ROBUST PRICING FOR CLOUD COMPUTING

DIRK BERGEMANN[†] AND RAHUL DEB[◊]

FEBRUARY 10, 2025

ABSTRACT: We study the robust sequential screening problem of a monopolist seller of multiple cloud computing services facing a buyer who has private information about his demand distribution for these services. At the time of contracting, the buyer knows the distribution of his demand of various services and the seller simply knows the mean of the buyer's total demand. We show that a simple "committed spend mechanism" is robustly optimal: it provides the seller with the highest profit guarantee against all demand distributions that have the known total mean demand. This mechanism requires the buyer to commit to a minimum total usage and a corresponding base payment; the buyer can choose the individual quantities of each service and is free to consume additional units (over the committed total usage) at a fixed marginal price. This result provides theoretical support for prevalent cloud computing pricing practices while highlighting the robustness of simple pricing schemes in environments with complex uncertainty.

KEYWORDS: Sequential Screening, Second-Degree Price Discrimination, Mechanism Design, Cloud Computing, Commitment Spend Contract.

JEL CLASSIFICATION: D44, D82, D83.

1. INTRODUCTION

1.1. *Motivation and Results*

Cloud computing has dramatically changed how individuals and businesses access computational resources, leading to fundamental theoretical questions about optimal pricing strategies in this market. Cloud computing providers offer several different services, perhaps most importantly: (*i*) input/output, (*ii*) compute, (*iii*) memory/storage, and (*iv*) software. A key challenge in designing cloud computing pricing mechanisms is that providers face significant uncertainty about customers' demands for these different services. Customers, themselves, may not know their exact demand at the time of contracting (as this will realize over a period of time) but they have knowledge about the *distribution* of their demands of each service. This information is private and difficult for providers to verify or elicit. Providers must design pricing mechanisms that balance efficiency, profit, robustness and simplicity.

In practice, contracts for cloud computing frequently take the form of what are known as "committed spend mechanisms." Such a mechanism (or contract) commits the customer to spend a certain minimum amount $\underline{t}$ over a given time horizon. The contract is legally binding so the consumer cannot pay less than this minimum even if their realized demand for the cloud computing services is lower than expected. In return, the customer is allocated a *total* quantity $\underline{q}$ of services and they are free to use any combination of individual quantities of different services that add up to less than this total. The contract additionally specifies a price p and, if a customer's total usage of services q exceeds $\underline{q}$, they are charged an additional amount $p(q - \underline{q})$ equal to their excess usage times the price.

We model this setting as a robust multidimensional sequential screening problem. A seller (the provider of cloud computing services) has K goods for sale and offers a contract. At the time of contracting, the buyer initially knows his distribution F of demands for these different services. Upon contracting with the seller, the buyer learns his demand $(\theta_1, \ldots, \theta_K)$ for the various services and chooses a quantity $(q_1, \ldots, q_K)$ of each that maximizes his utility $\theta_1 u(q_1) + \cdots + \theta_K u(q_K)$ less the cost of this bundle of services.

Formally, the seller offers a mechanism—an allocation and transfer—that takes as its arguments a distribution and a realized demand type. The buyer first reports his type distribution and then his type, once it is realized. The seller simply knows the average total buyer demand λ and

chooses a robustly optimal mechanism that maximizes her worst-case profit evaluated against all possible buyer type (joint) distributions F over $(\theta_1, \ldots, \theta_K)$ that have this known average total demand (so $\mathbb{E}_F[\theta_1 + \cdots + \theta_K] = \lambda$).

The central result of this paper is to show that a committed spend mechanism—that we characterize explicitly—is robustly optimal. The main applied contribution of this result is that it provides an intuitive explanation for why committed spend mechanisms are widely used in cloud computing, highlighting their robustness and, as we argue below, efficiency. This result bridges the gap between theoretical mechanism design and real-world pricing practices, offering insights for practitioners and policymakers. Importantly, the optimal mechanism does not depend on, and therefore does not require the seller to solicit, the buyer's type distribution. In this multidimensional environment, the type distribution can be complex and hard to elicit in practice. Lastly, we show that the price p for excess usage is simply the seller's cost. Consequently, conditional on the buyer having high-demand, the mechanism allocates efficiently.

1.2. *Related Literature*

This paper lies at the intersection of three literatures: sequential screening, multidimensional screening and robust mechanism design.

Because the buyer in our model does not know their exact demand at the time of contracting, we are in a sequential screening environment. Following Courty and Li (2000), the literature on sequential screening derives optimal contracts in a variety of different settings where buyers first learn their demand distribution for a single good and then their true demand is realized from this distribution in the second period. Because the buyer's private information realizes sequentially, the technical challenges associated with multidimensional screening can be skirted to a large extent. There are multiple goods in our model so all the technical challenges of multidimensional screening are present because the buyer learns his (multidimensional) type in the second period. We are nonetheless able to characterize the optimal mechanism because we employ a robust optimization criterion. Consequently, we view our modeling choices to be an important contribution of this paper as, in our opinion, our model is both a natural fit for the pricing of cloud computing and nonetheless is tractable enough to characterize the (robustly) optimal mechanism in a multidimensional environment.

The literature on multidimensional screening—in which the buyer knows his type at the moment of contracting—dates back till at least Adams and Yellen (1976). But, despite being a canonical problem, a general result characterizing the optimal mechanism (in a standard Bayesian environment) akin to Myerson (1981) (for the single good case) has proved elusive. Schmalensee (1984) and McAfee, McMillan, and Whinston (1989) showed that, for additive values and independent types, it is never optimal to sell items separately as the seller can do better by additionally offering the grand bundle at a discount (although this latter mechanism is also generally not optimal). Bakos and Brynjolfsson (1999) showed that bundling a large number of independent products can improve profit due to the better predictability of consumers' valuations; the seller can extract almost all the surplus as the number of goods gets large. Armstrong (2013) relaxed the assumption of additive values of two products and showed that even for potentially non-additive values, offering a bundle with a discounted price increases the profit whenever the demand for a single item is less elastic than the demand for the bundle.

The optimal mechanism for multiple products can take a very complex form. Even for two goods and additive values that are independently drawn from the same uniform distribution, Pavlov (2011) shows that the optimal mechanism can involve randomization. When the values are drawn from a beta distribution, Daskalakis, Deckelbaum, and Tzamos (2017) show that the optimal mechanism can feature an infinite menu of lotteries. In fact, Hart and Nisan (2019) show that profit maximizing mechanism that only involves pure bundling or selling goods separately, might leave the seller with a negligible fraction of the profit from the optimal mechanism. Thus, there is a disconnect between the complex optimal mechanisms predicted by theory and the simple mechanisms we observe in practice.

This disconnect can be reduced by changing the optimization criterion from Bayesian to robust. This was first observed by Carroll (2017) who considers a seller who has information about the buyer's marginal value distribution for each good but not the joint distribution. He shows that selling each good separately is a robustly optimal mechanism when the seller evaluates worst case profits against all joint distributions with the given marginals. Che and Zhong (forthcoming) and Deb and Roesler (2024) consider different robustness criteria to that of Carroll and establish conditions under which it is robustly optimal to sell bundles of goods. While bundling is commonly observed in practice, the mechanisms in these papers require the seller to randomize the

bundle prices which is not something that real world sellers do. Specifically, for our robustness criterion (that the seller knows the mean of the total demand but nothing further), Che and Zhong (forthcoming) show that the optimal mechanism involves pure bundling with a random bundle price.

In contrast to these papers, we consider a multidimensional *sequential screening* environment. Moreover, in our setting, there are *multiple units* of each good for sale. That is, while the papers mentioned in the previous paragraph are multidimensional versions of Myerson (1981), our setting is a multdimensional, sequential screening version of Maskin and Riley (1984). In addition to realism, the advantage of the sequential screening environment we study is that we show a committed spend mechanism is robustly optimal that, in particular, does not require any price randomization.

2. MODEL

We study the sequential screening problem of a single seller of $\mathcal{K} = \{1, \ldots, K\}$ with $K \geq 1$ cloud computing services facing a single (representative) buyer. The seller offers a menu of contracts where each contract determines the price that the buyer would pay for using different quantities of the K services. The buyer picks a contract from the menu, learns his computing needs and then determines his consumption of cloud computing services based on the prices of his chosen contract. In the remainder of this section, we describe this interaction formally.

The buyer privately knows the (joint) distribution $F \in \Delta([0,1]^K)$ that determines his type $\theta = (\theta_1, \ldots, \theta_K) \in [0,1]^K =: \Theta$ that, in turn, determines his demand for computing.[1] The value of a type θ for the quantity $q = (q_1 \ldots, q_K) \in \mathbb{R}_+^K$ of different computing services is given by

$$\sum_{i=1}^{K} \theta_i u(q_i),$$

where the utility function $u : \mathbb{R}_+ \to \mathbb{R}_+$ is assumed to be strictly increasing and strictly concave with $u'(0) = \infty$ and $\lim_{q_i \to \infty} u'(q_i) = 0$. We additionally assume that the buyer's preference is quasilinear in transfers.

---

[1]The restriction of the type space to $\Theta = [0,1]^K$ is simply a normalization to reduce notation. Instead of 1, we could have chosen the upper bound arbitrarily and no result would change.

The cost to the seller of providing the buyer with the services q is

$$c \sum_{i=1}^{K} q_i$$

where $c \in \mathbb{R}_{++}$.

The buyer's type distribution F lies in a set $\mathcal{F} \subseteq \Delta([0,1]^K)$ where

$$\mathcal{F} := \left\{ F \in \Delta([0,1]^K) \,\middle|\, \int_{\Theta} (\theta_1 + \cdots + \theta_K) dF(\theta) = \lambda \right\}.$$

In words, the distribution from which the buyer's type is drawn can be any distribution such that the expectation of the sum of the K dimensions is $\lambda \in (0, K)$.

The timing of the interaction between the buyer and seller is the following:

(1) The seller offers the buyer a mechanism (Q, T) where Q is the allocation and T is the transfer. We formally define a mechanism below.

(2) The buyer is privately informed about her distribution $F \in \mathcal{F}$, reports $F' \in \mathcal{F}$ to the mechanism.

(3) The buyer's value θ is drawn from distribution F. The buyer reports value $\theta' \in \Theta$.

(4) The buyer and seller receive their respective payoffs from the mechanism.

The revelation principle applies so it is without loss to assume that, in step 1, the seller offers the buyer a (direct) mechanism

$$Q : \mathcal{F} \times \Theta \to \mathbb{R}_+^K \quad \text{and} \quad T : \mathcal{F} \times \Theta \to \mathbb{R},$$

that satisfies incentive compatibility and individual rationality. The function $Q(F', \theta')$ determines the quantity of each service that the buyer receives when she reports F' in step 2 and θ' in step 3. $T(F', \theta')$ is the transfer from the buyer to the seller.

Incentive compatibility (or IC for short) requires that

$$\sum_{i=1}^{K} \theta_i u(Q_i(F, \theta)) - T(F, \theta) \geq \sum_{i=1}^{K} \theta_i u(Q_i(F, \theta')) - T(F, \theta') \text{ for all } \theta, \theta' \in \Theta \text{ and all } F \in \mathcal{F}$$

and

$$\mathbb{E}_F \left[\sum_{i=1}^{K} \theta_i u(Q_i(F, \theta)) - T(F, \theta) \right] \geq \mathbb{E}_F \left[\sum_{i=1}^{K} \theta_i u(Q_i(F', \theta)) - T(F', \theta) \right] \text{ for all } F, F' \in \mathcal{F}.$$

The incentive compatibility (IC) constraint can be written this way (as opposed to in terms of joint misreports of F and θ) because, as in standard sequential screening, once the buyer's value θ is realized, the distribution F no longer affects their utility.

Individual rationality (or IR for short) requires that

$$\mathbb{E}_F \left[\sum_{i=1}^{K} \theta_i u(Q_i(F,\theta)) - T(F,\theta) \right] \geq 0 \text{ for all } F \in \mathcal{F}.$$

We denote the set of all IC and IR mechanisms as $\mathcal{M}$.

The seller solves the following robust mechanism design problem

$$(Q^\star, T^\star) \in \operatorname*{argmax}_{(Q,T) \in \mathcal{M}} \inf_{F \in \mathcal{F}} \left\{ \mathbb{E}_F \left[T(F,\theta) - c \sum_{i=1}^{K} Q_i(F,\theta) \right] \right\}. \tag{1}$$

In words, the seller chooses a mechanism that maximizes her payoff again the worst-possible type distribution in the set $\mathcal{F}$ that the buyer might have. We refer to a solution $(Q^\star, T^\star)$ as a *robustly optimal mechanism*.

Before we present our main result in the next section, it is worth making a few remarks. First, note that the statement (1) of the seller's problem implicitly assumes that a robustly optimal mechanism exists. We will show that this is indeed the case.

Second, observe that the allocation Q is restricted to be deterministic; there is obviously no benefit in randomizing the transfer since this is enters linearly in the payoff of both the seller and the buyer. As we argue next, the restriction to deterministic allocations is without loss.

LEMMA 1. *Let*

$$\mathcal{M}^r := \{ (Q,T) \mid Q : \mathcal{F} \times \Theta \to \Delta(\mathbb{R}_+^K), \ T : \mathcal{F} \times \Theta \to \Delta(\mathbb{R}) \text{ and } (Q,T) \text{ is IC and IR} \}$$

be the set of incentive compatible and individually rational random mechanisms. Then,

$$\max_{(Q,T) \in \mathcal{M}^r} \inf_{F \in \mathcal{F}} \left\{ \mathbb{E}_F \left[T(F,\theta) - c \sum_{i=1}^{K} Q_i(F,\theta) \right] \right\} = \max_{(Q,T) \in \mathcal{M}} \inf_{F \in \mathcal{F}} \left\{ \mathbb{E}_F \left[T(F,\theta) - c \sum_{i=1}^{K} Q_i(F,\theta) \right] \right\}$$

or, in words, that the seller does not achieve a higher profit guarantee by employing a random mechanism.

PROOF. The proof of the lemma employs a simple argument: we take any mechanism that involves randomization and then construct an alternate deterministic mechanism that yields the seller a higher payoff.

Consider an IC and IR mechanism (Q, T) in which both the allocation $Q : \mathcal{F} \times \Theta \to \Delta(\mathbb{R}_+^K)$ and transfer $T : \mathcal{F} \times \Theta \to \Delta(\mathbb{R})$ can be random.

Construct a deterministic mechanism (Q', T') such that

$$u(Q'_i(F,\theta)) = \mathbb{E}_{Q_i(F,\theta)}[u(q_i)] \text{ and } T'(F,\theta) = \mathbb{E}_{T(F,\theta)}[t] \quad \text{for all } i \in \mathcal{K}, \ F \in \mathcal{F} \text{ and } \theta \in \Theta,$$

where the first expectation is taken over q_i with respect to the (marginal) allocation distribution $Q_i(F,\theta)$ and the second expectation is taken over t with respect to the transfer distribution $T(F,\theta)$.

By construction, we have that

$$\sum_{i=1}^{K} \theta_i u(Q'_i(F,\theta')) - T'(F,\theta') = \sum_{i=1}^{K} \theta_i \mathbb{E}_{Q_i(F,\theta')}[u(q_i)] - \mathbb{E}_{T(F,\theta')}[t],$$

where the left and right sides are the utilities of a buyer who reported distribution $F \in \mathcal{F}$ in step 2, had type $\theta \in \Theta$ realize but reported type $\theta' \in \Theta$ in step 3 from mechanisms (Q', T') and (Q, T) respectively. Consequently, the fact that (Q, T) is IC and IR implies that the deterministic mechanism (Q', T') is also IC and IR.

Lastly observe that, since u is concave, $u(Q'_i(F,\theta)) = \mathbb{E}_{Q_i(F,\theta)}[u(q_i)]$ implies that the certainty equivalent $Q'_i(F,\theta) \leq \mathbb{E}_{Q_i(F,\theta)}[q_i]$ for all $i \in \mathcal{K}, F \in \mathcal{F}$ and $\theta \in \Theta$. Consequently, the seller's expected cost

$$\int_{\Theta} c \sum_{i=1}^{K} Q'_i(F,\theta) dF(\theta) \leq \int_{\Theta} c \sum_{i=1}^{K} \mathbb{E}_{Q_i(F,\theta)}[q_i] dF(\theta)$$

is lower for mechanism (Q', T') for all distributions $F \in \mathcal{F}$.

Since, the expected transfers from both (Q', T') and (Q, T) are same by construction, this implies that the seller's worst case profit (taken over $F \in \mathcal{F}$) from (Q', T') is weakly greater than that from (Q, T). This shows that the seller has no incentive to choose a mechanism with a randomized allocation and transfer. ∎

In words, because the buyer's utility is strictly concave in the quantity of each service, the seller has no incentive to offer a random allocation rule $Q : \mathcal{F} \times \Theta \to \Delta(\mathbb{R}_+^K)$ since she could instead offer certainty equivalents at a lower cost without changing buyer incentives. We chose

not to allow random allocations while defining the model above as this needlessly complicates the presentation.

3. THE OPTIMALITY OF COMMITTED SPEND MECHANISMS

Before we present our main result, we need one additional definition.

A *committed spend mechanism* $(Q^b, T^b) \in \mathcal{M}$ is defined by three values $\underline{q} \in \mathbb{R}_+$, $\underline{t} \in \mathbb{R}_+$ and $p \in \mathbb{R}_{++}$ that determine the allocation and transfer as follows:

$$Q^b(F, \theta) \in \underset{(q_1, \dots, q_K)}{\operatorname{argmax}} \left\{ \sum_{i=1}^{K} \theta_i u(q_i) - \underline{t} - p \left(\sum_{i=1}^{K} q_i - \underline{q} \right) \right\} \text{ such that } \sum_{i=1}^{K} q_i \geq \underline{q},$$

$$T^b(F, \theta) = \underline{t} + p \left(\sum_{i=1}^{K} Q_i^b(F, \theta) - \underline{q} \right).$$

In words, a committed spend mechanism requires the buyer to commit to a consuming a minimal total (summed across services) quantity $\underline{q}$ and making a minimal payment $\underline{t}$. The buyer can choose to consume a higher total quantity and pay for the additional quantity consumed at a unit price p. The buyer is free to pick how she allocates the total quantity across different services. Note that, our assumptions on u imply that Q^b is well defined as there is always a solution to the maximization problem that defines the quantity consumed. Indeed, the solution is unique for all $\theta \neq 0$.

As we discussed earlier, cloud computing is typically priced using a committed spend mechanism. Our main result shows that there is a robustly optimal mechanism of this form!

THEOREM 1. *The committed spend mechanism* $(Q^{b\star}, T^{b\star})$ *defined by*

$$\underline{q}^\star = K u'^{-1} \left(\frac{cK}{\lambda} \right), \quad \underline{t}^\star = \lambda u \left(\frac{q^\star}{K} \right) \text{ and } p^\star = c$$

is a robustly optimal mechanism; that is, it is a solution to the principal's problem (1).

This result states that there is a robust optimal mechanism in which the buyer commits to using at least a total $\underline{q}^\star$ of the various cloud computing services and paying at least $\underline{t}^\star$. The buyer has the option of consuming a higher quantity, should he need too, and additional units are priced by the seller at her cost c.

Observe a few nice properties of this robustly optimal mechanism. Such mechanisms are observed in practice and do not have the unrealistic feature of requiring the buyer to report his type distribution F. Moreover, the mechanism is such that, whenever the buyer consumes a higher total quantity than $\underline{q}^\star$, his chosen quantity is the socially efficient level since it maximizes $\sum_{i=1}^{K} \theta_i u(q_i) - c \sum_{i=1}^{K} q_i$! As the proof of Theorem 1 in the next section shows, there are multiple robustly optimal mechanisms and this feature makes the committed spend mechanism particularly appealing.

4. THE PROOF OF THEOREM 1

The result is proved in several steps. We first establish a robustly optimal mechanism for the special case of a single service ($K = 1$). We then use that to derive a robustly optimal mechanism for the case of multiple services ($K > 1$). This mechanism is not a committed spend mechanism. In the final step, we show that there is also a committed spend mechanism that is robustly optimal and strictly more efficient.

The first lemma derives a robustly optimal mechanism for a single service (that is, $K = 1$).

LEMMA 2. *Suppose $K = 1$. There is a robust optimal mechanism $Q^*(F, \theta) = q^*$, $T^*(F, \theta) = \lambda u(q^*)$ for all $F \in \mathcal{F}$ and $\theta \in [0, 1]$ where $q^* = u'^{-1}\left(\frac{c}{\lambda}\right)$.*

PROOF. Let $\delta_\lambda \in \mathcal{F}$ denote the Dirac distribution that assigns probability one to the mean λ.

Take any IC and IR mechanism $(Q, T) \in \mathcal{M}$. Observe that because (Q, T) is IR, we have

$$\mathbb{E}_{\delta_\lambda} \left[\theta u(Q(\delta_\lambda, \theta)) - T(\delta_\lambda, \theta) \right] \geq 0 \iff T(\delta_\lambda, \lambda) \leq \lambda u(Q(\delta_\lambda, \lambda)).$$

Now define a new mechanism $(\hat{Q}, \hat{T})$ such that

$$\hat{Q}(F, \theta) = Q(\delta_\lambda, \lambda), \ \hat{T}(F, \theta) = \lambda u(Q(\delta_\lambda, \lambda)) \text{ for all } F \in \mathcal{F} \text{ and } \theta \in [0, 1].$$

In words, the mechanism $(\hat{Q}, \hat{T})$ is constant in both F and θ with the allocation and transfer being $Q(\delta_\lambda, \lambda)$ and $\lambda u(Q(\delta_\lambda, \lambda))$ respectively.

$(\hat{Q}, \hat{T})$ is clearly IC since it is a constant mechanism and so misreporting F and θ makes no difference to either the allocation or the transfer. Also,

$$\mathbb{E}_F\left[\theta u(\hat{Q}(F,\theta)) - \hat{T}(\theta)\right] = \mathbb{E}_F[\theta]u(Q(\delta_\lambda,\lambda)) - \lambda u(Q(\delta_\lambda,\lambda)) = 0 \quad \text{for all } F \in \mathcal{F}$$

because $\mathbb{E}_F[\theta] = \lambda$. Consequently, $(\hat{Q}, \hat{T})$ is also IR and so $(\hat{Q}, \hat{T}) \in \mathcal{M}$.

Therefore, for any $(Q,T) \in \mathcal{M}$, we have

$$\inf_{F \in \mathcal{F}} \mathbb{E}_F\left[T(F,\theta) - cQ(F,\theta)\right] \leq \mathbb{E}_{\delta_\lambda}\left[T(\delta_\lambda,\theta) - cQ(\delta_\lambda,\theta)\right]$$

$$= T(\delta_\lambda,\lambda) - cQ(\delta_\lambda,\lambda)$$

$$\leq \lambda u(Q(\delta_\lambda,\lambda)) - cQ(\delta_\lambda,\lambda)$$

$$= \inf_{F \in \mathcal{F}} \mathbb{E}_F\left[\hat{T}(F,\theta) - c\hat{Q}(F,\theta)\right].$$

In words, relative to any mechanism $(Q,T) \in \mathcal{M}$, the principal gets a higher profit guarantee offering the mechanism $(\hat{Q}, \hat{T}) \in \mathcal{M}$ instead. Thus, there is a robustly optimal mechanism that takes the form $(\hat{Q}, \hat{T})$ (were a robustly optimal mechanism to exist).

Let,

$$q^* = \operatorname*{argmax}_{q}\{\lambda u(q) - cq\}$$

and so, from the first-order condition, we have

$$q^* = u'^{-1}\left(\frac{c}{\lambda}\right).$$

In words, this states that, if we were to consider mechanisms that always offer the constant quantity q along with a transfer $\lambda u(q)$, the highest possible profit is obtained by choosing the quantity $q^* = u'^{-1}\left(\frac{c}{\lambda}\right)$. Combining this observation with the above argument, we have shown that there is a robustly optimal mechanism

$$Q^*(F,\theta) = q^*, \ T^*(F,\theta) = \lambda u(q^*) \text{ for all } F \in \mathcal{F} \text{ and } \theta \in [0,1]$$

which completes the proof of the lemma. ∎

11

We now move to the case where there are multiple services offered by the seller (so $K > 1$). We begin with some notation. Let

$$\Theta^d := \{(\tilde{\theta}, \ldots, \tilde{\theta}) \mid \tilde{\theta} \in [0,1]\} \subset \Theta$$

denote the subset of types that lie on the diagonal. We use $\mathcal{F}^d \subset \mathcal{F}$ to denote the subset of (perfectly correlated) distributions that are defined on the set Θ^d, that is, $F^d(\Theta^d) = 1$ for all $F^d \in \mathcal{F}^d$.[2]

As an intermediate step, we now derive a robustly optimal mechanism for the case where the buyer's type distribution F^d lies in $\mathcal{F}^d$ and consequently, the buyer's type lies on the diagonal Θ^d. Formally, we now solve

$$(Q^{d^\star}, T^{d^\star}) \in \operatorname*{argmax}_{(Q,T) \in \mathcal{M}} \inf_{F^d \in \mathcal{F}^d} \left\{ \mathbb{E}_{F^d} \left[T(F^d, \theta) - c \sum_{i=1}^{K} Q_i(F^d, \theta) \right] \right\}, \tag{2}$$

the solution to which is described in the next lemma.

LEMMA 3. *There is a solution to the problem* (2) *given by* $Q_i^{d^\star}(F, \theta) = q^{d^\star}$, $T^{d^\star}(F, \theta) = \lambda u(q^{d^\star})$ *for all* $F \in \mathcal{F}$ *and* $\tilde{\theta} \in \Theta$ *where* $q^{d^\star} = u'^{-1}\left(\frac{cK}{\lambda}\right) q^*$.

PROOF. Take any IC and IR mechanism $(Q, T) \in \mathcal{M}$ and any distribution $F^d \in \mathcal{F}^d$. Define an alternate mechanism (Q^d, T^d) where the allocation for types along the diagonal is given by the solution to

$$u(Q_i^d(F^d, (\tilde{\theta}, \ldots, \tilde{\theta}))) = \frac{1}{K} \sum_{j=1}^{K} u(Q_j(F^d, (\tilde{\theta}, \ldots, \tilde{\theta}))) \quad \text{for all} \quad \tilde{\theta} \in [0,1], \ i \in \mathcal{K}$$

and the allocation

$$Q^d(F^d, \theta) = Q^d \left(F^d, \left(\frac{\theta_1 + \cdots + \theta_K}{K}, \ldots, \frac{\theta_1 + \cdots + \theta_K}{K} \right) \right)$$

for all other types $\theta \in \Theta \backslash \Theta^d$ is defined to be the allocation corresponding to the diagonal type $(\tilde{\theta}, \ldots, \tilde{\theta})$ where $K\tilde{\theta} = \theta_1 + \cdots + \theta_K$.

The transfers T^d are the same

$$T^d(F^d, \theta^d) = T(F^d, \theta^d)$$

[2] In a slight abuse of notation, we use F to refer both to a cumulative distribution and a measure. The meaning should be clear from the argument (type or set of types respectively).

for types $\theta^d \in \Theta^d$ that lie on the diagonal. For all other types $\theta \in \Theta \backslash \Theta^d$, the transfer

$$T^d(F^d, \theta) = T\left(F^d, \left(\frac{\theta_1 + \cdots + \theta_K}{K}, \ldots, \frac{\theta_1 + \cdots + \theta_K}{K}\right)\right)$$

is the same as the diagonal type $(\tilde{\theta}, \ldots, \tilde{\theta})$ where $K\tilde{\theta} = \theta_1 + \cdots + \theta_K$.

For any distribution $F \in \mathcal{F} \backslash \mathcal{F}^d$ that is not in the set $\mathcal{F}^d$, let $G_F^d \in \mathcal{F}^d$ be the distribution such that

$$G_F^d(\tilde{\theta}, \ldots, \tilde{\theta}) = F(\{(\theta \in \Theta \mid \theta_1 + \cdots + \theta_K \leq K\tilde{\theta})\}) \quad \text{for all } \tilde{\theta} \in [0, 1].$$

In words, $G_F^d \in \mathcal{F}^d$ is the distribution supported on the diagonal Θ^d that induces the same distribution on the sum $\theta_1 + \cdots + \theta_K$ as F. Note that the distribution $G_F^d \in \mathcal{F}^d$ always exists for all $F \in \mathcal{F}$.

Now, for all $F \in \mathcal{F} \backslash \mathcal{F}^d$, we define

$$Q^d(F, \theta) = Q^d(G_F^d, \theta) \text{ and } T^d(F, \theta) = T^d(G_F^d, \theta)$$

where $Q^d(G_F^d, \theta)$ and $T^d(G_F^d, \theta)$ are defined above.

Observe that, for any reported $F^d \in \mathcal{F}^d$, if type $(\tilde{\theta}, \ldots, \tilde{\theta})$ reports as type $(\tilde{\theta}', \ldots, \tilde{\theta}')$, their payoff in the mechanism (Q^d, T^d)

$$\sum_{i=1}^K \tilde{\theta} u(Q_i^d(F^d, (\tilde{\theta}', \ldots, \tilde{\theta}'^d(F^d, (\tilde{\theta}', \ldots, \tilde{\theta}')) = \tilde{\theta} \sum_{i=1}^K u(Q_i(F^d, (\tilde{\theta}', \ldots, \tilde{\theta}'^d, (\tilde{\theta}', \ldots, \tilde{\theta}'))$$

is the same as that from the original mechanism (Q, T). Moreover, for any type $\theta \in \Theta \backslash \Theta^d$ that is not on the diagonal, the allocation Q^d and transfer T^d is the same as the type $\left(\frac{\theta_1 + \cdots + \theta_K}{K}, \ldots, \frac{\theta_1 + \cdots + \theta_K}{K}\right)$. Moreover, for all types, the same quantity of each service is allocated. Consequently, the payoff that type θ gets from misreporting as any other type $\theta' \in \Theta$ is the same as the payoff that type $\left(\frac{\theta_1 + \cdots + \theta_K}{K}, \ldots, \frac{\theta_1 + \cdots + \theta_K}{K}\right)$ receives by misreporting as $\left(\frac{\theta_1' + \cdots + \theta_K'}{K}, \ldots, \frac{\theta_1' + \cdots + \theta_K'}{K}\right)$. Therefore, because the original mechanism (Q, T) is IC, the buyer has no incentive to misreport their type θ (in step 3) after reporting a distribution $F^d \in \mathcal{F}^d$ in step 2.

Since we defined $Q^d(F, \theta) = Q^d(G_F^d, \theta)$ and $T^d(F, \theta) = T^d(G_F^d, \theta)$, the above argument also implies that the buyer has no incentive to misreport their type θ if they report a distribution $F \in \mathcal{F} \backslash \mathcal{F}^d$ in step 2.

Observe that

$$\mathbb{E}_{F^d}\left[\sum_{i=1}^{K}\theta_i u(Q_i^d(\tilde{F}^d,\theta)) - T^d(\tilde{F}^d,\theta)\right] = \mathbb{E}_{F^d}\left[\sum_{i=1}^{K}\theta_i u(Q_i(\tilde{F}^d,\theta)) - T(\tilde{F}^d,\theta)\right] \text{ for all } F^d, \tilde{F}^d \in \mathcal{F}^d$$

because we have already argued that the type θ is truthfully reported in step 3. This implies that since, (Q,T) is IC, the buyer whose type distribution is $F^d, \tilde{F}^d \in \mathcal{F}^d$ has no incentive to misreport as $F^d, \tilde{F}^d \in \mathcal{F}^d$. Moreover, since (Q,T) is IR, the buyer's expected utility from reporting both F^d and θ truthfully is at least zero.

For any $F \in \mathcal{F} \backslash \mathcal{F}^d$ and $F' \in \mathcal{F}$, observe that

$$\mathbb{E}_F\left[\sum_{i=1}^{K}\theta_i u(Q_i^d(F'^d(F',\theta))\right] = \mathbb{E}_{G_F^d}\left[\sum_{i=1}^{K}\theta_i u(Q_i^d(G_{F'}^d,\theta)) - T^d(G_{F'}^d,\theta)\right].$$

In words, the expected utility that a buyer with type distribution F receives from the mechanism (Q^d, T^d) when she misreports as type distribution F' is the same as the utility that a buyer with type distribution G_F^d gets from misreporting as $G_{F'}^d$. We have already argued above that such a misreport does not benefit the buyer.

Thus, taken together, we have shown that the mechanism (Q^d, T^d) is IC, IR and provides the same expected utility as (Q,T) whenever the buyer's type distribution is $F^d \in \mathcal{F}^d$.

Now, note that the concavity of u implies that

$$u(Q_i^d(\tilde{\theta},\ldots,\tilde{\theta})) \le u\left(\frac{1}{K}\sum_{j=1}^{K}Q_j(\tilde{\theta},\ldots,\tilde{\theta})\right) \implies KQ_i^d(\tilde{\theta},\ldots,\tilde{\theta}) \le \sum_{j=1}^{K}Q_j(\tilde{\theta},\ldots,\tilde{\theta}) \tag{3}$$

where the second inequality follows from the fact that u is strictly increasing.

Thus, for any $F^d \in \mathcal{F}^d$, the principal's payoff from $(\hat{Q}^d, T^d)$

$$\mathbb{E}_{F^d}\left[T^d(\tilde{\theta},\ldots,\tilde{\theta}) - c\sum_{i=1}^{K}Q_i^d(\tilde{\theta},\ldots,\tilde{\theta})\right] \ge \mathbb{E}_{F^d}\left[T(\tilde{\theta},\ldots,\tilde{\theta}) - c\sum_{i=1}^{K}Q_i(\tilde{\theta},\ldots,\tilde{\theta})\right]$$

is higher than the original mechanism (Q,T) because the transfer is the same for all types in the support of F^d and, the total cost is lower (due of inequality 3).

Thus, in order to solve the problem 2, it is without loss for the seller to consider mechanisms where the quantity $Q_1(\cdot) = \cdots = Q_K(\cdot)$ is the same for all goods. This, in turn, implies that the seller's problem collapses to that of a single good. The characterization of the solution to the seller's problem (2) then follows from Lemma 2. ∎

14

We use Lemma 3 to characterize a robustly optimal mechanism for the seller. This is not a committed spend mechanism but we will show that a committed spend mechanism is also robustly optimal and strictly more efficient.

LEMMA 4. *There is a robust optimal mechanism $Q_i^*(F, \theta) = q^*$, $T^*(F, \theta) = \lambda u(q^*)$ for all $F \in \mathcal{F}$ and all $\theta \in \Theta$ where $q^* = u'^{-1}\left(\frac{cK}{\lambda}\right)$*

PROOF. Observe that

$$\max_{(Q,T) \in \mathcal{M}} \inf_{F \in \mathcal{F}} \mathbb{E}_F \left[T(F, \theta) - c \sum_{i=1}^{K} Q_i(F, \theta) \right] \leq \max_{(Q,T) \in \mathcal{M}} \inf_{F^d \in \mathcal{F}^d} \mathbb{E}_{F^d} \left[T(\theta) - c \sum_{i=1}^{K} Q_i(\theta) \right]$$

$$= \lambda u(q^*) - cKq^*$$

$$= \inf_{F \in \mathcal{F}} \mathbb{E}_F \left[T^*(F, \theta) - c \sum_{i=1}^{K} Q_i^*(F, \theta) \right],$$

where the inequality follows from the fact that the infimum is being taken over a smaller set, the first equality follows from Lemma 3 and the last inequality follows from the definition of (Q^*, T^*) in the statement of the lemma. The robust optimality of (Q^*, T^*) follows. ∎

Theorem 1 is an immediate consequence of Lemma 4 and we complete the proof below.

PROOF OF THEOREM 1. Consider the committed spend mechanism defined in the statement of the theorem and recall that a committed spend mechanism does not depend on the reported type distribution $F \in \mathcal{F}$. We now compute the seller's profit from each type $\theta \in \Theta$ and show that it is the same as the profit the seller obtains from the mechanism (Q^*, T^*).

For each $F \in \mathcal{F}$ and $\theta \in \Theta$, observe that the seller's payoff is

$$T^{b\star}(F, \theta) - c \sum_{i=1}^{K} Q_i^{b\star}(F, \theta) = \lambda u\left(\frac{q^\star}{K}\right) + c \left(\sum_{i=1}^{K} Q_i^{b\star}(F, \theta) - \underline{q}^\star\right) - c\underline{q}^\star - c \left(\sum_{i=1}^{K} Q_i^{b\star}(F, \theta) - \underline{q}^\star\right)$$

$$= \lambda u\left(\frac{q^\star}{K}\right) - c\underline{q}^\star$$

$$= \lambda u(q^*) - cKq^*$$

$$= T^*(F, \theta) - c \sum_{i=1}^{K} Q_i^*(F, \theta).$$

The committed spend mechanism is IC by definition. Note that for all $F \in \mathcal{F}$ and all $\theta \in \Theta$, the buyer's payoff is at least as high as his payoff from (Q^*, T^*). Consequently, $(Q^{b\star}, T^{b\star})$ is also IR

(that is, $(Q^{b^\star}, T^{b^\star}) \in \mathcal{M}$) and its robust optimality follows from the robust optimality of (Q^*, T^*). This completes the proof of the theorem. ∎

5. Conclusion

This paper has examined the problem of pricing in cloud computing through the lens of robust mechanism design. By formulating a sequential screening model, we have characterized the optimal mechanism for a seller facing a buyer with private information about their type distribution. Our main result demonstrates that a committed spend mechanism—requiring a minimum commitment and pricing additional usage at cost—is robustly optimal.

Three key insights emerge from our analysis. First, the ability to achieve optimal profits with mechanisms that do not require buyers to report their demand distributions indicates that simpler pricing schemes may be preferable to more complex ones. Second, the flexibility afforded to buyers in allocating their usage across services aligns with practical needs while preserving optimality. Finally, the optimality of setting marginal prices equal to marginal costs for usage above the minimum commitment suggests that efficiency considerations should guide pricing of incremental consumption.

The findings have significant implications for the design of cloud computing contracts. committed spend mechanisms are not only theoretically optimal but also align closely with practical pricing schemes observed in the industry. Their robustness to information asymmetry and efficiency in resource allocation make them particularly appealing in dynamic and uncertain environments. Moreover, the simplicity of these mechanisms enhances their feasibility for implementation, reducing the informational burden on both buyers and sellers.

Future research could explore extensions of this framework to multi-seller markets, dynamic pricing scenarios, and environments with additional sources of uncertainty. By doing so, we can further deepen our understanding of optimal pricing strategies in cloud computing and related industries.

To conclude, this paper demonstrates that simple committed spend mechanisms are optimal for pricing cloud computing services under demand uncertainty. Our results provide theoretical support for common industry practices while highlighting how elementary pricing structures can achieve optimal outcomes in complex environments. The robustness of committed

spend mechanisms—their ability to achieve optimal profits without detailed knowledge of demand patterns—may help explain their widespread adoption in cloud computing markets.

REFERENCES

ADAMS, W. J., AND J. L. YELLEN (1976): "Commodity Bundling and the Burden of Monopoly," *Quarterly Journal of Economics*, 90(3), 475–498.

ARMSTRONG, M. (2013): "A More General Theory of Commodity Bundling," *Journal of Economic Theory*, 148(2), 448–472.

BAKOS, Y., AND E. BRYNJOLFSSON (1999): "Bundling Information Goods: Pricing, Profits, and Efficiency," *Management Science*, 45(12), 1613–1630.

CARROLL, G. (2017): "Robustness and Separation in Multidimensional Screening," *Econometrica*, 85(2), 453–488.

CHE, Y.-K., AND W. ZHONG (forthcoming): "Robustly-Optimal Mechanism for Selling Multiple Goods," *Review of Economic Studies*.

COURTY, P., AND H. LI (2000): "Sequential screening," *Review of Economic Studies*, 67(4), 697–717.

DASKALAKIS, C., A. DECKELBAUM, AND C. TZAMOS (2017): "Strong Duality for a Multiple-Good Monopolist," *Econometrica*, 85(3), 735–767.

DEB, R., AND A.-K. ROESLER (2024): "Multi-Dimensional Screening: Buyer-Optimal Learning and Informational Robustness," *Review of Economic Studies*, 91(5), 2744–2770.

HART, S., AND N. NISAN (2019): "Selling Multiple Correlated Goods: Revenue Maximization and Menu-Size Complexity," *Journal of Economic Theory*, 183, 991–1029.

MASKIN, E., AND J. RILEY (1984): "Monopoly with Incomplete Information," *RAND Journal of Economics*, 15(2), 171–196.

MCAFEE, R. P., J. MCMILLAN, AND M. D. WHINSTON (1989): "Multiproduct Monopoly, Commodity Bundling, and Correlation of Values," *Quarterly Journal of Economics*, 104(2), 371–383.

MYERSON, R. B. (1981): "Optimal auction design," *Mathematics of Operations Research*, 6(1), 58–73.

PAVLOV, G. (2011): "Optimal Mechanism for Selling Two Goods," *BE Journal of Theoretical Economics*, 11(1).

SCHMALENSEE, R. (1984): "Gaussian Demand and Commodity Bundling," *Journal of Business*, pp. S211–S230.