Service Engineering: Data-Based Science & Teaching in support of Service Management

Avi Mandelbaum

Technion, Haifa, Israel

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

Stanford GSB, October 2006

Based on joint work with Sergey Zeltyn

Technion Statistics / SEE Lab: Paul Feigin, Valery Trofimov, RA's

Main Messages

1. **Simple Useful Models** at the Service of **Complex Realities**.

Main Messages

1. Simple Useful Models at the Service of **Complex Realities**.

Note: Useful must be Simple; Simple often rooted in **Deep** analysis.

Main Messages

1. Simple Useful Models at the Service of **Complex Realities**.

Note: Useful must be Simple; Simple often rooted in **Deep** analysis.

2. Data-Based Research & Teaching is a Must & Fun.

Supported by **DataMOCCA** = Data **MO**dels for **C**all **C**enter **A**nalysis.
Initiated with Wharton, developed at Technion, available for adoption.

Main Messages

1. Simple Useful Models at the Service of **Complex Realities**.

Note: Useful must be Simple; Simple often rooted in **Deep** analysis.

2. Data-Based Research & Teaching is a Must & Fun.

Supported by **DataMOCCA** = Data **MO**dels for **C**all **C**enter **A**nalysis.
Initiated with Wharton, developed at Technion, available for adoption.

3. Back to the **Basic-Research Paradigm** (Physics, Biology, . . .):
Measure, Model, Experiment, Validate, Refine, etc.

Main Messages

1. Simple Useful Models at the Service of **Complex Realities**.

Note: Useful must be Simple; Simple often rooted in **Deep** analysis.

2. Data-Based Research & Teaching is a Must & Fun.

Supported by **DataMOCCA** = Data **MO**dels for **C**all **C**enter **A**nalysis.
Initiated with Wharton, developed at Technion, available for adoption.

3. Back to the **Basic-Research Paradigm** (Physics, Biology, . . .):
Measure, Model, Experiment, Validate, Refine, etc.

4. Ancestors & **Practitioners** often knew/apply the “**right answer**":
simply did/do not have our tools/desire/need to prove it so.

Supported by **Erlang (1915), Palm (1945)**,..., thoughtful managers.

Main Messages

1. **Simple Useful Models** at the Service of **Complex Realities**.

Note: Useful must be Simple; Simple often rooted in **Deep** analysis.

2. **Data-Based** Research & Teaching is a Must & Fun.

Supported by **DataMOCCA** = Data **MO**dels for **C**all **C**enter **A**nalysis.
Initiated with Wharton, developed at Technion, available for adoption.

3. Back to the **Basic-Research Paradigm** (Physics, Biology, . . .):
Measure, Model, Experiment, Validate, Refine, etc.

4. **Ancestors** & **Practitioners** often knew/apply the "**right answer**":
simply did/do not have our tools/desire/need to prove it so.

Supported by **Erlang (1915), Palm (1945)**,..., thoughtful managers.

5. **Scientifically-based design principles and tools (software)**,
that support the balance of service **quality**, process **efficiency** and business **profitability**, from the (often-conflicting) views of **customers, servers, managers**: **Service Engineering**.

Background Material (Downloadable)

- Technion's "**Service-Engineering" Course** ($\geq$ 1995):
  http://ie.technion.ac.il/serveng

Background Material (Downloadable)

- Technion's "**Service-Engineering" Course** ($\geq$ 1995): http://ie.technion.ac.il/serveng
- Gans (U.S.A.), Koole (Europe), and M. (Israel): "Telephone Call Centers: Tutorial, Review and Research Prospects." MSOM, 2003.
- Brown, Gans, M., Sakov, Shen, Zeltyn, Zhao: "Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective." JASA, 2005.
- Trofimov, Feigin, M., Ishay, Nadjharov: "DataMOCCA: Models for Call/Contact Center Analysis." Technion Report, 2004-2006.
- M. "Call Centers: Research Bibliography with Abstracts." Version 7, December 2006.

The First Prerequisite: Data & Measurements

Empirical “Axiom": The data one needs is **never** there for one to use – always problems with historical data (eg. lacking, contaminated, averaged, . . .)

The First Prerequisite: Data & Measurements

Empirical "Axiom": The data one needs is **never** there for one to use – always problems with historical data (eg. lacking, contaminated, averaged, . . .)

Averages Prevalent.

But I need data at the level of the **Individual Transaction**: For each service transaction, its **operational history** – time-stamps of events. (Towards integrating with marketing / financial history.)

The First Prerequisite: Data & Measurements

Empirical “Axiom": The data one needs is **never** there for one to use – always problems with historical data (eg. lacking, contaminated, averaged, . . .)

Averages Prevalent.

But I need data at the level of the **Individual Transaction**: For each service transaction, its **operational history** – time-stamps of events. (Towards integrating with marketing / financial history.)

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)
- Future: Hospitals (via RFID)

Measurements: Face-to-Face Services

23 Bar-Code Readers at a Bank Branch

Bank – 2nd Floor Measurements

Measurements: Telephone Call-by-Call Data (Log-File)

vru+line	call_id	customer_id	priority	type	date	vru_entry	vru_exit	vru_time	q_start	q_exit	q_time	outcome	ser_start	ser_exit	ser_time	server
AA0101	44749	27644400	2	PS	990901	11:45:33	11:45:39	6	11:45:39	11:46:58	79	AGENT	11:46:57	11:51:00	243	DORIT
AA0101	44750	12887816	1	PS	990905	14:49:00	14:49:06	6	14:49:06	14:53:00	234	AGENT	14:52:59	14:54:29	90	ROTH
AA0101	44967	58660291	2	PS	990905	14:58:42	14:58:48	6	14:58:48	15:02:31	223	AGENT	15:02:31	15:04:10	99	ROTH
AA0101	44968	0	0	NW	990905	15:10:17	15:10:26	9	15:10:26	15:13:19	173	HANG	00:00:00	00:00:00	0	NO_SERVER
AA0101	44969	63193346	2	PS	990905	15:22:07	15:22:13	6	15:22:13	15:23:21	68	AGENT	15:23:20	15:25:25	125	STEREN
AA0101	44970	0	0	NW	990905	15:31:33	15:31:47	14	00:00:00	00:00:00	0	AGENT	15:31:45	15:34:16	151	STEREN
AA0101	44971	41630443	2	PS	990905	15:37:29	15:37:34	5	15:37:34	15:38:20	46	AGENT	15:38:18	15:40:56	158	TOVA
AA0101	44972	64185333	2	PS	990905	15:44:32	15:44:37	5	15:44:37	15:47:57	200	AGENT	15:47:56	15:49:02	66	TOVA
AA0101	44973	3.06E+08	1	PS	990905	15:53:05	15:53:11	6	15:53:11	15:56:39	208	AGENT	15:56:38	15:56:47	9	MORIAH
AA0101	44974	74780917	2	NE	990905	15:59:34	15:59:40	6	15:59:40	16:02:33	173	AGENT	16:02:33	16:26:04	1411	ELI
AA0101	44975	55920755	2	PS	990905	16:07:46	16:07:51	5	16:07:51	16:08:01	10	HANG	00:00:00	00:00:00	0	NO_SERVER
AA0101	44976	0	0	NW	990905	16:11:38	16:11:48	10	16:11:48	16:11:50	2	HANG	00:00:00	00:00:00	0	NO_SERVER
AA0101	44977	33689787	2	PS	990905	16:14:27	16:14:33	6	16:14:33	16:14:54	21	HANG	00:00:00	00:00:00	0	NO_SERVER
AA0101	44978	23817067	2	PS	990905	16:19:11	16:19:17	6	16:19:17	16:19:39	22	AGENT	16:19:38	16:21:57	139	TOVA
AA0101	44764	0	0	PS	990901	15:03:26	15:03:36	10	00:00:00	00:00:00	0	AGENT	15:03:35	15:06:36	181	ZOHARI
AA0101	44765	25219700	2	PS	990901	15:14:46	15:14:51	5	15:14:51	15:15:10	19	AGENT	15:15:09	15:17:00	111	SHARON
AA0101	44766	0	0	PS	990901	15:25:48	15:26:00	12	00:00:00	00:00:00	0	AGENT	15:25:59	15:28:15	136	ANAT
AA0101	44767	58859752	2	PS	990901	15:34:57	15:35:03	6	15:35:03	15:35:14	11	AGENT	15:35:13	15:35:15	2	MORIAH
AA0101	44768	0	0	PS	990901	15:46:30	15:46:39	9	00:00:00	00:00:00	0	AGENT	15:46:38	15:51:51	313	ANAT
AA0101	44769	78191137	2	PS	990901	15:56:03	15:56:09	6	15:56:09	15:56:28	19	AGENT	15:56:28	15:59:02	154	MORIAH
AA0101	44770	0	0	PS	990901	16:14:31	16:14:46	15	00:00:00	00:00:00	0	AGENT	16:14:44	16:16:02	78	BENSION
AA0101	44771	0	0	PS	990901	16:38:59	16:39:12	13	00:00:00	00:00:00	0	AGENT	16:39:11	16:43:35	264	VICKY
AA0101	44772	0	0	PS	990901	16:51:40	16:51:50	10	00:00:00	00:00:00	0	AGENT	16:51:49	16:53:52	123	ANAT
AA0101	44773	0	0	PS	990901	17:02:19	17:02:28	9	00:00:00	00:00:00	0	AGENT	17:02:28	17:07:42	314	VICKY
AA0101	44774	32387482	1	PS	990901	17:18:18	17:18:24	6	17:18:24	17:19:01	37	AGENT	17:19:00	17:19:35	35	VICKY
AA0101	44775	0	0	PS	990901	17:38:53	17:39:05	12	00:00:00	00:00:00	0	AGENT	17:39:04	17:40:43	99	TOVA
AA0101	44776	0	0	PS	990901	17:52:59	17:53:09	10	00:00:00	00:00:00	0	AGENT	17:53:08	17:53:09	1	NO_SERVER
AA0101	44777	37635950	2	PS	990901	18:15:47	18:15:52	5	18:15:52	18:16:57	65	AGENT	18:16:56	18:18:48	112	ANAT
AA0101	44778	0	0	NE	990901	18:30:43	18:30:52	9	00:00:00	00:00:00	0	AGENT	18:30:51	18:30:54	3	MORIAH
AA0101	44779	0	0	PS	990901	18:51:47	18:52:02	15	00:00:00	00:00:00	0	AGENT	18:52:02	18:55:30	208	TOVA
AA0101	44780	0	0	PS	990901	19:19:04	19:19:17	13	00:00:00	00:00:00	0	AGENT	19:19:15	19:20:20	65	MEIR
AA0101	44781	0	0	PS	990901	19:39:19	19:39:30	11	00:00:00	00:00:00	0	AGENT	19:39:29	19:41:42	133	BENSION
AA0101	44782	0	0	NW	990901	20:08:13	20:08:25	12	00:00:00	00:00:00	0	AGENT	20:08:28	20:08:41	13	NO_SERVER
AA0101	44783	0	0	PS	990901	20:23:51	20:24:05	14	00:00:00	00:00:00	0	AGENT	20:24:04	20:24:33	29	BENSION
AA0101	44784	0	0	NW	990901	20:36:54	20:37:14	20	00:00:00	00:00:00	0	AGENT	20:37:13	20:38:07	54	BENSION
AA0101	44785	0	0	PS	990901	20:50:07	20:50:16	9	00:00:00	00:00:00	0	AGENT	20:50:15	20:51:32	77	BENSION
AA0101	44786	0	0	PS	990901	21:04:41	21:04:51	10	00:00:00	00:00:00	0	AGENT	21:04:50	21:05:59	69	TOVA
AA0101	44787	0	0	PS	990901	21:25:00	21:25:13	13	00:00:00	00:00:00	0	AGENT	21:25:13	21:28:03	170	AVI
AA0101	44788	0	0	PS	990901	21:50:40	21:50:54	14	00:00:00	00:00:00	0	AGENT	21:50:54	21:51:55	61	AVI
AA0101	44789	9103060	2	NE	990901	22:05:40	22:05:46	6	22:05:46	22:09:52	246	AGENT	22:09:51	22:13:41	230	AVI
AA0101	44790	14558621	2	PS	990901	22:24:11	22:24:17	6	22:24:17	22:26:16	119	AGENT	22:26:15	22:27:28	73	VICKY
AA0101	44791	0	0	PS	990901	22:46:27	22:46:37	10	00:00:00	00:00:00	0	AGENT	22:46:36	22:47:03	27	AVI
AA0101	44792	67158097	2	PS	990901	23:05:07	23:05:13	6	23:05:13	23:05:30	17	AGENT	23:05:29	23:06:49	80	VICKY
AA0101	44793	15317126	2	PS	990901	23:28:52	23:28:58	6	23:28:58	23:30:08	70	AGENT	23:30:07	23:35:03	296	DARMON
AA0101	44794	0	0	PS	990902	00:10:47	00:12:05	78	00:00:00	00:00:00	0	HANG	00:00:00	00:00:00	0	NO_SERVER
AA0101	44795	0	0	PS	990902	07:16:52	07:17:01	9	00:00:00	00:00:00	0	AGENT	07:17:01	07:17:44	43	ANAT
AA0101	44796	0	0	PS	990902	07:50:05	07:50:16	11	00:00:00	00:00:00	0	AGENT	07:50:16	07:53:03	167	STEREN

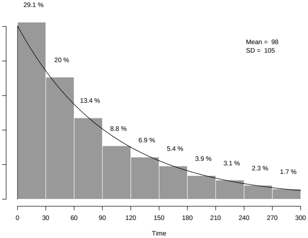
Averages Prevalent

ACD Report: Health Insurance

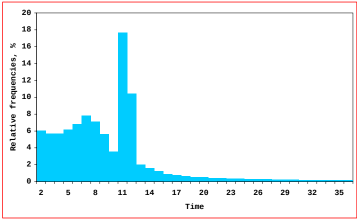
Time	Calls	Answered	Abandoned%	ASA	AHT	Occ%	# of agents
Total	20,577	19,860	**3.5%**	**30**	307	95.1%	
8:00	332	308	7.2%	27	302	87.1%	59.3
8:30	653	615	5.8%	58	293	96.1%	104.1
9:00	866	796	8.1%	63	308	97.1%	140.4
9:30	1,152	1,138	1.2%	28	303	90.8%	211.1
10:00	1,330	1,286	3.3%	22	307	98.4%	223.1
10:30	1,364	1,338	1.9%	33	296	99.0%	222.5
11:00	1,380	1,280	7.2%	34	306	98.2%	222.0
11:30	1,272	1,247	2.0%	44	298	94.6%	218.0
12:00	1,179	1,177	0.2%	1	306	91.6%	218.3
12:30	1,174	1,160	1.2%	10	302	95.5%	203.8
13:00	1,018	999	1.9%	9	314	95.4%	182.9
13:30	**1,061**	**961**	**9.4%**	**67**	**306**	**100.0%**	**163.4**
14:00	1,173	1,082	7.8%	78	313	99.5%	188.9
14:30	**1,212**	**1,179**	**2.7%**	**23**	**304**	**96.6%**	**206.1**
15:00	1,137	1,122	1.3%	15	320	96.9%	205.8
15:30	1,169	1,137	2.7%	17	311	97.1%	202.2
16:00	1,107	1,059	4.3%	46	315	99.2%	187.1
16:30	914	892	2.4%	22	307	95.2%	160.0
17:00	**615**	**615**	**0.0%**	**2**	**328**	**83.0%**	**135.0**
17:30	420	420	0.0%	0	328	73.8%	103.5
18:00	49	49	0.0%	14	180	84.2%	5.8

Beyond Averages: Waiting Times in a Call Center

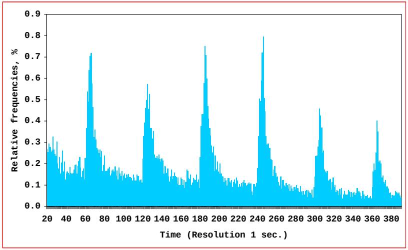
Small Israeli Bank



Large U.S. Bank



Medium Israeli Bank



The Second Prerequisite: Models

Through **Examples** Only.

Each example starts with a **Complex Reality** and ends with a **useful** insight due to a **Simple Model**.

The Second Prerequisite: Models

Through **Examples** Only.

Each example starts with a **Complex Reality** and ends with a **useful** insight due to a **Simple Model**.

"**Theorem**": A useful model must be simple (yet not too simple).

Models in decreasing simplicity-levels:

- **Conceptual**: Service Networks = **Queueing Networks**
- **Descriptive**: Averages, Histograms
- **Explanatory**: Comparative, Regression
- **Analytical/Mathematical**: Little's Law, Fluid Models, Queueing Models, Diffusion Refinements.

The Second Prerequisite: Models

Through **Examples** Only.

Each example starts with a **Complex Reality** and ends with a **useful** insight due to a **Simple Model**.

“**Theorem"**: A useful model must be simple (yet not too simple).

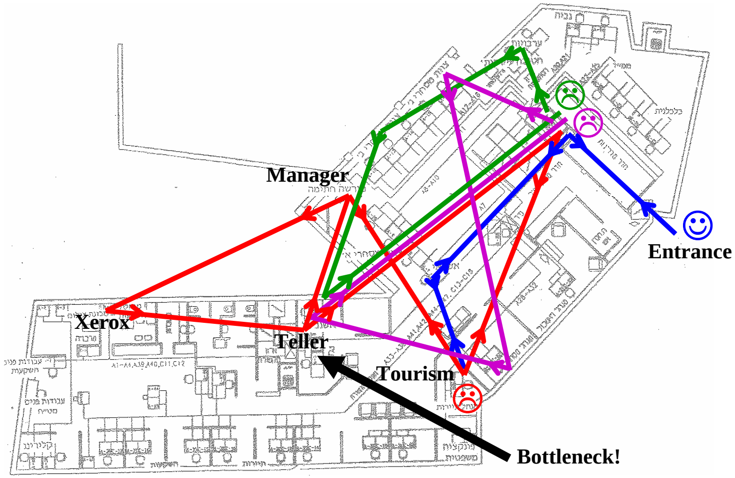
Models in decreasing simplicity-levels:

- **Conceptual**: Service Networks = **Queueing Networks**
- **Descriptive**: Averages, Histograms
- **Explanatory**: Comparative, Regression
- **Analytical/Mathematical**: Little’s Law, Fluid Models, Queueing Models, Diffusion Refinements.

“Corollary": To be useful, a simple model sometimes requires deep analysis.

Conceptual Model: Face-to-Face Services

Bank Branch = Queueing Network



Descriptive Model: Transition Probabilities (Averages)

Transition Frequencies Between Units in The Private and Business Sections:

	From Unit / To Unit	Private Banking: Bankers	Private Banking: Authorized Personal	Private Banking: Compens - - ations	Private Banking: Tellers	Business: Tellers	Business: Overdrafts	Business: Authorized Personal	Business: Full Service	Exit
Private Banking	Bankers		1%	1%	4%	4%	0%	0%	0%	90%
	Authorized Personal	12%		5%	4%	6%	0%	0%	0%	73%
	Compensations	7%	4%		18%	6%	0%	0%	1%	64%
	Tellers	6%	0%	1%		1%	0%	0%	0%	90%
Services	Tellers	1%	0%	0%	0%		1%	0%	2%	94%
	Overdrafts	2%	0%	1%	1%	19%		5%	8%	64%
	Authorized Personal	2%	1%	0%	1%	11%	5%		11%	69%
	Full Service	1%	0%	0%	0%	8%	1%	2%		88%
	Entrance	13%	0%	3%	10%	58%	2%	0%	14%	0%

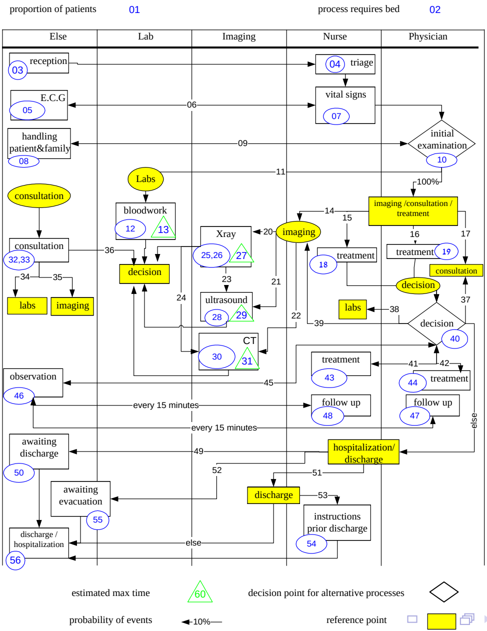
Legend:

0%-5%	5%-10%	10%-15%	>15%

Dominant Paths - Business:

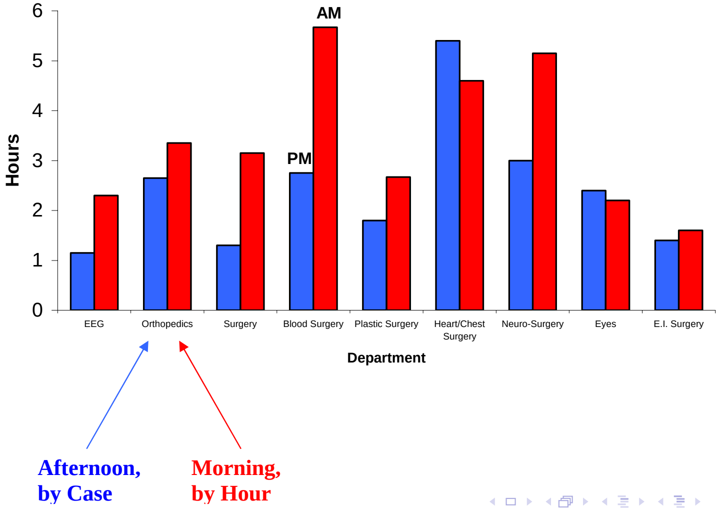
Unit / Parameter	Station 1 Tourism	Station 2 Teller	Total Dominant Path
Service Time	12.7	4.8	17.5
Waiting Time	8.2	6.9	15.1
Total Time	20.9	11.7	32.6
Service Index	**0.61**	**0.41**	**0.53**

Conceptual Model: Hospital (ED) Network (Sinreich)



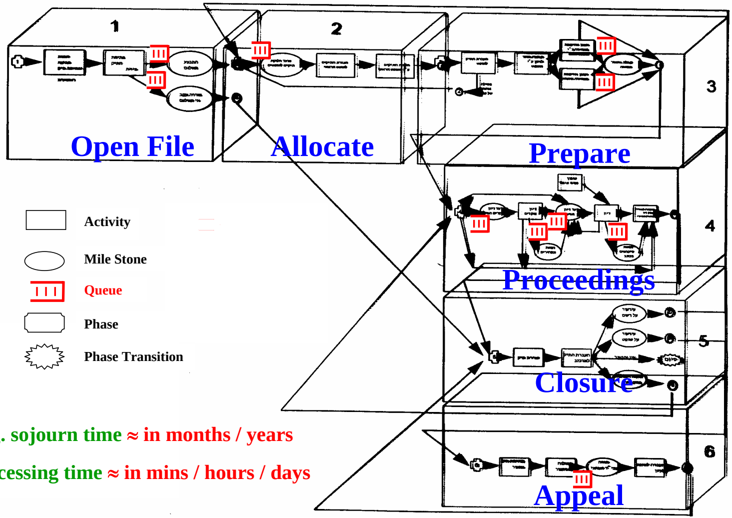
Descriptive Model: Service Times (Averages) or, Even "Doctors" Can Manage

Operations Time - Morning (by Hour) vs. Afternoon (by Case):



Conceptual Model: The "Production of Justice"

The Labor-Court Process in Haifa, Israel

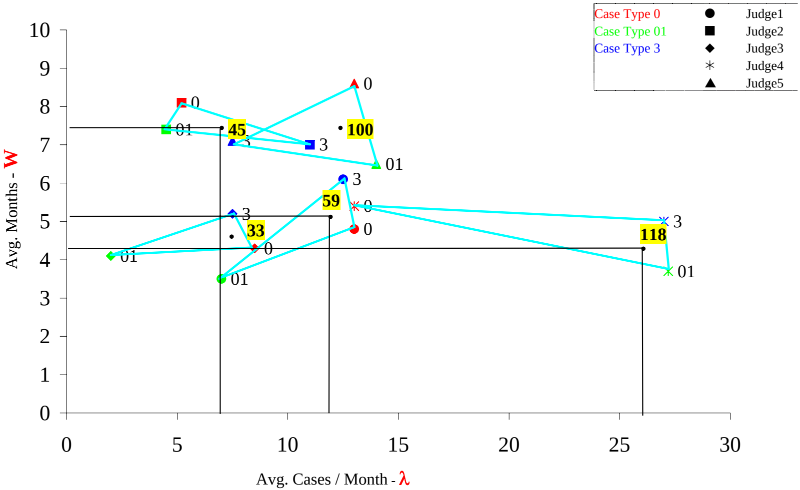


Avg. sojourn time ≈ in months / years

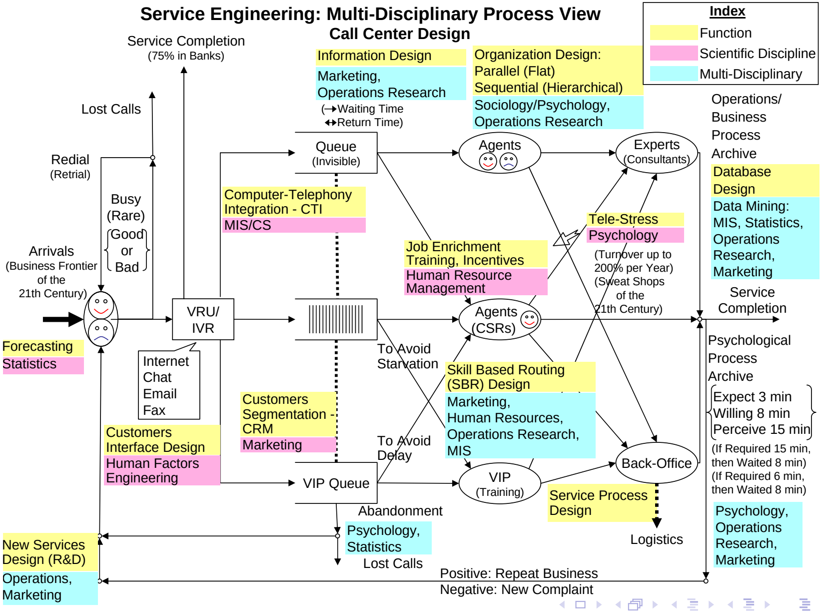
Processing time ≈ in mins / hours / days

Analytical Model: Little's Law in Court (still Averages)

Judges: The Best/Worst (Operational) Performer



Call-Center Network: Gallery of Models



The "Phases of Waiting" for Service

Common Experience:

- Expected to wait 5 minutes, Required to 10
- Felt like 20, Actually waited 10 (hence Willing $\geq$ 10)

An attempt at "Modeling the Experience":

1.	Time that a customer	**expects** to wait	
2.		**willing** to wait	((Im)Patience: τ)
3.		**required** to wait	(Offered Wait: V)
4.		**actually** waits	($W_q = \min(\tau, V)$)
5.		**perceives** waiting.	

Experienced customers $\Rightarrow$ Expected = Required

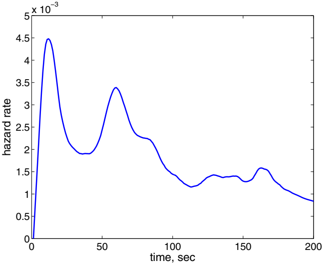
"Rational" customers $\Rightarrow$ Perceived = Actual.

Then left with (τ, V).

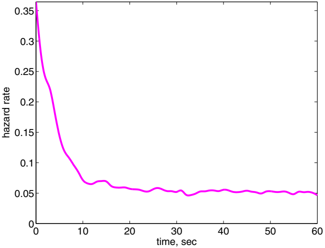
Call Center Data: Hazard Rates (Un-Censored)

(Im)Patience Time τ

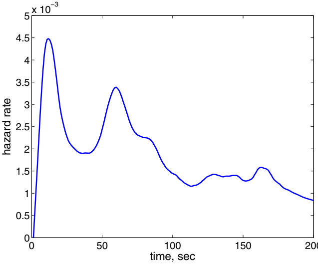
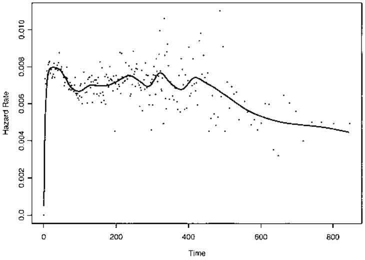
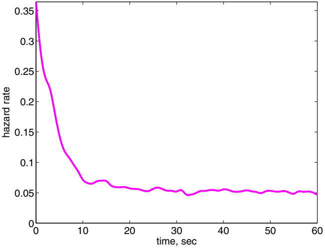
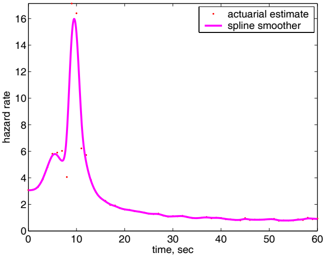
Israel



U.S.



Call Center Data: Hazard Rates (Un-Censored)

	(Im)Patience Time τ	Required/Offered Wait V
Israel		
U.S.		

A Patience Index

Quantifying (Im)Patience: “Willing to wait 15 min" = Patient?

$$\textbf{Theoretical}\ \text{Patience-Index} \ \triangleq\ \frac{\textbf{Willing}\ \text{to wait}}{\textbf{Expected}\ \text{to wait}} \ =\ \frac{\mathrm{E}[\tau]}{\mathrm{E}[V]},$$

“assuming" Experienced;

A Patience Index

Quantifying (Im)Patience: “Willing to wait 15 min" = Patient?

$$\textbf{Theoretical}\text{ Patience-Index } \triangleq \frac{\textbf{Willing}\text{ to wait}}{\textbf{Expected}\text{ to wait}} = \frac{\mathsf{E}[\tau]}{\mathsf{E}[V]},$$

“assuming" Experienced; further “assuming" that τ and V are Exponential, the M-L estimate of Index is the easily-measurable:

$$\textbf{Empirical}\text{ Patience-Index } \triangleq \frac{\textbf{\% Served}}{\textbf{\% Abandoning}}$$

A Patience Index

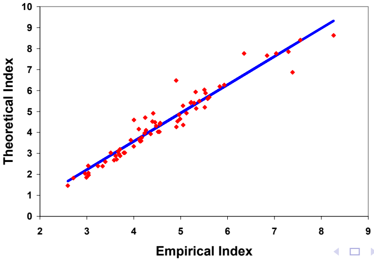
Quantifying (Im)Patience: "Willing to wait 15 min" = Patient?

$$\textbf{Theoretical}\ \text{Patience-Index} \triangleq \frac{\textbf{Willing}\ \text{to wait}}{\textbf{Expected}\ \text{to wait}} = \frac{E[\tau]}{E[V]},$$

"assuming" Experienced; further "assuming" that τ and V are Exponential, the M-L estimate of Index is the easily-measurable:

$$\textbf{Empirical}\ \text{Patience-Index} \triangleq \frac{\textbf{\% Served}}{\textbf{\% Abandoning}}$$

Index Validation: Theoretical vs. Empirical



Predicting Performance

Model **Primitives**:

- **Arrivals** to service (random process)
- **(Im)Patience** while waiting τ (r.v.)
- **Service** times (r.v.)
- **# Servers** / **Agents** (parameter / r.v.)

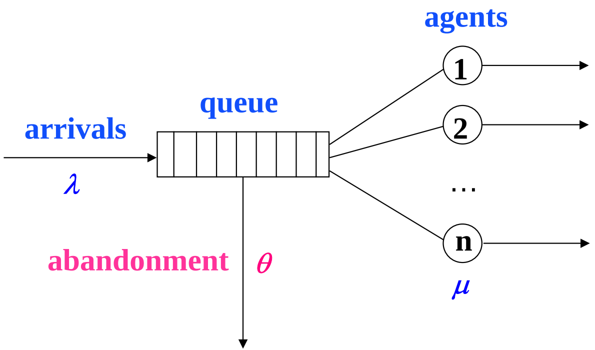
Model **Output**: **Offered-Wait** V (r.v.)

Operational Performance Measure calculable in terms of (τ, V).

- eg. Average Wait = $E[\min\{\tau, V\}]$
- eg. % Abandonment = $P\{\tau < V\}$

Application: **Staffing – How Many Agents?** (When? Who?)

The Basic Staffing Model: Erlang-A (M/M/n +M)



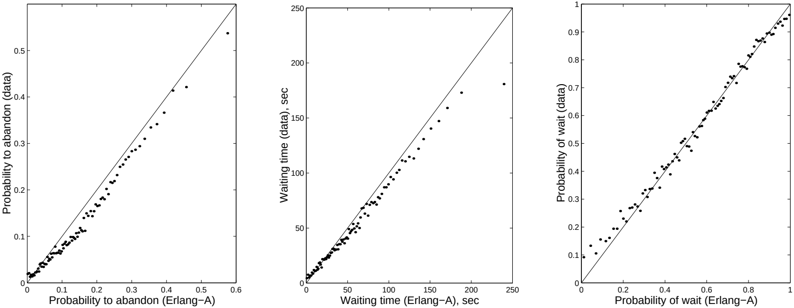
Erlang-A Parameters:

- λ – **Arrival** rate (Poisson)
- μ – **Service** rate (Exponential)
- θ – **Impatience** rate (Exponential)
- n – Number of **Service-Agents**.

Erlang-A: Fitting a Simple Model to a Complex Reality

Hourly Performance vs. Erlang-A Predictions (1 year)

% Abandon | **E[*Wait*]** | **%{*Wait* > 0}**



- Empirically-Based & Theoretically-Supported Estimation of (Im)Patience: $\hat{\theta} = P\{Ab\}/\mathrm{E}[W_q]$)
- Small Israeli Bank (more examples in progress)

Testing the Erlang-A Primitives

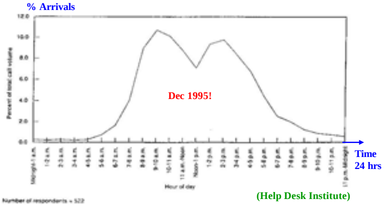
- **Arrivals**: Poisson?
- **Service-durations**: Exponential?
- **(Im)Patience**: Exponential?

Validation: Support? Refute?

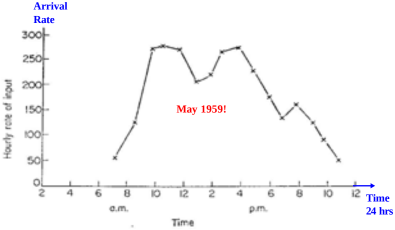
Arrivals to Service: only Poisson-Relatives

Arrival Rate to Three Call Centers

Dec. **1995** (U.S. 700 Helpdesks)



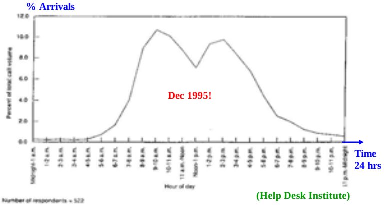
May **1959** (England)



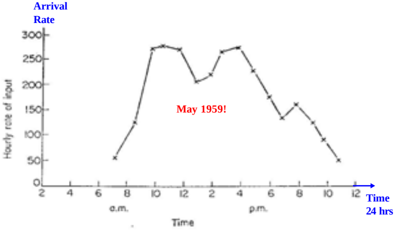
Arrivals to Service: only Poisson-Relatives

Arrival Rate to Three Call Centers

Dec. **1995** (U.S. 700 Helpdesks)



May **1959** (England)



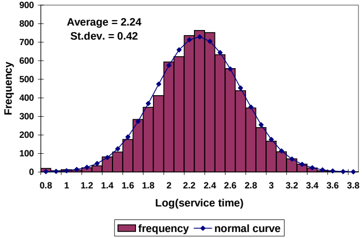
November **1999** (Israel)

Observation:

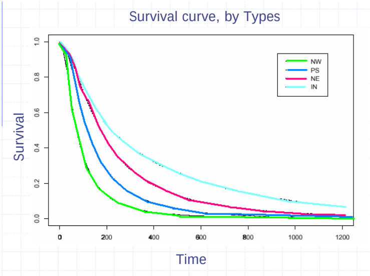
Peak Loads at **10:00 & 15:00**

Service Durations: LogNormal Prevalent

Israeli Bank Log-Histogram



Survival-Functions by Service-Class

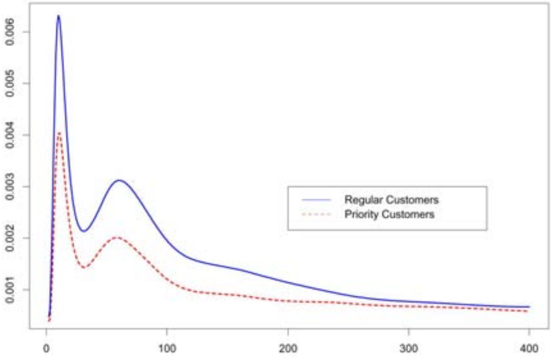


- **New** Customers: 2 min (NW);
- **Regulars**: 3 min (PS);
- **Stock**: 4.5 min (NE);
- Tech-Support: 6.5 min (IN).

Observation: **VIP** require **longer service** times.

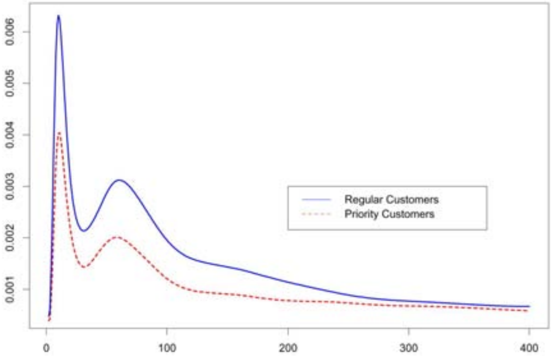
(Im)Patience while Waiting (Palm 1943-53)

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



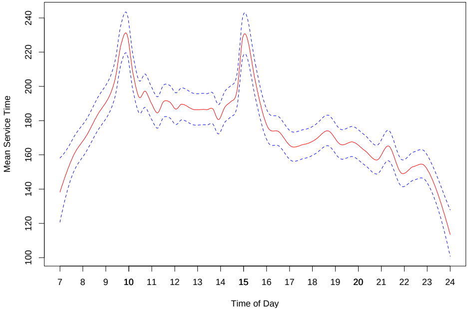
- **Peaks** of abandonment at times of **Announcements**
- **Call-by-Call Data (DataMOCCA)** required (& Un-Censoring).

Observation: **VIP** are **more patient** (Needy)

A "Service-Time" Puzzle at a Small Israeli Bank

Inter-related Primitives

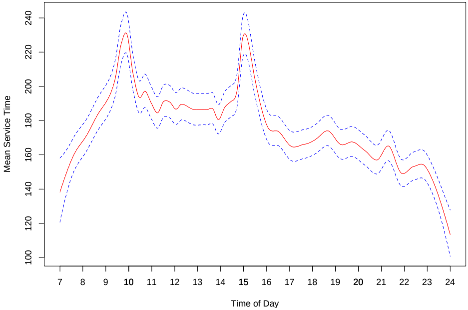
Average Service Time over the Day – Israeli Bank



Prevalent: **Longest services** at **peak-loads** (10:00, 15:00). **Why?**

A “Service-Time" Puzzle at a Small Israeli Bank
Inter-related Primitives

Average Service Time over the Day – Israeli Bank



Prevalent: **Longest services** at **peak-loads** (10:00, 15:00). **Why?**

Explanations:

- Prevalent: Service protocol different (longer) during peak times.
- Operational: The **needy** abandon **less** during peak times; hence the **VIP** remain on line, with their **longer** service times.

Erlang-A: Simple, but Not Too Simple

Experience:

- Arrival process **not pure Poisson** (time-varying, σ^2 too large)
- Service times **not exponential** (typically close to lognormal)
- Patience times **not exponential** (various patterns observed).
- Customers and Servers **not homogeneous** (classes, skills)

Erlang-A: Simple, but Not Too Simple

Experience:

- Arrival process **not pure Poisson** (time-varying, σ^2 too large)
- Service times **not exponential** (typically close to lognormal)
- Patience times **not exponential** (various patterns observed).
- Customers and Servers **not homogeneous** (classes, skills)

Questions naturally arise:

1. Why does Erlang-A practically work? justify **robustness**.
2. When does it fail? chart **boundaries**.
3. Generalize: time-variation, SBR, networks, uncertainty , ...

Erlang-A: Simple, but Not Too Simple

Experience:

- Arrival process **not pure Poisson** (time-varying, σ^2 too large)
- Service times **not exponential** (typically close to lognormal)
- Patience times **not exponential** (various patterns observed).
- Customers and Servers **not homogeneous** (classes, skills)

Questions naturally arise:

1. Why does Erlang-A practically work? justify **robustness**.
2. When does it fail? chart **boundaries**.
3. Generalize: time-variation, SBR, networks, uncertainty , . . .

Answers via **Asymptotic Analysis**, as load- and staffing-levels $\uparrow$:

- **E**fficiency-**D**riven (**ED**) regime: Fluid models (Whitt; Harrison, Zeevi; Bassamboo).
- **Q**uality- and **E**fficiency-**D**riven (**QED**) regime: Diffusion refinements (Erlang, 1913; Halfin-Whitt, 1981; present surge).

DataMOCCA = Data MOdels for Call Center Analysis

- **Technion**: P. Feigin, V. Trofimov, Statistics / SEE Laboratory.
- **Wharton**: L. Brown, N. Gans, H. Shen (UNC).
- **industry**:
  - U.S. Bank: **2.5 years, 220M calls, 40M by 1000 agents**.
  - Israeli Cellular: **2.5 years, 110M calls, 25M calls by 750 agents; ongoing**.

DataMOCCA = Data MOdels for Call Center Analysis

- **Technion**: P. Feigin, V. Trofimov, Statistics / SEE Laboratory.
- **Wharton**: L. Brown, N. Gans, H. Shen (UNC).
- **industry**:
  - U.S. Bank: **2.5 years, 220M calls, 40M by 1000 agents**.
  - Israeli Cellular: **2.5 years, 110M calls, 25M calls by 750 agents; ongoing**.

Project Goal: Designing and Implementing a (universal) data-base/data-repository and interface for storing, retrieving, analyzing and displaying **Call-by-Call-based Data / Information**.

DataMOCCA = Data MOdels for Call Center Analysis

- **Technion**: P. Feigin, V. Trofimov, Statistics / SEE Laboratory.
- **Wharton**: L. Brown, N. Gans, H. Shen (UNC).
- **industry**:
  - U.S. Bank: **2.5 years, 220M calls, 40M by 1000 agents**.
  - Israeli Cellular: **2.5 years, 110M calls, 25M calls by 750 agents; ongoing**.

Project Goal: Designing and Implementing a (universal) data-base/data-repository and interface for storing, retrieving, analyzing and displaying **Call-by-Call-based Data / Information**.

System Components:

- **Clean Databases**: operational-data of individual calls / agents.
- **Graphical Online Interface**: easily generates graphs and tables, at varying resolutions (seconds, minutes, hours, days, months).

DataMOCCA = Data MOdels for Call Center Analysis

- **Technion**: P. Feigin, V. Trofimov, Statistics / SEE Laboratory.
- **Wharton**: L. Brown, N. Gans, H. Shen (UNC).
- **industry**:
  - U.S. Bank: **2.5 years, 220M calls, 40M by 1000 agents**.
  - Israeli Cellular: **2.5 years, 110M calls, 25M calls by 750 agents; ongoing**.

Project Goal: Designing and Implementing a (universal) data-base/data-repository and interface for storing, retrieving, analyzing and displaying **Call-by-Call-based Data / Information**.

System Components:

- **Clean Databases**: operational-data of individual calls / agents.
- **Graphical Online Interface**: easily generates graphs and tables, at varying resolutions (seconds, minutes, hours, days, months).

Free for academic adoption: **7GB tables**, or **20GB raw zipped**, for each call center – **ask for my mini-HD**.

Call Centers: Hierarchical Operational View

Forecasting Customers (Statistics), Agents (HRM)

Staffing: **Queueing Theory** (Erlang-A based)

Service Level, Costs

FTE's (Seats)
per unit of time

Shifts: IP, Combinatorial Optimization; LP

Union constraints, Costs

Shift structure

Rostering: Heuristics, AI (Complex)

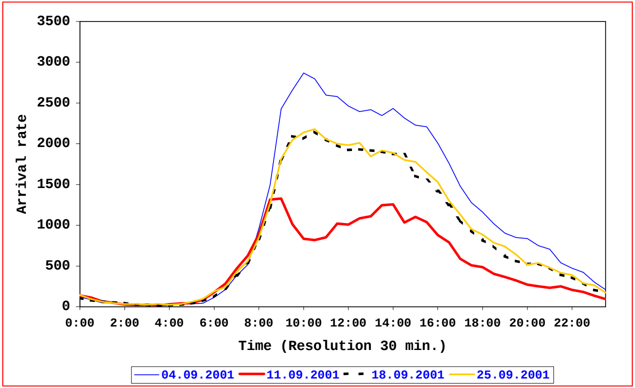
Individual constraints

Agents Assignments

Skills-based Routing: Stochastic Control (of Q's)

Arrivals to Service: Predictable vs. Random

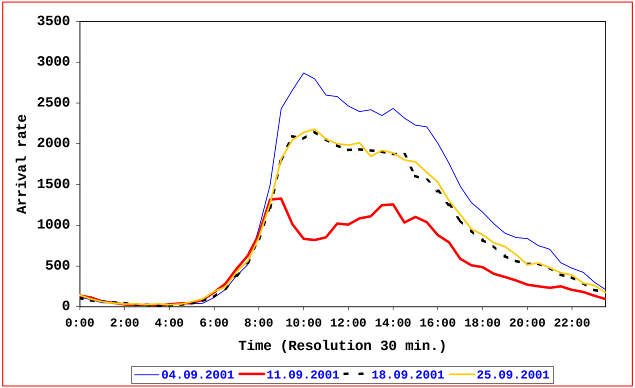
Arrival Rates on Tuesdays in a September – U.S. Bank



- **Tuesday**, September 4th: **Heavy**, following Labor Day.
- **Tuesdays**, September 18 & 25: **Normal**.

Arrivals to Service: Predictable vs. Random

Arrival Rates on Tuesdays in a September – U.S. Bank

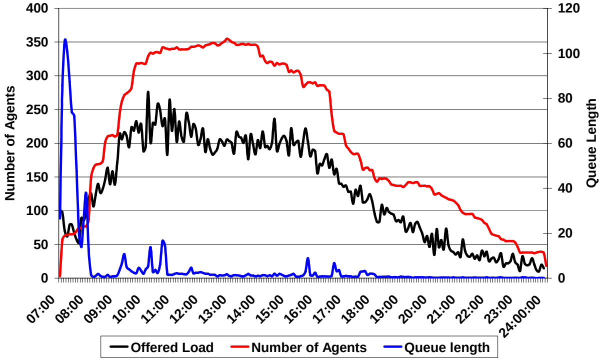


- **Tuesday**, September 4th: **Heavy**, following Labor Day.
- **Tuesdays**, September 18 & 25: **Normal**.
- **Tuesday, September 11th, 2001.**

When: Shift Scheduling

Integer Programming, given piecewise-constant Staffing Levels.

U.S. Bank: Queue-length and Staffing on May 3, 2002

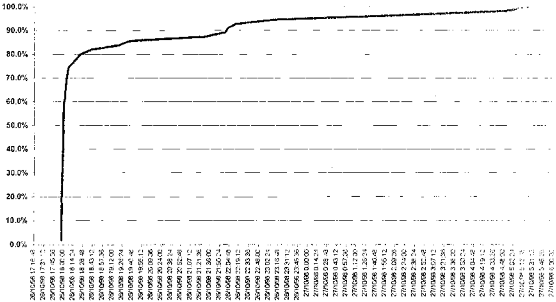


Shift scheduling **matters**: **consistent under-staffing at 7:00am**.

Who: Rostering

Assigning **individual agents** to **shifts**.
Typically, heuristics (AI) to accommodate individual constraints.

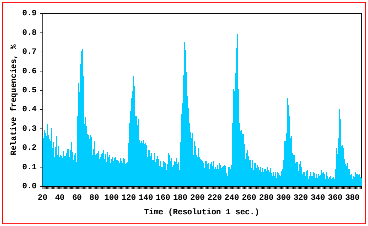
Israeli Technical Support Call Center: Online Shift Bidding



Shift-bidding starts at **18:00**.

- **60% of "successful"** agents are registered till **18:00**.
- **80% till 18:24**; 90% till 22:00; registration closed at 5:23am.

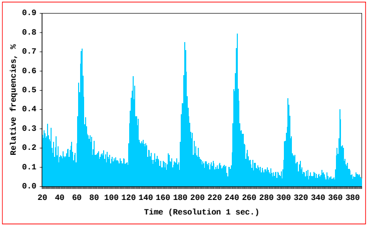
A “Waiting-Times" Puzzle at a Medium Israeli Bank



Peaks Every 60 Seconds. Why?

- Human: **Voice-announcement** every 60 seconds.
- System: **Priority-upgrade** (unrevealed) every 60 seconds.

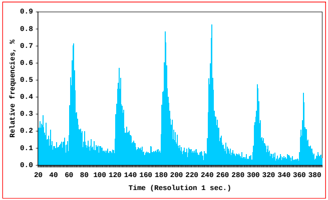
A “Waiting-Times" Puzzle at a Medium Israeli Bank



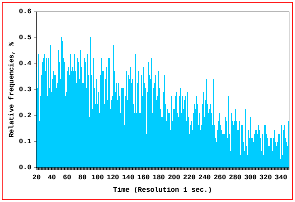
Peaks Every 60 Seconds. Why?

- Human: **Voice-announcement** every 60 seconds.
- System: **Priority-upgrade** (unrevealed) every 60 seconds.

Served Customers



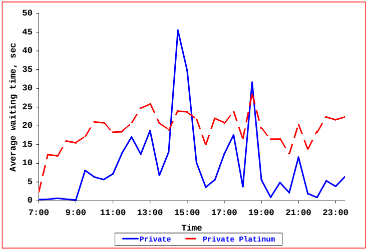
Abandoning Customers



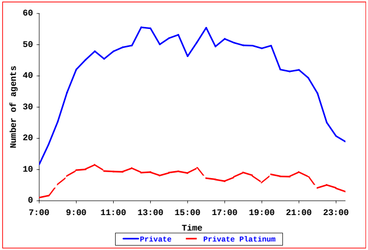
Priorities, Economies-of-Scale, SBR

Regular vs. VIP Customers: Cellular – March 23, 2004

Average Wait



Staffing Level

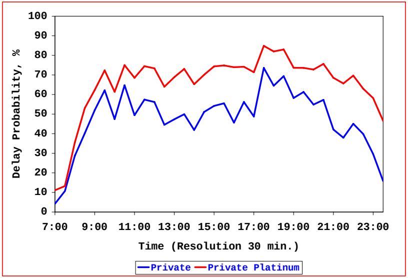


- **Design:** VIP-dedicated agents, Regular-dedicated Agents.
- **VIP's** are **not** served better than **Regular's**
- **Solutions:** Add VIP agents (costly), or Re-Design.

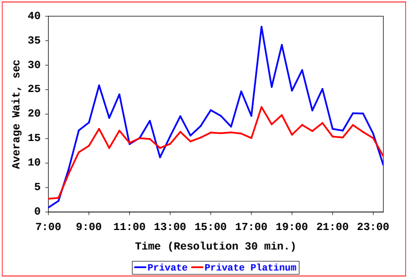
Priorities and Routing Protocols I

Regular vs. VIP Customers: Cellular – October 2004

Delay Probability



Average Wait



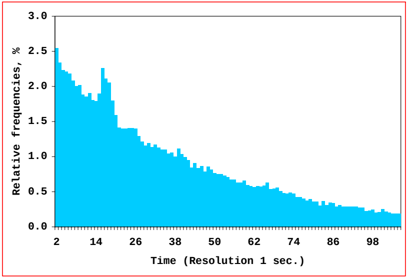
More **VIPs** delayed than **Regulars**, **yet** their average wait is shorter.

What changed since last March?

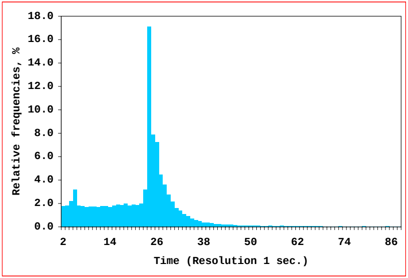
Priorities and Routing Protocols II

Waiting-Time Histograms: Cellular – October 2004

Regular Customers

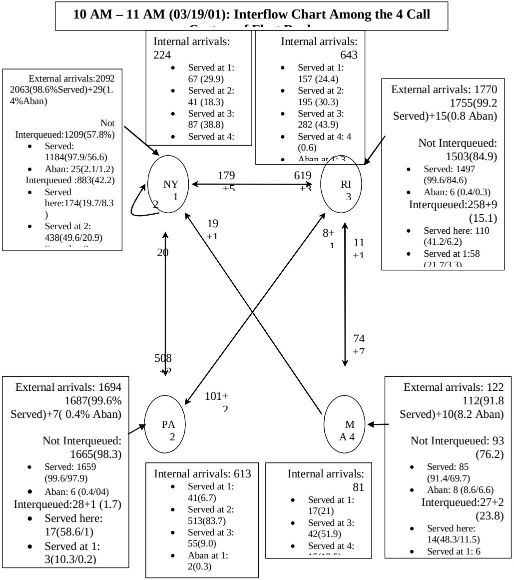


VIP (Platinum) Customers



After **25 seconds** of wait, **VIP** customers are **routed** with **high priority** to Regular agents. Hence, almost **no long waiting times** for VIP's.

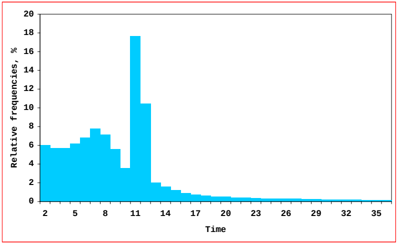
Network Balancing via Inter-Queues at a U.S. Bank



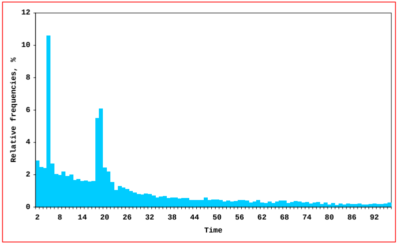
Balancing Protocols and Performance Level

U.S. Bank: Histograms of Waiting Times

Retail



Business



Peak for **Retail** service at **10 seconds** – **Why?**
After 10 seconds of wait, **Retail** customers sent into the **inter-queue**.

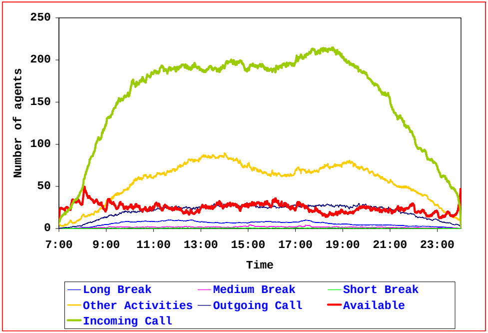
Business customers – peak at **5 seconds**, for the same reason.
Second peak – unclear, maybe priority-upgrade.

The Planning-Reality Gap: Agent Status

Erlang-A Model $\Rightarrow$ optimal **Staffing Level** n.

n = number-of-agents that show up? serve? **No!**

Israeli Bank, Agent Status: Monthly Averages



n (FTE) = Busy with “Incoming Calls" + “Available" for service.

Data-Based Service-Research (with DataMOCCA, even before tenure)

- Contrast with EmpOM: Industry / Company / Survey Data (Social Sciences)
- Converge to: Measure, Model, Validate, Experiment, Refine (**Physics, Biology,** ...).

Data-Based Service-Research (with DataMOCCA, even before tenure)

- Contrast with EmpOM: Industry / Company / Survey Data (Social Sciences)
- Converge to: Measure, Model, Validate, Experiment, Refine (**Physics, Biology,** . . .).
- Prerequisites:
  - OR, OM, IE, (Mktg.) - for **Design**
  - CS, IS, Stat. - for **Implementation**.

Data-Based Service-Research (with DataMOCCA, even before tenure)

- Contrast with EmpOM: Industry / Company / Survey Data (Social Sciences)
- Converge to: Measure, Model, Validate, Experiment, Refine (**Physics, Biology,** . . .).
- Prerequisites:
  - OR, OM, IE, (Mktg.) - for **Design**
  - CS, IS, Stat. - for **Implementation**.
- Outcomes: Relevance, Credibility; Interest, Fun;
  Call Centers as a Pilot (eg. for Healthcare). Moreover,
  - **Teaching**: Class, Homework (Experimental Data Analysis); Cases.
  - **Research**: Validate Existing (Queueing) Theory/Laws and Suggest New Models/Research.
  - **Practice**: OM Tools (Scenario Analysis), Mktg. (Trends, Benchmarking).

Live Demonstration of DataMOCCA

5-7 minutes, to emphasize **"online"** capabilities.

U.S. Bank

- Daily Reports: October 2003, weekdays; typically takes 10-20 sec till a first output, but this is because of PowerPoint/Windows. Then do few additional Daily Reports, say Monday, Tuesday,... (starting with STATCCA, as opposed to by minimizing the powerpoint screnn) - this will be now happening very fast.
- Time-Series: Number of agents, for ALL classes, all months, weekdays. (Including total). Shows scale, trends. Then do Service Durations, indicating that 1 second of 1000 agents could cost $500M per year. Could also do Unhandled (lower middle entry in list), for only Retail and Premium - Premium is worse, and deteriorating,
- Daily Summaries:
  - Tuesdays in September 2001; September 11th; shown during lecture under the heading "Predictable or Random";
  - 30 sec scale - stoch. variability, 1 hour scale - the "right" scale;
  - % to mean, to show very similar shape over 3 Tuesdays. Suggests the model $\lambda(t) = \lambda_0(t) \cdot Z$, for $0 \le t \le T$.