Differential Pricing of 5G Network Slices for Heterogeneous Customers

Siying Chen
Viterbi School of Engineering
University of Southern California
Los Angeles, USA
siyingc@usc.edu

Bhaskar Krishnamachari
Viterbi School of Engineering
University of Southern California
Los Angeles, USA
bkrishna@usc.edu

***Abstract*—5G wireless networks promise to enable new kinds of cellular use cases, by offering different network slices to users with different needs. While pricing network bandwidth is relatively straightforward when all users care about data rate, more sophisticated pricing strategies are likely to emerge when some of the customers acquire a network slice with the desire to optimize for a different metric. We consider how a 5G wireless provider may set per-class differential prices to maximize its profit when offering different network slices to different classes of customers. We formulate the problem using fundamental economic principles and present a Drift-Plus-Penalty algorithm to solve the problem in a dynamic setting. We show, through simulation and analysis for a two-class network with latency and throughput-oriented customers, that some surprising phenomena may occur under certain conditions, such as an overall reduction of total resources sold when the relative number of latency-oriented customers grows.**

***Index Terms*— 5G network slices, differential pricing, DPP**

I. Introduction

With telecommunication technologies developing, 5G wireless networks are being developed and deployed to provide enhanced data rates, latency, density, and reliability. They are anticipated to play an increasingly role in society, enabling the emergence of a wide array of new applications such as mixed-reality, industrial IoT, connected Vehicles, smart cities and more [1]. One of the key novel techniques introduced in 5G systems is network slicing: this allows network resources to be divided into different slices on a per-service basis in order to dynamically accommodate heterogeneous demands.

In the network slicing system, the different slices can offer one or more different main functions based on special requirements. For example, an emergency communication slice may need to provide alert broadcast and distress call with low-rate but requiring latency on the order of seconds; and in a slice serving autonomous vehicles, high data rates may be needed [2]. Each slice consists of virtual network functions which could support and build up the network services, and it is created on-demand based on user requests to the provider. After that, the slices with corresponding functional blocks are made operative [3].

978-1-7281-3783-4/20/$31.00 ©2020 IEEE

There is a growing literature addressing network slicing architecture and technique, like SDN-NFV integration [2], [3]. Some have focused on utility maximization and tried to construct a business model in order to maximize the revenue, but emphasize resource allocation based on a single fixed given price instead of price setting [4].

From an economics perspective, the profit-maximization strategy of the seller is extracting all surplus of tenants. In other words, the seller will try to set the price as close as possible to the valuation of the item in users' opinion. However, in real life, it is really difficult to realize arbitrarily fine-grained price discrimination in the monopolist market. In most cases, the seller can profit more if consumers compete against each other like in an auction [5]. There are some works on optimal auctions and mechanisms designing [6], [7], [8]. In the telecommunication history, there did exist lots of charging principles with high complexity, especially price discrimination [9], [10]. However, the drawback is obvious as well: the mechanism is too complex to use in practice, and consumers dislike it compared with paying for simple prices [9].

In [9], [10], Odlyzko admits the high profits that complicated pricing forms can bring, but he predicts the trending of pricing form should be toward simplicity. Not surprisingly, there are a few works trying to solve the price problem by setting a simple price instead of complex mechanism designing in network economics in a traditional network system [11], [12], [13]. For example, in [11], for some common utility functions, Ozdaglar proves the profits of simple pricing scheme, which only charges a flat entry fee to all potential users, approximating the profits of revenue-maximizing prices in communication networks. Furthermore, Ozdaglar provides a new approach called marginal user principle to improve the provider's revenue, and with setting simple entry price as well [12]. However, there was only single service – rate, in the traditional model.

We believe it is necessary to revisit the question of pricing and develop price schemes that are well suited for different service needs or metrics, such as bandwidth, latency, reliability, energy efficiency and so on in 5G system. Concretely, we advocate a middle ground that balances profit maximization and resource efficiency with relative simplicity that would favor customer acceptance - allow a coarse-grained level of price differentiation, so that customers with fundamentally different metrics are charged different prices, but all customers within the same class of utilities are charged the same price (per unit resource).

To study this problem concretely, we build a model for a 5G infrastructure provider that must allocate resources to two classes of tenants: one that cares about bandwidth, and one that cares about latency. We focus on this two-class model for simplicity and ease of expositions, but the ideas presented in this work can be extended in a straightforward manner to more classes.

The paper is organized as follows. First we show how the provider could obtain the best demand curve from tenants' different utility functions for each class. Then, we try to find the point that the marginal revenue equals to the marginal cost to increase the total profits. In this case, the provider could determine the lower bound of entry price, and the upper bound of resource allocation for each slice in certain type. We then formulate the problem more generally to allow for joint resource optimization to determine a single price for each class. We present a Drift-plus-penalty (DPP) algorithm that allows for pricing in the presence of dynamic, unpredictable, arrivals. Finally, we simulate both static and dynamic conditions and analyze the resource allocation and profits obtained in different cases. We show that, as expected, having a differential price, one for each class, improves the provider's profit.

II. Model

A. Network Model

The network is logically divided into different network slices according to different demands, such as bandwidth, latency, reliability, and so on.

Let $\mathbb{S}$ denote the set of logical slices and r^s denote the network rate resource of slice $s, \forall s \in \mathbb{S}$. Suppose the $\mathcal{S}^s$ be the set of tenants who choose type s slice and define R_tot be the total rate resource that the provider has.

$$\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{S}^s} r_i^s \le R_{tot} \tag{1}$$

In general, we assume the group of tenants who request to slice $s \in \mathcal{S}$ has an increasing and concave utility function $U^s(r^s)$, which could be seen as the amount of money they would like to pay to the allocated rate r^s. Moreover, each tenant in the same group may have different utility gain coefficient $\gamma_i, \forall i \in \mathcal{S}^s$.

For ease of exposition, let us henceforth suppose the tenants just have two types of demand and classify them into two groups accordingly: the first group of tenants cares about latency while another is interested in bandwidth.

The first group of tenants demands steady rate allocation in order to ensure low latency in every time slot; and the second group has elastic demand that they would like to pay if the average throughput during a long period of time is above a certain threshold.

Assume there are M potential tenants in the first "latency" group and let $\mathcal{S}^{\mathcal{L}} = \{1, 2, ..., M\}$ denote the set of tenants; similarly, assume the "bandwidth" group size is N and the set $\mathcal{S}^{\mathcal{B}} = \{1, 2, ..., N\}$, where $\mathbb{S} = \mathcal{S}^{\mathcal{L}} \cup \mathcal{S}^{\mathcal{B}}$.

Define the utility function of latency and bandwidth: $U^L(r), U^B(r)$. As for these tenants, it is natural that they have different utility gains:

$$U_i^L(r_i) = \alpha_i U^L(r_i); \tag{2}$$

$$U_j^B(r_j) = \beta_j U^B(r_j); \tag{3}$$

where α, β is the utility gain coefficient of users, and it follows normal distribution:

$$\alpha_i \sim \mathcal{N}(\bar{\alpha},\ \sigma_\alpha^2), \forall i \in \mathcal{S}^{\mathcal{L}}$$
$$\beta_j \sim \mathcal{N}(\bar{\beta},\ \sigma_\beta^2), \forall j \in \mathcal{S}^{\mathcal{B}}$$

B. Business Model

First of all, We construct the business model with the following players:

- Infrastructure Provider: the owner of the network resource. The resources are virtualized and allocated to tenants corresponding to demands.
- Tenant: requests the virtual network resources from the InP, and uses these resources to serve users.
- User: purchases the network service from the tenants for themselves

In this paper, we only consider the game between infrastructure provider and tenants: tenants need to purchase the service that can satisfy their demands while the provider would like to maximize the revenue.

Let K denote the number of tenants with the same demand, and the tenant set is $\mathcal{S}^s = \{1, 2, , ..., K\}$. The utility function of tenant k is $U_k^s(r) = \gamma_k U^s(r), \forall k \in \mathcal{S}^f$, which is an increasing concave function.

In tenants' opinion, everyone has valued the service and would purchase it if and only if the value is greater or equal to the unit price, where the valuation function $V(r)$ is based on the utility function.

$$V_k^s(r) = \frac{U_k^s(r)}{r} > Price \tag{4}$$

In the market, we would like to find a best valuation function $V_{best}^s(r)$ of the certain slice type among tenants' different valuation functions $\{V_k^s(r)\}, k \in \mathcal{S}^f$, which is called the demand curve – represents the relationship between the unit rate price and amount of rate needed.

$$Price^s(r) = V_{best}^s(r) = f(\gamma_1, ..., \gamma_n)V^s(r) \tag{5}$$

The revenue $RV(r)$ and economic profits $\pi(r)$ that the infrastructure provider sales the certain type slice:

$$RV^s(r) = Price^s(r) \cdot r \tag{6}$$

$$\pi^s(r) = RV^s(r) - Cost^s(r) \tag{7}$$

Per classical microeconomics, the necessary condition for rate allocation that maximizes profits is that the marginal revenue must equal the marginal cost $MRV = MC$ [5]. In this model, we assume the marginal cost is fixed, i.e. $MC^s(r) = constant$.

$$MRV^s(r) = \frac{\mathrm{d}RV^s(r)}{\mathrm{d}r} \tag{8}$$

$$MC^s(r) = \frac{\mathrm{d}Cost^s(r)}{\mathrm{d}r} \tag{9}$$

Hence the provider could maximize the profits :

$$MRV^s(r) = MC^s(r) \tag{10}$$

and the total rate allocated to type "A" slice: $r \leq r^s_{max}$, where r^s_{max} is the solution of (10); such that $r^s_k \leq r^s_{max}$

Correspondingly, the price setting

$$Price^s \geq p^s_{min} = Price^s(r^s_{max}) \tag{11}$$

Similarly, if there are more than one type of slices, we could use the same method to ensure the price range for each type. Take latency and bandwidth we mentioned before as examples:

$$\begin{aligned} MRV^L(r^L_{max}) &= MC^L(r^L_{max}) \\ r^L_i &\leq r^L_{max} \\ Price^L &\geq p^L_{min}; \end{aligned} \tag{12}$$

$$\begin{aligned} MRV^B(r^B_{max}) &= MC^B(r^B_{max}) \\ r^B_j &\leq r^B_{max} \\ Price^B &\geq p^B_{min}; \end{aligned} \tag{13}$$

C. Allocation Rules

We have determined the pricing strategy to each type slice and the infrastructure provider could price differently with constraints to maximize the revenue. However, the above decoupled formulation doesn't clarify how resources should be split across the different slices, which requires a joint formulation. In general, the profit maximization problem can be expressed as a joint utility maximization:

$$\begin{aligned} max \quad & \sum_{s\in\mathcal{S}} \sum_{i\in\mathcal{S}^s} U^s_i(r^s_i) \\ s.t. \quad & r_0 \leq r^s_i \leq r^s_{max}, \forall i \in \mathcal{S}^s, s \in \mathbb{S} \\ & \sum_{s\in\mathbb{S}} \sum_{i\in\mathcal{S}^s} r^s_i \leq R_{tot} \end{aligned} \tag{14}$$

In our model, we have classified two groups and defined the tenants sets and utility functions.

$$\begin{aligned} max \quad & \sum_{i\in\mathcal{S}^L} U^L_i(r^L_i) + \sum_{j\in\mathcal{S}^B} U^B_j(r^B_j) \\ s.t. \quad & r^L_0 \leq r^L_i \leq r^L_{max}, \forall i \in \mathcal{S}^L \\ & r^B_0 \leq r^B_j \leq r^B_{max}, \forall j \in \mathcal{S}^B \\ & \sum_{i\in\mathcal{S}^L} r^L_i + \sum_{j\in\mathcal{S}^B} r^B_j \leq R_{tot} \end{aligned} \tag{15}$$

where r^L_0 and r^B_0 are the minimum demands of tenants in latency group and bandwidth group.

One approach to solving this problem is to use Lagrange multiplier theory with KKT constraints. In this paper, we instead present a drift-plus-penalty (DPP) algorithm to solve it [14], and we can simplify it (15) by using virtual queues and defining an appropriate domain set. This approach has the benefit of being simple to implement in a dynamic setting and not requiring prior knowledge of arrival rates of different classes.

- Define the domain set $\mathcal{R}$

$$\begin{aligned} \mathcal{R} = \{ & r^L_0 \leq r^L_i \leq r^L_{max}, \forall i \in \mathcal{S}^L; \\ & r^B_0 \leq r^B_j \leq r^B_{max}, \forall j \in \mathcal{S}^B \} \end{aligned} \tag{16}$$

- Define virtual queue $Q(t)$:

$$Q(t+1) = \max\left[Q(t) + g(r^L_i(t), r^B_j(t)) - R_{tot}, 0\right] \tag{17}$$

  where $Q(0) = 0$,

$$g(r^L_i(t), r^B_j(t)) = \sum_{i\in\mathcal{S}^L} r^L_i(t) + \sum_{j\in\mathcal{S}^B} r^B_j(t);$$

- The problem equals to choose $r^s(t)$ to minimize

$$-Wf(r^L_i(t), r^B_j(t)) + Q(t)g(r^L_i(t), r^B_j(t)); \tag{18}$$

  where

$$f(r^L_i(t), r^B_j(t)) = \sum_{i\in\mathcal{S}^L} U^L_i(r^L_i(t)) + \sum_{j\in\mathcal{S}^B} U^B_j(r^B_j(t))$$

  and W is a non-negative parameter that affects the amount minimization of f.
- Observe the virtual queue and choose $.r^L_i(t), r^B_j(t) \in \mathcal{R}$ to solve it.

III. ANALYSIS

A. Demand Curve Fitting

In our assumption above, the potential tenants in set $\mathcal{S^L} = \{1, 2, ..., M\}$ who care about latency should satisfy (2)

$$\begin{aligned} U^L_i(r_i) &= \alpha_i U^L(r_i); \\ \alpha_i &\sim \mathcal{N}(\bar{\alpha},\ \sigma^2_\alpha), \forall i \in \mathcal{S^L} \end{aligned}$$

They would like to purchase if and only if the valuation of the object is greater or equal to the price:

$$V^L_i(r_i) - Price > 0;$$

Suppose every tenant has the same latency demand $r^L_0 < r_{max}$, the valuation function could be expressed as:

$$V^L_i(r_i) = \frac{\alpha_i U^L(r_i)}{r_i} = \frac{\alpha_i U^L(r_0)}{r_0} = V\alpha_i; \tag{19}$$

where $V = U^L(r_0)/r_0$ is a constant. Hence the valuation function follows the normal distribution:

$$V^L_i(r_i) \sim \mathcal{N}(V\bar{\alpha}, V^2\sigma^2_\alpha) \tag{20}$$

In order to maximize the profit, it is necessary to set an appropriate demand curve to set price so that we can attract enough potential tenants.

For example, in Fig.2, there are four tenants with different utility gain coefficients, which means for the same service,

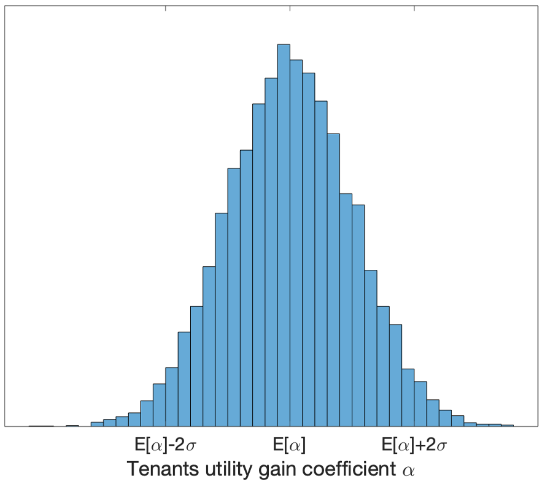

Fig. 1. Distribution of tenants' gain coefficient

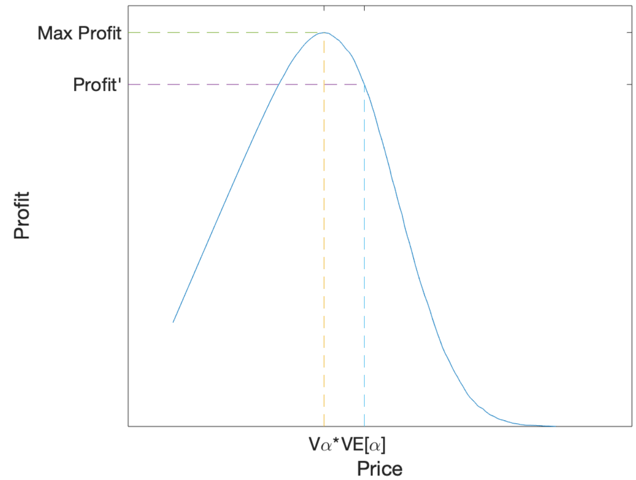

Fig. 3. Profit-α, Finding the best curve fitting

Fig. 2. Example of curve fitting

they would like to pay different amount of money because they gain differently. Assume the optimal valuation function, which could be regarded as the demand curve, is the valuation function of tenant 2:

$$V^L_{best}(r) = f(\alpha_1, ..., \alpha_4)V^L(r) = \alpha_2 V^L(r)$$

If we set the unit price same as the $V^L_{best}(r_{max})$, where r_{max} is the solution of (10), then we have:

$$\begin{cases} all\ tenants\ will\ purchase & if\ r \leq r' \\ tenants\ 2 \sim 4\ will\ purchase & if\ r' \leq r \leq r_{max} \end{cases}$$

In a more general case, assume the number of arrival potential tenants A_t follows a Poisson process of rate λ_L at each time slot t; s.t.

$$A_t = \lambda(t) \sim Poisson(\lambda_L) \tag{21}$$

The revenue of time slot t:

$$\begin{aligned} RV(t) &= TotalSold \cdot UnitPrice \\ &= (Pr_{[V^L_i(r_0) > V^*(r_0)]} \cdot A_t \cdot r_0) \cdot V^*(r_0) \\ &= [1 - \phi(V\alpha^*)]A_t r_0 V\alpha^* \end{aligned} \tag{22}$$

where $\phi(V\alpha)$ is the cumulative distribution function (CDF) of the normal distribution (20).

The average revenue:

$$\begin{aligned} E[RV] &= \frac{1}{T}\sum_{t=1}^{T} RV(t) \\ &= (1 - \phi(V\alpha^*))\lambda_L r_0 V\alpha^* \end{aligned} \tag{23}$$

Maximize the revenue: $\partial E[RV]/\partial V\alpha^* = 0$

$$\Rightarrow 1 - \phi(V\alpha^*) - V\alpha^*\phi'(V\alpha^*) = 0 \tag{24}$$

Hence, we could find the best utility function, called demand curve, of tenants who care about latency:

$$\begin{aligned} Price^L(r) &= V^L_{best}(r) = \alpha^* V^L(r) \\ U^L_{best}(r) &= V^L_{best}(r) * r = \alpha^* U^L(r) \end{aligned} \tag{25}$$

where α^* is the solution of (24), $\alpha^* \approx 0.8\bar{\alpha} < \bar{\alpha}$.

$$1 - \phi(V\alpha^*) - V\alpha^*\phi'(V\alpha^*) = 0$$

Similarly, as for tenants in the bandwidth group:

$$\begin{aligned} Price^B(r) &= V^B_{best}(r) = \beta^* V^B(r) \\ U^B_{best}(r) &= V^B_{best}(r) * r = \beta^* U^B(r) \end{aligned} \tag{26}$$

where β^* is the solution of

$$1 - \phi(V\beta^*) - V\beta^*\phi'(V\beta^*) = 0$$

and $\phi(V\beta)$ is the CDF of $\mathcal{N}(V\bar{\beta}, V^2\sigma^2_\beta)$

B. Lyapunov Optimization

Based on the lower bound of price from price discrimination, we could determine the resource allocation and set the price in order to maximize the total profits. In practice, it is related to the arrival requests at each time slot. We have assumed it follows a Poisson process (21) s.t.

$$A_t^L = \lambda^L(t) \sim Poisson(\lambda_L);$$
$$A_t^B = \lambda^B(t) \sim Poisson(\lambda_B);$$

Previously, we have found the best demand curve (25), (26) that profits most in such circumstance. In the ideal case, suppose the arrived tenant in the same group has the same demand r_0^L, r_0^B and everyone would purchase.

Then the problem (15) is converted to

$$\begin{aligned} Max \quad & A_t^L U_{best}^L(r^L(t)) + A_t^B U_{best}^B(r^B(t)) \\ s.t. \quad & A_t^L r^L(t) + A_t^B r^B(t) \leq R_{tot} \\ & r^L(t) \leq r_{max}^L \\ & r^B(t) \leq r_{max}^B \\ & r^L(t) \geq r_0^L \\ & r^B(t) \geq r_0^B \end{aligned} \tag{27}$$

Define: Set $\mathcal{R}$

$$\mathcal{R} = \{r_0^L \leq r^L(t) \leq r_{max}^L; r_0^B \leq r^B(t) \leq r_{max}^B\}$$

Lyapunov function $L(t)$ and Lyapunov drift $\Delta(t)$ [14]

$$L(t) = \frac{1}{2}Q(t)^2 \tag{28}$$

$$\Delta(t) = L(t+1) - L(t) \tag{29}$$

where $Q(t)$ is the virtual queues defined in (17).

$$Q(t+1) = \max[Q(t) + A_t^L r^L(t) + A_t^B r^B(t) - R_{tot}, 0]$$

The $L(t)$ is a parameter to decide the virtual queue size so the lower the value of $L(t)$ is, the more stable the system is. The drift-plus penalty inequality:

$$\begin{aligned} \Delta(t) - W(A_t^L U_{best}^L(r^L(t)) + A_t^B U_{best}^B(r^B(t))) \leq \\ B - W(A_t^L U_{best}^L(r^L(t)) + A_t^B U_{best}^B(r^B(t))) \\ + Q(t)[A_t^L r^L(t) + A_t^B r^B(t) - R_{tot}] \end{aligned} \tag{30}$$

where W is a non-negative weight parameter and B is the upper bound of $\frac{1}{2}(A_t^L r^L(t) + A_t^B r^B(t) - R_{tot})^2$.

Moreover, the problem is seperable to minimize the following equation with constraint $r^L(t), r^B(t) \in \mathcal{R}$ and it is separable:

$$\begin{aligned} [-W A_t^L U_{best}^L(r^L(t)) + Q(t) A_t^L r^L(t)] \\ +[-W A_t^B U_{best}^B(r^B(t)) + Q(t) A_t^B r^B(t)] \end{aligned} \tag{31}$$

For example, take $U^L(r) = a - b/(r-\lambda) + c\lambda$, $U^B(r) = dlog(1+r)$. Then we can select optimized $r^L(t)$ and $r^B(t)$:

$$r^L(t) = \left[(\frac{Wb\alpha^*}{Q(t)})^{\frac{1}{2}} + \lambda\right]_{r_0^L}^{r_{max}^L} \tag{32}$$

$$r^B(t) = \left[\frac{Wd\beta^*}{Q(t)} - 1\right]_{r_0^B}^{r_{max}^B} \tag{33}$$

C. Price Setting

The following can then be used to set the price for each slice based on the resource allocation from DPP algorithm.

$$\begin{aligned} Price^L(r^L) = V_{best}^L(\bar{r}^L(t)) \\ Price^B(r^B) = V_{best}^B(\bar{r}^B(t)) \end{aligned} \tag{34}$$

IV. Simulation

In this section, we simulate the cases that the resource allocated, price setting and profits with number of latency group tenants increasing while total number of both groups is fixed.

We generated two groups of tenants whose utility gain coefficients follow same normal distribution: $\alpha \sim \mathcal{N}(5,1); \beta \sim \mathcal{N}(5,1)$; and have same group size $M = N = 1000, \mathcal{S}^L = \mathcal{S}^B = \{1, 2, ..., 1000\}$.

Set fixed marginal cost $MC = 0.5$ and two marginal revenue function, we have calculated the $r_{max}^L < r_{max}^B$, so that we have $p_{min}^L > p_{min}^B$, which means allocating resource to latency group is more profitable than to bandwidth group.

A. Static Case

In the ideal static case, we have a fixed total tenants number; assume the tenants in the same group have the same demand r_0^L, r_0^B and the price satisfies the purchase constraint (4) for everyone.

Fix the total arrivals $A = A^L + A^B = 1000$, take $A^L \in \{100, 200, ..., 900\}$. From Fig.4, with A^L increasing, we see, as expected, that the resource allocated to latency group increases while the resource allocated to bandwidth group decreases.

Moreover, surprisingly, we observe that in this case the total resource consumed actually decreases slightly when A^L is large enough. We present below an analysis of the condition under which this would happen.

We have calculated the r_{max}^L and r_{max}^B where $r_{max}^L < r_{max}^B$, which are the solutions of equation (10) (MRV = MC). The resource sold $R_{sold} = A^L r^L + A^B r^B$. There are three cases that can occur:

1) $R_{tot} < r_{max}^L \cdot A = r_{max}^L \cdot (A^L + A^B) < r_{max}^L \cdot A^L + r_{max}^B \cdot A^B$; In this case, we can always sell the total resource $R_{sold} = R_{tot}, \forall \ A^L \in [0, N]$; and we could have $r^L \leq r_{max}^L, r^B \leq r_{max}^B$;
2) $r_{max}^L \cdot A < R_{tot} < r_{max}^B \cdot A$; in this case, there exists a threshold A_{th}, s.t.:

$$R_{tot} = A_{th} \cdot r_{max}^L + (A - A_{th}) \cdot r_{max}^B;$$

- If $A^L << A_{th}$, we have $A^L \cdot r_{max}^L + A^B \cdot r_{max}^B > R_{tot}$:

$$\Rightarrow R_{sold} = A^L R^L + A^B R^B = R_{tot}$$

where $R^L \leq r_{max}^L, R^B \leq r_{max}^B$ and can be solved by resource allocation algorithm;
- With A^L increasing and $A^L \to A_{th}$, $R^L \to r_{max}^L$ and $R^B \to r_{max}^B$; $R_{sold} = R_{tot}$;

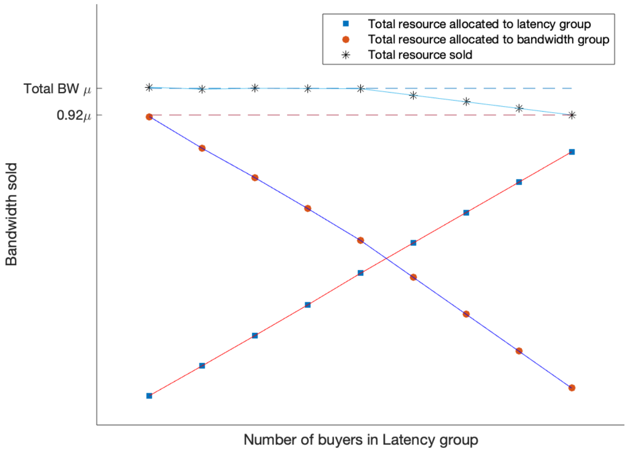

Fig. 4. Resource allocation

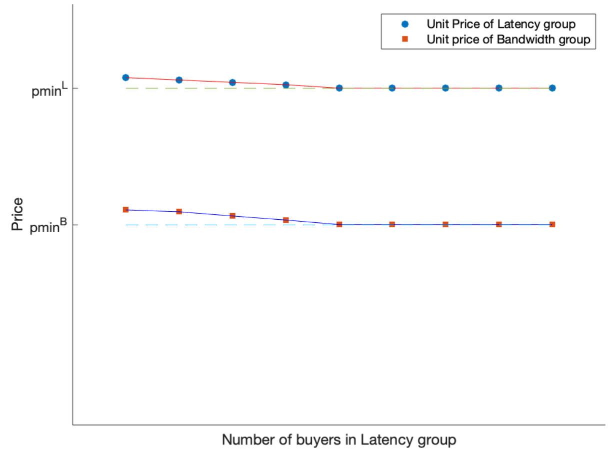

Fig. 6. Price set to each slice

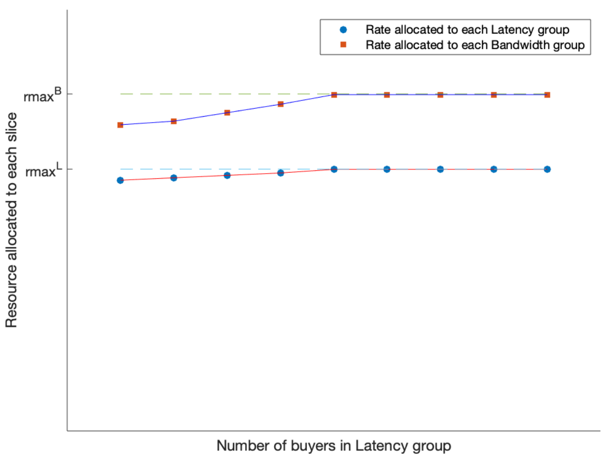

Fig. 5. Resource allocated to each slice

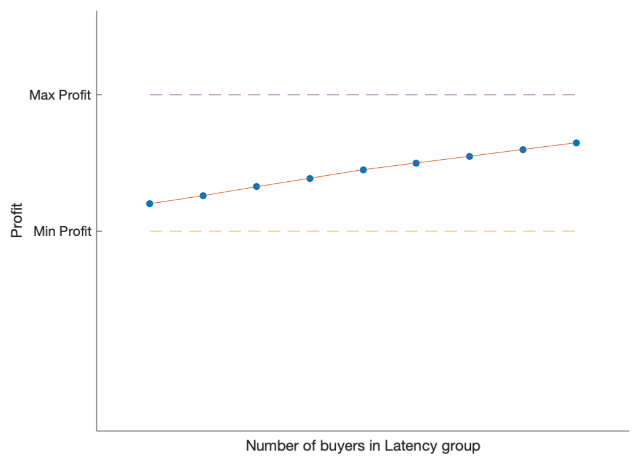

Fig. 7. Total profits

- If $A^L > A_{th}$, the resource sold would decrease because $R_{tot} > A^L r^L_{max} + A^B r^B_{max}$; This is the surprising case observed in our simulations.

3) $R_{tot} > r^B_{max} \cdot A > r^B_{max} \cdot (A^L + A^B) > r^L_{max} \cdot A^L + r^B_{max} \cdot A^B$; In this case, the provider has enough resource so that he could allocate r_{max} to every slice: $r^L = ramx^L, r^B = r^B_{max}$; but $R_{sold} < R_{tot}, \forall\ A^L \in [0, A]$.

In Fig.6, with resource allocated increasing slightly, the unit price of each slice decreases because $Price(r) = U(r)/r$, which is a decreasing convex function of r.

In Fig.7, it shows the trend of total profits earned. We have known that $p^L_{min} > p^B_{min}$, which means it is more profitable to allocate resource to latency group. With A^L increasing, the total profit $= A^L r^L_{max} p^L_{min} + A^B r^B_{max} p^B_{min}$ increases, although the provider cannot sell all resource.

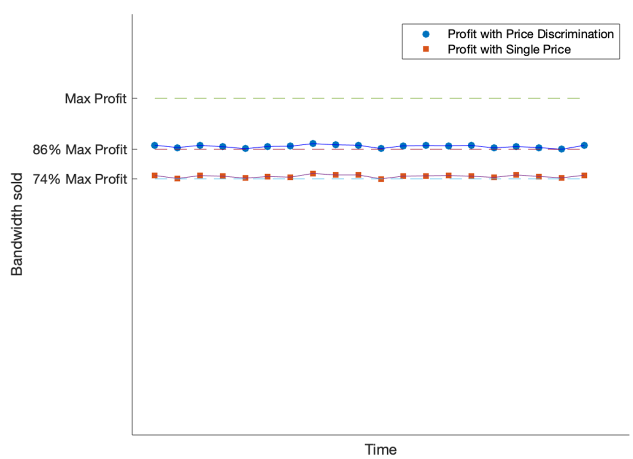

Fig. 8. Profit from Price Discrimination vs. Single Price

B. Dynamic Case

In the ideal static case, the total buyers are fixed and everyone will purchase at every time slot, so that we define the profit in such case as "Max Profit".

In dynamic case, the number of arrival requests follows the Poisson distribution $\lambda_L = \lambda_B = N/2; A_t^L \sim Poisson(500), A_t^B \sim Poisson(500)$.

Take $A_t^L = \lambda_L$ and $A_t^B = \lambda_B$ into DPP algorithm to converge the best resource allocation strategy, and we can finish corresponding price setting. The profit earned by price discrimination is 86% of the max profit while the profit earned by best single price is only 74% of the max.

V. Conclusion and Future Work

In this paper, we have presented the price setting principle based on simple price discrimination, where the provider sets different entry prices according to different services provided. We did not only focus on utility optimization but also considered demand curve fitting based on different potential tenants' gain coefficient, to determine the optimum price for each group.

The approach we have presented improves profits compared to the single-price scheme while keeping complexity relatively low (single price for each type of service).

In the future, we would like to extend our work into more general and complex settings, such as more complex variations and combinations of utility functions, and understand their implication on resource allocation and price-setting. Evaluating the proposed approach of per-class differential pricing of network slices with subjective survey-type feedback on the complexity of the service and price offerings may also be of interest to the 5G industry.

VI. Acknowledgement

This material is based upon work supported by Defense Advanced Research Projects Agency (DARPA) under Contract No. HR001117C0053. Any views, opinions, and/or findings expressed are those of the author(s) and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

References

[1] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5g: Survey and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, 2017.

[2] X. Zhou, R. Li, T. Chen, and H. Zhang, "Network slicing as a service: enabling enterprises' own software-defined cellular networks," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 146–153, 2016.

[3] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira, "Network slicing for 5g with sdn/nfv: Concepts, architectures, and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 80–87, 2017.

[4] D. Bega, M. Gramaglia, A. Banchs, V. Sciancalepore, K. Samdanis, and X. Costa-Perez, "Optimising 5g infrastructure markets: The business of network slicing," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.

[5] C. M. Snyder, W. Nicholson, and R. Stewart, *Microeconomic theory: Basic principles and extensions*. South-Western Cengage Learning, 2012.

[6] R. B. Myerson, "Optimal auction design," *Mathematics of operations research*, vol. 6, no. 1, pp. 58–73, 1981.

[7] J. Cremer and R. P. McLean, "Full extraction of the surplus in bayesian and dominant strategy auctions," *Econometrica: Journal of the Econometric Society*, pp. 1247–1257, 1988.

[8] Z. Han, D. Niyato, W. Saad, T. Başar, and A. Hjørungnes, *Game theory in wireless and communication networks: theory, models, and applications*. Cambridge university press, 2012.

[9] A. Odlyzko, "The evolution of price discrimination in transportation and its implications for the internet," *Review of Network Economics*, vol. 3, no. 3, 2004.

[10] ——, "Internet pricing and the history of communications," *Computer networks*, vol. 36, no. 5-6, pp. 493–517, 2001.

[11] S. Shakkottai, R. Srikant, A. Ozdaglar, and D. Acemoglu, "The price of simplicity," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1269–1276, 2008.

[12] D. Acemoglu, A. Ozdaglar, and R. Srikant, "The marginal user principle for resource allocation in wireless networks," in *2004 43rd IEEE Conference on Decision and Control (CDC)(IEEE Cat. No. 04CH37601)*, vol. 2. IEEE, 2004, pp. 1544–1549.

[13] R. El Azouzi, E. Altman, and L. Wynter, "Telecommunications network equilibrium with price and quality-of-service characteristics," in *Teletraffic Science and Engineering*. Elsevier, 2003, vol. 5, pp. 369–378.

[14] L. Georgiadis, M. J. Neely, L. Tassiulas *et al.*, "Resource allocation and cross-layer control in wireless networks," *Foundations and Trends® in Networking*, vol. 1, no. 1, pp. 1–144, 2006.