The University of Sydney

**School of Electrical and Information Engineering**

**Master of Engineering (Research)**

# Fair Service for High-Concurrent Requests

**Zhanwen Li**

# Acknowledgement

First and foremost, I am indebted to David Levy and Shiping Chen, my supervisors, for giving me the chance to take on this research project, and their unwavering support and guidance. I thank them so much for their extensive experience, and incredibly broad vision to this thesis work. They are courageous at working on very hard problems, and they also have amazing ways of cutting to the core of complex subjects and focusing on the important details.

I owe a great deal of thanks to Alex Ng from the Macquarie University, who offered critical analysis for my research. I am grateful to have his comments on my project.

I must also thank the anonymous referees from the Conferences of Middleware 2006, MoDDM 2006 and ASWEC2007, who thoroughly reviewed my papers and gave very insightful feedback.

Other thanks are due to many other people at the University of Sydney. In particular, David Peterson inspired me to undertake this project at the beginning stage. David Wang, Justin Tang, Guillaume, Donghe Nam, Jeff Zhou and others have been involved in many aspects of this research, providing countless hours of discussion and debate during this study.

Despite my efforts in producing this work, this thesis would be seriously challenged grammatically if it was not for the contribution of Jeremy Wu, Joshua Ho and Liheng Li, who kindly edited and proofread this entire document. Thank you so much.

The biggest thank you must go to my family for supporting me to pursue my wild dream. I have enjoyed this year so much. Thank you.

Zhanwen Li

# Abstract

Fair Service for High-Concurrent Requests

By

Zhanwen Li

Master of Engineering (Research)

University of Sydney

Associate Professor David Levy

This thesis presents a new approach to ensuring fair service for highly concurrent requests. Our design uses the advantages of staged event-driven architecture (SEDA) to support high-concurrent loadings and makes use of control theory to manage the system performance.

In order to guarantee the quality of service is fairly made to each request, based on SEDA, the control system for fairness is developed as a combination of a global control framework and a set of local self-tune stags. The global control framework is used to control the performance of the whole staged network at the top-level, aimed at coordinating the performance of the stages in the network. On the other hand, each self-tune stage under the control framework is built on the thread pool model, and will use automatic control theory to adjust its performance locally in order to meet the overall target performance. The automatic control system in each stage consists of an automatic modeling mechanism and a feedback module, which optimizes the controller parameters

in the system automatically and guarantees the quality of performance (service rate here) for the stage at runtime.

Based on mathematical proof and simulation results, our designs are implemented in a SEDA-based web server running in a dynamic loading environment. Results demonstrate that the performance of the new system in the real world is almost the same as the theoretical results. It demonstrates that the design is able to adaptively ensure the quality of service to the high-concurrent requests fairly. Compared to the original SEDA design, our design is an effective and handy approach to significantly enhancing the performance of SEDA in a variety of aspects, including fairer service, faster convergent speed, better robustness, higher accuracy and ease of deployment in various practical applications.

# Figure List

# Table of Contents