathcal{T}$. □

APPENDIX V: PROOF OF DECREASING OF (33) WITH ξ

Proof: The second part of the right hand side of (33) is equal to $\frac{(\frac{C}{F}T+\eta)^{1-\alpha}}{(\frac{C}{F})^{1-\alpha}}(\sum_{t\in\Pi_2}\sigma^t(\frac{(\sigma^t)^{\frac{1}{\alpha}}}{\sum_t(\sigma^t)^{\frac{1}{\alpha}}})^{1-\alpha})\frac{1}{\sum_t\sigma^t}$. Thus, the increasing property of (33) with α is equivalent to the increasing property of $\frac{(\frac{C}{F}T+\eta)^{1-\alpha}}{(\frac{C}{F})^{1-\alpha}}(\frac{(\sigma^t)^{\frac{1}{\alpha}}}{\sum_t(\sigma^t)^{\frac{1}{\alpha}}})^{1-\alpha}, \forall t \in \Pi_2$ with respect to α, which is always true. It is because that $\frac{\frac{C}{F}T+\eta}{\frac{C}{F}}\frac{(\sigma^t)^{\frac{1}{\alpha}}}{\sum_t(\sigma^t)^{\frac{1}{\alpha}}} < 1$ always holds for $t \in \Pi_2$ according to its definition. □

[10] Notice that this lower bound of $(h)^{\frac{1}{\alpha}}$ is loose since we aggressively assume that $\widetilde{x}^t = \frac{C}{F}, \forall t$. Thus, the derived upper bound (33) is more tight when η is relatively large as shown in Figure 2. Meanwhile, the upper bound is more accurate when the time horizon length T is small.

APPENDIX VI: PROOF OF PROPOSITION 5

Proof: According to the feasible region of the usage-price $\widetilde{h}_f$, i.e., (22), there exists

$$V_{PS}^{usage} = \sum_f(\sum_t \widetilde{x}_f^t)\widetilde{h}_f \leq \sum_f(\sum_t \widetilde{x}_f^t)\min_t\{\frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}\} \leq \sum_f\sum_t \sigma_f^t(\widetilde{x}_f^t)^{1-\alpha_f}.$$

If $\alpha_f = \alpha, \forall f$, then $V_{PS}^{usage} \leq V_{PS}(1-\alpha)$. The equal sign holds when $\frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}}, \forall t$ is a flow dependent constant (i.e., $\frac{\sigma_f^t}{(\widetilde{x}_f^t)^{\alpha_f}} = \widetilde{h}_f, \forall t$), which corresponds to the case that $\Gamma_f^t = 0, \forall f, t$ for problem (RMP-PS). □

REFERENCES

[1] Telecom Regulartory Policy CRTC, "Review of the Internet Traffic Management Practices of Internet Service Providers", Oct. 21, 2009

[2] J.W. Roberts, "Internet Traffic, QoS, and Pricing", in *Proc. of The IEEE*, vol. 92, no. 9, pp. 1389-1399, Step. 2004

[3] P. Hande, M. Chiang, R. Calderbank, J. Zhang, "Pricing under Constraints in Access Networks: Revenue Maximization and Congestion Managment," in *Proc. IEEE INFOCOM'2010*

[4] D. Clark, W. Fang, "Explicit Allocation of Best-Effor Packet Delivery Service", in *IEEE/ACM Transactions on Networking*, vol. 6, no. 4, pp. 362-373, 1998

[5] M.A. Labrador, S. Baneerjee, "Packet Dropping Policies for ATM and IP Networks", in *IEEE Communication Survey*, vol. 2, no. 3, pp. 2-14, Third Quarter 1999.

[6] X. Zhang, S.F. Wu, F. Zhi, T. Wu, "Malicious packet dropping: how it might impact the TCP performance and how we can detect it", in *Proc. of ICNP'2000*, pp. 263-272, 2000

[7] S. Zander, G. Armitage, "Empirically Measuring the QoS Sensitivity of Interactive Online Game Players", in *Proc. ATNAC'2004*, pp. 511-518, 2004

[8] "Your Cell Phone Company's Dirty Little Secret", *Yahoo Finance*, Feb. 10, 2010

[9] A. Gupta, B. Jukic, M. Li, D. Stahl, and A. Whinston, "Estimating Internet Users' Demand Characteristics," *Computational Economics*, vol. 17, pp. 203-218, 2001

[10] W. Oi, "A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly," *Quarterly Journal of Economics*, vol. 85, pp. 77-96, 1971

[11] M. Falkner, M. Devetsikiotis, I. Lambadaris, "An Overview of Pricing Concepts for Broadband ISP Networks", *IEEE Communication Surveys*, vol. 3, no. 2, pp. 2-13, 2nd Quarter, 2000.

[12] M. Chiang, S. Low, A. Calderbank, J. Doyle, "Layering as Optimization Decomposition: A Mathematical Theory of Network Architectures", *Proceeding of the IEEE*, vol. 95, no. 1, pp. 255-312, Jan. 2007.

[13] J. Walrand, "Economics and Networks: An Introduction, Internet Tutorial", The University of California, Berkeley, 2009

[14] J.K. Mackie-Mason, H. Varian, "Pricing Congestible Network Resource", *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1141-1149, Sep. 1995.

[15] Q. Wang, D.M. Chiu, J.C.S Lui, "ISP Uplink Pricing in a Competitive Market", *in Proceedings of ICT2008*

[16] S.C.M. Lee, J.W.J. Jiang, D.M. Chiu, J.C.S. Lui, "Interaction of ISPs: Distributed Resource Allocation and Revenue Maximization", in *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 2, pp. 204-218, Feb, 2008

[17] S. Shakkottai, R. Srikant, "Economics of Network Pricing with Multiple ISPs", in *IEEE/ACM Transactions on Networking*, vol 14, no 6, pp. 1233-1245, 2006

[18] D. Wu, Y.T. Hou, and Y.Q. Zhang, "Scalable video coding and transport over broad-band wireless networks", in *Proceedings of the IEEE*, vol. 11, no. 6, pp. 1182-1193, Oct. 2001.

[19] US National Exchange of Carrier Association, https://www.neca.org/

[20] H.R. Varian, *Intermediate Microeconomics: A Modern Approach*.