Multi-Dimensional Contract Design for Mobile Data Plan with Time Flexibility

Zhiyuan Wang
The Chinese University of Hong Kong
wz016@ie.cuhk.edu.hk

Lin Gao
Harbin Institute of Technology
gaol@hit.edu.cn

Jianwei Huang
The Chinese University of Hong Kong
jwhuang@ie.cuhk.edu.hk

ABSTRACT

Mobile network operators (MNOs) have been offering mobile data plans with different data caps and subscription fees as an effective way of achieving price discrimination and improving revenue. Recently, some MNOs are investigating innovative data plans with *time flexibility* based on the *multi-cap scheme*. The rollover data plan and the credit data plan are such innovative data plans with time flexibility. In this paper, we study how the MNO optimizes its multi-cap data plan with time flexibility in the realistic asymmetric information scenario, where each user is associated with multi-dimensional private information, i.e., the data valuation and the network substitutability. Specifically, we consider a multi-dimensional contract-theoretic approach, and analyze the optimal data caps and the subscription fees design systematically. We find that each user's *willingness-to-pay* for a particular data cap can be captured by the slope of his indifference curve on the contract plane, and the feasible contract (satisfying the incentive compatibility and individual rationality conditions) will allocate larger data caps for users with higher willingness-to-pay. Furthermore, we conduct a market survey to estimate the statistical distribution of users' private information, and examine the performance of our proposed multi-dimensional contract design using the empirical data. Numerical results further reveal that the optimal contract may provide price discounts (i.e., negative subscription fees) to attract low valuation users to select a small-cap (possibly zero-cap) contract item. A data mechanism with better time flexibility brings users higher payoffs and the MNO more profit, hence increases the social welfare.

CCS CONCEPTS

• **Networks → Network economics**; • **Applied computing →** *Consumer products*;

KEYWORDS

Rollover data plan, Multi-dimensional contract, Time flexibility.

This work is supported in part by the General Research Funds (Project Number CUHK 14219016) established under the University Grant Committee of the Hong Kong Special Administrative Region, China, and in part by the National Natural Science Foundation of China (Grant Number 61771162).

Mobihoc '18, June 26–29, 2018, Los Angeles, CA, USA

ACM ISBN 978-1-4503-5770-8/18/06…$15.00
https://doi.org/10.1145/3209582.3209588

ACM Reference Format:
Zhiyuan Wang, Lin Gao, and Jianwei Huang. 2018. Multi-Dimensional Contract Design for Mobile Data Plan with Time Flexibility. In *Mobihoc '18: The Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing, June 26–29, 2018, Los Angeles, CA, USA*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3209582.3209588

1 INTRODUCTION

1.1 Background and Motivation

The growing competition in the telecommunication market is putting an increasing pressure on the mobile network operators (MNOs) in terms of the innovation of their data plan offerings [18]. The *traditional* data plan is a two-part tariff, which is defined as a monthly one-time subscription fee and a data cap. A linear price will be charged for any data usage exceeding the data cap. However, due to the stochastic nature of a user's data demand, it is often a challenging task for the user to decide the optimal choice of the monthly data cap to achieve the best trade-off between the possible *wasting data* within the data cap and the possible *overage usage* when consuming beyond the data cap. Such a challenge might discourage users from subscribing to the MNOs' services. This motivates MNOs to explore various innovative data mechanisms that provides more flexibility and hence encourage user subscriptions.

Several innovative data plans explore various diversities of the mobile data across different dimensions: time [9], location [14][6], and user [3, 19, 26]. In particular, rollover data plans allows users to explore time flexibility of data usage, by allowing the unused data from the previous month to roll over to the current month. Different rollover data plans can be classified according to the consumption priority between the rollover data and the monthly data cap. For example, the rollover data plan offered by AT&T requires the rollover data to be consumed after the current monthly data cap [1], while the rollover data plan offered by China Mobile requires the rollover data to be consumed before the current monthly data cap [2].

A common feature of various rollover data plans is that a user can only use the "remaining" data from the previous month(s). In a more general case, we can allow users to "borrow" their data quota from future months, which we call the *credit data plan* [22]. In fact, the rollover data plan and the credit data plan represent two different ways of exploiting the data dynamics across the time dimension, backward and forward, under a common framework of time flexibility. The traditional data plan is a degenerated case of such a framework with no time flexibility.

In reality, a profitable price discrimination strategy for the MNO is to offer different data caps, each corresponding to a different monthly subscription fee, to attract different types of users. This motivates us to study how the MNO optimizes the *multi-cap* mechanism design for a mobile data plan with *time flexibility*.

1.2 Related Literature

Although the rollover data plan has been widely implemented in practice, there are very few theoretical studies regarding such a novel data mechanism. Zheng *et al.* in [28] found that moderately price-sensitive users can benefit from subscribing to the rollover data plan. Wei *et al.* in [25] analyzed the impact of different rollover period lengths from the MNO's perspective. In our previous work [21, 23, 24], we focused on the optimal design and MNOs' competition for mobile data plan with time flexibility under the single-cap scheme.

As for the MNO's multi-cap offering with the rollover data, there is no existing systematic study in the literature. A key challenge for this problem is that different users make their data cap choices based on their individual preferences, which are often private and may involve multi-dimensional information. Hence the MNO needs to properly design multiple data caps to differentiate users without knowing their private information and maximize the MNO's profit. Such a problem naturally leads to a contract design analysis.

The contract problem with users' single-dimensional private information has been extensively studied (e.g., see [5, 8, 11–13, 27]). Users' multi-dimensional private information leads to a multi-dimensional contract design problem. Such a problem is often very challenging, because the related model often violates the single-crossing condition [4, 16], which makes it difficult to achieve the global incentive compatibility. In this paper we introduce users' willingness-to-pay by investigating their indifference curves, so that we can develop a tractable approach for the MNO to provide the global incentive to all user types and solve its optimal contract under multi-dimensional private information.

1.3 Key Results and Contributions

The key results and contributions are summarized as follows:

- *Multi-Cap Design for Mobile Data Plan with Time Flexibility:* To the best of our knowledge, this is the first work studying the MNO's optimal multi-cap design for mobile data plan with time flexibility, including the traditional data plan, two existing rollover data plans, and the credit data plan.
- *Contract with Multi-dimensional Private Information:* We develop a tractable approach to analytically solve the MNO's contract problem in the presence of multi-dimensional private information. Specifically, our analysis shows that the slope of each user type's indifference curve on the contract plane corresponds to his willingness-to-pay, and a feasible contract (satisfying the Incentive Compatibility and Individual Rationality conditions) should allocate the data caps according to users' willingness-to-pay.
- *Empirical Implementation:* We conduct a market survey to estimate the statistical distribution of users' private information, and examine the performance of our proposed multi-dimensional contract design using the empirical data.
- *Performance and Insights:* The simulation based on the empirical results shows the optimal contract may offer some low-valuation users a small cap (possibly zero cap) together with a price discount (i.e., negative subscription fee); and a data mechanism with better time flexibility can increase users' payoffs and the MNO's profit under the optimal contract, hence the social welfare.

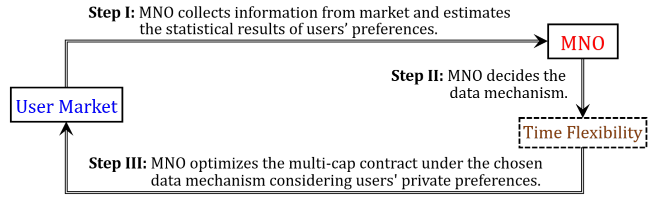


Figure 1: System model for the MNO's multi-cap design.

The rest of the paper is organized as follows. In Section 2, we introduces the system model. In Section 3, we we study the four data mechanisms with time flexibility in details. In Section 4, we analyze and solve the MNO's optimal contract. In Section 5, we provide numerical results to validate our analysis. Finally, we conclude this paper in Section 6.

2 SYSTEM MODEL

We formulate the MNO's multi-cap data plan design as a three-step process as shown in Fig. 1. In Step I, the MNO collects data from the user market to estimate the *statistical information* of users' individual preferences (i.e., a user's type), which are often *private* and *multi-dimensional* information (hence difficult to predict on a per user basis) . In Step II, the MNO chooses a data mechanism to provide subscribers with time flexibility. Then in Step III, the MNO proceeds with multi-cap contract design to make users truthfully reveal their types and maximize the MNO's profit. Generally speaking, the MNO should periodically (e.g., every year) repeat the three steps to capture users' varying requirements (due to, for example, technology changes).

Furthermore, the MNO should extract as many dimensions of the user type as possible to characterize users' private information precisely, which leads to a contract problem with multi-dimensional private information. As mentioned as Section 1, a multi-dimensional contract is very challenging to solve. To cope with users' multi-dimensional private information, in this paper, we propose to characterize each type of users' willingness-to-pay by investigating their indifference curves. To provide a clear demonstration, in this paper, we use a two-dimensional user type to illustrate our approach. In reality, the MNO can further introduce more dimensions and solve the multi-dimensional contract using a similar method.

In the following, we first describe four data mechanisms offering different levels of time flexibility. Then we introduce users' two-dimensional characteristics and summarize users' payoffs under different data mechanism in a unified model. Finally, we formulate the MNO's optimal contract problem.

2.1 Data Mechanisms

Various mobile data plans can be characterized by the tuple $\mathcal{T}_i = \{Q_i, \Pi_i\}$, where a $\mathcal{T}_i$ subscriber pays a fixed lump-sum subscription fee Π_i for a data usage up to the cap Q_i, beyond which the MNO

Table 1: $\mathcal{T}_i \triangleq \{Q_i, \Pi_i\},\ i \in \{0,1,2,3\}$

Plan	Special Data	τ	Priority	$Q_i^e(\tau)$
$\mathcal{T}_0$	None	0	Cap	Q_0
$\mathcal{T}_1$	Rollover data	$\tau \in [0, Q_1]$	Cap⇒Rollover	$Q_1 + \tau$
$\mathcal{T}_2$	Rollover data	$\tau \in [0, Q_2]$	Rollover⇒Cap	$Q_2 + \tau$
$\mathcal{T}_3$	Credit data	$\tau \in [-Q_3, 0]$	Cap⇒Credit	$2Q_3 + \tau$

will charge additional fee π for each unit of data consumption.[1] Here $i \in \{0,1,2,3\}$ represents different data mechanisms that offer subscribers time flexibility on their data consumption over time.

The key differences among the four data mechanisms are the *special data*[2] and *consumption priority*, both of which will affect the subscriber's expected overage usage. The special data can enlarge a user's *effective data cap* within which no additional fee involved. And the consumption priorities of the special data and the monthly data cap further affect how much the effective cap can be enlarged. In Table 1, we use τ to denote a user's data surplus ($\tau \geq 0$) or data deficit ($\tau < 0$) from the previous month, and use $p_i(\tau)$ to denote the probability mass function of τ. More specifically,

- $\mathcal{T}_0$ denotes the traditional data plan. The subscriber has no special data, and the effective cap of each month is $Q_0^e(\tau) = Q_0$;
- $\mathcal{T}_1$ denotes the rollover data plan (offered by AT&T). The special data of the current month is the rollover data surplus $\tau \in [0, Q_1]$ from the previous month, which is consumed *after* the current monthly data cap. Thus, the effective cap of the current month is $Q_1^e(\tau) = Q_1 + \tau$;
- $\mathcal{T}_2$ denotes the rollover data plan (offered by China Mobile). The special data of the current month is the rollover data surplus $\tau \in [0, Q_2]$ from the previous month, which is consumed *prior* to the current monthly data cap. Thus, the effective cap of the current month is $Q_2^e(\tau) = Q_2 + \tau$;
- $\mathcal{T}_3$ denotes the credit data plan. The special data of the previous month is the credit data deficit $\tau \in [-Q_3, 0]$ borrowed from the current month, which is consumed *after* the previous monthly data cap. Thus, the effective cap of the current month is $Q_3^e(\tau) = 2Q_3 + \tau$ (including the data that can be borrowed from the next month).

The rollover data plan and the credit data plan correspond to exploiting the time flexibility in the backward fashion and forward fashion, respectively. The traditional data plan is a special case with no time flexibility.

As we mentioned above, the time flexibility can enlarge the subscriber's effective data cap. According to Table 1, the effective data cap of the traditional data plan $\mathcal{T}_0$ is always Q_0, while the maximal value of the effective data cap is $2Q_i$ for $\mathcal{T}_1$, $\mathcal{T}_2$, and $\mathcal{T}_3$. The larger the effective data cap is, the less additional payment is incurred, which will further change users' subscription choices.

[1]For simplicity, we assume that all data plans have the same additional unit usage fee π. This is often true in practice. For example, for AT&T, π = \$15/GB.

[2]The special data refers to the data that a user can use free of charge on top of the monthly data cap, e.g., the rollover data from the previous month.

2.2 User Model

Next we characterize users' stochastic data demand, and introduce users' two-dimensional characteristics: θ for the valuation of unit data and β for the network substitutability.

To capture the stochastic nature of a user's data demand over time, we model a user's data demand as a discrete random variable with a probability mass function $f(d)$, a mean value of $\bar{d}$, and a finite integer support $\{0, 1, 2, ..., D\}$.[3] Here the data demand is measured in the minimum data unit (e.g, 1KB or 1MB according to the MNO's billing practice). Accordingly, we denote θ as a user's utility from one unit of data consumption, i.e., his valuation for unit data [15][20].

Furthermore, a user's data consumption behavior might change after exceeding the *effective cap*, since it incurs additional payment. Intuitively, the user will still continue to consume data in this case, but may reduce his data consumption by utilizing alternative networks (such as WiFi) instead. Therefore, we follow [17] by incorporating users' network substitutability β as one of the user's characteristics. Mathematically speaking, $\beta \in [0, 1]$ denotes the fraction of overage usage shrink. A larger β value represents more overage usage cut (thus, a better substitutability). A user's mobility pattern can significantly influence the availability of alternative networks, which will further change a user's data plan choice [23]. For example, a businessman who is always on the road may have a poor network substitutability (hence a small value of β), thus, prefers to a large data cap; while a student can take advantage of the school Wi-Fi network (hence a large value of β), thus, a small cap is more suitable.

Different from our previous work [23], in this paper, we consider a realistic *asymmetric information* scenario, i.e., θ and β are each user's private information that the MNO cannot know precisely. As a result, the MNO can use the contract-theoretic approach to cope with users' multi-dimensional private information and optimize its multi-cap data plans with time flexibility.

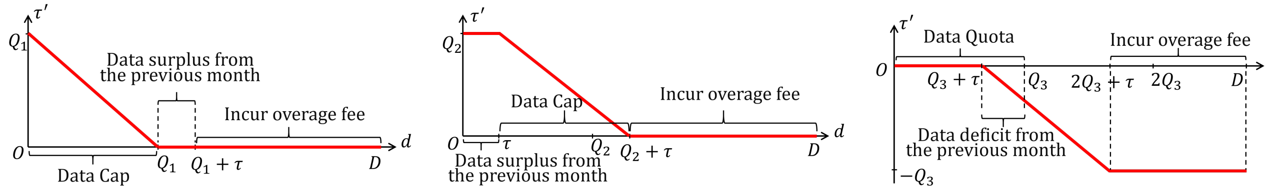
Therefore, a type-(θ, β) $\mathcal{T}_i = \{Q_i, \Pi_i\}$ subscriber's payoff with d units data demand and an effective cap $Q_i^e(\tau)$ is given by

$$\begin{aligned} S_i(\theta, \beta, d, \tau) =& \theta(d - \beta[d - Q_i^e(\tau)]^+) \\ & - \pi(1-\beta)[d - Q_i^e(\tau)]^+ - \Pi_i, \end{aligned} \tag{1}$$

where $[x]^+ = \max\{0, x\}$, $\theta(d - \beta[d - Q_i^e(\tau)]^+)$ is the user's utility due to data consumption, $\pi(1-\beta)[d - Q_i^e(\tau)]^+$ is the user's possible additional payment due to his data usage above the effective cap, and Π_i is the monthly lump-sum subscription fee. Here d and τ are two random variables, and we take the expectation over d and τ to obtain a user's expected payoff as

$$\begin{aligned} \bar{S}_i(\theta, \beta) &= \mathbb{E}_{d,\tau}\left\{S_i(\theta, \beta, d, \tau)\right\} \\ &= \theta[\bar{d} - \beta A_i(Q_i)] - \pi(1-\beta)A_i(Q_i) - \Pi_i, \end{aligned} \tag{2}$$

[3]In practice, the MNO can estimate users' demand distributions based on their historical data, and incorporate such a difference among users into the user type modeling. In this paper, we assume homogeneous demand distribution among users and focus on the user differences in data evaluation and network substitutability. Even for homogeneous distribution, the realized data demands are still different across users, which means some users end up as light users and other end up as heavy users.



(a) $\mathcal{T}_1$ (b) $\mathcal{T}_2$ (c) $\mathcal{T}_3$

Figure 2: Transition from τ (data surplus or deficit of the current month) to τ' (data surplus or deficit of the next month).

where $A_i(Q_i)$ is the type-$(\theta,0)$ $\mathcal{T}_i$ subscriber's expected overage data usage, as follows:

$$\begin{aligned} A_i(Q_i) =& \mathbb{E}_{d,\tau}\left\{[d - Q_i^e(\tau)]^+\right\} \\ =& \textstyle\sum_\tau \sum_d [d - Q_i^e(\tau)]^+ f(d) p_i(\tau), \end{aligned} \tag{3}$$

Note that the differences among the four data mechanisms are entirely captured by $A_i(Q_i)$ in (3). Specifically, $p_i(\tau)$ in (3) represents the distribution of the $\mathcal{T}_i$ subscriber's data surplus or deficit, which is the key difference among the four data mechanisms. Next we will explain how to compute $A_i(Q_i)$ in detail.

2.3 Time Flexibility

For $\mathcal{T}_0$ subscribers, the data surplus or deficit is always zero, i.e., $\tau = 0$, since it does not offer subscribers any special data. For $\mathcal{T}_1$, $\mathcal{T}_2$, and $\mathcal{T}_3$ subscribers, Fig. 2 illustrates the transition of users' data surplus or deficit τ between two successive months. To be specific, the horizontal axis corresponds to users' random data demand $d \in [0, D]$, and the vertical axis represents users' data surplus (or deficit) τ' of the next month given his data surplus (or deficit) τ in the current month and the data demand d. The differences among the three red curves in Fig. 2 indicate the differences among the three innovative data mechanisms.

To have a better understanding, next we take $\mathcal{T}_3$ as an example to illustrate how to compute $A_3(\cdot)$, and refer interested readers to the relevant Section 4 in [21] for more details.

For a $\mathcal{T}_3$ subscriber, the special data is the credit data borrowed from the next month, which is used after the current monthly data cap. Therefore, the effective data cap of a $\mathcal{T}_3$ subscriber with a data deficit $\tau \in [-Q_3, 0]$ consists of the remaining current monthly data cap (with a deficit τ) and the maximum credit data that he can borrow from the next month (which is Q_3), i.e., $Q_3^e(\tau) = 2Q_3 + \tau$. According to Fig.2(c), the data deficit in the next month, denoted by τ', is given by

$$\tau' = \begin{cases} 0, & \text{if } d \in [0, Q_3 + \tau], \\ Q_3 + \tau - d, & \text{if } d \in (Q_3 + \tau, 2Q_3 + \tau), \\ -Q_3, & \text{if } d \in [2Q_3 + \tau, D]. \end{cases} \tag{4}$$

Here we note that the data deficit τ' in the next month depends on both the data demand d and the data deficit τ in the current month, which indicates a Markov property of the data deficit τ in $\mathcal{T}_3$. The one-step transition probability of the data deficit τ is

$$p_3(\tau, \tau') = \begin{cases} \sum_{d=0}^{Q_3+\tau} f(d), & \text{if } \tau' = 0, \\ f(Q_3 + \tau - \tau'), & \text{if } \tau' \in (-Q_3, 0), \\ \sum_{d=2Q_3+\tau}^{D} f(d), & \text{if } \tau' = -Q_3. \end{cases} \tag{5}$$

We can derive the stationary distribution $p_3(\tau)$ based on the above transition probability. Thus, $A_3(Q_3)$ is given by

$$A_3(Q_3) = \sum_{\tau=-Q_3}^{0} \sum_{d=0}^{D} [d - Q_3^e(\tau)]^+ f(d) p_3(\tau). \tag{6}$$

The above analysis shows that data plans $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$ bring additional time flexibility to users comparing with $\mathcal{T}_0$. But how about the time flexibility among $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$? Following the convention in [21], we define the time flexibility of a data mechanism $\mathcal{T}_i$ as $\mathcal{F}(\mathcal{T}_i)$. Then we can define the following.

Definition 2.1 (Comparison of Time Flexibility). A data mechanism $\mathcal{T}_i$ has a better time flexibility than a data mechanism $\mathcal{T}_j$, i.e., $\mathcal{F}(\mathcal{T}_i) > \mathcal{F}(\mathcal{T}_j)$, if and only if $A_i(Q) < A_j(Q)$ for all Q.

In the above definition, we use the type-$(\theta,0)$ $\mathcal{T}_i$ subscriber's expected overage data usage $A_i(Q)$ to indicate the time flexibility of this data mechanism. Intuitively, a user experiences less overage usage under a data mechanism with more flexibility. Our previous work [21] has proved *Proposition 2.2*, which indicates that $\mathcal{T}_2$ and $\mathcal{T}_3$ offer subscribers the same and the best time flexibility, while $\mathcal{T}_0$ offers the worst.

Proposition 2.2. *The time flexibility of the four data mechanisms satisfies $\mathcal{F}(\mathcal{T}_0) < \mathcal{F}(\mathcal{T}_1) < \mathcal{F}(\mathcal{T}_2) = \mathcal{F}(\mathcal{T}_3)$. In other words, $A_0(Q) > A_1(Q) > A_2(Q) = A_3(Q)$ for any Q.*

2.4 MNO's Contract Formulation

Next we formulate the MNO's two-dimensional contract problem under the unified model with time flexibility. Before that, let us fist generalize the user's payoff in a unified expression, since a user's expected payoffs under different mechanisms have a similar expression. The only difference is the expected overage usage $A_i(Q_i)$. Thus, we are able to express the expected payoff of a type-(θ,β) $\mathcal{T}_i$ subscriber as

$$\bar{S}_i(Q_i, \Pi_i, \theta, \beta) = V_i(Q_i, \theta, \beta) - P_i(Q_i, \beta) - \Pi_i, \tag{7}$$

where $V_i(Q_i, \theta, \beta) \triangleq \theta[\bar{d} - A_i(Q_i)\beta]$ is the $\mathcal{T}_i$ subscriber's utility, and $P_i(Q_i, \beta) \triangleq \pi(1-\beta) A_i(Q_i)$ is the overage payment.

Therefore, we unify the four data mechanisms in a unified framework with time flexibility, and our later analysis for the contract design is based on this general framework. For notation simplicity, in Sections 3 and 4 we will ignore the data plan index i, and write a type-(θ,β) subscriber's payoff under $\{Q, \Pi\}$ as

$$\bar{S}(Q, \Pi, \theta, \beta) = V(Q, \theta, \beta) - P(Q, \beta) - \Pi. \tag{8}$$

In economics, the subscription fee can be viewed as a user's *sunk cost* (incurred in advance and often independent of the user's actual consumption), while the overage payment is the *prospective cost* (depending on the user's actual consumption). For notational convenience, we call the user's payoff without the sunk cost as the "virtual payoff", defined as

$$L(Q,\theta,\beta) = V(Q,\theta,\beta) - P(Q,\beta), \tag{9}$$

which will be used in Section 3.1.4.

The MNO offers a contract (with different combinations of data caps and corresponding subscription fees) to a population of users who are distinguished by two-dimensional private information: the data valuation θ and the network substitutability β. According to the statistical information from the user market, the MNO can flexibly divide users' into several categories.[4] For example, a set $\Theta = \{\theta_k : 1 \le k \le K\}$ of K valuation types and a set $\mathcal{B} = \{\beta_m : 1 \le m \le M\}$ of M network substitutability types. Thus, there are a total of KM types of users in the market, characterized by a joint probability mass function $q(\theta_k, \beta_m)$ for each k and m. Without loss of generality, we assume that users' types are indexed in the ascending sort order in both dimensions, i.e., $\theta_1 < \theta_2 < \ldots < \theta_K$ and $\beta_1 < \beta_2 < \ldots < \beta_M$. The case of MNO's single-cap design is a special case with $K = M = 1$.

According to the revelation principle [7], it is enough for the MNO to consider a class of contracts that enables users to truthfully reveal their types. In other words, it is enough for the MNO to design a contract, denoted by $\Phi(\Theta,\mathcal{B}) = \{\phi_{k,m}, 1 \le k \le K, 1 \le m \le M\}$ that consists of KM contract items $\phi_{k,m} = \{Q(\theta_k,\beta_m), \Pi(\theta_k,\beta_m)\}$, one for each user type. Formally, a contract is *feasible* if and only if it ensures each user selects the contract item intended for this type. It is obvious that a contract is *feasible* if and only if it satisfies Incentive Compatibility (IC) and Individual Rationality (IR) defined as follows:

Definition 2.3 (Incentive Compatibility). A contract is incentive compatible if each type-(θ_k,β_m) user maximizes its payoff by choosing the contract item $\phi_{k,m}$ intended for this user type, i.e.,

$$\bar{S}(\phi_{k,m},\theta_k,\beta_m) \ge \bar{S}(\phi_{l,n},\theta_k,\beta_m),\ \forall\, \phi_{l,n} \ne \phi_{k,m}. \tag{10}$$

Definition 2.4 (Individual Rationality). A contract is individually rational if each type-(θ_k,β_m) user achieves a non-negative payoff by choosing the contract item $\phi_{k,m}$ intended for this user type, denoted by $\phi_{k,m} \succcurlyeq 0$, i.e.,

$$\bar{S}(\phi_{k,m},\theta_k,\beta_m) \ge 0. \tag{11}$$

Our later analysis involves the concept of Pairwise Incentive Compatibility (PIC), defined as follows:

Definition 2.5 (Pairwise Incentive Compatibility). The contract items $\phi_{k,m}$ and $\phi_{l,n}$ are pairwise incentive compatible, denoted by $\phi_{k,m} \overset{\text{IC}}{\Longleftrightarrow} \phi_{l,n}$, if and only if

$$\begin{cases} \bar{S}(\phi_{k,m},\theta_k,\beta_m) \ge \bar{S}(\phi_{l,n},\theta_k,\beta_m), \\ \bar{S}(\phi_{l,n},\theta_l,\beta_n) \ge \bar{S}(\phi_{k,m},\theta_l,\beta_n). \end{cases} \tag{12}$$

Next we derive the MNO's revenue, cost, and profit under a feasible contract.

[4]In practice, the MNO can achieve the partition of user types through some data mining techniques such as k-means. The choices of parameters K and M determine the trade-off between contract complexity and profit.

The MNO's revenue from a subscriber consists of the subscription fee and the overage fee. Based on the above discussion of the feasible contract, the MNO's expected revenue $R(\Phi)$ under a *feasible* contract $\Phi(\Theta,\mathcal{B})$ is

$$R(\Phi) = \sum_{k=1}^{K}\sum_{m=1}^{M} q(\theta_k,\beta_m)\Big[\underbrace{\Pi(\theta_k,\beta_m)}_{\text{subscription}} + \underbrace{P(Q(\theta_k,\beta_m),\beta_m)}_{\text{overage}}\Big]. \tag{13}$$

Furthermore, we consider two kinds of costs suffered by the MNO, i.e., the capital expenditure (CapEx) and operational expenditure (OpEx).

The MNO's capital expenditure (CapEx) is mainly due to its investment on the network capacity. Imposing the data cap would help manage the network congestion and arrange the scarce network capacity. Motivated by this phenomenon, we model the MNO's CapEx caused by a type-(θ_k,β_m) subscriber as an increasing function $J(Q)$ on his data cap Q. Intuitively, a larger data cap corresponds to a severer network congestion that requires the MNO's more investment on the network in advance.

The MNO's operational expenditure (OpEx) is mainly due to the system management. After the MNO decides which data plan to implement, the subscribers' total data consumption will influence the MNO's operational expense. Therefore, the MNO's OpEx caused by a type-(θ_k,β_m) subscriber with data cap Q can be formulated as $c \cdot U(Q,\beta_m)$, where c is the MNO's marginal cost for the system management, and $U(Q,\beta_m) = \bar{d} - \beta_m A(Q)$ is the type-(θ_k,β_m) subscriber's data consumption.

Therefore, the MNO's expected cost $C(\Phi)$ under a *feasible* contract $\Phi(\Theta,\mathcal{B})$ can be calculated as

$$C(\Phi) = \sum_{k=1}^{K}\sum_{m=1}^{M} q(\theta_k,\beta_m)\Big[\underbrace{c\cdot U(Q(\theta_k,\beta_m),\beta_m)}_{\text{OpEx}} + \underbrace{J(Q(\theta_k,\beta_m))}_{\text{CapEx}}\Big], \tag{14}$$

The MNO's expected profit under a *feasible* contract $\Phi(\Theta,\mathcal{B})$ is the difference between its revenue and cost, which is given by

$$W(\Phi) = R(\Phi) - C(\Phi), \tag{15}$$

Now we formulate the MNO's optimal contract design with two-dimensional private information as follows:

Problem 1 (Optimal Contract Design).

$$\begin{aligned} &\max_{\Phi}\ W(\Phi) \\ &\text{s.t. } (10), (11). \end{aligned} \tag{16}$$

The core of contract design is to offer each type of users individual rationality and all types of users incentive compatibility, so that each user truthfully reveals his type and selects the contract item intended for his type. Reflected in Problem 1, the MNO needs to address a total of $(KM-1)KM$ IC constraints and a total of KM IR constraints. For those contract problems involving only one-dimensional private information, the single-crossing condition usually holds for monotonic user types. With single-crossing condition, the approach used in [5, 8, 12, 13, 27] can reduce the unbinding IC and IR constraints so that the problem becomes analytically tractable. However, the single-crossing condition usually does not hold for those contract problems involving multi-dimensional private information. As a result, the standard approach

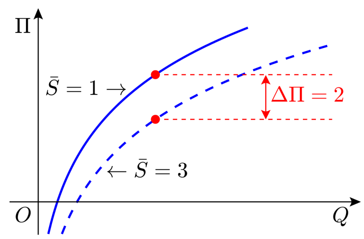


Figure 3: Two indifference curves of the same user type with two different expected payoffs, i.e., $\bar{S}=1$ and $\bar{S}=3$.

in [5, 8, 12, 13, 27] of reducing unbinding constraints is not applicable here. In Section 3, we will introduce our approach to handling the large number of constraints under multi-dimensional private information.

3 TWO-DIMENTIONAL CONTRACT DESIGN

Next we propose a general method to solve the optimal contract under users' two-dimensional private information: data valuation θ and network substitutability β. We first introduce a user's marginal rate of substitution (willingness-to-pay) to tackle the feasibility condition. Then we look at the contract optimality.

3.1 Contract Feasibility

Now we introduce a user's marginal rate of substitution related to his two-dimensional private information. Based on this concept, we further derive the necessary and sufficient conditions for a contract to be feasible.

3.1.1 Marginal Rate of Substitution (Willingness-to-Pay). In economics, a consumer's indifference curve connects those goods bundles that achieve the same consumer satisfaction (payoff). In our problem, we can plot a user's indifference curve over the contract plan (i.e., the data cap Q and the subscription fee Π) as in Fig. 3. On the (Q,Π) plane, a type-(θ,β) user's indifference curve with a fixed payoff $\bar{S}$ satisfies

$$\bar{S}=\theta[\bar{d}-\beta A(Q)]-\pi(1-\beta)A(Q)-\Pi. \tag{17}$$

Fig. 3 shows that the indifference curve is increasing concave, which indicates that the subscription fee would increase (with a diminishing marginal increment) as the data cap increases to maintain the same payoff. Moreover, as a user's indifference curve shifts downward, his payoff increases because of the decreasing subscription fee.

The slope of an indifference curve is also called the marginal rate of substitution (MRS), which is the rate at which a consumer is ready to give up one goods in exchange for another goods, while maintaining the same level of satisfaction. In our problem, we denote the MRS of a type-(θ,β) user on a data cap Q as

$$\sigma(Q,\theta,\beta)\triangleq\frac{\partial\Pi}{\partial Q}=-\left[\theta\beta+\pi(1-\beta)\right]\frac{\partial A(Q)}{\partial Q}, \tag{18}$$

which depends on the user's private information (θ,β) and the data cap Q. The MRS $\sigma(Q,\theta,\beta)$ indicates a type-(θ,β) user's *willingness-to-pay* for an additional unit of data on a data cap Q. In the rest of the paper, we will use the three phrases "marginal rate of substitution", "slope of the indifference curve", and "willingness-to-pay" interchangeably.

3.1.2 User Ordering based on Willingness-to-Pay. Now we sort the KM user types based on $\sigma(Q,\theta,\beta)$ in an ascending order for any fixed data cap Q as follows:

$$\Lambda^1,\ \Lambda^2,\ \ldots,\ \Lambda^{KM}, \tag{19}$$

where $\Lambda^i\triangleq\{\theta_k,\beta_m\}$ for some k and m, and $\sigma(Q,\Lambda^1)\leq\sigma(Q,\Lambda^2)\leq\ldots\leq\sigma(Q,\Lambda^{KM})$ for any fixed Q. Note that the user ordering in (19) does not change, even though the value of $\sigma(Q,\Lambda^i)$ would change with the data cap Q, In the following, we will use ϕ^i to represent the contract item intended for the type-Λ^i users.

To have a better understanding on the transformation, we use Fig. 4 to illustrate how (θ_k,β_m) maps to Λ^i. There are three different market modes depending on the relationship between the extreme valuations (θ_1 and θ_K) and the overage fee π: $\theta_1<\theta_K<\pi$ as in Fig. 4(a), $\theta_1<\pi<\theta_K$ as in Fig. 4(b), and $\pi<\theta_1<\theta_K$ as in Fig. 4(c). Specifically, the arrows in Fig. 4 point to the direction where the user's MRS $\sigma(\theta,\beta,Q)$ increases, the blue square denotes the *minimum willingness-to-pay* user type-Λ^1, and the red star denotes the *largest willingness-to-pay* user type-Λ^{KM}. The following proposition summarizes the mapping from (θ_k,β_m) to Λ^1 and Λ^{KM}.

Proposition 3.1. *Under the three market modes, the type-Λ^1 and type-Λ^{KM} users have their private information as follows:*

$$\begin{cases}\Lambda^1=\{\theta_1,\beta_M\},\Lambda^{KM}=\{\theta_K,\beta_1\}, & \text{if } \theta_1<\theta_K<\pi,\\ \Lambda^1=\{\theta_1,\beta_M\},\Lambda^{KM}=\{\theta_K,\beta_M\}, & \text{if } \theta_1<\pi<\theta_K,\\ \Lambda^1=\{\theta_1,\beta_1\},\ \Lambda^{KM}=\{\theta_K,\beta_M\}, & \text{if } \pi<\theta_1<\theta_K.\end{cases} \tag{20}$$

Furthermore, the green triangles in Fig. 4 denote the *smallest payoff* user type Λ^u, defined as

$$\Lambda^u\triangleq\arg\min_{\Lambda^i}\bar{S}(Q,\Pi,\Lambda^i),\ \forall\,(Q,\Pi). \tag{21}$$

Note that we can show that Λ^u does not depend on the value of (Q,Π), i.e., type-Λ^u user achieves the smallest payoff among the all user types for any given contract item. We summarize the mapping from (θ_k,β_m) to Λ^u as follows:

Proposition 3.2. *Under the three market modes, the type-Λ^u users have the private information as follows:*

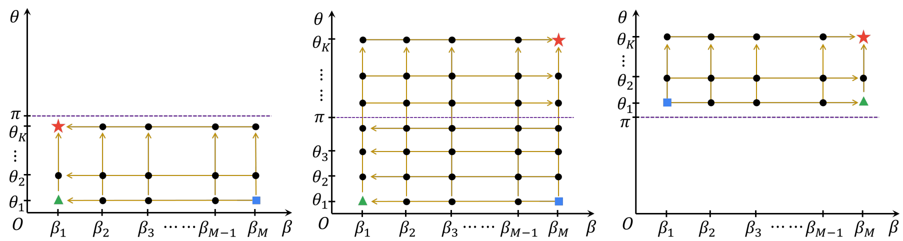
$$\begin{cases}\Lambda^u=\{\theta_1,\beta_1\}, & \text{if } \theta_1<\theta_K<\pi,\\ \Lambda^u=\{\theta_1,\beta_1\}, & \text{if } \theta_1<\pi<\theta_K,\\ \Lambda^u=\{\theta_1,\beta_M\}, & \text{if } \pi<\theta_1<\theta_K.\end{cases} \tag{22}$$

Next we derive the necessary an sufficient conditions for a contract to be feasible based on users' *willingness-to-pay*.

3.1.3 Necessary Conditions . Lemmas 3.3 and 3.4 provide two necessary conditions for a contract to be feasible.

Lemma 3.3. *For any feasible contract $\Phi(\Theta,\mathcal{B})$, $Q(\Lambda^i)<Q(\Lambda^j)$ if and only if $\Pi(\Lambda^i)<\Pi(\Lambda^j)$.*

Lemma 3.4. *For any feasible contract $\Phi(\Theta,\mathcal{B})$, if $\sigma(Q,\Lambda^i)>\sigma(Q,\Lambda^j)$ for any Q, then $Q(\Lambda^i)\geq Q(\Lambda^j)$.*



(a) $\theta_1 < \theta_K < \pi$ (b) $\theta_1 < \pi < \theta_K$ (c) $\pi < \theta_1 < \theta_K$

Figure 4: Three market modes.

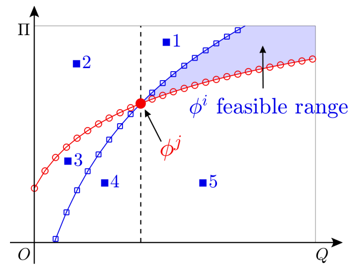


Figure 5: Illustration of Lemma 3.4.

Lemma 3.3 reveals that a larger data cap corresponds to a higher subscription fee. Lemma 3.4 shows that a user with a stronger *willingness-to-pay* for the data cap deserves a larger data cap in the feasible contract. Here we provide a proof sketch for Lemma 3.4 to show the key insights.

Proof Sketch: In Fig. 5, for a type-Λ^j user, the red dot denotes the contract item ϕ^j intended for this user, and the red circle curve l_j represents his indifference curve with a payoff equal to that of selecting ϕ^j. For a type-Λ^i user, the blue square curve l_i is his indifference curve with a payoff equal to this user choosing the red dot contract item ϕ^j (not intended for his type). It is obvious that l_i is steeper than l_j; mathematically speaking, $\sigma(Q, \Lambda^i) > \sigma(Q, \Lambda^j)$ for all Q (which is the condition in Lemma 3.4). In other words, comparing with the type-Λ^j users, the type-Λ^i users have a stronger willingness-to-pay under any given data cap. Moreover, as a user's indifference curve shifts downward, his payoff increases because of the decreasing subscription fee. Now we make the following claim.

Claim: To ensure the pairwise IC constraint $\phi^i \overset{\text{IC}}{\Longleftrightarrow} \phi^j$, the contract item ϕ^i intended for the type-Λ^i users must locate *below* (or on) the blue square curve l_i and *above* (or on) the red circle curve l_j, i.e., in the blue region.

We prove the claim by contradiction. If the claim is not true, then we need to consider the following two scenarios:

Scenario 1: ϕ^i is above the blue square curve l_i, such as the green squares labeled 1, 2, 3 in Fig. 5. In this case, the indifference curve l_i for the type-Λ^i should shift upward to a new curve l_i' (with a decreasing payoff) to touch one of the three green squares. Therefore, the type-Λ^i user can achieve a higher payoff (comparing with selecting ϕ^i) by selecting the red dot contract item ϕ^j, which violates the pairwise IC constraint for type-Λ^i user.

Scenario 2: ϕ^i is below the red circle curve $\hat{l}_j$, such as the green squares labeled 3, 4, 5 in Fig. 5. In this case, the indifference curve l_j for the type-Λ^j user should shift downward to l_j' (with an increasing payoff) to touch one of the three green squares. Therefore, the type-Λ^j user can achieve a higher payoff by selecting the green square contract item ϕ^i, which violates the pairwise IC constraint for type-Λ^j users.

The above analysis proves the claim, i.e., the contract item ϕ^i must locate in the blue area, which is on the right of the dash line. Thus, $Q(\Lambda^i) \geq Q(\Lambda^j)$, as Lemma 3.4 implies. ■

According to Lemma 3.3 and Lemma 3.4, we summarize the *necessary conditions* for a feasible contract as follows:

Theorem 3.5 (Necessary Conditions). *The feasible contract $\Phi(\Theta, \mathcal{B})$ has the following structure*

$$\begin{cases} Q(\Lambda^1) \leq Q(\Lambda^2) \leq \ldots \leq Q(\Lambda^{KM}), \\ \Pi(\Lambda^1) \leq \Pi(\Lambda^2) \leq \ldots \leq \Pi(\Lambda^{KM}). \end{cases} \tag{23}$$

3.1.4 Sufficient Conditions . Now we derive the sufficient conditions for the feasible contract though the following two transitivity properties for Pairwise Incentive Compatibility (PIC) and Individual Rationality (IR).

Lemma 3.6 (PIC-Transitivity). *Suppose the necessary conditions in Theorem 3.5 hold, then for any $i_1 < i_2 < i_3$, the following is true*

$$\text{if } \phi^{i_1} \overset{\text{IC}}{\Longleftrightarrow} \phi^{i_2}, \text{ and } \phi^{i_2} \overset{\text{IC}}{\Longleftrightarrow} \phi^{i_3}, \text{ then } \phi^{i_1} \overset{\text{IC}}{\Longleftrightarrow} \phi^{i_3}. \tag{24}$$

The above PIC transitivity property makes the contract design more tractable. It reveals that we can reduce a total of $KM(KM-1)/2$ IC constraints to a total of $KM-1$ PIC constraints for the neighbor user type pairs, i.e., $\phi^i \overset{\text{IC}}{\Longleftrightarrow} \phi^{i+1}$, $i = 1, 2, \ldots, KM-1$.

Now we present the IR transitivity in the following lemma.

Lemma 3.7 (IR-Transitivity). *Suppose the necessary conditions in Theorem 3.5 and all IC conditions hold, then the following is true,*

$$\text{if } \phi^u \geqslant 0, \text{ then } \phi^i \geqslant 0, \ \forall\, i \neq u.$$

Recall that user type-Λ^u achieves the smallest payoff among all the user types for any given contract item. Lemma 3.7 implies that once we can guarantee all the IC conditions, then we only need to further ensure that the IR constraint for the smallest type Λ^u. This allows us to reduce a total of KM IR constraints to one IR constraint of the smallest user type, i.e., $\phi^u \geqslant 0$.

Before we present the sufficient conditions for the feasible contract, we first introduce a user's *virtual payoff increment.* Recall that $L(Q, \Lambda^i)$ denotes the type-Λ^i user's virtual payoff as defined in (9). We define $\gamma(\Lambda^i)$ and $\eta(\Lambda^i)$ as the type-Λ^i user's virtual payoff increments between selecting the contract item ϕ^i and the contract items intended for his neighbor user types (i.e., ϕ^{i-1} and ϕ^{i+1}), as follows

$$\begin{aligned} \gamma(\Lambda^i) &= L(Q(\Lambda^i), \Lambda^i) - L(Q(\Lambda^{i-1}), \Lambda^i), \\ \eta(\Lambda^i) &= L(Q(\Lambda^i), \Lambda^i) - L(Q(\Lambda^{i+1}), \Lambda^i). \end{aligned} \tag{25}$$

Based on Lemmas 3.3~3.7, we derive the following *sufficient conditions* for a contract to be feasible.

Theorem 3.8 (Sufficient Conditions). *The contract $\Phi(\Theta, \mathcal{B})$ is feasible if the following conditions hold,*

(1) $Q(\Lambda^1) \le Q(\Lambda^2) \le \ldots \le Q(\Lambda^{KM})$,
(2) for $i = u$,

$$\Pi(\Lambda^u) \le L(Q(\Lambda^u), \Lambda^u). \tag{26}$$

(3) for all $i = u+1, u+2, \ldots, KM$,

$$\begin{cases} \Pi(\Lambda^i) \le \Pi(\Lambda^{i-1}) + \gamma(\Lambda^i), \\ \Pi(\Lambda^i) \ge \Pi(\Lambda^{i-1}) - \eta(\Lambda^{i-1}), \end{cases} \tag{27}$$

(4) for all $i = 1, 2, \ldots, u-1$,

$$\begin{cases} \Pi(\Lambda^i) \le \Pi(\Lambda^{i+1}) + \eta(\Lambda^i), \\ \Pi(\Lambda^i) \ge \Pi(\Lambda^{i+1}) - \gamma(\Lambda^{i+1}), \end{cases} \tag{28}$$

Next we discuss the intuitions of Theorem 3.8. Condition 1 satisfies the necessary conditions in Theorem 3.5. Condition 2 guarantees the IR condition for the type-Λ^u users, i.e., $\phi^u \succcurlyeq 0$, which is sufficient for the IR conditions of all other user types according to Lemma 3.7. Condition 3 and Condition 4 guarantee the pairwise IC condition for the neighbor user types, i.e., $\phi^i \overset{\text{IC}}{\Longleftrightarrow} \phi^{i+1}$, $\forall\, 1 \le i \le KM-1$, which is sufficient for the global IC according to Lemma 3.6. Specifically, the first inequality in (27) ensures that the type-Λ^i user will not select the contract item ϕ^{i-1}, i.e., $\bar{S}(Q(\Lambda^i), \Lambda^i) \ge \bar{S}(Q(\Lambda^{i-1}), \Lambda^i)$; the second inequality ensures the type-Λ^{i-1} user will not select the contract item ϕ^i, i.e., $\bar{S}(Q(\Lambda^{i-1}), \Lambda^{i-1}) \ge \bar{S}(Q(\Lambda^i), \Lambda^{i-1})$. Similar intuitions apply to (28).

With the necessary and sufficient conditions of a feasible contract, next we will analyze the optimality of the contract.

3.2 Contract Optimality

We analyze the optimality of the contract in two steps. First, we derive the MNO's optimal (profit-maximizing) prices $\{\Pi^*(\Lambda^i), 1 \le i \le KM\}$ given a feasible choice of data caps, i.e., $Q(\Lambda^1) \le Q(\Lambda^2) \le \ldots \le Q(\Lambda^{KM})$ in Problem 2. Second, we substitute the optimal prices to the MNO's profit function and derive the optimal data cap assignment $\{Q^*(\Lambda^i), 1 \le i \le KM\}$ in Problem 3.

Problem 2 (Optimal Prices).

$$\max \sum_{i=1}^{KM} q(\Lambda^i)\Big[\Pi(\Lambda^i) + P(Q(\Lambda^i), \Lambda^i) - c \cdot U(Q(\Lambda^i), \Lambda^i) - J(Q(\Lambda^i))\Big] \tag{29}$$

s.t. (26), (27), (28)

var : $\{\Pi(\Lambda^i), 1 \le i \le KM\}$.

In Problem 1, the decision variables are the subscription fees. We characterize the optimal prices $\{\Pi^*(\Lambda^i), 1 \le i \le KM\}$ in the following theorem.

Theorem 3.9 (Optimal Pricing Policy). *Given a set of feasible data caps $Q(\Lambda^1) \le Q(\Lambda^2) \le \ldots \le Q(\Lambda^{KM})$. The optimal pricing policy for the MNO, denoted by $\{\Pi^*(\Lambda^i), 1 \le i \le KM\}$, is unique and given by*

$$\begin{cases} \Pi^*(\Lambda^i) = L(Q(\Lambda^i), \Lambda^i), & \text{if } i = u, \\ \Pi^*(\Lambda^i) = \Pi^*(\Lambda^{i-1}) + \gamma(\Lambda^i), & \text{if } i > u, \\ \Pi^*(\Lambda^i) = \Pi^*(\Lambda^{i+1}) + \eta(\Lambda^i), & \text{if } i < u. \end{cases} \tag{30}$$

Comparing Theorem 3.8 and Theorem 3.9, we note that, given a set of feasible data caps, the optimal price $\Pi^*(\Lambda^i)$ is the highest price satisfying the IC and IR conditions.

Before we consider the optimal data caps, we first introduce the *virtual payoff difference*. For the given data cap Q, the virtual payoff differences between the type-Λ^i user and his neighbor user types (i.e., Λ^{i-1} and Λ^{i+1}) are defined as

$$\begin{aligned} \xi_i(Q) &= L(Q, \Lambda^i) - L(Q, \Lambda^{i-1}), \\ \rho_i(Q) &= L(Q, \Lambda^i) - L(Q, \Lambda^{i+1}). \end{aligned} \tag{31}$$

By substituting the optimal prices (30) derived in Theorem 3.9 into the objective function of Problem 2, then we can get the following optimization problem over the data caps.

Problem 3 (Optimal Data Caps).

$$\begin{aligned} \max \; & \sum_{i=1}^{KM} G_i(Q(\Lambda^i)) - q(\Lambda^i)\left[c \cdot U(Q(\Lambda^i), \Lambda^i) + J(Q(\Lambda^i))\right] \\ \text{s.t.} \; & 0 \le Q(\Lambda^1) \le Q(\Lambda^2) \le \ldots \le Q(\Lambda^{KM}) \\ \text{var}: \; & \{Q(\Lambda^i), 1 \le i \le KM\}, \end{aligned} \tag{32}$$

In Problem 3, $G_i(Q)$ is given by

$$G_i(Q) \triangleq \begin{cases} q(\Lambda^i)V(Q, \Lambda^i), & i = 1, KM, \\ q(\Lambda^i)V(Q, \Lambda^i) + h^i \xi_i(Q), & 1 < i < u, \\ q(\Lambda^i)V(Q, \Lambda^i) + h^i \xi_i(Q) + h_i \rho_i(Q), & i = u, \\ q(\Lambda^i)V(Q, \Lambda^i) + h_i \rho_i(Q), & u < i < KM, \end{cases} \tag{33}$$

where $h^i = \sum_{t=1}^{i-1} q(\Lambda^t)$, $h_i = \sum_{t=i+1}^{KM} q(\Lambda^t)$.

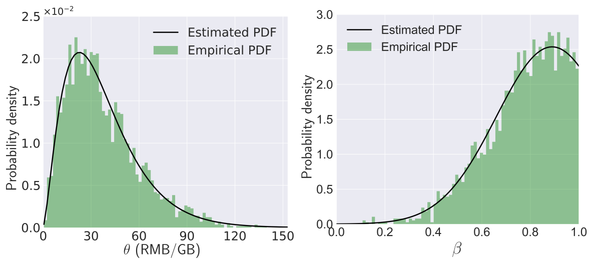
Note that the objective function in Problem 3 has a separable structure over i. To take advantage of such a structure, we first relax the monotonicity constraints, and optimize over each decision variable separately as follows:

$$Q^\dagger(\Lambda^i) = \arg\max_{Q \ge 0} G_i(Q) - q(\Lambda^i)\left[c \cdot U(Q, \Lambda^i) + J(Q)\right], \tag{34}$$

If the solution $\{Q^\dagger(\Lambda^i), 1 \le i \le KM\}$ is feasible, i.e., satisfying the monotonicity constraints $Q^\dagger(\Lambda^1) \le Q^\dagger(\Lambda^2) \le \ldots \le Q^\dagger(\Lambda^{KM})$, then it is the optimal solution. If not, however, we will use the Dynamic Algorithm in [8] to adjust the solution to make it feasible and generate a new *adjusted solution* $\{\tilde{Q}(\Lambda^i), 1 \le i \le KM\}$. The intuition behind the Dynamic Algorithm is to first find a consecutive infeasible subsequence, e.g., $Q^\dagger(\Lambda^n) \ge Q^\dagger(\Lambda^{n+1}) \ge \ldots \ge Q^\dagger(\Lambda^m)$ where $n < m$ and $Q^\dagger(\Lambda^n) > Q^\dagger(\Lambda^m)$, then generate the *adjusted solution* $\{\tilde{Q}(\Lambda^i), 1 \le i \le KM\}$ as follows:

$$\tilde{Q}(\Lambda^i) = \begin{cases} \arg\max\limits_{Q \ge 0} \sum_{j=n}^{m} G_j(Q) - q(\Lambda^j)[c \cdot U(Q, \Lambda^j) + J(Q)], & \text{if } n \le i \le m, \\ Q^\dagger(\Lambda^i) & \text{otherwise}. \end{cases} \tag{35}$$

The algorithm runs iteratively until there is no infeasible subsequence. Such an adjusted solution $\{\tilde{Q}(\Lambda^i), 1 \le i \le KM\}$ must be feasible. Therefore, we can summarize the optimal solutions of Problem 3, denoted by $\{Q^*(\Lambda^i), 1 \le i \le KM\}$ as follows:



(a) $\theta \sim \text{Gamma}(4.5, 0.11)$ (b) $\beta \sim \text{TN}(0.91, 0.22, 0, 1)$

Figure 6: Fitting the PDF of θ and β.

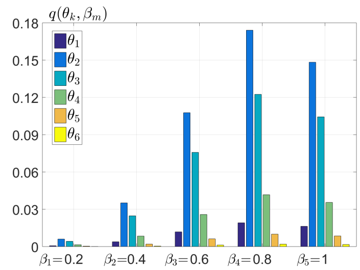


Figure 7: User type distribution.

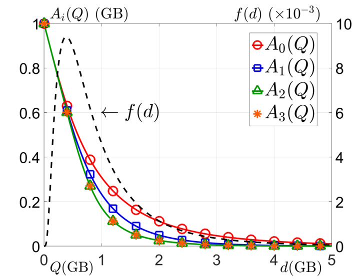


Figure 8: $A_i(Q)$ *vs* Q and $f(d)$ *vs* d.

Theorem 3.10 (Optimal Data Caps). *Suppose that Problem 3 is convex. Then the optimal data caps are given by:*

(1) $Q^*(\Lambda^i) = Q^\dagger(\Lambda^i)$, if $\{Q^\dagger(\Lambda^i), 1 \le i \le KM\}$ in (34) satisfies the monotonicity constraints (32);
(2) $Q^*(\Lambda^i) = \tilde{Q}(\Lambda^i)$, if $\{Q^\dagger(\Lambda^i), 1 \le i \le KM\}$ in (34) does not satisfy the monotonicity constraints (32).

Theorem 3.10 shows that the solution $\{Q^*(\Lambda^i), 1 \le i \le KM\}$ is globally optimal if Problem 3 is convex under some distribution of users' private information (θ, β); otherwise, it is locally optimal.

4 NUMERICAL RESULTS

Now we numerically demonstrate the MNO's multi-cap data plan design. To carry out the user preference estimation in Step I, we conduct a market survey regarding the telecommunication market in mainland China. We begin in Section 4.1 by analyzing the distribution of users' characteristics. Then in Section 4.2 we simulate the MNO's optimal contract under different data mechanisms based on the estimated user preferences.

4.1 Empirical Results

The empirical PDFs of users' data valuation θ and network substitutability β are shown by the green bars in Fig. 6(a) and Fig. 6(b), respectively. We observe that a large proportion of users' data valuations θ locates between $[10, 50]$; most people would like to shrink approximately 80% ∼ 100% overage usage through alternative networks. Moreover, we find that the Pearson correlation coefficient between θ and β is less than 0.05, which allows us to fit the two distributions separately.

Next we estimate the data valuation PDF by assuming a gamma distribution, and the one of network substitutability by assuming a truncated normal distribution. We show the distribution fits in Fig. 6, where the parameters are chosen to minimize the least-squares divergence between the estimated and empirical PDF.

In reality, it may not be realistic to design a contract with hundreds of completely different contract items. Therefore, the MNO would extract some typical user types for the purpose of contract design based on the fitted distributions. For example, we consider six types of data valuation and five types of network substitutability, i.e., $\Theta = \{10, 32, 54, 76, 98, 120\}$ and $\mathcal{B} = \{0.2, 0.4, 0.6, 0.8, 1\}$, for the rest of the simulations. The corresponding distribution $q(\theta_k, \beta_m)$ is shown in Fig. 7.

4.2 Optimal Contract

Now we simulate the MNO's optimal contract under four data mechanisms $\mathcal{T}_i, i = 0, 1, 2, 3$, based on the distribution of the selected user types as mentioned above.

Following the data analysis results in [10], we suppose that users' monthly data demand follows a truncated log-normal distribution with mean $\bar{d} = 1000$ on the interval $[0, 5000]$, i.e., the average data demand is 1GB and the potential maximal data demand is $D = 5$GB. Fig. 8 shows the probability mass function $f(d)$ and the expected overage usage under the four data mechanisms, which indicates that $A_0(Q) \ge A_1(Q) \ge A_2(Q) = A_3(Q)$ for any Q. It validates our previous analysis results in Proposition 2.2, i.e., $\mathcal{T}_2$ and $\mathcal{T}_3$ offer the best and the same time flexibility, while $\mathcal{T}_0$ offers the worst.

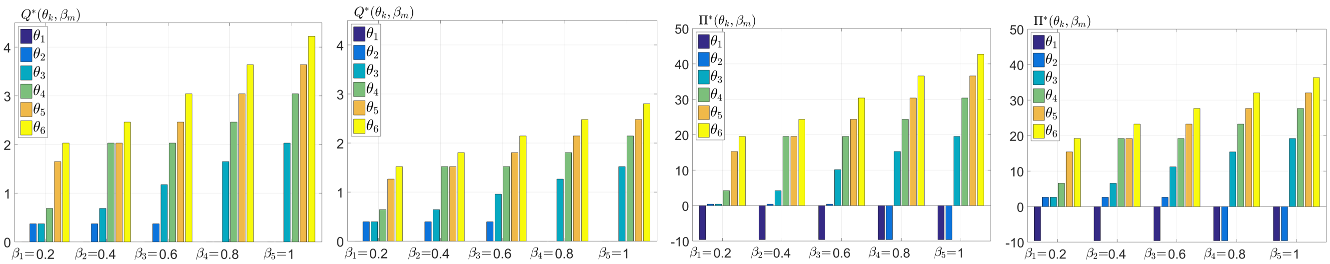
Furthermore, we use the per-unit fee in the telecommunication market of mainland China, i.e., $\pi = 40$ RMB/GB, and assume that the MNO's marginal operational cost is $c = 2 \times 10^{-3}$ RMB/MB and CapEx takes the form of $J(Q) = Q/10^3$ RMB. Moreover, the minimum data unit is set to 1MB.

Fig. 9 plots the data caps and the corresponding subscription fees in the MNO's optimal contract under $\mathcal{T}_2$. We have the following observations:

1). The optimal contract may offer some low valuation users a zero cap (i.e., pay-as-you-go pricing), together with a price discount (i.e., a negative subscription fee). The price discount can ensure the subscription of the low valuation users (satisfying the IR condition). Meanwhile, the pay-as-you-go pricing can reduce the MNO's CapEx due to the zero data cap.

2). Under the optimal contract, the better time flexibility enables the MNO to offer a smaller data cap for the same type of users with a relatively lower subscription fee. For example, in Fig. 9, the optimal contract item for type-(θ_6, β_3) users (i.e., type-Λ^{27} users) is $\mathcal{T}_0 = \{3.1 \text{ GB}, 30.5 \text{ RMB}\}$ and $\mathcal{T}_2 = \{2.2 \text{ GB}, 27.4 \text{ RMB}\}$, respectively. In this case, the MNO not only reduces its cost (including CapEx and OpEx) but also maintains its overall market share, by providing incentives to all users types through the optimal contract.

Table 2 further investigates the impact of the time flexibility on users' payoff and the MNO's profit. From the users' perspective, a data mechanism with a better flexibility can reduce their payments and increase their payoffs. As for the MNO, a data mechanism with a better flexibility can lead to a higher revenue and a smaller cost, hence a larger overall MNO profit. In a nutshell, among the four data mechanisms, $\mathcal{T}_2$ and $\mathcal{T}_3$ are the most beneficial in terms of users' payoff and the MNO's profit, hence the social welfare.



(a) Data cap under $\mathcal{T}_0$ (b) Data cap under $\mathcal{T}_2$ (c) Subscription fee under $\mathcal{T}_0$ (d) Subscription fee under $\mathcal{T}_2$

Figure 9: Optimal contract under data mechanism $\mathcal{T}_2$.

Table 2: Comparison between the four mechanisms.

Plans	Users		MNO		
	Overage Payment	Payoff	Revenue	Cost	profit
$\mathcal{T}_0$	4.50	9.4	15.9	3.4	12.5
$\mathcal{T}_1$	3.96	11.7	17.2	3.1	14.1
$\mathcal{T}_2$	3.24	12.2	19.1	2.8	16.3
$\mathcal{T}_3$	3.24	12.2	19.1	2.8	16.3

5 CONCLUSION

In this paper we studied the MNO's optimal multi-cap data plan design under four data mechanisms. We first unified different data mechanisms in the same general framework, where different mechanisms differ in terms of the users' expected payoffs. Then we derived the optimal contract under two-dimensional private information, i.e., data valuation and network substitutability. Our analysis revealed that the slope of a user's indifference curve on the contract plane corresponds to his willingness-to-pay, and the feasible contract (satisfying IC and IR conditions) would offer a larger data cap to the user with the stronger willingness-to-pay.

There are some possible directions to extend this paper. An interesting direction is to relax our assumption on homogeneous data demand distribution, which leads to an optimal contract design with users' three-dimensional private information. Another direction is to consider the competitive market and analyze the impact of the time flexibility on MNOs' competition.

REFERENCES

[1] *AT&T Rollover Data Plan.* https://www.att.com.

[2] *China Mobile Rollover Data Plan.* http://www.10086.cn.

[3] Matthew Andrews. 2016. Understanding the effects of quota trading on mobile usage dynamics. In *14th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt).*

[4] Aloisio Araujo and Humberto Moreira. 2010. Adverse selection problems without the Spence–Mirrlees condition. *Journal of Economic Theory* 145, 3 (2010), 1113–1141.

[5] Lingjie Duan, Lin Gao, and Jianwei Huang. 2014. Cooperative spectrum sharing: A contract-based approach. *IEEE Transactions on Mobile Computing* 13, 1 (2014), 174–187.

[6] Lingjie Duan, Jianwei Huang, and Biying Shou. 2015. Pricing for local and global Wi-Fi markets. *IEEE Transactions on Mobile Computing* 14, 5 (2015), 1056–1070.

[7] Drew Fudenberg and Jean Tirole. 1991. Game theory. (1991).

[8] Lin Gao, Xinbing Wang, Youyun Xu, and Qian Zhang. 2011. Spectrum trading in cognitive radio networks: A contract-theoretic modeling approach. *IEEE Journal on Selected Areas in Communications* 29, 4 (2011), 843–855.

[9] Sangtae Ha, Soumya Sen, Carlee Joe-Wong, Youngbin Im, and Mung Chiang. 2012. TUBE: Time-dependent pricing for mobile data. *ACM SIGCOMM Computer Communication Review* 42, 4 (2012), 247–258.

[10] Anja Lambrecht, Katja Seim, and Bernd Skiera. 2007. Does uncertainty matter? Consumer behavior under three-part tariffs. *Marketing Science* 26, 5 (2007), 698–710.

[11] Yuan Luo, Lin Gao, and Jianwei Huang. 2015. Spectrum reservation contract design in TV white space networks. *IEEE Transactions on Cognitive Communications and Networking* 1, 2 (2015), 147–160.

[12] Qian Ma, Lin Gao, Ya-Feng Liu, and Jianwei Huang. 2016. A contract-based incentive mechanism for crowdsourced wireless community networks. In *14th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt).*

[13] Qian Ma, Lin Gao, Ya-Feng Liu, and Jianwei Huang. 2018. Incentivizing Wi-Fi Network Crowdsourcing: A Contract Theoretic Approach. *arXiv preprint arXiv:1802.09391* (2018).

[14] Qian Ma, Ya-Feng Liu, and Jianwei Huang. 2016. Time and location aware mobile data pricing. *IEEE Transactions on Mobile Computing* 15, 10 (2016), 2599–2613.

[15] Richard TB Ma. 2016. Usage-based pricing and competition in congestible network service markets. *IEEE/ACM Transactions on Networking* 24, 5 (2016), 3084–3097.

[16] R Preston McAfee and John McMillan. 1988. Multidimensional incentive compatibility and mechanism design. *Journal of Economic Theory* 46, 2 (1988), 335–354.

[17] Soumya Sen, Carlee Joe-Wong, and Sangtae Ha. 2012. The economics of shared data plans. In *Annual Workshop on Information Technologies and Systems.*

[18] Soumya Sen, Carlee Joe-Wong, Sangtae Ha, and Mung Chiang. 2013. A survey of smart data pricing: Past proposals, current plans, and future trends. *ACM Computing Surveys (CSUR)* 46, 2 (2013), 15.

[19] Xuehe Wang, Lingjie Duan, and Rui Zhang. 2016. User-initiated data plan trading via a personal hotspot market. *IEEE Transactions on Wireless Communications* 15, 11 (2016), 7885–7898.

[20] Xin Wang, Richard TB Ma, and Yinlong Xu. 2017. The role of data cap in optimal two-part network pricing. *IEEE/ACM Transactions on Networking* 25, 6 (2017), 3602–3615.

[21] Zhiyuan Wang, Lin Gao, and Jianwei Huang. Exploring Time Flexibility in Wireless Data Plan. *submitted to IEEE Transactions on Mobile Computing.*

[22] Zhiyuan Wang, Lin Gao, and Jianwei Huang. 2017. A Contract-Theoretic Design of Mobile Data Plan with Time Flexibility. In *12th Workshop on the Economics of Networks, Systems and Computation (NetEcon).*

[23] Zhiyuan Wang, Lin Gao, and Jianwei Huang. 2017. Pricing optimization of rollover data plan. In *15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt).*

[24] Zhiyuan Wang, Lin Gao, and Jianwei Huang. 2018. Pricing Competition of Rollover Data Plan. In *16th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt).*

[25] Yi Wei, Junlin Yu, Lok Tat, M., and Lin Gao. 2016. A Novel Mobile Data Plan Design from the Perspective of Data Period. In *IEEE International Conference on Communication Systems (ICCS).*

[26] Junlin Yu, Man Hon Cheung, Jianwei Huang, and H Vincent Poor. 2017. Mobile data trading: Behavioral economics analysis and algorithm design. *IEEE Journal on Selected Areas in Communications* 35, 4 (2017), 994–1005.

[27] Yanru Zhang, Lingyang Song, Walid Saad, Zaher Dawy, and Zhu Han. 2015. Contract-based incentive mechanisms for device-to-device communications in cellular networks. *IEEE Journal on Selected Areas in Communications* 33, 10 (2015), 2144–2155.

[28] Liang Zheng and Carlee Joe-Wong. 2016. Understanding rollover data. In *IEEE INFOCOM Workshop on Smart Data Pricing.*