Strategic Pricing and Resource Allocation: Framework and Applications

Shaolei Ren
Electrical Engineering Department
University of California, Los Angeles

Ph.D. Advisor: Prof. Mihaela van der Schaar
Outline

- Limitations and Opportunities
- Two Pricing Algorithms
- Future Work
Information and Communication Technology

We’ve entered the information age...
Information and Communication Technology

- Wireless communications
- Social networking
- Cloud computing
- Smart grid
Information and Communication Technology

How to make a technology more profitable?
Limitations and Opportunities

- Limitations
  - Engineering approach: not *proactively* reshape the user demand
Limitations and Opportunities

- **Limitations**
  - Engineering approach: not *proactively* reshape the user demand
  - Economics approach: treat engineering as a "black box"

---

**Pricing**
Limitations and Opportunities

- **Limitations**
  - Engineering approach: not *proactively* reshape the user demand
  - Economics approach: treat engineering as a "black box"

- **Opportunities**
  - Integrated design of pricing and resource management
    - User heterogeneity
    - Possibly random environment
    - Repeated interactions
Outline

Pricing
“A group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content.”

--- A. Kaplan, et. al.
A group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content.

---

A. Kaplan, et. al.
User-Generated Content Platforms

- **Content Producers**
- **Content Viewers**
- **Platform Owner (a.k.a. Intermediary)**

Intra-group negative externalities

Social incentive

12
User-Generated Content Platforms

- Content Producers
- Platform Owner (a.k.a. Intermediary)
- Content Viewers

Content

Platform Owner

($$)
User-Generated Content Platforms

Content Producers

Content

Platform Owner
(a.k.a. Intermediary)

Content

Content Viewers
User-Generated Content Platforms

Problem:
1. To pay or to charge content producers for maximizing the platform’s profit?
2. What’s the payment rate?

Characteristics of UGC platforms
• Intra-group negative externalities
• Content substitution and content viewers’ “love for variety”
• Content producer heterogeneity (e.g., content quality, production costs)
## Existing Research

<table>
<thead>
<tr>
<th>Research</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implicit incentive mechanism (e.g., rating) to incentivize high-quality content</td>
<td>[6]</td>
</tr>
<tr>
<td>Pricing in two-sided markets for general settings</td>
<td>[1]-[4][9][10]</td>
</tr>
<tr>
<td>Pricing in two-sided markets for specific settings (e.g., credit card, broadband)</td>
<td>[5][7][8][11]</td>
</tr>
</tbody>
</table>

## Existing Research

### Research

<table>
<thead>
<tr>
<th>Implicit incentive mechanism</th>
<th>Neglect the power of explicit mechanism (e.g., pricing)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pricing in two-sided markets</td>
<td>1. Very few consider intra-group negative externalities</td>
</tr>
<tr>
<td></td>
<td>2. Neglect the content substitution and content</td>
</tr>
<tr>
<td></td>
<td>viewers’ “love for variety”</td>
</tr>
</tbody>
</table>

**Shaolei Ren, J. Park, and M. van der Schaar, “Maximizing Profit on User-Generated Content Platforms with Heterogeneous Participants” IEEE Infocom 2012 (acceptance ratio: 18%).**

---

Three-Stage Game

- Online UGC platform modeled as a three-stage game played in the following order:

  **Platform**
  - Set payment rate

  **Content Producers**
  - Decide whether or not to produce content

  **Content Viewers**
  - Decide which content to view
Platform

- **Profit per content view:** $b$
  - Advertising revenue minus operational cost
  - Exogenously given and constant (may be negative)

- **Payment rate:** $\theta$
  - Paid to content producers per content view
  - Negative $\theta \rightarrow$ charge content producers

- **Total content views:** $\bar{x}(\theta)$
  - Determined by users (i.e. content producers and viewers), given the payment rate

Platform’s profit: $\Pi(\theta) = (b - \theta) \cdot \bar{x}(\theta)$
Content Producers

- **Continuum model**
  - Total mass of potential content producers normalized to one
  - Content producers indexed by $i$

- **Content producer $i$**
  - Content quality $q_i \geq 0$
  - Content production cost $c > 0$
  - Binary decision $y_i \in \{0, 1\}$

Payoff: $\pi_i(\theta, y) = (\theta + s) \cdot x_i(y) - c$, if $y_i = 1$
Content Viewers

- Representative agent model
  - All content viewers are consolidated as a representative content viewer

- Decision-making model
  - Total content views $T > 0$
  - Outside activities
    - Aggregate content quality $q_a$
  - Decisions $x = \left( (x_i)_{i \in [0,1]}, x_a \right) \in R^{[0,1]}_+ \times R_+$

$$\max_x U(x)$$

Optimally allocate attention to maximize payoff
Equilibrium

- At the equilibrium
  - Platform’s profit is maximized
  - Content producers’ production decisions do not change
  - Representative content viewer’s payoff is maximized

- Definition of equilibrium

**Definition:** \((\theta^*, y^*(\theta^*), x^*(\theta^*, y^*))\) is an equilibrium if

(i) \(x^*(\theta^*, y^*)\) maximizes the representative content viewer’s payoff

(ii) For each content producer \(i\), \(y_i^*(\theta^*)\) is the optimal production decision

(iii) \(\theta^*\) maximizes the platform’s profit, i.e.,

\[
(b - \theta^*) \cdot \bar{x}(\theta^*) \geq (b - \theta) \cdot \bar{x}(\theta) \quad \text{for} \ \theta \in R
\]
Three-Stage Game

Platform
- Set payment rate

Content Producers
- Decide whether or not to produce content

Content Viewers
- Decide which content to view
Optimal Content Viewing

Decision making rule

\[
\max_x U(x) \quad \text{s.t.} \quad \int_0^1 x_i \, di + x_a \leq T
\]

Example payoff function (quality-augmented Dixit-Stiglitz utility function)

\[
U(x) = \left[ \int_0^1 q_i x_i \frac{\sigma-1}{\sigma} \, di + q_a x_a \frac{\sigma-1}{\sigma} \right]^{\frac{\sigma}{\sigma-1}}
\]

where \( \sigma > 1 \) measures elasticity of substitution between different pieces of content.

- **Optimal** \( x^*(y) \)

\[
x_i^*(y) = \frac{T q_i^{\sigma}}{q_a^{\sigma} + \int_0^1 y_j q_j^{\sigma} \, dj} \quad y_i \quad \text{for} \quad i \in [0,1]
\]

1. **Intra-group (negative) externalities**: more content available decreases the payoff of an individual content producer.
2. **“Love for variety”**: more diversified content makes the content viewers better off.
Three-Stage Game

Platform
- Set payment rate

Content Producers
- Decide whether or not to produce content

Content Viewers
- Decide which content to view
Lemma 1: Let $y^*(\theta)$ be an equilibrium strategy profile of content producers. If $y_i^*(\theta) = 1$ and $y_j \geq y_i$, then $y_j^*(\theta) = 1$. 

Non-cooperative game

- $y^*(\theta)$ is the (Nash) equilibrium of the subgame played by content producers

Intuitions

- Higher content quality yields a higher payoff for content producers
Content Production Subgame

- **Non-cooperative game**
  - $y^*(\theta)$ is the (Nash) equilibrium of the subgame played by content producers

- **Intuitions**
  - Higher content quality yields a higher payoff for content producers

- **Marginal content quality $q_m$**
  - Content quality threshold below which content producers will not produce content

**Proposition 1:** There exists a unique equilibrium in the content production subgame given any payment rate $\theta$.

Deterministic outcome given the platform’s payment rate.
Three-Stage Game

Platform
- Set payment rate

Content Producers
- Decide whether or not to produce content

Content Viewers
- Decide which content to view
Optimal Payment Rate

- **Profit-maximizing payment rate**
  \[ \max_{\theta} (b - \theta) \cdot \bar{x}(\theta) \]

- **Consider quality-augmented Dixit-Stiglitz utility function**

**Theorem 1:**
There exists a unique optimal payment rate \( \theta^* \) maximizing the platform's equilibrium profit. The equilibrium marginal content quality given \( \theta^* \), denoted by \( q_m^{**} \), is the unique root of the following equation in the variable \( q_m \):

\[ - \frac{T(b+s)q_a^\sigma}{[(\sigma+1)q_a^\sigma+1-q_m^{\sigma+1}]} + \frac{c(\sigma+q_m^{\sigma+1})}{(\sigma+1)^3 q_m^{2\sigma+1}} = 0, \]

and \( \theta^* \) is given by

\[ \theta^* = \frac{c[(\sigma+1)q_a^\sigma+1-(q_m^{**})^{\sigma+1}]}{T(\sigma+1)(q_m^{**})^\sigma} - S. \]

Solve the optimal marginal content quality

Apply “marginal user principle” to derive the optimal payment rate
To Pay or To Charge?

- Structural property of the optimal payment rate

**Theorem 2:**
The optimal payment rate $\theta^*$ is positive if and only if

$$\frac{\sigma}{q_m(0)} + [q_m^*(0)]^\sigma < \frac{c(b+s)(\sigma+1)q_\sigma}{Ts^2}.$$ 

- Insights
  - As $b$ increases, the platform has a stronger incentive to increase content production by paying the content producers.
  - As $c$ increases, the content producers incur a larger production cost, and the platform should pay content producers for content production.
  - As $s$ increases, the content producers obtain larger surplus by attracting content views, and the platform extracts their surplus by charging them.
Asymptotic Cases

$\bullet$ $q_\alpha \to 0$

- The platform is virtually a monopolist, and all the content views go to the content on the platform
- Optimal payment rate $\theta^* \to -s$, and the maximum profit $\Pi(\theta^*) \to (b + s)T$

$\bullet$ $\sigma \to \infty$

- Perfectly substitutable content
- $q_\alpha < 1$: optimal payment rate $\theta^* \to -s$, and the maximum profit $\Pi(\theta^*) \to (b + s)T$

Fully extract the content producers’ surplus!
Impacts of Outside Content

Setting

\[ s = 0.4, \ c = 1.0, \ b = 1.0 \]

\[ \sigma = 2.0 \]

Profit increase

10% profit increase!
Impacts of Production Cost

Setting

\[ s=0.4, \quad b=1.0, \quad q_a = 1.5 \]
\[ \sigma = 2.0 \]

Profit increase

Double the profit!
Heterogeneous Production Costs

- **K production costs**
  - $c_1, c_2, \cdots, c_K$
  - The mass of type-$k$ content producers is $n_k \geq 0$ such that $\sum_k n_k = 1$
  - Equilibrium marginal content quality vector
    \[ q_m = (q_{m,1}, q_{m,2}, \cdots, q_{m,K}) \]

- **Main results**
  - Uniqueness of equilibrium marginal content quality is proved
  - Develop an iterative algorithm to derive the optimal payment rate $\theta^*$ based on marginal user principle (details omitted)
Summary

- **Proposed a profit-maximizing payment scheme**
  - Derive the optimal payment rate and show when it is optimal to charge or reward

- **Insight**
  - **Charging** content producers may also maximize the platform’s profit
    - Strong social incentives
    - Low content production costs
Outline

Pricing
### Existing Solutions

<table>
<thead>
<tr>
<th>Solution</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heuristic algorithms and trace-based simulations to show cost saving by scheduling workloads among multiple data centers</td>
<td>[1]</td>
</tr>
<tr>
<td>Dynamic sizing</td>
<td>[2]-[4]</td>
</tr>
<tr>
<td>Instantaneous and static optimization to minimize energy/delay cost</td>
<td>[5][6]</td>
</tr>
<tr>
<td>Online algorithm in a stochastic environment to explore electricity price diversity</td>
<td>[7][8]</td>
</tr>
</tbody>
</table>

#### Challenges
- Environment may be arbitrarily random
- Long-term performance is important (e.g., profit, delay)
- Exploit the benefits of demand-side management

---

## Existing Solutions

<table>
<thead>
<tr>
<th>Solution</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heuristic algorithms and trace-based simulations to show cost saving by scheduling workloads among multiple data centers</td>
<td>[1]</td>
</tr>
<tr>
<td>Dynamic sizing</td>
<td>[2]-[4]</td>
</tr>
<tr>
<td>Instantaneous and static optimization</td>
<td>[5][6]</td>
</tr>
<tr>
<td>Online algorithm in a stochastic environment</td>
<td>[7][8]</td>
</tr>
</tbody>
</table>

### Challenges

- Environment may be arbitrarily random
- Long-term performance is important (e.g., profit, delay)
- Exploit the benefits of demand-side management

### Solution Ref.

- **No analytic performance guarantees**
  - 1. Myopic optimization without foresightness
  - 2. No long-term performance guarantee

- Not applicable for practical environment which is neither i.i.d. nor Markovian

---


---

Control Decisions

- **Service provider**
  - \( p(t) \): price for batch services
  - \( d(t) \): number of jobs processed in the data center
Control Decisions

- **Service provider**
  - $p(t)$: price for batch services
  - $d(t)$: number of jobs processed in the data center

Queue dynamics:
$$q(t + 1) = \max[q(t) - d(t), 0] + b(t)$$

Indirectly controlled by $p(t)$
Problem Formulation

Offline problem formulation

- Decisions $z(t)$ are made at the beginning of every time slot
  - Price for batch services and # of jobs processed in the data center

$\max \quad h(z)$

s.t.,

$$\bar{b} \leq d$$

$$d(t) \leq W(t)$$

where $\bar{b}$ is the average service demand, $d$ is the average number of processed jobs, and $W(t)$ is the number of available servers for batch services
Problem Formulation

- **Offline problem formulation**
  - Decisions \( z(t) \) are made at the beginning of every time slot
    - Price for batch services and # of jobs processed in the data center

\[
\max_{z(t), t=0,1,\ldots,t_{\text{end}-1}} h(z)
\]

s.t.,
\[
\bar{b} \leq d(t) \leq W(t)
\]

where \( \bar{b} \) is the average service demand, \( d \) is the average number of processed jobs, and \( W(t) \) is the number of available servers for batch services

Future information required!

Average profit

All arrival jobs need to be processed

Computing resource constraint
Online Algorithm

- **Intuitions**
  - Use prices to regulate the service demand
    - To avoid excessive delays, set higher prices to reduce the demand when the queue is longer
  - Opportunistically utilize low electricity prices

![Price vs Time Graph](image-url)

- Process more jobs!
Online Algorithm — Dyn-SP

- **Step 1: Observe current information**
  - Electricity price, renewable energy supply, interactive service demand, queue length

- **Step 2: Determine the price**
  - Choose \( p(t) \) to minimize \( b(p(t)) \cdot [q(t) - V \cdot p(t)] \)

- **Step 3: Schedule the jobs**
  - Choose \( d(t) \) to minimize
    \[
    V \cdot \phi(t) \cdot \left[ d(t) + f(d(t)) - [y(t) - a(t) - f(a(t))]^+ \right]^+ - q(t)d(t)
    \]

- **Step 4: Update job queue**
  - \( q(t + 1) = \max[q(t) - d(t), 0] + b(t) \)
Role of $V$

- **Impact of $V$ on scheduling decision**
  - “Step 3” solves a linear programming problem
  - Grid power is used to process batch jobs only when
    \[ \phi(t) \leq \frac{q(t)}{V(1 + \gamma)} \]
    where $\gamma$ is the cooling system power consumption for one unit of servers

- **Insight**
  - Large $V$: jobs are processed only when electricity price is sufficiently low
    - Low energy cost but may increase the queue length (and hence, also delay)
  - Small $V$: jobs are processed more frequently
    - Small delay but may lose the chance of opportunistically utilizing low electricity prices
Role of $V$

Impact of $V$ on scheduling decision

- “Step 3” solves a linear programming problem
- Grid power is used to process batch jobs only when

$$\phi_t \leq q(t) V(1 + \gamma)$$

where $\gamma$ is the cooling system power consumption for one unit of servers

Insight

- Large $V$: jobs are processed only when electricity price is sufficiently low
  - Low energy cost but may increase the queue length (and hence, also delay)
- Small $V$: jobs are processed more frequently
  - Small delay but may lose the chance of opportunistically utilizing low electricity prices

Dynamic threshold determined by queue length

How “good” is the algorithm?

1. Benchmark: offline algorithm with future information
2. Performance bound
Theorem: Suppose that some mild boundedness conditions (details in the dissertation) are satisfied, then

a. At any time, the queue length is upper bounded
   \[ q(t) \leq V \cdot C \]

b. The average profit achieved by Dyn-SP satisfies
   \[ H_r^*(T) - \bar{h}^* \leq \frac{B + D(T - 1)}{V} \]

where \( B, C, D \) are certain constants and \( V \) is the control parameter.

Key insights

- Queue length bounded by \( V \cdot C \)
  - Average delay performance is closely related to queue length

- For arbitrarily random environment, the average profit is within \( \frac{B + D(T - 1)}{V} \) of the optimal offline algorithm with future \( T \)-slot information.
Performance Evaluation

Remark:
1. Trade queueing delay for energy cost saving and profit increase
2. Tune $V$ to get desired performance
Performance Evaluation

Pricing can effectively reshape the demand: significant profit increase compared to existing solutions

- Applicable for arbitrarily random environment
- Long-term performance guarantees

Setting: $V = 10$

Electrical Engineering
Other Projects

- Optimal pricing plan and capacity investment in wireless markets
  - Users have heterogeneous valuations of QoS and data service demand
  - Unlimited plan versus capped data plan
  - Monopoly and duopoly

- Real-time multimedia stream mining in mobile clouds
  - Migrate data-intensive and computation-intensive tasks from mobile devices to cloud
  - Minimize energy consumption at the cloud subject to stream mining performance requirement
What’s Next?

Future Work
Future Work

- Robust Pricing
  - Pricing with inaccurate demand function
  - Price-anticipation users
- Non-uniform pricing
- Social-welfare maximization
Cyber-Physical Systems

- Sustainability
- Privacy
- Security
- Reliability
Acknowledgement

- **UCLA**
  - Prof. Mihaela van der Schaar
  - Prof. William Zame
  - Prof. Ali H. Sayed
  - Prof. Jason Speyer

- **SUNY Buffalo**
  - Prof. Nicholas Mastronarde

- **Microsoft Research**
  - Dr. Yuxiong He, Dr. Sameh Elnikety, Dr. Phil Chou

- **Intel**
  - Dr. Fangwen Fu

- **Others**
  - Prof. Jaeok Park, Prof. Pingyi Fan, Prof. Khaled Ben Letaief, etc.
Selected Publications

Journal


Conference

Thanks!

E-mail: rsl@ucla.edu
http://www.ee.ucla.edu/~rsl/