Empirical Analysis of Service Centers: A Queueing-Science Perspective

Pablo Liberman, SEE center, Technion

pabloliberman@gmail.com

http://ie.technion.ac.il/Labs/Serveng/

(Joint work with Itai Gurvich, Avishai Mandelbaum and the Technion’s SEE center)

Contents

DataMOCCA

Data MOdels for Call Center Analysis

Project Partners:

Technion: **Paul Feigin, Avi Mandelbaum, Sergey Zeltyn (IBM Research)**

SEElab: **Valery Trofimov, Ella Nadjharov, Igor Gavako, Katerina Kutsyy**

Students, RA's

Wharton: **Larry Brown, Noah Gans, Haipeng Shen (N. Carolina), Linda Zhao**

Students, Wharton Financial Institutions Center

Companies: **US Bank, IL Telecom (both Aspect-Based), IL Bank (Genesys), …**

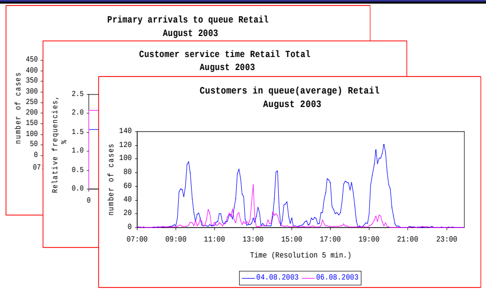
Project DataMOCCA

<u>Goal</u>: Designing and Implementing a (universal) data-base/data-repository and interface for storing, retrieving, analyzing and displaying Call-by-Call-based data/information

<u>Enables the Study of</u>:

- Customers	Wait Time, Abandonment, Retrials
- Service Providers / Agents	Service Duration, Activity Profile
- Managers / System	Loads, Queue Lengths, Trends

DataMOCCA: Present



System Components:

1. Clean databases: operational histories of individual calls, agents and sometimes customers (ID's).
2. SEEStat (friendly powerful interface): online access to (mostly) operational and (some) administrative data; flexible customization, statistics

Currently Four Databases:

- US Bank (220/40M calls, 1000 agents, 2.5 years;7-48GB; 3GB DVD).
- Israeli Telecom (800 agents, 3.5 years; 55GB; ongoing).
- Israeli Bank (500 agents, 1.5 years; ongoing).
- Israeli Hospital (1000 beds, 45 medical units, about 75000 patients hospitalized yearly , 4.9 years; ongoing)

DataMOCCA Interface: SEEStat

- Daily / monthly / yearly reports & flow-charts for a complete operational view.

- Graphs and tables, in customized resolutions (month, days, hours, minutes, seconds) for a variety of (pre-designed) operational measures (arrival rates, abandonment counts, service- and wait-time distribution, utilization profiles,…).

- Graphs and tables for new user-defined measures.

- Direct access to the raw (cleaned) data: export, import.

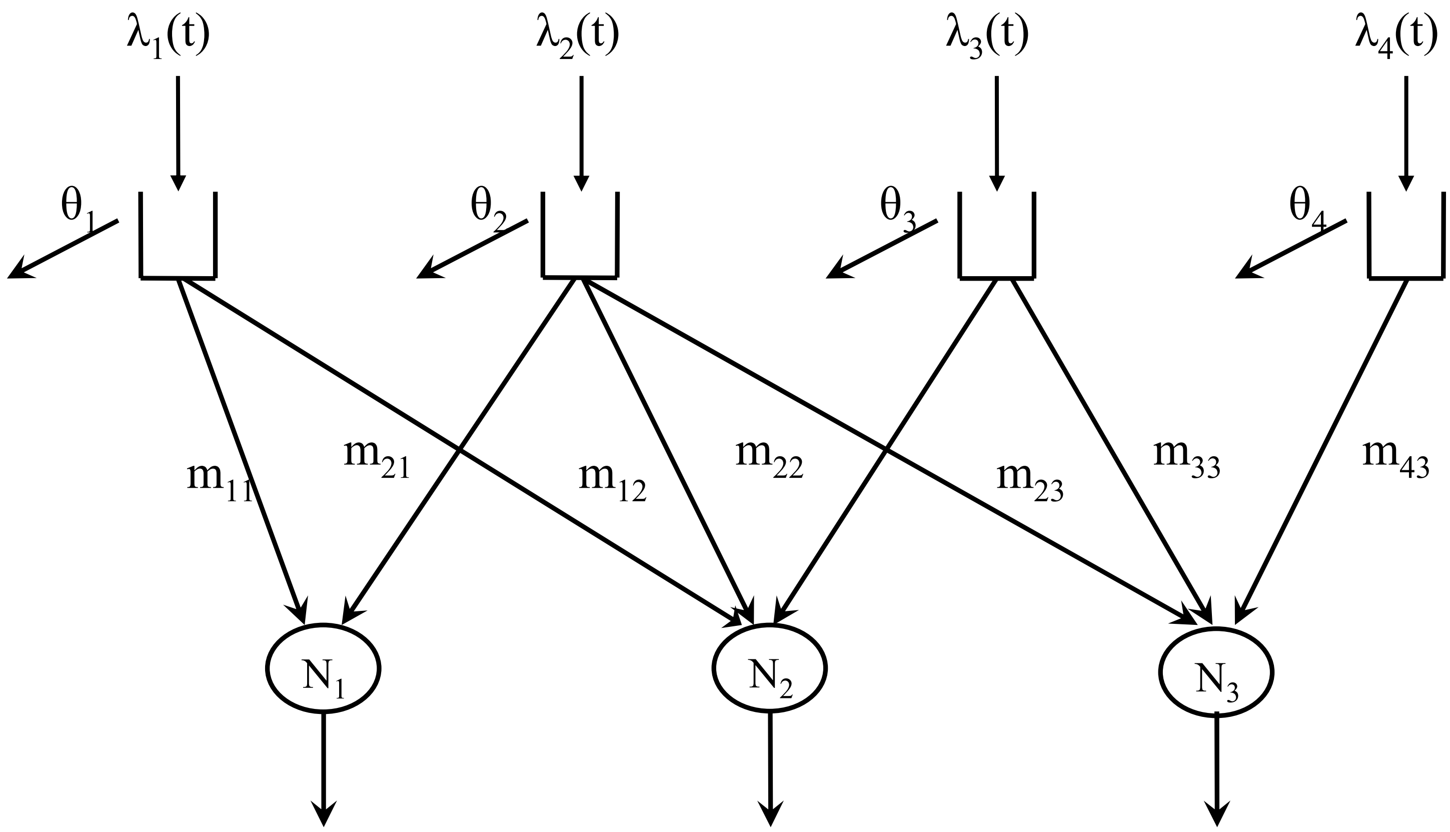
- Web-access (Research, Applications): http://ie.technion.ac.il/Labs/Serveng/

DataMOCCA History: Research & Teaching

- Large well-run call centers beyond conventional Queueing Theory:
  - Both Quality and Efficiency Driven (vs. tradeoff)
  - Multi-Disciplinary view: OR/OM, HRM (Psychology), Marketing, MIS
- Research: Asymptotic analysis of the Palm/Erlang-A model and beyond, in the Halfin-Whitt regime = QED Regime
- Teaching: Service Management + Industrial Engineering = Service Engineering / Science
- Technion IE&M core course: http://ie.technion.ac.il/serveng/
- Mini courses (INSEAD, Warton, Columbia, Stanford)
- PhD, MSc; Projects

Empirical Research - Basics

Workforce Management in Multi-Class Settings



- **Primitives:** arrivals, service time, patience
- **Decisions:** Operational objectives, routing rules, design, staffing and scheduling.

Starting Point: A Staffing Problem Formulation

- i customer classes
- j agent groups (skills)

$$
\begin{aligned}
&\text{minimize } \sum_{j=1}^{J} c_j N_j \\
&s.t. \qquad \mathrm{P}\{W_i(N,\pi) > T_i\} \le \alpha_i,\ i = 1,\dots,I, \\
&\qquad\qquad \mathrm{P}\{Ab_i\} \le \beta_i,\ i = 1,\dots,I, \\
&\qquad\qquad \mathrm{N} \in \mathbb{N}_+^J,\ \pi \in \Pi
\end{aligned}
$$

Variants:

- Time varying, uncertain arrival rates
- Alternative QoS constraints and time resolution

Inputs: arrivals, service times, patience, operational objectives

Output: Routing rule and staffing levels

Data Sources – Transactional Data

US Bank	Total	Average per Weekday
Total # of Arriving Calls	218,047,488	271,006
# Requesting Agents Service	41,646,142	55,445

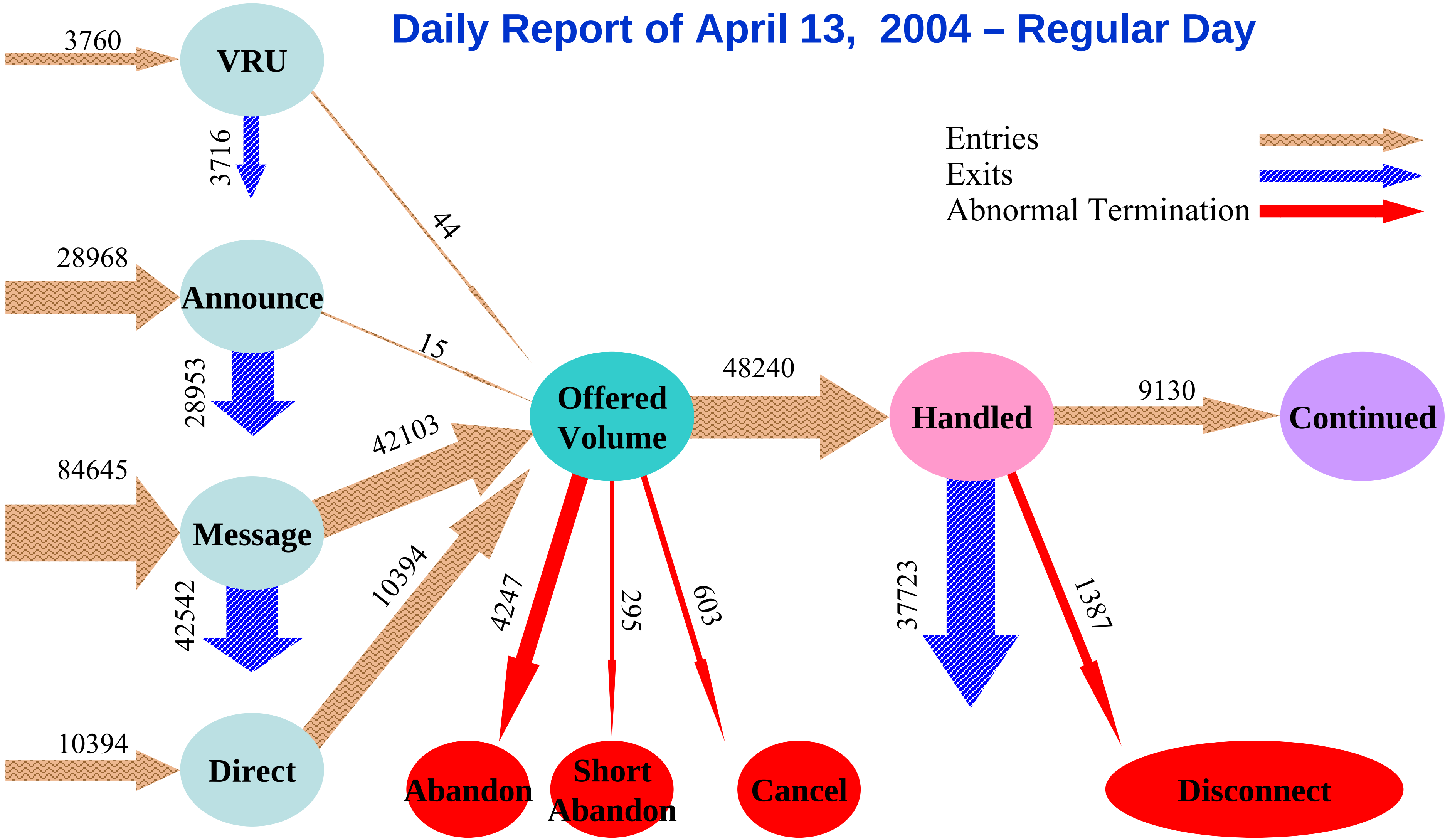
IL Bank	Total	Average per Weekday
Total # of Arriving Calls	25,990,281	90,696
# Requesting Agents Service	9,041,425	31.987

IL Tel	Total	Average per Weekday
Total # of Arriving Calls	112,488,410	224,976
# Requesting Agents Service	26,464,119	52,928

Arrivals (Demand)

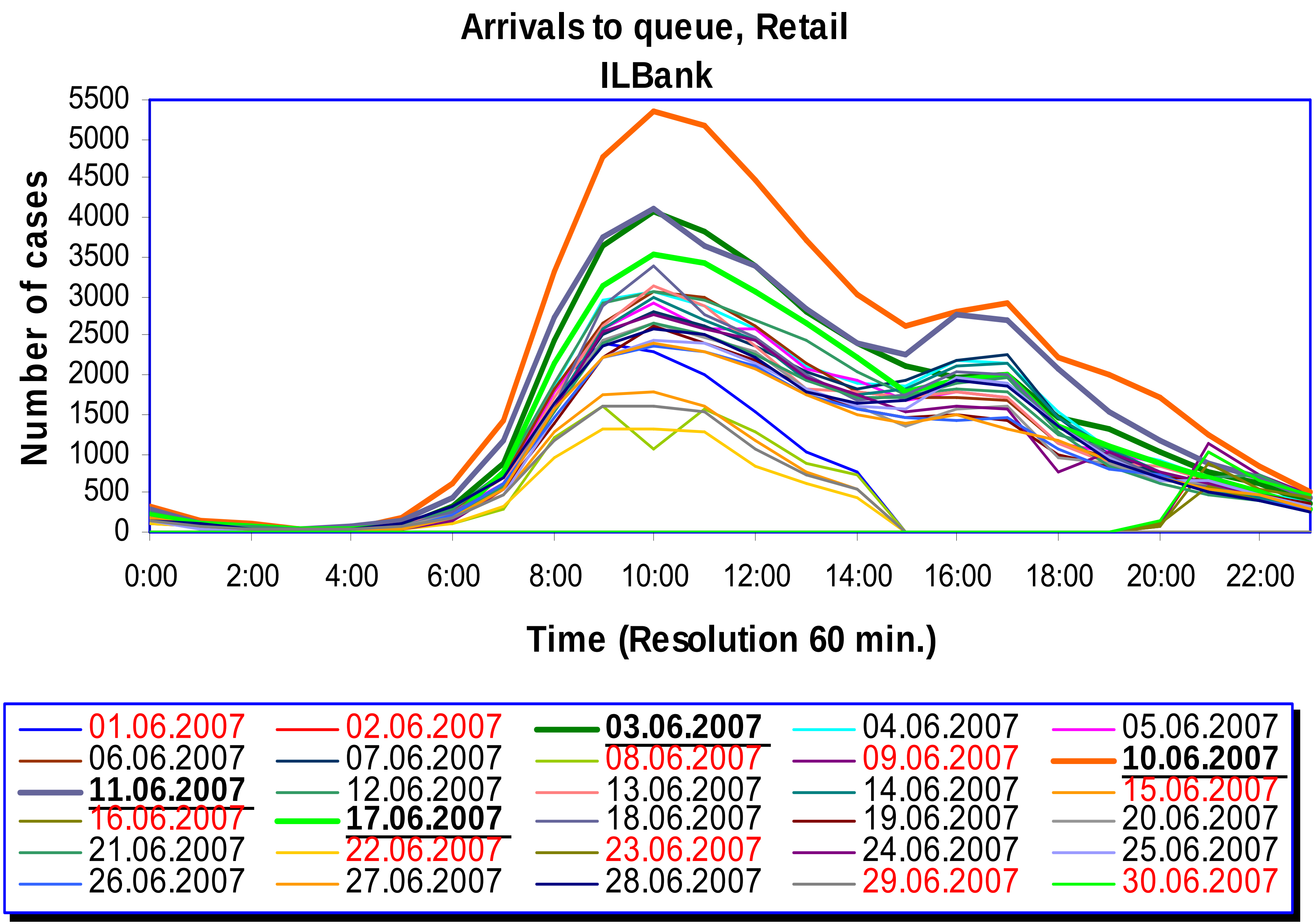
- **Flow Chart Examples**
- Unpredictable and Predictable Variability
- Over Dispersion Model
- Unpredictable Variability: The Multi-Class Case

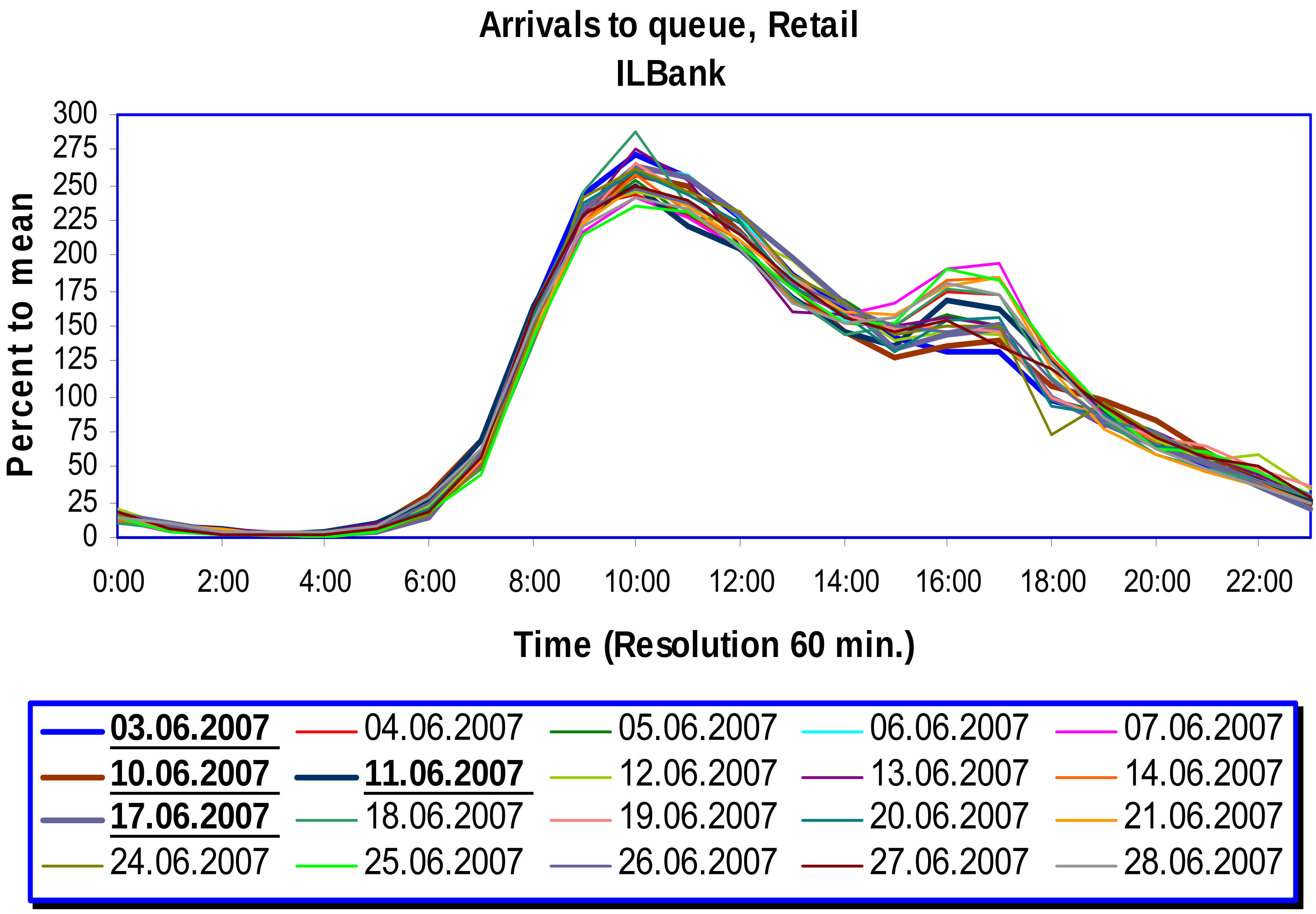
Daily Report- Flow Chart



Arrivals (Demand)

- Flow Chart Examples
- **Unpredictable and Predictable Variability**
- Over Dispersion Model
- Unpredictable Variability: The Multi-Class Case


Arrivals to queue, Retail
ILBank
Number of cases
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0:00
2:00
4:00
6:00
8:00
10:00
12:00
14:00
16:00
18:00
20:00
22:00
Time (Resolution 60 min.)
01.06.2007
02.06.2007
03.06.2007
04.06.2007
05.06.2007
06.06.2007
07.06.2007
08.06.2007
09.06.2007
10.06.2007
11.06.2007
12.06.2007
13.06.2007
14.06.2007
15.06.2007
16.06.2007
17.06.2007
18.06.2007
19.06.2007
20.06.2007
21.06.2007
22.06.2007
23.06.2007
24.06.2007
25.06.2007
26.06.2007
27.06.2007
28.06.2007
29.06.2007
30.06.2007


Arrivals to queue, Retail
ILBank
Percent to mean
300
275
250
225
200
175
150
125
100
75
50
25
0
0:00
2:00
4:00
6:00
8:00
10:00
12:00
14:00
16:00
18:00
20:00
22:00
Time (Resolution 60 min.)
03.06.2007
04.06.2007
05.06.2007
06.06.2007
07.06.2007
10.06.2007
11.06.2007
12.06.2007
13.06.2007
14.06.2007
17.06.2007
18.06.2007
19.06.2007
20.06.2007
21.06.2007
24.06.2007
25.06.2007
26.06.2007
27.06.2007
28.06.2007

Brown et. al. (05'), and others:

Class-i arrivals doubly-stochastic Poisson with instantaneous rate

$$\Lambda_i(t) = \Lambda_i \ \lambda_i(t) \qquad\qquad \int_0^T \lambda_i(t)\,dt = 1$$

Λ_i = total arrivals [Unpredictable Variability]

$\lambda_i(t)$ = the "shape" of weekday [Predictable Variability]

Arrivals (Demand)

- Flow Chart Examples
- Unpredictable and Predictable Variability
- **Over Dispersion Model**
- Unpredictable Variability: The Multi-Class Case

A Model for Call Arrivals - Over Dispersion Model

- The standard assumption is that the arrival process is Poisson with known parameters.

- In practice, however, we observe larger variability of the arrival process than we expected from the Poisson hypothesis.

Random Arrival Rate Model:

$$\Lambda = \lambda + \lambda^{c} \cdot X, \qquad c \leq 1,$$

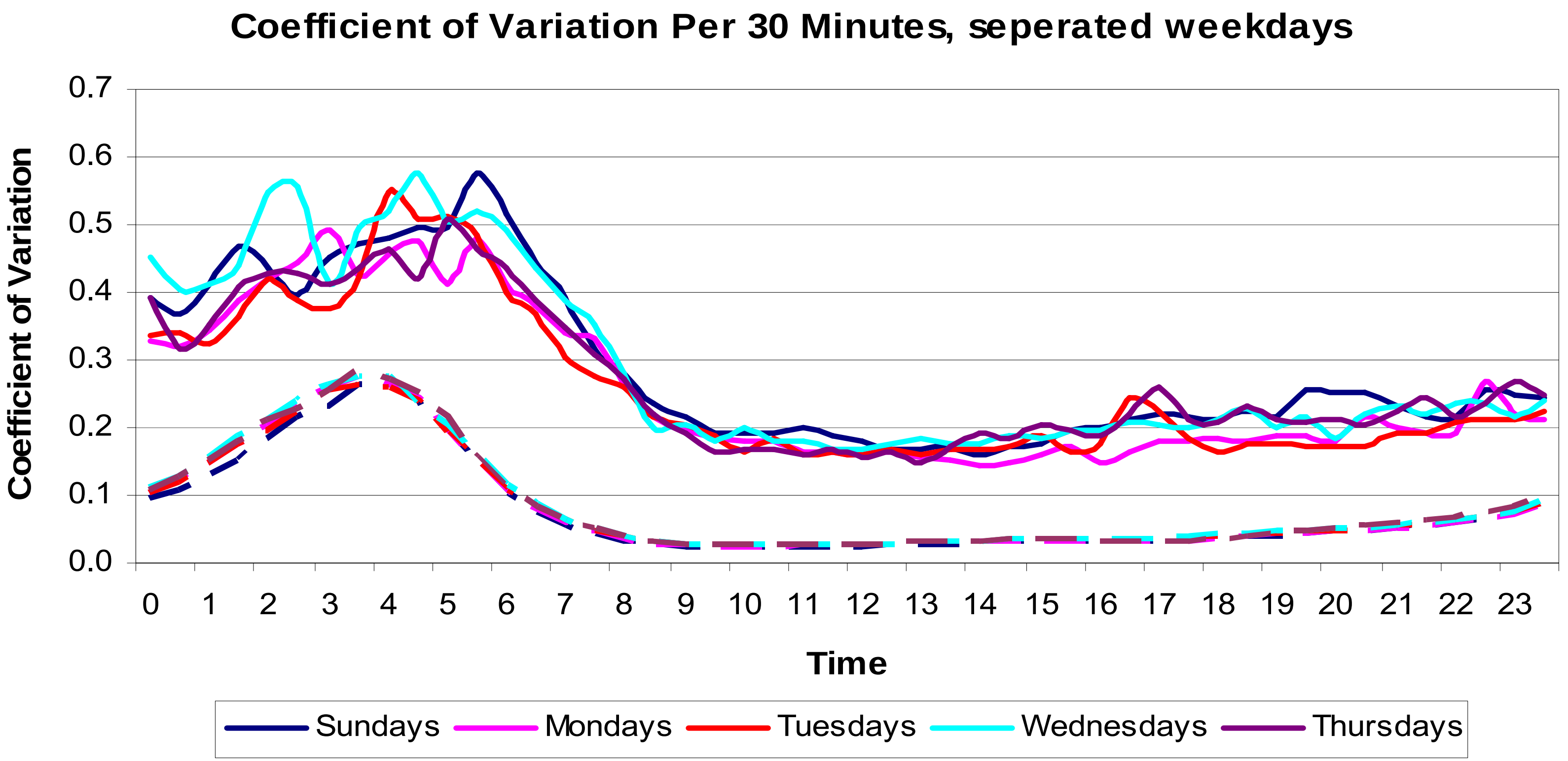
- X is a random variable with E[X] = 0, which captures the magnitude of the stochastic deviation from the expected arrival rate.

- c determines the order of over dispersion.

[Maman S., M.Sc. Thesis (09')]

Example 1: ILBank

- Arrival counts to the Retail queue are studied.
- 263 regular weekdays ranging between April 2007 and April 2008.
- Each day is divided into 48 half-hour intervals.



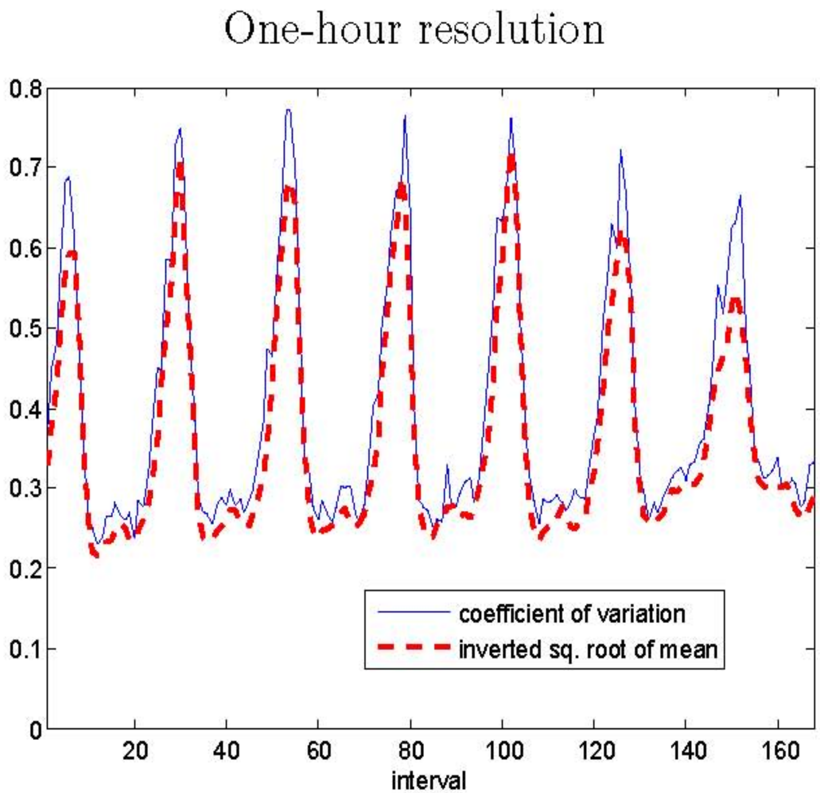
Poisson CV = $1/\sqrt{\text{mean arrival rate}}$

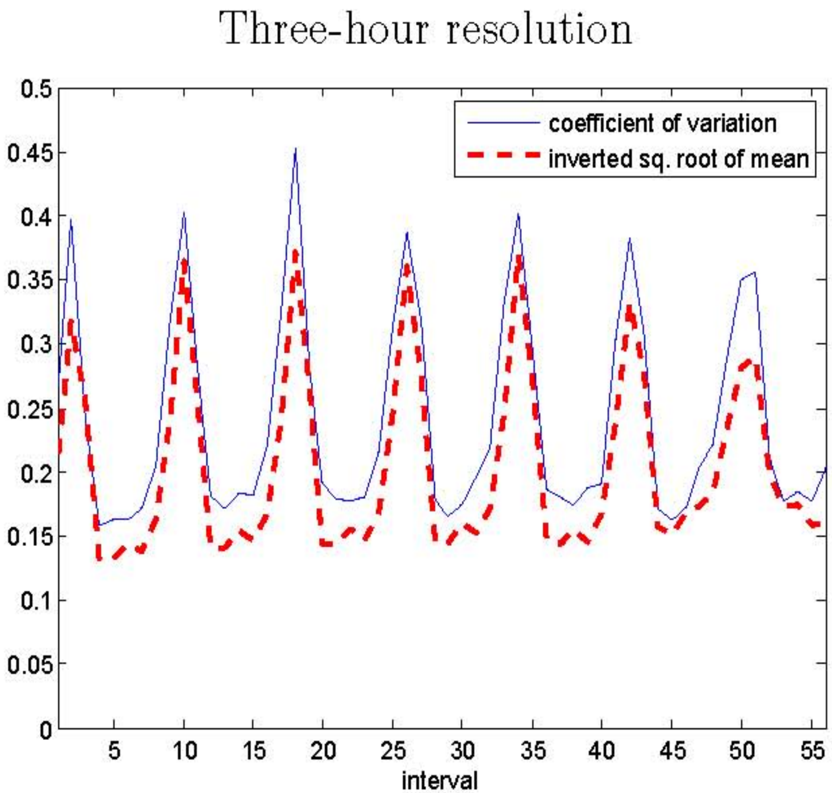
Sampled CV's >> Poisson CV's ⇨ Over-Dispersion

The over-dispersion is of order. $\lambda^c,\ c \approx 0.8 - 0.85$

Example 2: ILHospital - Emergency Department

- 194 weeks between from January 2004 till October 2007 (five war weeks are excluded from data).





Moderate over-dispersion.
The over-dispersion is of order: $\lambda^c,\ c \approx 0.5$
Less variability than in call centers.

Arrivals (Demand)

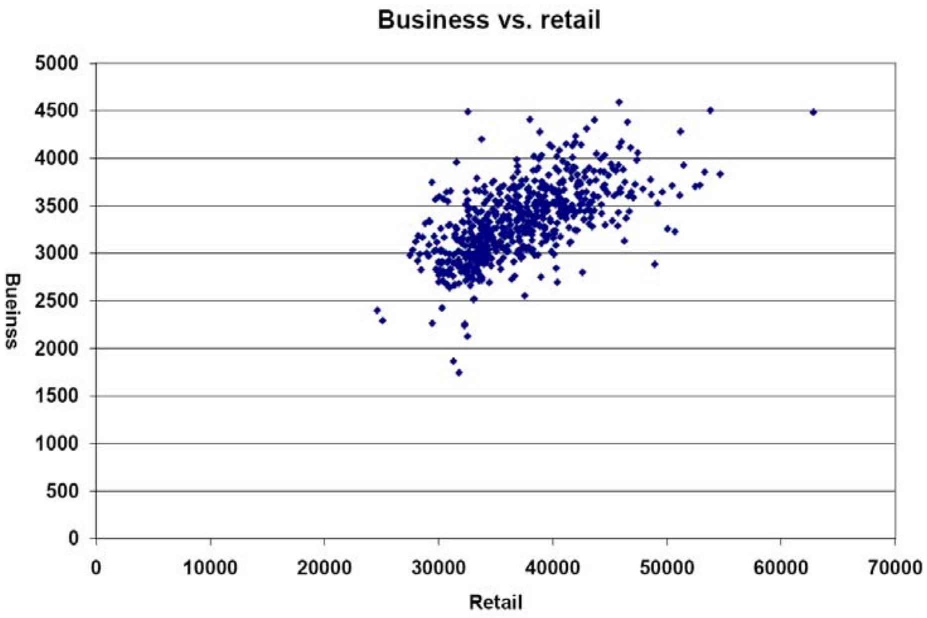
- Flow Chart Examples
- Unpredictable and Predictable Variability
- Over Dispersion Model
- **Unpredictable Variability: The Multi-Class Case**

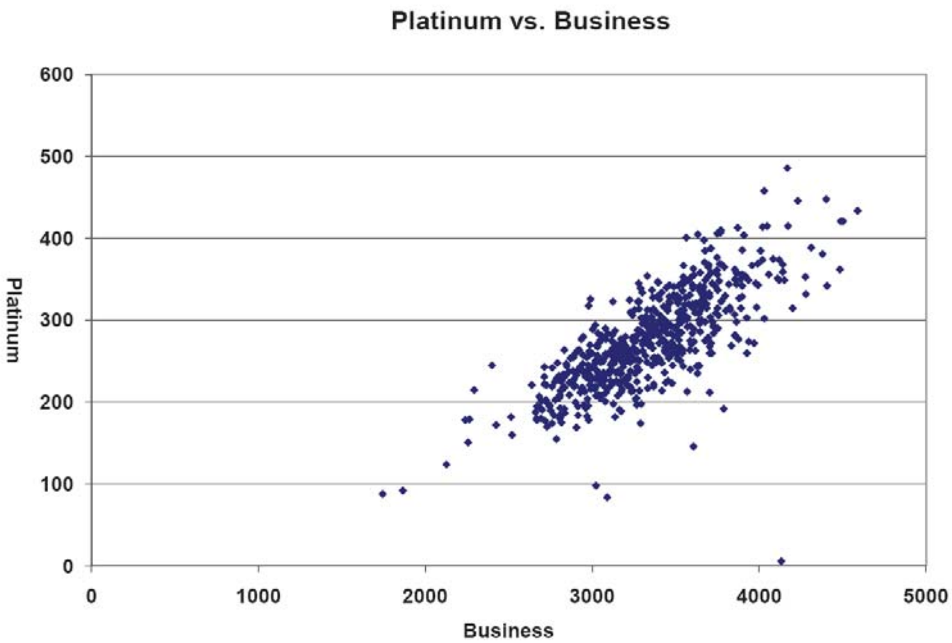
Unpredictable Variability: The Multi-Class Case

Unpredictable variability: $\Lambda = (\Lambda_1, \ldots, \Lambda_I)$

Pairs: $(\Lambda_{\text{Retail}}, \Lambda_{\text{Business}})$ and $(\Lambda_{\text{Business}}, \Lambda_{\text{Platinum}})$

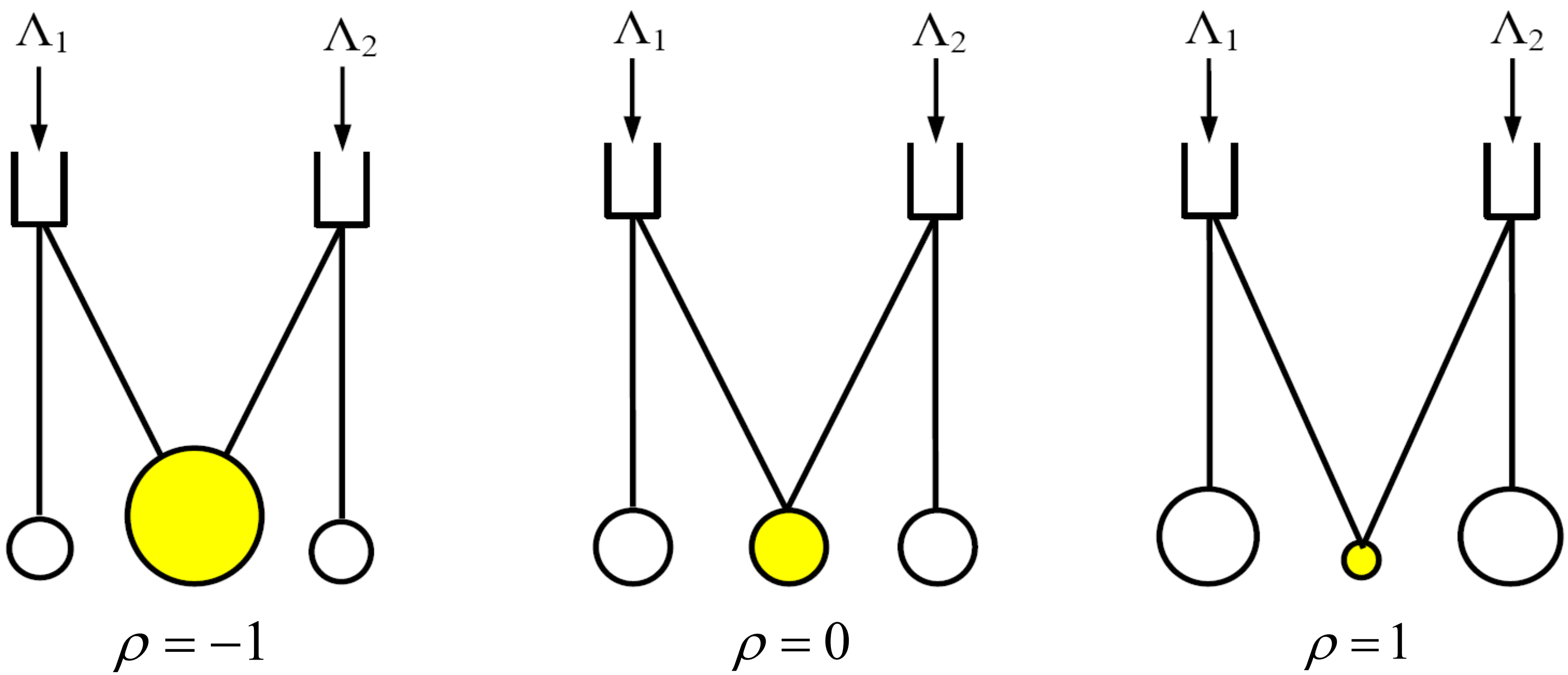
USBank data, Arrivals correlations, 600 weekdays





- Positive correlation (not independent)
- Impact on design decisions?

Unpredictable Variability - Impact on Design



Flexibility is less beneficial as correlation increases

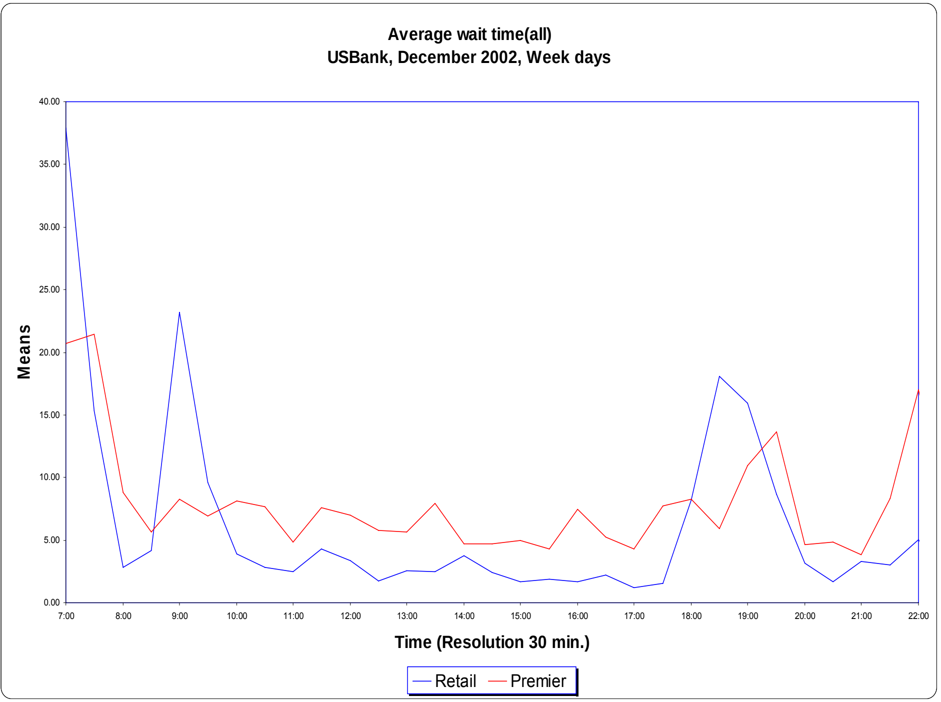
e.g. Van Mieghem (98’): Newsvendor context

Waiting Time

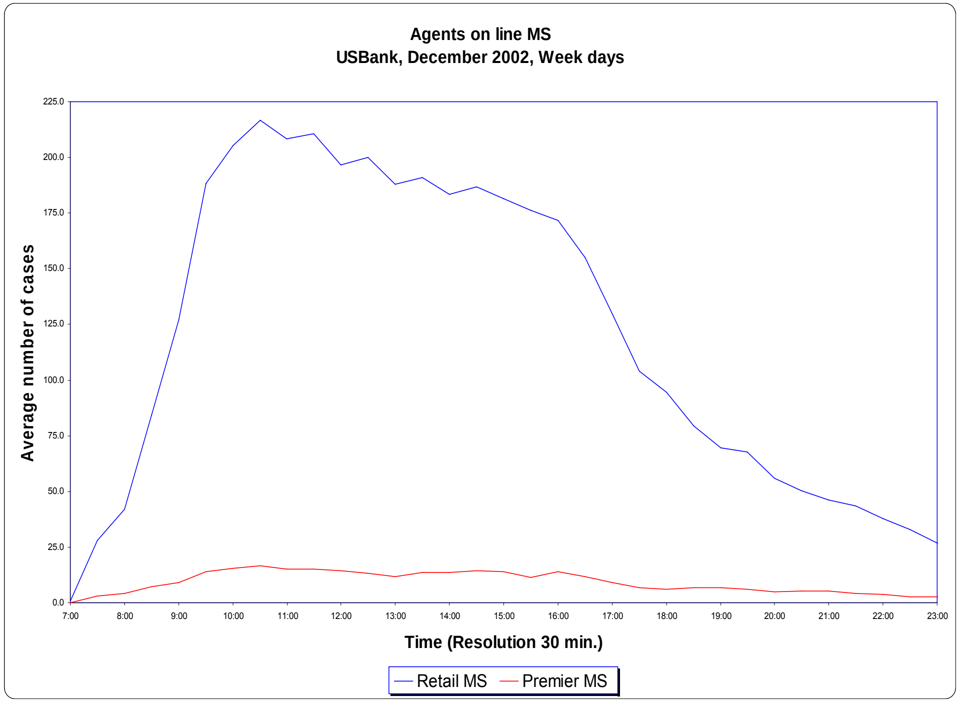
Waiting Time: Priorities and Routing Protocols

U.S. Bank Regular vs. VIP Customers

Average Waiting Time



Staffing Level

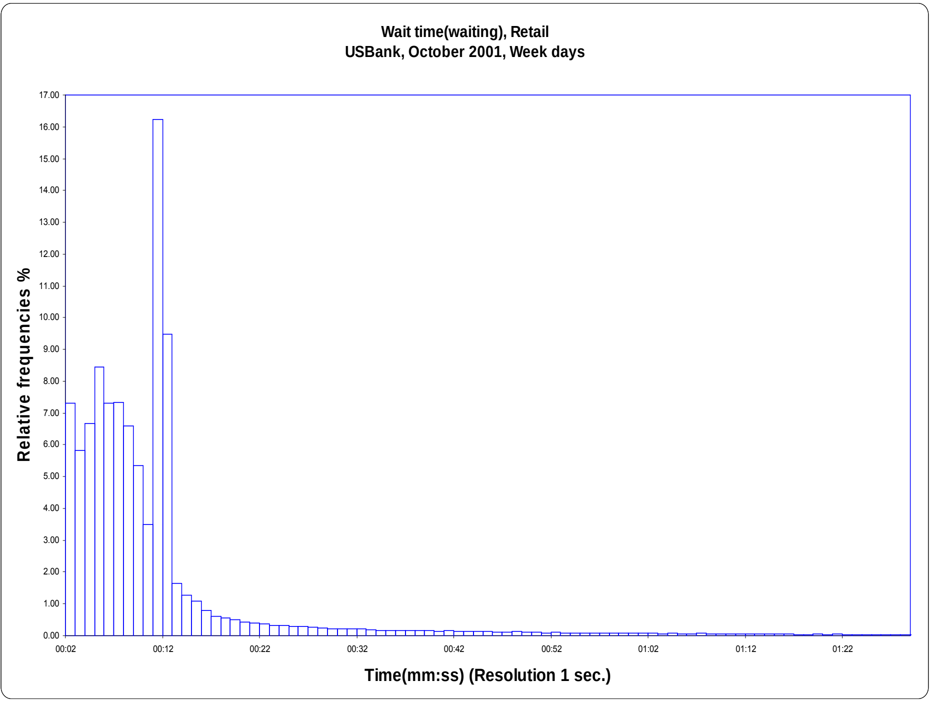


Premier customers do not get a better service level.

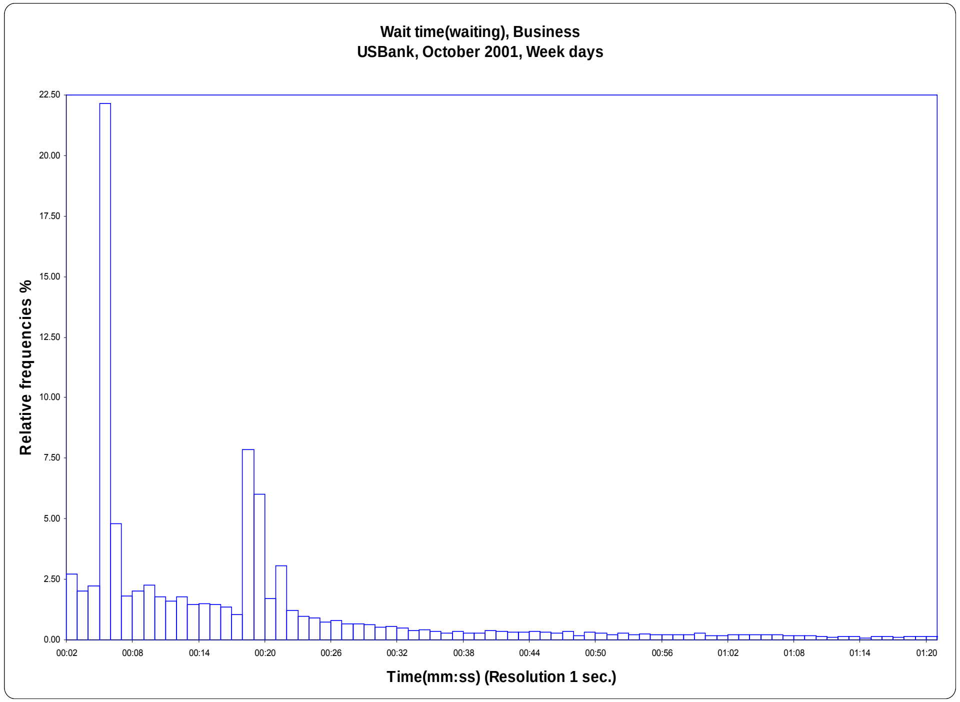
Number of agents assigned to Premier is small and they do not get enough help from regular agents.

U.S. Bank, Histogram of Waiting Times, October 2001

Retail



Business

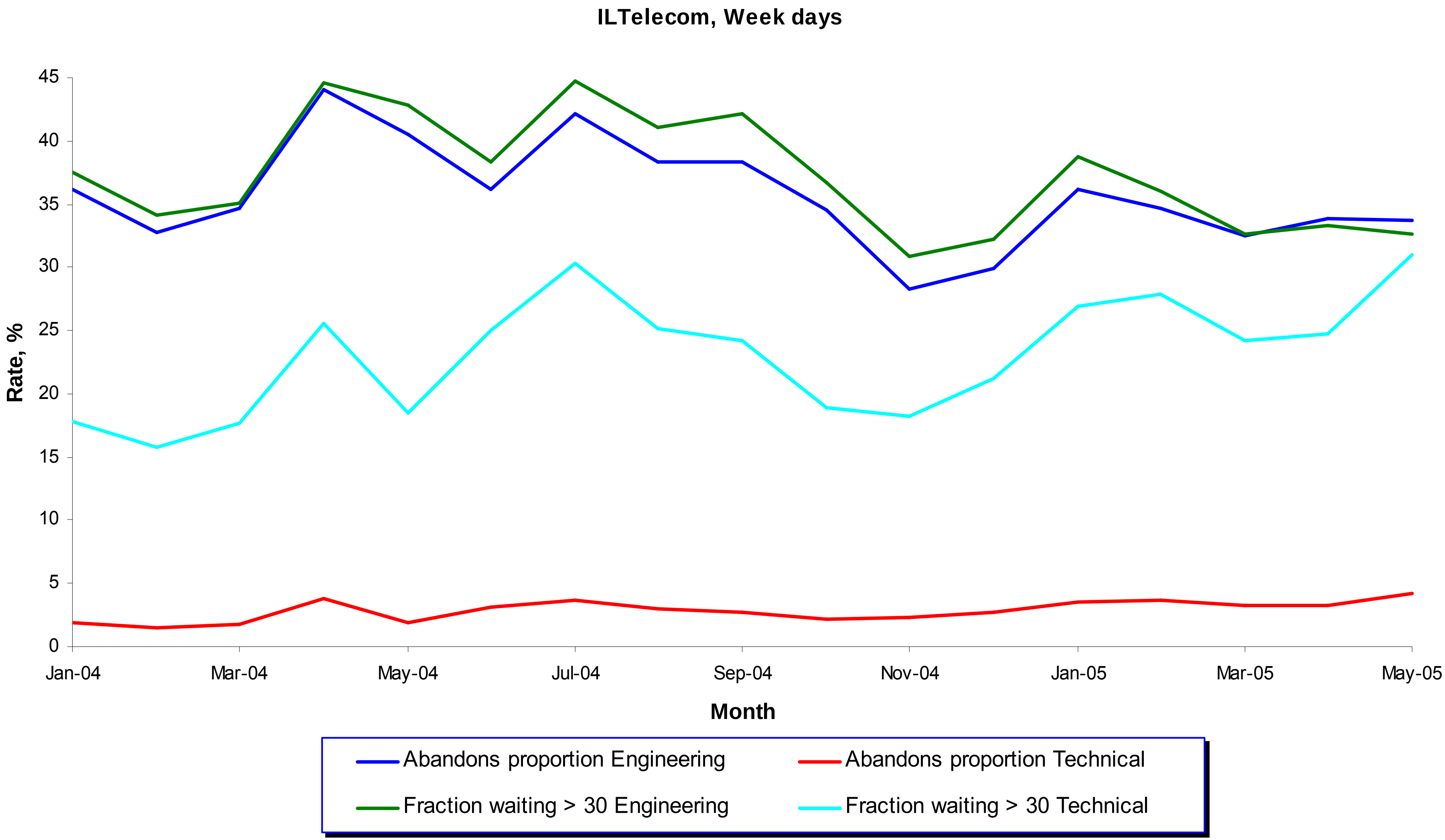


Why do we observe a peak for **Retail** service (10 seconds)?
After 10 seconds of wait, **Retail** customers sent to interqueue.

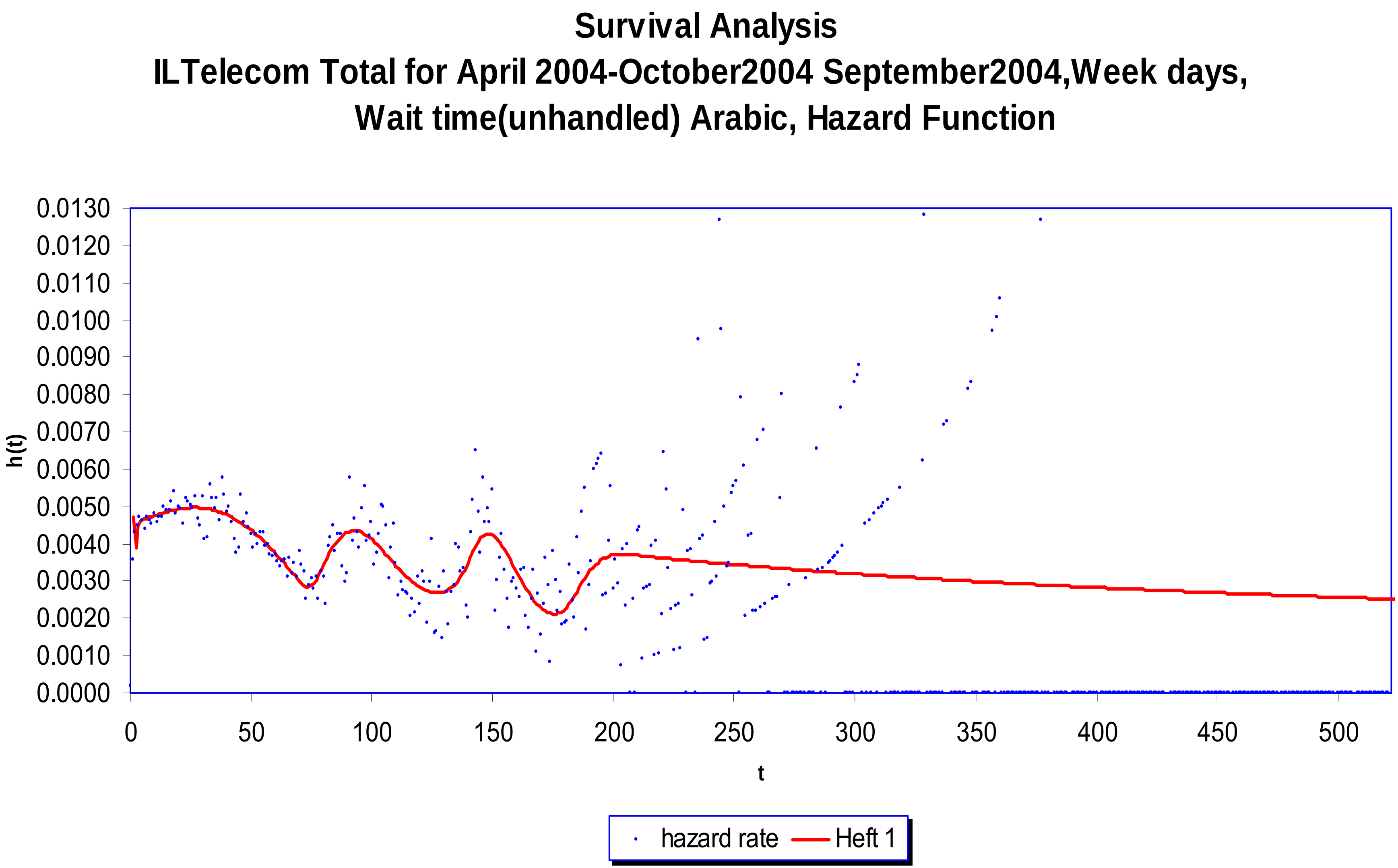
Business customers – peak at 5 seconds, for the same reason.
Second peak – unclear, maybe priority- upgrade.

Abandonment (Impatience)

Pre-designed Operational Measures: %Abandonment, TSF



Patience Time: Hazard Rate

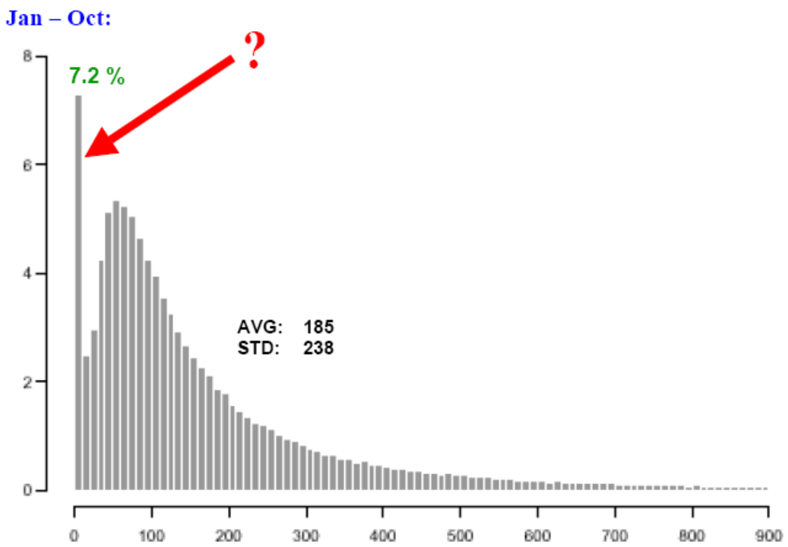


Service Time (Process)

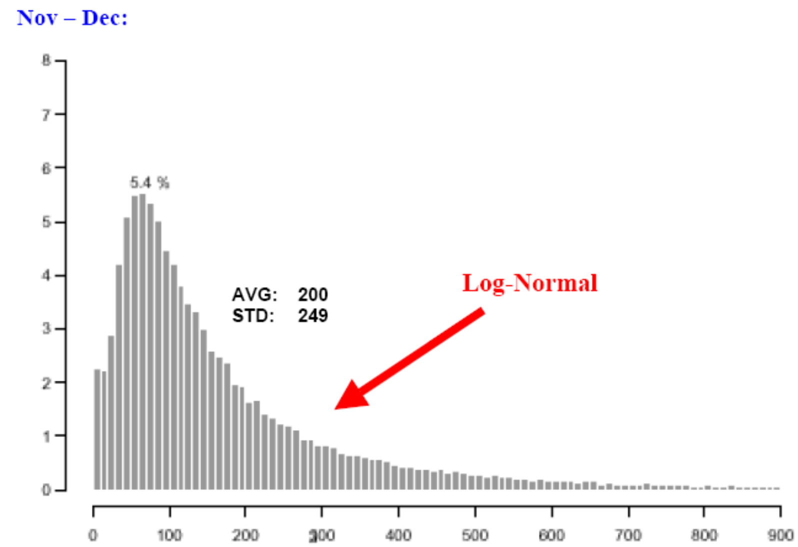
Service Time: Beyond Averages (+Human Factor)

Histogram of Service Times in an Israeli Call Center

January-October

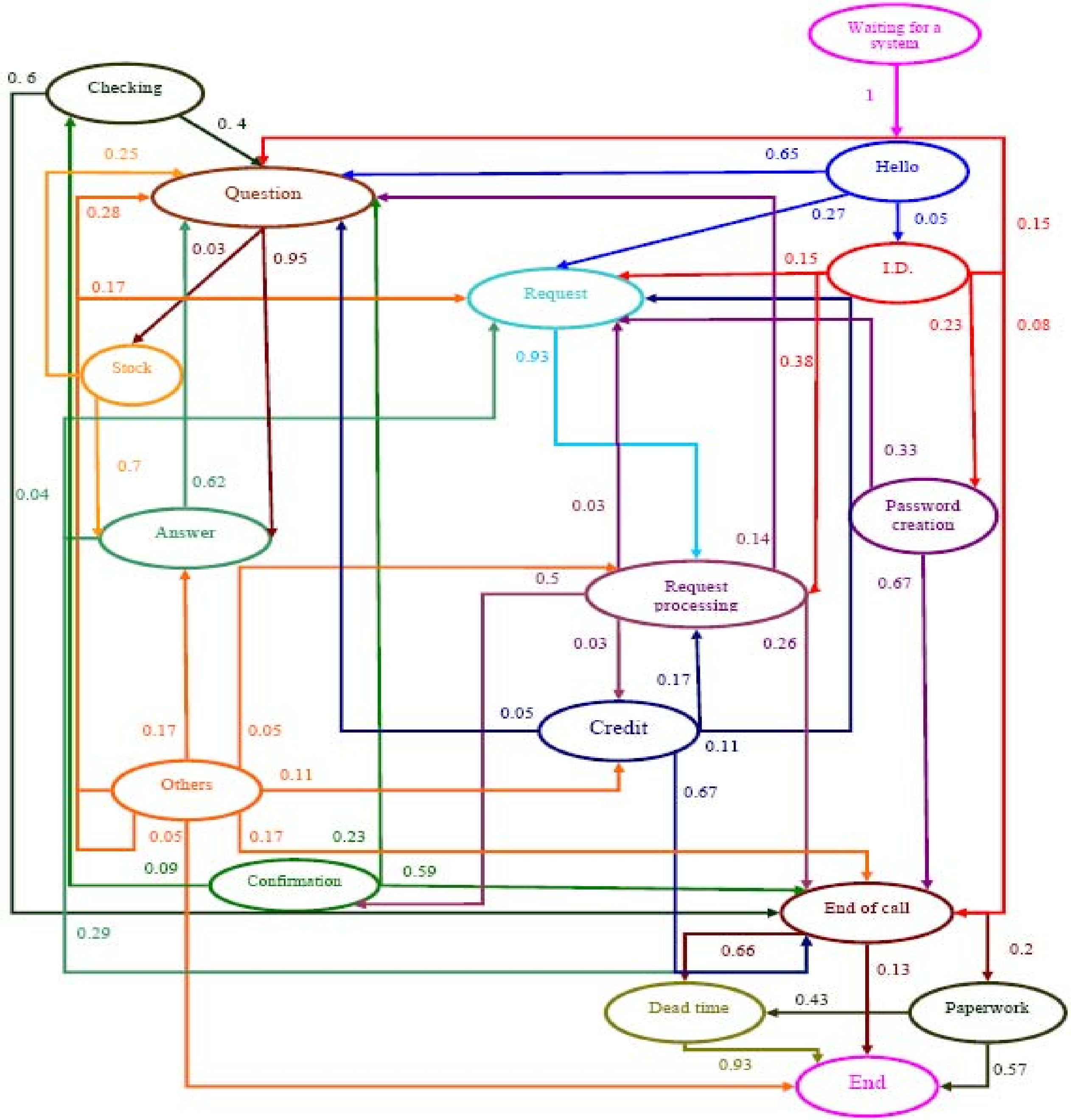


November-December



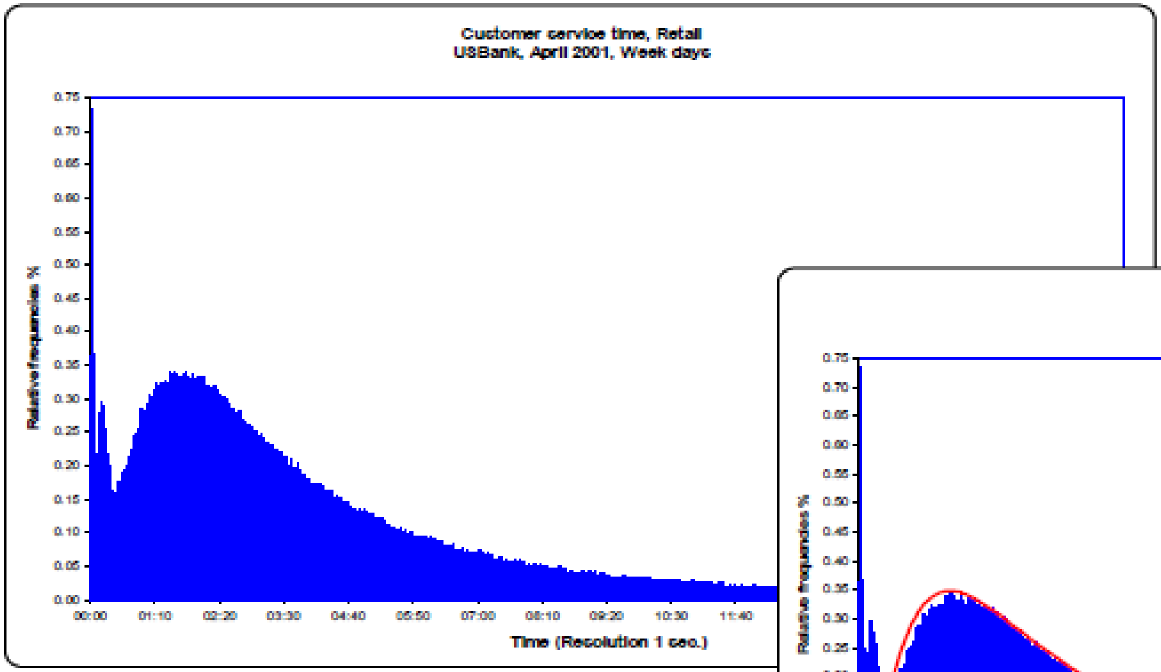
- 7.2% Short-Services: Agents' "abandon" (improve bonus, rest)
- Distributions, not only Averages, must be measured (seconds)
- Lognormal service times prevalent in call centers (Why?)

Service Time Process- Flow Chart Diagram

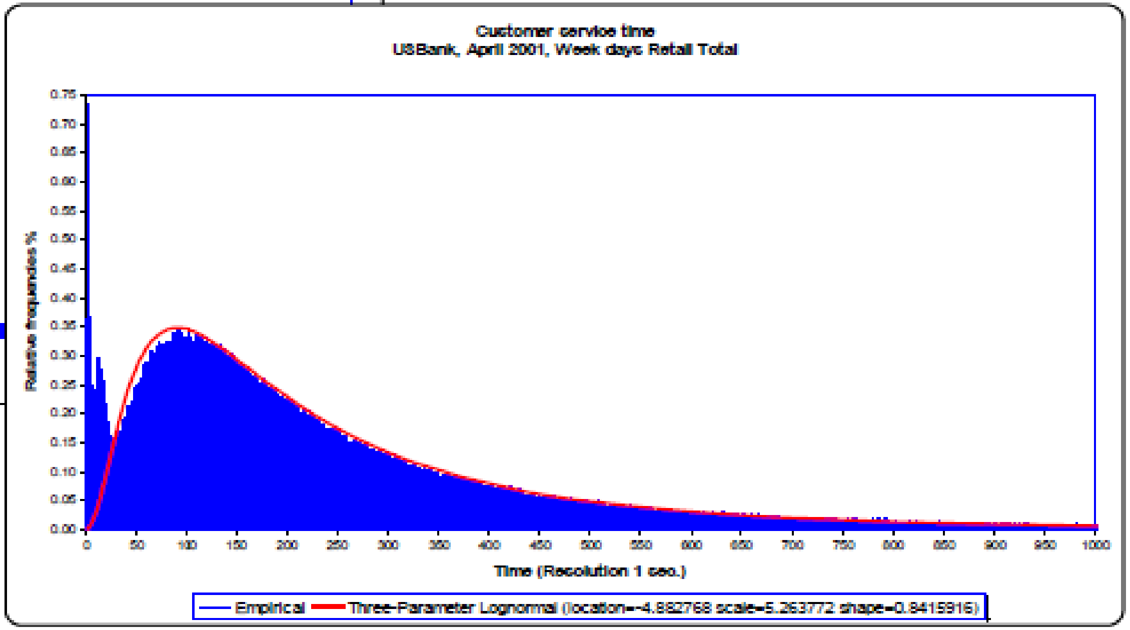


Service Time: Fitting Distribution

Histograms

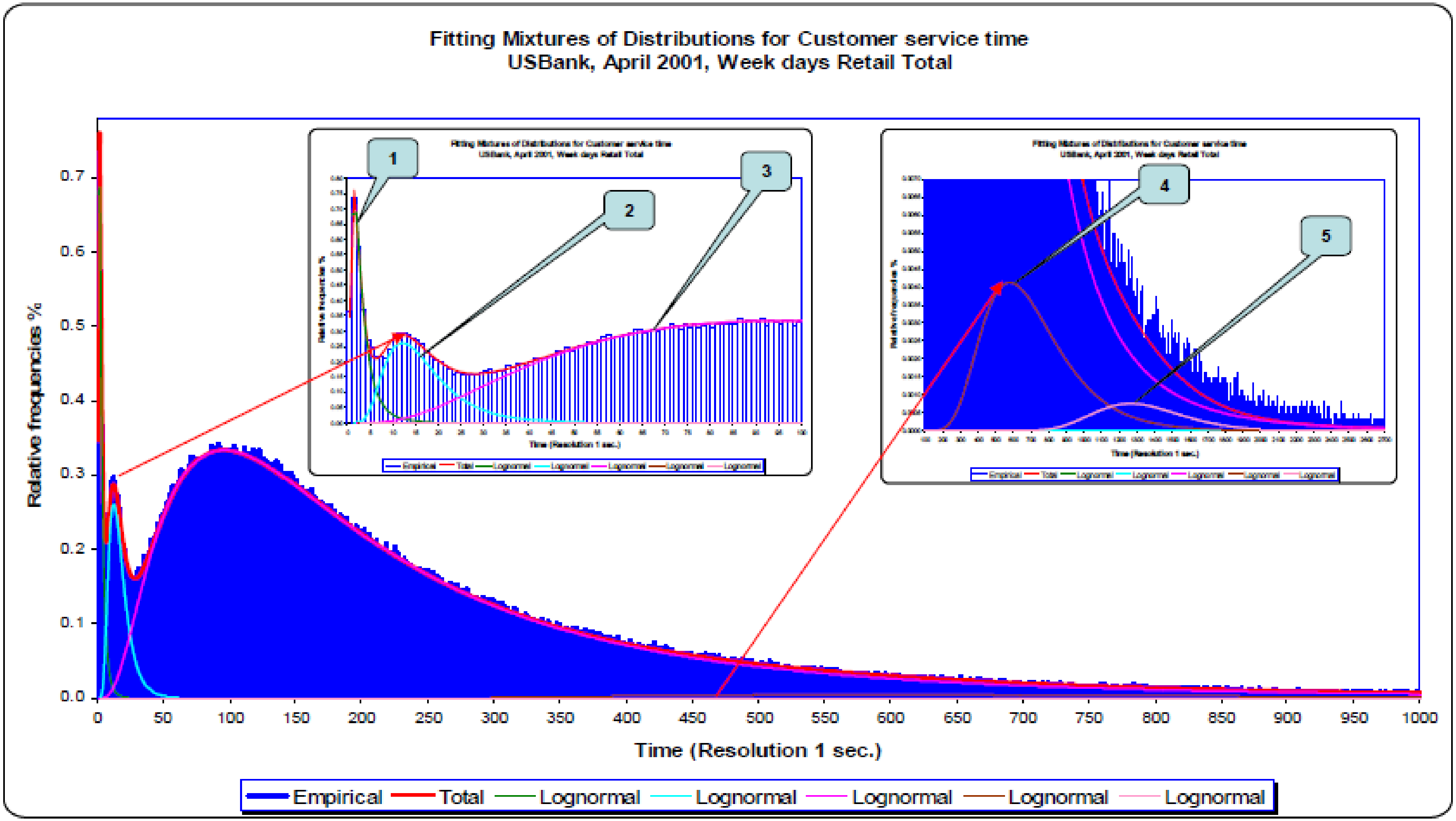


Fitting Distribution



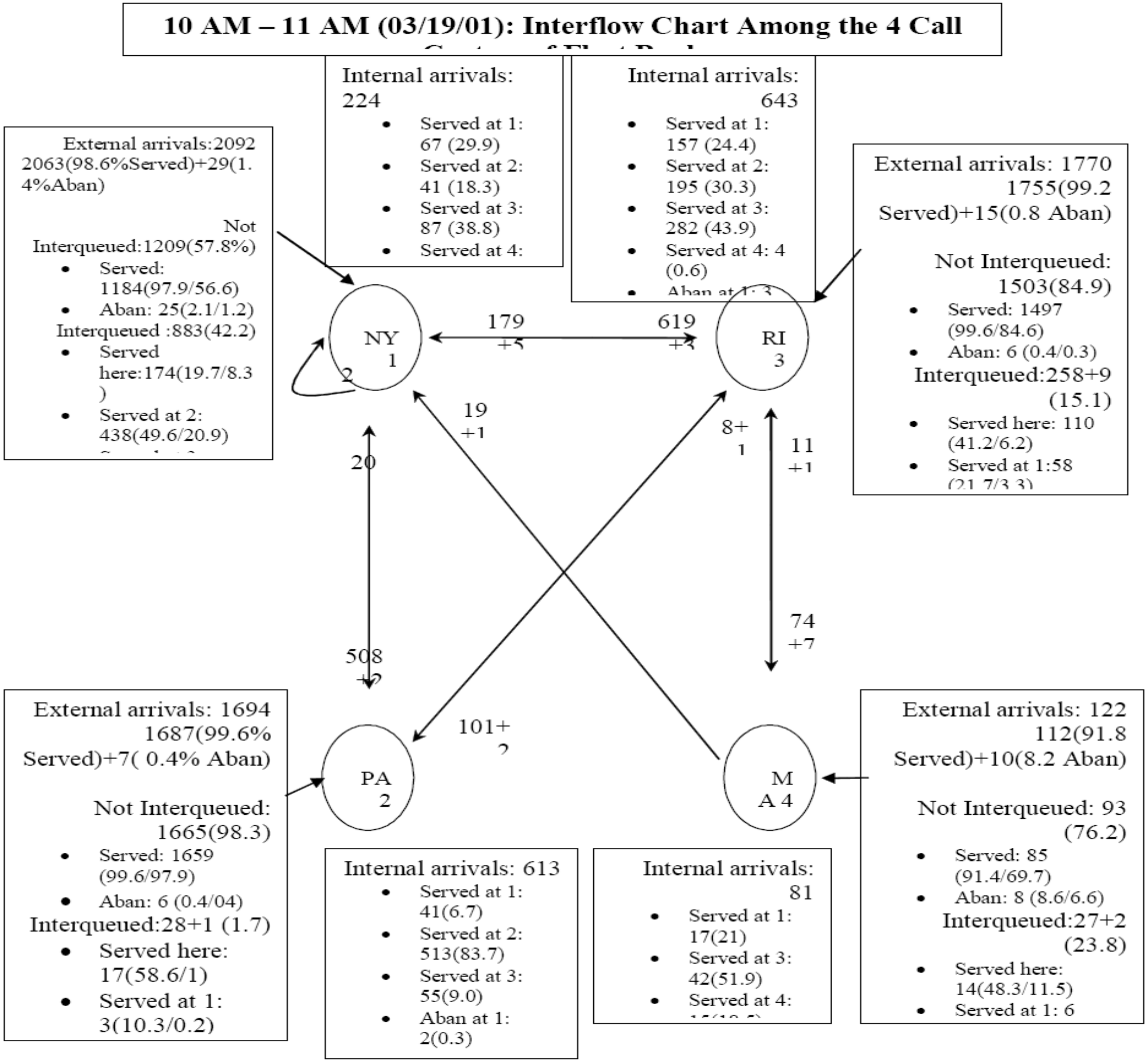
Service Time: Fitting Distribution

Fitting Mixture of 5 Lognormal Components

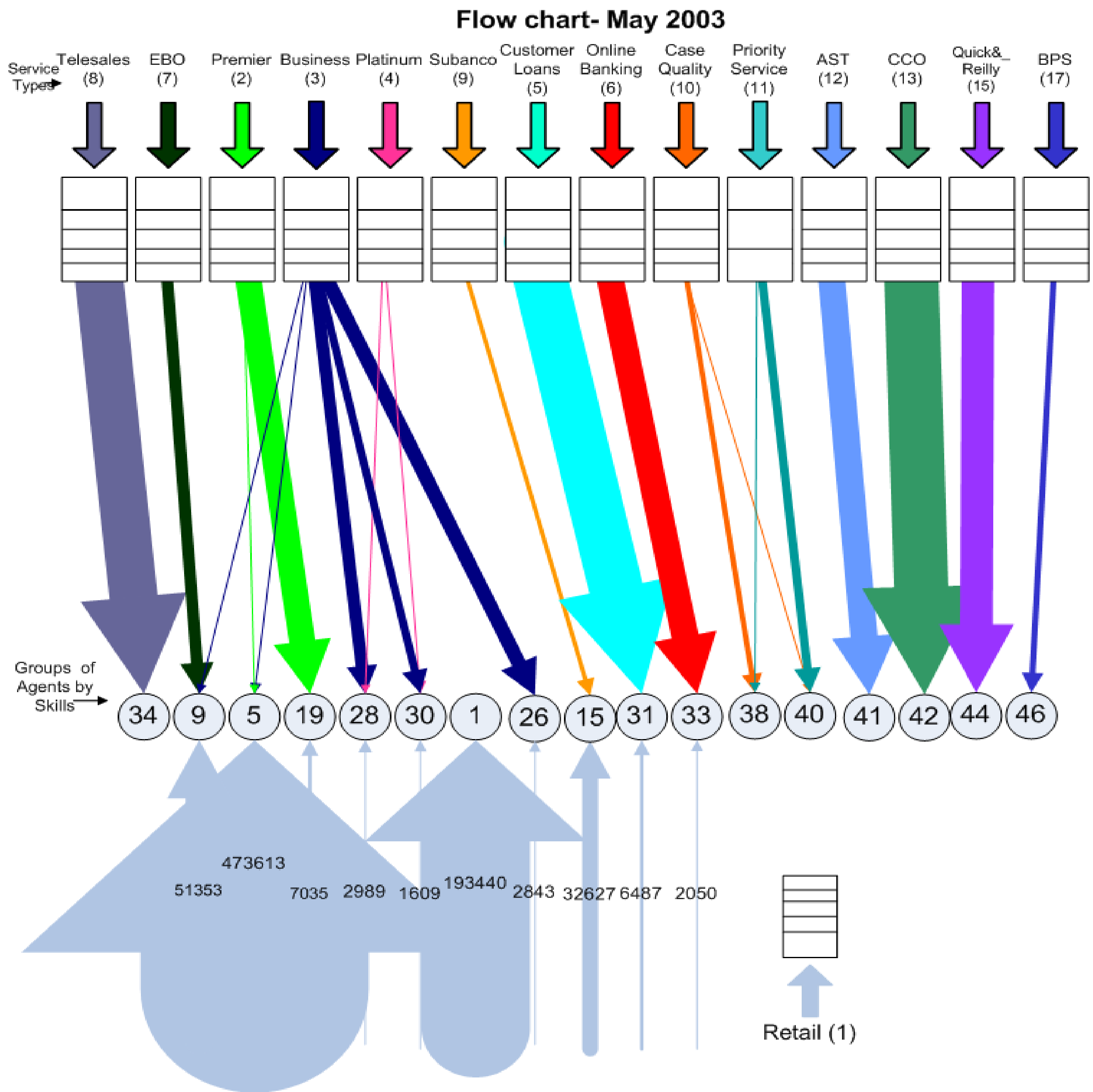


System Design

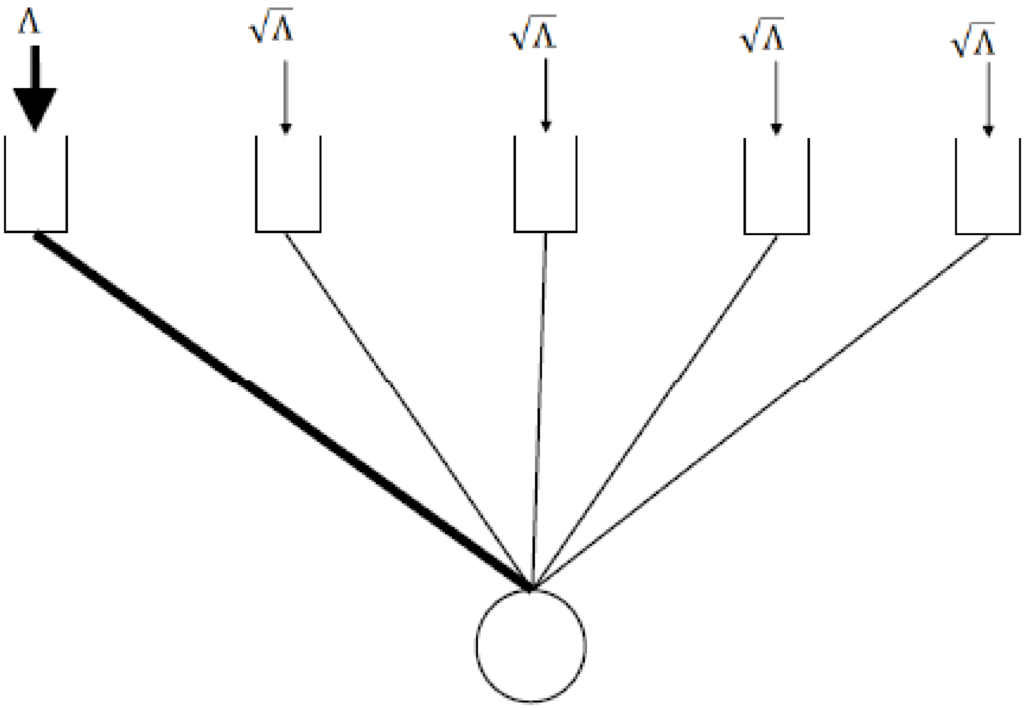
System Design- Interqueue Model



One “Trunk” and Multiple Thin Streams



One "Trunk" and Multiple Thin Streams

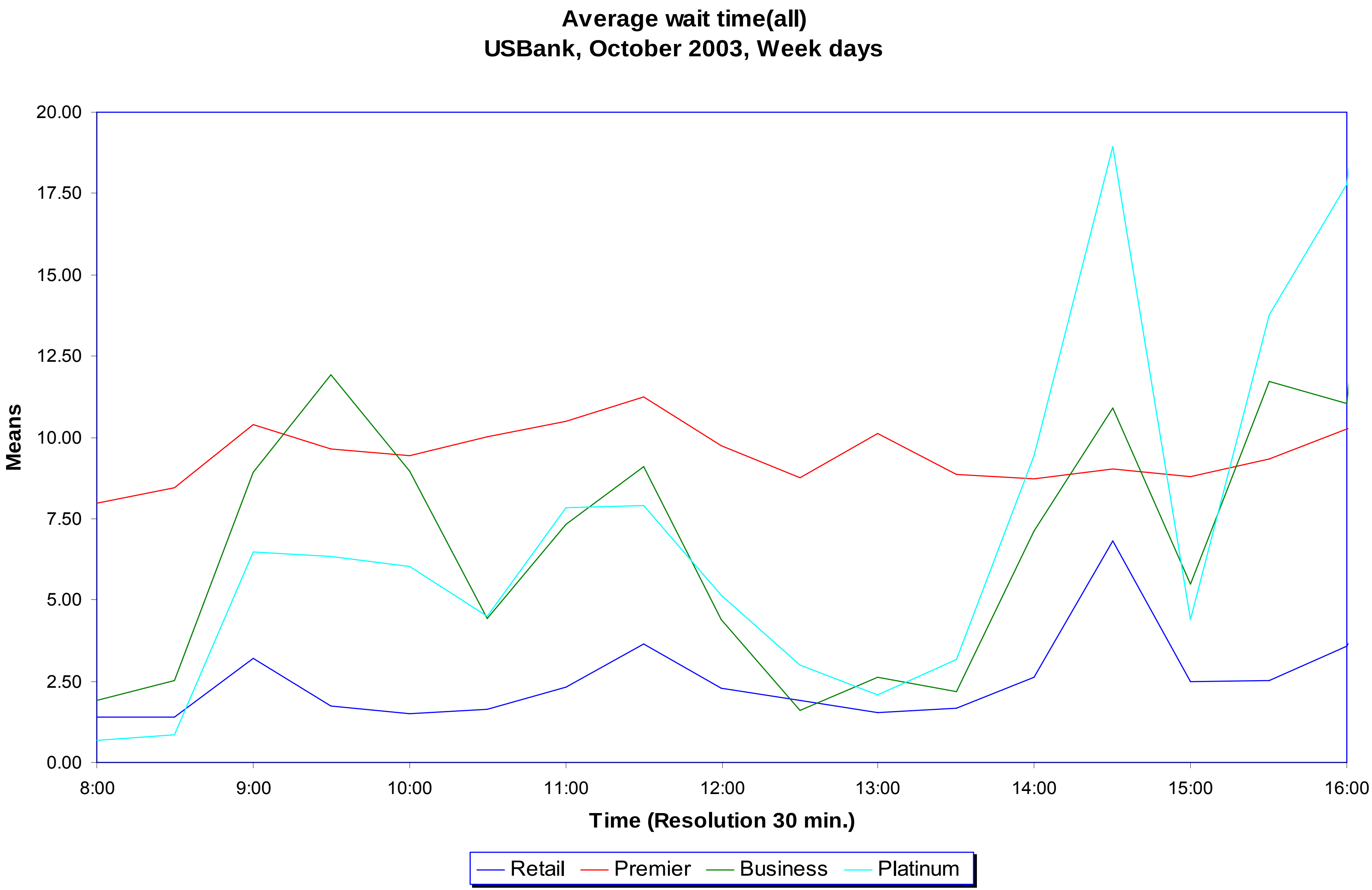


- # of thin class ≈ √Λ
- Under static priority for thin:

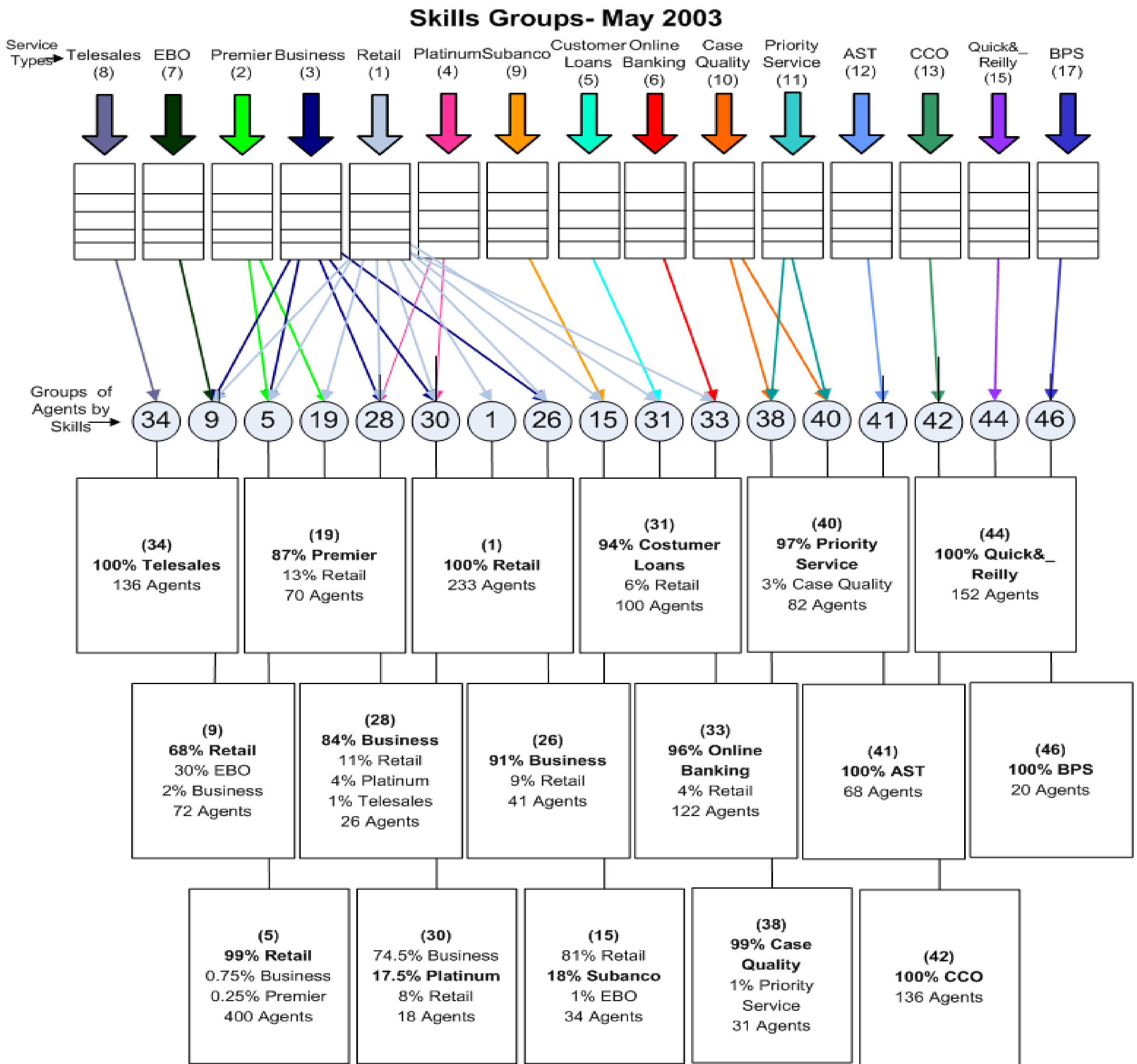
$$E[W_{thin}] > E[W_{thick}]$$

- Individual thin class has no impact
- On aggregate, thin classes do have impact

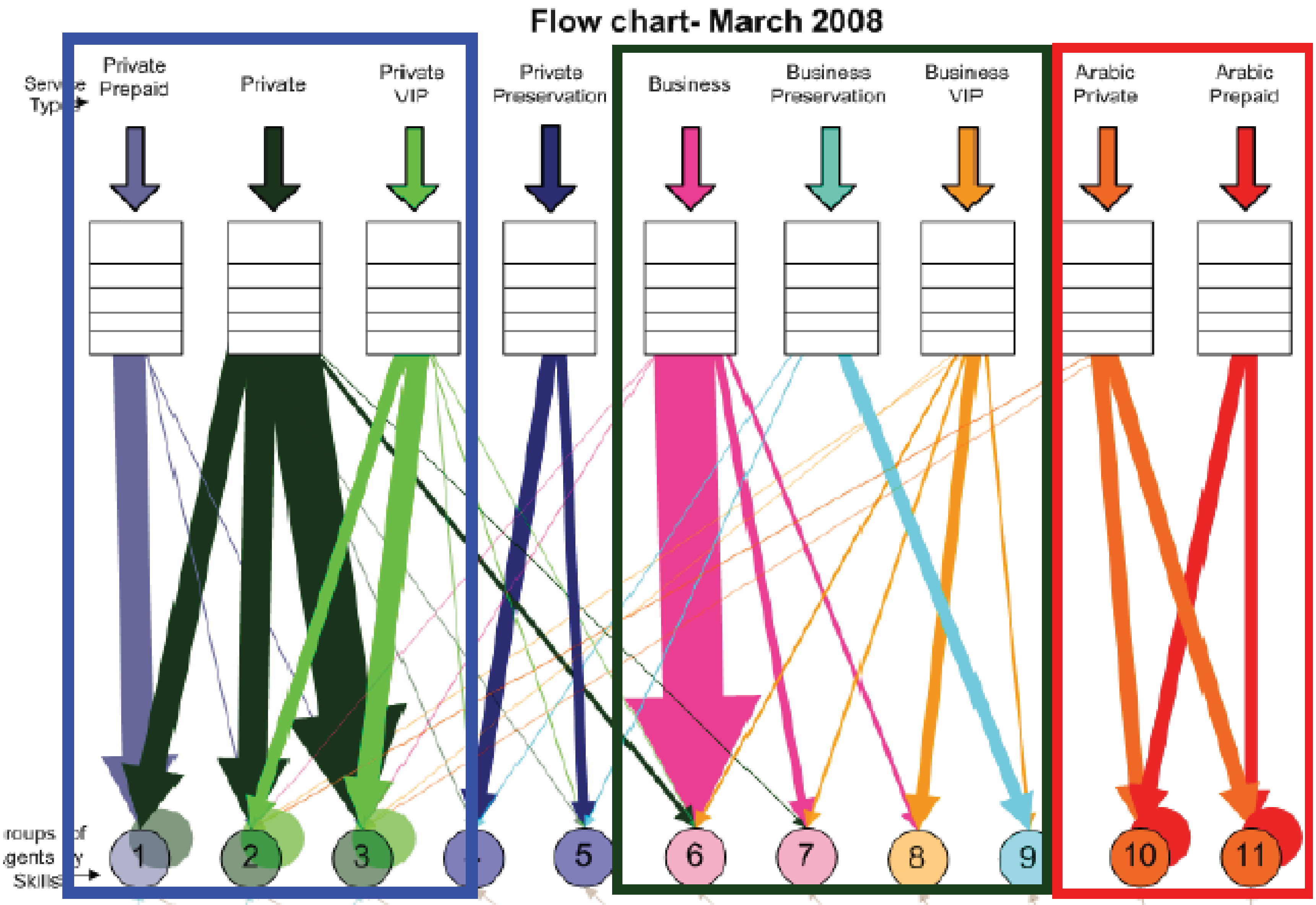
One “Trunk” and Multiple Thin Streams



System Design: The Agent Perspective



System Design: The Simplification Process



Service Rate, Class or Pool Dependent?

Agents Group\Service Class	Private Prepaid	Private	Private VIP
Private Prepaid	163.1	236.1	
Private - Private VIP (1)		243.5	195.1
Private - Private VIP (2)		244	201.4

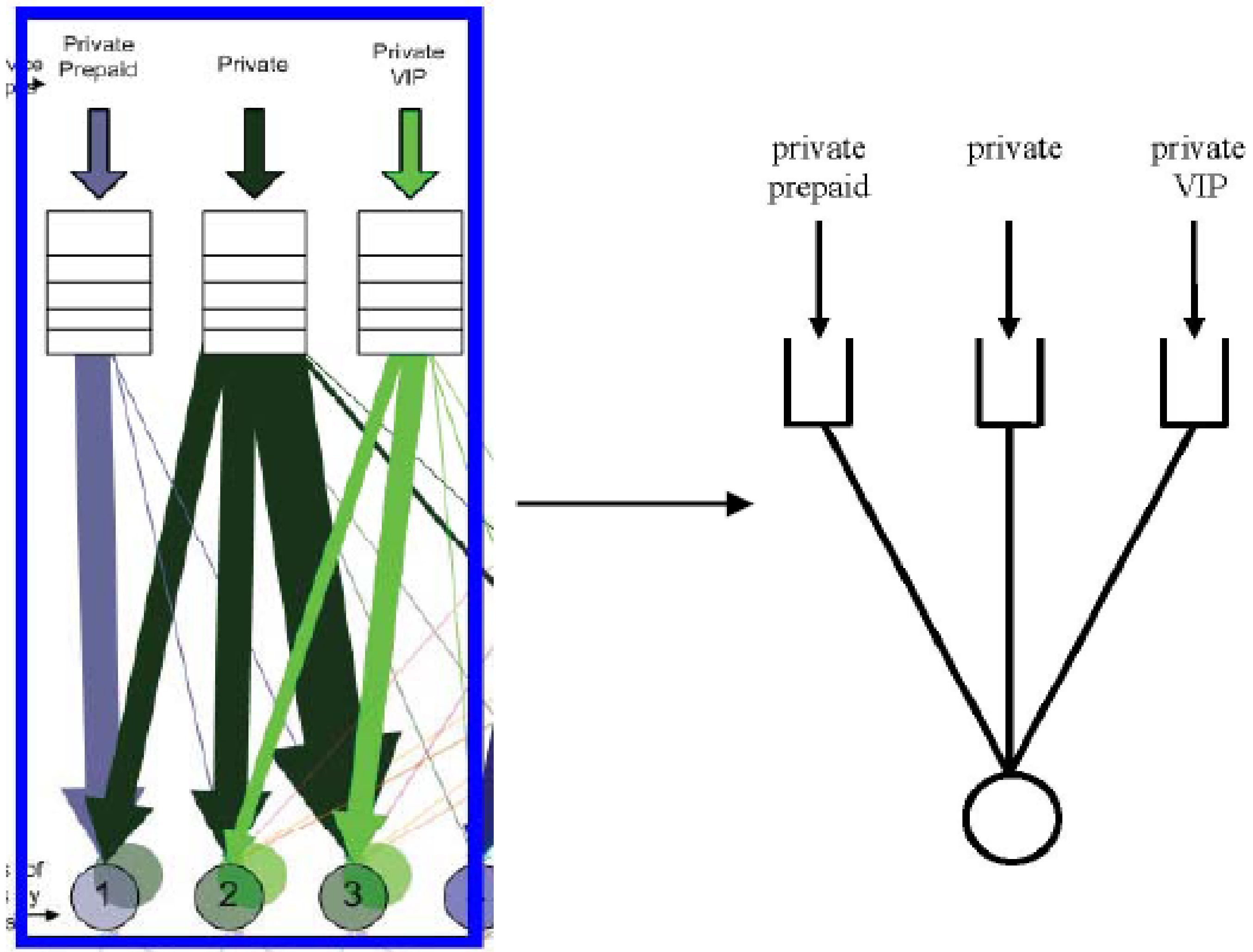
- **Class dependent service rate**

Agents Group\Service Class	Business	Business VIP	Business Preservation
Business (1)	276.9	261.5	
Business (2)	336.7	334.5	
Business VIP	315.9	280.5	
Business Preservation		386.2	634.1

- **Pool dependent service rate**

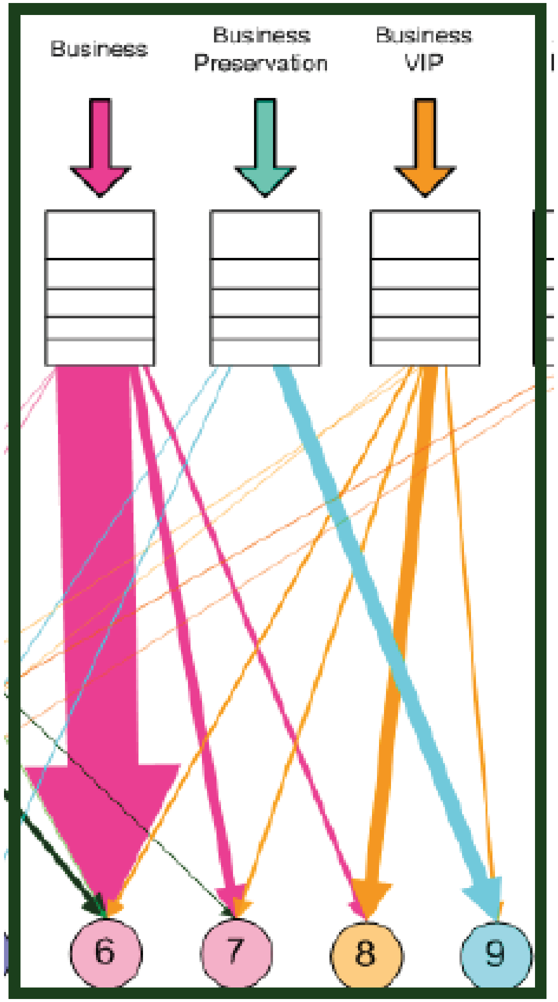
System Design: The Simplification Process

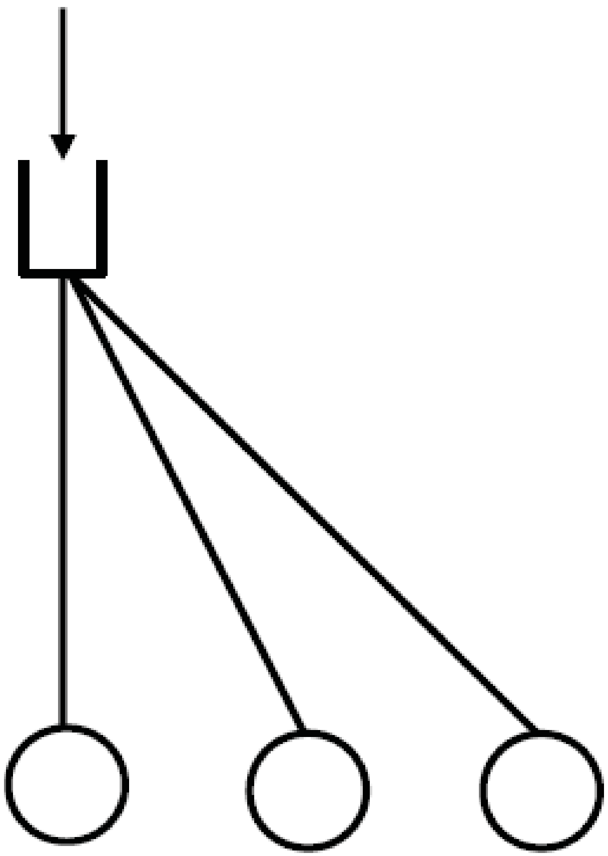
The Class-Dependent Case



In what sense the reduction work? [Atar et. al.]

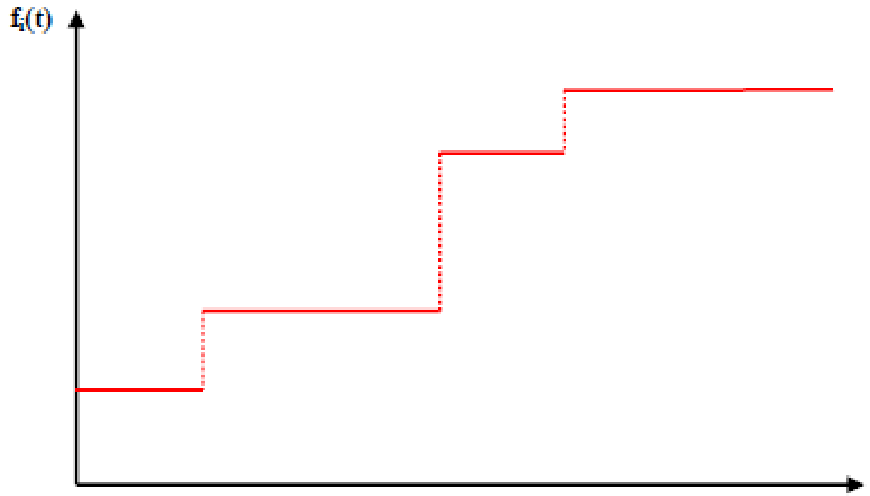
The pool-dependent case





SBR- Skills-Based Routing

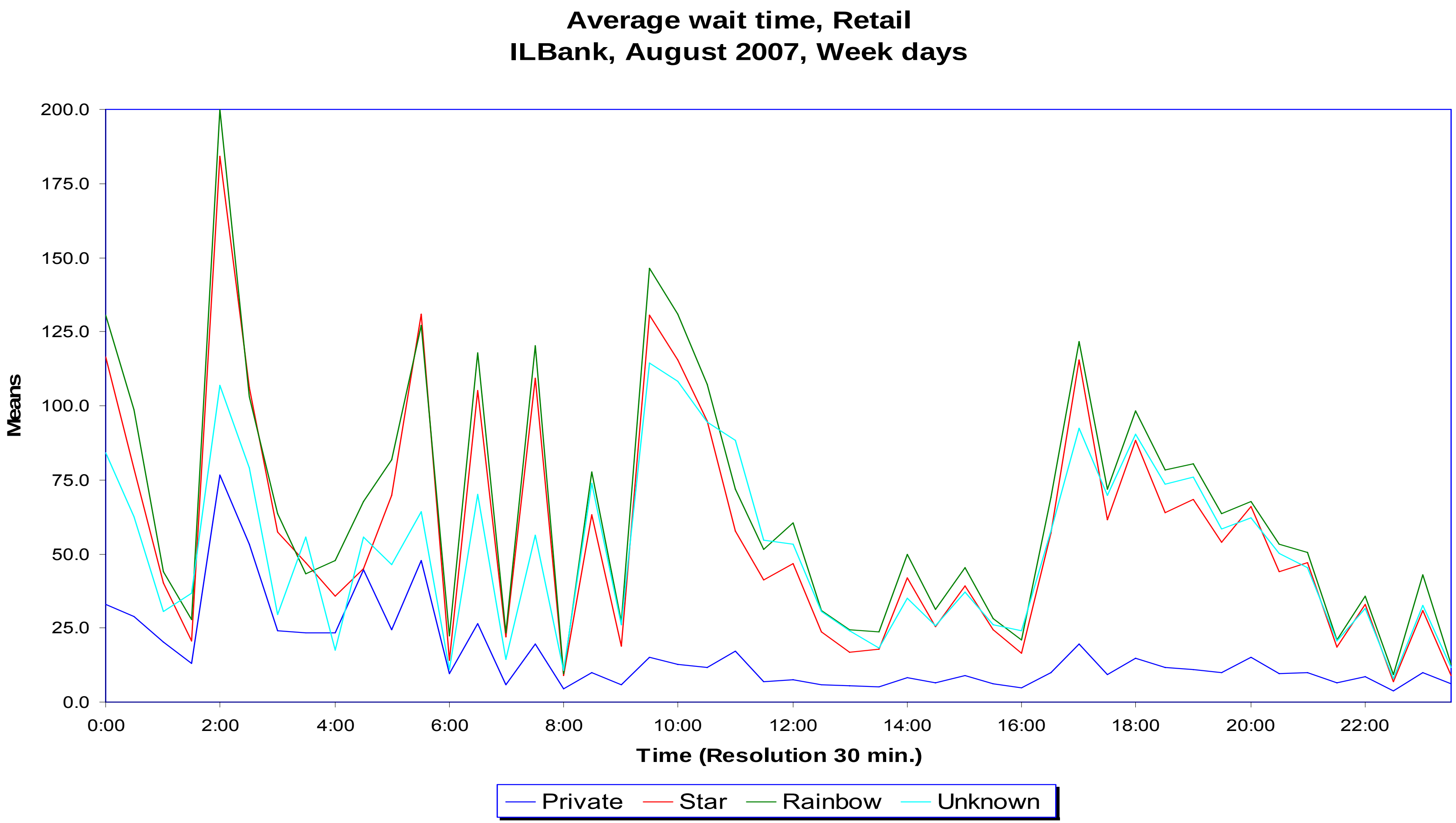
SBR - The ILBank Routing Rule



Customer type	ntranc	10	20	30	40	60	120	180	240	300	360	420	480	540	600	660
private	84	84	104	104	104	119	134	149	164	179	194	209	224	239	254	269
star	50	50	60	60	60	75	90	105	120	135	150	169	180	195	210	225
rainbow	28	28	34	34	34	49	64	79	94	109	124	139	154	169	184	199
unidentified	11	11	11	11	11	22	32	42	52	62	72	82	92	102	112	122

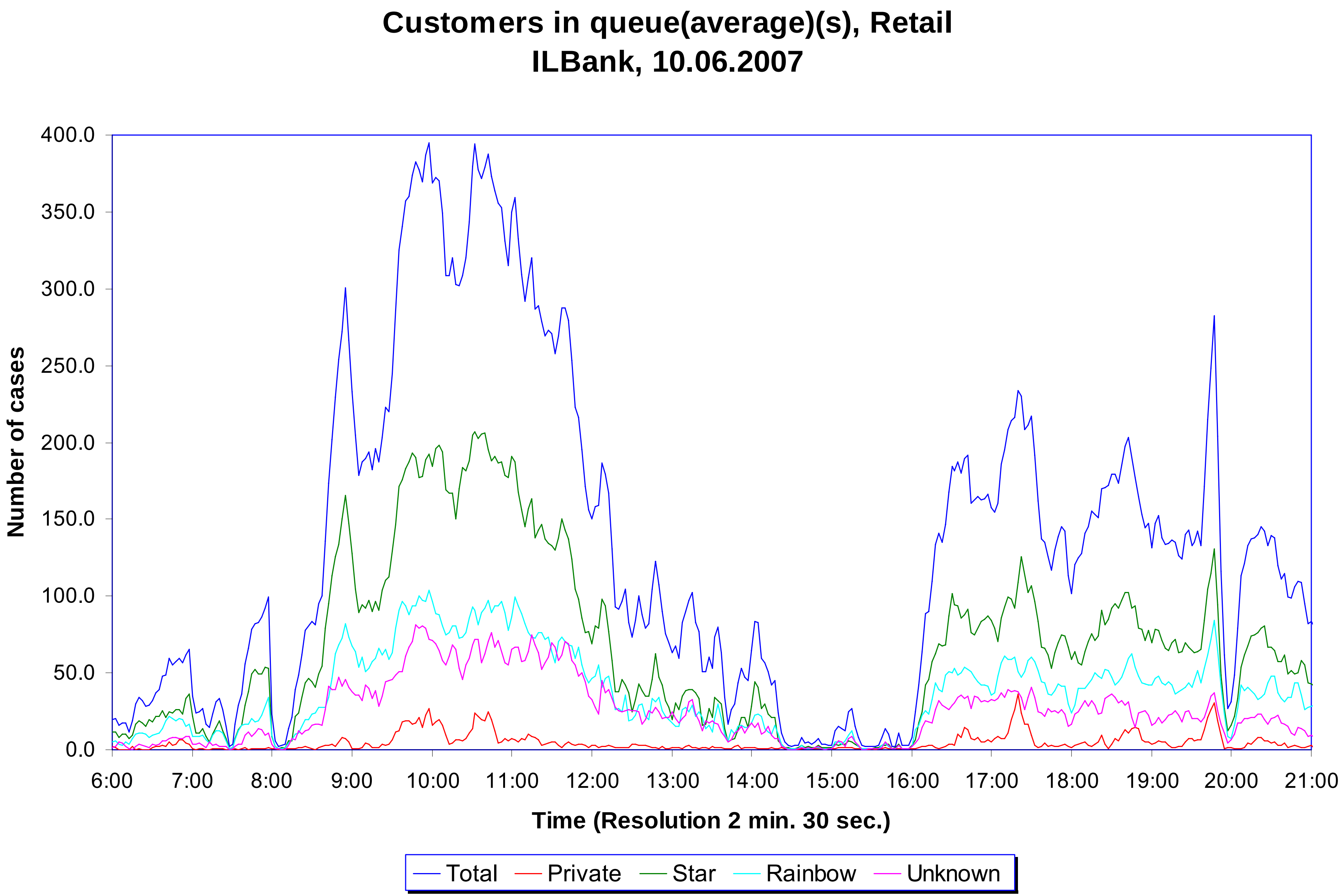
private>star> rainbow>unidentified

ILBank- Performance of Index-Based Routing: Proportionality



- Private class wait as in static priority
- Index-based provides a spectrum of differentiation

ILBank- Performance of Index-Based Routing: Proportionality



ILBank- Fixed Queue Ratios

$$\frac{Q_i(t)}{\sum_{k=1}^{I} Q_k(t)} \approx p_i$$

- Asymptotic proportionality in heavy-traffic simplifies staffing [e.g. Gurvich & Whitt]
- QoS “reduced” to constraint on aggregate queue.

Questions:

- How do parameters map to proportions?
- Can the rule be adjusted to meet multiple targets?

e.g. dedicated service + abandonment fraction

Expanding to Hospital

The Hospital Project

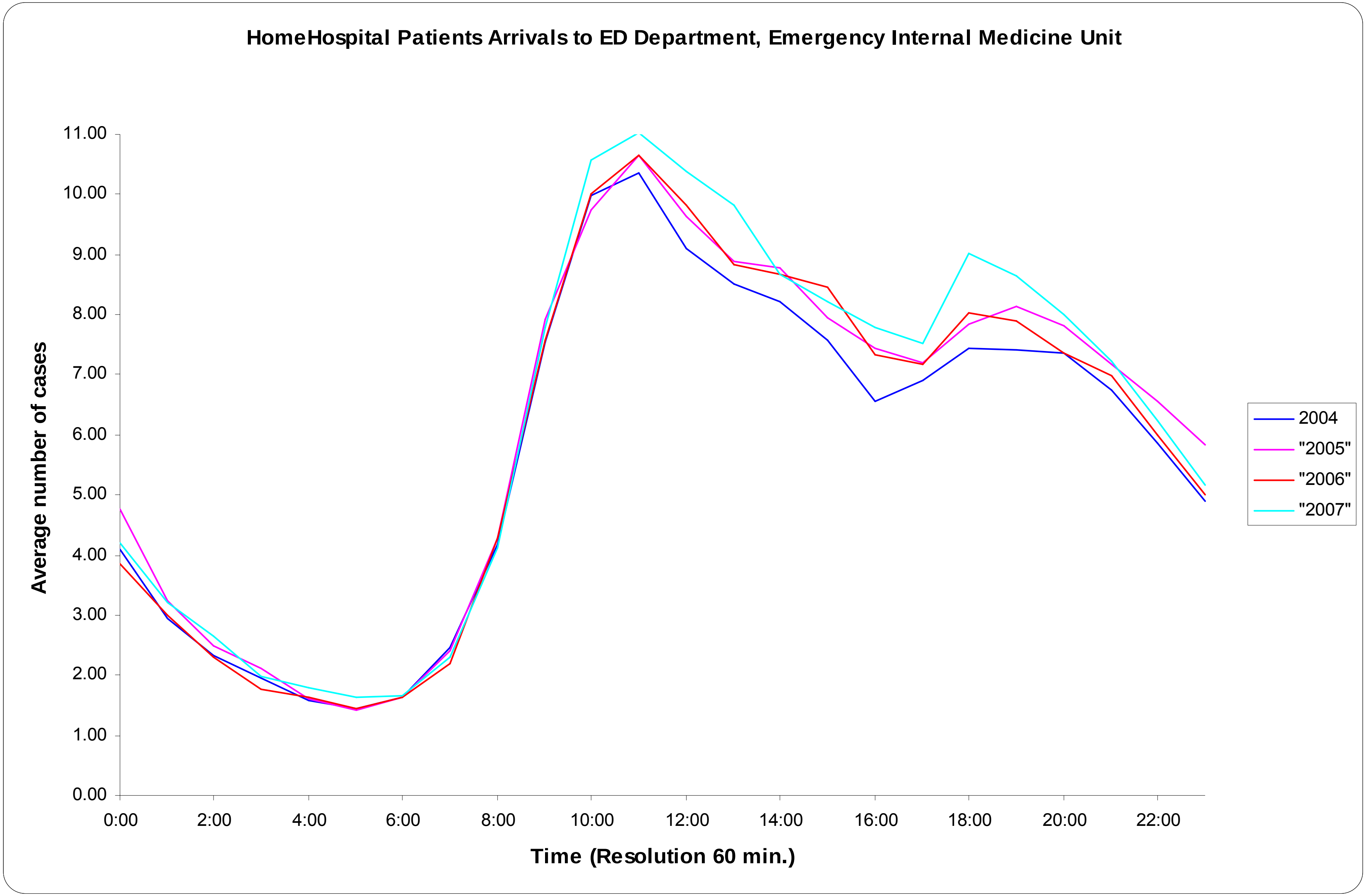
Goal: Flow analysis, from the Operational, Clinical and Financial view-points, through the multi-disciplinary eyes of **OR, IE, CS, HFE**.

- Tracks flow of individual patients, similarly to DataMOCCA, at the Emergency Departments of a large Israeli Hospital (have older data from 8 ED's).

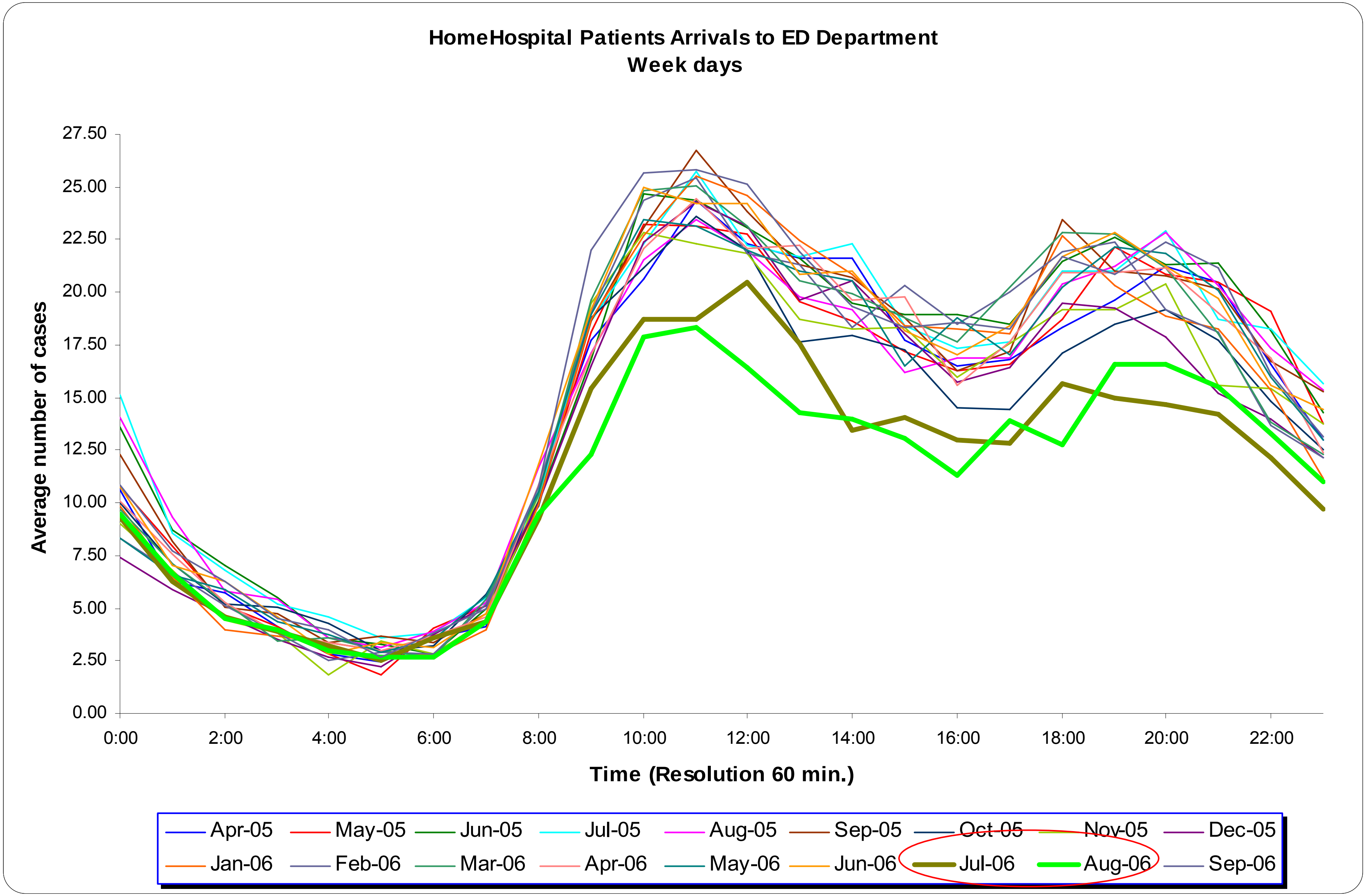
- Move on to Operating Rooms, Trauma, Pre-Natal, and Wards, jointly with IBM(within OCR = Open Collaborative Research).

Near Future: Based on Radio-Frequency ID's (RFID), a patients-tracking system, hopefully combined with nurses' and doctors' flow-data.

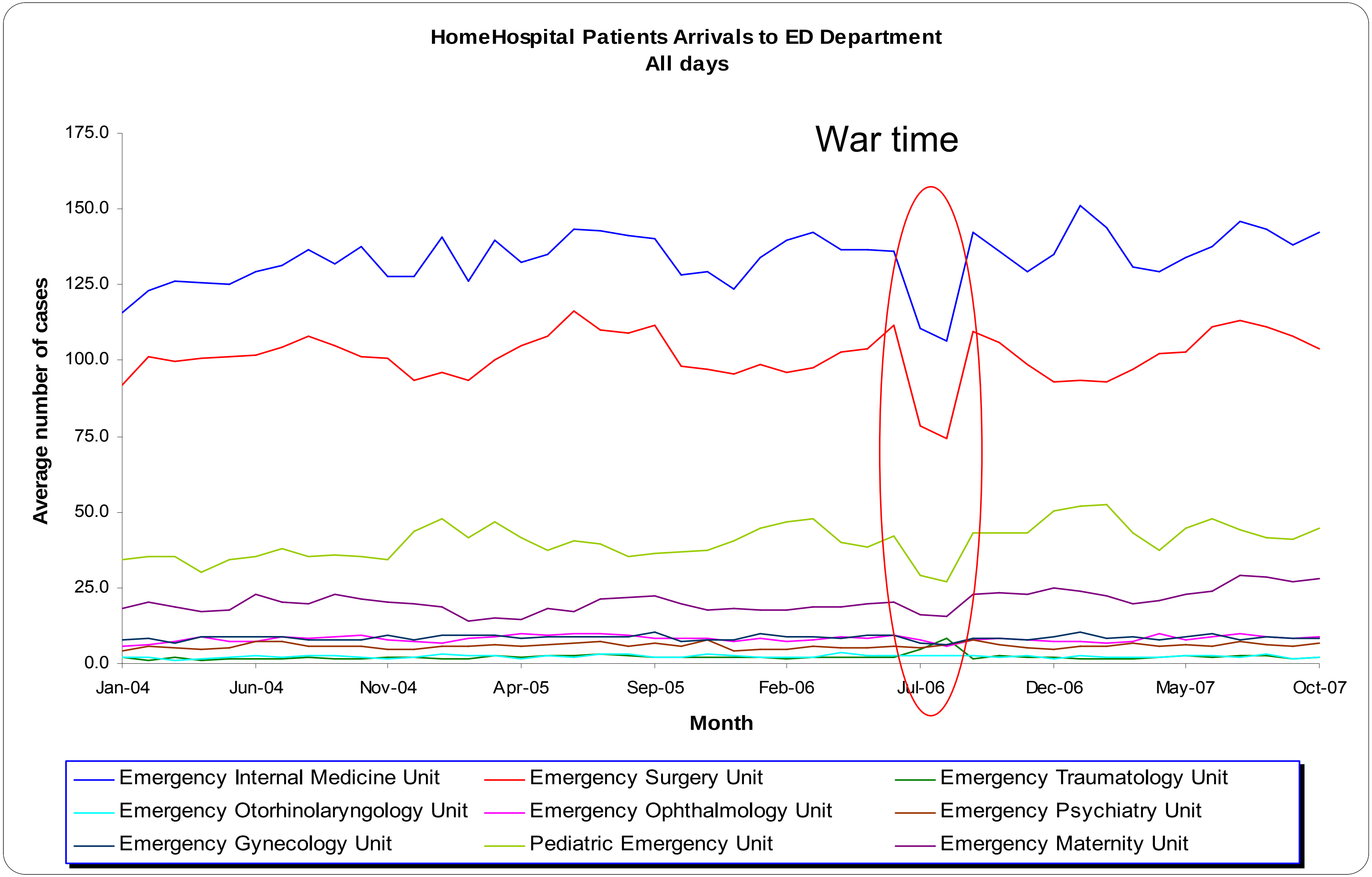
The Hospital Project



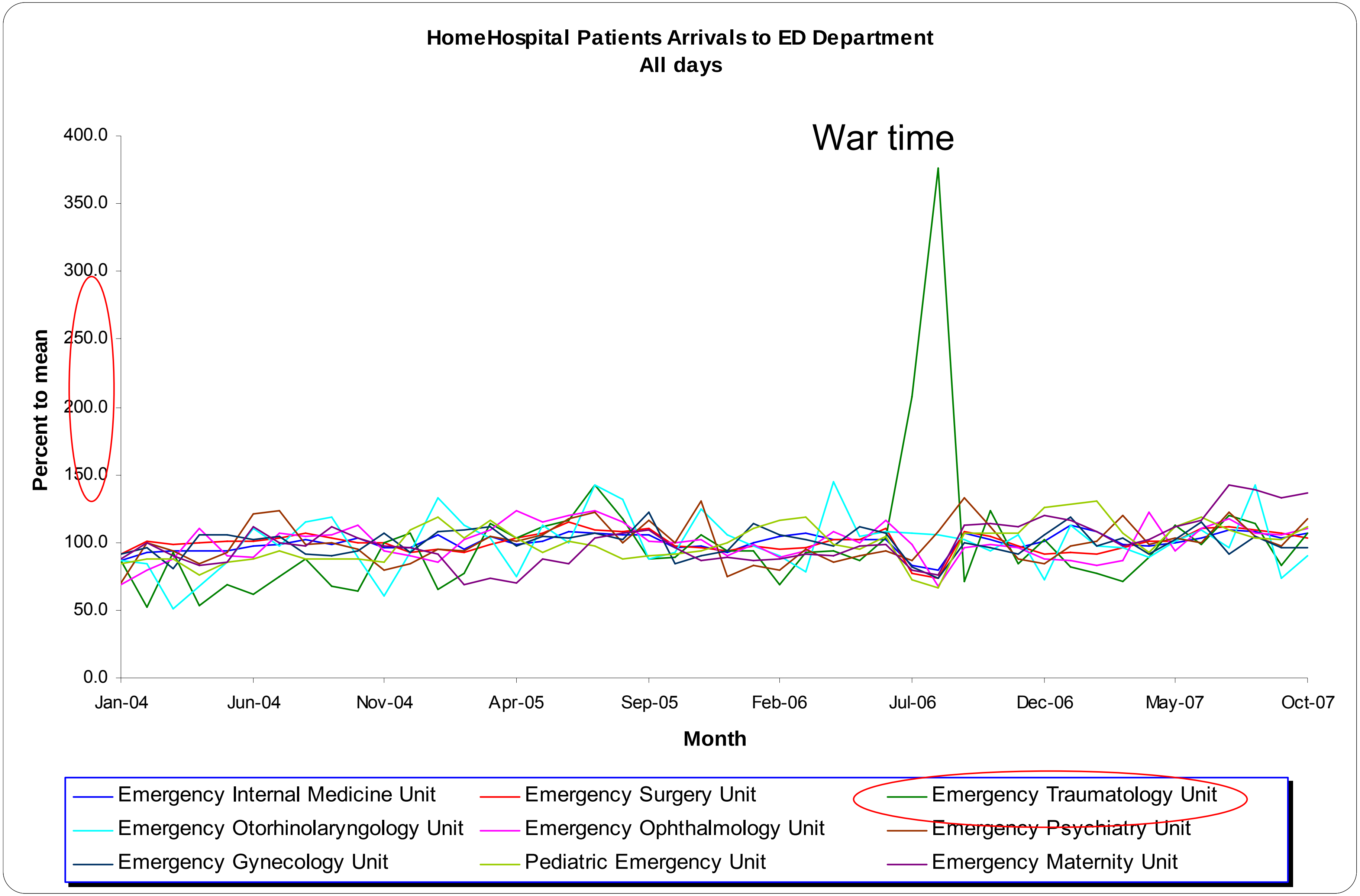
The Hospital Project



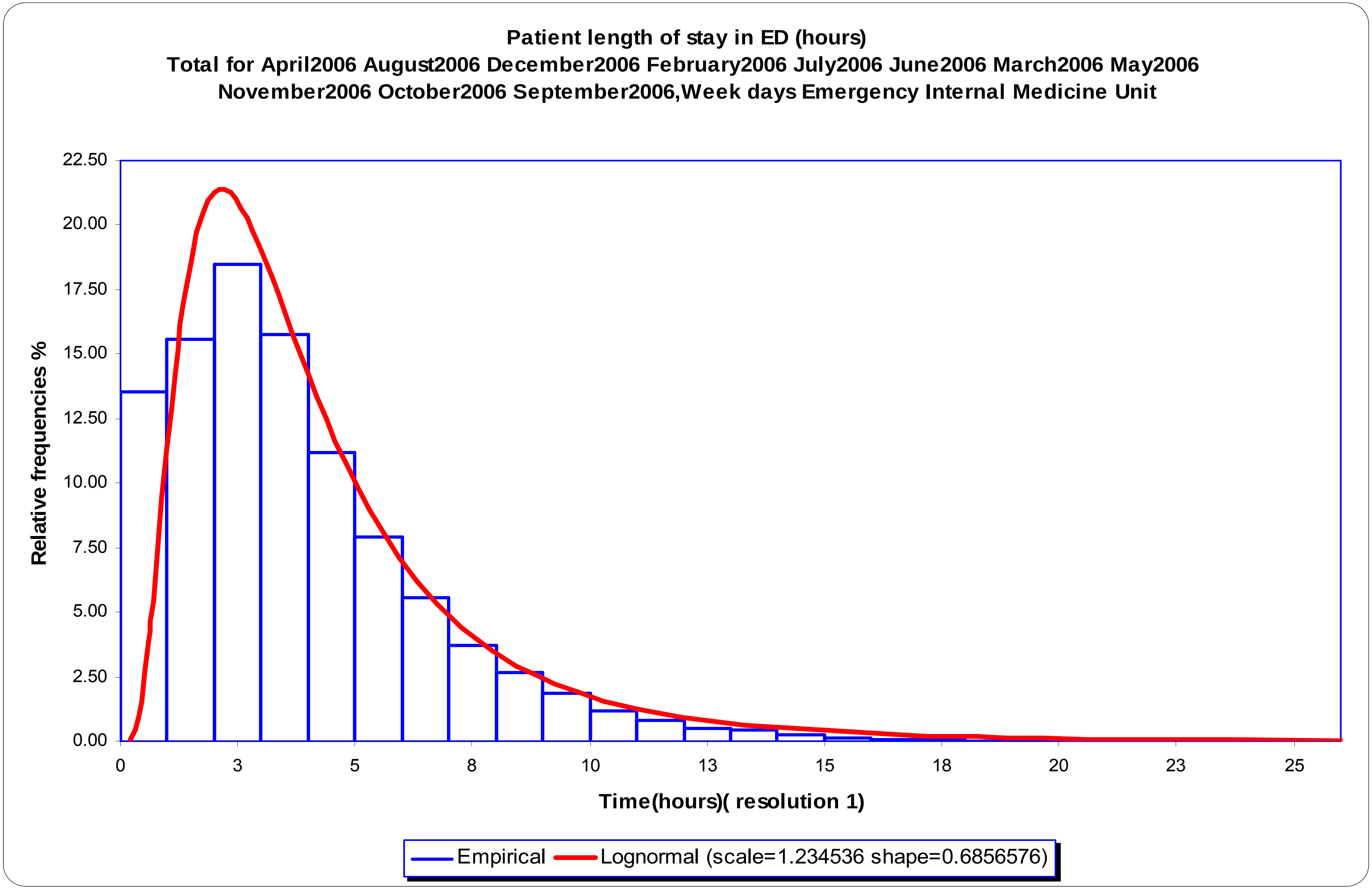
The Hospital Project



The Hospital Project

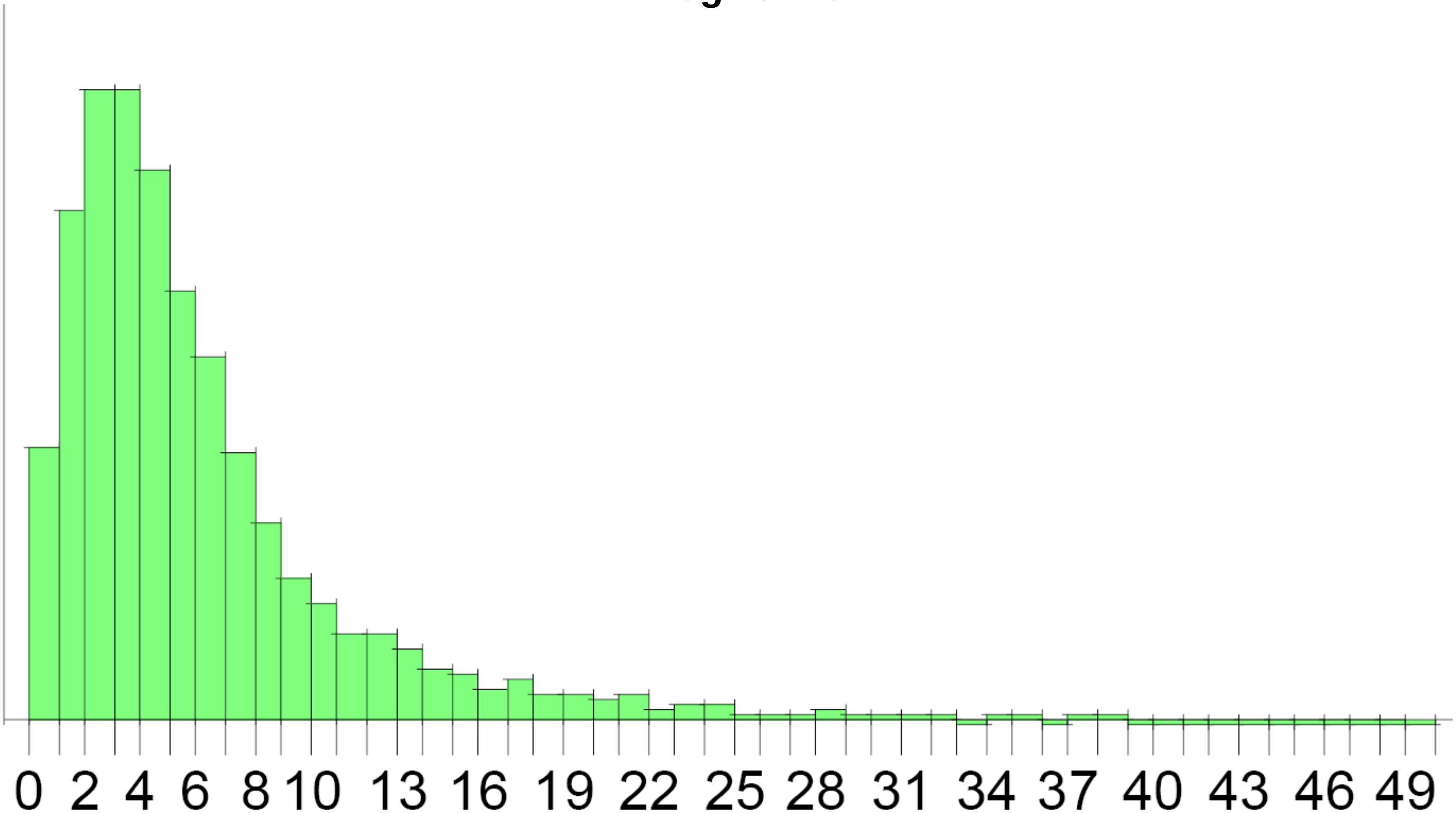


The Hospital Project



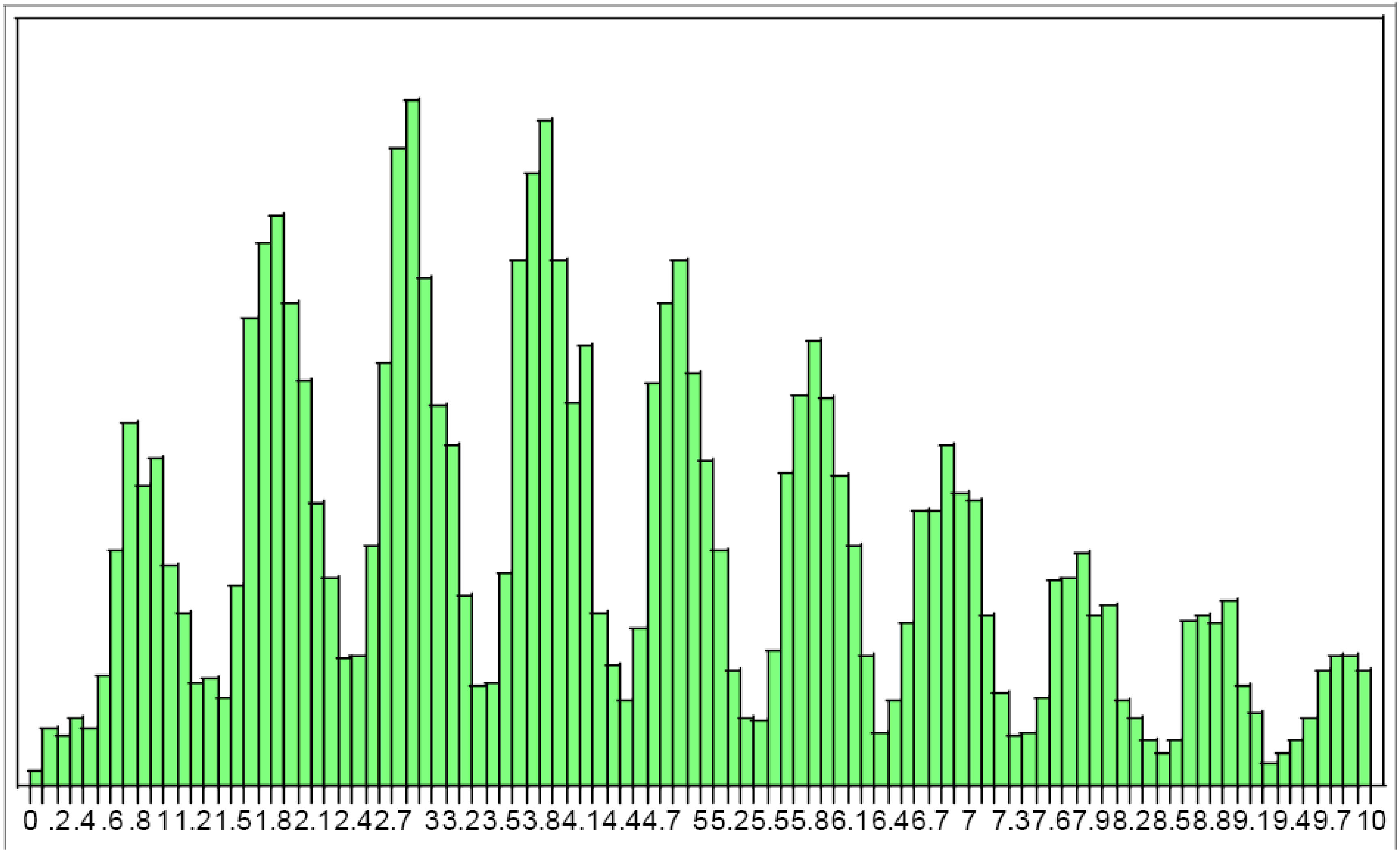
Length-of-Stay in Days: Internal Medicine Unit

LogNormal

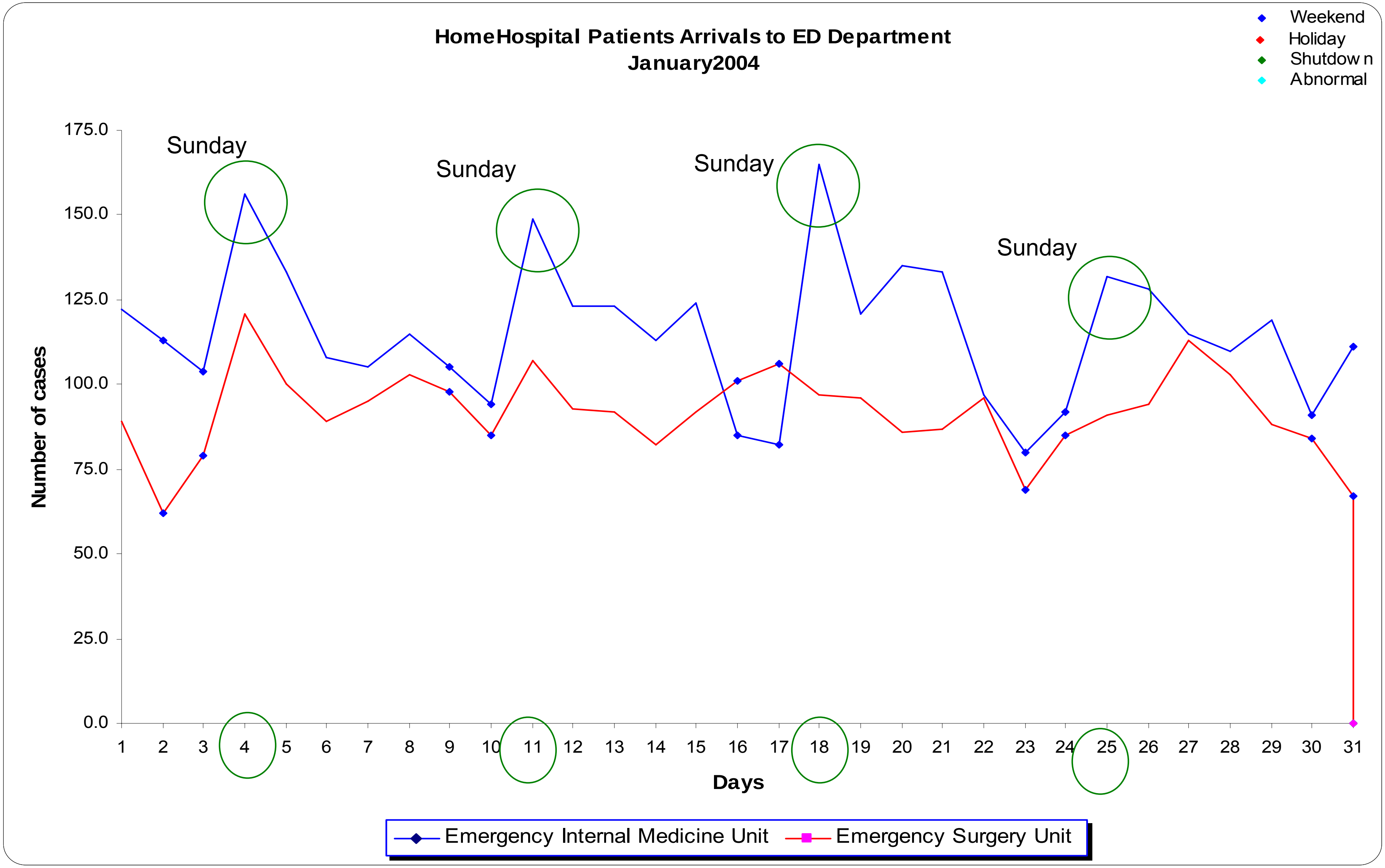


Length-of-Stay in Hours: Internal Medicine Unit

LogNormal



The Hospital Project



Empirical Analysis of Service Centers: A Queueing-Science Perspective

Pablo Liberman, SEE center, Technion

pabloliberman@gmail.com

http://ie.technion.ac.il/Labs/Serveng/

Questions?