# EDITORIAL

# PERFORMANCE ENGINEERING OF COMPUTER AND COMMUNICATION SYSTEMS

IRFAN AWAN and ROD FRETWELL

*Network and Performance Engineering Research Group*
*Department of Computing, University of Bradford, Bradford, BD7 1DP, UK*
*Email: {i.awan, r.j.fretwell}@scm.brad.ac.uk*

Due to significant development in the software industries and communication networks over the recent years, performance engineering has become a major concern for the design and development of software for complex systems and effective protocols for communication networks.

The UK Performance Engineering Workshop is a popular forum for contributions on performance modelling and evaluation of Computer and Communication systems. Papers in this special issue are the extended versions of selected papers that were presented at the 20th UK Performance Engineering Workshop held in the University of Bradford, Bradford.

Performance Engineering is a fast developing subject with exiting developments on several fronts, as is shown by the variety of topics, techniques and approaches in the eight papers that have been selected for this special issue.

In the first paper Filippopoulos and Karatza present a tractable dynamic programming model of two heterogeneous clusters with co-allocation of jobs. In their work, they computed the optimal policy with respect to whether to co-allocate the next job to start execution or not. The optimal policy is tabulated and compared to the one that does not employ co-allocation at all. The results indicate that co-allocation can be beneficial even when the communication speed between different clusters is slow compared to the speed within the same cluster. The authors have also proposed a heuristic policy and compared its performance with the optimal policy by simulation. They demonstrated that the heuristic policy performs close to optimal for a wide range of parameters.

Dalton, Rungta, Shankar and Webster present the performance implications of implementing cross-cutting concerns as *aspects*. Aspects are frequently cited as an ideal solution to large cross-cutting concerns such as logging and capture of first failure data. The flexibility of the AspectJ™ language with its use of wildcards and combined with a powerful runtime reflection API allows the creation of simple and extensible aspects. This work outlines the performance implications of weaving aspects with existing programs and the relative run-time performance of woven programs in terms of throughput and scalability. They focussed on the new developments in AspectJ technology and deeper empirical analysis.

The performance of a memory system depends greatly on the algorithms used for memory allocation and garbage collection. Basch and Borozan present results of profiling and analysis of behaviour of memory objects in Java. These results can be useful in improving the existing or in the creation of new garbage collection algorithms. Their work proposed a set of parameters that should be traced in benchmark programs and the organisation of an analysis environment.

PEPA is a Markovian process algebra that is easy to use and has a wide user community. Bradley presents extension of the PEPA to semi-Markov PEPA. The extension incorporates generally distributed sojourn-times for action duration so that models can be made more realistic. In this work, Bradley discusses how semi-Markov PEPA models are analysed through Dingle and Knottenbelt's semi-Markov DNAmaca tool and present a small example for the analysis of passage times.

Brenner, Fernandes and Sales present the advantages to be had by extending Classical Tensor Algebra

(CTA), also known as Kronecker Algebra, to allow the definition of functions, i.e. functional dependencies among its operands. Such extended tensor algebra have been called Generalized Tensor Algebra (GTA). Stochastic Automata Networks (SAN) and Superposed Generalized Stochastic Petri Nets (SGSPN) formalisms use such Kronecker representations. The advantages of GTA do not imply any reduction or augmentation in the scope of application, since there is an equivalence of representation between SAN, which uses GTA, and SGSPN, which uses only CTA. Two modelling examples are presented in order to draw comparisons between the memory needs and CPU time required for the generation and solution of problems using both formalisms and to show the computational advantages in using GTA.

Bewick, Pereira and Merabti describe a novel connection admission control mechanism for smoothed video traffic using Network Constrained Smoothing algorithm. Their approach is two-pronged: traffic smoothing according to the critical links that are heavily loaded and connection renegotiation. The proposed approach alters the routing on a per-interval basis when admission control cannot accept new connection beyond certain threshold. They have shown using simulation results that how their method improves aggregate traffic manageability and copes with the delays along different network paths.

Kouvatsos, Assi and Ould-Khaoua investigated performance analysis of wormhole-routed hypercubes with bursty traffic and finite buffers. They proposed an open queueing network model (QNM) for wormhole-routed hypercubes with finite buffers and deterministic routing subject to a compound Poisson arrival process (CPP) with geometrically distributed batches or, equivalently, a generalised exponential (GE) interarrival time distribution. The GE/G/1/K queue and appropriate GE-type flow formulae have been adopted, as cost-effective building blocks, in a queue-by-queue decomposition of the entire network. The validity of the analytic approximations has been demonstrated against simulation. Their numerical study shows that the wormhole-routed hypercubes suffer progressive performance degradation with increasing traffic variability (burstiness).

In the final paper, Awan and Younas present an effective differentiated scheduling scheme to improve the web services performance. Their work is an investigation into the transmission cost of XML messages which is affected by the network latency. The main aim of their research is to minimize network latency of message communication of Web services by employing pre-emptive resume scheduling of active networks. The fundamental principle in this approach is the provision of preferential treatment to some messages as compared to others. This approach assigns different priorities to distinct classes of XML messages given the fact that some messages may tolerate longer delays than others.

# Biographies

**Irfan Awan** received his PhD degree ('97) in Performance Analysis of Queueing Network Models with priorities and blocking from the University of Bradford – UK. He is a senior lecturer in the Department of Computing, University of Bradford which he joined in 1999. He is member of the Network and Performance Engineering Research Group and co-tutor for the MSc Mobile Computing Course. Dr. Awan's research has mainly focussed on developing cost effective analytical models for measuring the performance of complex queueing networks with finite capacities and priorities. He has produced over 70 publications and edited proceedings of the 20th UK Performance Engineering Workshop. He has also authored several special issues of the international journals and is a member of various programme committees and steering committees for International conferences. He is a member of ILT and BCS.

.

**Rod Fretwell** has a background in information systems. In a 25-year career he analysed organisational and technical problems, developed systems and led software development teams. For several years he was systems consultant to an international development agency. He gained a formal qualification in Computing (MSc, Bradford,

1993) and joined the University of Bradford first as a research assistant and then as lecturer. He received his PhD in 2002 for his thesis on traffic correlation in high-speed networks. He is a member of the Network and Performance Engineering research group.