CORRECTIVE MAINTENANCE OPTIMIZATION IN AN AIR FORCE

Hoong Chuin LAU — Kok Yong NEO — Wee Chong WAN

The Logistic Institute – Asia Pacific
National University of Singapore
Singapore 117543, Republic of Singapore

Email: lauhc@comp.nus.edu.sg — *Email: tlinky@nus.edu.sg* — *Email: tliwwc@nus.edu.sg*

ABSTRACT

Successful military mission planning and execution depend critically on equipment serviceability and resupply. Due to the stochastic nature of demands, the forecast of optimal spares and resources needed to guarantee the level of serviceability is a complex problem, especially in a multi-echelon setting. In this paper, we propose a decision-support concept and software tool known as ***Corrective Maintenance Optimizer (CMO)*** *that helps to optimize system availability, through proper allocation of spare parts, both strategically and operationally.*

Keywords: Corrective Maintenance, Decision Support System, Military OR

1. INTRODUCTION

Spare parts are vital and expensive assets in military services such as an air force. Ineffective management of these resources could result in a high operating costs or a low operational availability. Repairable items (i.e. items that can be repaired from a breakdown or failure) are of particular importance and this is typically evident in an air force that relies on numerous heavily utilized and relatively expensive equipment. For example in US Air Force, repair items alone occupy 2/3 of total inventory asset values (which amounts to US$24b in 1998) [3]. Any failure to re-supply either through procurement of new equipment or repair failed components will affect the functionality of aircrafts and impose perilous threat to combat missions. To ensure an adequate supply of repair items, large amounts of spare components must be stocked. However, spares consume capital and space and many of which become obsolete quickly. The goal of the military planner is then to balance between cost and availability in order to sustain the demands of combat operations.

The logistics infrastructure for corrective maintenance usually consists of a centralized depot, a number of operating bases, and possibly a hierarchy of intermediate repair stations in between. When a failure occurs, the malfunction part is removed and replaced with a functional spare part from the base stock if it is available. The malfunction part is then sent to the nearest repair unit, where it is repaired and eventually sent to the base as a spare part. The problem is then to consider trade-off between the stocking levels at the operating bases and the centralized depot subject to budgetary constraints. A well-known solution technique is the *Multi-Echelon Technique for Recoverable Item Control* proposed by Sherbrooke [14,15]. In this seminal work, the author proposed two components. The first component is to characterize the performance of the multi-echelon system under a specification of stock levels. The second is to determine optimal levels that minimize inventory costs subject to resource constraints. Unfortunately the correctness of the technique relies on unrealistic assumptions such as a steady state demand rate and infinite repair resources.

Many recent research have made inroad in addressing these issues and are consequently incorporated by the authors of this paper to develop the *Corrective Maintenance Optimizer (CMO)*. Our extended METRIC model is able to handle more complicated scenarios, including multi-indenture items and time-dependent demand rates. We also realize that many factors such as logical structure of the systems, components age and aging processes cannot be expressed analytically. Such real-life complexities can be handled by approximation or correction factors, but may sometimes lead to highly inaccurate results. Simulation offers the best possible remedy as rules can be incorporated to better reflect the reality of life. On the other hand, simulation is crippled by long execution time even with today's modern computers, which renders them inappropriate in time-critical situations.

This observation has motivated us to develop a hybrid concept that addresses these shortcomings. We proposed an analytical-simulation hybrid approach with the proposition that the vast majority of the possible scenarios can be tackled by the analytical model, while the few difficult cases can subject to more rigorous evaluation by high-fidelity simulation models. A tool was developed utilizing the above hybrid approach and the results are verified using practical instances. Our results demonstrate that the proposed concept and tool achieve both efficiency and flexibility, making it a realistic candidate for real-world deployment. This paper is organized as follows. Section 2 briefly reviews some related works. Section 3 presents the system design. Section 4 describes the experimental verification. Finally, conclusions and future work are given in section 5.

2. RELATED WORK

Multi-echelon inventory modeling and mathematical models in reliability analysis are the two fundamental techniques behind maintenance resources optimization that have growth rapidly in the last decades. The following are the reviews of a few selected influential inventory models. METRIC (Multi-Echelon Technique for Recoverable Item Control) is the basis for many multi-echelon repairable-item models. It is a pioneer study for multi-echelon, single-indenture and multi-item optimization models presented by Sherbrooke [14]. The main issue with METRIC is the assumption that waiting times at the depot are not correlated, and thus the depot can be seen as an infinite server system. This assumption is often invalid as repair facilities are often capacity-constrained.

The OPUS series are well-known spare optimization software developed commercially by Systecon AB [13]. While OPUS has solved successfully many spares optimization problems, the key limitation is that it can only handle steady state problems. It also assumed that the repair resources are infinite, which further reduce its usability. SPAR [16] is yet another popular commercial software tool, using Monte-Carlo simulation. SPAR is able to handle components with time-dependent demand rates. However, it performs optimization inefficiently due to the long simulation runtime.

3. SYSTEM DESIGN

3.1 Overview

Consider a simplified scenario of a two-level echelon structure with a centralized depot connected to a stock supporting site, a repair site and an arbitrary number of operating sites. The corrective maintenance workflow begins with a malfunction system at one of the operating sites. The failed part (commonly known as the *Line Replaceable Unit* (*LRU*)) will be removed from the system and replaced with a functional spare LRU if there is an available spare part at the operating site. Otherwise the operational site will place an order to the depot immediately. If there is an available spare at the supporting site, the depot will reallocate the spare to the operational site at once. In the case when the stock is unavailable, the operating base will wait for a functional spare to become available. At the same time, the malfunction LRU will be sent from the operational base to the repairing site, where it will be repaired with a certain success probability. A successfully repaired component will be stored in the supporting site and subsequently be recycled. Should the repair fail, the irreparable LRU will be sent to the depot for discard. The centralized depot holds the ultimate decision in determining whether to procure a new spare or to wait for a spare to be repaired, in accordance to the budget set by the decision makers. The optimization problem is then to maximize operational availability subject to budgetary constraints.

3.2 Search Algorithm

As discussed above, the corrective maintenance problem is to determine the optimal levels that minimize inventory costs subject to budgetary constraints. We apply marginal analysis to determine the next item to stock based on the marginal or incremental value of each item. The output result is the Cost-Effective curve (C/E Curve) (See Figure 4) where each point on the curve indicates optimal allocation under different target availability.

Under the multi-echelon setting, we work with the notion of *stock positions*. We define a *stock position* as the spare item, repair resources and the station where it is stored. In addition, the LRUs are partitioned into *resource groups*, each of which only contains the LRUs that require this particular resource. The simplified algorithm is outlined as follows:

1. Split the problem into sub-problems, one for each resource group.
2. For each resource group, enumerate all feasible resource allocations alternative at the base.
3. For each alternative, optimize the spares allocations at the base and repair resources at operational sites to give a C/E curve.
4. Merge the C/E curves by finding the convex hull.
5. For each allocation on the C/E curve in (4), optimize the spare allocations at the operational sites, resulting in a C/E curve.
6. Merge the curves to find a C/E Curve for a resource group.
7. Merge the curves for all the resource groups to find a total C/E Curve.

3.3 CMO Modes of Operation

The uniqueness of CMO lies in its three modes of operation. Each of these operations is intended to cater to different problem requirements and has its own set of benefits and drawbacks. The combination of the three modes present a scheme in which the strengths of one mode cover the weaknesses of the other when correctly deployed. In the following sections, we describe the operation in each mode and recommend some of the situations in which each mode is best deployed.

3.3.1 Analytical Mode (The Projector)

The analytical model of CMO is based on a modification of the METRIC model developed by Sherbrooke, with extension to handle multi-indenture and time varying demand with passivation. The details of this model are found in [10]. Essentially, the role of the projector is to project the performance of the multi-echelon system under a specification of stock levels, which will be system availability in our scenarios. Given a particular demand pattern and spare allocation, the projector's can compute the availability attained by this allocation.

We present a simplified version of the projector algorithm as follows:

1. Compute the demand rate arising from unreliability for each LRU m, at each time period t, and at each base unit h.
2. Compute the demand rate for each LRU m, at each time period, at the intermediate organizations i.
3. Compute the Probability Density Function for Repair Pipeline for each LRU m, period t at every Organization i from Unit h.
4. Compute the Cumulative Distribution Function for Coverage Time for each LRU m, period t at every Unit h from Unit h.
5. Compute the Cumulative Distribution Function for Buffer Time for each LRU m, period t at every Unit h.
6. Compute the Cumulative Distribution Function for Runout Time for each LRU m, period t at every Unit h.
7. Compute the EBO (Expected Back Order) for each LRU m, period t at every Unit h.
8. Computation of Availability of System l at Unit h.

The analytical mode works under the following assumptions:

1. There are infinite repair resources, i.e. a failed system can be repaired at once.
2. All LRUs are repairable at the depot, i.e. there is no irreparable item.
3. Continuous resupply, i.e. an LRU can be sent up or down the echelon immediately at any time. The transport time for each item between two sites is a constant.
4. FCFS (First Come First Serve) replenishment policy.
5. The remove-and-replace time for each item follows an exponential distribution.
6. The repair time for each item follows an exponential distribution.
7. No lateral supply, i.e. no supply or shipment across sites within the same echelon.

The advantage of this mode is the relatively fast computation speed. For instance a problem size of 5 layers echelon with 400 components runs within minutes, whereas the other two modes can achieve similar results with a much longer runtime.

3.3.2 Simulation Mode

The Simulation Mode involves a simulation of the maintenance workflow and the conduct of experiments as the workflow is simulated over time. In CMO, we apply Monte Carlo simulation. Many simulations are then performed and the desired result is taken as an average over the number of observations.

Apparently, the strength of Simulation Mode lies in its capability in modeling fairly complex problems. The drawback of this approach is the large running time needed, as expected from most simulation techniques. This is especially noticeable when the problem size is very large whereby hours or even days have to be dedicated, dominating considerable amount of computing resources. In addition, it is often necessary to run the simulation mode several times on a same problem in order to take the average results of these runs so as to remove the stochastic noises in the results, thus improving the statistical accuracy.

3.3.3 Hybridized Analyze-Simulate mode

Although the analytical model can handle many scenarios with relatively good accuracy, there are times when it becomes inadequate and simulation mode is often required to breach the gap. A logical extension is then to hybridize the two modes such that run-time can be reduced without making too many assumptions on the problem. Hence the Analyze-Simulate mode is introduced to integrate simulation into the analytical model. Simulation in the Analyze-Simulate mode serves two purposes. The first is to verify and possibly rectify any inconsistency between the analytical and simulation result. As discussed in section 3.3.1, the result in analytical model would match the simulation result to a small degree of imperfection if and only if the problem is modeled with the required assumptions. However, if one or more of the assumption is violated, there will be an error with deviation dependent on the degree of violation. Hence, if a deviation within a certain threshold is allowed, the analytical model could in fact be deployed. In case where threshold is exceeded, simulations are run over certain allocations in the C/E optimization curves to verify on correctness of the result (within the allowed deviation).

In addition, the simulated results are updated into the analytical model to rectify the error, as well as to reduce the cumulative errors for the subsequent points in the C/E curve. In short, the optimization aspect of the problem is handled by analytical method, while simulation is used to characterize transient behavior, provide intermediate feedback and confirm the final solution. In this way, more complex problems such as evaluation of maintenance support concept, impact of combat damage, workshop loading can also be performed via Analyst-Simulate mode of operation, achieving high effectiveness and efficiency.

The second purpose of Analyst-Simulate is in circumstances when the problem can be decomposed into disjoint sub-problems. For example, if lateral re-supply (not supported in analytical model) is only applied in certain time period within the planning horizon, then Analyst-Simulate will apply simulation only on the required time interval where lateral re-supply is needed and then integrate the result back into the Analytical model. As simulation is applied only over a certain interval, this reduces the required runtime significantly.

3.3.4 System Design

Figure 1 gives an overview of the CMO system infrastructure.

The User Interface Layer implements a GUI for input of scenario data and output of results in both graphical and tabular format.

The Optimization Layer contains the optimization engine together with the following modules:

- The Projector, that measures availability given an initial resource allocation.
- The Acquisition Manager that generate the initial resource allocation for all the sites.
- The Search Algorithm Manager that uses the projector to search for optimal solutions.
- The Controller handles the coordination among the different modules.
- The Verifier is the interface with the simulator.

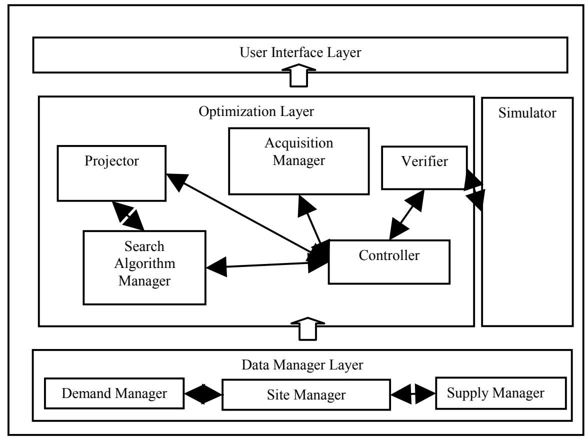

Figure 1: System Infrastructure

The Data Manager Layer models data as object-oriented classes. It consists of the following modules:

- Site Manager manages the site dependant information
- Demand Manager manages the modeling of. demand arising from unreliability failure and combat damage.
- Supply Manager manages the various supporting resources such as repair resources, spare stocks and transporters.

4. IMPLEMENTATION AND RESULTS

The analytical engine is developed in C++ while the simulation engine is developed using Extend [6]. The simulation model employs a Mote Carlo simulation and is developed to complement analytical methods. The two engines interact through dynamic link library calls.

4.1 Interface

Figure 2 shows a screenshot of the graphical user interface of the system.

A – Data Input: The data is entered using an excel-like grid. Essential input data includes:

- Echelon Structure – Operation bases, intermediate support bases, depots.
- Component Data – Part types, cost, failure rate.
- System Data – System deployment, component indenture.
- Mission Profile – Mission length, utilization rate, critical period
- Supply and Logistics Data – current inventories, supply lead times, transportation times, Not-Reparable at this Station (NRTS) rates/probabilities.

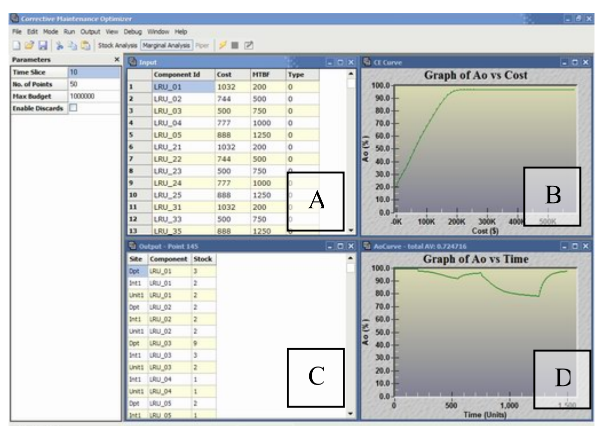

Figure 2: Graphical user interface

B – Output (Availability-Cost Curve): An example of the output from the optimization of spares allocation. The optimal system availability is plotted against the cost incurred.
C – Output (Stock allocation): This shows the optimal stock allocation for particular system availability.

D – Output (Availability-Time Curve): This shows how the availability changes according to time-varying demand for a particular stock allocation.

4.2 Results

4.2.1 Projector

For the purpose of validating availability estimates generated by the analytic model, test cases are developed with the assumptions of the analytic model. The Air Force typically deals with large expensive equipments with many sub-components, and a relatively simple maintenance facilities network. Thus we formulate our test scenario with the above factors in mind.

The test scenario contains of a two-level echelon structure. One system type consisting of 330 LRU types is deployed at the base units over a period of 180 days (4320 hours). The system has time-varying demand during the deployment. Figure 3 shows the analytical availability for this scenario against simulation result, and the time-varying utilization rate (UR).

From Figure 3, we can see that the analytical result matches very closely with the simulation. The maximum absolute difference is less than 0.1%. The average runtime of the analytical engine is 1min while that of the simulation engine is about 20min. Table 1 shows results for additional test scenarios.

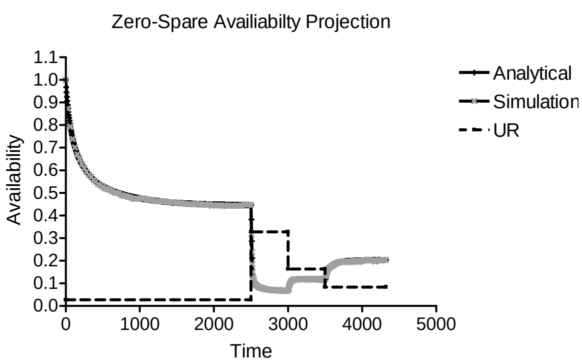

Figure 3: Result of analytical availability projection

Table 1: Verification results for the projector

Test Scenario	Difference
2-Echelon, 330LRU, 0 Spares	<0.1%
2-Echelon, 330LRU, ∞ Spares	<0.1%
2-Echelon, 330LRU, Allocation 1	1.30%
2-Echelon, 330LRU, Allocation 2	<0.1%
4-Echelon, 46LRU, 0 Spares	<0.1%
4-Echelon, 46LRU, 0 Spares	<0.1%

4.2.2 Search Algorithm

We verified the results against OPUS [13]. The same test scenario as section 4.2.is used for scenario 1 and another test case with 4 echelons and 230 LRUs is used for scenario 2. However a constant utilization rate is assumed due to limitations in the OPUS software. Figure 4 shows the Cost-Effective curve for the two scenarios generated by CMO and OPUS. Observe that the two outputs are almost identical, which provide evidence of the optimizer's accuracy. We can use OPUS to approximate time-varying demand by average the demand rate over time. However this approximation tends to underestimate the availability due to the simplified assumptions. We verified the results using the same test case from section 4.2.1. Figure 5 shows the Cost-Effective curve generated by CMO and OPUS approximation for the test case with time-varying demand.

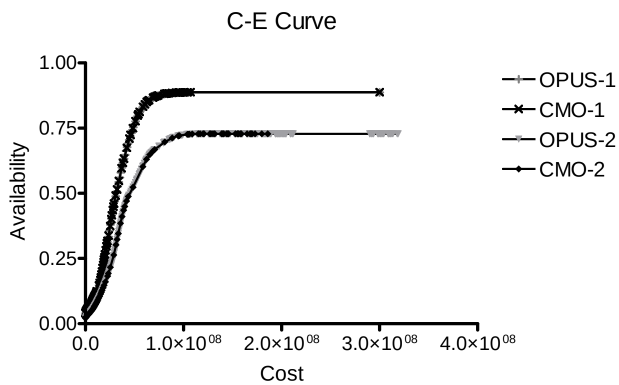

Figure 4: Result of analytical engine for optimization

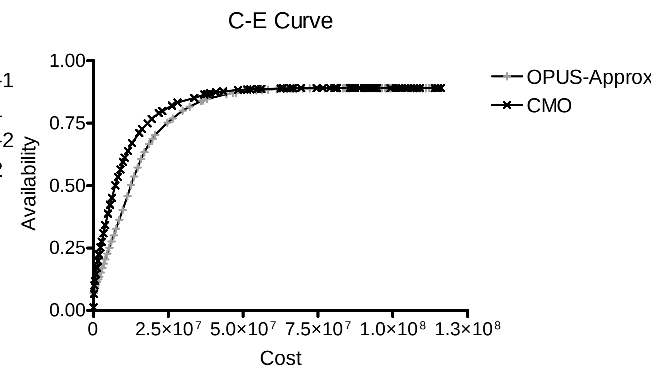

Figure 5: Result of analytical engine for optimization with time varying demand

5. CONCLUSION

A hybrid analytical-simulation optimizer for corrective maintenance has been presented in this paper. The system provides both efficiency and flexibility in dealing with complex scenarios. The analytical engine is in close agreement with the simulation model and thus can be applied in most air force scenarios tested.

6. REFERENCES

[1] Alfredsson P., “Optimization of Multi-Echelon Repairable Item Inventory Systems with Simultaneous Location of Repair Facilities”, *European Journal of Operational Research*, 99 584-595, 1997

[2] Balana, A., Gross D., Soland R., “Optimal Provisioning for Single-Echelon Repairable Item Inventory Control in a Time-Varying Environment”, *IIE Transactions* 21(3) 202-212, 1989

[3] “Department of Defence’s Materiel and Distribution Management Fact Book Fiscal Year 1998”, p28, 1988

[4] Díaz A. and Fu M., “Models for Multi-echelon Repairable Item Inventory Systems with Limited Repair Capacity”, *European Journal of Operational Research*, 97 480-492, 1997

[5] “Extend v6 User’s Guide”, Imagine That, Inc. 2002

[6] Gross D., Miller D., Soland R., “A closed queueing Network Model for Multi-Echelon Repairable Item Provisioning”, *IIE* Transactions, 15 (4) 344-352, 1983

[7] Hillestad R.J., “Dyna-METRIC: Dynamic Multi-Echelon Technique for Recoverable Item Control”, The RAND Corporation, July 1982

[8] Jung W., “Recoverable Inventory Systems with Time-Varying Demand”, *Production and Inventory Management Journal*, 34(1) 77-81, 1993

[9] Kaplan A. “Incorporating Redundancy Considerations into Stockage Models”, *Naval Research Logistics*, 36, 625-638, 1989

[10] Lau H. C., Song H., See C. T. and Cheng S. Y. Evaluation of Time-Varying Availability in Multi-Echelon Spare Parts Systems with Passivation, *European J. Operational Research*. In Press

[11] Muckstadt J., “A model for multi-item, multi-echelon, multi-indenture inventory system”, *Management Science*, 20 472-481, 1973

[12] Nelson B. L., “Stochastic Modeling: Analysis & Simulation”, McGraw-Hill International Edition, 1995

[13] “OPUS10 User’s Reference – Logistics Support and Spares Optimization version 3”, Systecon AB, May 1998

[14] Sherbrooke C. C., “METRIC: A multi-echelon technique for recoverable item control”, *Operations Research*, 16(2) 122-141, 1968

[15] Sherbrooke C. C., “Optimal Inventory Modeling of System: Multi-Echelon Techniques”, John Wiley & Sons, New York, 1992

[16] Spar 2002, http://www.clockwork-group.com