

**Data-Based**  
**Service Engineering** (Science, Management)  
in **Call Centers, Hospitals, . . .**

Avishai Mandelbaum

Technion, Haifa, Israel

<http://ie.technion.ac.il/serveng>

**16th IE&M Conference**, Tel-Aviv, March 2010

## Research Partners

- ▶ **Students:**

Aldor\*, Baron\*, Carmeli, Feldman, Garnett\*, Gurvich\*, Khudiakov\*, Maman\*, Marmor, Reich, Rosenshmidt\*, Shaikhet\*, Senderovic, Tseytlin\*, Yom-Tov, Zaied, Zeltyn\*, Zohar\*, Zviran, ...

- ▶ **Empirical/Statistical Analysis:**

Brown, Gans, Zhao; Shen; Ritov, Goldberg; Allon, Bassamboo, Gurvich; Armony, ...

- ▶ **Theory:**

Armony, Atar, Feigin, Gurvich, Jelenkovic, Kaspi, Massey, Momcilovic, Reiman, Shimkin, Stolyar, Wasserkrug, Whitt, Zeltyn, ...

- ▶ **Industry:**

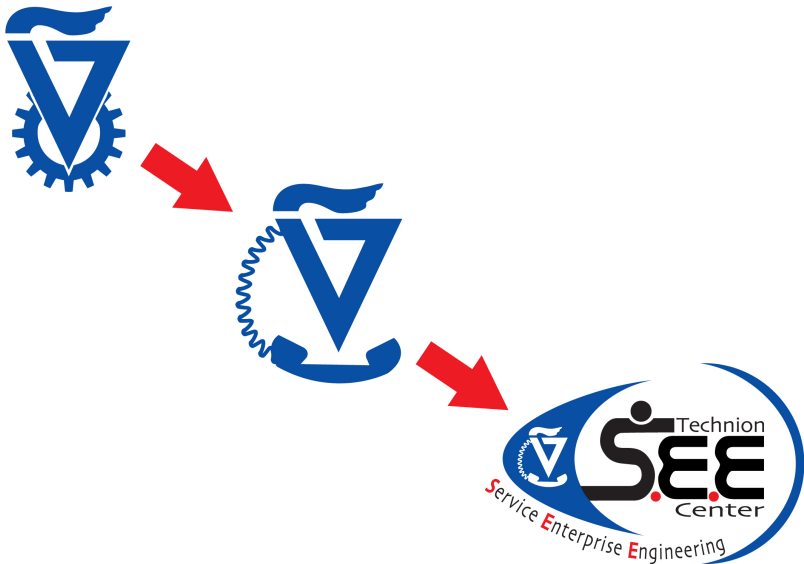
IBM Research (OCR: Carmeli, Vortman, Wasserkrug, Zeltyn), Rambam Hospital, Hapoalim Bank, Mizrahi Bank, Pelephone Cellular, ...

- ▶ **Technion SEE Center / Laboratory:**

Feigin; Trofimov, Nadjharov, Gavako, Kutsyy; Liberman, Koren, Rom; Research Assistants, ...

# The Technion SEE Center / Laboratory

## Data-Based Research & Teaching



## History, Resources (Downloadable)

- ▶ Math. + C.S. + Stat. + O.R. + Mgt.  $\Rightarrow$  **IE&M**
- ▶ **“Service-Engineering” Course** ( $\geq 1995$ ):  
<http://ie.technion.ac.il/serveng> - website  
[http://ie.technion.ac.il/serveng/References/teaching\\_paper.pdf](http://ie.technion.ac.il/serveng/References/teaching_paper.pdf)



## History, Resources (Downloadable)

- ▶ Math. + C.S. + Stat. + O.R. + Mgt.  $\Rightarrow$  **IE&M**
- ▶ **“Service-Engineering” Course** ( $\geq 1995$ ):  
<http://ie.technion.ac.il/serveng> - website  
[http://ie.technion.ac.il/serveng/References/teaching\\_paper.pdf](http://ie.technion.ac.il/serveng/References/teaching_paper.pdf)
- ▶ **SEELab** ( $\geq 2007$ ), following StatLab ( $\geq 2000$ ):  
Data, Reports, Tutorials.  
<http://ie.technion.ac.il/Labs/Serveng>
- ▶ **OCR Project** ( $\geq 2008$ ):  
IBM Research + Rambam Hospital + Technion IE&M  
[http://ie.technion.ac.il/Labs/Serveng/closed/OCR\\_Documents.php](http://ie.technion.ac.il/Labs/Serveng/closed/OCR_Documents.php)

## History, Resources (Downloadable)

- ▶ Math. + C.S. + Stat. + O.R. + Mgt.  $\Rightarrow$  **IE&M**
- ▶ **“Service-Engineering” Course** ( $\geq 1995$ ):  
<http://ie.technion.ac.il/serveng> - website  
[http://ie.technion.ac.il/serveng/References/teaching\\_paper.pdf](http://ie.technion.ac.il/serveng/References/teaching_paper.pdf)
- ▶ **SEELab** ( $\geq 2007$ ), following StatLab ( $\geq 2000$ ):  
Data, Reports, Tutorials.  
<http://ie.technion.ac.il/Labs/Serveng>
- ▶ **OCR Project** ( $\geq 2008$ ):  
IBM Research + Rambam Hospital + Technion IE&M  
[http://ie.technion.ac.il/Labs/Serveng/closed/OCR\\_Documents.php](http://ie.technion.ac.il/Labs/Serveng/closed/OCR_Documents.php)
- ▶ **Technion IE&M** = Outsourcing Knowledge (Research, Practice)  
**e.g.** Search Google Scholar for **<Call Centers>**:  
First 5 entries originated at the Technion.

# The Case for Service Science / Engineering

- ▶ **Service Science / Engineering** (vs. Management) are emerging **Academic Disciplines**. For example, universities (world-wide), IBM (SSME, a la Computer-Science), USA NSF (SEE), Germany IAO (ServEng), ...

# The Case for Service Science / Engineering

- ▶ **Service Science / Engineering** (vs. Management) are emerging **Academic Disciplines**. For example, universities (world-wide), IBM (SSME, a la Computer-Science), USA NSF (SEE), Germany IAO (ServEng), ...
- ▶ **Simple** models that explain **fundamental phenomena**, which are **common** across applications:
  - Call Centers
  - Hospitals
  - Justice
  - Transportation
  - ...

# The Case for Service Science / Engineering

- ▶ **Service Science / Engineering** (vs. Management) are emerging **Academic Disciplines**. For example, universities (world-wide), IBM (SSME, a la Computer-Science), USA NSF (SEE), Germany IAO (ServEng), ...
- ▶ **Simple** models that explain **fundamental phenomena**, which are **common** across applications:
  - Call Centers
  - Hospitals
  - Justice
  - Transportation
  - ...
- ▶ **What Can Be Done** vs. **How To** (*Pause for a Commercial*)

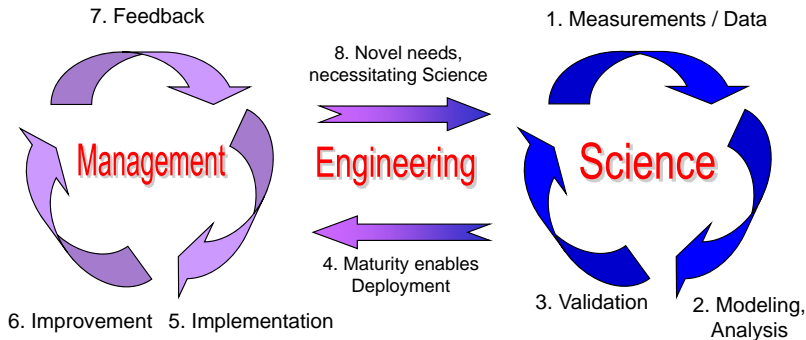
# Expanding the Scientific Paradigm

**Service Engineering** vs. **Industrial** Engineering  
**Human Complexity**

# Expanding the Scientific Paradigm

**Service Engineering** vs. **Industrial Engineering**

**Human Complexity**  $\Rightarrow$  **Scientific Paradigm** (Physics ... Economics)  
and beyond (with IBM Research):



# Started with Call Centers, Expanded to Hospitals

## Call Centers - U.S. (Israel) Stat.

- ▶ \$200 – \$300 billion annual expenditures (0.5)
- ▶ 100,000 – 200,000 call centers (500)
- ▶ “Window” into the company, for better or worse
- ▶ Over 3 million agents = **2% – 4% workforce** (11K)



# Started with Call Centers, Expanded to Hospitals

## Call Centers - U.S. (Israel) Stat.

- ▶ \$200 – \$300 billion annual expenditures (0.5)
- ▶ 100,000 – 200,000 call centers (500)
- ▶ “Window” into the company, for better or worse
- ▶ Over 3 million agents = **2% – 4% workforce** (11K)

## Healthcare - similar and unique challenges:

- ▶ Cost-figures far more staggering
- ▶ Risks much higher
- ▶ ED (initial focus) = hospital-window
- ▶ Over 3 million nurses

# Call-Center Environment: Service Network



## Call-Centers: “Sweat-Shops of the 21st Century”



# ER / ED Environment: Service Network

## Acute (Internal, Trauma)



## Walking



## Multi-Trauma



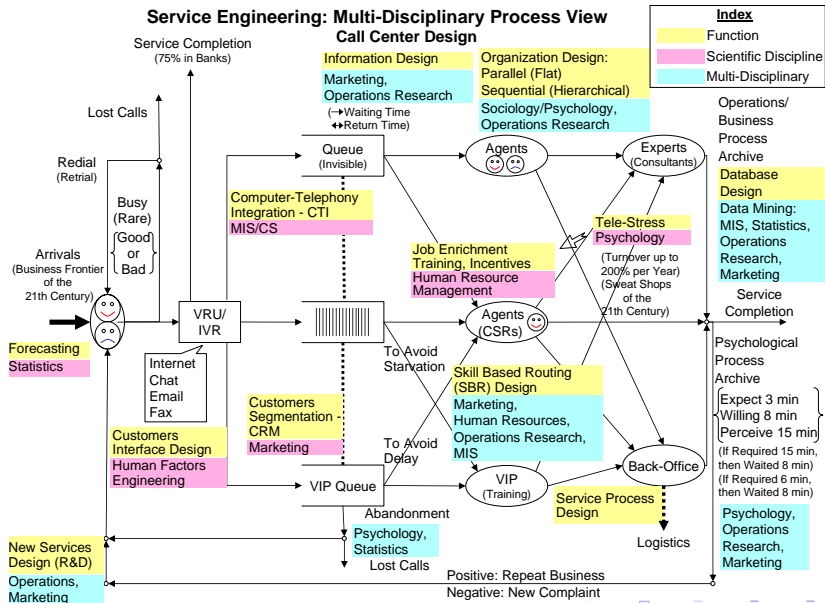
## ED-Environment in Israel



## ED-Queue in a “Good” Beijing Hospital

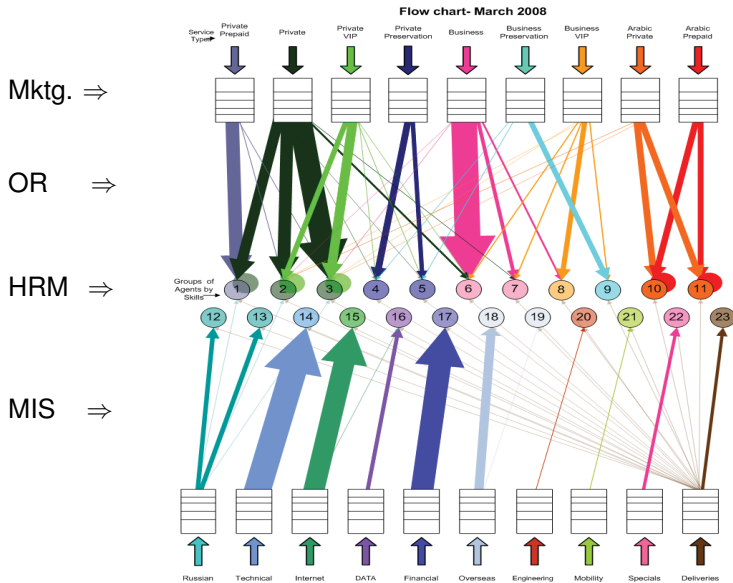


# Call-Center: Multi-Disciplinary ServEng View



# Skills-Based Routing in Call Centers

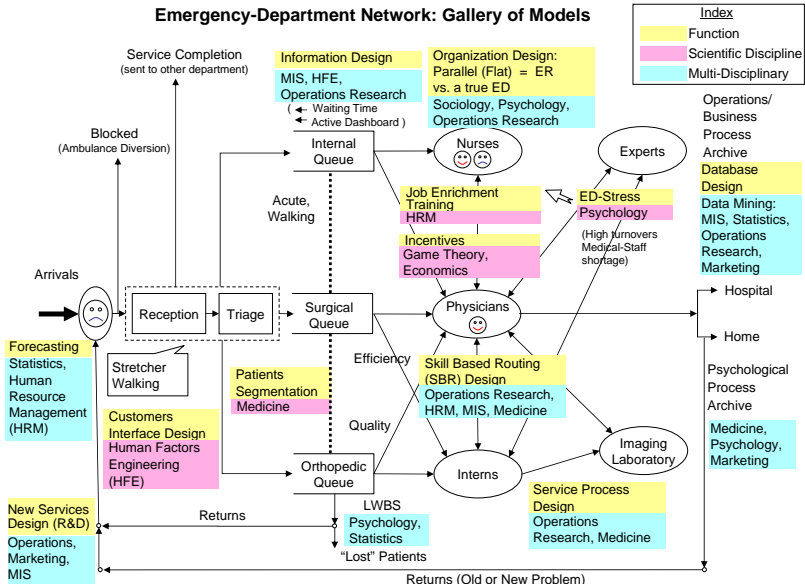
## EDA and OR, with I. Gurvich and P. Lieberman





# Emergency-Dept.: Multi-Disciplinary ServEng View

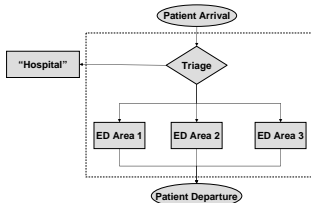
## Emergency-Department Network: Gallery of Models



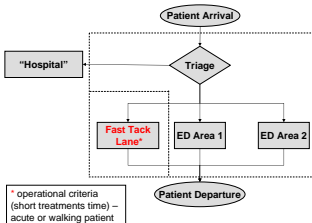
## ED Design, with B. Golany and Y. Marmor (PhD)

Routing: **Triage (Clinical)**, **Fast-Track (Operational)**, ... (via DEA)

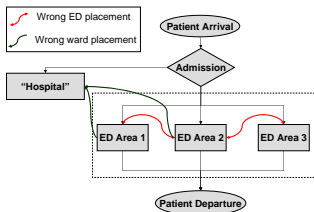
e.g. Fast Track most suitable when elderly dominate



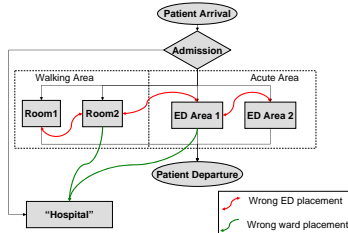
(a) Triage Model



(b) Fast-Track Model



(c) Illness-based Model



(d) Walking-Acute Model

## ED-to-IW Routing: A Hospital Bottleneck

Israeli Large Hospital (1/5/06 to 30/10/08, excluding 1-3/07)

	Ward A	Ward B	Ward C	Ward D
ALOS (days)	6.37	4.47	5.36	5.56
Avg Occupancy Rate	97%	95%	86%	92%
Avg # Patients per Month	206	187	210	210
Standard bed capacity	45	30	44	42
Avg # Patients /Bed/Month	4.57	6.25	4.77	4.77
Returns (within 3 months)	15.4%	15.6%	16.2%	14.8%

## ED-to-IW Routing: A Hospital Bottleneck

Israeli Large Hospital (1/5/06 to 30/10/08, excluding 1-3/07)

	Ward A	Ward B	Ward C	Ward D
ALOS (days)	6.37	4.47	5.36	5.56
Avg Occupancy Rate	97%	95%	86%	92%
Avg # Patients per Month	206	187	210	210
Standard bed capacity	45	30	44	42
Avg # Patients /Bed/Month	4.57	6.25	4.77	4.77
Returns (within 3 months)	15.4%	15.6%	16.2%	14.8%

- ▶ The “fastest” + smallest **Ward B** subject to highest **workload** = **bed-occupancy, bed-turnover (flux): unfair !**
- ▶ Calls for ED-to-IW routing, which is both **efficient and fair** (w/ **Tseytlin** (M.Sc.), Tseytlin & **Momcilovic**, Tseytlin & **Zviran**): **Markov exact, QED approx. (natural), partial information.**

## On “Fairness” in Hospitals (“Justice-Table”)

- ▶ **Patients Fairness** (prevalent): Least delays, hence higher priority to “faster” wards.
- ▶ **Personnel Fairness**: Nurses (doctors) share equal **Workload**.

## On “Fairness” in Hospitals (“Justice-Table”)

- ▶ **Patients Fairness** (prevalent): Least delays, hence higher priority to “faster” wards.
- ▶ **Personnel Fairness**: Nurses (doctors) share equal **Workload**.
- ▶ **Bed-Occupancy = Bed-Turnover  $\times$  ALOS**,  
by **Little’s Law**.

## Fair & Efficient ED-to-IW Routing

- ▶ **Tunable Routing**, customized to preferences, with **Y. Tseytlin** and **P. Momcilovic**:

Route to ward with **highest (weighted) idleness-ratio**, i.e.  
 $\# \text{ idle beds in ward} / \# \text{ idle-beds in total}.$

## Fair & Efficient ED-to-IW Routing

- ▶ **Tunable Routing**, customized to preferences, with **Y. Tseytlin** and **P. Momcilovic**:

Route to ward with **highest (weighted) idleness-ratio**, i.e.  
 $\# \text{ idle beds in ward} / \# \text{ idle-beds in total}$ .

- ▶ For example, can be tuned so that faster wards have lower occupancy (**nurses happy**) and higher turnover (**management happy**).



## Fair & Efficient ED-to-IW Routing

- ▶ **Tunable Routing**, customized to preferences, with **Y. Tseytlin** and **P. Momcilovic**:

Route to ward with **highest (weighted) idleness-ratio**, i.e.

$\# \text{ idle beds in ward} / \# \text{ idle-beds in total}$ .

- ▶ For example, can be tuned so that faster wards have lower occupancy (**nurses happy**) and higher turnover (**management happy**).
- ▶ Last, but not least - Workload is both **Operational + Cognitive**:  
5 minutes taking temperature vs. 5 minutes saving life.  
e.g. Two Maternity Wards perceive unfairness, hence psychological:
  - ▶ Ward 1: complications **before** birth
  - ▶ Ward 2: complications **after** birth
  - ▶ **Fair routing** of **normal** births? (Just starting, with **A. Rafaeli**)

## Prerequisite I: Data

**Averages Prevalent** (and could be useful / interesting).

But I need data at the level of the **Individual Transaction**:

For each service transaction (during a phone-service in a call center, or a patient's visit in a hospital, or browsing in a website, or ...), its

**operational history** = time-stamps of events .

## Prerequisite I: Data

**Averages Prevalent** (and could be useful / interesting).

But I need data at the level of the **Individual Transaction**:

For each service transaction (during a phone-service in a call center, or a patient's visit in a hospital, or browsing in a website, or ...), its

**operational history** = time-stamps of events .

Sources: **"Service-floor"** (vs. Industry-level, Surveys, ...)

- ▶ **Administrative** (Court, via "paper analysis")
- ▶ **Face-to-Face** (Bank, via bar-code readers)
- ▶ **Telephone** (Call Centers, via ACD / CTI, IVR/VRU)
- ▶ **Hospitals** (Emergency Departments, ...)

## Prerequisite I: Data

**Averages Prevalent** (and could be useful / interesting).

But I need data at the level of the **Individual Transaction**:

For each service transaction (during a phone-service in a call center, or a patient's visit in a hospital, or browsing in a website, or ...), its

**operational history** = time-stamps of events .

Sources: **"Service-floor"** (vs. Industry-level, Surveys, ...)

- ▶ **Administrative** (Court, via "paper analysis")
- ▶ **Face-to-Face** (Bank, via bar-code readers)
- ▶ **Telephone** (Call Centers, via ACD / CTI, IVR/VRU)
- ▶ **Hospitals** (Emergency Departments, ...)
- ▶ Expanding:
  - ▶ Hospitals, via **RFID**, with **B. Carmeli, S. Israelit, Y. Marmor**
  - ▶ Operational + Financial + Contents (Marketing, Clinical)
  - ▶ Internet, chat (multi-media)

## Prerequisite II: Models (The Fluid View)

“Laws of Large Numbers” capture **Predictable** Variability (Averages)

## Prerequisite II: Models (The Fluid View)

“Laws of Large Numbers” capture **Predictable** Variability (Averages)



## Flow Design and Control: Transportation (Fluid) Network



# The Fluid View: Labor-day Queueing at Niagara Falls

Stochastic Individualism Averaged out by the LLNs (Scale)





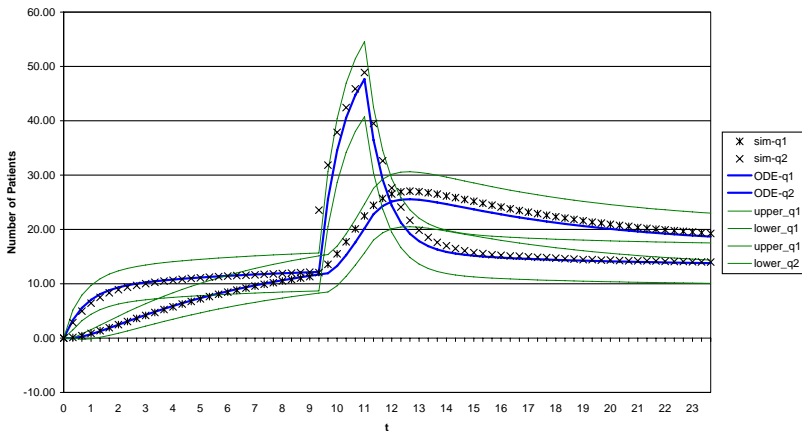
# Fluid Models: Preparing for Mass-Casualty Events

# Fluid Models: Preparing for Mass-Casualty Events

e.g. **Erlang-R** = **R**eEntrant Patients, with **G. Yom-Tov** (PhD).

**5-fold Rise** in Inflow-Rate, between 9am -11am:

Delta = 0.2; Mu = 1; p = 0.25; s = 50; Lambda=10 (t<9 or t>11), Lambda=50 (9<t<11)



## Prerequisite II: Models (Stochastic)

**Traditional Queueing Theory** predicts that **Service-Quality** and **Servers' Efficiency** must be traded off against each other.

e.g. **Single-server** queue (M/M/1) in **Heavy-Traffic**:  
**91%** server's utilization goes with

$$\text{Congestion Index} = \frac{E[\text{Wait}]}{E[\text{Service}]} = 10,$$

and only **9%** of the customers are served immediately upon arrival.

## Prerequisite II: Models (Stochastic)

Traditional Queueing Theory predicts that **Service-Quality** and **Servers' Efficiency** must be traded off against each other.

e.g. **Single-server** queue (M/M/1) in **Heavy-Traffic**:  
**91%** server's utilization goes with

$$\text{Congestion Index} = \frac{E[\text{Wait}]}{E[\text{Service}]} = 10,$$

and only **9%** of the customers are served immediately upon arrival.

**Yet**, heavily-loaded queueing systems with **Congestion Index = 0.1** (Waiting one order of magnitude less than Service) are prevalent:

- ▶ **Call Centers**: Wait "**seconds**" for **minutes** service;
- ▶ **Transportation**: Search "**minutes**" for **hours** parking;
- ▶ **Hospitals**: Wait "**hours**" in ED for **days** hospitalization in IW's;

## Prerequisite II: Models (Stochastic)

Traditional Queueing Theory predicts that **Service-Quality** and **Servers' Efficiency** must be traded off against each other.

e.g. **Single-server** queue (M/M/1) in **Heavy-Traffic**:  
**91%** server's utilization goes with

$$\text{Congestion Index} = \frac{E[\text{Wait}]}{E[\text{Service}]} = 10,$$

and only **9%** of the customers are served immediately upon arrival.

**Yet**, heavily-loaded queueing systems with **Congestion Index = 0.1** (Waiting one order of magnitude less than Service) are prevalent:

- ▶ **Call Centers**: Wait "**seconds**" for **minutes** service;
- ▶ **Transportation**: Search "**minutes**" for **hours** parking;
- ▶ **Hospitals**: Wait "**hours**" in ED for **days** hospitalization in IW's;

and, moreover, a significant fraction are not delayed in queue. (For example, in well-run call-centers, **50%** served "immediately", along with over **90%** agents' utilization, is not uncommon ) **?** **QED**

## Operational Regimes: Conceptual Framework

**$R$ : Offered Load** not too small.

def.  $R = \text{Arrival-rate} \times \text{Average-Service-Time}$

e.g.  $R = 25 \text{ calls/min.} \times 4 \text{ min./call} = 100$

$N = \text{\#Agents}$  ?

## Operational Regimes: Conceptual Framework

**$R$ : Offered Load** not too small.

def.  $R$  = Arrival-rate  $\times$  Average-Service-Time

e.g.  $R$  = 25 calls/min.  $\times$  4 min./call = **100**

**$N$**  = #Agents **?**

**QD Regime:**  **$N \approx R + \delta R$** ,  $0.1 < \delta < 0.25$  (e.g.  $N = 115$ )

- ▶ Essentially **no** delays

## Operational Regimes: Conceptual Framework

**R: Offered Load** not too small.

def.  $R = \text{Arrival-rate} \times \text{Average-Service-Time}$

e.g.  $R = 25 \text{ calls/min.} \times 4 \text{ min./call} = 100$

$N = \text{\#Agents}$  ?

**QD Regime:**  $N \approx R + \delta R$ ,  $0.1 < \delta < 0.25$  (e.g.  $N = 115$ )

- ▶ Essentially **no** delays

**ED Regime:**  $N \approx R - \gamma R$ ,  $0.1 < \gamma < 0.25$  (e.g.  $N = 90$ )

- ▶ **Garnett, M. & Reiman 2003**
- ▶ Essentially **all** customers are delayed
- ▶ Wait same order as service-time;  $\gamma\%$  Abandon (10-25%).



# Operational Regimes: Conceptual Framework

**R: Offered Load** not too small.

def.  $R = \text{Arrival-rate} \times \text{Average-Service-Time}$

e.g.  $R = 25 \text{ calls/min.} \times 4 \text{ min./call} = 100$

**N** = #Agents ?

**QD Regime:**  $N \approx R + \delta R$ ,  $0.1 < \delta < 0.25$  (e.g.  $N = 115$ )

- ▶ Essentially **no** delays

**ED Regime:**  $N \approx R - \gamma R$ ,  $0.1 < \gamma < 0.25$  (e.g.  $N = 90$ )

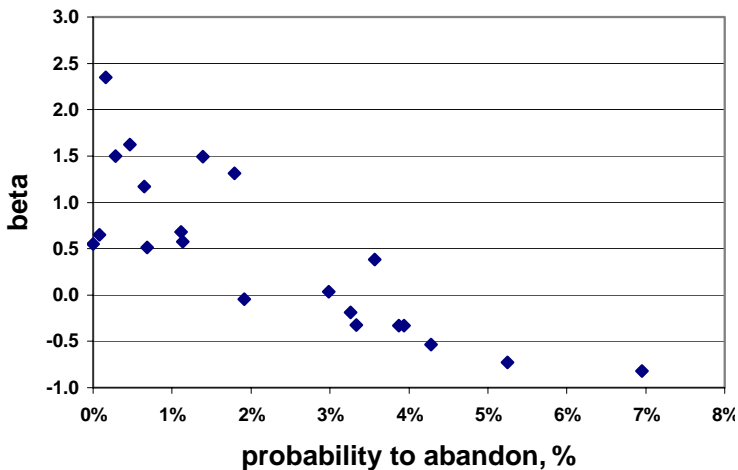
- ▶ **Garnett, M. & Reiman 2003**
- ▶ Essentially **all** customers are delayed
- ▶ Wait same order as service-time;  $\gamma\%$  Abandon (10-25%).

**QED Regime:**  $N \approx R + \beta\sqrt{R}$ ,  $-1 < \beta < +1$  (e.g.  $N = 100$ )

- ▶ Erlang 1913/24, Halfin & Whitt 1981, Garnett et. al.
- ▶ %Delayed between 25% and 75%
- ▶ Wait one-order below service-time (sec vs. min); 1-5% Abandon

## QED: Practical Support

QOS parameter  $\beta = (N - R)/\sqrt{R}$  vs. %Abandonment



# Operational Regimes: Rules-of-Thumb, with S. Zeltyn

Constraint	P{Ab}		E[W]		P{W > T}	
Offered Load	Tight	Loose	Tight	Loose	Tight	Loose
	1-10%	$\geq 10\%$	$\leq 10\%E[\tau]$	$\geq 10\%E[\tau]$	$0 \leq T \leq 10\%E[\tau]$ $5\% \leq \alpha \leq 50\%$	$T \geq 10\%E[\tau]$ $5\% \leq \alpha \leq 50\%$
Small (10's)	QED	QED	QED	QED	QED	QED
Moderate-to-Large (100's-1000's)	QED	ED, QED	QED	ED, QED if $\tau \stackrel{d}{=} \exp$	QED	ED+QED

## Operational Regimes: Rules-of-Thumb, with S. Zeltyn

Constraint	P{Ab}		E[W]		P{W > T}	
	Tight	Loose	Tight	Loose	Tight	Loose
	1-10%	$\geq 10\%$	$\leq 10\%E[\tau]$	$\geq 10\%E[\tau]$	$0 \leq T \leq 10\%E[\tau]$ $5\% \leq \alpha \leq 50\%$	$T \geq 10\%E[\tau]$ $5\% \leq \alpha \leq 50\%$
Offered Load						
Small (10's)	QED	QED	QED	QED	QED	QED
Moderate-to-Large (100's-1000's)	QED	ED, QED	QED	ED, QED if $\tau \stackrel{d}{=} \exp$	QED	ED+QED

**ED:**  $N \approx R - \gamma R \quad (0.1 \leq \gamma \leq 0.25).$

**QD:**  $N \approx R + \delta R \quad (0.1 \leq \delta \leq 0.25).$

**QED:**  $N \approx R + \beta \sqrt{R} \quad (-1 \leq \beta \leq 1).$

**ED+QED:**  $N \approx (1 - \gamma)R + \beta \sqrt{R} \quad (\gamma, \beta \text{ as above}).$

## Operational Regimes: Rules-of-Thumb, with S. Zeltyn

Constraint	P{Ab}		E[W]		P{W > T}	
	Tight	Loose	Tight	Loose	Tight	Loose
	1-10%	$\geq 10\%$	$\leq 10\%E[\tau]$	$\geq 10\%E[\tau]$	$0 \leq T \leq 10\%E[\tau]$ $5\% \leq \alpha \leq 50\%$	$T \geq 10\%E[\tau]$ $5\% \leq \alpha \leq 50\%$
Offered Load						
Small (10's)	QED	QED	QED	QED	QED	QED
Moderate-to-Large (100's-1000's)	QED	ED, QED	QED	ED, QED if $\tau \stackrel{d}{=} \exp$	QED	ED+QED

**ED:**  $N \approx R - \gamma R \quad (0.1 \leq \gamma \leq 0.25).$

**QD:**  $N \approx R + \delta R \quad (0.1 \leq \delta \leq 0.25).$

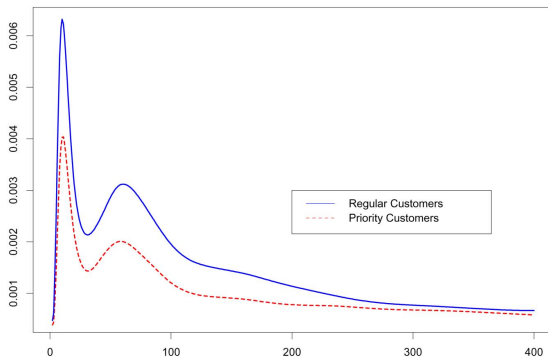
**QED:**  $N \approx R + \beta \sqrt{R} \quad (-1 \leq \beta \leq 1).$

**ED+QED:**  $N \approx (1 - \gamma)R + \beta \sqrt{R} \quad (\gamma, \beta \text{ as above}).$

**WFM:** How to determine specific staffing level  $N$  ? e.g.  $\beta$ .

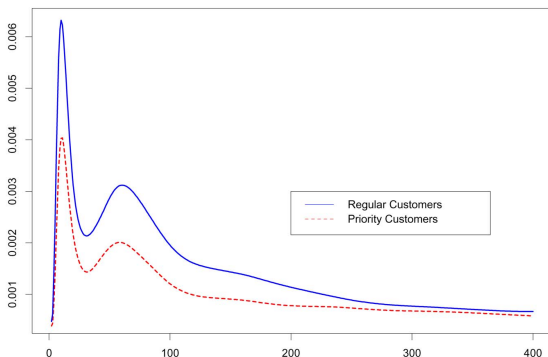
## (Im)Patience while Waiting (Palm 1943-53)

Irritation  $\propto$  Hazard Rate of (Im)Patience Distribution  
**Regular** over **VIP** Customers – Israeli Bank



## (Im)Patience while Waiting (Palm 1943-53)

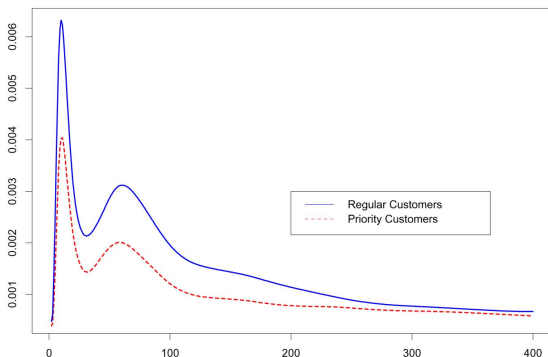
Irritation  $\propto$  Hazard Rate of (Im)Patience Distribution  
**Regular** over **VIP** Customers – Israeli Bank



- **Call-by-Call Data (SEELab) required (& Un-Censoring)**

## (Im)Patience while Waiting (Palm 1943-53)

Irritation  $\propto$  Hazard Rate of (Im)Patience Distribution  
**Regular** over **VIP** Customers – Israeli Bank

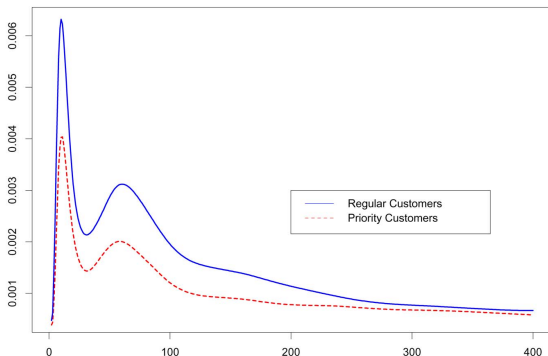


- **Call-by-Call Data (SEELab)** required (& **Un-Censoring**)
- **Peaks** of abandonment at times of **Announcements**



## (Im)Patience while Waiting (Palm 1943-53)

Irritation  $\propto$  Hazard Rate of (Im)Patience Distribution  
**Regular** over **VIP** Customers – Israeli Bank

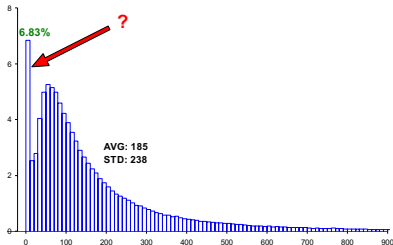


- ▶ **Call-by-Call Data (SEELab)** required (& **Un-Censoring**)
- ▶ **Peaks** of abandonment at times of **Announcements**
- ▶ **VIP** are **more patient** (Needy)

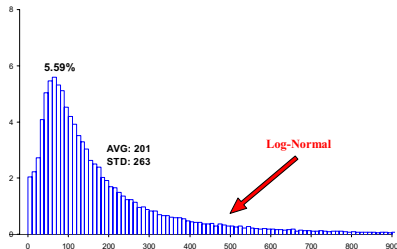
# Beyond Averages: The Human Factor

## Histogram of Service-Time in a (Small Israeli) Bank

### January-October



### November-December

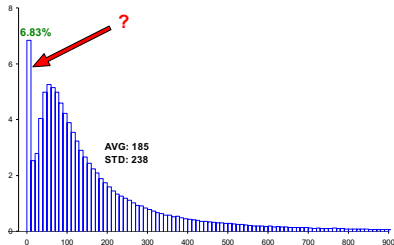


### ► 6.8% Short-Services:

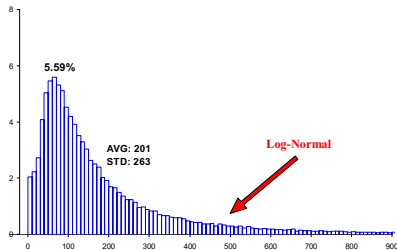
# Beyond Averages: The Human Factor

## Histogram of Service-Time in a (Small Israeli) Bank

### January-October



### November-December

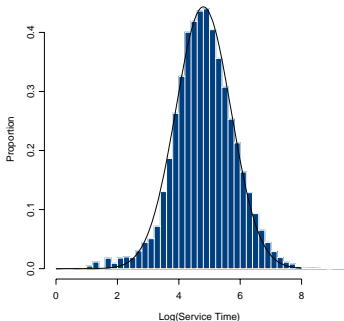


- ▶ **6.8% Short-Services:** Agents' "Abandon" (improve bonus, rest), lead by **incentives**
- ▶ **Distributions** must be measured (in seconds)
- ▶ **LogNormal** service times common in call centers

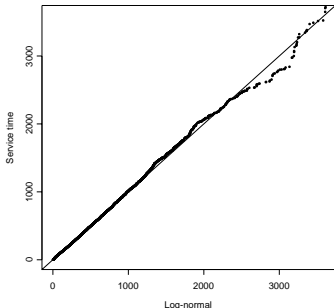
# Validating LogNormality of Service-Times

Israeli Call Center, Nov-Dec, 1999

Log(Service Times)

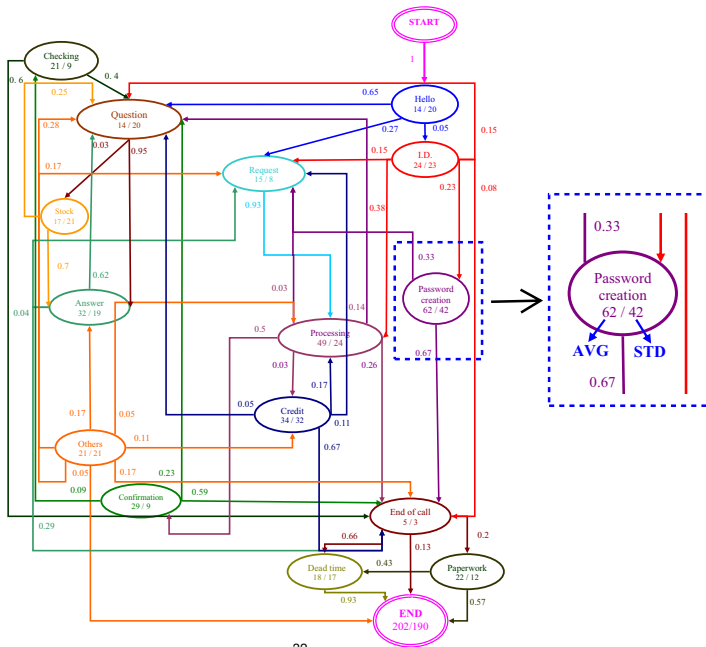


LogNormal QQPlot



- ▶ **Practically Important:** (mean, std)(log) capture Service-Times
- ▶ **Theoretically Challenging:** Why LogNormal ?
- ▶ Simple-model of a complex-reality? The **Service Process:**

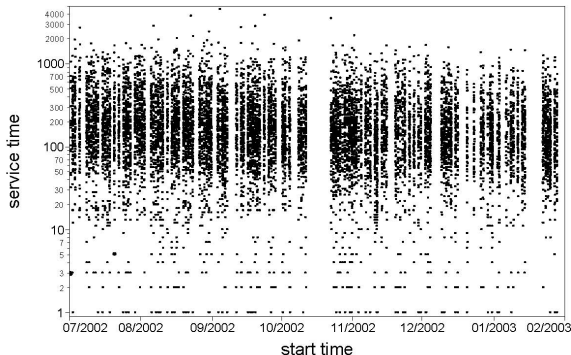
## Work Design (Time Study)



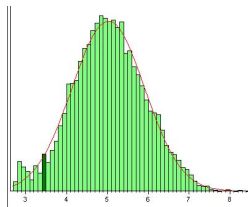
# Individual Agents: Service-Time, Variability

## Agent 14115

### Service-Time Evolution: 6 month



### Log(Service-Time)

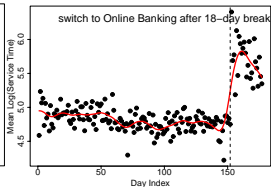
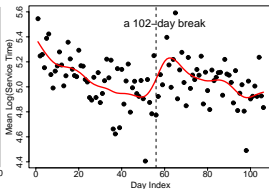
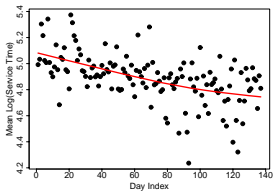


- ▶ **Learning**: Noticeable decreasing-trend in service-time
- ▶ **LogNormal** Service-Time, individually and collectively

# Individual Agents: Learning, Forgetting, Switching

Daily-Average Log(Service-Time), over 6 months

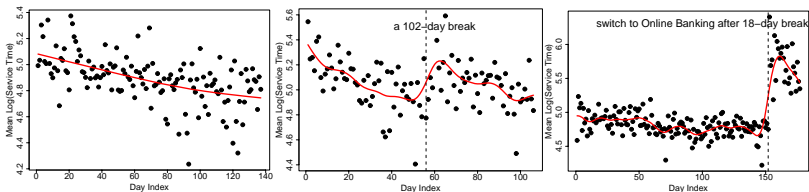
Agents 14115, 14128, 14136



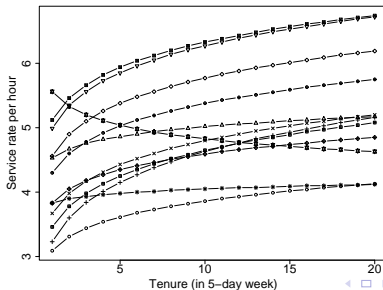
# Individual Agents: Learning, Forgetting, Switching

Daily-Average Log(Service-Time), over 6 months

Agents 14115, 14128, 14136



Weakly Learning-Curves for 12 Homogeneous(?) Agents





## Why Bother?

In large call centers:

**+One Second** to Service-Time implies **+Millions** in costs, annually

## Why Bother?

In large call centers:

**+One Second** to Service-Time implies **+Millions** in costs, annually

- ▶ **Classical IE with New-age Technology:**
  - ▶ Work Design, Time and "Motion" Studies (w/ **Khudiakov** (PhD))
  - ▶ "Worker" Design, Learning, Forgetting, . . . (w/ **Gans & Shen**)

# Why Bother?

In large call centers:

**+One Second** to Service-Time implies **+Millions** in costs, annually

- ▶ **Classical IE with New-age Technology:**

- ▶ Work Design, Time and "Motion" Studies (w/ **Khudiakov** (PhD))
- ▶ "Worker" Design, Learning, Forgetting, . . . (w/ **Gans & Shen**)

- ▶ **Service-Process Model** helps the bank:

- ▶ **Technology Management** - Old MIS system has slow response & cumbersome protocols, which gives rise to phases with little or no added-value: Justify replacement value
- ▶ **Cross-Selling** - Potentially more money at the cost of longer services: Justify value, which is congestion-dependent

# Why Bother?

In large call centers:

**+One Second** to Service-Time implies **+Millions** in costs, annually

- ▶ **Classical IE with New-age Technology:**
  - ▶ Work Design, Time and "Motion" Studies (w/ **Khudiakov** (PhD))
  - ▶ "Worker" Design, Learning, Forgetting, . . . (w/ **Gans & Shen**)
- ▶ **Service-Process Model** helps the bank:
  - ▶ **Technology Management** - Old MIS system has slow response & cumbersome protocols, which gives rise to phases with little or no added-value: Justify replacement value
  - ▶ **Cross-Selling** - Potentially more money at the cost of longer services: Justify value, which is congestion-dependent
  - ▶ **Learning**: Predict individual future performance, which is important in a high-turnover environment
  - ▶ **Heterogeneity**: Quantify operational consequences (WFM, SBR)

# Why Bother?

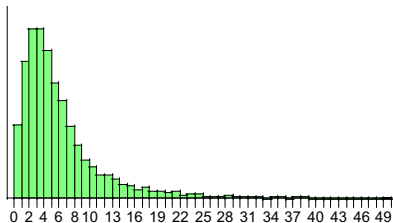
In large call centers:

**+One Second** to Service-Time implies **+Millions** in costs, annually

- ▶ **Classical IE with New-age Technology:**
  - ▶ Work Design, Time and "Motion" Studies (w/ **Khudiakov** (PhD))
  - ▶ "Worker" Design, Learning, Forgetting, . . . (w/ **Gans & Shen**)
- ▶ **Service-Process Model** helps the bank:
  - ▶ **Technology Management** - Old MIS system has slow response & cumbersome protocols, which gives rise to phases with little or no added-value: Justify replacement value
  - ▶ **Cross-Selling** - Potentially more money at the cost of longer services: Justify value, which is congestion-dependent
  - ▶ **Learning:** Predict individual future performance, which is important in a high-turnover environment
  - ▶ **Heterogeneity:** Quantify operational consequences (WFM, SBR)
- ▶ **IVR Process Model:** **75% services**, same method, easier data

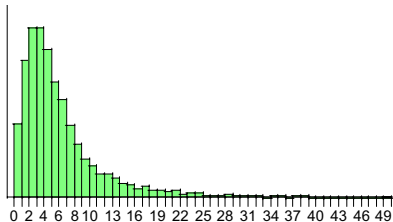
# Beyond Averages: Length-of-Stay in a Hospital

Israeli Hospital, in Days: LN

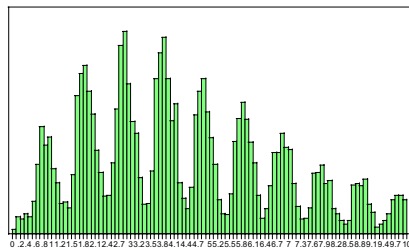


# Beyond Averages: Length-of-Stay in a Hospital

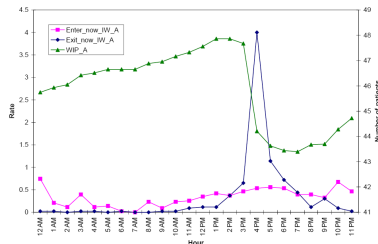
Israeli Hospital, in Days: LN



Israeli Hospital, in Hours



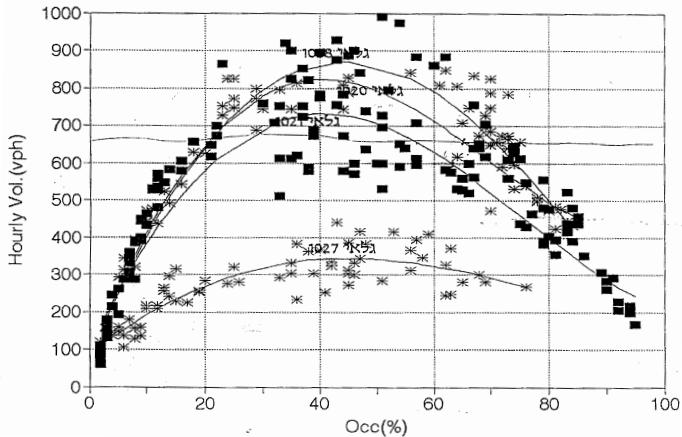
## Explanation: Releases around 3pm



# Transportation: Throughput (Flow) vs. Occupancy

Free-Flow → Critical-Occupancy → Congestion (Human)

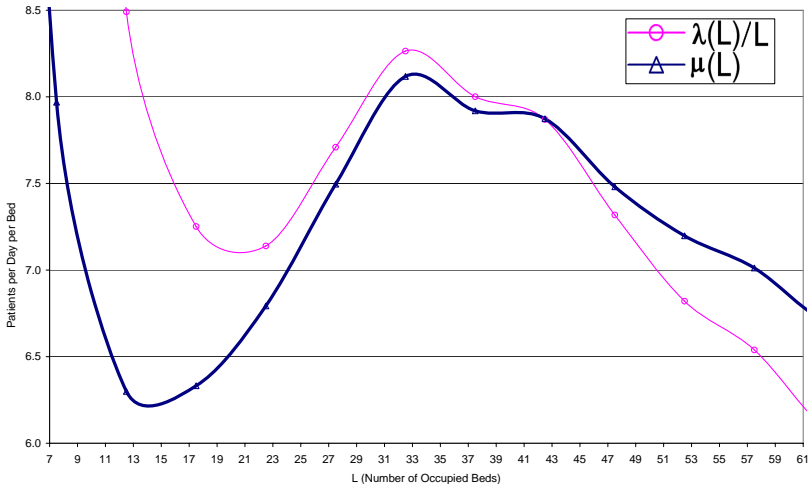
HERZEL - BALFUR  
KN010103-4 1020-1-7-8 27-28/9/93





# ED: Throughput (Flow) vs. Occupancy (Human)

## Congestion-Dependent Flow-Rates: Light, Regular, Heavy



## Empirical Analysis of an ED:

Y. Marmor (PhD), Y. Tseytlin (MSc), G. Yom-Tov (PhD), Mor Armony.

# The Technion SEE Center / Laboratory

## Data-Based Service Engineering

