Workload Modeling for Computer Systems Performance Evaluation

Reliable performance evaluations require the use of representative workloads. This is no easy task because modern computer systems and their workloads are complex, with many interrelated attributes and complicated structures. Experts often use sophisticated mathematics to analyze and describe workload models, making these models difficult for practitioners to grasp. This book aims to close this gap by emphasizing the intuition and the reasoning behind the definitions and derivations related to the workload models. It provides numerous examples from real production systems, with hundreds of graphs. Using this book, readers will be able to analyze collected workload data and clean it if necessary, derive statistical models that include skewed marginal distributions and correlations, and consider the need for generative models and feedback from the system. The descriptive statistics techniques covered are also useful for other domains.

Dror G. Feitelson is a professor of computer science at the Hebrew University of Jerusalem. He is a founding co-organizer of a series of international workshops on job-scheduling strategies for parallel processing and of the ACM Experimental Computer Science Workshop. He maintains the Parallel Workloads Archive, a widely used community resource with logs of activity on parallel supercomputers.

Cambridge University Press
978-1-107-07823-9 - Workload Modeling for Computer Systems Performance Evaluation
Dror G. Feitelson
Frontmatter
More information

WORKLOAD MODELING FOR COMPUTER SYSTEMS PERFORMANCE EVALUATION

Dror G. Feitelson
The Hebrew University of Jerusalem

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

32 Avenue of the Americas, New York, NY 10013-2473, USA

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107078239

First published 2015

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data
Feitelson, Dror G.
Workload modeling for computer systems performance evaluation / Dror G. Feitelson, The Hebrew University of Jerusalem.
pages cm
Includes bibliographical references and index.
ISBN 978-1-107-07823-9 (hardback)
1. Electronic digital computers – Evaluation. 2. Electronic digital computers – Workload – Mathematical models. I. Title.
QA76.9.E94F43 2015
004–dc23 2014023640

ISBN 978-1-107-07823-9 Hardback

To T. P.,

Who taught me that one needs to be aware of one's limitations

in order to make the load workable

Contents

Cambridge University Press
978-1-107-07823-9 - Workload Modeling for Computer Systems Performance Evaluation
Dror G. Feitelson
Frontmatter
More information

Preface

In 1994 I wrote a long survey about parallel job scheduling [229]. This work described and classified the scheduling schemes of 76 systems, as well as many others that were proposed but never implemented, backed by 638 references. In retrospect, one of the things that struck me was that practically any paper that proposed a new scheme also proved it to be better than competing schemes. On reflection, my conclusion was that the source of the problem was in different assumptions and mindsets, including about the properties of the workloads that would run on these systems. The operational conclusion was that it may be more important to understand the workloads than to design new scheduling schemes.

At about the same time, in work on parallel I/O, I was exposed to the Charisma I/O traces collected by David Kotz and Nils Nieuwejaar [516]. Among the voluminous data on I/O operations were a few records about the jobs to which they belonged. This led to an interaction with Bill Nitzberg who provided me with data regarding three months of jobs from the NASA Ames iPSC/860 system, and then to the publication of the first analysis of such a workload log [244]. Several years later, this log became one of the first to be included in the Parallel Workloads Archive [533]. This archive has been instrumental in facilitating research based on real data rather than on baseless assumptions.

Fast forward to 2014. It is now widely accepted that workload characterization and workload modeling are very important for reliable performance evaluations of computer systems. If the workload is wrong, the results will be wrong too – not in the mathematical sense, but in the sense that they will not apply to the situation at hand. Regrettably, workloads are sometimes (and maybe often) still treated as an afterthought, despite a lot of work that has been done on this topic.

At least part of the problem is that there is a gap between what is studied in basic probability and statistics courses and what needs to be used in workload modeling and performance evaluation. In particular, topics such as heavy-tailed distributions and self-similarity are advanced statistical concepts that are not covered in basic courses. To make matters worse, books and research papers on these topics tend to start at a level of mathematical sophistication that is beyond that achieved in basic probability and statistics courses. This makes much of the relevant material inaccessible to many practitioners who want to use the ideas but not spend a lot of time studying the underlying theory.

One goal of this book is to fill this gap. Specifically, I attempt to make definitions and techniques accessible to practitioners by emphasizing the intuition behind them. Although math is used to avoid misunderstandings and explain derivations, this is typically done at a rather elementary level, forgoing mathematical rigor in the interest of making the material more understandable. The book does assume a basic working knowledge of probability, and beyond that it provides a relatively detailed discussion that does not assume the reader can fill in the details. Moreover, we specifically avoid a full and detailed discussion of all the latest bleeding-edge research on advanced statistics.

Another problem with the workloads used in performance evaluation studies is that they are often based on assumptions rather than measurements. Therefore, another goal of this book is to encourage and promote the experimental aspects of computer science. To further this goal, the book emphasizes the use of real data and contains numerous examples based on real datasets. The datasets used are listed in the Appendix and linked from the book's website.

Using real data to illustrate various concepts is also a means to help build an intuition of what definitions mean and how real data behaves – including cases where data tends to misbehave. This is extremely important, because mathematical techniques will provide some sort of results even when they are misapplied. Developing an intuition regarding your data is therefore an important first step in successful evaluations, and knowing how to look at the data and, in particular, how to create illuminating statistical graphs is an important skill.

In developing my ideas about computer workloads and their modeling I was privileged to work with several outstanding students. The ones who contributed the most to this subject were Uri Lublin, Dan Tsafrir, David Talby, Edi Shmueli, Yoav Etsion, and Netanel Zakay.

By far the most mathematically advanced material is contained in Chapter 7 on self-similarity. In writing about this material (and understanding it myself) I received immense help from Benjamin Yakir, both in explaining the mathematical procedures and in bringing to light the insights behind them. Thanks also to Thomas Mikosch for setting me straight in some places. Daniel Nevo did his best to proofread the statistical parts of the text and tried to convince me to make it more rigorous. Naturally errors and misrepresentations remain my responsibility.

Heartfelt thanks are due to all those who have made their workload data available on the Internet for the benefit of the research community. I hope that in the future this will be taken for granted, and much more data will be available for use. The book's website is at http://www.cs.huji.ac.il/%7Efeit/wlmod and includes links to data sources. Updates and errata will be posted there as well.

The book can be used as the basis for a course on workload modeling or as a supplementary text for a course on performance evaluation. However, it is intended for use by practitioners no less than by academics. I wrote it because I found no source from which I myself could learn and understand the more advanced concepts, based on data and intuition rather than formal proofs. I do not know of any other book like it. I hope you find it useful.

Dror G. Feitelson
Jerusalem, February 2014