Dynamic Routing of Prioritized Warranty Repairs

Feng Chen, Vidyadhar G. Kulkarni
Department of Statistics and Operations Research
University of North Carolina at Chapel Hill

Abstract

Recent years have seen a strong trend toward outsourcing warranty repair services to outside vendors. In this paper we consider the problem of dynamically routing warranty repairs to service vendors when warranties have priority levels. Each time an item under warranty fails, it is sent to one of many service vendors for repair. Items covered by higher priority warranty receive higher priority in repair service. The manufacturer pays a fixed fee per repair and incurs a linear holding cost while an item is undergoing or waiting for repair. The objective is to minimize the manufacturer's long-run cost. Due to the complexity of the problem, it is very unlikely that there exist tractable ways to find the optimal solutions. Therefore, we propose four heuristic routing procedures that are applicable to real-life problems. We evaluate the heuristics using simulation. The simulation results show that the index-based "generalized join the shortest queue" policy, which applies a sinlge policy improvement step to an inital state-independent policy, perfoms the best among all four heuristics.

Key words: Optimal allocation, Warranty outsourcing, Index policies, Multi-priority, Dynamic routing.

1 Introduction

Warranty has been playing an increasingly important role in product sales and services since the second half of twentieth century. In 2004, the 25 largest manufactures in the United States spent a total of about $15 billion on warranty claims. Warranty claims processing consumed 2.5% $\sim$ 4.5% of revenues across all industries (Byrne [4]). It has been shown that warranty improvements can not only save cost but also boost revenues, enhance customer satisfaction and loyalty, and even drive up the product quality.

There is an extensive literature on the subject of warranty. For a comprehensive reference, see Blischke and Murthy [3]. They discuss a variety of warranty policies including standard consumer product warranties such as the free replacement and pro rata, as well as warranties used in large volume or specialized transactions. Analytical models dealing with cost and optimization problems from both the manufacturer's and the buyer's point of view are developed. Methods of collecting and analyzing relevant data are also addressed. A literature review until 2002 is given by Murthy and Djamaludin [17]. For recent development, among others, see Dimitrov, Chukova, and Khalil [6], Yeh, Chen, and Chen [22], and Manna, Pal, and Kulandaiyan [16].

There has been a strong trend towards outsourcing various business operations in recent years, especially in the IT industry. According to IDC (a Framingham, Massachusetts-based market research firm), spending on IT outsourcing reached $56 billion in 2000 and is expected to top $100 billion by 2005. As a major component of the manufacturing and retail industry, warranty repair services have experienced the rising outsourcing tide as well. Outsourcing warranty repairs offers the original equipment manufacturer the opportunity to reduce operating cost and capital investment, focus on their core business, increase speed to market, and faster customer response time.

Typically a manufacturer outsources repair work to several vendors, in which case the manufacturer faces the problem of how to distribute the workload among vendors in a cost-effective manner. We categorize warranty repair routing problems by the following four aspects.

(i) Based on the priority levels, we have either *single-priority problems* or *multi-priority problems*. Multi-priority problem arises when the manufacturer provides warranties with priority levels that specify different turnaround times. Warranty with higher priority level guarantees a shorter turnaround time. It is given to very important customers (e.g. customers that make frequent or large purchases from the manufacturer) to keep them happy, or sold to customers who are willing to pay more for a shorter repair time. To meet the specified turnaround times, products covered by a higher priority warranty are given higher priority in repair service.

(ii) Based on the number of items under warranty, we have either *fixed-population problems*, or *variable-population problems*. Fixed-population problem arises when we are dealing with warranty repairs for a batch of items sold at once, in which case no items enter or leave the warranty population of interest during the warranty period. More often, the items are sold in a continuous fashion. Thus the number of items under warranty increases when a new sale occurs and decreases when the warranty expires on an existing item, in which case we have a variable-population problem.

(iii) Based on the assignment rule, we have either *assign-at-purchase policies* or *assign-at-*

failure policies. The former requires an item to be assigned to a vendor at the time of purchase and sent to that vendor for repair each time it fails. This can be done by printing the repair vendor's phone number on the warranty card and instructing customers to call that number for repair services. The latter allows the items to be assigned to different vendors at the time of failure. In this case, a routing center's phone number is printed on the warranty card. A repair vendor's information is provided when the customer calls with a request for repair.

(iv) Based on the available information, we can use either *state-independent policies*, *partially state-dependent policies*, or *fully state-dependent policies*. State-independent policies do not use any real-time information of the system, i.e., the same rule is applied to every assignment. Partially state-dependent policies use only the real-time information of the warranty population, which includes the number of items under each type of warranty and the remaining warranty length of each item. This information can be easily collected by keeping a record of the purchases made in the past W time, where W is the warranty length. If the warranty periods are assumed to be i.i.d. exponential random variables, then only the warranty population size is necessary. Fully state-dependent policies use real-time information of both the warranty population and the vendors. Real-time information of vendors means the number of items at each vendor at any time. Collecting this information requires real-time communication between the manufacturer and the vendors, which may need a more complicated information system and cost extra depending on the level of information needed.

The warranty repair allocation problem has the simplest structure when considering single-priority and fixed-population. In this case, the assign-at-purchase model reduces to a resource allocation problem with separable objective function. Note that only state-independent policies are applicable in assign-at-purchase model. This problem has been extensively studied in the literature. When the objective is convex, a simple greedy algorithm first proposed by Gross [11] can be used to solve the problem optimally. See Ibaraki and Katoh [12] for a comprehensive reference for the resource allocation problems. Opp, Kulkarni, and Glazebrook [18] discuss the application of the greedy algorithm to the warranty repair allocation problem. Ding and Glazebrook [7] consider a goodwill cost model that takes explicit account of the delays experienced by customers. They show that simple greedy approaches work well. The assign-at-failure model for single-priority and fixed-population problem is studied by Opp, Kulkarni, and Glazebrook [18]. They argue that optimally solving real-life size problems is numerically intractable. They develop index-based, fully state-dependent heuristic policies to find near-optimal solutions.

When priorities are considered, the objective function is no longer separable. Buczkowski, Kulkarni, and Hartmann [5] study the assign-at-purchase model for multi-priority, fixed-population problems. They formulate the problem as a minimum cost network flow problem and provide an efficient algorithm to solve it.

In this paper, we consider the multi-priority, variable-population problem, and we study assign-at-failure policies. Given the difficulty of the problem even without considering priority and finite constant warranty length (see Opp, Kulkarni, and Glazebrook [18]), trying to find the optimal solution is unrealistic. Hence we turn our attention to heuristic allocation procedures. One natural way of obtaining an approximate solution is to simplify the problem by assuming exponential warranty length and formulate it as an Markov decision process (MDP). The optimal policy for the resulting MDP can be expected to work reasonably well for the original problem. However, the curse of dimensionality (of the state space) makes solving the Bellman equations of the MDP impractical even for small-size problems. We present four heuristics that are applicable to large problems, then evaluate and compare them using simulation. Among the four heuristics, the Generalized Join the Shortest Queue (GJSQ) policy performs the best and is of our primary interest.

The rest of this paper is organized as follows: We introduce notation and describe the problem in details in Section 2. In Section 3, we present four heuristic algorithms. The detailed derivation for the indices for GJSQ policy is given in the Appendix. A simulation study is provided in Section 4. We conclude in Section 5.

2 Problem Description

Consider a manufacturer offering two types of warranties for its product. Type 1 warranty guarantees a shorter repair turn around time than type 2 warranty. For example, the turn around times could be three days for items covered by type 1 warranty (type 1 items) and two weeks for items covered by type 2 warranty (type 2 items). In order to meet the guaranteed turn around times, type 1 items are given priority in repairs over type 2 items. To simplify analysis, we assume the priority is preemptive resume. Sales of type k items form a Poisson process with rate λ_k, denoted by $PP(\lambda_k)$, $k = 1, 2$. Warranty length for either type is a constant W. The manufacturer outsources the warranty repairs to V vendors (one of them could be the manufacturer's facility itself). The life times of the items are i.i.d. $\exp(\beta)$ random variables. When an item fails while it is under warranty, it is sent to one of the V vendors for repair. There is one repair person at each vendor. The repair times are i.i.d. $\exp(\mu_i)$ at vendor i. The manufacturer pays vendor i a fixed fee c_i each time a repair is assigned to vendor i, $i = 1, \cdots, V$. While a type k item is awaiting or under repair at vendor i, the manufacturer incurs a holding cost (good will cost) at rate h_{ki}, $k = 1, 2, i = 1, \cdots, V$. We assume items covered by higher priority warranty incur holding cost at a larger rate, i.e., $h_{1i} \geq h_{2i}$, $i = 1, \cdots, V$. This situation agrees with the way the well-known c-mu rule assigns priorities to multiple classes of jobs at a single service station (higher priority is given to the class with larger holding cost rate). Items are as good as new after repair. The goal of the manufacturer is to assign repairs to outside vendors in such a way that the expected long-run average cost is minimized.

3 Heuristics

3.1 Optimal State-Independent Policy (OSI)

We first consider state-independent policies, i.e., stationary policies that do not depend on the real-time system state. We confine ourselves to a specific, yet natural, type of state-independent policy, namely, a Bernoulli splitting policy. Under this policy, a type k repair is assigned to vendor i with probability p_{ki}, where $\sum_{i=1}^{V} p_{ki} = 1$, $k = 1, 2$. Let $\boldsymbol{p}_k = (p_{k1}, p_{k2}, \cdots, p_{kV})$, $k = 1, 2$. Then the Bernoulli splitting policy can be denoted by $(\boldsymbol{p}_1, \boldsymbol{p}_2)$. We aim to find an optimal Bernoulli splitting policy that minimizes the long-run average cost. In order to compute the long-run average cost of a Bernoulli splitting policy with splitting probabilities $(\boldsymbol{p}_1, \boldsymbol{p}_2)$, we simplify the real system by assuming that failures of type k items under warranty occur according to a $PP(\phi_k)$, where $\phi_k = \lambda_k W \beta$ (or, equivalently, the number of type k functioning items under warranty is a constant $\lambda_k W$). Then type k repairs arrive at vendor i according to a $PP(\phi_k p_{ki})$. We assume $\sum_{i=1}^{V} \mu_i > \phi_1 + \phi_2$, i.e., the total arrival rate of failed items is less than the total service rate. As a result, there must exist policies $(\boldsymbol{p}_1, \boldsymbol{p}_2)$ such that $\phi_1 p_{1i} + \phi_2 p_{2i} < \mu_i$, $i = 1, ..., V$. We only consider such stable policies for the rest of the paper.

Because of their preemptive resume priority, type 1 items simply do not see type 2 items in the repair queue. So the expected number of type 1 items at vendor i is

$$L_{1i}(p_{1i}) = \phi_1 p_{1i}/(\mu_i - \phi_1 p_{1i}). \tag{1}$$

Obviously, the expected number of all items at vendor i is $(\phi_1 p_{1i} + \phi_2 p_{2i})/(\mu_i - (\phi_1 p_{1i} + \phi_2 p_{2i}))$. Hence the expected number of type 2 items at vendor i is

$$L_{2i}(p_{1i}, p_{2i}) = \frac{\phi_1 p_{1i} + \phi_2 p_{2i}}{\mu_i - (\phi_1 p_{1i} + \phi_2 p_{2i})} - \frac{\phi_1 p_{1i}}{\mu_i - \phi_1 p_{1i}}. \tag{2}$$

Let

$$f_{1i}(x) = \begin{cases} (h_{1i} - h_{2i})\frac{x}{\mu_i - x}, & \text{if } x < \mu_i \\ \infty, & \text{if } x \geq \mu_i, \end{cases} \tag{3}$$

$$f_{2i}(x) = \begin{cases} c_i x + h_{2i}\frac{x}{\mu_i - x}, & \text{if } x < \mu_i \\ \infty, & \text{if } x \geq \mu_i, \end{cases} \tag{4}$$

and

$$f_i(x_1, x_2) = f_{1i}(x_1) + f_{2i}(x_1 + x_2). \tag{5}$$

Then the long-run average cost rate at vendor i is $f_i(\phi_1 p_{1i}, \phi_2 p_{2i})$.

Therefore, the optimal Bernoulli splitting policy $(\boldsymbol{p}_1^*, \boldsymbol{p}_2^*)$ can be obtained by solving the following optimization problem:

$$\begin{aligned} \min \quad & \sum_{i=1}^{V} f_i(\phi_1 p_{1i}, \phi_2 p_{2i}) \\ \text{s.t.} \quad & \sum_{i=1}^{V} p_{ki} = 1, \quad k = 1, 2 \\ & p_{ki} \geq 0, \quad k = 1, 2, \quad i = 1, 2, \cdots, V. \end{aligned} \tag{6}$$

Note that the objective function is separable in terms of pairs (p_{1i}, p_{2i}), i.e., it is a sum of functions of two variables (p_{1i}, p_{2i}) each. Each single term can be further decomposed as in (5). To take advantage of the structure of this problem and apply simple and efficient algorithm, we solve the discretized version of the above optimization problem as described in the following. Suppose $\lambda_k W$ are integers, $k = 1, 2$. Otherwise, take their integer parts. Associate a pair of integers (y_{1i}, y_{2i}) with each vendor, where $\sum_{i=1}^{V} y_{ki} = \lambda_k W$, and let $p_{ki} = \frac{y_{ki}}{\lambda_k W}$, $k = 1, 2, i = 1, 2, \cdots, V$. This can be interpreted in the following way: Assume that there are $\lambda_k W$ type k items under warranty. Assign y_{ki} of them to vendor i and always send them to vendor i for repair upon failure. In terms of (y_{1i}, y_{2i}), the long-run average cost rate at vendor i is $f_i(y_{1i}\beta, y_{2i}\beta)$, where f_i is defined in (5).

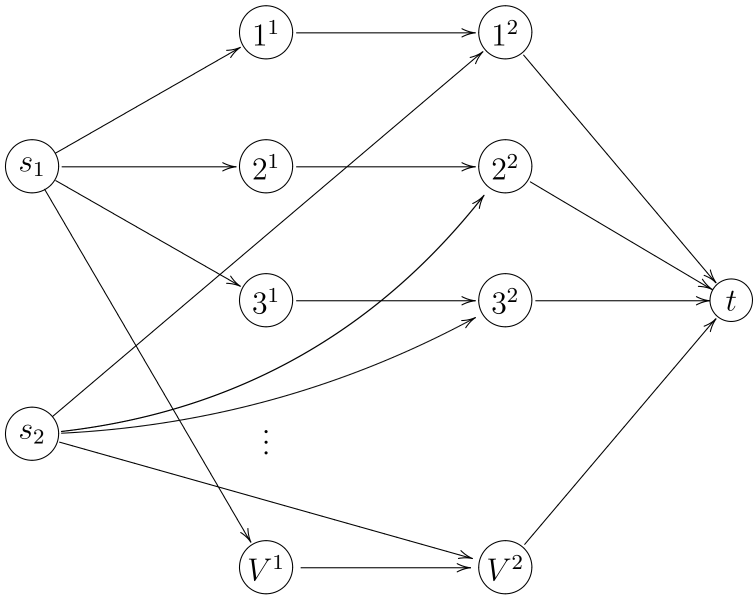

Figure 1: Network Model of Two-priority Problem

Therefore, the discretized version of (6) can be written as

$$\begin{aligned} \min \quad & \sum_{i=1}^{V} f_{1i}(y_{1i}\beta) + f_{2i}(y_{1i}\beta + y_{2i}\beta) \\ \text{s.t.} \quad & \sum_{i=1}^{V} y_{ki} = \lambda_k W, \quad k = 1, 2 \\ & y_{ki} \geq 0 \ \text{ and integer}, \quad k = 1, 2, \quad i = 1, 2, \cdots, V. \end{aligned} \tag{7}$$

Optimization problem (7) can be formulated as a minimum cost network flow problem which can be solved by a *Successive Shortest Path Algorithm* with complexity $O(V+(\lambda_1+\lambda_2)W \log V)$ (see Buczkowski, Kulkarni, and M. Hartman [5], and Ahuja, Magnanti, and Orlin [1]). We provide the formulation and algorithm here for ready reference.

Figure 1 shows the network model for (7). There are two source nodes s_1, s_2 with supplies of $\lambda_1 W$ and $\lambda_2 W$, respectively. There is one sink node t with a demand of $(\lambda_1 + \lambda_2)W$. All other nodes are transshipment nodes with 0 demand and 0 supply. The arc properties are summarized in table 1.

Define $\boldsymbol{y}_k = (y_{k1}, \cdots, y_{kV}), k = 1, 2,$ $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$ and $\delta f(y\beta) = f(y\beta) - f((y-1)\beta), y \geq 1$. The following algorithm can be used to solve the network problem.

Successive Shortest Path Algorithm:

Initialize $\boldsymbol{y} := 0$;

Arc	Capacity	Flow	Cost
(s_i, j^i)	$\lambda_i W$	y_{ij}	0
(j^1, j^2)	$\lambda_1 W$	y_{1j}	$f_{1j}(y_{1j}\beta)$
(j^2, t)	$(\lambda_1+\lambda_2)W$	$y_{1j}+y_{2j}$	$f_{2j}((y_{1j}+y_{2j})\beta)$

Table 1: Arc Properties for the Network Representation of (7)

while $\sum_{i=1}^{V} y_{2i} < \lambda_2 W$ do

- compute $\min_{j=1,\ldots,V} \delta f_{2j}((y_{2j}+1)\beta)$,
- increment y_{2k} by 1, where $k \in \arg\min_{j=1,\ldots,V} \delta f_{2j}((y_{2j}+1)\beta)$;

end

while $\sum_{i=1}^{V} y_{1i} < \lambda_1 W$ do

- compute $d_i = \delta f_{1i}((y_{1i}+1)\beta) + \delta f_{2i}((y_{1i}+y_{2i}+1)\beta), i = 1, ..., V$ and $d_{V+1} = \min_{y_{2j}>0} \delta f_{1j}((y_{1j}+1)\beta) + \min_{j=1,\ldots,V} \delta f_{2j}((y_{1j}+y_{2j}+1)\beta)$.
- let $q \in \arg\min_{j=1,..,V+1} d_j$.
  If $q \in \{1, ..., V\}$, increment y_{1q} by 1.
  If $q = V+1$, let $k \in \arg\min_{y_{2j}>0} \delta f_{1j}((y_{1j}+1)\beta)$ and $p \in \arg\min_{j=1,\ldots,V} \delta f_{2j}((y_{1j}+y_{2j}+1)\beta)$. Increment y_{1k} and y_{2p} by 1 and decrement y_{2k} by 1 unit.

end

Denote the optimal solution to (7) by (y_{1i}^*, y_{2i}^*), then we have the following approximate solution to (6):

$$p_{1i}^* = \frac{y_{1i}^*}{\lambda_1 W}, \quad p_{2i}^* = \frac{y_{2i}^*}{\lambda_2 W}, \quad i = 1, 2, \cdots, V. \tag{8}$$

Note that p_{ki}^* is an optimal solution to (6) to a degree of accuracy of $\frac{1}{\lambda_k W}$, $k = 1, 2$, $i = 1, \ldots, V$. The expected numbers of items under warranty, $\lambda_k W$, are usually large in real problems, in which case, the discretized solution is very close to the real-valued optimal solution.

3.2 Generalized Join the Shortest Queue Policy (GJSQ)

We continue to assume that the number of type k functioning items under warranty is a constant $\lambda_k W$ and failures of type k items under warranty occur according to a $PP(\phi_k)$, where

$\phi_k = \lambda_k W \beta$, $k = 1, 2$. Therefore, the original warranty repair allocation problem reduces to the problem of routing items arriving according to two independent Poisson streams to several vendors where service is provided according to a predetermined priority policy.

A generalized model of this situation is studied by Ansell, Glazebrook, and Kirkbride [2]. In their model, jobs from a number of different classes arrive according to independent Poisson processes. Jobs are either generic or dedicated, and they are routed to a set of service stations. Dedicated jobs can be processed only by a specified station, while generic jobs can be processed at any station. Jobs are served according to a static priority policy at each station. A holding cost is incurred at a class-dependent rate while a job is in the system. The objective is to minimize the long-run average holding cost rate. The authors develop a dynamic routing heuristic by applying a single policy improvement step to an initial static policy (see also Krishnan [13] and Tijms [21] for this approach). They name the resulting index-based heuristic "generalized join the shortest queue" policy. We provide their main result here for ready reference.

Denote the set of generic jobs by G and the set of dedicated jobs to station i by D_i. Denote the number of class k jobs that are currently awaiting or under service at station i by x_{ki}, $k \in G \cup D_i$, $i = 1, \cdots, V$. Let $\boldsymbol{x}^i = \{x_{ki} | k \in G \cup D_i\}$. For each class k, the GJSQ policy associates with each station i an index I_{ki} which is a linear function of the number of jobs of each class at station i, i.e.,

$$I_{ki}(\boldsymbol{x}^i) = \sum_{l \in G \cup D_i} \theta^i_{kl} x_{li} + \delta^i_k, \quad k \in G \cup D_i, \quad i = 1, \cdots, V,$$

where the coefficients θ^i_{kl} and δ^i_k are constants. The GJSQ policy routes an incoming class k job to the station with the smallest index.

Although structurally our problem is a special case of that studied by Ansell, Glazebrook, and Kirkbride [2] (we consider two generic classes and no dedicated classes), our approach differs from theirs in the following ways: In addition to holding cost, we also allow a station-dependent fixed cost per assignment, which is not considered by Ansell, Glazabrook, and Kirkbride [2]. Furthermore, we are able to give tractable closed-form expressions for the coefficients following complicated queueing theoretic calculations, while in Ansell, Glazabrook, and Kirkbride [2] the coefficients are given as a solution to an infinite set of recursive equations.

Following Ansell, Glazabrook, and Kirkbride [2], we show that the linear structure of the indices continues to hold in the presence of fixed costs. In particular, for our problem there exist two indices $I_{1j}(x_{1j}, x_{2j})$ and $I_{2j}(x_{1j}, x_{2j})$ for each vendor $j = 1, \ldots, V$ of the following form

$$I_{1j}(x_{1j}, x_{2j}) = A_{1j} + B_{1j}x_{1j} + C_{1j}x_{2j}, \tag{9}$$

$$I_{2j}(x_{1j}, x_{2j}) = A_{2j} + B_{2j}x_{1j} + C_{2j}x_{2j}. \tag{10}$$

After some lengthy algebra (see the Appendix), we get the closed-form expressions for the coefficients of the indices as given in the following theorem. We introduce the following notations before stating the theorem:

$$\phi_{kj} = \phi_k p_{kj}, \tag{11}$$

$$\eta_j = \frac{1}{\mu_j - \phi_{1j}}, \tag{12}$$

$$\xi_j = \frac{1}{\mu_j - \phi_{1j} - \phi_{2j}}, \tag{13}$$

where $k = 1, 2$, $j = 1, \ldots, V$.

Theorem 1 *Assume $\mu_j > \phi_{1j} + \phi_{2j}$ and let f_j, ϕ_{kj}, η_j and ξ_j be as given in (5), (11), (12), and (13), respectively. Then the coefficients of the indices defined in (9) and (10) are given by*

$$\begin{aligned}
A_{1j} &= c_j + [c_j(\phi_{1j} + \phi_{2j}) + h_{1j}]\eta_j + f_j(\phi_{1j}, \phi_{2j})\xi_j \\
&\quad + (\phi_{1j}h_{1j} + \frac{1}{2}\phi_{2j}h_{2j})\eta_j^2 + [c_j(\phi_{1j} + \phi_{2j})\phi_{2j} + \phi_{2j}h_{2j}]\eta_j\xi_j \\
&\quad + \frac{1}{2}(\phi_{1j} + \mu_j)\phi_{2j}h_{2j}\eta_j^3 + (\phi_{1j}\phi_{2j}h_{1j} + \frac{1}{2}\phi_{2j}^2h_{2j})\eta_j^2\xi_j \\
&\quad + \frac{1}{2}(\phi_{1j} + \mu_j)\phi_{2j}^2h_2\eta_j^3\xi_j + \phi_{2j}^2\mu_j h_{2j}\eta_j^2\xi_j^2, \\
B_{1j} &= h_{1j}\eta_j + \phi_{2j}h_{2j}\eta_j^2 + \phi_{2j}^2h_{2j}\eta_j^2\xi_j, \\
C_{1j} &= h_{2j}\eta_j + \phi_{2j}h_{2j}\eta_j\xi_j,
\end{aligned}$$

and

$$\begin{aligned}
A_{2j} &= c_j + [c_j(\phi_{1j} + \phi_{2j}) + h_{2j} + f_j(\phi_{1j}, \phi_{2j})]\xi_j + \phi_{1j}h_{1j}\eta_j\xi_j + \phi_{2j}\mu_j h_{2j}\eta_j\xi_j^2, \\
B_{2j} &= h_{2j}\eta_j + \phi_{2j}h_{2j}\eta_j\xi_j, \\
C_{2j} &= h_{2j}\xi_j.
\end{aligned}$$

Proof: See Appendix. ■

4 Simulation Study

Although we have the optimal splitting probabilities for the OSI policy and the closed-form expressions for the indices of the GJSQ policy, calculating the expected costs of these policies for the original warranty repair allocation problem is analytically intractable. Therefore, we use simulation to evaluate the performance of these two policies and compare them with two other heuristics. All four heuristics are described below.

- **Join the Shortest Queue policy (JSQ)**: An incoming failed item is sent to the vendor with the shortest repair queue (i.e., the least number of items of both types). If more than one vendor has the shortest queue, among those the item is sent to the one with the smallest fixed cost. In case a tie still exists, it is broken arbitrarily.

- **Optimal State-Independent policy (OSI)**: Incoming failed items are routed according to the optimal Bernoulli splitting probabilities $(\boldsymbol{p}_1^*, \boldsymbol{p}_2^*)$ as given in (8).

- **Tracking (T)**: An incoming failed type k item is sent to the vendor at which the expected number of type k items under policy $(\boldsymbol{p}_1^*, \boldsymbol{p}_2^*)$ minus the number of existing type k items is the largest, i.e., keep the number of failed items at each vendor as close as possible to the expected number of failed items under the OSI policy.

- **Generalized Join the Shortest Queue policy (GJSQ)**: An incoming failed type k item is sent to the vendor with the smallest type k index. The indices are defined in (15) and (16).

Our simulation programs were written in SIMSCRIPT II.5, and we use LABATCH.2 (Fishman [8]) to calculate the 95% confidence intervals of the average cost. Each simulation collects data from 2,000 independent replications. Each replication runs for a duration of 5 years and outputs the average yearly cost.

Following Opp, Kulkarni, and Glazebrook [18], we use Gini coefficient (Gini [9]) as a measure of the uniformity of the optimal state-independent allocation, which is connected to the performance of the GJSQ policy. The Gini coefficient is widely used in the economic literature as a measure of income inequality. It is a number between 0 and 1, where 0 corresponds to perfect equality and 1 corresponds to perfect inequality. We use it to measure the inequality of the distribution of repairs among vendors.

The Gini coefficient is calculated using the Lorenz curve (Lorenz [15]), which is a graphical representation of income inequality. In the context of warranty repair allocation, it can be explained as follows. Let x-axis correspond to the percentage of vendors and y-axis correspond to the percentage of repair allocation. The Lorenz curve is a piecewise linear function that contains point (x, y) if the bottom $x\%$ of vendors have $y\%$ of the total repairs (see Figure 2). In the case of perfect equality, every vendor gets the same number of repairs and the Lorenz curve becomes the 45° line, which is called the *perfect equality line*. The Gini coefficient is the ratio between the area enclosed by the perfect equality line and the Lorenz curve, and the total area under the perfect equality line.

We illustrate the concepts of Lorenz curve and Gini coefficient using a small warranty repair allocation example. Suppose three vendors provide repair services for two types of items. Sales of type 1 items form a $PP(100)$ and sales of type 2 items form a $PP(300)$. The warranty length and failure rate are the same for both types of items. The optimal state-independent policy is $(\boldsymbol{p}_1^*, \boldsymbol{p}_2^*)$, with $\boldsymbol{p}_1^* = (0.2, 0.6, 0.2)$ and $\boldsymbol{p}_2^* = (0.1, 0.5, 0.4)$. We measure the distribution inequality among vendors in terms of total number of repairs (of both types) assigned. Therefore, on average the percentage of total repairs assigned to vendor 1, 2, and 3 are 12.5%, 52.5%, and 35%, respectively. Sorting the vendors in ascent order of repair assignment, we can see that the lowest (33.3%) vendor gets 12.5% of the total assignment, the lowest two (66.7%) vendors get 47.5% of the total assignment, and the lowest three (100%) vendors get 100% of the total assignment. Therefore, the Lorenz curve is a piecewise linear function that connects points $(0, 0), (33.3, 12.5), (66.7, 47.5)$, and $(100, 100)$ as shown in Figure 2.

In general, suppose there are V vendors providing repair services for K class of items. Failures from class k items occur according to $PP(\phi_k)$, $k = 1, 2, \cdots, K$, and the optimal state-independent policy is $(\boldsymbol{p}_1^*, \boldsymbol{p}_2^*, \cdots, \boldsymbol{p}_K^*)$, where $\boldsymbol{p}_k^* = (p_{k1}^*, p_{k2}^*, \cdots, p_{kV}^*)$. Then the Gini coefficient can be calculated using the following formula:

$$G = \frac{\sum_{i=j+1}^{V} \sum_{j=1}^{V} | \sum_{k=1}^{K} \phi_k p_{ki}^* - \sum_{k=1}^{K} \phi_k p_{kj}^* |}{V \sum_{k=1}^{K} \phi_k}.$$

Next we present the simulation results as a function of the Gini coefficient of the optimal state-independent allocation.

We simulate a system with 2 types of items and 3 vendors. Sales of each type of items occur according to a Poisson process with rate 200 items per year. Both types of items are covered under warranty for 1 year and have a failure rate of 1.5 failures per item per year. The holding

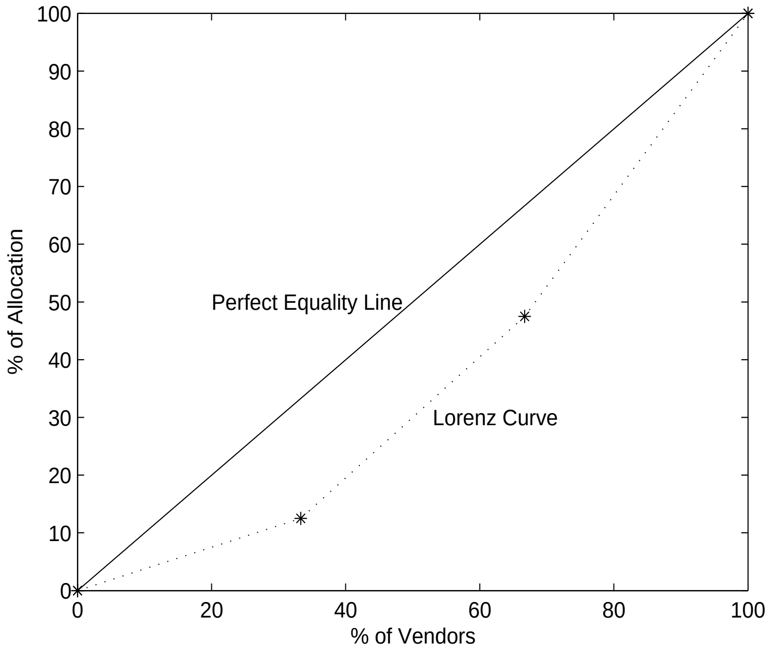

Figure 2: Lorenz curve and perfect equality line

cost rates are the same across all vendors with $h_{1i} = 500, h_{2i} = 300, i = 1, 2, 3$. The fixed cost at each vendor is randomly generated from distribution $U[20, 150]$. A total service rate is randomly generated from distribution $U[605, 1210]$ and is randomly distributed among 3 vendors. Note that the total service rate is guaranteed to be larger than the expected total failure rate, therefore the system is always stable. Since there are only 3 vendors, the Gini coefficient of the optimal state-independent allocation ranges from 0 to $\frac{2}{3}$. 30 random examples are generated for each of the following Gini coefficient ranges: $[0, 0.1), [0.1, 0, 2), \cdots, [0.5, 0.6)$, and 20 random examples are generated for Gini coefficient rage $[0.6, \frac{2}{3}]$. Call cases with Gini coefficient < 0.5 *non-extreme cases*, and those with Gini coefficient ≥ 0.5 *extreme cases*.

Tables 2, 3, and 4 summarize the cost reductions by using the GJSQ policy instead of the other three heuristics. These tables show the minimum, maximum, and mean percent cost reductions among all 200 cases, among the 150 non-extreme cases, and among the 50 extreme cases. As we can see that on average the GJSQ policy performs better than the other heuristics, and there are instances for which the GJSQ policy provides remarkably significant savings over the other heuristics. The cost reduction provided by the GJSQ policy is even larger (except for the maximum and mean reductions over JSQ) when restricted to the non-extreme cases. There are a small number of instances for which the GJSQ policy costs slightly more than the other heuristics, most of which are extreme cases.

Figure 3, 4, and 5 plot the percent cost reductions against Gini coefficients of the optimal state-independent allocations. A straight line is fitted to the data points using the ROBUST-FIT function provided by MATLAB, which uses robust linear regression that is less sensitive to outliers in the data as compared with ordinary least squares regression. Plots 4 and 5 (corresponding to OSI and T, respectively) show a downward trend in savings as Gini coefficient increases. Plot 3 (corresponding to JSQ) shows a slightly upward trend. These observations are consistent with the results summarized in the tables.

Table 2 shows the comparison results between the GJSQ policy and the JSQ policy. Among all 200 cases, the GJSQ policy provides an average cost saving of 9.85% over the JSQ policy. Since the JSQ policy tends to allocate items evenly among vendors, it is expected to perform very poorly in extreme cases. In another word, the GJSQ policy has a greater chance to provide large cost reduction over the JSQ policy in extreme cases. This intuition explains the fact that, when restricted to non-extreme cases, the minimum reduction improves but the maximum and mean reductions reduce. Plot 3 shows the upward trend although it is not statistically significant. Out of the 150 non-extreme cases, the GJSQ policy provides positive cost reductions in 139 cases. Out of the 50 extreme cases, the GJSQ policy provides positive cost reductions in 20 cases. The negative cost savings are all relatively small (-2.13% in the worst case), while the positive cost savings can be very large (up to 63.5%). As a result, although the GJSQ policy gives negative savings in 60% of the extreme cases, the average cost saving among extreme cases is still positive (14.30%).

	All cases	Cases with Gini coefficient < 0.5	Cases with Gini coefficient ≥ 0.5
Min. reduction	-2.15%	-1.19%	-2.15%
Max. reduction	63.54%	48.83%	63.50%
Mean reduction	9.85%	8.37%	14.30%

Table 2: Cost reduction of GJSQ over JSQ

Table 3 shows the comparison results between the GJSQ policy and the OSI policy. Among all 200 cases, the GJSQ policy provides an average cost saving of 3.49% over the OSI policy. Among the 150 non-extreme cases, the average cost saving is 4.57%. Plot 4 shows the downward trend which is statistically significant at 99% level. The GJSQ policy provides positive cost reductions in all non-extreme cases. Out of the 50 extreme cases, the GJSQ policy provides positive cost reductions in 47 cases. One may argue that the GJSQ policy should never perform worse than the OSI policy, since it improves on top of the optimal state-independent policy by applying a single step of policy improvement. However, when calculating the indices, we ignored the dynamics of the system by assuming the number of functioning items under warranty stays constant and is always the expected number of items under warranty in steady state. This simplifying assumption as well as the error introduced by simulation cause the seemingly lawbreaking behavior.

	All cases	Cases with Gini coefficient < 0.5	Cases with Gini coefficient ≥ 0.5
Min. reduction	-3.05%	0.098%	-3.05%
Max. reduction	18.24%	18.24%	7.79%
Mean reduction	3.49%	4.57%	0.24%

Table 3: Cost reduction of GJSQ over OSI

Table 4 shows the comparison results between the GJSQ policy and the T policy. Among all 200 cases, the GJSQ policy provides an average cost saving of 4.23% over the T policy. Among the 150 non-extreme cases, the average cost saving is 4.87%. Plot 5 shows the downward trend which is statistically significant at 99% level. The GJSQ policy provides positive cost reductions in all non-extreme cases. Out of the 50 extreme cases, the GJSQ policy provides positive cost reductions in 33 cases.

	All cases	Cases with Gini coefficient < 0.5	Cases with Gini coefficient ≥ 0.5
Min. reduction	-2.16%	0.007%	-2.16%
Max. reduction	21.98%	21.98%	15.65%
Mean reduction	4.23%	4.87%	2.31%

Table 4: Cost reduction of GJSQ over T

From the above observations we can see that the GJSQ policy is a very robust and efficient algorithm. It beats the other heuristics on average even when considering only the extreme cases. It can provide significant cost savings over the other heuristics in many cases (up to 63.54% over JSQ, 18.24% over OSI, and 21.98% over T). In the worst case among our 800 random examples, the GJSQ policy costs only 3.05% more.

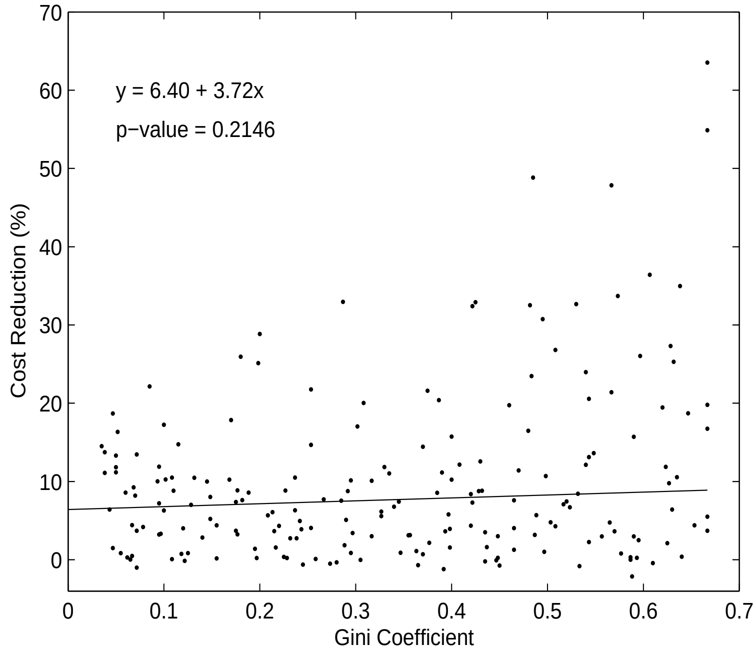

Figure 3: Cost reduction of GJSQ over JSQ

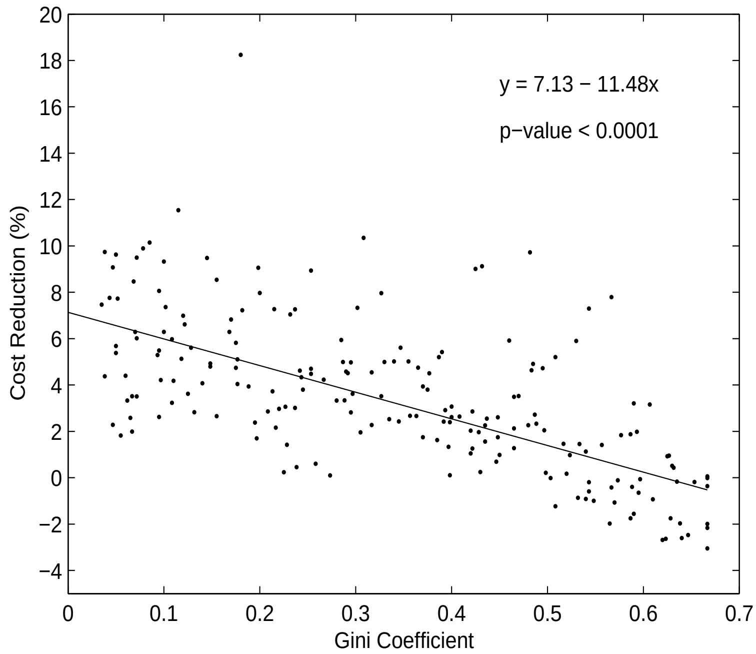

Figure 4: Cost reduction of GJSQ over OSI

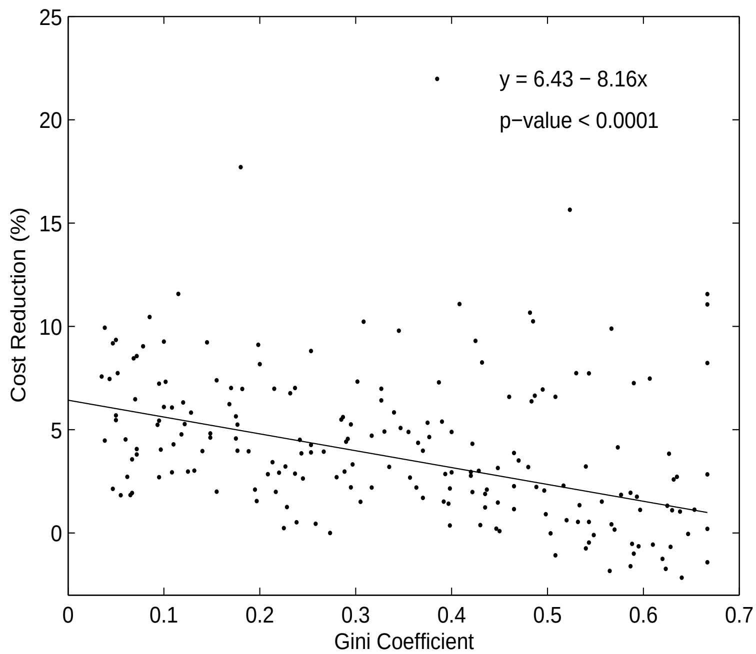

Figure 5: Cost reduction of GJSQ over T

5 Conclusions

In this paper, we addressed the problem of dynamically routing prioritized warranty repairs to multiple vendors. The problem is too complicated to be solved optimally. We presented four heuristics that are applicable to real-life size problems. Among the four heuristics, the index-based GJSQ policy is of our primary interest. The GJSQ policy is developed by applying a single policy improvement step to a judiciously chosen initial state-independent policy. We derived colosed-form expressions for the indicies of the GJSQ policy. We then evaluated and compared the heuristics using simulation. The simulation results suggest that the GJSQ policy is a robust, efficient algorithm to use in practice over a large parameter range. It can provide a significant cost reduction over the other heuristics, especially when the optimal state-independent allocation is relatively uniform among vendors.

6 Appendix: Proof of Theorem 1

We consider the case where there are two types of items and V vendors. Denote by x_{ki} the number of type k items that are currently awaiting or under repair at vendor i, $k = 1, 2$, $i = 1, \cdots, V$. The system state can be written as $(\boldsymbol{x}_1, \boldsymbol{x}_2)$, where $\boldsymbol{x}_k = [x_{k1}, x_{k2}, \cdots, x_{kV}]$,

$k = 1, 2$. Let $\boldsymbol{x}^i = [x_{1i}, x_{2i}]$, $i = 1, \cdots, V$.

Since the optimality of the initial static policy is not crucial to the performance of the GJSQ policy, which is also asserted by Ansell, Glazabrook, and Kirkbride [2], we use the solution $(\boldsymbol{p}_1^*, \boldsymbol{p}_2^*)$ given in (8) as the initial state-independent policy. Following Ansell, Glazabrook, and Kirkbride [2], we improve the policy $(\boldsymbol{p}_1^*, \boldsymbol{p}_2^*)$ by applying one step of policy-improvement, which works as follows. When a type k item fails, we send it to the vendor where the cost increment caused by assigning one type k repair is the smallest assuming that policy $(\boldsymbol{p}_1^*, \boldsymbol{p}_2^*)$ is applied forever afterwards. In another word, for each vendor j we calculate the difference in total cost over an infinite horizon between starting in state $\boldsymbol{x}^j + \boldsymbol{e}_k$ and in state $\boldsymbol{x}^j$ under policy $(\boldsymbol{p}_1^*, \boldsymbol{p}_2^*)$, where $\boldsymbol{e}_k$ is the unit vector whose k^{th} component is 1 with zeros elsewhere. We send the item to the vendor with the smallest index of the required priority type.

Denote by g_i^* the expected cost rate at vendor i in steady state under policy $(\boldsymbol{p}_1^*, \boldsymbol{p}_2^*)$, then

$$g_i^* = f_i(\phi_1 p_{1i}^*, \phi_2 p_{2i}^*), \tag{14}$$

where f_i is defined in (5). Let $T_i(x_{1i}, x_{2i})$ be the time it takes vendor i to reach state $(0, 0)$ for the first time from initial state (x_{1i}, x_{2i}) and let $\tau_i(x_{1i}, x_{2i}) = E[T_i(x_{1i}, x_{2i})]$. Let $J_i(x_{1i}, x_{2i})$ be the expected cost incurred by vendor i starting from initial state (x_{1i}, x_{2i}) until the first time vendor i reaches state $(0, 0)$. Following Ansell, Glazabrook, and Kirkbride [2], we have the index for type 1 items at vendor j is

$$I_{1j}(x_{1j}, x_{2j}) = c_j + J_j(x_{1j} + 1, x_{2j}) - J_j(x_{1j}, x_{2j}) - g_j^*[\tau_j(x_{1j} + 1, x_{2j}) - \tau_j(x_{1j}, x_{2j})], \tag{15}$$

and the index for type 2 items at vendor j is

$$I_{2j}(x_{1j}, x_{2j}) = c_j + J_j(x_{1j}, x_{2j} + 1) - J_j(x_{1j}, x_{2j}) - g_j^*[\tau_j(x_{1j}, x_{2j} + 1) - \tau_j(x_{1j}, x_{2j})], \tag{16}$$

where $j = 1, \cdots, V$.

Next, we derive closed-form expressions for the indices defined in (15) and (16). Consider a fixed vendor, i.e., vendor i for a fixed i. For notational simplicity, we drop the vendor suffix i, e.g., $x_{ki}, \phi_{ki}, \mu_i, h_{ki}$, and c_i are written as x_k, ϕ_k, μ, h_k, and c, respectively, $k = 1, 2$. Thus, the single vendor can be viewed as the following queueing system. Two types of failed items arrive at a single-server queue according to $PP(\phi_k)$, $k = 1, 2$, respectively. Repair times are i.i.d. $\exp(\mu)$ for both types of items. Type 1 items have preemptive resume priority in service over type 2 items. Each failed type k item incurs a holding cost at rate h_k throughout its sojourn time at the vendor and a fixed cost c, $k = 1, 2$. We are interested in computing $J(x_1, x_2)$, the expected total cost incurred by this queueing system starting from initial state (x_1, x_2) until the first time it reaches state $(0, 0)$. We use the following notations in the rest of this section:

$X_k(t)$ = number of type k items in the system at time t, $k = 1, 2$;

$B_1 = \min\{t \geq 0 : X_1(t) = 0 | X_1(0) = 1\}$, i.e., the busy period for serving type 1 items initiated by a single type 1 item;

$B_2 = \min\{t \geq 0 : X_2(t) = 0 | X_1(0) = 0, X_2(0) = 1\}$, i.e., the busy period for serving both types of items initiated by a single type 2 item;

S_2 = the service completion time of a type 2 item accounting for interruptions from type 1 items. Thus if a type 2 item starts service at time 0, it will complete service at time S_2;

$L_1 = \lim_{t\to\infty} E(X_1(t))$, i.e., the expected number of type 1 items in the system;

$L_{1B} = E[\int_0^{B_1} X_1(t)dt]/E(B_1)$, i.e., the expected number of type 1 items in the system during B_1;

$L_2 = \lim_{t\to\infty} E(X_2(t))$, i.e., the expected number of type 2 items in the system;

C_{11} = the expected holding cost incurred by the type 1 items during B_1;

C_{12} = the expected holding cost incurred by the type 1 items during S_2;

$T_1(x_1) = \min\{t \geq 0 : X_1(t) = 0 | X_1(0) = x_1\}$;

$T(x_1, x_2) = \min\{t \geq 0 : X_1(t) = 0, X_2(t) = 0 | X_1(0) = x_1, X_2(0) = x_2\}$;

$\tau_1(x_1) = E(T_1(x_1))$;

$\tau(x_1, x_2) = E(T(x_1, x_2))$;

$H_1(x_1, x_2)$ = the expected total holding cost incurred by the system starting from initial state (x_1, x_2) until time $T_1(x_1)$;

$H_2(x_1, x_2)$ = the expected total holding cost incurred by the system from time $T_1(x_1)$ until time $T(x_1, x_2)$;

$H(x_1, x_2)$ = the expected total holding cost incurred by the system starting from initial state (x_1, x_2) until time $T(x_1, x_2)$.

Also let $\rho_1 = \frac{\phi_1}{\mu}$, $\rho_2 = \phi_2 E(S_2)$, $\eta = \frac{1}{\mu - \phi_1}$, $\xi = \frac{1}{\mu - \phi_1 - \phi_2}$.

Assuming $\phi_1 > 0$ (otherwise, the problems reduces to a single-priority problem), from Prabhu [19] (Chapter 3, Theorem 1) we have

$$E(B_1) = \frac{1}{\mu - \phi_1}, \tag{17}$$

and

$$Var(B_1) = \frac{\phi_1 + \mu}{(\mu - \phi_1)^3}. \tag{18}$$

Obviously,

$$H(x_1, x_2) = H_1(x_1, x_2) + H_2(x_1, x_2)$$

The following two lemmas compute $H_1(x_1, x_2)$ and $H_2(x_1, x_2)$, respectively.

Lemma 1

$$H_1(x_1, x_2) = [\frac{1}{2}h_1\eta + \phi_1 h_1 \eta^2 + \frac{1}{2}(\phi_1 + \mu)\phi_2 h_2 \eta^3]x_1 + \frac{1}{2}(h_1\eta + \phi_2 h_2 \eta^2)x_1^2 + h_2\eta x_1 x_2. \tag{19}$$

Proof Since type 1 items do not see type 2 items in the repair queue, we have

$$L_1 = \frac{\rho_1}{1 - \rho_1} = \frac{\phi_1}{\mu - \phi_1}. \tag{20}$$

Note that ρ_1 is the fraction of time that the server is busy serving type 1 items, therefore

$$L_{1B} = \frac{L_1}{\rho_1} = \frac{\mu}{\mu - \phi_1}. \tag{21}$$

By definitions of L_{1B} and B_1, we have

$$C_{11} = h_1 L_{1B} E(B_1), \tag{22}$$

where $E(B_1)$ is given by (17).

To simplify analysis, we assume that the vendor follows Last-Come-First-Served (LCFS) preemptive service discipline within class 1 items. The assumption of LCFS service discipline is valid because we are interested in total cost, which is independent of the order of service within each class.The assumption of preemption is valid because of the exponential service times. Then $H_1(x_1, x_2)$ can be written as the sum of four parts as follows.

$$H_1(x_1, x_2) = C_{11}x_1 + h_1 E(B_1)\sum_{i=1}^{x_1}(i-1) + h_2 E(B_1)x_1 x_2 + \frac{1}{2}\phi_2 h_2 E(T_1^2(x_1)). \tag{23}$$

The first term includes the holding cost incurred by every initial type 1 item during the busy period initiated by itself and the holding cost incurred by all type 1 items arrive during this busy period. The second term is the holding cost incurred by the initial type 1 items before the busy periods initiated by themselves, since the ith initial type 1 item waits for an expected $(i-1)B_1$ amount of time before its service starts, $i = 1, 2, \cdots, x_1$. The third term is the holding cost incurred by x_2 initial type 2 items during $[0, T_1(x_1))$, since the expected waiting time for each of them is $x_1 B_1$. The last term is the holding cost incurred by newly arrived type 2 items during $[0, T_1(x_1))$, since, conditioned on $T_1(x_1) = t$, the expected number of type 2 items arrived during $[0, t)$ is $\phi_2 t$ and the expected waiting time for each of them during $[0, t)$ is $\frac{1}{2}t$.

Note that $T_1(x_1)$ is the busy period for serving type 1 items initiated by x_1 type 1 items. Thus

$$E(T_1^2(x_1)) = Var(T_1(x_1)) + E^2(T_1(x_1)) = x_1 Var(B_1) + (x_1 E(B_1))^2, \tag{24}$$

where $E(B_1)$ and $Var(B_1)$ are given by (17) and (18).

Substituting (17), (21), (22), and (24) into (23), after some algebra one can show (19) holds. ■

Lemma 2

$$\begin{aligned} H_2(x_1, x_2) &= [\phi_1\phi_2 h_1 \eta^2 \xi + \frac{1}{2} h_2 \xi (2\phi_2 \eta + (\mu + \phi_1)\phi_2^2 \eta^3) + \phi_2^2 h_2 \mu \eta^2 \xi^2] x_1 \\ &+ (\phi_1 h_1 \eta \xi + \frac{1}{2} h_2 \xi + \phi_2 h_2 \mu \eta \xi^2) x_2 + \frac{1}{2}\phi_2^2 h_2 \eta^2 \xi x_1^2 + \phi_2 h_2 \eta \xi x_1 x_2 + \frac{1}{2} h_2 \xi x_2^2. \end{aligned} \tag{25}$$

Proof Because of its lower priority, a type 2 item's service may be interrupted by newly arrived type 1 items. The expected number of interruptions during one service completion time is $\frac{\phi_1}{\mu}$ and each interruption lasts for B_1 amount of time. Hence,

$$C_{12} = \frac{\phi_1}{\mu} h_1 L_{1B} E(B_1). \tag{26}$$

Since the service of a type 2 item can only be interrupted by newly arrived type 1 items and items of both types require the same service time, S_2 has the same distribution as B_1. So

$$E(S_2) = E(B_1) = \eta, \tag{27}$$

and

$$Var(S_2) = Var(B_1) = (\phi_1 + \mu)\eta^3.$$

Type 2 items view the system as an $M/G/1$ queue with $PP(\phi_2)$ arrival and i.i.d. service times with mean $E(S_2)$ and variance $Var(S_2)$. From Kulkarni [14] (Theorem 7.11), we know

$$L_2 = \rho_2 + \frac{\rho_2^2}{2(1-\rho_2)}(1 + \frac{Var(S_2)}{E^2(S_2)}) = \phi_2 \eta + \phi_2^2 \mu \eta^2 \xi. \tag{28}$$

From Prabhu [19] (Chapter 7, Theorem 8), we have

$$E(B_2) = \frac{1}{\mu - \phi_1 - \phi_2}. \tag{29}$$

The system state at time $T_1(x_1)$ can be written as $(0, x_2 + K)$, where K is the number of type 2 items arrive during $[0, T_1(x_1))$. For a fixed K, denote by H_{2K} the holding cost incurred by the queueing system starting from state $(0, x_2 + K)$ until state $(0, 0)$ is reached. Assuming LCFS preemptive service discipline within class 2 items, H_{2K} can be written as the sum of three parts as follows.

$$H_{2K} = (x_2 + K)\frac{E(B_2)}{E(S_2)}C_{12} + (x_2 + K)h_2 E(B_2)\frac{L_2}{\rho_2} + h_2 E(B_2)\sum_{i=1}^{x_2+K}(i-1). \tag{30}$$

The first term is the holding cost incurred by all type 1 items during this period, since $\frac{E(B_2)}{E(S_2)}$ is the expected total number of type 2 items served during B_2. The second term includes the holding cost incurred by every existing type 2 item during the busy period initiated by itself and the holding cost incurred by all type 2 items arrive during this busy period, since $\frac{L_2}{\rho_2}$ is the average number of type 2 items in the system during a busy period initiated by a single type 2 item. The third term is the holding cost incurred by the $x_2 + K$ existing type 2 customers before the busy periods initiated by themselves, since the ith existing type 2 customer waits for an expected $(i-1)E(B_2)$ amount of time before its service starts.

Substituting (26), (27), (28), and (29) into (30), after some algebra, we get

$$\begin{aligned} H_{2K} &= K\phi_1 h_1 \eta\xi + \frac{1}{2}h_2 K(K+1)\xi + K h_2 \phi_2 \mu\eta\xi^2 \\ &\quad + [h_1\phi_1\eta\xi + \frac{1}{2}h_2(2K+1)\xi + h_2\phi_2\mu\eta\xi^2]x_2 + \frac{1}{2}h_2\xi x_2^2 \end{aligned} \tag{31}$$

For K, the number of type 2 items arrive during the busy period started by x_1 type 1 items, we have

$$E(K) = x_1 E(B_1)\phi_2, \tag{32}$$

and

$$\begin{aligned} E(K^2) &= E[E(K^2|T_1(x_1))] = E[Var(K|T_1(x_1)) + E^2(K|T_1(x_1))] \\ &= E(\phi_2 T_1(x_1) + \phi_2^2 T_1^2(x_1)) = \phi_2 E(T_1(x_1)) + \phi_2^2 E(T_1^2(x_1)). \end{aligned}$$

Plugging in (24), we get

$$E(K^2) = [\phi_2\eta + \phi_2^2(\phi_1 + \mu)\eta^3]x_1 + \phi_2^2\eta^2 x_1^2. \tag{33}$$

Obviously, $H_2(x_1,x_2) = E_K(H_{2K})$. Taking expectation on both sides of (31) with respect to K and plugging in (32) and (33), one can show that (25) holds. ■

The above results allow us to computer $J(x_1,x_2)$ as given in the following lemma.

Lemma 3

$$J(x_1,x_2) = Ax_1 + Bx_2 + Cx_1^2 + Dx_1x_2 + Ex_2^2, \tag{34}$$

where

$$\begin{aligned}
A &= [c(\phi_1+\phi_2)+\frac{1}{2}h_1]\eta + \phi_1 h_1\eta^2 + [c(\phi_1+\phi_2)\phi_2 + \phi_2 h_2]\eta\xi + \frac{1}{2}(\phi_1+\mu)\phi_2 h_2\eta^3 \\
&\quad +\phi_1\phi_2 h_1\eta^2\xi + \frac{1}{2}(\phi_1+\mu)\phi_2^2 h_2\eta^2\xi^2 + \phi_2^2\mu h_2\eta^2\xi^2, \\
B &= [\frac{1}{2}h_2 + c(\phi_1+\phi_2)]\xi + \phi_1 h_1\eta\xi + \phi_2 h_2\mu\eta\xi^2, \\
C &= \frac{1}{2}(h_1\eta + \phi_2 h_2\eta^2 + \phi_1^2 h_2\eta^2\xi), \\
D &= h_2\eta + \phi_2 h_2\eta\xi, \\
E &= \frac{1}{2}h_2\xi.
\end{aligned}$$

Proof Denote by $C(x_1,x_2)$ the expected total fixed cost generated by this queueing system starting from initial state (x_1,x_2) until state $(0,0)$ is reached for the first time. Then

$$C(x_1,x_2) = c(\phi_1+\phi_2)[x_1E(B_1) + (x_2+E(K))E(B_2)] = c(\phi_1+\phi_2)[(1+\phi_2\xi)\eta x_1 + \xi x_2]. \tag{35}$$

The total cost $J(x_1,x_2)$ can be written as the sum of three parts as follows

$$J(x_1,x_2) = H_1(x_1,x_2) + H_2(x_1,x_2) + C(x_1,x_2). \tag{36}$$

Substituting (19), (25), and (35) into (36), after some algebra, one can show (34) holds. ■

We also have

$$\tau(x_1,x_2) = x_1E(B_1) + (x_2+E(K))E(B_2).$$

Plugging in (17), (29), and (32), we get

$$\tau(x_1,x_2) = \frac{x_1+x_2}{\mu-\phi_1-\phi_2}. \tag{37}$$

Substituting (14), (34), and (37) into (15) and (16), Theorem 1 follows after some algebra.

References

[1] Ahuja, R. K., T. L. Magnanti, and J. B. Orlin. *Network Flows.* Prentice Hall, New Jersey, 1993.

[2] Ansell, P. S., K. D. Glazebrook, and C. Kirkbride. *Generalised "join the shortest queue" policies for the dynamic routing of jobs to multiclass queues.* Journal of the Operational Research Society, 2003, 54, 379-389.

[3] Blischke, W. R. and D. N. P. Murthy. *Warranty Cost Analysis.* Marcel Dekker, New York, 1994.

[4] Byrne, P. M. *Making warranty management manageable.* Logistics Management (Highlands Ranch, Co.), August 2004, v43 i8.

[5] Buczkowski, P., V. Kulkarni, and M. Hartmann. *Outsourcing prioritized warranty repairs.* International Journal of Quality & Reliability Management, 22(7): 699 - 714, 2005.

[6] Dimitrov, B., S. Chukova, and Z. Khalil. Warranty costs: an age-dependent failure/repair model. *Naval Research Logistics* vol.51, no.7 : 959-76, Oct. 2004.

[7] Ding, L and Glazebrook, K. D. *A static allocation model for the outsourcing of warranty repairs.* Journal of the Operational Research Society, 56(7): 825-835, 2005.

[8] Fishman, G. S. *Discrete-Event Simulation: Modeling, Programming, and Analysis* Springer-Verlag, New York, 2001.

[9] Gini, C. Variabilità e mutabilità, *Studi economico-giuridici pubblicati per cura della Facoltà di Giurisprudenza della R.Università di Cagliari* III (2a), 1912. Reprinted in Gini (1939) with additional footnotes, pp.189-358.

[10] Gini, C. *Memorie di Metodologica Statistica*, vol.I, *Variabilità e concentrazione.* Milano: Dott.A.Giuffrè Editore, 1939.

[11] Gross, O. A class of discrete type minimization problems. Technical Report RM-1644, RAND Corporation, 1956.

[12] Ibaraki, T. and N. Katoh. *Resource Allocation Problems: Algorithmic Approaches.* MIT Press, Cambridge, MA, 1988.

[13] Krishnan, K. R. Joining the right queue: a Markov decision rule. *Proceedings of the 28th IEEE Conference on Decision and Control*, 1863-1868, 1987.

[14] Kulkarni, V. G. *Modeling and analysis of stochastic systems*. Chapman & Hall, London, UK, 1995.

[15] Lorenz, M. O. *Methods for measuring the concentration of wealth*. Journal of the American Statistical Association, 9:209-219, 1905.

[16] Manna, D. K., S. Pal, and A. Kulandaiyan. Warranty cost estimation of a multi-module product. *International Journal of Quality Reliability Management*, vol.21, no.1 : 102-17, 2004.

[17] Murthy, D. N. P. and I. Djamaludin. New product warranty: a literature review. *International Journal of Production Economics*, vol.79, no.3 : 231-60, 11 Oct. 2002.

[18] Opp, M., V. G. Kulkarni, and K. Glazebrook. Outsourcing Warranty Repairs: Dynamic Allocation. *Nav. Res. Log.* **52**, 381-398, 2005.

[19] Prabhu, N. V. *Foundations of Queueing Theory*. Kluwer Academic Publishers, Boston, 1997.

[20] Puterman, M. L. *Markov Decision Processes*. John Wiley & Sons, New York, 1994.

[21] Tijms, H. C. *Stochastic Models: An Algorithmic Approach*. John Wiley & Sons, New York, 1994.

[22] Yeh, R. H., G. Chen, and M. Chen. Optimal age-replacement policy for nonrepairable products under renewing free-replacement warranty. *IEEE Transactions on Reliability* vol.54, no.1 : 92-7, March 2005.