Congestion Equilibrium for Differentiated Service Classes

Richard T. B. Ma
School of Computing
National University of Singapore

Allerton Conference 2011
Outline

- Characterize Congestion Equilibrium
- Modeling Differentiated Service Classes
- Solve Congestion Equilibrium
- Applications
Competitive Market Equilibrium

- In a competitive economy, buyers decide how much to buy and producers decide how much to produce.

- A market competitive equilibrium is characterized by price:
  - Higher prices induce lower demand/consumption
  - Higher prices induce higher supply/production

- Prices can be thought of indicators of congestion in system ➔ a congestion equilibrium generalization.
An Internet Ecosystem Model

- Three parties system $(M, \mu, \mathcal{N})$: 1) Content Providers (CPs), 2) ISPs, and 3) Consumers.

- $\mu$: capacity of a bottleneck ISP.
- $\lambda_i$: throughput rate of CP $i \in \mathcal{N}$.
- $M$: number of end customers using the ISP.
What drives traffic demands?

- User drives traffic rates from the CPs.
- User demand depends on the level of system congestion denoted as $\Gamma$.

- Given a fixed congestion $\Gamma$, we characterize
  $$\lambda_i(M, \mu, N) = \lambda_i(\Gamma) = \alpha_i M \rho_i(\Gamma)$$

- Assumption 1: $\rho_i(\cdot)$ is non-negative, continuous and non-increasing on $[0, \tilde{\theta}_i]$ with
  $$\rho_i(0) = \tilde{\theta}_i \text{ and } \lim_{\Gamma \to \infty} \rho_i(\Gamma) = 0.$$
Unconstrained Demand $\hat{\theta}_i$

- **Google Search**
  - Search Page 20 KB
  - Search Time .25 sec
  - Unconstrained demand 600 KBps

- **Netflix**
  - HD quality Stream
  - Unconstrained demand 6 MBps
Interpretation of $\rho_i(\cdot)$

$$\lambda_i(\Gamma) = \alpha_i M \rho_i(\Gamma)$$

- $\alpha_i$ is the % of users that are interested in content of CP $i$.

- $\rho_i(\Gamma)$ can be interpreted as the per-user achievable throughput rate, which can be written as
  
  $$\rho_i(\Gamma) = d_i(\Gamma) \theta_i(\Gamma),$$

  where $\theta_i(\Gamma) \in [0, \hat{\theta}_i]$, is the throughput of an active user and $d_i(\Gamma) \in [0,1]$ is the % of users that are active under $\Gamma$. 
What affects congestion $\Gamma$?

- Let $\Lambda = (\lambda_1, \dots, \lambda_N)$ be the rates of the CPs.
- $\Gamma$ of system $(M, \mu, \mathcal{N})$ is characterized by
  - Throughput rates $\Lambda$ and system capacity $\mu$
  - Higher throughput induces severer congestion
  - Larger capacity relieves congestion

- Assumption 2: For any $\mu_1 \leq \mu_2$ and $\Lambda_1 \leq \Lambda_2$, $\Gamma(\cdot)$ is a continuous function that satisfies $\Gamma(\Lambda, \mu_1) \geq \Gamma(\Lambda, \mu_2)$ and $\Gamma(\Lambda_1, \mu) \leq \Gamma(\Lambda_2, \mu)$. 
Unique Congestion Equilibrium

Definition: A pair \( (\Lambda, \Gamma) \) is a congestion equilibrium of the system \( (M, \mu, \mathcal{N}) \) if
\[
\lambda_i(M, \mu, \mathcal{N}) = \alpha_i M \rho_i(\Gamma) \quad \forall i \in \mathcal{N} \text{ and } \Gamma = \Gamma(\Lambda, \mu)
\]

Theorem 1: Under assumption 1 and 2, system \( (M, \mu, \mathcal{N}) \) has a unique congestion equilibrium.

Intuition:
- A1: decreasing monotonicity of demand
- A2: increasing monotonicity of congestion
  - System balances at a unique level of congestion
Further Characterization

- Assumption 3 (Independent of Scale): \( \Gamma(\Lambda, \mu) = \Gamma(\xi \Lambda, \xi \mu) \ \forall \xi > 0 \).

- Theorem 2: Under assumption 1 to 3, if \((\Lambda, \mu)\) is the unique equilibrium of \((M, \mu, N)\), then for any \(\xi > 0\), \((\xi \Lambda, \mu)\) is the unique equilibrium of \((\xi M, \xi \mu, N)\).

- Equilibrium \((\Lambda, \Gamma)\) can be expressed as a function of the per capita capacity \(\nu \equiv \frac{\mu}{M}\).
Equilibrium as a Function of $\nu$

- Congestion in equilibrium $\Gamma_\mathcal{N}(M, \mu) \overset{\text{def}}{=} \Gamma(M, \mu, \mathcal{N})$ is a homogenous function of degree 0, i.e.
  \[ \Gamma_\mathcal{N}(\nu) = \Gamma_\mathcal{N}(\xi M, \xi \mu) \quad \forall \xi > 0. \]

- $\Gamma_\mathcal{N}(\nu)$ is a continuous non-increasing function of $\nu$ that satisfies
  \[ \Gamma_\mathcal{N}_1(\nu) \leq \Gamma_\mathcal{N}_2(\nu) \quad \forall \mathcal{N}_1 \subseteq \mathcal{N}_2. \]

- Rates in equilibrium $\Lambda_\mathcal{N}(M, \mu) \overset{\text{def}}{=} \Lambda(M, \mu, \mathcal{N})$ is a homogenous function of degree $-1$, i.e.
  \[ \Lambda_\mathcal{N}(M, \mu) = \xi^{-1} \Lambda_\mathcal{N}(\xi M, \xi \mu) \quad \forall \xi > 0. \]
Interpretations of Congestion

- The concept of congestion is very broad
  - depends on the system resource mechanism
  - can be functions of delay, throughput and etc.

1. System mechanism: $M/M/1$, FIFO queue; Congestion metric: queueing delay;

$$\Gamma(\Lambda, \mu) = \Gamma_N = \frac{1}{\mu - \lambda_N}$$
Interpretations of Congestion

2. System mechanism: Proportional rate control, i.e. $\theta_i: \theta_j = \hat{\theta}_i: \hat{\theta}_j$ for all $i, j \in \mathcal{N}$; Congestion metric: throughput ratio;
$$\Gamma(\Lambda, \mu) = \Gamma_{\mathcal{N}} = \frac{\hat{\theta}_i}{\theta_i} - 1 \forall i \in \mathcal{N}$$

3. System mechanism: End-to-end congestion control, e.g. max-min fair mechanism; Congestion metric: function of throughput;
$$\Gamma(\Lambda, \mu) = \Gamma_{\mathcal{N}} = \frac{1}{\max\{\theta_i: i \in \mathcal{N}\}}$$
### PMP-like Differentiations

- $\kappa$ percentage of capacity dedicated to premium content providers
- $c$ per unit traffic charge for premium content

<table>
<thead>
<tr>
<th>Class</th>
<th>Capacity</th>
<th>Charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Premium Class</td>
<td>$\kappa \mu$</td>
<td>$c $/unit traffic</td>
</tr>
<tr>
<td>Ordinary Class</td>
<td>$(1 - \kappa) \mu$</td>
<td>$0$</td>
</tr>
</tbody>
</table>
Two-stage Game \((\mathcal{M}, \mu, \mathcal{N}, \mathcal{I})\)

- **Players:** ISP \(\mathcal{I}\) and the set of CPs \(\mathcal{N}\)
- **Strategies:** ISP chooses a strategy \(s_{\mathcal{I}} = (k, c)\).
  CPs choose service classes with \(s_{\mathcal{N}} = (\mathcal{O}, \mathcal{P})\).
- **Rules:** 1\textsuperscript{st} stage, ISP announces \(s_{\mathcal{I}}\). 2\textsuperscript{nd} stage, CPs simultaneously reach a joint decision \(s_{\mathcal{N}}\).
- **Outcome:** set \(\mathcal{P}\) of CPs shares capacity \(\kappa \mu\)
  and set \(\mathcal{O}\) of CPs share capacity \((1- \kappa)\mu\).
Payoffs (Surplus)

- **Content Provider Payoff:**
  \[ u_i(\lambda_i) = \begin{cases} 
  v_i \lambda_i & \text{if } i \in \Omega, \\
  (v_i - c) \lambda_i & \text{if } i \in \mathcal{P}.
\end{cases} \]

- **ISP Payoff:**
  \[ c \sum_{i \in \mathcal{P}} \lambda_i = c \lambda_{\mathcal{P}} \]

- **Consumer Surplus:**
  \[ \sum_{i \in \mathcal{N}} \phi_i \lambda_i \]
CPs’ strategy

- Choose which service class to join
- Congestion-taking assumption: Competitive congestion equilibrium in each service class

<table>
<thead>
<tr>
<th>Capacity</th>
<th>Charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1-\xi)n</td>
<td>$0</td>
</tr>
</tbody>
</table>

\[c \in [0, \infty) \] for all unit traffic
Best response, Nash equilibrium

- Lemma: Given $(\mathcal{O}, \mathcal{P})$, CP $i$'s best response to join the premium service class if
  \[(v_i - c) \rho_i(\Gamma_{\mathcal{P}\cup\{i\}}(\kappa v)) \geq v_i \rho_i(\Gamma_{\mathcal{O}\cup\{i\}}((1 - \kappa) v)).\]

- Nash equilibrium:
  \[
  \frac{v_i - c}{v_i} \begin{cases} 
  \leq \frac{\rho_i(\Gamma_{\mathcal{O}}((1 - \kappa) v))}{\rho_i(\Gamma_{\mathcal{P}\cup\{i\}}(\kappa v))} & \text{if } i \in \mathcal{O}, \\
  > \frac{\rho_i(\Gamma_{\mathcal{O}\cup\{i\}}((1 - \kappa) v))}{\rho_i(\Gamma_{\mathcal{P}}(\kappa v))} & \text{if } i \in \mathcal{P}.
  \end{cases}
  \]
Competitive equilibrium vs Nash

- Under the congestion-taking assumption:
  - Competitive equilibrium:
    \[
    \frac{v_i - c}{v_i} \begin{cases}
    \leq & \frac{\rho_i(\Gamma_\mathcal{O}((1 - \kappa)v))}{\rho_i(\Gamma_\mathcal{P}(\kappa v))} & \text{if } i \in \mathcal{O}, \\
    > & \frac{\rho_i(\Gamma_\mathcal{O}((1 - \kappa)v))}{\rho_i(\Gamma_\mathcal{P}(\kappa v))} & \text{if } i \in \mathcal{P}.
    \end{cases}
    \]

- Advantages of competitive equilibrium:
  - Does not assume “common knowledge”
  - Like the price-taking assumption, valid for large number of players (CPs)
Solving Competitive Equilibrium

- Each CP has a binary choice, state space size is $2^{|\mathcal{N}|}$, exhaustive search not feasible

- If for any $\Gamma_1$ and $\Gamma_2$, $\rho_i(\cdot)$ satisfies
  \[ \frac{\rho_i(\Gamma_1)}{\rho_i(\Gamma_2)} = F_i(G(\Gamma_1, \Gamma_2)), \]
  where $F_i$ is continuous and invertible

- Sort the CPs by $F_i^{-1}\left(\frac{v_i-c}{v_i}\right)$ and use binary search to find a competitive equilibrium
Solving Competitive Equilibrium

- A general searching method in the “congestion space”

  - Initialize at step 0, assume the congestion in service classes to be \( \Gamma[0] = \left( \Gamma_0^0, \Gamma_P^0 \right) \).
  
  - At step \( t \), take previous congestion \( \Gamma[t-1] \), calculate induced equilibrium \( (\mathcal{O}_t, \mathcal{P}_t) \).
  
  - Update the congestion level \( \Gamma[t] \) based on the previous estimate \( \Gamma[t-1] \) and the induced congestion level \( (\Gamma_0^t, \Gamma_P^t) \).
Finding competitive equilibrium

1. Initialize $\Gamma[0] = (\Gamma_0^0, \Gamma_P^0)$; $t = 0$;
2. Calculate induced equilibrium $(\mathcal{O}_0, \mathcal{P}_0)$;
3. Do
4. $\Gamma'[t] = (\Gamma_0[t], \Gamma_P[t])$;
5. $\Gamma[t + 1] = \Gamma[t] + g[t](\Gamma'[t] - \Gamma[t])$;
6. $t = t + 1$;
7. Calculate the induced equilibrium $(\mathcal{O}_t, \mathcal{P}_t)$;
8. Until $t > T$ or $(\mathcal{O}_t, \mathcal{P}_t) = (\mathcal{O}_{t-1}, \mathcal{P}_{t-1})$;
9. Return $(\mathcal{O}_t, \mathcal{P}_t)$;

Parameters: gain $g[t]$ and maximum steps $T$. 
Applications

- Congestion equilibrium serves a building block of more complicated game models
- Analyze strategic behavior of a monopolistic ISP
- Analyze strategic behavior of ISPs under oligopolistic competition
- Compare social welfare under different policy regime, e.g. Network Neutrality Vs. non neutral policies.