Dynamic scheduling of a multiclass fluid model with transient overload

Junxia Chang[*] Hayriye Ayhan[†] J. G. Dai[‡] Cathy H. Xia[§]

April 26, 2004

Abstract

We study the optimal dynamic scheduling of different requests of service in a multiclass stochastic fluid model that is motivated by recent and emerging computing paradigms for Internet services and applications. In particular, our focus is on environments with specific performance guarantees for each class under a profit model in which revenues are gained when performance guarantees are satisfied and penalties are incurred otherwise. Within the context of the corresponding fluid model, we investigate the dynamic scheduling of different classes of service under conditions where the workload of certain classes may be overloaded for a transient period of time. Specifically, we consider the case with two fluid classes and a single server whose capacity can be shared arbitrarily among the two classes. We assume that the class 1 arrival rate varies with time and the class 1 fluid can more efficiently reduce the holding cost. Under these assumptions, we characterize the optimal server allocation policy that minimizes the holding cost in the fluid model when the arrival rate function for class 1 is known. Using the insights gained from this deterministic case, we study the stochastic fluid system when the arrival rate function for class 1 is random and develop various policies that are optimal or near optimal under various conditions. In particular, we consider two different types of heavy traffic regimes and prove that our proposed policies are strongly asymptotically optimal. Numerical examples are also provided to demonstrate further that these policies yield good results in terms of minimizing the expected holding cost.

Keywords: Stochastic fluid model, transient overload, e-commerce, quality-of-service, service-level-agreement

1 Introduction

Recent advances in Internet services and other emerging applications have created new computing and networking paradigms in which a set of e-commerce businesses contract with a common hosting provider of Internet applications and services for their respective customers. In such an environment, the hosting service provider needs to meet a diverse set of requirements of the various e-commerce businesses and customers. To address these diverse requirements and leverage

[*]School of ISyE, Georgia Institute of Technology, Atlanta, GA 30332-0205; email: junxiach@isye.gatech.edu

[†]School of ISyE, Georgia Institute of Technology, Atlanta, GA 30332-0205; email: hayhan@isye.gatech.edu. Research supported in part by National Science Foundation grants DMI-9908161 and DMI-9984352.

[‡]School of ISyE, Georgia Institute of Technology, Atlanta, GA 30332-0205; email: dai@isye.gatech.edu. Research supported in part by National Science Foundation grants DMI-9813345 and DMI-0300599, and by TLI-AP, a partnership between National University of Singapore and Georgia Institute of Technology.

[§]IBM T. J. Watson Research Center, P.O. Box 704, Yorktown Heights, NY 10598; email: cathyx@us.ibm.com.

potential economies of scale, the hosting service provider will often deploy a cluster of servers to effectively share the computing and networking resources required to support the desired Internet applications and services. A number of computer industry companies such as HP, IBM and Intel are already providing such hosting services and it appears that more companies will be doing so in the future.

To differentiate the diverse requirements of e-commerce businesses and customers, it is necessary to introduce the notion of different service classes. These service classes typically have distinct levels of importance to the hosting service provider, the businesses and their customers. Moreover, many of these service classes require specific Quality-of-Service (QoS) performance guarantees; failures to deliver such levels of QoS can have a significant impact on the e-commerce businesses and customers. For example, customers may easily lose patience and discontinue using the service if its responsiveness is perceived to be too long. Hence, as part of the contract between the service provider and each business, the hosting service provider agrees to guarantee a certain level of QoS for each class of service, and in return each e-commerce business agrees to pay the service provider for satisfying these QoS performance guarantees. Such Service-Level-Agreements (SLA) are included in service contracts between each business and the service provider, and they specify both performance targets or QoS guarantees, and financial consequences for meeting or failing to meet these targets. A service level agreement may also depend on the anticipated level of per-class workload from the customers of the business.

Thus, it is critical for the hosting service provider to dynamically allocate its server resources to optimize performance and profit measures in cluster-based computing environments with SLA contracts containing QoS performance guarantees. This is also an important issue for the continued growth and success of Internet services and applications. Therefore, in this paper we focus on a particularly important class of dynamic scheduling problems that arise in these computing environments. However, it is important to note that while our analysis and results are motivated by such environments, they apply more generally to a wide variety of emerging computing environments with SLA-based QoS performance guarantees.

Previous studies that address QoS performance guarantees have focused mostly on throughput or mean response time measures. However, a crucial issue for Internet applications and services concerns the per-request efficiency with which the differentiated services are handled, since delays experienced by customers can result in lost revenue and customers for a business as described above. Furthermore, more standard performance metrics such as throughput and mean response time may not fully capture such QoS performance guarantees. In order to address these issues, we consider a general class of SLAs in which a threshold is defined for each class of service such that the hosting service provider gains revenues when the QoS level experienced by the class stays at or below the threshold, but the service provider pays penalties to the corresponding businesses when this threshold is exceeded. Then the optimal control problem focuses on allocating server resources in order to maximize the profit of hosting the collection of e-commerce sites under these SLA constraints.

Another big challenge of the problem concerns the diverse workloads of different e-commerce businesses and their variation over time. It is common in the computing environments of interest to have the workload of certain classes in each e-commerce site alternate between a period during which the arriving workload exceeds the allocated capacity, and a period during which the arriving workload is less than this capacity, even though the average load is within the allocated capacity; e.g., see [2]. These periods of transient overload can have a significant effect on the performance experienced by the different classes of service. This in turn can have a critical impact on the penalties that the hosting service provider is required to pay each e-commerce business according to the SLA contract between them. Hence, it is crucial to include these important workload characteristics in the analysis of the optimal control problem.

Our problem falls within the general class of optimal resource control problems with the foregoing non-conventional performance metrics and workload characteristics. Several researchers have studied the issue of workloads with transient overload, but their studies have focused on single-class workloads and specific scheduling strategies, such as admission control (e.g., [8]) and direct modifications to the Internet server scheduling mechanism (e.g., [2, 5]). On the contrary, our focus in this paper is on the optimal dynamic scheduling of a multiclass system with transient overload. Furthermore, little has been done to consider the issue of maximizing profit in these computing paradigms under non-conventional performance metrics. The primary exception is the study in [11], which develops queueing-theoretic bounds and approximations to formulate the resource control optimization problem and then develops efficient algorithms to compute the optimal solution. This study is the one that is most relevant to our research, but it differs from the present study in several important aspects. Our focus in this paper is on computing the optimal dynamic scheduling policy and gaining insights into its fundamental properties, as opposed to computing the steady-state solution, and to do so under a workload with transient overload, which is not considered in [11].

Our primary concern in this paper is to investigate the preceding optimal server resource control problem as a dynamic scheduling problem. Our motivation behind considering a fixed time horizon is that in reality many web sites exhibit [12] regular daily access patterns, typically there is one single peak period each day, the low period load is far below the system capacity so that the system usually starts empty the next day. Distributed architectures with separate machines for different geographical locations are also common in practice in order to improve the response time for accessing data over the Internet. This again validates the single period model. Hence, the traffic from the previous period does not have an effect on the next period. Our approach is based on formulating the problem as a multiclass stochastic fluid model and employing optimal control theory [13, 14] to search for the optimal control policy that maximizes the total revenue over a fixed time horizon. Even though recent studies of a similar spirit for different dynamic scheduling problems include [1, 3, 6, 17], to the best our knowledge, no optimal scheduling policy is known for the general problem considered herein. As mentioned above, we focus on minimizing the penalty of the hosting service provider by dynamically scheduling its server resources among the fluid classes in a system that can be overloaded for a transient period. In order to capture the QoS performance guarantees in the SLA contracts, we introduce a threshold value for each fluid class such that a holding cost is incurred only if the amount of fluid of a certain class exceeds its threshold value. Preliminary results on this problem can be found in an earlier work [4]. In this paper, we consider the specific case of two fluid classes and a single server whose capacity can be shared arbitrarily among the two classes. We assume that the class 1 arrival rate changes with time and the class 1 fluid can more efficiently reduce the holding cost and develop the optimal server resource allocation policy that minimizes the holding cost in the corresponding fluid model when the arrival rate function for class 1 is known. We then study the stochastic fluid system when the arrival rate function for class 1 is random and propose various policies that are optimal or near optimal under various conditions. In particular, we consider two different types of heavy traffic regimes and prove that our proposed policies are *strongly asymptotically optimal* in the following sense: the difference between its performance and the optimality is *bounded* from above by a constant even as the optimal value itself goes to infinity. This notion of strong asymptotic optimality has also been considered in [15, 18], as a measure to evaluate the closeness to optimality of approximating control policies. Numerical examples are also provided to demonstrate further that these policies yield good results in terms of minimizing the expected holding cost.

The outline of the paper is as follows. We define our multiclass fluid model in §2. Deterministic instance of the model is analyzed in §3 where we provide the optimal control policy. §4 and §5 consider the stochastic instance of the model. In §4, we present a discrete review policy and show that it is asymptotically optimal as the expected length of the high period tends to infinity. Other policies that are asymptotically optimal are further discussed in §5. Our concluding remarks are provided in §7. Throughout, proofs are relegated to the Appendix.

2 The stochastic fluid model

This paper focuses on the following stochastic fluid system that serves two classes of fluid. Each class fluid continuously arrives at its buffer whose capacity is assumed to be infinite. Both classes are served by a single server whose service capacity can be shared arbitrarily among the two classes. When the server devotes full effort to class i, it processes class i fluid at rate μ_i, $i = 1, 2$.

Class 2 fluid arrives at a constant rate λ_2 throughout the time horizon under consideration. Class 1 fluid has a high arrival rate λ_1^h during the first part of the time interval and a low arrival rate λ_1^l in the rest of the time interval. Naturally, $\lambda_1^l \leq \lambda_1^h$. The durations of the first and second time intervals are denoted by H and L, respectively. Both H and L are random. Some of their statistics like mean remaining life times are assumed to be known. These assumptions will be spelled in more precise terms later in the paper. We call the time interval $[0, H)$ the high load period and the time interval $[H, H+L)$ the low load period.

We use $Z_i(t)$ to denote the fluid level in class i at time t, and $T_i(t)$ to denote the cumulative amount of time in $[0, t]$ that the server spends on class i fluid, $i = 1, 2$. The dynamics of the fluid model is given by the following equations

$$Z_i(t) = Z_i(0) + \int_0^t \lambda_i(s)\, ds - \mu_i T_i(t), \quad t \in [0, H+L), \tag{1}$$

$$T_i(0) = 0, \quad T_i(t) \text{ is a nondecreasing function of } t, \tag{2}$$

$$t - (T_1(t) + T_2(t)) \text{ is a nondecreasing function of } t, \tag{3}$$

where $\lambda_i(s)$ is the arrival rate to class i at time s. Since the class 1 arrival rate function $\lambda_1(\cdot)$ is random, the fluid level process Z is random as well. The allocation process $T = \{(T_1(t), T_2(t)), t \geq 0\}$ reflects how the server spends its service capacity among two classes and it is called a scheduling or a service policy.

Let $h_i > 0$ and $\theta_i \geq 0$ be constants, $i = 1, 2$. For a real number x, define $x^+ = \max(x, 0)$. Consider the integral

$$\int_0^{H+L} \sum_{i=1}^{2} h_i \left(Z_i(t) - \theta_i\right)^+ \, dt \tag{4}$$

which is called the total cost of the system. Then one interprets h_i as the holding cost per unit time when the fluid level in class i exceeds θ_i. If the fluid level in class i is below θ_i, the fluid does not accumulate cost for the system. Clearly, the cost depends on initial fluid level $z = Z(0)$, and allocation T employed. Since H and L are random variables, the cost is also random. The focus of this paper is to find an allocation T to minimize the expected total cost for each initial point z. We assume that working on class 1 can more efficiently reduce holding costs. Namely,

$$h_1 \mu_1 > h_2 \mu_2. \tag{5}$$

If the assumption in (5) is violated, the optimal policy is a generalization of the well-known $c\mu$ rule (see for example, Smith [16], Klimov [9] and Green and Stidham [7]). Details of such an optimal policy are discussed in Appendix A.

When $\theta_i = 0$ for $i = 1, 2$, the optimal policy is again given by the $c\mu$-rule. That is the server gives priority to class i with highest $h_i\mu_i$. To the best our knowledge, the optimal policy for our general problem is not known. In the special case when H and L are deterministic, and are known at the beginning of the time window, we will present an optimal policy. Using this policy, we will construct heuristic policies, known as discrete review policies, for controlling the system. We will present numerical experiments showing that these policies perform well. We will establish asymptotic results guaranteeing good performance of these policies in certain parameter regions. We will also identify other policies that are asymptotically optimal in certain parameter regions.

For any feasible allocation T, it follows that $T(t)$ is Lipschitz continuous in t. Thus, T is absolutely continuous and has derivatives almost everywhere. Therefore, specifying an allocation T is equivalent to specifying its derivative $\dot{T}(t)$ for almost every t in $(0, H + L)$. (For a function f, $\dot{f}(t)$ denotes the derivative of f at time t. Whenever $\dot{f}(t)$ is used, the derivative of f at time t is assumed to exist.) Clearly, any feasible allocation T should be non-anticipating. Namely, $\dot{T}(t)$ depends only on the information available up to time t.

For future reference, we also define the traffic intensities of the system. The system load per unit of time contributed by class 1 fluid is $\rho_1^h = \lambda_1^h/\mu_1$ for the high load period and $\rho_1^l = \lambda_1^l/\mu_1$ for the low load period. The system load per unit of time contributed by class 2 fluid is constant and given by $\rho_2 = \lambda_2/\mu_2 > 0$. The overall system load is $\rho^h = \rho_1^h + \rho_2$ for the high load period and $\rho^l = \rho_1^l + \rho_2$ for the low load period. When $\rho^h > 1$ and $\rho^l < 1$, the total system work increases in the high load period and decreases in the low load period. In this case, the high load period is also called the overload period. Thus, when $\rho^h > 1$ and $\rho^l < 1$ the system experiences an overload period followed by an under-load period, a phenomenon known as transient overload in literature; see, for example, [2]. Although understanding transient overload is the primary motivation of this paper, except explicitly stated otherwise, we do not assume $\rho^h > 1$.

3 Optimal policies in the deterministic case

In this section, we present the optimal policy when the lengths of the high period and the low period are known. Thus, H and L are deterministic quantities. The optimality of this policy is proven in Appendix B. For convenience, we first define the following policy.

Definition 1. The following policy referred as the *Low-period-policy* is implemented in the low period, i.e, when $H < t \le H + L$.

- If $Z_1(t) > \theta_1$, full capacity is given to class 1, i.e. $\dot{T}_1(t) = 1, \dot{T}_2(t) = 0$.
- If $Z_1(t) = \theta_1,\ Z_2(t) > \theta_2$, class 1 fluid is kept at its threshold value θ_1, while the remaining capacity is used to serve class 2, i.e. $\dot{T}_1(t) = \rho_1^l,\ \dot{T}_2(t) = 1 - \rho_1^l$.
- If $Z_1(t) < \theta_1,\ Z_2(t) > \theta_2$, then full capacity is given to class 2, i.e. $\dot{T}_1(t) = 0,\ \dot{T}_2(t) = 1$.
- If $Z_1(t) \le \theta_1,\ Z_2(t) \le \theta_2$, then the policy is not unique and $\dot{T}_1(t)$ and $\dot{T}_2(t)$ can be chosen from any solution satisfying $\dot{T}_1(t) \ge \rho_1^l,\ \dot{T}_2(t) \ge \rho_2$ and $\dot{T}_1(t) + \dot{T}_2(t) \le 1$.

The optimal policy depends on the system load. In the next three subsections, we will describe the optimal policy under all load conditions. In the first case, $\rho_1^h > 1, \rho^l \le 1$; in the second case, $\rho^h > 1,\ \rho_1^h \le 1,\ \rho^l \le 1$; and in the last case, $\rho^h \le 1, \rho^l \le 1$.

3.1 The case $\rho_1^h > 1,\ \rho^l \leq 1$

Suppose that $\rho_1^h > 1$ and $\rho^l \leq 1$. Then the optimal policy has the following structure:

(OPT)

$$\begin{aligned}
\forall t \in (0,\ s_1): &\quad \dot{T}_2(t) = 1,\ \dot{T}_1(t) = 0;\\
\forall t \in (s_1,\ s_2): &\quad \dot{T}_2(t) = u_2,\ \dot{T}_1(t) = u_1, u_1 + u_2 = 1;\\
\forall t \in (s_2,\ H): &\quad \dot{T}_2(t) = 0,\ \dot{T}_1(t) = 1;\\
\forall t \in (H,\ H+L): &\quad \text{Low-period-policy.}
\end{aligned}$$

Thus, the optimal policy gives fixed priority to class 2 in the interval 0 to s_1, employs processor sharing in the interval s_1 to s_2 and gives fixed priority to class 1 in the interval s_2 to H. Specific values of s_1, s_2, u_1, and u_2 depend on the initial fluid levels and the length of the high and the low periods. Before discussing the computation of s_1, s_2, u_1 and u_2 for all possible cases, we introduce the notation used in our developments:

$$d_1 = \theta_1 - Z_1(0), \quad \psi_1 = \tfrac{d_1/\mu_1}{\rho_1^h - 1}, \quad \tilde{\psi}_1 = \frac{d_1/\mu_1}{\rho_1^h}, \tag{6}$$

$$d_2 = \theta_2 - Z_2(0), \quad \psi_2 = \tfrac{d_2/\mu_2}{\rho_2}, \quad \tilde{\psi}_2 = \frac{-d_2/\mu_2}{1-\rho_2}. \tag{7}$$

The quantities ψ_1, ψ_2, $\tilde{\psi}_1$ and $\tilde{\psi}_2$ have the following interpretations. Quantity ψ_1 is the time that class 1 increases to its threshold θ_1 under the policy that gives fixed priority to class 1 if the initial fluid level of class 1 is below θ_1 and if the high period is long enough. Quantity $\tilde{\psi}_1$ is the time class 1 increases to its threshold θ_1 under the policy that gives fixed priority to class 2 if the initial fluid level of class 1 is below θ_1 and if the high period is long enough. Quantity ψ_2 is the time class 2 increases to its threshold θ_2 under the policy that gives fixed priority to class 1 if the initial fluid level of class 2 is below θ_2. Finally, $\tilde{\psi}_2$ is the time class 2 decreases to its threshold θ_2 under the policy that gives fixed priority to class 2 if the initial fluid level of class 2 is above θ_2. Clearly, d_1 and d_2 denote the initial deviation of the fluid levels from the desired thresholds for classes 1 and 2, respectively.

We also define

$$a_1 = \frac{d_1/\mu_1 + d_2/\mu_2}{\rho_1^h + \rho_2 - 1}, \quad a_2 = \frac{1-\eta\xi}{1-\eta}\psi_1^+ - \frac{\eta(1-\xi)}{1-\eta}\psi_2^+, \tag{8}$$

$$B = \frac{1-\eta\xi}{1-\eta}\psi_1^+ - \frac{(1-\rho_1^l)[1+\eta(\rho_1^h-1)] + (1-\eta)(\rho_1^h-1)}{(\rho_1^h-1)(\rho_1^h-\rho_1^l)(1-\eta)}\tilde{\psi}_2^+, \tag{9}$$

where

$$\xi = \frac{(\rho_1^h - 1)}{(\rho_1^h - \rho_1^l)} \text{ and } \eta = \frac{h_2\mu_2}{h_1\mu_1}.$$

Quantities a_1, a_2 and B have the following interpretations. Quantity a_1 is the critical value such that if the high period is longer than a_1 then under any policy either class 1 fluid level will exceed its threshold θ_1 or class 2 fluid level will exceed its threshold θ_2. Quantity a_2 is the critical value such that if the high period is longer than a_2 and the low period is long enough to reduce the fluid level of class 1 to its threshold θ_1 then fixed priority to class 1 is the optimal policy in the high period. Finally, B is the critical value such that if the high period is longer than B and the low period is long enough to reduce the fluid level of class 1 to its threshold θ_1 then the optimal policy never uses processor sharing in the high period. Finally, for the sake of simplicity, we define

$$\gamma_1 = \frac{\eta(\rho_1^h-1)(\rho_1^h+\rho_2-1)}{(1-\rho_1^l)[\rho_2+\eta(\rho_1^h-1)]+(1-\eta)\rho_2(\rho_1^h-1)}, \quad \gamma_2 = \frac{\eta\rho_1^h(\rho_1^h-1)}{(1-\rho_1^l)[1+\eta(\rho_1^h-1)]+(1-\eta)(\rho_1^h-1)}, \quad \gamma_3 = \frac{\rho_1^h-1}{1-\rho_1^l}.$$

We now provide a more detailed description of the optimal policy by considering all possible cases of the initial load. As can be seen below, Cases 1 and 3 are simple and have no subcases (i.e the policy is independent of the length of H and L). However, Cases 2 and 4 have many subcases. Hence, for the sake of clarity, we provide pictorial representations of Cases 2 and 4 in Figures 1 to 3. In particular, we present the corresponding case for each value of H and L and demonstrate that we consider all possible values for the length of the high and low periods. Depending on the relationship between $\tilde{\psi}_1$ and $\tilde{\psi}_2$, we provide the corresponding pictorial representation of Case 2, respectively in Figures 1 and 2. Figure 3 is the pictorial representation of Case 4.

- Case 1: $Z_1(0) \geq \theta_1$. In this case, the optimal policy is given by (OPT) with $s_1 = s_2 = 0$. Note that when setting $s_1 = s_2 = 0$, the (OPT) policy gives fixed priority to class 1 throughout the high period.

- Case 2: $Z_1(0) < \theta_1$, $Z_2(0) > \theta_2$. Computation of s_1, s_2, u_1 and u_2 depends on the length of the high and the low periods.

  - Case 2.1: If

$$a_1 \leq H \leq B, \quad L \geq \gamma_1(H - a_1), \tag{10}$$

then s_1, s_2, u_1 and u_2 are computed by solving

$$Z_2(0) + (\lambda_2 - \mu_2)s_1 = \theta_2, \tag{11}$$
$$Z_1(0) + \lambda_1^h s_1 = Z_1(s_1), \tag{12}$$
$$Z_2(s_1) + (\lambda_2 - \mu_2 u_2)(s_2 - s_1) = \theta_2, \tag{13}$$
$$Z_1(s_1) + (\lambda_1^h - \mu_1 u_1)(s_2 - s_1) = Z_1(s_2), \tag{14}$$
$$u_1 + u_2 = 1, \tag{15}$$
$$Z_1(s_2) + (\lambda_1^h - \mu_1)(t_1 - s_2) = \theta_1, \tag{16}$$
$$Z_1(t_1) + (\lambda_1^h - \mu_1)(H - t_1) = Z_1(H), \tag{17}$$
$$Z_1(H) + (\lambda_1^l - \mu_1)(t_2 - H) = \theta_1, \tag{18}$$
$$\mu_1 h_1(t_2 - t_1) = \mu_2 h_2(t_2 - s_2). \tag{19}$$

Note that equations (11) to (18) describe the evolution of the fluid levels of class 1 and class 2 from time 0 to t_2 under the optimal policy, where t_2 represents the time epoch at which the class 1 fluid level in the low period reaches its threshold value as indicated in equation (18). In particular, equations (11) and (12) describe the evolution of fluid levels from time 0 to s_1 when higher priority is given to class 2. At s_1, class 2 fluid level is reduced to its threshold θ_2 from above. Equations (13) to (15) describe the evolution of the fluid levels from s_1 to s_2 under the processor sharing policy. In $[s_1, s_2]$, class 2 fluid level remains at its threshold θ_2. Equations (16) to (18) describe the evolution of class 1 fluid level from s_2 to t_2 under the policy that gives higher priority to class 1. Equation (16) implies that at time t_1, class 1 fluid level increases to its threshold θ_1. Equation

(17) records the class 1 fluid level at the end of the high period. Equation (19) ensures that the profit gained by serving class 1 is equal to the profit lost by not serving class 2. Under the conditions given in (10), it will be shown in Appendix B that equations (11) to (19) have a unique solution with $0 \le s_1 \le s_2 \le t_1 \le H \le t_2 \le H + L$ and $u_1,\ u_2 \ge 0$.

- Case 2.2: If

$$L \le \gamma_1(H - a_1), \quad a_1 \le H, \quad H + L \le \tilde{\psi}_1 + \frac{1 + \eta(\rho_1^h - 1)}{(1-\eta)(\rho_1^h - 1)}(\tilde{\psi}_1 - \tilde{\psi}_2),$$

  then we set $t_2 = H + L$ and compute s_1, s_2, u_1, u_2 and t_1 by solving equations (11)-(17) and (19).

- Case 2.3: If

$$\max\{B, \tilde{\psi}_1\} \le H \le a_2, \quad L \ge \gamma_2(H - \tilde{\psi}_1),$$

  then we set $s_1 = s_2$ and solve the equations (12) and (16)–(19) for s_2, t_1 and t_2.

- Case 2.4: If

$$L \le \gamma_2(H - \tilde{\psi}_1), \quad \max\left\{\tilde{\psi}_1, \tilde{\psi}_1 + \frac{1 + \eta(\rho_1^h - 1)}{(1-\eta)(\rho_1^h - 1)}(\tilde{\psi}_1 - \tilde{\psi}_2)\right\} \le H + L \le \frac{\psi_1}{1-\eta},$$

  then we set $s_1 = s_2$ and $t_2 = H + L$ and compute s_2 and t_1, by solving equations (12), (16)-(17) and (19).

- Case 2.5: If $H \le \max\{a_1, \tilde{\psi}_1\}$, then the optimal policy is given by (OPT) with $s_1 = \min\{\tilde{\psi}_2,\ H\}$, $s_2 = H$, $u_2 = \rho_2$, and $u_1 = 1 - \rho_2$.

- Case 2.6: If $H \ge a_2$ and $H + L \ge (1-\eta)^{-1}\psi_1$, then the optimal policy is given by (OPT) with $s_1 = s_2 = 0$.

- Case 3: $Z_1(0) < \theta_1$, $Z_2(0) \le \theta_2$, $\psi_1 \le \psi_2$. In this case, the optimal policy is given by (OPT) with $s_1 = s_2 = 0$.

- Case 4: $Z_1(0) < \theta_1$, $Z_2(0) \le \theta_2$, $\psi_1 \ge \psi_2$. In this case, $s_1 = 0$. However, the computation of s_2, u_1 and u_2 depends on the lengths of the high and the low periods as discussed below.

  - Case 4.1: If $a_1 \le H \le a_2$, $L \ge \gamma_1(H - a_1)$, then s_2, u_1, u_2, t_1 and t_2 are computed by solving equations (13)–(19) with $s_1 = 0$.

  - Case 4.2: If

$$H \ge a_1,\ H + L \le \psi_1 + \frac{\eta}{1-\eta}(\psi_1 - \psi_2),\ L \le \gamma_1(H - a_1),$$

    then we set $t_2 = H + L$, and solve the equations (13)-(17) and (19) with $s_1 = 0$ to compute s_2, u_1, u_2 and t_1.

  - Case 4.3: If $H \le a_1$, then the optimal policy is given by (OPT) upon setting $s_1 = 0$, $s_2 = H$, selecting u_2 as any value in the interval $[(\rho_2 - d_2(\mu_2 H)^{-1})^+,\ d_1(\mu_1 H)^{-1} - (\rho_1^h - 1)]$ and setting $u_1 = 1 - u_2$.

  - Case 4.4: If $H \ge a_2$, $H + L > \psi_1 + \eta(1-\eta)^{-1}(\psi_1 - \psi_2)$, then the optimal policy is given by (OPT) with $s_1 = s_2 = 0$.

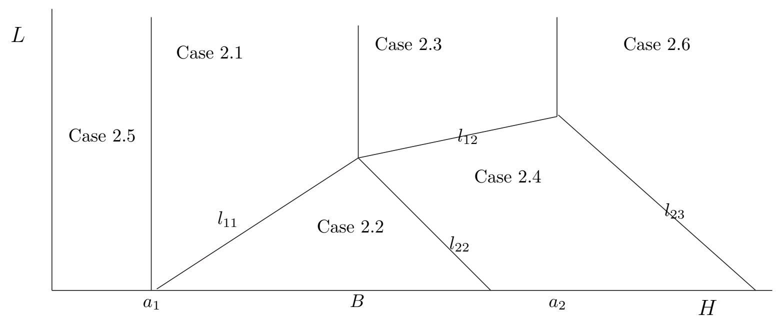

Figure 1: Pictorial representation for the case $Z_1(0) \leq \theta_1, Z_2(0) \geq \theta_2$ and $\tilde{\psi}_1 \geq \tilde{\psi}_2$, where $l_{11} : L = \gamma_1(H - a_1)$, $l_{12} : L = \gamma_2(H - \tilde{\psi}_1)$, $l_{13} : L = \gamma_3(H - \psi_1)$, $l_{22} : H + L = \tilde{\psi}_1 + \frac{1+\eta(\rho_1^h - 1)}{(1-\eta)(\rho_1^h - 1)}(\tilde{\psi}_1 - \tilde{\psi}_2)$, $l_{23} : H + L = \frac{\psi_1}{1-\eta}$.

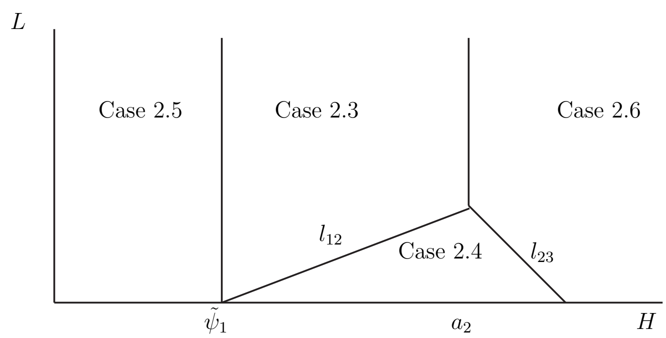

Figure 2: Pictorial representation for the case $Z_1(0) \leq \theta_1, Z_2(0) \geq \theta_2$ and $\tilde{\psi}_1 \leq \tilde{\psi}_2$, where $l_{12} : L = \gamma_2(H - \tilde{\psi}_1)$, $l_{13} : L = \gamma_3(H - \psi_1)$, $l_{23} : H + L = \frac{\psi_1}{1-\eta}$.

As mentioned above, we prove the optimality of this policy in Appendix B. However, in order to give the reader an intuitive explanation, we consider one of the cases above, for example Case 3. We claim that if $Z_1(0) < \theta_1$, $Z_2(0) \leq \theta_2$, $\psi_1 \leq \psi_2$, then the optimal policy is given by (OPT) with $s_1 = s_2 = 0$. In order to see this, first consider the case $H \geq \psi_1$. Under the policy with $s_1 = s_2 = 0$, class 1 fluid level reaches its threshold θ_1 at time ψ_1, and class 2 fluid level reaches its threshold θ_2 at time ψ_2. Note that for any $t \geq \psi_1$, we have

$$\mu_1 h_1(t - \psi_1) \geq \mu_2 h_2(t - \psi_2),$$

since $\psi_2 \geq \psi_1 \geq 0$ and $\mu_1 h_1 > \mu_2 h_2$. Thus, it is more profitable to give fixed priority to class 1 until the class 1 fluid level decreases to its threshold in the low period. If $H < \psi_1$, then again the optimal policy is given by (OPT) upon setting $s_1 = s_2 = 0$ (i.e. giving fixed priority to class 1 in the high period), which yields a total cost of 0.

The following corollary follows from the description of the optimal policy.

Corollary 2. *If*
(i) $Z_1(0) \geq \theta_1$ *or,*
(ii) $Z_1(0) \leq \theta_1$, $Z_1(0) \leq \theta_2$ *and* $0 \leq \psi_1 \leq \psi_2$,

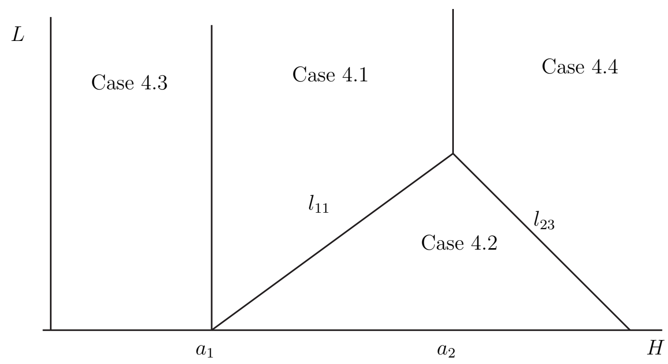

Figure 3: Pictorial representation for the case $Z_1(0) \le \theta_1, Z_2(0) \le \theta_2$ and $\psi_2 \le \psi_1$, where $l_{11} : L = \gamma_1(H - a_1)$, $l_{13} : L = \gamma_3(H - \psi_1)$, $l_{23} : H + L = \psi_1 + \frac{\eta}{1-\eta}(\psi_1 - \psi_2)$.

then the policy with

$$\begin{aligned} \forall t \in (0, H) \qquad & \dot{T}_1(t) = 1,\, \dot{T}_2(t) = 0; \\ \forall t \in (H, H + L) \qquad & \textit{Low-period-policy} \end{aligned}$$

is optimal for all $H \ge 0$ and $L \ge 0$.

Note that if the initial fluid levels satisfy the conditions in (i) or (ii), the policy described in Corollary 2 is optimal even when the length of the high period and the length of the low period are random variables.

3.2 The case $\rho^h > 1,\ \rho_1^h \le 1,\ \rho^l \le 1$.

In this case the optimal policy has the following structure:

$$\begin{aligned} \forall t \in (0,\ s_1): \qquad & \dot{T}_2(t) = 1,\ \dot{T}_1(t) = 0; \\ \forall t \in (s_1,\ s_2): \qquad & \dot{T}_2(t) = \rho_2 - \frac{(\theta_2 - Z_2(s_1))/\mu_2}{a_1(s_1)},\ \dot{T}_1(t) = 1 - \dot{T}_2(t); \\ \forall t \in (s_2,\ s_3): \qquad & \dot{T}_2(t) = 0,\ \dot{T}_1(t) = 1; \\ \forall t \in (s_3,\ H): \qquad & \dot{T}_2(t) = 1 - \rho_1^h,\ \dot{T}_1(t) = \rho_1^h; \\ \forall t \in (H,\ H + L): \qquad & \text{Low-period-policy}; \end{aligned}$$

where

$$a_1(s_1) = \frac{(\theta_1 - Z_1(s_1))/\mu_1 + (\theta_2 - Z_2(s_1))/\mu_2}{\rho_1^h + \rho_2 - 1},$$

and s_1, s_2, s_3 are given as

$$\begin{aligned} s_1 &= \max\{t :\ 0 \le t \le H, Z_2(t) \ge \theta_2,\ Z_1(t) \le \theta_1\}, \\ s_2 &= \max\{t :\ s_1 \le t \le H, Z_1(t) \le \theta_1\}, \\ s_3 &= \max\{t :\ s_2 \le t \le H, Z_1(t) \ge \theta_1\}. \end{aligned}$$

with the convention that $\max\{t : x \le t \le y, t \in A\} = x$ if $A = \emptyset$.

3.3 The case $\rho^h \leq 1,\ \rho^l \leq 1$

In this case the optimal policy has the following structure:

$$\begin{array}{rl} \forall t \in (0, H) & \text{Low-period-policy except replace } \rho_1^l \text{ by } \rho_1^h; \\ \forall t \in (H, H+L) & \text{Low-period-policy.} \end{array}$$

Remark 3. The policies described in Sections 3.2 and 3.3 can be implemented without knowing the length of the high and the low periods. Hence, these policies are also optimal when the length of the high period and the length of the low period are random variables.

4 Discrete review policies in the stochastic case

Throughout the rest of this paper, we shall consider the stochastic instance of the fluid model described in Section 2. Recall that the system starts with a high period, followed by a low period. The duration of the high period H, and the duration of the low period L are independent random variables. For this stochastic fluid control problem, the optimal policy when $\rho^h > 1,\ \rho_1^h \leq 1,\ \rho^l \leq 1$ is given in Section 3.2 and the optimal policy when $\rho^h \leq 1,\ \rho^l \leq 1$ is given in Section 3.3 (see Remark 3). We therefore focus only on the case when

$$\rho_1^h > 1,\ \rho^l \leq 1.$$

To specify the control policy in this case, we shall always consider the following four subcases which were first introduced in Section 3 and are summarized below:

$$\begin{array}{lll} \text{Case 1:} & Z_1(0) \geq \theta_1, & (20) \\ \text{Case 2:} & Z_1(0) < \theta_1,\ Z_2(0) > \theta_2, & (21) \\ \text{Case 3:} & Z_1(0) < \theta_1,\ Z_2(0) \leq \theta_2,\ \psi_1 \leq \psi_2, & (22) \\ \text{Case 4:} & Z_1(0) < \theta_1,\ Z_2(0) \leq \theta_2,\ \psi_1 \geq \psi_2. & (23) \end{array}$$

In this section, we present a discrete review policy that is asymptotically optimal as the expected length of the high period tends to infinity. Under our discrete review policy, the state of the system is observed at intervals of length τ which is a predetermined positive number. Note that no assumptions are imposed on τ. Given τ, the distribution of the high period and the mean of the low period, the discrete review policy is implemented as follows. Let H_0 and L_0 denote the actual values of the high period and the low period respectively. The state of the system is observed at times $t = 0, \tau, 2\tau, \ldots, M\tau$, where

$$M = \min\{n \in \mathbb{N} : n\tau \geq H_0\}.$$

Note that we do not assume that we know H_0 initially. We assume that the system can detect the end of the high period by observing a sudden drop in the arrival rate of class 1 fluid. At each time t, we observe the fluid level of both classes, i.e., $Z_1(t)$ and $Z_2(t)$. We then predict the remaining high period $\tilde{H}(t)$ and the low period $\tilde{L}(t)$ using one of the methods described below. If $t < M\tau$, we implement the policy described in Section 3 from t to $t + \tau$ using $\tilde{H}(t)$ as the length of the high period, $\tilde{L}(t)$ as the length of the low period, and $Z_1(t)$ and $Z_2(t)$ as the initial fluid levels. If $t = M\tau$, we implement the Low-period-policy from t until the end of the low period.

At time t, we either set

$$\tilde{H}(t) \quad = \quad \mathbb{E}[H|H > t] - t, \tag{24}$$

or

$$\tilde{H}(t) = \min\{x \geq 0 : \mathbb{P}(H > x + t | H > t) = p\}, \tag{25}$$

where p will be specified later. Note that in (24) remaining high period is estimated by its expected value, and in (25) remaining high period is set equal to x which guarantees that the probability that the remaining high period is larger than x is p. While implementing the discrete review policy in the numerical examples of Section 6, we use both of these methods to estimate the remaining high period and we set $p = 0.25, 0.5$ and 0.75. On the other hand, the remaining low period is always set equal to its mean. Hence, $\tilde{L}(t) = \mathbb{E}[L]$.

We now show that our discrete review policy is asymptotically optimal as the expected length of the high period tends to infinity. Given the actual values of the high and low periods, let $c(H_0, L_0)$ be the holding cost under the optimal policy described in Section 3. The closed form expression for $c(H_0, L_0)$ is given in Appendix C. Similarly, let $c^{\text{DR}}(H_0, L_0)$ denote the holding cost under our discrete review policy when the length of the high period is H_0 and the length of the low period is L_0.

Proposition 4. *There exist $D > 0$ and $\beta_1 \geq 0$ (which depend on the arrival rates, service rates, initial fluid levels, threshold values and holding costs per unit time) such that if*

$$\tilde{H}(0) \geq D,$$

then the discrete review policy is equivalent to giving fixed priority to class 1 in the high period, and we have

$$c^{\text{DR}}(H_0, L_0) - c(H_0, L_0) \leq \beta_1 \tag{26}$$

for all $H_0 \geq 0$ and $L_0 \geq 0$.

Proof. We provide the proof for the discrete review policy where $\tilde{H}(t)$ is calculated based on the method given in (24). The proof for the discrete review policy implemented with the method given in (25) is similar.

With a slight abuse of notation, we use $d_i(t)$ and $\psi_i(t)$, $i = 1, 2$ to denote the quantities defined in (6) and (7) at time t when fluid levels are $Z_i(t)$, $i = 1, 2$. Similarly, let $a_i(t), i = 1, 2$ denote the corresponding quantities given in (8) at time t. Hence, $d_i(0) = d_i$, $\psi_i(0) = \psi_i$ and $a_i(0) = a_i$ for $i = 1, 2$. Let

$$D = \max\Big\{a_2(0), \;\; \psi_1(0) + \frac{\eta}{1-\eta}(\psi_1(0) - \psi_2(0))\Big\}. \tag{27}$$

We first show by induction that for all $0 \leq n \leq M - 1$, the discrete review policy sets $\dot{T}_1(t) = 1$, $\dot{T}_2(t) = 0$ for all $t \in [n\tau, (n+1)\tau)$. Hence the discrete review policy is equivalent to giving fixed priority to class 1 in the high period $[0, H_0)$.

First consider $t = 0$. Note that for Case 1 and Case 3, it follows immediately from Corollary 2 that the discrete review policy gives fixed priority to class 1, i.e. $\dot{T}_1(t) = 1$, $\dot{T}_2(t) = 0$ for all $t \in [0, \tau)$.

For Case 2, note that $\psi_2 = \psi_2(0) \leq 0$, then $D \geq a_2$ and $D \geq \psi_1 + \eta(1-\eta)^{-1}\psi_1 = (1-\eta)^{-1}\psi_1$. Hence, $\tilde{H}(t) \geq D$ (where D is given in (27)) which implies that the condition of Case 2.6 in Section 3.1 is satisfied, where the discrete review policy gives fixed priority to class 1 in the interval $[0, \tau)$.

For Case 4, $\tilde{H}(t) \geq D$ (where D is given in (27)) which implies that the condition of Case 4.4 in Section (3.1) is satisfied, where the discrete review policy gives fixed priority to class 1 in $[0, \tau)$.

Therefore the claim is true for $n = 0$. Now assume that under the discrete review policy fixed priority is given to class 1 until $t = n\tau$ for $1 \leq n \leq M - 1$. Then the fluid levels of the two classes at time $t = n\tau$ are $Z_1(n\tau) = Z_1(0) + n\tau(\lambda_1^h - \mu_1)$, and $Z_2(n\tau) = Z_2(0) + n\tau\lambda_2$, respectively. It is easily checked from (6),(7) and (8) that

$$\psi_1(n\tau) = \psi_1(0) - n\tau, \quad \psi_2(n\tau) = \psi_2(0) - n\tau, \quad a_2(n\tau) = a_2(0) - n\tau.$$

To specify the discrete review policy at time $t = n\tau$, again we consider Cases 1 to 4 given in (20) to (23) separately. Note that the conditions of these four cases should now be evaluated at time $t = n\tau$ based on $Z_i(n\tau)$ and $\psi_i(n\tau)$, $i = 1, 2$.

Again under Case 1 and Case 3, Corollary 2 applies, hence, the discrete review policy sets $\dot{T}_1(t) = 1$, $\dot{T}_2(t) = 0$ and gives fixed priority to class 1 for all $t \in [n\tau, (n+1)\tau)$.

Under Case 2, since $\tilde{H}(0) = \mathbb{E}[H]$ and

$$\tilde{H}(n\tau) \geq \mathbb{E}[H] - n\tau \geq D - n\tau = \max\{a_2(n\tau), \ \psi_1(n\tau) + \frac{\eta}{1-\eta}(\psi_1(n\tau) - \psi_2(n\tau))\}, \tag{28}$$

it follows from Case 2.6 in Section 3.1 that the discrete review policy gives fixed priority to class 1 in the interval $[n\tau, (n+1)\tau)$.

Similarly, for Case 4, (28) implies that conditions of Case 4.4 in Section (3.1) hold, hence the discrete review policy gives fixed priority to class 1 in the interval $[n\tau, (n+1)\tau)$.

This then completes the induction and we therefore conclude that the discrete review policy sets $\dot{T}_1(t) = 1$, $\dot{T}_2(t) = 0$ for all $0 \leq t \leq H_0$. The result in (26) then follows from Proposition 6 in Section 5. □

Remark 5. The proof for other methods are the same except $\mathbb{E}[H]$ is replaced by $\tilde{H}(0)$ in (28).

5 Other policies that are asymptotically optimal

Throughout this section, we assume that $\rho_1^h > 1$ and $\rho^l \leq 1$. We are interested in two heavy traffic regimes. In the first one, the expected length of the high period tends to infinity. In the second one, traffic intensity of class 2 (i.e. ρ_2) tends to $1 - \rho_1^l$ when ρ_1^l is fixed and the low period is infinitely long. Under both these regimes, we are interested in finding the asymptotically optimal policies.

Consider the policy that gives fixed priority to class 1 in the high period and uses the Low-period-policy in the low period. For the rest of the paper, we will refer to this policy as FP1. We shall use $c^{FP1}(H_0, L_0)$ to denote the holding cost of the FP1 policy when the length of the high period is H_0 and the length of the low period is L_0. Recall that $c(H_0, L_0)$ denotes the holding cost of the optimal control policy (as specified in Section 3) when the lengths of the high and the low periods are known and equal to H_0 and L_0, respectively. We have the following proposition.

Proposition 6. *There exists $\beta_2 \geq 0$, which does not depend on the duration of the high period and low period, such that*

$$c^{\text{FP1}}(H_0, L_0) - c(H_0, L_0) \leq \beta_2.$$

for all $H_0 \geq 0$ and $L_0 \geq 0$.

Proof. We need to consider the holding costs under Cases 1 to 4 separately. Note that for Case 1 and Case 3, Corollary 2 applies and the optimal policy is FP1, hence we can take $\beta_2 = 0$ for these two cases.

Now consider Case 4. Note that the optimal policy (as described in Section 3) is the same as FP1 policy in Case 4.4, and differs from FP1 only under Cases 4.1, 4.2 and 4.3; that is, the two costs differ only when (H_0, L_0) belongs to Δ_1 or Δ_2 where

$$\begin{aligned}\Delta_1 &= \Big\{(h,l) \in \mathbb{R}_+^2 : a_1 \le h,\, h+l \le \psi_1 + \frac{\eta}{1-\eta}(\psi_1 - \psi_2),\, l \le \gamma_1(h-a_1)\Big\} \\ &\quad \bigcup \Big\{(h,l) \in \mathbb{R}_+^2 : a_1 \le h \le a_2,\, \gamma_1(h-a_1) \le l \le \gamma_4(h-a_1)\Big\},\end{aligned}$$

$$\Delta_2 = \Big\{(h,l) \in \mathbb{R}_+^2 : a_1 < h < a_2,\, l > \gamma_4(h-a_1)\Big\} \bigcup \Big\{(h,l) \in \mathbb{R}_+^2 : h < a_1\Big\},$$

and $\gamma_4 = (\rho_1^h + \rho_2 - 1)(1 - \rho_2 - \rho_1^l)^{-1}$.

Expressions for $c^{\text{FP1}}(H_0, L_0)$ and $c(H_0, L_0)$ are given in Appendix C. Suppose $(H_0, L_0) \in \Delta_1$. Note that holding cost function is quadratic in H_0 and L_0 under both the optimal policy and the FP1 policy (See Appendix C). Moreover, since Δ_1 is a bounded region, the difference between the holding cost of these two policies must be bounded.

Now suppose that $(H_0, L_0) \in \Delta_2$. Since $\tilde{\psi}_2 < 0$ in Case 4, it follows from the holding cost expressions in (OPT:12) and (OPT:15) in Appendix C that $c(H_0, L_0)$ depends only on H_0. Then since H_0 is bounded, $c(H_0, L_0)$ is bounded. Consider the cost function $c^{\text{FP1}}(H_0, L_0)$ when $(H_0, L_0) \in \Delta_2$. Note that since $\psi_2 \le a_1$ in Case 4, and H, L are always nonnegative, Δ_2 can be written as $\Delta_{21} \cup \Delta_{22}$ where

$$\begin{aligned}\Delta_{21} &= \big\{(h,l) \in \mathbb{R}_+^2 :\, \psi_2 \le h \le a_2,\, l \ge \gamma_4(h-a_1)\big\} \text{ and} \\ \Delta_{22} &= \big\{(h,l) \in \mathbb{R}_+^2 :\, h \le \psi_2\big\}.\end{aligned}$$

If $(H_0, L_0) \in \Delta_{22}$, $c^{\text{FP1}}(H_0, L_0) = 0$ (see the holding cost expression in (FP1:9) in Appendix C). When $\psi_2 \le a_1 \le \psi_1$, Δ_{21} is a subset of

$$\begin{aligned}&\big\{(h,l) \in \mathbb{R}_+^2 :\, \psi_1 \le h,\, l \ge \gamma_4(h-a_1)\big\} \\ \bigcup\ &\Big\{(h,l) \in \mathbb{R}_+^2 :\, \hat{\psi} \le h \le \psi_1,\, l \ge \gamma_4(h-a_1)\Big\} \\ \bigcup\ &\Big\{(h,l) \in \mathbb{R}_+^2 :\, \psi_2 \le h \le \hat{\psi},\, l \ge \gamma_4(h-a_1)\Big\},\end{aligned}$$

which correspond to the conditions given in (FP1:1), (FP1:4) and (FP1:7) (note that $\gamma_5 \ge \gamma_4$) respectively. In these three cases, the cost function $c^{\text{FP1}}(H_0, L_0)$ is a quadratic function of H_0 only (not a function of L_0). But since $H_0 \le a_2$, $c^{\text{FP1}}(H_0, L_0)$ and $c(H_0, L_0)$ are bounded when $(H_0, L_0) \in \Delta_{21}$.

Therefore, if $(H_0, L_0) \in \Delta_1$ or $(H_0, L_0) \in \Delta_2$, the difference between $c^{\text{FP1}}(H_0, L_0)$ and $c(H_0, L_0)$ is bounded and the result of Proposition 6 holds for Case 4.

The proof for Case 2 is similar and thus omitted. □

We next consider the case that the traffic intensify of class 2 tends to $1 - \rho_1^l$ (i.e. the system is always heavily loaded) and the expected length of the low period tends to infinity. Again we consider Cases 1 to 4 given in (20) to (23), separately. We know from Corollary 2 that in Cases 1 and 3, FP1 policy is optimal. Hence, we only consider Cases 2 and 4. We start with Case 4.

Definition 7. Assume conditions of Case 4. We define the π^{a_1} policy as follows:

$$\begin{aligned}
&\forall t \in (0, a_1 \wedge H), && \dot{T}_2(t) = \rho_2 - \frac{\theta_2 - Z_2(0)}{a_1 \mu_1}, \quad \dot{T}_1(t) = 1 - \dot{T}_2(t); \\
&\forall t \in (a_1 \wedge H, H), && \dot{T}_2(t) = 0, \quad \dot{T}_1(t) = 1; \\
&\forall t \in (H, H+L), && \text{Low-period-policy.}
\end{aligned}$$

Under Case 4, since initially both class 1 and class 2 fluid levels are below their threshold values, π^{a_1} policy starts with processor sharing. In the processor sharing serving scheme, $\dot{T}_1(t)$ and $\dot{T}_2(t)$ are chosen such that the time that class 2 fluid level reaches its threshold is delayed while ensuring that the cost accumulated from class 1 in the high period is not too high. Moreover, this choice of $\dot{T}_1(t)$ and $\dot{T}_2(t)$ guarantees that class 1 and class 2 reach their thresholds from below at the same time if H is long enough to do so. Thus, during the processor sharing period, the π^{a_1} policy gives as much proportion of service as possible to class 2 while maintaining class 1 below its threshold. Note that if the traffic intensity in the low period is close to 1 and the low period is long, the holding cost for class 2 fluid in the low period can be high. Hence, it is important to reduce the amount of class 2 fluid at the beginning of the low period without incurring too much cost from class 1 fluid. We will show in Proposition 10 that when $\rho_2 \to 1 - \rho_1^l$ and $\mathbb{E}[L] \to \infty$, π^{a_1} is strongly asymptotically optimal under the assumptions of Case 4. We use a notion of strongly asymptotically optimal (as introduced in [15]) in the following sense:

Definition 8. Consider a control problem where the performance measure $J(u, \alpha)$ is a function of the control policy u and parameter α. Let the optimal control policy be $u^*(\alpha)$, and suppose $J(u^*(\alpha), \alpha) \to \infty$ as $\alpha \to \alpha_0$. A control policy $\hat{u}$ is called *strongly asymptotically optimal* if there exists $K < \infty$ such that

$$J(\hat{u}(\alpha), \alpha) - J(u^*(\alpha), \alpha) \leq K, \text{ as } \alpha \to \alpha_0.$$

We will also use the following notation.

Definition 9. For $f : \mathbb{R} \to \mathbb{R}$, we write

$$f(r) \quad = \quad \mathcal{O}(1) \text{ as } r \to r_0$$

to mean that there exists a constant $M > 0$ such that $|f(r)| < M$ as $r \to r_0$.

Let $c^{a_1}(H, L)$ denote the holding cost under policy π^{a_1} when the length of the high period is H and the length of the low period is L. The closed form expression for $c^{a_1}(H, L)$ is given in Appendix C.

Proposition 10. *Assume conditions of Case 4. Suppose H and L are random variables with $\mathbb{E}[H^2] < \infty$. If $\mathbb{E}[L] \to \infty$ and $\rho_2 \to (1 - \rho_1^l)$ (where ρ_1^l is fixed), then*

$$\mathbb{E}[c^{a_1}(H, L) - c(H, L)] \quad = \quad \mathcal{O}(1),$$

and π^{a_1} is strongly asymptotically optimal.

Proof. We start with characterizing the difference between $c^{a_1}(H, L)$ and $c(H, L)$. We have the following cases based on the regions that (H, L) belongs to.

- If $H \geq a_2$ and $L \geq \gamma_3(H-a_1)$ (note that $a_2 \geq a_1$), then (H,L) either satisfies the condition of $(a_1:1)$ or $(a_1:2)$ (for holding cost under π^{a_1} policy given in Appendix C). If (H,L) satisfies the condition of $(a_1:1)$, then (H,L) satisfies the condition of (OPT:1)(for holding cost under the optimal policy given in Appendix C). If (H,L) satisfies the condition of $(a_1:2)$, then (H,L) satisfies the condition of (OPT:2). When we compare the cost difference, we either compare $c^{a_1}(H,L)$ given by $(a_1:1)$ and $c(H,L)$ given by (OPT:1) or we compare $c^{a_1}(H,L)$ given by $(a_1:2)$ and $c(H,L)$ given by (OPT:2). These two cases bear the same difference, which is

$$\begin{aligned} & c^{a_1}(H,L) - c(H,L) \\ = \;& \frac{1}{2}\mu_2 h_2 \Bigg\{ \left[\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}\right][(H-a_1)^2-(H-\psi_1)^2] \\ & \quad + \frac{\rho_2(\rho_1^h-\rho_1^l)^2}{(1-\rho_1^l)(1-\rho_1^l-\rho_2)}(H-a_1)^2 - \frac{\rho_2(\rho_1^h-1)^2}{(1-\rho_1^l)(1-\rho_1^l-\rho_2)}(H-\psi_1)^2 \\ & \quad - \frac{2\rho_2(\rho_1^h-1)}{1-\rho_2-\rho_1^l}(H-\psi_2)(H-\psi_1) - \frac{\rho_2(1-\rho_1^l)}{1-\rho_2-\rho_1^l}(H-\psi_2)^2 \Bigg\}. \end{aligned}$$

The items that possibly go to infinity are those that have $(1-\rho_2-\rho_1^l)$ in the denominator. We next show that $(1-\rho_2-\rho_1^l)$ is cancelled off from the denominator after combining all these items. First, consider the last three terms. Factoring out $\rho_2[(1-\rho_1^l)(1-\rho_1^l-\rho_2)]^{-1}$ and applying the formula $a^2+2ab+b^2=(a+b)^2$, we can combine them into

$$-\frac{\rho_2}{(1-\rho_1^l)(1-\rho_1^l-\rho_2)}[(\rho_1^h-1)(H-\psi_1)+(1-\rho_1^l)(H-\psi_2)]^2.$$

Adding this value to the second term (i.e. $[\rho_2(\rho_1^h-\rho_1^l)^2][(1-\rho_1^l)(1-\rho_1^l-\rho_2)]^{-1}(H-a_1)^2$), taking the common factor $\rho_2[(1-\rho_1^l)(1-\rho_1^l-\rho_2)]^{-1}$ out and applying the formula $a^2-b^2=(a+b)(a-b)$, we can combine all the terms with $(1-\rho_2-\rho_1^l)$ in the denominator into

$$\begin{aligned} \frac{\rho_2}{(1-\rho_1^l)(1-\rho_1^l-\rho_2)} \Big\{ & [(\rho_1^h-\rho_1^l)(H-a_1)+(\rho_1^h-1)(H-\psi_1)+(1-\rho_1^l)(H-\psi_2)] \\ & \times[(\rho_1^h-\rho_1^l)(H-a_1)-(\rho_1^h-1)(H-\psi_1)-(1-\rho_1^l)(H-\psi_2)] \Big\}. \end{aligned}$$

From the definitions of a_1, ψ_1, and ψ_2, we know that $a_1 = ((\rho_1^h-1)\psi_1+\rho_2\psi_2)(\rho_1^h+\rho_2-1)^{-1}$. Plugging in this expression of a_1, we can further simplify the second line of the above expression into

$$-\frac{(\rho_1^h-1)(1-\rho_2-\rho_1^l)}{(\rho_1^h+\rho_2+1)}(\psi_1-\psi_2).$$

Thus, $(1-\rho_2-\rho_1^l)$ is cancelled off from the denominator and we have

$$\begin{aligned} & c^{a_1}(H,L) - c(H,L) \\ = \;& \frac{1}{2}\mu_2 h_2 \Bigg\{ \frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(2H-a_1-\psi_1)(\psi_1-a_1) \\ & \quad - \frac{\rho_2(\rho_1^h-\rho_1^l)(\rho_1^h-1)}{(1-\rho_1^l)(\rho_1^h+\rho_2-1)}(2H-a_1-\frac{(\rho_1^h-1)\psi_1+(1-\rho_1^l)\psi_2}{\rho_1^h-\rho_1^l})(\psi_1-\psi_2) \Bigg\}. \\ \leq \;& \frac{1}{2}h_2\mu_2 \left\{ \frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(2H-a_1-\psi_1)(\psi_1-a_1) \right\}, \end{aligned}$$

where the inequality follows from the fact that the second term is not positive since $0 \le \psi_2 \le a_1 \le \psi_1 \le a_2$ and $H \ge a_2$. At the same time, since $0 \le \psi_1 - a_1 \le H$ and $0 \le (2H - a_1 - \psi_1) \le 2H$, we obtain

$$\begin{aligned} c^{a_1}(H,L) - c(H,L) &\le \frac{1}{2}h_2\mu_2\left(\frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{\eta(1-\rho_1^l)}\right)2H^2 \\ &= \frac{1}{2}h_1\mu_1\left(\frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{(1-\rho_1^l)}\right)2H^2 \end{aligned} \tag{29}$$

where the equality follows from the definition of η.

- If $a_1 \le H \le a_2$ and $L \ge \gamma_3(H - a_1)$, then (H,L) either satisfies the condition given by $(a_1 : 1)$ or $(a_1 : 2)$ (for holding cost under π^{a_1} policy) since $\gamma_3 \ge \gamma_1$. If (H,L) satisfies $(a_1 : 1)$, then it satisfies (OPT:12) (for holding cost under the optimal policy in case 4). If (H,L) satisfies $(a_1 : 2)$, then it satisfies (OPT:13) (for holding cost under the optimal policy in case 4). When we compare the cost difference, we either compare $c^{a_1}(H,L)$ given by $(a_1 : 1)$ and $c(H,L)$ given by (OPT:12) or we compare $c^{a_1}(H,L)$ given by $(a_1 : 2)$ and $c(H,L)$ given by (OPT:13). Note that $c^{a_1}(H,L)$ under $(a_1 : 2)$ is different from that under $(a_1 : 1)$ by the same amount as $c(H,L)$ under (OPT:13) from (OPT:12). Hence, these two cases bear the same cost difference. We are going to compare the holding cost under the conditions of $(a_1 : 1)$ with that one under the conditions of (OPT:12).

  First, we simplify the holding cost expression for (OPT:12). Note that conditions of Case 4 imply that $\tilde{\psi}_2 = 0$. Hence

$$\begin{aligned} c(H,L) &= \frac{1}{2}h_2\mu_2\left\{\frac{(1-\rho_1^l)(\rho_1^h - \rho_1^l)}{\eta(\rho_1^h - 1)}(t_2 - H)^2 + \rho_2(t_2 - s_2)^2 \right. \\ &\quad \left. + (1 - \rho_2 - \rho_1^l)\left[\frac{\rho_1^h + \rho_2 - 1}{1 - \rho_2 - \rho_1^l}(H - a_1) - (t_2 - H)\right]^2\right\} \end{aligned}$$

  where t_2 and s_2 are given in Appendix C. Using these expressions of t_2 and s_2, we obtain

$$\begin{aligned} t_2 - H &= \frac{\eta(\rho_1^h - 1)(\rho_1^h + \rho_2 - 1)}{(1-\rho_1^l)(\rho_2 + \eta(\rho_1^h - 1)) + (1-\eta)\rho_2(\rho_1^h - 1)}(H - a_1), \\ t_2 - a_1 &= t_2 - H + H - a_1 \\ &= \frac{(\rho_1^h - \rho_1^l)[\rho_2 + \eta(\rho_1^h - 1)]}{(1-\rho_1^l)(\rho_2 + \eta(\rho_1^h - 1)) + (1-\eta)\rho_2(\rho_1^h - 1)}(H - a_1), \\ t_2 - s_2 &= \frac{(\rho_1^h + \rho_2 - 1)}{\rho_2 + \eta(\rho_1^h - 1)}(t_2 - a_1) \\ &= \frac{(\rho_1^h - \rho_1^l)(\rho_1^h + \rho_2 - 1)}{(1-\rho_1^l)(\rho_2 + \eta(\rho_1^h - 1)) + (1-\eta)\rho_2(\rho_1^h - 1)}(H - a_1). \end{aligned}$$

  For notational convenience, let γ_{1d} denote the denominator in the above expressions. We have

$$\begin{aligned} c(H,L) &= \frac{1}{2}h_2\mu_2\left\{\frac{(1-\rho_1^l)(\rho_1^h - \rho_1^l)\eta(\rho_1^h - 1)(\rho_1^h + \rho_2 - 1)^2}{\gamma_{1d}^2}(H - a_1)^2 \right. \\ &\quad + \frac{\rho_2(\rho_1^h - \rho_1^l)^2(\rho_1^h + \rho_2 - 1)^2}{\gamma_{1d}^2}(H - a_1)^2 \end{aligned}$$

$$
+\frac{\rho_2^2(\rho_1^h-\rho_1^l)^2(\rho_1^h+\rho_1^l-1)^2}{(1-\rho_1^l-\rho_2)\gamma_{1d}^2}(H-a_1)^2\Big\}
$$

$$
= \frac{1}{2}h_2\mu_2\left\{\frac{(1-\rho_1^l)(\rho_1^h-\rho_1^l)(\rho_1^h+\rho_2-1)^2}{(1-\rho_1^l-\rho_2)\gamma_{1d}}(H-a_1)^2\right\}.
$$

Then the difference between $c^{a_1}(H,L)$ and $c(H,L)$ is given as

$$
\begin{aligned}
& c^{a_1}(H,L)-c(H,L)\\
= & \frac{1}{2}\mu_2h_2\Big\{\Big(\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}+\frac{\rho_2(\rho_1^h-\rho_1^l)^2}{(1-\rho_1^l)(1-\rho_1^l-\rho_2)}\Big)(H-a_1)^2\\
& -\Big(\frac{(1-\rho_1^l)(\rho_1^h-\rho_1^l)(\rho_1^h+\rho_2-1)^2}{(1-\rho_1^l-\rho_2)\gamma_{1d}}\Big)(H-a_1)^2\Big\}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
& \frac{\rho_2(\rho_1^h-\rho_1^l)^2}{(1-\rho_1^l)(1-\rho_1^l-\rho_2)}-\frac{(1-\rho_1^l)(\rho_1^h-\rho_1^l)(\rho_1^h+\rho_2-1)^2}{(1-\rho_1^l-\rho_2)\gamma_{1d}}\\
= & \frac{\rho_1^h-\rho_1^l}{(1-\rho_1^l)(1-\rho_1^l-\rho_2)\gamma_{1d}}\Big(\rho_2(\rho_1^h-\rho_1^l)\gamma_{1d}-(1-\rho_1^l)^2(\rho_1^h+\rho_2-1)^2\Big).
\end{aligned}
$$

Plugging in the expression of γ_{1d}, we simplify the expression in the parentheses above as follows

$$
\begin{aligned}
& \rho_2(\rho_1^h-\rho_1^l)\gamma_{1d}-(1-\rho_1^l)^2(\rho_1^h+\rho_2-1)^2\\
= & \rho_2(\rho_1^h-\rho_1^l)[(1-\rho_1^l)(\rho_2+\eta(\rho_1^h-1))+(1-\eta)\rho_2(\rho_1^h-1)]-(1-\rho_1^l)^2(\rho_1^h+\rho_2-1)^2\\
= & \rho_2^2(\rho_1^h-\rho_1^l)(1-\rho_1^l)+\rho_2(\rho_1^h-\rho_1^l)(1-\rho_1^l)\eta(\rho_1^h-1)\\
& +\rho_2^2(\rho_1^h-\rho_1^l)(\rho_1^h-1)-\rho_2^2(\rho_1^h-\rho_1^l)\eta(\rho_1^h-1)\\
& -(1-\rho_1^l)^2(\rho_1^h+\rho_2-1)^2\\
= & \Big(\rho_2^2(\rho_1^h-\rho_1^l)^2-(1-\rho_1^l)^2(\rho_1^h+\rho_2-1)^2\Big)\\
& +\Big(\rho_2(\rho_1^h-\rho_1^l)(1-\rho_1^l)\eta(\rho_1^h-1)-\rho_2^2(\rho_1^h-\rho_1^l)\eta(\rho_1^h-1)\Big)\\
= & \Big(-[\rho_2(\rho_1^h-\rho_1^l)+(1-\rho_1^l)(\rho_1^h+\rho_2-1)](1-\rho_1^l-\rho_2)(\rho_1^h-1)\Big)\\
& +\Big(\rho_2(\rho_1^h-\rho_1^l)(\rho_1^h-1)\eta(1-\rho_1^l-\rho_2)\Big)\\
= & -(1-\rho_2-\rho_1^l)(\rho_1^h-1)[\rho_2(\rho_1^h-\rho_1^l)(1-\eta)+(1-\rho_1^l)(\rho_1^h+\rho_2-1)].
\end{aligned}
$$

Again plugging in the expression of γ_{1d}, we finally have

$$
\begin{aligned}
& c^{a_1}(H,L)-c(H,L)\\
= & \frac{1}{2}\mu_2h_2\Big\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(H-a_1)^2\\
& -\frac{(\rho_1^h-\rho_1^l)(\rho_1^h-1)[\rho_2(\rho_1^h-\rho_1^l)(1-\eta)+(1-\rho_1^l)(\rho_1^h+\rho_2-1)]}{(1-\rho_1^l)[(1-\rho_1^l)(\rho_2+\eta(\rho_1^h-1))+(1-\eta)\rho_2(\rho_1^h-1)]}(H-a_1)^2\Big\}\\
\leq & \frac{1}{2}\mu_2h_2\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(H-a_1)^2.
\end{aligned}
$$

Since in this case $0 \le H - a_1 \le H$,

$$\begin{aligned} c^{a_1}(H,L) - c(H,L) &\le \frac{1}{2}\mu_2 h_2 \frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{\eta(1-\rho_1^l)} H^2 \\ &= \frac{1}{2}\mu_1 h_1 \frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{(1-\rho_1^l)} H^2 \end{aligned} \tag{30}$$

where the last equality follows from the definition of η.

- If $H \ge a_1$, $L \le \gamma_3(H - a_1)$, then (H, L) satisfies the condition given by $(a_1 : 3)$. Plugging in the corresponding cost expressions, we have

$$\begin{aligned} & c^{a_1}(H,L) - c(H,L) \le c^{a_1}(H,L) \\ = \; & \frac{1}{2}h_2\mu_2\Big\{\frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{\eta(1-\rho_1^l)}(H-a_1)^2 + \rho_2(H+L-a_1)^2 \\ & -\frac{(1-\rho_1^l)}{\eta}\left[\frac{\rho_1^h - 1}{1-\rho_1^l}(H-a_1) - L\right]^2\Big\} \\ \le \; & \frac{1}{2}h_2\mu_2\Big\{\frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{\eta(1-\rho_1^l)}(H-a_1)^2 + \rho_2(H+L-a_1)^2\Big\}. \end{aligned}$$

  Since $H \ge a_1$ and $0 \le L \le \gamma_2(H - a_1)$, $0 \le (H + L - a_1) \le (1 + \gamma_3)(H - a_1)$. Then

$$c^{a_1}(H,L) - c(H,L) \le \frac{1}{2}h_2\mu_2\Big(\frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{\eta(1-\rho_1^l)} + (1+\gamma_3)\Big)(H-a_1)^2.$$

  Moreover, since $0 \le H - a_1 \le H$, and $1 + \gamma_3 = (\rho_1^h - \rho_1^l)(1 - \rho_1^l)^{-1}$,

$$\begin{aligned} c^{a_1}(H,L) - c(H,L) &\le \frac{1}{2}h_2\mu_2\Big(\frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{\eta(1-\rho_1^l)} + \frac{\rho_1^h - \rho_1^l}{1-\rho_1^l}\Big)H^2 \\ &\le \frac{1}{2}h_1\mu_1\Big(\frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{(1-\rho_1^l)} + \frac{\rho_1^h - \rho_1^l}{1-\rho_1^l}\Big)H^2 \end{aligned} \tag{31}$$

  where the second inequality follows from the definition of η and our assumption that $h_2\mu_2 < h_1\mu_1$.

- If $H \le a_1$, we have from the cost expressions under (OPT:15) (note that $\tilde{\psi}_2 < 0$ in Case 4) and $(a_1 : 4)$,

$$c^{a_1}(H,L) = c(H,L) = 0. \tag{32}$$

Putting these cases together, from (29) to (32) and our assumption that $E[H^2] \le \infty$, we have the desired result.

□

We next consider Case 2 given in (21) and we define the following policy.

Definition 11. Assume conditions of Case 2. We define the FP2-FP1 policy as follows:

$$\begin{array}{rcl} \forall t \in (0,H), & \text{if } Z_2(t) > \theta_2,\ Z_1(t) < \theta_1 & \text{then } \dot{T}_2(t) = 1,\ \dot{T}_1(t) = 0; \\ \forall t \in (0,H) & \text{if } Z_2(t) = \theta_2,\ Z_1(t) < \theta_1 & \text{then } \dot{T}_2(t) = \rho_2,\ \dot{T}_1(t) = 1-\rho_2; \\ \forall t \in (0,H) & \text{if } Z_1(t) \ge \theta_1 & \text{then } \dot{T}_2(t) = 0,\ \dot{T}_1(t) = 1; \\ \forall t \in (H,H+L) & & \text{Low-period-policy.} \end{array}$$

Note that FP2-FP1 policy is similar to the π^{a_1} policy. However, since initially class 2 fluid is above its threshold level, FP2-FP1 policy starts with giving fixed priority to class 2. Let $c^{\mathrm{FP2-FP1}}(H, L)$ denote the holding cost under the FP2-FP1 policy when the length of the high period is H and the length of the low period is L. The closed form expression for $c^{\mathrm{FP2-FP1}}(H, L)$ is given in Appendix. C.

Proposition 12. *Assume conditions of Case 2. Suppose H and L are random variables with $\mathbb{E}[H^2] < \infty$. If $\mathbb{E}[L] \to \infty$ and $\rho_2 \to 1 - \rho_1^l$ (where ρ_1^l is fixed), then*

$$\mathbb{E}[c^{\mathrm{FP2-FP1}}(H, L) - c(H, L)] \quad = \quad \mathcal{O}(1),$$

and FP2-FP1 policy is strongly asymptotically optimal.

Proof. First suppose $\tilde{\psi}_1 \le \tilde{\psi}_2$, then $B \le a_1 \le \tilde{\psi}_1 \le \tilde{\psi}_2 \le a_2$. Define

$$\begin{aligned}
\Delta_1 &= \Big\{(h, l) \in \mathbb{R}_+^2 : h \ge a_2,\, l \ge \gamma_3(h - \tilde{\psi}_1)\Big\}, \\
\Delta_2 &= \Big\{(h, l) \in \mathbb{R}_+^2 : \tilde{\psi}_1 \le h \le a_2,\, l \ge \gamma_3(h - \tilde{\psi}_1)\Big\}, \\
\Delta_3 &= \Big\{(h, l) \in \mathbb{R}_+^2 : h \le \tilde{\psi}_1\Big\}, \\
\Delta_4 &= \Big\{(h, l) \in \mathbb{R}_+^2 : \tilde{\psi}_1 \le h,\, l \le \gamma_3(h - \tilde{\psi}_1)\Big\}.
\end{aligned}$$

Clearly, $\Delta_1 \cup \Delta_2 \cup \Delta_3 \cup \Delta_4 = \mathbb{R}_+^2$.

- If $(H, L) \in \Delta_1$, then $a_2 \ge \tilde{\psi}_1$, and $(H,\, L)$ belongs to either the region considered in (FP2-FP1:1) or the region in (FP2-FP1:2) for FP2-FP1 policy (see Appendix C). Comparing the conditions of (FP2-FP1:1) with those of (OPT:1) and conditions of (FP2:FP1:2) with those of (OPT:2), one can conclude that if (H, L) belongs to the region given by (FP2-FP1:1), then (H, L) belongs to the region given by (OPT:1); if (H, L) belongs to the region given by (FP2-FP1:2), then (H, L) belongs to the region given by (OPT:2) since $\tilde{\psi}_1 \le \psi_1$. To compare the cost difference under the optimal policy and the FP2-FP1 policy, we either compare the cost under FP2-FP1 policy given by (FP2-FP1:1) with the optimal cost given by (OPT:1), or the cost under FP2-FP1 policy given by (FP2-FP1:2) with the optimal cost given by (OPT:2). Note that the cost under FP2-FP1 policy in (FP2-FP1:2) is different from that in (FP2:FP1:1) by the same amount as optimal cost in (OPT:2) is different from that in (OPT:1). These two cases yield the same cost difference between FP2-FP1 policy and the optimal policy. The cost difference is equal to $c^{\mathrm{FP2-FP1}}(H, L) - c(H, L)$ where $c^{\mathrm{FP2-FP1}}(H, L)$ is given in (FP2-FP1:1) and $c^{\mathrm{FP2-FP1}}(H, L)$ is given in (OPT:1), respectively. Note that in this case, $\psi_1^- = 0$ and $\psi_2^- = -\psi_2$ (where $a^- = \max\{-a, 0\}$ for a real number a). We have

$$\begin{aligned}
c(H, L) \quad = \quad & \frac{1}{2} h_2 \mu_2 \Big\{ \underbrace{\frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{\eta(1 - \rho_1^l)}(H - \psi_1)^2}_{f_3} - \rho_2 \psi_2^2 \\
& + \underbrace{\frac{(1 - \rho_2 - \rho_1^l)^2}{\rho_2}\Big(\frac{\rho_1^h + \rho_2 - 1}{1 - \rho_1^l - \rho_2}(H - a_1) - \frac{\rho_1^h - 1}{1 - \rho_1^l}(H - \psi_1)\Big)^2}_{f_2} \\
& \underbrace{(1 - \rho_1^l - \rho_2)\Big(\frac{\rho_1^h + \rho_2 - 1}{1 - \rho_1^l - \rho_2}(H - a_1) - \frac{(\rho_1^h - 1)}{(1 - \rho_1^l)}(H - \psi_1)\Big)^2}_{f_1} \Big\}
\end{aligned}$$

and

$$c^{\text{FP2}-\text{FP1}}(H,L) = \frac{1}{2}h_2\mu_2\Big\{\underbrace{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(H-\tilde{\psi}_1)^2}_{g_3}+(1-\rho_2)(2\tilde{\psi}_2-\tilde{\psi}_1)\tilde{\psi}_1$$

$$+2\frac{(1-\rho_2)(\rho_1^h-\rho_1^l)}{1-\rho_1^l}(\tilde{\psi}_2-\tilde{\psi}_1)(H-\tilde{\psi}_1)+\underbrace{\rho_2\Big(\frac{\rho_1^h-\rho_1^l}{1-\rho_1^l}(H-\tilde{\psi}_1)\Big)^2}_{g_2}$$

$$+\underbrace{(1-\rho_2-\rho_1^l)\Big(\frac{\rho_1^h+\rho_2-1}{1-\rho_2-\rho_1^l}(H-a_1)-\frac{\rho_1^h-1}{1-\rho_1^l}(H-\tilde{\psi}_1)\Big)^2}_{g_1}\Big\}.$$

With some algebra, we have

$$g_1-f_1 = -\Big(\frac{(\rho_1^h-1)}{(1-\rho_1^l)}(\psi_1-\tilde{\psi}_1)\Big)$$
$$\Big(2(\rho_1^h+\rho_2-1)(H-a-1)-\frac{(1-\rho_2-\rho_1^l)(\rho_1^h-1)}{(1-\rho_1^l)}(2H-\psi_1-\tilde{\psi}_1)\Big).$$

From the definitions of a_1, ψ_1 and ψ_2, we have $(\rho_1^h+\rho_2-1)(H-a_1)=(\rho_1^h-1)(H-\psi_1)+\rho_2(H-\psi_2)$. Plugging in this expression, we obtain

$$g_1-f_1 = \frac{(1-\rho_1^l-\rho_2)(\rho_1^h-1)^2}{(1-\rho_1^l)^2}(2H-\psi_1-\tilde{\psi}_1)(\psi_1-\tilde{\psi}_1)$$
$$-\frac{2(\rho_1^h-1)^2}{(1-\rho_1^l)}(H-\psi_1)(H-\tilde{\psi}_1)-\frac{2\rho_2(\rho_1^h-1)}{(1-\rho_1^l)}(H-\psi_2)(\psi_1-\tilde{\psi}_1).$$

Similarly,

$$g_2-f_2$$
$$=\frac{1}{\rho_2}\Big(\frac{\rho_2(\rho_1^h-1)}{1-\rho_1^l}(H-\tilde{\psi}_1)+(\rho_1^h+\rho_2-1)(H-a_1)-\frac{(1-\rho_2-\rho_1^l)(\rho_1^h-1)}{1-\rho_1^l}(H-\psi_1)\Big)$$
$$\times\Big(\frac{\rho_2(\rho_1^h-1)}{1-\rho_1^l}(H-\tilde{\psi}_1)-(\rho_1^h+\rho_2-1)(H-a_1)+\frac{(1-\rho_2-\rho_1^l)(\rho_1^h-1)}{1-\rho_1^l}(H-\psi_1)\Big).$$

Since $(\rho_1^h+\rho_2-1)(H-a_1)=(\rho_1^h-1)(H-\psi_1)+\rho_2(H-\psi_2)$,

$$g_2-f_2 = \frac{\rho_2(\rho_1^h-1)^2}{(1-\rho_1^l)^2}(2H-\psi_1-\tilde{\psi}_1)(\psi_1-\tilde{\psi}_1)$$
$$-\frac{2\rho_2(\rho_1^h-1)}{(1-\rho_1^l)}(H-\psi_1)(H-\psi_2)-\rho_2(H-\psi_2)^2.$$

Finally,

$$g_3-f_3 = \frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(2H-\psi_1-\tilde{\psi}_1)(\psi_1-\tilde{\psi}_1).$$

Then

$$\begin{aligned}
& (g_1-f_1)+(g_2-f_2)+(g_3-f_3) \\
= & \Big(\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}+\frac{(\rho_1^h-1)^2}{(1-\rho_1^l)}\Big)(2H-\psi_1-\tilde{\psi}_1)(\psi_1-\tilde{\psi}_1)-\rho_2(H-\psi_2)^2 \\
- & \frac{2\rho_2(\rho_1^h-1)}{(1-\rho_1^l)}(H-\psi_2)(H-\tilde{\psi}_1)-\frac{2(\rho_1^h-1)^2}{(1-\rho_1^l)}(H-\psi_1)(\psi_1-\tilde{\psi}_1).
\end{aligned}$$

Note that $\psi_1 \geq \tilde{\psi}_1$. We also have $a_2 \geq \psi_1 \geq \tilde{\psi}_1$ and $H \geq a_2$ which imply that $H-\tilde{\psi}_1 \geq 0$ and $(H-\psi_2)(H-\tilde{\psi}_1) \geq -\psi_2(H-\tilde{\psi}_1)$. Then

$$\begin{aligned}
& (g_1-f_1)+(g_2-f_2)+(g_3-f_3) \\
\leq & \Big(\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}+\frac{(\rho_1^h-1)^2}{(1-\rho_1^l)}\Big)(2H-\psi_1-\tilde{\psi}_1)(\psi_1-\tilde{\psi}_1) \\
& -\rho_2(H-\psi_2)^2+\frac{2\rho_2(\rho_1^h-1)}{(1-\rho_1^l)}\psi_2(H-\tilde{\psi}_1).
\end{aligned}$$

Note that

$$\begin{aligned}
c^{\mathrm{FP2-FP1}}(H,L)-c(H,L) &= \frac{1}{2}h_2\mu_2\Big\{(g_1-f_1)+(g_2-f_2)+(g_3-f_3)+(1-\rho_2)(2\tilde{\psi}_2-\tilde{\psi}_1)\tilde{\psi}_1 \\
&\quad +\frac{2(1-\rho_2)(\rho_1^h-\rho_1^l)}{(1-\rho_1^l)}(\tilde{\psi}_2-\tilde{\psi}_1)(H-\tilde{\psi}_1)+\rho_2\psi_2^2 \\
&\leq \frac{1}{2}h_2\mu_2\Big\{(g_1-f_1)+(g_2-f_2)+(g_3-f_3)+2(1-\rho_2)\tilde{\psi}_2\tilde{\psi}_1 \\
&\quad +\frac{2(1-\rho_2)(\rho_1^h-\rho_1^l)}{(1-\rho_1^l)}\tilde{\psi}_2(H-\tilde{\psi}_1)+\rho_2\psi_2^2\Big\}.
\end{aligned}$$

Plugging in the upper bound for $(g_1-f_1)+(g_2-f_2)+(g_3-f_3)$, we obtain

$$\begin{aligned}
& c^{\mathrm{FP2-FP1}}(H,L)-c(H,L) \\
\leq & \frac{1}{2}h_2\mu_2\Big\{\Big(\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}+\frac{(\rho_1^h-1)^2}{(1-\rho_1^l)}\Big)(2H-\psi_1-\tilde{\psi}_1)(\psi_1-\tilde{\psi}_1) \\
& -\rho_2(H-\psi_2)^2+\frac{2\rho_2(\rho_1^h-1)}{(1-\rho_1^l)}\psi_2(H-\tilde{\psi}_1) \\
& +2(1-\rho_2)\tilde{\psi}_2\tilde{\psi}_1 \\
& +\frac{2(1-\rho_2)(\rho_1^h-\rho_1^l)}{(1-\rho_1^l)}\tilde{\psi}_2(H-\tilde{\psi}_1)+\rho_2\psi_2^2\Big\}.
\end{aligned}$$

Since $(1-\rho_2)\tilde{\psi}_2=-\rho_2\psi_2$, we can simplify the above upper bound as follows

$$\frac{1}{2}h_2\mu_2\Big\{\Big(\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}+\frac{(\rho_1^h-1)^2}{(1-\rho_1^l)}\Big)(2H-\psi_1-\tilde{\psi}_1)(\psi_1-\tilde{\psi}_1)-\rho_2H^2\Big\}.$$

Then we have

$$c^{\mathrm{FP2-FP1}}(H,L)-c(H,L)$$

$$
\begin{aligned}
&\leq \frac{1}{2}h_2\mu_2\Big\{\Big(\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}+\frac{(\rho_1^h-1)^2}{(1-\rho_1^l)}\Big)(2H-\psi_1-\tilde{\psi}_1)(\psi_1-\tilde{\psi}_1)\Big\}\\
&\leq \frac{1}{2}h_1\mu_1\Big\{\Big(\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{(1-\rho_1^l)}+\frac{(\rho_1^h-1)^2}{(1-\rho_1^l)}\Big)(2H-\psi_1-\tilde{\psi}_1)(\psi_1-\tilde{\psi}_1)\Big\}
\end{aligned}
$$

where the last inequality follows from the definition of η and our assumption that $h_1\mu_1 > h_2\mu_2$.

- If $(H, L) \in \Delta_2$, since $B \leq \tilde{\psi}_1$ and $\gamma_3 \geq \gamma_2$, (H, L) belongs to the union of the regions considered in (FP2-FP1:1) and (FP2-FP1:2) for FP2-FP1 policy (see Appendix C). If (H, L) belongs to Δ_2 and the region considered in (FP2-FP1:1), then (H, L) belongs to the region considered in (OPT:4) for the optimal policy; if (H, L) belongs to Δ_2 and the region considered in (FP2-FP1:2), then (H, L) belongs to the region considered in (OPT:5) for the optimal policy. Note that the cost under FP2-FP1 policy in (FP2-FP1:2) is different from that in (FP2:FP1:1) by the same amount as optimal cost in (OPT:5) is different from that in (OPT:4). Thus, it suffices to compare the cost under FP2-FP1 policy when (H, L) belongs to the region in (FP2-FP1:1) with the cost under the optimal policy when (H, L) belongs to the region given by (OPT:4).

  Note that if (H, L) belongs to Δ_2 and the region given in (FP2-FP1:1), we have $H - \tilde{\psi}_1 \geq 0$ and $0 \leq \tilde{\psi}_1$. Hence, from the cost expressions given in Appendix C, we have

$$
\begin{aligned}
c^{\mathrm{FP2-FP1}}(H,L) \;\leq\; & \frac{1}{2}h_2\mu_2\Big\{\Big(\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}+\frac{\rho_2(\rho_1^h-\rho_1^l)^2}{(1-\rho_1^l)^2}\Big)(H-\tilde{\psi}_1)^2\\
&+2(1-\rho_2)\tilde{\psi}_2\tilde{\psi}_1+\frac{2(1-\rho_2)(\rho_1^h-\rho_1^l)}{1-\rho_1^l}\tilde{\psi}_2(H-\tilde{\psi}_1)\\
&+(1-\rho_2-\rho_1^l)\Big(\frac{(\rho_1^h+\rho_2-1)}{(1-\rho_2-\rho_1^l)}(H-a_1)-\frac{\rho_1^h-1}{1-\rho_1^l}(H-\tilde{\psi}_1)\Big)^2\Big\}.
\end{aligned}
$$

  Plugging in the above upper bound for $c^{\mathrm{FP2-FP1:1}}(H, L)$, we have

$$
\begin{aligned}
&c^{\mathrm{FP2-FP1}}(H,L)-c(H,L)\\
\leq\; & \frac{1}{2}h_2\mu_2\Big\{\Big(\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}+\frac{\rho_2(\rho_1^h-\rho_1^l)^2}{(1-\rho_1^l)^2}\Big)(H-\tilde{\psi}_1)^2\\
&+2(1-\rho_2)\tilde{\psi}_2\tilde{\psi}_1+\frac{2(1-\rho_2)(\rho_1^h-\rho_1^l)}{1-\rho_1^l}\tilde{\psi}_2(H-\tilde{\psi}_1)\\
&+(1-\rho_2-\rho_1^l)\Big(\frac{(\rho_1^h+\rho_2-1)}{(1-\rho_2-\rho_1^l)}(H-a_1)-\frac{\rho_1^h-1}{1-\rho_1^l}(H-\tilde{\psi}_1)\Big)^2\\
&-\Big(\frac{(1-\rho_1^l)(\rho_1^h-\rho_1^l)}{\eta(\rho_1^h-1)}(t_2-H)^2+\rho_2(t_2-s)^2\\
&+(1-\rho_2)(2\tilde{\psi}_2-s)s+2(1-\rho_2)(\tilde{\psi}_2-s)(t_2-s)\\
&+(1-\rho_2-\rho_1^l)\Big(\frac{\rho_1^h+\rho_2-1}{1-\rho_2-\rho_1^l}(H-a_1)-(t_2-H)\Big)^2\Big)\Big\}
\end{aligned}
$$

  where $s = s_1 = s_2$ and t_2 are computed through solving equations (12) to (19) and are given in Appendix C as

$$
s = \frac{d_1/\mu_1-(\rho_1^h-1)(1-\eta)t_2}{1+\eta(\rho_1^h-1)},
$$

$$t_2 = \frac{(\rho_1^h - \rho_1^l)(1 + \eta(\rho_1^h - 1)H - \eta(\rho_1^h - 1)d_1/\mu_1}{\gamma_{2d}},$$

where $\gamma_{2d} = (1-\rho_1^l)(1+\eta(\rho_1^h-1)) + (1-\eta)(\rho_1^h-1)$. The above upper bound can be further simplified as

$$\begin{aligned}
& c^{\mathrm{FP2-FP1}}(H,L) - c(H,L) \\
\leq\ & \frac{1}{2}h_2\mu_2\Big\{\Big(\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)} + \frac{\rho_2(\rho_1^h-\rho_1^l)^2}{(1-\rho_1^l)^2}\Big)(H-\tilde{\psi}_1)^2 \\
& +2(1-\rho_2)\tilde{\psi}_2\tilde{\psi}_1 + \frac{2(1-\rho_2)(\rho_1^h-\rho_1^l)}{1-\rho_1^l}\tilde{\psi}_2(H-\tilde{\psi}_1) \\
& + \underbrace{(1-\rho_2-\rho_1^l)\Big(\frac{(\rho_1^h+\rho_2-1)}{(1-\rho_2-\rho_1^l)}(H-a_1) - \frac{\rho_1^h-1}{1-\rho_1^l}(H-\tilde{\psi}_1)\Big)^2}_{g_1} \\
& -(1-\rho_2)(2\tilde{\psi}_2 - s)s - 2(1-\rho_2)(\tilde{\psi}_2 - s)(t_2 - s) \\
& - \underbrace{(1-\rho_2-\rho_1^l)\Big(\frac{\rho_1^h+\rho_2-1}{1-\rho_2-\rho_1^l}(H-a_1) - (t_2 - H)\Big)^2}_{f_1}\Big\}.
\end{aligned}$$

We begin with simplifying $g_1 - f_1$ as

$$\begin{aligned}
g_1 - f_1 =\ & \Big((t_2 - H) - \frac{\rho_1^h-1}{1-\rho_1^l}(H-\tilde{\psi}_1)\Big) \\
& \times\Big(2(\rho_1^h+\rho_2-1)(H-a_1) - (1-\rho_2-\rho_1^l)\Big(\frac{\rho_1^h-1}{1-\rho_1^l}(H-\tilde{\psi}_1) + (t_2-H)\Big)\Big)
\end{aligned}$$

With some algebra, we obtain

$$t_2 - H = \frac{\eta\rho_1^h(\rho_1^h-1)}{\gamma_{2d}}(H-\tilde{\psi}_1) = \gamma_2(H-\tilde{\psi}_1),$$

and

$$(t_2 - H) - \frac{\rho_1^h-1}{1-\rho_1^l}(H-\tilde{\psi}_1) = -\frac{(1-\eta)(\rho_1^h-\rho_1^l)(\rho_1^h-1)}{\gamma_{2d}(1-\rho_1^l)}(H-\tilde{\psi}_1) \leq 0.$$

Since $(\rho_1^h+\rho_2-1)(H-a_1) = (\rho_1^h-1)(H-\psi_1) + \rho_2(H-\psi_2)$, we have

$$\begin{aligned}
g_1 - f_1 =\ & -\frac{(1-\eta)(\rho_1^h-\rho_1^l)(\rho_1^h-1)}{\gamma_{2d}(1-\rho_1^l)}(H-\tilde{\psi}_1)(2(\rho_1^h-1)(H-\psi_1) + 2\rho_2(H-\psi_2)) \\
& -(1-\rho_2-\rho_1^l)\Big((t_2-H) - \frac{\rho_1^h-1}{1-\rho_1^l}(H-\tilde{\psi}_1)\Big)\Big((t_2-H) + \frac{\rho_1^h-1}{1-\rho_1^l}(H-\tilde{\psi}_1)\Big) \\
\leq\ & -\frac{2(1-\eta)(\rho_1^h-\rho_1^l)(\rho_1^h-1)^2}{\gamma_{2d}(1-\rho_1^l)}(H-\tilde{\psi}_1)(H-\psi_1) \\
& -\frac{2(1-\eta)(\rho_1^h-\rho_1^l)(\rho_1^h-1)\rho_2}{\gamma_{2d}(1-\rho_1^l)}(H-\tilde{\psi}_1)(H-\psi_2) \\
& +\frac{(1-\rho_2-\rho_1^l)(\rho_1^h-1)^2}{(1-\rho_1^l)^2}(H-\tilde{\psi}_1)^2.
\end{aligned}$$

Since $(H-\tilde{\psi}_1)(H-\psi_2) \geq -\psi_2(H-\tilde{\psi}_1)$,

$$\begin{aligned} g_1 - f_1 \quad \leq \quad & -\frac{2(1-\eta)(\rho_1^h-\rho_1^l)(\rho_1^h-1)^2}{\gamma_{2d}(1-\rho_1^l)}(H-\tilde{\psi}_1)(H-\psi_1) \\ & +\frac{2(1-\eta)(\rho_1^h-\rho_1^l)(\rho_1^h-1)\rho_2}{\gamma_{2d}(1-\rho_1^l)}\psi_2(H-\tilde{\psi}_1) \\ & +\frac{(1-\rho_2-\rho_1^l)(\rho_1^h-1)^2}{(1-\rho_1^l)^2}(H-\tilde{\psi}_1)^2. \end{aligned}$$

Plugging in this expression into the upper bound of $c^{\mathrm{FP2-FP1}}(H,L)-c(H,L)$, we get

$$\begin{aligned} c^{\mathrm{FP2-FP1}}(H,L)-c(H,L) \quad \leq \quad & \frac{1}{2}h_2\mu_2\Big\{\Big(\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}+\frac{(\rho_1^h-\rho_1^l)^2}{(1-\rho_1^l)}\Big)(H-\tilde{\psi}_1)^2 \\ & -\frac{2(1-\eta)(\rho_1^h-\rho_1^l)(\rho_1^h-1)^2}{\gamma_{2d}(1-\rho_1^l)}(H-\tilde{\psi}_1)(H-\psi_1) \\ & +\Big(\frac{2(1-\eta)(\rho_1^h-\rho_1^l)(\rho_1^h-1)\rho_2}{\gamma_{2d}(1-\rho_1^l)}\psi_2(H-\tilde{\psi}_1) \\ & +2(1-\rho_2)\tilde{\psi}_2\tilde{\psi}_1+\frac{2(1-\rho_2)(\rho_1^h-\rho_1^l)}{1-\rho_1^l}\tilde{\psi}_2(H-\tilde{\psi}_1) \\ & -(1-\rho_2)(2\tilde{\psi}_2-s)s-2(1-\rho_2)(\tilde{\psi}_2-s)(t_2-s)\Big)\Big\}. \end{aligned}$$

Since $s=(t_2-H)+H-(t_2-s)$, we can further simplify the last line of the above upper bound as

$$\begin{aligned} & -(1-\rho_2)(2\tilde{\psi}_2-s)s-2(1-\rho_2)(\tilde{\psi}_2-s)(t_2-s) \\ = \quad & -2(1-\rho_2)\tilde{\psi}_2(t_2-H)-2(1-\rho_2)\tilde{\psi}_2H+(1-\rho_2)s^2+2(1-\rho_2)s(t_2-s) \\ = \quad & -2(1-\rho_2)\gamma_2\tilde{\psi}_2(H-\tilde{\psi}_1)-2(1-\rho_2)\tilde{\psi}_2H+(1-\rho_2)s^2+2(1-\rho_2)s(t_2-s). \end{aligned}$$

Plugging in this result, the expressions for γ_2 and γ_{2d} and noting that $\rho_2\psi_2=-(1-\rho_2)\tilde{\psi}_2$, we can simplify the part in the parenthesis of the above upper bound of $c^{\mathrm{FP2-FP1}}(H,L)-c(H,L)$ and get

$$\begin{aligned} & \Big(\frac{2(1-\eta)(\rho_1^h-\rho_1^l)(\rho_1^h-1)\rho_2}{\gamma_{2d}(1-\rho_1^l)}\psi_2(H-\tilde{\psi}_1) \\ & +2(1-\rho_2)\tilde{\psi}_2\tilde{\psi}_1+\frac{2(1-\rho_2)(\rho_1^h-\rho_1^l)}{1-\rho_1^l}\tilde{\psi}_2(H-\tilde{\psi}_1) \\ & -(1-\rho_2)(2\tilde{\psi}_2-s)s-2(1-\rho_2)(\tilde{\psi}_2-s)(t_2-s)\Big) \\ = \quad & (1-\rho_2)s^2+2(1-\rho_2)s(t_2-s). \end{aligned}$$

Hence,

$$\begin{aligned} c^{\mathrm{FP2-FP1}}(H,L)-c(H,L) \quad \leq \quad & \frac{1}{2}h_2\mu_2\Big\{\Big(\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}+\frac{(\rho_1^h-\rho_1^l)^2}{(1-\rho_1^l)}\Big)(H-\tilde{\psi}_1)^2 \\ & -\frac{2(1-\eta)(\rho_1^h-\rho_1^l)(\rho_1^h-1)^2}{\gamma_{2d}(1-\rho_1^l)}(H-\tilde{\psi}_1)(H-\psi_1) \end{aligned}$$

$$
\begin{aligned}
&+(1-\rho_2)s^2+2(1-\rho_2)s(t_2-s)\\
\leq\ & \frac{1}{2}h_1\mu_1\Big\{\Big(\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{(1-\rho_1^l)}+\frac{(\rho_1^h-\rho_1^l)^2}{(1-\rho_1^l)}\Big)(H-\tilde{\psi}_1)^2\\
&+\frac{2(1-\eta)(\rho_1^h-\rho_1^l)(\rho_1^h-1)^2}{\gamma_{2d}(1-\rho_1^l)}(H-\tilde{\psi}_1)|(H-\psi_1)|\\
&+(1-\rho_2)s^2+2(1-\rho_2)s(t_2-s)\Big\}.
\end{aligned}
$$

Since s and t_2 are finite linear functions of H as $\rho_2 \to 1-\rho_1^l$, the above upper bound is a finite quadratic function of H as $\rho_2 \to 1-\rho_1^l$.

- If $(H,L) \in \Delta_3$, then $(H,\ L)$ belongs to the union of the regions considered in (FP2-FP1:4), (FP2-FP1:5), (FP2-FP1:6), (FP2-FP1:7) and (FP2-FP1:8) for FP2-FP1 policy given in Appendix C. If (H,L) belongs to the region in (FP2-FP1:4), then (H,L) is in the region considered in (OPT:9) for the optimal policy. If (H,L) belongs to the region in (FP2-FP1:5), then (H,L) is in the region given in (OPT:10). If (H,L) belongs to the region in (FP2-FP1:6), then (H,L) is in the region considered in (OPT:11). If (H,L) belongs to the region in (FP2-FP1:7), then (H,L) is in the region given in (OPT:7). Finally, if (H,L) belongs to the region in (FP2-FP1:8), then (H,L) is in the region given in (OPT:8). Using the corresponding holding cost expressions, we have

$$
c^{\mathrm{FP2-FP1}}(H,L)-c(H,L) \quad = \quad 0.
$$

- Finally, if $(H,L) \in \Delta_4$, then from the holding cost expression in (FP2-FP1:3) given in Appendix C, we have

$$
\begin{aligned}
c^{\mathrm{FP2-FP1}}(H,L) \quad =\quad & \frac{1}{2}h_2\mu_2\Big\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(H-\tilde{\psi}_1)^2+\rho_2(H+L-\tilde{\psi}_1)^2\\
&+(1-\rho_2)(2\tilde{\psi}_2-\tilde{\psi}_1)\tilde{\psi}_1+2(1-\rho_2)(\tilde{\psi}_2-\tilde{\psi}_1)(H+L-\tilde{\psi}_1)\\
&-\frac{1-\rho_1^l}{\eta}\left[\frac{\rho_1^h-1}{1-\rho_1^l}(H-\tilde{\psi}_1)-L\right]^2\Big\}\\
\leq\quad & \frac{1}{2}h_2\mu_2\Big\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(H-\tilde{\psi}_1)^2+\rho_2(H+L-\tilde{\psi}_1)^2\\
&+2(1-\rho_2)\tilde{\psi}_2\tilde{\psi}_1+2(1-\rho_2)\tilde{\psi}_2(H-\tilde{\psi}_1)+2(1-\rho_2)\tilde{\psi}_2 L\Big\}. \qquad (33)
\end{aligned}
$$

Note that $L \leq \gamma_3(H-\tilde{\psi}_1)$. For notational convenience define

$$
g_4 \quad = \quad 2(1-\rho_2)\tilde{\psi}_2\tilde{\psi}_1+2(1-\rho_2)(1+\gamma_3)\tilde{\psi}_2(H-\tilde{\psi}_1).
$$

Then

$$
c^{\mathrm{FP2-FP1}}(H,L) \quad \leq \quad \frac{1}{2}h_2\mu_2\Big\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(H-\tilde{\psi}_1)^2+\rho_2(H+L-\tilde{\psi}_1)^2+g_4\Big\}.
$$

If $(H,L) \in \Delta_4$, then $\tilde{\psi}_1 \leq \tilde{\psi}_2$. Since $a_2 \geq \psi_1 > \tilde{\psi}_1 \geq a_1$, $\gamma_3 \leq \gamma_4$, and $\gamma_1(H-a_1) \leq \gamma_3(H-\tilde{\psi}_1)$, one can conclude that (H,L) belongs to the union of the regions considered in (OPT:2), (OPT:3), (OPT:5) and (OPT:6) of Appendix C. We are going to compare the cost difference between the FP2-FP1 policy and the optimal policy by employing the cost expressions for each region separately.

If (H, L) is in the region given in (OPT:2), then

$$\begin{aligned} & (1-\rho_2-\rho_1^l)\left(\frac{\rho_1^h+\rho_2-1}{1-\rho_1^l-\rho_2}(H-a_1)-\frac{\rho_1^h-1}{1-\rho_1^l}(H-\psi_1)\right)^2 \\ \geq\ & (1-\rho_2-\rho_1^l)\left(\frac{\rho_1^h+\rho_2-1}{1-\rho_1^l-\rho_2}(H-a_1)-L\right)^2. \end{aligned}$$

Since in this case $\psi_1^- = 0$ and $\psi_2^- = -\psi_2$,

$$c(H,L) \ \geq\ \frac{1}{2}h_2\mu_2\Big\{-\rho_2\psi_2^2+\frac{(1-\rho_2-\rho_1^l)^2}{\rho_2}\Big(\frac{\rho_1^h+\rho_2-1}{1-\rho_2-\rho_1^l}(H-a_1)-\frac{\rho_1^h-1}{1-\rho_1^l}(H-\psi_1)\Big)^2\Big\}.$$

Replacing $(\rho_1^h+\rho_2-1)(H-a_1)$ by $(\rho_1^h-1)(H-\psi_1)+\rho_2(H-\psi_2)$, we have

$$\begin{aligned} & \Big((\rho_1^h+\rho_2-1)(H-a_1)-\frac{(1-\rho_2-\rho_1^l)(\rho_1^h-1)}{1-\rho_1^l}(H-\psi_1)\Big)^2 \\ =\ & \rho_2^2\Big(\frac{\rho_1^h-1}{1-\rho_1^l}(H-\psi_1)+(H-\psi_2)\Big)^2. \end{aligned}$$

Thus,

$$\begin{aligned} c(H,L) &\geq \frac{1}{2}h_2\mu_2\Big\{-\rho_2\psi_2^2+\rho_2(H-\psi_2)^2+\frac{2\rho_2(\rho_1^h-1)}{1-\rho_1^l}(H-\psi_2)(H-\psi_1)\Big\} \\ &\geq \frac{1}{2}h_2\mu_2\Big\{-2\rho_2\psi_2H+\frac{2\rho_2(\rho_1^h-1)}{1-\rho_1^l}(H-\psi_2)(H-\psi_1)\Big\} \end{aligned}$$

Note that in the region considered in (OPT:2), $H \geq a_2 \geq \psi_1$. Moreover since $H-\psi_1 \geq 0$, $(H-\psi_2)(H-\psi_1) \geq -\psi_2(H-\psi_1)$ and

$$\begin{aligned} c(H,L) &\geq \frac{1}{2}h_2\mu_2\Big\{-2\rho_2\psi_2H-\frac{2\rho_2(\rho_1^h-1)}{1-\rho_1^l}\psi_2(H-\psi_1)\Big\} \\ &= \frac{1}{2}h_2\mu_2 f_4, \end{aligned}$$

where

$$f_4 \ =\ -2\rho_2\psi_2H-\frac{2\rho_2(\rho_1^h-1)}{1-\rho_1^l}\psi_2(H-\psi_1).$$

Since $\rho_2\psi_2 = -(1-\rho_2)\tilde{\psi}_2$, with some algebra one can verify that $g_4 - f_4 = 0$ and

$$\begin{aligned} c^{\mathrm{FP2-FP1}}(H,L)-c(H,L) &\leq \frac{1}{2}h_2\mu_2\Big\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(H-\tilde{\psi}_1)^2+\rho_2(H+L-\tilde{\psi}_1)^2\Big\} \\ &\leq \frac{1}{2}h_1\mu_1\Big\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{(1-\rho_1^l)}(H-\tilde{\psi}_1)^2+(1+\gamma_3)^2(H-\tilde{\psi}_1)^2\Big\}, \end{aligned}$$

which is a finite quadratic function of H as $\rho_2 \to 1-\rho_1^l$.

If (H,L) is in the region given in (OPT:3), then since $\psi_1^- = 0$ and $\psi_2^- = -\psi_2$, we have

$$\begin{aligned} c(H,L) &\geq \frac{1}{2}h_2\mu_2\Big\{\rho_2(H+L-\psi_2)^2 - \rho_2\psi_2^2 - \frac{1-\rho_1^l}{\eta}\Big(\frac{\rho_1^h-1}{1-\rho_1^l}(H-\psi_1)-L\Big)^2\Big\} \\ &\geq \frac{1}{2}h_2\mu_2\Big\{-2\rho_2\psi_2(H+L) - \frac{1-\rho_1^l}{\eta}\Big(\frac{\rho_1^h-1}{1-\rho_1^l}(H-\psi_1)-L\Big)^2\Big\}. \end{aligned}$$

From this result, the upper bound given in (33) and the fact that $\rho_2\psi_2 = -(1-\rho_2)\tilde{\psi}_2$, we have

$$\begin{aligned} c^{\mathrm{FP2-FP1}}(H,L) - c(H,L) &\leq \frac{1}{2}h_2\mu_2\Big\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(H-\tilde{\psi}_1)^2 + \rho_2(H+L-\tilde{\psi}_1)^2 \\ &\quad + \frac{1-\rho_1^l}{\eta}\Big(\frac{\rho_1^h-1}{1-\rho_1^l}(H-\psi_1)-L\Big)^2\Big\}. \end{aligned}$$

Note that $0 \leq L \leq \gamma_3(H-\tilde{\psi}_1)$. Plugging in the expression of γ_3, we can further simplify the above upper bound as

$$\begin{aligned} c^{\mathrm{FP2-FP1}}(H,L) - c(H,L) &\leq \frac{1}{2}h_2\mu_2\Big\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(H-\tilde{\psi}_1)^2 + \frac{\rho_2(\rho_1^h-\rho_1^l)^2}{(1-\rho_1^l)^2}(H-\tilde{\psi}_1)^2 \\ &\quad + \frac{(\rho_1^h-1)^2}{\eta(1-\rho_1^l)}((H-\psi_1)^2 + (H-\tilde{\psi}_1)^2)\Big\}. \end{aligned}$$

Note that $\eta = h_2\mu_2(h_1\mu_1)^{-1} \leq 1$ and $\rho_2 \leq 1$. Then

$$\begin{aligned} c^{\mathrm{FP2-FP1}}(H,L) - c(H,L) &\leq \frac{1}{2}h_1\mu_1\Big\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{(1-\rho_1^l)}(H-\tilde{\psi}_1)^2 + \frac{(\rho_1^h-\rho_1^l)^2}{(1-\rho_1^l)^2}(H-\tilde{\psi}_1)^2 \\ &\quad + \frac{(\rho_1^h-1)^2}{(1-\rho_1^l)}((H-\psi_1)^2 + (H-\tilde{\psi}_1)^2)\Big\}. \end{aligned}$$

If (H,L) is in the region given in (OPT:5), we have

$$\begin{aligned} & (1-\rho_2-\rho_1^l)\Big(\frac{\rho_1^h+\rho_2-1}{1-\rho_2-\rho_1^l}(H-a_1)-(t_2-H)\Big)^2 \\ - \;& (1-\rho_2-\rho_1^l)\Big(\frac{\rho_1^h+\rho_2-1}{1-\rho_2-\rho_1^l}(H-a_1)-L\Big)^2 \\ = \;& (H+L-t_2)\Big(2(\rho_1^h+\rho_2-1)(H-a_1) - (1-\rho_2-\rho_1^l)(t_2-H+L)\Big) \\ = \;& 2\rho_2(H+L-t_2)(H-\psi_2) + 2(\rho_1^h-1)(H+L-t_2)(H-\psi_1) \\ & +(1-\rho_2-\rho_1^l)((t_2-H)^2-L^2) \\ \geq \;& -2\rho_2\psi_2(H+L-t_2) + 2(\rho_1^h-1)(H+L-t_2)(H-\psi_1) - (1-\rho_2-\rho_1^l)L^2, \end{aligned}$$

where the inequality follows from the fact that $H+L-t_2 \geq 0$ since $t_2 - H = \gamma_2(H-\tilde{\psi}_1)$ and $L \geq \gamma_2(H-\tilde{\psi}_1)$ in (OPT:5). Thus,

$$\begin{aligned} c(H,L) &\geq \frac{1}{2}h_2\mu_2\Big\{(1-\rho_2)(2\tilde{\psi}_2-s)s + 2(1-\rho_2)(\tilde{\psi}_2-s)(t_2-s) \\ &\quad -2\rho_2\psi_2(H+L-t_2) + 2(\rho_1^h-1)(H+L-t_2)(H-\psi_1) - (1-\rho_2-\rho_1^l)L^2\Big\}. \end{aligned}$$

Plugging in the upper bound for $c^{\text{FP2}-\text{FP1}}(H,L)$ given in (33) and noting that $\rho_2\psi_2 = -(1-\rho_2)\tilde{\psi}_2$, we have

$$\begin{aligned}
c^{\text{FP2}-\text{FP1}}(H,L) - c(H,L) \;\le\;& \frac{1}{2}h_2\mu_2\Big\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(H-\tilde{\psi}_1)^2 + \rho_2(H+L-\tilde{\psi}_1)^2 \\
&+(1-\rho_2)s^2 + 2(1-\rho_2)s(t_2-s) \\
&-2(\rho_1^h-1)(H+L-t_2)(H-\psi_1) + (1-\rho_2-\rho_1^l)L^2\Big\} \\
\le\;& \frac{1}{2}h_1\mu_1\Big\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{(1-\rho_1^l)}(H-\tilde{\psi}_1)^2 + (H+L-\tilde{\psi}_1)^2 \\
&+(1-\rho_2)s^2 + 2(1-\rho_2)s(t_2-s) \\
&-2(\rho_1^h-1)(H+L-t_2)(H-\psi_1) + (1-\rho_2-\rho_1^l)L^2\Big\}.
\end{aligned}$$

Note that s and t_2 are all finite linear functions of H as $\rho_2 \to 1-\rho_1^l$. Moreover, $\gamma_2(H-\tilde{\psi}_1) \le L \le \gamma_2(H-\tilde{\psi}_1)$ since $(H,L) \in \Delta_4$ and (OPT:5). Thus, $c^{\text{FP2}-\text{FP1}}(H,L) - c(H,L)$ is bounded by a quadratic function of H as $\rho_2 \to 1-\rho_1^l$.

Finally, if (H,L) is in the region given in (OPT:6), then

$$\begin{aligned}
c(H,L) \;\ge\;& \frac{1}{2}h_2\mu_2\Big\{(1-\rho_2)(2\tilde{\psi}_2 - s)s + 2(1-\rho_2)(\tilde{\psi}_2 - s)(H+L-s) \\
&-\frac{1-\rho_1^l}{\eta}\Big(\frac{\rho_1^h-1}{1-\rho_1^l}(H-t_1) - L\Big)^2\Big\}.
\end{aligned}$$

Plugging in the upper bound of $c^{\text{FP2}-\text{FP1}}(H,L)$ given in (33), we obtain

$$\begin{aligned}
c^{\text{FP2}-\text{FP1}}(H,L) - c(H,L) \;\le\;& \frac{1}{2}h_2\mu_2\Big\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(H-\tilde{\psi}_1)^2 + \rho_2(H+L-\tilde{\psi}_1)^2 \\
&+(1-\rho_2)s^2 + 2(1-\rho_2)s(H+L-s) \\
&+\frac{1-\rho_1^l}{\eta}\Big(\frac{\rho_1^h-1}{1-\rho_1^l}(H-t_1) - L\Big)^2\Big\} \\
\le\;& \frac{1}{2}h_1\mu_1\Big\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{(1-\rho_1^l)}(H-\tilde{\psi}_1)^2 + \rho_2(H+L-\tilde{\psi}_1)^2 \\
&+(1-\rho_2)s^2 + 2(1-\rho_2)s(H+L-s) \\
&+\frac{1-\rho_1^l}{\eta}\Big(\frac{\rho_1^h-1}{1-\rho_1^l}(H-t_1) - L\Big)^2\Big\}.
\end{aligned}$$

Note that s and t_2 are all finite linear functions of H as $\rho_2 \to 1-\rho_1^l$. Moreover, since (H,L) belongs to (OPT:6) (and also Δ_4), $\gamma_2(H-\tilde{\psi}_1) \le L \le \gamma_2(H-\tilde{\psi}_1)$. Thus, $c^{\text{FP2}-\text{FP1}}(H,L) - c(H,L)$ is bounded by a quadratic function of H as $\rho_2 \to 1-\rho_1^l$.

Note that since $0 \le \rho_2 \le 1$ and L is bounded, the above upper bounds could be further simplified so that they do not depend on ρ_2 and L. Therefore, since $\mathbb{E}[H^2] < \infty$, the proof for the case $\tilde{\psi}_1 \le \tilde{\psi}_2$ is complete. The proof for the case $\tilde{\psi}_1 \ge \tilde{\psi}_2$ is similar and it is omitted □

6 Numerical results

In this section, we provide numerical examples to demonstrate the performance of the discrete review policy described in Section 4 in systems with random high and low periods. Ideally, once the exact lengths of the high and low periods (H and L) are known, one can follow the optimal policy in the deterministic case described in Section 3. Recall that $c(H, L)$ denotes the total holding cost under the optimal policy when the lengths of the high and low periods are known. Since one can not observe the true lengths of the either periods until they end, such a policy is not implementable. However, the quantity $\mathbb{E}[c(H, L)]$ can be used as a lower bound of the cost function since no other policy can outperform such a policy with perfect knowledge of H and L. We will use this lower bound (which will be referred as LB) as a guideline to evaluate the performance of other implementable policies.

While implementing the discrete review policy, we use both of the methods given in (24) and (25) to estimate the remaining high period and set $p = 0.25, 0.5$ and 0.75. Recall that the remaining low period is always set equal to its mean. The discrete review policy implemented with the method in (24) (i.e. the remaining high period is set equal to its expected value) will be called DSview1, and the discrete review policies implemented with the method given in (25) with $p = 0.25, 0.5$ and 0.75 will be called DSview2, DSview3, and DSview4, respectively. We compare the expected holding cost of these four policies with the lower bound LB, the expected holding cost of the FP1 policy and the expected holding cost of the π^{a_1} policy.

Even though we have considered several systems, in the interest of space we report our findings from two sets of examples referred as System I and System II respectively. In System I, parameters are set as follow: $\theta_1 = 50$, $\theta_2 = 100$, $h_1 = 2$, $h_2 = 1$, $Z_1(0) = 0$, $Z_2(0) = 90$, $\rho_1^h = 2$, $\rho_1^l = 0.1$ and $\rho_2 = 0.4$. In System II, $\rho_1^l = 0$ and $\rho_2 = 0.95$ and the remaining parameters remain the same. We consider four different distributions (referred as Case A, Case B, Case C and Case D respectively) for the length of the high (H) and the low (L) periods, In Case A, both H and L are Erlang-2 random variables. In Case B, both H and L are exponential random variables. In Cases C and D, both H and L are hyper-exponential random variables with squared coefficient of variation 2 and 10, respectively. Note that the squared coefficient of variation of the distributions in Case A and Case B are 1/2 and 1, respectively. In our experiments, $\mathbb{E}[H]$ attains the values: 5, 12.5, 25, 37.5 and 50 and $\mathbb{E}[L]$ attains the values: 12.5, 25, 50 and 1000.

Under a specified distribution with fixed values of $\mathbb{E}[H]$ and $\mathbb{E}[L]$, we generate 500,000 sets of H and L values. For each set of H and L values, we compute $c(H, L)$ (lower bound), $c^{FP1}(H, L)$, $c^{a_1}(H, L)$ and the holding costs of the four discrete review policies. We then compute the average holding costs over 500,000 replications. In all our numerical experiments, while implementing the discrete review policies, we set τ equal to 0.1. The value of τ is determined by simulating the systems that we consider under the discrete review policies with different τ values and eventually picking the τ value which yields a good holding cost performance while keeping the run times reasonably short. Tables 1 through 4 display the average value of the lower bound on holding cost and the percentage difference off the lower bound of the average holding cost of the FP1, π^{a_1}, DSview1, DSview2, DSview3 and DSview4 policies.

As Tables 1 through 4 show, discrete review policies have a good holding cost performance. The largest percentage difference between the holding cost of discrete review policies and the lower bound on the holding cost is approximately 21%. Moreover, the discrete review policies are more robust than the FP1 and the π^{a_1} policies. Note that the average holding cost under the discrete review policies is much less than the average holding cost under the FP1 policy in Cases A and B when $\mathbb{E}[H]$ is small to moderate. The same result also holds for Case C when $\rho_1^l = 0$ and $\rho_2 = 0.95$.

However, as the variability increases, FP1 policy outperforms all other policies. In particular, in Case D the holding cost under the FP1 policy is less than the holding cost under all discrete review policies except when $E[L]$ is large (see Table 4). Discrete review policies outperform π^{a_1} policy in Cases C and D. When the system variability is low, for systems with $\rho_1^l = 0.1$ and $\rho_2 = 0.4$, the discrete review policies outperform the π^{a_1} policy. For systems with $\rho_1^l = 0$ and $\rho_2 = 0.95$, the same observation holds for the DSview1, DSview3 and DSview4 policies. If $\rho_1^l = 0$ and $\rho_2 = 0.95$, DSview2 has higher holding cost than π^{a_1} policy in Cases A and B when $\mathbb{E}[H]$ is small and $\mathbb{E}[L]$ is not large or when $\mathbb{E}[L]$ is large.

In systems with $\rho_1^l = 0.1$ and $\rho_2 = 0.4$, in general DSview4 policy has a poor performance compared to the other discrete review policies. It performs well only for small values of $\mathbb{E}[H]$ in Case A. On the other hand, DSview2 significantly outperforms DSview1 and DSview3 policies in Cases A and B and in Case C when $\mathbb{E}[H]$ is not large. In Case C, as $\mathbb{E}[H]$ increases, DSview1 policy starts dominating the other discrete review policies. On the other hand, in Case D, DSview1 policy always outperforms the other discrete review policies in systems with $\rho_1^l = 0.1$ and $\rho_2 = 0.4$. The same assertion holds for systems with $\rho_1^l = 0$ and $\rho_2 = 0.95$ except when $E[L]$ and $E[H]$ are both large (see Table 4).

In systems with $\rho_1^l = 0$ and $\rho_2 = 0.95$, the performances of DSview2 and DSview4 policies depend on the expected length of the low period. Even though the DSview4 policy shows poor performance (compared to the other discrete review policies) when $\mathbb{E}[L]$ is small, its performance improves (in particular in Cases A and B) as $\mathbb{E}[L]$ gets large. On the other hand, even though DSview2 policy has one of the best performances among the discrete review policies when $\mathbb{E}[L]$ is small, its performance deteriorates in Cases A and B as $\mathbb{E}[L]$ gets large. However, in Cases C and D, DSview1 and DSview2 policies always have better holding cost performance than the other discrete review policies.

In conclusion, discrete review policies yield good holding cost performance and they are robust with respect to the system parameters. Among the discrete review policies, one can employ the DSview2 policy (in order to reduce the total holding cost) if class 2 is not heavily loaded and the coefficient of variation of the high period is not large. However, if the coefficient of variation of the high period is large, DSview1 policy seems to outperform the other discrete review policies. On the other hand, if class 2 is heavily loaded, DSview1 policy has a good overall policy. If the system variability is high (as in Case D), FP1 policy has the best performance among all policies.

7 Summary and conclusions

We studied the dynamic scheduling of different classes of service in a fluid model of computing paradigms for Internet services that may be overloaded for a transient period. We focused on minimizing the penalty of the hosting service provider by scheduling its server resources among various e-commerce sites under Service-Level-Agreement (SLA) contracts with specific Quality-of-Service (QoS) performance guarantees for each class of service.

Our focus in this paper was on a system with two fluid classes and a single server whose capacity can be shared arbitrarily among the two classes. To capture the QoS performance guarantees in the SLA contracts, we introduced a threshold value for each fluid class such that a holding cost is incurred only if the amount of fluid of a certain class exceeds its threshold value. We assumed that the class 1 arrival rate changes with time and the class 1 fluid can more efficiently reduce the holding cost. Under these assumptions, our objective is to specify the optimal server allocation policy that minimizes the total holding cost.

We first considered the case that the arrival rate function for class 1 fluid is known. In this deterministic setting we could completely characterize the optimal server allocation policy that minimizes the holding cost. We then studied the stochastic fluid system when the arrival rate function for class 1 is random. Using the key insights gained from the optimal policy in the deterministic setting, we developed simple (heuristic) server allocation policies. These policies called "discrete review policies" are not only easy to implement but also shown to be strongly asymptotically optimal for the two heavy traffic regimes considered in this paper. Moreover, numerical studies have also demonstrated that discrete review policies yield good holding cost performance in general (not only in the asymptotic sense) and they are robust with respect to the system parameters such as load and class 1 arrival rate function.

Appendices

A An Optimal policy if $h_1\mu_1 \leq h_2\mu_2$

Under the assumption that the class 2 has constant arrival rate λ_2 and $\rho_2 < 1$, if $h_1\mu_1 \leq h_2\mu_2$, the optimal policy is a generalization of the $c\mu$ rule. Such an optimal policy is given below. The optimality of this policy can be proven using the techniques in Appendix B as is done when the assumption in (5) holds.

- If $Z_2(t) > \theta_2$, full capacity is given to class 2, i.e. $\dot{T}_1(t) = 0$, $\dot{T}_2(t) = 1$.
- If $Z_2(t) = \theta_2$ and $Z_1(t) > \theta_1$, enough capacity is given to class 2 such that class 2 fluid level is kept at θ_2 and the remaining capacity is used to serve class 1, i.e. $\dot{T}_1(t) = 1 - \rho_2$, $\dot{T}_2(t) = \rho_2$.
- If $Z_2(t) < \theta_2$ and $Z_1(t) \geq \theta_1$, full capacity is given to class 1, i.e. $\dot{T}_1(t) = 1$, $\dot{T}_2(t) = 0$.
- If $Z_2(t) < \theta_2$ and $Z_1(t) < \theta_1$, and the system is in the high load period ($t < H$), full capacity is given to class 1, i.e. $\dot{T}_1(t) = 1$, $\dot{T}_2(t) = 0$.
- If $Z_2(t) \leq \theta_2$ and $Z_1(t) \leq \theta_1$, and the system is in the low period ($H < t < H + L$), enough capacity is given to each class such that the fluid levels of both classes are kept below their threshold values. We have multiple choices in this case, one is to let $\dot{T}_1(t) \geq \rho_1^l$, $\dot{T}_2(t) \geq \rho_2$ such that $\dot{T}_1(t) + \dot{T}_2(t) \leq 1$.

B Proofs for the optimality of the policies in Section 3

Before we prove the optimality of the policies given in Section 3, we provide a lemma related to the Pontryagin maximum principle. Originally, this lemma was given in Seierstad and Sydsaeter [14] but the version stated here was tailored for our problem. For completeness, we also provide the proof of the lemma.

Consider an optimal control problem as follows,

$$\max \int_{B_0}^{B_1} f^0(x(t), u(t), t)\, dt \tag{34}$$

such that

$$\dot{x}(t) = f(x(t), u(t), t), \tag{35}$$

$$x(B_0) = x_0, \tag{36}$$

$$x(B_1) \geq x_1, \tag{37}$$

$$u(t) \in \mathsf{U} \text{ where } \mathsf{U} \subset \mathbb{R}^r \text{ and } (x(t), u(t)) \in \mathbb{R}^n \times \mathbb{R}^r, \tag{38}$$

where $f^0(x(t), u(t), t)$, and $f(x(t), u(t), t)$ are continuous functions of t over $[B_0,\ B_1]$ except at finite number of points.

We say that $(x(t), u(t))$ is an *admissible pair* if $x(t)$ is absolutely continuous, $u(t)$ is piecewise continuous, and they satisfy (35) to (38). We want to find an optimal admissible pair $(x(t),\ u(t))$ that maximizes integral in (34). In the following lemma, for vectors a and b, $a \cdot b$ denotes the usual inner product of a and b.

Lemma 13. *Let $(\bar{x}(t), \bar{u}(t))$ be an admissible pair for the problem given in (34) to (38). Suppose there exists a continuous function $p(t) = (p_1(t),\ p_2(t), \ldots, p_n(t))$ on $[B_0,\ B_1]$ such that it has a piecewise continuous derivative $\dot{p}(t)$, the continuity of $\dot{p}(t)$ is violated only at finite number of points, and $p(t)$ satisfies*

$$p_i(B_1) \geq 0, \text{ and } p_i(B_1) = 0 \text{ if } \bar{x}_i(B_1) > x_1^i,\ \forall i = 1, \ldots, n. \tag{39}$$

In addition, the Hamiltonian function

$$H(x(t), u(t), p(t), t) = f^0(x(t), u(t), t) + p(t) \cdot f(x(t), u(t), t) \tag{40}$$

satisfies the following

$$H(\bar{x}(t), \bar{u}(t), p(t), t) - H(x(t), u(t), p(t), t) \geq \dot{p}(t) \cdot (x(t) - \bar{x}(t)) \tag{41}$$

for all admissible pairs $(x(t), u(t))$, for all $t \in [B_0, B_1]$ except at finite number of points. Then $(\bar{x}(t), \bar{u}(t))$ is an optimal pair for problem (34) to (38).

Proof. We use Δ to denote the following

$$\Delta = \int_{B_0}^{B_1} f^0(\bar{x}(t), \bar{u}(t), t)\mathrm{d}t - \int_{B_0}^{B_1} f^0(x(t), u(t), t)\, dt.$$

Then the optimality of $(\bar{x}(t), \bar{u}(t))$ is equivalent to $\Delta \geq 0$ for all admissible pairs $(x(t), u(t))$.

According to (40) we have

$$\begin{aligned} \Delta &= \int_{B_0}^{B_1} [H(\bar{x}(t), \bar{u}(t), p(t), t) - H(x(t), u(t), p(t), t)]\, dt \\ &\quad + \int_{B_0}^{B_1} p(t) \cdot [f(x(t), u(t), t) - f(\bar{x}(t), \bar{u}(t), t)]\, dt. \end{aligned}$$

It then follows from (35) and (41) that

$$\Delta \geq \int_{B_0}^{B_1} \dot{p}(t) \cdot [x(t) - \bar{x}(t)]\, dt + \int_{B_0}^{B_1} p(t) \cdot [\dot{x}(t) - \dot{\bar{x}}(t)]\, dt.$$

Assume that $B_0 = \xi_0 < \xi_1 < \cdots \xi_k < \xi_{k+1} = B_1$, are all the possible discontinuity points of $\dot{p}(t)$, $\dot{x}(t)$ and $\dot{\bar{x}}(t)$. So the right hand side of the above inequality can be written as

$$\begin{aligned} &\sum_{i=0}^{k} \Big\{ \int_{\xi_i}^{\xi_{i+1}} \dot{p}(t) \cdot [x(t) - \bar{x}(t)]\mathrm{d}t + \int_{\xi_i}^{\xi_{i+1}} p(t) \cdot [\dot{x}(t) - \dot{\bar{x}}(t)\, dt] \Big\} \\ =\ &\sum_{i=0}^{k} \int_{\xi_i}^{\xi_{i+1}} \frac{d}{dt}[p(t) \cdot (x(t) - \bar{x}(t))] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=0}^{k} \Big[p(\xi_{i+1}) \cdot (x(\xi_{i+1}) - \bar{x}(\xi_{i+1})) - p(\xi_i) \cdot (x(\xi_i) - \bar{x}(\xi_i)) \Big] \\
&= p(B_1) \cdot (x(B_1) - \bar{x}(B_1)) \\
&\geq 0,
\end{aligned}
$$

where the last equality is due to the continuity of $p(t), x(t), \bar{x}(t)$ and (36), and the last inequality is based on (37) and (39). Hence, $\Delta \geq 0$, and the optimality of $(\bar{x}(t), \bar{u}(t))$ is proven. □

We next prove that the policy specified in Section 3 is optimal for our original problem described in Section 2 with deterministic high and low periods. First, replacing $\dot{T}_i(t)$ by $u_i(t)$, notice that our original control problem is equivalent to

$$
\max \quad \int_0^{H+L} \sum_{i=1}^{2} -h_i \left(Z_i(t) - \theta_i\right)^+ \, dt. \tag{42}
$$

$$
\begin{aligned}
\text{such that} \quad & \dot{Z}_i(t) = \lambda_i(t) - \mu_i u_i(t) \qquad i = 1,2 && (43)\\
& Z_i(t) \geq 0 \qquad \forall t \in [0, H+L], \quad i = 1,2 && (44)\\
& u_i(t) \geq 0 \qquad \forall t \in [0, H+L], \quad i = 1,2 && (45)\\
& u_1(t) + u_2(t) \leq 1 \qquad \forall t \in [0, H+L], && (46)
\end{aligned}
$$

where $\lambda_1(t) = \lambda_1^h,\ \forall t \in (0, H)$, and $\lambda_1(t) = \lambda_1^l,\ \forall t \in (H, H+L)$, and $\lambda_2(t) = \lambda_2,\ \forall t \in (0, H+L)$.

Hereafter, we are going to use $u^*(t)$ to denote the proposed policy given in Section 3, and $Z^*(t)$ to denote the fluid level under this policy.

Based on Lemma 13, in order to prove the optimality of (Z^*, u^*), it suffices to construct continuous functions $p_i(t),\ i = 1,2$, with piecewise continuous derivatives such that $(Z^*(t), u^*(t), p(t))$ satisfies (39) and (41). In what follows, we illustrate the basic idea of the construction and proof by focusing on only one special case in Section 3. Notice that other cases can be proved similarly.

B.1 Proof for the optimality of the policy in section 3.1.

Before introducing our construction of p's, we first describe the fluid level evolution of both classes under the policy u^* specified in Section 3.1.

Notice that under the policy u^*, class 1 will have higher priority starting from time s_2 until time t in the low period such that $Z_1^*(t) \leq \theta_1$. Corresponding to this policy, we define two critical time instances for class 1 as follow

$$
\begin{aligned}
t_1 &= \max\{t : s_2 \leq t \leq H,\ Z_1^*(t) \leq \theta_1\}, && (47)\\
t_2 &= \max\{t : H \leq t \leq H+L,\ Z_1^*(t) \geq \theta_1\}, && (48)
\end{aligned}
$$

where t_1 is the time that class 1 increases to its threshold from below in the high period if the duration of high period is long enough and t_2 is the time that class 1 decreases to its threshold from above in the low period if the duration of the low period is long enough.

Similarly, we define two critical time instances for class 2

$$
\begin{aligned}
\tilde{s}_2 &= \max\{t : s_2 \leq t \leq t_2,\ Z_2^*(t) \leq \theta_2\}, && (49)\\
\tilde{t}_2 &= \max\{t : t_2 \leq t \leq H+L,\ Z_2^*(t) \geq \theta_2\}, && (50)
\end{aligned}
$$

where $\tilde{s}_2$ is the time that class 2 increases to its threshold from below during the time interval that class 1 has higher priority, i.e. during interval $[s_2, t_2]$ and $\tilde{t}_2$ is the time that class 2 decreases to its threshold from above in the low period if the duration of the low period is long enough. Note that after class 1 decreases to its threshold from above in the low period at t_2, the Low-period-policy gives enough capacity to class 2 to decrease class 2 fluid level.

Based on the definition of s_1, s_2 (described in Section 3) and the definition of t_1, $t_2, \tilde{s}_2$, $\tilde{t}_2$, we claim the following holds:

Claim 1:

$$\begin{aligned} s_1 \le s_2 \le t_1 \le H \le t_2 \le H + L, \\ s_1 \le s_2 \le \tilde{s}_2 \le t_2 \le \tilde{t}_2 \le H + L, \end{aligned}$$

Claim 2:

$$\begin{aligned}
\forall t \in (0,\, s_1) &\qquad Z_1^*(t) < \theta_1,\; Z_2^*(t) > \theta_2, \\
\forall t \in (s_1,\, s_2) &\qquad Z_1^*(t) < \theta_1,\; Z_2^*(t) \le \theta_2, \\
\\
\forall t \in (s_2,\, t_1) &\qquad Z_1^*(t) < \theta_1, \\
\forall t \in (t_1,\, t_2) &\qquad Z_1^*(t) > \theta_1, \\
\forall t \in (t_2,\, H+L) &\qquad Z_1^*(t) \le \theta_1, \\
\\
\forall t \in (s_2,\, \tilde{s}_2) &\qquad Z_2^*(t) < \theta_2, \\
\forall t \in (\tilde{s}_2,\, \tilde{t}_2) &\qquad Z_2^*(t) > \theta_2, \\
\forall t \in (\tilde{t}_2,\, H+L) &\qquad Z_2^*(t) \le \theta_2.
\end{aligned}$$

For ease of readability, we defer the proof of the claims to the end and next show how to construct the auxiliary functions $p(t)$.

It follows from the Pontryagin maximal principle that the optimal policy has to satisfy $\dot{p}_i(t) = \frac{\partial}{\partial Z_i} H(Z(t), p(t), t)$ at the differentiable points, where the Hamiltonian function is given by

$$H(Z(t), u(t), p(t), t) = \sum_{i=1}^{2} \left(-h_i \left(Z_i(t) - \theta_i \right)^+ + p_i(t)(\lambda_i(t) - \mu_i u_i(t)) \right). \tag{51}$$

We therefore construct $p_i(t)$, $i = 1, 2$ (in a backward fashion) as follows:

$$\begin{aligned}
&\qquad p_i(H+L) = 0;\; i = 1, 2, \\
\forall t \in (\tilde{t}_2,\, H+L): &\qquad \dot{p}_1(t) = 0,\; \dot{p}_2(t) = 0, \\
\forall t \in (t_2,\, \tilde{t}_2): &\qquad \dot{p}_1(t) = \frac{\mu_2 h_2}{\mu_1},\; \dot{p}_2(t) = h_2, \\
\\
\forall t \in (t_1, t_2): &\qquad \dot{p}_1(t) = h_1, \\
\forall t \in (s_2,\, t_1): &\qquad \dot{p}_1(t) = 0, \\
\\
\forall t \in (\tilde{s}_2,\, t_2): &\qquad \dot{p}_2(t) = h_2, \\
\forall t \in (s_2,\, \tilde{s}_2): &\qquad \dot{p}_2(t) = 0,
\end{aligned}$$

$$\forall t \in (s_1,\, s_2): \qquad \dot{p}_1(t) = 0;\, \dot{p}_2(t) = 0,$$
$$\forall t \in (0,\, s_1): \qquad \dot{p}_1(t) = 0;\, \dot{p}_2(t) = h_2.$$

Based on the above construction, we have the following properties stated as Claim 3, whose proof is also deferred to the end of this subsection.

Claim 3:

$$\begin{aligned}
\forall t \in (t_2,\, H+L): &\qquad \mu_1 p_1(t) = \mu_2 p_2(t) \le 0;\\
\forall t \in (s_2,\, t_2): &\qquad \mu_1 p_1(t) < \mu_2 p_2(t) \le 0;\\
\forall t \in (s_1,\, s_2): &\qquad \mu_1 p_1(t) = \mu_2 p_2(t) \le 0;\\
\forall t \in (0,\, s_1): &\qquad 0 \ge \mu_1 p_1(t) > \mu_2 p_2(t).
\end{aligned}$$

Based on Lemma 13, the optimality follows once we show that $(Z^*(t), u^*(t), p(t))$ satisfies (39) and (41). From the construction of $p_i(t)$, (39) holds immediately. It remains to show that (41) holds in each time interval throughout $(0, H+L)$ under all four cases given in (20) to (23). Here, we focus only on Case 2.1 to illustrate the basic idea. The other cases can be proved similarly.

Consider, for example, the first time interval $(0, s_1)$. The policy in this period is $u_1^*(t) = 0$, $u_2^*(t) = 1$, and from Claim 2 we have $Z_1^*(t) < \theta_1$, $Z_2^*(t) > \theta_2$. Note that no other admissible policy can reduce more class 2 fluid level than u^*, thus under any admissible policy $u_i(t)$, the fluid level will satisfy $Z_1(t) < \theta_1$ and $Z_2(t) > \theta_2$ for $t \in (0, s_1)$. Plugging this in (51), we have the left hand side of (41) equal to

$$h_2(Z_2(t) - Z_2^*(t)) + \sum_{i=1}^{2} -\mu_i p_i(t)(u_i^*(t) - u_i(t)).$$

Based on Claim 3, for all t in $(0, s_1)$, we have $-\mu_2 p_2(t) \ge -\mu_1 p_1(t) \ge 0$. Therefore,

$$\sum_{i=1}^{2} -\mu_i p_i(t)(u_i^*(t) - u_i(t)) \ge -\mu_1 p_1(t)(u_1^*(t) + u_2^*(t) - u_1(t) - u_2(t)).$$

Note that $u_1^*(t) + u_2^*(t) = 1$, and the admissible $u_i(t)$, $i = 1, 2$ satisfies $u_1(t) + u_2(t) \le 1$, so the right hand side of the above inequality is non-negative. It follows immediately that (41) holds for all time t in the interval $(0, s_1)$.

Repeating this procedure for the remaining intervals, we can similarly prove that (41) holds for all time t in $(0, H+L)$. Hence the optimality of the proposed policy is guaranteed.

We now prove the three claims we made earlier. Again, we focus only on Case 2.1 to illustrate the basic idea. The other cases can be proved similarly.

- **Proof for Claim 1 and Claim 2 in Case 2.1.** Recall that in Case 2.1, we assume that $Z_1^*(0) < \theta_1$, $Z_2^*(0) > \theta_2$, and condition (10) holds.

In this case, s_1 and s_2 are solved using the equations given in (11) to (19). Simultaneously, we also compute u_1, u_2, t_1 and t_2. They can all be expressed in terms of initial fluid levels $Z_i^*(0)$, $i = 1, 2$, durations of the high and low periods H and L, the arrival rates λ_1^h, λ_1^l, and λ_2, service rates $\mu_i, i = 1, 2$, and holding cost rates h_i, $i = 1, 2$.

Since $Z_2^*(0) > \theta_2$ and $\rho_2 < 1$ (i.e $\lambda_2 < \mu_2$), it follows from (11) that $s_1 > 0$ (s_1 is the time that class 2 decreases to its threshold when it has higher priority). Since $Z_2^*(s_1) = \theta_2$, it follows from

(13) that $u_2 = \rho_2 > 0$. Hence, from (15) $u_1 = 1 - \rho_2 > 0$. One can check that the requirement $t_2 \le H + L$ is equivalent to $L \ge \gamma_1(H - a_1)$. In addition, $t_1 \le H \le t_2$ is equivalent to $a_1 \le H$, and $s_1 \le s_2$ is equivalent to $H \le B$. So, in Case 2.1 of Section 3.1, condition (10) guarantees that we have $0 \le s_1 \le s_2 \le t_1 \le H \le t_2 \le H + L$ and $u_1 > 0$, $u_2 > 0$.

Under the proposed policy, we know that $\lambda_1^h > \mu_1$. Hence, the fluid level $Z_1^*(t)$ increases in the interval $(0, H)$ and $Z_1^*(0) < \theta_1$ and $Z_1^*(t_1) = \theta_1$ (see (16)). Thus, for any $t \in (0, t_1)$, we know that $Z_1^*(t) < \theta_1$ and for any $t \in (t_1, H)$, $Z_1^*(t) > \theta_1$. Under the proposed policy, in the low period, the fluid level $Z_1^*(t)$ decreases until it hits its threshold at t_2 (see (18)). Hence, for any $t \in (t_1, t_2)$, $Z_1^*(t) > \theta_1$. Then we can see that t_1 and t_2 obtained from the set of equations of Case 2.1 coincide with their definitions given in (47) and (48). Hence, the first inequality of claim 1 holds. From the definition of $\tilde{s}_2$ and $\tilde{t}_2$, we can immediately see that the second inequality of claim 1 also holds.

We now prove Claim 2. While proving Claim 1, we have already shown that $Z_1^*(t)$ satisfies the inequalities in Claim 2 for all $t < t_2$. Since $\lambda_2 < \mu_2$ and $u_2 = \rho_2$, under the proposed policy, $Z_2^*(t)$ decreases in the interval $(0, s_1)$, until it reaches θ_2 at s_1 (see (11)). It is kept at its threshold θ_2 in the interval (s_1, s_2) since $\lambda_2 = \mu_2 u_2$. Then it increases in the interval (s_2, H) since class 1 has higher priority. Since $Z_1^*(t) > \theta_1$ in the interval (H, t_2), under the proposed Low-period-policy, class 1 still has higher priority and class 2 fluid continues to increase until class 1 fluid decreases to its threshold at t_2. Hence,

$$\begin{aligned}
\forall t \in (0, s_1), &\qquad Z_2^*(t) > \theta_2,\ Z_2^*(s_1) = \theta_2,\\
\forall t \in (s_1,\, s_2), &\qquad Z_2^*(t) = \theta_2,\ Z_2^*(s_2) = \theta_2,\\
\forall t \in (s_2,\, t_2), &\qquad Z_2^*(t) > \theta_2, Z_2^*(t_2) \ge \theta_2.
\end{aligned}$$

After t_2, under the proposed Low-period-policy, if $Z_2^*(t_2) > \theta_2$, then class 1 fluid is going to be kept at its threshold by setting $u_1^*(t) = \rho_1^l$, and class 2 fluid is going to decrease by holding service capacity at $u_2^*(t) = 1 - \rho_1^l > \rho_2$ until class 2 fluid reaches its threshold from above at $\tilde{t}_2$ (see the definition of $\tilde{t}_2$ given in (50)). After $\tilde{t}_2$, $u_1^*(t) > \rho_1^l$ and $u_2^*(t) > \rho_2$. So, fluid levels of both classes are going to decrease and are maintained below their thresholds. Hence,

$$\begin{aligned}
\forall t \in (t_2,\, \tilde{t}_2), &\qquad Z_2^*(t) > \theta_2,\ Z_1^*(t) = \theta_1,\\
\forall t \in (\tilde{t}_2,\, H + L), &\qquad Z_2^*(t) \le \theta_2,\ Z_1^*(t) \le \theta_1.
\end{aligned}$$

This completes the proofs of Claims 1 and 2.

• **Proof for Claim 3 in Case 2.1.** From the proofs of Claims 1 and 2, we know that in this case $\tilde{s}_2 = s_2$.

From the construction of $p_i(t)$, $i = 1, 2$, we know that they are piecewise linear functions. To compare their values, it is sufficient to compare them at the end points of each interval. Since $p_i(H + L) = 0$ and $\dot{p}_i(t) \ge 0$ at all differentiable points, we know $p_i(t) \le 0$, $i = 1, 2$, for all $t \in [0, H+L]$. Note that since $p_1(H+L) = p_2(H+L) = 0$ and $\mu_1\dot{p}_1(t) = \mu_2\dot{p}_2(t)$ for $t \in (t_2, H+L)$, we have $\mu_1 p_1(t) = \mu_2 p_2(t)$ for $t \in [t_2,\, H + L]$. Based on the derivatives, we then have

$$\forall t \in [t_1,\, t_2], \qquad \mu_i p_i(t) = \mu_i p_i(t_2) + \mu_i h_i (t - t_2),\ \ i = 1, 2.$$

Using the fact that $\mu_1 h_1 > \mu_2 h_2$, $\mu_1 p_1(t_2) = \mu_2 p_2(t_2)$ and noting $t - t_2 < 0$ for $t \in (t_1,\, t_2)$, we have

$$\forall t \in (t_1,\, t_2), \qquad \mu_2 p_2(t) > \mu_1 p_1(t).$$

Based on the derivatives of $p(t)$, we have

$$\forall t \in [s_2,\, t_1], \qquad \mu_1 p_1(t) = \mu_1 p_1(t_1),$$
$$\forall t \in [s_2,\, t_2], \qquad \mu_2 p_2(t) = \mu_2 p_2(t_2) + \mu_2 h_2 (t - t_2).$$

¿From (19) and $\mu_1 p_1(t_2) = \mu_2 p_2(t_2)$, we have $\mu_1 p_1(s_2) = \mu_2 p_2(s_2)$. Combining this with $\mu_1 p_1(t_1) \le \mu_2 p_2(t_1)$, we have

$$\forall t \in (s_2,\, t_1), \qquad \mu_1 p_1(t) \le \mu_2 p_2(t).$$

¿From $\mu_1 p_1(s_2) = \mu_2 p_2(s_2)$ and $\dot{p}_i(t) = 0$, $i = 1, 2$, for $t \in (s_1,\, s_2)$, we can immediately see that

$$\forall t \in [s_1,\, s_2], \qquad \mu_1 p_1(t) = \mu_2 p_2(t) = \mu_2 p_2(s_2).$$

For $t \in (0, s_1)$, based on the derivatives of $p(t)$, we have

$$\forall t \in [0,\, s_1], \qquad \mu_2 p_2(t) = \mu_2 p_2(s_1) + \mu_2 h_2 (t - s_1),$$
$$\forall t \in [0,\, s_1], \qquad \mu_1 p_1(t) = \mu_1 p_1(s_1).$$

Note that $\mu_i p_i(t)$ has the same value at s_1 for $i = 1, 2$ and for $t \in (0,\, s_1)$, $\dot{p}_2(t) = h_2 > 0 = \dot{p}_1(t)$, then we have

$$\forall t \in (0,\, s_1), \qquad \mu_1 p_1(t) > \mu_2 p_2(t).$$

This completes the proof of Claim 3.

B.2 Proof for the optimality of the policy in Section 3.2

We will only construct the auxiliary function $p_i(t)$, $i = 1, 2$. To complete the proof of (41), one only needs to go through the routine procedure as described in Section B.1. We define t_2, $\tilde{s}_2$ and $\tilde{t}_2$ in the same way as in (48), (49) and (50) but now they are defined under the policy given in Section 3.2. According to the definition of the break points s_i, $i = 1, 2, 3$, $\tilde{s}_2$, t_2, and $\tilde{t}_2$, we can specify the fluid level evolution for each time interval, and the derivatives of $p_i(t)$, $i = 1, 2$. In the equations given below, if the right hand side of an interval is not strictly larger than the left side of the interval, then that interval does not exist but this does not affect our definition of the derivatives of $p_i(t)$ and the fluid level description $Z_i^*(t)$ for $i = 1, 2$. We have

$$\begin{aligned}
\forall t \in (0,\, s_1): &\qquad Z_1^*(t) < \theta_1,\ Z_2^*(t) > \theta_2,\ \dot{p}_1(t) = 0,\ \dot{p}_2(t) = h_2,\\
\forall t \in (s_1,\, s_2): &\qquad Z_1^*(t) < \theta_1,\ Z_2^*(t) \le \theta_2,\ \dot{p}_1(t) = 0,\ \dot{p}_2(t) = 0,\\
\\
\forall t \in (s_2,\, \tilde{s}_2): &\qquad Z_2^*(t) < \theta_2,\ \dot{p}_2(t) = 0,\\
\forall t \in (\tilde{s}_2,\, \tilde{t}_2): &\qquad Z_2^*(t) > \theta_2,\ \dot{p}_2(t) = h_2,\\
\\
\forall t \in (s_2,\, s_3): &\qquad Z_1^*(t) > \theta_1,\ \dot{p}_1(t) = h_1,\\
\forall t \in (s_3,\, H): &\qquad Z_1^*(t) = \theta_1,\ \dot{p}_1(t) = \mu_2 \dot{p}_2(t)/\mu_1,\\
\forall t \in (H,\, t_2): &\qquad Z_1^*(t) > \theta_1,\ \dot{p}_1(t) = h_1,\\
\forall t \in (t_2,\, \tilde{t}_2): &\qquad Z_1^*(t) = \theta_1,\ \dot{p}_1(t) = \mu_2 \dot{p}_2(t)/\mu_1,\\
\\
\forall t \in (\tilde{t}_2,\, H + L): &\qquad Z_1^*(t) \le \theta_1,\ Z_2^*(t) \le \theta_2,\ \dot{p}_i(t) = 0,\ i = 1, 2,
\end{aligned}$$

and we let $p_i(H + L) = 0$, $i = 1, 2$. Thus, we can construct continuous and piecewise linear functions $p_i(t)$, $i = 1, 2$ which have the specified derivatives in each interval and satisfy (39).

B.3 Proof for the optimality of the policy in Section 3.3

As in the proof of the optimality of the policies given in Sections 3.1 and 3.2, the proof involves constructing the functions $p_i(t)$, $i = 1, 2$ based on the Pontryagin maximal principle and is omitted.

C Holding cost expressions

In this section, we provide expressions for the holding cost under various policies when the length of the high period is H and the length of the low period is L. These expressions are used extensively in Section 5. We only consider the cases given in Section 3, i.e., we assume $\rho_1^h > 1$ and $\rho_1^l + \rho_2 < 1$.

We know from Corollary 2 that for Case 1 and Case 3, specified in (20) and (22) respectively, FP1 policy is optimal. Hence, we focus only on Cases 2 and 4 given in (21) and (23) respectively. In order to see the performance of the policies considered in Section 5, we first provide the holding cost expressions under the optimal policy. These expressions serve as the lower bound for all the other policies. Then we also provide the holding cost expressions under FP1 policy for Cases 2 and 4. In addition, we also provide the holding cost expressions under FP2-FP1 policy for Case 2, and the holding cost expressions under π^{a_1} policy for Case 4. These expressions help evaluate the performance of these two policies when $\rho_2 + \rho_1^l \to 1$.

When H and L are known, the optimal policy is given in Section 3.1. In order to compute the holding cost expression under a given policy, we observe the evolution of the fluid levels of both classes under this policy. Given the fluid levels, holding cost incurred by class 1 and class 2 can be computed easily. For example, for Case 2, when H and L satisfy the conditions of Case 2.6 (in Section 3.1), i.e., $H > a_2$, $H + L > \psi_1(1 - \eta)^{-1}$, the optimal policy is to set $s_1 = s_2 = 0$ which is equivalent to the FP1 policy. We know that fluid levels of both classes will increase, and at $t_1 = \psi_1$, class 1 fluid reaches its threshold from below and starts to incur cost. Fluid levels of both classes continue to increase linearly until the beginning of the low period. In the low period, fluid level of class 1 begins to decrease and class 2 fluid continues to increase until class 1 fluid decreases to its threshold, which happens at t_2. After t_2, class 1 fluid is kept at its threshold and class 2 fluid begins to decrease and reaches its threshold at $\tilde{t}_2$. We know that after $\tilde{t}_2$, both classes will be kept below their thresholds. Note that when $H > a_2$, under the optimal policy, $L \geq \tilde{t}_2 - H$ is equivalent to $L \geq \gamma_4(H - a_1)$ (which is OPT:1 below) and $H \geq a_2$, $L \geq \gamma_4(H - a_1)$ imply that the conditions of Case 2.6, i.e., $H \geq a_2$, $H + L > \psi_1(1 - \eta)^{-1}$ are satisfied. So, we can compute the holding cost when $H \geq a_2$ and $L \geq \gamma_4(H - a_1)$. We obtain the holding cost expressions for the other cases in a similar way.

C.1 Cost under the optimal policy

While computing the holding cost under the optimal policy, we combine Cases 2 and 4 whenever $\psi_2^- = 0$ (in Case 4), where $a^- = \max\{-a, 0\}$. However, we have to divide each case into several subcases in order to obtain closed form expressions for the holding cost. As a result, we have 17 subcases labeled (OPT:1) to (OPT:17). Recall that t_1 is the time that class 1 increases to its threshold from below in the high period, and t_2 is the time that class 1 decreases to its threshold from above in the low period if the low period is long enough, and $\tilde{t}_2$ is the time that class 2 decreases to its threshold from above if the low period is long enough. Also, recall that $\tilde{\psi}_1 \geq \tilde{\psi}_2$ is equivalent to $B \geq a_1 \geq \tilde{\psi}_1 \geq \tilde{\psi}_2$.

1. Assume that the conditions of Case 2.6 (or Case 4.4) are satisfied and $L \geq \tilde{t}_2$. In Case 2.6 (and Case 4.4), the optimal policy sets $s_1 = s_2 = 0$, i.e. implements the FP1 policy. If $L \geq \tilde{t}_2$

is also satisfied, then the low period is long enough so that the fluid levels of both classes reach their thresholds. This is equivalent to

$$\text{(OPT:1)} \qquad H \geq a_2,\ L \geq \gamma_4(H - a_1),$$

where γ_4 is given in the proof of Proposition 6 in Section 5, and the holding cost is

$$\begin{aligned} c(H,L) &= \frac{1}{2}h_2\mu_2\Big\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(H-\psi_1)^2 - \frac{(\rho_1^h-1)}{\eta}(\psi_1^-)^2 - \rho_2(\psi_2^-)^2 \\ &\quad + \frac{(1-\rho_2-\rho_1^l)^2}{\rho_2}\Big[\frac{\rho_1^h+\rho_2-1}{1-\rho_1^l-\rho_2}(H-a_1) - \frac{(\rho_1^h-1)}{(1-\rho_1^l)}(H-\psi_1)\Big]^2 \\ &\quad + (1-\rho_1^l-\rho_2)\Big[\frac{\rho_1^h+\rho_2-1}{1-\rho_1^l-\rho_2}(H-a_1) - \frac{(\rho_1^h-1)}{(1-\rho_1^l)}(H-\psi_1)\Big]^2\Big\}. \end{aligned}$$

2. Assume that the conditions of Case 2.6 (or Case 4.4) are satisfied and $t_2 \leq L \leq \tilde{t}_2$. As mentioned above, in Case 2.6 (and Case 4.4), the optimal policy implements the FP1 policy. If $t_2 \leq L \leq \tilde{t}_2$, then the low period is long enough such that class 1 fluid level reaches its threshold, but class 2 fluid is still above its threshold when the low period is over. This is equivalent to

$$\text{(OPT:2)} \qquad H \geq a_2,\ \gamma_3(H-\psi_1) \leq L \leq \gamma_4(H-a_1),$$

and the holding cost is

$$\begin{aligned} c(H,L) &= \frac{1}{2}h_2\mu_2\Big\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(H-\psi_1)^2 - \frac{(\rho_1^h-1)}{\eta}(\psi_1^-)^2 - \rho_2(\psi_2^-)^2 \\ &\quad + \frac{(1-\rho_2-\rho_1^l)^2}{\rho_2}\Big[\frac{\rho_1^h+\rho_2-1}{1-\rho_1^l-\rho_2}(H-a_1) - \frac{(\rho_1^h-1)}{(1-\rho_1^l)}(H-\psi_1)\Big]^2 \\ &\quad + (1-\rho_1^l-\rho_2)\Big[\frac{\rho_1^h+\rho_2-1}{1-\rho_1^l-\rho_2}(H-a_1) - \frac{(\rho_1^h-1)}{(1-\rho_1^l)}(H-\psi_1)\Big]^2\Big\} \\ &\quad - \frac{1}{2}h_2\mu_2(1-\rho_2-\rho_1^l)\Big[\frac{(\rho_1^h+\rho_2-1)}{(1-\rho_2-\rho_1^l)}(H-a_1) - L\Big]^2. \end{aligned}$$

3. Assume that the conditions of Case 2.6 (or Case 4.4) are satisfied and $L \leq t_2$. The optimal policy sets $s_1 = s_2 = 0$, i.e. implements the FP1 policy. Since $L \leq t_2$, at the end of the low period, both classes will be above their thresholds. This is equivalent to

$$\text{(OPT:3)} \qquad H \geq a_2,\ L \leq \gamma_3(H-\psi_1),\ H+L \geq \psi_1^+ + \frac{\eta}{1-\eta}(\psi_1^+ - \psi_2^+),$$

and the optimal cost is

$$\begin{aligned} c(H,L) &= \frac{1}{2}h_2\mu_2\Big\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(H-\psi_1)^2 + \rho_2(H+L-\psi_2)^2 \\ &\quad - \frac{1-\rho_1^l}{\eta}\Big[\frac{\rho_1^h-1}{1-\rho_1^l}(H-\psi_1) - L\Big]^2 - \frac{(\rho_1^h-1)}{\eta}(\psi_1^-)^2 - \rho_2(\psi_2^-)^2\Big\}. \end{aligned}$$

4. Assume that the conditions of Case 2.3 are satisfied and $L \geq \tilde{t}_2 - H$. In Case 2.3 optimal policy sets $s_1 = s_2$. Let $s_1 = s_2 = s$. Note that $L \geq \tilde{t}_2 - H$ is equivalent to $L > \gamma_4(H - a_1)$, which means that the low period is long enough so that fluid levels of both classes reach their thresholds. Thus, if

$$\text{(OPT:4)} \qquad \max(\tilde{\psi}_1, B) \leq H \leq a_2,\ L \geq \gamma_4(H - a_1),$$

then the holding cost is

$$\begin{aligned} c(H,L) &= \frac{1}{2}h_2\mu_2 \Big\{ \frac{(1-\rho_1^l)(\rho_1^h - \rho_1^l)}{\eta(\rho_1^h - 1)}(t_2 - H)^2 + \rho_2(t_2 - s)^2 \\ &\quad + (1-\rho_2)(2\tilde{\psi}_2 - s)s + 2(1-\rho_2)(\tilde{\psi}_2 - s)(t_2 - s) \\ &\quad + (1 - \rho_2 - \rho_1^l)\Big[\frac{\rho_1^h + \rho_2 - 1}{1 - \rho_2 - \rho_1^l}(H - a_1) - (t_2 - H)\Big]^2 \Big\}, \end{aligned}$$

where

$$\begin{aligned} s &= \frac{d_1/\mu_1 - (\rho_1^h - 1)(1-\eta)t_2}{1 + \eta(\rho_1^h - 1)}, \\ t_1 &= \frac{(1-\eta)t_2 + \eta d_1/\mu_1}{1 + \eta(\rho_1^h - 1)}, \\ t_2 &= \frac{(\rho_1^h - \rho_1^l)H - \eta(\rho_1^h - 1)d_1(\mu_1(1 + \eta(\rho_1^h - 1)))^{-1}}{(1 - \rho_1^l) + (1-\eta)(\rho_1^h - 1)(1 + \eta(\rho_1^h - 1))^{-1}}, \end{aligned}$$

5. Assume that the conditions of Case 2.3 are satisfied and $L \leq \tilde{t}_2 - H$. When H and L satisfy the conditions of Case 2.3, it implies that $L \geq t_2$, i.e., the low period is long enough so that class 1 reaches its threshold. However, since $L \leq \tilde{t}_2 - H$, the low period is not long enough for class 2 to reach its threshold. At the end of the low period, class 2 fluid is still above its threshold, but class 1 is below its threshold. Hence, if

$$\text{(OPT:5)} \qquad \max(\tilde{\psi}_1, B) \leq H \leq a_2,\ \gamma_2(H - \tilde{\psi}_1) \leq L \leq \gamma_4(H - a_1)$$

then the holding cost is

$$\begin{aligned} c(H,L) &= \frac{1}{2}h_2\mu_2 \Big\{ \frac{(1-\rho_1^l)(\rho_1^h - \rho_1^l)}{\eta(\rho_1^h - 1)}(t_2 - H)^2 + \rho_2(t_2 - s)^2 \\ &\quad + (1-\rho_2)(2\tilde{\psi}_2 - s)s + 2(1-\rho_2)(\tilde{\psi}_2 - s)(t_2 - s) \\ &\quad + (1 - \rho_2 - \rho_1^l)\Big[\frac{\rho_1^h + \rho_2 - 1}{1 - \rho_2 - \rho_1^l}(H - a_1) - (t_2 - H)\Big]^2 \Big\} \\ &\quad - \frac{1}{2}h_2\mu_2(1 - \rho_2 - \rho_1^l)\Big[\frac{(\rho_1^h + \rho_2 - 1)}{(1 - \rho_2 - \rho_1^l)}(H - a_1) - L\Big]^2, \end{aligned}$$

where t_2, t_1, s are the same as in the previous case.

6. Assume that the conditions of Case 2.4 are satisfied. In Case 2.4, the optimal policy sets $s_1 = s_2 = s$ and $t_2 = H + L$. In this case, at the end of the low period the fluid levels of both classes are above their thresholds. Thus, if

$$\text{(OPT:6)} \qquad L \leq \gamma_2(H - \tilde{\psi}_1),\ \max\{\tilde{\psi}_1,\ \tilde{\psi}_1 + \frac{1 + \eta(\rho_1^h - 1)}{(1-\eta)(\rho_1^h - 1)}(\tilde{\psi}_1 - \tilde{\psi}_2)\} \leq H + L \leq \frac{\psi_1}{1 - \eta},$$

then the holding cost is

$$\begin{aligned} c(H,L) &= \frac{1}{2}h_2\mu_2\left\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(H-t_1)^2+(1-\rho_2)(2\tilde{\psi}_2-s)s\right. \\ &\quad +2(1-\rho_2)(\tilde{\psi}_2-s)(H+L-s)+\rho_2(H+L-s)^2 \\ &\quad \left.-\frac{(1-\rho_1^l)}{\eta}\Big[\frac{(\rho_1^h-1)}{(1-\rho_1^l)}(H-t_1)-L\Big]^2\right\}, \end{aligned}$$

where

$$\begin{aligned} s &= \frac{d_1/\mu_1-(1-\eta)(\rho_1^h-1)(H+L)}{1+\eta(\rho_1^h-1)}, \\ t_1 &= \frac{\eta d_1/\mu_1+(1-\eta)(H+L)}{1+\eta(\rho_1^h-1)}. \end{aligned}$$

7. Assume that the conditions of Case 2.5 are satisfied, $\tilde{\psi}_1 \le \tilde{\psi}_2$, class 2 decreases to its threshold before class 1 increases to its threshold, and the low period is long enough to decrease class 2 fluid to its threshold. When conditions of Case 2.5 are satisfied, and $\tilde{\psi}_1 \le \tilde{\psi}_2$, we have $H \le \tilde{\psi}_1 \le \tilde{\psi}_2$ and the optimal policy sets $s_1 = s_2 = H$. In the high period, class 2 has higher priority and in the low period, class 2 has higher priority until class 1 fluid increases to its threshold or class 2 fluid decreases to its threshold. Under this policy, let t_1' be the time that class 1 fluid increases to its threshold in the low period if the low period is long enough. Then, $t_1' = H + \rho_1^h/\rho_1^l(\tilde{\psi}_1 - H)$. If $\tilde{\psi}_2 \le t_1'$, then class 2 fluid decreases to its threshold in the low period, at $\tilde{\psi}_2$, before class 1 fluid increases to its threshold. In this case, after $\tilde{\psi}_2$, no class will incur cost under the Low-period-policy. Hence, if

$$\text{(OPT:7)} \qquad H \le \tilde{\psi}_1 \le \tilde{\psi}_2 \le \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H) + H, \quad H + L > \tilde{\psi}_2,$$

then the holding cost is

$$c(H,L) = \frac{1}{2}h_2\mu_2(1-\rho_2)\tilde{\psi}_2^2.$$

8. Assume that the conditions of Case 2.5 are satisfied, $\tilde{\psi}_1 \le \tilde{\psi}_2$, class 2 fluid decreases to its threshold before class 1 fluid increases to its threshold, but the low period is not long enough for class 2 fluid to reach its threshold. Hence, $H + L \le \tilde{\psi}_2$. If

$$\text{(OPT:8)} \qquad H \le \tilde{\psi}_1 \le \tilde{\psi}_2 \le \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H) + H, \quad H + L \le \tilde{\psi}_2,$$

then the holding cost is

$$c(H,L) = \frac{1}{2}h_2\mu_2(1-\rho_2)(2\tilde{\psi}_2 - H - L)(H+L).$$

9. Assume that the conditions of Case 2.5 are satisfied, $\tilde{\psi}_1 \le \tilde{\psi}_2$, class 1 fluid level increases to its threshold before class 2 fluid level decreases to its threshold, and the low period is long enough for class 2 fluid to reach its threshold. Since the conditions of Case 2.5 are satisfied

and $\tilde{\psi}_1 \le \tilde{\psi}_2$, we have $H \le \tilde{\psi}_1 \le \tilde{\psi}_2$. Following the optimal policy, we set $s_1 = s_2 = H$. Class 2 has higher priority in the high period and also in the low period before class 1 fluid reaches its threshold at t_1'. So, if $\tilde{\psi}_2 \ge t_1'$, it means that class 2 is still above its threshold when class 1 increases to its threshold in the low period. Based on the Low-period-policy after t_1', server will spend just enough effort ($u_1 = \rho_1^l$) to keep class 1 at its threshold, and use the remaining effort ($u_2 = 1 - \rho_1^l > \rho_2$) to serve class 2. Let $\tilde{t}_2$ be the time that class 2 fluid level decreases to its threshold, then $L + H \ge \tilde{t}_2$, which is equivalent to $L \ge \gamma_4(H - a_1)$. So, if

$$\text{(OPT:9)} \qquad H \le \tilde{\psi}_1 \le \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H) + H \le \tilde{\psi}_2, \quad L \ge \gamma_4(H - a_1),$$

then the holding cost is

$$\begin{aligned} c(H,L) &= \frac{1}{2} h_2\mu_2 \Big\{ (1-\rho_2)\Big[2\tilde{\psi}_2 - \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H) - H\Big]\Big[\frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H) + H\Big] \\ &\quad + (1-\rho_2-\rho_1^l)\Big[\frac{\rho_1^h + \rho_2 - 1}{1-\rho_2-\rho_1^l}(H-a_1) - \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H)\Big]^2 \Big\}. \end{aligned}$$

10. Assume that the conditions of Case 2.5 are satisfied, $\tilde{\psi}_1 \le \tilde{\psi}_2$, class 1 fluid level increases to its threshold before class 2 fluid level decreases to its threshold, and the low period is not long enough for class 2 fluid to reach its threshold, but long enough for class 1 fluid to reach its threshold. Hence, $H + L \ge t_1'$, which is equivalent to $L \ge \rho_1^h/\rho_1^l(\tilde{\psi}_1 - H)$. If

$$\text{(OPT:10)} \qquad H \le \tilde{\psi}_1 \le \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H) + H \le \tilde{\psi}_2, \quad \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H) \le L \le \gamma_4(H - a_1),$$

then the holding cost is

$$\begin{aligned} c(H,L) &= \frac{1}{2} h_2\mu_2 \Big\{ (1-\rho_2)\Big[2\tilde{\psi}_2 - \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H) - H\Big]\Big[\frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H) + H\Big] \\ &\quad + (1-\rho_2-\rho_1^l)\Big[\frac{\rho_1^h + \rho_2 - 1}{1-\rho_2-\rho_1^l}(H-a_1) - \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H)\Big]^2 \Big\} \\ &\quad - \frac{1}{2} h_2\mu_2(1-\rho_2-\rho_1^l)\Big[\frac{\rho_1^h + \rho_2 - 1}{1-\rho_2-\rho_1^l}(H-a_1) - L\Big]^2. \end{aligned}$$

11. Assume that the conditions of Case 2.5 are satisfied, $\tilde{\psi}_1 \le \tilde{\psi}_2$, the low period is neither long enough for class 1 fluid to increase to its threshold nor long enough for class 2 fluid to decrease to its threshold. However, if the low period were long enough class 1 fluid would increase to its threshold before class 2 would decrease to its threshold. Hence, $H + L \le t_1'$. If

$$\text{(OPT:11)} \qquad H \le \tilde{\psi}_1 \le \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H) + H \le \tilde{\psi}_2, \quad L \le \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H),$$

then the holding cost is

$$c(H,L) = \frac{1}{2} h_2\mu_2(1-\rho_2)(2\tilde{\psi}_2 - H - L)(H + L).$$

12. Assume that the conditions of Case 2.1 (or Case 4.1) are satisfied and the low period is long enough for class 2 fluid to decrease to its threshold. Recall that when conditions of Case 2.1 (or Case 4.1) are satisfied, we denote the time that class 2 decreases to its threshold as $\tilde{\psi}_2$. Moreover, $H + L \geq \tilde{\psi}_2$ is equivalent to $L > \gamma_4(H - a_1)$ which implies that $L \geq \gamma_1(H - a_1)$. In this case, at the end of the low period, fluid levels of both classes will be below their thresholds. Notice that $a_1 \leq B$ implies that $\tilde{\psi}_2 < \tilde{\psi}_1$. Hence, if

$$\text{(OPT:12)} \qquad \begin{array}{ll} a_1 \leq H \leq B, \quad L \geq \gamma_4(H - a_1), \text{ for Case 2} \\ a_1 \leq H \leq a_2, \quad L \geq \gamma_4(H - a_1), \text{ for Case 4} \end{array}$$

then the holding cost is

$$\begin{aligned} c(H, L) &= \frac{1}{2} h_2 \mu_2 \left\{ \frac{(1 - \rho_1^l)(\rho_1^h - \rho_1^l)}{\eta(\rho_1^h - 1)} (t_2 - H)^2 + (1 - \rho_2) s_1^2 + \rho_2 (t_2 - s_2)^2 \right. \\ &\qquad \left. + (1 - \rho_2 - \rho_1^l) \Big[\frac{\rho_1^h + \rho_2 - 1}{1 - \rho_2 - \rho_1^l} (H - a_1) - (t_2 - H) \Big]^2 \right\}, \end{aligned}$$

where

$$\begin{aligned} s_1 &= \tilde{\psi}_2^+, \\ s_2 &= \frac{(d_1/\mu_1 + d_2/\mu_2) - (\rho_1^h - 1)(1 - \eta) t_2}{\rho_2 + \eta(\rho_1^h - 1)}, \\ t_1 &= \frac{\eta(d_1/\mu_1 + d_2/\mu_2) + \rho_2(1 - \eta) t_2}{\rho_2 + \eta(\rho_1^h - 1)}, \\ t_2 &= \frac{(\rho_1^h - \rho_1^l) H - (\rho_1^h - 1)\eta(d_1/\mu_1 + d_2/\mu_2)(\rho_2 + \eta(\rho_1^h - 1))^{-1}}{(1 - \rho_1^l) + (\rho_1^h - 1)(1 - \eta)\rho_2(\rho_2 + \eta(\rho_1^h - 1))^{-1}}. \end{aligned}$$

13. Assume that the conditions of Case 2.1 (or Case 4.1) are satisfied but the low period is not long enough for class 2 fluid to decrease to its threshold. Hence, $L \leq \gamma_4(H - a_1)$. At the end of the low period, class 2 is still above its threshold but class 1 is below its threshold. Hence, if

$$\text{(OPT:13)} \qquad \begin{array}{ll} a_1 \leq H \leq B, \quad \gamma_1(H - a_1) \leq L \leq \gamma_4(H - a_1), \text{ for Case 2} \\ a_1 \leq H \leq a_2, \quad \gamma_1(H - a_1) \leq L \leq \gamma_4(H - a_1), \text{ for Case 4} \end{array}$$

then

$$\begin{aligned} c(H, L) &= \frac{1}{2} h_2 \mu_2 \left\{ \frac{(1 - \rho_1^l)(\rho_1^h - \rho_1^l)}{\eta(\rho_1^h - 1)} (t_2 - H)^2 + (1 - \rho_2) s_1^2 + \rho_2 (t_2 - s_2)^2 \right. \\ &\qquad \left. + (1 - \rho_2 - \rho_1^l) \Big[\frac{\rho_1^h + \rho_2 - 1}{1 - \rho_2 - \rho_1^l} (H - a_1) - (t_2 - H) \Big]^2 \right\} \\ &\quad - \frac{1}{2} h_2 \mu_2 (1 - \rho_2 - \rho_1^l) \Big[\frac{\rho_1^h + \rho_2 - 1}{1 - \rho_2 - \rho_1^l} (H - a_1) - L \Big]^2 \end{aligned}$$

where s_1, s_2, t_1, t_2 are the same as given in Case 2.1 (Case 4.1).

14. Assume that conditions of Case 2.2 (or Case 4.2) are satisfied. Note that

$$\tilde{\psi}_1 + (1+\eta(\rho_1^h - 1))((1-\eta)(\tilde{\psi}_1 - \tilde{\psi}_2^+)(\rho_1^h - 1))^{-1} \geq a_1$$

implies that $\tilde{\psi}_1 \geq \tilde{\psi}_2$. Since the low period is not long enough for class 1 fluid to decrease to its threshold, $t_2 = H + L$. With some algebra we have

$$\text{(OPT:14)} \qquad a_1 \leq H, \quad L \leq \gamma_1(H - a_1), \quad H + L \leq \tilde{\psi}_1 + \frac{1+\eta(\rho_1^h - 1)}{(1-\eta)(\rho_1^h - 1)}(\tilde{\psi}_1 - \tilde{\psi}_2^+) - \frac{\eta}{1-\eta}\psi^+,$$

and the holding cost is

$$\begin{aligned} c(H,L) &= \Bigg\{ \frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{\eta(1-\rho_1^l)}(H - t_1)^2 + (1-\rho_2)s_1^2 \\ &\quad + \rho_2(H + L - s_2)^2 - \frac{(1-\rho_1^l)}{\eta}\Big[\frac{\rho_1^h - 1}{1-\rho_1^l}(H - t_1) - L\Big]^2 \Bigg\} \end{aligned}$$

where

$$\begin{aligned} s_1 &= \tilde{\psi}_2^+, \\ s_2 &= \frac{(d_1/\mu_1 + d_2/\mu_2) - (\rho_1^h - 1)(1-\eta)t_2}{\rho_2 + \eta(\rho_1^h - 1)}, \\ t_1 &= \frac{\eta(d_1/\mu_1 + d_2/\mu_2) + \rho_2(1-\eta)t_2}{\rho_2 + \eta(\rho_1^h - 1)}, \\ t_2 &= H + L. \end{aligned}$$

15. Assume that conditions of Case 2.5 are satisfied, $\tilde{\psi}_2 \leq \tilde{\psi}_1$ and class 2 reaches its threshold from below in the high period or conditions of Case 4.3 are satisfied. Recall that $\tilde{\psi}_2 \leq \tilde{\psi}_1$ implies that $\tilde{\psi}_2 \leq \tilde{\psi}_1 \leq a_1$. Since conditions of Case 2.5 and $\tilde{\psi}_2 \leq \tilde{\psi}_1$ are satisfied, we have $H \leq a_1$. According to the optimal policy, class 2 has higher priority in the high period as long as its fluid level is above its threshold, and class 2 fluid reaches its threshold at $\tilde{\psi}_2$. After $\tilde{\psi}_2$, server will allocate enough capacity to keep class 2 fluid level below its threshold and the remaining capacity will be allocated to class 1. In this case, class 1 fluid will never reach its threshold in the high period. Similarly, for Case 4.3, under the optimal policy class 1 and class 2 fluids will stay below their thresholds in the high period. Hence, if

$$\text{(OPT:15)} \qquad \tilde{\psi}_2 \leq \tilde{\psi}_1, \quad \tilde{\psi}_2 \leq H \leq a_1,$$

then the holding cost is

$$c(H,L) = \frac{1}{2}h_2\mu_2(1-\rho_2)s_1^2.$$

16. Assume that conditions of Case 2.5 are satisfied, $\tilde{\psi}_2 \leq \tilde{\psi}_1$, and class 2 fluid does not reach its threshold in the high period, but it reaches its threshold in the low period. If

$$\text{(OPT:16)} \qquad \tilde{\psi}_2 \leq \tilde{\psi}_1, \quad H \leq \tilde{\psi}_2 \leq H + L,$$

then the holding cost is

$$c(H,L) = \frac{1}{2}h_2\mu_2(1-\rho_2)\tilde{\psi}_2^2.$$

17. Assume that conditions of Case 2.5 are satisfied $\tilde{\psi}_2 \le \tilde{\psi}_1$, and class 2 fluid does not reach its threshold in the low period. Hence, if

$$\text{(OPT:17)} \qquad \tilde{\psi}_2 \le \tilde{\psi}_1, \quad H + L \le \tilde{\psi}_2, \tag{52}$$

then the holding cost is

$$c(H, L) \;=\; \frac{1}{2} h_2 \mu_2 (2\tilde{\psi}_2 - H - L)(H + L).$$

C.2 Cost under the FP1 policy when $\psi_2 < \psi_1$ (Case 2 and Case 4)

Note that when $\psi_2 < \psi_1$, under the FP1 policy, class 2 fluid increases to its threshold before class 1 fluid increases to its threshold in the high period if the high period is long enough, i.e. if $H \ge \psi_2$. In order to compute the holding cost under the FP1 policy, we consider 9 different cases labeled (FP1:1) to (FP1:9).

1. Assume that class 1 fluid increases to its threshold in the high period and decreases to its threshold in the low period and class 2 fluid also decreases to its threshold in the low period. Hence, if

$$\text{(FP1:1)} \qquad H \ge \psi_1, \quad L \ge \gamma_4 (H - a_1)$$

then the holding cost under FP1 policy is

$$\begin{aligned} c^{FP1}(H, L) \;=\; & \frac{1}{2} h_2 \mu_2 \Big\{ \frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{\eta(1 - \rho_1^l)} (H - \psi_1)^2 - \frac{(\rho_1^h - 1)}{\eta} (\psi_1^-)^2 - \rho_2 (\psi_2^-)^2 \\ & + \frac{(1 - \rho_2 - \rho_1^l)^2}{\rho_2} \Big[\frac{\rho_1^h + \rho_2 - 1}{1 - \rho_1^l - \rho_2} (H - a_1) - \frac{(\rho_1^h - 1)}{(1 - \rho_1^l)} (H - \psi_1) \Big]^2 \\ & + (1 - \rho_1^l - \rho_2) \Big[\frac{\rho_1^h + \rho_2 - 1}{1 - \rho_1^l - \rho_2} (H - a_1) - \frac{(\rho_1^h - 1)}{(1 - \rho_1^l)} (H - \psi_1) \Big]^2 \Big\}. \end{aligned}$$

2. Assume that class 1 fluid increases to its threshold in the high period and decreases to its threshold in the low period but class 2 fluid does not decrease to its threshold at the end of the low period. Hence, if

$$\text{(FP1:2)} \qquad H \ge \psi_1, \qquad \gamma_3 (H - \psi_1) \le L \le \gamma_4 (H - a_1)$$

the holding cost under FP1 policy is

$$\begin{aligned} c^{FP1}(H, L) \;=\; & \frac{1}{2} h_2 \mu_2 \Big\{ \frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{\eta(1 - \rho_1^l)} (H - \psi_1)^2 - \frac{(\rho_1^h - 1)}{\eta} (\psi_1^-)^2 - \rho_2 (\psi_2^-)^2 \\ & + \frac{(1 - \rho_2 - \rho_1^l)^2}{\rho_2} \Big[\frac{\rho_1^h + \rho_2 - 1}{1 - \rho_1^l - \rho_2} (H - a_1) - \frac{(\rho_1^h - 1)}{(1 - \rho_1^l)} (H - \psi_1) \Big]^2 \\ & + (1 - \rho_1^l - \rho_2) \Big[\frac{\rho_1^h + \rho_2 - 1}{1 - \rho_1^l - \rho_2} (H - a_1) - \frac{(\rho_1^h - 1)}{(1 - \rho_1^l)} (H - \psi_1) \Big]^2 \Big\} \\ & - \frac{1}{2} h_2 \mu_2 (1 - \rho_2 - \rho_1^l) \Big[\frac{(\rho_1^h + \rho_2 - 1)}{(1 - \rho_2 - \rho_1^l)} (H - a_1) - L \Big]^2. \end{aligned}$$

3. Assume that class 1 fluid increases to its threshold in the high period but does not decrease to its threshold in the low period. Hence, if

$$\text{(FP1:3)} \qquad H \geq \psi_1, \qquad L \leq \gamma_3(H - \psi_1), \tag{53}$$

then the holding cost under FP1 policy is

$$\begin{aligned} c^{FP1}(H,L) &= \frac{1}{2}h_2\mu_2 \left\{ \frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{\eta(1-\rho_1^l)}(H-\psi_1)^2 + \rho_2(H+L-\psi_2)^2 \right. \\ &\quad \left. - \frac{(\rho_1^h - 1)}{\eta}(\psi_1^-)^2 - \rho_2(\psi_2^-)^2 - \frac{1-\rho_1^l}{\eta}\left[\frac{\rho_1^h - 1}{1-\rho_1^l}(H-\psi_1) - L\right]^2 \right\}. \end{aligned}$$

4. Assume that class 1 fluid does not increase to its threshold in the high period, but class 2 fluid increases to its threshold in the high period. In the low period, before class 2 fluid decreases to its threshold, class 1 fluid increases to its threshold, and the server allocates enough capacity to maintain class 1 fluid at its threshold until class 2 fluid decreases to its threshold. The low period is long enough for class 2 fluid to decrease to its threshold. At the end of the low period, fluid levels of both classes are below their thresholds. Hence, if

$$\text{(FP1:4)} \qquad \hat{\psi} \leq H \leq \psi_1, \qquad L \leq \gamma_4(H - a_1)$$

where

$$\hat{\psi} = \frac{(\rho_1^h - 1)(1-\rho_2)}{\rho_2\rho_1^l + (\rho_1^h - 1)(1-\rho_2)}\psi_1 + \frac{\rho_2\rho_1^l}{\rho_2\rho_1^l + (\rho_1^h - 1)(1-\rho_2)}\psi_2,$$

then the holding cost under FP1 policy is

$$\begin{aligned} c^{FP1}(H,L) &= \frac{1}{2}h_2\mu_2 \left\{ -\rho_2(\psi_2^-)^2 + \frac{\rho_2(1-\rho_1^l)}{1-\rho_1^l-\rho_2}(H-\psi_2)^2 \right. \\ &\quad + \frac{2\rho_2(\rho_1^h - 1)}{1-\rho_2-\rho_1^l}(H-\psi_1)(H-\psi_2) \\ &\quad \left. + \frac{(1-\rho_2)(\rho_1^h - 1)^2}{\rho_1^l(1-\rho_2-\rho_1^l)}(H-\psi_1)^2 \right\}. \end{aligned}$$

5. Assume that class 1 fluid does not increase to its threshold in the high period, but class 2 fluid increases to its threshold in the high period. In the low period, before class 2 fluid decreases to its threshold, class 1 fluid increases to its threshold, and the server allocates just enough capacity to maintain class 1 fluid at its threshold until class 2 fluid decreases to its threshold. The low period is not long enough for class 2 fluid to decrease to its threshold. At the end of the low period, class 1 fluid is below its threshold and class 2 fluid is still above its threshold. Hence, if

$$\text{(FP1:5)} \qquad \hat{\psi} \leq H \leq \psi_1, \qquad -\gamma_5(\psi_1 - H) \leq L \leq \gamma_4(H - a_1)$$

where $\gamma_5 = -(\rho_1^h - 1)(\rho_1^l)^{-1}$ then the holding cost under FP1 policy is

$$c^{FP1}(H,L) = \frac{1}{2}h_2\mu_2 \left\{ -\rho_2(\psi_2^-)^2 + \frac{\rho_2(1-\rho_1^l)}{1-\rho_1^l-\rho_2}(H-\psi_2)^2 \right.$$

$$
\begin{aligned}
&+\frac{2\rho_2(\rho_1^h-1)}{1-\rho_2-\rho_1^l}(H-\psi_1)(H-\psi_2)\\
&+\frac{(1-\rho_2)(\rho_1^h-1)^2}{\rho_1^l(1-\rho_2-\rho_1^l)}(H-\psi_1)^2\Big\}\\
-&\frac{1}{2}h_2\mu_2(1-\rho_2-\rho_1^l)\Big[\frac{\rho_1^h+\rho_2-1}{1-\rho_2-\rho_1^l}(H-a_1)-L\Big]^2.
\end{aligned}
$$

6. Assume that class 1 does not increase to its threshold in the high period, but class 2 increases to its threshold in the high period. If the low period were long enough, class 1 fluid would increase to its threshold before class 2 fluid decreases to its threshold. However, the low period is not long enough and class 1 fluid is still below its threshold and class 2 is above its threshold at the end of the low period. Hence, if

$$
\text{(FP1:6)} \qquad \hat{\psi} \le H \le \psi_1, \qquad L \le -\gamma_5(\psi_1 - H)
$$

then the holding cost under FP1 policy is

$$
\begin{aligned}
c^{FP1}(H,L) &= \frac{1}{2}h_2\mu_2\left\{\frac{\rho_2}{1-\rho_2}(H-\psi_2)^2-\rho_2(\psi_2^-)^2\right\}\\
&-\frac{1}{2}h_2\mu_2(1-\rho_2)\Big[\frac{\rho_2}{1-\rho_2}(H-\psi_2)-L\Big]^2.
\end{aligned}
$$

7. Assume that class 1 fluid does not increase to its threshold in the high period, but class 2 fluid increases to its threshold in the high period. In the low period, class 2 fluid decreases to its threshold before class 1 fluid increases to its threshold. The low period is long enough such that at the end of the low period, both class 1 and class 2 fluids are below their thresholds. Hence, if

$$
\text{(FP1:7)} \qquad \psi_2 \le H \le \hat{\psi}, \qquad L \ge \gamma_6(H-\psi_2),
$$

where $\gamma_6 = \rho_2(1-\rho_2)^{-1}$ then the holding cost under FP1 policy is

$$
c^{FP1}(H,L) = \frac{1}{2}h_2\mu_2\Big\{\frac{\rho_2}{1-\rho_2}(H-\psi_2)^2-\rho_2(\psi_2^-)^2\Big\}.
$$

8. Assume that class 1 fluid does not increase to its threshold in the high period, but class 2 fluid increases to its threshold in the high period. If the low period were long enough, class 2 fluid would decrease to its threshold before class 1 fluid increases to its threshold. However, the low period is not long enough. So, at the end of the low period, class 2 fluid is still above its threshold and class 1 fluid is below its threshold. Hence, if

$$
\text{(FP1:8)} \qquad \psi_2 \le H \le \hat{\psi}, \qquad L \le \gamma_6(H-\psi_2),
$$

then the holding cost under FP1 policy is

$$
\begin{aligned}
c^{FP1}(H,L) &= \frac{1}{2}h_2\mu_2\left\{\frac{\rho_2}{1-\rho_2}(H-\psi_2)^2-\rho_2(\psi_2^-)^2\right\}\\
&-\frac{1}{2}h_2\mu_2(1-\rho_2)\Big[\frac{\rho_2}{1-\rho_2}(H-\psi_2)-L\Big]^2.
\end{aligned}
$$

9. Assume that neither class 1 fluid nor class 2 fluid reaches its threshold in the high period. Hence, if

$$\text{(FP1:9)} \qquad H \le \psi_2,$$

then the cost under FP1 policy is

$$c^{FP1}(H,L) \;=\; 0.$$

C.3 Cost under the π^{a_1} policy when $0 \le \psi_2 < \psi_1$ (Case 4)

Note that in this case if the high period is long enough (i.e. if $H \ge a_1$), under the π^{a_1} policy, class 1 and class 2 fluids reach their thresholds at the same time, namely, at a_1. In order to compute the holding cost under the π^{a_1} policy, we consider 4 different cases labeled (a_1:1) to (a_1:4).

1. Assume that fluid levels of both classes increase to their thresholds at the same time in the high period, and the low period is long enough to decrease fluid levels of both classes below their thresholds. Thus, at the end of the low period, both class 1 and class 2 fluids are below their thresholds. Hence, if

$$(a_1{:}1) \qquad H \ge a_1, \qquad L \ge \gamma_4(H-a_1),$$

then the holding cost under π^{a_1} policy is

$$c^{a_1}(H,L) \;=\; \frac{1}{2}h_2\mu_2\left(\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)} + \frac{\rho_2(\rho_1^h-\rho_1^l)^2}{(1-\rho_1^l)(1-\rho_1^l-\rho_2)}\right)(H-a_1)^2.$$

2. Assume that both classes increase to their thresholds at the same time in the high period but the low period is not long enough for class 2 fluid to decrease to its threshold. At the end of the low period, class 1 fluid is below its threshold but class 2 fluid is still above its threshold. Hence, if

$$(a_1{:}2) \qquad H \ge a_1, \qquad \gamma_3(H-a_1) \le L \le \gamma_4(H-a_1),$$

then the holding cost under π^{a_1} policy is

$$\begin{aligned} c^{a_1}(H,L) \;=\; & \frac{1}{2}h_2\mu_2\left(\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)} + \frac{\rho_2(\rho_1^h-\rho_1^l)^2}{(1-\rho_1^l)(1-\rho_1^l-\rho_2)}\right)(H-a_1)^2 \\ & -\frac{1}{2}h_2\mu_2(1-\rho_2-\rho_1^l)\left[\frac{\rho_1^h+\rho_2-1}{1-\rho_2-\rho_1^l}(H-a_1)-L\right]^2. \end{aligned}$$

3. Assume that fluid levels of both classes increase to their thresholds at the same time in the high period but the low period is not long enough for either class 1 or class 2 fluid to decrease to its threshold. At the end of the low period, both class 1 and class 2 fluids are above their thresholds. Hence, if

$$(a_1{:}3) \qquad H \ge a_1, \qquad L \le \gamma_3(H-a_1),$$

then the cost under π^{a_1} policy is

$$\begin{aligned} c^{a_1}(H,L) \;=\; & \frac{1}{2}h_2\mu_2\left\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(H-a_1)^2 + \rho_2(H+L-a_1)^2\right. \\ & \left. -\frac{(1-\rho_1^l)}{\eta}\left[\frac{\rho_1^h-1}{1-\rho_1^l}(H-a_1)-L\right]^2\right\}. \end{aligned}$$

4. Assume that fluid levels of both classes are still below their thresholds at the end of the high period. Hence, if

$$(a_1{:}4) \qquad H \le a_1,$$

then the holding cost under π^{a_1} policy is

$$c^{a_1}(H, L) = 0.$$

C.4 Cost under the FP2-FP1 policy when $\psi_2 < 0 < \psi_1$ (Case 2)

Case 2 has two subcases: $\tilde{\psi}_1 \le \tilde{\psi}_2$ and $\tilde{\psi}_1 \ge \tilde{\psi}_2$. Recall that $\tilde{\psi}_1$ ($\tilde{\psi}_2$) is the time that class 1 fluid increases (class 2 fluid decreases) to its threshold from below (from above) in the high period if class 2 has higher priority and if the high period is long enough. So, if $\tilde{\psi}_1 \le \tilde{\psi}_2 \le H$, class 1 fluid increases to its threshold before class 2 fluid decreases to its threshold. However, if $\tilde{\psi}_2 \le \tilde{\psi}_1 \le H$, then class 1 fluid is still below its threshold when class 2 fluid reaches its threshold in the high period.

1. Assume that in the high period class 1 fluid increases to its threshold (at $\tilde{\psi}_1$) before class 2 fluid decreases to its threshold, after $\tilde{\psi}_1$, class 1 has higher priority in the high period. Suppose that the low period is long enough to reduce fluid levels of both classes below their thresholds. Thus, at the end of the low period, fluid levels of both classes are below their thresholds. Hence, if

$$(\text{FP2-FP1:1}) \qquad \tilde{\psi}_1 \le \tilde{\psi}_2, \quad H \ge \tilde{\psi}_1, \quad L \ge \gamma_4(H - a_1),$$

then the holding cost under FP2-FP1 policy is

$$\begin{aligned} c^{\text{FP2-FP1}}(H, L) \;=\; & \frac{1}{2} h_2 \mu_2 \Bigg\{ \frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{\eta(1 - \rho_1^l)}(H - \tilde{\psi}_1)^2 + (1 - \rho_2)(2\tilde{\psi}_2 - \tilde{\psi}_1)\tilde{\psi}_1 \\ & + 2\frac{(1 - \rho_2)(\rho_1^h - \rho_1^l)}{1 - \rho_1^l}(\tilde{\psi}_2 - \tilde{\psi}_1)(H - \tilde{\psi}_1) + \rho_2 \Big[\frac{\rho_1^h - \rho_1^l}{1 - \rho_1^l}(H - \tilde{\psi}_1)\Big]^2 \\ & + (1 - \rho_2 - \rho_1^l)\Big[\frac{\rho_1^h + \rho_2 - 1}{1 - \rho_2 - \rho_1^l}(H - a_1) - \frac{\rho_1^h - 1}{1 - \rho_1^l}(H - \tilde{\psi}_1)\Big]^2 \Bigg\}. \end{aligned}$$

2. Assume that in the high period class 1 fluid increases to its threshold (at $\tilde{\psi}_1$) before class 2 fluid decreases to its threshold, after $\tilde{\psi}_1$, class 1 has higher priority in the high period. Suppose that the low period is long enough for class 1 fluid to decrease below its threshold, but not long enough for class 2 fluid to decrease to its threshold. Thus, at the end of the low period, class 1 fluid level is at its threshold but class 2 fluid is still above its threshold. Hence, if

$$(\text{FP2-FP1:2}) \qquad \tilde{\psi}_1 \le \tilde{\psi}_2, \quad H \ge \tilde{\psi}_1,\ \gamma_3(H - \tilde{\psi}_1) \le L \le \gamma_4(H - a_1),$$

then the holding cost under FP2-FP1 policy is

$$\begin{aligned} c^{\text{FP2-FP1}}(H, L) \;=\; & \frac{1}{2} h_2 \mu_2 \Bigg\{ \frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{\eta(1 - \rho_1^l)}(H - \tilde{\psi}_1)^2 + (1 - \rho_2)(2\tilde{\psi}_2 - \tilde{\psi}_1)\tilde{\psi}_1 \\ & + 2\frac{(1 - \rho_2)(\rho_1^h - \rho_1^l)}{1 - \rho_1^l}(\tilde{\psi}_2 - \tilde{\psi}_1)(H - \tilde{\psi}_1) + \rho_2 \Big[\frac{\rho_1^h - \rho_1^l}{1 - \rho_1^l}(H - \tilde{\psi}_1)\Big]^2 \end{aligned}$$

$$+(1-\rho_2-\rho_1^l)\Big[\frac{\rho_1^h+\rho_2-1}{1-\rho_2-\rho_1^l}(H-a_1)-\frac{\rho_1^h-1}{1-\rho_1^l}(H-\tilde{\psi}_1)\Big]^2\Big\}$$
$$-\frac{1}{2}h_2\mu_2(1-\rho_2-\rho_1^l)\Big[\frac{\rho_1^h+\rho_2-1}{1-\rho_2-\rho_1^l}(H-a_1)-L\Big]^2.$$

3. Assume that in the high period class 1 fluid increases to its threshold (at $\tilde{\psi}_1$) before class 2 fluid decreases to its threshold, after $\tilde{\psi}_1$, class 1 has higher priority in the high period. Suppose that low period is not long enough for class 1 or class 2 fluid to reach its threshold. Thus, at the end of the low period, both class 1 and class 2 fluid levels are above their thresholds. Hence, if

$$\text{(FP2-FP1:3)} \qquad \tilde{\psi}_1 \le \tilde{\psi}_2, \quad H \ge \tilde{\psi}_1, \quad L \le \gamma_3(H-\tilde{\psi}_1),$$

then the holding cost under FP2-FP1 policy is

$$c^{\text{FP2}-\text{FP1}}(H,L) = \frac{1}{2}h_2\mu_2\Big\{\frac{(\rho_1^h-1)(\rho_1^h-\rho_1^l)}{\eta(1-\rho_1^l)}(H-\tilde{\psi}_1)^2+(1-\rho_2)(2\tilde{\psi}_2-\tilde{\psi}_1)\tilde{\psi}_1$$
$$+\rho_2(H+L-\tilde{\psi}_1)^2+2(1-\rho_2)(\tilde{\psi}_2-\tilde{\psi}_1)(H+L-\tilde{\psi}_1)$$
$$-\frac{1-\rho_1^l}{\eta}\Big[\frac{\rho_1^h-1}{1-\rho_1^l}(H-\tilde{\psi}_1)-L\Big]^2\Big\}.$$

4. Assume that if the high period were long enough, class 1 fluid would increase to its threshold (at $\tilde{\psi}_1$) before class 2 fluid decreases to its threshold. However, the length of the high period is shorter than $\tilde{\psi}_1$. Thus, at the end of the high period, class 2 fluid is above its threshold but class 1 fluid is still below its threshold. In the low period, according to FP2-FP1 policy, class 1 fluid increases to its threshold at $\rho_1^h(\tilde{\psi}_1-H)(\rho_1^l)^{-1}+H$. Suppose that this happens before class 2 fluid decreases to its threshold. Then the server allocates enough capacity to class 1 to maintain class 1 fluid it at its threshold level and the remaining capacity is allocated to serving class 2. Moreover, assume that $L \ge \gamma_4(H-a_1)$, i.e. the low period is long enough for class 2 fluid to reach its threshold. Hence, if

$$\text{(FP2-FP1:4)} \qquad H \le \tilde{\psi}_1 \le \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1-H)+H \le \tilde{\psi}_2, \quad L \ge \gamma_4(H-a_1),$$

then the holding cost under FP2-FP1 policy is

$$c^{\text{FP2}-\text{FP1}}(H,L) = \frac{1}{2}h_2\mu_2\Big\{(1-\rho_2)\Big[2\tilde{\psi}_2-\frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1-H)-H\Big]\Big[\frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1-H)+H\Big]$$
$$+(1-\rho_2-\rho_1^l)\Big[\frac{\rho_1^h+\rho_2-1}{1-\rho_2-\rho_1^l}(H-a_1)-\frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1-H)\Big]^2\Big\}.$$

5. Assume that all the assumptions of (FP2-FP1:4) hold except $L \le \gamma_4(H-a_1)$, i.e. the low period is not long enough for class 2 fluid to reach its threshold. Thus, at the end of the low period, class 1 fluid is at its threshold and class 2 fluid is still above its threshold. Hence, if

$$\text{(FP2-FP1:5)} \qquad H \le \tilde{\psi}_1 \le \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1-H)+H \le \tilde{\psi}_2, \quad \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1-H) \le L \le \gamma_4(H-a_1),$$

then the holding cost under FP2-FP1 policy is

$$\begin{aligned} c^{\mathrm{FP2-FP1}}(H,L) &= \frac{1}{2}h_2\mu_2\Big\{(1-\rho_2)\Big[2\tilde{\psi}_2 - \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H) - H\Big]\Big[\frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H) + H\Big] \\ &\quad +(1-\rho_2-\rho_1^l)\Big[\frac{\rho_1^h+\rho_2-1}{1-\rho_2-\rho_1^l}(H-a_1) - \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H)\Big]^2\Big\} \\ &\quad -\frac{1}{2}h_2\mu_2(1-\rho_2-\rho_1^l)\Big[\frac{\rho_1^h+\rho_2-1}{1-\rho_2-\rho_1^l}(H-a_1) - L\Big]^2. \end{aligned}$$

6. Assume that the assumptions of (FP2-FP1:5) hold except the low period is not long enough for either class 1 fluid or class 2 fluid to reach its threshold. Thus, at the end of the low period, class 1 fluid is below its threshold and class 2 fluid is above its threshold. Hence, if

$$\text{(FP2-FP1:6)} \qquad H \le \tilde{\psi}_1 \le \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H) + H \le \tilde{\psi}_2, \quad L \le \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H),$$

then the cost under FP2-FP1 policy is

$$c^{\mathrm{FP2-FP1}}(H,L) = \frac{1}{2}h_2\mu_2(1-\rho_2)(2\tilde{\psi}_2 - H - L)(H+L).$$

7. Assume that if the high period were long enough, class 1 fluid would increase to its threshold before class 2 fluid decreases to its threshold. However, the high period is not long enough for class 1 fluid to increase to its threshold. Suppose that at the end of the high period, class 1 fluid is still below its threshold, and class 2 fluid is still above its threshold. Moreover, assume that the low period is long enough for class 2 fluid to decrease to its threshold and in the low period, class 2 fluid decreases to its threshold earlier than class 1 fluid increases to its threshold. Hence, if

$$\text{(FP2-FP1:7)} \qquad H \le \tilde{\psi}_1 \le \tilde{\psi}_2 \le \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H) + H, \quad L + H \ge \tilde{\psi}_2,$$

then the holding cost under FP2-FP1 policy is

$$c^{\mathrm{FP2-FP1}}(H,L) = \frac{1}{2}h_2\mu_2(1-\rho_2)\tilde{\psi}_2^2.$$

8. Assume that all the assumptions of (FP2-FP1:7) hold except that the low period is not long enough for class 2 fluid to decrease to its threshold. Hence, if

$$\text{(FP2-FP1:8)} \qquad H \le \tilde{\psi}_1 \le \tilde{\psi}_2 \le \frac{\rho_1^h}{\rho_1^l}(\tilde{\psi}_1 - H) + H, \quad L + H \le \tilde{\psi}_2$$

then the holding cost under FP2-FP1 policy is

$$c^{\mathrm{FP2-FP1}}(H,L) = \frac{1}{2}h_2\mu_2(1-\rho_2)(2\tilde{\psi}_2 - H - L)(H+L).$$

9. Assume that class 2 fluid decreases to its threshold before class 1 fluid increases to its threshold. Hence, $\tilde{\psi}_1 \geq \tilde{\psi}_2$. After $\tilde{\psi}_2$, the server allocates just enough capacity to keep class 2 fluid at its threshold, i.e. $u_2 = \rho_2$, and the remaining capacity is allocated to class 1, i.e $u_1 = 1 - \rho_2$. If $H \geq a_1$, then class 1 fluid reaches its threshold at a_1 and after a_1, class 1 has higher priority until class 1 fluid decreases to its threshold again in the low period. After class 1 fluid decreases to its threshold in the low period, we have $u_1 = \rho_1^l$ and $u_2 = 1 - \rho_1^l$. Moreover, assume that $L \geq \gamma_4(H - a_1)$. Thus, at the end of the low period, fluid levels of both classes are below their thresholds. Hence, if

$$\text{(FP2-FP1:9)} \qquad \tilde{\psi}_1 \geq \tilde{\psi}_2, \quad H \geq a_1, \quad L \geq \gamma_4(H - a_1),$$

then the holding cost under FP2-FP1 policy is

$$\begin{aligned} c^{\text{FP2}-\text{FP1}}(H, L) &= \frac{1}{2} h_2 \mu_2 \Big\{ \frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{\eta(1 - \rho_1^l)} (H - a_1)^2 + \rho_2 \Big[\frac{\rho_1^h - \rho_1^l}{1 - \rho_1^l} (H - a_1) \Big]^2 \\ &\quad + (1 - \rho_2)\tilde{\psi}_2^2 + (1 - \rho_2 - \rho_1^l) \Big[\frac{\rho_1^h + \rho_2 - 1}{1 - \rho_2 - \rho_1^l} (H - a_1) - \frac{\rho_1^h - 1}{1 - \rho_1^l} (H - a_1) \Big]^2 \Big\}. \end{aligned}$$

10. Assume that all the assumptions of (FP2-FP1:9) hold except $L \leq \gamma_4(H - a_1)$. Thus, at the end of the low period, class 1 fluid is below its threshold, but class 2 fluid is above its threshold. Hence, if

$$\text{(FP2-FP1:10)} \qquad \tilde{\psi}_1 \geq \tilde{\psi}_2, \quad H \geq a_1, \quad \gamma_3(H - a_1) \leq L \leq \gamma_4(H - a_1),$$

then the holding cost under FP2-FP1 policy is

$$\begin{aligned} c^{\text{FP2}-\text{FP1}}(H, L) &= \frac{1}{2} h_2 \mu_2 \Big\{ \frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{\eta(1 - \rho_1^l)} (H - a_1)^2 + \rho_2 \Big[\frac{\rho_1^h - \rho_1^l}{1 - \rho_1^l} (H - a_1) \Big]^2 \\ &\quad + (1 - \rho_2)\tilde{\psi}_2^2 + (1 - \rho_2 - \rho_1^l) \Big[\frac{\rho_1^h + \rho_2 - 1}{1 - \rho_2 - \rho_1^l} (H - a_1) - \frac{\rho_1^h - 1}{1 - \rho_1^l} (H - a_1) \Big]^2 \Big\} \\ &\quad - \frac{1}{2} h_2 \mu_2 (1 - \rho_2 - \rho_1^l) \Big[\frac{\rho_1^h + \rho_2 - 1}{1 - \rho_2 - \rho_1^l} (H - a_1) - L \Big]^2. \end{aligned}$$

11. Assume that all the assumptions of (FP2-FP1:10) hold except that the low period is not long enough for class 1 fluid to decrease to its threshold, i.e. $L \leq \gamma_3(H - a_1)$. Thus, at the end of the low period, class 1 and class 2 fluids are above their thresholds. Hence, if

$$\text{(FP2-FP1:11)} \qquad \tilde{\psi}_1 \geq \tilde{\psi}_2, \quad H \geq a_1, \quad L \leq \gamma_3(H - a_1),$$

then the holding cost under FP2-FP1 policy is

$$\begin{aligned} c^{\text{FP2}-\text{FP1}}(H, L) &= \frac{1}{2} h_2 \mu_2 \Big\{ \frac{(\rho_1^h - 1)(\rho_1^h - \rho_1^l)}{\eta(1 - \rho_1^l)} (H - a_1)^2 + \rho_2 (H + L - a_1)^2 \\ &\quad + (1 - \rho_2)\tilde{\psi}_2^2 - \frac{1 - \rho_1^l}{\eta} \Big[\frac{\rho_1^h - 1}{1 - \rho_1^l} (H - a_1) - L \Big]^2 \Big\}. \end{aligned}$$

12. Assume that all the assumptions of (FP2-FP1:9) hold except $H \leq a_1$, i.e. high period is not long enough for class 1 fluid to increase to its threshold. Thus, at the end of the high period, class 2 and class 1 fluids are below their thresholds. Hence, if

$$\text{(FP2-FP1:12)} \qquad \tilde{\psi}_1 \geq \tilde{\psi}_2, \quad \tilde{\psi}_2 \leq H \leq a_1,$$

then the holding cost under FP2-FP1 policy is

$$c^{\mathrm{FP2-FP1}}(H, L) \;\; = \;\; \frac{1}{2} h_2 \mu_2 (1 - \rho_2) \tilde{\psi}_2^2.$$

13. Assume that class 2 fluid decreases to its threshold before class 1 fluid increases to its threshold but $H \leq \tilde{\psi}_2$. Then, at the end of the high period, class 2 fluid is above its threshold and class 1 fluid is below its threshold. Suppose that in the low period class 2 has higher priority and class 2 fluid decreases to its threshold at $\tilde{\psi}_2$ and class 1 fluid remains below its threshold. Thus, at the end of the low period both classes are below their thresholds. Hence, if

$$\text{(FP2-FP1:13)} \qquad \tilde{\psi}_1 \geq \tilde{\psi}_2, \quad H \leq \tilde{\psi}_2, \quad H + L \geq \tilde{\psi}_2,$$

then the holding cost under FP2-FP1 policy is

$$c^{\mathrm{FP2-FP1}}(H, L) \;\; = \;\; \frac{1}{2} h_2 \mu_2 (1 - \rho_2) \tilde{\psi}_2^2.$$

14. Assume that all assumptions of (FP2-FP1:13) hold except that the low period is not long enough for class 2 fluid to decrease to its threshold. Hence, if

$$\text{(FP2-FP1:14)} \qquad \tilde{\psi}_1 \geq \tilde{\psi}_2, \quad H \leq \tilde{\psi}_2, \quad H + L \leq \tilde{\psi}_2,$$

then the holding cost under the FP2-FP1 policy is

$$c^{\mathrm{FP2-FP1}}(H, L) \;\; = \;\; \frac{1}{2} h_2 \mu_2 (1 - \rho_2)(2\tilde{\psi}_2 - H - L)(H + L).$$

References

[1] F. Avram, D. Bertsimas and M. Ricard. Fluid models of sequencing problems in open queueing networks: an optimal control approach. *Stochastic Networks*, eds. F.P. Kelly and R.J. Williams pp.199–234, 1995.

[2] N. Bansal and M. Harchol-Balter. Analysis of M/G/1/SRPT under Transient Overload. *Performance Evaluation Review*, 29:3, pp.12–14, 2001.

[3] N. Bauerle and U. Rieder. Optimal control of single-server fluid networks. *Queueing Systems*, 35, pp.185–200, 2000.

[4] J.X. Chang, H. Ayhan, J. Dai, Z. Liu, M.S. Squillante and C.H. Xia. Optimal dynamic scheduling in a multiclass fluid model of internet servers with transient overload. *Proceedings of 42nd IEEE Conference on Decision and Control*, pp.721–727, 2003.

[5] H. Chen and P. Mohapatra. Session-based overload control in QoS-aware Web servers. *Proceedings of INFOCOM*, 23-27, vol.2, pp.516–524, 2002.

[6] A. Gajrat and A. Hordijk. Fluid approximation of a controlled multiclass tandem network. *Queueing Systems*, 35, pp.349–380, 2000.

[7] T.C. Green and S. Stidham Sample-path conservation laws, with applications to scheduling queues and fluid systems. *Queueing Systems*, 36, pp.175–199, 2000.

[8] A. Iyengar, E. MacNair and T. Nguyen. An analysis of web server performance. *Proceedings of IEEE Global Telecommunications Conference*, vol.3, pp.1943–1947, 1997.

[9] G.P. Klimov. Time sharing service systems I. *Theory of Probability and Its Applications*, 19:3, pp.532–551, 1974.

[10] Z. Liu, N. Niclausse, and C. Jalpa-Villanueva. Traffic model and performance evaluation of Web servers. *Performance Evaluation*, 46, pp.77–100, 2001.

[11] Z. Liu, M.S. Squillante, and J.L. Wolf. Optimal control of resource allocation in e-business environments with strict quality-of-service performance guarantees. *Proceedings of the 41st IEEE Conference on Decision and Control*, vol.4, pp.4431–4439, 2002.

[12] Z. Liu, M.S. Squillante, C.H. Xia, S.-Z. Yu, and L. Zhang. Profile-based Traffic Characterization of Commercial Web Sites, *Proceedings of the 18th International Teletraffic Congress (ITC18)*, Berlin, Germany 2003, pp. 231-240.

[13] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze and E.F. Mishchenko. *The Mathematical Theory of Optimal Processes*, Interscience Publishers, New York, 1962.

[14] A. Seierstad, and K. Sydsater. Sufficient Conditions in Optimal Control Theory. *International Economic Review*, vol.18, No.2, pp.367-391, 1977.

[15] J. G. Shanthikumar and S. H. Xu. Strongly Asymptotically Optimal Design and Control of Production and Service Systems. *IIE Transactions*, 32:9, pp.881–890, 2000.

[16] W. E. Smith. Various optimizers for single-stage production. *Naval Res. Logist. Quart.*, 3, pp.59–66, 1956.

[17] G. Weiss. Scheduling and Control of Manufacturing Systems – a Fluid Approach. *Proceedings of the 37th Allerton Conference*, pp.577–586, 1999.

[18] C.H. Xia and J.G. Shanthikumar. Asymptotic Optimal Control of Multi-class $G/G/1$ Queues with Feedback. To appear in *Stochastic Modeling and Optimization of Manufacturing Systems and Supply Chains*, J.G. Shanthikumar, D.D. Yao and W.H.M. Zijm (eds), 2003.

				Percentage Differences off the Lower Bound					
System	Case	$\mathbb{E}[H]$	LB	FP1	π^{a_1}	DSview1	DSview2	DSview3	DSview4
I	A	5	0.00	0.0026*	0.00	0.00	0.00	0.00	0.00
		12.5	1.98	100.77	21.36	19.76	13.34	16.70	19.34
		25	132.16	9.48	15.59	12.66	7.09	10.52	13.35
		37.5	705.40	2.42	11.71	7.66	2.79	6.29	9.31
		50	1883.35	0.90	9.24	3.73	1.36	3.30	6.67
	B	5	0.03	332.61	22.97	16.94	15.29	18.56	21.09
		12.5	20.68	17.92	16.70	10.53	8.99	12.12	14.67
		25	407.05	2.43	10.90	4.90	3.50	6.41	8.88
		37.5	1535.52	0.75	8.01	2.14	0.93	3.59	6.01
		50	3516.83	0.33	6.32	0.61	0.35	1.95	4.33
	C	5	29.51	3.27	11.36	5.58	4.29	7.03	9.41
		12.5	523.38	0.33	5.65	2.05	1.84	3.01	4.26
		25	2962.87	0.17	3.18	0.97	1.08	1.95	2.61
		37.5	7544.71	0.11	2.34	0.45	0.54	1.32	1.88
		50	14337.9	0.07	1.89	0.13	0.17	0.92	1.46
	D	5	3230.38	0.0004	0.73	0.12	0.18	0.23	0.28
		12.5	21190.3	0.002	0.30	0.07	0.18	0.22	0.26
		25	86143.0	0.004	0.15	0.04	0.10	0.12	0.14
		37.5	194926	0.004	0.10	0.02	0.05	0.08	0.10
		50	347597	0.003	0.09	0.005	0.03	0.06	0.07
II	A	5	0.00	67296.6	14.87	13.84	15.74	11.54	12.75
		12.5	14.68	452.27	21.58	18.18	15.69	15.18	18.19
		25	344.88	46.81	22.15	15.58	12.67	13.42	17.60
		37.5	1332.03	12.86	19.44	10.78	8.82	9.61	14.29
		50	3073.30	4.96	16.66	6.54	4.75	6.14	11.05
	B	5	0.43	2061.88	18.97	13.30	14.39	13.82	16.32
		12.5	67.00	104.28	21.63	13.29	13.74	14.41	17.90
		25	756.57	15.74	18.41	8.85	8.76	10.35	14.30
		37.5	2404.51	5.11	14.96	5.19	4.84	6.79	10.82
		50	5085.14	2.27	12.41	2.71	2.31	4.27	8.29
	C	5	59.54	21.69	18.05	9.19	8.99	10.56	14.23
		12.5	747.00	4.34	11.20	5.02	4.82	6.92	9.12
		25	3870.27	2.01	6.82	2.77	2.95	4.53	5.79
		37.5	9600.58	1.06	5.12	1.54	1.67	3.07	4.20
		50	17991.4	0.62	4.19	0.81	0.88	2.18	3.28
	D	5	3908.27	0.01	1.62	0.36	0.56	0.71	0.87
		12.5	25373.9	0.05	0.67	0.23	0.49	0.58	0.63
		25	102840.0	0.05	0.35	0.12	0.23	0.29	0.32
		37.5	232504.0	0.04	0.25	0.06	0.13	0.18	0.22
		50	414419.0	0.02	0.20	0.03	0.07	0.13	0.17

Table 1: Average holding costs when $\mathbb{E}[L] = 12.5$. * indicates the actual value of the average holding cost for the FP1 policy.

System	Case	$\mathbb{E}[H]$	LB	Percentage Differences off the Lower Bound FP1	π^{a_1}	DSview1	DSview2	DSview3	DSview4
I	A	5	0.00	0.0026*	0.00	0.00	0.00	0.00	0.00
		12.5	2.36	86.93	20.92	19.43	13.34	16.57	19.04
		25	160.28	8.03	15.49	12.69	7.13	10.62	13.35
		37.5	847.81	2.06	11.68	7.73	2.62	6.37	9.35
		50	2230.50	0.78	9.22	3.77	1.28	3.33	6.71
	B	5	0.03	294.06	22.96	17.21	15.60	18.78	21.18
		12.5	24.52	15.61	16.48	10.55	9.00	12.10	14.55
		25	478.94	2.13	10.83	4.92	3.46	6.43	8.86
		37.5	1774.28	0.67	7.98	2.13	0.85	3.60	6.01
		50	3998.12	0.30	6.30	0.58	0.31	1.95	4.33
	C	5	33.96	2.95	11.36	5.65	4.30	7.10	9.46
		12.5	578.39	0.31	5.63	2.05	1.84	3.01	4.25
		25	3174.51	0.17	3.19	0.97	1.08	1.95	2.61
		37.5	7964.52	0.11	2.35	0.45	0.54	1.33	1.90
		50	15003.8	0.07	1.92	0.13	0.17	0.94	1.48
	D	5	3270.14	0.0004	0.73	0.12	0.18	0.23	0.29
		12.5	21300.0	0.002	0.30	0.07	0.18	0.22	0.26
		25	86382.7	0.004	0.15	0.04	0.10	0.12	0.14
		37.5	195315	0.004	0.11	0.02	0.05	0.08	0.10
		50	348157	0.003	0.09	0.005	0.03	0.06	0.07
II	A	5	0.01	65776	9.40	8.74	12.53	7.46	8.04
		12.5	21.73	471.27	16.69	14.19	14.51	12.28	14.20
		25	462.35	50.90	19.88	14.39	13.51	12.77	16.03
		37.5	1687.87	14.39	18.44	10.73	9.88	9.76	13.79
		50	3756.04	5.66	16.17	6.80	5.35	6.45	10.93
	B	5	0.65	2113.81	14.14	10.65	12.36	10.62	12.25
		12.5	91.84	112.58	18.42	12.36	13.64	12.82	15.39
		25	949.06	17.91	17.16	9.07	9.56	10.13	13.49
		37.5	2879.99	5.99	14.39	5.58	5.57	6.88	10.52
		50	5914.70	2.72	12.11	3.06	2.76	4.43	8.18
	C	5	74.93	23.78	16.53	9.07	9.29	10.06	13.17
		12.5	852.74	5.35	10.81	5.10	4.93	6.78	8.83
		25	4212.98	2.57	6.73	2.93	3.08	4.51	5.72
		37.5	10251.4	1.37	5.11	1.70	1.82	3.09	4.19
		50	18997.4	0.79	4.21	0.95	1.01	2.22	3.30
	D	5	3962.43	0.014	1.63	0.37	0.56	0.71	0.87
		12.5	25524.5	0.07	0.67	0.23	0.49	0.58	0.63
		25	103166	0.07	0.35	0.13	0.23	0.29	0.32
		37.5	233036	0.05	0.25	0.07	0.13	0.19	0.22
		50	415185	0.03	0.20	0.04	0.08	0.13	0.17

Table 2: Average holding costs when $\mathbb{E}[L] = 25$. * indicates the actual value of the average holding cost for the FP1 policy.

System	Case	$\mathbb{E}[H]$	LB	Percentage Differences off the Lower Bound FP1	π^{a_1}	DSview1	DSview2	DSview3	DSview4
I	A	5	0.00	0.0027∗	0.00	0.00	0.00	0.00	0.00
		12.5	2.69	76.98	20.01	18.63	12.91	15.96	18.26
		25	191.80	6.78	14.76	12.16	6.84	10.22	12.77
		37.5	1029.76	1.72	11.23	7.50	2.40	6.19	9.04
		50	2712.54	0.65	8.95	3.69	1.18	3.26	6.55
	B	5	0.04	264.86	22.08	16.66	15.12	18.14	20.41
		12.5	28.61	13.63	15.88	10.28	8.77	11.76	14.07
		25	570.02	1.82	10.47	4.80	3.33	6.28	8.61
		37.5	2106.81	0.58	7.76	2.07	0.76	3.53	5.87
		50	4708.96	0.26	6.16	0.54	0.27	1.92	4.25
	C	5	39.83	2.57	11.18	5.61	4.25	7.06	9.35
		12.5	659.90	0.28	5.55	2.03	1.82	2.98	4.20
		25	3511.68	0.15	3.16	0.96	1.08	1.94	2.60
		37.5	8655.56	0.10	2.35	0.45	0.54	1.33	1.90
		50	16120.7	0.06	1.93	0.13	0.17	0.94	1.50
	D	5	3346.52	0.0004	0.73	0.12	0.18	0.23	0.29
		12.5	21512.7	0.002	0.30	0.07	0.18	0.22	0.26
		25	86834.6	0.004	0.15	0.04	0.10	0.12	0.14
		37.5	196039	0.004	0.11	0.02	0.05	0.08	0.10
		50	349187	0.003	0.09	0.006	0.03	0.06	0.08
II	A	5	0.01	60834.20	5.81	5.40	8.63	4.67	4.97
		12.5	34.48	472.51	11.30	9.65	11.70	8.57	9.66
		25	679.44	54.28	15.35	11.35	12.94	10.41	12.50
		37.5	2336.96	16.06	15.52	9.46	10.62	8.84	11.76
		50	4988.97	6.52	14.34	6.55	5.93	6.30	9.84
	B	5	1.02	2040.44	9.61	7.63	9.41	7.37	8.36
		12.5	137.10	116.42	13.77	10.06	12.01	9.93	11.59
		25	1299.36	19.96	14.39	8.50	9.77	8.92	11.41
		37.5	3739.20	7.00	12.77	5.71	6.30	6.49	9.43
		50	7410.93	3.27	11.10	3.41	3.31	4.39	7.58
	C	5	102.12	25.17	13.99	8.35	9.05	8.88	11.25
		12.5	1044.08	6.34	9.93	4.98	4.87	6.34	8.15
		25	4836.52	3.24	6.43	3.06	3.16	4.36	5.48
		37.5	11441.8	1.77	4.98	1.90	1.97	3.06	4.10
		50	20846.7	1.04	4.16	1.14	1.19	2.23	3.26
	D	5	4605.41	0.02	1.62	0.37	0.56	0.71	0.87
		12.5	25815.6	0.11	0.66	0.24	0.50	0.58	0.63
		25	103803	0.10	0.35	0.14	0.24	0.29	0.33
		37.5	234065	0.07	0.25	0.08	0.14	0.19	0.23
		50	416648	0.04	0.21	0.05	0.08	0.14	0.18

Table 3: Average holding costs when $\mathbb{E}[L] = 50$. * indicates the actual value of the average holding cost for the FP1 policy.

System	Case	$\mathbb{E}[H]$	LB	Percentage Differences off the Lower Bound FP1	π^{a_1}	DSview1	DSview2	DSview3	DSview4
I	A	5	0.00	0.0027*	0.00	0.00	0.00	0.00	0.00
		12.5	3.04	68.30	18.55	17.29	12.03	14.85	16.96
		25	246.67	5.29	12.83	10.61	5.98	8.94	11.13
		37.5	1463.13	1.21	9.39	6.32	1.91	5.22	7.59
		50	4161.72	0.43	7.31	3.05	0.89	2.68	5.38
	B	5	0.04	231.28	21.14	16.13	14.67	17.51	19.60
		12.5	37.56	10.58	14.15	9.28	7.92	10.59	12.59
		25	846.81	1.25	8.86	4.11	2.80	5.38	7.32
		37.5	3386.34	0.36	6.41	1.70	0.53	2.96	4.88
		50	7981.57	0.15	5.03	0.40	0.16	1.58	3.50
	C	5	62.00	1.70	9.53	4.87	3.64	6.11	8.03
		12.5	1174.26	0.16	4.61	1.71	1.53	2.51	3.52
		25	6481.53	0.09	2.64	0.80	0.90	1.63	2.17
		37.5	15825.0	0.06	1.98	0.36	0.45	1.13	1.60
		50	28922.2	0.04	1.64	0.09	0.13	0.81	1.28
	D	5	5151.63	0.0002	0.69	0.12	0.17	0.22	0.27
		12.5	27731.1	0.001	0.29	0.07	0.17	0.22	0.25
		25	101086	0.004	0.15	0.04	0.10	0.12	0.14
		37.5	218685	0.004	0.11	0.02	0.05	0.08	0.10
		50	380521	0.003	0.09	0.006	0.03	0.06	0.08
II	A	5	0.03	42604.30	2.25	2.09	3.43	1.80	1.92
		12.5	158.84	297.99	2.56	2.19	3.30	1.99	2.19
		25	4177.56	31.79	2.77	2.08	4.00	2.03	2.27
		37.5	15194.7	9.47	2.80	1.85	4.69	1.89	2.14
		50	31848.9	4.03	2.77	1.64	2.95	1.69	1.94
	B	5	3.88	1275.31	2.75	2.33	3.15	2.16	2.40
		12.5	721.66	69.32	3.03	2.57	3.69	2.30	2.57
		25	7654.53	11.62	3.03	2.40	3.76	2.08	2.43
		37.50	21436.6	4.30	2.90	2.10	3.41	1.74	2.18
		50	40180.6	2.15	2.77	1.74	2.17	1.41	1.93
	C	5	525.66	17.23	3.56	2.79	3.88	2.51	2.91
		12.5	4768.37	4.52	3.15	1.94	1.97	2.11	2.61
		25	18169.3	2.99	2.56	1.65	1.61	1.80	2.20
		37.5	37865.5	1.91	2.25	1.39	1.33	1.45	1.87
		50	62926.3	1.25	2.05	1.13	1.10	1.18	1.62
	D	5	6885.53	0.03	1.38	0.32	0.48	0.61	0.74
		12.5	34368.7	0.29	0.63	0.28	0.47	0.55	0.60
		25	123211	0.36	0.34	0.24	0.24	0.29	0.32
		37.5	265930	0.25	0.26	0.19	0.16	0.19	0.23
		50	462363	0.17	0.21	0.15	0.12	0.15	0.18

Table 4: Average holding costs when $\mathbb{E}[L] = 1000$. * indicates the actual value of the average holding cost for the FP1 policy.