Malaysian Journal of Science **26** (SKSM14 Special Issue): 85 - 93 (2007)

Performance Modelling and Evaluation of a Telecommunication Contact Centre: A Queuing Theory Approach

Balambigai Balakrishnan[1] and **Susila Munisamy**[2]

Department of Applied Statistics, Faculty of Economics and Administration, University of Malaya, 50603 Kuala Lumpur, Malaysia.
[1]balambigaip@yahoo.com (corresponding author), [2]susila@um.edu.my

ABSTRACT A call centre is a centralised office of a company that mainly handles incoming telephone calls from customers via telephone. The call centre basically functions as a primary contact point between customers and their service providers. Call centres are highly technology driven. However, surprisingly, most of the costs incurred in a call centre are due to human resources. Customer service agents who handle the calls form most of the human resource component in a contact centre. An important goal of the call centre is to provide a good level of customer service. A good customer service level will ensure customer satisfaction so that customers will return. The consequence of making customers wait too long may be lost profit from lost business opportunities. In this context, queuing models are important to determine the appropriate number of customer service agents that strike a balance between the two conflicting objectives of cost reduction and provision of good service. In this research, we have attempted to build a queuing model to evaluate the performance of a call centre that belongs to one of Malaysia's leading telecommunication service providers and to allocate its human resources. The telephone calls coming in to the call centre are the customers in large queuing system, and the customer service agents are the servers. The telephone calls that arrived on a Monday during the peak period fitted the assumptions underlying the Erlang C or the M/M/s model in terms of arrival pattern and service time. The model built then was used to analyse the call centres' operating characteristics and to decide various numbers of staffs required to achieve different management objectives.

ABSTRAK Sebuah pusat perhubungan adalah suatu jabatan pusat yang terutamanya mengendalikan panggilan telefon yang diterima dari pelanggan. Ia biasanya berfungsi sebagai satu titik hubungan primer di antara pelanggan dan pembekal perkhidmatan. Sebuah pusat perhubungan beroperasi dengan menggunakan teknologi tinggi. Akan tetapi, kebanyakan kos operasi sebuah pusat perhubungan adalah berasaskan tenaga manusia. Bilangan ejen perkhidmatan pelanggan yang melayani panggilan yang diterima merupakan sumber utama komponen kos sebuah pusat perhubungan. Perkhidmatan pelanggan yang baik dapat memastikan kepuasan pelanggan supaya pelanggan kembali untuk mendapatkan perkhidmatan. Jika pelanggan menunggu terlalu lama, syarikat boleh kehilangan keuntungan yang disebabkan oleh kehilangan peluang perniagaan. Dalam keadaan ini, pendekatan teori giliran adalah sangat penting untuk menentukan bilangan pekerja ejen perkhidmatan pelanggan yang sesuai, yang dapat mengimbangkan pencapaian dua objektif yang bertentangan iaitu: penjimatan kos dan perkhidmatan yang terbaik. Dalam kajian ini kita berusaha untuk membina sebuah model yang dapat menilai prestasi operasi pusat perhubungan salah satu syarikat telekomunikasi yang terbesar di Malaysia. Panggilan telefon yang memasuki pusat perhubungan merupakan pelanggan yang menunggu giliran, dan ejen perkhidmatan pelanggan adalah pelayan. Panggilan yang diterima pada suatu hari Isnin untuk jangka masa sejam sewaktu masa sibuk menepati andaian Erlang C ataupun model M/M/S dari segi corak ketibaan dan masa layanan. Model yang dibentuk, seterusnya, digunakan untuk menganalisa ciri-ciri operasi pusat perhubungan yang dikaji dan menentukan bilangan ejen perkhidmatan yang diperlukan untuk mencapai objektif pengurusan yang berbeza.

(Contact centre, queuing theory, Erlang C, M/M/s)

INTRODUCTION

A call centre (also known as a contact centre), is the most important way by which an organisation interacts with its customer[1]. Contact centres provide primarily tele-services, where they answer incoming calls from customers via telephone [1]. Contact centres are established in a diverse range of businesses, in large, small and medium organisations. Mail order catalogue firms, utility companies, banks, departmental stores, insurance companies, airlines, emergency road service operators and many others get connected to their customers via contact centres. Telecommunication service providers also use contact centres to answer customer enquiries regarding phone service billings, reporting of faulty services, ordering of new features and benefits, changing customer profile (including address) and cancellation of services.

In a typical commercial contact centre, about 65 percent of costs are due to staffing, 25 percent of costs are for networking and communication, and the remaining 10 percent of costs are associated with maintenance and other overheads [2]. Thus, the major component of contact centre costs is staffing [3]. In addition to having an inaccurate number of agents, the costs can also be incurred by the high contact centre agents' turnover rate. Poor contact centre management increases the stress level and leads to drastic resignation of jobs among employees [4]. As a result, the contact centre faces increased expenses caused by ongoing training programmes to replace employees. The cost concern resulting from staffing can be tackled via the quantitative models which is generally analytical [5] and at times empirical [6]. The stochastic model especially the queuing model has been the standard model used to investigate the quantitative aspect of the performance of a contact centre. The 80/20 rule is an example of quantitative based service quality in the contact centre industry. This indicates that at least 80 percent of the customers must not wait more than 20 seconds in the telephone queue [5]. Such targets are generally set by upper management and the contact centre managers are called to defend their budget which will enable them to achieve the target set. Thus this study aims to evaluate and improve the performance of the contact centre under study using the M/M/s model.

The layout of this paper is as follows. The next section provides information on the background of the contact centre of the telecommunication service provider in this study. This is followed by a literature review on the use of queuing theory in studies on call centres. The section following that explains the data collected and the methodology of the study, and this is then followed by a section that describes the performance model used and discusses the results of the model. The final section concludes the paper and offers some recommendations.

CONTACT CENTRE OF THE TELECOMMUNICATION SERVICE PROVIDER

The contact centre in this study belongs to one of Malaysia's leading telecommunication service providers. The contact centre handles inbound and outbound calls, faxes, e-services and mail, and operates 7 days a week, 24 hours a day. The incoming calls make up more than two–thirds of total calls handled at the contact centre. Figure 1 presents the flow of an incoming call through the system.

An incoming call generally goes through three stages: Interactive Voice Response (IVR), queue and service, although some calls skip the queuing stage and go directly to the service stage. When a customer dials a given number they are instantly connected to an Interactive Voice Response (IVR) (also called Voice Response Unit or VRU). The IVR also enables customers to complete some self-service transactions (for example, for balance enquiries, customers may be told to "press five"). After completing the self-service transaction a customer may end the call at this point. If the customer/caller opts to speak to an agent or a Customer Service Representative (CSR), he presses the specified number on the telephone keypad and either gets connected immediately to an agent or joins the tele-queue waiting for an agent to become available. When a call exits from an IVR to join a queue it is recorded as calls arrived, entered or received. Impatient callers may not wait and leave the

[1] Initially set up to answer calls, these centres have grown to provide many other services such as responding to faxes, email and e-services, and are now known as contact centres. In this study both the terms are used interchangeably.

queue abruptly. These calls are categorised as abandoned calls. The calls that reach the agent and receive the service are referred to as calls answered or served. Customers in a tele-queue are normally served on a first come first-served (FCFS) basis. However, some callers classified as priority customers by the contact centre by-pass this system and get connected immediately to the next available agent. For example, a customer who spends more than RM500 a month is classified as a priority customer at the contact centre under study.

REVIEW OF LITERATURE

Waiting in line is almost an everyday practice. Customers wait to receive services. However, making customers, employees, jobs or even telephone calls wait very long in a queue can have serious consequences. Thus, performance of waiting line or queue[2] is an important concern to many organisations. Queuing theory, which was conceived by A.K. Erlang in 1917 (see [7], [8]), is the theory that deals with performance of waiting line [9]. This theory uses queuing models which consist of mathematical formulas and relationship to represent the various types of queuing system. The queuing models are used to study the performance indicators of a waiting line. Performance of waiting line is normally measured in terms of average number of customers waiting in queue and the system (which includes customers in the queue and being served), the average time spent in the queue and the system, the percentage of time the servers are busy (utilisation rate) and others [9,10,11]. The main objective of a queuing model is to determine how much service capacity should be provided to a queue to avoid excessive waiting, which in turn will contribute to economic gain [10,11]. In other words, in designing queuing models, organisations aim to achieve a balance between offering quality service to customers (short queues requiring many servers) and economic considerations (not too many servers).

Queuing is a common feature in a wide range of fields covering from many daily-life situations to more technical environments such as computer networks [12] and telecommunications systems [13]. Much of the early work was motivated by practical problems concerning telephone traffic. At later stages, queuing has been extensively applied to real problems arising in manufacturing [14] public transportation and service operations [15]. The queuing literature has grown, looking for theoretic and algorithmic tools, and mathematical models of queuing phenomena. As a result of this, despite the potential applicability of queuing theory, the gap between theoretical developments and real applications has also grown.

Research on contact centres and their performance has gained tremendous popularity in the past years. More and more contact centres are analysing their performance based on sound scientific principles. Queuing theory has become one of the central research themes of operations research [6,14] that studies contact centre performance.

One such study that contributed empirically and analytically in the area of contact centre is based on the full call-detail record of 450,000 telephone calls of a small call centre of a bank over a twelve month period [16]. This analysis was guided by queuing theory and is divided into two parts. In the first section, each of the parameters namely; the arrival process, the service time-distribution and the distribution of customer patience were analysed using different statistical techniques. This analysis prompted development of new statistical methods and approaches, especially the use of non-parametric techniques to test the service time distributions, in the second section.

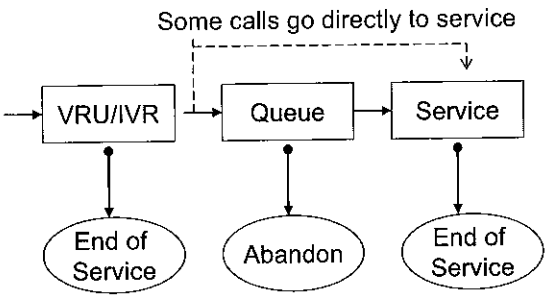

Figure 1. Flow of an incoming call

To sum up, although contact centre offers ample opportunities for research, it is astonishing that very little information is available, especially based on analytical model and validated by real data. The contact centre data from a small bank in

[2] The management scientist refers to waiting line as queue [11]. In this study we use both the term interchangeably.

Israel is one of the most highly used data in modeling exercises based on queuing theory. This data is made available at the website http://iew3.technion.ac.il/serveng/callcentredata/index.html [17]. Availability of data could have been one of the major reasons for the lack of practical studies involving queuing theory generally and contact centre specifically.

METHODOLOGY

Primary data was collected detailing call by call history for a period of one week (from 2 February 2005 till 7 February 2005). This data was used to explore the underlying patterns of incoming calls to the call centre. Having done that, the data for a period of one hour, between 11 am to 12 pm on a Monday (peak hours) was selected for the purpose of performance evaluation.

Queuing theory, which is a branch of mathematics, is used to evaluate the performance of waiting line or queue for the contact centre under study. Queuing models can be built based on different parameters of a queuing system. The main three parameters explored in building a queuing model are the arrival distribution, the service time distribution and number of servers. The queuing models are used to describe the behavior of the queuing system (for example, in terms of average number of customers in the waiting line, average number of customers in the system, average time each customer spend in the waiting line etc, to determine the level of service to provide and to evaluate alternative configurations (including number of servers) to providing service [18].

The M/M/s model is a multiple server model that assumes Poisson arrival rate and exponential service time [7, 9]. This model, which is also known as Erlang C is the most commonly used model in workforce management of contact centres [17]. This model was also used in this study. The performance indicators for this model are presented in Appendix 1. The performance models were implemented using Queuing ToolPak 4.0 which is a Microsoft Excel add-in tool [18].

THE PERFORMANCE MODEL

The queuing model used to analyse the performance of the call centre is M/M/60/FCFS/∞/∞ or commonly known as the M/M/s or Erlang C model. The data fitted the assumption of Poisson arrivals and exponentially distributed service time required by this model. We found that the contact centre received a total of 954 calls in an hour which is at a rate of 15.9 calls per minute, had a service rate of 16.2 calls per hour (for every single server) and a total of 60 servers. The queue capacity and the population were found to be infinite and the callers were treated on a first come first serve (FCFS) basis. In addition to this, the model also satisfied the condition that the mean service rate, μ, is greater than the mean arrival rate, λ. In other words, the utilisation factor must be less than 1. This is required for the queuing system to be stable, that is, to prevent the queue from growing indefinitely. The utilisation factor calculates the fraction of time each server is busy in a queuing system (the symbol rho = ρ is used to represent this). The value of ρ at the call centre when 60 servers are used is 0.981. The other performance indicators for a range of number of servers are presented in Table 1.

Since the main objective of this study is to equip the contact centre under study with the right number of agents or servers, we first attempt to look at the effect of the different number of servers on the performance indicators. The changes in the performance indicators were measured for a range of servers beginning from 59 to 85. When the contact centre uses 59 servers, the utilisation factor is found to be 1 which is an indication that the contact centre will be operating at full capacity. However, when the total servers used are below 59, many of the performance indicators were not available due to the fact the system becomes unstable and the queue grows indefinitely.

Firstly, the relationship between number of servers and the utilisation factor is tested. Negative linear relationship is found between the utilisation factor and the number of servers (as indicated by the formulae for utilisation factor). As the number of servers increase, the utilisation factor decreases steadily. This indicates that by keeping the arrival rate and service rate constant, a higher number of servers reduces the workload and relaxes the contact centre environment.

Table 1 :Performance indicators for ATSP with λ = 954 calls an hour and μ = 16.2 calls per hour and varying number of servers

Servers	59	60	61	62	63	64	65	66	67
Utilization factor (ρ)	1.00	0.98	0.97	0.95	0.93	0.92	0.91	0.89	0.88
Probability of no calls in the system (P_0)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Average number of calls in the queue (L_q)	520.76	44.29	19.69	11.21	7.06	4.70	3.22	2.25	1.59
Average number of calls in the system (L)	579.65	103.18	78.58	70.10	65.95	63.58	62.11	61.14	60.48
Average time in the queue (W_q) (seconds)	1965.14	167.14	74.32	42.30	26.66	17.72	12.15	8.49	6.00
Average time in the system (W) (seconds)	2187.36	389.36	296.54	264.53	248.88	239.94	234.37	230.71	228.23
Probability that an arriving call has to wait to be served (P_w)	0.98	0.84	0.71	0.59	0.49	0.41	0.33	0.27	0.22
Percentage of calls answered within 20 seconds	3	24	42	55	66	74	81	86	89
Servers	**68**	**69**	**70**	**71**	**72**	**73**	**74**	**75**	**76**
Utilization factor (ρ)	0.87	0.85	0.84	0.83	0.82	0.81	0.80	0.79	0.77
Probability of no calls in the system (P_0)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Average number of calls in the queue (L_q)	1.13	0.81	0.58	0.41	0.29	0.21	0.15	0.10	0.07
Average number of calls in the system (L)	60.02	59.70	59.47	59.30	59.18	59.10	59.04	58.99	58.96
Average time in the queue (W_q) (seconds)	4.27	3.05	2.18	1.55	1.11	0.78	0.55	0.39	0.27
Average time in the system (W) (seconds)	226.49	225.27	224.40	223.78	223.33	223.01	222.78	222.61	222.49
Probability that an arriving call has to wait to be served (P_w)	0.18	0.14	0.11	0.08	0.07	0.05	0.04	0.03	0.02
Percentage of calls answered within 20 seconds	92	94	96	97	98	99	99	99	100
Servers	**77**	**78**	**79**	**80**	**81**	**82**	**83**	**84**	**85**
Utilization factor (ρ)	0.76	0.75	0.75	0.74	0.73	0.72	0.71	0.70	0.69
Probability of no calls in the system (P_0)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Average number of calls in the queue (L_q)	0.05	0.03	0.02	0.02	0.01	0.01	0.00	0.00	0.00
Average number of calls in the system (L)	58.94	58.92	58.91	58.90	58.90	58.90	58.89	58.89	58.89
Average time in the queue (W_q) (seconds)	0.19	0.13	0.09	0.06	0.04	0.03	0.02	0.01	0.01
Average time in the system (W) (seconds)	222.41	222.35	222.31	222.28	222.26	222.25	222.24	222.23	222.23
Probability that an arriving call has to wait to be served (P_w)	0.02	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00
Percentage of calls answered within 20 seconds	100	100	100	100	100	100	100	100	100

Next, the relationship between the utilisation factor and average time in queue or queue in seconds is tested. Up until 80 percent of utilisation of the capacity, the waiting time for a call remains close to 0 seconds. As soon as it goes beyond 80 percent, the amount of time each call spends in the queue grows rather quickly. This can be solved either by increasing the number of servers or reducing the time spent on each call or by controlling the number of calls arriving at the contact centre. Next, the relationship between the number of servers and the probability of an arriving call having to wait to be served is found to be negative exponential in nature. When there are 80 servers, each call arriving at the contact centre has 0 probability of having to wait to be served, in other words, absolutely no calls has to wait when there are 81 servers waiting to serve the incoming calls. In a nutshell, the M/M/s model indicates that by holding the arrival rate and service rate constant, different number of servers can leave different impact on the service provided. At the beginning, as the number of servers increase the performance indicators improved tremendously. However, after reaching the optimal point (in this case approximately at 65 servers), the impact of the additional servers reduces tremendously.

Service rate is influenced by the service time, which is the amount of time spent on each call and the server utilisation rate. Service time tend to vary depending on the complexity of the call received. Thus, it is important to look at number of servers required for various service times. Using Erlang C or the M/M/s model developed, the number of servers required was calculated for various service time ranging from 120 seconds to 360 seconds. It was found that the relationship between the number of servers required and average service time is linear for all the three scenarios. However, there are more servers required for the same amount of service time when the agent utilisation rate is lower. For example, a call that takes 6 minutes to answer requires 136 servers for 70 percent agent utilisation, 119 for 80 percent agent utilisation and 106 servers for 90 percent agent utilisation. The difference between 70 percent and 90 percent agent utilisation is 30 servers per hour. This indicates the vulnerability of the contact centre and the tremendous effect of different number of servers on the smooth running of a contact centre on an hourly basis.

As we have seen, the number of servers, service rate (and time) and arrival rate are very important parameters in the M/M/s model. With the current 60 servers, the contact centre is operating at its full capacity leaving a lot of room for improvement in the level of service provided to customers.

CONCLUSION AND RECOMMENDATION

As said earlier, one of the performance objectives of the contact centre under study is to answer 80 percent of the incoming calls within 20 seconds. Following Table 1, it was found that placing 64 servers to answer calls reduces average time in the queue to 17.72 seconds. However, Table 1 indicates only 74 percent of the calls will be answered within 17.72 seconds. The 80 percent target is achieved only when 65 servers are on duty, at this stage the average time in the queue also has been reduced to 12.15 seconds. In addition, Table 1 indicates that the utilisation rate is rather high at 91 percent when 65 servers are placed to answer calls. The high utilisation factor could result in low quality of calls answered. Calls may not be answered in full or cut short due to pressure faced by the servers. Thus, to provide a level of service that achieves both 80 percent utilisation factor and 80 percent calls answered in 20 seconds, the most appropriate number of servers would be 74 (the average time in the queue will be 0.6 seconds only). In short, the number of servers may change according to the choice of objectives to be achieved by the contact centre under study. In summary, the insight gained from Table 1 is very useful for improving the level of service and efficiency of the contact centre being studied. Next, Table 2 summarises various objectives and the number of servers required to achieve the objectives.

The first objective only takes the customer welfare into consideration. The agent utilisation rate is found to be rather high when the 80/20 target is achieved. This will cause employee burn out and lead to higher staff turnover rate. Thus, the second objective, takes both the customer and employee welfare into consideration. This objective is highly recommended to the contact centre when the performance of the contact centre is modeled based on Erlang C. However, employing 74 servers per hour instead of the current 60 will cost more. Thus, the additional servers can be made available during the peak

hour only, for example, beginning from 10 am to 2 pm. In short, the right number of servers to be employed entirely depends on the objective of the company. However, it is important to note that these recommendations are purely developed based on the data collected between 11 am to 12 pm on Monday, 7 February 2005.

In addition, the contact centre under study can attempt to increase the mean service rate which is currently at 16.1 customers per hour. This can be done by making a creative design change with the application of technology. More questions can be included in the IVR to enable the servers and the callers to have a clear idea of the purpose of the call. This enables the servers to focus on the problem solving rather than spending time on understanding the problem. The contact centre also can attempt to reduce the number of arrivals or more importantly distribute the call arrivals fairly throughout the day. Off-peak hour promotions can be carried out to encourage the callers to call after office hours. An important point to note is that the cost of promotional efforts must be less than employing new customer service representatives and should be able to improve customer satisfaction. More customers can be encouraged to use the IVR, thus reducing the burden on the customer service representatives fully. Last but not least, currently, there are ample data available in the contact centre. These data needs to be managed to realise its full benefits. Thus, it is important to employ a team of analyst, with contact centre and management science knowledge to reap the benefit from this data.

Table 2. Number of servers required to achieve various objectives by Erlang C

Objective	Number of servers
Time in queue less than 20 seconds	64
80/20 – at least 80 percent of the calls are answered within 20 seconds	65
80 percent utilization rate	73
80/20 rule and 80 percent utilization rate	74

λ = 15.9 calls per minute, μ = 16.2 calls per hour

APPENDIX 1

M/M/S Model or Erlang C

λ = the mean number of arrivals per time period (the mean arrival rate)
μ = the mean number of services per time period (the mean service rate)
s = number of servers
ρ = utilization factor
P_0 = The probability that no units are in the system
L_q = The average number of units in the waiting line
L = The average number of units in the system
W_q = The average time a unit spends in the waiting line
W = The average time a unit spends in the system
P_w = The probability that an arriving unit has to wait for service
P_n = The probability of n units in the system

Performance Indicators

$$\rho = \frac{\lambda}{s\mu}$$

$$P_0 = \frac{1}{\sum_{n=0}^{s-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!}\left(\frac{s\mu}{s\mu-\lambda}\right)}$$

$$L_q = \frac{(\lambda/\mu)^s \lambda\mu}{(s-1)!(s\mu-\lambda)^2} P_0$$

$$L = L_q + \frac{\lambda}{\mu}$$

$$W_q = \frac{L_q}{\lambda}$$

$$P_n = \frac{(\lambda/\mu)^n}{n!} P_0 \quad \text{for n} \leq \text{s}$$

$$P_n = \frac{(\lambda/\mu)^n}{s!\,s^{(n-k)}} P_0 \quad \text{for n} > \text{s}$$

REFERENCES

1. Mandelbaum, A. Sakov, A and Zeltyn, S. (2001). Empirical analysis of call center. Technical Report. (Available: Online). www.ie.technion.ac.il/serveng/course/096324
2. Antipov, A and Meade, N. (2002). Forecasting call frequency at a financial services call center. *The Journal of Operational Research Society* **53** (9): 953-960.
3. Gans, N., Koole, G and Mandelbaum, A. (2003). Telephone call centers: a tutorial and literature review. Invited review paper. *Manufacturing and Service Operations Management* **5**(2): 79 – 141. (Available: Online). http://iew3.technion.ac.il/serveng/References/references.html
4. Tuten, L. and Neidermeyer, E. (2004). Performance, satisfaction and turnover in call centers: The effects of stress and optimism. *Journal of Business Research* **57**(1): 26-34.
5. Koole, G and Mandelbaum, A. (2002). Queuing models of call centers, an introduction. *Annals of Operations Research* **113**: 41–59.
6. Whitt, W. (2002). *Stochastic-Processes Limits*. Springer. New York.
7. Erlang, A.K. (1911). The Theory of Probability and Telephone Conversations.
8. Erlang, A.K (1917). Solutions of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges. *Electroteknikeren* **13**: 5–13 [in Danish]. *Nyt Tidsskrift Mat. B*, **20**: 33–39.
9. Anderson, D., Sweeney, D and Williams, T. (2003). *An Introduction to Management Science Quantitative Approaches to Decision Making*. Tenth Edition. Thomson, Australia.
10. Reid, D and Sanders, R. (2004). *Operations Management: An Integrated Approach*. John Wiley and Sons, New York.
11. Hillier, F and Hillier, M. (2004). *Introduction to Management Science: A Modeling and Case Study Approach with spreadsheet*. 2nd edition. McGraw-Hill.
12. Kleinrock, L. (1976). *Queuing Systems: Computer Applications*. Vol. 2, John Wiley & Sons, New York.
13. Ross, K. W. (1995). *Multiservice Loss Models for Broadband Telecommunication Networks*. Springer-Verlag, London.

14. Buzacott, J.A. and Shanthikumar, J.G. (1993). *Stochastic Models of Manufacturing Systems*, Prentice Hall.
15. Hall, R.W. (1991). *Queuing Methods: For Services and Manufacturing*, Prentice Hall.
16. Brown, L. D., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., and Zhao, L. (2002). Statistical Analysis of a Telephone Call Center: A Queuing Science Perspective. Technical report, University of Pennsylvania.
(Available online).
http://iew3.technion.ac.il/serveng/References/references.html
http://iew3.technion.ac.il/serveng/callcenterdata/index.html (available online).
17. Ragsdale, C. (2001). *Spreadsheet Modeling and Decision Analysis: A Practical Introduction to Management Science.* 3rd Edition. South-Western College Publishing.
http://www.bus.ualberta.ca/aingolfsson/qtp (Available online).
18. Mandelbaum, A and Zeltyn, S. (2005). *Service Engineering in Action: The Palm/Erlang-A Queue, with Applications to Call Centers.* Draft, March 2005. (Available online).
http://iew3.technion.ac.il/serveng/References/references.html