Optimization of On-line Appointment Scheduling

Brian Denton
Edward P. Fitts Department of Industrial and Systems Engineering
North Carolina State University

Tsinghua University, Beijing, China

May, 2012

Acknowledgements

Ayca Erdogan, School of Medicine, Stanford University

Alex Gose, NC State University

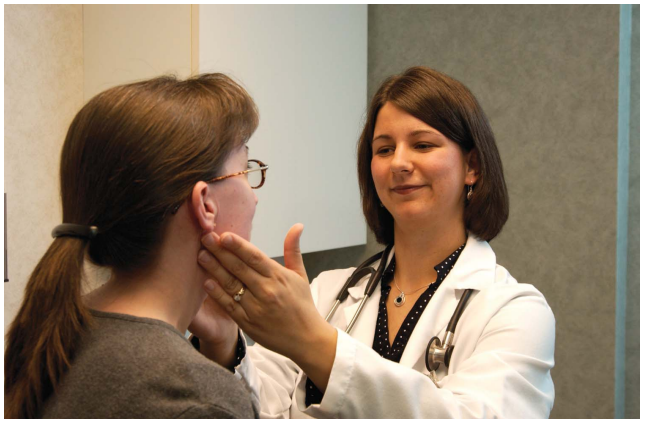
Supported by National Science Foundation: CMMI Service Enterprise Systems Grant 0620573

Appointment Scheduling Systems

- Interface between healthcare providers and patients
- Arises in many healthcare contexts
  - Primary care
  - Radiation Oncology
  - Surgery
  - Outpatient Procedures
  - Chemotherapy

Scheduling Challenges

- Competing criteria
  - Patient waiting time
  - Provider idle time and overtime
- Complicating Factors
  - Uncertain service durations
  - Uncertain patient demand
    - No-shows
    - Urgent Add-ons



Research Questions

Given a probabilistic arrival process for customer appointment requests to a single server, in which appointments must be quoted on-line:

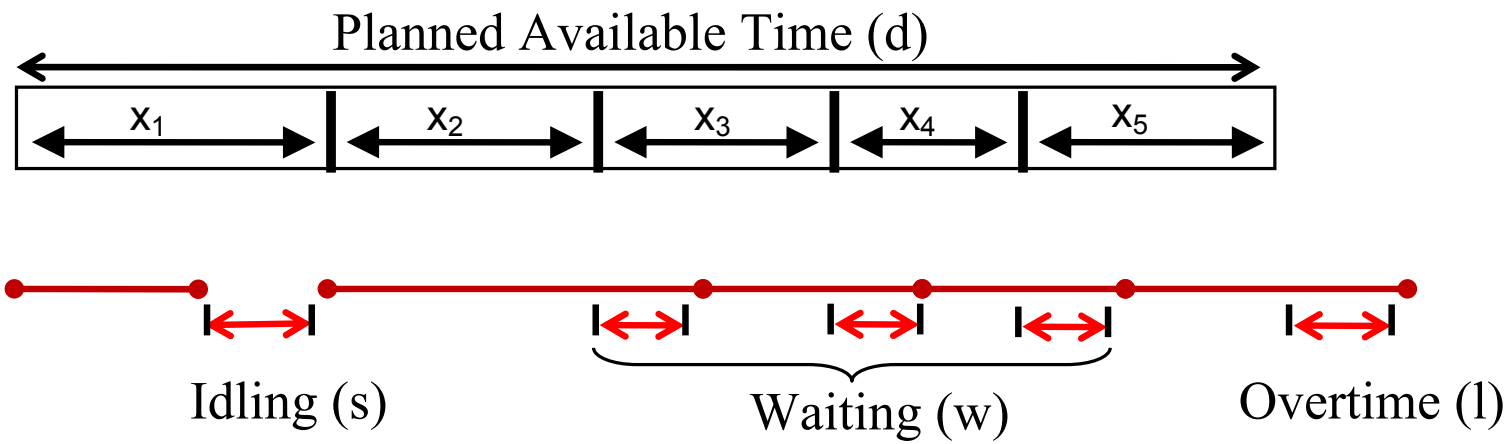
- What is the structure of the optimal appointment schedule?
- How can problems be classified into easy and hard?
- How important is it to find optimal schedules?

Presentation Outline

- Introduction
- Problems
  - Static Appointment Scheduling
  - Dynamic Appointment Scheduling
  - Dynamic Appointment Sequencing and Scheduling
- Conclusions
- Other Research

Static Appointment Scheduling Problem

Problem: Schedule n customers with uncertain service times during a fixed length of day, d



Common Heuristics

Mean Service Times:

$$\begin{aligned} a_1 &= 0 \\ a_i &= a_{i-1} + \mu_{i-1}, \quad \forall i \end{aligned}$$

Hedging:

$$\begin{aligned} a_1 &= 0 \\ a_i &= a_{i-1} + \mu_{i-1} + \kappa\sigma_{i-1}, \quad \forall i \end{aligned}$$

Ho, C., H. Lau. 1992. Minimizing Total Cost in Scheduling Outpatient Appointments, *Management Science* 38(12).

Cayirli, T., E. Veral. 2003. Outpatient Scheduling in Health Care: A Review of Literature, *Production and Operations Management* 12.

Literature Review

Queuing Analysis

- Bailey and Welch (1952)
- Jansson (1966)
- Sabria and Daganzo (1989)

Heuristics

- White and Pike (1964)
- Soriano (1966)
- Ho and Lau (1992)

Optimization

- Weiss (1990)
- Wang (1993)
- Denton and Gupta (2003)

Two-Stage Stochastic Linear Program

- First stage decisions

  x_i: Time allowance for customer i

- Second stage decisions

  $w_i(\omega)$: Customer i waiting time
  $\ell(\omega)$: Server overtime w.r.t. length of session d

- Random service durations:

  $Z_i(\omega)$: Random service time for customer i

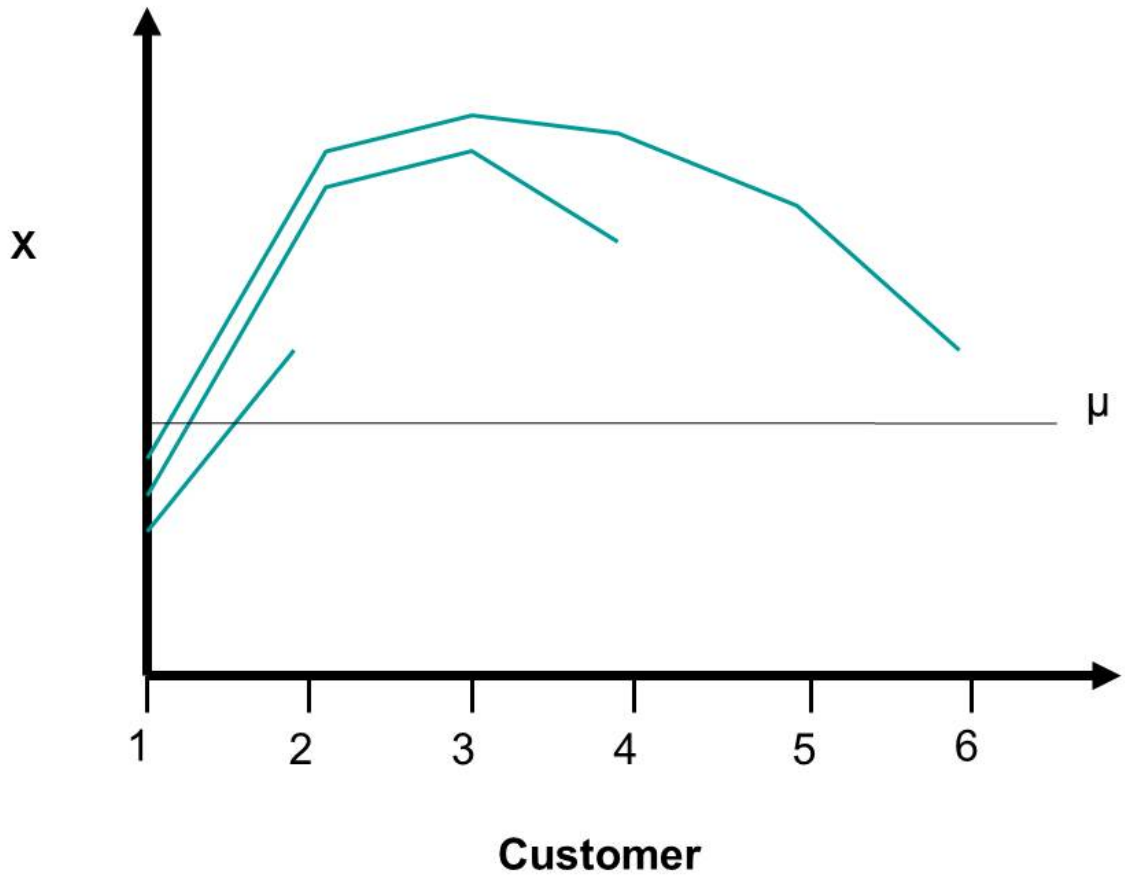
Model Formulation

$$\min E_\omega[\sum_{i=2}^{n} c_i^w w_i(\omega) + c^\ell \ell(\omega)]$$

$$\begin{aligned}
s.t. \quad & w_2(\omega) && \geq Z_1(\omega) - x_1, \forall \omega \\
& - w_2(\omega) + w_3(\omega) && \geq Z_2(\omega) - x_2, \forall \omega \\
& \ddots \quad \ddots && \vdots \\
& - w_{n-1}(\omega) + w_n(\omega) && \geq Z_{n-1}(\omega) - x_{n-1}, \forall \omega \\
& - w_n(\omega) + \ell(\omega) && \geq Z_n(\omega) + \sum_{i=1}^{n-1} x_i - d, \forall \omega
\end{aligned}$$

$$\mathbf{x} \geq \mathbf{0}, \ \mathbf{w}(\omega), \ell(\omega) \geq \mathbf{0}, \forall \omega$$

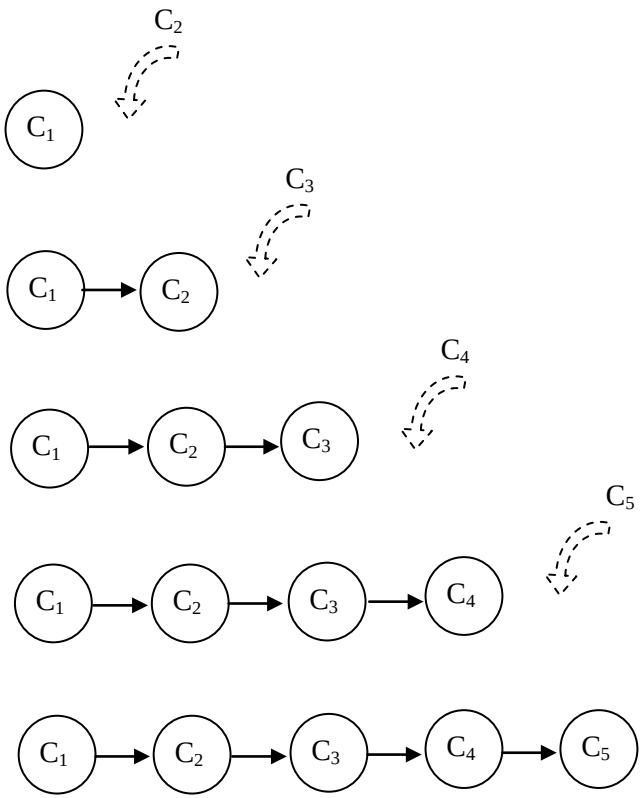
Example: 6 Customers



Denton, B.T. and Gupta D., 2003, "A Sequential Bounding Approach for Optimal Appointment Scheduling," *IIE Transactions*, 35, 1003-1016

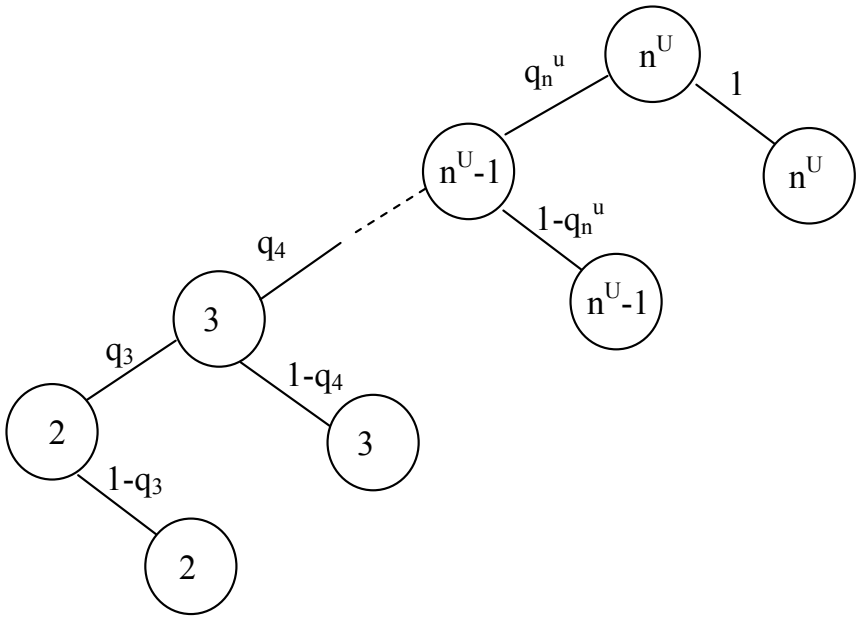
Dynamic Appointment Scheduling

Problem: Up to n^U customers are scheduled dynamically as they request appointments. Appointment requests are probabilistic.



Multi-stage Stochastic Program

Appointment requests are defined by a multi-stage scenario tree:



$$\min_{x_1}\{(1-q_3)Q_2(x_1)+\min_{x_2}\{q_3(1-q_4)Q_3(x_2)+\cdots+\min_{x_{n^U-1}}\{(\prod_{i=3}^{n^U})(q_i)Q_{n^U}(x_{n^U-1})\}\cdots\}\}$$

Model Formulation: Stage j

$$
\begin{aligned}
Q_j(x_j, \omega_j) \quad = \quad & \min_{\mathbf{w},\ell} \{ \sum_{i=2}^{j+1} c_i^w w_{j,i}(\omega_j) + c^\ell \ell_{j+1}(\omega_j) \} \\
s.t \quad & w_{j,2}(\omega_j) && \geq Z_1(\omega_j) - x_1 \\
& -w_{j,2}(\omega_j) + w_{j,3}(\omega_j) && \geq Z_2(\omega_j) - x_2 \\
& \quad \ddots \quad \ddots && \quad \vdots \\
& -w_{j,j}(\omega_j) + w_{j,j+1}(\omega_j) && \geq Z_j(\omega_j) - x_j \\
& -w_{j,j+1}(\omega_j) + \ell_{j+1}(\omega_j) && \geq Z_{j+1}(\omega_j) + \sum_{i=1}^{j} x_i - d \\
& w_{j,i}(\omega_j) \geq 0 \ \forall i, \ell_j(\omega_j) \geq 0.
\end{aligned}
$$

Model Properties

Motivation for first come first serve (FCFS) appointment sequence:

Proposition

For $n^U = 2$ with i.i.d. service durations, and identical waiting costs, the optimal sequence is FCFS.

Counter-examples exist for non i.i.d. and nonidentical waiting costs.

Solution Methods

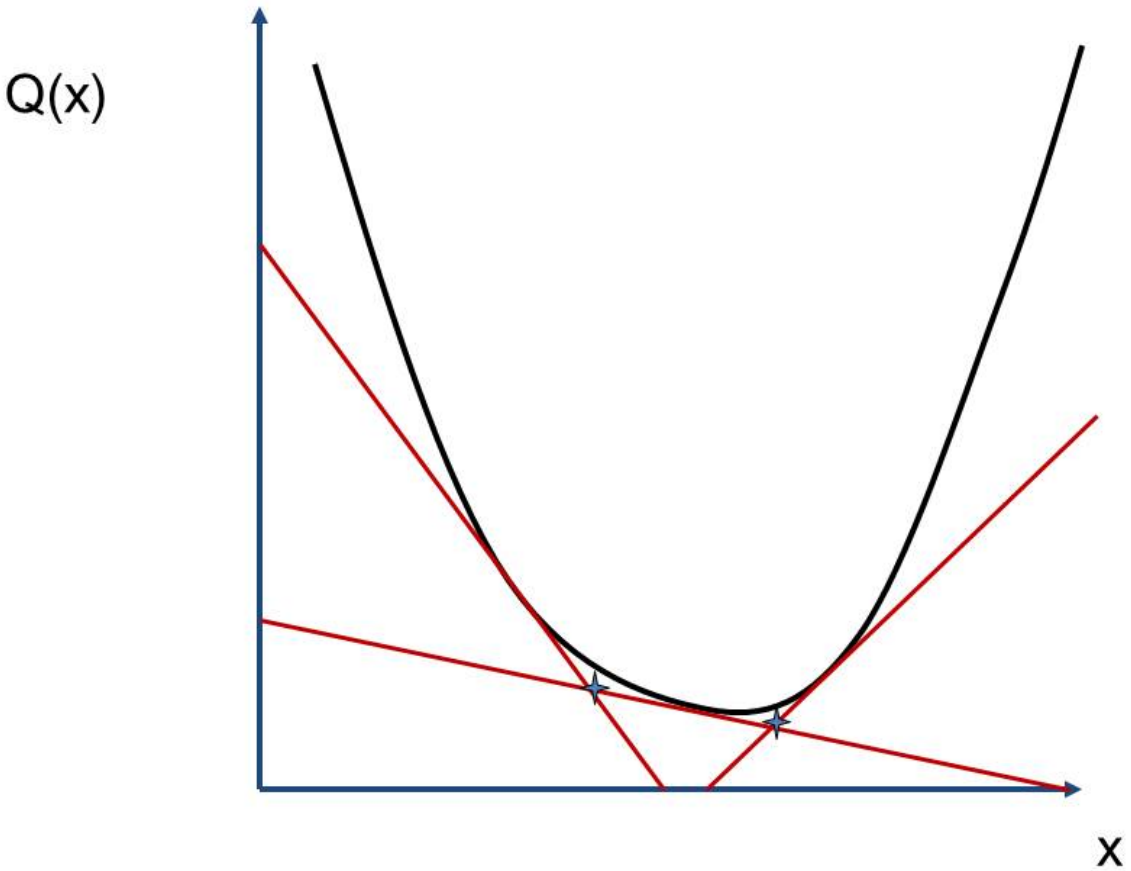
Variants of *nested decomposition*:

- Fast-forward-fast-back implementation
- Multi-cut method
- 2 variable method for master problems
- Valid inequalities based on relaxations of the mean value problem

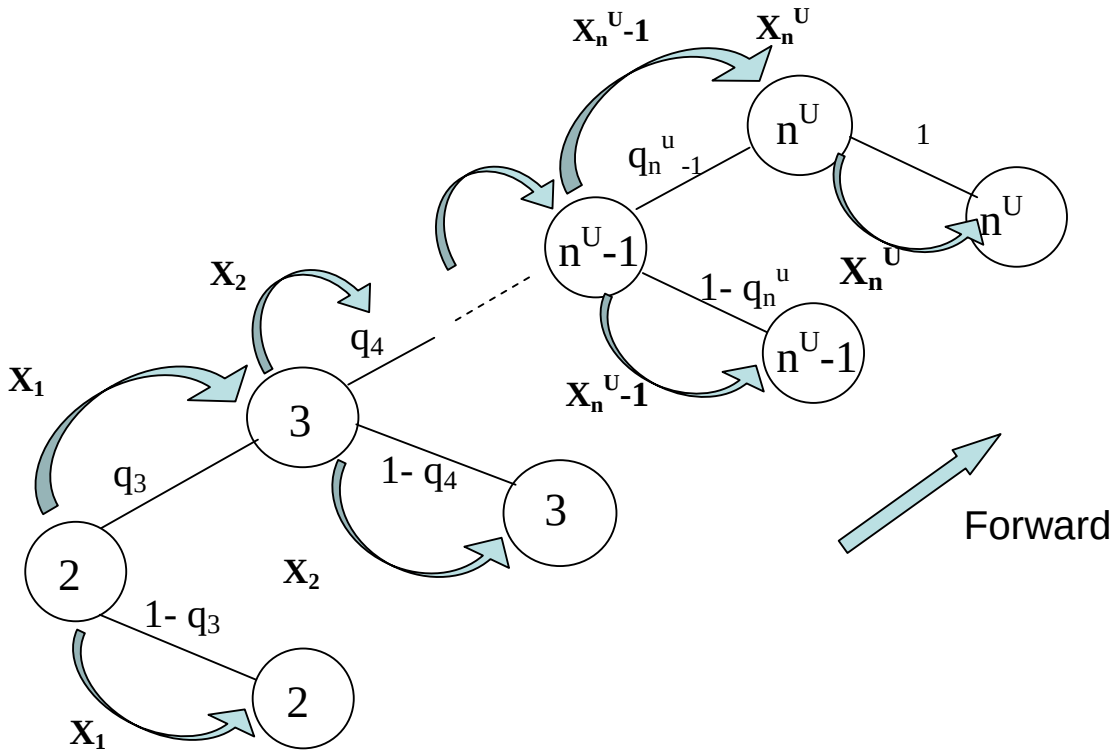
Outer Linearization

Outerlinearize the recourse function:

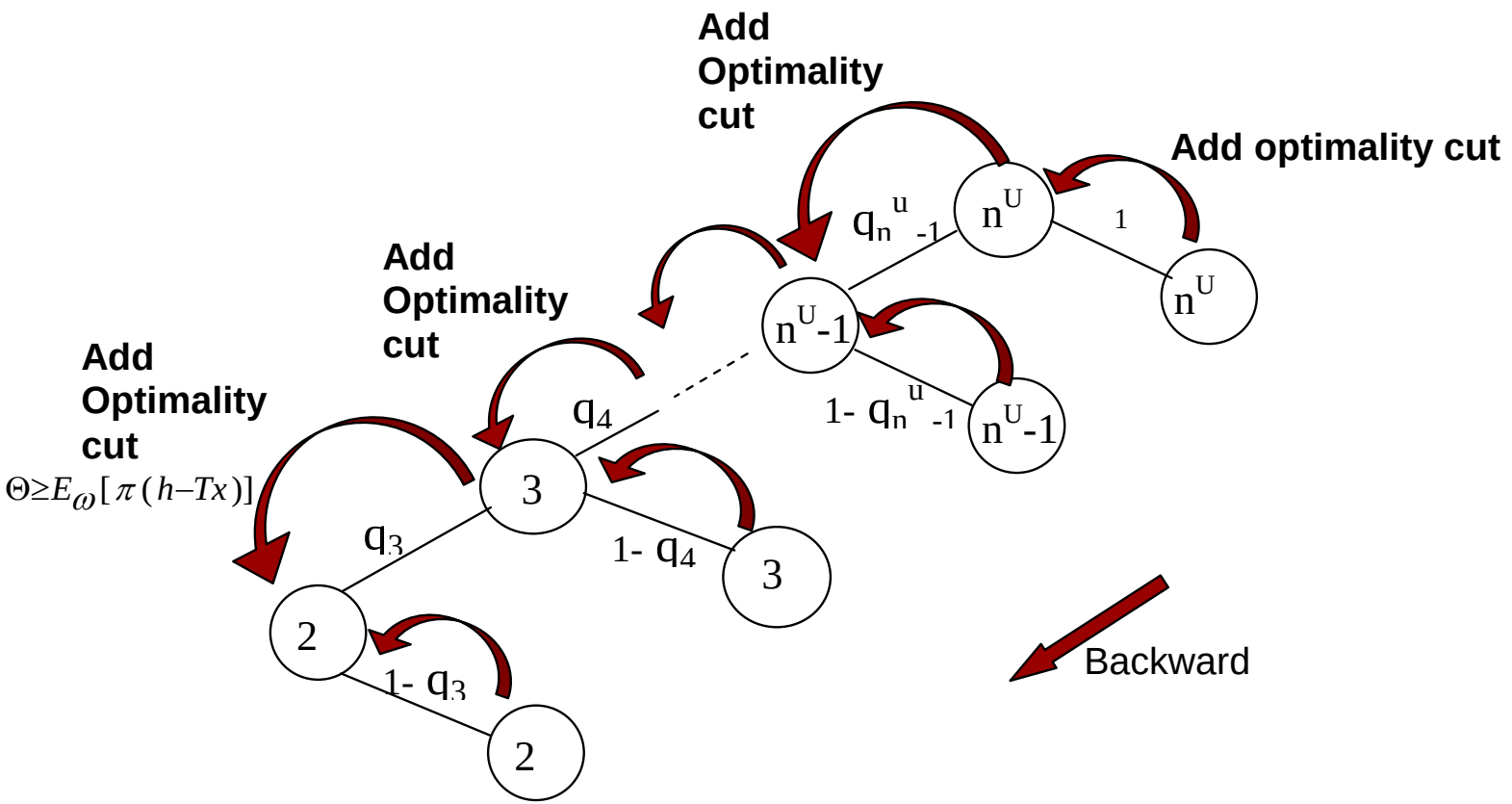
$$\min\{\theta \mid \theta \geq Q(x)\}$$



Methodology: Nested Decomposition Method

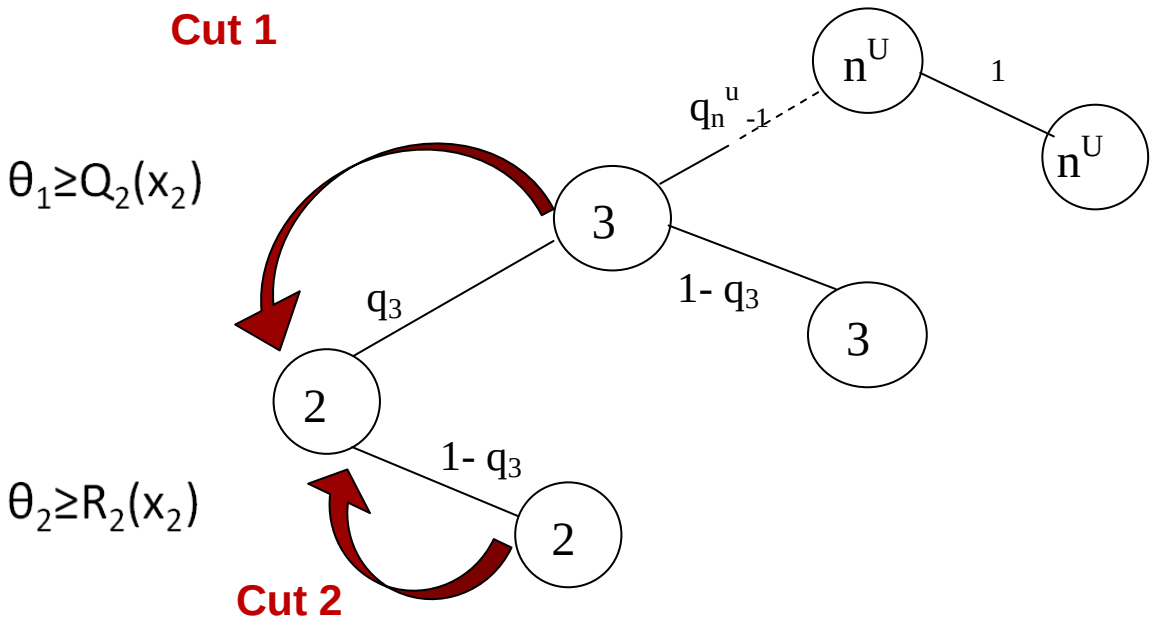


Methodology: Nested Decomposition Method



Multi-Cut Method

Separate cuts from master problems and subproblems (similar to multi-cut approach proposed by Birge and Louveaux (1985)

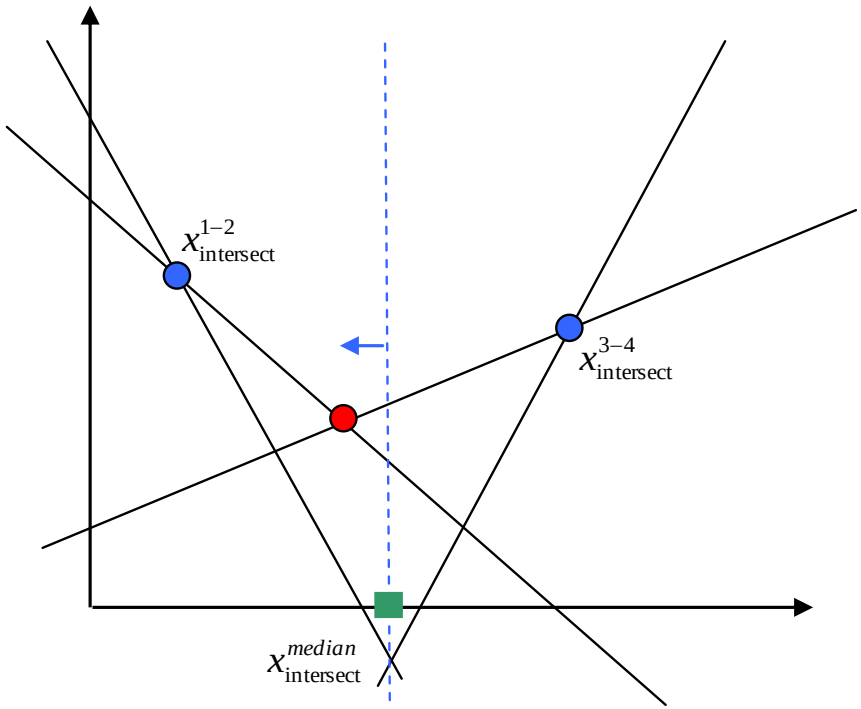


Two-variable LPs

- Master problems at each stage are two-variable LPs (x_j and θ_j)

$$\alpha_j x_j + \theta_j \geq \beta - (\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_{j-1} x_{j-1})$$

- Solve LPs with a modified version of the algorithm proposed by Dyer (1984)



Valid Inequalities

Proposition

The optimal solution to the mean value problem is $\bar{x}_i = \mu_i, \forall i$.

- Constraints based on mean value problem

$$\theta_j \geq Q_j(x, \bar{\xi})$$

- Similar to valid inequalities proposed by Batun et al. (2011)

Solution Methods

Several adaptations of *nested decomposition* were compared:

- Standard nested decomposition (ND)
- Multi-cut ND
- Two-variable ND
- ND with mean value valid inequalities (VI)

Comparisons of Methods

		Number of Iterations			CPU Time (seconds)		
		$n^U = 10$ (d=200)	$n^U = 20$ (d=400)	$n^U = 30$ (d=600)	$n^U = 10$ (d=200)	$n^U = 20$ (d=400)	$n^U = 30$ (d=600)
$\frac{c^\ell}{c^w} = 10$	ND	244	432	438	3.42	23.26	49.68
	Multi-cut ND	186	244	202	2.63	13.52	23.21
	Two-variable ND	254	406	362	3.56	24.06	43.59
	ND with VIs	232	370	442	3.65	20.83	51.79
$\frac{c^\ell}{c^w} = 1$	ND	192	330	392	2.75	16.77	42.46
	Multi-cut ND	106	184	174	1.55	9.81	19.85
	Two-variable ND	186	290	284	2.54	16.32	31.82
	ND with VIs	188	306	364	2.98	16.89	42.50
$\frac{c^\ell}{c^w} = 0.1$	ND	190	302	422	2.55	14.54	43.48
	Multi-cut ND	96	176	162	1.33	8.79	17.45
	Two-variable ND	186	290	384	2.37	15.49	42.95
	ND with VIs	174	284	412	2.62	14.86	45.70

2 QuadCore Intel® Xeon® Processor 2.50GHz CPU, 16GB Ram, CPLEX 11.0

Value of Stochastic Solution (VSS)

Table: VSS for test instances with $Z_i \sim U(20, 40)$ and $q_i = 0.5$ for add-on requests.

Number of Customers (Routine, Add-on)	VSS (%)		
	$d = 200$		
	$\frac{c^\ell}{c^w} = 10$	$\frac{c^\ell}{c^w} = 1$	$\frac{c^\ell}{c^w} = 0.1$
(0,30)	9.63	65.59	95.15
(10,30)	1.40	19.63	79.41
(20,30)	0.50	23.63	80.33

Example: Scheduling an Endoscopy Suite

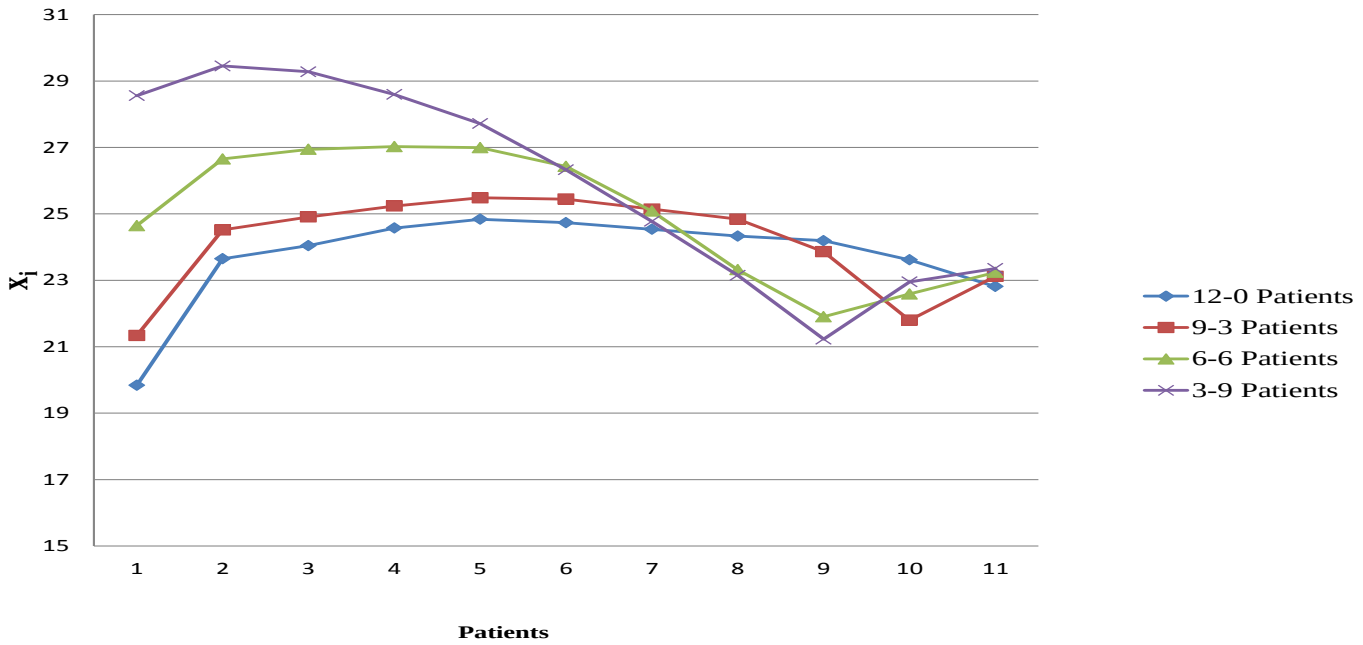


Figure: Service times based on colonoscopy times for an outpatient endoscopy practice: $Z_i \sim \text{Lognormal}(23.55, 11.89)$, $\forall i$.

Example: Multi-Procedure Room Endoscopy Practice

Endoscopy Practice:

- 2 intake rooms
- 2 procedure rooms
- 4 recovery rooms
- Service timed based on empirical data

Table: Expected waiting time and overtime according to different schedules

	Heuristic			Stochastic Program Based Schedule		
	$\frac{c^\ell}{c^w} = 10$	$\frac{c^\ell}{c^w} = 1$	$\frac{c^\ell}{c^w} = 0.1$	$\frac{c^\ell}{c^w} = 10$	$\frac{c^\ell}{c^w} = 1$	$\frac{c^\ell}{c^w} = 0.1$
Expected total cost	975.19	111.72	253.71	878.03	104.58	162.65
Expected waiting time	15.78			16.28	10.54	5.06
Expected overtime	95.94			86.17	94.05	111.97

Dynamic Appointment Sequencing and Scheduling

The **appointment request sequence** and the **appointment arrival sequence** are not necessarily the same.

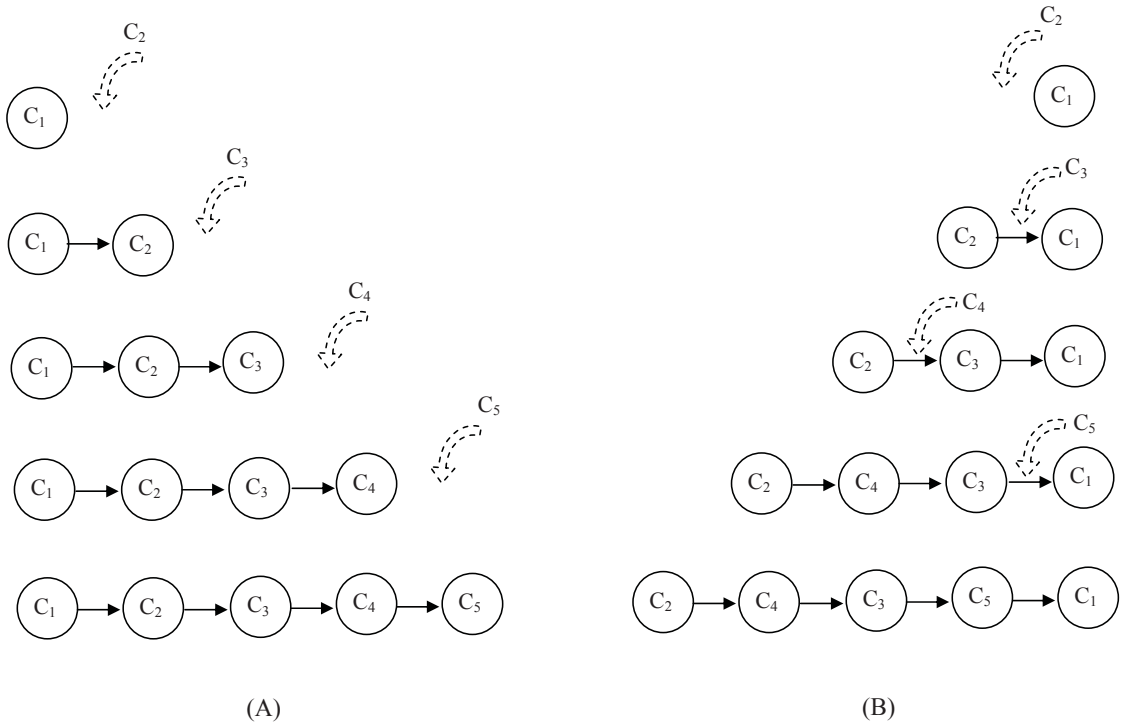


Figure: (A) FCFS; (B) Example of the general case.

Two Stage Stochastic Integer Program

Minimize {Cost of Indirect Waiting + E_ω[Direct Waiting + Overtime]}

First Stage Decisions:

- Customer sequencing (binary)
- Service time allowances (continuous, sequence dependent)
- Appointment times (continuous, sequence dependent)

Second Stage Decisions:

- Waiting time (continuous, sequence dependent)
- Overtime (continuous)

Two Stage Stochastic Integer Program

First Stage Decisions:

$o_{j,i,i'}$: binary sequencing variable where $o_{jii'} = 1$ if customer i immediately precedes i' at stage j, and $o_{ii'j} = 0$ otherwise

$x_{j,i,i'}$: time allowance for customer i given that i immediately precedes i' at stage j

$a_{j,i,i'}$: appointment time of customer i', given that i immediately precedes i' at stage j

Second Stage Decisions:

$w_{j,i,i'}(\omega)$: waiting time of customer i' given that customer i immediately precedes i' at stage j under duration scenario ω

$s_{j,i,i'}(\omega)$: server idle time between customer i and i', given that i immediately precedes i' at stage j

$\ell_j(\omega)$: overtime at stage j with respect to session length d

First Stage Problem

$$\min \sum_{j=1}^{n} p_j [\sum_{i=1}^{j} \sum_{i'=1}^{j} c^a_{i'} a_{j,i,i'}] + Q(\mathbf{o}, \mathbf{x})$$

s.t.

$$\sum_{i'=1}^{j+1} o_{j,i,i'} = 1, \quad \sum_{i'=0}^{j+1} o_{j,i',i} = 1 \qquad \forall j, i = 1, 2, \ldots, j$$

$$\sum_{i=0}^{j+1} \sum_{i=0}^{j+1} o_{j,i,i'} = j + 1 \qquad \forall j$$

$$o_{j,i,j} + o_{j,j,i'} - 2(o_{j-1,i,i'} - o_{j,i,i'}) \geq 0 \qquad \forall j, \ \forall i, i' < j$$

$$x_{j,i,i'} \leq M_1 o_{j,i,i'}, \quad a_{j,i,i'} \leq M_1 o_{j,i,i'} \qquad \forall j, i, i'$$

$$\sum_{i'=1}^{j+1} x_{j,i,i'} = \sum_{i'=1}^{j+1} a_{j,i,i'} - \sum_{i'=1}^{j+1} a_{j,i',i} \qquad \forall j, \ i$$

$$x_{j,i,i'}, \ a_{j,i,i'} \geq 0, \ o_{j,i,i'} \in \{0, 1\} \qquad \forall j, i, i', \forall j$$

Second Stage Subproblem

$$Q(\mathbf{o}, \mathbf{x}, \omega) = \min \quad E_\omega[\sum_{i=1}^{j}\sum_{i'=1}^{j}(c_{i'}^{w} w_{j,i,i'}(\omega) + c^{\ell}\ell_j(\omega)]$$

$s.t.$

$$w_{j,i,i'}(\omega) \le M_2(\omega) o_{j,i,i'} \qquad \forall i, i', j, \omega$$

$$s_{j,i,i'}(\omega) \le M_3(\omega) o_{j,i,i'} \qquad \forall i, i', j, \omega$$

$$-\sum_{i'=1}^{j} w_{j,i',i}(\omega) + \sum_{i'=1}^{j} w_{j,i,i'}(\omega) - \sum_{i'=1}^{j} s_{j,i,i'}(\omega) = Z_i(\omega) - \sum_{i'=1}^{j} x_{j,i,i'} \qquad \forall i, j, \omega$$

$$\ell_j(\omega) \ge \sum_{i=1}^{j}\sum_{i'=1}^{j} s_{j,i,i'}(\omega) + \sum_{i=1}^{j} Z_i(\omega) + \sum_{i'=1}^{j} x_{j,0,i'} - d \qquad \forall j, \omega$$

$$w_{j,i,i'}(\omega), s_{j,i,i'}(\omega), \ell_j(\omega) \ge 0, \qquad \forall j, i, i', \omega.$$

Model Properties

The addition of indirect waiting costs results in conditions under which FCFS is not optimal:

Proposition

For $n^U = 2$ with i.i.d. service durations if

$$c_2^a \geq c_1^w$$

then the optimal sequence is LCFS.

Methodology

Compared *L-shaped method* and *Integer L-shaped method*

- Fast solution to second stage subproblems
- Presolve
- Warm start
- Branch-and-cut vs. dynamic search
- MIP cuts (MIR, implied bound cuts, etc.)
- Mean value problem based valid inequalities

Computational Performance

No. of Customers	Class	Type of Customers	L-Shaped Method			
			CPU Time		# of Iterations	
			Average	Max	Average	Max
5	2.1	5 Add on Customers	449	484	192.9	202
	2.2	3 Routine + 2 Add on	2247.71	2546	608.7	660
7	2.3	7 Add on Customers	15000*	15000*	283	290
	2.4	4 Routine 3 Add on	15000*	15000*	241	247
10	2.5	10 Add on Customers	15000*	15000*	92	97
	2.6	7 Routine 3 Add on	15000*	15000*	93	102

Computational Performance

Table: Gap at the time of termination for the instances that are not solved to optimality

Problem Size	Instance No	Patient Type	Best Gap	
			L-Shaped Method	L-Shaped Method (mean value based cuts)
7 (uniform)	2.3	7 Add on Patients	107.12%	optimal
	2.4	4 Routine 3 Add on	174.62%	1.95%
10 (uniform)	2.5	10Add on Patients	240.11%	7.26%
	2.6	7 Routine 3 Add on	375.32%	1.99%
7 (lognormal)	3.3	7 Add on Patients	223.32%	21.99%
	3.4	4 Routine 3 Add on	335.02%	15.71%
10 (lognormal)	3.5	10Add on Patients	338.37%	31.53%
	3.6	7 Routine 3 Add on	517.307%	13.07%

Example 1: Structure of the Optimal Solution

Table: Examples with varying direct/indirect cost for instance 3.6 (7 routine, 3 add on, lognormal service times) parameters

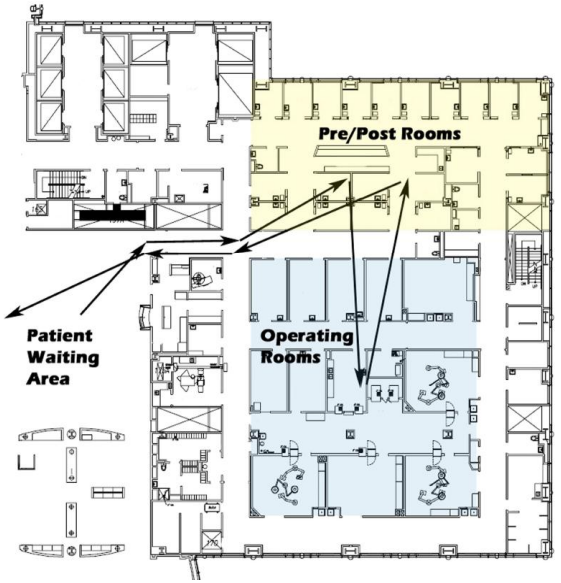
Instance No	c^a Routine	c^w Routine	c^a Add-on	c^w Add-on	c^L	Optimal Sequence	CPU Time		# of Iterations	
							ave	max	ave	max
1	0	1	0.1	0.1	10	R-R-R-R-R-R-R-A-A-A	12295.5	14980	55.2	598
2	0	1	10	10	10	A-A-A-R-R-R-R-R-R-R	1174.8	1852	163.5	209
3	0	1	50	50	10	A-A-A-R-R-R-R-R-R-R	418.2	613	94.9	122
4	0	1	100	100	10	A-A-A-R-R-R-R-R-R-R	257.6	522	67.4	112
5	0	1	250	250	10	A-A-A-R-R-R-R-R-R-R	117.2	290	36	73
6	0	1	500	500	10	A-A-A-R-R-R-R-R-R-R	52.5	112	18.1	36
7	0	1	750	750	10	A-A-A-R-R-R-R-R-R-R	28.1	48	10.3	17
8	0	1	1000	1000	10	A-A-A-R-R-R-R-R-R-R	19.4	30	7.1	10

Conclusions

- VSS can be as high as 95% and as low as 0.5%
- Large instances of dynamic scheduling problem can be solved efficiently but sequencing and scheduling is much harder
- FCFS generally optimal when probabilities of add-on customers are low and/or indirect cost of waiting is low
- Placement of add on customers is frequently “all or nothing”

Other Research

- Complex service systems with multiple servers and stages of service
- Uncertain service time, demand, and patient/provider behavior



Applications:

- Hospital surgery practices
- Outpatient procedure and treatment centers

Questions?

Brian Denton
bdenton@ncsu.edu
http://www.ise.ncsu.edu/bdenton/