Parametric Analysis of Optimal Static Load Balancing in Distributed Computer Systems

YONGBING ZHANG*, HISAO KAMEDA* and KENTARO SHIMIZU*

A model of optimal static load balancing problems in a distributed computer system, which consists of a set of heterogeneous host computers connected by a communications network, is considered. We consider the overall and individually optimal policies. The first is for minimizing the overall mean job response time. The second is for determining the equilibrium so that no user has any incentive to change his processing node to improve his expected response time. Tantawi and Towsley showed the conditions that the solution of the overall optimal policy satisfies. In this paper we show the conditions that the solution of the individually optimal policy satisfies and show the existence of the solution. Then we examine the effects of some of the main system parameters on the performance variables of the two policies. In the parametric analysis and numerical examination we show that there exists a striking parallelism between the characteristics of the two policies even though they are entirely different from each other. Some anomalous or counter-intuitive phenomena are also observed.

1. Introduction

Distributed computer systems possess many potentially attractive features. One of these is the capability to share processing of jobs in the event of overloads. This study focuses on the issue of balancing loads between nodes of a distributed system in response to imbalances in loads. Load balancing policies may be either static or adaptive. It seems clear that adaptive policies [3, 4, 9, 10] may be more effective than static policies whereas the former may have more overhead than the latter. Furthermore, it seems that there currently exists no optimal adaptive policy that is sophisticated enough to be generally applicable to various environments and analytically tractable. Static policies are very useful to estimate the performance measure when constructing a new system or updating an existing system. Thus, in this paper, we focus on static policies.

We can think of two optimal static load balancing policies which have contrastive performance objectives for load balancing. We call the one the *overall optimal policy* which optimizes the mean job response time and the other the *individually optimal policy* whereby job scheduling is determined so that every job may feel that its own expected response time is minimum if it knows the expected node and communication delays. The motivation of the overall optimal policy seems clear. We may say that the individually optimal policy balances loads from the user's viewpoint.

Related to these two policies, many studies [1, 2, 7, 11] have been concerned with the overall and individually optimal static and adaptive policies for various specific systems. Tantawi and Towsley [13] considered an overall optimal policy in a model of a distributed computer system that consists of a set of heterogeneous host computers connected by a single channel communications network such as satellite networks and local area networks. They derived the conditions that the optimal solution should satisfy and proposed an algorithm that determines the optimal solution. We call the solution the *optimum*. Kim and Kameda [5] showed an algorithm improved over that of Tantawi and Towsley [12, 13]. Furthermore, they [6] extended the model of Tantawi and Towsley [13] to the case of multiple job classes.

In this paper, we study an individually optimal policy in the same model as the Tantawi and Towsley single class model. We show the conditions that the solution of the individually optimal policy satisfies. Then we show that there exists a striking *parallelism* between the above mentioned solutions of the overall and individually optimal policies. Furthermore, we study the characteristics of the overall and individually optimal policies and parametric analysis, that is, the effects of varying the system parameters, such as a planned equipment upgrade and long term fluctuations of loads, on the performance variables of these policies.

*Department of Computer Science and Information Mathematics, University of Electro-Communications, Tokyo 182, Japan.

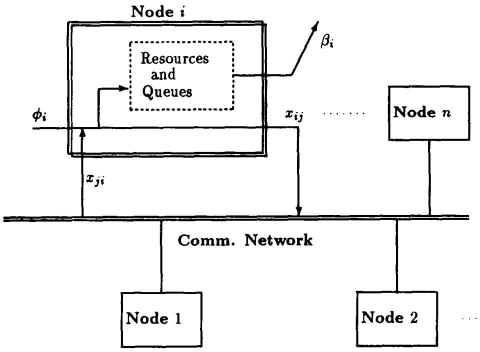

Fig. 1 A model of a distributed computer system.

2. Model Description

We consider a distributed computer system model that consists of n nodes (host computers) connected by a single channel communications network as shown in Fig. 1. The key assumptions of the model are the same as those of Tantawi and Towsley [13]. Nodes may be heterogeneous, i.e., they may have different configurations, number of resources, and processing capacities. We assume that the expected communication delay from node i to node j is independent of source-destination pair (i, j). Let us have the following notation.

- n Number of nodes
- ϕ_i External arrival rate to node i
- Φ Total external job arrival rate, i.e., $\Phi=\Sigma_{i=1}^n \phi_i$
- x_{ij} Job flow rate from node i to node j
- β_i Job processing rate (load) at node i, i.e., $\beta_i=\Sigma_{l=1}^n x_{li}$
- β $[\beta_1, \beta_2, \cdots, \beta_n]$
- λ Total traffic through the network, i.e., $\lambda=\Sigma_i \Sigma_{j,j\neq i} x_{ij}$
- $F_i(\beta_i)$ Expected node delay of jobs processed at node i (We assume that it is differentiable, increasing, and convex with respect to β_i)
- $G(\lambda)$ Expected communication delay of jobs (We assume that it is source-destination independent, differentiable, nondecreasing, and convex with respect to λ)
- $T(\beta)$ Overall mean job response time, i.e., the mean length of the time period that starts when a job arrives in the system and ends when it leaves the system.

Jobs arrive at each node according to a time-invariant Poisson process. A job that arrives at node i(origin node) may either be processed at node i or be transferred to another node j(processing node). After the job is processed at node j, a response is sent back to the origin node. Also we assume that a transferred job from node i to node j receives its service at node j and is not transferred to other nodes. We can write the overall mean job response time as the sum of the mean node delay and the mean communication delay, that is

$$T(\beta)=\frac{1}{\Phi}\left[\sum_{i=1}^{n}\beta_i F_i(\beta_i)+\lambda G(\lambda)\right], \tag{1}$$

subject to

$$\sum_{i=1}^{n}\beta_i=\Phi, \tag{2}$$

$$\beta_i\geq 0, \quad i=1, 2, \cdots, n, \tag{3}$$

where the network traffic λ may be expressed in terms of variable β_i as

$$\lambda=\frac{1}{2}\sum_{i=1}^{n}|\phi_i-\beta_i|. \tag{4}$$

We identify nodes in the following way: (1) idle source (R_d): The node does not process any jobs, i.e., $\beta_i=0$. (2) active source (R_a): The node sends jobs and does not receive any jobs, i.e., $\phi_i>\beta_i>0$. (3) neutral (N): The node processes jobs locally without sending or receiving jobs, i.e., $\beta_i=\phi_i$. (4) sink (S): The node receives jobs from other nodes but does not send any jobs out, i.e., $\beta_i>\phi_i$.

3. Optimal Solutions

3.1 The Solution of the Overall Optimal Policy

By the overall optimal policy we mean the policy whereby load is balanced so as to minimize the overall mean job response time, that is, solving problem (1) with constraints (2) and (3).

Now we introduce two functions, the *incremental node delay* $f_i(\beta_i)$ and the *incremental communication delay* $g(\lambda)$, as follows.

$$f_i(\beta_i)=\frac{d}{d\beta_i}\beta_i F_i(\beta_i), \quad g(\lambda)=\frac{d}{d\lambda}\lambda G(\lambda). \tag{5}$$

Define the inverse of the incremental node delay f_i^{-1} by

$$f_i^{-1}(x)=\begin{cases} a, & f_i(a)=x, \\ 0, & f_i(0)\geq x. \end{cases}$$

According to the results of Tantawi and Towsley [13], we have the following theorem by which we can determine β that implements the overall optimal policy.

Theorem 3.1 *The optimal solution, β, to problem (1) satisfies the relations*

$$f_i(\beta_i)\geq\alpha+g(\lambda), \quad \beta_i=0 \quad (i\in R_d), \tag{6}$$

$$f_i(\beta_i)=\alpha+g(\lambda), \quad 0<\beta_i<\phi_i \quad (i\in R_a), \tag{7}$$

$$\alpha\leq f_i(\beta_i)\leq\alpha+g(\lambda), \quad \beta_i=\phi_i \quad (i\in N), \tag{8}$$

$$\alpha=f_i(\beta_i), \quad \beta_i>\phi_i \quad (i\in S), \tag{9}$$

subject to the total flow constraint

$$\sum_{i\in S}f_i^{-1}(\alpha)+\sum_{i\in R_a}f_i^{-1}(\alpha+g(\lambda))+\sum_{i\in N}\phi_i=\Phi, \tag{10}$$

where α is the Lagrange multiplier.

3.2 The Solution of the Individually Optimal Policy

According to the individually optimal policy, jobs are scheduled so that every job may feel that its own expected response time is minimum if it knows the expected node delay at each node and the expected communication delay. In other words, when the individually optimal policy is realized, the expected response time of a job cannot be improved further when the scheduling decisions for other jobs are fixed, and the system reaches an equilibrium. By the following definition, we define the equilibrium conditions.

Definition: *β is said to satisfy the equilibrium conditions for the individually optimal policy, if the following relations hold*:

$$F_i(\beta_i) \geq R + G(\lambda), \quad \beta_i = 0 \qquad (i \in R_d), \tag{11}$$

$$F_i(\beta_i) = R + G(\lambda), \quad 0 < \beta_i < \phi_i \qquad (i \in R_a), \tag{12}$$

$$R \leq F_i(\beta_i) \leq R + G(\lambda), \quad \beta_i = \phi_i \qquad (i \in N), \tag{13}$$

$$R = F_i(\beta_i), \quad \beta_i > \phi_i \qquad (i \in S), \tag{14}$$

subject to the total flow constraint:

$$\sum_{i \in S} F_i^{-1}(R) + \sum_{i \in R_a} F_i^{-1}(R + G(\lambda)) + \sum_{i \in N} \phi_i = \Phi. \tag{15}$$

We call β the solution of the individually optimal policy if it satisfies the equilibrium conditions (11)-(14). We also call such a β the *equilibrium*.

Remark. In the above definition, R and $R + G(\lambda)$ represent the expected response time of jobs arriving at sinks and the expected response time of jobs sent to sinks, respectively. $F_i(\beta_i)$ denotes the expected node delay of node i. For example, let us examine a job that arrives at an idle node. According to eq. (11), we see that the job should be sent to a sink node. If the job decides to receive service locally, its expected response time cannot be improved because the expected node delay of the idle node is greater than the expected response time of jobs sent to sinks. Furthermore, if the job decides to be sent to another sink node, its expected response time can also not be improved because the expected node delays of all sinks are equal according to eq. (14). For a job arriving at an active or a neutral node, we see that the expected response time of the job cannot be improved according to eqs. (12) and (13).

For the individually optimal policy, we have the following theorem.

Theorem 3.2 *There exists one and only one β that satisfies the equilibrium conditions (11)-(14). Such β is the solution of the individually optimal policy.*

Proof. As pointed out in [8] by Magnanti, we can express the individually optimization problem by using an equivalent overall optimization problem. We denote the expected node and communication delays of the equivalent overall optimization problem by $F_i^*(\beta_i)$ and $G^*(\lambda)$, respectively, and define $F_i^*(\beta_i)$ and $G^*(\lambda)$ as follows.

$$F_i^*(\beta_i) = \frac{1}{\beta_i}\int_0^{\beta_i} F_i(\beta_i)\,d\beta_i, \quad F_i^*(0) = F_i(0),$$

$$G^*(\lambda) = \frac{1}{\lambda}\int_0^{\lambda} G(\lambda)\,d\lambda, \quad G^*(0) = G(0).$$

$F_i^*(\beta_i)$ and $G^*(\lambda)$ are different from $F_i(\beta_i)$ and $G(\lambda)$, respectively, but they have relationship with $F_i(\beta_i)$ and $G(\lambda)$ as defined above. By using the similar way as in problem (1), we can formulate the equivalent overall optimization problem as follows.

$$\min T^*(\beta) = \frac{1}{\Phi}\left[\sum_{i=1}^{n} \beta_i F_i^*(\beta_i) + \lambda G^*(\lambda)\right]. \tag{16}$$

subject to

$$\sum_{i=1}^{n} \beta_i = \Phi, \tag{17}$$

$$\beta_i \geq 0, \quad i = 1, 2, \cdots, n. \tag{18}$$

where the network traffic λ is expressed in terms of variables β_i as $\lambda = \frac{1}{2}\sum_{i=1}^{n} |\phi_i - \beta_i|$.

Noting problems (1) and (16), we see that they are the similar problems except the differences between $F_i^*(\beta_i)$, $G^*(\lambda)$ and $F_i(\beta_i)$, $G(\lambda)$. Here, we need to check the convexity of problem (16). Since $dF_i^*/d\beta_i > 0$ and $d^2F_i^*/d\beta_i^2 > 0$ for all i, and $dG^*/d\lambda \geq 0$ and $d^2G^*/d\lambda^2 \geq 0$ for λ, we may immediately conclude that $T^*(\beta)$ is a strictly convex function of variables β_i. Furthermore, the variables β_i belong to a convex polyhedron. Thus we may conclude that if the problem is feasible at all, then any local minimum is a global minimum for $T^*(\beta)$.

To problem (16), by using the similar way as that of Tantawi and Towsley [12], we have the following relations.

$$F_i(\beta_i) \geq R + G(\lambda), \quad \beta_i = 0 \qquad (i \in R_d), \tag{19}$$

$$F_i(\beta_i) = R + G(\lambda), \quad 0 < \beta_i < \phi_i \qquad (i \in R_a), \tag{20}$$

$$R \leq F_i(\beta_i) \leq R + G(\lambda), \quad \beta_i = \phi_i \qquad (i \in N), \tag{21}$$

$$R = F_i(\beta_i), \quad \beta_i > \phi_i \qquad (i \in S), \tag{22}$$

subject to the total flow constraint

$$\sum_{i \in S} F_i^{-1}(R) + \sum_{i \in R_a} F_i^{-1}(R + G(\lambda)) + \sum_{i \in N} \phi_i = \Phi. \tag{23}$$

By noting the equilibrium conditions (11)-(14) for the individually optimal policy in Section 3, we see that the conditions that the solution of the problem (16) satisfies are equivalent to those equilibrium conditions. Therefore we conclude that the individually optimal policy has one and only one solution and that its solution satisfies the conditions (11)-(14). □

Remark. We can observe that the solutions of the overall and individually optimal policies have a striking parallelism in the forms of the solutions of the two policies. The parallelism between Theorems 3.2 and 3.1 gives us an intuitive explanation of one in terms of the other. That is, the overall optimal policy would be realized by an individually optimal policy, if the values of the incremental node and communication delays were given as the expected node and communication delays, and vice versa. Therefore, we may implement the individually optimal policy by using that $f_i(\beta_i)$ and $g(\lambda)$ are replaced with $F_i(\beta_i)$ and $G(\lambda)$ in the algorithm of [5].

4. Parametric Analysis

In this section, we study the effects of the system parameters on the behavior of the system in the optimum and in the equilibrium while node partition remains the same, respectively. We consider the communication time t, the node i processing time $u_i (i=1, 2, \cdots, n)$, and the node i job arrival rate $\phi_i (i=1, 2, \cdots, n)$ as system parameters. We use a vector $\mathbf{p}$ to denote $[t, u_1, u_2, \cdots, u_n, \phi_1, \phi_2, \cdots, \phi_n]$. $f_i(\beta_i, u_i)$ and $g(\lambda, t)$ denote the incremental node and communication delays, respectively. We assume that $f_i(\beta_i, u_i)$ is a convex function with respect to β_i, and an increasing function with respect to u_i. Similarly, we assume that $g(\lambda, t)$ is a convex function with respect to λ, and an increasing function with respect to t. Similarly for $F_i(\beta_i, u_i)$ and $G(\lambda, t)$. Parameters λ, β_i, and α in the optimum are determined when $\mathbf{p}$ is given and we write them as $\lambda(\mathbf{p})$, $\beta_i(\mathbf{p})$, and $\alpha(\mathbf{p})$. Similarly we may write parameters λ, β_i, and R in the equilibrium as $\lambda(\mathbf{p})$, $\beta_i(\mathbf{p})$, and $R(\mathbf{p})$.

We define an inverse function e_i of the incremental node delay f_i as follows.

$$e_i(\alpha, u_i)=\beta_i, \quad \text{iff} \quad f_i(\beta_i, u_i)=\alpha. \tag{24}$$

Similarly, we can define an inverse function E_i of the expected node delay F_i as follows.

$$E_i(R, u_i)=\beta_i, \quad \text{iff} \quad F_i(\beta_i, u_i)=R. \tag{25}$$

We assume that the expected communication delay $G(\lambda, t)$ and the incremental communication delay $g(\lambda, t)$ increase with t.

4.1 Parametric Analysis of the Overall Optimal Policy

We analyze the behavior of the performance variables of the overall optimal policy as follows.

Theorem 4.1 *The following relations hold for the incremental node delay $\alpha(\mathbf{p})$ at sinks.*

$$\frac{\partial \alpha(\mathbf{p})}{\partial t}<0. \tag{26}$$

$$\begin{aligned}\frac{\partial \alpha(\mathbf{p})}{\partial u_i}&>0, \; i\in S\cup R_a,\\ &=0, \; i\in N\cup R_d.\end{aligned} \tag{27}$$

$$\begin{aligned}\frac{\partial \alpha(\mathbf{p})}{\partial \phi_i}&>0, \; i\in S\cup R_a,\\ &=0, \; i\in N\cup R_d.\end{aligned} \tag{28}$$

Proof. Given in Appendix A. □

Remark. This theorem implies that the incremental node delay at sinks will decrease as the communication time increases, and that it will increase with the increase in the processing time or in the arrival rate, at a sink or at an active source.

Corollary 4.2 *The following relations hold for the network traffic $\lambda(\mathbf{p})$.*

$$\frac{\partial \lambda(\mathbf{p})}{\partial t}<0. \tag{29}$$

$$\begin{aligned}\frac{\partial \lambda(\mathbf{p})}{\partial u_i}&<0, \; i\in S,\\ &=0, \; i\in N\cup R_d,\\ &>0, \; i\in R_a.\end{aligned} \tag{30}$$

$$\begin{aligned}\frac{\partial \lambda(\mathbf{p})}{\partial \phi_i}&<0, \; i\in S,\\ &=0, \; i\in N\cup R_d,\\ &>0, \; i\in R_a.\end{aligned} \tag{31}$$

Proof. We can derive these from Theorem 4.1. □

Remark. This corollary implies that the network traffic decreases with the increase in the communication time, or with the increase in the processing time or the arrival rate at a sink, and that it increases with the increase in the processing time or in the arrival rate at an active source.

Denote the overall mean job response time in the optimum under the overall optimal policy by $T(\mathbf{p})$. Then we have the following theorem.

Theorem 4.3 *The following relations hold for the overall mean job response time in the optimum, $T(\mathbf{p})$.*

$$\frac{\partial T(\mathbf{p})}{\partial t}>0, \tag{32}$$

$$\begin{aligned}\frac{\partial T(\mathbf{p})}{\partial u_i}&>0, \quad i\in S\cup R_a\cup N,\\ &=0, \quad i\in R_d.\end{aligned} \tag{33}$$

Proof. From eq. (1) and relations (7) and (9), we have

$$\begin{aligned}\frac{\partial T(\mathbf{p})}{\partial t}&=\frac{1}{\Phi}\left[\sum_{i=1}^{n}\frac{\partial(\beta_i F_i)}{\partial \beta_i}\frac{\partial \beta_i}{\partial t}+\frac{\partial(\lambda G(\lambda))}{\partial \lambda}\frac{\partial \lambda}{\partial t}+\lambda\frac{\partial G}{\partial t}\right]\\ &=\frac{\lambda}{\Phi}\frac{\partial G}{\partial t}.\end{aligned}$$

Therefore we have relation (32).

We have (33) in the similar way as above. □

Remark. This theorem implies that the overall mean job response time in the optimum will increase with the increase in the communication time, or with the increase in the node processing times at sinks, sources, or neutrals.

4.2 Parametric Analysis of the Individually Optimal Policy

For the individually optimal policy, we have the following theorems.

Theorem 4.4 *The following relations hold for the expected node delay* $R(\mathbf{p})$ *at sinks.*

$$\frac{\partial R(\mathbf{p})}{\partial t}<0. \tag{34}$$

$$\begin{aligned}\frac{\partial R(\mathbf{p})}{\partial u_i}&>0,\ i\in S\cup R_a,\\ &=0,\ i\in N\cup R_d.\end{aligned} \tag{35}$$

$$\begin{aligned}\frac{\partial R(\mathbf{p})}{\partial \phi_i}&>0,\ i\in S\cup R_a,\\ &=0,\ i\in N\cup R_d.\end{aligned} \tag{36}$$

Proof. By using the similar way as Theorem 4.1, we can derive equations on $\partial R(\mathbf{p})/\partial t$, $\partial R(\mathbf{p})/\partial u_i$, and $\partial R(\mathbf{p})/\partial \phi_i$. Then we can derive the above relations. □

Remark. This theorem implies that the expected node delay at sinks will decrease as the communication time increases, and that it will increase with the increase in the processing time or in the arrival rate, at a sink or at an active source. These agree with our intuition. The corresponding results on the optimum can be derived simply by replacing R with α in eqs. (34), (35), and (36).

Corollary 4.5 *The following relations hold for the network traffic* $\lambda(\mathbf{p})$.

$$\frac{\partial \lambda(\mathbf{p})}{\partial t}<0. \tag{37}$$

$$\begin{aligned}\frac{\partial \lambda(\mathbf{p})}{\partial u_i}&<0,\ i\in S,\\ &=0,\ i\in N\cup R_d,\\ &>0,\ i\in R_a.\end{aligned} \tag{38}$$

$$\begin{aligned}\frac{\partial \lambda(\mathbf{p})}{\partial \phi_i}&<0,\ i\in S,\\ &=0,\ i\in N\cup R_d,\\ &>0,\ i\in R_a.\end{aligned} \tag{39}$$

Proof. We can derive these from Theorem 4.4. □

Remark. This corollary implies that the network traffic decreases with the increase in the communication time, or with the increase in the processing time or the arrival rate at sinks, and that it increases with the increase in the processing time or in the arrival rate at an active source. These agree with our intuition. The corresponding results on the optimum have the same form as relations (37), (38), and (39).

Theorem 4.6 *The following relations hold for the expected response time for jobs arriving at active sources* $\hat{R}(\mathbf{p})=R(\mathbf{p})+G(\lambda(\mathbf{p}),t)$.

$$\frac{\partial \hat{R}(\mathbf{p})}{\partial t}>0. \tag{40}$$

$$\begin{aligned}\frac{\partial \hat{R}(\mathbf{p})}{\partial u_i}&>0,\ i\in S\cup R_a,\\ &=0,\ i\in N\cup R_d.\end{aligned} \tag{41}$$

$$\begin{aligned}\frac{\partial \hat{R}(\mathbf{p})}{\partial \phi_i}&>0,\ i\in S\cup R_a,\\ &=0,\ i\in N\cup R_d.\end{aligned} \tag{42}$$

Proof. Relation (40) can be derived in the same way as Theorem 7 of [12].

Note that

$$\frac{\partial \hat{R}(\mathbf{p})}{\partial u_i}=\frac{\partial R(\mathbf{p})}{\partial u_i}+\frac{\partial G(\lambda,t)}{\partial \lambda}\frac{\partial \lambda(\mathbf{p})}{\partial u_i}. \tag{43}$$

From Theorem 4.4, Corollary 4.5, and eq. (43) we easily see that

$$\begin{aligned}\frac{\partial \hat{R}(\mathbf{p})}{\partial u_i}&>0,\quad i\in R_a,\\ &=0,\quad i\in N\cup R_d.\end{aligned}$$

For $i\in S$, we have from Theorem 4.6 and eq. (43)

$$\frac{\partial \hat{R}(\mathbf{p})}{\partial u_i}=\left(\frac{\partial R(\mathbf{p})}{\partial u_i}\right)\frac{1}{1+B(\mathbf{p})(\partial G(\lambda,t)/\partial\lambda)}>0. \tag{44}$$

where $B(\mathbf{p})=\sum_{i\in R_a}(\partial E_i(\hat{R},u_i)/\partial\hat{R})|_{\hat{R}=\hat{R}(\mathbf{p})}$. Therefore we have relation (41).

Relation (42) is derived similarly as above. □

Denote the overall mean job response time in the equilibrium under the individually optimal policy by $T(\mathbf{p})$. Then we have

$$\begin{aligned}\Phi T(\mathbf{p})=&\sum_{i\in S}\phi_i R(\mathbf{p})+\sum_{R_a\cup R_d}\phi_i(R(\mathbf{p})+G(\lambda(\mathbf{p}),t))\\ &+\sum_{i\in N}\phi_i F_i(\phi_i).\end{aligned} \tag{45}$$

Theorem 4.7 *The following relations hold for the overall mean job response time in the equilibrium,* $T(\mathbf{p})$.

$$\frac{\partial T(\mathbf{p})}{\partial u_i}>0, i\in S\cup R_a\cup N,$$

$$=0, i\in R_d. (46)$$

$$\frac{\partial T(\mathbf{p})}{\partial \phi_i}>0, i\in R_a\cup R_d\cup N. (47)$$

Proof. From eq. (45) we have

$$\begin{aligned}\frac{\partial T(\mathbf{p})}{\partial u_i}&=\frac{1}{\Phi}\left[\sum_{i\in S}\phi_i\frac{\partial R(\mathbf{p})}{\partial u_i}+\sum_{R_a\cup R_d}\phi_i\frac{\partial \hat{R}(\mathbf{p})}{\partial u_i}\right],\ i\in S\cup R_a,\\ &\phi_i\frac{\partial F_i}{\partial u_i},\quad i\in N.\end{aligned}$$

Therefore we have relation (46) by noting Theorems 4.4 and 4.6.

Relation (47) is derived similarly as above. □

Remark. This theorem implies that the overall mean job response time in the equilibrium will increase with the increase in the node processing times at sinks, sources, or neutrals, or with the increase in the job arrival rates at sources or neutrals. These agree with our intuition.

5. Anomalous Behavior of the Optimum and the Equilibrium

In the previous section, we have analyzed the effects of the system parameters on the performance variables of the overall and individually optimal policies. Note that $\partial T(\mathbf{p})/\partial\phi_i$ is not presented in Section 4.1 and that $\partial T(\mathbf{p})/\partial t$ and $\partial T(\mathbf{p})/\partial\phi_i$ for $i\in S$ are not presented in Section 4.2. Let us examine these in the following.

5.1 The Overall Optimal Policy

From eq. (1) and relations (7) and (9), we have

$$\begin{aligned}\frac{\partial T(\mathbf{p})}{\partial\phi_i}&=\frac{1}{\Phi^2}\left[\Phi\left(\sum_{j=1}^{n}\frac{\partial(\beta_jF_j)}{\partial\beta_j}\frac{\partial\beta_j}{\partial\phi_i}\right.\right.\\&\qquad\left.\left.+\frac{\partial(\lambda G(\lambda))}{\partial\lambda}\frac{\partial\lambda}{\partial\phi_i}\right)-\Phi T(\mathbf{p})\right]\\&=\frac{1}{\Phi^2}\left[\Phi\phi_i\frac{\partial F_i(\beta_i,u_i)}{\partial\beta_i}\bigg|_{\beta_i=\phi_i}\right.\\&\qquad+\sum_{j=1}^{n}\beta_j(F_i(\beta_i,u_i)-F_j(\beta_j,u_j))\\&\qquad\left.-\lambda G(\lambda,t)\right]\quad i\in N\\&\frac{1}{\Phi^2}\left(\sum_{j=1}^{n}\beta_j(\alpha+g-F_j(\beta_j,u_j))-\lambda G(\lambda,t)\right)\quad i\in R_a\cup R_d\\&\frac{1}{\Phi^2}\left(\sum_{j=1}^{n}\beta_j(\alpha-F_j(\beta_j,u_j))-\lambda G(\lambda,t)\right)\quad i\in S\end{aligned}$$

Let us see the case $i\in S$ as an example. When communication speed is slow, i.e., the communication time is large, and sinks are lightly loaded, we may have $\alpha<F_j(\beta_j,u_j)$ and $G(\lambda)\gg 0$. It is $\partial T/\partial\phi_i<0$, in this case, so that the overall mean job response time have a chance to decrease as the job arrival rate at a sink increases.

5.2 The Individually Optimal Policy

We have in deriving the theorem 4.4,

$$\frac{\partial R(\mathbf{p})}{\partial t}=\frac{-B(\mathbf{p})(\partial G(\lambda,t)/\partial t)}{A(\mathbf{p})+B(\mathbf{p})+A(\mathbf{p})B(\mathbf{p})(\partial G(\lambda,t)/\partial\lambda)},$$

$$\frac{\partial\hat{R}(\mathbf{p})}{\partial t}=\frac{A(\mathbf{p})(\partial G(\lambda,t)/\partial t)}{A(\mathbf{p})+B(\mathbf{p})+A(\mathbf{p})B(\mathbf{p})(\partial G(\lambda,t)/\partial\lambda)}.$$

From eq. (45) and the above relations, we have

$$\begin{aligned}\frac{\partial T(\mathbf{p})}{\partial t}&=\frac{1}{\Phi}\left\{\sum_{i\in S}\phi_i\frac{\partial R(\mathbf{p})}{\partial t}+\sum_{R_a\cup R_d}\phi_i\frac{\partial\hat{R}(\mathbf{p})}{\partial t}\right\}\\&=\frac{1}{\Phi}\frac{\Sigma_{R_a\cup R_d}\phi_iA(\mathbf{p})-\Sigma_{i\in S}\phi_iB(\mathbf{p})}{A(\mathbf{p})+B(\mathbf{p})+A(\mathbf{p})B(\mathbf{p})(\partial G(\lambda,t)/\partial\lambda)}\\&\quad\times\frac{\partial G(\lambda,t)}{\partial t}.\end{aligned}$$

Note that if all sinks are congested, $A(\mathbf{p})=\Sigma_{i\in S}(\partial E_i(R,u_i)/\partial R)$ will be small, i.e. $A(\mathbf{p})\approx 0$. Thus, for example, if all sinks are nearly saturated or the arrival rates $\Sigma_{i\in S}\phi_i$ at sinks are high, we may observe an anomalous behavior of the equilibrium such that the overall mean job response time decreases even though the communication time increases.

Similarly from the theorem 4.4, and eq. (45), we have for $i\in S$

$$\begin{aligned}\frac{\partial T(\mathbf{p})}{\partial\phi_i}&=\frac{1}{\Phi^2}\frac{1}{A(\mathbf{p})+B(\mathbf{p})+A(\mathbf{p})B(\mathbf{p})(\partial G/\partial\lambda)}\\&\quad\times\left\{\Phi\left(\sum_{i\in S}\phi_i+\sum_{R_a\cup R_d}\phi_i+\sum_{i\in S}\phi_iB(\mathbf{p})\frac{\partial G}{\partial\lambda}\right)\right.\\&\quad-[\sum_{R_a\cup R_d}\phi_iG+\sum_{i\in N}\phi_i(F_i(\phi_i)-R(p))]\\&\quad\left.\times[A(\mathbf{p})+B(\mathbf{p})+A(\mathbf{p})B(\mathbf{p})(\partial G/\partial\lambda)]\right\}.\end{aligned}$$

Although the above relation is quite complicated, we may see that it has chances to be negative. Thus, in some cases, the overall mean job response time of the equilibrium may have chances to decrease as the job arrival rate at a sink increases.

6. Numerical Examination

We have examined numerically the effects of the system parameters in several examples of a distributed computer system that consists of four host computers (nodes) connected via a single channel. Each node is modeled as a central-server model as shown in Figure 2. Server 0 is a CPU that processes jobs according to the processor sharing discipline. Servers 1, 2, $\cdots$, d are I/O devices which process jobs according to FCFS. Let $p_{i,0}$ and $p_{i,j}$, $j=1,2\cdots,d$, be the probabilities that, after departing from the CPU, a job leaves node i or requests I/O service at device j, $j=1,2,\cdots,d$, respectively.

The expected node delay of a job in such a node model is given as

$$F_i(\beta_i)=\sum_{j=0}^{d}\frac{q_{i,j}}{\mu_{i,j}-q_{i,j}\beta_i},\tag{48}$$

where $q_{i,0}=1/p_{i,0}$ and $q_{i,j}=p_{i,j}/p_{i,0}$, and $\mu_{i,j}$ is the service rate of at server j, $j=0,1,\cdots,d$ of node i. We consider processor sharing M/G/1 model for the single channel communications network. The expected communication delay is given by

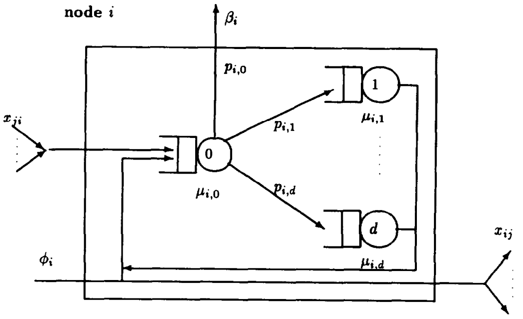

Fig. 2 Node model.

Table 1. Parameters of node models.

node	Processing rates of servers(jobs/sec)				Probabilities of a job leaving CPU			
	$\mu_{i,0}$	$\mu_{i,1}$	$\mu_{i,2}$	$\mu_{i,3}$	$p_{i,0}$	$p_{i,1}$	$p_{i,2}$	$p_{i,3}$
1	1500	450	450	450	0.1	0.3	0.3	0.3
2	100	45	45	-	0.1	0.45	0.45	-
3	100	30	30	30	0.1	0.3	0.3	0.3
4	120	54	54	-	0.1	0.45	0.45	-

$$G(\lambda)=\frac{t}{1-\lambda t}, \tag{49}$$

where t is the mean communication time for sending and receiving a job. We used the algorithm which is developed by Kim and Kameda [5] in our numerical calculation program.

We have observed (results not presented here), in most cases, that the results of the numerical examination agree with our intuition and that the overall mean job response time of the equilibrium is close to that of the optimum. We also observed that, in most cases, the individually optimal policy is more sensitive to the system parameters than the overall optimal policy. Table 1 shows an example of the set of the values of processing rates $\mu_{i,j}$ and the transition probabilities $p_{i,j}$. The results using Table 1 are given in Figs. 3, 4, and 5.

In Figures 3, 4, and 5, 'OOP' and 'IOP' denote the overall mean job response times of the overall and individually optimal policies, respectively. The solid line shows the overall mean job response time ($T(\beta)$) of the optimum under the overall optimal policy. The dotted line shows the overall mean job response time ($T(\beta)$) of the equilibrium under the individually optimal policy.

Figures 3 and 4 show how the overall mean job response times ($T(\beta)$) of the optimum and of the equilibrium vary as the communication time (t) changes. The values of ϕ_2, ϕ_3, and ϕ_4 are fixed to be 7, 7, and 7.5 (jobs/sec), respectively. In Fig. 3, ϕ_1 equals 80 (jobs/sec). From this figure, we can observe that the overall mean job response time of the equilibrium is close to that of the optimum. This is what we noted at the end of Section 3. The right-most end of each curve shows the case where all nodes are neutral, i.e., the case of no load balancing. In this figure, we see how static load balancing improves the mean response time in an example.

In Fig. 4, ϕ_1 equals 140 (jobs/sec). We can observe such an *anomalous* behavior that the overall mean job response time of the equilibrium decreases even though the communication time increases up to a certain value. This is what we noted in Section 5.2. Furthermore, we can even find such a seemingly extraordinary case that the overall mean job response time of the equilibrium is minimum only when all nodes are neutral (no load balancing).

Figure 5 shows how the overall mean job response times ($T(\beta)$) of the optimum and of the equilibrium vary as the job arrival rate of node 1 (ϕ_1) changes from 0 to 146 (jobs/sec). The values of ϕ_2, ϕ_3, ϕ_4 and t are fixed to be 9, 9, 11.5 (jobs/sec) and 0.3 (sec), respectively. In Fig. 5, we can observe that the overall mean job response time of the equilibrium is close to that of the optimum. Under both of the two policies, however, we can observe anomalous behaviors that the overall mean job response times of the optimum and of the equilibrium decrease even though the total external arrival rate increases up to a certain value. These are what we noted in Section 5.

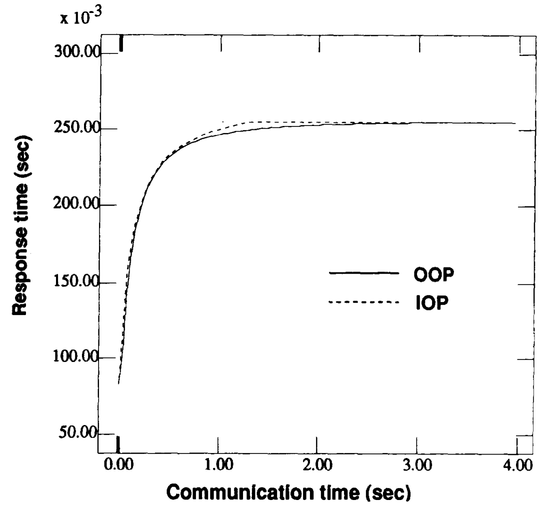

Fig. 3 The overall mean job response times ($T(\beta)$) when changing the communication time (t) under the overall and individually optimal policies in the case where $\phi_1=80$, $\phi_2=7$, $\phi_3=7$, and $\phi_4=7.5$.

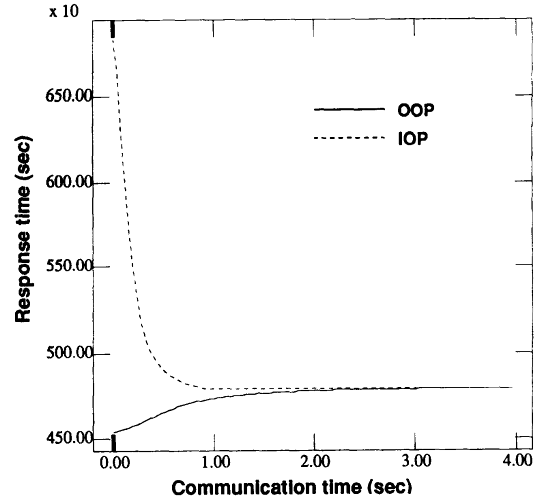

Fig. 4 The overall mean job response times ($T(\beta)$) when changing the communication time (t) under the overall and individually optimal policies in the case where $\phi_1=140$, $\phi_2=7$, $\phi_3=7$, and $\phi_4=7.5$.

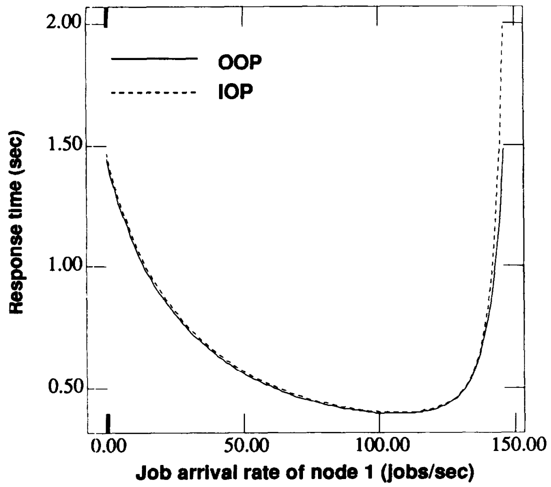

Fig. 5 The overall mean job response times ($T(\beta)$) when changing the job arrival rate of node 1 (ϕ_1) from 0 to 146 (jobs/sec) under the overall and individually optimal policy in the case where $t=0.3$ (sec) and $\phi_2=9$, $\phi_3=9$, $\phi_4=11.5$.

7. Conclusion

We studied two contrasting policies, the overall and individually optimal policies, for statistically balancing the load on a set of heterogeneous host computers connected by a communications network. We showed the conditions that the solution of the individually optimal policy satisfies. We studied the effects of the system parameters on the performance variables of the two policies. The main results of this paper can be summarized as follows:

- We found that the two policies have very similar characteristics even though they are of the nature entirely different from each other. We observed, in most numerical examples, that the overall mean job response time of the equilibrium is near to that of the optimum. We also observed that the two policies can be implemented in a similar way.
- We observed, however, an anomalous phenomenon, that there are cases where in the equilibrium, the overall mean job response time decreases even though the communication time increases. We can improve the overall mean job response time in this situation, by decreasing the ratio of jobs sent to sinks in order to reduce the congestion in sinks. However, it may degrade the mean response time of jobs in sources. It is necessary, therefore, to consider the trade-off between the mean response time of jobs in sources and the overall mean job response time.
- We also observed another anomalous phenomenon that there are cases where in the optimum and in the equilibrium, the overall mean job response time decreases even though the job arrival rates increase.
- From the above results, we may conclude that the overall mean job response time is not a performance measure perfect in all respects and that it is necessary to take account of other performance measures. This problem seems to be worth while to study.

One of the extensions to this work is the consideration of multiple class cases where different classes may have different processing requirements and node assignments. In the multiple class cases, neither the overall optimal policy nor the individually optimal policy may generally have a unique solution. Thus the parametric analysis may encounter many difficulties. In the parametric analysis of the multiple class cases, one of the most important tasks is to determine the meaningful performance variables which are uniquely determined.

Appendix A. Proof of Theorem 4.1

Let us define

$$\delta_i(X)=\begin{cases}1, & i\in X,\\ 0, & i\notin X.\end{cases}$$

Relation (26) itself is Theorem 5 of [12].

Lemma A.1 *For a given set of sinks S and for all i,*

$$\frac{\partial\lambda(\mathbf{p})}{\partial t}=C(\mathbf{p})\frac{\partial\alpha(\mathbf{p})}{\partial t}, \tag{A.1}$$

$$\frac{\partial\lambda(\mathbf{p})}{\partial u_i}=C(\mathbf{p})\frac{\partial\alpha(\mathbf{p})}{\partial u_i}+\delta_i(S)\frac{\partial e_i}{\partial u_i}, \tag{A.2}$$

$$\frac{\partial\lambda(\mathbf{p})}{\partial\phi_i}=C(\mathbf{p})\frac{\partial\alpha(\mathbf{p})}{\partial\phi_i}-\delta_i(S), \tag{A.3}$$

where

$$C(\mathbf{p})=\sum_{i\in S}\frac{\partial e_i(\alpha,u_i)}{\partial\alpha}\bigg|_{\alpha=\alpha(\mathbf{p})}. \tag{A.4}$$

Proof. Eq. (A.1) itself is eq. (17) of [12]. To derive eq. (A.2) we have from relation (9) and definition (24)

$$e_i(\alpha(\mathbf{p}),u_i)=\beta_i(\mathbf{p}),\quad i\in S. \tag{A.5}$$

Therefore we have

$$\frac{\partial\beta_i(\mathbf{p})}{\partial u_j}=\frac{\partial e_i(\alpha,u_i)}{\partial\alpha}\bigg|_{\alpha=\alpha(\mathbf{p})}\frac{\partial\alpha(\mathbf{p})}{\partial u_j}+\frac{\partial e_i(\alpha,u_i)}{\partial u_j}\bigg|_{\alpha=\alpha(\mathbf{p})}. \tag{A.6}$$

By noting that $\lambda(\mathbf{p})=\Sigma_{i\in S}(\beta_i(\mathbf{p})-\phi_i)$, we easily have eq. (A.2) from eq. (A.6).

We have (A.3) in the similar way as above. □

Lemma A.2 *For a given set of active sources R_a and for all i,*

$$\frac{\partial\lambda(\mathbf{p})}{\partial t}=-\frac{(\partial\alpha(\mathbf{p})/\partial t)+(\partial g(\lambda,t)/\partial t)}{(1/D(\mathbf{p}))+(\partial g(\lambda,t)/\partial\lambda)}, \tag{A.7}$$

$$\frac{\partial\lambda(p)}{\partial u_i}=-\frac{(\partial\alpha(\mathbf{p})/\partial u_i)+\delta_i(R_a)(\partial e_i(\alpha,u_i)/\partial u_i)}{(1/D(\mathbf{p}))+(\partial g(\lambda,t)/\partial\lambda)}, \tag{A.8}$$

$$\frac{\partial \lambda(\mathbf{p})}{\partial \phi_i} = -\frac{(\partial \alpha(\mathbf{p})/\partial \phi_i) - \delta_i(R_a)}{(1/D(\mathbf{p})) + (\partial g(\lambda, t)/\partial \lambda)}, \tag{A.9}$$

where

$$\hat{\alpha}(\mathbf{p}) = \alpha(\mathbf{p}) + g(\lambda(\mathbf{p}), t), \tag{A.10}$$

$$D(\mathbf{p}) = \sum_{i \in R_a} \left. \frac{\partial e_i(\hat{\alpha}, u_i)}{\partial \hat{\alpha}} \right|_{\hat{\alpha} = \hat{\alpha}(\mathbf{p})}. \tag{A.11}$$

Proof. Eq. (A.7) itself is eq. (20) of [12]. To derive eq. (A.8) we have from relation (7) by using definition (24)

$$\beta_i(\mathbf{p}) = e_i(\alpha(\mathbf{p}) + g(\lambda, t), u_i), \quad i \in R_a.$$

Then we have

$$\frac{\partial \beta_i(\mathbf{p})}{\partial u_j} = \frac{\partial e_i(\hat{\alpha}, u_i)}{\partial \hat{\alpha}} \left(\frac{\partial \alpha}{\partial u_j} + \frac{\partial g}{\partial \lambda} \frac{\partial \lambda}{\partial u_j} \right) + \frac{\partial e_i(\hat{\alpha}, u_i)}{\partial u_j}. \tag{A.12}$$

Note that $\lambda(\mathbf{p}) = \Sigma_{i \in R_a}(\phi_i - \beta_i(\mathbf{p})) + \Sigma_{i \in R_d} \phi_i$. From these, we have eq. (A.8).

We have (A.9) in the similar way as above. □

By combining each equation in Lemma A.1 with the corresponding equation in Lemma A.2, we can derive equations on $\partial\alpha(\mathbf{p})/\partial u_i$, and $\partial\alpha(\mathbf{p})/\partial\phi_i$. Then we can derive relations (27)-(28).

References

1. Bell, C. E. and Stidham, S. Individually versus social optimization in the allocation of customers to alternate servers. *Management Sci.* **29**, 7 (July 1983), 831-839.

2. Dafermos, S. The traffic assignment problem for multiclass-user transportation networks, *Transportation Science 6* (1972), 73-87.

3. Eager, D. L., Lazowska, E. D. and Zahorjan, J. Adaptive load sharing in homogeneous distributed systems. *IEEE Trans. Softw. Eng.* **12**, 5 (May 1986), 662-675.

4. Eager, D. L., Lazowska, E. D. and Zahorjan, J. A comparison of receiver-initiated and sender-initiated adaptive load sharing. *Performance Evaluation 6* (1986), 53-68.

5. Kim, C. and Kameda, H. An algorithm for optimal static load balancing in distributed computer systems, *IEEE Trans. Comput.* (to appear).

6. Kim, C. and Kameda, H. Optimal static load balancing of multiclass jobs in a distributed computer system, *Proc. 10th Int. Conf. on Distributed Computing Systems* (Paris, 1990), IEEE Computer Society Press, Los Alamitos, CA, 562-569.

7. Lippman, S. A. and Stidham, S. Individual versus social optimization in exponential congestion systems. *Oper. Res.* **25**, 2 (1977), 233-247.

8. Magnanti, T. L. Models and algorithms for predicting urban Traffic equilibria, *Transportation Planning Models* (M. Florian, ed.) Elsevier Science Publishers B. V., North-Holland (1984), 153-185.

9. Mirchandaney, R., Towsley, D. and Stankovic, J. A. Analysis of the effects of delays on load sharing, *IEEE Trans. Comput.* **38**, 11 (Nov. 1989), 1513-1525.

10. Mirchandaney, R., Towsley, D. and Stankovic, J. A. Adaptive load sharing in heterogeneous distributed systems, *Journal of Parallel and Distributed Computing* 9 (1990), 331-346.

11. Stidham, S. Socially and individually optimal control of arrival to a GI/M/1 queue. *Management Sci. 24* (1978), 1598-1610.

12. Tantawi, A. N. and Towsley, D. A general model for optimal static load balancing in star network configurations, *Proc. of PERFORMANCE' 84* (Paris, Dec. 19-21, 1984) North-Holland, New York, 277-291.

13. Tantawi, A. N. and Towsley, D. Optimal static load balancing in distributed computer systems, *J. ACM* **32**, 2 (April 1985), 445-465.

(Received April 1, 1991; revised August 29, 1991)