

SIMULATION OF A STOCHASTIC MODEL FOR A SERVICE SYSTEM

Catherine Brickner
Dennis Indrawan
Derrick Williams
Srinivas R. Chakravarthy

Department of Industrial and Manufacturing Engineering
Kettering University,
Flint, MI 48504, USA

ABSTRACT

In this paper we simulate a queueing model useful in a service system with the help of ARENA simulation software. The service calls (henceforth referred to as customers) arrive to a processing center according to a Markovian arrival process (MAP). There is a buffer of finite size to hold the customers. Any customer finding the buffer is considered lost. An arriving customer belongs to one of three types, and the admitted customer is served by one of many dedicated servers (exclusively set aside for each of the three types of customers) or by one of many flexible servers who are capable of servicing all types of customers. The flexible servers are used only when the respective dedicated servers are all busy. A priority scheme is used to select the type of customer from the buffer when a flexible server is called for servicing the waiting customers. The processing times are assumed to be of phase type. Simulated results are discussed.

1. INTRODUCTION

Service systems play an important role in day-to-day situations in practice. Companies selling products are realizing that a customer's choice in a product is not only influenced by the value the product is offering but also the support provided by the company after the sale. The support crew of a company involves specialists (or dedicated) who have expertise in serving one type of customers and generalists (or flexible) who have expertise to serve more than one type of customers. While flexible servers provide much needed flexibility to the company, they also cost more to the company as the training needed for such servers is much more intensive as compared to dedicated servers.

Recently, Chakravarthy and Agnihotri (2005) studied a service system with two types of customers with their own dedicated servers as well as (common) flexible servers. Assuming Poisson arrivals and exponential service times whose parameters may depend on the type of customers, they developed guidelines for managers to decide when to have total flexibility, total specialization, and a combination of these two using matrix-analytic methods. We refer the reader to Chakravarthy and Agnihotri (2005) for a literature survey on strategic use of flexibility in manufacturing as well as in service industries.

In this paper we consider a service system in which three types of customers arrive according to a Markovian arrival process (MAP), a versatile point process useful in modeling practical systems. A brief description of this process is given in the next section. There are four types of servers: three groups of dedicated servers and one group of flexible servers, who offer services to admitted customers. The service times are assumed to be of phase type (PH-type) that may possibly depend on the type of services as well

as on the type of servers (dedicated or flexible). A brief introduction to PH distributions is given in the next section.

The model presented here is motivated by an example in electronics stores such as Best buy. Best buy offers electronic sales, repairs, and returns for customers throughout the world. There are four groups of associates to address the three services offered by Best Buy. The first group, sales associates, has responsibilities that relate to selling products and services to new and existing customers who arrive to the store. They are not qualified to perform repairs or handle returns. They are usually not highly trained on electronics, but only to explain the features of the products. This is the reason why the median pay for retail sales associates is relatively low at \$9.49/hr according to [payscale.com](http://www.payscale.com/research/US/Employer=Best_Buy/Hourly_Rate) (<http://www.payscale.com/research/US/Employer=Best_Buy/Hourly_Rate> as of 9/7/2009).

The second group of servers corresponds to repair associates. These associates have the responsibility over handling of repairs and maintenance of electronics brought into the store. It is because of their experience and pay that they are dedicated to doing repair and maintenance work on electronics that come into the store. Repair associates are usually trained more than other associates because they need to know how electronics work along with the features of the products. This is the reason why the median pay for repair associates is generally much higher than sales associates. In some cases their pay is even higher than the sales associate supervisors. The median pay for repair associates is between \$12.40 and \$16.91 per hour according to [payscale.com](http://www.payscale.com) as of September 2009.

The third group consists of customer service associates. Generally these associates have the responsibility of handling product returns and replacements. These associates require little training such as the sales associates. They are of higher value to the company as they help to retain customers, keep the cost of fraud return low, and determine the liability cost of defective inventory from manufacturers. Their median pay is sometimes higher than that of sales associates. The median pay of customer service associates is \$9.67/hr according to [payscale.com](http://www.payscale.com) as of September 2009.

The final group of associates is the floaters who often start as particular type of associates and transfer to various types of associate groups. Since these associates have been trained in various types of jobs they can easily be utilized to handle responsibilities of various associate groups. Often these associates are supervisors and work with all groups and are therefore called flexible servers. It is because of their extensive experience and qualifications gained throughout the organization that their wages are higher. The rate is \$19.00/hr on the average.

One can easily find similar examples in practice for which the current model can be applied. The paper is organized as follows. In Section 2, we describe the mathematical model along with a brief description of the MAP. The simulation aspects of this model are given in Section 3, and illustrative simulated numerical examples are presented in Section 4.

2. MODEL DESCRIPTION

Customers arrive to a service system according to a MAP, a versatile point process introduced by Neuts (1979). In practice, we come across many situations where the arrival processes are not necessarily renewal processes. For example, consider a queueing network consisting of, say, 2 nodes. The output of node 1 will form the input to node 2. If we consider non-Poisson arrivals, say, to the first node, the output process will not necessarily be a renewal process. Thus, MAP is convenient as well as proper processes to model under these conditions. A MAP is constructed as follows.

Consider an irreducible continuous-time Markov chain (CTMC) with m transient states. At the end of a sojourn in state i , that is exponentially distributed with parameter λ_i , there are two possibilities. The first possibility corresponds to an "event" (or an arrival) and the CTMC can visit any of the m transient states including the state from which this event occurred with probability p_{ij} . The second possibility corresponds to no arrival and the CTMC can visit any of the $m-1$ transient states (all m except i) with probability q_{ij} .

Thus, the CTMC can go from state i to state i only through an arrival. Define matrices $D_0 = (d_{ij}^{(0)})$ and $D_1 = (d_{ij}^{(1)})$ such that $d_{ii}^{(0)} = -\lambda_i, 1 \leq i \leq m, d_{ij}^{(0)} = \lambda_i q_{ij}, j \neq i, 1 \leq i, j \leq m, d_{ij}^{(1)} = \lambda_i p_{ij}, 1 \leq i, j \leq m,$ with $\sum_j p_{ij} + \sum_{j \neq i} q_{ij} = 1, \text{ for } 1 \leq i \leq m.$ By assuming that D_0 to be a nonsingular matrix, the successive times between “events” (or arrivals) will be finite with probability one and that the process will not terminate.

Thus, a MAP is described by the two parameter matrices (D_0, D_1) of order m such that the transitions corresponding to no arrivals are governed by D_0 and the transitions corresponding to arrivals are governed by D_1 . The underlying CTMC has the generator given by $Q = D_0 + D_1$. This representation of MAP is a special case of batch Markovian arrival process (BMAP). This BMAP was originally introduced by Neuts (1979) as a versatile Markovian point process (VMPP) in 1979. MAP includes several well-known point processes such as Poisson, Erlang, and hyperexponential. For full details on MAP and its applications to stochastic models we refer the reader to the survey papers by Lucantoni (1991) and Chakravarthy (2001). The fundamental rate (the rate of arrivals per unit of time), λ , is given by $\lambda = \pi D_1 e$, where π is the steady state probability vector of the generator Q governing the underlying CTMC satisfying $\pi Q = 0, \pi e = 1$ and where e is a column vector of 1’s with dimension m .

A phase type distribution (PH-distribution) is obtained as the time until absorption in a finite state continuous time Markov chain with n transient states and one absorbing states. The generator, Q , of this Markov chain is of the form given by

$$Q = \begin{bmatrix} S & S^0 \\ 0 & 0 \end{bmatrix},$$

where S is an $n \times n$ matrix, S^0 is a column vector of order n such that $S e + S^0 = \mathbf{0}$, where e is a column vector of 1’s of dimension n . In this case, we say that $Y \sim \text{PH}(\beta, S)$ of order n . The mean and variance of Y are given, respectively, by $\beta(-S)^{-1}e$, and $2\beta(-S)^{-2}e - [\beta(-S)^{-1}e]^2$.

PH-distributions include well-known distributions such as exponential, (generalized) Erlang, and hyperexponentials as very special cases. For details on PH-distributions and their properties we refer the reader to Neuts (1995).

An arriving customer is of type $i, i = 1, 2, 3$, with probability p_i . Thus, the rate of arrivals of type i customers to the system is λp_i . There is a finite buffer of size K for the arriving customers to wait if all servers are busy attending to other customers. Any arrival finding the buffer to be full is considered lost. For our model in this paper, we assume that the buffer is common to all three types of customers. However, it is very easy to modify this assumption to include separate buffer for each type of customers. There are d_i dedicated type i servers as well as d_f flexible servers to attend to the customers. While dedicated type i servers can offer services only to type i customers, the flexible servers can offer services to any type. An arriving type i customer finding a free dedicated type i server will get into service with one of the dedicated servers. However, if all dedicated type i servers are busy and if at least one flexible server is available at that instant, then one of the free flexible servers will offer service. Otherwise, the arrival will enter into the finite buffer of size K and wait for the next available server should there be a space in the buffer; otherwise the arrival will be lost. Waiting customers are served on a first-in-first-out basis within their types by either their respective dedicated servers or by flexible servers. When services are offered by freed flexible servers there is a priority involved in the selection of the type of the customers. Any waiting type 1 customer has the highest priority followed by type 2 and type 3. Thus, a waiting type 3 will be served by a flexible server when and only when there are no type 1 or type 2 customers waiting in the buffer. Note that this priority rule applies only when a flexible server becomes free as dedicated servers can only serve their respective customers. Services within a type are offered on a first-come-first-served basis by both dedicated and flexible servers.

The service times of type i , $1 \leq i \leq 3$, customers are assumed to be of phase type with representation given by $(\beta(i), S(i))$ of order n_i if services are offered by dedicated servers and $(\beta(i,F), S(i,F))$ of order n_{iF} if services are offered by flexible servers. Let μ_i , for $i=1, 2, 3$, denote the rate of services when offered by dedicated type i servers. That is, $\mu_i = [\beta(i) (-S(i))^{-1}e]^{-1}$. The rate of services when offered by flexible servers to type i customers is denoted by μ_{iF} , for $i, 1 \leq i \leq 3$. That is, $\mu_{iF} = [\beta(i,F) (-S(i,F))^{-1}e]^{-1}$.

By keeping track of the phase of the arrival process, the number of type i customers in the system including the number of flexible servers busy with type i customers, the model described above is amenable to study through the classical algorithmic methods due to Neuts (1989, 1995). However, we have chosen to simulate this service system using ARENA, a simulation software, as the state space for the current model grows exponentially and the book-keeping becomes very intensive due to the type of priorities associated with services involving flexible servers. Furthermore, the computation of the distributions of the waiting time in the system of various types of customers is very complicated to describe analytically. Thus, simulation will not only help to compare analytical results with those of the simulated ones but also get a feel for the waiting time distributions.

3. SIMULATION OF THE MODEL

In this section we will outline how ARENA is used to simulate the service system under study. The ARENA modules for developing the model under study are displayed in Figures 1 through 4. The purpose of this section is to bring out the qualitative aspects of the service system under consideration through some interesting simulated (numerical) examples.

SIMULATION OF A STOCHASTIC MODEL FOR A SERVICE SYSTEM
Catherine Brickner, Dennis Indrawan, Derrick Williams, and
Srinivas Chakravarthy

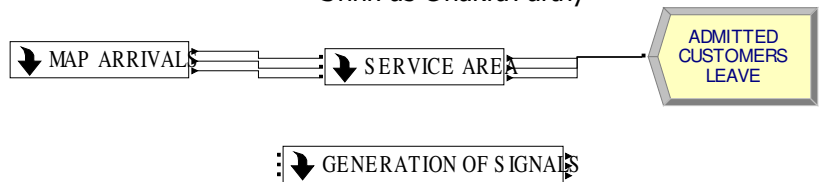


Figure 1: Main model

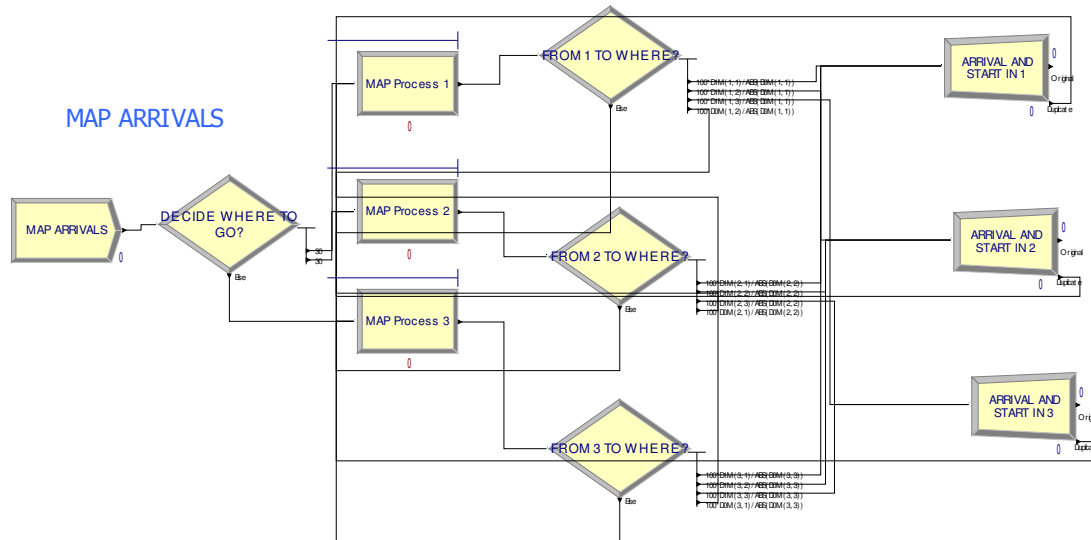


Figure 2: MAP arrivals

▶
▶
▶

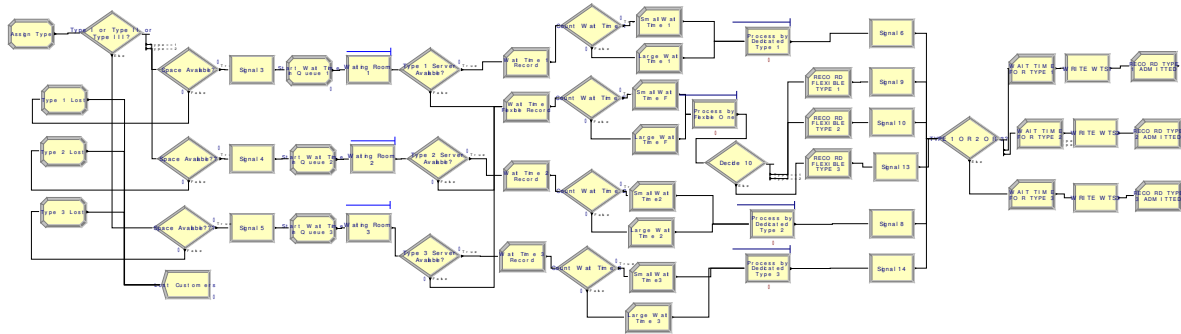


Figure 3: Service system

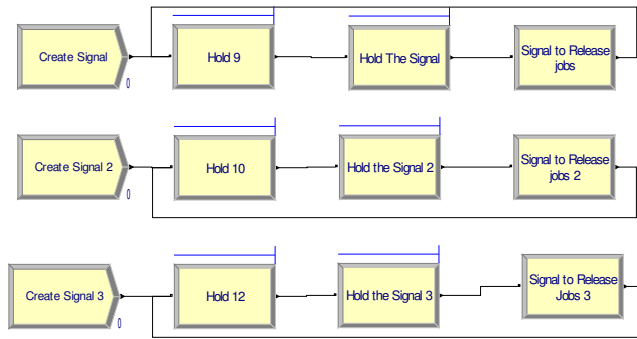


Figure 4: Generation of signals to release customers

For our numerical discussions, we consider five different arrival processes and three different service time distributions. The five arrival processes with parameter matrices D_0 and D_1 are as follows.

ERLA: Erlang of order 5:

$$D_0 = \begin{bmatrix} -5 & 5 & 0 & 0 & 0 \\ 0 & -5 & 5 & 0 & 0 \\ 0 & 0 & -5 & 5 & 0 \\ 0 & 0 & 0 & -5 & 5 \\ 0 & 0 & 0 & 0 & -5 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

EXPA: Exponential: $D_0 = -1$ and $D_1 = 1$.

HEXA: Hyperexponential: This is the mixture of two exponential with mixing probabilities 0.7, 0.2, and 0.1, and with parameters 12.7, 1.27, and 0.127. Here

$$D_0 = \begin{bmatrix} -12.7 & 0 & 0 \\ 0 & -1.27 & 0 \\ 0 & 0 & -0.127 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 12.7 & 0 & 0 \\ 0 & 1.27 & 0 \\ 0 & 0 & 0.127 \end{bmatrix}.$$

minute, and $\mu_{1F} = \mu_{2F} = \mu_{3F} = 0.2/3$ per minute. For the resources (number of dedicated and flexible servers), we take $d_1 = d_2 = d_3 = d_f = 3$, and the buffer size, $K = 300$.

We define the following system performance measures for our discussion on the simulated results.

$P_{Lost}^{(i)}$, $1 \leq i \leq 3$, is the probability that an arriving customer of type i is lost.

$P_{Lost}^{(all)}$ is the probability that an arriving customer (irrespective of the type) is lost. [Note that this probability is the sum of the probabilities of type 1, type 2, and type 3 customers are lost. However, this measure's half-width value cannot be obtained from those of the individual types.]

$\mu_{WTS}^{(i)}$, $1 \leq i \leq 3$, is the mean waiting time (in hrs) in the system of a type i customer.

$\sigma_{WTS}^{(i)}$, $1 \leq i \leq 3$, is the standard deviation (in hrs) of the waiting time in the system of a type i customer.

$\mu_{Busy}^{(i)}$, $1 \leq i \leq 3$, is mean number of busy dedicated type i servers.

$\mu_{Busy}^{(F)}$, is mean number of busy flexible servers.

$P_{Flexible}^{(i)}$, $1 \leq i \leq 3$, is the probability that a type i customer will be served by a flexible server.

For the above five arrival processes by fixing the other parameters to be as indicated earlier, we ran our simulation model for 10000 units and five replications. Due to space limitations only selected performance measures from the above set with the half-widths of the intervals are displayed for ERLS, EXPS, and HEXS, in Tables 1 through 3 below. However, we will summarize the results for other measures here.

First we consider the renewal arrivals (namely, for ERLA, EXPA, and HEXA cases).

The measure $P_{Lost}^{(i)}$, $1 \leq i \leq 3$, appears to increase with increasing variability of the arrivals times. This appears to be the case even if the variability in the service times increases.

The measures $\mu_{WTS}^{(1)}$ and $\mu_{WTS}^{(2)}$ appear to increase with increasing variability of the arrival times for the three service cases considered.

The measure $\mu_{WTS}^{(3)}$ appears to decrease with increasing variability of the arrivals only for ERLS and EXPS services; however, for HEXS services this measure appears to increase with increasing variability of the arrivals.

The measure $P_{Flexible}^{(1)}$ appears to increase with increasing variability of the arrival times for all three service cases studied.

The measures $P_{Flexible}^{(i)}$, $2 \leq i \leq 3$, appear to decrease with increasing variability of the arrival times for all three service cases studied.

The mean waiting time in the system for admitted type 3 customers has the largest value (for all five arrival processes) as compared to other two types of customers. This is as expected since type 3 customers have the lowest priority when served by flexible servers. This is true for all three service cases considered.

Whenever the flexible servers are used they seem to be busy mostly with type 1 customers, followed by type 3, and type 2 customers.

Now we compare the results of MNCA and MPCA arrivals. With respect to these two correlated processes, we see some interesting results. Note that these two processes have the same mean and variance. However, MNCA is negatively correlated while MPCA has positive correlation.

It appears that MPCA has a smaller value as compared to MNCA for $P_{Flexible}^{(i)}$, $2 \leq i \leq 3$, $\mu_{Busy}^{(i)}$, $1 \leq i \leq 3$, and $\mu_{Busy}^{(F)}$. In the case of the measure $\mu_{WTS}^{(3)}$ MPCA has a lower value only for ERLS and EXPS cases. For HEXS services, MPCA has a larger value for $\mu_{WTS}^{(3)}$ as compared to MNCA. For all other measures MPCA has a higher value as compared to MNCA.

Finally, it is worth pointing out some interesting contrasts between HEXA and MPCA cases. Note first that HEXA has a larger variance and MPCA has the largest (positive) correlation among all five arrival processes. The measures $\mu_{WTS}^{(1)}$ and $\mu_{WTS}^{(2)}$ appear to be larger for MPCA as compared to HEXA for all three service cases; the measure $\mu_{WTS}^{(3)}$ is higher for HEXA for all three services. The measures $\mu_{Busy}^{(i)}$, $1 \leq i \leq 3$, appear to be larger for HEXA than for MPCA in all cases. This illustrates the crucial role played by the correlated, especially the positive one, arrivals in stochastic modeling. We have seen such a crucial role played by the correlated arrivals in our other stochastic models analyzed using analytical and computational modeling tools.

Table 1: Selected system performance measures for Erlang services

Measure	ERLA	EXPA	HEXA	MNCA	MPCA
$P_{Lost}^{(all)}$	0.06152	0.06663	0.11485	0.06674	0.12444
$\mu_{WTS}^{(1)}$	0.23537	0.24704	0.39379	0.24784	1.12680
$\mu_{WTS}^{(3)}$	13.21500	13.13100	12.62000	13.11800	4.98260
$P_{Flexible}^{(1)}$	0.55386	0.56773	0.66735	0.56858	0.61635
$\mu_{Busy}^{(2)}$	1.75130	1.75580	1.72640	1.75430	1.70740

Table 2: Selected system performance measures for exponential services

Measure	ERLA	EXPA	HEXA	MNCA	MPCA
$P_{Lost}^{(all)}$	0.06247	0.06800	0.10563	0.06734	0.13305
$\mu_{WTS}^{(1)}$	0.26777	0.27871	0.41761	0.27831	1.12240
$\mu_{WTS}^{(3)}$	12.97300	12.89300	12.23100	12.89700	5.13620
$P_{Flexible}^{(1)}$	0.56225	0.56868	0.64950	0.56966	0.61416
$\mu_{Busy}^{(2)}$	1.75110	1.74700	1.72370	1.74520	1.70900

Table 3: Selected system performance measures for hyperexponential services

Measure	ERLA	EXPA	HEXA	MNCA	MPCA
$P_{Lost}^{(all)}$	0.00002	0.00022	0.02443	0.00011	0.10324
$\mu_{WTS}^{(1)}$	0.27206	0.28476	0.39878	0.28410	0.89044
$\mu_{WTS}^{(3)}$	1.53010	2.16580	6.69770	2.08980	3.63660
$P_{Flexible}^{(1)}$	0.50571	0.51399	0.58836	0.51897	0.56901
$\mu_{Busy}^{(2)}$	1.77980	1.79150	1.80050	1.79300	1.69790

Now we look at the fitted distributions of the waiting time distributions of various types of customers. Using the simulated data, ARENA has the option to identify the best fit (based on the least sum of squares due to error) among many distributions. In Table 4 we list the fitted distributions for various types for the

five arrival processes and three types of services considered. An examination of this table reveals the following.

It is worth noting that the waiting time distribution of an admitted type 3 customer is fitted using beta distribution for more scenarios. As is known, beta distribution has the ability to fit a variety of shapes in the data, and due to type 3 customers having the lowest priority (when served by flexible servers) the waiting times have more variability.

For MPCA arrivals, the fitted distributions for the waiting times for type 1 as well as for type 2 customers are beta or lognormal indicating again the variability caused by the positively correlated arrivals. Normally the waiting time distribution in a queueing model is skewed to the right (to accommodate for some customers having to wait unusually longer than the others) and this can be seen in ARENA identifying lognormal distribution to be the best fit for many combinations.

Table 4: Identification of the fitted distributions for the waiting times in the system

Arrival process	Service	Type 1	Type 2	Type 3
ERLA	ERLS	GAMM(0.0346, 6.8)	LOGN(0.186, 0.0688)	NORM(13.2, 0.972)
	EXPS	-0.001 + GAMM(0.191, 1.41)	-0.001 + GAMM(0.175, 1.13)	18 BETA(30.9, 12)
	HEXS	-0.001 + LOGN(0.3, 0.626)	-0.001 + LOGN(0.219, 0.509)	-0.001 + 17 BETA(0.584, 5.9)
EXPA	ERLS	LOGN(0.248, 0.106)	LOGN(0.191, 0.0735)	NORM(13.1, 0.938)
	EXPS	-0.001 + GAMM(0.197, 1.42)	-0.001 + GAMM(0.177, 1.14)	18 BETA(30.9, 12.2)
	HEXS	-0.001 + LOGN(0.316, 0.65)	LOGN(0.25, 0.665)	-0.001 + 19 * BETA(0.643, 5)
HEXA	ERLS	GAMM(0.114, 3.46)	LOGN(0.264, 0.146)	16 BETA(29.2, 7.76)
	EXPS	-0.001 + 4 BETA(1.39, 11.9)	GAMM(0.242, 1.14)	18 BETA(16.8, 7.93)
	HEXS	0.001 + WEIB(0.357, 0.942)	-0.001 + LOGN(0.391, 0.974)	NORM(6.7, 3.83)
MNCA	ERLS	LOGN(0.248, 0.107)	LOGN(0.191, 0.0734)	NORM(13.1, 1.01)
	EXPS	-0.001 + WEIB(0.291, 1.32)	GAMM(0.182, 1.11)	NORM(12.9, 1.2)
	HEXS	-0.001 + LOGN(0.315, 0.641)	-0.001 + LOGN(0.235, 0.552)	-0.001 + 17 BETA(0.629, 4.49)
MPCA	ERLS	LOGN(1.24, 1.92)	LOGN(0.773, 1.07)	14 BETA(1.05, 2.04)
	EXPS	-0.001 + 6.8 BETA(0.819, 4.27)	6 BETA(0.672, 4.66)	TRIA(0, 0.188, 15)
	HEXS	GAMM(1.31, 0.678)	-0.001 + 13 BETA(0.571, 8.71)	-0.001 + 18 BETA(0.723, 3.07)

Now we consider two optimization problems of interest. For these optimization problems, the objective function is defined as the expected total cost per hour and is calculated based on various costs (a) for hiring dedicated and flexible servers, (b) for customers waiting in the system, and (c) for lost customers due to finite buffer. The expected total cost per unit of time, F , is then given by (recall that λ is given in minutes and the mean waiting times are in hours)

$$F = \sum_{i=1}^3 c_{1i} d_i + 19 d_f + 60 \lambda \sum_{i=1}^3 c_{2i} p_i (1 - P_{Lost}^{(i)}) \mu_{WTS}^{(i)} + 60 \lambda \sum_{i=1}^3 c_{3i} p_i P_{Lost}^{(i)}, \tag{1}$$

where the costs (per hour) are as given in the following table.

Table 5: Various costs (per hour)

$(c_{1,1}, c_{1,2}, c_{1,3})$	$(c_{2,1}, c_{2,2}, c_{2,3})$	$(c_{3,1}, c_{3,2}, c_{3,3})$
(9.49, 12.4, 9.67)	(4, 3, 2)	(10, 8, 6)

OPT1: Optimization problem 1: We first fix $\lambda = 1/\text{min}$, $p_1 = 0.4$, $p_2 = 0.2$, $p_3 = 0.4$, $\mu_1 = \mu_2 = \mu_3 = 0.10/\text{min}$, $\mu_{1F} = \mu_{2F} = \mu_{3F} = 0.2/3$ per min, $d_1 = d_2 = d_f = 3$, and $K = 300$. We want to find an optimum value for d_3 for the five arrival processes and for the three service cases such that the function F given in (1) is minimized subject to $\mu_{WTS}^{(3)} \leq 0.5$ hr. Using one replicate of length 10,000 hours, the output from ARENA is summarized in Table 6. Along with the optimum values for the number of dedicated type 3 servers, four key measures $\mu_{WTS}^{(1)}$, $\mu_{WTS}^{(2)}$, $P_{Flexible}^{(1)}$, and $P_{Flexible}^{(2)}$ are compared for the current case with the one for which $d_1 = d_2 = d_3 = d_f = 3$ (see Tables 1 through 3). MPCA has the largest expected total cost per hour among all.

The measures, $\mu_{WTS}^{(1)}$ and $\mu_{WTS}^{(2)}$ are large as compared to their corresponding values displayed in Tables 1 through 3, for all the five arrival processes. Note that the values in Tables 1 through 3 are for $d_1 = d_2 = d_3 = d_f = 3$. At first it might look counter-intuitive see high values when the number of dedicated type 3 is increased. However, a closer look will indicate that as the number of dedicated type 3 servers is increased more type 3 customers will be taken into service by their dedicated servers leaving room for additional type 1 and type 2 customers. This results in an increase for the mean waiting times for type 1 and type 2 customers.

Due to an increase in the values for d_3 we see a decrease in the flexible servers attending type 3 customers and see an increase in the flexible servers attending both type 1 and type 2 customers.

Table 6: Optimum values for d_3 and $F[(d_3^*, F^*)]$ for various arrivals and services

SERVICES	ERLA	EXPA	HEXA	MNCA	MPCA
ERLS	(4, 203.18)	(4, 209.98)	(6, 264.05)	(4, 210.93)	(5, 422.45)
EXPS	(5, 228.58)	(5, 217.99)	(5, 266.36)	(5, 218.70)	(5, 421.49)
HEXS	(4, 209.08)	(4, 213.81)	(6, 243.11)	(4, 213.13)	(6, 366.38)

OPT2: Optimization problem 2: We first fix $\lambda = 1/\text{min}$, $p_1 = 0.4$, $p_2 = 0.2$, $p_3 = 0.4$, $\mu_1 = \mu_2 = \mu_3 = 0.10/\text{min}$, $\mu_{1F} = \mu_{2F} = \mu_{3F} = 0.2/3$ per min, and $K = 300$. We want to find optimum values for d_1 , d_2 , d_3 , and d_f , for the five arrival processes such that the function F given in (1) is minimized subject to $\mu_{WTS}^{(i)} \leq 0.5$, $1 \leq i \leq 3$, $2 \leq d_1, d_2 \leq 3$, $2 \leq d_3 \leq 5$, $2 \leq d_f \leq 4$. Using “Optquest” in ARENA with one replicate of length 10,000 hours the results are summarized in Table 7.

Table 7: Optimum values for the number of servers

ARRIVALS	ERLS	EXPS	HEXS
ERLA	(3, 2, 4, 3, 196.20)	(3, 2, 5, 3, 209.40)	(3, 2, 5, 3, 205.24)
EXPA	(3, 2, 5, 3, 205.97)	(3, 2, 5, 3, 215.66)	(3, 3, 5, 3, 208.31)
HEXA	(3, 3, 5, 4, 264.64)	(3, 3, 5, 4, 264.39)	(3, 3, 5, 3, 244.26)
MNCA	(3, 2, 5, 3, 206.41)	(3, 2, 5, 3, 244.26)	(3, 2, 5, 3, 206.14)
MPCA	Infeasible	Infeasible	Infeasible

A quick look at this table reveals the following observations

For MPCA arrivals there is no feasible solution indicating that the need for a large number of servers in order to guarantee the mean waiting time in the system for any type customers to be no larger than 0.5 hr. Even though type 2 customers arrive less on the average, for HEXA requires more dedicated type 2 servers due to its high variability.

As expected the maximum number of available dedicated type 3 servers are used in most cases.

4. CONCLUDING REMARKS

In this paper we studied a service system model in which three types of customers receive services from a pool of dedicated and flexible servers. Using MAP arrivals and phase type services, the complex model is studied through simulation using ARENA. Interarrivals times having a larger variation and or a positive correlation yield higher values for the mean time spent in the system. Furthermore, we observed that servicing calls that arrive in a positively correlated manner incur the largest expected total cost per hour to the system. Additional experimentation to look into more interesting qualitative behavior of the model is in progress and the results will be reported elsewhere.

REFERENCES

- Chakravarthy, S. R. 2001. The Batch Markovian Arrival Process: A Review and Future Work. In *Advances in Probability Theory and Stochastic Processes*, Eds., A. Krishnamoorthy, N. Raju, and V. Ramaswami, 21-49. Notable Publications, Inc., New Jersey, USA.
- Chakravarthy, S. R., and S. R. Agnihotri. 2005. Optimal workforce mix in service systems with two types of customers, *Production and Operations Management*, 14, 218-231.
- Lucantoni, D. 1991. New results on the single server queue with a batch Markovian arrival process. *Stochastic Models*, 7, 1-46.
- Neuts, M. F. 1995. *Matrix-Geometric Solutions in Stochastic Models -An Algorithmic Approach*. Dover Publications, Mineola, NY, USA. (originally published by Johns Hopkins University Press, 1981).
- Neuts, M. F. 1979. A versatile Markovian point process. *J. Appl. Prob.*, 16, 764-779.
- Neuts, M. F. 1989. *Structured Stochastic Matrices of M/G/1 type and their Applications*. Marcel Dekker, New York, NY, USA.
- "PayScale-Best Buy Employer Wages, Hourly Wage Rate". PayScale.com. 09/07/2009 <http://www.payscale.com/research/US/Employer=Best_Buy/Hourly_Rate>.

AUTHOR BIOGRAPHIES

CATHERINE A. BRICKNER is currently an Industrial Engineer at American Express, in Greensboro, North Carolina. She recently graduated from Kettering University in Flint, Michigan with a Bachelors degree from Industrial Engineering and a Quality Assurance concentration. Her interests include queuing and process control. Her email address for these proceedings is <catherine.brickner@gmail.com>.

DENNIS INDRAWAN is a senior undergraduate student majoring in Industrial Engineering at Kettering University, Flint, Michigan, USA. His areas of interest are in the field of probability, simulation, Six Sigma quality control, and financial risk management and company valuation. On campus, he is actively involved in the professional student chapter of Institute of Industrial Engineers at Kettering University and is a member of Alpha Pi Mu, the Industrial Engineering Honor Society and Tau Beta Pi, the National Engineering Honor Society. He expects to graduate with a Summa Cum Laude and plans to continue his academics in the field of Financial Engineering. His email address for these proceedings is <indr4284@kettering.edu>.

DERRICK WILLIAMS recently graduated from Kettering University with a Bachelors degree from Industrial Engineering.

SRINIVAS R. CHAKRAVARTHY is a Professor of Operations Research and Statistics in the Department of Industrial and Manufacturing Engineering at Kettering University, Flint, Michigan, USA. His research interests are in the areas of algorithmic probability, queuing, reliability, inventory, and simulation.

Brickner, Indrawan, Williams and Chakravarthy

He has published more than 80 papers in leading journals and presented papers at national and international conferences. He co-organized the First and Second International Conferences on Matrix-analytic Methods in Stochastic Models in 1995 and 1998. He received Kettering University Alumni's Outstanding Teacher Awards (1990 and 2001) and Kettering University's Outstanding Researcher (1996) and Distinguished Researcher (2003) Awards. He is on the scientific advisory committee of several international conferences, and also on the editorial board of a few journals. His email address for these proceedings is schakrav@kettering.edu.