Definition of a Robustness Metric for Resource Allocation

Shoukat Ali[†], Anthony A. Maciejewski[‡], Howard Jay Siegel[‡§], and Jong-Kook Kim[†]

[†]Purdue University
School of Electrical and
Computer Engineering
West Lafayette, IN 47907-1285 USA
{alis, jongkook}@purdue.edu

Colorado State University
[‡]Department of Electrical and
Computer Engineering
[§]Department of Computer Science
Fort Collins, CO 80523-1373 USA
{hj, aam}@colostate.edu

Abstract

Parallel and distributed systems may operate in an environment that undergoes unpredictable changes causing certain system performance features to degrade. Such systems need robustness to guarantee limited degradation despite fluctuations in the behavior of its component parts or environment. This research investigates the robustness of an allocation of resources to tasks in parallel and distributed systems. The main contributions of this paper are (1) a mathematical description of a metric for the robustness of a resource allocation with respect to desired system performance features against perturbations in system and environmental conditions, and (2) a procedure for deriving a robustness metric for an arbitrary system. For illustration, this procedure is employed to derive robustness metrics for two example distributed systems. Such a metric can help researchers evaluate a given resource allocation for robustness against uncertainties in specified perturbation parameters.

1. Introduction

Parallel and distributed systems may operate in an environment where certain system performance features degrade due to unpredictable circumstances, such as sudden machine failures, higher than expected system load, or inaccuracies in the estimation of system parameters (e.g., [4, 5, 8, 10, 12, 13, 19, 17, 16, 18, 20, 22, 24]). An important question then arises: given a system design, what extent of departure from the assumed circumstances will cause a performance feature to be unacceptably degraded? That is, how robust is the system? Before answering this question one needs to clearly define robustness. Robustness has been defined in different ways by different researchers. According to [18], robustness is the degree to which a system can function correctly in the presence of inputs different from those assumed. In a more general sense, [13] states that a robust system continues to operate correctly across a wide range of operational conditions. Robustness, according to [8] and [17], guarantees the maintenance of certain desired system characteristics despite fluctuations in the behavior of its component parts or its environment. The concept of robustness, as used in this research, is similar to that in [8] and [17]. Like [17], this work emphasizes that robustness should be defined for a given set of system features, with a given set of perturbations applied to the system. This research investigates the robustness of resource allocation in parallel and distributed systems, and accordingly customizes the definition of robustness.

This research was supported by the DARPA/ITO Quorum Program through the Office of Naval Research under Grant No. N00014-00-1-0599, and by the Colorado State University George T. Abell Endowment. Some of the equipment used was donated by Intel and Microsoft.

Parallel and distributed computing is the coordinated use of different types of machines, networks, and interfaces to meet the requirements of widely varying application mixtures and to maximize the system performance or cost-effectiveness. An important research problem is how to determine a mapping (matching of applications to resources and ordering their execution (e.g., [7, 21, 25])) so as to maximize robustness of desired system features against inaccuracies in estimated system parameters and changes in the environment. This research addresses the design of a robustness metric for mappings.

A mapping is defined to be robust *with respect to specified system performance features against perturbations in specified system parameters* if degradation in these features is limited when the perturbations occur. For example, if a mapping has been declared to be robust with respect to satisfying a throughput requirement against perturbations in the system load, then the system, configured under that mapping, should continue to operate without a throughput violation when the system load increases. The immediate question is: what is the *degree* of robustness? That is, for the example given above, how much can the system load increase before a throughput violation occurs? This research addresses this question, and others related to it, by formulating the mathematical description of a metric that evaluates the robustness of a mapping with respect to certain system performance features against perturbations in system components and environmental conditions. In addition, this work outlines a four-step procedure for deriving a robustness metric for an arbitrary system. For illustration, the four-step procedure is employed to derive robustness metrics for two example distributed systems. The robustness metric and the four-step procedure for its derivation are the main contributions of this paper. Although measures of robustness have been studied in the literature (e.g., [5, 10, 9, 18, 19, 15, 20, 22]), those measures were developed for specific systems (see [1] for a discussion of the related work). Unlike those efforts, this paper presents a general mathematical formulation of a robustness metric that could be applied to a variety of parallel and distributed systems by following the four-step derivation procedure.

The rest of the paper is organized as follows. Section 2 describes the four-step derivation procedure mentioned above. It also defines a generalized robustness metric. Derivations of this metric for two example parallel and distributed systems are given in Section 3. Section 4 presents some experiments that highlight the usefulness of the robustness metric. Section 5 concludes the paper. A glossary of the notation used in this paper is given in Table 1.

Table 1. Glossary of notation.

Φ	the set of all performance features
ϕ_i	the i-th element in Φ
$\left\langle \beta_i^{\min}, \beta_i^{\max} \right\rangle$	a tuple that gives the bounds of the tolerable variation in ϕ_i
Π	the set of all perturbation parameters
$\boldsymbol{\pi}_j$	the j-th element in Π
$n_{\boldsymbol{\pi}_j}$	the number of elements in $\boldsymbol{\pi}_j$
μ	a mapping
$r_\mu(\phi_i,\ \boldsymbol{\pi}_j)$	the robustness radius of mapping μ with respect to ϕ_i against $\boldsymbol{\pi}_j$
$\rho_\mu(\Phi,\ \boldsymbol{\pi}_j)$	the robustness of mapping μ with respect to set Φ against $\boldsymbol{\pi}_j$
$\mathcal{A}$	the set of applications
$\mathcal{M}$	the set of machines

2. Generalized Robustness Metric

The key contribution of the work presented here is the design of a general mathematical methodology for deriving the range of uncertainty in system parameters within which a desired level of quality of service (QoS) can be guaranteed. Central to achieving this goal is the development of a general, mathematically precise, definition of robustness that can be applied in a wide variety of scenarios, for a wide variety of application-specific measures of performance, and a wide variety of system parameters, whose behavior is uncertain, but whose values will affect system performance. This paper presents a general four-step procedure for deriving such a robustness metric for any desired computing environment. The procedure is referred to in this paper as the FePIA procedure, where the abbreviation stands for identifying the performance features, the perturbation parameters, the impact of perturbation parameters on performance features, and the analysis to determine the robustness. Specific examples illustrating the application of the FePIA procedure to sample systems are given in the next section.

Each step of the FePIA procedure is now described.

1) Describe quantitatively the requirement that makes the system robust. Based on this *robustness requirement*, determine the system performance features that should be limited in variation to ensure that the robustness requirement is met. For example, the robustness requirement could be that the makespan[1] of the mapping should not increase more than 30% beyond its predicted value, where the predicted value is the value expected in the absence of uncertainty. In that case, the system performance features that should be limited in variation are the finishing times for all machines in the system. Mathematically, let Φ be the set of such system features. For each element $\phi_i \in \Phi$, quantitatively describe the tolerable variation in ϕ_i. Let $\left\langle \beta_i^{\min}, \beta_i^{\max} \right\rangle$ be a tuple that gives the bounds of the tolerable variation in the system feature ϕ_i. For the makespan example, ϕ_i is the time the i-th machine finishes its assigned applications, and its corresponding $\left\langle \beta_i^{\min}, \beta_i^{\max} \right\rangle$ tuple could be $\langle 0,\ 1.3 \times \text{(predicted makespan value)} \rangle$.

2) Determine the system and environment perturbation parameters against which the robustness of the system features described in item 1 is sought. These are called the perturbation parameters (these are similar to hazards in [5]). For the makespan example above, the resource allocation (and its associated predicted makespan) was based on the estimated application execution times. It is desired that

[1]Makespan of a set of applications is the completion time for the entire set.

the makespan be robust (stay within 130% of its estimated value) with respect to uncertainties in these estimated execution times. Mathematically, let Π be the set of such system and environment parameters. It is assumed that the elements of Π are vectors. Let $\boldsymbol{\pi}_j$ be the j-th element of Π. For the makespan example, $\boldsymbol{\pi}_j$ could be the vector composed of the actual application execution times, i.e., the i-th element of $\boldsymbol{\pi}_j$ is the actual execution time of the i-th application on the machine it was assigned.

3) Identify the impact of the perturbation parameters in item 2 on the system performance features in item 1. For the makespan example, the sum of the actual execution times for all of the applications assigned a given machine is the time when that machine completes its applications. Mathematically, for every $\phi_i \in \Phi$, determine the relationship $\phi_i = f_{ij}(\boldsymbol{\pi}_j)$, if any, that relates ϕ_i to $\boldsymbol{\pi}_j$. In this expression, f_{ij} is a function that maps $\boldsymbol{\pi}_j$ to ϕ_i. For the makespan example, ϕ_i is the finishing time for machine m_i, and f_{ij} would be the sum of execution times for applications assigned to machine m_i. Note that this expression assumes that each $\boldsymbol{\pi}_j \in \Pi$ affects a given ϕ_i independently. The case where multiple perturbation parameters can affect a given ϕ_i simultaneously is discussed in [1]. The rest of this discussion will be developed assuming only one element in Π.

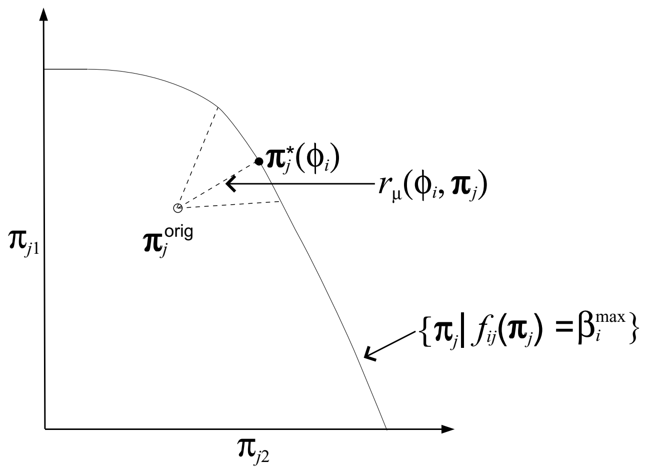
4) The last step is to determine the smallest collective variation in the values of perturbation parameters identified in step 2 that will cause any of the performance features identified in step 1 to violate the robustness requirement. Mathematically, for every $\phi_i \in \Phi$, determine the *boundary values of* $\boldsymbol{\pi}_j$, i.e., the values satisfying the boundary relationships $f_{ij}(\boldsymbol{\pi}_j) = \beta_i^{\min}$ and $f_{ij}(\boldsymbol{\pi}_j) = \beta_i^{\max}$. (If $\boldsymbol{\pi}_j$ is a discrete variable then the boundary values correspond to the closest values that bracket each boundary relationship.) These relationships separate the region of robust operation from that of non-robust operation. Find the smallest perturbation in $\boldsymbol{\pi}_j$ that causes any $\phi_i \in \Phi$ to exceed the bounds $\langle \beta_i^{\min}, \beta_i^{\max} \rangle$ imposed on it by the robustness requirement.

Specifically, let $\boldsymbol{\pi}_j^{\text{orig}}$ be the value of $\boldsymbol{\pi}_j$ at which the system is originally assumed to operate. However, due to inaccuracies in the estimated parameters and/or changes in the environment, the value of the variable $\boldsymbol{\pi}_j$ might differ from its assumed value. This change in $\boldsymbol{\pi}_j$ can occur in different "directions" depending on the relative differences in its individual components. Assuming that no information is available about the relative differences, all values of $\boldsymbol{\pi}_j$ are possible. Figure 1 illustrates this concept for a single feature, ϕ_i, and a 2-element perturbation vector $\boldsymbol{\pi}_j \in \mathbf{R}^2$. The curve shown in Figure 1 plots the set of boundary points $\{\boldsymbol{\pi}_j |\ f_{ij}(\boldsymbol{\pi}_j) = \beta_i^{\max}\}$ for a mapping μ. For this figure, the set of boundary points $\{\boldsymbol{\pi}_j |\ f_{ij}(\boldsymbol{\pi}_j) = \beta_i^{\min}\}$ is given by the points on the π_{j1}-axis and π_{j2}-axis.

The region enclosed by the axes and the curve gives the values of $\boldsymbol{\pi}_j$ for which the system is robust with respect to ϕ_i. For a vector $\mathbf{x} = [x_1\, x_2\, \cdots\, x_n]^{\mathrm{T}}$, let $\|\mathbf{x}\|_2$ be the ℓ_2-norm (Euclidean norm) of the vector, and be defined by $\sqrt{\sum_{r=1}^{n} x_r^2}$. The point on the curve marked as $\boldsymbol{\pi}_j^{\star}(\phi_i)$ has the feature that the Euclidean distance from $\boldsymbol{\pi}_j^{\text{orig}}$ to $\boldsymbol{\pi}_j^{\star}(\phi_i)$, $\|\boldsymbol{\pi}_j^{\star}(\phi_i) - \boldsymbol{\pi}_j^{\text{orig}}\|_2$, is the smallest over all such distances from $\boldsymbol{\pi}_j^{\text{orig}}$ to a point on the curve. An important interpretation of $\boldsymbol{\pi}_j^{\star}(\phi_i)$ is that the value $\|\boldsymbol{\pi}_j^{\star}(\phi_i) - \boldsymbol{\pi}_j^{\text{orig}}\|_2$ gives the largest Euclidean distance that the variable $\boldsymbol{\pi}_j$ can change in *any* direction from the assumed value of $\boldsymbol{\pi}_j^{\text{orig}}$ without the performance feature ϕ_i exceeding the tolerable variation. Let the distance $\|\boldsymbol{\pi}_j^{\star}(\phi_i) - \boldsymbol{\pi}_j^{\text{orig}}\|_2$ be called the robustness radius, $r_\mu(\phi_i,\ \boldsymbol{\pi}_j)$, of ϕ_i against $\boldsymbol{\pi}_j$. Mathematically,

$$r_\mu(\phi_i,\ \boldsymbol{\pi}_j) = \min_{\substack{\boldsymbol{\pi}_j:\,(f_{ij}(\boldsymbol{\pi}_j)=\beta_i^{\max})\vee \\ (f_{ij}(\boldsymbol{\pi}_j)=\beta_i^{\min})}} \|\boldsymbol{\pi}_j - \boldsymbol{\pi}_j^{\text{orig}}\|_2. \quad (1)$$

This work defines $r_\mu(\phi_i,\ \boldsymbol{\pi}_j)$ to be the robustness metric for the robustness of mapping μ with respect to performance feature ϕ_i against the perturbation parameter $\boldsymbol{\pi}_j$.



$\boldsymbol{\pi}_j^{\star}(\phi_i)$
$r_\mu(\phi_i, \boldsymbol{\pi}_j)$
$\boldsymbol{\pi}_j^{\text{orig}}$
π_{j1}
$\{\pi_j | f_{ij}(\pi_j) = \beta_i^{\max}\}$
π_{j2}

Figure 1. The possible directions of increase of the perturbation parameter π_j, and the direction of the smallest increase. The curve plots the set of points, $\{\boldsymbol{\pi}_j |\ f_{ij}(\boldsymbol{\pi}_j) = \beta_i^{\max}\}$. The set of boundary points, $\{\boldsymbol{\pi}_j |\ f_{ij}(\boldsymbol{\pi}_j) = \beta_i^{\min}\}$ is given by the points on the π_{j1}-axis and π_{j2}-axis.

The metric definition can be extended easily for all $\phi_i \in \Phi$. It is simply the minimum of all robustness radii. Mathematically, let $\rho_\mu(\Phi,\ \boldsymbol{\pi}_j)$ be the *robustness of mapping* μ

with respect to the performance feature set Φ *against the perturbation parameter* $\boldsymbol{\pi}_j$. Then,

$$\rho_\mu(\Phi,\ \boldsymbol{\pi}_j) = \min_{\phi_i \in\ \Phi} \left(r_\mu(\phi_i,\ \boldsymbol{\pi}_j)\right) \tag{2}$$

3. Example Derivations

3.1. Independent Application Allocation

The first example derivation of the robustness metric is for a system that maps a set of independent applications on a set of machines [7]. In this system, it is required that the makespan (defined as the completion time for the entire set of applications) be robust against errors in application execution time estimates.

A brief description of the system model is now given. The applications are assumed to be independent, i.e., no communications between the applications are needed. The set $\underline{\mathcal{A}}$ of applications is to be mapped on a set $\underline{\mathcal{M}}$ of machines so as to minimize the makespan. Each machine executes a single application at a time (i.e., no multi-tasking), in the order in which the applications are assigned. Let $\underline{C_{ij}}$ be the estimated time to compute (ETC) for application $\underline{a_i}$ on machine $\underline{m_j}$. It is assumed that $\underline{C_{ij}}$ values are known for all i, j, and a mapping μ is determined using the ETC values. In addition, let $\underline{F_j}$ be the time at which m_j finishes executing all of the applications mapped on it.

Assume that unknown inaccuracies in the ETC values are expected, requiring that the mapping μ be robust against them. More specifically, it is required that, for a given mapping, its actual makespan value $\underline{M}$ (calculated considering the effects of ETC errors) may be no more than τ times its predicted value, $\underline{M^{\text{orig}}}$. The predicted value of the makespan is the value calculated assuming the ETC values are accurate. Following step 1 of the FePIA procedure in Section 2, the system performance features that should be limited in variation to ensure the makespan robustness are the finishing times of the machines. That is,

$$\Phi = \left\{F_j \mid 1 \le j \le |\mathcal{M}|\right\}. \tag{3}$$

According to step 2 of the FePIA procedure, the perturbation parameter needs to be defined. Let $\underline{C_i^{\text{orig}}}$ be the ETC value for application a_i on the machine where it is mapped. Let $\underline{C_i}$ be equal to the actual computation time value (C_i^{orig} plus the estimation error). In addition, let $\boldsymbol{C}$ be the vector of the C_i values, such that $\boldsymbol{C} = [C_1\ \ C_2\ \cdots\ C_{|\mathcal{A}|}]$. Similarly, $\underline{\boldsymbol{C}^{\text{orig}}} = [C_1^{\text{orig}}\ \ C_2^{\text{orig}}\ \cdots\ C_{|\mathcal{A}|}^{\text{orig}}]$. The vector $\boldsymbol{C}$ is the perturbation parameter for this analysis.

In accordance with step 3 of the FePIA procedure, F_j has to be expressed as a function of $\boldsymbol{C}$. To that end,

$$F_j(\boldsymbol{C}) = \sum_{i:\, a_i \text{ is mapped to } m_j} C_i. \tag{4}$$

Note that the finishing time of a given machine depends only on the actual execution times of the applications mapped to that machine, and is independent of the finishing times of the other machines. Following step 4 of the FePIA procedure, the set of boundary relationships corresponding to Equation 3 is given by $\left\{F_j(\boldsymbol{C}) = \tau M^{\text{orig}} \mid 1 \le j \le |\mathcal{M}|\right\}$.

For a two-application system, $\boldsymbol{C}$ corresponds to $\boldsymbol{\pi}_j$ in Figure 1. Similarly, C_1 and C_2 correspond to π_{j1} and π_{j2}, respectively. The terms $\boldsymbol{C}^{\text{orig}}$, $F_j(\boldsymbol{C})$, and τM^{orig} correspond to $\boldsymbol{\pi}_j^{\text{orig}}$, $f_{ij}(\boldsymbol{\pi}_j)$, and $\beta_i^{\max}$, respectively. The boundary relationship '$F_j(\boldsymbol{C}) = \tau M^{\text{orig}}$' corresponds to the boundary relationship '$f_{ij}(\boldsymbol{\pi}_j) = \beta_i^{\max}$.'

From Equation 1, the robustness radius of F_j against $\boldsymbol{C}$ is given by

$$r_\mu(F_j,\ \boldsymbol{C}) = \min_{\boldsymbol{C}:\ F_j(\boldsymbol{C}) = \tau M^{\text{orig}}} \|\boldsymbol{C} - \boldsymbol{C}^{\text{orig}}\|_2 \tag{5}$$

That is, if the Euclidean distance between any vector of the actual execution times and the vector of the estimated execution times is no larger than $r_\mu(F_j,\ \boldsymbol{C})$, then the finishing time of machine m_j will be at most τ times the predicted makespan value.

Note that the right hand side in Equation 5 can be interpreted as the perpendicular distance from the point $\boldsymbol{C}^{\text{orig}}$ to the hyperplane described by the equation $\tau M^{\text{orig}} - F_j(\boldsymbol{C}) = 0$. Using the point-to-plane distance formula [23], Equation 5 reduces to

$$r_\mu(F_j,\ \boldsymbol{C}) = \frac{\tau M^{\text{orig}} - F_j(\boldsymbol{C}^{\text{orig}})}{\sqrt{\text{number of applications mapped to } m_j}} \tag{6}$$

The robustness metric, from Equation 2, is

$$\rho_\mu(\Phi,\ \boldsymbol{C}) = \min_{F_j\ \in\ \Phi} r_\mu(F_j,\ \boldsymbol{C}) \tag{7}$$

That is, if the Euclidean distance between any vector of the actual execution times and the vector of the estimated execution times is no larger than $\rho_\mu(\Phi,\ \boldsymbol{C})$, then the actual makespan will be at most τ times the predicted makespan value.

Two observations can be made with respect to Equation 7. (These observations are specific to this system only, and do not necessarily apply to the other example system discussed later.) Let $\underline{\boldsymbol{C}^\star}$ be the value of $\boldsymbol{C}$ that minimizes $\|\boldsymbol{C} - \boldsymbol{C}^{\text{orig}}\|_2$.
(1) At the point $\boldsymbol{C}^\star$, the actual execution times for all applications are the same as the estimated times, except for the applications mapped on the machine that finishes last at $\boldsymbol{C}^\star$. This is because the finishing times are independent of each other, *and* the minimization is constrained only by the finishing time of one machine – the machine that finishes last.

(2) At the point $\boldsymbol{C}^\star$, the ETC errors for all applications mapped on the machine that finishes last are the same. This is because the weight of each such application towards determining the finishing time is the same (Equation 4).

Note that the calculation of the robustness metric above did not make any assumptions about the distribution of the estimation errors. In addition, note that $\rho_\mu(\Phi,\ \boldsymbol{C})$ has the units of $\boldsymbol{C}$, namely time.

3.2. The HiPer-D System

The second example derivation of the robustness metric is for a HiPer-D [11, 14] like system that maps a set of continuously executing, communicating applications to a set of machines. It is required that the system be robust with respect to certain QoS attributes against unforeseen increases in the "system load."

The HiPer-D system model used here was developed in [2], and is summarized here for reference. The system consists of heterogeneous sets of sensors, applications, machines, and actuators. Each machine is capable of multitasking, executing the applications mapped to it in a round robin fashion. Similarly, a given network link is multitasked among all data transfers using that link. Each sensor produces data periodically at a certain rate, and the resulting data streams are input into applications. The applications process the data and send the output to other applications or to actuators. The applications and the data transfers between them are modelled with a directed acyclic graph, shown in Figure 2. The figure also shows a number of *paths*

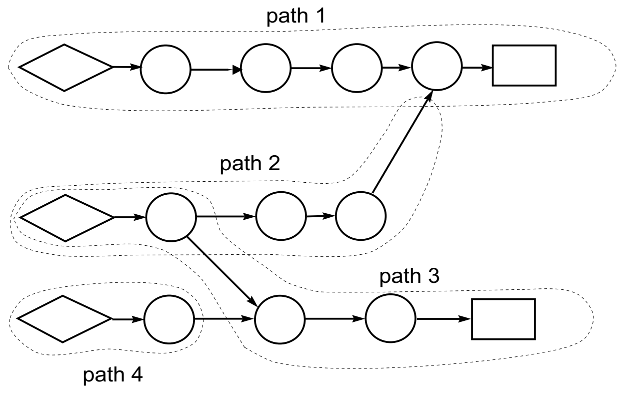


Figure 2. The DAG model for the applications (circles) and data transfers (arrows). The diamonds and rectangles denote sensors and actuators, respectively. The dashed lines enclose each path formed by the applications.

(enclosed by dashed lines) formed by the applications. A path is a chain of producer-consumer pairs that starts at a sensor (the driving sensor) and ends at an actuator (if it is a "trigger path") or at a multiple-input application (if it is an "update path"). Let $\mathcal{P}$ be the set of all paths, and $\mathcal{P}_k$ be the list of applications that comprise the k-th path. Note that an application may be present in multiple paths. As in Subsection 3.1, $\mathcal{A}$ is the set of applications.

The sensors constitute the interface of the system to the external world. Let the maximum periodic data output rate from a given sensor be called its output data rate. The minimum throughput constraint states that the computation or communication time of any application in $\mathcal{P}_k$ is required to be no larger than the reciprocal of the output data rate of the driving sensor for $\mathcal{P}_k$. For application $a_i \in \mathcal{P}_k$, let $R(a_i)$ be set to the output data rate of the driving sensor for $\mathcal{P}_k$. In addition, let T_{ij}^{c} be the computation time for application a_i mapped to machine m_j. Also, let T_{ip}^{n} be the time to send data from application a_i to application a_p. Because this analysis is being carried out for a specific mapping, the machine where a given application is mapped is known. It is assumed that a_i is mapped to m_j, and the machine subscript for T_{ij}^{c} is omitted in the ensuing analysis for clarity unless the intent is to show the relationship between execution times of a_i at various possible machines.

The maximum end-to-end latency constraint states that, for a given path $\mathcal{P}_k$, the time taken between the instant the driving sensor outputs a data set until the instant the actuator or the multiple-input application fed by the path receives the result of the computation on that data set must be no greater than a given value, $L_k^{\max}$. Let L_k be the actual (as opposed to the maximum allowed) value of the end-to-end latency for $\mathcal{P}_k$. The quantity L_k can be found by adding the computation and communication times for all applications in $\mathcal{P}_k$ (including any sensor or actuator communications). Let $\mathcal{D}(a_i)$ be the set of successor applications of a_i. Then,

$$L_k = \sum_{\substack{i:\ a_i \in \mathcal{P}_k \\ p:\ (a_p \in \mathcal{P}_k) \wedge (a_p \in \mathcal{D}(a_i))}} \left[T_i^{\mathrm{c}} + T_{ip}^{\mathrm{n}} \right]. \qquad (8)$$

It is desired that a given mapping μ of the system be robust with respect to the satisfaction of two QoS attributes: the latency and throughput constraints. Following step 1 of the FePIA procedure in Section 2, the system performance features that should be limited in variation are the latency values for the paths and the computation and communication time values for the applications. The set Φ is given by

$$\Phi = \{T_i^{\mathrm{c}} \mid 1 \le i \le |\mathcal{A}|\} \bigcup \{T_{ip}^{\mathrm{n}} \mid (1 \le i \le |\mathcal{A}|) \wedge (p, a_p \in \mathcal{D}(a_i))\} \bigcup \{L_k \mid 1 \le k \le |\mathcal{P}|\} \qquad (9)$$

This system is expected to operate under uncertain outputs from the sensors, requiring that the mapping μ be robust against unpredictable increases in the sensor outputs. Let $\underline{\lambda_z}$ be the output from the z-th sensor in the set of sensors, and be defined as the number of objects present in the most recent data set from that sensor. The system workload, $\boldsymbol{\lambda}$, is the vector composed of the load values from all sensors. Let $\boldsymbol{\lambda}^{\text{orig}}$ be the initial value of $\boldsymbol{\lambda}$, and λ_i^{orig} be the initial value of the i-th member of $\boldsymbol{\lambda}^{\text{orig}}$. Following step 2, the perturbation parameter $\boldsymbol{\pi}_j$ is identified to be $\boldsymbol{\lambda}$.

Step 3 of the FePIA procedure requires that the impact of $\boldsymbol{\lambda}$ on each of the system performance features be identified. The computation times of different applications (and the communication times of different data transfers) are likely to be of different complexities with respect to $\boldsymbol{\lambda}$. Assume that the dependence of T_i^{c} and T_{ip}^{n} on $\boldsymbol{\lambda}$ is known (or can be estimated) for all i, p. Given that, T_i^{c} and T_{ip}^{n} can be re-expressed as functions of $\boldsymbol{\lambda}$ as $T_i^{\text{c}}(\boldsymbol{\lambda})$ and $T_{ip}^{\text{n}}(\boldsymbol{\lambda})$, respectively. Then Equation 8 can be used to express L_k as a function of $\boldsymbol{\lambda}$.

Following step 4 of the FePIA procedure, the set of boundary relationships corresponding to Equation 9 is given by

$$\{T_i^{\text{c}}(\boldsymbol{\lambda}) = 1/R(a_i) \mid 1 \le i \le |\mathcal{A}|\} \bigcup$$
$$\left\{T_{ip}^{\text{n}}(\boldsymbol{\lambda}) = 1/R(a_i) \mid (1 \le i \le |\mathcal{A}|) \wedge (p, a_p \in \mathcal{D}(a_i))\right\} \bigcup$$
$$\left\{L_k(\boldsymbol{\lambda}) = L_k^{\max} \mid 1 \le k \le |\mathcal{P}|\right\}.$$

Then, using Equation 1, one can find, for each $\phi_i \in \Phi$, the robustness radius, $r_\mu(\phi_i,\ \boldsymbol{\lambda})$. Specifically,

$$r_\mu(\phi_i,\ \boldsymbol{\lambda}) =$$

$$\begin{cases} \min\limits_{\boldsymbol{\lambda}:\ T_x^{\text{c}}(\boldsymbol{\lambda})=1/R(a_x)} \|\boldsymbol{\lambda} - \boldsymbol{\lambda}^{\text{orig}}\|_2 & \text{if } \phi_i = T_x^{\text{c}} \quad \text{(10a)} \\ \min\limits_{\boldsymbol{\lambda}:\ T_{xy}^{\text{n}}(\boldsymbol{\lambda})=1/R(a_x)} \|\boldsymbol{\lambda} - \boldsymbol{\lambda}^{\text{orig}}\|_2 & \text{if } \phi_i = T_{xy}^{\text{n}} \quad \text{(10b)} \\ \min\limits_{\boldsymbol{\lambda}:\ L_k(\boldsymbol{\lambda})=L_k^{\max}} \|\boldsymbol{\lambda} - \boldsymbol{\lambda}^{\text{orig}}\|_2 & \text{if } \phi_i = L_k \quad \text{(10c)} \end{cases}$$

The robustness radius in Equation 10a is the largest increase (Euclidean distance) in load in any direction (i.e., for any combination of sensor load values) from the assumed value that does not cause a throughput violation for the computation of application a_x. The robustness radii in Equations 10b and 10c are the similar values for the communications of application a_x and the latency of path $\mathcal{P}_k$, respectively. The robustness metric, from Equation 2, is given by

$$\rho_\mu(\Phi,\ \boldsymbol{\lambda}) = \min_{\phi_i \in\ \Phi} \left(r_\mu(\phi_i,\ \boldsymbol{\lambda})\right). \quad (11)$$

For this system, $\rho_\mu(\Phi,\ \boldsymbol{\lambda})$ is the largest increase in load in any direction from the assumed value that does not cause a latency or throughput violation for any application or path. Note that $\rho_\mu(\Phi,\ \boldsymbol{\lambda})$ has the units of $\boldsymbol{\lambda}$, namely objects per data set. In addition, note that although $\boldsymbol{\lambda}$ is a discrete variable, it has been treated as a continuous variable in this section because the number of possible discrete values $\boldsymbol{\lambda}$ can take is infinite. However, because $\rho_\mu(\Phi,\ \boldsymbol{\lambda})$ should not have fractional values, one can take the floor of the right hand side in Equation 11. A different method for handling a discrete perturbation parameter is discussed in [1].

This research assumes that the optimization problems given in Equations 10a, 10b, and 10c can be solved to find the respective global minima. Note that an optimization problem of the form

$$\min_x f(x), \text{ subject to the constraint } g(x) = 0,$$

where $f(x)$ and $g(x)$ are both convex functions, can be easily solved to find the global minimum [6]. Because all norms are convex functions, the optimization problem posed in Equation 11 reduces to a convex optimization problem if $T_x^{\text{c}}(\boldsymbol{\lambda})$ and $T_{xy}^{\text{n}}(\boldsymbol{\lambda})$ are convex functions. Many commonly encountered complexity functions are convex. (A notable exception is $\log x$.) For example, all of the following functions are convex over the domain of positive real numbers ($x > 0$): e^{px} for $p \in \mathbf{R}$; x^p for $p \ge 1$; and $x \log x$. In addition, positive multiples of convex functions and sums of convex functions are also convex functions. Note that if the $T_x^{\text{c}}(\boldsymbol{\lambda})$ and $T_{xy}^{\text{n}}(\boldsymbol{\lambda})$ functions are not convex, then it is assumed that heuristic techniques can be used to find near-optimal solutions.

4. Experiments

4.1. Overview

The experiments in this section seek to establish the utility of the robustness metric in distinguishing between mappings that perform similarly in terms of a commonly used metric, such as makespan. Two different systems were considered: the independent task allocation system discussed in Subsection 3.1 and the HiPer-D system outlined in Subsection 3.2. Experiments were performed for a system with five machines and 20 applications. A total of 1000 mappings were generated by assigning a randomly chosen machine to each application, and then each mapping was evaluated with the robustness metric and the commonly used metric.

4.2. Independent Application Allocation

For the system in Subsection 3.1, the ETC values were generated by sampling a Gamma distribution. The mean was arbitrarily set to 10, the task heterogeneity was set to 0.7, and the machine heterogeneity was also set to 0.7 (the heterogeneity of a set of numbers is the standard deviation divided by the mean). See [3] for a description of a method for generating random numbers with given mean and heterogeneity values.

The mappings were evaluated for robustness, makespan, and load balance index (defined as the ratio of the finishing time of the machine that finishes first to the makespan). The larger the value of the load balance index, the more balanced the load (the largest value being 1). The tolerance, τ, was set to 20% (i.e., the actual makespan could be no more than 1.2 times the predicted value). In this context, a robustness value of x for a given mapping means that the mapping can endure any combination of ETC errors without the makespan increasing beyond 1.2 times its predicted value as long as the Euclidean norm of the errors is no larger than x seconds.

Figure 3 shows the robustness of a mapping against its makespan. It can be seen that sharp differences exist in the robustness of some mappings that have very similar values of makespan. A similar conclusion could be drawn from the robustness against load balance index plot (not shown here). Both of the above conclusions highlight the fact that the robustness metric can be used to differentiate between mappings that perform similarly with respect to two popular performance metrics.

It can be seen in Figure 3 that some mappings are clustered into groups, such that for all mappings within a group, the robustness increases linearly with the makespan. This can be explained using Equation 6. Let $m(\boldsymbol{C})$ be the machine that determines the makespan at $\boldsymbol{C}$. Let $n(m_j)$ be the number of applications mapped to machine m_j. If $m(\boldsymbol{C}^{\text{orig}})$ has the largest number of applications mapped to it, then it is also the machine that determines the robustness of the mapping (because it has the smallest robustness radius, Equation 6). Now consider the set $S_1(x)$ of mappings such that $x = n(m(\boldsymbol{C}^{\text{orig}})) = \max\limits_{j:m_j \in \mathcal{M}} n(m_j)$ for each mapping in the set. For mappings in $S_1(x)$, the robustness is directly proportional to M^{orig} (Equation 6). Each distinct straight line in Figure 3 corresponds to $S_1(x)$ for some $x \in \{1 \cdots |\mathcal{A}|\}$. The explanation for the outlying points is as follows. Let $S_2(x)$ be the union of $S_1(x)$ and the set of mappings for which $x = n(m(\boldsymbol{C}^{\text{orig}})) \neq \max\limits_{j:m_j \in \mathcal{M}} n(m_j)$. The outlying points belong to the latter set, $S_2(x) - S_1(x)$. Note that all such outlying points lie "below" the line specified by $S_1(x)$. For a mapping that corresponds to an outlying point, the machine that determines the robustness is not $m(\boldsymbol{C}^{\text{orig}})$; it is some other machine for which the robustness radius is smaller than the robustness radius for $m(\boldsymbol{C}^{\text{orig}})$.

4.3. The HiPer-D System

For the model in Subsection 3.2, the experiments were performed for a system that consisted of 19 paths, where the end-to-end latency constraints of the paths were uniformly

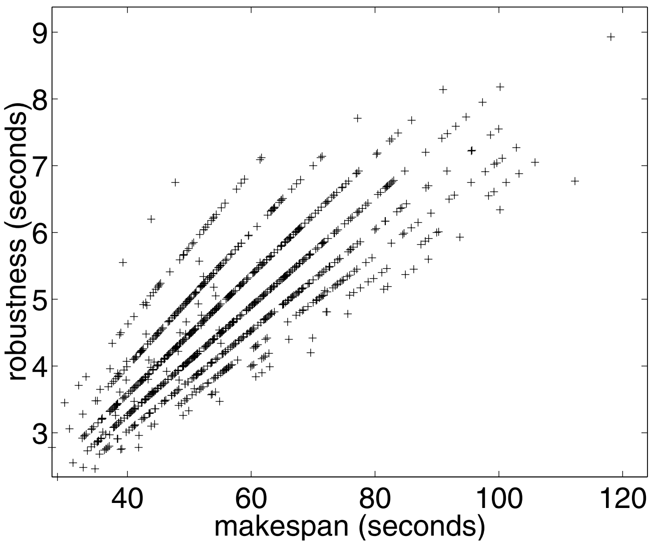


Figure 3. The plot of robustness against makespan for 1000 randomly generated mappings. While robustness and makespan are generally correlated, for any given value of makespan there are a number of mappings that differ significantly in terms of their actual robustness.

sampled from the range [750, 1250]. The system had three sensors (with rates 4×10^{-5}, 3×10^{-5}, and 8×10^{-6}), and three actuators. The experiments made the following simplifying assumptions. The computation time function, $T^{c}_{ij}(\boldsymbol{\lambda})$, was assumed to be of the form $\sum_{1 \leq z \leq 3} b_{ijz}\lambda_z$, where $b_{ijz} = 0$ if there is no route from the z-th sensor to application a_i. Otherwise, b_{ijz} was sampled from a Gamma distribution with a mean of 10 and task and machine heterogeneity values of 0.7 each. For simplicity in the presentation of the results, the communication times were all set to zero. These assumptions were made only to simplify the experiments, and are *not* a part of the formulation of the robustness metric. The salient point in this example is that the utility of the robustness metric can be seen even when simple complexity functions are used.

The mappings were evaluated for robustness and "slack." In this context, a robustness value of x for a given mapping means that the mapping can endure any combination of sensor loads without a latency or throughput violation as long as the Euclidean norm of the increases in sensor loads (from the assumed values) is no larger than x. Slack has been used in many studies as a performance measure (e.g., [9, 18]) for mapping in parallel and distributed systems, where a mapping with a larger slack is considered better. In this study, slack is defined mathematically as follows. Let the fractional value of a given QoS attribute be the value of the attribute as a percentage of the maximum allowed value. Then the percentage slack for a given QoS attribute is the fractional value subtracted from 1. The system-wide percentage slack is the minimum value of percentage slack taken over all QoS constraints, and can be expressed mathematically as

$$\min\left(\min_{k:\,\mathcal{P}_k\in\mathcal{P}}\left(1-\frac{L_k(\boldsymbol{\lambda})}{L_k^{\max}}\right),\right.$$
$$\left.\min_{i:\,a_i\in\mathcal{A}}\left(1-\frac{\max\left(T_i^{\mathrm{c}}(\boldsymbol{\lambda}),\ \max_{a_p\in\mathcal{D}(a_i)}T_{ip}^{\mathrm{n}}(\boldsymbol{\lambda})\right)}{1/R(a_i)}\right)\right).$$

Figure 4 shows the robustness of a mapping against its slack. It can be seen that while mappings with a larger slack are more robust in general, sharp differences exist in the robustness of some mappings that have very similar values of slack. In particular, examine the two mappings shown in Table 2. Although the slack values are approximately the same, the robustness of B is about 3.3 times that of A. While both A and B perform similarly when compared using slack, B is a much better mapping considering its robustness against increases in sensor loads. That is, using slack as a measure of how much increase in sensor load a system can tolerate may cause system designers to grossly misjudge the systems capability. The results show that slack, as defined here, is not a good indicator of the robustness of the system as to how many objects per data set could be processed by the system before a QoS violation occurred. As can be seen in Figure 4, there is a set of mappings with slack values ranging from approximately 0.2 to approximately 0.5, but all these mappings have the same robustness value of approximately 250. These mappings are virtually indistinguishable from each other with respect to how many objects per data set could be processed by the system before a QoS violation occurred. The experiments in this section illustrate that, even for very simple computing environments, a commonly used measure of the availability of computing resources is not a reliable measure of the system's ability to maintain a desired QoS in the presence of uncertainty, and that an explicit measure of robustness should be used in the manner specified by the FePIA procedure presented here.

Table 2 also shows the actual sensor load values at which a throughput or latency constraint is reached for a given mapping. For reference, the computation time functions $T_{ij}^{\mathrm{c}}(\boldsymbol{\lambda})$ for this experiment are given in Table 2.

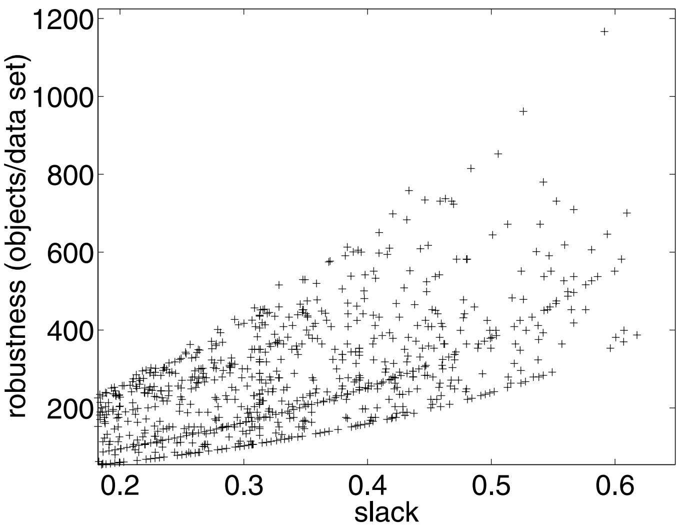


Figure 4. The plot of robustness against slack for 1000 randomly generated mappings. While robustness and slack are generally correlated, for any given value of slack there are a number of mappings that differ significantly in terms of their actual robustness, a problem that is exacerbated for large values of slack.

5. Conclusions

This paper has presented a mathematical description of a new metric for the robustness of a resource allocation with respect to desired system performance features against perturbations in system and environmental conditions. In addition, the research describes a four-step procedure, the FePIA procedure, to methodically derive the robustness metric for a variety of parallel and distributed resource allocation systems. For illustration, the FePIA procedure is employed to derive robustness metrics for two example distributed systems. The experiments conducted in this research for two example parallel and distributed systems illustrate the utility of the robustness metric in distinguishing between the mappings that perform similarly otherwise.

Acknowledgments: The authors thank Sahra Sedigh-Ali, Martin Nagel, and Safoora Sarvestani for their valuable comments.

References

[1] S. Ali. *Robust Resource Allocation in Dynamic Distributed Heterogeneous Computing Systems*. PhD thesis, School of Electrical and Computer Engineering, Purdue University, to appear, 2003.

Table 2. For the system in Subsection 3.2, these two mappings perform similarly in terms of their slack values, but are significantly different in their robustness values. The initial sensor load values are λ_1 = 962, λ_2 = 380, and λ_3 = 240. The final sensor load values at which a throughput or latency constraint is reached for a given mapping are given in the table (λ_1^*, λ_2^*, and λ_3^*). The computation time function $T_{ij}^{c}(\lambda)$ for each application on the machine where it is mapped is also shown. For $T_{ij}^{c}(\lambda)$ functions, the number outside the parenthesis is the "multi-tasking factor," and equals $1.3n(m_j)$ where $n(m_j) \geq 2$ is the number of applications mapped to m_j. The function inside the parenthesis gives the complexity of the execution time as a function of λ, and may be different for a given application mapped to different machines (e.g., as for a_2).

		mapping A	mapping B
robustness		353 objects/data set (based on Euclidean distance)	1166 objects/data set (based on Euclidean distance)
slack		0.5961	0.5914
$\lambda_1^*, \lambda_2^*, \lambda_3^*$		962, 380, 593	962, 1546, 240
application assignments	m_1	$a_5, a_9, a_{12}, a_{17}, a_{20}$	$a_3, a_4, a_5, a_{17}, a_{18}, a_{20}$
	m_2	a_6, a_{16}	$a_2, a_{11}, a_{14}, a_{19}$
	m_3	a_1, a_3, a_7	a_1, a_7, a_{13}
	m_4	$a_2, a_4, a_{10}, a_{13}, a_{15}, a_{19}$	a_9, a_{12}, a_{15}
	m_5	$a_8, a_{11}, a_{14}, a_{18}$	a_6, a_8, a_{10}, a_{16}
computation time functions	a_1	$3.90(4\lambda_3)$	$3.90(4\lambda_3)$
	a_2	$7.80(5\lambda_2)$	$5.20(2\lambda_2)$
	a_3	$3.90(6\lambda_1)$	$7.80(11\lambda_1)$
	a_4	$7.80(\lambda_1)$	$7.80(4\lambda_1 + 2\lambda_2)$
	a_5	$6.50(3\lambda_1 + \lambda_3)$	$7.80(3\lambda_1 + \lambda_3)$
	a_6	$2.60(\lambda_3)$	$5.20(\lambda_3)$
	a_7	$3.90(5\lambda_2)$	$3.90(5\lambda_2)$
	a_8	$5.20(6\lambda_2)$	$5.20(6\lambda_2)$
	a_9	$6.50(20\lambda_3)$	$3.90(3\lambda_3)$
	a_{10}	$7.80(5\lambda_2 + 7\lambda_3)$	$5.20(3\lambda_2 + 3\lambda_3)$
	a_{11}	$5.20(10\lambda_1 + 8\lambda_2 + 6\lambda_3)$	$5.20(10\lambda_1 + 4\lambda_2 + 8\lambda_3)$
	a_{12}	$6.50(26\lambda_1)$	$3.90(24\lambda_1)$
	a_{13}	$7.80(19\lambda_1 + 8\lambda_2)$	$3.90(23\lambda_1 + 6\lambda_2)$
	a_{14}	$5.20(11\lambda_1)$	$5.20(7\lambda_1)$
	a_{15}	$7.80(13\lambda_1 + 17\lambda_2 + 9\lambda_3)$	$3.90(13\lambda_1 + 17\lambda_2 + 9\lambda_3)$
	a_{16}	$2.60(2\lambda_2)$	$5.20(7\lambda_2)$
	a_{17}	$6.50(3\lambda_1 + 5\lambda_3)$	$7.80(3\lambda_1 + 5\lambda_3)$
	a_{18}	$5.20(3\lambda_1 + 19\lambda_2 + 11\lambda_3)$	$7.80(6\lambda_1 + 2\lambda_2 + 10\lambda_3)$
	a_{19}	$7.80(9\lambda_1 + 13\lambda_2)$	$5.20(4\lambda_1 + 8\lambda_2)$
	a_{20}	$6.50(3\lambda_1 + 14\lambda_2 + 18\lambda_3)$	$7.80(3\lambda_1 + 14\lambda_2 + 18\lambda_3)$

[2] S. Ali, J.-K. Kim, Y. Yu, S. B. Gundala, S. Gertphol, H. J. Siegel, A. A. Maciejewski, and V. Prasanna. Greedy heuristics for resource allocation in dynamic distributed real-time heterogeneous computing systems. In *The 2002 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2002), Vol. II*, pages 519–530, June 2002.

[3] S. Ali, H. J. Siegel, M. Maheswaran, D. Hensgen, and S. Sedigh-Ali. Representing task and machine heterogeneities for heterogeneous computing systems. *Tamkang Journal of Science and Engineering*, 3(3):195–207, invited, Nov. 2000.

[4] P. M. Berry. Uncertainty in scheduling: Probability, problem reduction, abstractions and the user. IEE Computing and Control Division Colloquium on Advanced Software Technologies for Scheduling, Digest No: 1993/163, Apr. 26, 1993.

[5] L. Bölöni and D. C. Marinescu. Robust scheduling of metaprograms. *Journal of Scheduling*, 5(5):395–412, Sep. 2002.

[6] S. Boyd and L. Vandenberghe. *Convex Optimization,* to appear in 2003, available at `http://www.stanford.edu/class/ee364/index.html#book`.

[7] T. D. Braun, H. J. Siegel, N. Beck, L. L. Bölöni, M. Maheswaran, A. I. Reuther, J. P. Robertson, M. D. Theys, B. Yao, D. Hensgen, and R. F. Freund. A comparison of eleven static heuristics for mapping a class of independent tasks onto heterogeneous distributed computing systems. *Journal of Parallel and Distributed Computing*, 61(6):810–837, June 2001.

[8] J. M. Carlson and J. Doyle. Complexity and robustness. *Proceedings of National Academy of Science (PNAS)*, 99(1):2538–2545, Feb. 2002.

[9] A. Davenport, C. Gefflot, and J. Beck. Slack-based techniques for robust schedules. In *6th European Conference on Planning (ECP-2001)*, pages 7–18, Sep. 2001.

[10] J. DeVale. *High Performance Robust Computer Systems*. PhD thesis, Department of Electrical and Computer Engineering, Carnegie Mellon University, 2001.

[11] S. Gertphol, Y. Yu, S. B. Gundala, V. K. Prasanna, S. Ali, J.-K. Kim, A. A. Maciejewski, and H. J. Siegel. A metric and mixed-integer-programming-based approach for resource allocation in dynamic real-time systems. In *16th International Parallel and Distributed Processing Symposium (IPDPS 2002)*, Apr. 2002.

[12] M. L. Ginsberg, A. J. Parkes, and A. Roy. Supermodels and robustness. In *15th National Conference on Artificial Intelligence, American Association for Artificial Intelligence (AAAI)*, pages 334–339, July 1998.

[13] S. D. Gribble. Robustness in complex systems. In *8th Workshop on Hot Topics in Operating Systems (HotOS-VIII)*, pages 21–26, May 2001.

[14] R. Harrison, L. Zitzman, and G. Yoritomo. High performance distributed computing program (HiPer-D)—engineering testbed one (T1) report, Nov. 1995. Technical Report.

[15] E. Hart, P. Ross, and J. Nelson. Producing robust schedules via an artificial immune system. In *The 1998 International Conference on Evolutionary Computing*, pages 464–469, May 1998.

[16] E. Jen. Definitions. Santa Fe Institute Robustness Site, RS-2001-009, `http://discuss.santafe.edu/robustness`, posted 10-22-01.

[17] E. Jen. Stable or robust? What is the difference? In *Complexity*, to appear.

[18] M. Jensen. Improving robustness and flexibility of tardiness and total flowtime job shops using robustness measures. *Journal of Applied Soft Computing*, 1(1):35–52, June 2001.

[19] M. T. Jensen and T. K. Hansen. Robust solutions to job shop problems. In *The 1999 Congress on Evolutionary Computation*, pages 1138–1144, July 1999.

[20] V. Leon, S. Wu, and R. Storer. Robustness measures and robust scheduling for job shops. *IEE Transactions*, 26(5):32–43, Sep. 1994.

[21] M. Maheswaran, S. Ali, H. J. Siegel, D. Hensgen, and R. F. Freund. Dynamic mapping of a class of independent tasks onto heterogeneous computing systems. *Journal of Parallel and Distributed Computing*, 59(2):107–131, Nov. 1999.

[22] M. Sevaux and K. Sörensen. Genetic algorithm for robust schedules. In *8th International Workshop on Project Management and Scheduling (PMS 2002)*, pages 330–333, Apr. 2002.

[23] G. F. Simmons. *Calculus With Analytic Geometry, Second Edition*. McGraw-Hill, New York, NY, 1995.

[24] K. Swanson, J. Bresina, and M. Drummond. Just-in-case scheduling for automatic telescopes. In *Knowledge-Based Artificial Intelligence Systems in Aerospace and Industry,* W. Buntine and D. H. Fisher, eds., pages 10–19. Proceedings of SPIE, Vol. 2244, SPIE Press, Bellingham, WA, 1994.

[25] L. Wang, H. J. Siegel, V. P. Roychowdhury, and A. A. Maciejewski. Task matching and scheduling in heterogeneous computing environments using a genetic-algorithm-based approach. *Journal of Parallel and Distributed Computing*, 47(1):8–22, Nov. 1997.