ETM
ENGINEERING & TECHNOLOGY MANAGEMENT

CALL CENTER OPTIMIZATION AND STAFF SCHEDULING UTILIZING THE ERLANG A MODEL

ETM 540 Operation Research

Winter Term 2013

Team Members

Jing Jiang

Zeina Boules

Mel Cooper

Instructor

Dr. Tim Anderson

Table of contents

1. Introduction

Call centers enable organizations to manage and communicate with large numbers of people or customers in an efficient and cost-effective manner. For many organizations the efficiency of call center operations has a significant impact on the organization’s success. Some call centers are indeed the primary source of generating revenue for the organization such as a catalogue company or an airline.

The costs of creating and maintaining a call center can be significant in terms of the initial investment and ongoing recurring costs management. In perhaps most cases, the two highest costs components are staffing and telecommunications service. This report provides an optimization model for call center staffing and telecom costs (specifically telecommunications trunking) and these components are directly related.

The purpose of this report is to examine traffic for a generic call center, conduct an optimization analysis of staffing and telecom trunking, and identify the thresholds that signal the implementation of changes in telecommunications trucking and staffing to address significant changes in traffic load.

For many organizations, traffic demand varies considerably throughout the year for a variety of reasons. Most commonly, seasonal changes such as the winter holidays bring about increases/decreases in traffic loads. Often such changes are of a magnitude that increases in staffing and telecommunications facilities are a foregone conclusion and year after year organizations implement the necessary changes to meet their needs. However, other seasonal changes are less clear and require detailed analysis to assess the best course of action, particularly in call center environments where multiple products demand specialized agents to meet customer service needs. We are proposing an optimization study that will create a best practices optimization model that call center managers can use to assess their call center costs for these components and help identify areas of potential improvement.

Telecommunications optimization

- Call queuing/Routing tree optimization for product specific call center agents
- Optimization analysis of best practices to identify traffic load thresholds indicating provisioning or reduction of additional facilities (trunk/lines).

1.1. Staffing

- Optimization analysis of best practices to identify optimal hiring mix of permanent and seasonal staff thresholds based on scalable demand levels.
- Optimization analysis to identify the optimal hiring mix of product specific call center agents based on demand.
- Optimization analysis to identify daily staffing demand based on traffic demand ion a multi-product environment with product specific agents.

This project will provide a model for call center managers to improve efficiency and lower costs. The two primary areas of focus, staffing and telecom facility management, require consideration of a complex number of variables. Ensuring that a sufficient number of telephone lines are maintained and available to support operations is generally the responsibility of a telecom manager or telecom analyst. Because these facilities are critical to business operations, over-provisioning is a common problem and can result on thousands of dollars of unnecessary operational costs. Under-provisioning facilities however can also have a negative impact resulting in longer wait times, higher call abandonment rates, and poor customer service perceptions.

Call Center managers are typically responsible for determining staffing levels and scheduling. Traffic load or the number of calls processed, staffing costs and the call center's goals or targets for call completion, percentage of blocked or abandoned calls are the primary variables necessary to determine staffing levels.

Both the call center manager and a telecom manager must work together to determine how call queuing is managed. Call queuing is based on business goals or target wait times, call answer rates.

The challenge for the telecom manager is determining the optimal number of lines ensuring service availability while at the same time avoiding over-provisioning. Minimizing the monthly recurring charges of telecommunications facilities or telephone lines can result in significant costs avoidance. The monthly charge for a single telephone line varies by supplier, typically ranging from $35 to $60 per line. A mid-size business may have hundreds of telephone lines and as such, the total annual costs can be significant. For example, at fifty dollars per line ($50) for 250 lines, the annual cost is $150,000 per year. Large global organizations may spend millions of dollars per year for these services.

2. Telephone network

Determining the optimal number of lines an organization needs to support its business operations can be a very complex process for large business. Network optimization must consider a variety of factors including but not limited to the nature and type of the business application supported, as well as technical issues such as whether or not lines are dedicated to specific applications or shared among multiple applications.

Most telephone lines are used for either direct dial outbound calls including local, domestic and international long distance and/or toll free inbound calls such as 800/888 numbers that are often used for customer service or telemarketing operations.

Determining how many lines are needed is based, in part, on the business application and type of calls. Business applications might include support for call center operations (customer service, sales, etc.), intra-company communications, and external outbound calls supporting day-to-day operations. To address an organization's needs, suppliers offer several different options to meet the demand for telephone lines. These options include scalable and dynamically allocated high bandwidth access facilities, such as T1s and T3s, which are more efficient and may help ensure service availability.

While a small company might order individual telephone lines to meet its needs, larger organizations typically order T1s or T3s – high bandwidth access facilities where a single T1 provides 24 individual lines, and a single T3 equals 672 individual lines. When an organization's needs exceed roughly 16 lines that cost $50 each per month or $800 total, a single T1 providing 24 lines may be more economical depending on pricing.

Figure 1, below, illustrates a typical configuration where an organization maintains 3 individual access facilities (T1s), each of which provides 24 individual lines and is dedicated solely to a specific application (Customer Service, Telemarketing, Individual phones). In effect this configuration ensures that the customer service and the telemarketing department each have a minimum (and maximum) of 24 lines available at all times. The monthly charge for a T1 facility ranges from $300 to $1200 per month.

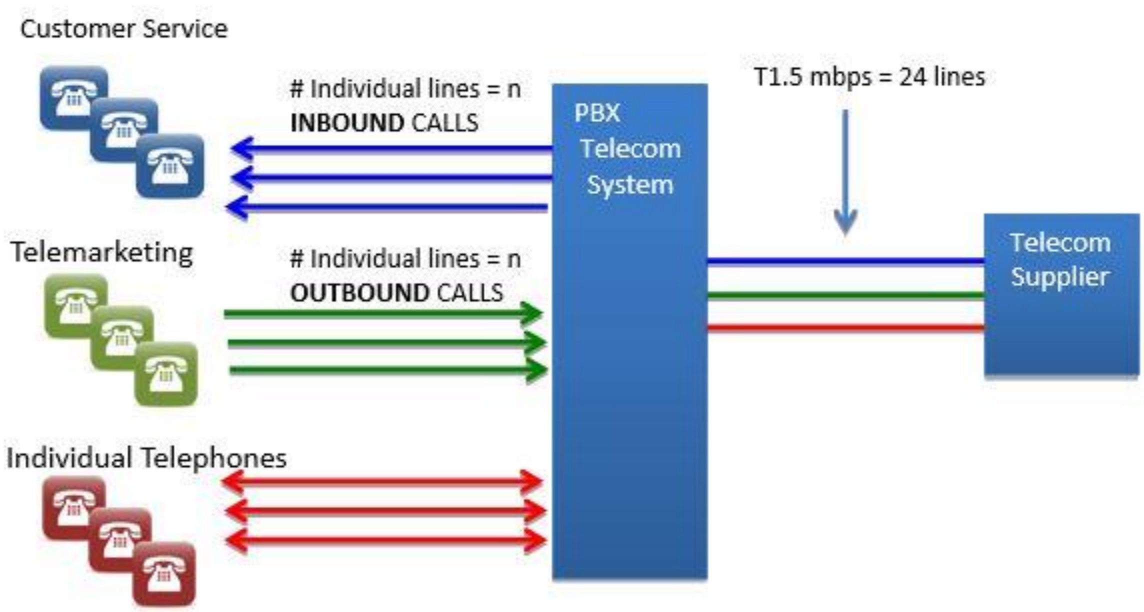


Figure 1: A typical configuration where an organization maintains 3 individual access facilities

While the above design ensures availability of telecom lines, it may not be the most efficient configuration. Telecom suppliers can configure a single T1 such that the 24 individual lines are dynamically allocated based on demand. If all 24 lines or channels are dynamically allocated then they are utilized on a first come – first served basis. Since this could result in problems where for example an outbound sales promotion might utilize the majority of the lines thereby blocking potential inbound customer service calls, a hybrid configuration can be provisioned. A hybrid configuration includes both shared and dedicated lines, increasing efficiency by sharing resources while also ensuring a number of dedicated resources (lines).

Determining the optimal number of lines an organization needs depends on many factors including:

- The average monthly/daily number of calls processed.
- Significant changes in demand throughout the year such as seasonal changes or other business cycles that have a major impact on call activity such as sales promotions or product recalls.
- Industry specific standards of operation or the need to provide a level of service comparable to ensure a business is competitive.

- The use of automated attendants also called 'call prompters' (i.e. press 1 for sales, 2 for customer service, & 0 for the operator).
- The application or type of calls processed. For example, an emergency 911 call center would likely seek to provision sufficient lines to attain 0% blockage and avoid any blocked calls (busy signals).
- The organization's business goals. For example, an airline may advertise that callers will always reach a 'real person', rather than an automated attendant, or seek to ensure all calls are answered within a pre-determined time period.
- Staffing levels for high volume users such as telemarketing call centers – having sufficient lines to accommodate all calls is wasteful of the organization doesn't have people to answer them.
- The recurring costs of maintaining telephone facilities/lines and the organization's telecommunications equipment capacity.

The key aspect of telecom optimization is the organization's 'grade of service' goals. An organization's grade of service is typically measured in number of calls blocked or conversely the percentage of calls it can process.

For example, an emergency 911 call center would likely seek to accommodate every caller and as such, its goal would be to provision enough lines to attain nearly zero blocked calls during normal business operations. Nevertheless, no plan can accommodate every potential event. A disaster such as a hurricane for example could result in an excessive number of callers that exceed the telephone systems capabilities. As such, any call center that is critical to business operations should include disaster planning to address such potentialities. For a 911 emergency call center for example, Disaster planning would consider other factors such as the maximum capacity of the community's emergency services (hospitals, police, fire, medical) limitations, call center supplemental staffing capacity and limitations, and of course the costs of services.

Clearly, an organization's approach to telecommunication facilities optimization is often unique to the organization's needs.

3. Optimization Methodology

Once an organization has determined its grade of service goals, calculating the number of facilities or lines necessary to meets its goals are based on the following:

- The busy hour period – the number of calls completed during the organization's busiest hour of a calendar month.
- The total number of calls processed in an average month. Typically several months are averaged to determine the average monthly calls.
- The average call duration measured in minutes during normal business hours.
- The number of business days during a calendar month (typically 22)
- An industry constant is also used in the calculations and is based on the type of business.

4. Inbound Call Centers

The diagram below shows the queue model of a typical call center. The customers are callers and the resources are telephone (trunk) lines and agents. Tele-queues consist of callers that are waiting for available staff to process their call. The number of trunk lines should be greater than the number of agents. If all trunk lines are occupied, subsequent incoming calls will receive busy signals – are blocked from entering the system. If trunks are available but no agent is available then the call can be held 'in cue,' where they are typically informed via automate voice that they are waiting (on hold) for the next available agent and their call will be connected as soon as possible. If there is no agent available to answer the call immediately, the caller can either determine to continue to remain on hold until their call is answered or to hang up. The latter behavior is called balking or call abandonment. Typically, telecom cueing systems are configured to process callers on a first-come, first-served basis. Exceptions to this are when customers are segmented, based for example of the revenue they generate, identified via the caller's telephone number, and given priority routing.

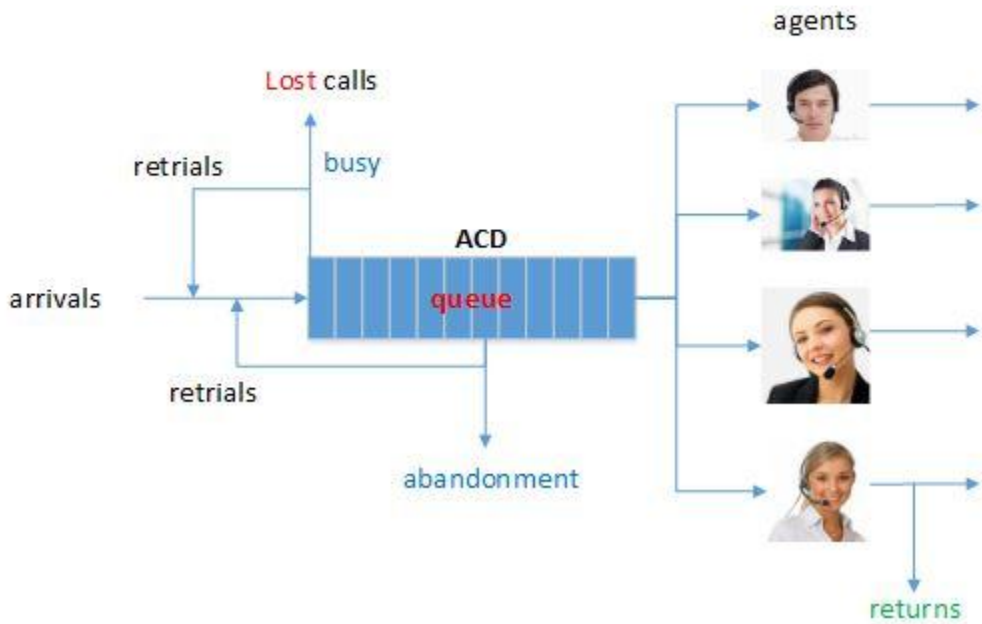


Figure 2: Call Center Diagram

Once the caller gets served and exits the system, the respondent resource such as the agent and the trunk line will be available again. Calls that do not get served in queuing order because for example, they are routed to an agent that becomes unavailable (takes a break, end of shift, etc.) become retrials and are returned to the front of the cue. The remaining blocked and abandoned calls are lost[1-3].

4.1. Erlang models for call center analysis

Many queuing models were developed to optimize and analyze the call centers. Kendall notation was developed to describe queuing system [2] – A/B/X/Y, where A indicates the arrival distribution, B indicates the service pattern, X indicates the number of agents, Y indicates the system capacity.

We evaluated the most common traffic analysis models Erlang A, B and C, and concluded that the Erlang A model is best suited for our analysis because it takes into account parameters that Erlang B and C do not. The Erlang A model is: M/M/N/K + A - where the first M indicates Markovian interarrival times (following exponential distribution), the second M indicators Markvian service times (following exponential distribution), N indicates the number of agents, K indicates the queue length, and A indicates the system with abandonment[4,5,7-16].

The Erlang C (M/M/N) model [7, 11]] is simple and widely used, but it doesn't acknowledge abandonment. The model assumes that callers are patient and never hang up or abandon calls. It also assumes the waiting queue length is infinite. In an actual call center environment, an infinite number of trunk lines is simply cost prohibitive and callers have a limited amount of patience and of course a percentage will hang up or abandon the call.

Erlang B (M/M/N/N) model [7, 11] adds blocking. This model presumes that the number of telephone trunk lines is the same as the number of agents in the center. Based on this model, if there is no agent available, the caller will be blocked. This model doesn't allow callers to wait in a holding queue. Utilizing this model would require a call center to add an agent for every line or trunk installed, and does not take into account call cueing capabilities.

Based upon our analysis, the Erlang A model is the ideal mode to be used for the call center data analysis, call forecasting, and staff scheduling. It considers not only call blocking but also call abandonments.

In the Erlang A model, there are four parameters we must consider to conduct our analysis [7, 11]:

- λ – arrival rate (calls per unit of time), Poisson arrival rate
- μ – service rate (1/ μ is the average duration of service), Exponential service rate
- N – the number of agents
- θ – individual abandonment rate (1/θ is the average patience). Patience time is exponentially distributed with mean θ^{-1}.

We also have to mention another parameter - Offered load. Its definition is:

$$\alpha = \frac{\lambda}{\mu}$$

In the Erlang A model, patience time is exponentially distributed with mean θ^{-1}, then[11]

$$H(x) = \frac{1}{\theta} \cdot (1 - e^{-\theta x})$$

Define the incomplete Gamma function

$$\gamma(x, y) \triangleq \int_0^y t^{x-1} e^{-t}\, dt, x > 0, y \geq 0$$

$$\varepsilon \triangleq \frac{\sum_{j=0}^{N-1} \frac{1}{j!}\left(\frac{\lambda}{\theta}\right)^{j}}{\frac{1}{(N-1)!}\left(\frac{\lambda}{\mu}\right)^{N-1}}$$

Then

$$J = \frac{exp\left\{\frac{\lambda}{\theta}\right\}}{\theta} \cdot \left(\frac{\lambda}{\theta}\right)^{\frac{N\mu}{\theta}} \cdot \gamma\left(\frac{N\mu}{\theta}, \frac{\lambda}{\theta}\right)$$

$$J(t) = \frac{exp\left\{\frac{\lambda}{\theta}\right\}}{\theta} \cdot \left(\frac{\lambda}{\theta}\right)^{\frac{N\mu}{\theta}} \cdot \gamma\left(\frac{N\mu}{\theta}, \frac{\lambda}{\theta} e^{-\theta t}\right)$$

$$J_H = \frac{J}{\theta} - \frac{exp\left\{\frac{\lambda}{\theta}\right\}}{\theta^2} \cdot \left(\frac{\lambda}{\theta}\right)^{\frac{N\mu}{\theta}+1} \cdot \gamma\left(\frac{N\mu}{\theta} + 1, \frac{\lambda}{\theta}\right)$$

$$J_H = \frac{J(t)}{\theta} - \frac{exp\left\{\frac{\lambda}{\theta}\right\}}{\theta^2} \cdot \left(\frac{\lambda}{\theta}\right)^{\frac{N\mu}{\theta}+1} \cdot \gamma\left(\frac{N\mu}{\theta} + 1, \frac{\lambda}{\theta} e^{-\theta t}\right)$$

Then

$$P\{Ab\} = \frac{1 + (\lambda - \mathrm{N}\mu)\mathrm{J}}{\epsilon + \lambda\mathrm{J}}$$

$$P\{Sr\} = \frac{\epsilon + \mathrm{N}\mu\mathrm{J} - 1}{\epsilon + \lambda\mathrm{J}}$$

$$P\{W\} = \frac{\lambda\mathrm{J}_H}{\epsilon + \lambda\mathrm{J}}, W - waiting\ time$$

$$P\{Q\} = \frac{\lambda^2\mathrm{J}_H}{\epsilon + \lambda\mathrm{J}}, Q - queue\ length$$

In fact, we also need to give a safety coefficient for the number of the trunk lines, Garnett O et al. [7] proposed the safety staffing rule to determine to the coefficient.

In an actual call center, agent skill levels may be different. This difference may require the call center routes the call based upon the question categories or specific products, further increasing the complexity of staffing and trunking optimization. In our analysis, we assume customers and agents are all homogeneous. In the call center evaluation or planning, we need to:

Minimize the Required Staff and Telephone Lines while Meeting the Service level Agreement (SLA)

- $P(Delay \leq \tau) \geq 0.80$
  A common rule of thumb is the goal that at least 80% of the customers to be served within τ seconds.

- $P(Blocking) \leq b$
  b is the specified the blocking service level target

- $P(Abandoment) \leq c$
  c is the specified the abandonment level target

There are additional important measures we consider include:

- Delay probability P{W > 0}

- Average wait E[W]

- ASA (Average Speed of Answer)
  Used extensively in call centers; usually defined as E[W|Sr];

- Agents' occupancy

$$\rho = \frac{\lambda \cdot (1 - \mathrm{P}\{\mathrm{Ab}\})}{N\mu}$$

- Average queue-length E[Q].

4.2. A Call center data analysis

For the purposes of our analysis, we gathered and analyzed traffic data from a non-profit call center over a period of nine months to obtain the key parameters, averages and determined the busy hour to which we apply our Erlang model. In addition, we created target service levels, limitations and constraints, and applied the Erlang A model to obtain the optimal number of agents and trunks. Based upon inputs such as λ – arrival rate - from historical data and the standard fixed parameters or segmentation/intervals of the data set; 60 minutes.

We based our calculations, analysis and forecast on the standard rules/guidelines outlined below [14]:

Arrivals (λ)

- Typically Poisson, time-varying rates, constant at 15/30/60 min scale

- Significant uncertainty concerning future rates – prediction

- Predict separately daily volumes and fraction of arrivals per time interval

Services (μ)

- Trunking is consistent from day to day – estimates based on actually traffic utilization.

- Trunk utilization varies depending on time-of-day, however, because our analysis is based on the busy hour (busiest hour of the day, of the month, based on eight months of data) staffing and trunking recommendations have a high level of reliability to meet our business targets such as percent of calls blocked, abandoned, etc.).

- Typically, service time ≠ talk time (Our analysis utilized actually call data to determine an average length of call, and calculated agent after-call work time based on the number of calls processed and the average length of agents work day).

talk time + wrap-up time (after-call work) + . . . or

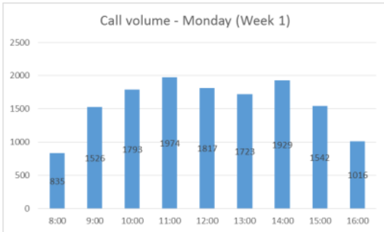
$$\frac{Total\ Working\ Time - Total\ Idle\ Time}{Number\ of\ Served\ Customers}$$

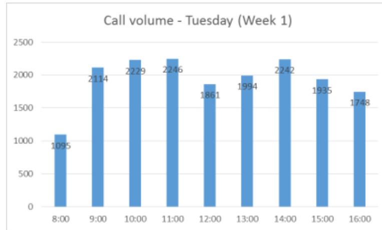
Number of agents (N)

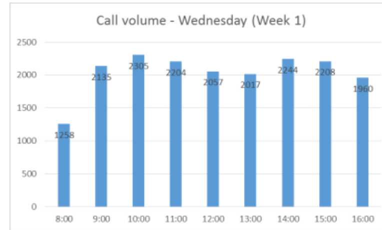
- Based on workforce management (WFM) - performance goals.
One gets number of FTE's (Full Time Equivalent positions).

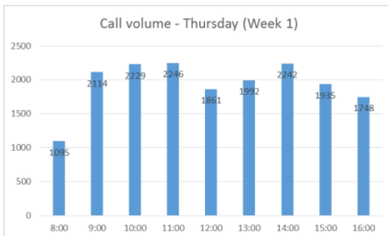
- Agents on schedule = FTE's · RSF (Rostered Staff Factor) (RSF > 1).
Reasons: absenteeism, unscheduled breaks, . . .

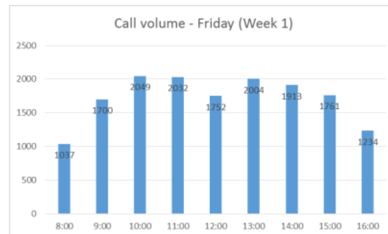
The Figures below are calls per intervals. Although we compiled 8 months of data, due to time limitations, our analysis is based primarily January, 2003, traffic/call data.

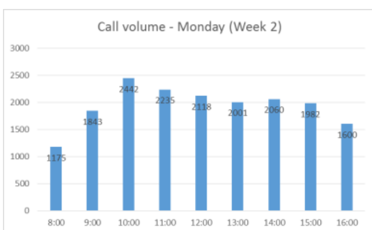

Call volume - Monday (Week 1)
2500 2000 1500 1000 500 0
835 1526 1793 1974 1817 1723 1929 1542 1016
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00

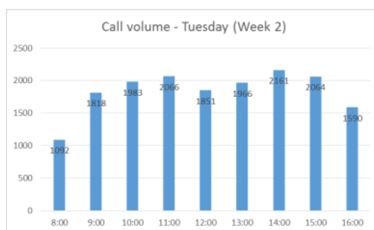

Call volume - Tuesday (Week 1)
2500 2000 1500 1000 500 0
1095 2114 2229 2246 1861 1994 2242 1935 1748
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00

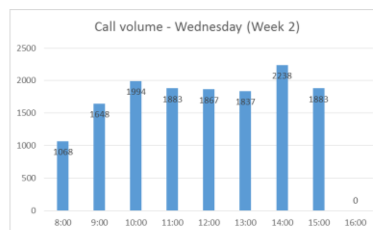

Call volume - Wednesday (Week 1)
2500 2000 1500 1000 500 0
1258 2135 2305 2204 2057 2017 2244 2208 1960
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00

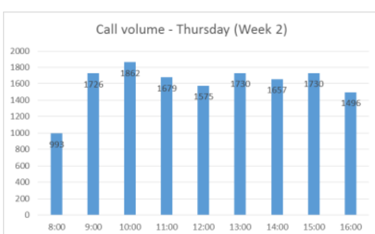

Call volume - Thursday (Week 1)
2500 2000 1500 1000 500 0
1095 2114 2229 2246 1861 1992 2242 1935 1748
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00

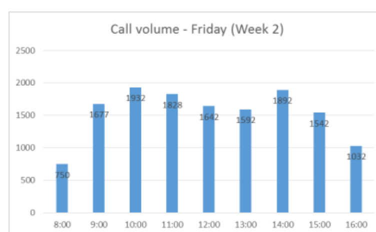

Call volume - Friday (Week 1)
2500 2000 1500 1000 500 0
1037 1700 2049 2032 1752 2004 1913 1761 1234
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00

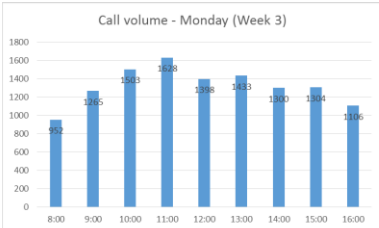

Call volume - Monday (Week 2)
3000 2500 2000 1500 1000 500 0
1175 1843 2442 2235 2118 2001 2060 1982 1600
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00

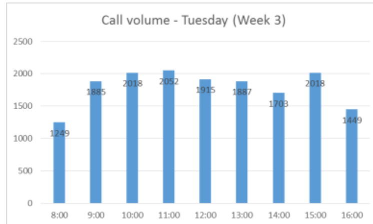

Call volume - Tuesday (Week 2)
2500 2000 1500 1000 500 0
1092 1818 1983 2066 1851 1966 2161 2064 1590
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00

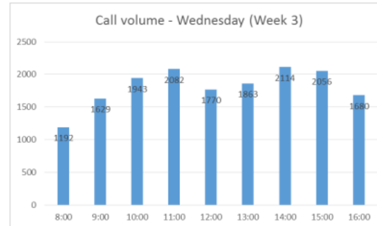

Call volume - Wednesday (Week 2)
2500 2000 1500 1000 500 0
1068 1648 1994 1883 1867 1837 2238 1883 0
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00

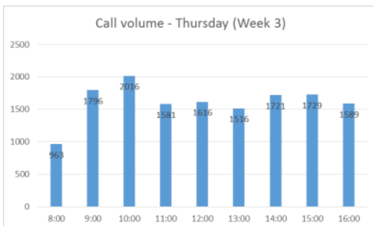

Call volume - Thursday (Week 2)
2000 1800 1600 1400 1200 1000 800 600 400 200 0
993 1726 1862 1679 1575 1730 1657 1730 1496
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00

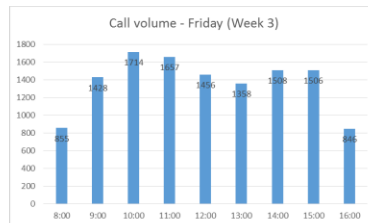

Call volume - Friday (Week 2)
2500 2000 1500 1000 500 0
750 1677 1932 1828 1642 1592 1892 1542 1032
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00

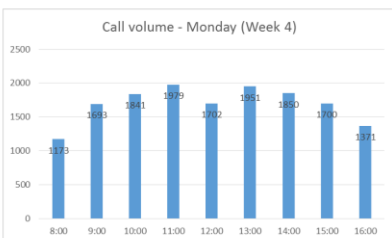


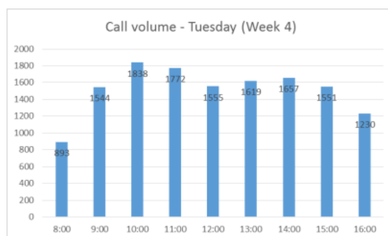


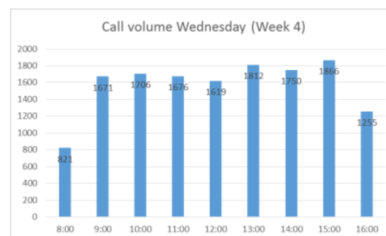


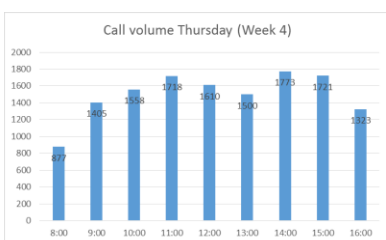


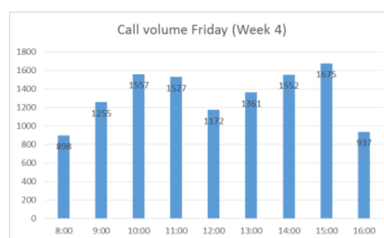












Based upon the call volume diagram, we can ascertain a number of calling patterns during a typical week:

- The middle of the week generates the most traffic (Wednesday), Mondays and Fridays are typically the slowest days of the week.

- The busiest time of the day are consistently between 10:00am – 12:00pm and 2:00pm-4:00pm.
- Most of the traffic is generated at the beginning rather than the end of the month suggesting that calls are generated from an end of the month billing cycle, and calls continually trapper off throughout the month until the next billing cycle.

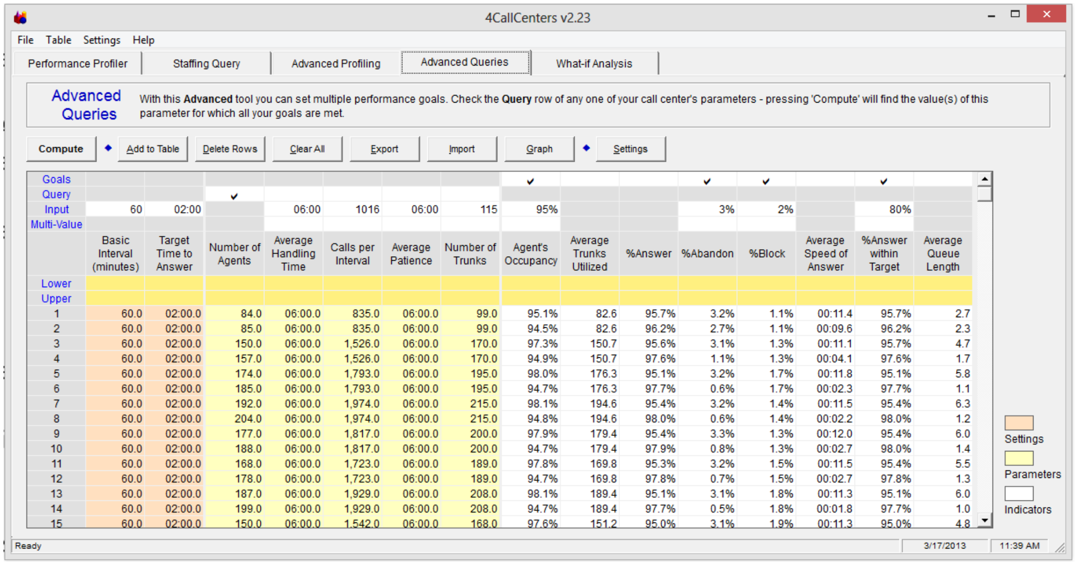
4.3. The 4CallCenters software

To conduct our analysis we used a free Erlang-A calculation software [21] "4CallCenters" to analyze/optimize the call center we evaluated. This software can be downloaded and free to use. Once we entered our call data, we created the target parameters as listed below to generate trunking & staffing recommendations. Our targets and parameters are as follows:

- The time interval to determine the busy hour is 60 minutes
- The target time to answer is 2 minutes, and 80% callers should get answered in 2 minutes.
- The maximum wait time is 6 minutes
- The service time (average length of call) is 6 minutes
- The agent's occupancy rate is above 95%
- The blocking rate is 2%
- The abandonment rate is 3%

Utilizing our call volume data, we input the number of calls during respondent intervals and our parameters and ran the software to generate recommendations. Based on output data, we found that to the optimal number of agents and trunks changed significantly based on subtle changes in our target parameters, and as such ran multiple reports/scenarios. For example, the first run, you may get the optimal number of agents, but the initial number of trunks may not be an optimal value. This resulted in abandonment and blocking rates are above the target we defined. As a result adjusted the number of trunks to meet our targets, and re-ran the analysis.

Table 1: 4Call Center



4CallCenters v2.23

File Table Settings Help

Performance Profiler | Staffing Query | Advanced Profiling | Advanced Queries | What-if Analysis

Advanced Queries

With this **Advanced** tool you can set multiple performance goals. Check the **Query** row of any one of your call center's parameters - pressing 'Compute' will find the value(s) of this parameter for which all your goals are met.

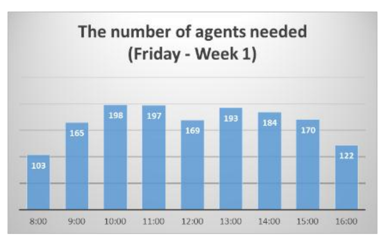
Compute | Add to Table | Delete Rows | Clear All | Export | Import | Graph | Settings

	Basic Interval (minutes)	Target Time to Answer	Number of Agents	Average Handling Time	Calls per Interval	Average Patience	Number of Trunks	Agent's Occupancy	Average Trunks Utilized	%Answer	%Abandon	%Block	Average Speed of Answer	%Answer within Target	Average Queue Length
Goals								✔			✔	✔		✔	
Query			✔												
Input	60	02:00		06:00	1016	06:00	115	95%			3%	2%		80%	
Multi-Value															
Lower															
Upper															
1	60.0	02:00.0	84.0	06:00.0	835.0	06:00.0	99.0	95.1%	82.6	95.7%	3.2%	1.1%	00:11.4	95.7%	2.7
2	60.0	02:00.0	85.0	06:00.0	835.0	06:00.0	99.0	94.5%	82.6	96.2%	2.7%	1.1%	00:09.6	96.2%	2.3
3	60.0	02:00.0	150.0	06:00.0	1,526.0	06:00.0	170.0	97.3%	150.7	95.6%	3.1%	1.3%	00:11.1	95.7%	4.7
4	60.0	02:00.0	157.0	06:00.0	1,526.0	06:00.0	170.0	94.9%	150.7	97.6%	1.1%	1.3%	00:04.1	97.6%	1.7
5	60.0	02:00.0	174.0	06:00.0	1,793.0	06:00.0	195.0	98.0%	176.3	95.1%	3.2%	1.7%	00:11.8	95.1%	5.8
6	60.0	02:00.0	185.0	06:00.0	1,793.0	06:00.0	195.0	94.7%	176.3	97.7%	0.6%	1.7%	00:02.3	97.7%	1.1
7	60.0	02:00.0	192.0	06:00.0	1,974.0	06:00.0	215.0	98.1%	194.6	95.4%	3.2%	1.4%	00:11.5	95.4%	6.3
8	60.0	02:00.0	204.0	06:00.0	1,974.0	06:00.0	215.0	94.8%	194.6	98.0%	0.6%	1.4%	00:02.2	98.0%	1.2
9	60.0	02:00.0	177.0	06:00.0	1,817.0	06:00.0	200.0	97.9%	179.4	95.4%	3.3%	1.3%	00:12.0	95.4%	6.0
10	60.0	02:00.0	188.0	06:00.0	1,817.0	06:00.0	200.0	94.7%	179.4	97.9%	0.8%	1.3%	00:02.7	98.0%	1.4
11	60.0	02:00.0	168.0	06:00.0	1,723.0	06:00.0	189.0	97.8%	169.8	95.3%	3.2%	1.5%	00:11.5	95.4%	5.5
12	60.0	02:00.0	178.0	06:00.0	1,723.0	06:00.0	189.0	94.7%	169.8	97.8%	0.7%	1.5%	00:02.7	97.8%	1.3
13	60.0	02:00.0	187.0	06:00.0	1,929.0	06:00.0	208.0	98.1%	189.4	95.1%	3.1%	1.8%	00:11.3	95.1%	6.0
14	60.0	02:00.0	199.0	06:00.0	1,929.0	06:00.0	208.0	94.7%	189.4	97.7%	0.5%	1.8%	00:01.8	97.7%	1.0
15	60.0	02:00.0	150.0	06:00.0	1,542.0	06:00.0	168.0	97.6%	151.2	95.0%	3.1%	1.9%	00:11.3	95.0%	4.8

Settings
Parameters
Indicators

Ready 3/17/2013 11:39 AM

After we ran the program, we obtained the optimal number of agents needed for each time interval. The software provided both the lower bound and upper bound number. The graphs below show only lower bound number. Please refer to the attachment for additional details.

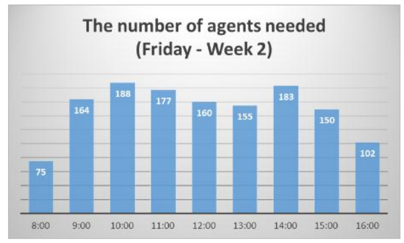



The number of agents needed
(Tuesday - Week 2)
108 177 192 200 179 189 208 199 155
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00


The number of agents needed
(Monday - Week 2)
115 179 235 216 206 195 198 192 156
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00


The number of agents needed
(Wednesday - Week 2)
108 177 192 200 179 189 208 199 105
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00


The number of agents needed
(Thursday - Week 2)
98 168 180 164 154 167 161 168 146
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00

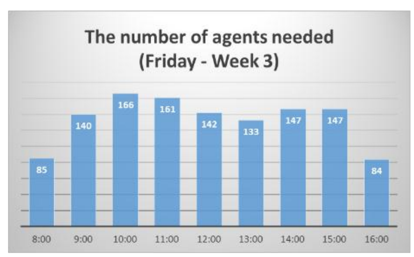

The number of agents needed
(Friday - Week 2)
75 164 188 177 160 155 183 150 102
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00


The number of agents needed
(Monday - Week 3)
94 123 146 158 137 140 127 128 109
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00


The number of agents needed
(Tuesday - Week 3)
123 183 195 198 186 183 165 195 141
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00


The number of agents needed
(Wednesday - Week 3)
118 159 188 202 172 180 204 198 164
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00

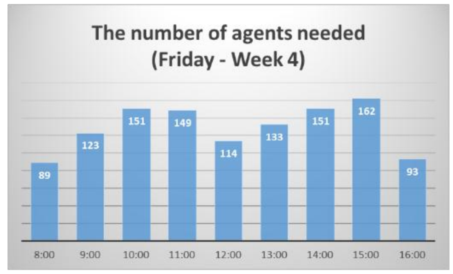

The number of agents needed
(Thursday - Week 3)
95 174 195 155 156 147 167 168 155
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00


The number of agents needed
(Friday - Week 3)
85 140 166 161 142 133 147 147 84
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00


The number of agents needed
(Monday - Week 4)
115 165 179 192 165 189 179 165 134
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00


The number of agents needed
(Tuesday - Week 4)
88 151 179 173 151 157 161 151 121
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00


The number of agents needed
(Wednesday - Week 4)
82 162 165 163 157 175 169 181 123
8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00



5. Staffing

After we obtained the optimal number of agents needed based on our analysis in the respondent time interval, the subsequent task is to how to schedule resources to cover timeslots to meet desired service levels. In fact, the scheduling process includes two activities – shifts and schedules. A schedule is to assign an employee to a set of daily shifts over the course of a week or month. Both shifts and schedules are often constrained by rules such as union rules or legal rules. As a result actual scheduling methodology can be very complex, and they were discussed and researched widely [17-20].

We use the Covering Optimization Model to schedule the call center staff and minimize the cost. In the call center, the main cost is the cost of agents. We need to find suitable number of full time agents and contract-based agents to cover all shifts. Our call center for this report has:

- Nine working hours each day from 8:00am to 5:00 pm
- There are two types of agents – full time employees and part time employees.
- Full time employee can work 9 hours per day with a pay rate of $15 per hour, but can only work 4 days a week. They can take one day off in the week, but the number of employees in day-off cannot be 25% of total full-time employee.
- Part-time employees can work one of three shifts with a pay rate of 10 per hour. Each shift is three hours, and they can start from 8:00am, 9:00am, 10am, 11am, and 12:00pm, 1:00pm, 2:00pm.
- Every 50 employees should have a manager with a pay rate of $30, and the manager can answer calls but should be counted as an agent.

We need to schedule this call center and minimize the cost (staff cost) and meet the service requirement. The mathematical expressions for the above consideration as follow:

Objective function:

Minimize $\sum_{i=1}^{5}\sum_{j=1}^{9}(15 \cdot X_{ij} + 10 \cdot Y_{ij} + 30 \cdot Z_{ij})$

Subject to:

$50Z_{ij} - X_{ij} - Y_{ij} \geq 0 \quad \forall\, i,j$

$A_{ij}M_{ij} + B_{ij}T_{ij} + C_{ij}W_{ij} + D_{ij}R_{ij} + E_{ij}F_{ij} = X_{ij} \quad \forall\, i,j$

$Q - 0.25X_{ij} \leq 0 \; Q \in M_{ij}, T_{ij}, W_{ij}, R_{ij}, F_{ij}$

$$X_{ij} + \sum_{k=1}^{3} Y_{i,j-k+1} \geq L_{ij} \quad \forall\, i,j$$

$X_{ij}, Y_{ij}, Z_{ij}, M_{ij}, T_{ij}, W_{ij}, R_{ij}, F_{ij} \geq 0$

$X_{ij}, Y_{ij}, Z_{ij}, M_{ij}, T_{ij}, W_{ij}, R_{ij}, F_{ij} \in \{Interger\}$

Where:

X_{ij} - The number of full-time employees on day i at hour j

Y_{ij} - The number of part-time employees on day i at hour j

Z_{ij} - The number of managers on day i at hour j

M_{ij} – The number of full-time employees who take Monday off

A_{ij} - The mask coefficient taking Monday off

T_{ij} – The number of full-time employees who take Tuesday off

B_{ij} - The mask coefficient taking Tuesday off

W_{ij} – The number of full-time employees who take Wednesday off

C_{ij} - The mask coefficient taking Wednesday off

R_{ij} – The number of full-time employees who take Thursday off

D_{ij} - The mask coefficient taking Thursday off

F_{ij} – The number of full-time employees who take Friday off

E_{ij} - The mask coefficient taking Friday off

Another business consideration is to have full-time employees meet the minimum number of required agents, which will be determined as:

Subject to:

$X_{ij} \geq \min\{L_{ij}\}$

Where L_{ij} – lower bound of required agent

The diagram below is the screen shot of the optimal schedule. Please refer to the attachment for the schedule file.

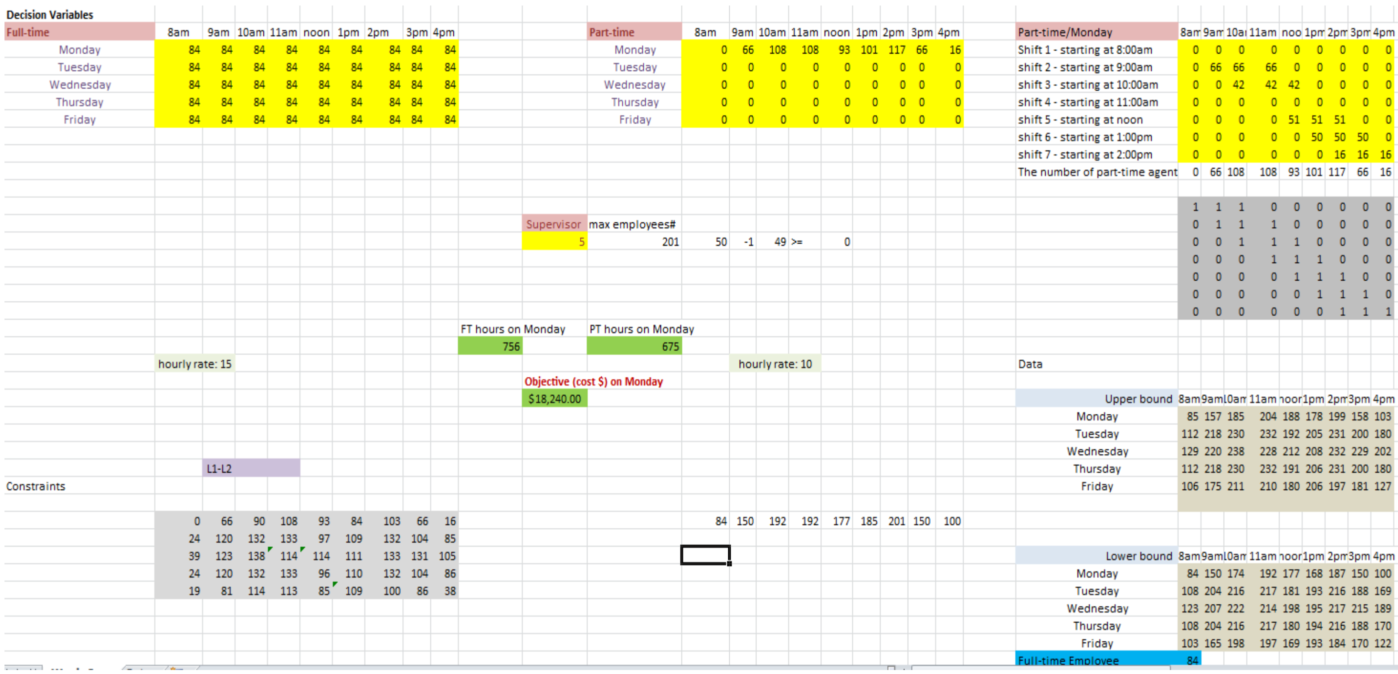


Figure 3 : Excel Model

6. Conclusion

In this paper, we successfully analyzed a typical inbound call center utilizing the Erlang A model to determine the optimal number of trunks and staffing. Our analysis takes into consideration the typical call center environment and may require some adjustments to address issues such as seasonal traffic demands, and call centers' that have calls that are routed based on product or task specific demands. Based upon the number of agents, we use linear programing and covering model to schedule the call center shifts to minimize the call center cost. To address telecom trunking demand we based our analysis on the industry standard Erlang A model that takes into consideration both call blocking and abandonment rates.

Notably, due to the relatively large number of call arrivals even small changes in nearly any parameter such as targeted blocking or call completion rates may have a significant impact on the resulting number of agents or trunks required to meet business goals. As with any optimization analysis, a post-implementation analysis should be conducted to determine if anticipated targets are meet and if not, adjustments may be required to ensure the most cost efficient solution. In addition, telecom suppliers often modify services pricing and as such, maintaining an optimal configuration may require adjustments to the analysis to reflect price changes.

7. Limitation and future researches

In the project, we assume all agents either part time or full time have the same skill levels. In fact, full time agents should have higher skill levels than do part time agents because full time agents should have much more training than do part time agents. The ACD routing should base upon the callers requirement to route the call to the right agents.

We also have no enough time to do deep research into the call forecasting. The call forecasting is very important for call center planning and staff scheduling. Hiring more than needed agents or purchased more than needed trunked waste money; not enough agents may cause callers to abandon calls and lose business; not enough trunk lines may increase the blocking rate.

We should analyze all eight month data to estimate the traffic trend, but we just didn't have enough time to do so.

We also would like to do sensitivity analysis such as making a small change in the arrival rate, service rate, and the performance will be impacted.

The data we obtained were based upon the hourly interval, so we are not able to do 15 minutes and 30 minutes data analysis.

In the future, we should base upon the model and apply it to an actual call center, and see it correctly meet the actual trend and service level.

8. References

1. Raik Stolletz, "Performance Analysis and Optimization of Inbound Call Centers", New York: Springer-Verlag, 2003, pp 1-200
2. Cliff T. Ragsdale, "Spreadsheet Modeling & Decision Analysis 6e**"**, Mason, OH:South-Western Cengage Learning, 2012, pp 653-675
3. Gans N., Koole G., and Mandelbaum A. "Telephone Call Centers: Tutorial, Review and Research Prospects", Invited review paper by *Manufacturing and Service Operations Management* (M&SOM), 5 (2), pp. 79–141, 2003.
4. Mandelbaum A., Sakov A., and Zeltyn S., "Empirical Analysis of a Call Center", Technical Report, 2000
5. Garnett O., Mandelbaum A., and Reiman M., "Designing a Call Center with Impatient Customers", *Manufacturing and Service Operations Management*, **4(3)**, 208-227, 2002
6. Borst S., Mandelbaum A., and Reiman M, "Dimensioning Large Call Centers", *Operations Research*, 52(1), pp. 17-34, 2004
7. Mandelbaum A. and Zeltyn S., "Service Engineering in Action: The Palm/Erlang-A Queue, with Applications to Call Centers", *(2007) Advances in Services Innovations*, pp. 17-48, Spath D., Fähnrich, K.-P. (Eds.), Springer-Verlag
8. Aldor S. and Feigin P.D, "Modeling Arrivals to the Agents Queue in a Call Center*", Technical Report*, September 2007
9. Armony M, and Mandelbaum A, "Routing and Staffing in Large-Scale Service Systems: The Case of Homogeneous Impatient Customers and Heterogeneous Servers". *Operations Research* Vol. 59, No. 1, January–February 2011
10. Brown L., Gans N., Mandelbaum A., Sakov A., Shen H., Zeltyn S. and Zhao L., "Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective", (2005) Journal of the American Statistical Association, Vol 100: 36-50
11. Mandelbaum A., Zeltyn S, "The Palm/Erlang-A Queue, with Applications to Call Centers", Internet:ie.technion.ac.il/serveng/References/Erlang_A.pdf, Feb 4, 2013 [Feb 10, 2013]
12. Mandelbaum A., Zeltyn S, "M/M/n+G Queue: Summary of Performance Measures", Internet:ie.technion.ac.il/serveng/References/MMNG_formulae.pdf, Feb 4, 2013 [Feb 10, 2013]
13. Zeltyn S., "Dimensioning Call Centers with Abandonment: Constraint Satisfaction and Cost Minimization", *ORSIS annual meeting*, Naharia, May 2006
14. Rodney B. Wallace and Ward Whitt, "A Staffing Algorithm for Call Centers with Skill-Based Routing", *MSOM No,7, pp. 276-294, Fall 2005*
15. Pierre L'Ecuyer, "Modeling and Optimization Problems in Contact Centers", pp. 145 -156, QEST 2006
16. L. Brown, et all, "Statistical Analysis of a Telephone Call Center", *Journal of the American Statistical Association***,** Vol 100, No. 469, pp. 36-50, 2005
17. Aldor-Noiman S., Feigin P.D., and Mandelbaum A., "Workload Forecasting for a Call Center: Methodology and a Case Study*", The Annals of Applied Statistics*, vol. 3, No.4, pp. 1403-1447, 2009
18. Armony M. and Mandelbaum A, "Design, staffing and control of large service systems: The case of a single customer class and multiple server types", Draft, March 2004
19. Mehmet Tolga Cezik, and Pierre L'Ecuyer, "Staffing Multiskill Call Centers via Linear Programming and Simulation", *Management Science*, Vol. 54, No. 2 310-323, Feb. 2008

20. Willie B. Henderson and Willian L. Berry, "Heuristic Methods for Telephone Operator Shift Scheduling: An Experimental Analysis", *Management Science*, Vol 22, No. 12, pp. 1372-1380, 1976
21. 4CallCenters Software, Available at http://ie.technion.ac.il/serveng/4CallCenters/Downloads.htm

Appendix 1

The output zip file including all csv output files

Appendix 2

The excel sheet file