Managing appointment scheduling under patient choices

Peter van de Ven

Joint work with: Nan Liu (Columbia University) and Bo Zhang (IBM Research)

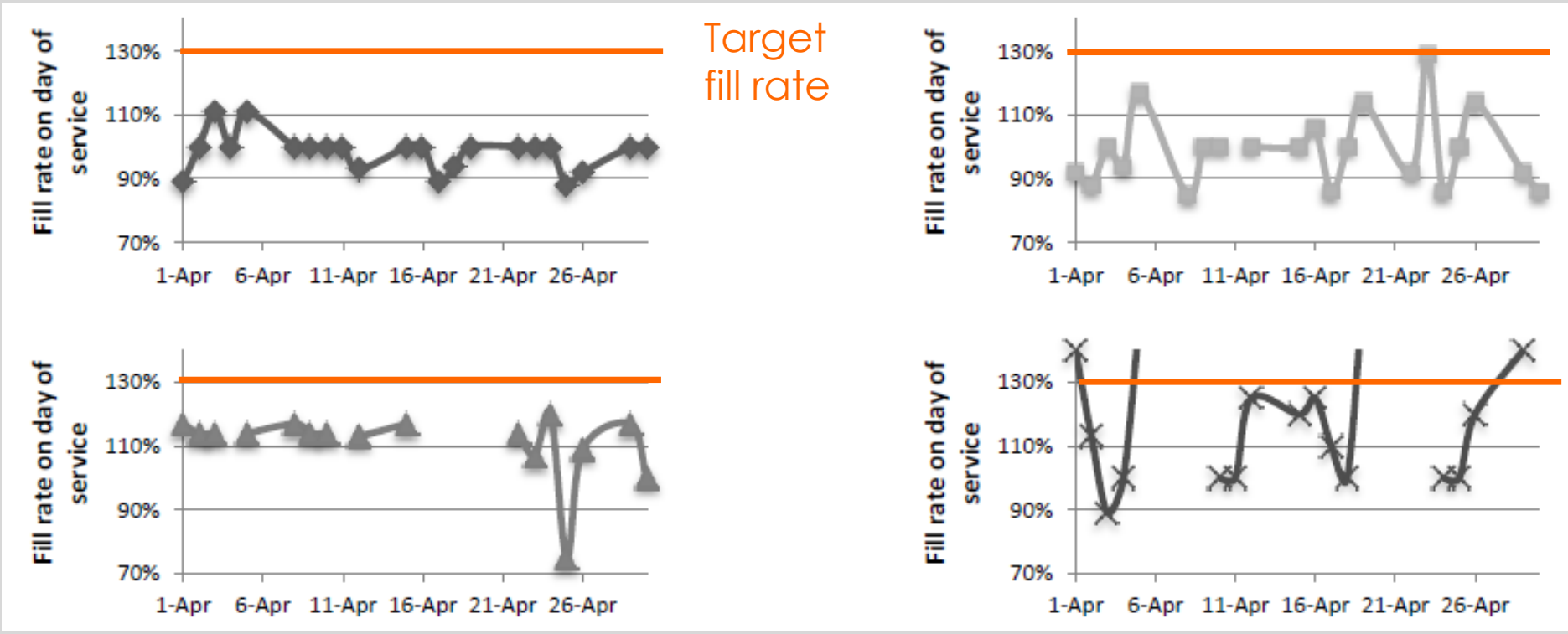
online decision
interaction between providers
and patients

Appointment scheduling in two steps:

1. determine appointment template
2. assign patients to appointment slots

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
9:00	established	established	established	established	established	established
9:20	walk-in	walk-in	walk-in	walk-in	walk-in	established
9:40	established	established	established	established	established	walk-in
10:00	walk-in	walk-in	walk-in	walk-in	walk-in	walk-in
10:20	walk-in	walk-in	walk-in	walk-in	walk-in	established
10:40	established	established	established	established	established	walk-in
11:00	established	established	established	established	established	established
11:20	walk-in	walk-in	walk-in	walk-in	walk-in	established
11:40	established	established	established	established	established	walk-in
12:00	walk-in	walk-in	walk-in	walk-in	walk-in	established
12:20	walk-in	walk-in	walk-in	walk-in	walk-in	established

Performance metric



Fill rate measures fraction of slots used

- as low as 75%
- fluctuation over time

Patient preference survey:

- Amazon Mechanical Turk, 301 participants
- imagine want to book non-urgent appointment, indicate all acceptable time slots
- early day (8am-10am), mid-day (10am-3pm), late day (3pm-8pm)

	Count	Percentage
ED only	23	8.5%
MD only	58	21.3%
LD only	31	11.4%
ED and MD	40	14.7%
ED and LD	21	7.7%
MD and LD	43	15.8%
ED, MD and LD	56	20.6%
Total	272	100%

Observe that preference is significant and heterogeneous

	Count	Percentage
ED only	23	8.5%
MD only	58	21.3%
LD only	31	11.4%
ED and MD	40	14.7%
ED and LD	21	7.7%
MD and LD	43	15.8%
ED, MD and LD	56	20.6%
Total	272	100%

Given these patient preferences, how can we

- model the scheduling process?
- tweak it to maximize fill rate?

Related literature

Patient scheduling:

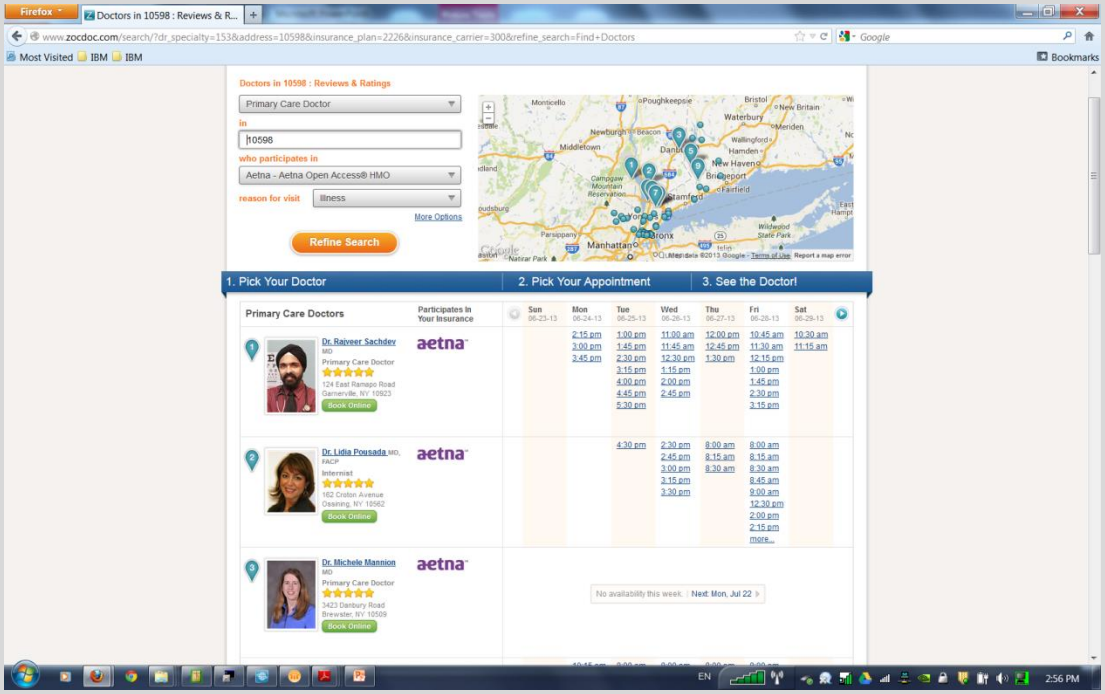
- Gupta and Wang (2008, 2011): patient reveals preference first
- Feldman *et al.* (2014): no appointment template

Related to assortment optimization

- Golrezaei et al. (2014), Bernstein et al. (2015), Gallego et al. (2016)
- we use no information on the patient type.

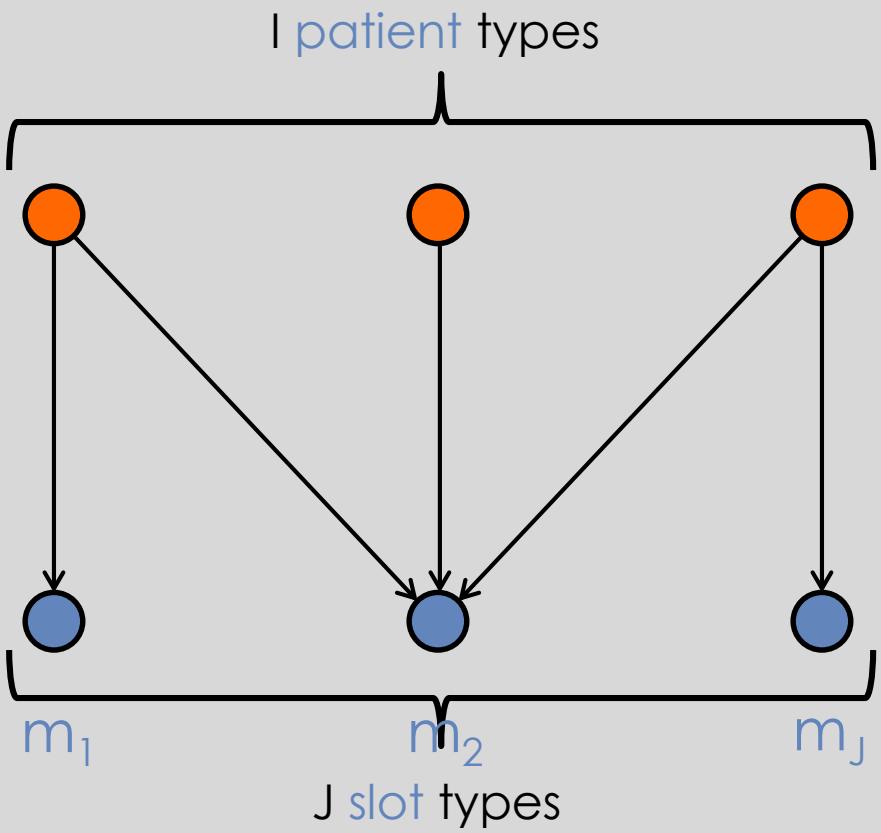
Patient-provider interaction

traditional scheduling

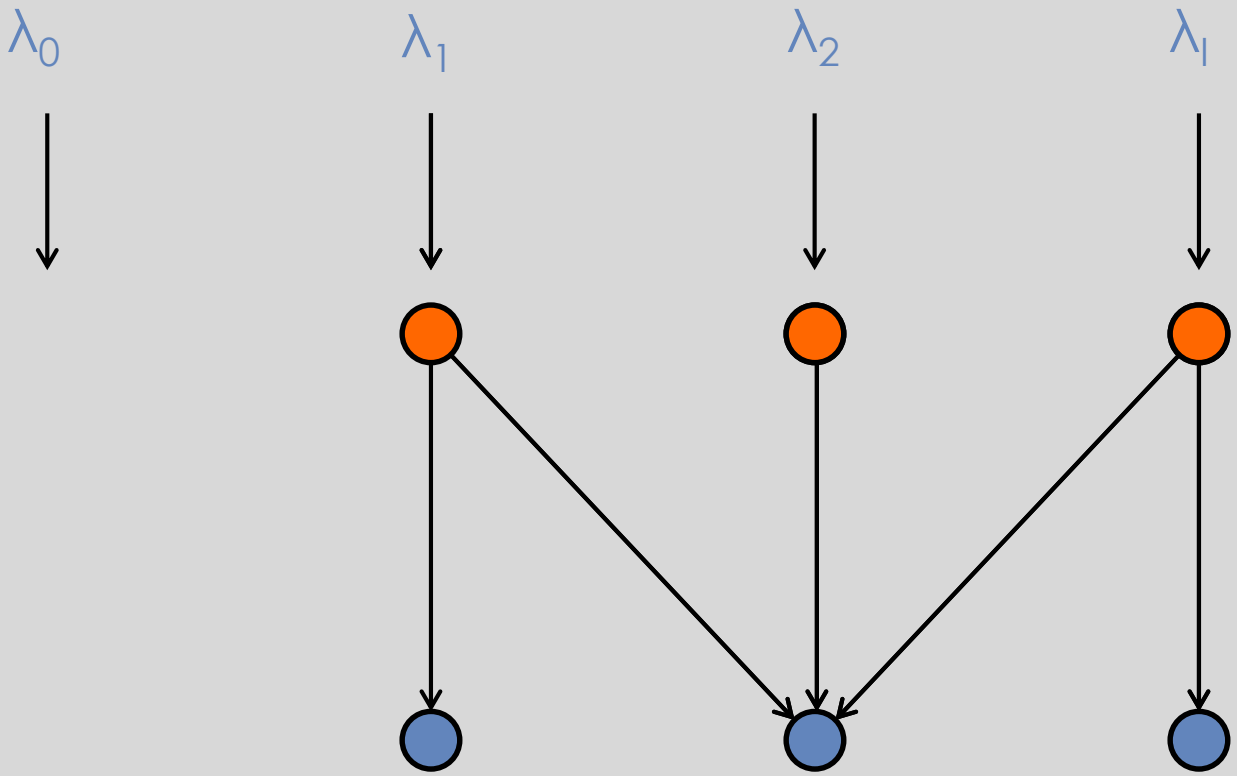


electronic scheduling

electronic scheduling model



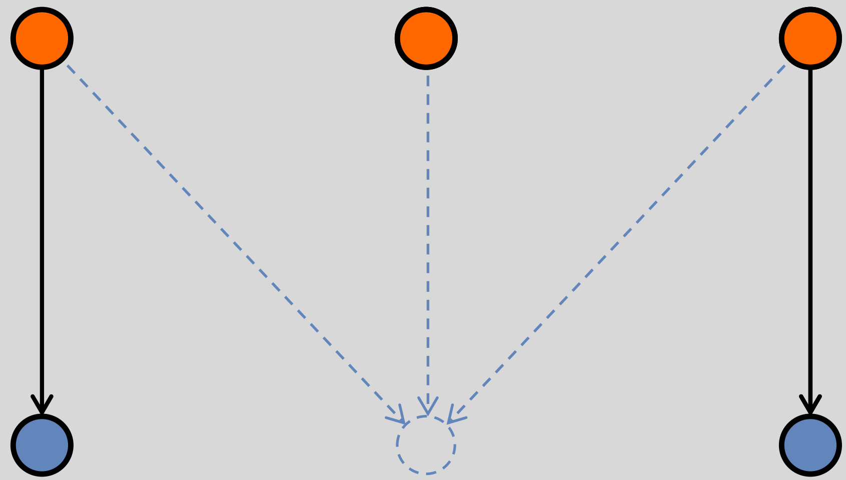
IxJ incidence matrix Ω captures patient preferences



N stages with at most one arriving patient each

- i.i.d. patient probability λ_i per stage

$$\lambda_0 = 1 - \sum_{i=1}^{I} \lambda_i$$



Upon patient arrival, the scheduler offers a subset of slots S

- based on stage j and capacity left $\mathbf{m} = (m_1,\ldots,m_J)$
- no knowledge of patient type

Patient selects uniformly at random among offered slots

- reduce capacity by 1
- no appointment if no acceptable slot offered

maximizing fill rate

Question: how to choose $S(n, \boldsymbol{m})$ to maximize fill rate?

number of stages and capacity left

only non-empty slots can be offered

Our decision variable is whether to open each slot

$$S(n, \boldsymbol{m}) \subseteq \{j = 1, \dots, J : m_j \geq 1\}$$

and the question can be written as an MDP

$$V_n(\boldsymbol{m}) = \max_S \Big(\sum_{j=0}^{J} p_j(\boldsymbol{m}, S)[\mathbf{1}_{(j>0)} + V_{n-1}(\boldsymbol{m} - \boldsymbol{e}_j)] \Big)$$

reward if slot is taken

transition probability depend on choice and state

Greedy policy

The greedy policy π_0 (offering all slots) is a natural heuristic

- Denote $V_{n,\pi_0}(\boldsymbol{m})$ its expected fill rate

Theorem

$\forall n, \boldsymbol{m}, \Omega, V_{n,\pi_0}(\boldsymbol{m}) \leq V_n(\boldsymbol{m}) \leq 2V_{n,\pi_0}(\boldsymbol{m})$

Greedy policy II

Greedy is asymptotically optimal in certain regimes $J \to \infty$

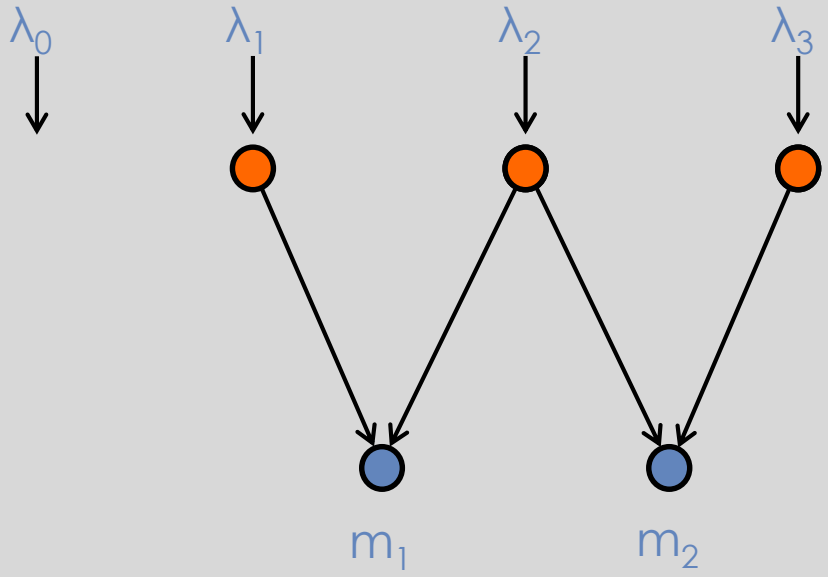
- I^J, N^J, m^j grow linearly in J
- number of slots per patient is bounded

Theorem

Under certain conditions

$$\lim_{J\to\infty} \frac{V_N(\boldsymbol{m}) - V_{N,\pi_0}(\boldsymbol{m})}{V_{N,\pi_0}(\boldsymbol{m})} = 0$$

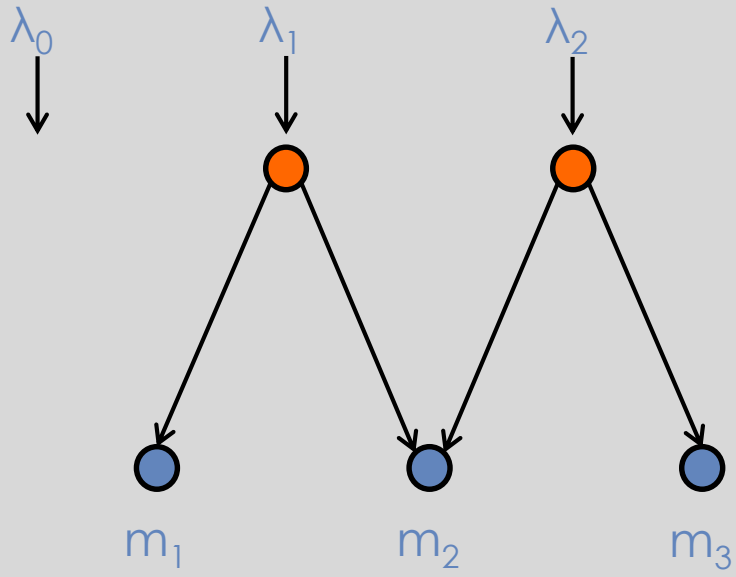
Specific model instances



"W" network

Proposition

For the "W" network, the greedy policy is optimal



"M" network

Proposition

For the "M" network, the following policy is optimal

$$S^*(\boldsymbol{m}, n) = \begin{cases} \{1,3\} & \text{if } m_1 > 0 \text{ and } m_3 > 0 \\ \text{`greedy'} & \text{otherwise} \end{cases}$$

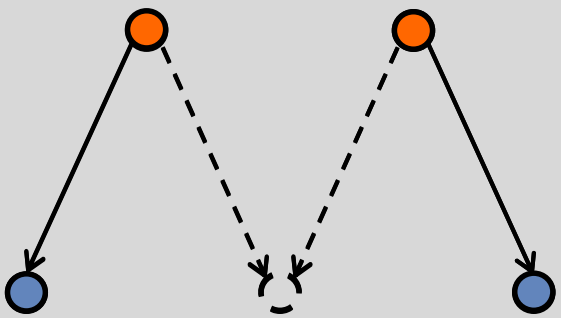
Relative improvement of the optimal policy over greedy policy

Table 1 **Capacity utilization improvement under the M model (optimal non-sequential vs. greedy).**

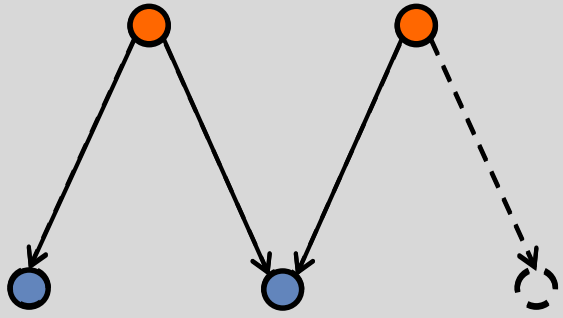
N	# of Scenarios	$(\lambda_1, \lambda_2) = (1/2, 1/2)$			$(\lambda_1, \lambda_2) = (1/3, 2/3)$			$(\lambda_1, \lambda_2) = (1/4, 3/4)$		
		Max	Average	Median	Max	Average	Median	Max	Average	Median
20	45	4.6%	3.7%	3.8%	4.3%	3.4%	3.5%	3.7%	3.0%	3.1%
30	91	5.0%	3.8%	3.9%	4.7%	3.6%	3.6%	4.0%	3.2%	3.3%
40	153	5.3%	3.9%	3.9%	4.9%	3.7%	3.7%	4.2%	3.4%	3.4%
50	231	5.6%	3.9%	3.9%	5.1%	3.8%	3.8%	4.3%	3.4%	3.5%

Some observation on the optimal policies:

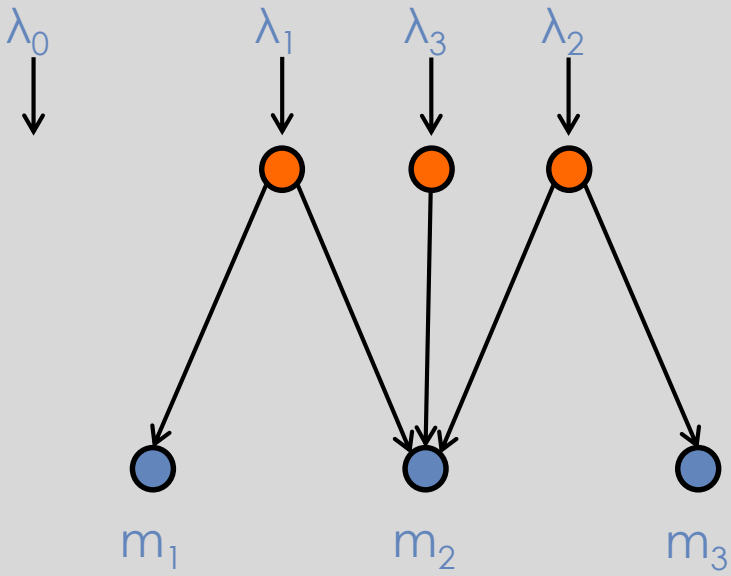
- work-conserving
- save versatile slots for later
- independent from **m**, λ and n



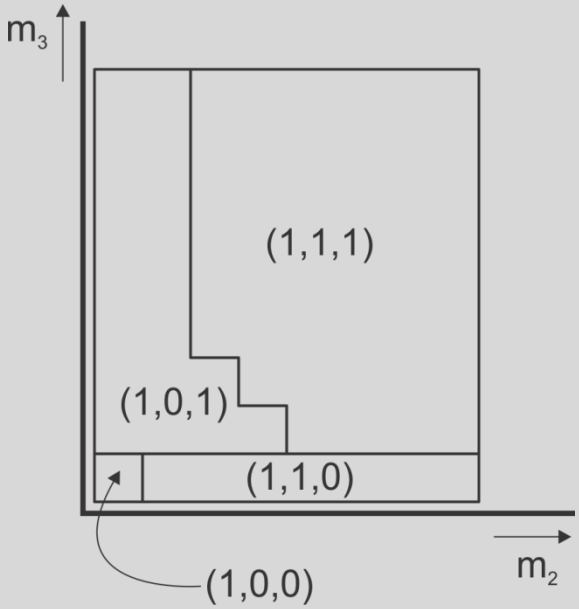
good



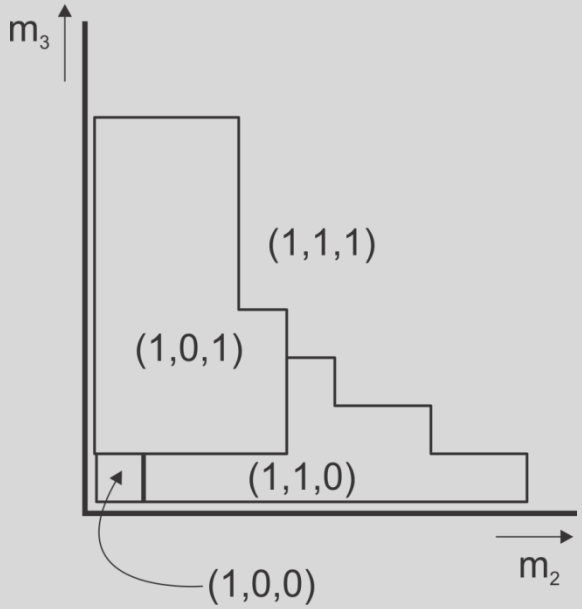
bad



"M+1" network



λ3 large



λ3 small

Greedy does well in the `M+1' model and beyond

N	# of Scenarios	$(\lambda_1,\lambda_2,\lambda_3)=(9/20,9/20,1/10)$			$(\lambda_1,\lambda_2,\lambda_3)=(2/5,2/5,1/5)$			$(\lambda_1,\lambda_2,\lambda_3)=(3/10,3/10,2/5)$		
		Max	Average	Median	Max	Average	Median	Max	Average	Median
20	45	3.1%	2.1%	2.0%	2.0%	1.1%	1.0%	0.7%	0.3%	0.2%
30	91	3.5%	2.1%	2.0%	2.3%	1.2%	1.0%	0.8%	0.3%	0.2%
40	153	3.8%	2.2%	2.0%	2.5%	1.2%	1.0%	0.8%	0.2%	0.1%
50	231	4.0%	2.2%	2.1%	2.7%	1.2%	1.0%	0.8%	0.2%	0.1%

Summary

Preprint: Nan Liu, Peter van de Ven, and Bo Zhang (2016). Managing Appointment Scheduling under Patient Choices

Online (non-sequential) scheduling

- offer multiple slots in sequence
- greedy does not do well, develop heuristic
- median improvement 8-9%, revenue gains $45k - $120k

Proposal: (semi)parametric model for appointment scheduling

- find ways to improve these scheduling mechanisms