Pricing surplus server capacity for mean waiting time sensitive customers

A proof of conjecture

Manu K. Gupta

Institut de Recherche en Informatique de Toulouse (IRIT),
Toulouse, France.

Joint work with N. Hemachandra

Outline

- Joint pricing and scheduling problem
- Conjecture
- Outline of the proof

Motivation

- Firm can 'lease its facilities' to new customers without affecting the service level of inhouse customers.

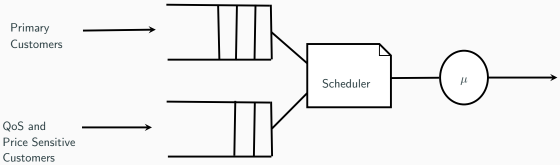


Figure 1: Model abstraction

Motivation

- Firm can 'lease its facilities' to new customers without affecting the service level of inhouse customers.
- Firms could be a container depot, a mobile service provider, a large manufacturing plant, etc.

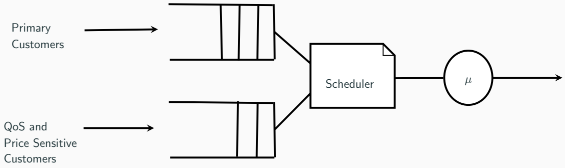
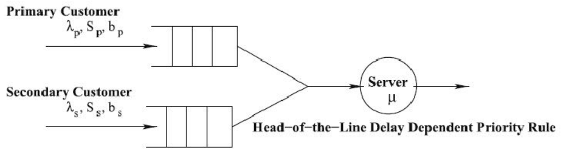


Figure 1: Model abstraction

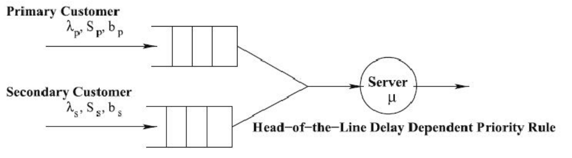
Joint pricing and scheduling problem[1]



- Model to price server's surplus capacity in M/G/1 queue.

[1] S. K. Sinha, N. Rangaraj, and N. Hemachandra. "Pricing surplus server capacity for mean waiting time sensitive customers". In: *European Journal of Operational Research* 205 (2010), pp. 159–171.

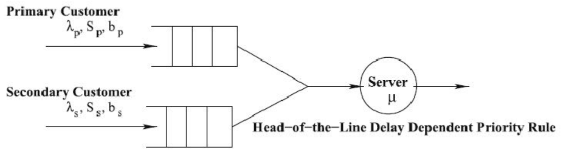
Joint pricing and scheduling problem[1]



- Model to price server's surplus capacity in M/G/1 queue.
- Surplus capacity utilized by introducing new (secondary) customers.

[1] S. K. Sinha, N. Rangaraj, and N. Hemachandra. "Pricing surplus server capacity for mean waiting time sensitive customers". In: *European Journal of Operational Research* 205 (2010), pp. 159–171.

Joint pricing and scheduling problem[1]



- Model to price server's surplus capacity in M/G/1 queue.
- Surplus capacity utilized by introducing new (secondary) customers.
- Primary are the existing customers and their mean waiting time is promised below S_p.

[1] S. K. Sinha, N. Rangaraj, and N. Hemachandra. "Pricing surplus server capacity for mean waiting time sensitive customers". In: *European Journal of Operational Research* 205 (2010), pp. 159–171.

Major Challenges

- Sensitivity of secondary class to both unit admission price and service level.

Major Challenges

- Sensitivity of secondary class to both unit admission price and service level.
- Admission control of secondary class customers.

Major Challenges

- Sensitivity of secondary class to both unit admission price and service level.
- Admission control of secondary class customers.
- Scheduling across classes.

Major Challenges

- Sensitivity of secondary class to both unit admission price and service level.
- Admission control of secondary class customers.
- Scheduling across classes.
    - Strict priority may not suffice.

Major Challenges

- Sensitivity of secondary class to both unit admission price and service level.
- Admission control of secondary class customers.
- Scheduling across classes.
    - Strict priority may not suffice.
- A dynamic priority scheduling scheme is needed.

Major Challenges

- Sensitivity of secondary class to both unit admission price and service level.
- Admission control of secondary class customers.
- Scheduling across classes.
    - Strict priority may not suffice.
- A dynamic priority scheduling scheme is needed.
    - Delay dependent priority scheme.

Major Challenges

- Sensitivity of secondary class to both unit admission price and service level.
- Admission control of secondary class customers.
- Scheduling across classes.
  - Strict priority may not suffice.
- A dynamic priority scheduling scheme is needed.
  - Delay dependent priority scheme.
- Decisions:

Major Challenges

- Sensitivity of secondary class to both unit admission price and service level.
- Admission control of secondary class customers.
- Scheduling across classes.
  - Strict priority may not suffice.
- A dynamic priority scheduling scheme is needed.
  - Delay dependent priority scheme.
- Decisions:
  - Price charged and service level offered to secondary class

Major Challenges

- Sensitivity of secondary class to both unit admission price and service level.
- Admission control of secondary class customers.
- Scheduling across classes.
    - Strict priority may not suffice.
- A dynamic priority scheduling scheme is needed.
    - Delay dependent priority scheme.
- Decisions:
    - Price charged and service level offered to secondary class
    - Scheduling across classes

Delay dependent priority scheduling scheme[2]

Prioritize jobs based on the following instantaneous priority:

$$q_p(t) = delay \times b_p$$

[2] Leonard Kleinrock. “A delay dependent queue discipline”. In: *Naval Research Logistics Quarterly* 11 (1964), pp. 329–341.

Delay dependent priority scheduling scheme[2]

Prioritize jobs based on the following instantaneous priority:

$$q_p(t) = delay \times b_p$$

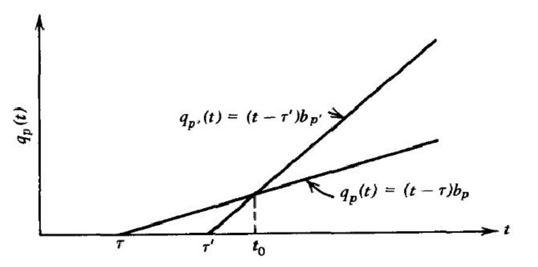


Figure 2: Illustration of delay dependent priority

[2] Leonard Kleinrock. "A delay dependent queue discipline". In: *Naval Research Logistics Quarterly* 11 (1964), pp. 329–341.

Delay dependent priority scheduling scheme[2]

Prioritize jobs based on the following instantaneous priority:

$$q_p(t) = delay \times b_p$$

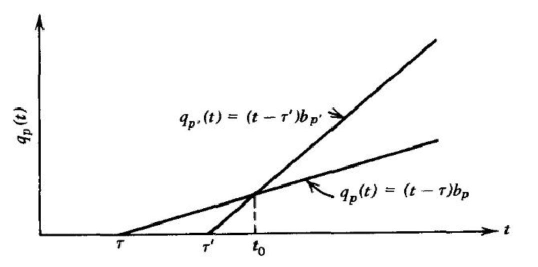


Figure 2: Illustration of delay dependent priority

Ties (at t_0) are broken according to FCFS

[2] Leonard Kleinrock. "A delay dependent queue discipline". In: *Naval Research Logistics Quarterly* 11 (1964), pp. 329–341.

Notations:

λ_p	Arrival rate for primary class customers
λ_s	Arrival rate of secondary customers
S_p	Promised mean waiting time of primary class customers
S_s	Promised mean waiting time of secondary class customers
μ	Mean service rate of server
σ^2	Variance of service time
θ	Unit admission price charged to secondary customers

Notations:

λ_p	Arrival rate for primary class customers
λ_s	Arrival rate of secondary customers
S_p	Promised mean waiting time of primary class customers
S_s	Promised mean waiting time of secondary class customers
μ	Mean service rate of server
σ^2	Variance of service time
θ	Unit admission price charged to secondary customers

Linear demand function

$$\lambda_s = a - b\theta - cS_s$$

Notations:

λ_p	Arrival rate for primary class customers
λ_s	Arrival rate of secondary customers
S_p	Promised mean waiting time of primary class customers
S_s	Promised mean waiting time of secondary class customers
μ	Mean service rate of server
σ^2	Variance of service time
θ	Unit admission price charged to secondary customers

Linear demand function

$$\lambda_s = a - b\theta - cS_s$$

- Decision variables:
  - Unit admission price (θ), service level for new class (S_s), scheduling parameter (β) and arrival rate for secondary class (λ_s)

Notations:

λ_p	Arrival rate for primary class customers
λ_s	Arrival rate of secondary customers
S_p	Promised mean waiting time of primary class customers
S_s	Promised mean waiting time of secondary class customers
μ	Mean service rate of server
σ^2	Variance of service time
θ	Unit admission price charged to secondary customers

Linear demand function

$$\lambda_s = a - b\theta - cS_s$$

- Decision variables:
  - Unit admission price (θ), service level for new class (S_s), scheduling parameter (β) and arrival rate for secondary class (λ_s)
- Objective is to maximize total revenue.

Optimization problem P0

$$\textit{Maximize } \theta\lambda_s \tag{1}$$

Subject to

Optimization problem P0

$$\textit{Maximize } \theta\lambda_s \tag{1}$$

Subject to

Primary service level: $$W_p(\lambda_s, \beta) \leq S_p \tag{2}$$

Optimization problem P0

$$\textit{Maximize } \theta\lambda_s \quad (1)$$

Subject to

$$\text{Primary service level:} \quad W_p(\lambda_s, \beta) \le S_p \quad (2)$$

$$\text{Secondary service level:} \quad W_s(\lambda_s, \beta) \le S_s \quad (3)$$

Optimization problem P0

$$\textit{Maximize } \theta\lambda_s \tag{1}$$

Subject to

$$\text{Primary service level:} \quad W_p(\lambda_s, \beta) \leq S_p \tag{2}$$

$$\text{Secondary service level:} \quad W_s(\lambda_s, \beta) \leq S_s \tag{3}$$

$$\text{Stability:} \quad \lambda_s \leq \mu - \lambda_p \tag{4}$$

Optimization problem P0

$$\textit{Maximize } \theta\lambda_s \tag{1}$$

Subject to

$$\text{Primary service level:} \quad W_p(\lambda_s, \beta) \leq S_p \tag{2}$$

$$\text{Secondary service level:} \quad W_s(\lambda_s, \beta) \leq S_s \tag{3}$$

$$\text{Stability:} \quad \lambda_s \leq \mu - \lambda_p \tag{4}$$

$$\text{Demand Constraint:} \quad \lambda_s \leq a - b\theta - cS_s \tag{5}$$

Optimization problem P0

$$Maximize\ \theta\lambda_s \tag{1}$$

Subject to

$$\text{Primary service level:} \quad W_p(\lambda_s, \beta) \leq S_p \tag{2}$$

$$\text{Secondary service level:} \quad W_s(\lambda_s, \beta) \leq S_s \tag{3}$$

$$\text{Stability:} \quad \lambda_s \leq \mu - \lambda_p \tag{4}$$

$$\text{Demand Constraint:} \quad \lambda_s \leq a - b\theta - cS_s \tag{5}$$

$$\lambda_s, \theta, S_s, \beta \geq 0 \tag{6}$$

- Constraint (3) and (5) will be binding.
- P0 is decomposed in P1 (with finite β) and P2 (with infinite β).

Optimization problem P1 and P2

$$\textbf{P1:} \max_{\lambda_s,\beta} \frac{1}{b}\left(a\lambda_s - \lambda_s^2 - c\lambda_s W_s(\lambda_s,\beta)\right) \tag{7}$$
$$W_p(\lambda_s,\beta) \le S_p \tag{8}$$
$$\lambda_s \le \mu - \lambda_p \tag{9}$$
$$\lambda_s, \beta \ge 0 \tag{10}$$

$$\textbf{P2:} \max_{\lambda_s} \frac{1}{b}[a\lambda_s - \lambda_s^2 - c\lambda_s \tilde{W}_s(\lambda_s)] \tag{11}$$
$$\tilde{W}_p(\lambda_s) \le S_p, \tag{12}$$
$$\lambda_s \le \mu - \lambda_p, \tag{13}$$
$$\lambda_s \ge 0. \tag{14}$$

- Solution by using KKT conditions

Solution of problem P0

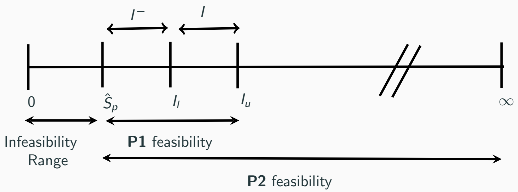
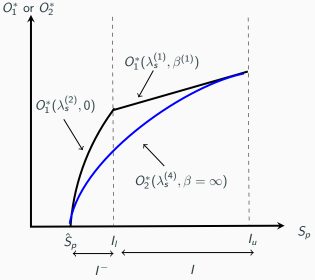


Figure 3: Range of S_p with optimal solutions coming from problem P1 and P2

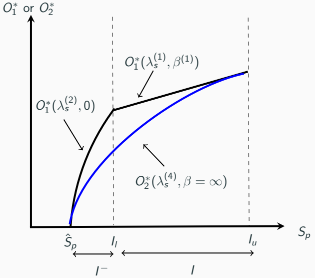
- Search for global optima
- Comparison of optimal objectives of problem P1, O_1^*, and problem P2, O_2^*, in service level range $I \cup I^-$

Comparison of objectives



[3] S. K. Sinha, N. Rangaraj, and N. Hemachandra. "Pricing surplus server capacity for mean waiting time sensitive customers". In: *European Journal of Operational Research* 205 (2010), pp. 159–171.

Comparison of objectives



Conjecture[3]**.** *For $S_p \in I^-$, the optimal solution of P0 is given by optimal solution of P1.*

[3] S. K. Sinha, N. Rangaraj, and N. Hemachandra. "Pricing surplus server capacity for mean waiting time sensitive customers". In: *European Journal of Operational Research* 205 (2010), pp. 159–171.

- A finite step algorithm[4] for optimal solution.

[4] S. K. Sinha, N. Rangaraj, and N. Hemachandra. "Pricing surplus server capacity for mean waiting time sensitive customers". In: *European Journal of Operational Research* 205 (2010), pp. 159–171.

[5] N. Hemachandra and B. S. Raghav. *On a conjecture and performance of a two class delay dependent priority queue,* tech. rep. http : //www.ieor.iitb.ac.in/files/TechRepo_COR_16042013.pdf. IIT Bombay, 2012.

- A finite step algorithm[4] for optimal solution.
    - Conjecture holds true.

[4] S. K. Sinha, N. Rangaraj, and N. Hemachandra. "Pricing surplus server capacity for mean waiting time sensitive customers". In: *European Journal of Operational Research* 205 (2010), pp. 159–171.

[5] N. Hemachandra and B. S. Raghav. *On a conjecture and performance of a two class delay dependent priority queue,* tech. rep. http : //www.ieor.iitb.ac.in/files/TechRepo_COR_16042013.pdf. IIT Bombay, 2012.

- A finite step algorithm[4] for optimal solution.
    - Conjecture holds true.
- Sufficient conditions[5] for conjecture to hold.

[4] S. K. Sinha, N. Rangaraj, and N. Hemachandra. "Pricing surplus server capacity for mean waiting time sensitive customers". In: *European Journal of Operational Research* 205 (2010), pp. 159–171.

[5] N. Hemachandra and B. S. Raghav. *On a conjecture and performance of a two class delay dependent priority queue,* tech. rep. http : //www.ieor.iitb.ac.in/files/TechRepo_COR_16042013.pdf. IIT Bombay, 2012.

Main result

Theorem

Optimal solution for optimization problem P_2 i.e. O_2^* is increasing concave in interval $I^- \cup I$ while O_1^* is increasing concave in I^- and linearly increasing in I

Steps in the proof

Main result

Theorem

Optimal solution for optimization problem P_2 i.e. O_2^* is increasing concave in interval $I^- \cup I$ while O_1^* is increasing concave in I^- and linearly increasing in I

Steps in the proof

- A claim on range of optimal solution "$I^- \subset J^-$"

Main result

Theorem

Optimal solution for optimization problem P_2 i.e. O_2^* is increasing concave in interval $I^- \cup I$ while O_1^* is increasing concave in I^- and linearly increasing in I

Steps in the proof

- A claim on range of optimal solution "$I^- \subset J^-$"
- Exploit the structure of optimal solution.

Main result

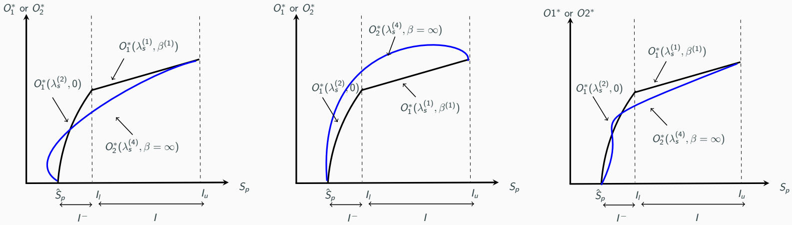
Theorem

Optimal solution for optimization problem P_2 i.e. O_2^* is increasing concave in interval $I^- \cup I$ while O_1^* is increasing concave in I^- and linearly increasing in I

Steps in the proof

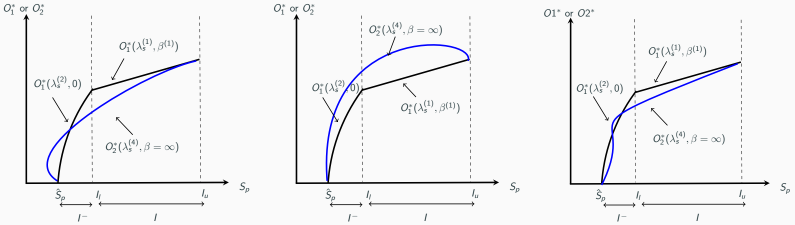
- A claim on range of optimal solution "$I^- \subset J^-$"
- Exploit the structure of optimal solution.
- $\lambda_s^{(4)}$ is an increasing function of S_p.

A proof by contradiction



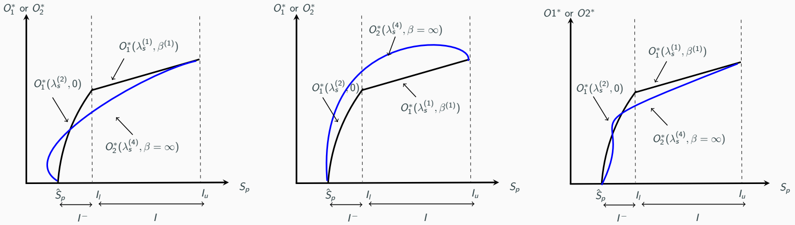
- Contradiction from infeasibility and $O_2^*(\lambda_s^i, \infty) < O_1^*(\lambda_s^f, \beta^f)$ at $\hat{S}_p + \epsilon$.

A proof by contradiction



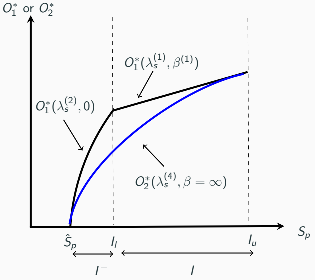
- Contradiction from infeasibility and $O_2^*(\lambda_s^i, \infty) < O_1^*(\lambda_s^f, \beta^f)$ at $\hat{S}_p + \epsilon$.
- Contradiction from $O_2^*(\lambda_s^i, \infty) < O_1^*(\lambda_s^f, \beta^f)$ for $S_p \in I$.

A proof by contradiction



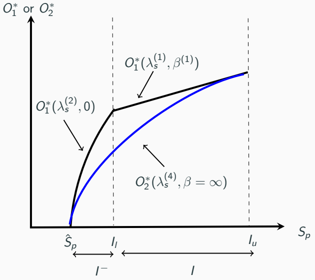
- Contradiction from infeasibility and $O_2^*(\lambda_s^i, \infty) < O_1^*(\lambda_s^f, \beta^f)$ at $\hat{S}_p + \epsilon$.
- Contradiction from $O_2^*(\lambda_s^i, \infty) < O_1^*(\lambda_s^f, \beta^f)$ for $S_p \in I$.
- Contradiction from concavity of O_2^*.

Possible way for O_1^* and O_2^*



- The proof of conjecture

Possible way for O_1^* and O_2^*



- The proof of conjecture
- Implications: Validation of finite step algorithm

Thank you!

http://manugupta-or.github.io