Modeling Flows of Engines Through Repair Facilities When Inputs Are Time-Varying

John A. Muckstadt

January, 2015

1 Introduction

In this paper we present two types of models that can be employed to analyze the time dependent behavior of flows of engines through maintenance facilities. These models first focus on engine repair when the length of the planning horizon is assumed to be short, say a few weeks at most. Then we discuss how to extend one of our models to include both repairs of engines and components. We first consider depot-only models and then two-echelon models. We begin by providing an overview of our two basic modeling types.

Our first modeling type is a queueing model in which there are many engines requiring repair at a single location, which we call a depot. These engines arrive for repair according to a non-homogeneous Poisson process. In this model, we assume all engines begin the repair process upon their arrival. More generally, we assume there are an infinite number of servers in all parts of the repair facility. While this assumption is obviously unrealistic, the model provides a way to determine the time-varying requirements for many resource types. Service times are assumed known for each engine in each workcenter in this depot-only model. Our goal is to determine the probability distribution for the number of engines in each portion of the repair process at any point in time. As we will see, these distributions can be used to estimate the probability distribution of demand for resources at a point in time, including components. We then extend this model to represent the flows of serviceable and reparable engines in a two-echelon environment consisting of a depot and a set of bases. Engines fail at bases according to a non-homogeneous Poisson process. Failed engines are repaired at either the bases or depot depending only on the type of maintenance needed to return the engine to a serviceable condition. Repair and transportation times vary with time but are assumed to be known and constant.

The second type of model is a deterministic one. Time is divided into periods. Again we have two types of models. One type is a single location (depot-only) model and the other is a two-echelon system. There are many engines of various types arriving for repairs. But, we also consider the arrival of components requiring repair and the quantities of them needed to complete the repair of engines. Arrival times and quantities for each engine and component type are assumed to be known.We assume capacities for service exist in each workcenter of the repair facility, which can vary over time. We construct an optimization model for setting priorities for repair for each engine and component type in each workcenter and for measuring the effect of these capacities on engine repair.

The single location deterministic model can be used to estimate engine flow times and repair facility resource requirements. Using a mechanism to generate many engine and component arrival scenarios, it is possible to approximate the effects of the uncertainty of the arrival processes in our deterministic model. As a consequence, it is possible to construct probability distributions for waiting times and resource requirements. The output of the deterministic model can be used to estimate the length of the service times in the stochastic models.

We then extend the deterministic depot model to a two echelon environment, say, a depot and a set of intermediate repair facilities supported by the depot. There are many variations on the model we present. The methods we develop here can be modified easily to represent alternative environments. In the model we discuss in detail, we focus on component or module repair at the depot and engine assembly and test at the intermediate repair facilities. The model's output will determine (1) what components or modules to repair at the depot in each time period, (2) which intermediate repair facilities should receive modules or components from serviceable depot stock in each time period, and (3) which engines to assemble and test in each time period at each intermediate repair facility.

Ultimately, once resource levels are established, a more detailed simulation model could be employed to gain greater insights into the system's time dependent behavior for both the single and multiple location systems.

We now discuss these two types of models in detail.

2 Model 1 - Non-homogeneous Engine Arrival Process

In this section we develop two types of continuous time models. The first type is a single location model, which we call the depot model. In section 2.2 we extend the analysis to a two-echelon setting consisting of a depot and set of operating locations, which we will call bases.

2.1 Single Location Models

Let us now construct the first depot model. In this model, we consider only engines. We assume that there are K engine types arriving for repair and that each engine type has a specified sequence of workcenters to be visited within the repair facility. There is no uncertainty in the sequence of workcenters to be visited for any engine entering the repair facility. The sequence is determined by the repairs that are required. Hence, K could be a large number. More will be said about the definition of an engine type subsequently.

We assume that the arrival process for engines of type k is a non-homogeneous Poisson process. Such a process arises based on a set of mathematical assumptions, which we will summarize shortly.

For ease of exposition, we begin by assuming that there is only one work-center in the repair facility. Let

- $a_k(t)$ be the arrival rate for engines of type k at time t to the repair facility, where $a_k(t)$ is an integrable function of t,
- $m_k(t) = \int_0^t a_k(u)du$, the expected number of arrivals of engines of type k during the interval $[0, t]$, and
- $N_k(t)$ be the the total (random) number of engines of type k arriving for repair over the interval $[0, t]$.

The process $\{N_k(t), t \geq 0\}$ is a non-homogeneous Possion process with rate function $a_k(t)$ provided that

1. $N(0) = 0$,
2. $N(t), t \geq 0$ has independent increments,
3. $P\{\text{two or more events occurring in } (t, t+h)\} = o(h)$, and
4. $P\{\text{exactly one event occurring in } (t, t+h)\} = a_k(t) \cdot h + o(h)$.

(Note: a function f is said to be $o(h)$ if $\lim_{t \to 0} \frac{f(t)}{t} = 0$.) Based on these assumptions we show in the Appendix that

$$P[N_k(t) = n] = e^{-m_k(t)} \frac{(m_k(t))^n}{n!}.$$

Next, let

- $X_k(t)$ be a random variable indicating the number of type k engines in repair at time t, and
- $G_k(t)$ be the cumulative distribution function for the length of time an engine of type k is in the repair facility.

Note that we implicitly assume that the length of time in the repair facility for an engine is independent of the number of engines of all types that are present at the time that the engine arrives or of engines that will arrive during that engine's repair process. This is equivalent to assuming that there are an infinite number of servers available at all times to repair engines.

We also assume that $G_k(t) = 0$ for $t \leq g_k$ and $G_k(t) = 1$ for $t \geq b_k$. Thus the time required to repair an engine of type k is in the interval $[g_k, b_k]$ with probability one.

Given the definition of $G_k(t)$, we redefine $N_k(t)$ to be the number of type k engines that have arrived during the interval $[t - b_k, t]$.

Let us now determine the probability distribution for the random variable $X_k(t)$, the number of type k engines in the repair facility at time t. To do this we use a conditional probability argument. Suppose $N_k(t) = n_k$. Consider any one of these arrivals. The probability density function for the time u during the interval $[t - b_k, t]$ that this engine arrived is $a_k(u)/m_k(t)$, for all $u \in [t - b_k, t]$. This can be seen as follows.

Suppose an engine arrives during the interval $(u, u + h]$. Then that engine did not arrive during the interval $[t - b_k, u]$ nor in the interval $(u + h, t]$. Then the probability that the engine arrived in the interval $(u, u + h)$ given that the engine arrived in the interval $(t - b_k, t]$ is

$$\frac{e^{-\int_{t-b_k}^{u} a_k(y)dy} \cdot a_k(u) \cdot h \cdot e^{-\int_{u+h}^{t} a_k(y)dy}}{m_k(t)e^{-m_k(t)}},$$

where $m_k(t) = \int_{t-b_k}^{t} a_k(s)ds$.To obtain the density function, we divide by h and take the limit as $h \to 0$, which yields the desired result.

Clearly

$$P\left[X_k(t) = j\right] = \sum_{n_k \geq j} P\left[X_k(t) = j | N_k(t) = n_k\right] \cdot P\left[N_k(t) = n_k\right].$$

We know that

$$P\left[N_k(t) = n_k\right] = e^{-m_k(t)} \frac{(m_k(t))^{n_k}}{n_k!}.$$

For any of the n_k arrivals, we determine the probability that an engine remains in the system, which we denote by p_k, as follows. Suppose the engine arrived at time $u \in [t - b_k, t]$. Then the probability that the engine remains in the repair facility at time t is $1 - G_k(t - u)$. Hence the unconditional probability an engine arriving for repair in the interval $[t - b_k, t]$ remains in repair at time t is

$$p_k = \int_{t-b_k}^{t} (1 - G_k(t - u)) \cdot \frac{a_k(u)}{m_k(t)} du.$$

Since repair times are independent of each other,

$$P\left[X_k(t) = j | N_k(t) = n_k\right] = \begin{pmatrix} n_k \\ j \end{pmatrix} p_k^j (1 - p_k)^{n_k - j}.$$

Combining these observations yields

$$P\left[X_k(t) = j\right] = \frac{(m_k(t) \cdot p_k)^j \, e^{-p_k m_k(t)}}{j!}.$$

Thus the total number of type k engines in repair at time t is Poisson distributed. Furthermore, since the arrivals of engines of different types occur independently, and repair times occur independently, the total number of engines in repair is also Poisson distributed! Thus

$$P\left[\sum_k X_k(t) = j\right] = \frac{e^{-\sum_k p_k m_k(t)} \left(\sum_k p_k m_k(t)\right)^j}{j!}.$$

Suppose $C_k(t)$ is a random variable that denotes the number of engines of type k that arrived during $(t - b_k, t]$ that have completed their repair by time t. Again, using a conditional probability approach, we can see that

$$P\left[C_k(t) = c\right] = e^{-m_k(t)(1-p_k)} \frac{\left(m_k(t)\left(1-p_k\right)\right)^c}{c!}.$$

We now consider a special case. Suppose the repair facility still consists of a single workcenter and that the length of time to complete repair for an engine of type k in this center is a constant, L_k. We assume this time represents the sum of the time to repair the engine and any delay time. In this case $m_k(t) = \int_0^{L_k} a_k(t-u)du$ and

$$P\left[X_k(t) = j\right] = \frac{\left(m_k(t)\right)^j e^{-m_k(t)}}{j!}.$$

Next suppose all arriving engines begin their repair in a tear down workcenter (TD). Following tear down, each engine follows a known sequence of repair stages and workcenters in which work is performed. We assume all engines of type k, $k = 1, ..., K$, have the same repair path before departing the repair facility. Thus engine types can refer to different engines (eg. F100) and/or to different types of repair required to return the engine to a desired operational status. An operational state reflects the expected number of operating hours until the next off-aircraft repair will be required for the engine.

Let I represent the set of workcenters providing repair and let L_{ik} be the length of time an engine of type k is in repair in workcenter i. Again this time includes an estimate of repair and delay or waiting time. To motivate the general ideas, let us assume engines of type k go to a specific workcenter, W, following work in the TD. Then the arrival rate at time t to this workcenter for these engines is $a_k(t - L_{TD,k})$. The arrival process to workcenter W for type k engines is therefore a non-homogeneous Poisson process as well. Assuming independence across engine types, the arrival process to the workcenter in total across engine types is also a non-homogeneous Poisson process.

Let $X_{Wk}(t)$ represent the random variable measuring the number of engines of type k in workcenter W at time t. Also, let $m_{Wk}(t) = \int_{t-L_{TD,k}-L_{Wk}}^{t-L_{TD,k}} a_k(u)du$. Then the probability that j engines of type k are in workcenter W at time t is

$$P\left[X_{Wk}(t) = j\right] = \frac{\left(m_{Wk}(t)\right)^j e^{-m_{Wk}(t)}}{j!},$$

again a Poisson distribution. Assuming independence of the arrival processes among engine types, the total number of engines in workcenter W at time t is Poisson distributed with mean $\sum_k m_{Wk}(t)$.

We can extend this type of analysis to any workcenter. To do so requires calculating $m_{ik}(t)$ for each location and each engine type. Since each engine type may follow a different path through the repair facility, we must know what locations are visited by each engine and in what sequence. For example, suppose an engine of type k enters workcenter A but has been in workcenters in set I_{Ak} before entering A. Then

$$m_{Ak}(t) = \int_{t-\sum_{i\in I_{Ak}} L_{ik}-L_{Ak}}^{t-\sum_{i\in I_{Ak}} L_{ik}} a_k(u)du.$$

Knowing these $m_{ik}(t)$ values permits us to calculate the probability distribution for the number of engines in each location by engine type and in total.

To this point we have assumed that the length of time an engine of type k is in workcenter i does not depend on the time the engine arrives to the repair facility. The amount of congestion in a repair facility may vary over time. Thus the length of time an engine is in a workcenter may be time dependent. Both repair and delay times may be time dependent. Although we are assuming an infinite capacity exists throughout time at each workcenter in our mathematical model, we will want to account for these time dependent delays, at least approximately. There are several ways to do so. One way is to revise the definition of an engine type. In addition to having a well defined repair path through the repair facility, engines of type k also will have a time frame during which they arrive for repair. Thus $a_k(t)$ is positive only over a well defined interval $[t_{k1}, t_{k2}]$. L_{ik} is the sum of the repair and delay times for a redefined engine of type k at workcenter i. The probability distribution corresponding to the number of engines of type k in workcenter i at time t remains a Poisson distribution. The probability that there are no engines of that type in a workcenter will be 1 for many workcenters and many points in time, however, because of our definition of an engine type.

The reason for wanting to know, even approximately, these distributions is so that we can assess the quantities of each resource type that will be needed to repair engines in each workcenter over time. Since requirements may differ by engine type, these distributions can be particularly useful. Note that if there is a probability distribution associated with the quantity of a resource needed to repair each engine of type k, then the total amount of that resource that is needed at a point in time to serve engines of type k follows a compound Poisson distribution. Thus it is possible to compute the mean and variance of the total amount of each resource type that is required in each workcenter over time, where the total is over all engine types.

There are two types of resources that are likely to be required to perform the tasks undertaken in a workcenter for each engine type. One type of resource consists of equipment, machines, and workers. These resources are ones that cannot store their capacity over time. That is, if a machine or worker is idle during an hour, then that hour of time is lost. It cannot be used to meet needs in the future. A second type of resource is parts used in the repair process such as engine components or subassemblies. If they are not consumed previously, they can be used at any future time to assist in completing repair tasks in an appropriate workcenter.

Suppose E_{rik} is the rate at which a resource r of the first type is consumed while an engine of type k is being worked on while in workcenter i. Then the probability distribution for the amount of resource of type r being consumed at time t by engines of type k in workcenter i is computed as follows. Let $R_{rik}(t)$ be the random variable corresponding to this quantity at time t. Then $P\{R_{rik}(t) = l\} > 0$ only if there are x engines of type k in workcenter i at time t where $x \cdot E_{rik} = l$ or $x = l/E_{rik}$, where x is a non-negative integer. We know that the number of engines of type k in workcenter i is Poisson distributed. So we can easily compute distributions for $R_{rik}(t)$. We would also want to know the distribution of $R_{ri}(t) = \sum_k R_{rik}(t)$. Unless the values of the E_{rik} are the same across engine types, determining the probability distribution for $R_{ri}(t)$ requires a considerably greater amount of computation, even assuming independence across engine types. Another approach is to assume the amount of time required of resources of type r in workcenter i is a random variable which depends only on the total number of engines of all types in workcenter i. Then we compute the distribution $R_{ri}(t)$ as follows.

$$P\{R_{ri}(t) \leq e\} = \sum_n P\{R_{ri}(t) \leq e | X_i(t) = n\} P\{X_i(t) = n\}$$

where $X_i(t)$ is the random variable for the number of engines in workcenter i at time t, which is Poisson distributed. The probability $P\{R_{ri}(t) \leq e | X_i(t) = n\}$ is problematic. For example, it could be represented by a triangular distribution whose parameters will depend on n as well as the range of time over which the density function is positive. Without having data, ascertaining an appropriate form for this distribution is not possible, unfortunately. Ascertaining an appropriate form for this distribution requires access to data and further study.

For the second type of resource, we would construct a cumulative distribution of the number of engines of each type that would enter a workcenter by time t in the planning horizon. Suppose there are n_{qk} components of type q on an engine of type k that are inspected and possibly removed from the engine in a workcenter. Suppose p_{qk} is the probability that a component of type q will need to be removed from an engine of type k. Assuming independence among individual components, the number of components that would be removed would be binomially distributed. Using the same argument as presented previously, the cumulative number of each component type being removed from engines of type k through time t can be determined. When $n_{qk} = 1$, then this probability distribution is a Poisson distribution. When $n_{qk} > 1$, then this distribution is a compound Poisson distribution. Assuming independence among engine types, the total number of components of each type that would be required through time t would have a Poisson distribution when $n_{qk} = 1$.

In summary, this single location model provides a computationally tractable way to estimate the number of engines in repair at a point in time and the corresponding resources needed to complete these repairs.

We now extend these ideas to a two echelon system.

2.2 A Two-Echelon Model

In this section we study a particular two-echelon system consisting of a depot and a collection of operating locations or bases supported by the depot. For simplicity, we restrict attention to a single engine type to minimize the notation required. The ideas we present can easily be extended to the multiple engine type case.

Let i denote the system location. When $i = 0$ we refer to the depot. For $i \geq 1$, we are referring to base i. We assume engines fail at base i according to a non-homogeneous Poisson process. Let $a_i(t)$ be the arrival (failure) rate at base i at time t, where $a_i(t)$ is an integrable function. Each failure requires repair. That repair occurs either at the base at which the failure occurred or at the depot. We assume the location at which the repair occurs depends only on the nature of the failure. We assume the probability that a failed engine at base i at time t will be repaired there is $r_i(t)$. Thus $1 - r_i(t)$ is the probability that the failed engine will be repaired at the depot.

We assume stock levels exist for each location. We assume these stock levels are known and constant over the planning horizon for each location. By a stock level, we mean the amount on-hand plus in base repair plus in depot repair plus in transit to and from the depot. That is, the stock level is the inventory position.

We assume that whenever a failure occurs at a base that a serviceable engine is withdrawn from the base's on-hand serviceable stock and is installed on the failed aircraft. If no such serviceable stock is available, then a backorder will exist.

Suppose an engine fails and is to be repaired at the base. We assume, again for ease of exposition, that the base's repair workcenter is a single stage facility. Once repair of the failed engine is completed, the serviceable engine is placed into the base's serviceable inventory.

Suppose the failed engine requires depot level maintenance. Then the depot is responsible for resupplying the base at which the failure occurred. If the depot has serviceable stock on-hand, then a serviceable engine is sent to the base. At the same time, the failed engine is sent from the base to the depot. If the depot does not have serviceable stock on-hand, then a depot backorder occurs. In our model, we assume the depot resupplies bases on a first-come-first-serve basis.

For simplicity, we assume that there is a single work center for repairing engines in the depot. (We develop a model in Section 3.2 that considers multiple phases to engine repairs in workcenters.) Also, we define the depot repair cycle time for an engine to be the time from the failed engine's removal from the aircraft until it completes repair at the depot.

Based on these and further assumptions, our goal is to construct the probability distributions for the number of engines in resupply, that is, the number of engines in base repair or due to arrive from the depot at each base i at each instant t.

We now introduce additional notation and state some additional assumptions.

Let $L_i(t)$ represent the base i repair cycle time at time t, a constant, where we assume $L_i(t) + t \geq L_i(s) + s$, $s < t$,

$A_i(t)$ represent the order and ship time for an engine from the depot for a failure occurring at base i at time t, which is known and deterministic, where we assume $A_i(t) + t \geq A_i(s) + s$, $s < t$,

$D(t)$ represent the depot repair cycle time for a failure occurring at any base at time t, where $D(t) + t \geq D(s) + s$, $s < t$,

$a_0(t) = \sum_i (1 - r_i(t)) a_i(t)$, the rate at which engine failures occur that require depot repair,

s_i represent the stock level at location i ($i = 0$ is the depot),

$B_0(t : s_0)$ represent the expected number of backorders at the depot at time t given the depot stock level s_0,

$B_i(t : s_i)$ represent the expected number of backorders at base i at time t given the base's stock level s_i,

$X_0(t)$ be a random variable describing the number of engines in the depot's repair cycle at time t, and

$X_i(t)$ be a random variable describing the number of engines in resupply at base i that are backordered at the depot at time t.

We have assumed that both the depot and base repair cycle times and the depot to base order and ship times are known constants, but are time dependent. This assumption is clearly restrictive. But it is important to note that these parameter values are time dependent. Thus it is possible to represent portions of time when there is no repair capacity available at a location, no ability to ship failed engines to a depot or no capability to receive serviceable ones from the depot. Note that our assumptions imply that there is no crossing of repair cycles or resupply times.

Let us now focus on the depot.

2.2.1 Depot Analysis

Our goal in this section is to determine the probability distribution for $X_0(t)$, the random variable representing the number of engines in the depot's repair cycle at time t.

Suppose we consider an arbitrary point in time t. Let $\widetilde{t} = \inf\{u : D(u) + u > t\}$. All demands for resupply for failed engines that must be resupplied to bases by the depot that occurred prior to time $\widetilde{t}$ will have been shipped to bases by time t. Furthermore, all failed engines entering the depot repair cycle subsequent to time $\widetilde{t}$ are still in the depot's repair cycle. This is the case since if $\overline{u} > u > \widetilde{t}$, then $D(\overline{u}) + \overline{u} > D(u) + u > \widetilde{t}$. Thus $P[X_0(t) = k] = e^{-m_0(\widetilde{t},t)} m_0(\widetilde{t}, t)^k / k!$, where $m_0(\widetilde{t}, t) = \int_{\widetilde{t}}^{t} \sum_i a_i(u)(1 - r_i(u)) du$. When $k > s_0$, then backorders exist at time t. The expected number of depot backorders at time t is

$$B_0(t : s_0) = \sum_{k > s_0} (k - s_0) P[X_0(t) = k].$$

Recall that our goal is to determine the probability distribution for the random variable $X_i(t)$. Suppose $X_i(t) = k$, $k > 0$. For this to occur, there exists a $u \in \left(\widetilde{t}, t\right)$ such that the total requests for engines from the depot in the interval $\left(\widetilde{t}, u\right)$ is $s_0 - 1$, there is a replenishment demand placed on the depot at time u, and there are k demands for depot resupply from base i during the interval $(u, t]$. All s_0 demands for resupply of serviceable engines by bases that arose in $(\widetilde{t}, u]$ will have been shipped to bases by time t (based on our first-come-first-serve policy assumption). But, all resupply requests placed subsequent to time $u \in \left(\widetilde{t}, t\right)$ will be backordered at time t. Thus, for $k > 0$,

$$P\left[X_i(t) = k\right] = \int_{\widetilde{t}}^{t} e^{-m_0(t,u)} \frac{m_0(\widetilde{t}, u)^{s_0-1}}{(s_0-1)!} a_0(u) e^{-m_i(u,t)} \frac{m_i(u,t)^k}{k!} du$$

where $m_i(u,t) = \int_u^t \left(1 - r_i(v)\right) a_i(v) dv$.

Note that $X_i(t) = 0$ if one of the following situations arises. First, $X_0(t) < s_0$ and second, depot demand in $(\widetilde{t}, u)$ is $s_0 - 1$, a base failure requiring depot repair occurs at time u, $u \in (\widetilde{t}, t)$ and there are no base i engine failures in $[u, t]$ that require depot repair. Therefore,

$$\begin{aligned} P\left[X_i(t) = 0\right] &= P\left[X_0(t) < s_0\right] \\ &\quad + \int_{\widetilde{t}}^{t} e^{-m_0(t,u)} \frac{m_0(\widetilde{t}, u)^{s_0-1}}{(s_0-1)!} a_0(u) e^{-m_i(u,t)} du. \end{aligned}$$

2.2.2 Base Analysis

We now turn our attention to the bases. Our goal is to establish the probability distribution for the random variable $\widetilde{X}_i(t)$, which represents the total number of engines in resupply at time t.

Let $\bar{t} = \inf\left\{u : A_i(u) + u > t\right\}$ and $\bar{\bar{t}} = \inf\left\{u : L_i(u) + u > t\right\}$. Furthermore, let $X_i^b(t)$ represent the random variable that corresponds to engine failures at base i occurring in $(\bar{\bar{t}}, t]$ that require base level repair and $X_i^d(t)$ represent the random variable that measures the number of engines that require depot repair that fail at base i during $(\bar{t}, t]$. Then

$$\widetilde{X}_i(t) = X_i^b(t) + X_i^d(t) + X_i(\bar{t}).$$

Since the random variables are independent, $X_i^b(t) + X_i^d(t)$ has a non-homogeneous Poisson distribution with mean

$$\widetilde{m}_i(t) = \int_{\bar{t}}^{t} \left(1 - r_i(v)\right) a_i(v) dv + \int_{\bar{\bar{t}}}^{t} r_i(v) a_i(v) dv.$$

Then

$$\begin{aligned} P\left[\widetilde{X}_i(t) = k\right] &= \sum_{j=0}^{k} P\left[X_i^b(t) + X_i^d(t) = j\right] \cdot P\left[X_i(\bar{t}) = k - j\right] \\ &= \sum_{j=0}^{k} e^{-\widetilde{m}_i(t)} \frac{\widetilde{m}_i(t)^j}{j!} \cdot P\left[X_i(\bar{t}) = k - j\right] \end{aligned}$$

where we showed how to determine $P\left[X_i(\bar{t}) = k - j\right]$ in Section 2.2.1.

We can now determine the expected number of backorders at base i at time t as follows.

$$B_i\left(t : s_i\right) = \sum_{k > s_i} \left(k - s_i\right) P\left[\widetilde{X}_i(t) = k\right].$$

Note that these calculations are much more complex than those presented in Section 2.1. However, this analysis is important to carry out since the relationship between the parameters ($a_i(t)$, $A_i(t)$, $r_i(t)$, and $L_i(t)$) and stock levels (s_i) and system performance need to be established. While we have considered the resupply times to be time varying, we need to determine what these values might be given that capacity limitations will determine them. We show how these values might be estimated using the type of analysis developed in the subsequent section.

3 Model Type 2 - Deterministic Flow Models

In our discussion pertaining to the models developed in Section 2, we made two critical assumptions. First, we assumed that the arrival process for each engine type to the depot was a non-homogeneous Poisson process and that these processes were independent among engine types. Second, we assumed we knew the repair time with certainty for each engine type in each depot or base workcenter (the infinite capacity assumption). We now present a modeling approach that differs significantly from our previous one. Finally, we will describe how the models developed in this and the preceding sections can provide insights about the dynamic behavior of the system.

We begin by stating the key assumptions underlying this model. In Section 2, we viewed the system's behavior continuously through time. In the models we develop in this section we assume time is divided into periods, perhaps periods of a day in duration. In Model 1, the engine arrival process for each of the K engine types was assumed to be a non-homogeneous Poisson process. Here we assume that the number of engines of each type arriving for repair is known in each period. As before, we assume the path through the repair facility for each engine type is known. In this model we assume the time an engine is in the repair facility is the sum of its repair time and delay time. We assume we know the repair times for each engine type in the appropriate workcenter. We use the model to determine the delay times. In Model 1 we implicitly estimated this sum in order to set the values of L_{ik}. In Model 1 we assumed infinite capacity existed everywhere. Here we assume capacity is limited in each workcenter in each time period. We also assume component stocks are known and, initially, in time-dependent limited amounts. Then we show how to include component repair in the model. Finally, we extend these latter ideas to a two-echelon repair and assembly system.

To begin our model development, let us assume the depot repair facility consists of only two workcenters, a tear down workcenter TD and a workcenter W. All engines are first worked on in TD. After that, they are transferred to workcenter W. Following repair in W, they are placed in the inventory of serviceable engines. For simplicity, let us first assume that there is only one type of engine. Note, as mentioned, we also do not consider component repair decisions at this point, but will do so subsequently.

Let us now present some nomenclature. Let

- a_t be the number of arrivals to the TD in period t,
- L_{TD} be the tear down time in the TD,
- L_W be the repair time in workcenter W,
- $C_{TD,t}$ be the capacity in TD in period t,
- $C_{W,t}$ be similarly defined for workcenter W,
- $X_{TD,t}$ be the number of engines entering TD in period t,
- $W_{TD,t}$ be the number of engines waiting to enter TD at the end of period t,
- $X_{W,t}$ be the number of engines entering workcenter W during period t,
- $I_{TD,t}$ be the number of engines in TD at the end of period t,
- $I_{W,t}$ be the number of engines in repair in W at the end of period t,
- $Z_{TD,t}$ be the number of engines completing TD work in period t,
- $J_{TD,t}$ be the cumulative number of engines completing TD work through period t,
- b_q be the number of components of type q that are required to repair an engine, and
- d_{qt} be the number of units of component q that first become available in period t,

Given these definitions, we can represent the system's behavior over time using the following relationships and constraints:

$$0 \leq I_{TD,t} = I_{TD,t-1} + X_{TD,t} - Z_{TD,t} \leq C_{TD,t} \tag{1}$$

$$W_{TD,t} = W_{TD,t-1} + a_t - X_{TD,t} \geq 0 \tag{2}$$

$$Z_{TD,t} = X_{TD,t-L_{TD}} \tag{3}$$

$$J_{TD,t} = \sum_{j=0}^{t} Z_{TD,j} \tag{4}$$

$$J_{TD,t} - \sum_{j=1}^{t} X_{W,j} \geq 0 \tag{5}$$

$$I_{W,t} = I_{W,t-1} + X_{W,t} - X_{W,t-L_W} \le C_{W,t}. \tag{6}$$

Let us now include the effect of component inventories on the engine repair plan. Suppose X_{it} is the number of engines that enter workcenter i in period t. Then there must be enough components on hand to repair the engines. Thus we must include constraints that ensure we have enough of each needed component on hand before an engine begins repair in a workcenter. These constraints are material balance constraints. Recall that b_q units of component q are required to complete the repair of an engine. Let Q be the set of all component types. Assume Q_i is the set of components required to execute repairs in workcenter i. We assume $Q_i \cap Q_j = \varnothing$, for all distinct workcenters i and j, that is, components of certain types are used only in a single workcenter.

Suppose there are c_{q0} units of component q on hand at the beginning of period 1. Also, recall that d_{qt} is the number of these components that will first become available for use in period t. If X_{it} engines are to enter repair in workcenter i in period t, then $b_q X_{it}$ units of component q must be on hand to accomplish the repair. Hence for each component q and time period t we have the following material balance equations:

$$\begin{aligned} c_{qt} &= c_{q,t-1} + d_{qt} - b_q X_{TD,t} \ge 0, \text{ for } q \in Q_{TD}, \qquad (7) \\ c_{qt} &= c_{q,t-1} + d_{qt} - b_q X_{W,t} \ge 0, \text{ for } q \in Q_W. \end{aligned}$$

Also,

$$X_{W,t}, X_{TD,t}, c_{qt} \ge 0. \tag{8}$$

Observe that $J_{TD,t} - \sum_{j=1}^{t} X_{W,,j}$ is the number of engines waiting to enter workcenter W at the end of period t. Engines wait either because capacity is limited in W or there are too few components available to carry out the repair.

In Model 1 we focused solely on ascertaining the uncertainty associated with the number of engines in each workcenter. In Model 2 our objective is to determine a repair strategy that minimizes the effect associated with delay in initiating repair in each workcenter.

Suppose at the end of period 0 that there are a number of engines awaiting entry into TD ($W_{TD,0}$), a number of engines in TD that will complete work there at known times ($Z_{TD,t}$ for $t = 1, ..., L_{TD} - 1$), a number of engines waiting to enter workcenter W ($Z_{TD,0}$), and a number of engines in repair in W ($X_{W,0}, X_{W,-1}, ..., X_{W,-L_W+1}$).

Suppose there are inventory targets for the cumulative number of serviceable engines at the end of each period over the planning horizon. These targets are set to ensure a system availability level is achieved. Suppose S_t is the target for period t for the total number of engines that will depart the repair process through period t, that is, engines completing repair incremental to those in serviceable inventory at the beginning of period 1.

The cumulative number of engines completing repair, that is, departing workcenter W, during periods 1 through t is $\sum_{j=-L_W+1}^{t-L_W} X_{W,j}$.

Suppose for engines below the target value S_t there is an expected increase in the number of aircraft engine backorders, that is, aircraft not-operationally ready due to the lack of an engine. Let w_j represent the incremental expected number of backorders if the inventory target is missed by j engines. Thus, w_j measures the expected increment above being $j-1$ engines short. Due to the convexity of the backorder function, $w_{j+1} > w_j$.

Let $U_t \geq S_t - \sum_{j=-L_W+1}^{t-L_W} X_{W,j}$, $U_t \geq 0$, and $\sum_{j\geq 0} u_{jt} = U_t$, where $0 \leq u_{jt} \leq 1$. Then our objective is to determine the number of engines to begin repair in TD and W so as to

$$\text{minimize} \sum_{t\geq 1}\sum_{j} w_j u_{jt} \tag{9}$$

subject to constraints (1) through (8) and

$$\sum_{j\geq 0} u_{jt} \geq S_t - \sum_{j=-L_W+1}^{t-L_W} X_{W,j}, \tag{10}$$

$$0 \leq u_{jt} \leq 1. \tag{11}$$

Since this is the single engine type case, the model forces engines to flow through the repair facility quickly so that the target stock level can be achieved. Suppose we extend this model to the multiple engine type case. This will require making an allocation of capacity among engine types. We now extend our model to the multi-engine type case. Again, we do not consider making component repair decisions.

As we did in Model 1, suppose there are K types of engines arriving for repair. Here we define an engine type solely by the path the engine takes through the repair facility and the components required to complete the repair. Again, to minimize notation and complexity, let us assume that engines of all K types arrive at the TD and then move for further repair into workcenter W.

For each parameter and variable described previously we add a subscript denoting the engine type. For example, a_{tk} is the number of arrivals to the TD of engine type k in period t and $X_{TD,t,k}$ is the number of engines of type k entering TD in period t. Only the capacity parameters $C_{TD,t}$, C_{Wt}, and component stock increments d_{qt} do not contain the additional k subscript. We also may include additional constraints to limit the number of engines of each type in each workcenter in each time period or perhaps for just a subset of engine types and time periods.

Thus in the multi-engine type, TD and workcenter W environment, the constraints are

$$0 \leq I_{TD,t,k} = I_{TD,t-1,k} + X_{TD,t,k} - Z_{TD,t,k} \leq C_{TD,t,k}, \tag{12}$$

(the number of engines of type k is constrained in the TD in period t)

$$\sum_{k} \alpha_k I_{TD,t,k} \leq C_{TD,t}, \tag{13}$$

(where α_k is the amount of capacity consumed by an engine of type k per period in the TD)

$$W_{TD,t,k} = W_{TD,t-1,k} + a_{t,k} - X_{TD,t,k} \geq 0, \tag{14}$$

$$Z_{TD,t,k} = X_{TD,t-L_{TD,k},k}, \tag{15}$$

$$J_{TD,t,k} = \sum_{j=0}^{t} Z_{TD,j,k} \geq 0, \tag{16}$$

$$J_{TD,t,k} - \sum_{j=1}^{t} X_{W,j,k} \geq 0, \tag{17}$$

(the number of engines of type k that have completed processing in the TD but have not begun processing in W).

$$0 \leq I_{W,t,k} = I_{W,t-1,k} + X_{W,t,k} - X_{W,t-L_{W,k},k} \leq C_{W,t,k}, \tag{18}$$

$$\sum_k \beta_k I_{W,t,k} \leq C_{W,t}, \tag{19}$$

(where β_k is the amount of capacity consumed by an engine of type k per period in workcenter W)

$$c_{qt} = c_{q,t-1} + d_{qt} - \sum_k b_{qk} X_{TD,t,k} \geq 0, \text{ for } q \in Q_{TD}, \tag{20}$$

$$c_{qt} = c_{q,t-1} + d_{qt} - \sum_k b_{qk} X_{W,t,k} \geq 0, \text{ for } q \in Q_{W}, \tag{21}$$

$$X_{W,t,k}, X_{TD,t,k} \geq 0. \tag{22}$$

Note that expressions (13) and (19) could be modified easily to measure the way workcenter capacity consumption changes throughout the time an engine is in a workcenter. That is, α_k and β_k could be made to be time dependent. Also note that additional workcenter capacity constraints can be incorporated into the model to reflect, for example, separate constraints for labor and equipment.

Clearly the only constraints linking the engine types are the capacity and component constraints. If these constraints are inactive, then the problem reduces to solving K independent problems.

As in the single engine type case, suppose there are engines of each type awaiting entry into and in process within the TD and W workcenters at the beginning of period 1. Suppose $W_{TD,0,k}$ is the number of engines of type k waiting to enter the TD workcenter at the beginning of day 1. Also, for engines already in the TD, $Z_{TD,t,k}$ is the number of type k engines that will complete processing there in periods $t = 1, ..., L_{TD,k} - 1$. The number of engines of type k waiting to enter workcenter W at the beginning of period 1 is $Z_{TD,0,k}$. The number of engines of type k in repair in workcenter W initially is equal to $\sum_{j=-L_{W,k}+1}^{0} X_{W,j,k}$.

We again assume there are inventory targets for the cumulative number of serviceable engines of each type to have in stock initially plus completing repair through each period of the planning horizon. Let S_{tk} be the target number of engines of type k that should enter serviceable inventory by the end of period t. This quantity reflects the increment to those in serviceable status at the beginning of period 1.

The total number of type k engines that complete repair and enter the serviceable inventory during periods 1 through t is $\sum_{j=-L_{W,k}+1}^{t-L_{W,k}} X_{W,j,k}$. As in the single engine case, if the cumulative number of engines of type k that enter serviceable inventory is less than S_{tk} through period t, there will be an incremental number of expected system shortages. We define w_{jk} to be the weighted expected increment to system backorders associated with the jth unit short (incremental to being $j-1$ units short of the target) of engine type k. By 'weighted,' we mean that some engine types may have higher priorty than others and hence backorders of those engine types will be assessed a higher 'cost.' Convexity of the backorder function implies that $w_{j+1,k} > w_{j,k}$.

Let $U_{tk} \geq S_{tk} - \sum_{j=-L_{W,k}+1}^{t-L_{W,k}} X_{W,j,k}$, $U_{tk} \geq 0$, and $\sum_{j\geq 0} u_{jtk} = U_{tk}$, where $0 \leq u_{jtk} \leq 1$. Then our objective is to

$$\text{minimize} \sum_{t\geq 1}\sum_{k}\sum_{j} w_{jk}u_{jtk} \tag{23}$$

subject to constraints (12) through (22) and

$$\sum_{j\geq 0} u_{jtk} \geq S_{tk} - \sum_{j=-L_{W,k}+1}^{t-L_{W,k}} X_{W,j,k}, \tag{24}$$

$$0 \leq u_{jtk} \leq 1. \tag{25}$$

This two-workcenter, multi-engine type model can be extended to represent much more general routings of engines through multiple repair centers over time. Although the notation must be adjusted to accommodate the more general case, the resulting model will have the same structure as the one we presented. The expanded model will contain so-called "material balance" constraints which ensure the conservation of flow of engines of a given type through the repair facility. There will be capacity and component availability constraints that force delays to occur. The problem's size and complexity clearly increase but the underlying structure is the same as in the two workcenter case we discussed.

There are many policies that could be used to prioritize the flow through the repair facility. These priorities result in delays. Comparing the implications associated with any policy, such as a first-come-first-served policy, and the optimal one would be of interest in general.

The most important reasons for constructing and solving this linear program and the ones developed later in this paper are (1) to use the solution to infer what resource levels are needed throughout a planning horizon in each workcenter, (2) to determine the dual variable values corresponding to the solution, and

(3) to determine the effect of capacities and component availability on expected operational performance, that is, aircraft related backorders.

Staffing requirements vary throughout the workcenters over time. Hence knowing when engines of each type arrive to a workcenter and when they depart provides a better understanding of how effective a staffing schedule might be.

Equipment are critical to the repair of engines. Also, availability of component stocks over time may result in shortages of serviceable engines.

Thus by analyzing solutions we obtain estimates of the need for all resources. The dual variable values can provide information about the economic effects of increasing these resource levels, that is, capacities or component stock levels. Hence making increments to them can be justified using the dual variable values.

The values of the dual variables are related to the "costs" of delays. Typically, delays can result in an increased number of backorders. Thus these projections of incremental shortages are key to making capacity decisions.

3.1 Combined Engine-Component Repair Model

To this point, our second depot model has focused solely on making decisions about which engine types to enter into each workcenter in each period of the planning horizon given resource and component stock availabilities. We assumed the planning horizon was so short that component repair cycle times exceeded the planning horizon's length. We can easily extend this second model to consider the possibility of component repair over a longer planning horizon. Arrivals of both components and engines requiring repair would occur in each period of the planning horizon. There would be different workcenters for component repair and engine repair. There would be constraints that would limit the number of components of various types that could be processed in each period in the extended model. Thus choices of which components to repair would impact which engines could be repaired subsequently and hence would affect operational performance. The constraints required to model component repair are of the form found in our multi-engine type formulation. The resulting model would be considerably larger, although solvable using commercially available linear programming software. We now present this expanded model, again limiting the number of workcenters for both engines and components to two of each type. As before, we make this assumption solely to limit the notation needed and the exposition required.

We begin by presenting the notation used in this engine-component repair model. Let

- a^e_{tk} be the number of arrivals of engines of type k to the TD in period t,
- a^c_{tq} be the number of arrivals of components of type q in period t requiring repair,
- $L^e_{TD,k}$ be the teardown time for an engine of type k in the TD,
- L^c_{1q} be the repair time for a component of type q in component workcenter 1,

- $L^e_{W,k}$ be the type k engine repair time in workcenter W,
- L^c_{2q} be the repair time for a component of type q in component workcenter 2,
- $C^e_{TD,t}$ be the capacity in TD in period t,
- $C^e_{W,t}$ be similarly defined for workcenter W,
- $C^c_{j,t}$ be the capacity in component workcenter j in period t, $j = 1, 2$,
- $X^e_{TD,t,k}$ be the number of engines of type k entering the TD repair in period t,
- $X^c_{1,t,q}$be the number of components of type q entering component workcenter 1 in period t,
- $W^e_{TD,t,k}$ be the number of engines of type k waiting to enter TD at the end of period t,
- $W^c_{1,t,q}$ be the number of components of type q waiting to enter component workcenter 1 at the end of period t,
- $X^e_{W,t,k}$ be the number of engines of type k entering repair in workcenter W in period t,
- $X^c_{2,t,q}$ be the number of components of type q transferred to component workcenter 2 during period t,
- $I^e_{TD,t,k}$ be the number of engines of type k in repair in TD at the end of period t,
- $I^e_{W,t,k}$ be the number of engines of type k in repair in W at the end of period t,
- $I^c_{j,t,q}$ be the number of components of type q in component workcenter j at the end of period t, $j = 1, 2$,
- $Z^e_{TD,t,k}$ be the number of engines of type k completing repair in TD in period t,
- $Z^c_{1,t,q}$ be the number of components of type q completing repair in component workcenter 1 in period t,
- $J^e_{TD,t,k}$ be the cumulative number of engines of type k completing TD work through period t,
- $J^e_{W,t,k}$ be the cumulative number of engines of type k completing W repair through period t,
- $J^c_{j,t,q}$ be the cumulative number of components of type q completing repair in component workcenter j through period t, $j = 1, 2$,

- $b_{q,k}$ be the number of components of type q needed to repair an engine of type k,and
- c_{qt} be the number of serviceable components of type q on-hand at the end of period t.

We now represent the constraints describing the system's dynamics over the planning horizon.

$$0 \leq I^e_{TD,t,k} = I^e_{TD,t-1,k} + X^e_{TD,t,k} - Z^e_{TD,t,k} \leq C^e_{TD,t,k}, \tag{26}$$

$$\sum_k \alpha_k I^e_{TD,t,k} \leq C^e_{TD,t}, \tag{27}$$

(where α_k is the amount of TD capacity consumed by an engine of type k per period while in TD).

$$0 \leq I^c_{1,t,q} = I^c_{1,t-1,q} + X^c_{1,t,q} - Z^c_{1,t,q} \leq C^c_{1,t,q}, \tag{28}$$

$$\sum_k \gamma_q I^c_{1,t,q} \leq C^c_{1,t}, \tag{29}$$

(where γ_q is the amount of component workcenter 1 capacity consumed by a component of type q per period while in component workcenter 1).

$$W^e_{TD,t,k} = W^e_{TD,t-1,k} + a^e_{t,k} - X^e_{TD,t,k} \geq 0, \tag{30}$$

$$W^c_{1,t,q} = W^c_{1,t-1,q} + a^c_{t,q} - X^c_{1,t,q} \geq 0, \tag{31}$$

$$Z^e_{TD,t,k} = X^e_{TD,t-L^e_{TD,k},k}, \tag{32}$$

$$Z^c_{1,t,q} = X^c_{1,t-L^c_{1,q},q}, \tag{33}$$

$$J^e_{TD,t,k} = \sum_{s=0}^{t} Z^e_{TD,s,k} \geq 0, \tag{34}$$

$$J^e_{TD,t,k} - \sum_{s=1}^{t} X^e_{W,s,k} \geq 0, \tag{35}$$

$$J^c_{1,t,q} = \sum_{s=0}^{t} Z^c_{1,s,q} \geq 0, \tag{36}$$

$$J^c_{1,t,q} - \sum_{s=1}^{t} X^c_{2,s,q} \geq 0, \tag{37}$$

$$0 \leq I^e_{W,t,k} = I^e_{W,t-1,k} + X^e_{W,t,k} - X^e_{W,t-L^e_{W,k},k} \leq C^e_{W,t,k}, \tag{38}$$

$$\sum_k \beta_k I^e_{W,t,k} \leq C^e_{W,t}, \tag{39}$$

$$0 \le I^c_{2,t,q} = I^c_{2,t-1,q} + X^c_{2,t,q} - X^c_{2,t-L^c_{2,q},q} \le C^c_{2,t,q}, \tag{40}$$

$$\sum_k \beta_q I^c_{2,t,q} \le C^c_{2,t}, \tag{41}$$

(where β_k and β_q represent the amount of capacity consumed in the workcenters W and 2 for engines of type k and components of type q, respectively).

$$J^e_{W,t,k} = \sum_{s=1}^{t} X^e_{W,s-L^e_{W,k},k}, \tag{42}$$

$$c_{qt} = c_{q,t-1} + X^c_{2,t-L^c_{2,q},q} - \sum_k b_{qk} X^e_{TD,t,k} \ge 0, \;\; q \in Q_{TD}, \tag{43}$$

$$c_{qt} = c_{q,t-1} + X^c_{2,t-L^c_{2,q},q} - \sum_k b_{qk} X^e_{W,t,k} \ge 0, \;\; q \in Q_{W}, \tag{44}$$

$$X^c_{1,t,q}, X^c_{2,t,q}, X^e_{TD,t,k}, X^e_{W,t,k} \ge 0. \tag{45}$$

As before, we assume inventory targets exist for the cumulative number of engines of each type to complete repair through each period t of the planning horizon. Again, let S_{tk} be this target for engines of type k in period t and $U_{tk} \ge S_{tk} - \sum_{j=-L^e_{W,k}+1}^{t-L^e_{W,k}} X^e_{W,j,k}$, $U_{tk} \ge 0$, $\sum_{j\ge 0} u_{jtk} = U_{tk}$, and $0 \le u_{jtk} \le 1$. As earlier, our objective is to minimize

$$\sum_{t\ge 1} \sum_k \sum_j w_{jk} u_{jtk} \tag{46}$$

subject to constraints (26) through (45) and

$$\sum_{j\ge 0} u_{jtk} \ge S_{tk} - \sum_{j=-L^e_{W,k}+1}^{t-L^e_{W,k}} X^e_{W,j,k}, \tag{47}$$

$$0 \le u_{jtk} \le 1. \tag{48}$$

We have constructed the model based on the assumption that the arrival times and quantities for each engine and component type are known. We may have reasonably accurate estimates of these expected times knowing the operating cycles on engines, knowing which engines are on which aircraft, knowing future flying schedules, and knowing budgets for repair. However, the uncertainty in their values should be a part of the analysis. We can measure, at least partially, the effect of uncertainty by creating a large number of arrival scenarios using variance estimates derived from historical data. Each scenario would consist of the number of engines (and components) of each type arriving in each time period. One possibility is to solve the linear program for each scenario.

Rather than solving the linear program for each scenario, we could employ a better approach. Suppose the time horizon is divided into two segments. The first segment consists of periods for which each $a^e_{t,k}$ and $a^c_{t,q}$ values are known with certainty and for which the decisions made in these periods will be executed during each period in this segment. The second segment consists of the remainder of the planning horizon. For ease of discussion, let us assume that the first segment consists only of period 1. Then the decisions made in period 1 for each engine and component type will be executed in period 1. But "decisions" made for all subsequent periods will correspond to the individual scenarios that are being considered. Let us now construct the optimization problem corresponding to this two time segment construction.

Suppose there are S scenarios of future engine arrivals being considered and suppose p_s is the probability that scenario s will occur, $s = 1, ..., S$. For period 1 (again assuming the first segment consists of only a single period, $t = 1$), we have constraints (26) through (45), (47), and (48) with $t = 1$. For each scenario s, $s = 1, ..., S$, let each decision variable have a superscript s associated with it for all periods 2 through the end of the planning horizon. That is, for scenario s and period $t \geq 2$, for example, $I^{es}_{TD,t,k}$, $X^{es}_{TD,t,k}$, $Z^{es}_{TD,t,k}$, $W^{es}_{TD,t,k}$, $J^{es}_{TD,t,k}$, $I^{es}_{W,t,k}$, $X^{es}_{W,t,k}$, $J^{es}_{W,t,k}$, $c^{es}_{t,q}$, and $u^{es}_{j,t,k}$ are decision variables. Also, $a^{es}_{t,k}$ are the engine arrivals in period $t \geq 2$ of type k corresponding to scenario s and $a^{cs}_{t,q}$ are the component q arrivals for repair in period t in scenario s. Then for each scenario, we have an additional set of constraints of the form (26) through (45), plus (47) and (48). Our objective function for the two segment approach is to minimize

$$\sum_k \sum_j w_{jk} u_{j1k} + \sum_s \sum_{t \geq 2} \sum_k \sum_j p_s w_{jk} u^s_{jtk}.$$

This two segment model has potentially a large number of constraints. As stated, there are constraints corresponding to only the first segment. Given the values of the first segment variables, each scenario produces a set of values for the decision variables for each time period $t \geq 2$. These scenario specific variables must satisfy constraints of the form found in (26) through (45), plus (47) and (48) for $t \geq 2$.

Suppose we have a solution to the linear program. For the first time segment, we can estimate the resource requirements corresponding to the solution. That is, given the values of the first segment decision variables, we know exactly how much of each resource type will be needed to execute the plan. We will also know the resource requirements corresponding to each scenario in each time period in the second segment.

Recall that p_s is the probability that scenario s will occur. Suppose the solution to the linear program consumes $r^s_{R,t}$ units of the resource of type R in period t in scenario s. Then, knowing the values of p_s, we can determine the probability distribution for the amount of each resource type required in all time periods in the second segment.

Suppose we have carried out this scenario analysis and have obtained estimates of the mean, variance and distribution of the resource requirements in all periods and the expected costs that would be incurred. Remember how the linear program is constructed. We assume we know when all the arrivals of each engine type will occur throughout the planning horizon for each scenario when we solve the problem. In reality, we do not know exactly what will happen throughout the planning horizon. We may have reasonably good probabilistic estimates, but we do not know precisely what the arrival pattern will be. As a consequence, our estimated operational performance is a lower bound on the true expected number of weighted shortages. In stochastic programming terminology, we have constructed and solved what is called a two-stage "wait and see" type of model. That is, we solved the linear program after we knew the arrival pattern corresponding to each scenario. Nonetheless, this approach will provide a good estimate of the consequences of current decisions on future decisions and system performance. The approach provides useful probabilistic estimates of resource requirements and operational performance. We will also employ this scenario-based, two-stage optimization process to solve the two-echelon model we now present.

3.2 A Depot / Intermediate Repair Facility Model

We now extend the results of the previous section to a two-echelon environment consisting of a depot and a set of intermediate repair facilities or IRFs. There are many possible environments of this type we can represent using the deterministic modeling approach presented in Section 3.1. We will present only one such environment. In this environment, engines requiring repair arrive at a regional IRF from a collection of operating bases. We assume the engine is designed in a modular way, as is the F-100 engine. We assume that failures require the removal of defective modules from the engine, which occurs at an IRF. The defective modules are sent to the depot from the IRF for repair. The engine is ultimately returned to a serviceable state by withdrawing serviceable modules from the IRFs inventory, reassembling the engine, and testing it. The depot repairs failed modules and subsequently allocates the resultant serviceable modules to the IRFs based on need. The depot has limited capacity for each stage of module repair for each module type. Hence determining which modules to enter each repair stage in each period of the planning horizon impacts the depot's ability to respond to IRF needs.

A second depot decision establishes which serviceable modules of each type to send to each IRF in each period.

In each period at each IRF a decision is made establishing which engines to begin the assembly and testing process. Assembly and testing capacities are limited at an IRF. Module stock levels are also limited. Thus not all reparable engines may be entered into these processes every day.

Finally, failed engines arriving at an IRF must be disassembled to a level that depends on which modules have failed. This disassembly time varies by which module or modules have failed. Disassembly capacity is also limited in each time period. Consequently, there is another decision to be made: which engines should begin the disassembly process each day.

The proposed model will address each of the decisions we have mentioned. The goal is to make these decisions so as to meet serviceable engine inventory targets in each period of the planning horizon. As before, the model is planned to be executed daily. Only the immediate period's decisions would be implemented. Future "planned" decisions will be altered as more information about the system's dynamics is revealed.

3.2.1 Assumptions and Notation

We begin our model development by stating additional assumptions and presenting notation.

Let us focus first on the depot. We assume each failed module goes through three workcenters: teardown (TD), assembly (A), and test (T). Let's assume that modules do not share repair capacity. This is not a critical assumption, however. Let m represent a module type and M be the number of modules in the engine.

Let $a^d_{t,m}$ represent the number of modules of type m that arrive for repair at the depot in period t,

$L_{TD,m}$ represent the number of periods required to tear down a module of type m,

$L_{A,m}$ represent the number of periods required to assemble a module of type m,

$L_{T,m}$ represent the number of periods required to test an assembled module of type m,

$C_{TD,t,m}$ represent the maximum number of modules of type m that can be in the teardown process in period t,

$C_{A,t,m}$ and $C_{T,t,m}$ are defined similarly.

Given the values of these parameters, we must define decision variables that represent how many modules of each type that should enter and complete each of the repair stages in each period.

Let $X_{j,t,m}$ represent the number of modules of type m to enter repair phase j, $j = TD, A, T$, in period t,

$W_{j,t,m}$ represent the number of modules of type m waiting to enter repair phase j, $j = TD, A, T$, at the end of period t,

$I_{j,t,m}$ represent the number of modules of type m in repair phase j, $j = TD, A, T$, at the end of period t,

$Z_{j,t,m}$ represent the number of modules of type m completing repair phase j, $j = TD, A, T$, in period t, and

$c_{t,m}$ represent the number of modules of type m that are in serviceable depot stock at the end of period t.

Next we define parameters and variables corresponding to the allocation decisions made for each module. That is, the quantity of each module that will be shipped to IRF i in period t.

Let $Y_{i,t,m}$ represent the number of serviceable modules of type m to ship to IRF i in period t, and

l^d_i represent the depot-to-IRF i shipping delay time.

We now represent the constraints pertaining to the depot's system dynamics over the planning horizon.

The first set of constraints are material balance constraints that indicate the number of modules of type m in each depot repair phase at the end of each period.

$$0 \le I_{j,t,m} = I_{j,t-1,m} + X_{j,t,m} - Z_{j,t,m} \le C_{j,t,m},\ j = TD, A, T. \qquad (49)$$

The next set of constraints are also material balance constraints that indicate the number of modules of type m that are waiting to enter depot repair phase j, $j = TD$, A, and T.

$$0 \le W_{TD,t,m} = W_{TD,t-1,m} + a^d_{tm} - X_{TD,t,m} \qquad (50)$$

$$0 \le W_{A,t,m} = W_{A,t-1,m} + Z_{TD,t,m} - X_{A,t,m} \qquad (51)$$

$$0 \le W_{T,t,m} = W_{T,t-1,m} + Z_{A,t,m} - X_{T,t,m} \qquad (52)$$

where

$$Z_{TD,t,m} = X_{TD,t-L_{TD,m},m} \qquad (53)$$

$$Z_{A,t,m} = X_{A,t-L_{A,m},m}. \qquad (54)$$

Note: $W_{j,0,m}$ represents the system's initial conditions at the beginning of period 1, $j = TD, A, T$.

Finally, we represent the constraint corresponding to the inventory of serviceable inventory of module m at the depot at the end of period t.

$$c_{t,m} = c_{t-1,m} + Z_{T,t,m} - \sum_i Y_{i,t,m}. \qquad (55)$$

We now turn our attention to the IRFs and their relation to the depot.

We define engine failures according to the module(s) that must be removed from the engine. We define an engine failure to be of type k where this type corresponds to a particular module or combination of modules that must be removed from the engine which are ultimately sent to the depot for repair. Thus for each failed engine there is a set of modules m that require depot level repair.

We assume each failed engine's maintenance at an IRF is performed in three phases: teardown, assembly, and test (TD, A, and T).

We first define parameters used in the IRF portion of our model.

Let $a^e_{i,t,k}$ represent the number of engine failures of type k that arise in period t at IRF i,

$L^e_{j,i,k}$ represent the time required to perform phase j maintenance for an engine of type k at IRF i, $j = TD$, A, and T,

$C^e_{j,t,i}$ represent the maximum number of engines of all types that can be in maintenance phase j at IRF i in time period t,

l^i_d represent the shipping time for reparable modules to the depot from IRF i,

$b_{m,k}$ represent the number of modules of type m that will be removed from a reparable engine of type k and sent to the depot for repair.This is also the number of type m modules required to return an engine of type k to a serviceable condition, and

$c_{t,m,i}$ represent the number of serviceable modules of type m on-hand at the end of period t at IRF i.

We next define decision variables pertaining to the IRF portion of our model.

Let $X^e_{j,i,t,k}$ represent the number of engines of type k that enter phase j of the maintenance process at IRF i in period t where $j = TD$, A, or T,

$W^e_{j,i,t,k}$ represent the number of engines of type k that are waiting to enter phase j of the maintenance process at IRF i at the end of period t, $j = TD$, A, or T,

$I^e_{j,i,t,k}$ represent the number of engines of type k in phase j of the maintenance process at IRF i at the end of period t, $j = TD$, A, or T,

$Z^e_{j,i,t,k}$ represent the number of engines of type k completing repair in phase j at IRF i in period t, $j = TD$, A, or T,

$c^e_{t,i}$ represent the cumulative number of engines that have completed repair at IRF i by the end of period t.

Given the above parameters and decision variables, we now represent the constraints that describe the system's dynamics at the IRFs.

First, we have a material balance equation for engines in each maintenance phase at the end of a period for each type of reparable engine

$$0 \leq I^e_{j,i,t,k} = I^e_{j,i,t-1,k} + X^e_{j,i,t,k} - Z^e_{j,i,t,k}. \tag{56}$$

Second, we have a constraint representing the repair capacity in each phase of the maintenance process at each IRF in each time period

$$\sum_k I^e_{j,i,t,k} \leq C^e_{j,t,i}. \tag{57}$$

Third, we have a material balance equation for modules

$$c_{t,m,i} = c_{t-1,m,i} + Y_{i,t-l^d_i,m} - \sum_k b_{m,k} X^e_{A,i,t,k} \geq 0. \tag{58}$$

Fourth, we have constraints representing the number of engines of type k waiting to enter maintenance phase j at IRF i at the end of period t

$$0 \leq W^e_{TD,i,t,k} = W^e_{TD,i,t-1,k} + a^e_{i,t,k} - X^e_{TD,i,t,k} \tag{59}$$

$$0 \leq W^e_{A,i,t,k} = W^e_{A,i,t-1,k} + Z^e_{TD,i,t,k} - X^e_{A,i,t,k} \tag{60}$$

$$0 \leq W^e_{T,i,t,k} = W^e_{T,i,t-1,k} + Z^e_{A,i,t,k} - X^e_{T,i,t,k} \tag{61}$$

The engines completing phase j repair in a period t equals the number of that type that entered phase j maintenance a repair lead time earlier, that is,

$$Z^e_{j,i,t,k} = X^e_{j,i,t-L^e_{j,i,k},k},\ j = TD,\ A, \text{and } T. \tag{62}$$

Next, the material balance equation for cumulative serviceable engines at each IRF by the end of period t

$$c^e_{t,i} = c^e_{t-1,i} + \sum_k Z^e_{T,i,t,k}, \tag{63}$$

where $c^e_{0,i} = 0$ at IRF i.

Remember that arrivals of reparable modules at the depot depend on these modules being removed from engines in the teardown process at an IRF and shipped to the depot from that IRF. Hence,

$$a^d_{m,t} = \sum_i \sum_k b_{m,k} X^e_{TD,i,t-L^e_{TD,i,k}-l^i_d+1,k}. \tag{64}$$

When $t - L^e_{TD,i,k} - l^i_d + 1 \leq 0$, then the shipment occurred prior to period 1 and the value of $X^e_{TD,i,t-L^e_{TD,i,k}-l^i_d+1,k}$ is a known constant.

Finally, all decision variables must assume non-negative values.

$$X_{j,t,m}, W_{j,t,m}, I_{j,t,m}, c_{t,m}, Y_{i,t,m}, X^e_{j,i,t,k}, W^e_{j,i,t,k}, I^e_{j,i,t,k}, c^e_{t,i}, c_{t.m.i} \geq 0. \tag{65}$$

Other variables will by necessity be non-negative.

As we did in Section 3.1, we assume there are inventory targets for the cumulative number of engines completing repair through each time period of the planning horizon at each IRF. Let $S_{t,i}$ be this target for IRF i in period t. Again we define a variable

$$U_{t,i} \geq S_{t,i} - c^e_{t,i} \tag{66}$$

$$U_{t,i} \geq 0 \tag{67}$$

and variables

$$0 \leq u_{i,j,t} \leq 1 \tag{68}$$

with

$$U_{t,i} = \sum_j u_{i,j,t}. \tag{69}$$

Again, we define w_{ij} to be the weighted expected increment to IRF i backorders associated with the j^{th} unit short of the target (incremental to being $j-1$ units below the target). Engines in some areas of the world may have a higher priority than in other areas. The values of weights w_{ij} reflect the difference in the priorities.

Then our multi-echelon optimization model is

$$\text{minimize} \sum_{t\geq 1} \sum_i \sum_j w_{ij} u_{i,j,t} \tag{70}$$

subject to constraints (49)-(69).

We again propose a scenario-based approach when constructing and solving the above problem. This approach is the same as the one we introduced at the end of Section 3.1.

A scenario consists of knowns and unknowns. The knowns reflect our knowledge of all modules and engines in repair in each of the three maintenance phases at the depot and each IRF. We also know what modules are in-transit to and from the depot and when they will arrive. We further know what serviceable inventories are at each location and how many modules and engines are waiting to enter each maintenance phase at each location. We know what failures are occurring on day 1 of the planning horizon for each engine failure type k. But we do not know with certainty what failures of each type will occur on each future day at each IRF.

Suppose we generate a collection of possible failure scenarios for all future periods, that is, values for the $a^e_{i,t,k}$ parameters. For scenario s we would have $a^{es}_{i,t,k}$ represent the number of engine failures of type k requiring repair arising in period t at IRF i.

As we discussed in Section 3.1 we assume the time horizon is divided into two segments. The first consists of those periods for which we know the values of $a^e_{t,k}$. The second consists of the remainder of the planning horizon. The solution to the resulting linear program will indicate what decisions should be executed in the first segment of the future, recognizing the consequences of future requirements when making these decisions.

The discussion towards the end of Section 3.1 pertaining to the construction of a scenario-based linear program is appropriate here. Let p_s be the probability that scenario s will occur. Suppose the first segment consists of a single period. For each period $t \geq 2$ we have, for example, decision variables $X^{es}_{j,i,t,k}, W^{es}_{j,i,t,k}, I^{es}_{j,i,t,k}, Z^{es}_{j,i,t,k}$, for $j = TD$, A, and T for each IRF and similarly defined scenario dependent variables for modules at the depot. Modifying the variables in the constraints (66)-(69) to reflect the scenarios yields the following objective function

$$\text{minimize} \sum_i \sum_j w_{ij} u_{i,j,t} + \sum_s \sum_{t \geq 2} \sum_i \sum_j p_s w_{ij} u^s_{i,j,t}. \tag{71}$$

Constraints (49)-(69) exist for the first segment ($t = 1$) and one set of constraints (49)-(69) exists for each of the scenarios.

As we stated in Section 3.1, the scenario based approach provides an optimistic estimate, or upper bound, on average system performance and a lower bound on expected backorders. This is the case since each scenario contains explicit arrival patterns for all failure types throughout the planning horizon. Specifically, the timing of these failures is known at the time the period 1 decisions are being determined. Again, this is the so-called two-stage "wait-and-see" type of model. The model's solution should, however, provide good estimates of the system's future performance and useful probabilistic estimates of resource requirements at all locations over time.

4 Final Comments

In this note, we have developed two quite different approaches to modeling the flows of engines through repair facilities. These models could be used together to provide an effective way to represent the dynamics present in flows of engines through repair facilities.

Suppose we construct and solve the scenario-based linear programs. For each time period we get estimates of the total average time an engine spends in a workcenter across scenarios. This total time reflects both repair and delay times. These total times can be used in Model Type 1 to indicate the time dependent length of time an engine is in each workcenter. Hence, the "repair" times used in Model Type 1 incorporate expected delays resulting from capacity limitations and the interactions occurring between engine types and components in each workcenter over time as estimated using Model Type 2.

5 Appendix

Let us now show how to determine the probability distribution for the random variable $N_k(t)$, the total number of engines of type k arriving for repair in $[0, t]$.

Let $P_n(t) = P\left[N_k(t) = n\right]$. Suppose $n = 0$. Then

$$\begin{aligned} P_0(t+h) &= P\left[N_k(t) = 0; N_k(t+h) - N_k(t) = 0\right] \\ &= P\left[N_k(t) = 0\right] \cdot P\left[N_k(t+h) - N_k(t) = 0\right] \end{aligned}$$

(by independent increments). Then

$$P_0(t+h) = P_0(t)P_0(h),$$

where

$$P_0(h) = P\left[N_k(t+h) - N_k(t) = 0\right].$$

Then

$$\frac{P_0(t+h) - P_0(t)}{h} = \frac{P_0(t)P_0(h) - P_0(t)}{h} = -\frac{P_0(t)\left(1 - P_0(h)\right)}{h}$$

By assumption, $P_0(h) = 1 - a_k(t) \cdot h + o(h)$ so

$$\frac{P_0(t+h) - P_0(t)}{h} = -\frac{\left(a_k(t+h) - o(h)\right) P_0(t)}{h}$$

and

$$\frac{dP_0(t)}{dt} = -a_k(t)P_0(t).$$

The solution to this differential equation is

$$P_0(t) = ce^{-m_k(t)}$$

for some constant c. Since $P_0(0) = 1$, $c = 1$.

A similar analysis can be executed for $n \geq 1$. In this case, it is easy to show that

$$\frac{P_n(t+h) - P_n(t)}{h} = \frac{-a_k(t)hP_n(t) + a_k(t)hP_{n-1}(t)) + o(h)}{h}$$

which, upon letting $h \to 0$, yields

$$\frac{dP_n(t)}{dt} = -a_k(t)P_n(t) + a_k(t)P_{n-1}(t).$$

Rearranging expressions and multiplying both sides of the resulting equalities by $e^{m_k(t)}$ yields

$$e^{m_k(t)} \left[\frac{dP_n(t)}{dt} + a_k(t)P_n(t) \right] = a_k(t)e^{m_k(t)}P_{n-1}(t).$$

Observe that

$$\frac{d\left\{e^{m_k(t)}P_n(t)\right\}}{dt} = e^{m_k(t)}m_k^{'}(t)P_n(t) + e^{m_k(t)}\frac{dP_n(t)}{dt}.$$

Since $m_k^{'}(t) = a_k(t)$, the right hand side of this expression is $e^{m_k(t)}a_k(t)P_n(t) + e^{m_k(t)}\frac{dP_n(t)}{dt}$. Therefore,

$$\frac{d\left\{e^{m_k(t)}P_n(t)\right\}}{dt} = e^{m_k(t)}a_k(t)P_{n-1}(t).$$

Suppose $n = 1$. Since $P_0(t) = e^{-m_k(t)}$,

$$\frac{d\left\{e^{m_k(t)}P_1(t)\right\}}{dt} = a_{k(t)}$$

and

$$P_1(t) = (m_k(t) + c)\, e^{-m_k(t)}$$

for some c. Since $P_1(0) = 0$, $c = 0$ and $P_1(t) = m_k(t)e^{-m_k(t)}$.

By employing an induction argument, we can show that

$$P_n(t) = \frac{m_k(t)^n}{n!}e^{-m_k(t)}.$$

Earlier we stated the four properties that the stochastic process $N_k(t)$ must possess to be a non-homogeneous Poisson process. These properties were used to derive the probabilities $P_n(t)$. If we initially assumed that $P_n(t) = \frac{m_k(t)^n}{n!}e^{-m_k(t)}$, then we could show that the stochastic process $N_k(t)$ must satisfy the four properties stated earlier.