NWO/RGC JOINT RESEARCH SCHEME
COMPLETION REPORT

Project Reference Number

D-HK007/11T

Project Title

Analysis of Limited Resource Sharing Models

Particulars

	Hong Kong team				Dutch team
Name of Principal Investigator (with title)	English: Zhang, Jiheng Chinese:				Vlasiou, Maria
Name of Co-Investigator (if any)	English: Chinese:				Zwart, Bert
Institution or Institutional affiliation		CityU		HKU	Eindhoven University of Technology
		CUHK	✓	HKUST	
		HKBU		LU	
		HKIEd		PolyU	
Other project team members (if any)					

Funding Period

	1st year	2nd year (if applicable)
Start Date	July 1, 2011	July 1, 2012
Completion Date	June 30, 2012	June 30, 2013

Objective(s) as per original application

1. Develop and analyze mathematical models for limited resource systems.
2. Develop approximations for such limited resource sharing models.
3. Develop design rules for limited resource sharing systems.

[Please attach relevant document(s)]

i) Outline of proposed research and results obtained

We investigate a computer network consisting of two layers occurring in, for example, application servers. The first layer incorporates the arrival of jobs at a network of multi-server nodes, which we model as a many-server Jackson network. At the second layer, active servers at these nodes act now as customers who are served by a common CPU. Our main result shows a separation of time scales in heavy traffic: the main source of randomness occurs at the (aggregate) CPU layer; the interactions between different types of nodes at the other layer is shown to converge to a fixed point at a faster time scale; this also yields a state-space collapse property. Apart from these fundamental insights, we also obtain an explicit approximation for the joint law of the number of jobs in the system, which is provably accurate for heavily loaded systems and performs numerically well for moderately loaded systems. The obtained results for the model under consideration can be applied to thread-pool dimensioning in application servers, while the technique seems applicable to other layered systems too.

We also study proportional fairness scheduling policy, which is the most popular service mechanism to describe and analyse the performance of data network at flow level. Recently, several authors have shown that the invariant distribution of such networks admits a product form distribution under critical loading, assuming exponential job size distributions, leaving the case of general job size distributions as an open question. In this paper we show that product form in heavy traffic still holds for general distributions, thus settling the conjecture.

ii) Significance of research results

Our methodological contribution is that it is possible to rigorously establish a separation of time scales property in heavy traffic in an important class of layered networks, which makes these layered networks tractable. Although we focus on the Markovian case, we believe that such properties hold more generally as well; we provide some physical and numerical arguments to support this claim. The result on separation of time scales result essentially implies that the main source of randomness in heavy traffic can be observed at the CPU layer, thus making performance analysis much more tractable. Apart from supporting these claims by theorems, some numerical experiments suggest that the resulting approximations perform well. The results in our paper may be useful to create design rules, for example to dimension thread-pools.

For proportional fairness scheduleing policy, we solved an open question to obtain diffusion limit for processor sharing networks. The key to our methodology is a nice analysis on Lyapunov function to obtain uniform attraction for the underlying fluid models. We also identify a nice reflection mapping to reveal the geometric structure of the network. We carefully analyze the stochastic process to characterize the oscillation and contribute the framework of proving state space collapse for the stochastic processes.

iii) Research output

Maria Vlasiou, Jiheng Zhang, Bert Zwart and Rob van der Mei. *Separation of timescales in a two-layered network.* ITC, 2012.

Maria Vlasiou, Jiheng Zhang, Bert Zwart. *Proportional Fairness in Heavy Traffic: Insensitivity and Product Form.* Manuscript available. To be submitted soon.

iv) Potential for or impact on further research collaboration

We thank the grant for give us the financial freedom to visit each other. Many research ideas have been developed during our visits. Not all of them have been written in papers. But surely we will. There is a lot of potential for future collaboration.