

# A SIMULATION FRAMEWORK FOR NETWORKED QUEUE MODELS: ANALYSIS OF QUEUE BOUNDS IN A G/G/c SUPPLY CHAIN

MAHYAR AMOUZEGAR AND KHOSROW MOSHIRVAZIRI

*Received 4 April 2006; Accepted 18 May 2006*

Some limited analytical derivation for networked queue models has been proposed in the literature, but their solutions are often of a great mathematical challenge. To overcome such limitations, simulation tools that can deal with general networked queue topology must be developed. Despite certain limitations, simulation algorithms provide a mechanism to obtain insight and good numerical approximation to parameters of networked queues. This paper presents a closed stochastic simulation network model and several approximation and bounding schemes for G/G/c systems. The analysis was originally conducted to verify the integrity of simulation models used to develop alternative policy options conducted on behalf of the US Air Force. We showed that the theoretical bounds could be used to approximate mean capacities at various queues. In this paper, we present results for a G/G/8 system though similar results have been obtained for other networks of queues as well.

Copyright © 2006 M. Amouzegar and K. Moshirvaziri. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. Introduction

In this paper we consider a closed stochastic simulation system model used in the analysis of aircraft engines maintenance and repair options. In this analysis, we evaluated the cost and benefits of centralized maintenance versus a decentralized option. This analysis was prompted by the ongoing reorganization of the Air Force into an Air and Space Expeditionary Force (AEF). The main objective of this reorganization is to replace the forward presence of air power with a force that can deploy quickly from the continental United States (CONUS) in response to a crisis, commence operations immediately upon arrival, and sustain those operations as needed. To support the expeditionary force, support processes such as munitions, fuels, and maintenance also need to be transformed.

## 2 A simulation framework for networked queue models

AEF requires a combat support system capable of supporting an expanded range of operations from humanitarian and disaster relief to major combat and peacekeeping operations, which could take place in any of a number of different locations.

One of the critical processes for the Air Force is the intermediate maintenance for jet engines. This so-called intermediate maintenance facility (IMF) consists of several components, including the maintenance (repair and service) shop, the module shop, and the assembly and test cell. IMF is one of three levels of maintenance used by the Air Force to repair jet engines, especially those powering fighter aircraft.

- (i) Flightline maintenance consists mostly of inspections, diagnostics, engine removals, and some quick repairs that do not involve engine teardown.
- (ii) Service at IMF includes disassembly of the engines; substantial repairs to parts such as fans, low pressure turbines, and afterburners; and engine test cell runs.
- (iii) Depot maintenance involves the complete teardown and refurbishment of any repairable part in an engine. The rebuilding of an engine at the depot allows the engine's use of parameters (flight time, cycles, etc.) effectively to be reset at zero.

Traditionally, the IMF has been located at the operating base with the aircraft. This policy was reinforced by the planning for major wars in Europe and Korea: a unit would be moved to existing bases in theater in preparation for immediate action and could expect little resupply during the first few weeks of combat. Under traditional planning for wing deployment, therefore, the IMF is prepared to move along with the rest of the wing support, although not with the combat units themselves, who will use spares to replace engines until the IMF arrives and is up and running.

**1.1. Current practices and trends.** In recent years, the question of whether or not IMF operations should be centralized has been the subject of frequent discussion in the engine community. Many factors have favored centralization, including the increased complexity of engines and the large investment required for repair facilities. Other factors have mitigated against centralization, particularly the fact that, unlike other commodities such as avionics components, engines are heavy and bulky and thus require special packing to ship. Over the years, the Air Force has experienced a pattern of alternation between the partial centralization of the maintenance operations—in certain regions and for certain engine types—and the subsequent restoration of IMF to operating units. The requirements associated with expeditionary operations—including the ability to move quickly and the need to keep initial transportation requirements down—have raised new questions about the policy of locating the IMF at the operating base. This research aims to provide insights into this issue by determining whether engine maintenance support can best be provided from decentralized shops at the supported bases or from a centralized, off-base facility.

The operation and maintenance of engines comprise the sequence of events illustrated as an aggregate in Figure 1.1: planes fly (sorties) from main bases and remote operating locations across the globe to meet training and other requirements. After each sortie, aircraft engines are inspected on the flight line, and, depending on the accumulated flying hours and other factors, are given minor maintenance. Engines may also be removed from aircraft and sent to an intermediate maintenance facility for major service and repair. At

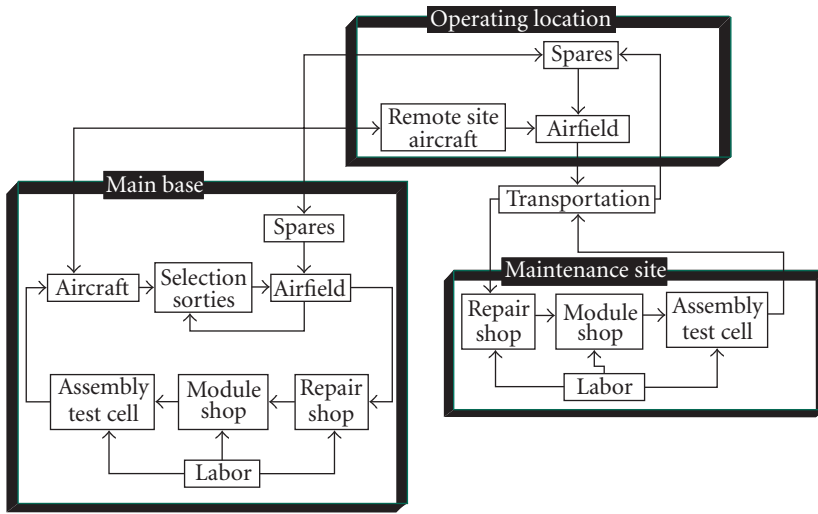


Figure 1.1. Operation and maintenance sequence.

this facility the engines are inspected, repaired, tested, and then returned to the flight line as serviceable spares. At each operating site there is a cache of serviceable spares to replace engines sent to IMF. However, there is only a limited inventory of such spares; there may be time where aircraft are grounded due to the engines availability. The ultimate goal is to increase the efficiency of the maintenance process while keeping the least number of spares as possible.

**1.2. Model formation.** The nature of this problem has lent itself to a closed loop networks of multiple servers/queues, some sequential and others parallel (over forty queues and servers). A queueing system is said to be closed if the servicing facility processes only a given group of permanent customers. When a customer needs service, it joins the queue and it is either served based on FIFO discipline or is given priority if it meets a certain criteria (e.g., a particular engine is required in the field faster than other type). The demand for service and duration of service depends on many variables and for this study we used historical data to compute the arrival and departure rates. The complexity of this problem led to a queueing model that could only be described with general arrival and service times or a  $G/G/c/n$  queueing system where  $n$ , the restriction on system capacity, varied depending on the process.  $G/G/c$  queue and its related families,  $M/G/c$ ,  $G/G/1$ , are too complex to analyze mathematically and there are very few closed-form results about such systems. However, several quite useful approximate and bounding results have been obtained. We used these approximations and bounds to create a robust simulation model for a large-scale engine maintenance system. These bounds and approximations were used in evaluating the robustness of our simulation model.

In the next section, we will describe some of the results associated with G/G/c. In the subsequent section, we will present the simulation model and some numerical results. We will end this paper with a few concluding remarks.

## 2. G/G/c system

The G/G/1 system and its theoretical results are used to derive what is presently known about the G/G/c system and thus will be discussed first. We consider a G/G/1 system consisting of a single server with independent and identically distributed interarrival times as well as service times and unlimited queueing capacity. Let  $X$  denote interarrival times and let  $f_x(x)$ ,  $1/\lambda$ , and  $\sigma_x^2$  denote the probability density function (pdf), the mean, and the variance of  $X$ , respectively. In addition, let  $S$ ,  $f_s(s)$ ,  $1/\mu$ , and  $\sigma_s^2$  represent those corresponding for the service times. Although there are no closed form solutions for this model, there are some useful bounds developed in recent years for the quantities  $L$ ,  $L_q$ ,  $W$ , and  $W_q$  (see [4, 7]).

For G/G/1 systems with no restrictions on the interarrival or on the service time pdf's, several bounds have been developed (see [8, 9]). These bounds, in essence, state that for the average steady-state waiting time in queue,  $W_q$ , we have

$$\frac{\rho^2(1+C_s^2)-2\rho}{2\lambda(1-\rho)} < W_q \leq \frac{\lambda(\sigma_x^2+\sigma_s^2)}{2(1-\rho)}, \quad (2.1)$$

where  $C_s = \sigma_s/\mu$  is the coefficient of variation for the service times, and  $\rho = \lambda/\mu$  is the utilization factor. For the stability of the system we must have  $\rho < 1$ . Note that the lower bound given above is not tight. This becomes obvious from the fact that, even at very high utilization rates, the bounds take negative values, unless  $C_s > 1$ . But for  $C_s$  to be greater than 1, it must be that the service time pdf must be "more random" than the negative exponential pdf which has its  $C_s = 1$ .

**2.1. Desired class property.** A tight simple lower bound is given in [9] for a class of G/G/1 queues, which includes most practical problems encountered in the real world. Thus, class requirement is that all queueing systems in it must have interarrival time pdf,  $f_x(x)$ , satisfying the following property:

$$E[X-t | X > t] \leq \frac{1}{\lambda} \quad \forall t \geq 0. \quad (2.2)$$

If it is known that any given interarrival gap lasted more than a time  $t$ , then the condition above requires that the expected length of the remaining time,  $X-t$ , in that gap be less than the unconditional expected length of the gap,  $E[X](=1/\lambda)$ . This is of course true for the negative exponential variable, and in that case the condition becomes equality. When the condition holds, then we have

$$U - \frac{1+\rho}{2\lambda} \leq W_q \leq U, \quad U = \frac{\lambda(\sigma_x^2 + \sigma_s^2)}{2(1-\rho)}. \quad (2.3)$$

The upper and lower bounds may now be derived using this and by applying Little's formula,  $L = \lambda W$ ,  $L_q = \lambda W_q$ , and the fact that  $W = 1/\mu + W_q$ . The following is easily

obtained:

$$\lambda \cdot U - \frac{1+\rho}{2} \leq L_q \leq \lambda \cdot U. \quad (2.4)$$

This implies that the difference between the upper and the lower bounds is  $(1 + \rho)/2$ , but  $0 < \rho < 1$ , so this difference is always between 0.5 and 1. Thus, we can find the average queue length to within an accuracy of between 0.5 and 1 (depending on the value of  $\rho$ ). Note that most “well-behaved” arrival time distributions satisfy the condition, including uniform, triangular, or beta-type pdf’s, which often are reasonably good approximations of many general interarrival time pdf’s. Only a few common continuous random variables, such as those in the hyperexponential family, which are “more random” (informally speaking) than the negative exponential random variable, do not satisfy the condition.

**2.2. Under heavy traffic.** Another important result that is available for the G/G/1 system is known as the heavy-traffic approximation (for more information see [5]). It applies for values of  $\rho$  near 1 and thus provides estimates for waiting times when it is known that waiting times are large. When  $\rho$  is near 1, the distribution of steady-state waiting time in queue in a G/G/1 system is approximately *negative exponential* with mean value  $W_q = U$ . The average waiting time for G/G/1 queueing systems is dominated by a  $(1 - \rho)^{-1}$  term under steady-state conditions, as the utilization ratio tends to 1. Consequently, the type of behavior that is normally seen in a simple M/M/1 system is also present for entirely general arrival- and service-time distributions, G/G/1.

**2.3. G/G/c bounds.** The only general results on G/G/c system [2] that have been obtained to date are in the form of quite relaxed upper and lower bounds on average steady-state queueing characteristics. These bounds are often computed by, first, comparing a G/G/c system with a G/G/1 system that has the same “service behavior” as the G/G/c system. That is, the single server in G/G/1 works  $c$  times as fast as each of the servers in G/G/c and by applying the earlier results on G/G/1, given in the previous section. The most useful and applicable bounds on the average waiting time in queue which have been derived to date for G/G/c systems, based on those of G/G/1, is

$$W_q^1 - \frac{(c-1)\mu E[S^2]}{2c} \leq W_q \leq \frac{[\sigma_X^2 + (1/c)\sigma_S^2 + ((c-1)/c^2)(1/\mu^2)]\lambda}{2(1-\lambda/c\mu)}, \quad (2.5)$$

where for each of the  $c$  servers,  $\mu$ ,  $\sigma_S^2$ , and  $E[S^2]$  are the rate, variance, and the second moment of service time, respectively.  $W_q^1$  denotes the mean waiting time for a G/G/1 system with a service time denoted by a random variable  $S^1 = S/c$  with service  $c$  times faster than that of each of the  $c$  servers in the G/G/c system, but with an identical arrival process. If  $W_q^1$  is known or is computed using the results discussed above, we can substitute an exact expression. Note that for the general M/G/1 system we have the following well-known

results, which can be used in deriving the G/G/c approximation bounds:

$$\begin{aligned}
 P_o &= 1 - \rho, & L &= \rho + \frac{\rho^2 + \lambda^2 \sigma_S^2}{2(1 - \rho)}, \\
 W &= \frac{L}{\lambda} = \frac{1}{\mu} + \frac{\rho^2 + \lambda^2 \sigma_S^2}{2(1 - \rho)}, \\
 W_q &= W - \frac{1}{\mu} = \frac{\rho^2 + \lambda^2 \sigma_S^2}{2(1 - \rho)} = \frac{\lambda[(1/\mu^2) + \sigma_S^2]}{2(1 - \rho)}, \\
 L_q &= \lambda W_q = \frac{\rho^2 + \lambda^2 \sigma_S^2}{2(1 - \rho)}.
 \end{aligned} \tag{2.6}$$

Thus, for example, for the M/G/c queueing system, one should use the exact expression for  $W_q^1$  given above with  $1/c\mu$  and  $\sigma_S^2/c^2$ , for the expected value and variance of the service times, respectively.

The corresponding heavy-traffic approximation for G/G/c systems has been derived [6]. This result implies that for  $\lambda c\mu$  approaching 1 in a G/G/c system, the waiting time in queue under steady-state conditions assumes a distribution that is approximately *negative exponential* with mean value

$$W_q = \frac{[\sigma_x^2 + (\sigma_S^2/c)]\lambda}{2(1 - \lambda/c\mu)}. \tag{2.7}$$

Note once more that expected waiting time is dominated by a  $(1 - \rho)$  term, as  $\rho$  approaches 1 ( $\rho = \lambda/c\mu$  for multiserver systems). We used the above results for G/G/c to have a point of reference for the simulation and tested the results against these theoretical backdrops.

### 3. An overview of the simulation model

In terms of modeling, we are interested in the flow of entities (e.g., spares, personnel), the state of the system (e.g., engine not serviceable, spares inventory), and the processes (e.g., service time, sortie rates). The structure of the model is based on a set of hierarchical, functional blocks that generate and modify entities, processes, and attributes. These blocks represent main bases, airfields, and intermediate maintenance shops.

In general, the simulation is based on the following sequence of events: aircraft are flown from main bases or remote sites to meet certain flying requirements. After each mission, the aircraft and their engines are inspected at the airfield and in most cases they are fully operational within hours. However, when engines accumulate enough flying hours, or when unscheduled maintenance is required, engines are removed from the planes and sent to an intermediate maintenance facility. Flightline maintenance includes servicing, repairs, cycle recording, and tracking, which are coordinated with the engine management branch (EMB) and IMF. On the flightline, installed aircraft engines are serviced on a daily basis, which includes servicing the oil, inspecting the chip detectors, and entering the intakes and augmentor to inspect for foreign object damage (FOD) and

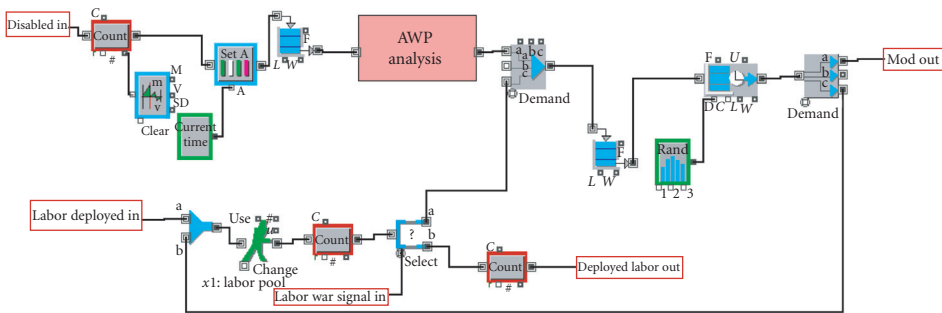


Figure 3.1. Intermediate maintenance shop.

external engine damage. In addition, engine cycles are recorded in the comprehensive engine management systems (CEMS) database. CEMS enables the EMB to monitor usage of engines and modules (when used) to determine the need for inspections and time change technical orders (TCTOs). The flightline also performs all engine removals and installations. After the flightline removes an engine for maintenance at the IMF, it sometimes performs sheet metal work on the engine bay and replaces some of the hydraulic lines and cables in the aircraft engine bay that have been damaged due to chafing, cracks, or heat.

The IMF is responsible for both scheduled and unscheduled off-equipment engine maintenance. Scheduled maintenance includes module time changes, TCTOs, and other inspections and repairs. Unscheduled maintenance consists primarily of performance-related problems that either cannot be corrected by the flightline or are beyond their capabilities per technical order. For unscheduled maintenance, the intermediate maintenance shop often performs a preliminary test cell run to troubleshoot the engine and identify other potential problems. The IMF is capable of replacing any module in a modular engine and also repairs some of the modules while sending others to the depot. It is also responsible for packing engines for transportation.

The IMF operates the engine test cell facility and functions. As part of this function, the IMF personnel transport engines, hook up cables, and fuel lines conduct pre- and post-run engine inspection, and disconnect cables and fuel lines. In many instances, the IMF also serves as a source of expertise to back up the flightline and provide quick response repair or cannibalizing key parts as needed. This organization is quite large (100–150 people for a fighter wing) and occupies an industrial space equipped with five or more work bays of 1500 square feet each, an overhead crane, supply storage, backshops for specialized repair activities, and a test cell. The test cell is typically located offsite in a “hush house” where a fully-assembled engine can be run at full power for testing purposes.

The general flow of IMF work is as follows with portion of the process depicted in Figure 3.1:

- (i) receive engine from the flightline;
- (ii) perform inspection and time change check;

## 8 A simulation framework for networked queue models

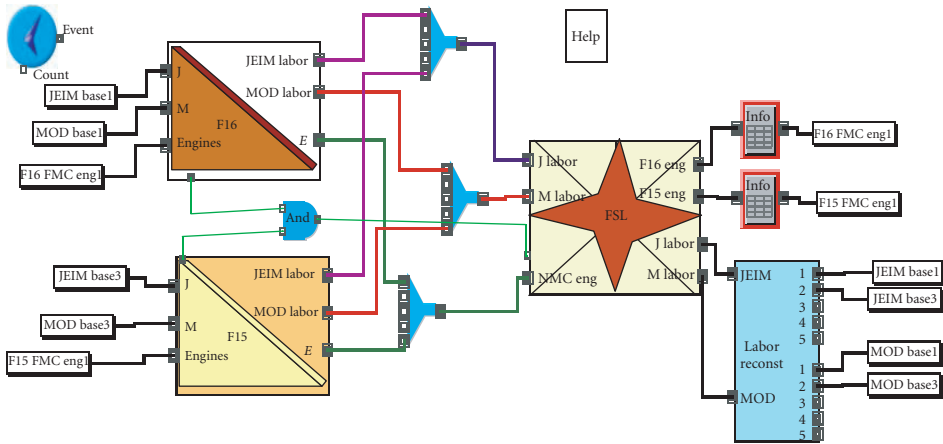


Figure 3.2. Top-level view of the simulation model.

- (iii) perform database history check;
- (iv) create a job in core automated maintenance system (CAMS);
- (v) assign engine to a crew;
- (vi) determine required repairs;
- (vii) decide on complete or partial disassembly;
- (viii) conduct other inspections;
- (ix) conduct teardown;
- (x) perform IMF repair and maintenance;
- (xi) perform module work, if needed, at module shop;
- (xii) assemble engine;
- (xiii) send to test cell “(hush house);”
- (xiv) conduct final inspection.

The first requirement for the model is the number and types of aircraft, and the number and the age of installed engines. The aircraft and engines are combined to form fully operational aircraft. They are sorted, based on the age of the engine, and are then queued for flying. After each sortie, the aircraft is sent to the airfield block where it is inspected and maintained. Each aircraft that passes the inspection is sent back to the pool of available aircraft. Some aircraft require minor repair, which is performed on the flight line. The number of engines pulled from the aircraft is a function of the age and the type of the engine. The detached engines are tagged according to the removal type (i.e., scheduled or unscheduled) and are sent to the IMF shop. Aircraft are then identified as not operational and are queued for the next available serviceable engine. These aircraft are either put back to service immediately, if there are serviceable spares available, or they await the arrival of engines from the maintenance shop. Figure 3.2 illustrates the top level view of the simulation model using block diagrams from Extend software (Extend is a registered



trademark of Imagine That, Inc.). This figure presents notional F-15 and F-16 jet fighters with a centralized maintenance facility.

At the maintenance facility, engines are queued in two parallel lines, the first is for the engines that require parts that are not available and the other is for engines that await maintenance. The modular engines that have been processed by the IMF shop are sent to the module shops. Engines that enter the module shop are separated into five modules. Engines that leave the module shop are sent to the assembly and test cell. In this section, engines are queued for assembly, the test cell, and the final inspection. After assembly and test cell, engines are sent to the spare engines pool to be installed on the aircraft to create fully operations aircraft. These aircraft leave this section to join the pool of other aircraft and the whole cycle starts again. Figure 1.1 illustrates this process for only one main and operating base. The model, however, has taken into account a problem with several such bases (for more information on the simulation model, see [1]).

**3.1. Simulation setup and data analysis.** We analyzed a number of possible support configurations for the IMF involving various combinations of centralized and decentralized locations. Centralized maintenance structures include forward support bases, while decentralized locations include home base support and maintenance at forward operating bases. Each structure was assessed under both a war and a peacetime scenario.

Here we describe in detail the specific IMF alternatives we evaluate in this analysis.

(i) *Decentralized-deployed.* In this alternative peacetime, maintenance is provided by IMFs located at each base. When part of a unit is deployed, part of that unit's IMF deploys to the appropriate forward bases as well.

(ii) *Decentralized-no deployment.* As with the previous alternative, each of the peacetime bases has its own IMF, but in this case the home IMF supports any deployed forces from its unit as well. The home base is sized so that it has the resources to support both peacetime and wartime flying.

(iii) *Decentralized-forward support location.* As with the previous two alternatives, each peacetime base has its own IMF, but when the units deploy, some of the IMF personnel (but not their equipment) deploy to a single overseas base in theater from which all deployed units are supported.

(iv) *US support location-forward support location.* In this alternative all units are supported in peacetime by a single centralized operation at home, which deploys personnel to an overseas base in theater when conflict occurs. In peacetime, the home IMF is staffed with the sum of the rail teams needed for deployment and those required to keep the nonengaged forces flying.

(v) *Home support location.* In this last alternative all units everywhere are supported by a single shop both during peacetime and in deployment.

During the simulations, we evaluated each of these alternatives using three broad metrics. The first is performance: does the alternative provide the required support for

operational flying? In peacetime this means being able to maintain the requisite flying for pilot training; in wartime it means being able to meet the required number of sorties day by day. The second metric is resources: what does the alternative require to provide adequate performance? For jet engines, one of the key resources is spare engines, which can provide a hedge against uncertainties. Other resources are personnel and transportation costs, and the evaluation provides an indication of the tradeoff between these two. The third metric is uncertainty: how well does the alternative respond to unforeseen events? For this metric, we evaluate how robust the Alternatives are to changes in the engine removal rate.

Many of the inputs to the model were provided by analysis of data drawn from the comprehensive engine management system (CEMS), reliability and maintainability management information system (REMIS), which rolls up data from the base-level core automated maintenance system (CAMS), as well as data in both electronic and paper form provided by the units we visited. The CEMS data provided information on total repair time for individual engines, engine not mission capable due to supply (ENMCS) times and transportation times for some engines were provided by some of the bases. REMIS provided a check on the CEMS data for overall engine repair and provided repair data for module work.

**3.2. Simulation results.** In this section, we will present some of the parameters used in our analysis and the results of the simulation runs. We will illustrate these parameters by running a scenario with 36 F-16s and 66 F-15s. The model is run for about two simulated years.

Engines are typically set on a rail and require a 5-person team per shift. The regular shift is about 8 hours and the shops operate at 2 shifts a day. During peak demand period, the shops may shift their operations to 24 hours a day, seven days a week with each shift as long as 12 hours. The capacity of the IMF is determined by the combination of rails and the personnel, a “rail team.” Other shops have different architecture but all are bounded by number of staff and the equipment. Airfields and the transportation network are bounded by the capacity of the flight line and the number of transporters, respectively. There are three smaller main bases with three-rail team capacity and a large one with 7-rail teams capacity; in other words, 3 and 7 parallel servers, respectively. There is also a remote facility with 8 rail teams. The other parts of the shop (e.g., the module shop) are sized accordingly.

On average about 119 customers entered the system (with variance of 253 and standard deviation of 15). At the end of the simulation run, about 105 customers were served. The IMF shop at the remote site (with 8 servers) reported an average wait time of 5.538461538462 days. Although the arrival and the service times varied widely, as they depend heavily on the other parts of the system, the reported wait time seemed reasonable and was consistent with the theoretical bounds. Using the Poisson distribution, we get a wait time of 2.06 days and 4.13 using the exponential distribution. Table 3.1 illustrates the theoretical bounds for a single server process in the inspection shop. The simulation model reported an average of 0.808499576845 for the queue length and 10 days for the average wait.

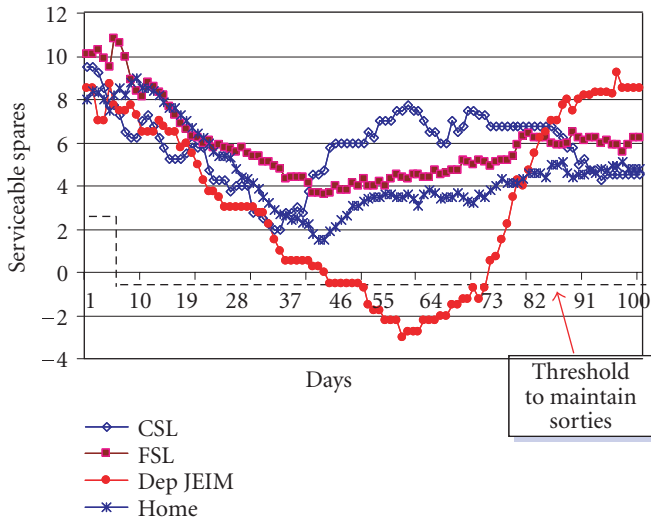


Figure 3.3. Deployed F-16 results.

Table 3.1. Sample results for a G/G/1 system.

Distribution	Utilization factor	Var (X)	Var (S)	Wait time		Length	
				LB	UB	LB	UB
Poisson	1.32	0.33	0.25	—	—	0.06130	—
Eponential	0.757576	0.1089	0.0625	0.78125	1.07125	3.125	3.24621
Uniform	—	0.02083	—	—	—	—	—
	0.08	3	0.003333	—	0.01050	—	0.00840
Normal	—	0.90090	—	—	—	—	—
	0.18018	1	5	2.58725	3.24225	2.51103	2.92094

Table 3.2 illustrates the arrival and departure rates for the sequence of servers in the maintenance process. Some customers bypass the first queue and enter the second queue with multiple servers. After the service, some customers, again, bypass the next server. In this section, there are five parallel servers and customers depending on their requirement must enter a particular server queue. Finally all customers enter the last server.

Table 3.3 illustrates the theoretical versus simulated bounds for the first queue, in the eight-server scenario discussed above.

Figure 3.3 presents the results from the deployment portion of the operation for the F-16 aircraft, comparing the centralized US support location (US), forward support location (FSL), deployment maintenance shops (Dep IMF), and home base locations (Home).

Table 3.2. Arrive and departure in the IMF.

Server (single)			Server (multiple)		Server (single) 5 parallel			Server (single)	
A	D	B	A	D	A	D	B	A	D
33	25	86	111	104	94	92	10	102	96

Table 3.3. Sample results for a G/G/8 system.

Distribution	Queue wait time		Length		Simulation results	
	LB	UB	LB	UB	W	L
Poisson arrivals	—	—	—	—	—	—
Exponential	3.21107	4.97622	1.22465	1.64215	5.04683	5.66852
Uniform	0.68181	1.3561	.02367	.15627	3.97198	3.20338
Normal	2.58725	3.24225	2.51103	2.92094	5.04280	5.89083

#### 4. Concluding remarks

We presented a closed stochastic simulation network model and several approximation and bounding options available in a G/G/c system. The model was implemented under several simulation environments, including Extend v6 [3]. The analysis was conducted to verify the integrity of the simulation model used to developed alternative policy options conducted on behalf of the US Air Force and presented in [1]. We showed that the theoretical bounds could be used to approximate mean capacities at various queues. In this paper only the results for G/G/8 was presented in order to avoid lengthy numerical tabulation of the results. However, such consistency was observed amongst the other queues.

#### References

- [1] M. Amouzegar, L. S. Galway, and A. Geller, *Supporting expeditionary aerospace forces: an analysis of jet engine intermediate maintenance options*, Tech. Rep. MR-1431-AF, RAND, California, 2001.
- [2] S. L. Brumelle, *Some inequalities for parallel-server queues*, *Operations Research* **19** (1971), 402–413.
- [3] B. Diamond, S. Lamperti, D. Krahl, and A. Nastasi, *Extend v6 User's Guide*, 2002.
- [4] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*, 3rd ed., Wiley Series in Probability and Statistics: Texts and References Section, John Wiley & Sons, New York, 1998.
- [5] J. F. C. Kingman, *On queues in heavy traffic*, *Journal of the Royal Statistical Society. Series B. Methodological* **24** (1962), 383–392.
- [6] L. Köllerström, *Heavy traffic theory for queues with several servers: I*, *Journal of Applied Probability* **11** (1974), 544–552.
- [7] R. Larson and A. Odoni, *Urban Operations Research*, Prentice-Hall, New Jersey, 1981.

- [8] W. G. Marchal, *Some simpler bounds on the mean queuing time*, Operations Research **26** (1978), no. 6, 1083–1088.
- [9] K. T. Marshall, *Some inequalities in queuing*, Operations Research **16** (1968), 651–665.

Mahyar Amouzegar: The RAND Corporation, Santa Monica, CA 90407-2138, USA;  
College of Engineering, California State University, Long Beach, CA 90840, USA  
*E-mail address:* mahyar@csulb.edu

Khosrow Moshirvaziri: Information Systems Department, California State University, Long Beach,  
CA 90840, USA  
*E-mail address:* moshir@csulb.edu