

Optimal Pricing Schemes for Public Wireless Networks

Yuanzhang Xiao, William R. Zame, and Mihaela van der Schaar

Email: yxiao@ee.ucla.edu

I. INTRODUCTION

II. SYSTEM MODEL

We consider a public wireless network with one access point (AP) and several users. The wireless network is setup by a service provider (SP) to enable Internet connections in public areas such as parks, libraries, and cafes. The service provider establishes appropriate pricing schemes to charge the users in order to cover its own cost in the setup and maintenance of the public wireless network. According to their demand and the prices, users choose the best pricing schemes for themselves. We can regard this public network as a three-layer network, with the three layers as the technology layer, the application layer, and the economic layer. In the next three subsections, we will describe these three layers and their interaction in details.

Since the arrival and departure of users can take place at any time instant, we use a continuous-time model to describe the system. We denote t as the index of time. According to their arrival and departure processes and utility functions, the users are inherently categorized into K types, where the user type is indexed by k . We define the total number of type- k users, no matter in service or not, as N_k . Note that the type is the inherent property of a user, which is the user's private information. However, since the users of the same type will choose the same payment plan, the service provider can partially observe the type of a user by the payment plan that user subscribes to. In other words, the service provider categorizes the users by their payment plans. We call this categorization as the user's *payment type*. We assume that there are L payment plans. Then we can define the mapping from the inherent type k of a user to its payment type l as

$$\pi : \{1, \dots, K\} \rightarrow \{1, \dots, L\}. \quad (1)$$

At time t , the system state is characterized by the number of users of different types in the system, denoted as $\mathbf{X}(t) = [X_1(t), X_2(t), \dots, X_K(t)]$, where $X_k(t)$ as the random variable indicating the number of type- k users in the system. The state space is

$$\mathcal{N} = \{\mathbf{n} = [n_1, \dots, n_K] \mid 1 \leq n_k \leq N_k\}. \quad (2)$$

We assume that the arrival process for type- k users is a Poisson process with rate $\lambda_k(\mathbf{X}(t))$ dependent on the current system state $\mathbf{X}(t)$, and that the service time of a user is exponentially distributed with mean $1/\mu_k$. Since

the arrival rate of a user is dependent on the system state, the arrival process is not exogenous but influenced by the admission control policy, which is a parameter in the technology layer described as follows.

A. The Technology Layer

The technology layer is characterized by the admission control policy, the scheduling policy, and the medium access control (MAC) protocol. Once the technology used in the network is determined, we can calculate the expected throughput and delay of a user, two instrumental parameters in the technology layer that influence the application layer and the economic layer.

1) The admission control policy: The admission control policy determines whether a user should be admitted into the system, given the current system state and the payment type of the incoming user. Here we only consider coordinate convex policies defined in [?]. We rewrite the definition of coordinate convex policies as follows.

Definition 1: A policy is coordinate convex if there exists a subset $\Omega \subseteq \mathcal{N}$ such that 1) $\mathbf{n} \in \Omega$ and $n_k > 0$ imply $\mathbf{n} - \mathbf{e}_k \in \Omega$ for $k = 1, \dots, K$, 2) for any current state $\mathbf{n} \in \Omega$, a type- k user is accepted if and only if $\mathbf{n} + \mathbf{e}_k \in \Omega$. Here \mathbf{e}_k is the all-zero vector except for a one as the k th element.

For this class of coordinate convex policies, there is a one-to-one mapping between the policy and a subset $\Omega \subseteq \mathcal{N}$ satisfying the first condition in the above definition [?].

The simplest admission control policy would be

$$\Omega = \left\{ [n_1, \dots, n_K] \in \mathcal{N} : \sum_{k=1}^K n_k \leq N_0 \right\}, \quad (3)$$

where any user is admitted as long as the total number of current users does not exceed the maximum number of users allowed in the system.

If the service provider limits the number of users subscribing to the l th payment plans in the system, it can use the following policy

$$\Omega = \left\{ [n_1, \dots, n_K] \in \mathcal{N} : \sum_{k:\pi(k)=l} n_k \leq N_l^* \right\}, \quad (4)$$

2) The scheduling policy: For all the system states, the service provider needs to specify the scheduling policy that determines how users can access the channels. A scheduling policy can be defined as a function $\Psi : \mathcal{S} \rightarrow \mathbb{R}^{|\mathcal{S}|}$, which is a mapping from \mathcal{S} , a partition of the set of payment types $\{1, 2, \dots, L\}$, to the resource, such as channels and bandwidth, allocated to each element in the partition.

If the service provider cannot or does not choose to partition the resource, the scheduling policy is specified by $\mathcal{S} = \{1, 2, \dots, L\}$ and $\Psi(\{1, 2, \dots, L\}) = B$, where B is the total amount of resources (bandwidth or the number of channels) available at the access point. In this case, all the users access the AP in a centralized or contention-based manner.

The service provider can schedule the users of different payment types to different partitions of resource. For example, if the service provider can tell the difference among different types of users except that between type-1 and type-2 users, it can use the scheduling policy with resource partition $\mathcal{S} = \{\{1, 2\}, \{3\}, \dots, \{K\}\}$ and

$\Psi(s) = \frac{B}{K-1}, \forall s \in \mathcal{S}$. In this case, the users of each type are allocated an equal amount of resource, except that type-1 and type-2 users are treated as one type of users. There is no conflict in accessing the channels among different types of users.

3) *The multiple-access control protocol:* The service provider also needs to prescribe the MAC protocol Φ that determines how the users can access the channel. We can use time-division multiple access (TDMA) or orthogonal frequency-division multiple access (OFDMA) as a centralized orthogonal multiple access scheme, where each user accessing the channel has a dedicated time or frequency bin. We can also use slotted Aloha or carrier-sensing multiple access (CSMA) as a decentralized multiple access scheme, where each user accessing the channel transmits with some probability.

Now we calculate the expected throughput and delay of each user under different MAC protocols and some typical admission control and scheduling policies used later in this paper. We denote $\Theta = (\Omega, \Psi, \Phi)$ as the parameter that specifies the admission control policy, the scheduling policy, and the MAC protocol. Then the expected throughput of a type- k user is denoted as $\tau_k(\mathbf{X}(t), \Theta)$ and the expected delay of a type- k user as $d_k(\mathbf{X}(t), \Theta)$. Here the expected delay is defined as the expected waiting time of a user before the first successful transmission starting from an arbitrarily chosen time.

If the service provider uses the combination of admission control and scheduling policies with the lowest complexity, the admission control policy would be

$$\Omega = \mathcal{N} = \left\{ \mathbf{n} = [n_1, \dots, n_K] \mid \sum_{k=1}^K n_k \leq N_0 \right\}, \quad (5)$$

and the scheduling policy would be

$$\mathcal{S} = \{1, 2, \dots, K\} \quad (6)$$

and

$$\Psi(\{1, 2, \dots, K\}) = B, \quad (7)$$

where B is the total bandwidth of the AP. In this case, the expected throughput and delay of a user in different MAC protocols under the above admission control and scheduling policies are listed as follows.

- If the SP uses centralized MAC protocols such as TDMA and OFDMA, the expected throughput of a type- k user is

$$\tau_k(\mathbf{X}(t), \Theta) = \frac{B}{N_0}, \quad (8)$$

and the expected delay in time slots is

$$d_k(\mathbf{X}(t), \Theta) = \frac{N_0}{2}. \quad (9)$$

- If the SP uses slotted Aloha with transmission probability f as the MAC protocol, the expected throughput of a type- k user is

$$\tau_k(\mathbf{X}(t), \Theta) = B \cdot f(1-f)^{\|\mathbf{X}(t)\|_1 - 1}, \quad (10)$$

where $\|\mathbf{X}(t)\|_1$ is the 1-norm of $\mathbf{X}(t)$, and the expected delay in time slots is [?]

$$d_k(\mathbf{X}(t), \Theta) = \frac{1}{f(1-f)^{\|\mathbf{X}(t)\|_1-1}} - 0.5. \quad (11)$$

- If the SP uses CSMA with a constant backoff window W as the MAC protocol, the expected throughput and the expected delay in time slots of a type- k user is equal to those of a type- k user under slotted Aloha with transmission probability $f = \frac{2}{1+W}$ [?].

B. The Application Layer

The application layer is characterized by the users' utility, which depends in part on the throughput and delay from the technology layer. Since users stay in the system for a finite time, we use an instantaneous utility rate u_k of a type- k user as a characterization of its quality of service (QoS) experienced at a certain time instant. At time t , the utility rate of a type- k user is $u_k(\tau_k(\mathbf{X}(t), \Theta), d_k(\mathbf{X}(t), \Theta))$, a function of the throughput $\tau_k(\mathbf{X}(t), \Theta)$ and the delay $d_k(\mathbf{X}(t), \Theta)$. We assume that u_k satisfies 1) u_k is increasing and concave in the throughput τ_k and 2) u_k is decreasing in the delay d_k . The expected utility rate \bar{u}_k is

$$\bar{u}_k(t, \Theta) = \mathbb{E}_{\mathbf{X}} \{ u_k(\tau_k(\mathbf{X}(t), \Theta), d_k(\mathbf{X}(t), \Theta)) \}. \quad (12)$$

Then the expected discounted utility of a type- k user during any period of time from T_0 to T_1 is

$$\int_{T_0}^{T_1} e^{-\rho_k t} \bar{u}_k(t + mT, \Theta) dt, \quad (13)$$

where $e^{-\rho_k}$ is the continuous-time discounting factor of type- k users.

Here we assume that each user evaluates its QoS experience for every period of time of length T . Therefore, we define the discounted utility for the m th period, i.e. for $t \in [mT, (m+1)T)$, as

$$\bar{U}_k(m, \Theta) = \int_{t=0}^T e^{-\rho_k t} \bar{u}_k(t + mT, \Theta) dt. \quad (14)$$

We also assume that the net utility of a user has a quasi-linear form, which means that the net sum utility of a user during the m th period is the sum utility $\bar{U}_k(m, \Theta)$ subtracted by its payment to the service provider during that period. The amount of payment is decided by the pricing scheme the user subscribes to, which is a component in the economic layer introduced in what follows.

C. The Economic Layer

1) *Payment plans:* The payment plan consists of one or several of six basic fees, namely the subscription fee p_{sub} , the connection fee p_{con} , the per-minute flat rate p_{min} , the per-bit flat rate p_{bit} , the surcharge rate q_{sub} during each period, and the surcharge rate q_{con} during each connection. In summary, we define the payment plan as a vector

$$\mathbf{p} = [p_{sub}, p_{con}, p_{min}, p_{bit}, q_{sub}, \delta_{sub}, q_{con}, \delta_{con}]^T, \quad (15)$$

where δ_{sub} or δ_{con} is the thresholds above which the surcharge is collected between two consecutive submissions of subscription fees or during one connection, respectively.

Since the users of the same type are statistically identical, they will choose the same payment plan and have the same expected payment. We denote $c_k(\mathbf{X}(t), \mathbf{p})$ as the instantaneous payment rate of a type- k user at time t , and $\bar{c}_k(t, \mathbf{p})$ as the expectation of $c_k(\mathbf{X}(t), \mathbf{p})$ over \mathbf{X} .

The subscription fee p_{sub} is charged periodically. Here we assume that all the users pay the subscription fee at the beginning of every period of T time, the period the same as that during which they evaluate the QoS experience. Thus, the payment rate of a type- k user by paying subscription fee is

$$c_{k,sub}(\mathbf{X}(t), p_{sub}) = p_{sub} \cdot I(t = mT), \quad m = 0, 1, \dots \quad (16)$$

Since no randomness is involved in the subscription fee, we have $\bar{c}_{k,sub}(t, p_{k,sub}) = p_{sub} \cdot I(t = mT)$.

The connection fee p_{con} is paid when a user connects the network. The payment of a type- k user by paying connection fee is

$$c_{k,con}(\mathbf{X}(t), p_{con}) = p_{con} \cdot I(\text{the user is not in the network and is connecting}) \cdot I(\mathbf{X}(t) + \mathbf{e}_k \in \Omega). \quad (17)$$

Thus, the expected payment of a type- k user by paying connection fee is

$$\begin{aligned} \bar{c}_{k,con}(t, p_{con}) &= \mathbb{E}_{\mathbf{X}} \{c_{k,con}(t, p_{con})\} \\ &= p_{con} \cdot \sum_{\mathbf{n}: \mathbf{n} + \mathbf{e}_k \in \Omega} P(\mathbf{X}(t) = \mathbf{n}) \cdot \frac{M_k - n_k}{M_k} \cdot \frac{\lambda_k}{M_k}. \end{aligned} \quad (18)$$

The per-minute flat rate p_{min} is charged for every time slot a user stays in the system while remains connected to the service provider. The payment of a type- k user by paying per-minute flat fee is

$$c_{k,min}(\mathbf{X}(t), p_{min}) = p_{min} \cdot I(\text{the user is in the network}). \quad (19)$$

Thus, the expected payment of a type- k user by paying per-minute flat fee is

$$\begin{aligned} \bar{c}_{k,min}(t, p_{min}) &= \mathbb{E}_{\mathbf{X}} \{c_{k,min}(\mathbf{X}(t), p_{min})\} \\ &= p_{min} \cdot \sum_{\mathbf{n} \in \Omega} P(\mathbf{X}(t) = \mathbf{n}) \cdot \frac{n_k}{M_k}. \end{aligned} \quad (20)$$

The per-bit flat rate p_{bit} is charged for every bit a user consumes. The payment of a type- k user by paying per-bit flat fee is

$$c_{k,bit}(\mathbf{X}(t), p_{bit}) = p_{bit} \cdot I(\text{the user is in the network}) \cdot \tau_k(\mathbf{X}(t), \Theta). \quad (21)$$

Thus, the expected payment of a type- k user by paying per-bit flat fee is

$$\begin{aligned} \bar{c}_{k,bit}(t, p_{bit}) &= \mathbb{E}_{\mathbf{X}} \{c_{k,bit}(\mathbf{X}(t), p_{bit})\} \\ &= p_{bit} \cdot \sum_{\mathbf{n} \in \Omega} P(\mathbf{X}(t) = \mathbf{n}) \cdot \frac{n_k}{M_k} \cdot \tau_k(\mathbf{n}, \Theta). \end{aligned} \quad (22)$$

The surcharge is charged for additional data usage or minutes exceeding a certain amount prescribed by the service plan. The users may have to pay the surcharge q_{sub} for the additional data usage for the period between two submissions of subscription fees, if the data usage exceeds δ_{sub} . They may also have to pay the surcharge q_{con}

for the additional data usage exceeding δ_{con} during every connection. The payment of a type- k user by paying surcharge is

$$\begin{aligned} & c_{k,sur}(\mathbf{X}(t), q_{sub}, \delta_{sub}, q_{con}, \delta_{con}) \\ &= q_{sub} \cdot (b_{k,sub} - \delta_{sub})^+ \cdot I(t = mT) \\ &+ q_{k,con} \cdot I(\text{the user is not in the network and is connecting}) \cdot I(\mathbf{X}(t) + \mathbf{e}_k \in \Omega) \cdot (b_{k,con} - \delta_{con})^+ \end{aligned} \quad (23)$$

where $b_{k,sub}$ and $b_{k,con}$ are the data or minute consumed during the period between consecutive payments of subscription fee and during each connection, respectively. Thus, the expected payment of a type- k user by paying surcharge is

$$\begin{aligned} \bar{c}_{k,sur}(t, q_{sub}, \delta_{sub}, q_{con}, \delta_{con}) &= \mathbb{E} \{c_{k,sur}(\mathbf{X}(t), q_{sub}, \delta_{sub}, q_{con}, \delta_{con})\} \\ &= q_{sub} \cdot (\bar{b}_{k,sub} - \delta_{sub})^+ \cdot I(t = mT) \\ &+ q_{con} \cdot \sum_{\mathbf{n}: \mathbf{n} + \mathbf{e}_k \in \Omega} P(\mathbf{X}(t) = \mathbf{n}) \cdot \frac{M_k - n_k}{M_k} \cdot \frac{\lambda_k}{M_k} \cdot (\bar{b}_{k,con} - \delta_{con})^+, \end{aligned} \quad (24)$$

where $m = 0, 1, \dots$, and $\bar{b}_{k,sub}$ and $\bar{b}_{k,con}$ are the expected amount of data or minute consumed during the period between consecutive payments of subscription fee and during each connection, respectively.

Finally, the expected payment of a type- k user at time t is

$$\bar{c}_k(t, \mathbf{p}) = \bar{c}_{k,sub}(t, p_{sub}) + \bar{c}_{k,con}(t, p_{con}) + \bar{c}_{k,min}(t, p_{min}) + \bar{c}_{k,bit}(t, p_{bit}) + \bar{c}_{k,sur}(t, q_{sub}, \delta_{sub}, q_{con}, \delta_{con}), \quad (25)$$

and the discounted expected cost of a type- k user in the m th period, i.e. for time $t \in [mT, (m+1)T]$, is

$$\bar{C}_k(m, \mathbf{p}) = \int_{t=0}^{T} e^{-\rho_k t} \bar{c}_k(t + mT, \mathbf{p}) dt. \quad (26)$$

2) *Revenue and cost of the service provider:* In order to meet its budget, the service provider needs to evaluate the cost and revenue during the operation of the system. Here we assume that the service provider assesses its profits every T time, namely at $t = mT, m = 1, 2, \dots$, the same time when the users evaluate their utility. Thus, the revenue of the service provider in the m th period can be written as

$$\bar{R}(m, \mathbf{p}) = \int_{t=0}^{T} e^{-\rho_0 t} \sum_{k=1}^K M_k \bar{c}_k(t + mT, \mathbf{p}) dt, \quad (27)$$

where $e^{-\rho_0}$ is the discounting factor of the service provider.

The total cost of the service provider during the m th period is denoted as a constant $\bar{C}_0(m)$.

III. PROBLEM FORMULATION

This problem can be modeled as a two-stage game. At the first stage, the service provider announces a set of L available payment plans $\mathcal{P} = \{\mathbf{p}^1, \dots, \mathbf{p}^L\}$. At the second stage, each user myopically chooses the optimal payment plan that maximizes its own utility. The design problem of the service provider is to find the optimal set of plans \mathcal{P} and the optimal combination of technology Θ , so that at the end of the second stage, the social welfare

or the total revenue is maximized. In the following, we will describe the decision process of the users and the design problem of the service provider in details.

At stage one, the service provider announces the set of available payment plans \mathcal{P} . Meanwhile, the SP may also reveal some information $\hat{\Theta}$ corresponding to the deployed technology Θ , in order to help the users to choose the appropriate payment plans. Usually, the service provider does not reveal the full information about the technology, if necessary at all. Thus, in general $\hat{\Theta} \neq \Theta$. Instead, the service provider may advertise the estimated throughput $\hat{\tau}_k$ or delay \hat{d}_k for the payment plans, where $\hat{\Theta} = (\hat{\tau}_k, \hat{d}_k)$, or may reveal no information about the admission control and scheduling policies, where $\hat{\Theta} = \emptyset$.

At stage two, each user chooses the optimal payment plan from the set of available payment plans \mathcal{P} according to the available information from its own knowledge or revealed by the service provider. We assume that a user only knows its own arrival and departure rates and its utility function. The user chooses the optimal plan by solving the following optimization problem

$$\begin{aligned} \tilde{\mathbf{p}}_k = \arg \max_{1 \leq l \leq L} & \sum_{m \in \mathcal{M}} \bar{U}_k(m, \hat{\Theta}) - \bar{C}_k(m, \mathbf{p}^l) \\ \text{s.t.} & \bar{U}_k(m, \hat{\Theta}) - \bar{C}_k(m, \mathbf{p}^l) \geq 0, \quad \forall m \in \mathcal{M}. \end{aligned} \quad (28)$$

for some periods of time $\mathcal{M} \subseteq \mathbb{Z}_{++}$. We can regard the constraints $\bar{U}_k(m, \hat{\Theta}) - \bar{C}_k(m, \mathbf{p}^l) \geq 0$ for $m \in \mathcal{M}$ as the individual rationality (IR) constraint for the users, and the choice of utility-maximizing $\tilde{\mathbf{p}}_k$ as the result of the incentive compatibility (IC) constraint.

Note that the information $\hat{\Theta}$ revealed by the service provider determines the interaction between users in the second stage. If the service provider only advertises the expect throughput and delay, there will be no interaction between users. If the service provider provides some information about the technology Θ , the utility of a user will depend on the actions, namely the chosen payment plans, of the other users. In the latter case, the outcome of the second stage will be the Nash equilibrium of the payment selection game.

Now we can introduce the design problem of the service provider. We assume that the service provider knows the expected utility of the users¹. Hence, given the payment plans chosen by all the users $\{\tilde{\mathbf{p}}_k\}_{k=1}^K$, the service provider can calculate the social welfare, defined as the total utility of all the users, and the total revenue of its own. The design problem of the service provider is to find the optimal set of payment plans \mathcal{P}^* , the combination of technology Θ^* , and the revelation of information related to technology $\hat{\Theta}^*$, so that after the users have chosen their best plans, the social welfare or the total revenue of the service provider is maximized, depending on if the service provider is benevolent or selfish. In addition, the budget constraint (BC) of the service provider should be satisfied.

The design problem of the service provider can be mathematically formulated as follows. For a benevolent service

¹The service provider cannot know the utility of users perfectly. It can only estimate it by survey or some other methods. Here we assume that it has perfect knowledge of the utility for simplicity

provider aiming at maximizing the social welfare, its design problem (PB) can be written as

$$(PB) : \max_{\mathcal{P}, \Theta, \hat{\Theta}} \sum_{m \in \mathcal{M}} \sum_{k=1}^K \alpha_k^{mT} \cdot \bar{U}_k(m, \Theta) - \bar{C}_k(m, \tilde{\mathbf{p}}_k) \quad (29)$$

$$s.t. \quad IC : \tilde{\mathbf{p}}_k = \arg \max_{1 \leq l \leq L} \sum_{m \in \mathcal{M}} \bar{U}_k(m, \hat{\Theta}) - \bar{C}_k(m, \mathbf{p}^l), \quad k = 1, \dots, K, \quad (30)$$

$$IR : \bar{U}_k(m, \hat{\Theta}) - \bar{C}_k(m, \tilde{\mathbf{p}}_k) \geq 0, \quad k = 1, \dots, K, \quad \forall m \in \mathcal{M}, \quad (31)$$

$$BC : \bar{R}(m, \tilde{\mathbf{p}}) \geq \bar{C}_0(m), \quad \forall m \in \mathcal{M}, \quad (32)$$

where $\tilde{\mathbf{p}} = [\tilde{\mathbf{p}}_1^T, \dots, \tilde{\mathbf{p}}_K^T]^T$. The solution \mathcal{P}^* to the above problem provides the users with a set of payment plans to choose from. After each user chooses the payment plan that maximizes its own expected net utility, the system reaches the maximum social welfare.

Similarly, for a selfish service provider aiming at maximizing its own revenue, its design problem (PS) can be written as

$$(PS) : \max_{\mathcal{P}, \Theta, \hat{\Theta}} \sum_{m \in \mathcal{M}} \bar{R}(m, \tilde{\mathbf{p}}) \quad (33)$$

$$s.t. \quad IC : \tilde{\mathbf{p}}_k = \arg \max_{1 \leq l \leq L} \sum_{m \in \mathcal{M}} \bar{U}_k(m, \hat{\Theta}) - \bar{C}_k(m, \mathbf{p}^l), \quad k = 1, \dots, K, \quad (34)$$

$$IR : \bar{U}_k(m, \hat{\Theta}) - \bar{C}_k(m, \tilde{\mathbf{p}}_k) \geq 0, \quad k = 1, \dots, K, \quad \forall m \in \mathcal{M}, \quad (35)$$

$$BC : \bar{R}(m, \tilde{\mathbf{p}}) \geq \bar{C}_0(m), \quad \forall m \in \mathcal{M}, \quad (36)$$

IV. DETAILED ANALYSIS ON A TYPICAL SCENARIO

In this section, we study a typical scenario, where there are two types of users. The two types of users can be differentiated by their application layers. For example, one type of users is email users, who requires low throughput and tolerates large delay, while another type of users is video users with stringent throughput and delay requirements.

In order to derive analytical results and gain insights into the problem, rather than considering the original problems (PB) and (PS), we consider the approximate problems (PB(∞)) and (PS(∞)) with objectives and constraints only imposed at the point the system reaches its steady state. This approximation is justified by the fact that we are more concerned about the outcome at the steady state.

The approximate problem for (PB) can be written as

$$(PB(\infty)) : \max_{\mathbf{p}} \sum_{k=1}^K \bar{V}_k(\infty, \Theta, \mathbf{p}) \cdot M_k \quad (37)$$

$$s.t. \quad IR(\infty) : \bar{V}_k(\infty, \Theta, \mathbf{p}) \geq 0, \quad k = 1, \dots, K, \quad (38)$$

$$BC(\infty) : \bar{R}(\infty, \mathbf{p}) \geq \bar{C}_0(\infty), \quad (39)$$

where $\bar{V}_k(\infty, \Theta, \mathbf{p}) = \lim_{m \rightarrow \infty} \bar{V}_k(m, \Theta, \tilde{\mathbf{p}})$, and $\bar{R}(\infty, \mathbf{p})$ and $\bar{C}_0(\infty)$ are defined similarly. Note that the limit exists since the equilibrium distribution exists for the system with the maximum number of users allowed.

Similarly, the approximate problem for (PS) can be written as

$$(\text{PS}(\infty)) : \max_{\mathbf{p}} \bar{R}(\infty, \mathbf{p}) \quad (40)$$

$$\text{s.t.} \quad \text{IR}(\infty) : \bar{V}_k(\infty, \Theta, \mathbf{p}) \geq 0, \quad k = 1, \dots, K, \quad (41)$$

$$\text{BC}(\infty) : \bar{R}(\infty, \mathbf{p}) \geq \bar{C}_0(\infty), \quad (42)$$

A. Subscription with or without bit surcharge

Now we study how the network performance is affected by the service provider's capability of measuring the data usage of the users. Specifically, we consider two scenarios. In the first scenario, the service provider cannot measure the data usage of the users, thus it can only charge a subscription fee for each user. This may be because the SP uses the CSMA protocol, which cannot provide the precise amount of data usage of each user. In the second scenario, the service provider can measure the data usage, thus it will collect a per-bit surcharge in addition to the subscription fee. In both cases, we solve the approximate problems $(\text{PB}(\infty))$ and $(\text{PS}(\infty))$ with

$$\mathbf{p} = [p_{\text{sub}}, 0, 0, 0, q_{\text{sub}}, \delta_{\text{sub}}, 0, 0]^T. \quad (43)$$

The following two theorems give us the feasibility and optimality conditions for the design problems in the above two scenarios, respectively. Before stating the theorems, we would like to define a auxiliary variable $\bar{\rho}_i = \frac{e^{\rho_i T} - 1}{\rho_i}$ to express the sum discounting factor in the following. Also, we would like to remind the users that the positive utility rate of a type- k user at the steady state is $\bar{u}_k(\infty, \Theta)$, which is a key parameter characterizing the QoS experience.

The feasibility and optimality conditions in the first scenario is stated as follows

Theorem 1: Suppose that the service provider cannot measure the data usage, thus it can only charge a subscription fee for each user. Then the design problems $(\text{PB}(\infty))$ and $(\text{PS}(\infty))$ have feasible solutions if and only if

$$\min \{\bar{\rho}_1 \cdot \bar{u}_1(\infty, \Theta), \bar{\rho}_2 \cdot \bar{u}_2(\infty, \Theta)\} \cdot (M_1 + M_2) \geq \bar{C}_0(\infty). \quad (44)$$

Furthermore, the solution to the approximate problem $(\text{PB}(\infty))$ is

$$p_{\text{sub}}^* = \frac{\bar{C}_0(\infty)}{M_1 + M_2}, \quad (45)$$

and the social welfare at the optimum is

$$\bar{\rho}_1 \cdot \bar{u}_1(\infty, \Theta) + \bar{\rho}_2 \cdot \bar{u}_2(\infty, \Theta) - \bar{C}_0(\infty). \quad (46)$$

Finally, the solution to the approximate problem $(\text{PS}(\infty))$ is

$$p_{\text{sub}}^* = \min \{\bar{\rho}_1 \cdot \bar{u}_1(\infty, \Theta), \bar{\rho}_2 \cdot \bar{u}_2(\infty, \Theta)\}, \quad (47)$$

and the social welfare at the optimum is

$$\bar{\rho}_1 \cdot \bar{u}_1(\infty, \Theta) + \bar{\rho}_2 \cdot \bar{u}_2(\infty, \Theta) - \min \{\bar{\rho}_1 \cdot \bar{u}_1(\infty, \Theta), \bar{\rho}_2 \cdot \bar{u}_2(\infty, \Theta)\} \cdot (M_1 + M_2). \quad (48)$$

The feasibility and optimality conditions in the second scenario is stated as follows

Theorem 2: Suppose that the service provider can measure the data usage, thus it can collect per-bit surcharge in addition to a subscription fee. Then the design problems $(PB(\infty))$ and $(PS(\infty))$ have feasible solutions if and only if

$$\bar{\rho}_1 \cdot \bar{u}_1(\infty, \Theta) \cdot M_1 + \bar{\rho}_2 \cdot \bar{u}_2(\infty, \Theta) \cdot M_2 \geq \bar{C}_0(\infty). \quad (49)$$

Furthermore, the social welfare at the optimum of the approximate problem $(PB(\infty))$ is

$$\bar{\rho}_1 \cdot \bar{u}_1(\infty, \Theta) + \bar{\rho}_2 \cdot \bar{u}_2(\infty, \Theta) - \bar{C}_0(\infty). \quad (50)$$

Finally, the social welfare at the optimum of the approximate problem $(PS(\infty))$ is 0.

Remark: From the above two theorems, we can see that the feasible region becomes larger if the service provider can measure the data usage and charge for the excessive bits used by the users. Intuitively, if the SP can only charge the same subscription fee for all the users, the high-usage users, such as the video users, will have the incentives to use unlimited amount of data, which will congest the network and result in a negative net utility for the low-usage users, such as the email users. By imposing the surcharge, we can charge less for the email users and more for the video users so that both types of users have positive net utility. However, the selfish SP can use the surcharge to maximize its own revenue, so that both types of users get zero utility.

B. Numerical Simulation

Now we use numerical simulations to observe more details about the impact of the technology on the system performance. The key parameters in the simulation are described as follows:

- The period during which the service provider and users evaluate the revenue and utility is $T = 360$ hours/month, namely 12 hours/day times 30 days/month.
- We only consider the first two months, since the system reaches the limiting states before the end of the first month.
- The total number of users allowed in the system is $N_0 = 8$, and we use $\Omega = \mathcal{N}$ as the admission control policy.
- The SP uses the simple scheduling policy with $\mathcal{S} = \{1, 2, \dots, K\}$ and $\Psi(\{1, 2, \dots, K\}) = B = 10$, where $B = 10$ is the total throughput of the AP.
- The arrival rate and the departure rate of type-1 users, the email users, are fixed at $\lambda_1 = 1$ and $\mu_1 = 1$.
- The total number of type-2 users, the video users, is $M_2 = 2$. Their departure rate is fixed at $\mu_2 = 1$.
- The cost of the service provider is fixed every month, namely $\bar{C}(m) = 250, \forall m$.
- All the users and the service provider has the same discounting factor $e^\rho = 0.9999$.

The utility rate of a type- k user at time t is

$$\begin{aligned}
& u_k(\tau_k(\mathbf{X}(t), \Theta), d_k(\mathbf{X}(t), \Theta)) \\
&= \lim_{\Delta t} \frac{\text{sum utility from } t \text{ to } t + \Delta t}{\Delta t} \\
&= \lim_{\Delta t} \frac{I(\text{the user is in the network}) \cdot f_k(\tau_k(\mathbf{X}(t), \Theta), d_k(\mathbf{X}(t), \Theta)) \cdot \Delta t}{\Delta t} \\
&\quad - \lim_{\Delta t} \frac{I(\text{the user is outside and trying to connect}) \cdot I(X_k(t) = N_k) \cdot \pi_k}{\Delta t}
\end{aligned} \tag{51}$$

where $I(\cdot)$ is the indicator function, $f_k(\cdot, \cdot)$ is the positive utility rate as a function of the throughput and delay, and π_k reflects the dissatisfaction when a type- k user is denied access to the network.

Then the expected utility rate of a type- k user at time t is

$$\begin{aligned}
& \bar{u}_k(t, \Theta) \\
&= \mathbb{E}_{\mathbf{X}} \{u_k(\tau_k(\mathbf{X}(t), \Theta), d_k(\mathbf{X}(t), \Theta))\} \\
&= \lim_{\Delta t} \frac{\sum_{n_k=1}^{N_k} P(X_k(t) = n_k) \cdot \frac{n_k}{N_{k,\text{tot}}} f_k(\tau_k(\mathbf{X}(t), \Theta), d_k(\mathbf{X}(t), \Theta)) \cdot \Delta t}{\Delta t} \\
&\quad - \frac{(\lambda_k \Delta t \frac{N_{k,\text{tot}} - N_k}{N_{k,\text{tot}}}) \cdot P(X_k(t) = N_k) \cdot \pi_k}{\Delta t} \\
&= \sum_{n_k=1}^{N_k} P(X_k(t) = n_k) \cdot \frac{n_k}{N_{k,\text{tot}}} f_k(\tau_k(\mathbf{X}(t), \Theta), d_k(\mathbf{X}(t), \Theta)) \\
&\quad - (\lambda_k \frac{N_{k,\text{tot}} - N_k}{N_{k,\text{tot}}}) \cdot P(X_k(t) = N_k) \cdot \pi_k.
\end{aligned} \tag{52}$$

The positive utility rate f_1 for the email users (type-1 users) can be (first defined in [?])

$$f_1(\tau_1(\mathbf{X}(t), \Theta), d_1(\mathbf{X}(t), \Theta)) = \log(1 + \tau_1(\mathbf{X}(t), \Theta)) - w_1 \cdot d_1(\mathbf{X}(t), \Theta), \tag{53}$$

where $w_1 = 0.01$ is a weighting factor indicating the importance of the delay to a type-1 user.

Similarly, the positive utility rate f_2 for the video users (type-2 users) can be

$$f_2(\tau_2(\mathbf{X}(t), \Theta), d_2(\mathbf{X}(t), \Theta)) = \log(1 + \log(1 + \tau_2(\mathbf{X}(t), \Theta))) - w_2 \cdot d_2(\mathbf{X}(t), \Theta), \tag{54}$$

where $w_2 = 0.1$.

In the simulation, we change the arrival rates of the video users and get the maximum number of email users that can be supported in the system for each arrival rate of video users. The simulation results are shown in Fig. 1 and Fig. 2.

REFERENCES

- [1] T. Basar and R. Srikant, "Revenue-maximizing pricing and capacity expansion in a many-users regime," in *Proceedings IEEE INFOCOM 2002*, pp. 1556-1563, 2002.
- [2] H. Shen and T. Basar, "Differentiated Internet pricing using a hierarchical network game model," in *Proc. 2004 American Control Conference*, pp. 2322-2327, 2004.

- [3] H. Shen and T. Basar, "Optimal nonlinear pricing for a monopolistic network service provider with complete and incomplete information," *IEEE J. Select. Areas Commun.*, vol. 25, pp. 1216-1223, Aug. 2007.
- [4] M. H. Manshaei, J. Freudiger, M. Felegyhazi, P. Marbach, and J.-P. Hubaux, "On wireless social community networks," in *Proceedings IEEE INFOCOM 2008*, 2008.
- [5] A. Mazloumian, M. H. Manshaei, M. Felegyhazi, and J.-P. Hubaux, "Optimal pricing strategy for wireless social community networks," in *Workshop on the Economics of Networks, Systems, and Computation. ACM*, 2008.
- [6] M. H. Manshaei, P. Marbach, and J.-P. Hubaux, "Evolution and market share of wireless community networks," in *Proceedings of the first IEEE International Conference on Game Theory for Networks*, 2009.
- [7] D. P. Bertsekas and R. G. Gallager, "Data Networks," Prentice-Hall, 1992.
- [8] W. A. Massey and W. Whitt, "An analysis of the modified offered-load approximatoin for the nonstationary Erlang loss model," *The Annals of Applied Probability*, vol. 4, pp. 1145-1160, 1994.
- [9] J. Aein and O. S. Kosovych, "Satellite capacity allocatoin," *Proceedings of IEEE*, vol. 65, pp. 332-342, 1977.
- [10] K. W. Ross and D. Tsang, "The stochastic knapsack problem," *IEEE Trans. on Commun.*, vol. 37, no. 7, pp. 740-747, 1989.
- [11] J. Park and M. van der Schaar, "Medium access control protocols with memory", *IEEE/ACM Trans. Networking*, to appear.
- [12] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp 535-547, Mar. 2000.