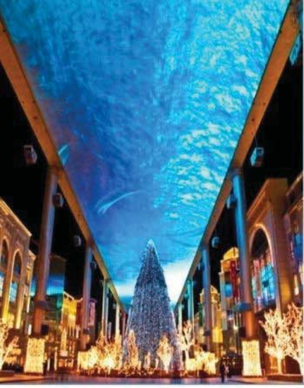
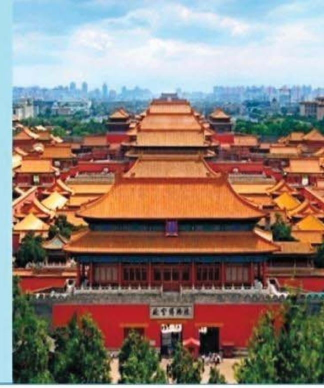
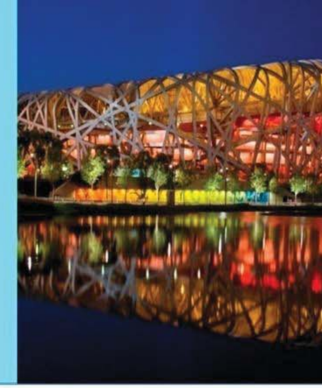
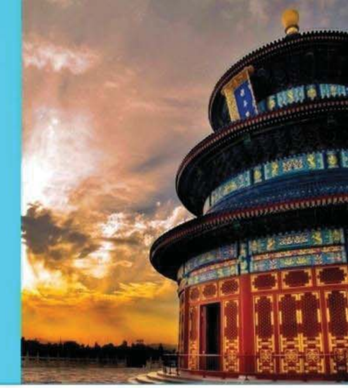


informs

INFORMS
INTERNATIONAL
Beijing 2012

Data-Based Service Networks:

A Research Framework for

Asymptotic Inference, Analysis & Control of Service Systems

Avi Mandelbaum

Technion, Haifa, Israel

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

INFORMS Beijing, June 2012

Research Partners

- **Students**:
  Aldor*, Baron*, Carmeli, Feldman*, Garnett*, Gurvich*, Huang, Khudiakov*, Maman*, Marmor*, Reich, Rosenshmidt*, Shaikhet*, Senderovic, Tseytlin*, Yom-Tov*, Yuviler, Zaied, Zeltyn*, Zychlinski, Zohar*, Zviran*, . . .
- **Theory**:
  Armony, Atar, Gurvich, Jelenkovic, Kaspi, Massey, Momcilovic, Reiman, Shimkin, Stolyar, Wasserkrug, Whitt, Zeltyn, . . .
- **Industry**:
  Mizrahi Bank (A. Cohen, U. Yonissi), Rambam Hospital (R. Beyar, S. Israelit, S. Tzafrir), IBM Research (OCR Project), Hapoalim Bank (G. Maklef, T. Shlasky), Pelephone Cellular, . . .
- **Technion SEE Center / Laboratory**:
  Feigin; Trofimov, Nadjharov, Gavako, Kutsy; Liberman, Koren, Plonsky, Senderovic; Research Assistants, . . .
- **Empirical/Statistical Analysis**:
  Brown, Gans, Zhao; Shen; Ritov, Goldberg; Gurvich, Huang, Liberman; Armony, Marmor, Tseytlin, Yom-Tov; Zeltyn, Nardi, Gorfine, . . .

Contents

- Service Systems: Call Centers, Hospitals, Websites, $\cdots$
- Prevalent paradigm of modeling and approximations
- Redefining the role of Asymptotics, via Data
- **ServNets: QNets, SimNets; FNets, DNets**
- **Ultimate Goal: Data-based creation and validation of ServNets, automatically in real-time**

Contents

- Service Systems: Call Centers, Hospitals, Websites, $\cdots$
- Prevalent paradigm of modeling and approximations
- Redefining the role of Asymptotics, via Data
- **ServNets: QNets, SimNets; FNets, DNets**
- **Ultimate Goal: Data-based creation and validation of ServNets, automatically in real-time**
- Why be Optimistic? Pilot at the Technion SEELab

Contents

- Service Systems: Call Centers, Hospitals, Websites, $\cdots$
- Prevalent paradigm of modeling and approximations
- Redefining the role of Asymptotics, via Data
- **ServNets: QNets, SimNets; FNets, DNets**
- **Ultimate Goal: Data-based creation and validation of ServNets, automatically in real-time**
- Why be Optimistic? Pilot at the Technion SEELab
- Related themes: Process Mining (BPM), Networks (Social, Biological,$\cdots$), "Almost-Complex" Systems, Simulation-based. . .,
- Scenic Route: Open Problems, New Directions, Uncharted Territories

On Asymptotic Research of Queueing Systems

Queueing asymptotics has grown to become a central research theme in Operations Research and Applied Probability, beyond just queueing theory. Its claim to fame has been the deep insights that it provides into the dynamics of Queueing Networks (**QNets**), and rightly so:

On Asymptotic Research of Queueing Systems

Queueing asymptotics has grown to become a central research theme in Operations Research and Applied Probability, beyond just queueing theory. Its claim to fame has been the deep insights that it provides into the dynamics of Queueing Networks (**QNets**), and rightly so:

- Kingman's invariance principle in conventional heavy-traffic
- Whitt's sample-path (functional) framework
- Reiman's network analysis via oblique reflection
- Bramson-Williams' framework for state-space collapse
- Laws' resource pooling
- Harrison's paradigm for asymptotic control (Wein; van Mieghem's $Gc\mu$)
- Dai's fluid-based stability
- Halfin-Whitt's (QED regime) ($\sqrt{}$-staffing for many-server queues)
- $P = NP$: Atar's equivalence of Preemptive and Non-Preemptive SBR; Stolyar, Gurvich
- Massey-Whitt's research of time-varying queues

Applying Queueing Asymptotics

Has it helped one approximate or simulate a service system more efficiently, estimate its parameters more accurately, teach it to our students more effectively, perhaps even manage the system better?

I am of the opinion that the answers to such questions have been too often negative, that positive answers must have theory and applications nurture each other, which is good, and my approach to make this good happen is by marrying theory with data.

Prevalent Asymptotic Approximations

System (Data)

Phenomenology

Accuracy

Models ⟷ Approximations

QNets SimNets

Data–Based Prevalent Asymptotic Approximations **Models**

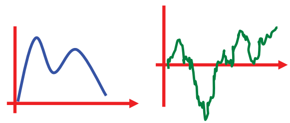
System (Data)

Phenomenology

Accuracy

Models ⟷ Approximations

QNets SimNets ⟷ FNet DNet **Models**



Data–Based Prevalent Asymptotic Approximations Models

System (Data)

Phenomenology

Value

Accuracy

Models

Approximations

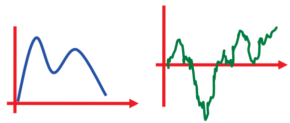
QNets

SimNets

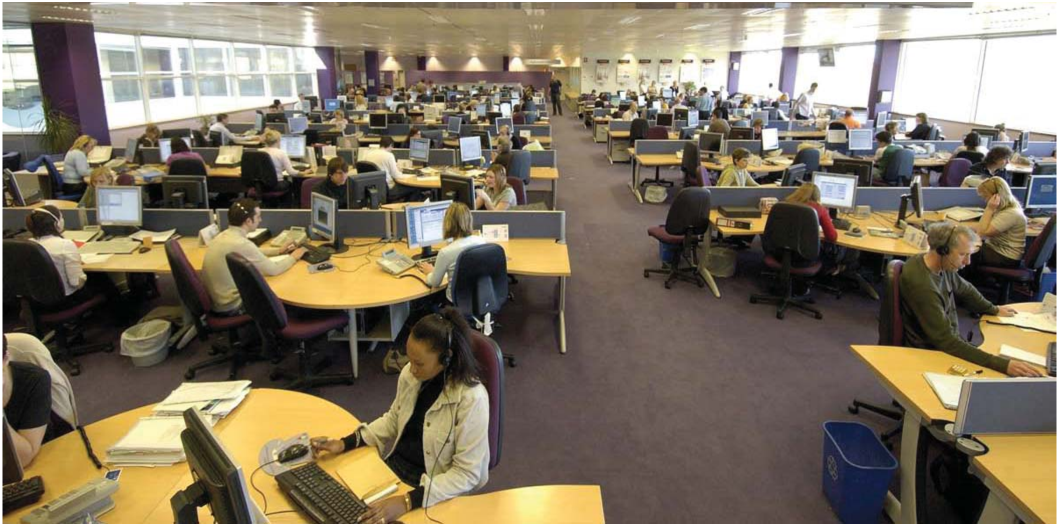
FNet

DNet

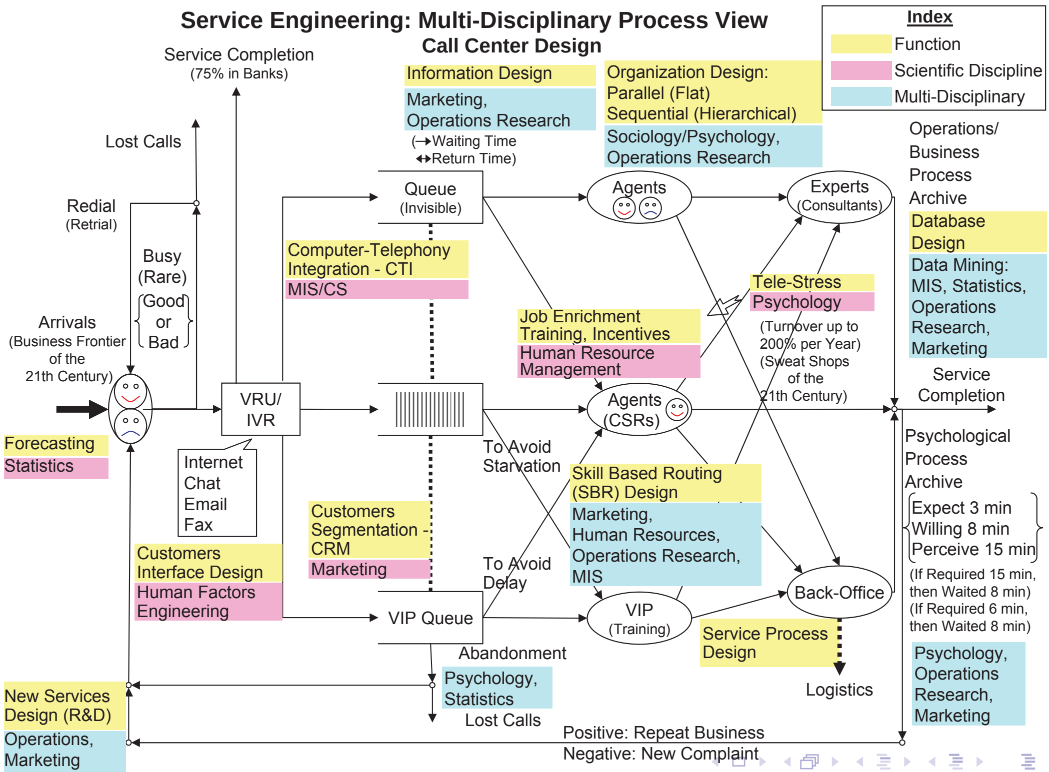
Models



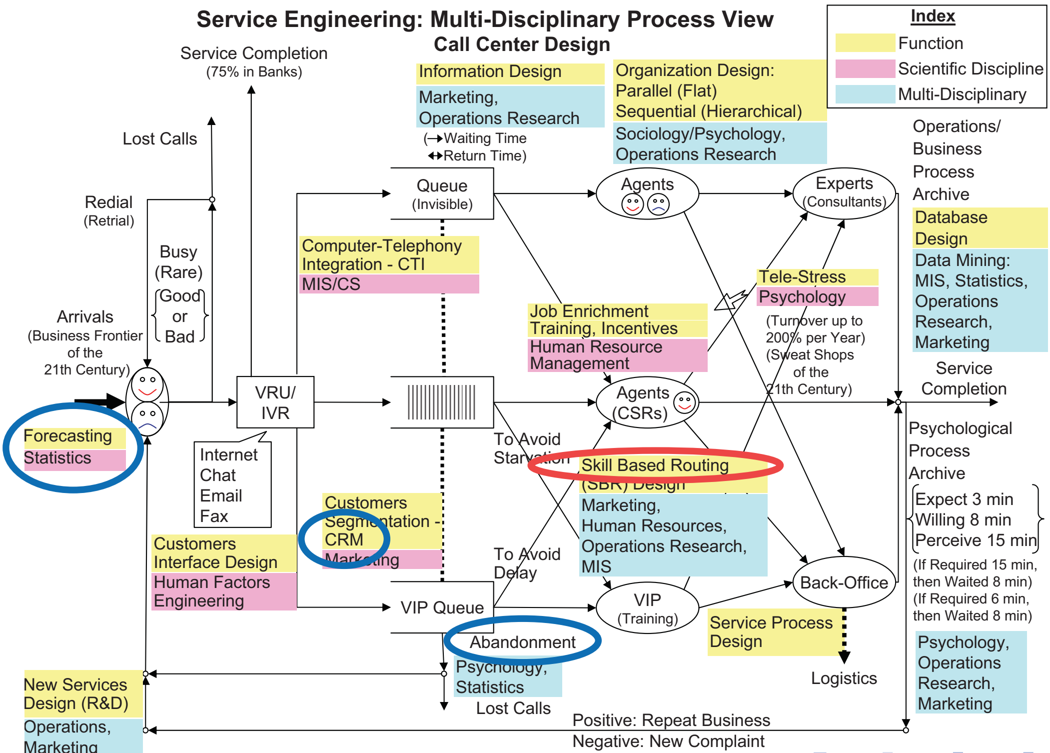
Call-Center Environment: Service Network



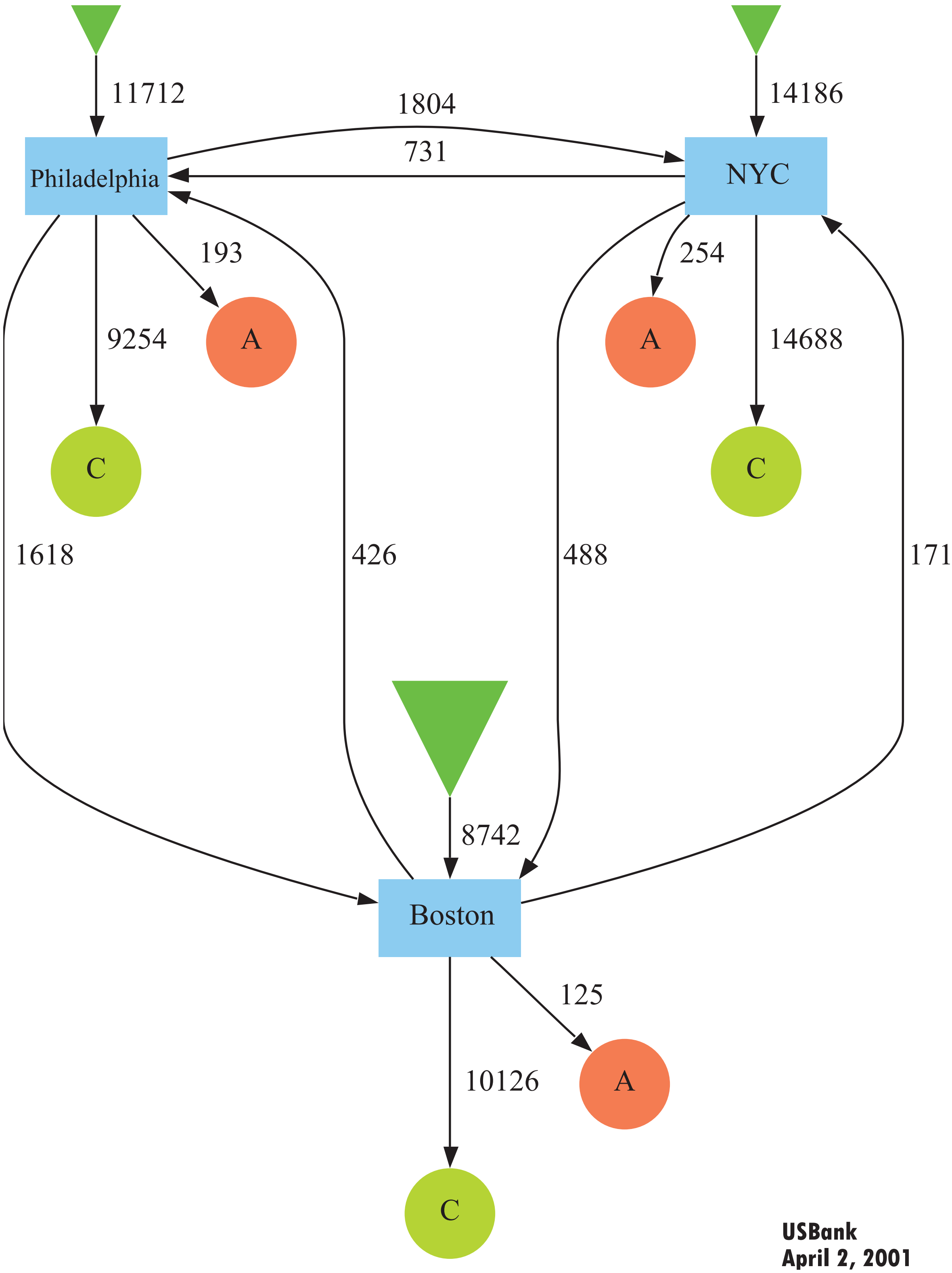
Call-Center Network: Gallery of Models



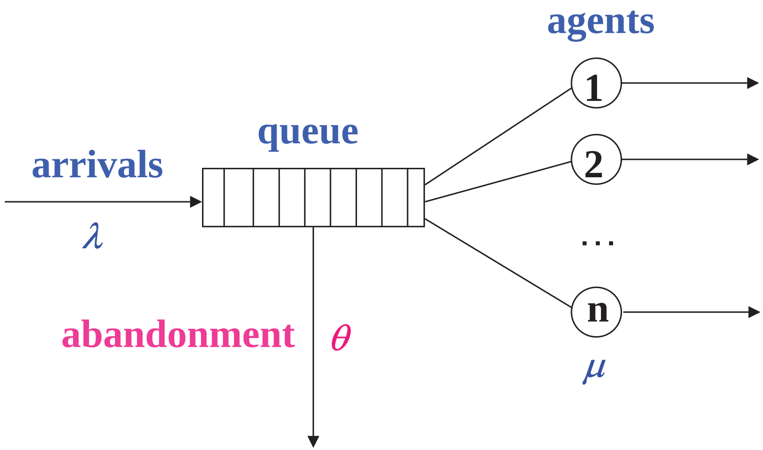
Call-Center Network: Gallery of Models




EBO
Business
Platinum
Subanco
Premier
Consumer Loans
Online Banking
Telesales
Case Quality
Priority Service
AST
CCO
Quick&Reilly
BPS
639
51
555
2372
597
91
195
325
54
1455
257
342
3660
2173
3251
487
32
12
619
2352
3668
4844
447
1_1
7
1_5
3_1
3_5
1_15
2_3
2_6
5_1
6
8
10_1
11
12
13
15
17
10430
2947
18993
40
384
98
157
Retail


11712
14186
1804
731
Philadelphia
NYC
193
254
9254
A
A
14688
C
C
1618
426
488
171
8742
Boston
125
10126
A
C

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



Erlang-A **(Palm 1940's)** = **Birth & Death Q**, with parameters:

- λ – **Arrival** rate (Poisson)
- μ – **Service** rate (Exponential; $E[S] = \frac{1}{\mu}$)
- θ – **Patience** rate (Exponential, $E[\text{Patience}] = \frac{1}{\theta}$)
- N – Number of **Servers** (Agents).

Erlang-A: Practical Relevance?

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).

Erlang-A: Practical Relevance?

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).

- Building Blocks need **not be independent** (eg. long wait associated with long service; with **w/ M. Reich and Y. Ritov**)
- Customers and Servers **not homogeneous** (classes, skills)
- Customers return for service (after busy, abandonment; dependently; **P. Khudiakov, M. Gorfine, P. Feigin**)
- $\cdots$, and more.

Erlang-A: Practical Relevance?

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).
- Building Blocks need **not be independent** (eg. long wait associated with long service; with **w/ M. Reich and Y. Ritov**)
- Customers and Servers **not homogeneous** (classes, skills)
- Customers return for service (after busy, abandonment; dependently; **P. Khudiakov, M. Gorfine, P. Feigin**)
- $\cdots$, and more.

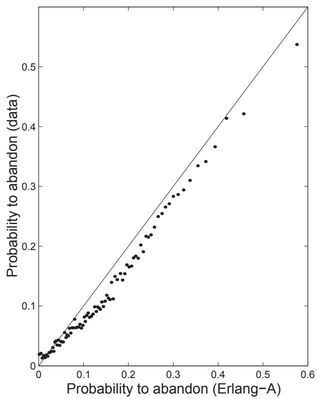
Question: **Is Erlang-A Relevant?**

YES ! **Fitting a Simple Model to a Complex Reality**, both **Theoretically** and **Practically**

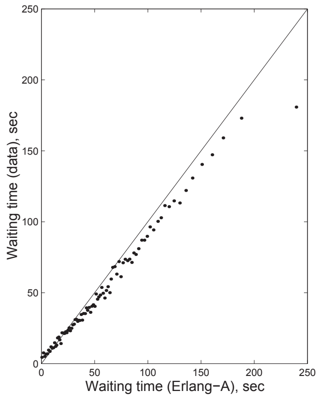
Erlang-A: Fitting a Simple Model to a Complex Reality

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

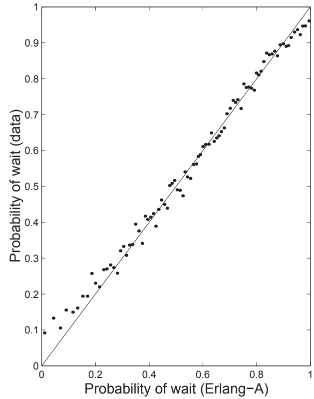
% Abandon



E[$Wait$]

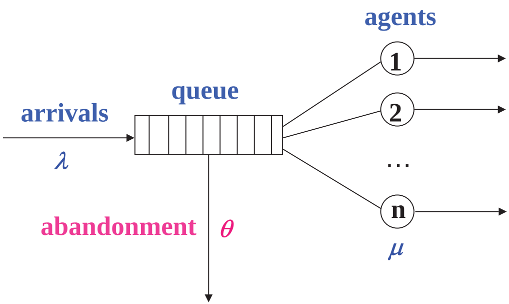


%$\{Wait > 0\}$



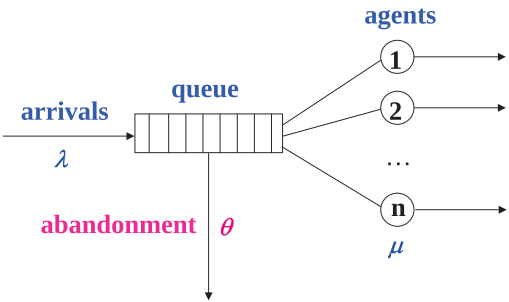
- Empirically-Based & Theoretically-Supported Estimation of (Im)Patience: $\hat{\theta} = P\{Ab\}/\mathrm{E}[W_q]$)
- Small Israeli Bank (more examples in progress)
- Hourly performance vs. Erlang-A predictions, 1 year: aggregated groups of 40 similar hours

Universal Approximations: Erlang-A (M/M/N + M)



w/ **I. Gurvich & J. Huang**

Universal Approximations: Erlang-A (M/M/N + M)

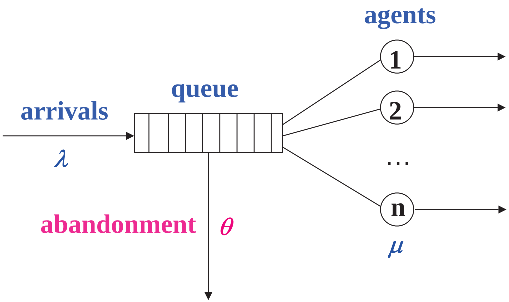


w/ **I. Gurvich & J. Huang**

- **QNet**: Birth & Death Queue, with B - D rates

$$F(q) = \lambda - \mu \cdot (q \wedge n) - \theta \cdot (q - n)^+, \quad q = 0, 1, \ldots$$

Universal Approximations: Erlang-A (M/M/N + M)



w/ **I. Gurvich & J. Huang**

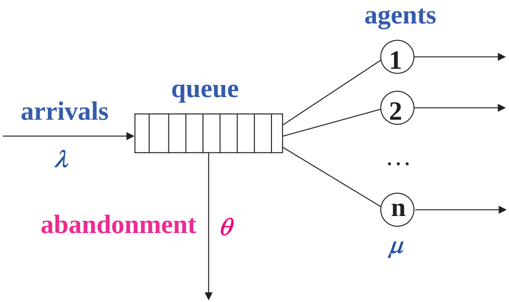
- **QNet**: Birth & Death Queue, with B - D rates

$$F(q) = \lambda - \mu \cdot (q \wedge n) - \theta \cdot (q - n)^+, \quad q = 0, 1, \ldots$$

- **FNet**: Dynamical (Deterministic) System – ODE

$$dx_t = F(x_t)dt, \quad t \geq 0$$

Universal Approximations: Erlang-A (M/M/N + M)



w/ **I. Gurvich & J. Huang**

- **QNet**: Birth & Death Queue, with B - D rates

$$F(q) = \lambda - \mu \cdot (q \wedge n) - \theta \cdot (q - n)^+, \quad q = 0, 1, \ldots$$

- **FNet**: Dynamical (Deterministic) System – ODE

$$dx_t = F(x_t)dt, \quad t \geq 0$$

- **DNet**: Universal (Stochastic) Approximation – SDE

$$dY_t = F(Y_t)dt + \sqrt{2\lambda}\, dB_t, \quad t \geq 0$$

eg. $\mu = \theta$, or $n = \infty$: $\quad \dot{x} = \lambda - \mu \cdot x, \quad Y = \text{OU process}$

Accuracy increases as $\lambda \uparrow \infty$ (no additional assumptions)

Universal Approximation: Why 2λ?

- Semi-martingale representation of the B&D process: Fluid + Martingale
- Predictable quadratic variation:

$$\int_0^t [\lambda + \mu(Q_s \wedge n) + \theta(Q_s - n)^+] ds$$

- In steady-state, arrival rate $\equiv$ departure rate:

$$\lambda = \mathbb{E}[\mu(Q_s \wedge n) + \theta(Q_s - n)^+]$$

- Expectation of the predictable quadratic variation:

$$\mathbb{E} \int_0^t [\lambda + \mu(Q_s \wedge n) + \theta(Q_s - n)^+] ds = 2\lambda t$$

- $\text{dMartingale}_t \approx \sqrt{2\lambda} \cdot \text{dBrownian}_t$

Value of Universal Approximation

- Tractable - closed-form stable expressions
- Accurate - more than heavy traffic limits
- Robust - all many-server regimes, and beyond, with hardly any assumptions
- Value
  - Performance Analysis
  - Optimization (Staffing)
  - Inference (**w/ G. Pang**)
  - Simulation (**w. J. Blanchet**)
- Limitation: Steady-State (but working on it)

Data–Based Prevalent Asymptotic Approximations Models

System (Data)

Phenomenology

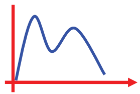
Value

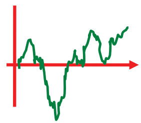
Accuracy

Models ⟷ Approximations

QNets SimNets ⟷ FNet DNet Models

- Inference SimLab
- Control
- Design
- Predictions





Data–Based Prevalent Asymptotic Approximations Models

System (Data)

Phenomenology

Value

Accuracy

Models ⟷ Approximations

QNets SimNets ⟷ FNet DNet Models

System = Coin Tossing, **Model** = Binomial ; de Moivre 1738

Approx. / Models: SLLN **(FNets)**, CLT **(DNets)** ; Laplace 1810

Value: Exceeds Value of originating stylized model

Normal, Brownian Motion ; Bachalier 1900

Poisson ; Poisson 1838

Asymptotic Landscape: 9 Operational Regimes, and then some

Erlang-A, w/ I. Gurvich & J. Huang

Erlang-A μ & θ fixed	Conventional scaling			Many-Server scaling			NDS scaling		
	Sub	Critical	Over	QD	QED	ED	Sub	Critical	Over
Offered load per server	$\frac{1}{1+\delta}$	$1-\frac{\beta}{\sqrt{n}}$	$\frac{1}{1-\gamma}$	$\frac{1}{1+\delta}$	$1-\frac{\beta}{\sqrt{n}}$	$\frac{1}{1-\gamma}$	$\frac{1}{1+\delta}$	$1-\frac{\beta}{n}$	$\frac{1}{1-\gamma}$
Arrival rate λ	$\frac{\mu}{1+\delta}$	$\mu-\frac{\beta}{\sqrt{n}}\mu$	$\frac{\mu}{1-\gamma}$	$\frac{n\mu}{1+\delta}$	$n\mu-\beta\mu\sqrt{n}$	$\frac{n\mu}{1-\gamma}$	$\frac{n\mu}{1+\delta}$	$n\mu-\beta\mu$	$\frac{n\mu}{1-\gamma}$
# servers	1			n			n		
Time-scale	n			1			n		
Impatience rate	θ/n			θ			θ/n		
Staffing level	$\frac{\lambda}{\mu}(1+\delta)$	$\frac{\lambda}{\mu}(1+\frac{\beta}{\sqrt{n}})$	$\frac{\lambda}{\mu}(1-\gamma)$	$\frac{\lambda}{\mu}(1+\delta)$	$\frac{\lambda}{\mu}+\beta\sqrt{\frac{\lambda}{\mu}}$	$\frac{\lambda}{\mu}(1-\gamma)$	$\frac{\lambda}{\mu}(1+\delta)$	$\frac{\lambda}{\mu}+\beta$	$\frac{\lambda}{\mu}(1-\gamma)$
Utilization	$\frac{1}{1+\delta}$	$1-\sqrt{\frac{\theta}{\mu}}\frac{h(\hat\beta)}{\sqrt{n}}$	1	$\frac{1}{1+\delta}$	$1-\sqrt{\frac{\theta}{\mu}}\frac{\hat h(\hat\beta)}{\sqrt{n}}$	1	$\frac{1}{1+\delta}$	$1-\sqrt{\frac{\theta}{\mu}}\frac{h(\hat\beta)}{n}$	1
$\mathbb{E}(Q)$	$\frac{1}{\delta(1+\delta)}$	$\sqrt{n}g(\hat\beta)$	$\frac{n\mu\gamma}{\theta(1-\gamma)}$	$\frac{1}{\delta}\varrho_n$	$\sqrt{n}g(\hat\beta)\alpha$	$\frac{n\mu\gamma}{\theta(1-\gamma)}$	$o(1)$	$ng(\hat\beta)$	$\frac{n^2\mu\gamma}{\theta(1-\gamma)}$
$\mathbb{P}(Ab)$	$\frac{1}{n}\frac{1}{\delta}\frac{\theta}{\mu}$	$\frac{\theta}{\sqrt{n\mu}}g(\hat\beta)$	γ	$\frac{1}{n}\frac{(1+\delta)}{\delta}\frac{\theta}{\mu}\varrho_n$	$\frac{\theta}{\sqrt{n\mu}}g(\hat\beta)\alpha$	γ	$o(\frac{1}{n^2})$	$\frac{\theta}{n\mu}g(\hat\beta)$	γ
$\mathbb{P}(W_q>0)$	$\frac{1}{1+\delta}$	≈ 1		ϱ_n	$\alpha\in(0,1)$	≈ 1	≈ 0	≈ 1	
$\mathbb{P}(W_q>T)$	$\frac{1}{1+\delta}e^{-\frac{\delta}{1+\delta}\mu T}$	$1+O(\frac{1}{\sqrt{n}})$	$1+O(\frac{1}{n})$	≈ 0		$f(T)$	≈ 0	$\frac{\bar\Phi(\hat\beta+\sqrt{\theta\mu T})}{\bar\Phi(\hat\beta)}$	$1+O(\frac{1}{n})$
Congestion $\frac{\mathbb{E}W_q}{\mathbb{E}S}$	$\frac{1}{\delta}$	$\sqrt{n}g(\hat\beta)$	$n\mu\gamma/\theta$	$\frac{1}{n}\frac{(1+\delta)}{\delta}\varrho_n$	$\frac{\alpha}{\sqrt{n}}g(\hat\beta)$	$\frac{\mu\gamma}{\theta}$	$o(\frac{1}{n})$	$g(\hat\beta)$	$n\mu\gamma/\theta$

- Conventional: Ward & Glynn (03, $G/G/1+G$)
- Many-Server:
  - QED: Halfin-Whitt (81), Garnett-M-Reiman (02)
  - ED: Whitt (04)
  - NDS: Atar (12)

Asymptotic Landscape: 9 Operational Regimes, and then some

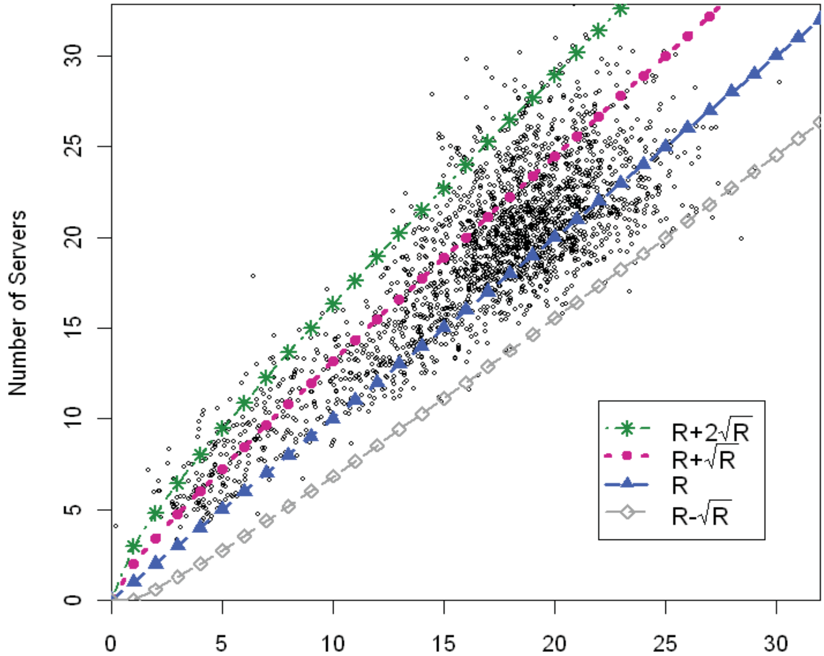
Erlang-A, w/ I. Gurvich & J. Huang

Erlang-A	Conventional scaling			Many-Server scaling			NDS scaling		
μ & θ fixed	Sub	Critical	Over	QD	QED	ED	Sub	Critical	Over
Offered load per server	$\frac{1}{1+\delta}$	$1-\frac{\beta}{\sqrt{n}}$	$\frac{1}{1-\gamma}$	$\frac{1}{1+\delta}$	$1-\frac{\beta}{\sqrt{n}}$	$\frac{1}{1-\gamma}$	$\frac{1}{1+\delta}$	$1-\frac{\beta}{n}$	$\frac{1}{1-\gamma}$
Arrival rate λ	$\frac{\mu}{1+\delta}$	$\mu-\frac{\beta}{\sqrt{n}}\mu$	$\frac{\mu}{1-\gamma}$	$\frac{n\mu}{1+\delta}$	$n\mu-\beta\mu\sqrt{n}$	$\frac{n\mu}{1-\gamma}$	$\frac{n\mu}{1+\delta}$	$n\mu-\beta\mu$	$\frac{n\mu}{1-\gamma}$
# servers	1			n			n		
Time-scale	n			1			n		
Impatience rate	θ/n			θ			θ/n		
Staffing level	$\frac{\lambda}{\mu}(1+\delta)$	$\frac{\lambda}{\mu}(1+\frac{\beta}{\sqrt{n}})$	$\frac{\lambda}{\mu}(1-\gamma)$	$\frac{\lambda}{\mu}(1+\delta)$	$\frac{\lambda}{\mu}+\beta\sqrt{\frac{\lambda}{\mu}}$	$\frac{\lambda}{\mu}(1-\gamma)$	$\frac{\lambda}{\mu}(1+\delta)$	$\frac{\lambda}{\mu}+\beta$	$\frac{\lambda}{\mu}(1-\gamma)$
Utilization	$\frac{1}{1+\delta}$	$1-\sqrt{\frac{\theta}{\mu}}\frac{h(\hat{\beta})}{\sqrt{n}}$	1	$\frac{1}{1+\delta}$	$1-\sqrt{\frac{\theta}{\mu}}\frac{\hat{h}(\hat{\beta})}{\sqrt{n}}$	1	$\frac{1}{1+\delta}$	$1-\sqrt{\frac{\theta}{\mu}}\frac{h(\hat{\beta})}{n}$	1
$\mathbb{E}(Q)$	$\frac{1}{\delta(1+\delta)}$	$\sqrt{n}g(\hat{\beta})$	$\frac{n\mu\gamma}{\theta(1-\gamma)}$	$\frac{1}{\delta}\varrho_n$	$\sqrt{n}g(\hat{\beta})\alpha$	$\frac{n\mu\gamma}{\theta(1-\gamma)}$	$o(1)$	$ng(\hat{\beta})$	$\frac{n^2\mu\gamma}{\theta(1-\gamma)}$
$\mathbb{P}(Ab)$	$\frac{1}{n}\frac{1}{\delta}\frac{\theta}{\mu}$	$\frac{\theta}{\sqrt{n}\mu}g(\hat{\beta})$	γ	$\frac{1}{n}\frac{(1+\delta)}{\delta}\frac{\theta}{\mu}\varrho_n$	$\frac{\theta}{\sqrt{n}\mu}g(\hat{\beta})\alpha$	γ	$o(\frac{1}{n^2})$	$\frac{\theta}{n\mu}g(\hat{\beta})$	γ
$\mathbb{P}(W_q>0)$	$\frac{1}{1+\delta}$	≈ 1		ϱ_n	$\alpha\in(0,1)$	≈ 1	≈ 0	≈ 1	
$\mathbb{P}(W_q>T)$	$\frac{1}{1+\delta}e^{-\frac{\delta}{1+\delta}\mu T}$	$1+O(\frac{1}{\sqrt{n}})$	$1+O(\frac{1}{n})$	≈ 0		$f(T)$	≈ 0	$\frac{\bar{\Phi}(\hat{\beta}+\sqrt{\theta\mu}T)}{\bar{\Phi}(\hat{\beta})}$	$1+O(\frac{1}{n})$
Congestion $\frac{\mathbb{E}W_q}{\mathbb{E}S}$	$\frac{1}{\delta}$	$\sqrt{n}g(\hat{\beta})$	$n\mu\gamma/\theta$	$\frac{1}{n}\frac{(1+\delta)}{\delta}\varrho_n$	$\frac{\alpha}{\sqrt{n}}g(\hat{\beta})$	$\frac{\mu\gamma}{\theta}$	$o(\frac{1}{n})$	$g(\hat{\beta})$	$n\mu\gamma/\theta$

- Conventional: Ward & Glynn (03, $G/G/1+G$)
- Many-Server:
  - QED: Halfin-Whitt (81), Garnett-M-Reiman (02)
  - ED: Whitt (04)
  - NDS: Atar (12)
- "Missing": ED+QED; Hazard-rate scaling (M/M/N+G); Time-Varying, Non-Parametric; Moderate- and Large-Deviation; Networks; Control

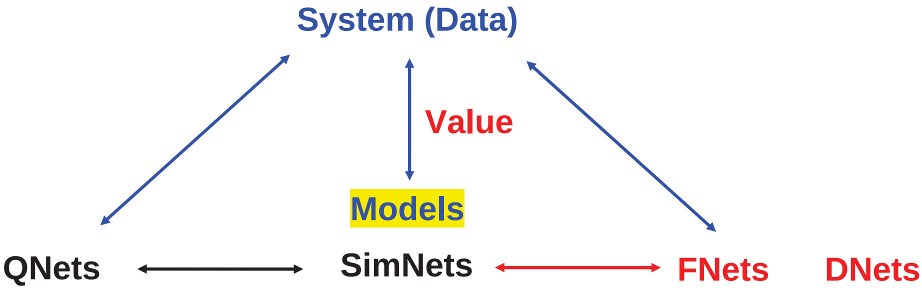
Protocols: Staffing (N) vs. Offered-Load (R $= \lambda \times E(S)$)

IL Telecom; June-September, 2004; w/ Nardi, Plonski, Zeltyn



2205 half-hour intervals (13 summer weeks, week-days)

Data–Based Asymptotic Framework

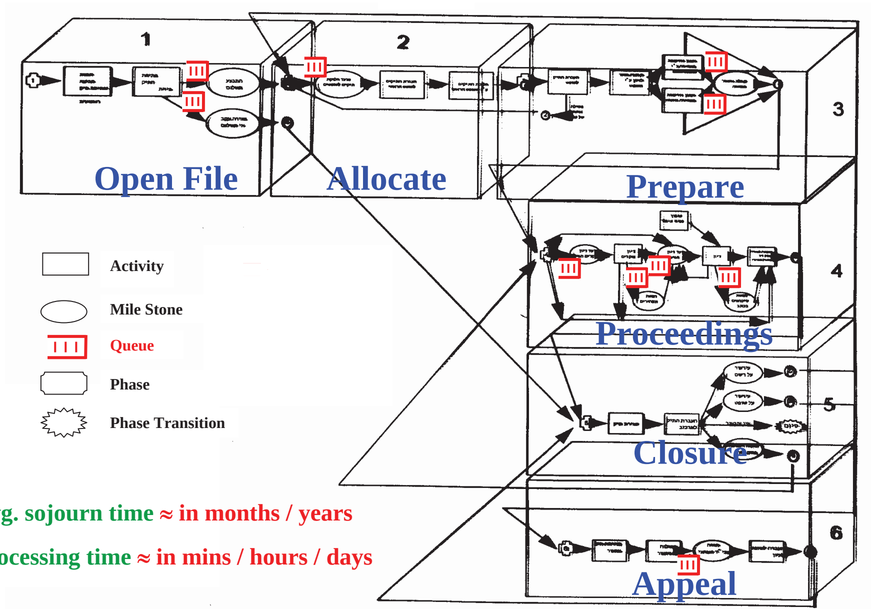


Scope of the Service Industry

Guangzhou Railway Station, Southern China

Operation Focus: The “Production of Justice"

The Labor-Court Process in Haifa, Israel



Avg. sojourn time ≈ in months / years

Processing time ≈ in mins / hours / days

- **Operational Performance surrogate for Financial, Psychological, Clinical,. . .**

Call Centers, Then Hospitals, Now Internet

Call Centers - U.S. Stat.

- \$200 – \$300 billion annual expenditures
- 100,000 – 200,000 call centers
- **"Window"** into the company, for better or worse
- Over 3 million agents = **2% – 4% workforce**

Call Centers, Then Hospitals, Now Internet

Call Centers - U.S. Stat.

- $200 – $300 billion annual expenditures
- 100,000 – 200,000 call centers
- **"Window"** into the company, for better or worse
- Over 3 million agents = **2% – 4% workforce**

Healthcare - similar and unique challenges:

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

Internet - . . .

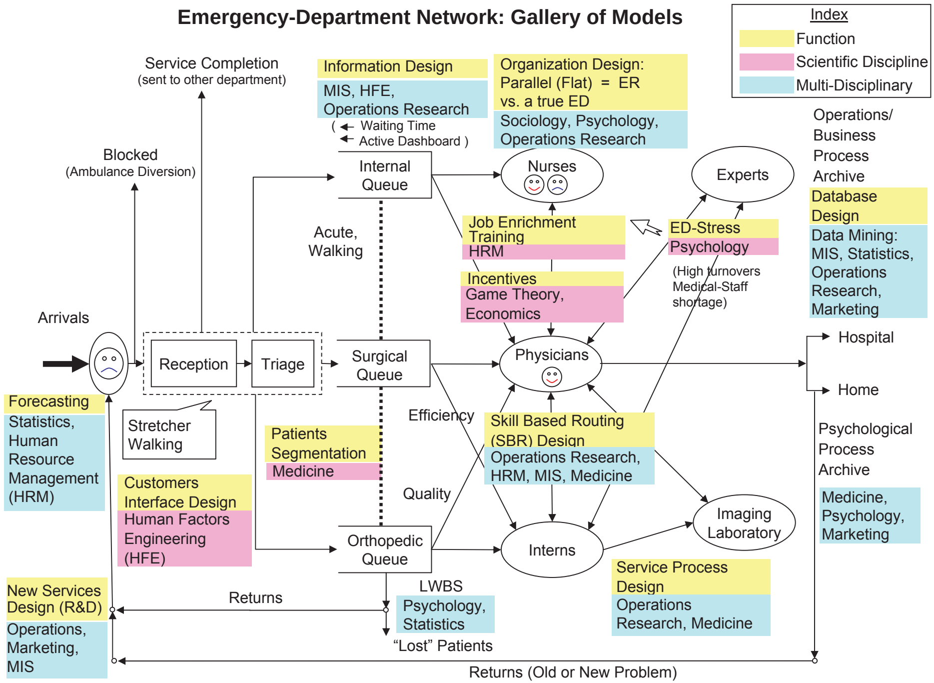
ER / ED Environment: Service Network

Acute (Internal, Trauma)

Walking

Multi-Trauma

Emergency-Department Network: Gallery of Models



- **Forecasting**, Abandonment = **LWBS**, SBR ≈ **Flow Control**

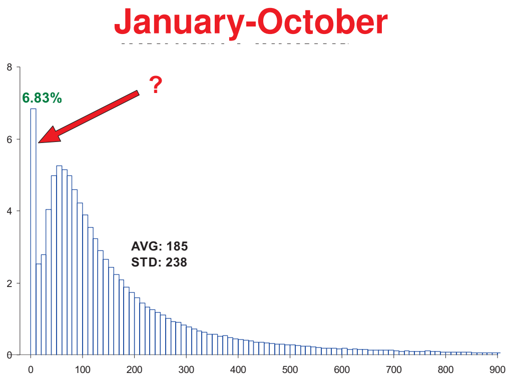
Prerequisite I: Data

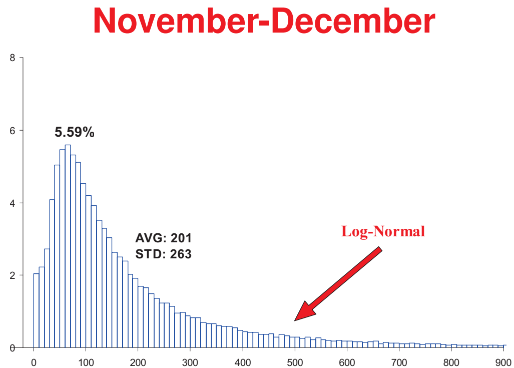
Averages Prevalent (and could be useful / interesting).

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 (events-log files).

Beyond Averages: The Human Factor

Histogram of Service-Time in an Israeli Call Center, 1999

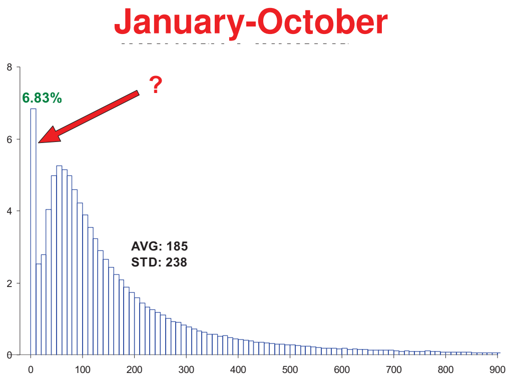


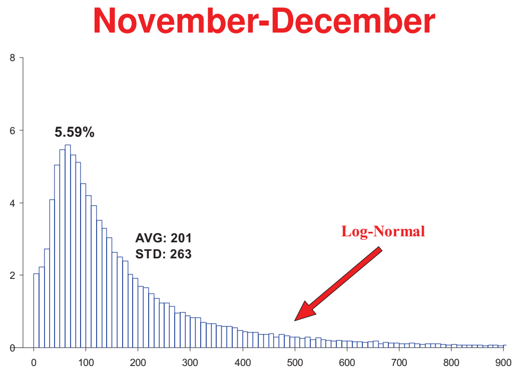


- 6.8% Short-Services:

Beyond Averages: The Human Factor

Histogram of Service-Time in an Israeli Call Center, 1999



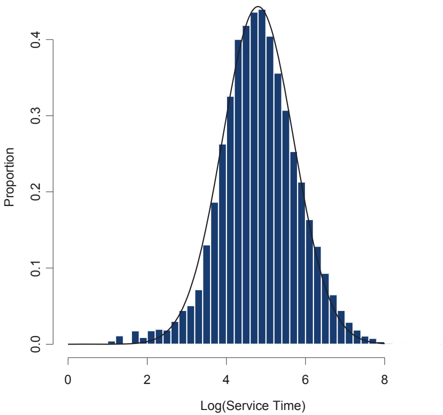


- **6.8% Short-Services:** Agents' "Abandon" (improve bonus, rest), (mis)lead by **incentives**
- **Distributions** must be measured (in **seconds** = **natural scale**)
- **LogNormal** service times common in call centers

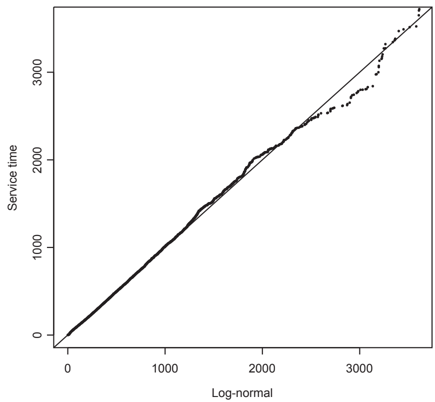
Durations: Phone Calls (2 Surprises)

Israeli Call Center, Nov–Dec, 1999

Log(Service Times)



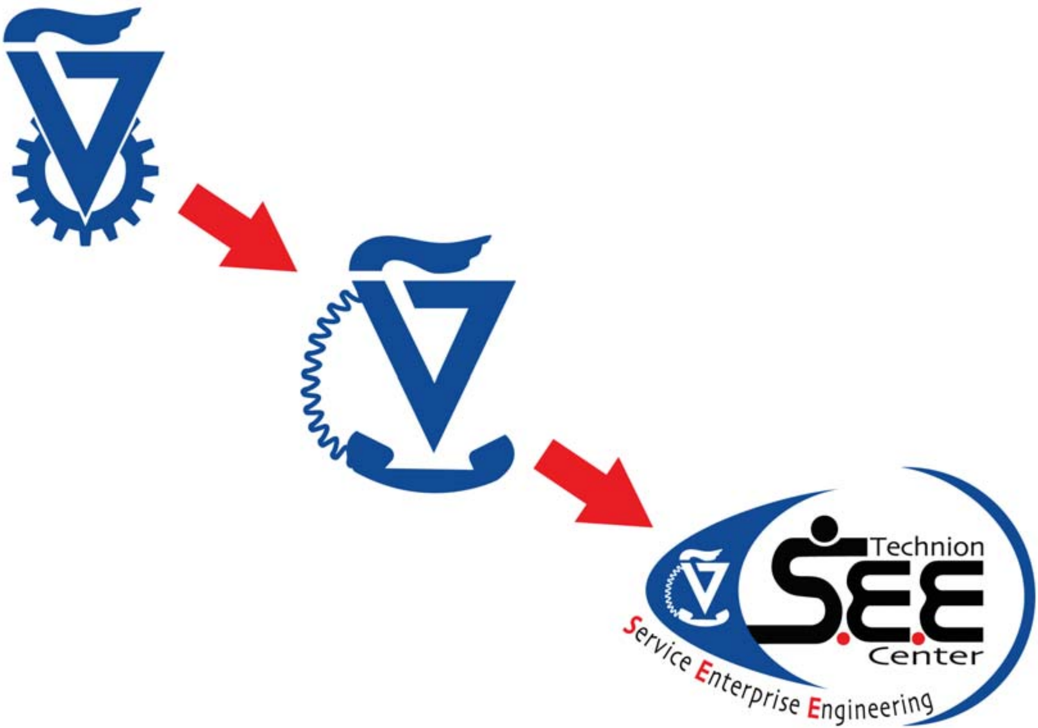
LogNormal QQPlot



- **Practically Important**: (mean, std)(log) characterization
- **Theoretically Intriguing**: Why LogNormal ? Naturally multiplicative but, in fact, also **Infinitely-Divisible** (Generalized Gamma-Convolutions)

Pause for a Commercial:

Pause for a Commercial: The Technion SEE Center



Technion SEE = Service Enterprise Engineering

SEELab: Data-repositories for research and teaching

- For example:
  - Bank Anonymous: **1 years, 350K calls by 15 agents** - in 2000. **Brown, Gans, Sakov, Shen, Zeltyn, Zhao** (JASA), paved the way for:
  - U.S. Bank: **2.5 years, 220M calls, 40M by 1000 agents.**
  - Israeli Cellular: **2.5 years, 110M calls, 25M calls by 750 agents.**
  - Israeli Bank: **from January 2010, daily-deposit at a SEESafe.**
  - Israeli Hospital: **4 years, 1000 beds; 8 ED's- Sinreich's data.**

Technion SEE = Service Enterprise Engineering

SEELab: **Data-repositories for research and teaching**

- For example:
  - Bank Anonymous: **1 years, 350K calls by 15 agents** - in 2000. **Brown, Gans, Sakov, Shen, Zeltyn, Zhao** (JASA), paved the way for:
  - U.S. Bank: **2.5 years, 220M calls, 40M by 1000 agents**.
  - Israeli Cellular: **2.5 years, 110M calls, 25M calls by 750 agents**.
  - Israeli Bank: **from January 2010, daily-deposit at a SEESafe**.
  - Israeli Hospital: **4 years, 1000 beds; 8 ED's- Sinreich's data**.

SEEStat: Environment for graphical **EDA** in real-time

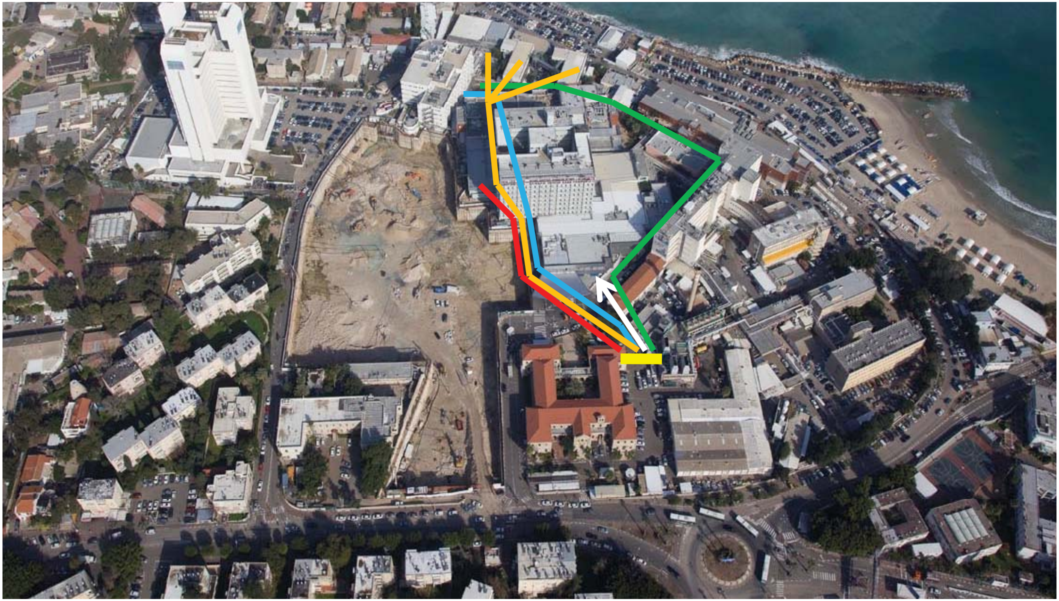
- **Universal Design, Internet Access, Real-Time Response**.

SEEServer: **Free for academic use**

Register; access U.S. Bank, Israeli Bank, Rambam Hospital.

eg. RFID-Based Data: Mass Casualty Event (MCE)

Drill: Chemical MCE, Rambam Hospital, May 2010



Focus on **severely wounded** casualties ($\approx$ 40 in drill)
Note: 20 observers support real-time control (helps validation)

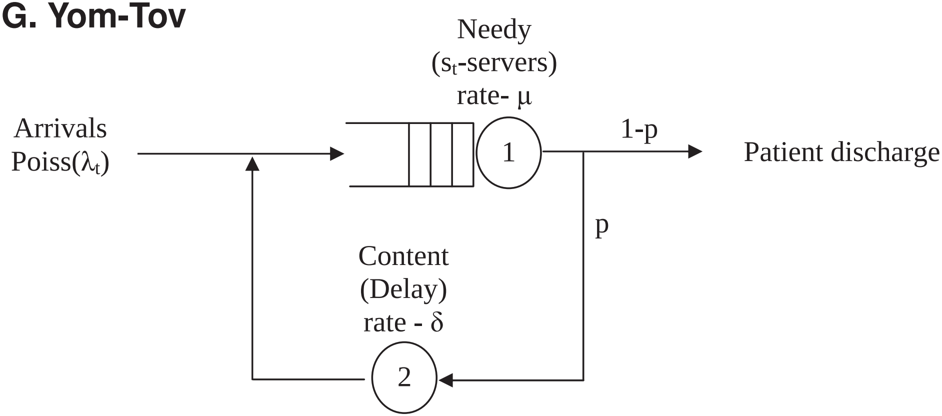
Data Cleaning: MCE with RFID Support

Data-base				Company report		comment
Asset id	order	**Entry** date	**Exit** date	Entry date	Exit date	
4	1	1:14:07 PM		1:14:00 PM		
6	1	12:02:02 PM	12:33:10 PM	12:02:00 PM	12:33:00 PM	
8	1	11:37:15 AM	12:40:17 PM	11:37:00 AM		**exit is missing**
10	1	12:23:32 PM	12:38:23 PM	12:23:00 PM		
12	1	12:12:47 PM	12:35:33 PM		12:35:00 PM	**entry is missing**
15	1	1:07:15 PM		1:07:00 PM		
16	1	11:18:19 AM	11:31:04 AM	11:18:00 AM	11:31:00 AM	
17	1	1:03:31 PM		1:03:00 PM		
18	1	1:07:54 PM		1:07:00 PM		
19	1	12:01:58 PM		12:01:00 PM		
20	1	11:37:21 AM	12:57:02 PM	11:37:00 AM	12:57:00 PM	
21	1	12:01:16 PM	12:37:16 PM	12:01:00 PM		
22	1	12:04:31 PM	12:20:40 PM			first customer is missing
22	2	12:27:37 PM		12:27:00 PM		
25	1	12:27:35 PM	1:07:28 PM	12:27:00 PM	1:07:00 PM	
27	1	12:06:53 PM		12:06:00 PM		
28	1	11:21:34 AM	11:41:06 AM	11:41:00 AM	11:53:00 AM	**exit time instead of entry time**
29	1	12:21:06 PM	12:54:29 PM	12:21:00 PM	12:54:00 PM	
31	1	11:40:54 AM	12:30:16 PM	11:40:00 AM	12:30:00 PM	
31	2	12:37:57 PM	12:54:51 PM	12:37:00 PM	12:54:00 PM	
32	1	11:27:11 AM	12:15:17 PM	11:27:00 AM	12:15:00 PM	
33	1	12:05:50 PM	12:13:12 PM	12:05:00 PM	12:15:00 PM	wrong exit time
35	1	11:31:48 AM	11:40:50 AM	11:31:00 AM	11:40:00 AM	
36	1	12:06:23 PM	12:29:30 PM	12:06:00 PM	12:29:00 PM	
37	1	11:31:50 AM	11:48:18 AM	11:31:00 AM	11:48:00 AM	
37	2	12:59:21 PM		12:59:00 PM		

- Imagine **"Cleaning" 60,000+ customers per day** (call centers) !
- "Psychology" of Data Trust and Transfer (e.g. 2 years till transfer)

The Basic Service-Network Model: Erlang-R

w/ G. Yom-Tov



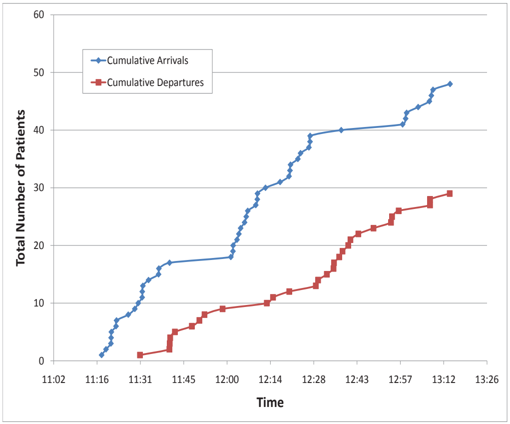
Erlang-R **(IE: Repairman Problem 50's; CS: Central-Server 60's)** =
2-station "Jackson" Network = (M/M/S, M/M/∞) :

- λ_t – **Time-Varying** **Arrival** rate
- S_t – Number of **Servers** (Nurses / Physicians).
- μ – **Service** rate (E[Service] $= \frac{1}{\mu}$)
- p – **ReEntrant** (Feedback) fraction
- δ – **Content-to-Needy** rate (E[Content] $= \frac{1}{\delta}$)

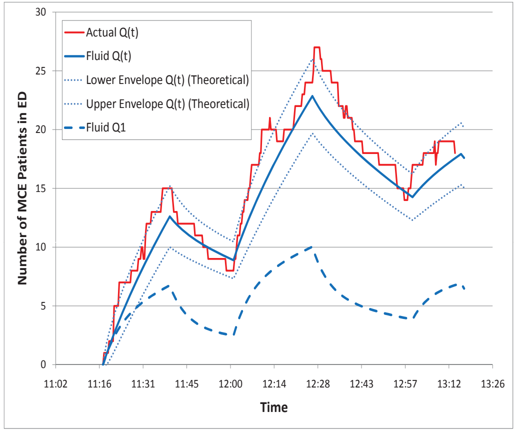
Erlang-R: Fitting a Simple Model to a Complex Reality

Chemical MCE Drill (Israel, May 2010)

Arrivals & Departures (RFID)



Erlang-R (Fluid, Diffusion)

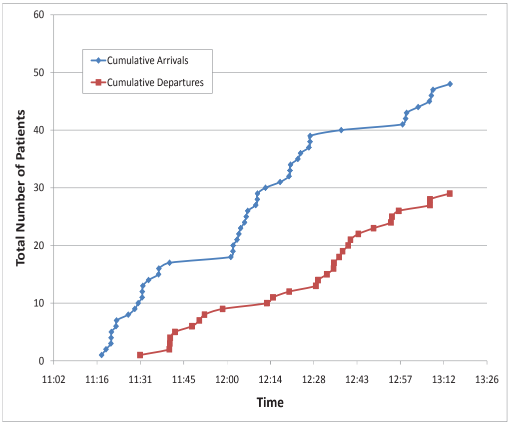


- **Recurrent/Repeated** services in MCE Events: eg. Injection every 15 minutes

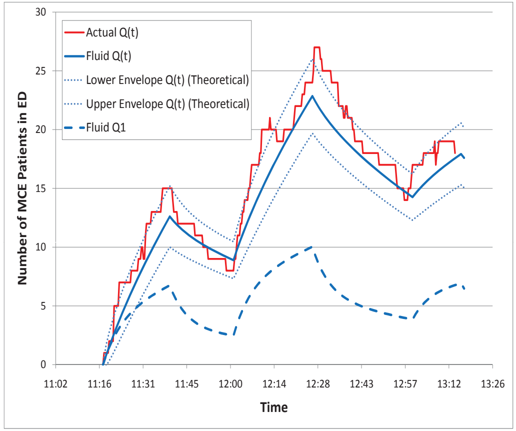
Erlang-R: Fitting a Simple Model to a Complex Reality

Chemical MCE Drill (Israel, May 2010)

Arrivals & Departures (RFID)



Erlang-R (Fluid, Diffusion)



- **Recurrent/Repeated** services in MCE Events: eg. Injection every 15 minutes
- **Fluid (Sample-path)** Modeling, via Functional Strong Laws of Large Numbers
- **Stochastic** Modeling, via Functional Central Limit Theorems
  - ED in **MCE**: Confidence-interval, usefully narrow for **Control**
  - ED in **normal** (**time-varying**) conditions: Personnel **Staffing**

The Asymptotic Framework: Erlang-R in the ED

System = Emergency Department (eg. Rambam Hospital)

- **SimNet** = Customized ED-Simulator (Marmor & Sinreich)
- **QNet** = Erlang-R (time-varying 2-station Jackson; **w/ Yom-Tov**)
- **FNet** = 2-dim dynamical system (Massey & Whitt)
- **DNet** = 2-dim Markovian Service Net (w/ Massey and Reiman)

The Asymptotic Framework: Erlang-R in the ED

System = Emergency Department (eg. Rambam Hospital)

- **SimNet** = Customized ED-Simulator (Marmor & Sinreich)
- **QNet** = Erlang-R (time-varying 2-station Jackson; **w/ Yom-Tov**)
- **FNet** = 2-dim dynamical system (Massey & Whitt)
- **DNet** = 2-dim Markovian Service Net (w/ Massey and Reiman)

Asymptotic Framework

- Start with Data
- Fit a simple QNet (time-varying Erlang-R)
  to a complex reality (ED Physicians)
- Develop FNet (Offered-Load of Physicians) and DNet
- Use QNet + FNet + DNet for Design ($\surd$-Staffing), Analysis (of Feedback), ...
- Virtual reality: SimNet of ED with $\surd$-staffing of Physicians)
- Validation: stable performance, confidence intervals, ...

ServNets: Data-Based Online Automatic Creation

w/ **V. Trofimov, E. Nadjharov, I. Gavako = Technion SEELab**

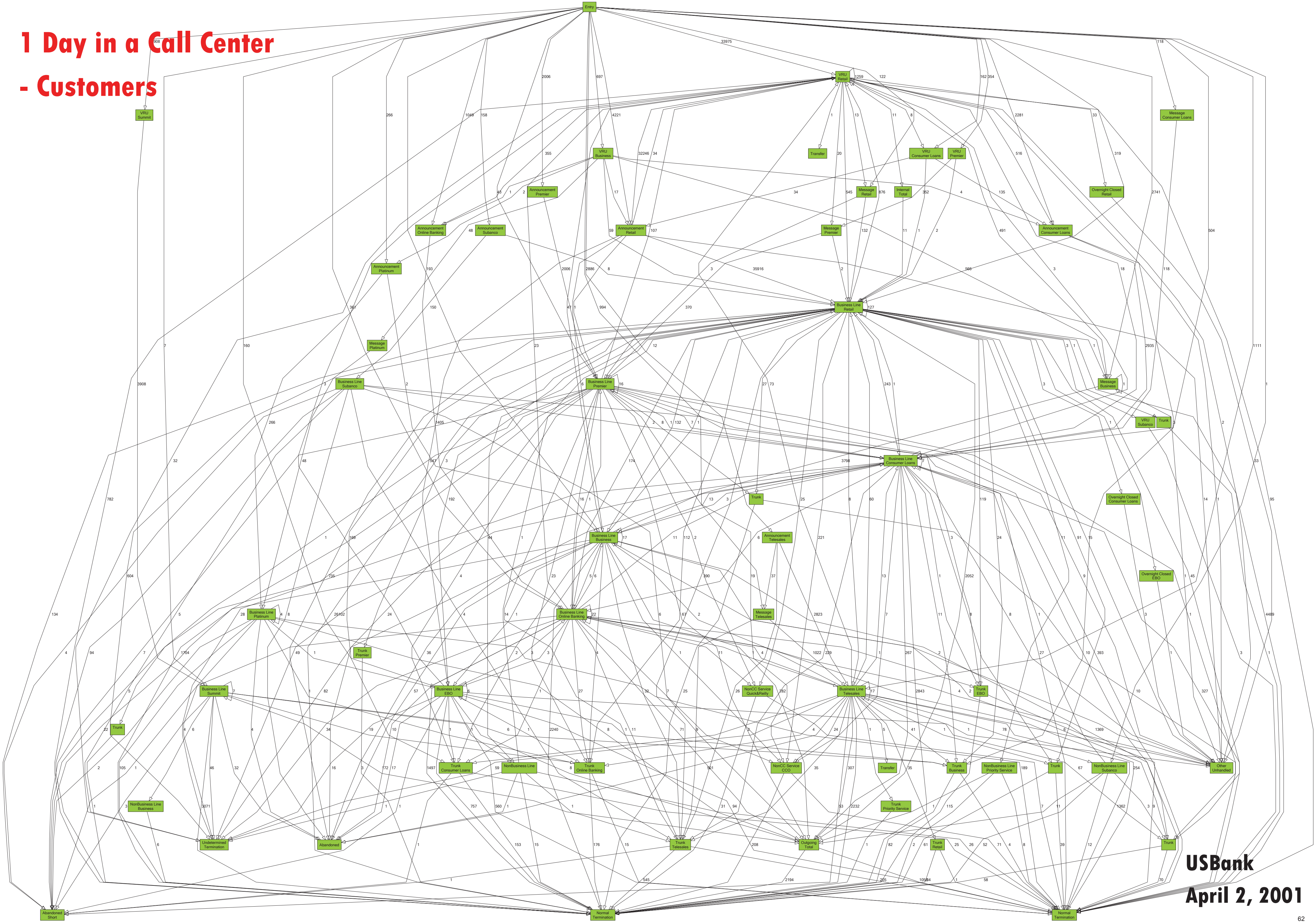
- Towards creating **ServNets = QNets, SimNets, FNets, DNets**
- SimNets of Service Systems = **Virtual Realities**, for Exploratory Data Analysis (EDA), Performance Analysis, Optimization, . . .
- SimNets **also** of QNEts, FNets, DNets, for “Valuation" (Validation, Accuracy)

ServNets: Data-Based Online Automatic Creation

w/ **V. Trofimov, E. Nadjharov, I. Gavako = Technion SEELab**

- Towards creating **ServNets = QNets, SimNets, FNets, DNets**
- SimNets of Service Systems = **Virtual Realities**, for Exploratory Data Analysis (EDA), Performance Analysis, Optimization, . . .
- SimNets **also** of QNEts, FNets, DNets, for “Valuation" (Validation, Accuracy)
- Ultimately: Empirical **Research Lab**, necessitated by the complexity of service systems and human-behavior:
  - Daily routine in Physics, Chemistry, Biology; Psychology
  - But also in Transportation (Science), and recently (Behavioral) Economics
  - **Why not in Service Science / Engineering / Management ?**

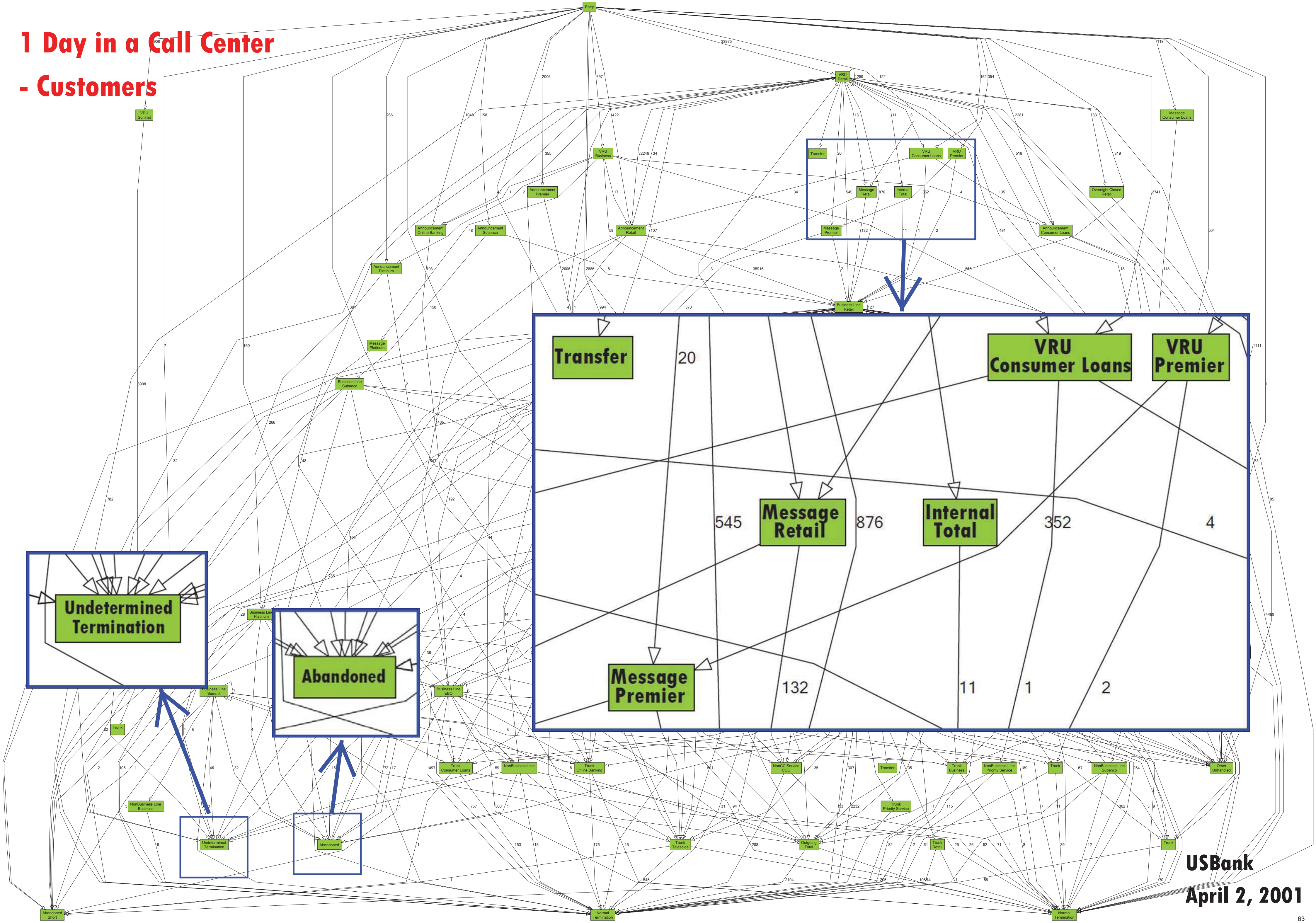
1 Day in a Call Center
- Customers



USBank

April 2, 2001

1 Day in a Call Center
- Customers

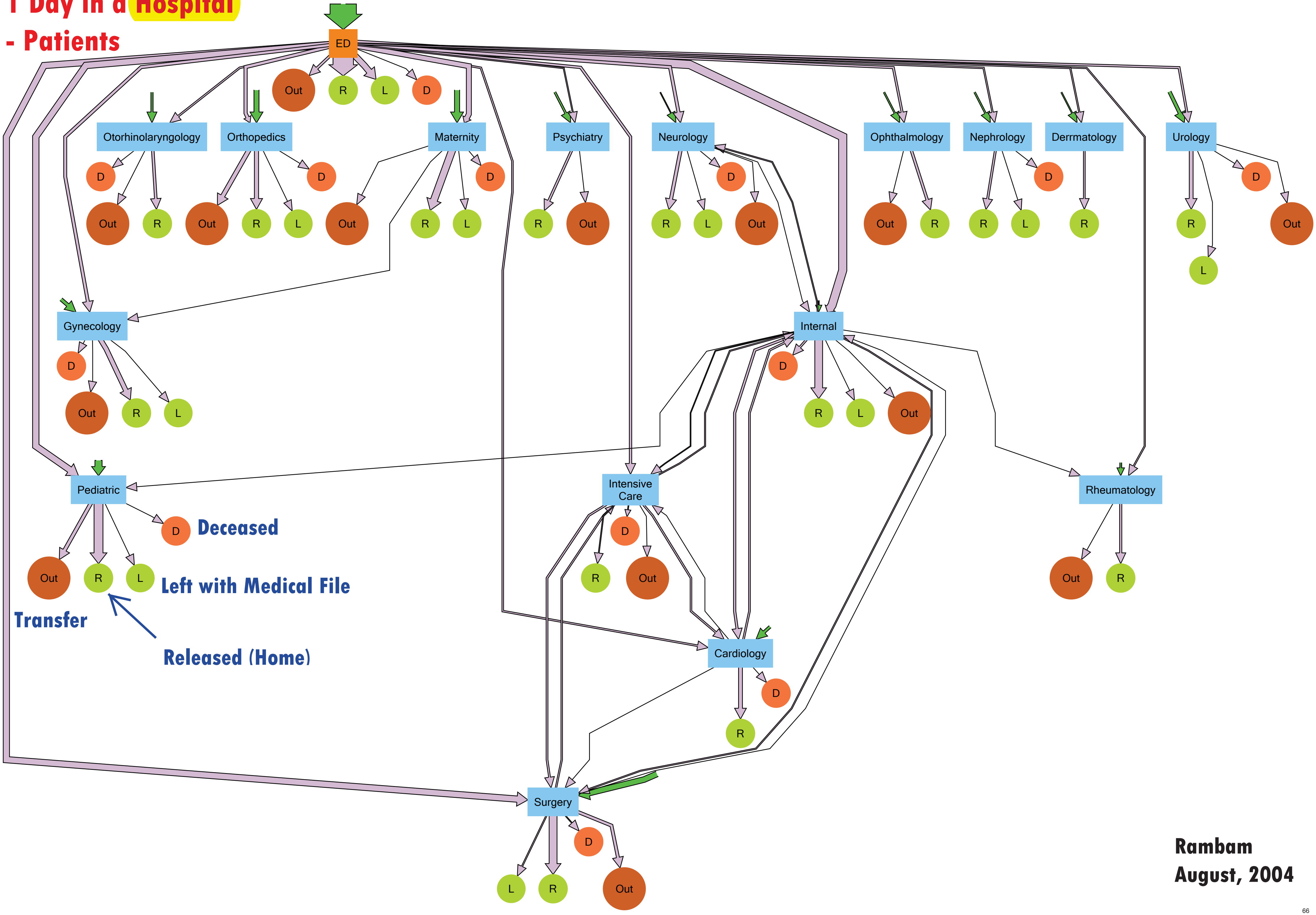


USBank
April 2, 2001

Server Network
- 1 Day
- All Agents
- Activities + Queues
Inbound
Available
Dialer
Call-Backs
Idle
Outgoing
Break
Logged Off
Logged In
Idle (185.11)
Available (7.36)
Sales (329.00)
Virtual Branches (195.46)
Pension (758.00)
General Banking (218.31)
Investments (252.68)
Wrap Up (8.71)
Securities (278.20)
Shifting (209.71)
Paperwork (140.62)
Logged Off
Consulting (562.54)
Short Break (12.97)
Sales (177.14)
Wrap Up (7.46)
Paperwork (140.57)
Consulting (208.00)
Managers VIP (227.05)
Medium Break (507.26
Paperwork (313.19)
Paperwork (140.97)
Consulting (175.42)
Private Call (27.11)
General Banking (136.89)
Wrap Up (8.11)
Internet Support (114.49)
Message (209.00)
Virtual Branches (133.67)
Managers VIP (149.36)
Long Break (1,384.28)
Internet Support (313.66)
General Banking (97.22)
Wrap Up (8.03)
Pension (307.14)
Internet Support (166.50)
Securities (532.00)
Investments (211.41)
ILDUBank
January 24, 2010
from 06:00 to 23:59:59


Server Network
- 1 Day
- All Agents
- Activities + Queues
Inbound
Available
Dialer
Call-Backs
Outgoing
Idle
Break
Sales (329.00)
Pension (758.00)
Securities
Consulting (562.54)
Short Break
Sales (177.14)
Wrap Up (7.46)
Paperwork (140.57)
Available (7.36)
Managers VIP (227.05)
Paperwork (313.19)
Paperwork (140.97)
Consulting (175.42)
Private Call (27.11)
Virtual Branches (133.67)
Managers VIP (149.36)
Internet Support (114.49)
Message (209.00)
General Banking (136.89)
Wrap Up (8.11)
Logged Off
Logged In
Idle (185.11)
Long Break (1,384.28)
Internet Support (313.66)
General Banking (97.22)
Wrap Up (8.03)
Pension (307.14)
Internet Support (166.50)
Securities (532.00)
Investments (211.41)
ILDUBank
January 24, 2010
from 06:00 to 23:59:59

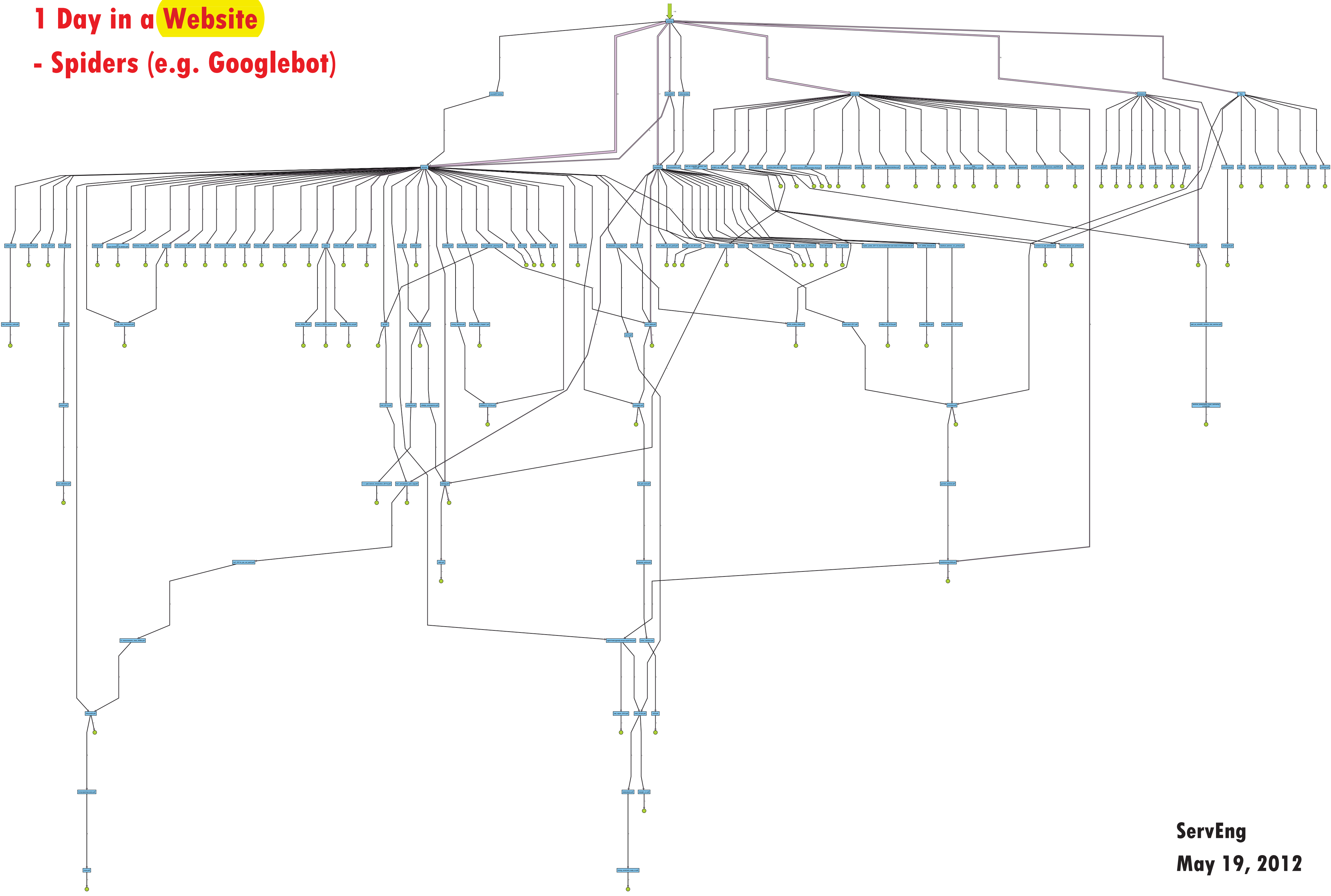
1 Day in a Hospital - Patients



Rambam
August, 2004

1 Day in a Website

- Spiders (e.g. Googlebot)

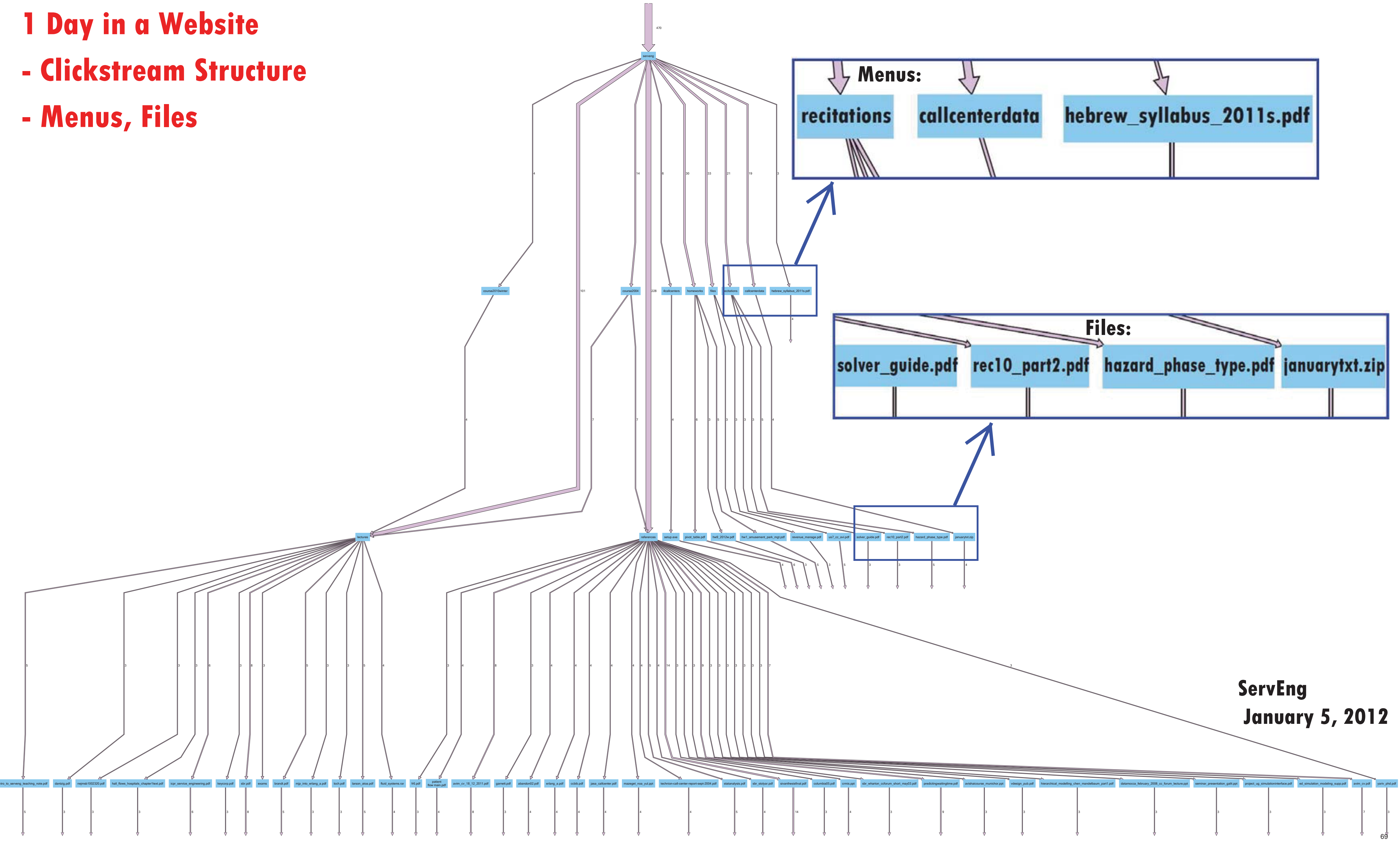


ServEng

May 19, 2012

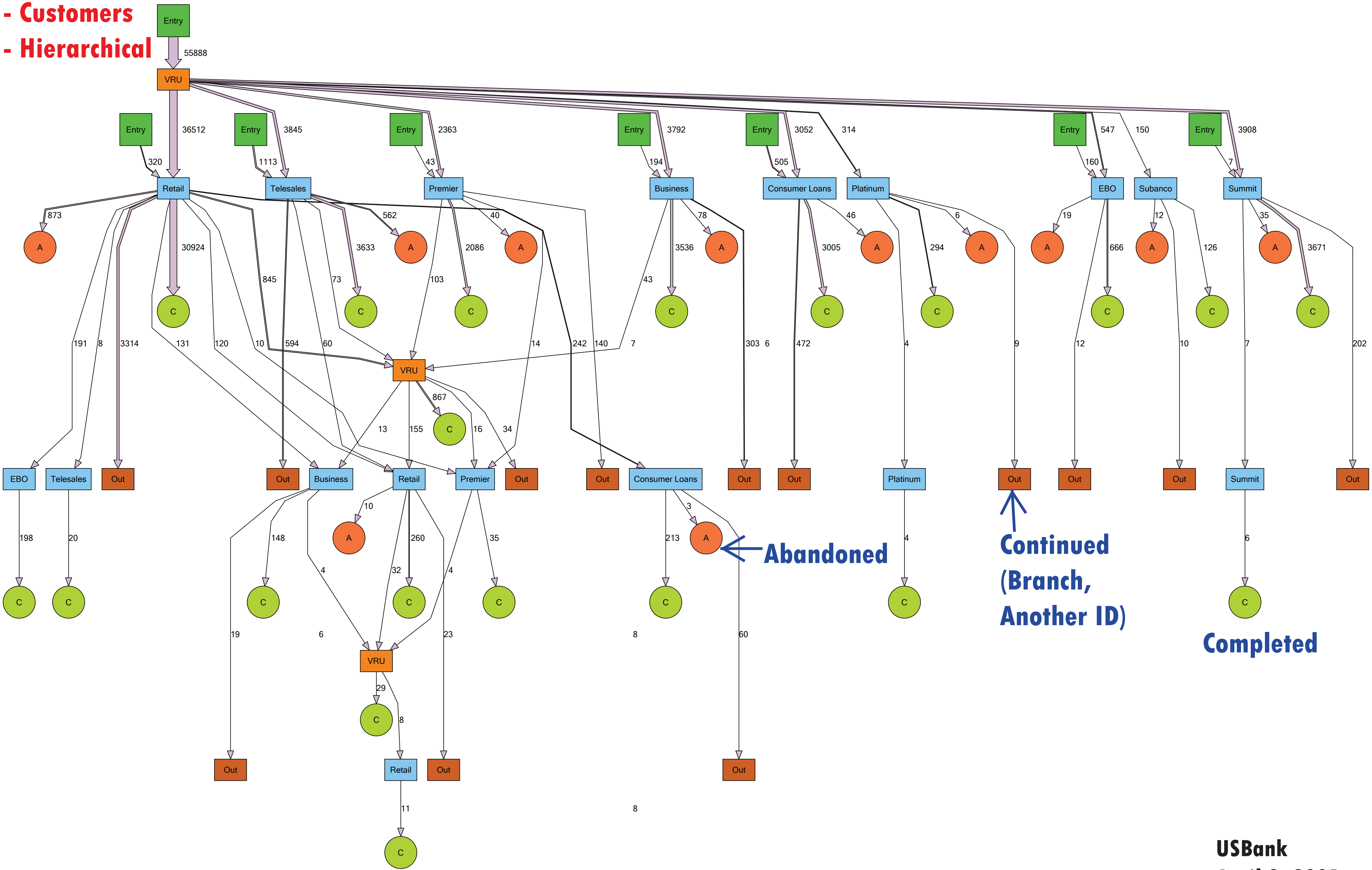
1 Day in a Website

- Clickstream Structure
- Menus, Files



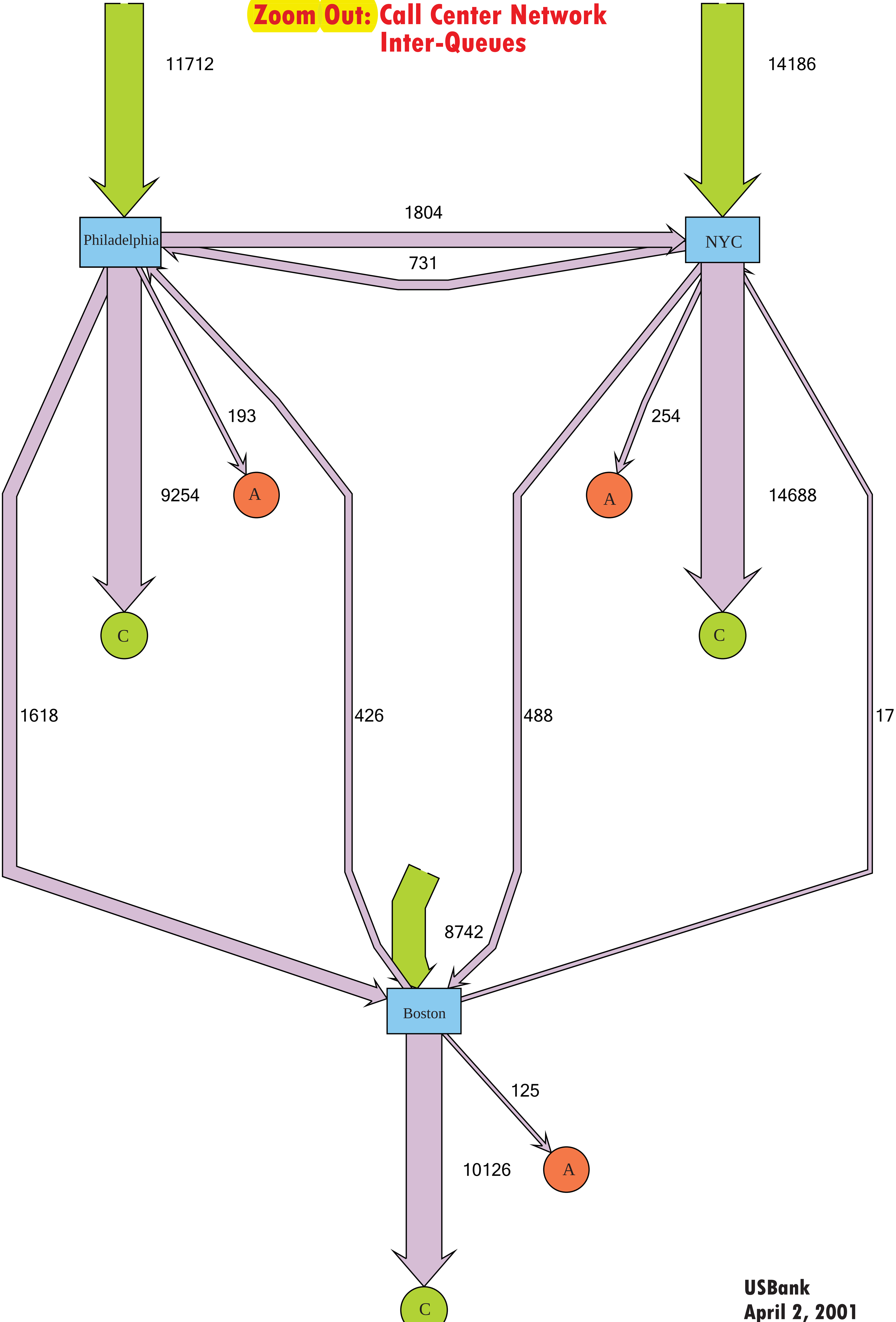
ServEng
January 5, 2012

1 Hour in a Call Center
- Customers
- Hierarchical



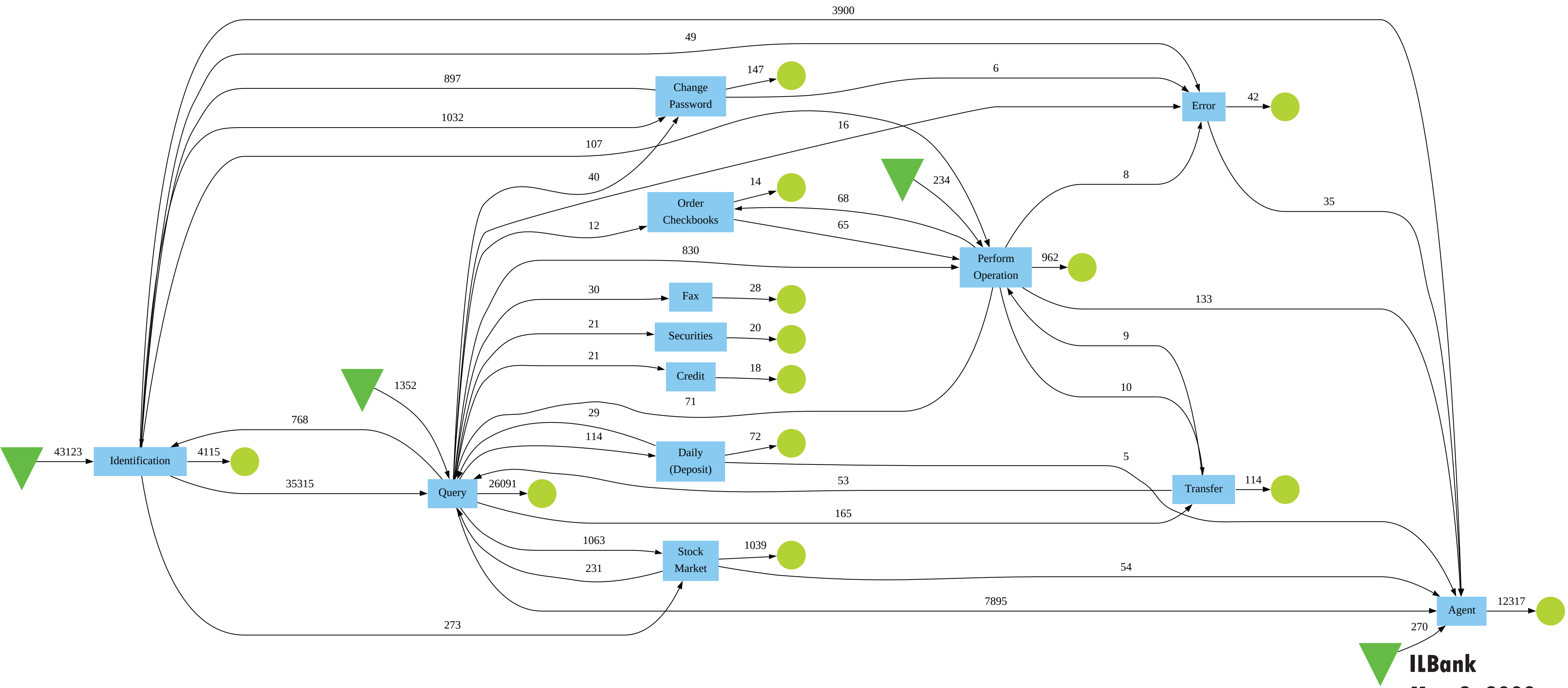
USBank
April 2, 2001

Zoom Out: Call Center Network Inter-Queues



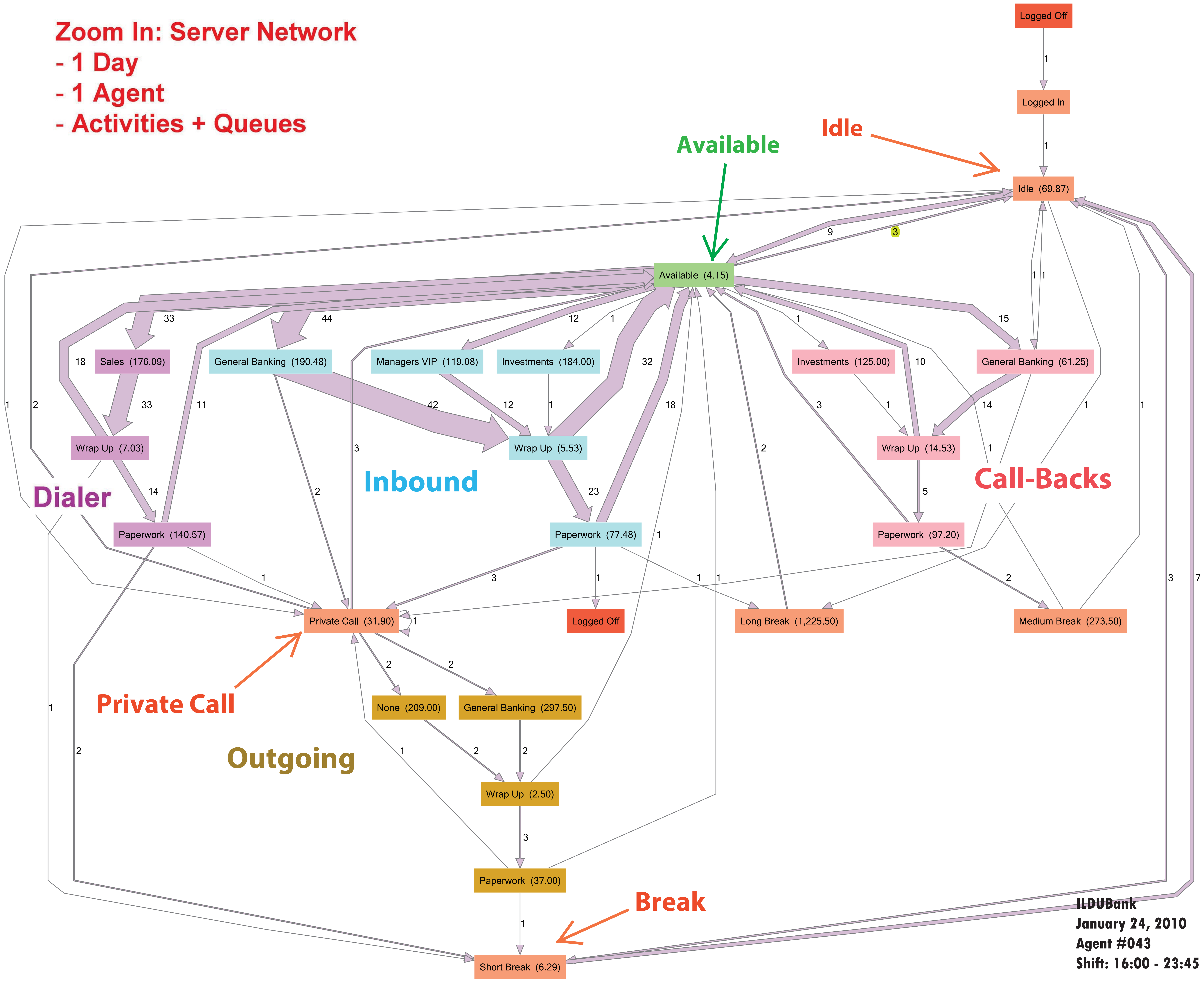
USBank
April 2, 2001

Zoom In: Interactive Voice Response (IVR)

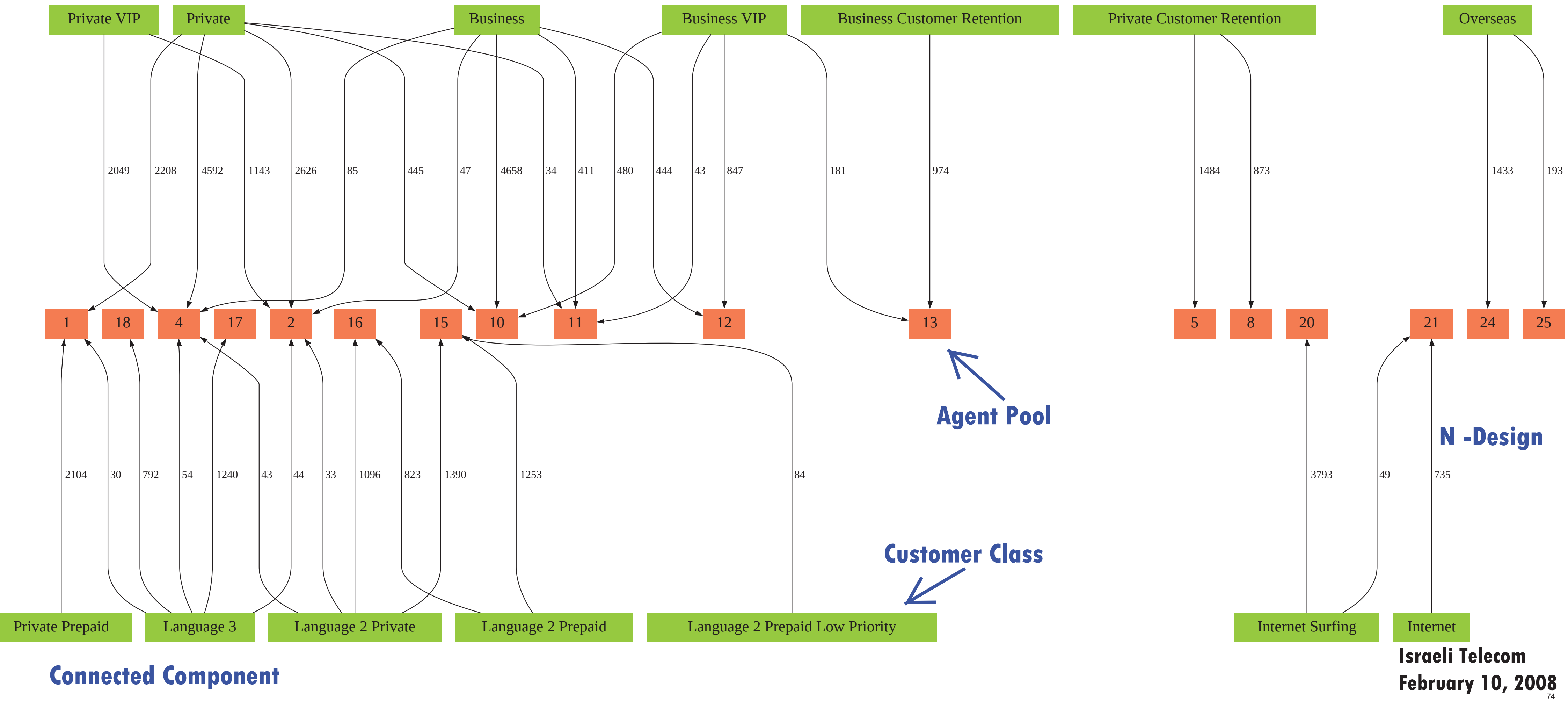


ILBank

May 2, 2008

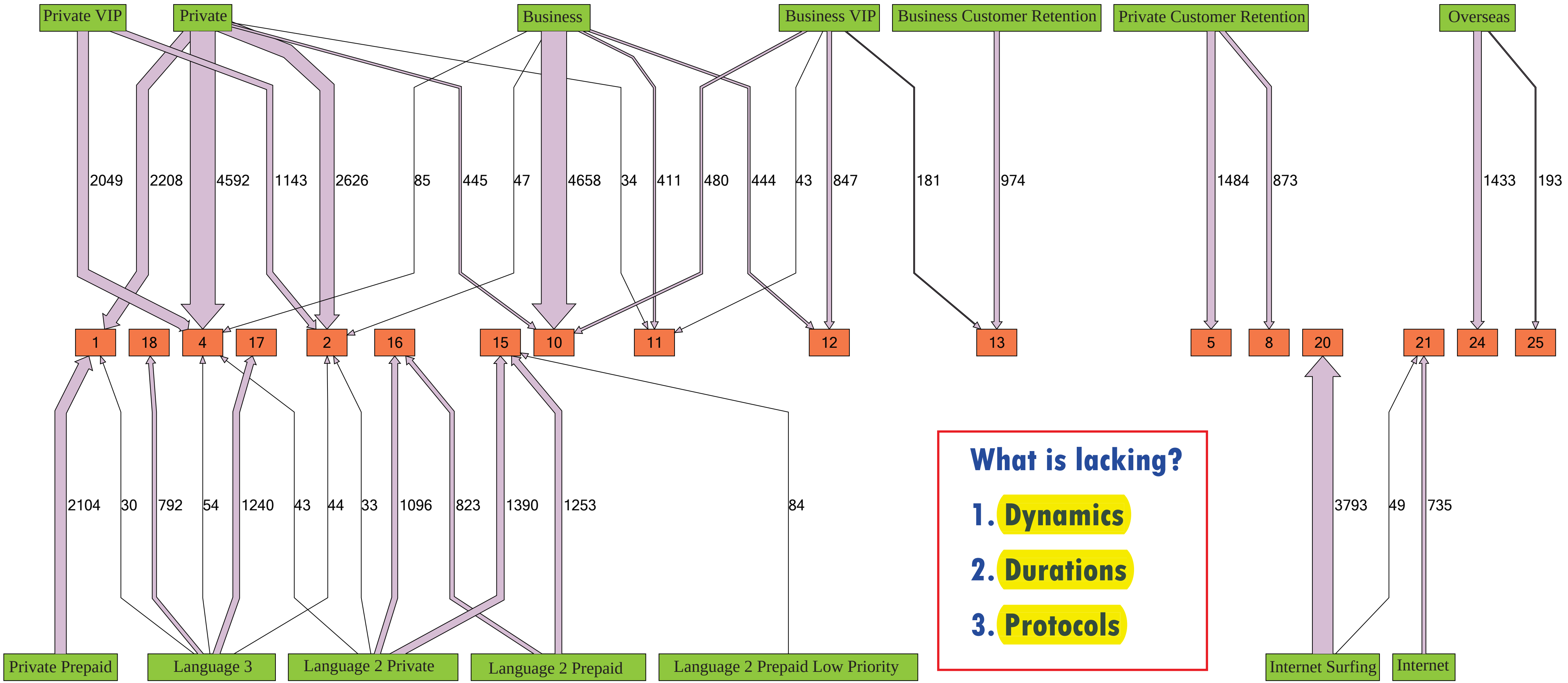

Zoom In: Server Network
- 1 Day
- 1 Agent
- Activities + Queues
Logged Off
Logged In
Idle
Idle (69.87)
Available
Available (4.15)
Sales (176.09)
General Banking (190.48)
Managers VIP (119.08)
Investments (184.00)
Investments (125.00)
General Banking (61.25)
Wrap Up (7.03)
Wrap Up (5.53)
Wrap Up (14.53)
Dialer
Inbound
Call-Backs
Paperwork (140.57)
Paperwork (77.48)
Paperwork (97.20)
Private Call (31.90)
Logged Off
Long Break (1,225.50)
Medium Break (273.50)
Private Call
None (209.00)
General Banking (297.50)
Outgoing
Wrap Up (2.50)
Paperwork (37.00)
Break
Short Break (6.29)
ILDUBank
January 24, 2010
Agent #043
Shift: 16:00 - 23:45

Zoom In: Skills - Based Routing (SBR)
Network Structure (Protocol)



Israeli Telecom
February 10, 2008

Goal: Data-Based Real-Time Simulation (SimNet)



What is lacking?

1. Dynamics
2. Durations
3. Protocols

Israeli Telecom
February 10, 2008

4 Hours in a Call Center - Dynamics

8 AM - 9 AM

USBank
April 2, 2001


Entry
VRU
Subanco
Premier
Telesales
Business
Retail
Summit
Consumer Loans
EBO
Platinum
A
C
Out

8 AM - 9 AM


Entry
VRU
Subanco
Premier
Telesales
Business
Retail
Summit
Consumer Loans
EBO
Platinum
A
C
Out

9 AM - 10 AM

Entry
VRU
Entry
Entry
Entry
Entry
Entry
Entry
Entry
Entry
Subanco
Premier
Telesales
Business
Retail
Summit
Consumer Loans
EBO
Platinum
A
A
A
A
A
A
A
A
A
C
C
C
C
C
C
C
C
C
VRU
C
Out
Out
Out
Premier
Retail
Out
Out
Business
Consumer Loans
EBO
Telesales
Out
Summit
Out
Out
Out
Out
Platinum
A
A
C
C
C
C
C
C
C
C
VRU
C
Out
Retail
Out
Out
Business
C
C

10 AM - 11 AM

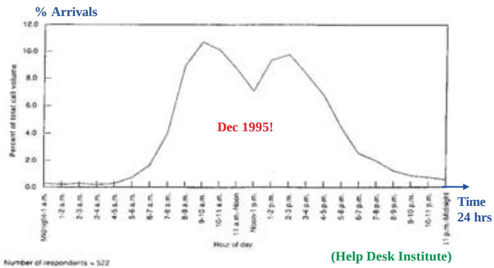

Entry
VRU
Entry
Entry
Entry
Entry
Entry
Entry
Entry
Subanco
Premier
Telesales
Business
Retail
Summit
Consumer Loans
EBO
Platinum
A
C
Out
Premier
Retail
Business
Consumer Loans
EBO
Telesales
Summit
Platinum

11 AM - 12 PM

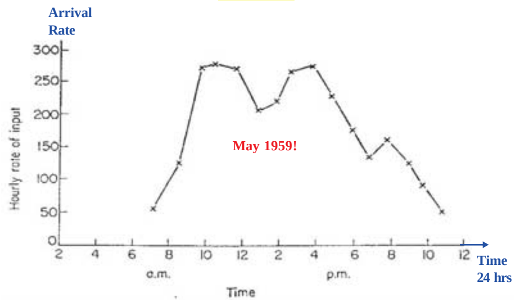
Dynamics: Time-Varying Arrival-Rates

2 Daily Peaks

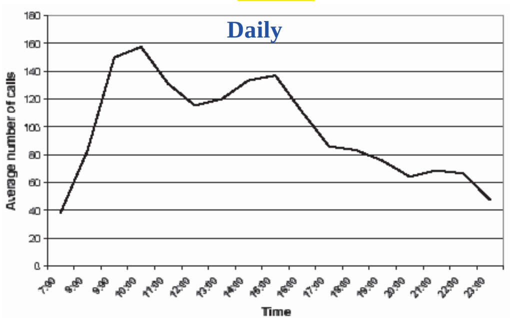
CC: Dec. **1995**, (USA, 700 Helpdesks)



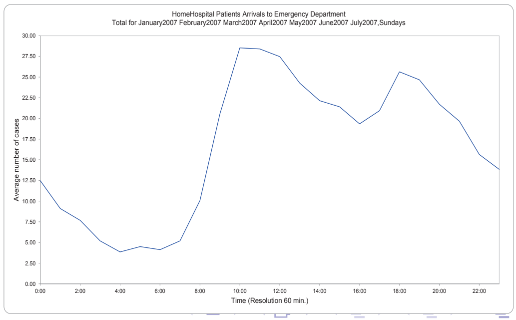
CC: May **1959** (England)



CC: Nov. 1999 (Israel)

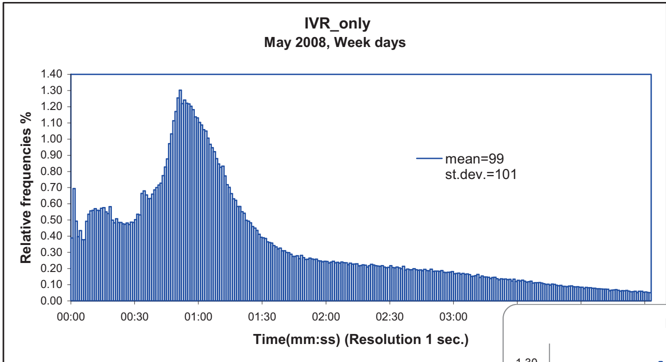


ED: Jan.–July 2007 (Israel)

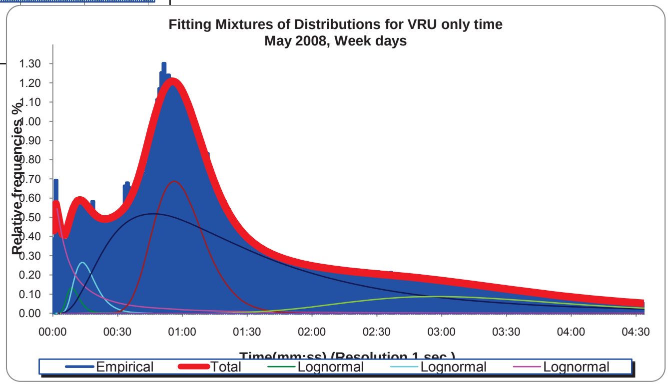


Durations: Answering Machine

Israeli Bank: IVR/VRU Only, May 2008



Mixture: 7 LogNormals

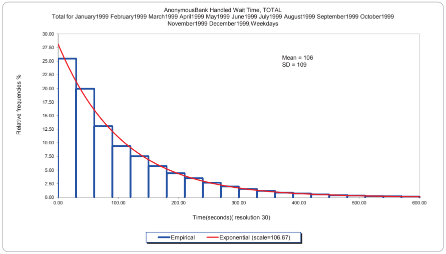


Durations: Waiting Times in a Call Center

$\Rightarrow$ Protocols

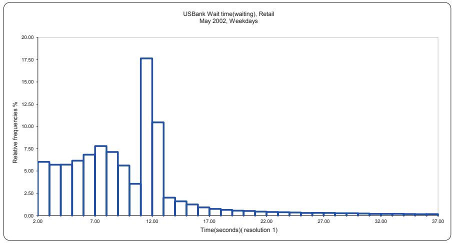
Exponential in Heavy-Traffic (min.)

Small Israeli Bank



Routing via Thresholds (sec.)

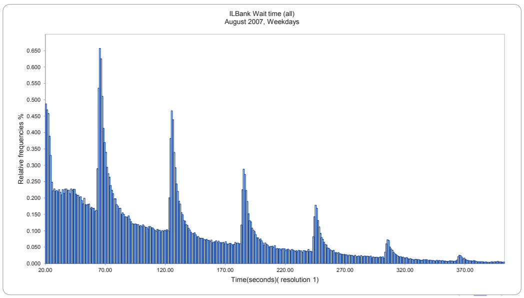
Large U.S. Bank



Scheduling Priorities (sec.)

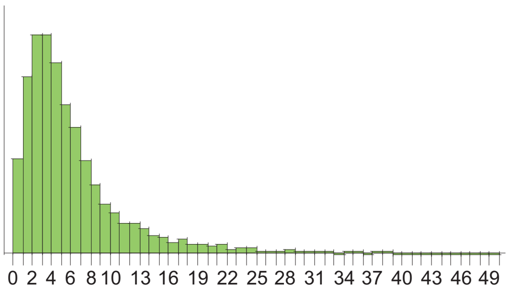
[compare Hospital LOS (hours)]

Medium Israeli Bank



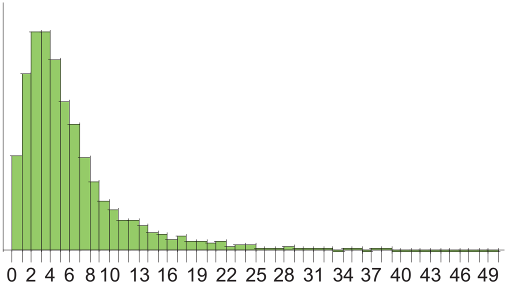
LogNormal & Beyond: Length-of-Stay in a Hospital

Israeli Hospital, in Days: LN

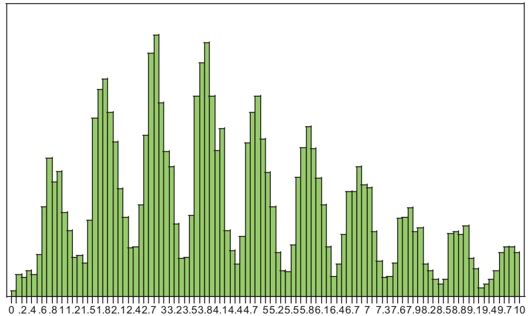


LogNormal & Beyond: Length-of-Stay in a Hospital

Israeli Hospital, in Days: LN



Israeli Hospital, in Hours: Mixture



Explanation: Patients released around **3pm** (1pm in Singapore)

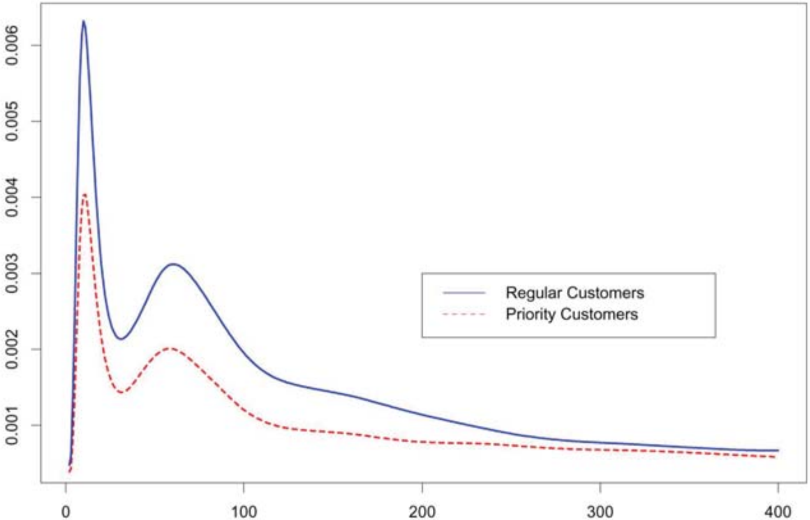
Why Bother ?

- Hourly Scale: Staffing,. . .
- Daily: Flow / Bed Control,. . .

Durations: (Im)Patience while Waiting (Psychology)

Palm: (1943–53): Irritation $\propto$ Hazard Rate

Regular over VIP Customers – Israeli Bank

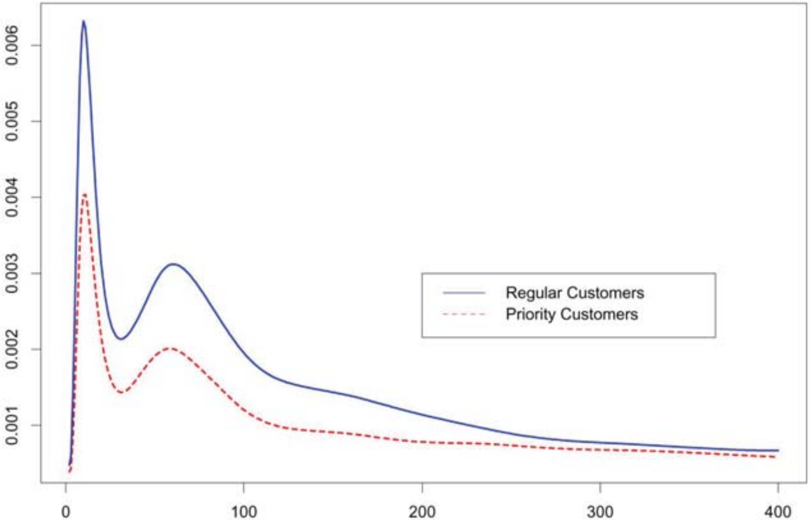


- Challenges: **Un-Censoring, Dependence, Smoothing** - requires **Call-by-Call Data**

Durations: (Im)Patience while Waiting (Psychology)

Palm: (1943–53): Irritation $\propto$ Hazard Rate

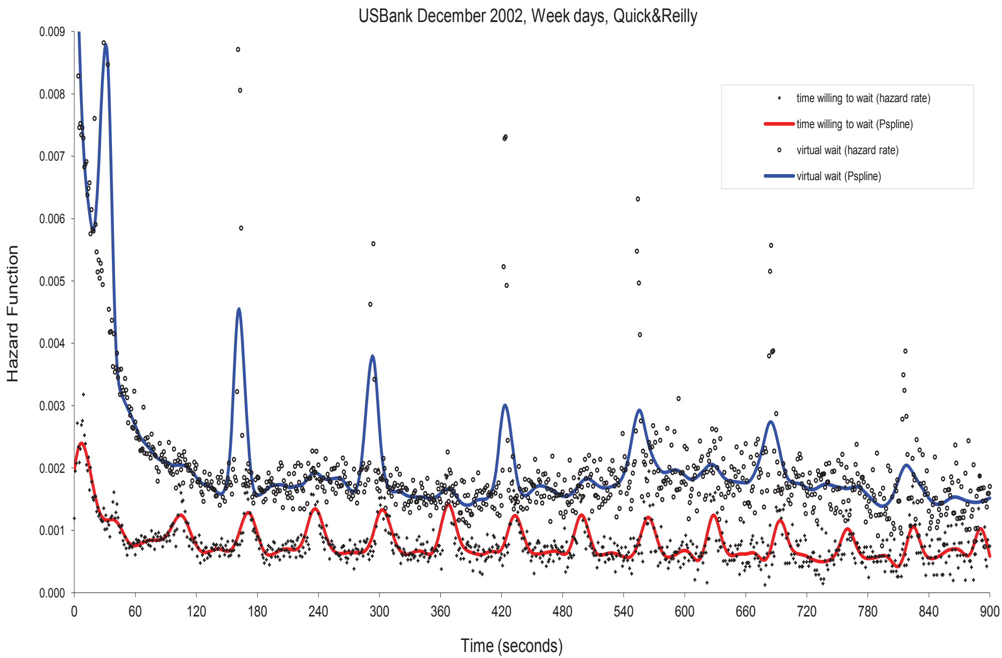
Regular over VIP Customers – Israeli Bank



- Challenges: **Un-Censoring, Dependence, Smoothing** - requires **Call-by-Call Data**
- **VIP** Customers are **more Patient** (Needy)
- **Peaks** of abandonment at times of **Announcements**

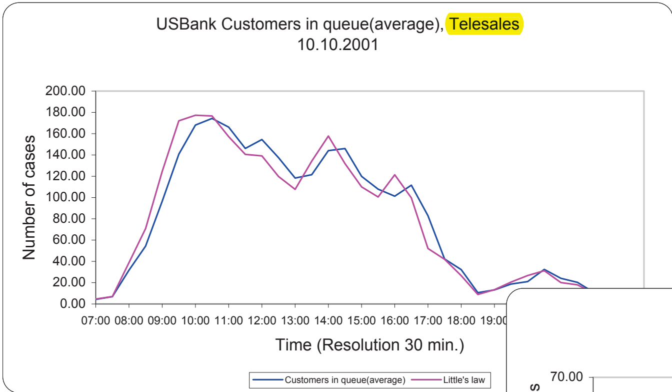
Protocols + Psychology

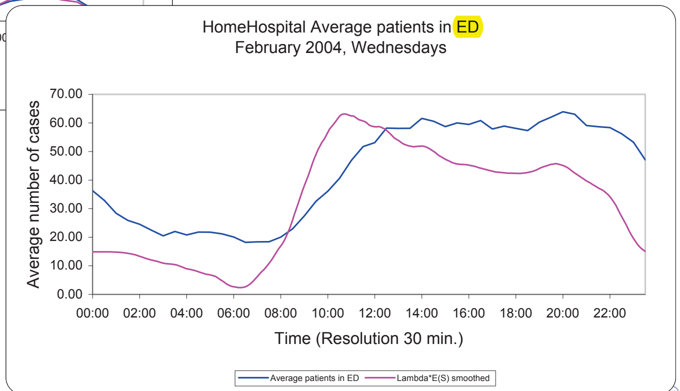
Patient Customers, Announcements, Priority Upgrades



Little's Law: Call Center & Emergency Department

Time-Gap: **# in System** lags behind **Piecewise-Little** ($L = \lambda \times W$)



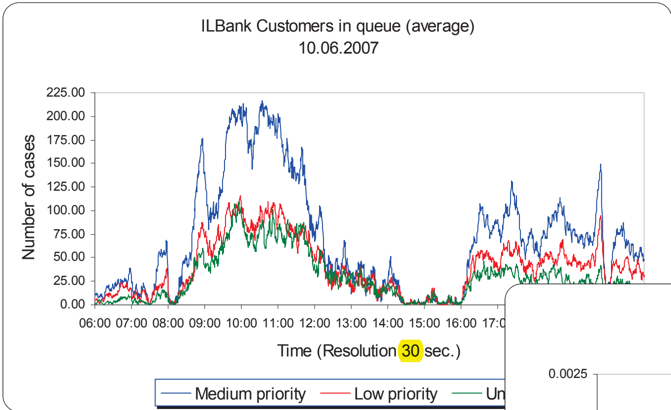


$\Rightarrow$ **Time-Varying Little's Law**

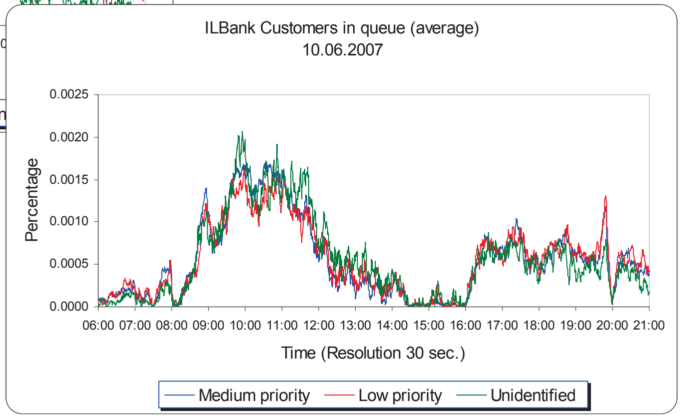
- **Berstemas & Mourtzinou;**
- **Fralix, Riano, Serfozo;** ...

Congestion Laws: Parsimonious Models

3 Queue-Lengths at 30 sec. resolution (ILBank, 10/6/2007)



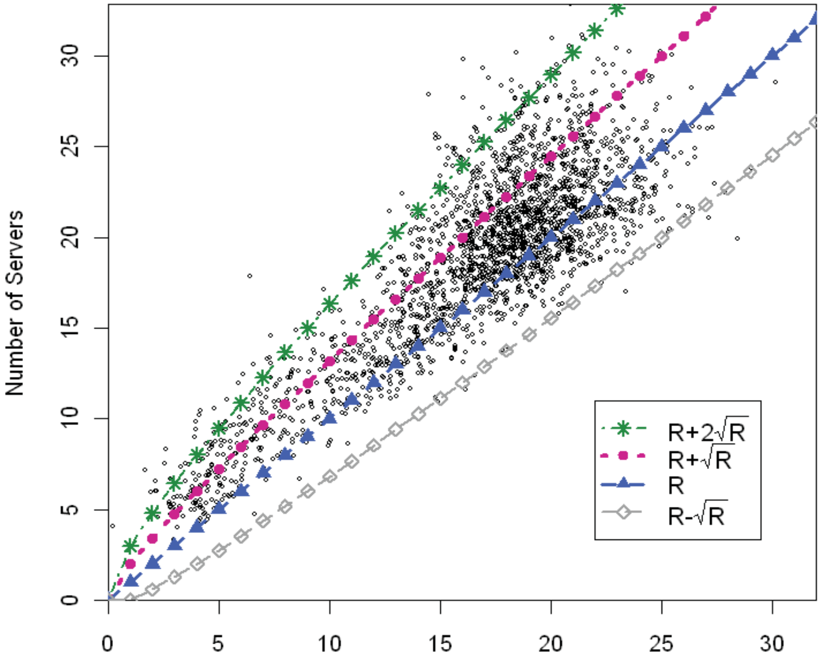
Queue "Shape"



- Area normalized to 100%
- **State-Space Collapse**

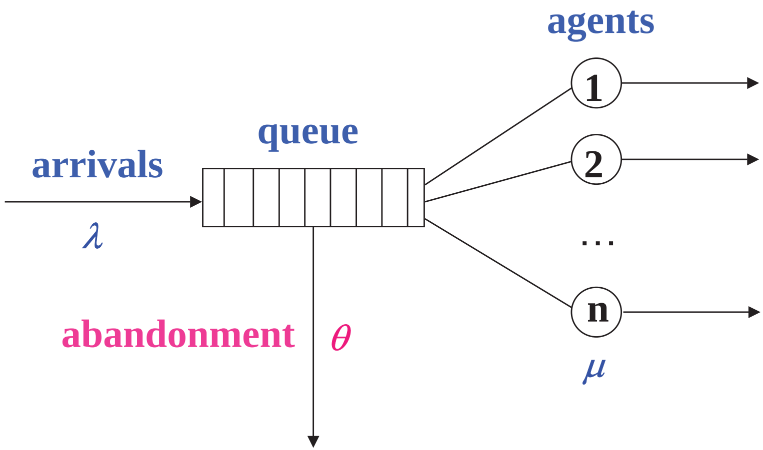
Protocols: Staffing (N) vs. Offered-Load (R $= \lambda \times E(S)$)

IL Telecom; June-September, 2004; w/ Nardi, Plonski, Zeltyn



2205 half-hour intervals (13 summer weeks, week-days)

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



Erlang-A **(Palm 1940's)** = **Birth & Death Q**, with parameters:

- λ – **Arrival** rate (Poisson)
- μ – **Service** rate (Exponential; $E[S] = \frac{1}{\mu}$)
- θ – **Patience** rate (Exponential, $E[\text{Patience}] = \frac{1}{\theta}$)
- N – Number of **Servers** (Agents).

QED Theory (Erlang '13; Halfin & Whitt '81; w/Garnett & Reiman '02)

Consider a sequence of **steady-state** M/M/N + M queues, $N = 1, 2, 3, \ldots$
Then the following points of view are **equivalent**, as $N \uparrow \infty$:

- **QED** $\quad \%\{\text{Cust Wait} > 0\} \approx \alpha, \qquad 0 < \alpha < 1;$

  or $\quad \%\{\text{Serv Idle} > 0\} \approx 1 - \alpha$

- **Customers** $\quad \{\text{Abandon}\} \approx \frac{\gamma}{\sqrt{N}}, \qquad 0 < \gamma;$

- **Agents** $\quad \text{OCC} \approx 1 - \frac{\beta+\gamma}{\sqrt{N}} \qquad -\infty < \beta < \infty;$

- **Managers** $\quad N \approx R + \beta\sqrt{R}, \quad R = \lambda \times E(S)$ not small;

Here $\boldsymbol{R}$ **= Offered Load**
eg. $\quad R = 25$ call/min. $\times$ 4 min./call $= 100$

Erlang-A: QED Approximations (Examples)

Assume **Offered Load** R not small ($\lambda \to \infty$).

Let $\hat{\beta} = \beta \sqrt{\frac{\mu}{\theta}}$, $h(\cdot) = \frac{\phi(\cdot)}{1 - \Phi(\cdot)}$ = hazard rate of $\mathcal{N}(0, 1)$.

- **Delay Probability:**

$$\mathrm{P}\{W_q > 0\} \approx \left[1 + \sqrt{\frac{\theta}{\mu}} \cdot \frac{h(\hat{\beta})}{h(-\beta)}\right]^{-1}.$$

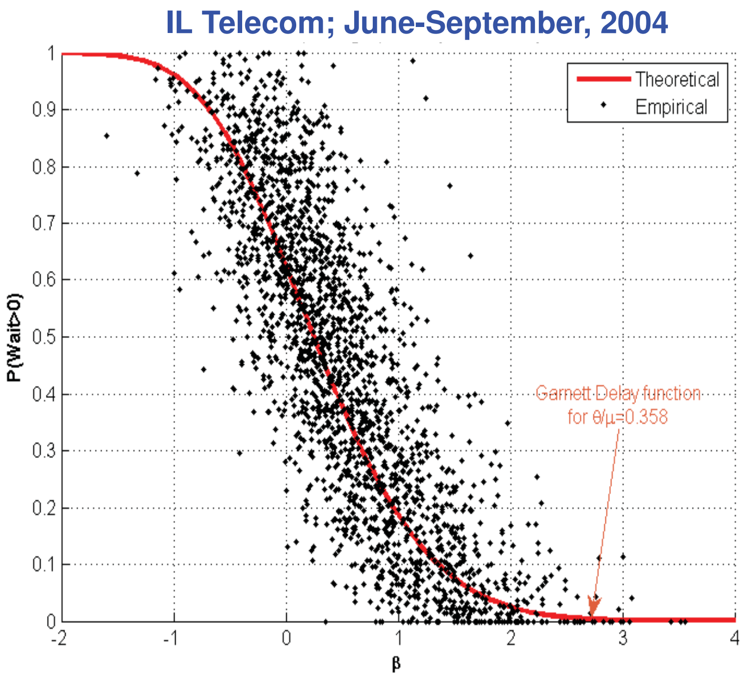
- **Probability to Abandon:**

$$\mathrm{P}\{\mathrm{Ab} | W_q > 0\} \approx \frac{1}{\sqrt{N}} \cdot \sqrt{\frac{\theta}{\mu}} \cdot \left[h(\hat{\beta}) - \hat{\beta}\right].$$

- $\boldsymbol{P\{Ab\} \propto \mathrm{E}[W_q]}$, both order $\frac{1}{\sqrt{N}}$:

$$\frac{\mathrm{P}\{\mathrm{Ab}\}}{\mathrm{E}[W_q]} = \theta \quad (\approx g(0) > 0).$$

QED Theory vs. Data: $P(W_q > 0)$



- ► 2205 half-hour intervals (13 summer weeks, week-days)
- ► Erlang-A approximations for the appropriate $\mu/\theta \approx 3$

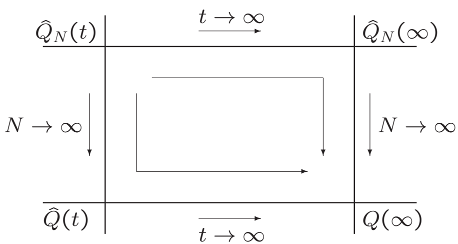
Process Limits (Queueing, Waiting)

- $\widehat{Q}_N = \{\widehat{Q}_N(t), t \geq 0\}$: stochastic process obtained by centering and rescaling:

$$\widehat{Q}_N = \frac{Q_N - N}{\sqrt{N}}$$

- $\widehat{Q}_N(\infty)$: stationary distribution of $\widehat{Q}_N$

- $\widehat{Q} = \{\widehat{Q}(t), t \geq 0\}$: process defined by: $\widehat{Q}_N(t) \xrightarrow{d} \widehat{Q}(t)$.



QED Q's in Practice, beyond Call Centers

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

For example, **M/M/1** (single-server queue): **91%** server's utilization goes with

$$\textbf{Congestion Index} = \frac{E[Wait]}{E[Service]} = \mathbf{10}.$$

QED Q's in Practice, beyond Call Centers

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

For example, **M/M/1** (single-server queue): **91%** server's utilization goes with

$$\textbf{Congestion Index} = \frac{E[Wait]}{E[Service]} = \mathbf{10}.$$

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

$\Rightarrow$ **QED Systems**

QED Intuition: Why $P\{W_q > 0\} \in (0, 1)$?

1. Why **subtle**: Consider a large service system (e.g. call center).
   - Fix λ and let $N \uparrow \infty$: $P\{W_q > 0\} \downarrow 0$, $P(I > 0) \uparrow 1$.

QED Intuition: Why $P\{W_q > 0\} \in (0, 1)$?

1. Why **subtle**: Consider a large service system (e.g. call center).
   - Fix λ and let $N \uparrow \infty$: $P\{W_q > 0\} \downarrow 0$, $P(I > 0) \uparrow 1$.
   - Fix N and let $\lambda \uparrow \infty$: $P\{W_q > 0\} \uparrow 1$, $P(I > 0) \downarrow 0$.

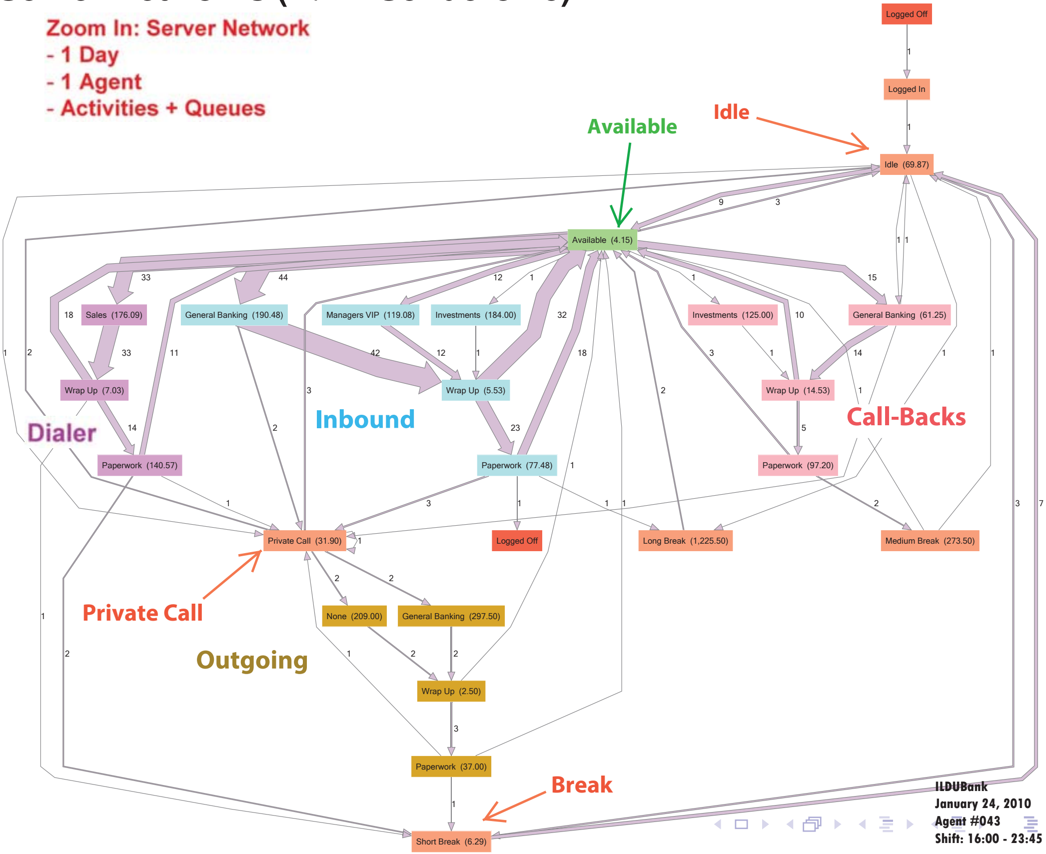
QED Intuition: Why $P\{W_q > 0\} \in (0,1)$?

1. Why **subtle**: Consider a large service system (e.g. call center).
   - Fix λ and let $N \uparrow \infty$: $P\{W_q > 0\} \downarrow 0$, $P(I > 0) \uparrow 1$.
   - Fix N and let $\lambda \uparrow \infty$: $P\{W_q > 0\} \uparrow 1$, $P(I > 0) \downarrow 0$.
   - $\Rightarrow$ **Must** have both λ and N increase simultaneously:
   - $\Rightarrow$ (CLT) **Square-root staffing**: $N \approx R + \beta\sqrt{R}$ $\left(\lambda \approx N\mu - \beta\sqrt{N\mu}\right)$

QED Intuition: Why $P\{W_q > 0\} \in (0, 1)$?

1. Why **subtle**: Consider a large service system (e.g. call center).
   - Fix λ and let $N \uparrow \infty$: $P\{W_q > 0\} \downarrow 0$, $P(I > 0) \uparrow 1$.
   - Fix N and let $\lambda \uparrow \infty$: $P\{W_q > 0\} \uparrow 1$, $P(I > 0) \downarrow 0$.
   - $\Rightarrow$ **Must** have both λ and N increase simultaneously:
   - $\Rightarrow$ (CLT) **Square-root staffing**: $N \approx R + \beta\sqrt{R}$
     $\left(\lambda \approx N\mu - \beta\sqrt{N\mu}\right)$
2. QED **Excursions**

Server Networks (w/ A. Senderovic)

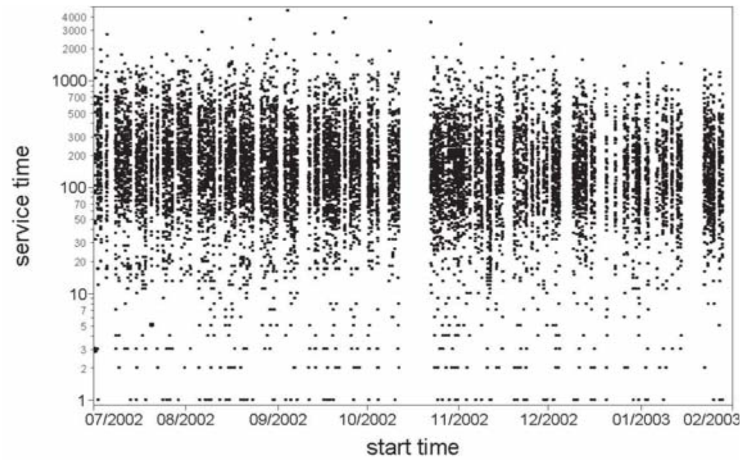


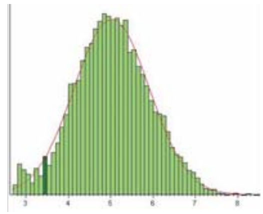
Individual Agents: Service-Duration, Variability

w/ **Gans, Liu, Shen & Ye**

Agent 14115

Service-Time Evolution: 6 month | **Log(Service-Time)**

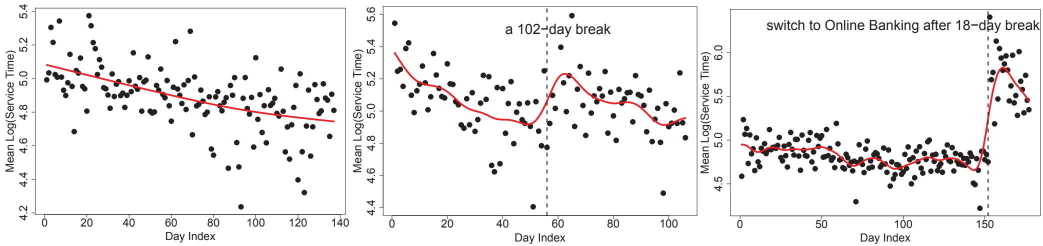




- **Learning**: Noticeable decreasing-trend in service-duration
- **LogNormal** Service-Duration, individually and collectively

Individual Agents: Learning, Forgetting, Switching

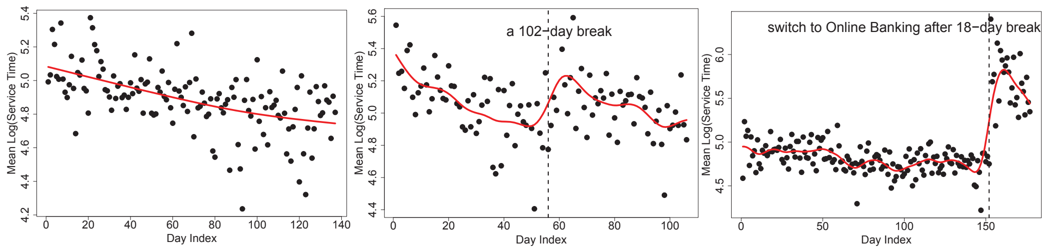
Daily-Average Log(Service-Time), over 6 months
Agents 14115, 14128, 14136



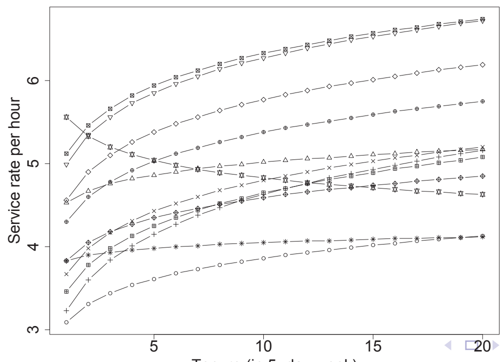
Individual Agents: Learning, Forgetting, Switching

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

Agents 14115, 14128, 14136

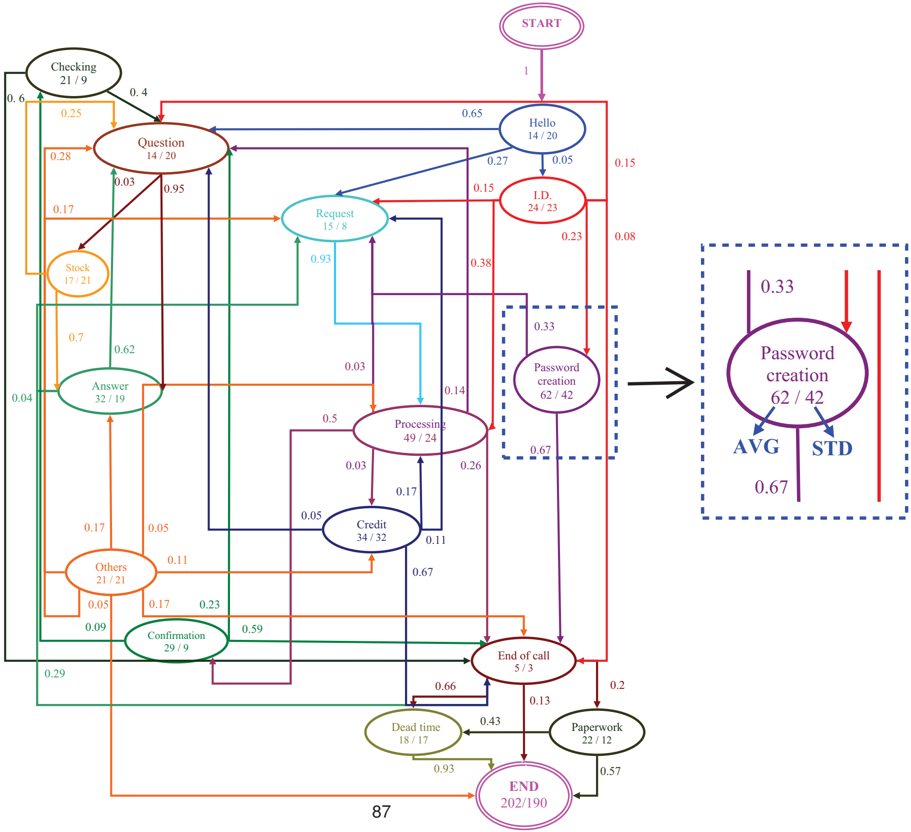


Weakly Learning-Curves for 12 Homogeneous(?) Agents



Tele-Service Process (Beyond present SEE Data)

Retail Service (Israeli Bank)



Statistics OR IE Psychol. MIS

The Technion SEE Center / Laboratory

Data-Based Service Science / Engineering

Technion Research Impact: An Example

Web Images Videos Maps News Shopping Gmail more ▼ Sign in

Google scholar | call centers | Search | Advanced Scholar Search | Scholar Preferences

Scholar | Articles and patents | anytime | include citations | Results **1** - **10** of about **2,510,000**. (**0.06** sec)

[PDF] Telephone **call centers**: Tutorial, review, and research prospects — psu.edu [PDF]
N Gans, G Koole, A Mandelbaum - Manufacturing and service operations …, 2003 - Citeseer
∗ The Wharton School, University of Pennsylvania,Philadelphia, PA 19104, USA gans@wharton.upenn.edu **...** † Vrije Universiteit, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands koole@few.vu.nl **...** ‡ Industrial Engineering and **...**
Cited by 455 - Related articles - View as HTML - All 32 versions

[PDF] Designing a **call center** with impatient customers — technion.ac.il [PDF]
O Garnet, A Mandelbaum, M … - Manufacturing & Service …, 2002 - iew3.technion.ac.il
ABSTRACT. The most common model to support workforce management of telephone **call centers** is the M/M/N/B model, in particular its special cases M/M/N (Erlang C, which models out busy-signals) and M/M/N/N (Erlang B, disallowing waiting). All of these models lack a **...**
Cited by 245 - Related articles - View as HTML - All 18 versions

Statistical Analysis of a Telephone **Call Center** — psu.edu [PDF]
L Brown, N Gans, A Mandelbaum, A Sakov, H … - Journal of the American …, 2005 - ASA
A **call center** is a service network in which agents provide telephone-based services. Customers who seek these services are delayed in tele-queues. This article summarizes an analysis of a unique record of **call center** operations. The data comprise a complete operational history **...**
Cited by 205 - Related articles - All 41 versions

Dimensioning large **call centers** — bell-labs.com [PDF]
S Borst, A Mandelbaum, MI Reiman - Operations research, 2004 - JSTOR
CWI, P. O. Box 94079, 1090 GB Amsterdam, The Netherlands, and Bell Labs, Lucent Technologies, Murray Hill, New Jersey 07974-0636, sem.borst@cwi.nl Avi Mandelbaum Faculty of Industrial Engineering and Management, Technion, Haifa 32000, Israel, avim@tx. **...**
Cited by 156 - Related articles - BL Direct - All 19 versions

Queueing models of **call centers**: An introduction — psu.edu [PDF]
G Koole, A Mandelbaum - Annals of Operations Research, 2002 - Springer
... the modern **call center** is a complex socio-technical system. It thus enjoys central features that

Data-Based Creation ServNets: some Technicalities

- ServNets = **QNets, SimNets, FNets, DNets**
- **Graph Layout**: Adapted from but significantly extends Graphviz (AT&T, 90's); eg. *edge-width*, which must be restricted to *poly-lines*, since there are "no parallel Bezier (Cubic) curves $(B_n(p) = E_p F[B(n,p)], 0 \leq p \leq 1)$
- **Algorithm**: Dot Layout (but with cycles), based on Sugiyama, Tagawa, Toda ('81): "Visual Understanding of **Hierarchical** System Structures"

Data-Based Creation ServNets: some Technicalities

- ServNets = **QNets, SimNets, FNets, DNets**
- **Graph Layout**: Adapted from but significantly extends Graphviz (AT&T, 90's); eg. *edge-width*, which must be restricted to *poly-lines*, since there are "no parallel Bezier (Cubic) curves $(B_n(p) = E_p F[B(n,p)], 0 \leq p \leq 1)$
- **Algorithm**: Dot Layout (but with cycles), based on Sugiyama, Tagawa, Toda ('81): "Visual Understanding of **Hierarchical** System Structures"
- Draws data directly from **SEELab** data-bases:
  - Relational DBs (Large! eg. USBank Full Binary = 37GB, Summary Tables = 7GB)
  - Structure: Sequence of events/states, which (due to size) partitioned (yet integrated) into days (eg. call centers) or months (eg. hospitals)
  - Differs from industry DBs (in call centers, hospitals, websites)

Skills-Based Routing in Call Centers

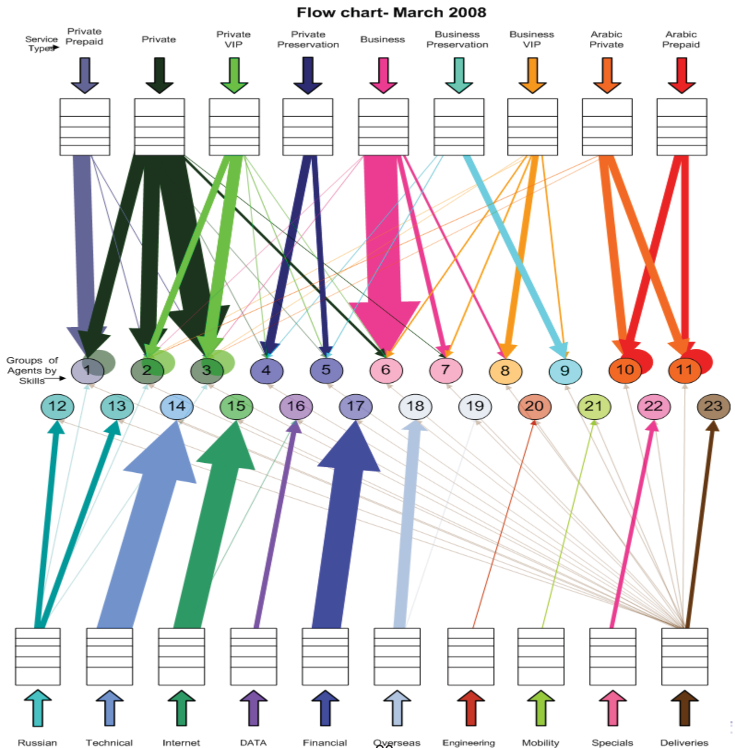
EDA and OR, with I. Gurvich and P. Liberman

Mktg. $\Rightarrow$

OR $\Rightarrow$

HRM $\Rightarrow$

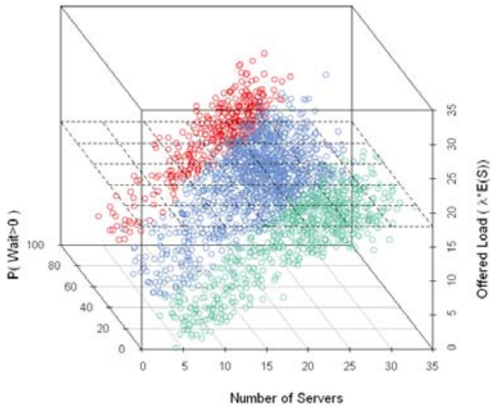
MIS $\Rightarrow$



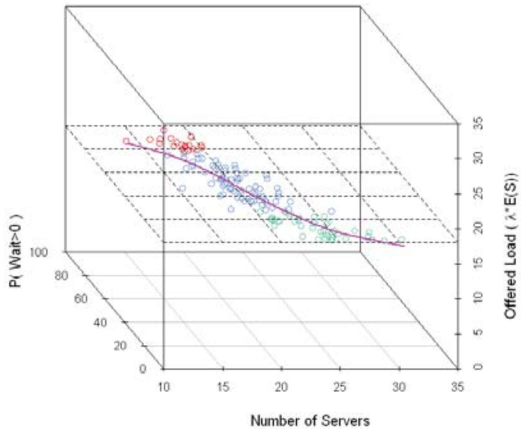
Operational Regimes in a Call Center: Q vs. E

Large Israeli Bank

P{$W_q > 0$} vs. (R, N)



R-Slice: P{$W_q > 0$} vs. N



3 Operational Regimes:

- **QD**: $\leq 25\%$
- **QED**: $25\% - 75\%$
- **ED**: $\geq 75\%$