

การปรับปรุงประสิทธิภาพของภาระงานที่ต่างแบบกันในเซิร์ฟเวอร์เสมือนจริงโดยใช้พฤติกรรม
ของผู้ใช้งาน



นายดุลยวิทย์ ปรางชุมพล



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ภาควิชาคณิตศาสตร์และวิทยาการ
คอมพิวเตอร์

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2556

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

IMPROVING HETEROGENEOUS WORKLOAD PERFORMANCE IN SERVER
VIRTUALIZATION BASED ON USER BEHAVIORS

Mr. Dulyawit Prangchumpol

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Computer Science and
Information Technology
Department of Mathematics and Computer Science
Faculty of Science
Chulalongkorn University
Academic Year 2013
Copyright of Chulalongkorn University

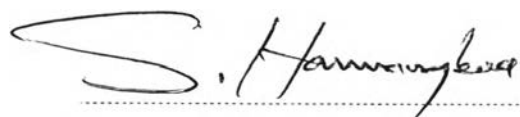
2211476350



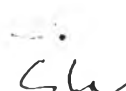

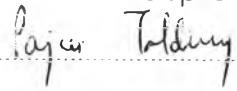
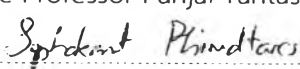

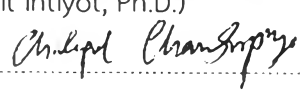
5173813423

Thesis Title IMPROVING HETEROGENEOUS WORKLOAD
PERFORMANCE IN SERVER VIRTUALIZATION
BASED ON USER BEHAVIORS
By Mr. Dulyawit Prangchumpol
Field of Study Computer Science and Information Technology
Thesis Advisor Associate Professor Peraphon Sophatsathit, Ph.D.
Thesis Co-Advisor Associate Professor Panjai Tantasanawong, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial
Fulfillment of the Requirements for the Doctoral Degree

..... Dean of the Faculty of Science
(Professor Supot Hannongbua, Dr.rer.nat.)

THESIS COMMITTEE

..... Chairman
(Professor Chidchanok Lursinsap, Ph.D.)
..... Thesis Advisor
(Associate Professor Peraphon Sophatsathit, Ph.D.)
..... Thesis Co-Advisor
(Associate Professor Panjai Tantasanawong, Ph.D.)
..... Examiner
(Assistant Professor Suphakant Phimoltares, Ph.D.)
..... Examiner
(Boonyarit Intiyot, Ph.D.)
..... External Examiner
(Chalermopol Charnsripinyo, Ph.D.)

ดุลยวิทย์ ปรางชุมพล : การปรับปรุงประสิทธิภาพของภาระงานที่ต่างแบบกันในเซิร์ฟเวอร์เสมือนจริงโดยใช้พฤติกรรมของผู้ใช้งาน. (IMPROVING HETEROGENEOUS WORKLOAD PERFORMANCE IN SERVER VIRTUALIZATION BASED ON USER BEHAVIORS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร. พีระพนธ์ โสฬศสถิตย์, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: รศ. ดร. ปานใจ ธารทัศนวงศ์, 52 หน้า.

องค์กรขนาดใหญ่มีเครื่องเซิร์ฟเวอร์จำนวนมากที่ต้องจัดการและดำเนินการเพื่อรองรับการบริการที่หลากหลาย สิ่งสำคัญคือการลดการใช้พลังงานของเครื่องเซิร์ฟเวอร์ในฝ่ายเทคโนโลยีสารสนเทศ เพื่อเตรียมรองรับกับปัญหาเหล่านี้ เทคนิคเวอร์ชวลไลเซชันได้ถูกนำมาใช้ในช่วงปัจจุบัน เซิร์ฟเวอร์เวอร์ชวลไลเซชันสามารถช่วยรวมเครื่องเซิร์ฟเวอร์ เพิ่มประสิทธิภาพการใช้ประโยชน์จากอุปกรณ์ ลดการใช้พลังงานขององค์กรและพื้นที่การใช้งานลงได้ อย่างไรก็ตาม การจัดการภาระงานที่แตกต่างกันของระบบนี้ก็เป็นเรื่องที่ท้าทาย งานวิจัยนี้นำเสนอแนวความคิดใหม่ในการจัดการภาระงานโดยขึ้นกับพฤติกรรมการใช้งาน โดยในการทดลองได้แบ่งออกเป็น 2 ส่วน ส่วนแรก สํารวจแนวโน้มพฤติกรรมการใช้บริการจากเครื่องเซิร์ฟเวอร์โดยใช้เทคนิคเหมืองข้อมูล ส่วนที่สอง ทำนายการใช้งานทรัพยากรหน่วยประมวลผลและหน่วยความจำบนพื้นฐานของการใช้งานจริงโดยใช้การทำนายจากหลายวิธี ได้แก่ การค้นหาความสัมพันธ์ แบบจำลอง ARIMA และเทคนิคการปรับเรียบแบบเอ็กโปเนนเชียล ประสิทธิภาพของตัวแบบในการทำนายได้แสดงให้เห็นระหว่างการใช้งานอย่างคุ่มค่ากับเวลาที่ตอบสนอง ผลการทดลองแสดงให้เห็นว่าพฤติกรรมของผู้ใช้งานมีความแตกต่างกันไปในแต่ละประเภทของบริการซึ่งมีผลต่อการกระจายภาระงานในแต่ละช่วงเวลา ประโยชน์ที่ได้รับจากแบบจำลองเวอร์ชวลไลเซชันคือการอนุญาตให้มีการจัดสรรทรัพยากรจากการทำนาย ซึ่งจะช่วยให้การจัดสรรทรัพยากรเกิดประสิทธิภาพ

ภาควิชา คณิตศาสตร์และวิทยาการ
คอมพิวเตอร์
สาขาวิชา วิทยาการคอมพิวเตอร์และ
เทคโนโลยีสารสนเทศ

ปีการศึกษา 2556

ลายมือชื่อนิสิต
ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก
ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม

5173813423 : MAJOR COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

KEYWORDS: SERVER VIRTUALIZATION / PREDICTING MODEL / RESOURCE

ALLOCATION / USER BEHAVIOR

DULYAWIT PRANGCHUMPOL: IMPROVING HETEROGENEOUS WORKLOAD PERFORMANCE IN SERVER VIRTUALIZATION BASED ON USER BEHAVIORS.

ADVISOR: ASSOC. PROF. PERAPHON SOPHATSATHIT, Ph.D., CO-ADVISOR:

ASSOC. PROF. PANJAI TANTASANAWONG, Ph.D., 52 pp.

Large enterprises have many servers which must be managed and operated to support heterogeneous service. The emphasis, however, is to reduce power consumption of these servers for IT applications. To cope with such dilemma, the virtualization technique is employed which has gained attention in recent years. Server virtualization can improve hardware utilization by consolidate servers and also reduce physical space and power consumption in data center. However, the challenge is management of heterogeneous workloads in this system. This research proposes a new concept for managing workloads based on user's behavior. The approach is divided into two parts. First, exploring the trend of user's behavior in utilizing the service of each server with the help of data mining technique. Next, predicting hardware resources, CPU, and memory based on their actual operating profile using various prediction algorithms such as association rule discovery, ARIMA model, and exponential smoothing technique. A prediction performance model is devised to demonstrate between utilization and response time. The results show that user's behavior are different in each type of service which affects the distribution of workload and time. The benefits obtained from the virtualization model permit assignment of resources from the prediction, thereby resource allocation can be carried out efficiently.

Department: Mathematics and
Computer Science

Field of Study: Computer Science and
Information Technology

Academic Year: 2013

Student's Signature

Advisor's Signature

Co-Advisor's Signature



ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Associate Professor Dr. Peraphon Sophatsathit and Associate Professor Dr. Panjai Tantasanawong who have provided me with a great opportunity to do my research work under their guidance. Their suggestions and comments have motivated me during my Ph.D. study.

I would like to thank dissertation committees, Professor Dr. Chidchanok Lursinsap, Assistant Professor Dr. Suphakant Phimoltares, Dr. Boonyarit Intiyot, and Dr. Chalernpol Charnsripinyo for their advices and guidance about research activities. Much of untold appreciations also go to those unaccounted individuals who have lent a helping hand to support this work.

Last but not least, I would like to express my sincere gratitude and deep appreciation to my parents, my family, and my friends for constant encouragement, love, and supports throughout my life.



CONTENTS

	Page
THAI ABSTRACT	iv
ENGLISH ABSTRACT	v
ACKNOWLEDGEMENTS	vi
CONTENTS	vii
List of Tables	x
List of Figures.....	xi
List of Figures.....	xii
CHAPTER I.....	1
INTRODUCTION.....	1
1.1 Problem Identification and Motivation.....	1
1.2 Research Objectives	2
1.3 Scope of the Study	2
1.4 Problem Statements	3
1.5 Research Contribution	3
1.6 Related Definitions	3
1.7 Organization of the Dissertation	4
CHAPTER II.....	5
LITERATURE REVIEW	5
2.1 Related Literature	5
2.2 Theoretical Background.....	6
2.2.1 Virtualization Backgrounds	7
2.2.2 Heavy-tailed Distribution	8
2.2.3 Exponential Smoothing Method.....	9
2.2.3.1 Simple Exponential Smoothing	9
2.2.3.2 Double Exponential Smoothing.....	10
2.2.3.3 Triple Exponential Smoothing.....	10
2.2.4 Association Rule Discovery	11



	Page
2.2.5 Autoregressive Integrated Moving Average (ARIMA).....	11
2.2.6 Resource Utilization	12
CHAPTER III.....	14
PROPOSED FRAMEWORK	14
3.1 User Behavior Analysis.....	14
3.1.1 Association Model for Analyzing User Behavior	19
3.1.2 Correlation between User Access and Workload.....	21
3.1.3 Workload Distribution	23
3.2 Consumption of Hardware Resources Analysis	26
3.2.1 Simulation for Resource Analysis and Prediction.....	26
3.2.2 Algorithm for Resources Prediction and Allocation.....	30
3.2.3 Problem Scenario	32
CHAPTER IV.....	34
EXPERIMENTAL RESULTS AND DISCUSSION.....	34
4.1 Data set of the experiment	34
4.2 Experiment 1: Results of User Behavior Analysis	34
4.3 Experiment 2: Results of Hardware Resources Consumption Analysis.....	36
4.4 Experiment 3: Comparison with Other Predicting Models	37
4.4.1 Comparing with Association Rule	37
4.4.2 Comparing with ARIMA.....	41
4.4.3 Summary of Comparison.....	42
4.5 Experiment 4: Effect of Compromising Factor on Resource Allocation	42
CHAPTER V	46
CONCLUSION AND FUTURE RESEARCH.....	46
5.1 Contributions and Implications of this Research.....	46
5.2 Future Research	46
REFERENCES	47



VITA..... 52



List of Tables

Table	Page
3.1 Examples of user access to the proxy at 10:00 AM.....	15
3.2 Examples of data size in the proxy at 10:00 AM.....	16
3.3 Examples of rules for prediction.....	20
3.4 Prediction model of proxy server.....	21
3.5 The coefficients of the correlations between user access and data size for each day of the week.....	22
4.1 Example of test model on Monday for proxy server.....	35
4.2 Sum square error with alpha values for simple exponential.....	37
4.3 Sum square error with alpha and gamma values for double exponential.....	37
4.4 Examples of association rules for predicting.....	38
4.5 Examples of predicting by association rule with confidence and support.....	38
4.6 Level of memory usage in 24 hours by association rules.....	39
4.7 Memory usage in percentage of 24 hours by association rules.....	39
4.8 MSE Comparison for method prediction.....	42
4.9 Resource utilization with allocation at u65 and u75.....	44
4.10 Resource utilization with different utilization boundary.....	45



2211476350

List of Figures

Figure	Page
2.1 Typical server virtualization architecture consisting of a pool of heterogeneous servers.....	7
2.2 An example of ARIMA model analyze resource usage in database server...	12
2.3 An example of the relationship between resource utilization and response time.....	13
2.4 Response time curves showing knee values.....	13
3.1 Example log file from proxy server.....	14
3.2 Example log file from web server.....	15
3.3 User access for the proxy in each day.....	16
3.4 User access for the proxy in each hour.....	17
3.5 User access for the web server in each day.....	17
3.6 User access for the web server in each hour.....	18
3.7 Workload for the proxy server.....	18
3.8 Workload for the web server.....	18
3.9 Data size (sorted by descending size) in Proxy server over 24 hour period..	23
3.10 Data size (sorted by descending size) in Web server over 24 hour period ..	24
3.11 Data size in proxy server over a 24 hour period (cumulative percentage)...	24
3.12 Pareto distribution from proxy server over a 24 hour period (cumulative percentage).....	25
3.13 Data size in web server over a 24 hour period (cumulative percentage)....	25
3.14 Pareto distribution from web server over a 24 hour period (cumulative percentage).....	26
3.15 Program VirtualBox for simulation server.....	27
3.16 Set-up server virtualization.....	28
3.17 CPU consumption in the proxy.....	28
3.18 Memory consumption in the proxy.....	29
3.19 CPU consumption in the web server.....	29
3.20 Memory consumption in the web server.....	29
4.1 Resource prediction by exponential smoothing method for three servers..	36
4.2 Comparison of resource prediction by association rules and the proposed method using double exponential smoothing method for three servers.....	40
4.3 ARIMA model resource prediction of three servers.....	41



List of Figures

Figure		Page
4.4	CPU allocation with compromising factor $u = 65\%$ for database server.....	43
4.5	Memory allocation with compromising factor $u = 65\%$ for database server	43

