INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN
ICED 03 STOCKHOLM, AUGUST 19-21, 2003

REAL-TIME CO-ORDINATED RESOURCE MANAGEMENT IN A COMPUTATIONAL ENVIRONMENT

G. Coates, A.H.B. Duffy, R.I. Whitfield, W. Hills

Abstract

Design co-ordination is an emerging engineering design management philosophy with its emphasis on timeliness and appropriateness. Furthermore, a key element of design co-ordination has been identified as resource management, the aim of which is to facilitate the optimised use of resources throughout a dynamic and changeable process. An approach to operational design co-ordination has been developed, which incorporates the appropriate techniques to ensure that the aim of co-ordinated resource management can be fulfilled. This approach has been realised within an agent-based software system, called the Design Co-ordination System (DCS), such that a computational design analysis can be managed in a coherent and co-ordinated manner.

The DCS is applied to a computational analysis for turbine blade design provided by industry. The application of the DCS involves resources, i.e. workstations within a computer network, being utilised to perform the computational analysis involving the use of a suite of software tools to calculate stress and vibration characteristics of turbine blades. Furthermore, the application of the system shows that the utilisation of resources can be optimised throughout the computational design analysis despite the variable nature of the computer network.

Keywords: Design management, resource management

1 Introduction

Co-ordination has been recognised as being relevant and important in a number of disciplines such as engineering design [1-3], distributed artificial intelligence [4-6] and organisational theory [7-9]. Indeed, while co-ordination has been studied from a number of discipline viewpoints, and alternative approaches have been proposed, it has been indicated that the final objective is the same in each case, i.e. the meshing of behaviour to promote co-operation and avoid conflict [10]. Based on extensive literature reviews [3,11], operational design co-ordination has been identified as consisting of six key elements: coherence, communication, task management, resource management, schedule management and real-time support. This paper focuses on resource management in real-time.

Design co-ordination has been described as covering aspects of organisation and management of resources, and control of the use of resources [12]. Similarly, the focus for supporting design co-ordination has been reported as being directed at the effective utilisation and integration of resources in order to optimise design activity [13]. The use of many resources to facilitate the efficient performance of tasks is an approach that has been reported as having benefits such as speeding up a process [5, 14]. However, with regard to computing environments, it has been recognised that the reduction in duration of many concurrent applications is not proportional to the increase in the number of processors [15]. On a related theme, it has been indicated that committing greater resources to a problem would not necessarily result in a proportional reduction of time to complete tasks [16]. Rather, "it is the capacity to co-ordinate the activity performed by each team member, taking into account the available resources and knowledge of their roles and effects, that enables a measured reduction in the duration of those activities to be achieved".

Distributed computing systems need resource management capabilities that can allocate resources to applications, monitor and control the use of resources, and re-allocate resources due to anomalies [17]. Thus, there is a need for research to develop new techniques that will manage resources in a uniform and co-ordinated way within a dynamic environment. While resource scheduling has been the focus of much research over recent years, network monitoring has been largely neglected [18]. Furthermore, it was indicated that the dynamic varying nature of network load can substantially impact resource performance. Similarly, it has been recognised that there is a need to be able to detect and recover from discrepancies between the expected state and actual state during execution [19]. As such, decision-making with regard to which resource should be used to run particular applications can be based on monitoring variations in network characteristics [20]. Despite the inherent inaccuracies of runtime measurements, and the added overhead of more frequent re-allocations, schedulers using these measurements can significantly outperform those that do not [21]. In addition, monitoring has been considered as almost always providing a reduction in missed deadlines [22]. However, a concern that has been voiced is that monitoring should be non-intrusive, i.e. not compromise the performance of resources [23].

The organisation of the remainder of this paper is as follows. In Section 2, an overview is given of the components of resource management incorporated within an agent-based software system, called the Design Co-ordination System (DCS). Section 3 then presents the aspects of resource management associated with an application of the DCS. A discussion of the results of the case study is given in Section 4. Finally, Section 5 presents some concluding remarks.

2 Real-time co-ordinated resource management

Within the DCS, a collection of agents act as members of a multi-functional team communicating and interacting in a coherent and co-ordinated fashion such that inter-related tasks can be undertaken in a structured manner. Simultaneously, resources, of varying performance efficiency, are allocated and utilised in an optimised fashion in accordance with multiple schedules. In the context of the DCS, resources are workstations within a computer network that can be utilised to undertake tasks, i.e. analysis tool executions.

The DCS operates in a real-time computer network of UNIX workstations, in which these resources may not perform as expected, the outcome of which will influence the performance of the computational design analysis. Hence, there is a requirement for resource management and monitoring to enable the detection of change such that, if appropriate, corrective action may be taken by performing re-scheduling.

Within the DCS, two types of agent are responsible for resource management, namely the Resource Manager and Resource Monitor(s). Jointly, these agents are dedicated to ensuring that the resources within the computer network can be continuously utilised in an optimised manner.

The Resource Manager is primarily responsible for maintaining accurate knowledge of all available resources within the computer network. This knowledge is stored in a resource model. A Resource Monitor is located on each workstation within the computer network and is responsible for sensing, forecasting and reporting resource performance efficiency. At regular time intervals throughout the operation of the DCS, each Resource Monitor observes the actual performance efficiency of their associated resource such that any significant deviation transgressing specified thresholds can be identified.

An overview of the operation of the Resource Manager and Resource Monitors is shown in Figure 1.

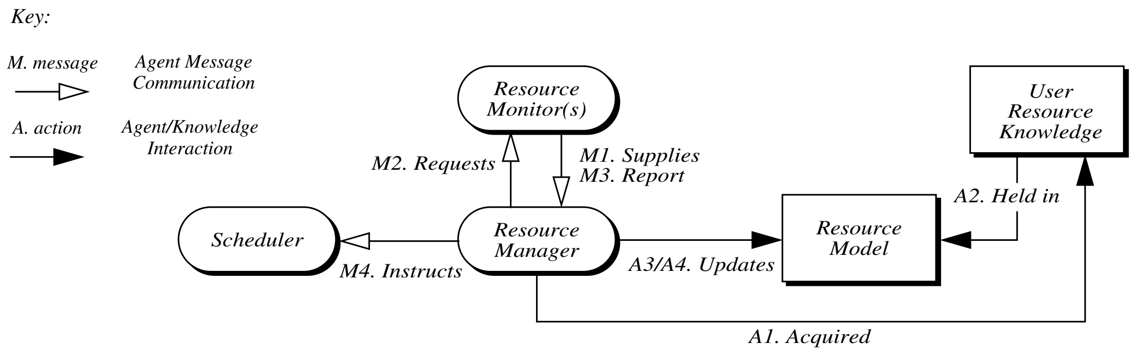

Figure 1. Overview of the operation of the Resource Manager and Resource Monitors

Initially, the Resource Manager acquires user supplied knowledge *(Action A1)* and holds it in the resource model *(Action A2)*. Knowledge of monitored resource performance efficiency is acquired by Resource Monitors throughout the computational design analysis. In addition, monitored performance efficiency changes dynamically, thus, reflecting the variability of the resource usage within the computer network. In the event of any violation of a resource's lower (R_{LT}) or upper (R_{UT}) performance efficiency threshold, a forecast of performance efficiency is calculated by the associated Resource Monitor. This forecast is then supplied to the Resource Manager *(Message M1)*. Following notification of a change in the forecasted performance efficiency of a resource, and after re-setting the appropriate upper and lower thresholds, the Resource Manager updates the resource model *(Action A3)*. The Resource Manager then requests *(Message M2)* that all other Resource Monitors determine and report *(Message M3)* an up-to-date forecast of the performance efficiency of their associated resource. This allows the Resource Manager to completely update the resource model *(Action A4)*. The Resource Manager then instructs *(Message M4)* the Scheduler to consider deriving a new schedule using knowledge from the resource model. Due to the resource model being up-to-date, any schedule produced will reflect the most recent performance efficiency forecast of the resources within the computer network.

The decision-making process regarding co-ordinated scheduling involves determining whether it is more economical time-wise to continue with the original schedule or derive and enact a revised schedule [24]. If it is decided that re-scheduling is inappropriate then the original schedule continues to be administered. However, if re-scheduling is appropriate, then a revised schedule must be derived. Furthermore, in order to maintain optimised resource utilisation during the period of re-scheduling, interim schedules must be constructed and then enacted.

3 Industrial case study

Siemens Power Generation Limited provided a practical case study to enable the application and evaluation of the Design Co-ordination System. The case study involves the use of a suite of analysis tools used in the selection of blades and bladepath, and the calculation of the associated stresses and vibration characteristics of the blades.

At the outset of the computational design analysis, the Resource Manager constructs a resource model, as shown in Table 1.

Table 1. Resource model

$\mathbf{R_I}$	$\mathbf{R_{FE}}$ (%)	$\mathbf{R_{LT}}$ (%)	$\mathbf{R_{UT}}$ (%)
1	98.9	50	100
2	99.6	50	100
3	96.1	50	100
4	91.6	50	100

Resources are arbitrarily assigned R_I of 1, 2, 3, and 4. Initial forecasts of performance efficiency, R_{FE}, for each resource are calculated by the respective Resource Monitor based on values of monitored performance efficiency, R_{ME}, observed over a period of time prior to scheduling. Values of lower, R_{LT}, and upper, R_{UT}, performance efficiency threshold are assigned by the user as 50% and 100% respectively.

Throughout the computational design analysis, Resource Monitors observe the various constituents of central processing unit (CPU) utilisation of their associated resource at 5 second intervals such that any significant deviation transgressing the specified thresholds can be identified. In particular, a Resource Monitor establishes what percentage of the current CPU utilisation of its associated resource is attributed to (i) user processes, R_{user}, i.e. computer programs being run by users, (ii) system processes, R_{system}, i.e. UNIX kernel code, and, (iii) idle, R_{idle}, i.e. not being utilised. R_{user} is divided into two components, namely the proportion of CPU utilisation attributed to analysis tools being run within the DCS, R_{DCS}, and other computer programs being run that are unrelated to the DCS, R_{other}.

Based on the CPU utilisation observed over a period of time, each Resource Monitor determines the monitored performance efficiency, R_{ME_t}, at corresponding time intervals t, of its associated resource using the equation:

$$R_{ME_t} = R_{CF} \times \left[R_{idle_t} + R_{DCS_t} + \frac{R_{other_t}}{n_{ps}}(1 + R_{system_t}) \right] \tag{1}$$

In equation (1), n_{ps} is the number of processes being executed on the resource. The resource coefficient, R_{CF}, is a relative measure of the processor speed of a resource such that the forecasted performance efficiencies determined for all resources are directly comparable. This is required for purposes of scheduling and re-scheduling.

Equation (1) has been developed from that used within the Network Weather Service [25]. The developed equation (i) caters for the replacement of a DCS user process as well as the addition of other processes, and (ii) further divides the resource utilisation attributed to users into that associated with the operation of the DCS and that not. Assumptions regarding this equation are that the replacement process will be entitled to all of the existing DCS and idle utilisation availability, and a fair, i.e. proportional, share of other user utilisation availability.

In Figure 2, the CPU utilisation and monitored performance efficiency of resource $R_I = 4$ over a period of the computational design analysis, are shown as observed by the associated Resource Monitor.

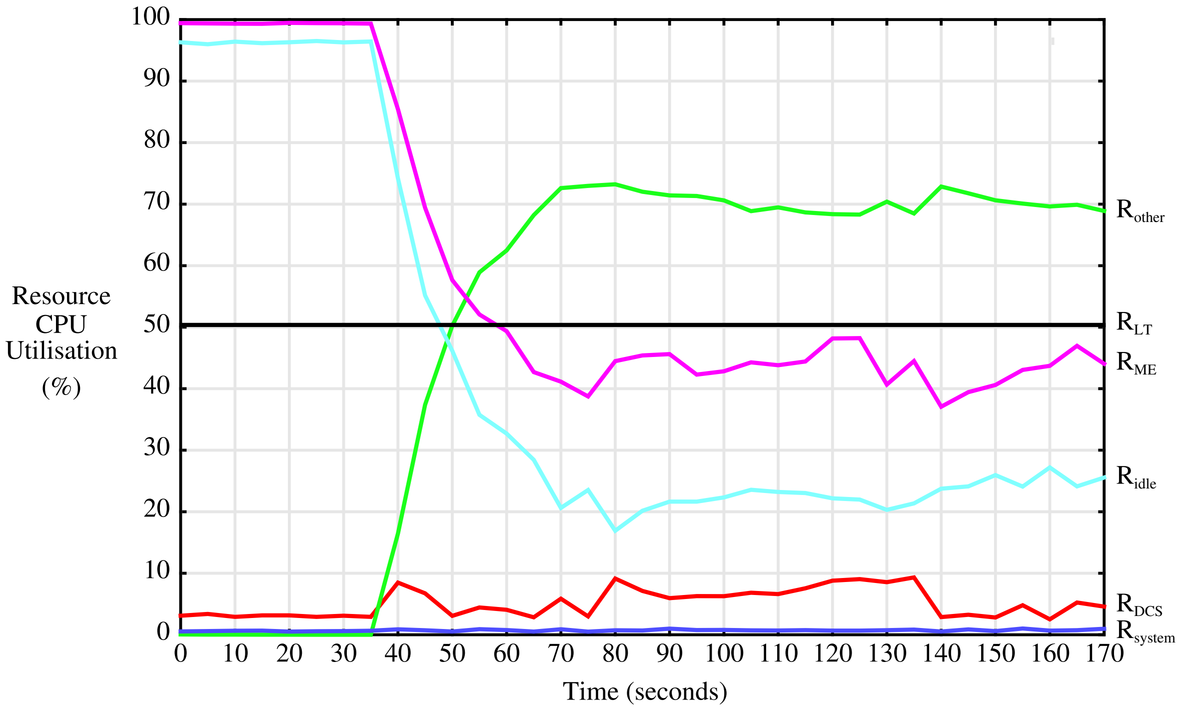

Figure 2. $R_I = 4$, Resource CPU utilisation versus time

In Figure 2, it can be seen that at approximately t = 60 seconds there is a deviation in monitored performance efficiency, R_{ME}, that transgresses its lower threshold of $R_{LT} = 50\%$, which instigates the consideration of re-scheduling. Further, for this resource, R_{ME} fluctuates between 38% and 48% during the remainder of the computational design analysis. Although not shown in Figure 2, for resources R_I = 1, 2, and 3, monitored performance efficiency remains at approximately 99% throughout the operation of the DCS.

Due to the monitored performance efficiency of resource $R_I = 4$ falling below the lower threshold of 50%, the associated Resource Monitor forecasts future performance efficiency. This is achieved by performing a regression analysis using orthogonal polynomials with recent values of monitored performance efficiency. Statistical regression analysis is employed since it has been reported as being suitable for prediction/forecasting purposes [26-28].

Since the order of the polynomial that best represents historical data was not known prior to the regression analysis being performed, the use of orthogonal polynomials and analysis of variance enables the best fitting polynomial to be determined dynamically. As such, forecasting is more efficient and less intrusive to computation, in comparison with using a number of pre-defined models and selecting the forecast with the least error [18, 23]. Forecasting is also less intrusive toward computation since it is only performed when necessary according to defined monitored performance efficiency thresholds being transgressed rather than each time a new network performance measurement was taken, as within other systems [18, 23]. Furthermore, regression analysis using orthogonal polynomials accounts for the order in which measurements of CPU utilisation are taken. Thus, the polynomial produced reflects the trend of the historical data, which is significant. This contrasts with the other systems mentioned that only use the mean and median of a fixed number of recent network performance measurements, i.e. a sliding window, thus not accounting for the order in which they were recorded.

The regression equations (2), (3), (4) and (5) correspond with resources R_I = 1, 2, 3 and 4 respectively. These equations were derived at the point in the computational design analysis when R_{LT} was transgressed for resource $R_I = 4$.

$$R_{FE,t} = -0.0441t + 99.703 \tag{2}$$

$$R_{FE,t} = 0.0007t + 99.229 \tag{3}$$

$$R_{FE,t} = -0.0392t + 99.555 \tag{4}$$

$$R_{FE,t} = -0.1979t^4 + 7.7977t^3 - 111.97t^2 + 683.18t - 1395.9 \tag{5}$$

The Resource Monitor, associated with resource $R_I = 4$, supplies the forecasted performance efficiency to the Resource Manager, which updates the resource model accordingly. The Resource Manager also re-sets the values of R_{LT} or R_{UT} to 10% and 70% respectively. The reason for re-setting these values is to avoid re-scheduling being considered again as a result of insignificant deviations in monitored performance efficiency about the previous lower threshold.

The Resource Manager also requests that the other three Resource Monitors determine and report forecasts of performance efficiency for their associated resource. The Resource Manager then updates all performance efficiency forecasts within the resource model, as shown in Table 2.

Table 2. Revised resource model

R_I	R_{FE} (%)	R_{LT} (%)	R_{UT} (%)
1	99.1	50	100
2	99.2	50	100
3	99.0	50	100
4	41.8	10	70

Subsequently, based on up-to-date knowledge of the forecasted performance efficiencies of the resources, the interim and revised schedules derived enabled the appropriate number and type of tasks to be assigned to each resource. As such, resources are able to continue being utilised in an optimised manner during and after re-scheduling, i.e. for the remainder of the computational design analysis.

4 Discussion

In the context of co-ordination, resource management has been defined as "organising and controlling resources to enable their continuous optimised utilisation throughout a changeable design development process" [3]. In order to assess co-ordinated resource management in the case study, the utilisation of the resources during the computational design analysis is considered in terms of the schedules used, i.e. original, interim and revised.

4.1 Resource utilisation within the original schedule

Based on the forecasts made by the Resource Monitors and maintained within the resource model by the Resource Manager, scheduling was performed using a genetic algorithm. As such, each resource was assigned the appropriate type and number of tasks. Consequently, resources were utilised in an optimised manner for a proportion of the original schedule. However, at a certain point in time, i.e. t = 60 seconds, the monitored performance efficiency of resource $R_I = 4$ began to decrease, which was observed by the associated Resource Monitor. Thus, while the monitored performance efficiencies of three of the four resources closely adhered to their original performance forecasted efficiencies, resource $R_I = 4$ started to affect their utilisation. Due to this decrease, tasks expected to be undertaken and completed on resource $R_I = 4$ were prolonged. Also, tasks dependent on the completion of these prolonged tasks were delayed, thus affecting the utilisation of all resources. While resource $R_I = 4$ continued to be utilised, re-scheduling was considered and, since appropriate, performed to avoid any further delays being encountered. As such, resources were utilised in an optimised manner according to the proportion of the original schedule that was enacted. However, the timely recognition of the need to consider re-scheduling ensured the future utilisation of resources could also optimised.

4.2 Resource utilisation within the interim schedule

In considering whether or not to re-schedule, an interim schedule was derived by applying a three-step procedure [24], rather than using a genetic algorithm. The application of this procedure resulted in tasks being included within the interim schedule that would ensure, as near as possible, resources would be utilised in an optimised manner during re-scheduling. The interim schedule reflected the workload able to be completed by each of the resources taking into account their respective newly forecasted performance efficiencies. That is, the cumulative datum duration of the tasks assigned to each resource were directly proportional to the forecasted performance efficiencies. Since re-scheduling was performed, the interim schedule enabled the continuation of optimised resource utilisation during this period. Furthermore, due to the appropriate of division of outstanding tasks being (i) re-scheduled and (ii) included within the interim schedule, the completion of (i) and (ii) was near co-incident, thus, minimising resource idle time.

4.3 Resource utilisation within the revised schedule

As with the original schedule, the revised schedule was derived using a genetic algorithm. Thus, the revised schedule reflected the tasks able to be completed by each of the resources considering their respective newly forecasted performance efficiencies. In the application of the DCS to the case study, the revised schedule was completed in the time estimated since the forecasts of resource performance efficiency approximately reflected those subsequently monitored. As such, resources were utilised in an optimised manner during the enactment of the revised schedule.

5 Concluding remarks

The need for co-ordinated resource management has been recognised within a number of disciplines such as engineering design and distributed artificial intelligence. Furthermore, the coherent organisation and control of resources has been identified as a means of enabling design activity to be performed efficiently.

Based on an approach to design co-ordination, an agent-oriented software system has been developed. Within an application of this system, aspects of resource management have been presented in order to demonstrate the optimised utilisation of resources, i.e. workstations within a computer network, throughout a computational design analysis. Through managing the resources, including their continuous monitoring, agents within the system reacted dynamically to significant events within the computer network. As a result, appropriate action was taken to maintain the optimised utilisation of resources such that tasks could be completed in the least time possible.

A challenge for the future is to develop and implement a means of co-ordinating the management of designers within engineering companies. Facilitating the optimised, or best, utilisation of design teams and their members will enable projects to be completed in less time, thus giving companies a competitive advantage.

Acknowledgements

The authors would like to thank Siemens Power Generation Limited for providing the practical case study.

References

[1] Andreasen, M.M., Duffy, A.H.B., MacCallum, K.J., Bowen, J. and Storm, T., “The Design Co-ordination Framework: key elements for effective product development”, Proceedings of the 1st International Engineering Design Debate, 1996, pp 151-174.

[2] Duffy, A.H.B., Andreasen, M.M. and O’Donnell, F.J., “Design Co-ordination”, Proceedings of ICED ‘99, Vol. 1, 1999, pp 113-118.

[3] Coates G., “An Approach to Operational Design Co-ordination”, Ph.D. Thesis, University of Newcastle upon Tyne, United Kingdom, 2001.

[4] Jennings, N.R., “Commitments and conventions: The foundation of coordination in multi-agent systems”, The Knowledge Engineering Review, Vol. 8(3), 1993, pp 223-250.

[5] Findler, N. V. and Elder, G. D., “Multiagent coordination and cooperation in a distributed dynamic environment with limited resources”, Artificial Intelligence in Engineering, Vol. 9, 1995, pp 229 - 238.

[6] Heck, P.S., “Steps Toward a Theory of Coordination for Intelligent Agent Societies”, Proceedings of COMPSAC ‘99, 1999, pp 299-300.

[7] Malone, T.W. and Crowston, K., “The Interdisciplinary Study of Coordination”, ACM Computing Surveys, Vol. 26(1), 1994, pp 87-119.

[8] Crowston, K., “A taxonomy of organizational dependencies and coordination mechanisms”, Technical Report 174, Massachusetts Institute of Technology, Centre for Co-ordination Science, 1996.

[9] Greenwood, R.M., Kawalak, P., Robertson, I. and Warboys, B.C., "Modern Systems Architecture: The Contribution of the Coordination Layer", Workshop Research Directions in Process Technology, 1997.

[10] Durfee, E.H., "Organisations, Plans, and Schedules: An Interdisciplinary Perspective on Coordinating AI Systems", Journal of Intelligent Systems, Special Issue on the Social Context of Intelligent Systems, 3(2-4), 1993.

[11] Coates, G., Whitfield, R.I., Duffy, A.H.B. and Hills, W., "Coordination Approaches & Systems - Part II: An Operational Perspective", Journal of Research in Engineering Design, Vol. 12(2), 2000, pp 73-89.

[12] MacCallum, K.J. and Carter, I.M., "An Opportunistic Software Architecture for Support of Design Coordination", Knowledge Based Systems Journal, Vol. 5(1), 1991, pp 55-65.

[13] Duffy, A.H.B., Andreasen, M.M., MacCallum, K.J. and Reijers, L.N., "Design Coordination for Concurrent Engineering", Journal of Engineering Design, Vol. 4(4), 1993, pp 251-265.

[14] Smith, R.G., "The Contract Net Protocol: High-Level Communication and Control in a Distributed Problem Solver", IEEE Transactions on Computers, Vol. C-29(12), 1980.

[15] Nguyen T.D., Vaswani R. and Zahorjan J., "Maximising Speedup through Self-Tuning of Processor Allocation", Proceedings of IPPS '96, 1996, pp 463-468.

[16] Coates, G., Duffy, A.H.B., Whitfield, R.I. and Hills, W., "A Methodology for Design Coordination in a Distributed Computing Environment", Proceedings of ICED '99, Vol. 2, 1999, pp 673-678.

[17] Davis, M.B. and Sydir, J.J, "Position Paper: Resource Management for Complex Distributed Systems", Proceedings of WORDS '96, 1996, pp 113-115.

[18] Kim, J. and Lilja, D.J., "A Network Status Predictor to Support Dynamic Scheduling in Network-Based Computing Systems", Technical Report No. HPPC-98-11, High-Performance Parallel Computing Research Group, University of Minnesota, 1998.

[19] Musliner, D.J., Durfee, E.H. and Shin, K.G., "Execution Monitoring and Recovery Planning with Time", Proceedings of the 7th Conference on Artificial Intelligence Applications, 1991, pp 385-388.

[20] Ranganathan, M., Acharya, A. and Salt, J., "Distributed Resource Monitors for Mobile Objects", Proceedings of IWOOOS '96, 1996, pp 19-23.

[21] Nguyen, T.D., Vaswani, R., and Zahorjan, J., "Using Runtime Measured Workload Characteristics in Parallel Processor Scheduling", Job Scheduling Strategies for Parallel Processing, D.G. Feitelson and L. Rudolph (Eds.), Vol. 1162 of Lecture Notes in Computing Science, pp 155-174, Springer-Verlag, 1996.

[22] Garvey, A. and Lesser, V., "Design-to-time Real-Time Scheduling", IEEE Transactions on Systems, Management and Cybernetics - Special Issue on Planning, Scheduling and Control, Vol. 23(6), 1994.

[23] Wolski, R., "Forecasting Network Performance to Support Dynamic Scheduling Using the Network Weather Service", Proceedings of HPDC '97, 1997, pp 316-325.

[24] Coates, G., Duffy, A.H.B., Whitfield, R.I., Hills, W. and Sen, P., "Real-Time Co-ordinated Scheduling using a Genetic Algorithm", Proceedings of ACDM '02, 2002, pp 381-392.

[25] Wolski, R., Spring, N. and Peterson, C., "Implementing a Performance Forecasting System for Metacomputing: The Network Weather Service", Proceedings of SC '97, 1997.

[26] Montgomery, D.C., and Smith, L.A., "Forecasting and Time Series Analysis", McGraw-Hill, 1976.

[27] Graybill F.A. and Iyer H.K., "Regression Analysis: Concepts and Applications", Duxberry Press, International Thomson Publishing, 1994.

[28] Draper, N.R. and Smith, H., "Applied Regression Analysis", John Wiley & Sons, 1998.

Corresponding author:
Graham Coates
Engineering Design Centre
University of Newcastle
Newcastle upon Tyne
NE1 7RU
United Kingdom
Tel: Int +44 191 222 5300
Fax: Int +44 191 261 6059
E-mail: graham.coates@ncl.ac.uk