การวิเคราะห์เปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่ในพร็อกซีแคช

นาย พรทวี วัฒนวิทูกูร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2542

ISBN 974-334-076-9

ลิขสิทธิ์ของ จุฬาลงกรณ์มหาวิทยาลัย.

I19190372

COMPARATIVE PERFORMANCE ANALYSIS OF PROXY CACHING REPLACEMENT ALGORITHMS

Mr. Porntawee Wattanavitukul

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Computer Science

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 1999

ISBN 974-334-076-9

หัวข้อวิทยานิพนธ์	การวิเคราะห์เปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่ในพร็อกซีแคช
โดย	นาย พรทวี วัฒนวิทูกูร
ภาควิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ ดร.ณัฐวุฒิ หนูไพโรจน์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.สมศักดิ์ ปัญญาแก้ว)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(อาจารย์ ดร.ยรรยง เต็งอำนวย)

.....อาจารย์ที่ปรึกษา
(อาจารย์ ดร.ณัฐวุฒิ หนูไพโรจน์)

.....กรรมการ
(อาจารย์ จารุมาตร ปิ่นทอง)

.....กรรมการ
(รองศาสตราจารย์ ดร.วันชัย ริ้วไพบูลย์)

พรทวี วัฒนวิทูกูร : การวิเคราะห์เปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่ในพร็อกซีแคช (COMPARATIVE PERFORMANCE ANALYSIS OF PROXY CACHING REPLACEMENT ALGORITHMS) อ.ที่ปรึกษา : ดร.ณัฐวุฒิ หนูไพโรจน์, 76 หน้า, ISBN 974-334-076-9

ปริมาณความต้องการใช้เว็บที่เพิ่มขึ้นอย่างรวดเร็วในช่วง 2-3 ปีที่ผ่านมาส่งผลให้ผู้ให้บริการอินเทอร์เน็ตแก้ปัญหาโดยการเพิ่มแบนด์วิดท์ในการเชื่อมต่อสู่อินเทอร์เน็ต งานวิจัยที่ผ่านมาได้เสนอการแก้ไขขั้นตอนวิธีการแทนที่หลายวิธีเพื่อที่จะปรับปรุงประสิทธิภาพของพร็อกซี ทั้งนี้การทดสอบที่ผ่านมาไม่ได้พิจารณาถึงช่วงเวลาในการใช้เว็บความจุของพร็อกซีแคชและผลกระทบของสภาพการใช้เว็บ ซึ่งจากการศึกษาเบื้องต้นพบว่าส่งผลโดยตรงต่อประสิทธิภาพของขั้นตอนวิธีการแทนที่ในพร็อกซีแคช ดังนั้นงานวิจัยนี้ได้ทำการศึกษาเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่ที่ได้ศึกษามาแล้วในงานวิจัยที่ผ่านมากับข้อมูลการใช้เว็บที่จุฬาลงกรณ์มหาวิทยาลัยและศึกษาแนวโน้มความสัมพันธ์ของสภาพการใช้เว็บกับประสิทธิภาพของขั้นตอนวิธีการแทนที่ในพร็อกซีแคช

การทดสอบใช้วิธีการจำลองการทำงานของพร็อกซีแคชเป็นเครื่องมือในการประเมินขั้นตอนวิธีการแทนที่แบบต่างๆ โดยใช้ข้อมูลการใช้เว็บจากจุฬาลงกรณ์มหาวิทยาลัยเพื่อพิจารณาประสิทธิภาพของขั้นตอนการแทนที่ รวมทั้งพิจารณาแนวโน้มความสัมพันธ์ระหว่างสภาพการใช้เว็บและประสิทธิภาพของขั้นตอนวิธีการแทนที่ ผลการจำลองการทำงานของพร็อกซีแคชพบว่าขั้นตอนวิธีการแทนที่ที่เหมาะสมกับสภาพการใช้เว็บช่วงเวลาที่มีการใช้เว็บมากและช่วงเวลาที่มีการใช้เว็บน้อยคือขั้นตอนการแทนที่แบบเดียวกันได้แก่ LRU-MIN สำหรับพร็อกซีแคชที่มีความจุน้อยและ SIZE สำหรับพร็อกซีแคชที่มีความจุมากในกรณีที่ใช้ค่าฮิตเรโชเป็นเครื่องวัด กรณีที่ใช้ค่าไบท์ฮิตเรโชเป็นเครื่องวัดพบว่า LRU-TH และ Perfect-LFU เป็นขั้นตอนวิธีการแทนที่ที่เหมาะสมกับพร็อกซีแคชที่มีความจุน้อย ส่วน LRU และ LRU-MIN เป็นขั้นตอนวิธีการแทนที่ที่เหมาะสมกับพร็อกซีแคชที่มีความจุมาก ดังนั้นการพิจารณาเลือกใช้ขั้นตอนวิธีการแทนที่ให้เหมาะสมกับสภาพการใช้เว็บขึ้นกับการเลือกเครื่องวัดประสิทธิภาพที่เหมาะสมสำหรับพร็อกซีแคช สำหรับสภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัยซึ่งต้องการลดปริมาณในการถ่ายโอนข้อมูลจากภายนอกเนื่องจากแบนด์วิดท์ออกสู่อินเทอร์เน็ตไม่เพียงพอ การพิจารณาปรับปรุงความต้องการใช้แบนด์วิดท์เพื่อรับข้อมูลเป็นสิ่งสำคัญดังนั้นการเลือกใช้ค่าไบท์ฮิตเรโชเป็นเครื่องวัดจึงเหมาะสมที่สุดสำหรับสภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา วิศวกรรมคอมพิวเตอร์	ลายมือชื่อนิสิต พรทวี วัฒนวิทูกูร
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์	ลายมือชื่ออาจารย์ที่ปรึกษา
ปีการศึกษา 2542	ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

4170427021 : MAJOR **COMPUTER SCIENCE**

KEY WORD: WWW / PROXY CACHING REPLACEMENT ALGORITHM / CACHE

PORNTAWEE WATTANAVITUKUL : COMPARATIVE PERFORMANCE ANALYSIS OF PROXY CACHING REPLACEMENT ALGORITHMS.

THESIS ADVISOR : Dr. NATAWUT NUPAIROJ, Ph. D. 76 pp. ISBN 974-334-076-9

Due to the rapidly increasing of web access demand in last 2 to 3 years, many Internet service providers decided to increase the bandwidth of their Internet links. Recent studies suggested several approaches to modify replacement algorithms to further improve the effectiveness of the proxies. However the period of web access time, cache size and the impact of web access condition have not been taken into consideration. The preliminary study show the high impact of period of web access time to proxy caching performance. In this study we present the comparative performance analysis of proxy caching replacement algorithms and the trend of relationship between web access conditions and replacement algorithms performance using web access condition at Chulalongkorn University.

In our experiment, the proxy caching simulation is used as a tool to evaluate various replacement algorithms using traces from Information Technology Chulalongkorn University. In addition we study the trend of relationship between web access condition and replacement algorithm performance. Our studies indicate that, using hit ratio as a metric, LRU-MIN outperforms others for small cache size while SIZE is considered the best for large cache size. When we consider byte hit ratio, LRU-TH and Perfect-LFU are the best algorithms for small cache size while LRU and LRU-MIN are the best algorithms for large cache size. As the result, selecting suitable replacement algorithm for the web access condition depends on choosing appropriate performance metric for proxy cache. For web access condition at Chulalongkorn University, which has limited bandwidth to the Internet, optimizing the bandwidth requirement of incoming data transfer is the most important goal. Thus, using byte hit ratio as a metric is appropriate under this circumstance.

ภาควิชา วิศวกรรมคอมพิวเตอร์	ลายมือชื่อนิสิต พรทวี วัฒนวิทยกุล
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์	ลายมือชื่ออาจารย์ที่ปรึกษา
ปีการศึกษา 2542	ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของ ดร.ณัฐวุฒิ หนูไพโรจน์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้ให้คำแนะนำและข้อคิดเห็นต่าง ๆ ในการวิจัยมาด้วยดีตลอด

ท้ายนี้ ผู้วิจัยใคร่ขอกราบขอบพระคุณ บิดา-มารดา ซึ่งสนับสนุนและให้กำลังใจแก่ผู้วิจัยเสมอจนสำเร็จการศึกษา

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญ ..... ช
สารบัญตาราง ..... ฌ
สารบัญรูปภาพ ..... ญ
บทที่
1 บทนำ 1
1.1 ความเป็นมาของปัญหา 1
1.2 งานวิจัยที่เกี่ยวข้อง 2
1.3 วัตถุประสงค์ของการวิจัย 3
1.4 ขอบเขตของการวิจัย 3
1.5 ขั้นตอนการทำวิจัย 4
1.6 ประโยชน์ที่คาดว่าจะได้รับ 5
2 ทฤษฎีและแนวความคิดที่ใช้ 6
2.1 เว็บแคช 6
2.2 ขั้นตอนวิธีการแทนที่ 7
2.3 สภาพการใช้เว็บ 7
2.4 การทดสอบแบบแฟคทอเรียล 10
3 การวิเคราะห์สภาพการใช้เว็บ 15
3.1 ข้อมูลการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย 16
3.2 ข้อมูลการใช้เว็บจากพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย 19
3.3 ข้อมูลการใช้เว็บจากเอ็นแอลเอเอ็นอาร์ 22
3.4 สรุปผลการวิเคราะห์สภาพการใช้เว็บ 25
4 การทดสอบเปรียบเทียบประสิทธิภาพขั้นตอนวิธีการแทนที่ 26
4.1 วิธีการทดสอบเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่ 26
4.2 เครื่องมือที่ใช้ในการจำลองการทำงานของพร็อกซีแคช 28
4.3 การทดสอบเปรียบเทียบขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัย 29
4.4 การทดสอบเปรียบเทียบขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บที่อื่นๆ 35

สารบัญ (ต่อ)

บทที่

4.5 แนวทางการนำผลการทดสอบไปประยุกต์ใช้กับสภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัย 38

4.6 สรุปผลการทดสอบเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่ 40

5 แนวโน้มความสัมพันธ์ระหว่างสภาพการใช้เว็บและประสิทธิภาพของขั้นตอนวิธีการแทนที่ 42

5.1 ปริมาณการเรียกขอ 42

5.2 ปริมาณการเรียกขอประเภท "Successful" 43

5.3 สัดส่วนของจำนวนเอกสารที่แตกต่างกันกับจำนวนการเรียกขอทั้งหมด 44

5.4 พารามิเตอร์จากกฎของ Zipf 45

5.5 ช่วงเวลาการใช้เว็บ 47

6 สรุปผลการวิจัยและข้อเสนอแนะ 48

รายการอ้างอิง 50

ภาคผนวก 52

ภาคผนวก ก 53

ภาคผนวก ข 54

ภาคผนวก ค 56

ภาคผนวก ง 70

ภาคผนวก จ 72

ประวัติผู้วิจัย 76

สารบัญตาราง

หน้า

ตารางที่ 2.1 ขั้นตอนวิธีการแทนที่และลำดับการแทนที่ของเอกสาร 7
ตารางที่ 2.2 การแยกประเภทของเอกสารตามสกุลของไฟล์ 8
ตารางที่ 2.3 พารามิเตอร์ของการกระจายแบบไฮบริด ล็อกนอมอล-พาเรโต 9
ตารางที่ 2.4 แสดงข้อมูลของการทดสอบแบบแฟคทอเรียลที่มีแฟคเตอร์ 2 ตัว 10
ตารางที่ 2.5 การวิเคราะห์ความแปรปรวนสำหรับการทดลองแบบแฟคทอเรียลที่มี 2 แฟคเตอร์ 12
ตารางที่ 2.6 เวลาที่ซีพียูใช้ในการประมวลผล (มิลลิวินาที) 13
ตารางที่ 2.7 ตารางการวิเคราะห์ความแปรปรวนเพื่อใช้ในการทดลองแบบแฟคทอเรียล 13
ตารางที่ 3.1 ช่วงเวลาที่มีการใช้เว็บมากและน้อย ของการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย วันที่ 9 – 16 ต.ค. 2542 17
ตารางที่ 3.2 ข้อมูลการใช้เว็บที่ใช้ในการทดสอบ และสภาพการใช้เว็บแต่ละช่วงเวลาของการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย วันที่ 9 – 16 ต.ค. 2542 18
ตารางที่ 3.3 ช่วงเวลาที่มีการใช้เว็บมากและน้อยของข้อมูลการใช้เว็บจากพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย วันที่ 10 – 14 ต.ค. 2542 20
ตารางที่ 3.4 ข้อมูลการใช้เว็บที่ใช้ในการทดสอบ และสภาพการใช้เว็บแต่ละช่วงเวลาของการใช้เว็บจากพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย วันที่ 10 – 14 ต.ค. 2542 21
ตารางที่ 3.5 ช่วงเวลาที่มีการใช้เว็บมากและน้อยของข้อมูลการใช้เว็บจากเอ็นแอลเอเอ็นอาร์ วันที่ 8 – 10 ต.ค. 2542 23
ตารางที่ 3.6 ข้อมูลการใช้เว็บที่ใช้ในการทดสอบ และสภาพการใช้เว็บแต่ละช่วงเวลาของการใช้เว็บจากเอ็นแอลเอเอ็นอาร์ วันที่ 8 – 10 ต.ค. 2542 24
ตารางที่ 4.1 ค่าเฉลี่ยของค่าฮิตเรโช และ ไบท์ฮิตเรโช จากแฟ้มที่บันทึกการเรียกขอ 38
ตารางที่ 4.2 แฟคเตอร์ที่ใช้ในการทดสอบแบบแฟคทอเรียลสำหรับการใช้ค่าฮิตเรโช เป็นเครื่องวัด 39
ตารางที่ 4.3 แฟคเตอร์ที่ใช้ในการทดสอบแบบแฟคทอเรียลสำหรับการใช้ค่าไบท์ฮิตเรโช เป็นเครื่องวัด 39
ตารางที่ 4.4 ผลการทดสอบแฟคทอเรียลเพื่อพิจารณาความสัมพันธ์ระหว่างขนาดของพร็อกซีแคชและขั้นตอนวิธีการแทนที่ที่มีต่อประสิทธิภาพของพร็อกซีแคช 39
ตารางที่ 4.5 ประสิทธิภาพของพร็อกซีแคชที่เพิ่มขึ้นโดยใช้ค่าฮิตเรโช เป็นเครื่องวัด 40
ตารางที่ 4.6 ขั้นตอนวิธีการแทนที่ที่เหมาะสมกับพร็อกซีแคชที่มีความจุและเครื่องวัดที่กำหนด 40
ตารางที่ 5.1 ค่าเฉลี่ยของข้อมูลการใช้เว็บช่วงที่มีการใช้เว็บมากและน้อย 47
ตารางที่ 6.1 ขั้นตอนวิธีการแทนที่ที่เหมาะสมกับพร็อกซีแคชที่มีความจุและเครื่องวัดที่กำหนด 49

สารบัญรูปภาพ

หน้า

รูปที่ 2.1 การแบ่งระดับของเว็บแคช 6
รูปที่ 2.2 กราฟแอลแอลซีดี 8
รูปที่ 2.3 การกระจายของไฟล์ขนาดเล็ก 9
รูปที่ 2.4 การแจกแจงของความถี่ในการเรียกขอเอกสาร 9
รูปที่ 2.5 การทดสอบแบบแฟคทอเรียลที่ไม่มีปฏิกิริยาระหว่างแฟคเตอร์ 12
รูปที่ 2.6 การทดสอบแบบแฟคทอเรียลที่มีปฏิกิริยาระหว่างแฟคเตอร์ 12
รูปที่ 3.1 ปริมาณการเรียกขอแต่ละชั่วโมงของข้อมูลการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย วันที่ 9 – 16 ต.ค.2542 16
รูปที่ 3.2 ปริมาณการเรียกขอเฉลี่ยแต่ละชั่วโมงของข้อมูลการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย วันที่ 9 – 16 ต.ค. 2542 17
รูปที่ 3.3 ปริมาณการเรียกขอแต่ละชั่วโมงของข้อมูลการใช้เว็บจากพร็อกซีเซิร์ฟเวอร์ระดับคณะของ จุฬาลงกรณ์มหาวิทยาลัย วันที่ 10 - 14 ต.ค. 2542 19
รูปที่ 3.4 ปริมาณการเรียกขอเฉลี่ยแต่ละชั่วโมงของข้อมูลการใช้เว็บจากพร็อกซีเซิร์ฟเวอร์ระดับคณะ ของจุฬาลงกรณ์มหาวิทยาลัย วันที่ 10 – 14 ต.ค. 2542 20
รูปที่ 3.5 ปริมาณการเรียกขอแต่ละชั่วโมงของข้อมูลการใช้เว็บจากเอ็นแอลเอเอ็นอาร์ วันที่ 8 - 10 ต.ค. 2542 22
รูปที่ 3.6 ปริมาณการเรียกขอเฉลี่ยแต่ละชั่วโมงของข้อมูลการใช้เว็บจากเอ็นแอลเอเอ็นอาร์ วันที่ 8 – 10 ต.ค. 2542 23
รูปที่ 4.1 แผนภาพแสดงการทำงานของโปรแกรมจำลองการทำงานของพร็อกซีแคช 28
รูปที่ 4.2 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศของจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าไบท์ฮิตเรโช 29
รูปที่ 4.3 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศของจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าฮิตเรโช 30
รูปที่ 4.4 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศของจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าไบท์ฮิตเรโช 31

สารบัญรูปภาพ (ต่อ)

รูปที่ 4.5 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศของจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าฮิตเรโช 31
รูปที่ 4.6 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย โดยพิจารณาจากค่าไบท์ฮิตเรโช 32
รูปที่ 4.7 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย โดยพิจารณาจากค่าฮิตเรโช 33
รูปที่ 4.8 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าไบท์ฮิตเรโช 34
รูปที่ 4.9 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าฮิตเรโช 34
รูปที่ 4.10 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย โดยพิจารณาจากค่าไบท์ฮิตเรโช 35
รูปที่ 4.11 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย โดยพิจารณาจากค่าฮิตเรโช 36
รูปที่ 4.12 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของเอ็นแอลเอเอ็นอาร์โดยพิจารณาจากค่า ไบท์ฮิตเรโช 37
รูปที่ 4.13 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของเอ็นแอลเอเอ็นอาร์โดยพิจารณาจากค่าฮิตเรโช 37
รูปที่ 5.1 ปริมาณการเรียกขอและประสิทธิภาพของขั้นตอนวิธีการแทนที่ 42
รูปที่ 5.2 ปริมาณการเรียกขอประเภท "Successful" และประสิทธิภาพของขั้นตอนวิธีการแทนที่ 43
รูปที่ 5.3 สัดส่วนของจำนวนเอกสารที่แตกต่างกันและประสิทธิภาพของขั้นตอนวิธีการแทนที่ 44
รูปที่ 5.4 ค่าอัลฟาและประสิทธิภาพของขั้นตอนวิธีการแทนที่ 45
รูปที่ 5.5 ค่าโอเมกาและประสิทธิภาพของขั้นตอนวิธีการแทนที่ 46

บทที่ 1

บทนำ

1.1. ความเป็นมาของปัญหา

ปัจจุบันความต้องการใช้บริการเว็บบนเครือข่ายอินเทอร์เน็ตมีปริมาณเพิ่มขึ้นมากเนื่องจากมีรูปแบบการใช้งานที่ง่ายและปริมาณของข้อมูลบนอินเทอร์เน็ตที่เพิ่มขึ้นอย่างรวดเร็ว โดยในระยะเวลา 2- 3 ปีที่ผ่านมามีการประมาณว่าปริมาณข้อมูลของเว็บบนอินเทอร์เน็ตเพิ่มขึ้นโดยเฉลี่ย 20 ล้านหน้าต่อเดือน [9] ซึ่งข้อมูลส่วนใหญ่อยู่ในทวีปอเมริกาและยุโรป ส่งผลให้ค่าใช้จ่ายในการเช่าสายเพื่อเชื่อมต่อออกสู่อินเทอร์เน็ตสำหรับประเทศไทยมีราคาสูงเนื่องจากระยะทางที่ไกลเมื่อเทียบกับประเทศที่อยู่ในทวีปอเมริกาและยุโรป

องค์กรที่ให้บริการด้านอินเทอร์เน็ต สถาบันการศึกษา และองค์กรที่ทำงานวิจัยทางด้านวิทยาศาสตร์หลายแห่งได้ติดตั้งซอฟต์แวร์หรือฮาร์ดแวร์เพื่อทำหน้าที่จัดเก็บข้อมูลที่ถูกถ่ายโอนผ่านโพรโตคอลเอชทีทีพี (HTTP) เพื่อลดค่าใช้จ่ายในการเชื่อมต่อสู่อินเทอร์เน็ต เนื่องจากซอฟต์แวร์ที่ทำหน้าที่ในการจัดการระบบเว็บแคชถูกพัฒนามาจากระบบแคชของซีพียู (CPU) โดยอาศัยขั้นตอนวิธีการแทนที่ (Replacement algorithm) มาจากการจัดการหน่วยความจำ ซึ่งได้มีการพิสูจน์ว่าขั้นตอนวิธีการแทนที่ที่เคยใช้ได้ผลดีในระบบแคชของซีพียู เช่น แอลอาร์ยู (LRU : Least Recently Used) ไม่เหมาะสมที่จะนำมาใช้กับการจัดการแคชข้อมูลบนอินเทอร์เน็ต เพราะลักษณะของข้อมูลในเว็บมีขนาดที่ไม่เท่ากันและลักษณะของการเรียกขอมาจากผู้ใช้หลายคน [1] [2]

ในช่วง 2 ปีที่ผ่านมาได้มีการวิจัยเพื่อเปรียบเทียบขั้นตอนวิธีการแทนที่ของเว็บแคชอย่างกว้างขวาง [1] [2] [10] [11] การเปรียบเทียบประสิทธิภาพนี้มีทั้งที่เปรียบเทียบขั้นตอนวิธีการแทนที่ที่มีอยู่แล้ว และเสนอขั้นตอนวิธีการแทนที่แบบใหม่ ซึ่งผู้เสนอเชื่อว่ามีประสิทธิภาพมากกว่าวิธีเดิม แต่การทดสอบเปรียบเทียบที่ผ่านๆ มาได้ทำการทดสอบภายใต้สภาพการใช้เว็บเพียงแห่งเดียว ทดสอบด้วยข้อมูลการใช้เว็บที่มีช่วงเวลาตั้งแต่ 2 วันจนถึง 1 ปี และไม่ได้คำนึงถึงลักษณะการใช้เว็บที่แตกต่างกันสำหรับแต่ละช่วงเวลาของวัน รวมทั้งไม่ได้พิจารณาถึงความสัมพันธ์ระหว่างสภาพการใช้เว็บกับประสิทธิภาพของขั้นตอนวิธีการแทนที่ ทำให้ผลการทดสอบแสดงได้เพียงประสิทธิภาพโดยรวมของเว็บแคชที่นำข้อมูลการใช้เว็บมาทดสอบที่เพิ่มขึ้นเมื่อใช้ขั้นตอนวิธีการแทนที่ที่ได้เสนอโดยใช้ค่าฮิตเรโช (hit ratio) และ ค่าไบท์ฮิตเรโช (byte hit ratio) เป็นเครื่องวัด (metric) และไม่สามารถประมาณประสิทธิภาพของขั้นตอนวิธีการแทนที่ในเว็บแคชที่อื่นๆได้

งานวิจัยนี้จะนำเสนอการทดสอบเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่สำหรับสภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัย โดยการทดสอบจะใช้การจำลองการทำงานของพร็อกซีแคชที่มีขั้นตอนวิธีการแทนที่แบบต่างๆ กับข้อมูลการใช้เว็บในช่วงเวลาครึ่งวันที่มีปริมาณการใช้เว็บมากและน้อยของข้อมูลการใช้เว็บที่รวบรวมมาจากพร็อกซีแคชจากสำนักเทคโนโลยีสารสนเทศของจุฬาลงกรณ์มหาวิทยาลัย รวมทั้งใช้ข้อมูลการใช้เว็บจากเอ็นแอลเอเอ็นอาร์ (NLANR) มาเป็นตัวเปรียบเทียบ ผลการทดสอบจะแสดงให้เห็นถึงประสิทธิภาพของขั้นตอนวิธีการแทนที่แต่ละแบบภายใต้สภาพการใช้เว็บของ

จุฬาลงกรณ์มหาวิทยาลัย โดยใช้ค่าฮิตเรโช และ ไบท์ฮิตเรโช เป็นเครื่องวัด และแสดงให้เห็นถึงความสัมพันธ์ระหว่างสภาพการใช้เว็บกับประสิทธิภาพของขั้นตอนวิธีการแทนที่แบบต่างๆ ซึ่งสามารถนำไปใช้ในการประมาณประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของเว็บแคชที่อื่นๆ ได้

1.2. งานวิจัยที่เกี่ยวข้อง

1.2.1. การหาลักษณะการใช้เว็บ

1.2.1.1. การแทนลักษณะปริมาณการเรียกขอเอกสารเดียวกันในช่วงเวลาที่จำกัด โดยการวิเคราะห์ LRU Stack Depth [12]

1.2.1.2. การแทนลักษณะการแจกแจงของขนาดไฟล์ด้วยการแจกแจงแบบล็อกนอมอล (lognormal) สำหรับไฟล์ขนาดเล็กและการแจกแจงแบบพาเรโต (pareto) สำหรับไฟล์ขนาดใหญ่ [4] [6]

1.2.1.3. การแทนลักษณะความถี่ในการเรียกขอเอกสารของการใช้เว็บโดยใช้กฎของ Zipf

ก. Steven Glassman [3] ได้นำเสนอการแทนลักษณะความถี่ในการเรียกขอเอกสารของข้อมูลการใช้เว็บจาก Digital Equipment Corporation's Palo Alto ด้วยการแจกแจงแบบ Zipf และ ปี พ.ศ. 2538 Carlos Cunha [4] พบว่าความถี่ในการเรียกขอเอกสารของการเรียกขอข้อมูลเป็นไปตามกฎของ Zipf โดยมีค่าอัลฟาเท่ากับ 0.986

ข. Norifumi Nishikawa [5] ใช้กฎของ Zipf กับการเรียกขอมายังเครื่องให้บริการ (server) ได้ค่าอัลฟาเท่ากับ 0.75 ซึ่งต่างไปจากการแจกแจงแบบ Zipf จึงได้เสนอการแจกแจงแบบ Zipf-Like

ค. Paul Barford [6] เปรียบเทียบค่าอัลฟาของข้อมูลการใช้เว็บในปี พ.ศ. 2538 และ พ.ศ. 2541 จากที่เดียวกัน ซึ่งค่าอัลฟาลดลงจาก 0.83 เป็น 0.65 เพราะพฤติกรรมของผู้ใช้ที่เปลี่ยนไปโดยมีความสนใจที่กว้างขวางมากขึ้นและประสิทธิภาพของระบบเครือข่ายอินเทอร์เน็ตที่ดีขึ้น

ง. Lee Breslau [7] สรุปการใช้กฎของ Zipf กับการใช้บริการเว็บว่าลักษณะความถี่ในการเรียกขอเอกสารของการเรียกขอที่มายังพร็อกซีเซิร์ฟเวอร์มีการแจกแจงแบบ Zipf-Like

1.2.2. ขั้นตอนวิธีการแทนที่

การพัฒนาขั้นตอนวิธีการแทนที่สำหรับเว็บแคชแบ่งออกได้เป็น 2 ประเภท คือ

1.2.2.1. ขั้นตอนวิธีการแทนที่ที่พัฒนาจากขั้นตอนวิธีการแทนที่ของระบบการจัดการหน่วยความจำ ได้แก่

ก. Marc Abrams [1] พัฒนา LRU-MIN และ LRU-Threshold โดยทั้ง 2 ขั้นตอนวิธีให้ประสิทธิภาพที่ดีกว่า LRU เนื่องจากมีการพิจารณาขนาดของไฟล์ในการแทนที่ด้วย และ สำหรับแคชที่มีเนื้อที่เก็บข้อมูลน้อย LRU-Threshold จะให้ประสิทธิภาพที่ดีกว่า LRU-MIN

ข. James E. Pitkow [14] เสนอแนะขั้นตอนวิธีการแทนที่ชื่อว่า Pitkow / Recker ที่พัฒนามาจากรูปแบบระบบความจำของมนุษย์มีหลักการทำงาน คือ แทนที่เอกสารที่ถูกเรียกใช้น้อยที่สุดก่อน โดยใช้ขั้นตอนวิธีการแทนที่แบบ LRU แต่ถ้าเอกสารทั้งหมดมีอายุในแคชไม่ถึงเวลาที่กำหนดไว้ (ช่วงเวลาที่วิเคราะห์ไว้คือ 1 วัน) จะแทนที่เอกสารที่มีขนาดใหญ่สุดก่อน อย่างไรก็ตามจากการวิเคราะห์พบว่าการ

กำหนดช่วงเวลาที่จะใช้ขั้นตอนวิธีการแทนที่แบบใดแบบหนึ่งส่งผลโดยตรงต่อประสิทธิภาพของระบบเว็บแคช ดังนั้นการใช้ขั้นตอนวิธีการแทนที่ Pitkow / Recker จึงเหมาะกับการทำระบบเว็บแคชที่สามารถปรับเปลี่ยนการทำงานให้เหมาะสมกับลักษณะการใช้เว็บได้ (Adaptive web caching)

ค. Lee Breslau [7] พัฒนาขั้นตอนวิธีการแทนที่แบบ Perfect-LFU และ In-cache-LFU ซึ่ง Perfect-LFU จะให้ประสิทธิภาพที่ดีกว่า LRU และ In-cache-LFU จะให้ประสิทธิภาพน้อยกว่า LRU แต่การใช้ Perfect –LFU กับสภาพการใช้เว็บจริงจำเป็นต้องใช้หน่วยความจำมากเพื่อใช้ในการเก็บสถิติการเรียกขอ

1.2.2.2. ขั้นตอนวิธีการแทนที่ที่พัฒนาขึ้นใหม่เพื่อให้เหมาะสมกับสภาพการใช้เว็บ

ก. Stephen Williams [15] ได้นำขั้นตอนวิธีการแทนที่ที่พิจารณาแทนที่เอกสารที่มีขนาดใหญ่ที่สุดก่อน คือ ขั้นตอนวิธีการแทนที่ SIZE ซึ่งมีการออกแบบที่ง่ายเนื่องจากมีตัวแปรที่ใช้ในการพิจารณาแทนที่เพียงตัวเดียว และให้ประสิทธิภาพสูงกว่าวิธีการแทนที่แบบอื่นๆ

ข. Pei Cao [2] พัฒนาขั้นตอนวิธีการแทนที่ GreedyDual-Size โดยขั้นตอนวิธีนี้จะใช้ขนาดเอกสารและความเร็วในการถ่ายโอนแต่ละเอกสารมาพิจารณาในการแทนที่ ซึ่งผลที่ได้จากการทดสอบแสดงให้เห็นว่าสามารถลดเวลาที่ใช้ในการถ่ายโอนข้อมูลน้อยลงกว่าขั้นตอนวิธีการแทนที่แบบ LRU

1.2.3. การทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ แบ่งออกได้เป็น 2 ประเภท คือ

1.2.3.1. การทดสอบประสิทธิภาพโดยการใช้ข้อมูลจริงในการทดสอบ (Trace-driven simulation)

การนำเสนอขั้นตอนวิธีการแทนที่จากข้อ 1.2.2. มีการทดสอบประสิทธิภาพโดยการใช้ข้อมูลจริงในการทดสอบเนื่องจากการใช้ข้อมูลการใช้เว็บจริงในการทดสอบสามารถทำได้ง่ายและสามารถแสดงให้เห็นถึงประสิทธิภาพที่เพิ่มขึ้นจากการเปลี่ยนแปลงขั้นตอนวิธีการแทนที่ของเว็บแคช

1.2.3.2. การทดสอบประสิทธิภาพโดยการใช้ข้อมูลที่สร้างขึ้นในการทดสอบ (Model-driven - simulation)

ก. Ying Shi [8] ทดสอบขั้นตอนวิธีการแทนที่ด้วยการจำลองการทำงานของเว็บแคชโดยใช้ข้อมูลการใช้เว็บที่สร้างขึ้นให้มีลักษณะของความถี่ในการเรียกขอและลักษณะของลำดับการเรียกขอเป็นไปตามข้อมูลการใช้บริการเว็บเซิร์ฟเวอร์ และนำเสนอขั้นตอนวิธีการแทนที่แบบ LWU (Least Weighted Usage) ซึ่งสามารถปรับให้เหมาะสมกับลักษณะการใช้เว็บที่มีลักษณะของความถี่ในการเรียกขอและลำดับในการเรียกขอแต่ละระดับได้

1.3. วัตถุประสงค์

เพื่อวิเคราะห์เปรียบเทียบประสิทธิภาพของวิธีการแทนที่ในพร็อกซีแคช

1.4. ขอบเขตการวิจัย

1.4.1. ข้อมูลการใช้เว็บที่ใช้ในการทดสอบเปรียบเทียบ คือ ข้อมูลการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศของจุฬาลงกรณ์มหาวิทยาลัยอย่างน้อย 7 วัน โดยใช้ข้อมูลจากพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัยและ เอ็นแอลเอเอ็นอาร์เป็นตัวเปรียบเทียบในบางกรณี

1.4.2. สภาพการใช้เว็บ ประกอบด้วยข้อมูล 2 ส่วนได้แก่

1.4.2.1. ช่วงเวลาของวันที่มีการใช้เว็บมากกับน้อยช่วงละ 12 ชั่วโมง

1.4.2.2. ลักษณะการใช้เว็บแบ่งออกเป็น

ก. ปริมาณการเรียกขอ

ข. ปริมาณการเรียกขอที่ได้รับการตอบกลับพร้อมส่งเอกสารกลับมา (successful request)

ค. สัดส่วนของจำนวนเอกสารที่แตกต่างกันกับจำนวนการเรียกขอทั้งหมด

ง. พารามิเตอร์จากกฎของ Zipf

1.4.3. ขั้นตอนวิธีการแทนที่ที่นำมาทดสอบ ได้แก่

1.4.3.1. FIFO (First-In-First-Out)

1.4.3.2. LRU (Least Recently Used)

1.4.3.3. In – cache - LFU (Least Frequently Used)

1.4.3.4. Perfect – LFU (Least Frequently Used)

1.4.3.5. SIZE

1.4.3.6. LRU-MIN

1.4.3.7. LRU-TH (LRU-Threshold)

1.4.4. การทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ ทำโดยใช้การจำลองการทำงานของพร็อกซีแคชโดยอาศัยข้อมูลการใช้เว็บจากข้อ 1.4.1 และขั้นตอนวิธีการแทนที่ในข้อ 1.4.3

1.5. ขั้นตอนและวิธีการดำเนินการวิจัย

1.5.1. ศึกษาหลักการทำงานของเว็บแคช การทดสอบเปรียบเทียบประสิทธิภาพแบบต่างๆ และการหาลักษณะการใช้เว็บจากข้อมูลการใช้เว็บ

1.5.2. เก็บข้อมูลการใช้เว็บ

1.5.3. พัฒนาเครื่องมือสำหรับคัดเลือกการเรียกขอ (request) ที่ส่งผลต่อการจัดการแคช และหาลักษณะการใช้เว็บจากข้อมูลการใช้เว็บ

1.5.4. หาลักษณะการใช้เว็บจากข้อมูลการใช้เว็บ

1.5.5. พัฒนาเครื่องมือที่ใช้จำลองการทำงานของพร็อกซีแคช และ จำลองการทำงานของระบบแคชด้วยวิธีการแทนที่ข้อมูลต่างๆ

1.5.6. วิเคราะห์และสรุปผล

1.5.7. จัดทำรายงาน

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1.6. ประโยชน์ที่จะได้รับ

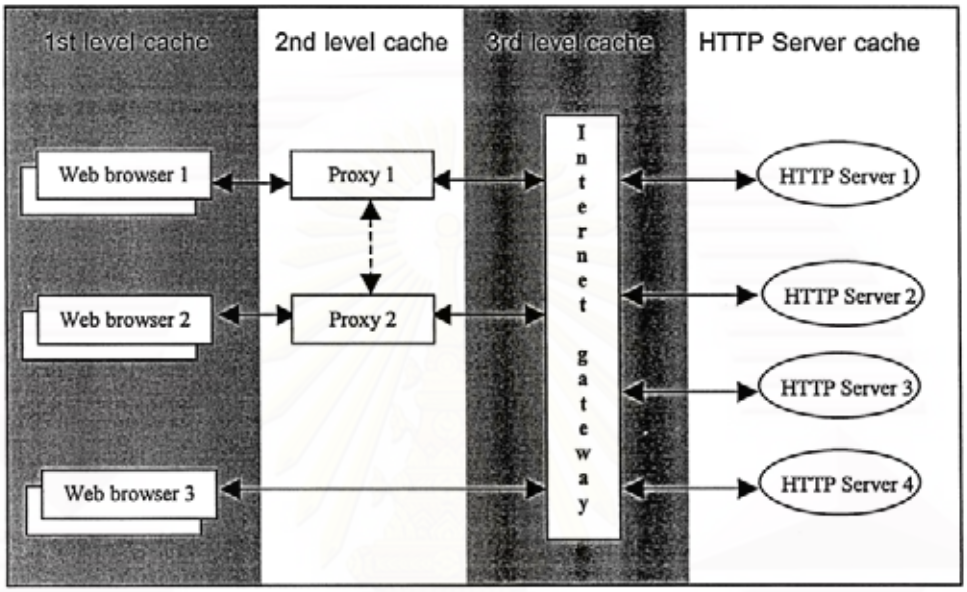
นำเสนอขั้นตอนวิธีการแทนที่ของพร็อกซีแคชที่เหมาะสมกับสภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัย ซึ่งสามารถนำไปใช้เพื่อปรับปรุงประสิทธิภาพของเว็บแคชได้

บทที่ 2

ทฤษฎีและแนวความคิดที่ใช้

2.1. เว็บแคช

เว็บแคช คือซอฟต์แวร์หรือฮาร์ดแวร์ที่ถูกนำมาใช้เพื่อช่วยลดปริมาณการถ่ายโอนข้อมูลเว็บผ่านระบบเครือข่ายโดยเว็บแคชแบ่งออกเป็นหลายระดับดังรูปที่ 2.1



รูปที่ 2.1 การแบ่งระดับของเว็บแคช

เว็บแคชแต่ละระดับมีการรองรับปริมาณการเรียกขอที่แตกต่างกันออกไป แต่มีหลักการทำงานที่เหมือนกัน คือ การทำสำเนาของข้อมูลที่เป็นที่นิยม มีขนาดใหญ่ หรือ น่าจะถูกเรียกใช้ในอนาคตไว้ใกล้ปลายทางที่สุดเพื่อลดเวลาการเดินทางของข้อมูล หรือช่วยลดภาระในการส่งข้อมูลสำหรับเว็บเซิร์ฟเวอร์ โดยใช้แนวคิดมาจากการออกแบบระบบจัดการหน่วยความจำของระบบปฏิบัติการและระบบเครือข่าย เนื่องจากเนื้อที่เก็บข้อมูล (disk space) ของเว็บแคชมีขนาดที่จำกัดและน้อยเมื่อเทียบกับปริมาณข้อมูลที่มีให้บริการบนอินเทอร์เน็ตที่มากและมีการเปลี่ยนแปลงตลอดเวลา ดังนั้นเว็บแคชจึงต้องมีขั้นตอนวิธีการแทนที่หรือลบข้อมูลที่มีความสำคัญน้อยออกเพื่อปรับปรุงให้ข้อมูลในแคชเป็นข้อมูลที่น่าจะถูกเรียกขอในอนาคต

2.2. ขั้นตอนวิธีการแทนที่

ระบบเว็บแคชทุกระดับจะมีขั้นตอนวิธีการแทนที่เพื่อใช้ในการพิจารณาเลือกแทนที่เอกสารในแคชเมื่อมีเอกสารใหม่เข้ามาในขณะที่แคชมีที่ว่างไม่เพียงพอ เพื่อให้เอกสารภายในแคชทันสมัยและมีความน่าจะเป็นสูงที่จะถูกเรียกขอในอนาคตอันใกล้ ขั้นตอนวิธีการแทนที่ที่ได้มีการศึกษาวิจัยเพื่อนำมาใช้กับระบบเว็บแคชมีดังนี้

ชื่อขั้นตอนวิธีการแทนที่	ลำดับการแทนที่เอกสาร
LRU (Least Recently Used)	แทนที่เอกสารที่ไม่ถูกเรียกใช้มานานที่สุดก่อน
Perfect-LFU (Least Frequently Used)	แทนที่เอกสารที่ถูกเรียกใช้น้อยที่สุดก่อน (เก็บสะสมความถี่ในการเรียกขอของเอกสารไว้ถึงแม้ว่าจะถูกลบออกจากแคชแล้ว)
In-cache LFU (Least Frequently Used)	แทนที่เอกสารที่ถูกเรียกใช้น้อยที่สุดก่อน (ลบความถี่ในการเรียกขอของเอกสารเมื่อเอกสารนั้นถูกลบออกจากแคช)
SIZE [15]	แทนที่เอกสารที่ขนาดใหญ่ที่สุดก่อน
LRU-MIN [1]	แทนที่เอกสารที่ไม่ถูกเรียกใช้มานานที่สุด โดยพิจารณาจากเอกสารที่ขนาดมากกว่า S ถ้าเนื้อที่เก็บข้อมูลไม่พอให้ พิจารณาจาก S/2 และ ลดค่า S ลงทีละครึ่ง
LRU-Threshold [1]	แทนที่เอกสารที่ไม่ถูกเรียกใช้มานานที่สุด และไม่ทำสำเนาของเอกสารที่ขนาดใหญ่กว่าค่าคงที่ที่กำหนดไว้

ตารางที่ 2.1 ขั้นตอนวิธีการแทนที่และลำดับการแทนที่ของเอกสาร

2.3. สภาพการใช้เว็บ

สภาพการใช้เว็บที่ได้มีการศึกษาในงานวิจัยที่ผ่านมามีดังนี้

2.3.1. สัดส่วนของการเรียกขอที่ได้รับการตอบกลับจากเว็บเซิร์ฟเวอร์หรือพร็อกซีเซิร์ฟเวอร์พร้อมทั้งส่งข้อมูลกลับไปยังผู้เรียกขอได้บางส่วนหรือทั้งหมดกับจำนวนการเรียกขอทั้งหมด ซึ่งการเรียกขอที่มีการส่งข้อมูลกลับนี้ จะถูกนำไปใช้สำหรับการพิจารณาลักษณะส่วนที่เหลือ

2.3.2. ปริมาณของข้อมูลประเภทต่างๆ จากการพิจารณาสกุลของไฟล์จากยูอาร์แอลที่เรียกขอดังตารางที่ 2.2 เพื่อใช้ในการพิจารณาว่าควรเก็บไว้ในแคชหรือไม่ เช่น ข้อมูลประเภท ซีจีไอ (CGI) ที่ผู้เรียกขอต้องการข้อมูลจากเว็บเซิร์ฟเวอร์โดยตรงจึงไม่เก็บข้อมูลไว้ในแคช

ประเภทของเอกสาร	สกุลของไฟล์
เอชทีเอ็มแอล(HTML)	"htm", "html"
รูปภาพ	"bmp", "gif", "jpg", "jpeg"
เสียง	"au"
วีดีโอ	"mpg", "mpeg"
ซีจีไอ	"cgi" และ ยูอาร์แอลที่มี "cgi-bin"
ฟอร์แมต (Formatted)	"ps", "pdf"

ตารางที่ 2.2 การแยกประเภทของเอกสารตามสกุลของไฟล์

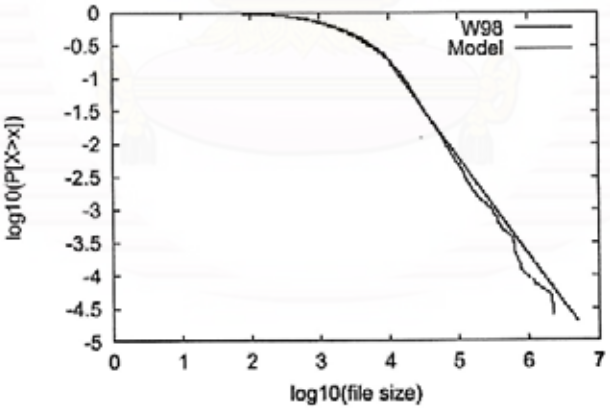
2.3.3. สัดส่วนของเอกสารที่เรียกขอที่แตกต่างกัน กับจำนวนการเรียกขอทั้งหมด

2.3.4. การกระจายของขนาดเอกสารที่ถูกเรียกขอ (file size distribution) จากการศึกษาของ Carlos Cunha [4] พบว่า ถ้ากำหนดให้

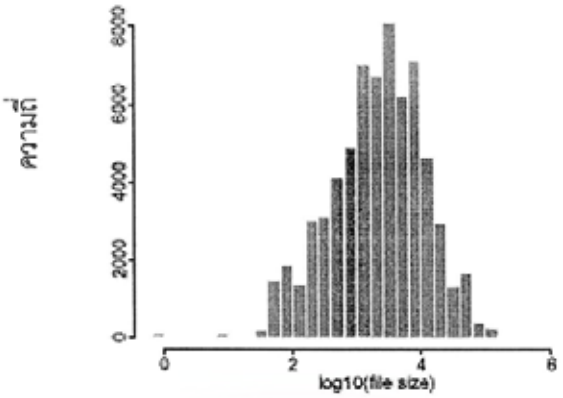
2.3.4.1. แกนตั้ง (Y) = log (1 – ความถี่สะสม / จำนวนการเรียกขอทั้งหมด)

2.3.4.2. แกนนอน (X) = log (ขนาดของไฟล์)

จะได้ กราฟมีลักษณะดังแผนภูมิในรูปที่ 2.2 และ 2.3



รูปที่ 2.2 กราฟแอลแอลซีดี (LLCD : Log-Log complementary distribution)



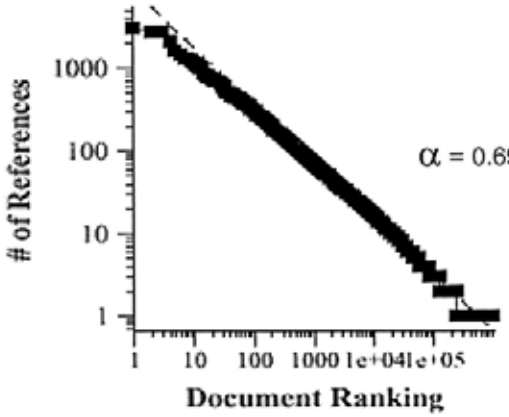
รูปที่ 2.3 การกระจายของไฟล์ขนาดเล็ก

กราฟแอลแอลซีดีของการกระจายแบบไฮบริด ล็อกนอมอล-พาเรโต ใช้สำหรับการหาลักษณะการกระจายของขนาดเอกสาร โดยมีพารามิเตอร์ที่ใช้แทนทั้งหมด 4 ตัวโดยลักษณะกระจายแบบ ไฮบริด ล็อกนอมอล-พาเรโต แบ่งออกได้เป็น 2 ส่วนคือ การกระจายของเอกสารขนาดเล็ก (น้อยกว่า 5000 ไบต์) สามารถแทนได้ด้วยการกระจายแบบล็อกนอมอลดังรูปที่ 2.3 และเอกสารขนาดใหญ่ (มากว่า 5000 ไบต์) สามารถแทนได้ด้วยการกระจายแบบพาเรโตจากการวิเคราะห์พบว่าพารามิเตอร์มีค่าดังตารางที่ 2.3

ส่วนประกอบ	โมเดล	พารามิเตอร์
Body	Lognormal	μ = 7.64 - 7.88, σ = 1.34-1.71
Tail	Pareto	k = 2,924 – 3,558, α = 1.18 – 1.38

ตารางที่ 2.3 พารามิเตอร์ของการกระจายแบบไฮบริด ล็อกนอมอล-พาเรโต

2.3.6. การแจกแจงของความถี่ในการเรียกขอเอกสารเรียงตามลำดับจากมากไปน้อย โดยใช้การแจกแจงแบบ Zipf-Like ซึ่งมีการพัฒนามาจากกฎของ Zipf ซึ่งมีลักษณะการแจกแจงดังแผนภูมิในรูปที่ 2.4 และพารามิเตอร์ที่ใช้แทนการแจกแจงแบบ Zipf มี 2 ค่า คือค่าโอเมกา และอัลฟา



รูปที่ 2.4 การแจกแจงของความถี่ในการเรียกขอเอกสาร

ค่าอัลฟาคือ พารามิเตอร์ที่ใช้ในการวิเคราะห์การกระจายของความถี่ในการเรียกขอเอกสาร โดยตามกฎของ Zipf ค่าอัลฟาจะมีค่าประมาณ 1 แต่การแจกแจงแบบ Zipf-Like ค่าอัลฟาจะอยู่ในช่วง 0 ถึง 1 ถ้า ค่าอัลฟาใกล้ 1 แสดงว่าความถี่ในการเรียกขอแต่ละเอกสารมีการกระจายมาก แต่ถ้าค่าอัลฟาน้อยหรือใกล้ 0 แสดงว่าความถี่ในการเรียกขอเอกสารมีการกระจายน้อย ดังนั้นถ้าข้อมูลการใช้เว็บที่สังเกตมีค่าใกล้ 1 แสดงว่าข้อมูลการใช้เว็บนี้น่าให้ประสิทธิภาพที่ดีกว่าในการทำระบบเว็บแคชเพื่อรองรับการเรียกขอ

ค่าโอเมกาเป็นพารามิเตอร์อีกตัวของการแจกแจงแบบ Zipf-Like ซึ่งเป็นจุดตัดแกน y ของกราฟที่แสดงการแจกแจงแบบ Zipf-Like โดยค่านี้จะขึ้นกับปริมาณการเรียกขอ ถ้าปริมาณการเรียกขอมากค่าโอเมกาก็จะมากขึ้นด้วย

2.4 การทดสอบแบบแฟคทอเรียล (the factorial experiment)

ในการออกแบบการทดลองเพื่อศึกษาผลกระทบของแฟคเตอร์ที่มีต่อประสิทธิภาพของระบบที่ศึกษาจำเป็นที่จะต้องเลือกทดสอบแฟคเตอร์ที่สำคัญและใช้เครื่องวัดที่เหมาะสมในการทดลองเพื่อให้จำนวนของการทดลองน้อยที่สุดและผลการทดลองสามารถให้ข้อมูลได้มากที่สุด โดยการสรุปผลของการทดลองจะใช้การทดสอบแบบแฟคทอเรียลมีประโยชน์ในการศึกษาความสัมพันธ์ของแฟคเตอร์ที่มีต่อประสิทธิภาพของระบบที่ได้ ซึ่งอาจจะมีเพียง 1 แฟคเตอร์หรือมากกว่า โดยการทดสอบจะใช้การวิเคราะห์ค่าความแปรปรวนของประสิทธิภาพ

การทดสอบแบบแฟคทอเรียลมีรูปแบบดังนี้

2.4.1 ข้อมูลการทดสอบมีรูปแบบดังตารางที่ 2.4

แฟคเตอร์ A	แฟคเตอร์ B				รวม	ค่าเฉลี่ย
	1	2	…	b		
1	X_{111}	X_{121}	…	X_{1b1}	$T_{1..}$	$\overline{X}_{1..}$
	X_{11n}	X_{12n}	…	X_{1bn}		
2	X_{211}	X_{221}	…	X_{2b1}	$T_{2..}$	$\overline{X}_{2..}$
	X_{21n}	X_{22n}	…	X_{2bn}		
:	:	:		:	:	:
a	X_{a11}	X_{a21}	…	X_{ab1}	$T_{a..}$	$\overline{X}_{a..}$
	X_{a1n}	X_{a2n}	…	X_{abn}		
รวม	$T_{.1.}$	$T_{.2.}$	…	$T_{.b.}$	$T_{...}$	
ค่าเฉลี่ย	$\overline{X}_{.1.}$	$\overline{X}_{.a.}$	…	$\overline{X}_{.b.}$		$\overline{X}_{...}$

ตารางที่ 2.4 แสดงข้อมูลของการทดสอบแบบแฟคทอเรียลที่มีแฟคเตอร์ 2 ตัว

2.4.2 โมเดลที่ใช้ในการทดสอบ คือ

$$x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

$i = 1, 2, \ldots, a$

$j = 1, 2, \ldots, b$

$k = 1, 2, \ldots, n$

โดย x_{ijk} คือ ข้อมูลที่ได้

μ คือ ค่าคงที่

α คือ ผลของแฟคเตอร์ A

β คือ ผลของแฟคเตอร์ B

$(\alpha\beta)$ คือ ผลของปฏิกิริยาระหว่างแฟคเตอร์ A และ B

e_{ijk} คือ ค่าผิดพลาดจากการทดลอง

2.4.3 การวิเคราะห์ความแปรปรวนสำหรับการทดสอบแบบแฟคทอเรียล

สมมติฐานที่ใช้ในการทดสอบ

การทดสอบว่าแฟคเตอร์ A มีผลต่อข้อมูลที่ได้ คือ

H_0 : ไม่ทุก α_i ที่ $= 0$

การทดสอบว่าแฟคเตอร์ B มีผลต่อข้อมูลที่ได้ คือ

H_0 : ไม่ทุก β_j ที่ $= 0$

สมการที่ใช้ในการคำนวณค่าผลบวกกำลังสอง คือ

$$SS_{total} = SS_A + SS_B + SS_{AB} + SS_{residual}$$

ค่าผลบวกกำลังสองคำนวณได้จาก

$$SS_{total} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn}$$

$$SS_A = \sum_{i=1}^{a}\frac{y_{i..}^2}{bn} - \frac{y_{...}^2}{abn}$$

$$SS_B = \sum_{j=1}^{b}\frac{y_{.j.}^2}{an} - \frac{y_{...}^2}{abn}$$

$$SS_{subtotals} = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{y_{ij.}^2}{n} - \frac{y_{...}^2}{abn}$$

$$SS_{AB} = SS_{subtotals} - SS_A - SS_B$$

$$SS_E = SS_T - SS_{AB} - SS_A - SS_B$$

การทดสอบสมมติฐานใช้ค่าอัตราส่วนของความแปรปรวน (variance ratio) ซึ่งสามารถคำนวณได้จากตารางที่ 2.5

แหล่งความแปรปรวน (source of variation)	องศาอิสระ (degree of freedom)	ผลบวกกำลังสอง (sum of squares)	ค่าเฉลี่ยกำลังสอง (mean square)	F_0 (variance ratio)
แฟคเตอร์ A	a-1	SS_A	$MS_A=SS_A/(a-1)$	MS_A/MS_E
แฟคเตอร์ B	b-1	SS_B	$MS_B=SS_B/(b-1)$	MS_B/MS_E
ระหว่างแฟคเตอร์	(a-1)(b-1)	SS_{AB}	$MS_{AB}=SS_{AB}/(a-1)(b-1)$	MS_{AB}/MS_E
ความคลาดเคลื่อน	ab(n-1)	SS_E	$MS_E=SS_E/ab(n-1)$	
ผลรวม	abn-1	SS_{total}		

ตารางที่ 2.5 การวิเคราะห์ความแปรปรวนสำหรับการทดลองแบบแฟคทอเรียลที่มี 2 แฟคเตอร์

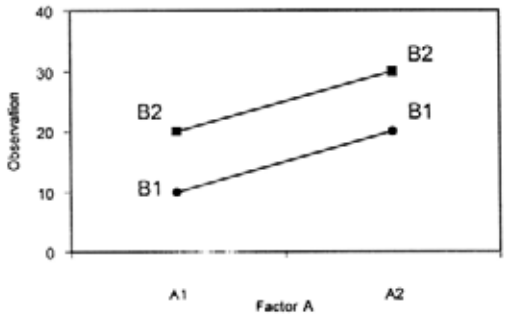
การสรุปผลจากการทดสอบว่าสมมติฐานเป็นจริงได้โดยค่าอัตราส่วนความแปรปรวนจะมีการแจกแจงแบบ F โดยมีองศาอิสระตามที่ระบุไว้ในตารางที่ 2.5 จะต้องมีค่ามากกว่าค่าอัตราส่วนความแปรปรวนที่ได้จากการเปิดตารางการแจกแจงแบบ F ผลการทดสอบแบ่งออกเป็น 2 ส่วน ได้แก่

ก. ผลกระทบของแฟคเตอร์ที่มีต่อข้อมูล

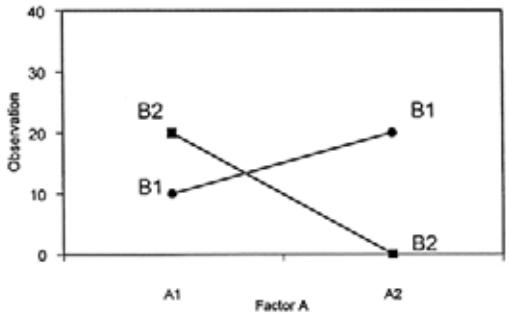
ถ้าความแปรปรวนของข้อมูลระหว่างระดับของแฟคเตอร์ใดมากกว่าความแปรปรวนของชุดภายในข้อมูลที่ใช้ในการทดสอบ จะสรุปว่าแฟคเตอร์นั้นๆ มีผลต่อข้อมูลที่ได้

ข. ผลของปฏิกิริยาระหว่างแฟคเตอร์

ถ้าความแปรปรวนระหว่างแฟคเตอร์มากกว่าความแปรปรวนของชุดภายในข้อมูลที่ใช้ในการทดสอบ จะสรุปว่าปฏิกิริยาระหว่างแฟคเตอร์มีผลต่อข้อมูลที่ได้ โดยปฏิกิริยาระหว่างแฟคเตอร์ หมายถึง ผลกระทบของการเปลี่ยนแปลงระดับของแฟคเตอร์ที่มีต่อค่าของข้อมูลที่ได้เนื่องจากการเปลี่ยนแปลงระดับของอีกแฟคเตอร์หนึ่ง ผลของปฏิกิริยาระหว่างแฟคเตอร์สามารถแสดงได้ดังแผนภูมิในรูปที่ 2.5 และรูปที่ 2.6



รูปที่ 2.5 การทดสอบแบบแฟคทอเรียลที่ไม่มีปฏิกิริยาระหว่างแฟคเตอร์



รูปที่ 2.6 การทดสอบแบบแฟคทอเรียลที่มีปฏิกิริยาระหว่างแฟคเตอร์

รูปที่ 2.5 แสดงการทดสอบแบบแฟคทอเรียลที่ไม่มีปฏิกิริยาระหว่างแฟคเตอร์ A และ แฟคเตอร์ B ลักษณะของแผนภูมิแสดงข้อมูลที่ได้จะเป็นเส้นขนานกันเมื่อเปลี่ยนระดับของแฟคเตอร์ A จาก A1 เป็น A2 และรูปที่ 2.6 แสดงการทดสอบแบบแฟคทอเรียลที่มีปฏิกิริยาระหว่างแฟคเตอร์ A และ แฟคเตอร์ B ลักษณะของแผนภูมิแสดงข้อมูลที่ได้จะเป็นเส้นตรงที่ไม่ขนานกันเมื่อเปลี่ยนระดับของแฟคเตอร์ A จาก A1 เป็น A2

ตัวอย่างการทดสอบแบบแฟคทอเรียล

การทดสอบประสิทธิภาพของซีพียู 3 รุ่น คือ A1 A2 และ A3 โดยการใช้ภาระในการทดสอบทั้งหมด 5 ชุด คือ B1 ถึง B5 และวัดเวลาที่ใช้ในการประมวลผลได้ผลการทดสอบดังตารางที่ 2.6 สำหรับการทดสอบแต่ละเงื่อนไขจะทำซ้ำ 3 ครั้ง

	A1	A2	A3
B1	3200	5120	8960
	3150	5100	8900
	3250	5140	8840
B2	4700	9400	19740
	4740	9300	19790
	4660	9500	19690
B3	3200	4160	7360
	3220	4100	7300
	3180	4220	7420
B4	5100	5610	22340
	5200	5575	22440
	5000	5645	22540
B5	6800	12240	28560
	6765	12290	28360
	6835	12190	28760

ตารางที่ 2.6 เวลาที่ซีพียูใช้ในการประมวลผล (มิลลิวินาที)

การวิเคราะห์ความแปรปรวนเพื่อใช้ในการทดสอบแฟคทอเรียลโดยใช้ระดับนัยสำคัญ 0.1 ได้ผลดังตารางที่ 2.7

แหล่งความแปรปรวน	องศาอิสระ	ผลบวกกำลังสอง	ค่าเฉลี่ยกำลังสอง	F_0	F_{crit}
แฟคเตอร์ A	2	1365256360	682628180	109924	2.488719
แฟคเตอร์ B	4	728826720	182206680	29340.85	2.142237
ระหว่างแฟคเตอร์	8	422501040	52812630	8504.449	1.884121
ความคลาดเคลื่อน	30	186300	6210		
ผลรวม	44	2516770420			

ตารางที่ 2.7 ตารางการวิเคราะห์ความแปรปรวนเพื่อใช้ในทดสอบแบบแฟคทอเรียล

การทดสอบแบบแฟคทอเรียลใช้การเปรียบเทียบค่า F_0 กับ F_{crit} ซึ่งได้จากการเปิดตารางการแจกแจงแบบ F พบว่าค่า F_0 ของแฟคเตอร์ A แฟคเตอร์ B และระหว่างแฟคเตอร์มากกว่า F_{crit} ดังนั้นจึงสรุปว่าซีพียูและภาระที่ใช้ในการทดสอบมีผลต่อเวลาในการประมวลผล และมีปฏิกิริยาระหว่างแฟคเตอร์ด้วย

บทที่ 3

การวิเคราะห์สภาพการใช้เว็บ

การวิเคราะห์สถิติการใช้งานอินเทอร์เน็ตในจุฬาลงกรณ์มหาวิทยาลัยพบว่าบริการเวิลด์ไวด์เว็บเป็นบริการที่มีอัตราการใช้บริการสูงที่สุด คือ 79.5 % ของการใช้งานทั้งหมดของจุฬาลงกรณ์มหาวิทยาลัย [16] ดังนั้นการศึกษาสภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัยจึงเป็นสิ่งสำคัญในการจำลองการทำงานของพร็อกซีแคชเพื่อใช้ปรับปรุงประสิทธิภาพ สภาพการใช้เว็บที่ใช้ในการพิจารณา ได้แก่

ก. ปริมาณการเรียกขอทั้งหมด คือ ปริมาณของการเรียกขอผ่านโพรโตคอลเอชทีทีพีทั้งหมดที่มีการบันทึกไว้ที่พร็อกซีเซิร์ฟเวอร์

ข. ปริมาณการเรียกขอที่ได้รับการตอบกลับแบบ "Successful" คือ การเรียกขอที่มีการบันทึกว่าได้รับการตอบกลับจากเว็บเซิร์ฟเวอร์ด้วยรหัสการตอบกลับ 200 เป็นการตอบกลับการเรียกขอจากเว็บเซิร์ฟเวอร์พร้อมทั้งส่งข้อมูลกลับไปให้กับไคลเอ็นต์ตามการเรียกขอนั้นๆ ซึ่งพร็อกซีแคชจะพิจารณาแทนที่เอกสารในแคชเมื่อการเรียกขอได้รับการตอบกลับจากเว็บเซิร์ฟเวอร์ด้วยรหัสการตอบกลับประเภทนี้

ค. ปริมาณการเรียกขอที่แตกต่างกัน คือ ปริมาณการเรียกขอที่มียูอาร์แอลไม่ซ้ำกันทั้งหมดจากการเรียกขอที่ได้รับการตอบกลับแบบ "Successful"

ง. พารามิเตอร์จากกฎของ Zipf คือ ค่าอัลฟาที่ได้จากการแจกแจงแบบ Zipf-like โดยวิธีการหาอยู่ในข้อที่ 2.3.6

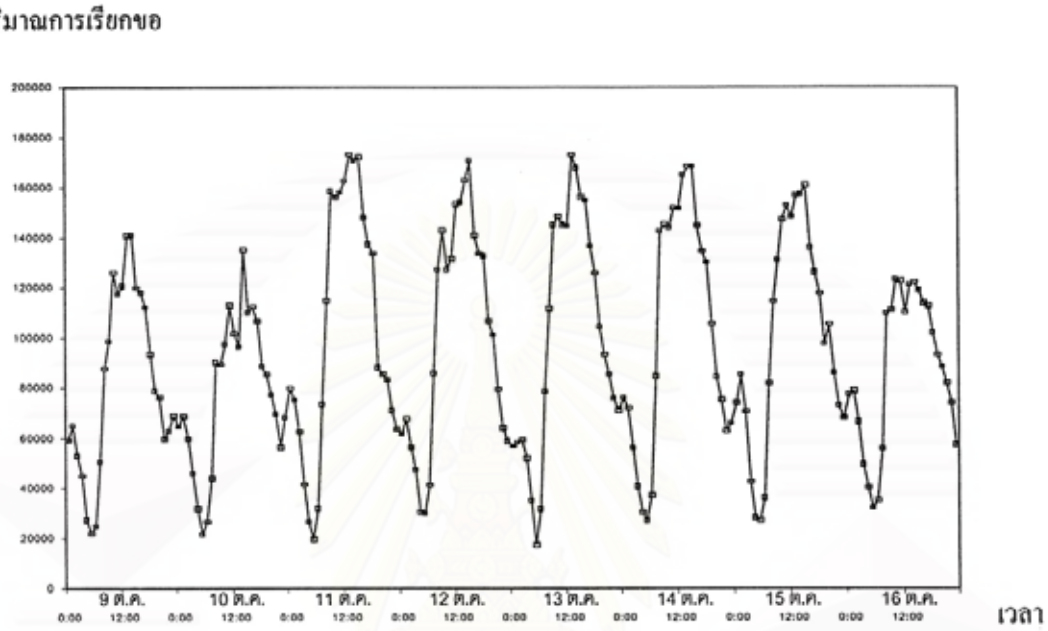
สภาพการใช้เว็บที่ได้พิจารณาเพิ่มเติมในงานวิจัยนี้ คือ ช่วงเวลาในการใช้เว็บ แบ่งออกเป็น 2 ช่วง คือ ช่วงที่มีการใช้เว็บมากและช่วงที่มีการใช้เว็บน้อย โดยการแบ่งออกเป็น 2 ช่วงนี้ใช้การพิจารณาปริมาณการเรียกขอเฉลี่ยแต่ละชั่วโมงของวันจากการเรียกขอทั้งหมดของพร็อกซีเซิร์ฟเวอร์แต่ละที่และแบ่งครึ่งไฟล์ที่บันทึกการเรียกขอแต่ละวันออกเป็นชั่วโมงที่มีการใช้เว็บมาก 12 ชั่วโมง และช่วงที่มีการใช้เว็บน้อยอีก 12 ชั่วโมง

การศึกษาสภาพการใช้เว็บแบ่งตามข้อมูลการใช้เว็บที่ศึกษาเป็น 3 ส่วน ได้แก่ ข้อมูลการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย ข้อมูลการใช้เว็บจากพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย และข้อมูลการใช้เว็บจากเอ็นแอลเอเอ็นอาร์

3.1. ข้อมูลการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย

ช่วงเวลาของการใช้เว็บที่ใช้ในการทดสอบได้แก่ ช่วงวันที่ 9 ตุลาคม พ.ศ. 2542 ถึงวันที่ 16 ตุลาคม พ.ศ. 2542 เนื่องจากเป็นช่วงเวลาที่ใกล้เคียงกับช่วงเวลาที่ทำการเก็บรวบรวมข้อมูลการใช้เว็บมาจากเอ็นแอลเอเอ็นอาร์และข้อมูลการใช้เว็บช่วงนี้เป็นข้อมูลที่สามารถใช้แทนการใช้เว็บในช่วงปิดภาคการศึกษาและการใช้เว็บในช่วงเปิดภาคการศึกษาส่วนใหญ่ได้โดยการทดสอบนี้อยู่ในภาคผนวก จ.

ข้อมูลการใช้เว็บมีปริมาณการเรียกขอแต่ละชั่วโมงดังแผนภูมิในรูปที่ 3.1



รูปที่ 3.1 ปริมาณการเรียกขอแต่ละชั่วโมงของข้อมูลการใช้เว็บ
จากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย วันที่ 9 – 16 ต.ค. 2542

ผลการวิเคราะห์พบว่าปริมาณการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัยในแต่ละวันจะมีรูปแบบของปริมาณการใช้เว็บที่ใกล้เคียงกัน คือมีแนวโน้มเพิ่มขึ้นในช่วงเช้าตั้งแต่ 6:00 น. ถึงเวลาประมาณ 15:00 น. ซึ่งเป็นเวลาที่มีปริมาณการใช้เว็บมากที่สุดโดยในช่วงเวลาพักเที่ยงจะมีปริมาณการใช้เว็บลดลงเล็กน้อย หลังจากเวลาประมาณ 15:00 น. จนถึงเวลา 6:00 น. ปริมาณการใช้เว็บมีแนวโน้มลดลงด้วยอัตราที่ใกล้เคียงกันกับที่เพิ่มขึ้นในช่วงเช้า โดยช่วงเวลาประมาณ 22:00 ถึงเวลาประมาณ 2:00 น. เป็นช่วงที่มีการใช้เว็บเพิ่มขึ้น ซึ่งคาดว่าเป็นช่วงเวลาหลังจากรายการโทรทัศน์หลังการรายงานข่าวซึ่งส่วนใหญ่เป็นรายการที่มีผู้ชมจำนวนมากจบ หลังจากช่วงเวลานี้ปริมาณการเรียกขอจะลดลงจนต่ำสุดที่เวลาประมาณ 6:00 น.

รูปที่ 3.2 ปริมาณการเรียกขอเฉลี่ยแต่ละชั่วโมงของข้อมูลการใช้เว็บ
จากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย วันที่ 9 – 16 ต.ค. 2542

การทดสอบเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่จะใช้ข้อมูลการใช้เว็บที่แบ่งออกเป็นช่วงเวลาที่มีการใช้เว็บมากและช่วงเวลาที่มีการใช้เว็บน้อย ช่วงเวลาละ 12 ชั่วโมง พิจารณาจากปริมาณการเรียกขอเฉลี่ยต่อชั่วโมงสำหรับแต่ละชั่วโมง ข้อมูลการใช้เว็บที่ใช้ในการทดสอบซึ่งมีปริมาณการเรียกขอเฉลี่ยต่อชั่วโมงดังแผนภูมิในรูปที่ 3.2 สามารถแบ่งช่วงเวลาการใช้เว็บโดยช่วงเวลาที่มีการใช้เว็บมากและช่วงเวลาที่มีการใช้เว็บน้อยดังตารางที่ 3.1

	ตั้งแต่	ถึง
ช่วงเวลาที่มีการใช้เว็บมาก	8:00 น.	20:00 น.
ช่วงเวลาที่มีการใช้เว็บน้อย	20:00 น.	8:00 น.

ตารางที่ 3.1 ช่วงเวลาที่มีการใช้เว็บมากและน้อย ของการใช้เว็บ
จากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย วันที่ 9 – 16 ต.ค. 2542

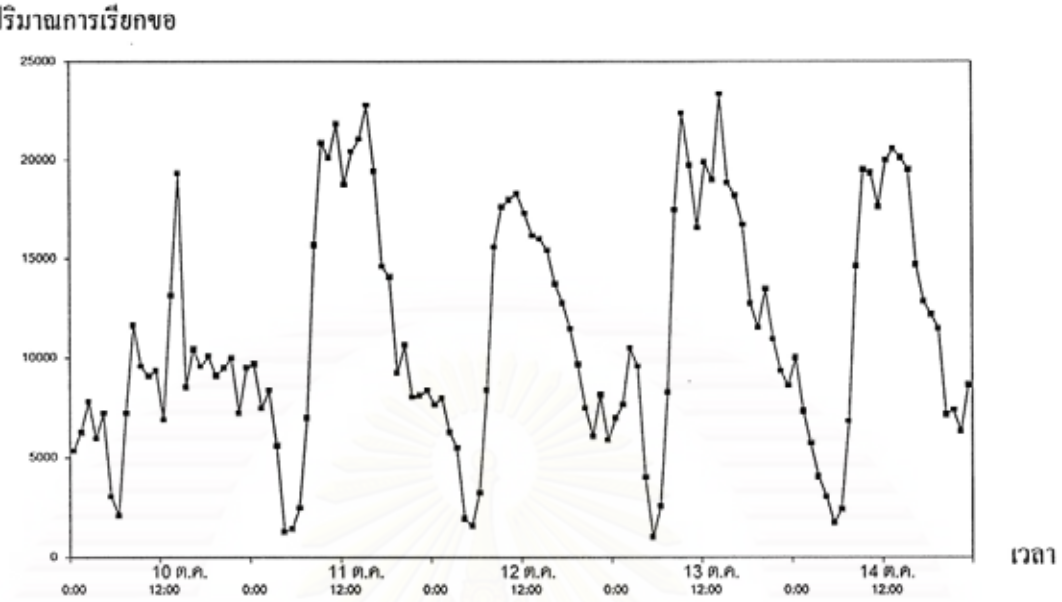
การใช้ช่วงเวลาการใช้เว็บตามตารางที่ 3.1 สามารถแบ่งข้อมูลการใช้เว็บที่ใช้ในการทดสอบเปรียบเทียบออกเป็น 15 ชุด ดังตารางที่ 3.2

	วันที่	ปริมาณการเรียกขอ	ปริมาณการเรียกขอที่ได้รับการตอบกลับแบบ "Successful"	ปริมาณการเรียกขอที่แตกต่างกัน	ค่าอัลฟา	ค่าโอเมกา
ช่วงที่มีการใช้เว็บมาก (8:00 น. - 20:00 น.)	9 ต.ค. 2542	1354159	1001401	396657	0.674	5304.358
	10 ต.ค. 2542	1225668	896668	357148	0.703	6191.965
	11 ต.ค. 2542	1772308	1275181	451274	0.703	8490.263
	12 ต.ค. 2542	1683198	1188228	450833	0.690	7064.015
	13 ต.ค. 2542	1713746	1240852	452043	0.691	7461.092
	14 ต.ค. 2542	1751179	1259421	460693	0.708	8595.517
	15 ต.ค. 2542	1646946	1034928	384688	0.692	7287.445
	16 ต.ค. 2542	1362212	1002265	376535	0.681	5703.448
ช่วงที่มีการใช้เว็บน้อย(20:00 น.- 8:00 น.)	9-10 ต.ค. 2542	629353	450735	221900	0.643	2095.900
	10-11 ต.ค. 2542	681554	478342	233673	0.663	2567.759
	11-12 ต.ค. 2542	723530	489099	246081	0.656	2419.417
	12-13 ต.ค. 2542	693486	499272	240552	0.669	2804.558
	13-14 ต.ค. 2542	750351	549007	255478	0.661	2937.457
	14-15 ต.ค. 2542	735886	521064	245336	0.670	2933.390
	15-16 ต.ค. 2542	769195	558185	266220	0.655	2773.076

ตารางที่ 3.2 ข้อมูลการใช้เว็บที่ใช้ในการทดสอบ และสภาพการใช้เว็บแต่ละช่วงเวลาของการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย วันที่ 9 – 16 ต.ค. 2542

3.2. ข้อมูลการใช้เว็บจากพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย

ช่วงเวลาของข้อมูลการใช้เว็บที่นำมาทดสอบคือช่วงวันที่ 10 ตุลาคม พ.ศ. 2542 ถึงวันที่ 14 ตุลาคม พ.ศ. 2542 มีปริมาณการเรียกขอแต่ละชั่วโมงดังแผนภูมิในรูปที่ 3.3



รูปที่ 3.3 ปริมาณการเรียกขอแต่ละชั่วโมงของข้อมูลการใช้เว็บจาก
พร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย วันที่ 10 - 14 ต.ค. 2542

ปริมาณการใช้เว็บผ่านพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัยมีแนวโน้มใกล้เคียงกับปริมาณการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัยเนื่องจากการใช้เว็บจากพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัยเป็นพร็อกซีแคชที่อยู่ระดับล่าง ข้อมูลการใช้เว็บจึงเป็นส่วนหนึ่งของข้อมูลการใช้เว็บที่รวบรวมมาจากสำนักเทคโนโลยีสารสนเทศซึ่งเป็นพร็อกซีแคชระดับที่สูงกว่า โดยปริมาณการใช้เว็บแต่ละชั่วโมงเพิ่มขึ้นตั้งแต่ช่วงเวลาประมาณ 6:00 น. จนถึงช่วงเวลาประมาณ 15:00 น. และหลังจาก 15:00 น. ปริมาณการใช้เว็บมีแนวโน้มลดลงจนถึงช่วงเวลาประมาณ 6:00 น. ซึ่งเป็นเวลาที่มีปริมาณการใช้เว็บน้อยที่สุด

รูปที่ 3.4 ปริมาณการเรียกขอเฉลี่ยแต่ละชั่วโมงของข้อมูลการใช้เว็บจาก
พร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย วันที่ 10 – 14 ต.ค. 2542

โดยการพิจารณาปริมาณการใช้เว็บจากแผนภูมิในรูปที่ 3.4 สามารถแบ่งช่วงเวลาที่มีการใช้เว็บมากและน้อยได้ดังตารางที่ 3.3 โดยผลการแบ่งช่วงเวลาที่มีการใช้เว็บมากและช่วงเวลาที่มีการใช้เว็บน้อยได้ช่วงเวลาเดียวกันกับช่วงเวลาที่มีการใช้เว็บมากและช่วงเวลาที่มีการใช้เว็บน้อยเหมือนกับการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัย เนื่องจากข้อมูลการเรียกขอผ่านพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย เป็นส่วนหนึ่งของการเรียกขอของการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัย

	ตั้งแต่	ถึง
ช่วงเวลาที่มีการใช้เว็บมาก	8:00 น.	20:00 น.
ช่วงเวลาที่มีการใช้เว็บน้อย	20:00 น.	8:00 น.

ตารางที่ 3.3 ช่วงเวลาที่มีการใช้เว็บมากและน้อยของข้อมูลการใช้เว็บจาก
พร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย วันที่ 10 – 14 ต.ค. 2542

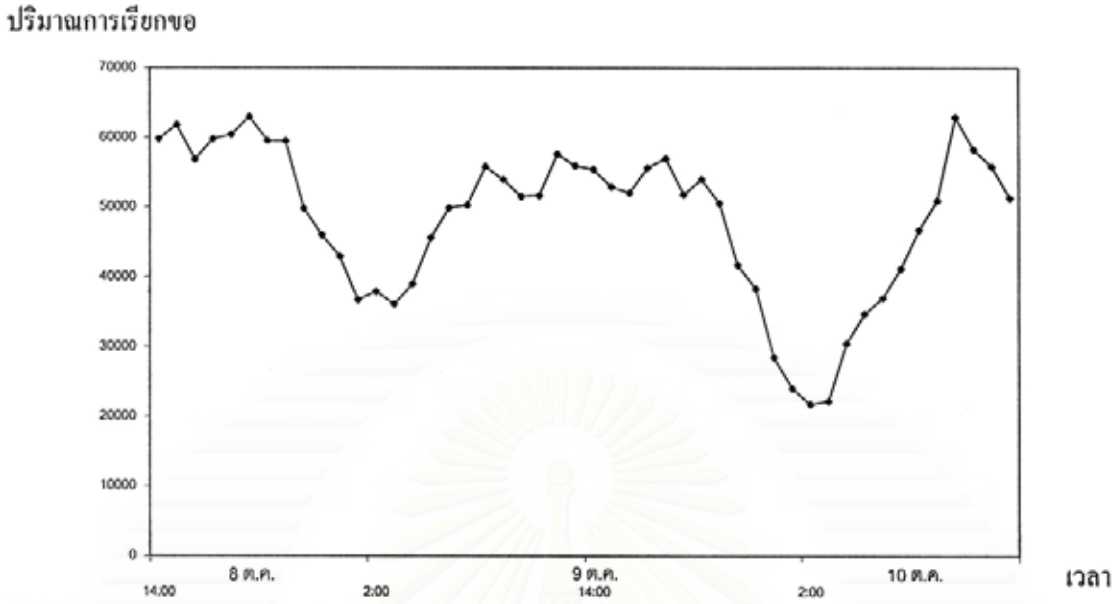
การใช้ช่วงเวลาการใช้เว็บตามตารางที่ 3.3 สามารถแบ่งข้อมูลการใช้เว็บที่ใช้ในการทดสอบเปรียบเทียบออกเป็น 9 ชุด ดังตารางที่ 3.4

	วันที่	ปริมาณการเรียกขอ	ปริมาณการเรียกขอที่ได้รับการตอบกลับแบบ "Successful"	ปริมาณการเรียกขอที่แตกต่างกัน	ค่าอัลฟา	ค่าโอเมกา
ช่วงที่มีการใช้เว็บมาก (8:00 น. - 20:00 น.)	10 ต.ค. 2542	126785	59412	36718	0.558	247.397
	11 ต.ค. 2542	218836	94318	50042	0.580	437.166
	12 ต.ค. 2542	181899	80719	45355	0.532	273.583
	13 ต.ค. 2542	216368	93561	50087	0.567	388.128
	14 ต.ค. 2542	202662	86499	47607	0.559	341.105
ช่วงที่มีการใช้เว็บน้อย (20:00 น. - 8:00 น.)	10-11 ต.ค. 2542	79469	36153	25880	0.555	158.688
	11-12 ต.ค. 2542	77720	38500	27558	0.538	142.117
	12-13 ต.ค. 2542	78169	36647	26593	0.592	203.375
	13-14 ต.ค. 2542	83554	42091	31003	0.522	132.900

ตารางที่ 3.4 ข้อมูลการใช้เว็บที่ใช้ในการทดสอบ และสภาพการใช้เว็บแต่ละช่วงเวลาของการใช้เว็บจากพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย วันที่ 10 – 14 ต.ค. 2542

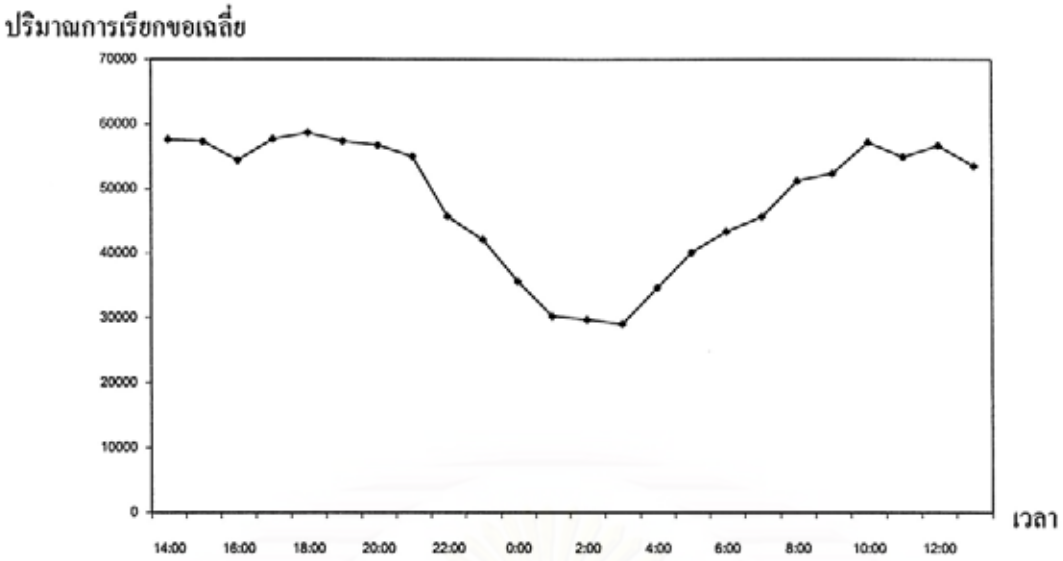
3.3. ข้อมูลการใช้เว็บจากเอ็นแอลเอเอ็นอาร์

ช่วงเวลาของข้อมูลการใช้เว็บที่นำมาทดสอบคือช่วงวันที่ 8 ตุลาคม พ.ศ. 2542 ถึงวันที่ 10 ตุลาคม พ.ศ. 2542 มีปริมาณการใช้เว็บเฉลี่ยแต่ละชั่วโมงดังแผนภูมิในรูปที่ 3.5



รูปที่ 3.5 ปริมาณการเรียกขอแต่ละชั่วโมงของข้อมูลการใช้เว็บจากเอ็นแอลเอเอ็นอาร์วันที่ 8 - 10 ต.ค. 2542

ข้อมูลการเรียกขอของเอ็นแอลเอเอ็นอาร์ที่นำมาใช้ในการทดสอบคือข้อมูลจากเซิร์ฟเวอร์ชื่อ "sv" เป็นเซิร์ฟเวอร์ที่รองรับการเรียกขอจากพร็อกซีแคชภายในประเทศอเมริกาเป็นส่วนใหญ่และรองรับการเรียกขอจากพร็อกซีเซิร์ฟเวอร์ภายนอก ได้แก่ ประเทศเกาหลี นิวซีแลนด์ ออสเตรเลีย และ ไต้หวัน เนื่องจากเวลาของแต่ละประเทศต่างกันออกไป ดังนั้นปริมาณการเรียกขอแต่ละชั่วโมงของวันดังแผนภูมิในรูปที่ 3.5 ซึ่งไม่มีรูปแบบที่ชัดเจนเหมือนสภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 3.6 ปริมาณการเรียกขอเฉลี่ยแต่ละชั่วโมงของข้อมูลการใช้เว็บจาก
เอ็นแอลเอเอ็นอาร์ วันที่ 8 – 10 ต.ค. 2542

โดยการพิจารณาปริมาณการใช้เว็บจากแผนภูมิในรูปที่ 3.6 สามารถแบ่งช่วงเวลาที่มีการใช้เว็บมากและน้อยได้ดังตารางที่ 3.5

	ตั้งแต่	ถึง
ช่วงเวลาที่มีการใช้เว็บมาก	10.00 น.	22.00 น.
ช่วงเวลาที่มีการใช้เว็บน้อย	22.00 น.	10.00 น.

ตารางที่ 3.5 ช่วงเวลาที่มีการใช้เว็บมากและน้อยของข้อมูลการใช้เว็บจาก
เอ็นแอลเอเอ็นอาร์ วันที่ 8 – 10 ต.ค. 2542

การใช้ช่วงเวลาการใช้เว็บตามตารางที่ 3.5 สามารถแบ่งข้อมูลการใช้เว็บที่ใช้ในการทดสอบเปรียบเทียบออกเป็น 3 ชุด ดังตารางที่ 3.6

	วันที่	ปริมาณการเรียกขอ	ปริมาณการเรียกขอที่ได้รับการตอบกลับแบบ "Successful"	ปริมาณการเรียกขอที่แตกต่างกัน	ค่าอัลฟา	ค่าโอเมกา
ช่วงที่มีการใช้เว็บมาก (10:00 น. - 22:00 น.)	9 ต.ค. 2542	645939	434420	353175	0.672	1562.145
ช่วงที่มีการใช้เว็บน้อย (22:00 น. - 10:00 น.)	8-9 ต.ค. 2542	543781	361196	281888	0.624	971.581
	9-10 ต.ค. 2542	417501	277273	232092	0.640	845.113

ตารางที่ 3.6 ข้อมูลการใช้เว็บที่ใช้ในการทดสอบ และสภาพการใช้เว็บแต่ละช่วงเวลาของการใช้เว็บจากเอ็นแอลเอเอ็นอาร์ วันที่ 8 – 10 ต.ค. 2542

3.4 สรุปผลการวิเคราะห์สภาพการใช้เว็บ

ปริมาณการใช้เว็บสำหรับแต่ละช่วงเวลาขึ้นกับสภาพแวดล้อมในการใช้เว็บและระดับของพร็อกซีแคชซึ่งจากผลการวิเคราะห์ช่วงเวลาการใช้เว็บของข้อมูลการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย และช่วงเวลาการใช้เว็บของข้อมูลการใช้เว็บจากพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย พบว่าช่วงเวลาที่มีการใช้เว็บมากและช่วงที่มีการใช้เว็บน้อยคือช่วงเวลาเดียวกันแต่มีปริมาณการเรียกขอที่ต่างกันเนื่องจากพร็อกซีเซิร์ฟเวอร์ที่สำนักเทคโนโลยีสารสนเทศรองรับปริมาณการใช้งานที่มากกว่า เมื่อพิจารณาพร็อกซีแคชที่เอ็นแอลเอเอ็นอาร์ พบว่าช่วงเวลาการใช้เว็บแตกต่างออกไปจากของจุฬาลงกรณ์มหาวิทยาลัย เนื่องจากอยู่ภายใต้สภาพการใช้เว็บที่ต่างกัน

การวิเคราะห์ปริมาณการเรียกขอที่ต่างกันพบว่าการใช้เว็บจากเอ็นแอลเอเอ็นอาร์มีปริมาณการเรียกขอที่แตกต่างกันมากถึง 81.02 เปอร์เซ็นต์เมื่อเทียบกับปริมาณการเรียกขอที่ได้รับการตอบกลับแบบ "Successful" โดยปริมาณการเรียกขอที่แตกต่างกันหรือปริมาณของเอกสารที่แตกต่างกันจากข้อมูลการใช้เว็บที่ศึกษา แสดงถึงขอบเขตของประสิทธิภาพในการแทนที่ของขั้นตอนวิธีการแทนที่ ส่วนการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศ และพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย มีปริมาณการเรียกขอที่แตกต่างกันเพียง 42.56 เปอร์เซ็นต์และ 63.22 เปอร์เซ็นต์ ตามลำดับ เนื่องจากพร็อกซีแคชที่เอ็นแอลเอเอ็นอาร์เป็นพร็อกซีแคชในระดับที่สูงกว่าซึ่งรองรับการเรียกขอจากพร็อกซีแคชในระดับที่ต่ำลงมา

การพิจารณาค่าเฉลี่ยของค่าอัลฟาจากการแจกแจงแบบ Zipf ของข้อมูลการใช้เว็บที่ใช้ในการทดสอบ ซึ่งค่าอัลฟานี้แสดงถึงการกระจายของความถี่ในการเรียกขอของเอกสาร โดยค่าอัลฟาเป็นคุณสมบัติของข้อมูลการเรียกขอแต่ละแห่ง พบว่าสำหรับการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศมีค่าใกล้เคียงกันกับการใช้เว็บจากเอ็นแอลเอเอ็นอาร์คือ 0.677 และ 0.645 ตามลำดับ เนื่องจากเป็นพร็อกซีแคชที่รองรับปริมาณการเรียกขอใกล้เคียงกันคือในช่วง 1 – 2 ล้านการเรียกขอต่อวัน ส่วนการใช้เว็บจากพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย มีค่าอัลฟาเพียง 0.556 เนื่องจากพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัยเป็นพร็อกซีแคชระดับล่างซึ่งรองรับปริมาณการเรียกขอที่น้อยกว่าคือประมาณ 0.27 ล้านการเรียกขอต่อวัน

บทที่ 4

การทดสอบเปรียบเทียบประสิทธิภาพขั้นตอนวิธีการแทนที่

4.1. วิธีการทดสอบเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่

การทดสอบเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่โดยการจำลองการทำงานของพร็อกซีแคช แบ่งออกได้ เป็น 2 ประเภท ได้แก่

4.1.1. การทดสอบเปรียบเทียบประสิทธิภาพโดยการใช้ข้อมูลจริงในการจำลองการทำงานของพร็อกซีแคชเพื่อทดสอบ (Trace-driven simulation)

วิธีการทดสอบทำโดยใช้โปรแกรมจำลองการทำงานของพร็อกซีแคชอ่านข้อมูลการใช้เว็บจากไฟล์ที่บันทึกการเรียกขอ และพิจารณาว่าเอกสารที่เรียกขอนั้นมีในพร็อกซีแคชหรือไม่ ถ้าพบว่ามีเอกสารอยู่ภายในพร็อกซีแคช ให้เพิ่มค่าการพบเอกสารเท่ากับ 1 และเพิ่มค่าปริมาณข้อมูลที่พบในแคชเท่ากับขนาดเอกสารที่เรียกขอ กรณีที่ไม่พบเอกสารภายในแคชจำเป็นที่จะต้องพิจารณาต่อว่า พื้นที่ภายในแคชมีเพียงพอที่จะเก็บเอกสารที่เรียกขอหรือไม่ ถ้ามีเพียงพอจะบันทึกชื่อเอกสารไว้ในแคชที่จำลองขึ้น แต่ถ้ามีพื้นที่ไม่เพียงพอ จำเป็นที่จะต้องพิจารณาแทนที่เอกสารภายในแคชตามขั้นตอนวิธีการแทนที่

ข้อดี

ก. ข้อมูลที่ใช้ในการทดสอบสามารถหาได้ง่าย

ข. ผลการทดสอบเปรียบเทียบแสดงให้เห็นถึงขั้นตอนวิธีการแทนที่ที่เหมาะสมกับสภาพการใช้เว็บที่นำมาศึกษา ดังนั้นจึงสามารถนำผลการทดสอบที่ได้ไปพัฒนาการทำงานของพร็อกซีแคชได้

ข้อเสีย

ก. ผลการทดสอบอาจนำไปใช้กับสภาพการใช้เว็บที่แตกต่างกันไม่ได้ เช่น ไม่สามารถนำผลการทดสอบกับสภาพการใช้เว็บขององค์กรที่ให้บริการด้านอินเทอร์เน็ตไปใช้กับสภาพการใช้เว็บของมหาวิทยาลัย เนื่องจากองค์กรที่ให้บริการด้านอินเทอร์เน็ตมีกลุ่มผู้ใช้ที่หลากหลายมากกว่าที่มหาวิทยาลัย ซึ่งผู้ใช้ส่วนใหญ่เป็น นักเรียน นักศึกษา และอาจารย์

4.1.2. การทดสอบประสิทธิภาพโดยการใช้ข้อมูลการใช้เว็บที่สร้างขึ้น (Model-driven - simulation)

การทดสอบประเภทนี้มีขั้นตอนการทดสอบเหมือนการทดสอบเปรียบเทียบประสิทธิภาพโดยการใช้ข้อมูลจริงในการจำลองการทำงานของพร็อกซีแคชเพื่อทดสอบ ส่วนที่แตกต่างกันคือ ส่วนของข้อมูลที่ใช้ในการจำลองการทำงานของพร็อกซีแคชที่สร้างขึ้นจากการวิเคราะห์การแจกแจงของข้อมูลทางสถิติของการเรียกขอ เช่น การแจกแจงของขนาดไฟล์ที่ถูกเรียกขอ การแจกแจงของความถี่การเรียกขอ

เอกสารเรียงตามลำดับความถี่ เป็นต้น และใช้พารามิเตอร์ของการแจกแจงนี้ในการสุ่มการเรียกขอ เพื่อใช้ในการทดสอบเปรียบเทียบได้

ข้อดี

ก. ผลการทดสอบทำให้ทราบว่าสภาพการใช้เว็บแต่ละประเภทเหมาะสมกับขั้นตอนวิธีการแทนที่แบบใด โดยการใช้ข้อมูลการใช้เว็บที่สร้างขึ้นให้มีลักษณะที่ใกล้เคียงกับสภาพการใช้เว็บนั้นๆ

ข. สามารถวางแผนการเลือกใช้ขั้นตอนวิธีการแทนที่ที่เหมาะสมกับการทำงานของพร็อกซีแคชในอนาคตได้ โดยการสังเกตจากแนวโน้มการเปลี่ยนแปลงสภาพการใช้เว็บจากค่าพารามิเตอร์ที่ใช้ในการสร้างข้อมูลการเรียกขอขึ้น

ข้อเสีย

ก. ข้อมูลการใช้เว็บที่สร้างขึ้นอาจมีคุณสมบัติไม่เพียงพอที่จะใช้ในการทดสอบขั้นตอนวิธีการแทนที่บางประเภท ทำให้ผลการทดสอบที่ได้คลาดเคลื่อนจากความเป็นจริง เช่น การใช้ข้อมูลการใช้เว็บที่สร้างขึ้นโดยไม่มีพารามิเตอร์ที่แสดงความถี่ในการเรียกขอเอกสารแต่ละเอกสารในการทดสอบประสิทธิภาพของขั้นตอนการแทนที่แบบ LFU ทำให้ผลการทดสอบที่ได้ไม่ถูกต้อง

การทดสอบเปรียบเทียบประสิทธิภาพในงานวิจัยนี้

การทดสอบเปรียบเทียบที่ใช้ คือ การทดสอบเปรียบเทียบประสิทธิภาพโดยการใช้ข้อมูลจริงในการจำลองการทำงานของพร็อกซีแคช

ข้อมูลที่ใช้ในการทดสอบคือข้อมูลการเรียกขอที่มีคุณสมบัติดังนี้

ก. รหัสการตอบกลับจากเว็บเซิร์ฟเวอร์แบบ "Successful" คือการเรียกขอที่ได้รับการตอบกลับพร้อมทั้งได้รับเอกสารตามที่เรียกขอ

ข. ยูอาร์แอลที่แสดงตำแหน่งของเอกสารจะต้องไม่มีเครื่องหมาย “?” และ “cgi-bin” เนื่องจากการทำงานของพร็อกซีแคชจะไม่เก็บเอกสารประเภทนี้ไว้ในพร็อกซีแคช

วิธีการคำนวณค่าฮิตเรโช และ ค่าไบท์ฮิตเรโช

การพบเอกสารภายในแคช (cache hit) คือ พบยูอาร์แอลของเอกสารที่ถูกเรียกขอเหมือนกับยูอาร์แอลของเอกสารในพร็อกซีแคช

ค่าฮิตเรโช และ ค่าไบท์ฮิตเรโช คำนวณได้จากสูตร

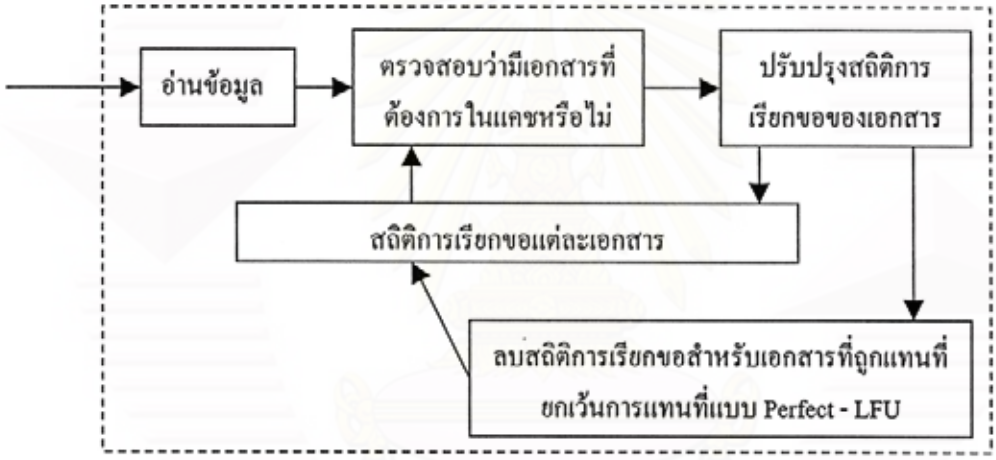
$$\text{ฮิตเรโช} = \frac{\text{ปริมาณการพบเอกสารภายในแคช}}{\text{ปริมาณของการเรียกขอที่ใช้ในการทดสอบ}}$$

$$\text{ไบท์ฮิตเรโช} = \frac{\text{ผลรวมของขนาดเอกสารจากการเรียกขอที่พบเอกสารในแคช}}{\text{ผลรวมของขนาดเอกสารจากการเรียกขอที่ใช้ในการทดสอบ}}$$

ค่าฮิตเรโช และ ค่าไบท์ฮิตเรโช สามารถใช้แทนประสิทธิภาพของพร็อกซีแคชได้ โดยค่าฮิตเรโช แสดงสัดส่วนของปริมาณการเรียกขอที่พร็อกซีแคชส่งเอกสารจากแคชไปยังไคลเอ็นต์ และ ไบท์ฮิตเรโช แสดงสัดส่วนของปริมาณข้อมูลของเอกสารที่พร็อกซีแคชส่งเอกสารจากแคชไปยังไคลเอ็นต์ การพัฒนาประสิทธิภาพของพร็อกซีแคชโดยการเพิ่มค่าฮิตเรโช ช่วยลดปริมาณการเชื่อมต่อไปยังเว็บเซิร์ฟเวอร์ที่อยู่ในเครือข่ายที่ห่างไกล ซึ่งเหมาะสมกับพร็อกซีเซิร์ฟเวอร์ที่รองรับปริมาณการเรียกขอจำนวนมากแต่การเรียกขอส่วนใหญ่เป็นการเรียกขอเอกสารที่มีขนาดเล็กและการเพิ่มค่า ไบท์ฮิตเรโช ช่วยลดปริมาณการถ่ายโอนข้อมูลผ่านระบบเครือข่าย ซึ่งเหมาะสมกับเครือข่ายที่มีการเชื่อมต่อออกสู่อินเทอร์เน็ตที่จำกัด เช่น การใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัย เป็นต้น

4.2. เครื่องมือที่ใช้ในการจำลองการทำงานของพร็อกซีแคช

งานวิจัยนี้ใช้การจำลองการทำงานของพร็อกซีแคชที่มีขั้นตอนวิธีการแทนที่แบบต่างๆ และใช้ข้อมูลการใช้เว็บในสภาพแวดล้อมจริงในการทดสอบเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่ โดยหลักการทำงานของโปรแกรมจำลองการทำงานของพร็อกซีแคชมีรูปแบบดังรูปที่ 4.1



รูปที่ 4.1 แผนภาพแสดงการทำงานของโปรแกรมจำลองการทำงานของพร็อกซีแคช

หลักการทำงานของโปรแกรมจำลองทำงานของพร็อกซีแคช แบ่งออกเป็น 4 ส่วน ได้แก่

4.2.1. อ่านข้อมูล

การอ่านข้อมูลการเรียกขอจากแฟ้มข้อมูลที่เก็บข้อมูลของการเรียกขอที่ได้รับการตอบกลับพร้อมส่งเอกสารกลับมาโดยข้อมูลการเรียกขอที่ใช้ในการจำลองการทำงานของพร็อกซีแคชมี 2 ส่วน คือ ยูอาร์แอล และ ขนาดของเอกสาร

4.2.2. ตรวจสอบว่ามีเอกสารที่ต้องการในแคชหรือไม่

การตรวจสอบใช้ยูอาร์แอลของการเรียกขอเปรียบเทียบกับยูอาร์แอลของเอกสารทั้งหมดในแคชว่ามียูอาร์แอลที่เหมือนกันหรือไม่

4.2.3. ปรับปรุงสถิติการเรียกขอเอกสาร

สำหรับทุกๆ เอกสารในแคชจะมีสถิติการเรียกขออยู่ เมื่อมีการเรียกขอที่มีอยู่ในแคชอยู่แล้ว ต้องปรับปรุงสถิติการเรียกขอเอกสารใหม่ และใช้ค่าเริ่มต้นสำหรับเอกสารที่มีการเรียกขอเป็นครั้งแรก

4.2.4. ลบสถิติการเรียกขอ

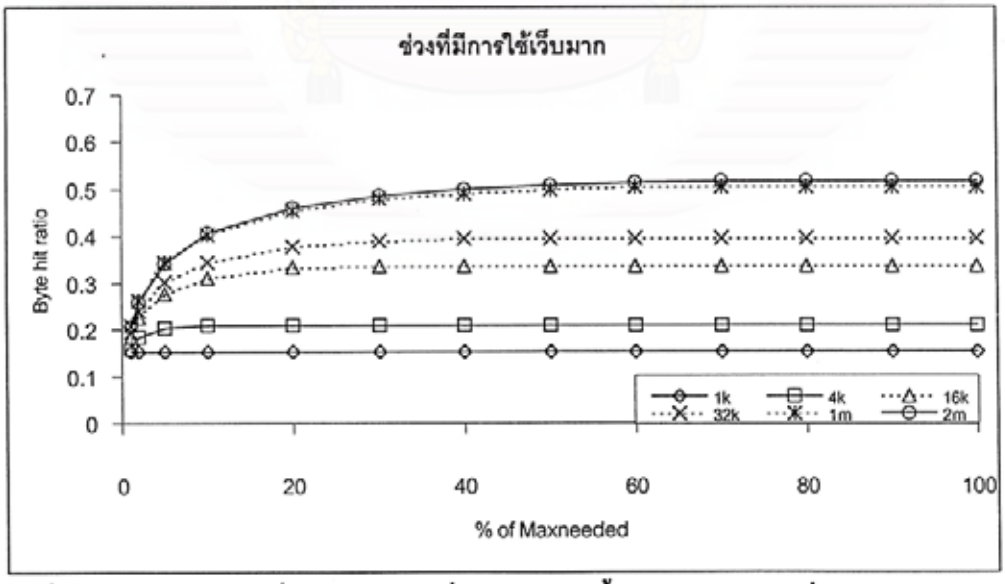
สำหรับเอกสารที่ถูกแทนที่ ยกเว้นการแทนที่แบบ Perfect – LFU กรณีที่ขนาดรวมของเอกสารทั้งหมดในแคชมากกว่าขนาดของแคช จำเป็นที่จะต้องแทนที่เอกสารในแคชโดยพิจารณาจากสถิติการเรียกขอเอกสารในแคชและประเภทของขั้นตอนวิธีการแทนที่

4.3. การทดสอบเปรียบเทียบเทียบขั้นตอนวิธีการแทนที่ภายใช้สภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัย

การทดสอบเปรียบเทียบขั้นตอนวิธีการแทนที่แบ่งออกเป็น 2 ส่วน คือ

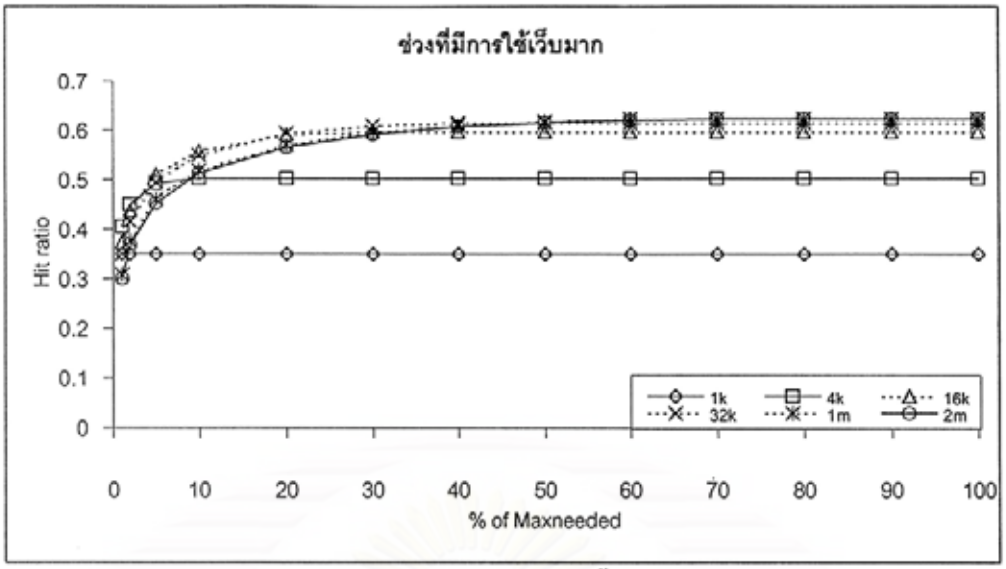
4.3.1. การหาพารามิเตอร์ที่เหมาะสมของขั้นตอนวิธีการแทนที่ LRU – TH

พารามิเตอร์ที่เหมาะสมของขั้นตอนวิธีการแทนที่ LRU – TH ได้จากผลการจำลองการทำงานของพร็อกซีแคชโดยใช้ข้อมูลการใช้เว็บที่แบ่งออกเป็น 2 ส่วน คือ ช่วงเวลาที่มีการใช้เว็บมากและช่วงเวลาที่มีการใช้เว็บน้อย และเครื่องวัดที่ใช้ในการเปรียบเทียบประสิทธิภาพมี 2 ค่า คือ ค่าฮิตเรโชและค่าไบท์ฮิตเรโช ที่ใช้ในการทดสอบ การทดสอบจะใช้ค่าขีดแบ่ง(threshold)ทั้งหมด 6 ค่า คือ 1 4 16 32 กิโลไบต์ 1 และ 2 เมกกะไบต์ ได้ผลการทดสอบดังแผนภูมิในรูปที่ 4.2 4.3 4.4 และ 4.5 การเลือกค่าขีดแบ่งที่เหมาะสมจะเลือกค่าขีดแบ่งที่น้อยที่สุดที่ให้ประสิทธิภาพใกล้เคียงกับประสิทธิภาพที่ได้จากค่าขีดแบ่งที่มากกว่าเนื่องจากการเลือกใช้ขั้นตอนวิธีการแทนที่แบบ LRU-TH ที่มีค่าขีดแบ่งน้อยส่งผลให้พร็อกซีแคชลดปริมาณเอกสารที่ต้องพิจารณาในการจัดเก็บหรือลดการทำงานของพร็อกซีแคชได้



รูปที่ 4.2 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศของจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าไบท์ฮิตเรโช

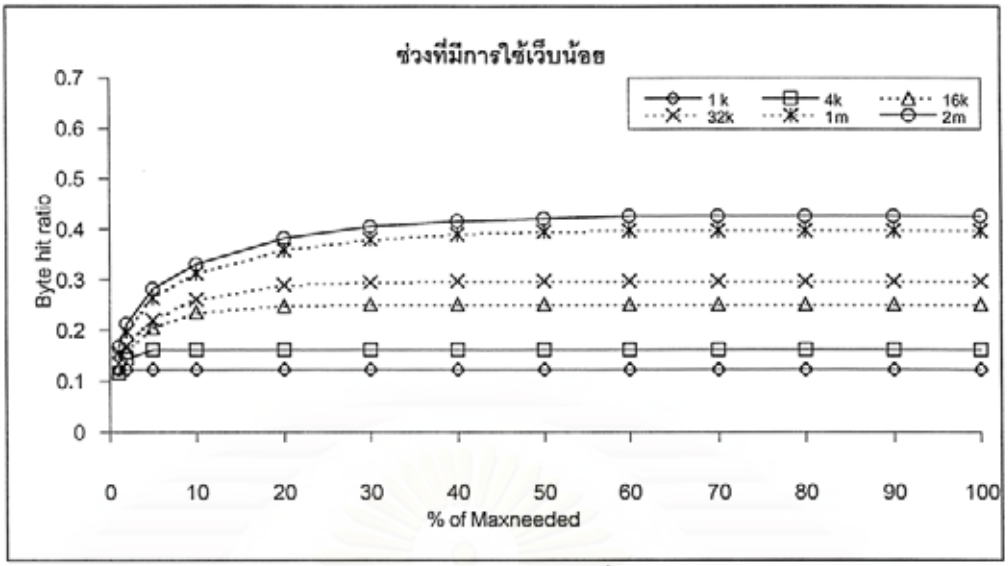
* Maxneeded คือ ขนาดของพร็อกซีแคชที่สามารถจัดเก็บเอกสารได้โดยไม่มีการแทนที่



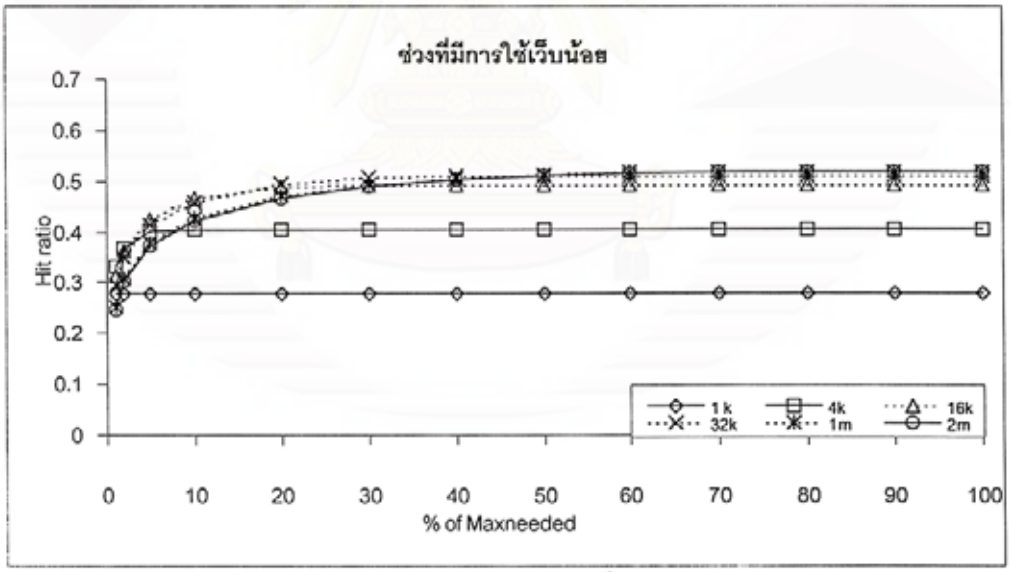
รูปที่ 4.3 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศของจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าฮิตเรโช

ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่แบบ LRU-TH เพื่อหาค่าขีดแบ่งที่เหมาะสมภายใต้สภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัยในช่วงที่มีการใช้เว็บมากโดยมีค่าฮิตเรโชและค่าไบท์ฮิตเรโชเป็นเครื่องวัดดังแผนภูมิในรูปที่ 4.2 และ 4.3 เมื่อพิจารณาประสิทธิภาพจากค่าฮิตเรโช ค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่แบบ LRU-TH ขึ้นกับความจุของพร็อกซีแคช คือ กรณีที่พร็อกซีแคชมีความจุน้อย ค่าขีดแบ่งที่เหมาะสมคือค่า 1 กิโลไบต์ หรือ 4 กิโลไบต์ และกรณีที่พร็อกซีแคชมีความจุมาก ค่าขีดแบ่งที่เหมาะสมคือค่า 1 เมกกะไบต์ หรือ 2 เมกกะไบต์ ดังแผนภูมิในรูปที่ 4.3 เนื่องจากในกรณีที่พร็อกซีแคชมีความจุน้อย การเลือกที่จะเก็บเฉพาะเอกสารที่มีขนาดเล็กส่งผลให้พร็อกซีแคชสามารถเก็บเอกสารได้จำนวนมาก ส่งผลให้ค่าฮิตเรโชที่ได้มากกว่าการเก็บเอกสารขนาดใหญ่ ซึ่งจำเป็นที่จะต้องแทนที่เอกสารขนาดเล็กจำนวนมากเพื่อจัดเก็บเอกสารขนาดใหญ่

การพิจารณาประสิทธิภาพจากค่าไบท์ฮิตเรโชพบว่าประสิทธิภาพของพร็อกซีแคชที่ได้ขึ้นกับค่าขีดแบ่งที่เลือกใช้ โดยการเลือกใช้ค่าขีดแบ่งที่มีค่ามากส่งผลให้ประสิทธิภาพของพร็อกซีแคชมากขึ้น แต่โดยการพิจารณาจากแผนภูมิในรูปที่ 4.2 พบว่าประสิทธิภาพที่ได้จากการเลือกใช้ขั้นตอนวิธีการแทนที่แบบ LRU-TH ที่มีค่าขีดแบ่ง 1 เมกกะไบต์และค่าขีดแบ่ง 2 เมกกะไบต์ใกล้เคียงกันเมื่อเทียบกับการเลือกใช้ค่าขีดแบ่ง 32 กิโลไบต์ ดังนั้นค่าขีดแบ่งที่เหมาะสมสำหรับสภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัยในช่วงที่มีการใช้เว็บมากคือ 1 เมกกะไบต์



รูปที่ 4.4 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศของจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าไบท์ฮิตเรโช



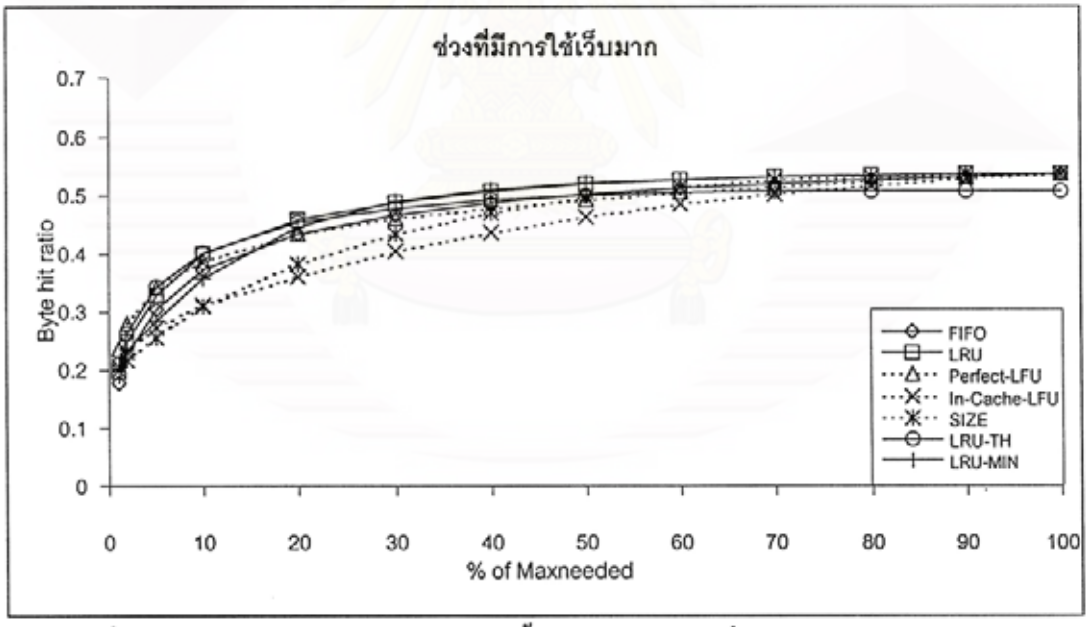
รูปที่ 4.5 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศของจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าฮิตเรโช

ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่แบบ LRU-TH เพื่อหาค่าขีดแบ่งที่เหมาะสมภายใต้สภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัยในช่วงที่มีการใช้เว็บน้อยแสดงให้เห็นว่าค่าขีดแบ่งที่เหมาะสมคือ 1 เมกกะไบต์ ถึงแม้ว่าที่ค่านี้ให้ประสิทธิภาพที่น้อยกว่าค่าขีดแบ่ง 2 เมกกะไบต์แต่เพียงเล็กน้อยเท่านั้นจากแผนภูมิในรูปที่ 4.4 เมื่อเทียบกับการลดภาระให้กับพร็อกซีแคชได้มากกว่าเนื่องจากไม่ต้องเก็บเอกสารที่มีขนาดตั้งแต่ 1 เมกกะไบต์ขึ้นไป

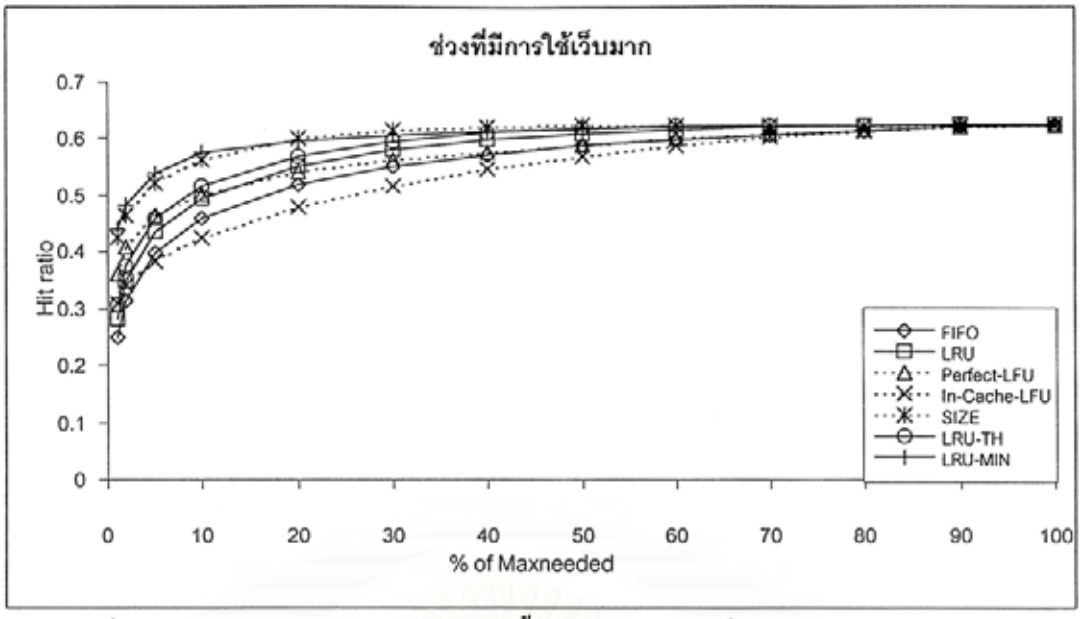
ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมของขั้นตอนวิธีการแทนที่แบบ LRU-TH สำหรับสภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัยในช่วงที่มีการใช้เว็บมากและในช่วงที่มีการใช้เว็บน้อยได้ผลการทดสอบเหมือนกันคือที่ค่าขีดแบ่ง 1 เมกกะไบต์ จึงสรุปได้ว่าช่วงเวลาในการใช้เว็บไม่มีผลต่อการเลือกค่าขีดแบ่งที่เหมาะสมสำหรับขั้นตอนวิธีการแทนที่แบบ LRU-TH

4.3.2 ประสิทธิภาพของขั้นตอนวิธีการแทนที่

การทดสอบเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่ทั้งหมด 7 แบบได้แก่ FIFO LRU Perfect LFU In-Cache LFU SIZE LRU-TH ที่มีค่าขีดแบ่ง 1 เมกกะไบต์ และ LRU-MIN โดยใช้ค่าฮิตเรโชเป็นเครื่องวัด และการทดสอบเปรียบเทียบโดยใช้ค่าไบท์ฮิตเรโชเป็นเครื่องวัดประสิทธิภาพ ได้ผลการทดสอบดังแผนภูมิในรูปที่ 4.6 4.7 4.8 และ 4.9



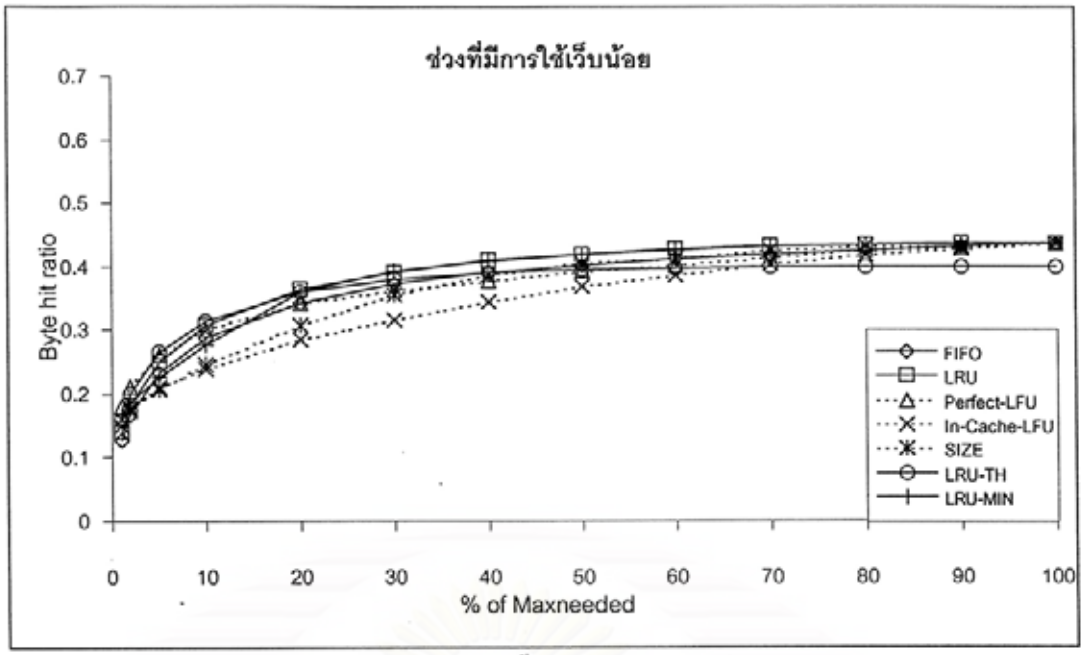
รูปที่ 4.6 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย โดยพิจารณาจากค่าไบท์ฮิตเรโช



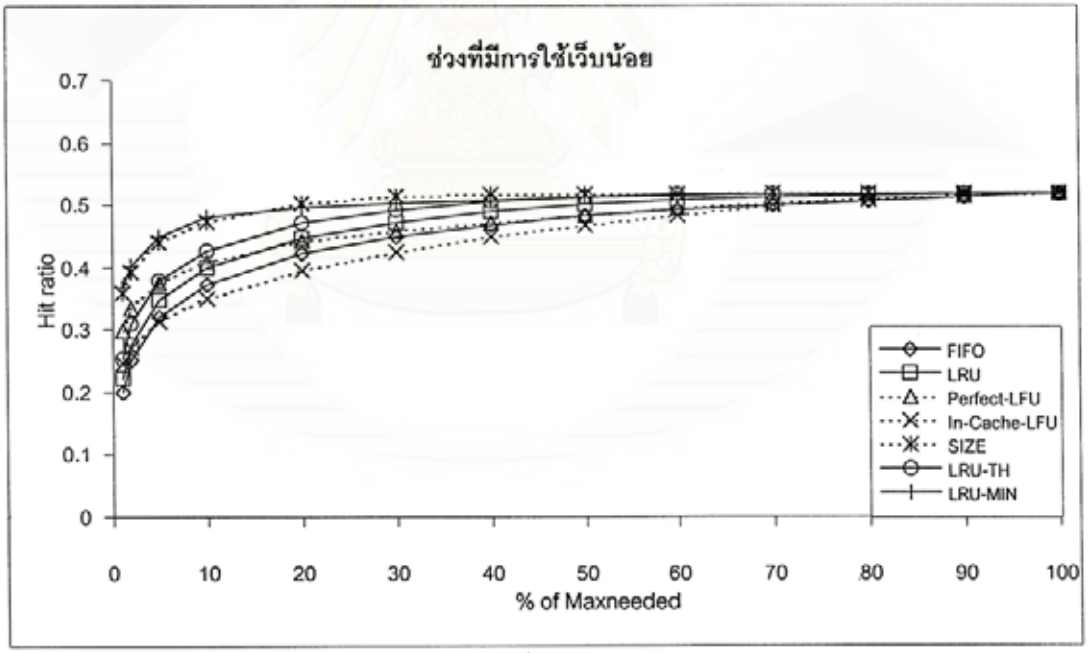
รูปที่ 4.7 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย โดยพิจารณาจากค่าฮิตเรโช

ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ทั้งหมด 7 แบบภายใต้สภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัยในช่วงที่มีการใช้เว็บมากโดยมีค่าฮิตเรโชและค่าไบท์ฮิตเรโชเป็นเครื่องวัดดังแผนภูมิในรูปที่ 4.6 และ รูปที่ 4.7 แสดงให้เห็นว่าขั้นตอนวิธีที่เหมาะสมกับช่วงที่มีการใช้เว็บมากโดยมีค่าไบท์ฮิตเรโชเป็นเครื่องวัดคือขั้นตอนวิธีการแทนที่แบบ Perfect-LFU และ LRU-TH สำหรับแคชที่มีความจุน้อยกว่า 10 เปอร์เซ็นต์ของความจุแคชที่ไม่มีการแทนที่ และขั้นตอนวิธีการแทนที่แบบ LRU-MIN สำหรับแคชที่มีความจุมากกว่า 10 เปอร์เซ็นต์ของความจุแคชที่ไม่มีการแทนที่ แต่ขั้นตอนวิธีการแทนที่แบบ LRU ซึ่งเป็นขั้นตอนการแทนที่ของพร็อกซีแคชในปัจจุบันมีประสิทธิภาพใกล้เคียงกับขั้นตอนวิธีการแทนที่ที่ให้ประสิทธิภาพสูงสุด ในขณะที่ขั้นตอนวิธีการแทนที่แบบ In-cache-LFU ซึ่งมีหลักการแทนที่ที่คล้ายกันกับขั้นตอนวิธีการแทนที่แบบ Perfect-LFU แต่มีการลบสถิติของปริมาณการเรียกขอเมื่อเอกสารถูกแทนที่เป็นขั้นตอนวิธีการแทนที่ที่ให้ประสิทธิภาพน้อยที่สุดเมื่อแคชมีความจุน้อยกว่า 10 เปอร์เซ็นต์ของความจุแคชที่ไม่มีการแทนที่ และขั้นตอนวิธีการแทนที่แบบ SIZE เป็นขั้นตอนวิธีการแทนที่ที่ให้ประสิทธิภาพน้อยที่สุดเมื่อแคชมีความจุมากกว่า 10 เปอร์เซ็นต์ของความจุแคชที่ไม่มีการแทนที่

การพิจารณาประสิทธิภาพของขั้นตอนวิธีการแทนที่โดยพิจารณาจากค่าฮิตเรโชพบว่าขั้นตอนวิธีการแทนที่ที่ให้ประสิทธิภาพสูงสุดคือขั้นตอนวิธีการแทนที่แบบ LRU-MIN สำหรับแคชที่มีความจุน้อยกว่า 10 เปอร์เซ็นต์ของความจุแคชที่ไม่มีการแทนที่ และขั้นตอนวิธีการแทนที่แบบ SIZE สำหรับแคชที่มีความจุมากกว่า 10 เปอร์เซ็นต์ของความจุแคชที่ไม่มีการแทนที่ เนื่องจากขั้นตอนวิธีการแทนที่แบบ SIZE และ LRU-MIN เป็นขั้นตอนวิธีการแทนที่ที่มีการพิจารณาลดปริมาณเอกสารที่ถูกแทนที่โดยการพิจารณาจากขนาดของเอกสารที่ถูกแทนที่



รูปที่ 4.8 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าไบท์ฮิตเรโช



รูปที่ 4.9 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าฮิตเรโช

ผลจากการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัยในช่วงที่มีการใช้เว็บน้อยโดยมีค่าฮิตเรโชและค่าไบท์ฮิตเรโชเป็นเครื่องวัดได้ผลดังแผนภูมิในรูปที่ 4.8 และ รูปที่ 4.9 แสดงให้เห็นว่าขั้นตอนวิธีการแทนที่ที่ให้ประสิทธิภาพสูงสุดในกรณีที่ใช้ค่าฮิตเรโชเป็นเครื่องวัดคือขั้นตอนวิธีการแทนที่แบบ Perfect-LFU และ LRU-TH สำหรับแคชที่มีความจุน้อยกว่า 10 เปอร์เซ็นต์ของความจุแคชที่ไม่มีการแทนที่ และขั้นตอนวิธีการแทนที่แบบ LRU-MIN สำหรับแคชที่มีความจุน้อยกว่า 10 เปอร์เซ็นต์ของความจุแคชที่ไม่มีการแทนที่ ซึ่งขั้นตอนวิธีที่มีประสิทธิภาพสูงสำหรับช่วงที่มีการใช้เว็บน้อยคือขั้นตอนวิธีเดียวกันกับช่วงที่มีการใช้เว็บมาก โดยค่าฮิตเรโชและค่าไบท์ฮิตเรโชที่ได้ต่ำกว่าในช่วงที่มีการใช้เว็บมากประมาณ 0.2

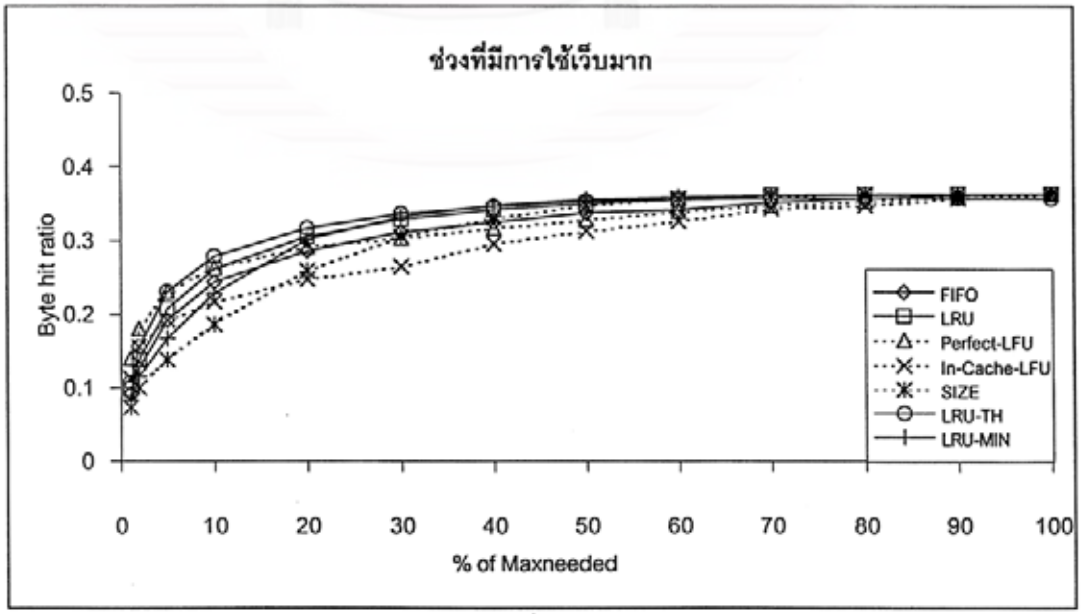
ขั้นตอนวิธีการแทนที่ที่ให้ประสิทธิภาพน้อยที่สุดเมื่อพิจารณาจากค่าฮิตเรโช และค่าไบท์ฮิตเรโช คือขั้นตอนวิธีการแทนที่แบบ In-cache LFU และ FIFO ดังนั้นในการปรับปรุงประสิทธิภาพการทำงานของพร็อกซีแคชให้เหมาะสมกับสภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัย ไม่ควรนำขั้นตอนวิธีการแทนที่ที่พิจารณาแทนที่เอกสารตามความถี่ในการถูกเรียกขอที่ผ่านมา หรือ In-cache LFU และขั้นตอนวิธีการแทนที่แบบ FIFO มาใช้

4.4. การทดสอบเปรียบเทียบขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บที่อื่นๆ

4.4.1. การทดสอบเปรียบเทียบเทียบขั้นตอนวิธีการแทนที่ภายใช้สภาพการใช้เว็บของพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย

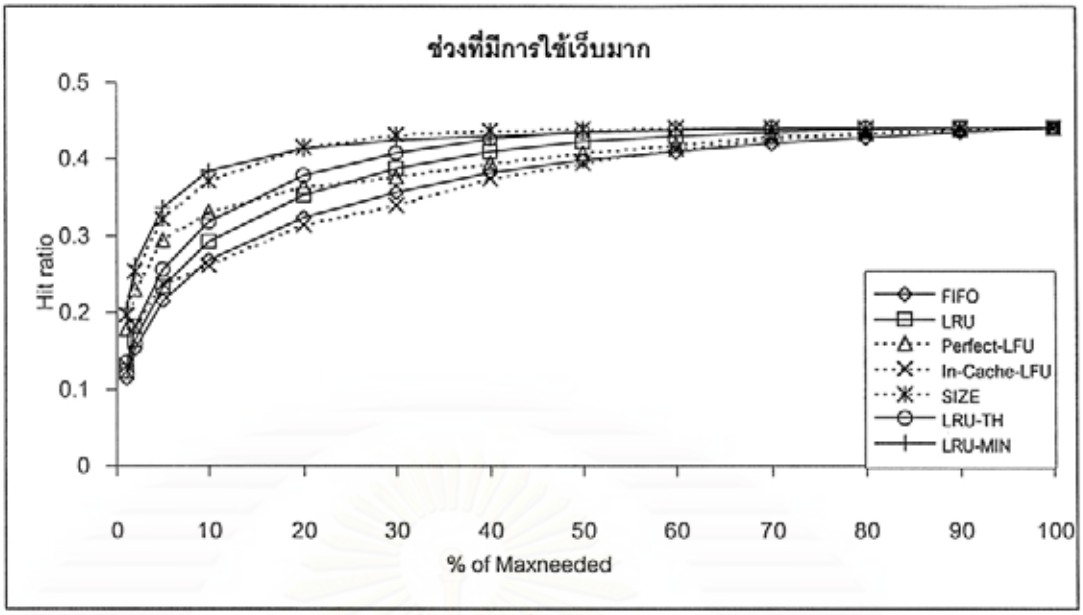
การทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมสำหรับขั้นตอนวิธีการแทนที่แบบ LRU-TH ภายใต้สภาพการใช้เว็บของพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย ได้ค่าขีดแบ่งที่เหมาะสมคือ 1 เมกกะไบต์ โดยแผนภูมิแสดงผลการทดสอบอยู่ในภาคผนวก ค.

ผลการทดสอบเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่โดยใช้ค่าฮิตเรโชและค่าไบท์ฮิตเรโช เป็นเครื่องวัด ได้ผลดังแผนภูมิในรูปที่ 4.10 และ 4.11



รูปที่ 4.10 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย โดยพิจารณาจากค่าไบท์ฮิตเรโช

I19190372



รูปที่ 4.11 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย โดยพิจารณาจากค่าฮิตเรโช

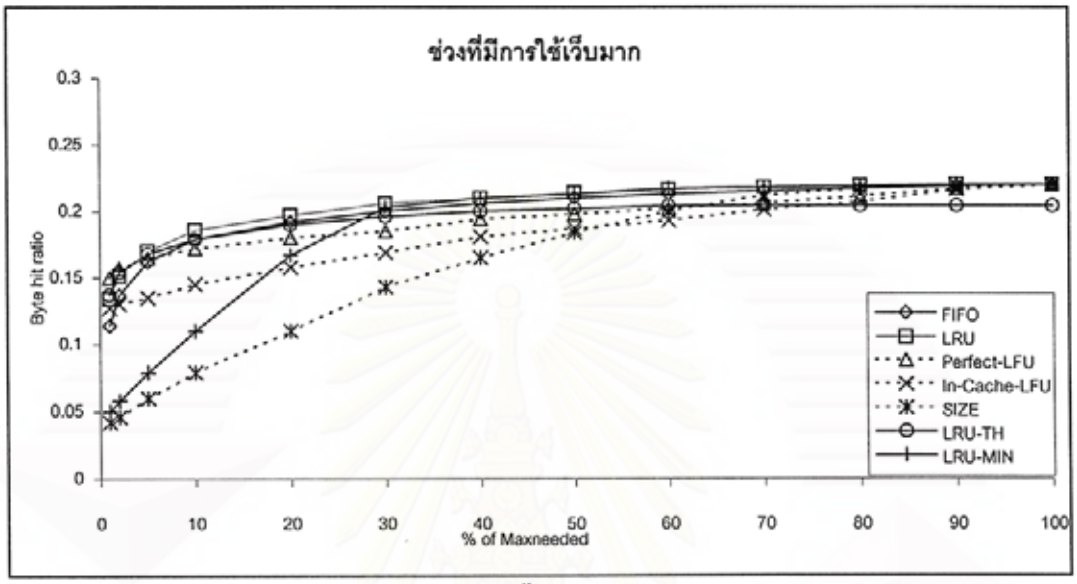
ผลการทดสอบจากแสดงให้เห็นว่าขั้นตอนวิธีการแทนที่ที่มีประสิทธิภาพสูงสุดภายใต้สภาพการใช้เว็บจากพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย ในช่วงที่มีการใช้เว็บมากและช่วงที่มีการใช้เว็บน้อยคือขั้นตอนการแทนที่แบบเดียวกันโดยแผนภูมิแสดงผลการทดสอบประสิทธิภาพของขั้นตอนการแทนที่ในช่วงที่มีการใช้เว็บน้อยอยู่ในภาคผนวก ค.

ขั้นตอนวิธีการแทนที่ที่มีประสิทธิภาพสูงสุดในช่วงเวลาที่มีการใช้เว็บมากและช่วงเวลาที่มีการใช้เว็บน้อย ได้แก่ ขั้นตอนวิธีการแทนที่ LRU-MIN และ SIZE ในกรณีที่ใช้ค่าฮิตเรโชเป็นเครื่องวัดประสิทธิภาพ และ ขั้นตอนวิธีการแทนที่แบบ LFU-Perfect LRU-TH (ค่าขีดแบ่ง = 1 เมกกะไบต์) และ LRU-MIN ในกรณีที่ใช้ไบท์ฮิตเรโชเป็นเครื่องวัด ผลการทดสอบพบว่าขั้นตอนวิธีการแทนที่ที่ให้ประสิทธิภาพสูงสุดเป็นขั้นตอนวิธีการแทนที่แบบเดียวกับของข้อมูลการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัยโดย ค่าฮิตเรโช และ ค่าไบท์ฮิตเรโช จะมีค่าน้อยกว่าเนื่องจากพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย เป็นพร็อกซีแคชที่รองรับปริมาณการเรียกขอต่อวันน้อยกว่าคือประมาณ 12.06 เปอร์เซ็นต์ของปริมาณการเรียกขอของจุฬาลงกรณ์มหาวิทยาลัยทั้งหมด

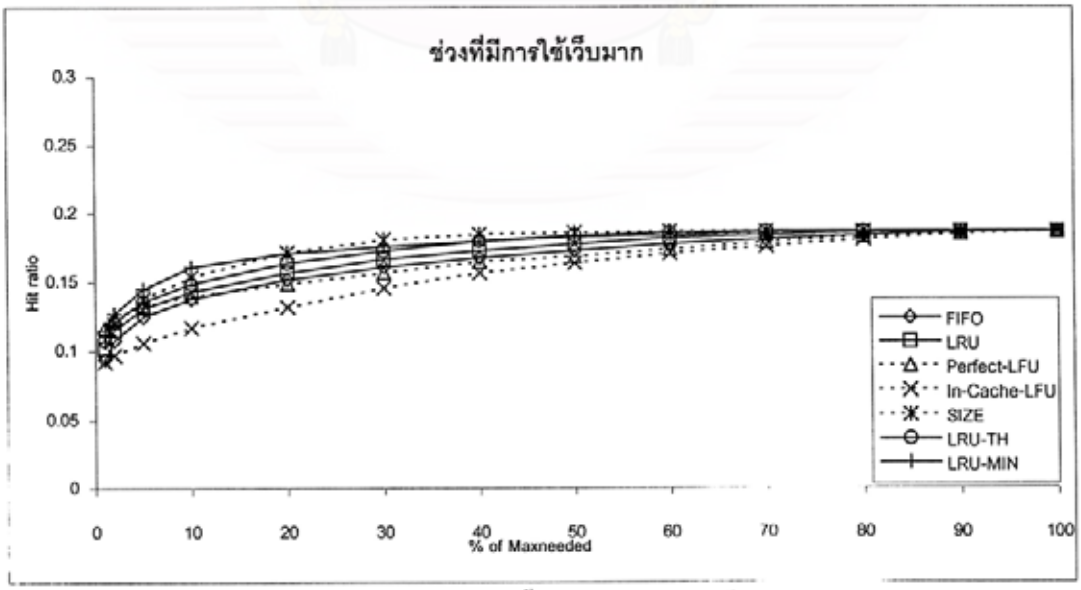
4.4.2. การทดสอบเปรียบเทียบเทียบขั้นตอนวิธีการแทนที่ภายใช้สภาพการใช้เว็บของเอ็นแอลเอเอ็นอาร์

การทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมสำหรับขั้นตอนวิธีการแทนที่แบบ LRU-TH ภายใต้สภาพการใช้เว็บของเอ็นแอลเอเอ็นอาร์ได้ค่าขีดแบ่งที่เหมาะสมคือ 1 เมกกะไบต์ โดยแผนภูมิแสดงผลการทดสอบอยู่ในภาคผนวก ค.

ผลการทดสอบเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่โดยใช้ค่าฮิตเรโช และค่าไบท์ฮิตเรโชเป็นเครื่องวัด ได้ผลดังแผนภูมิในรูปที่ 4.12 และ 4.13



รูปที่ 4.12 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของเอ็นแอลเอเอ็นอาร์โดยพิจารณาจากค่า ไบท์ฮิตเรโช



รูปที่ 4.13 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของเอ็นแอลเอเอ็นอาร์โดยพิจารณาจากค่าฮิตเรโช

ผลการทดสอบเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บจากเอ็นแอลเอเอ็นอาร์ในช่วงที่มีการใช้เว็บมากและช่วงที่มีการใช้เว็บน้อย พบว่าขั้นตอนวิธีการแทนที่ที่ให้ประสิทธิภาพสูงสุดเมื่อพิจารณาจากค่าไบท์ฮิตเรโชคือขั้นตอนการแทนที่แบบ Perfect-LFU และ LRU-TH ในกรณีที่แคชมีความจุน้อยกว่า 10 เปอร์เซ็นต์ของความจุแคชที่ไม่มีการแทนที่ และขั้นตอนวิธีการแทนที่แบบ LRU-MIN ในกรณีที่แคชมีความจุมากกว่า 10 เปอร์เซ็นต์ และโดยการพิจารณาประสิทธิภาพจากค่าฮิตเรโช พบว่าขั้นตอนวิธีการแทนที่แบบ LRU-MIN จะให้ประสิทธิภาพสูงสุดในกรณีที่แคชมีความจุน้อยกว่า 10 เปอร์เซ็นต์ของความจุแคชที่ไม่มีการแทนที่และขั้นตอนวิธีการแทนที่แบบ SIZE ในกรณีที่แคชมีความจุมากกว่า 10 เปอร์เซ็นต์ ซึ่งผลการทดสอบที่ได้นี้เหมือนกันกับขั้นตอนวิธีการแทนที่ที่เหมาะสมกับสภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัย แต่ค่าฮิตเรโชและค่าไบท์ฮิตเรโชมีค่าน้อยกว่าค่าที่ได้จากการทดสอบภายใต้สภาพการใช้เว็บจุฬาลงกรณ์มหาวิทยาลัยมากโดยค่าฮิตเรโชและค่าไบท์ฮิตเรโชสูงสุดที่ได้มีค่าประมาณ 0.2 เท่านั้น เนื่องจากเป็นพร็อกซีแคชที่อยู่ในระดับสูงกว่าพร็อกซีแคชที่จุฬาลงกรณ์มหาวิทยาลัย

4.5. แนวทางการนำผลการทดสอบไปประยุกต์ใช้กับสภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัย

ผลการทดสอบที่ได้จากข้อ 4.3. แสดงขั้นตอนวิธีการแทนที่ที่เหมาะสมกับสภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าฮิตเรโช และค่าไบท์ฮิตเรโช ซึ่งผลการทดสอบแสดงถึงความสัมพันธ์ระหว่างขนาดของพร็อกซีแคชที่มีผลต่อประสิทธิภาพของขั้นตอนวิธีการแทนที่แต่ละแบบ ดังนั้นการนำผลการทดสอบที่ได้ไปใช้ในสภาพการใช้เว็บจริงจะต้องพิจารณาทั้งขนาดของพร็อกซีแคชและเครื่องวัดที่ใช้ในการวัดประสิทธิภาพ

4.5.1. การประมาณขนาดของพร็อกซีแคชในปัจจุบันเทียบกับ maxneeded

การประมาณขนาดของพร็อกซีแคชใช้การคำนวณค่าเฉลี่ยของค่าฮิตเรโช และ ค่าไบท์ฮิตเรโช จากแฟ้มที่บันทึกการเรียกขอในแต่ละช่วงเวลาและนำมาเทียบกับผลการทดสอบเปรียบเทียบประสิทธิภาพที่ได้เพื่อประมาณขนาดของพร็อกซีแคชจากค่าฮิตเรโชและค่าไบท์ฮิตเรโชในปัจจุบันของพร็อกซีแคช

	ค่าฮิตเรโช	ค่าไบท์ฮิตเรโช
ช่วงที่มีการใช้เว็บมาก	0.458	0.388
ช่วงที่มีการใช้เว็บน้อย	0.405	0.298

ตารางที่ 4.1 ค่าเฉลี่ยของค่าฮิตเรโช และ ไบท์ฮิตเรโช จากแฟ้มที่บันทึกการเรียกขอ

จากตารางที่ 4.1 สามารถนำไปเปรียบเทียบกับผลการทดสอบเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่แบบ LRU พบว่าจะได้ค่า maxneeded ของการใช้เว็บจากจุฬาลงกรณ์มหาวิทยาลัยประมาณ 9.16 เปอร์เซ็นต์

4.5.2. การวิเคราะห์ความสัมพันธ์ระหว่างขนาดของพร็อกซีแคชและขั้นตอนวิธีการแทนที่ที่มีต่อประสิทธิภาพของพร็อกซีแคช

การปรับปรุงประสิทธิภาพของพร็อกซีแคชให้ได้ผลสูงสุด จะต้องพิจารณาความสัมพันธ์ระหว่างแฟคเตอร์ที่เกี่ยวข้องและประสิทธิภาพของพร็อกซีแคชที่ได้ ในช่วงเวลาที่มีการใช้เว็บมากและช่วงที่มีการใช้เว็บน้อย โดยการใช้การทดสอบแบบแฟคทอเรียลดังตารางที่ 4.2 และ 4.3

แฟคเตอร์	ระดับ
ขนาดพร็อกซีแคช (A)	10%, 30% และ 60% ของ maxneeded
ขั้นตอนวิธีการแทนที่ (B)	LRU, SIZE, LRU-MIN

ตารางที่ 4.2 แฟคเตอร์ที่ใช้ในการทดสอบแบบแฟคทอเรียลสำหรับการใช้ค่าฮิตเรโชเป็นเครื่องวัด

แฟคเตอร์	ระดับ
ขนาดพร็อกซีแคช (A)	10%, 30% และ 60% ของ maxneeded
ขั้นตอนวิธีการแทนที่ (B)	LRU, LRU-TH, LRU-MIN

ตารางที่ 4.3 แฟคเตอร์ที่ใช้ในการทดสอบแบบแฟคทอเรียลสำหรับการใช้ค่าไบท์ฮิตเรโชเป็นเครื่องวัด

ผลการวิเคราะห์ความสัมพันธ์ระหว่างขนาดของพร็อกซีแคชและขั้นตอนวิธีการแทนที่ที่มีต่อประสิทธิภาพของพร็อกซีแคชโดยใช้การทดสอบแบบแฟคทอเรียลที่มีสมมติฐานว่าระดับของแฟคเตอร์ที่ใช้มีผลต่อประสิทธิภาพของพร็อกซีแคชที่ได้ ที่ระดับนัยสำคัญเท่ากับ 0.05 แสดงให้เห็นว่า การปรับปรุงประสิทธิภาพของพร็อกซีแคชที่ให้ประสิทธิภาพสูงสุดสำหรับแต่ละช่วงการใช้เว็บได้ผลดังตารางที่ 4.4

เครื่องวัด	ช่วงการใช้เว็บ	ผลของแฟคเตอร์ A	ผลของแฟคเตอร์ B
ฮิตเรโช	มาก	มีผล	มีผล
	น้อย	มีผล	มีผล
ไบท์ฮิตเรโช	มาก	มีผล	ไม่มีผล
	น้อย	มีผล	ไม่มีผล

ตารางที่ 4.4 ผลการทดสอบแฟคทอเรียลเพื่อพิจารณาความสัมพันธ์ระหว่างขนาดของพร็อกซีแคชและขั้นตอนวิธีการแทนที่ที่มีต่อประสิทธิภาพของพร็อกซีแคช

ผลการทดสอบจากตารางที่ 4.4 สรุปได้ว่า การปรับปรุงประสิทธิภาพของพร็อกซีแคชทำได้โดยการเพิ่มความจุของพร็อกซีแคชหรือเปลี่ยนขั้นตอนวิธีการแทนที่ของพร็อกซีแคชจากขั้นตอนวิธีการแทนที่

แบบ LRU เป็น SIZE หรือ LRU-MIN เพื่อปรับปรุงค่าฮิตเรโชได้ โดยค่าฮิตเรโช ที่เพิ่มขึ้นเมื่อใช้ขั้นตอนวิธีการแทนที่แบบ SIZE และ LRU-MIN มีค่าดังตารางที่ 4.5

	LRU-MIN	SIZE
ช่วงที่มีการใช้เว็บมาก	13.91 %	16.13 %
ช่วงที่มีการใช้เว็บน้อย	20.30 %	18.80 %

ตารางที่ 4.5 ประสิทธิภาพของพร็อกซีแคชที่เพิ่มขึ้นโดยใช้ค่าฮิตเรโช เป็นเครื่องวัด

ตารางที่ 4.5 แสดงเปอร์เซ็นต์ของประสิทธิภาพที่สามารถปรับปรุงได้เทียบกับประสิทธิภาพ หรือเปอร์เซ็นต์ของค่าฮิตเรโชที่สามารถปรับปรุงได้ โดยการพิจารณาจากค่าที่ได้จากตารางที่ 4.5 สรุปได้ว่าขั้นตอนวิธีการแทนที่ที่เหมาะสมที่สุดสำหรับช่วงที่มีการใช้เว็บมากคือขั้นตอนวิธีการแทนที่แบบ SIZE และขั้นตอนวิธีการแทนที่ที่เหมาะสมที่สุดสำหรับช่วงที่มีการใช้เว็บน้อยคือขั้นตอนวิธีการแทนที่แบบ LRU-MIN

4.6. สรุปผลการทดสอบเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่

จากผลการจำลองการทำงานของพร็อกซีแคช สรุปได้ว่า ประสิทธิภาพของขั้นตอนวิธีการแทนที่ขึ้นกับปริมาณการเรียกขอและระดับของพร็อกซีแคช โดยภายใต้สภาพการใช้เว็บเดียวกันพร็อกซีแคชที่รองรับปริมาณการเรียกขอมากจะมีค่าฮิตเรโชและไบท์ฮิตเรโชที่สูงกว่าจากผลของการจำลองการทำงานของพร็อกซีแคชภายใต้สภาพการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศ และพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย และเมื่อพิจารณาข้อมูลการใช้เว็บจากพร็อกซีแคชที่เอ็นแอลเอเอ็นอาร์พบว่าให้ค่าฮิตเรโชและค่าไบท์ฮิตเรโชที่ต่ำกว่าข้อมูลการใช้เว็บจากพร็อกซีแคชที่สำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย เนื่องจากพร็อกซีแคชที่เอ็นแอลเอเอ็นอาร์เป็นพร็อกซีแคชระดับสูงเมื่อเทียบกับพร็อกซีแคชที่สำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย

ผลการจำลองการทำงานของพร็อกซีแคชโดยการใช้ข้อมูลการใช้เว็บจากทั้ง 3 แห่ง แสดงให้เห็นว่าขั้นตอนวิธีการแทนที่ที่เหมาะสมกับสภาพการใช้เว็บทั้ง 3 แห่งเป็นประเภทเดียวกันและขั้นตอนวิธีการแทนที่ที่เหมาะสมกับสภาพการใช้เว็บในช่วงที่มีการใช้เว็บมากและช่วงที่มีการใช้เว็บน้อยของแต่ละแห่งเป็นประเภทเดียวกัน โดยขั้นตอนวิธีการแทนที่ที่เหมาะสมขึ้นอยู่กับขนาดของพร็อกซีแคชและเครื่องวัดที่ใช้ในการวัดประสิทธิภาพ ขั้นตอนวิธีการแทนที่ที่เหมาะสมสำหรับสภาพการใช้เว็บทั้ง 3 แห่ง แสดงในตารางที่ 4.6

เครื่องวัด	ความจุ	
	น้อย	มาก
ค่าฮิตเรโช	LRU-MIN	SIZE
ค่าไบท์ฮิตเรโช	LRU-TH, Perfect-LFU	LRU-MIN

ตารางที่ 4.6 ขั้นตอนวิธีการแทนที่ที่เหมาะสมกับพร็อกซีแคชที่มีความจุและเครื่องวัดที่กำหนด

การทดสอบแบบแฟคทอเรียลกับผลการจำลองการทำงานของพร็อกซีแคชภายใต้สภาพการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย เพื่อศึกษาผลของความจุแคชและประเภทของขั้นตอนวิธีการแทนที่ที่มีต่อประสิทธิภาพของพร็อกซีแคได้ว่าการปรับปรุงประสิทธิภาพของพร็อกซีแคชโดยพิจารณาจากค่าฮิตเรโชสามารถทำได้โดยการเปลี่ยนขั้นตอนวิธีการแทนที่หรือเพิ่มความจุของพร็อกซีแคช และการปรับปรุงประสิทธิภาพของพร็อกซีแคชโดยพิจารณาจากค่าไบท์ฮิตเรโชทำได้โดยการเพิ่มความจุของพร็อกซีแคช

บทที่ 5

แนวโน้มความสัมพันธ์ระหว่างสภาพการใช้เว็บและประสิทธิภาพของขั้นตอนวิธีการแทนที่

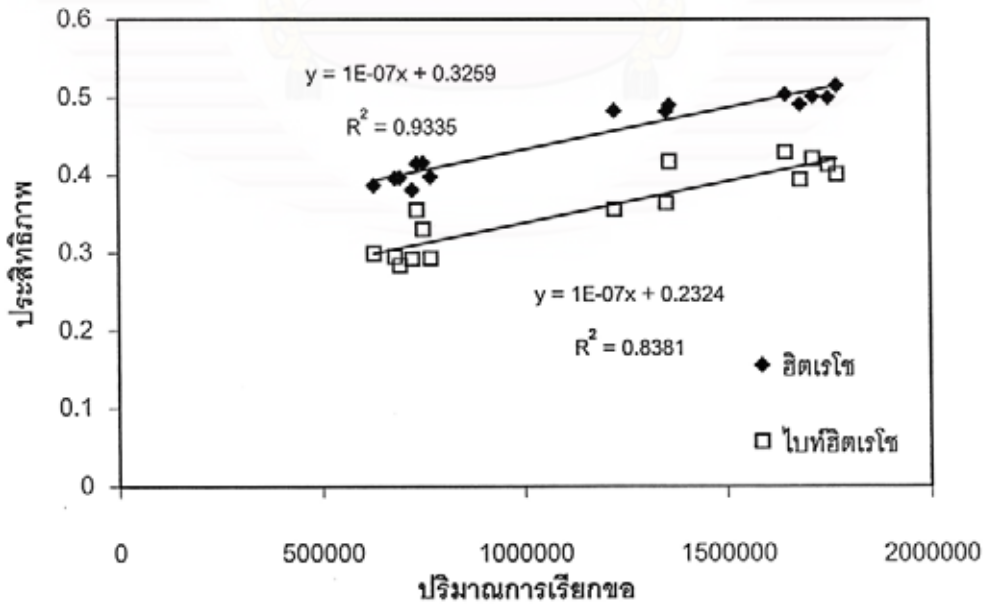
ส่วนหนึ่งของการพิจารณาเลือกใช้ขั้นตอนวิธีการแทนที่ที่เหมาะสมกับสภาพการใช้เว็บคือการศึกษาความสัมพันธ์ระหว่างสภาพการใช้เว็บและประสิทธิภาพของขั้นตอนวิธีการแทนที่ ซึ่งมีประโยชน์ในการประมาณประสิทธิภาพของพร็อกซีแคชจากข้อมูลของสภาพการใช้เว็บ การวิเคราะห์ใช้การเปรียบเทียบระหว่างสภาพการใช้เว็บที่หาได้จากแฟ้มที่บันทึกการเรียกขอ และผลการจำลองการทำงานของพร็อกซีแคชที่ได้ โดยเครื่องวัดที่ใช้ ได้แก่ ค่าฮิตเรโชและค่าไบท์ฮิตเรโช และสภาพการใช้เว็บที่ศึกษาได้แก่ ปริมาณการเรียกขอ ปริมาณการเรียกขอประเภท "Successful": สัดส่วนของจำนวนเอกสารที่แตกต่างกันกับจำนวนการเรียกขอทั้งหมด พารามิเตอร์จากกฎของ Zipf และช่วงเวลาการใช้เว็บ

การพิจารณาแนวโน้มความสัมพันธ์ระหว่างสภาพการใช้เว็บและประสิทธิภาพของขั้นตอนวิธีการแทนที่ใช้สมการเชิงเส้นแทนโดยค่าความชันของสมการเชิงเส้นแสดงสัดส่วนของการเปลี่ยนแปลงและค่า R^2 แสดงสัดส่วนของความแปรปรวนของข้อมูลที่สามารถใช้สมการเชิงเส้นในการแทนได้

ผลการวิเคราะห์แบ่งตามสภาพการใช้เว็บที่ศึกษาได้ 5 ส่วน ได้แก่

5.1. ปริมาณการเรียกขอ

ผลของการวิเคราะห์แนวโน้มความสัมพันธ์ระหว่างปริมาณการเรียกขอและประสิทธิภาพของขั้นตอนวิธีการแทนที่แบบ LRU โดยใช้ค่าฮิตเรโชและค่าไบท์ฮิตเรโชเป็นเครื่องวัด มีลักษณะดังแผนภูมิในรูปที่ 5.1

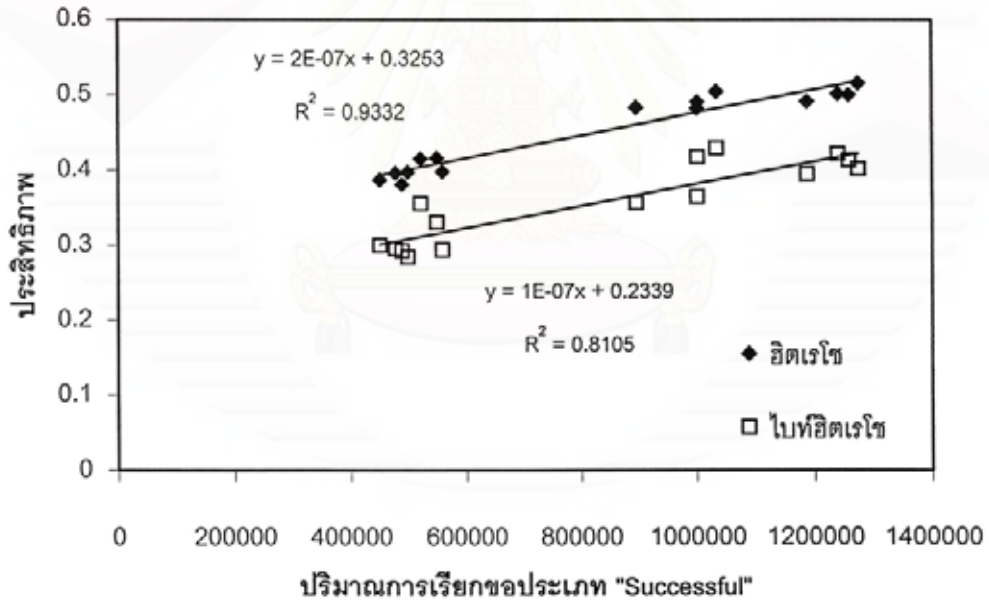


รูปที่ 5.1 ปริมาณการเรียกขอและประสิทธิภาพของขั้นตอนวิธีการแทนที่

ช่วงที่มีการใช้เว็บน้อยมีปริมาณการเรียกขออยู่ในช่วงประมาณ 630,000 ถึง 770,000 และช่วงที่มีการใช้เว็บมากมีปริมาณการเรียกขออยู่ในช่วงประมาณ 1,350,000 ถึง 1,770,000 จากแผนภูมิในรูปที่ 5.1 พบว่าข้อมูลแบ่งออกเป็น 2 ช่วงตามปริมาณการใช้เว็บโดยในช่วงที่มีการใช้เว็บน้อยมีค่าฮิตเรโชประมาณ 0.4 และค่าไบท์ฮิตเรโชประมาณ 0.3 ส่วนในช่วงที่มีการใช้เว็บมากมีค่าฮิตเรโชประมาณ 0.5 และค่าไบท์ฮิตเรโชประมาณ 0.4 ซึ่งสรุปได้ว่าประสิทธิภาพของขั้นตอนวิธีการแทนที่มีแนวโน้มเพิ่มขึ้นตามปริมาณการเรียกขอโดยแนวโน้มความสัมพันธ์ระหว่างปริมาณการเรียกขอและประสิทธิภาพของขั้นตอนวิธีการแทนที่มีลักษณะแปรผันตามกัน โดยมีสัดส่วนของค่าฮิตเรโชและค่าไบท์ฮิตเรโชต่อปริมาณการเรียกขอประมาณ 10^{-7} และค่า R^2 สำหรับสมการเชิงเส้นที่ใช้แทนแนวโน้มความสัมพันธ์ระหว่างปริมาณการเรียกขอและประสิทธิภาพของขั้นตอนวิธีการแทนที่มีค่า 0.9335 สำหรับการประมาณค่าฮิตเรโชและ 0.8381 สำหรับการประมาณค่าไบท์ฮิตเรโช

5.2. ปริมาณการเรียกขอประเภท "Successful"

ผลของการวิเคราะห์แนวโน้มความสัมพันธ์ระหว่างปริมาณการเรียกขอประเภท "Successful" และประสิทธิภาพของขั้นตอนวิธีการแทนที่แบบ LRU โดยใช้ค่าฮิตเรโชและค่าไบท์ฮิตเรโชเป็นเครื่องวัดมีลักษณะดังแผนภูมิในรูปที่ 5.2



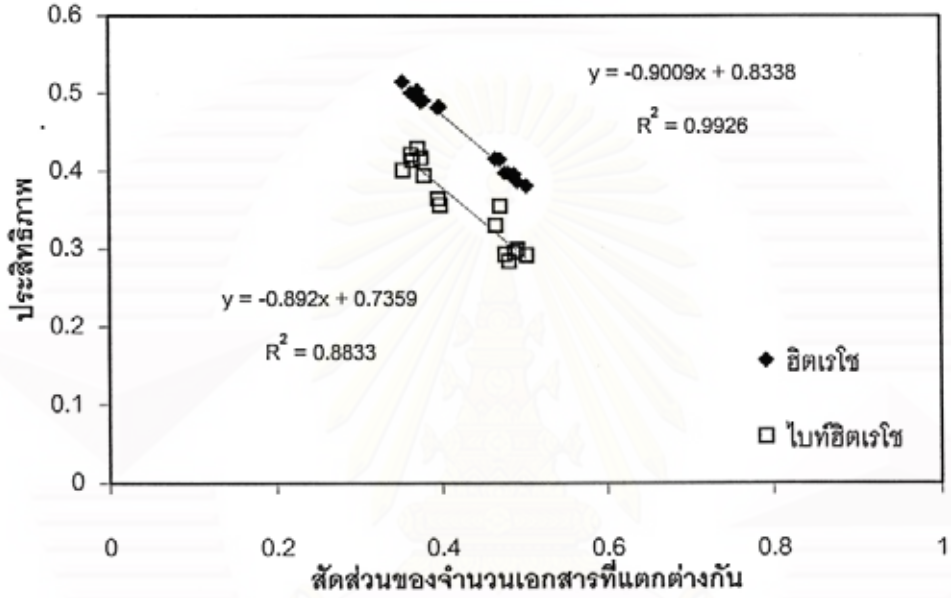
รูปที่ 5.2 ปริมาณการเรียกขอประเภท "Successful" และประสิทธิภาพของขั้นตอนวิธีการแทนที่

ปริมาณการเรียกขอประเภท "Successful" ในช่วงที่มีการใช้เว็บน้อยอยู่ในช่วงประมาณ 450,000 ถึง 560,000 และปริมาณการเรียกขอประเภท "Successful" ในช่วงที่มีการใช้เว็บมากอยู่ในช่วงประมาณ 900,000 ถึง 1,260,000 โดยแนวโน้มความสัมพันธ์ระหว่างปริมาณการเรียกขอประเภท "Successful" และประสิทธิภาพของขั้นตอนวิธีการแทนที่มีลักษณะแปรผันตามเช่นเดียวกับปริมาณการเรียกขอ สัดส่วนของ

ค่าฮิตเรโชและค่าไบท์ฮิตเรโชต่อปริมาณการเรียกขอประมาณ 10^{-7} และค่า R^2 สำหรับสมการเชิงเส้นที่ใช้แทนแนวโน้มความสัมพันธ์ระหว่างปริมาณการเรียกขอประเภท "Successful" และประสิทธิภาพของขั้นตอนวิธีการแทนที่มีค่า 0.9332 สำหรับการประมาณค่าฮิตเรโชและ 0.8105 สำหรับการประมาณค่าไบท์ฮิตเรโช

5.3. สัดส่วนของจำนวนเอกสารที่แตกต่างกันกับจำนวนการเรียกขอทั้งหมด

ผลของการวิเคราะห์แนวโน้มความสัมพันธ์ระหว่างสัดส่วนของจำนวนเอกสารที่แตกต่างกันกับจำนวนการเรียกขอทั้งหมด และประสิทธิภาพของขั้นตอนวิธีการแทนที่แบบ LRU โดยใช้ค่าฮิตเรโชและค่าไบท์ฮิตเรโชเป็นเครื่องวัด มีลักษณะดังแผนภูมิในรูปที่ 5.3



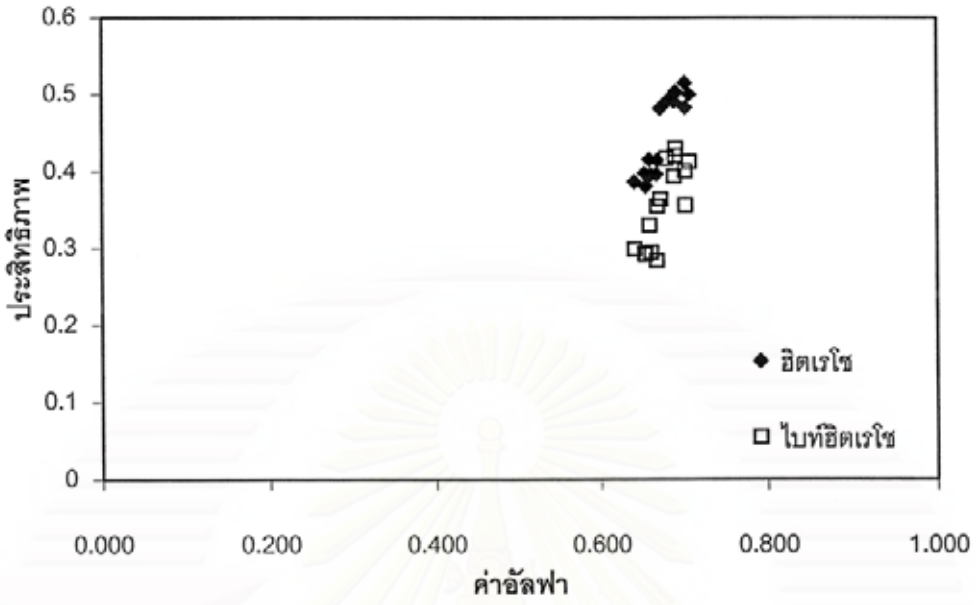
รูปที่ 5.3 สัดส่วนของจำนวนเอกสารที่แตกต่างกันและประสิทธิภาพของขั้นตอนวิธีการแทนที่

ค่าฮิตเรโชและไบท์ฮิตเรโชที่ได้จากการทดสอบข้อมูลการใช้เว็บที่มีสัดส่วนของจำนวนเอกสารที่แตกต่างกันเมื่อเทียบกันปริมาณการเรียกขอประเภท "Successful" ในที่มีการใช้เว็บน้อยและช่วงที่มีการใช้เว็บมากแสดงในแผนภูมิในรูปที่ 5.3 โดยสัดส่วนของจำนวนเอกสารที่แตกต่างกันสำหรับข้อมูลการใช้เว็บช่วงที่มีการใช้เว็บมากอยู่ในช่วง 0.3 ถึง 0.4 และสัดส่วนของจำนวนเอกสารที่แตกต่างกันสำหรับข้อมูลการใช้เว็บช่วงที่มีการใช้เว็บน้อยอยู่ในช่วง 0.4 ถึง 0.5 และแนวโน้มความสัมพันธ์ระหว่างสัดส่วนของจำนวนเอกสารที่แตกต่างกันกับจำนวนการเรียกขอทั้งหมดและประสิทธิภาพของขั้นตอนวิธีการแทนที่มีลักษณะแปรผกผัน โดยสัดส่วนของค่าฮิตเรโชและค่าไบท์ฮิตเรโชที่ลดลงต่อสัดส่วนของจำนวนเอกสารที่แตกต่างกัน มีค่าประมาณ 0.9 และค่า R^2 สำหรับสมการเชิงเส้นที่ใช้แทนแนวโน้มความสัมพันธ์ระหว่างสัดส่วนของจำนวนเอกสารที่แตกต่างกันและประสิทธิภาพของขั้นตอนวิธีการแทนที่มีค่า 0.9926 สำหรับการประมาณค่าฮิตเรโชและ 0.8833 สำหรับการประมาณค่าไบท์ฮิตเรโช

5.4. พารามิเตอร์จากกฎของ Zipf

ผลการวิเคราะห์แนวโน้มความสัมพันธ์ระหว่างพารามิเตอร์จากกฎของ Zipf แบ่งออกเป็น 2 ส่วน ได้แก่

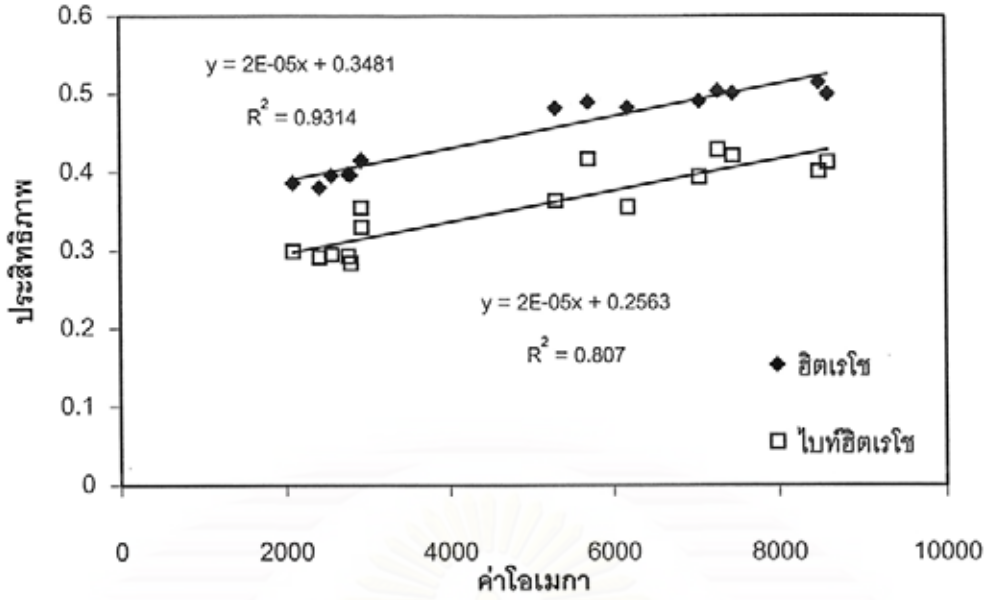
5.4.1 แนวโน้มความสัมพันธ์ระหว่างค่าอัลฟา และประสิทธิภาพของขั้นตอนวิธีการแทนที่



รูปที่ 5.4 ค่าอัลฟา และประสิทธิภาพของขั้นตอนวิธีการแทนที่

ผลการวิเคราะห์ความสัมพันธ์ระหว่างค่าอัลฟาและประสิทธิภาพของขั้นตอนวิธีการแทนที่ แสดงให้เห็นว่าค่าอัลฟาของสภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัยมีค่าเฉลี่ยเท่ากับ 0.68 โดยค่าอัลฟาที่ได้จากการทดสอบกับข้อมูลการใช้เว็บแต่ละช่วงเวลานี้อยู่ในช่วง 0.65 ถึง 0.71 จากผลการทดสอบที่ได้แสดงให้เห็นว่าประสิทธิภาพของขั้นตอนวิธีการแทนที่มีแนวโน้มเพิ่มขึ้นเมื่อค่าอัลฟามากขึ้นเนื่องจากงานวิจัยที่ผ่านมา [13] ได้กล่าวถึงค่าอัลฟาว่าเป็นค่าเฉพาะสำหรับแต่ละชุดข้อมูล จึงไม่ได้ใช้เส้นแนวโน้มแทนความสัมพันธ์ระหว่างค่าอัลฟาและประสิทธิภาพของขั้นตอนวิธีการแทนที่

5.4.2 แนวโน้มความสัมพันธ์ระหว่างค่าโอเมกา และประสิทธิภาพของขั้นตอนวิธีการแทนที่



รูปที่ 5.5 ค่าโอเมกา และประสิทธิภาพของขั้นตอนวิธีการแทนที่

ค่าโอเมกาของข้อมูลการใช้เว็บที่ใช้ในการทดสอบมีค่าอยู่ในช่วงประมาณ 2,000 ถึง 3,000 สำหรับช่วงที่มีการใช้เว็บน้อย และ ช่วงประมาณ 5,000 ถึง 9,000 สำหรับช่วงที่มีการใช้เว็บมาก ผลการวิเคราะห์ความสัมพันธ์ระหว่างค่าโอเมกาและประสิทธิภาพของขั้นตอนวิธีการแทนที่ แสดงให้เห็นว่าลักษณะของการกระจายของจุดข้อมูลของแผนภูมิในรูปที่ 5.5 แบ่งออกเป็น 2 กลุ่มเช่นเดียวกับการวิเคราะห์ความสัมพันธ์ระหว่างปริมาณการเรียกขอและประสิทธิภาพของขั้นตอนวิธีการแทนที่ ดังนั้นค่าโอเมกาที่ได้จากข้อมูลการใช้เว็บน่าจะมีผลมาจากปริมาณการเรียกขอของข้อมูลการใช้เว็บด้วย และประสิทธิภาพของขั้นตอนวิธีการแทนที่มีลักษณะแปรผันตามค่าโอเมกา โดยสัดส่วนของค่าฮิตเรโช และค่าไบท์ฮิตเรโชกับค่า โอเมกา ประมาณ 2×10^{-5} และค่า R^2 สำหรับสมการเชิงเส้นที่ใช้แทนแนวโน้มความสัมพันธ์ระหว่างค่าโอเมกาและประสิทธิภาพของขั้นตอนวิธีการแทนที่มีค่า 0.9314 สำหรับการประมาณค่าฮิตเรโชและ 0.807 สำหรับการประมาณค่าไบท์ฮิตเรโช

5.5. ช่วงเวลาการใช้เว็บ

	ช่วงที่มีการใช้เว็บมาก	ช่วงที่มีการใช้เว็บน้อย
ปริมาณการเรียกขอ	1563677	711908
ปริมาณการเรียกขอประเภท "Successful"	1112368	506529
สัดส่วนของจำนวนเอกสารที่แตกต่างกัน	0.376	0.483
ค่าอัลฟาจากกฎของ zipf	0.693	0.659
ค่าโอเมกาจากกฎของ zipf	7012	2647
ค่าฮิตเรโช	0.496	0.399
ค่าไบท์ฮิตเรโช	0.399	0.307

ตารางที่ 5.1 ค่าเฉลี่ยของข้อมูลการใช้เว็บช่วงที่มีการใช้เว็บมากและน้อย

สัดส่วนของปริมาณการเรียกขอช่วงที่มีการใช้เว็บมากต่อช่วงที่มีการใช้เว็บน้อยเท่ากับ 0.455 โดยสัดส่วนของปริมาณการเรียกขอประเภท "Successful" ต่อปริมาณการเรียกขอเท่ากับ 0.711 ในช่วงที่มีการใช้เว็บมากและ 0.712 ในช่วงที่มีการใช้เว็บน้อย ซึ่งมีค่าใกล้เคียงกันมากจึงสรุปได้ว่า สัดส่วนของปริมาณการเรียกขอประเภท "Successful" ต่อปริมาณการเรียกขอในช่วงที่มีการใช้เว็บมากและน้อยไม่แตกต่างกัน

สัดส่วนของจำนวนเอกสารที่แตกต่างกันแสดงถึงขอบเขตของประสิทธิภาพสูงสุดของขั้นตอนวิธีการแทนที่ที่สามารถปรับปรุงได้โดยสัดส่วนของเอกสารที่แตกต่างกันในช่วงที่มีการใช้เว็บมากเท่ากับ 0.376 หรือค่าฮิตเรโชสูงสุดที่เป็นไปได้ คือ 1-0.376 หรือ 0.624 และสัดส่วนของเอกสารที่แตกต่างกันในช่วงที่มีการใช้เว็บน้อยเท่ากับ 0.483 หรือค่าฮิตเรโชสูงสุดที่เป็นไปได้ คือ 1-0.483 หรือ 0.517 ซึ่งค่าฮิตเรโชเฉลี่ยสำหรับพร็อกซีแคชของจุฬาลงกรณ์มหาวิทยาลัยเท่ากับ 0.458 สำหรับช่วงที่มีการใช้เว็บมากและ 0.405 สำหรับช่วงที่มีการใช้เว็บน้อย ดังนั้นการปรับปรุงประสิทธิภาพของพร็อกซีแคชจึงสามารถทำได้โดยการเลือกใช้ขั้นตอนวิธีการแทนที่ที่เหมาะสม

ค่าไบท์ฮิตเรโชและค่าฮิตเรโชที่ได้จากการจำลองการทำงานของพร็อกซีแคชในช่วงที่มีการใช้เว็บมากและช่วงที่มีการใช้เว็บน้อยมีค่าต่างกันประมาณ 0.1 ซึ่งสรุปได้ว่าประสิทธิภาพของพร็อกซีแคชในช่วงที่มีการใช้เว็บมากสูงกว่าในช่วงที่มีการใช้เว็บน้อย

บทที่ 6

สรุปผลการวิจัยและข้อเสนอแนะ

จากผลของการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่แบบต่างๆ โดยจำลองการทำงานของพร็อกซีแคชซึ่งอาศัยข้อมูลการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัยและใช้ข้อมูลการใช้เว็บจากพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัยกับข้อมูลการใช้เว็บจากเอ็นแอลเอเอ็นอาร์ในการเปรียบเทียบ รวมทั้งศึกษาแนวโน้มความสัมพันธ์ระหว่างสภาพการใช้เว็บและประสิทธิภาพของขั้นตอนวิธีการแทนที่ในพร็อกซีแคช สามารถสรุปผลที่ได้รับจากการวิจัย ปัญหาข้อจำกัดที่พบ และเสนอแนะแนวทางเพื่อการพัฒนาระบบต่อไป ได้ดังนี้

สรุปผลการวิจัย

1. ประสิทธิภาพของขั้นตอนวิธีการแทนที่แต่ละแบบในช่วงที่มีการใช้เว็บมากมีค่าสูงกว่าช่วงที่มีการใช้เว็บน้อย โดยการเลือกใช้ขั้นตอนวิธีการแทนที่ให้เหมาะสมกับสภาพการใช้เว็บควรพิจารณาจากความจุของดิสก์ที่ใช้เป็นพร็อกซีแคชและประเภทของเครื่องวัดที่ต้องการปรับปรุง

2. ผลการทดสอบไม่พบขั้นตอนวิธีการแทนที่ใดที่ให้ค่าฮิตเรโช และค่าไบท์ฮิตเรโช สูงสุดพร้อมกันที่ความจุของพร็อกซีแคชแต่ละขนาดที่ใช้ในการจำลองการทำงานของพร็อกซีแคช ดังนั้นในการเลือกใช้ขั้นตอนวิธีการแทนที่ให้เหมาะสมจำเป็นที่จะต้องเลือกพิจารณาค่าเครื่องวัดที่คาดว่ามีผลต่อประสิทธิภาพของพร็อกซีแคชมากที่สุด สำหรับสภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัยจำเป็นที่จะต้องพิจารณาปรับปรุงค่าไบท์ฮิตเรโช เนื่องจากการปรับปรุงประสิทธิภาพโดยลดค่าไบท์ฮิตเรโชส่งผลโดยตรงต่อปริมาณข้อมูลที่รับมาจากภายนอกมหาวิทยาลัยซึ่งในปัจจุบันแบนด์วิดท์ของการเชื่อมต่อไปยังต่างประเทศที่มีอยู่ไม่เพียงพอต่อการใช้งานในมหาวิทยาลัย

3. ถ้าพิจารณาจากค่าฮิตเรโช จะได้ว่าขั้นตอนวิธีการแทนที่ที่เหมาะสมกับพร็อกซีแคชที่มีความจุน้อย (ประมาณ 1 – 10 เปอร์เซ็นต์ของขนาดรวมของเอกสารที่เรียกขอ) คือ ขั้นตอนวิธีการแทนที่แบบ LRU-MIN และขั้นตอนวิธีการแทนที่ที่เหมาะสมกับพร็อกซีแคชที่มีความจุมาก (มากกว่า 20 เปอร์เซ็นต์ของขนาดรวมของเอกสารที่เรียกขอ) คือ ขั้นตอนวิธีการแทนที่แบบ SIZE แต่ถ้าพิจารณาจากค่าไบท์ฮิตเรโช พบว่าขั้นตอนวิธีการแทนที่แบบ LRU-TH ที่มีค่าขีดแบ่ง 1 เมกกะไบต์เหมาะสมกับพร็อกซีแคชที่มีความจุน้อย และขั้นตอนวิธีการแทนที่แบบ LRU-MIN เหมาะสมกับพร็อกซีแคชที่มีความจุมาก แสดงได้ดังตารางที่ 6.1

เครื่องวัด	ความจุ	
	น้อย	มาก
ค่าฮิตเรโช	LRU-MIN	SIZE
ค่าไบท์ฮิตเรโช	LRU-TH, Perfect-LFU	LRU-MIN

ตารางที่ 6.1 ขั้นตอนวิธีการแทนที่ที่เหมาะสมกับพร็อกซีแคชที่มีความจุและเครื่องวัดที่กำหนด

ปัญหาและข้อจำกัดที่ได้พบจากการวิจัย

1. การจำลองการทำงานของพร็อกซีแคชจำเป็นต้องใช้หน่วยความจำในการเก็บข้อมูลที่มีอยู่ในพร็อกซีแคชและข้อมูลทางสถิติของข้อมูลที่อยู่ในแคชเพื่อความรวดเร็วในการประมวลผล โดยหน่วยความจำที่จำเป็นต้องใช้ในการจำลองการทำงานของพร็อกซีแคชแปรผันตรงตามขนาดของพร็อกซีแคช ดังนั้นจึงจำเป็นที่จะต้องใช้หน่วยความจำที่มากพอในการจำลองการทำงานของพร็อกซีแคช เช่น ข้อมูลการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัย 1 วันจำนวนประมาณ 1.5 ล้านการเรียกขอจำเป็นที่จะต้องใช้หน่วยความจำประมาณ 128 เมกกะไบต์ในการเก็บข้อมูลการเรียกขอและข้อมูลทางสถิติที่ใช้ในการแทนที่

2. ข้อมูลการใช้เว็บที่นำมาจากแคชเอ็นจิ้นของจุฬาลงกรณ์มหาวิทยาลัยบางช่วงเวลาไม่ได้บันทึกรหัสการตอบกลับจากเว็บเซิร์ฟเวอร์ ทำให้การจำลองการทำงานของพร็อกซีแคชมีความผิดพลาดเกิดขึ้น

3. การบันทึกการเรียกขอของแคชเอ็นจิ้นหรือพร็อกซีแคชมีข้อมูลที่จำกัดดังรูปแบบในภาคผนวก ก ซึ่งข้อมูลเหล่านี้ไม่เพียงพอที่จะใช้ในการจำลองการทำงานของพร็อกซีแคชเพื่อวิเคราะห์และปรับปรุงเวลาตอบสนองซึ่งมีผลต่อผู้ใช้เว็บโดยตรงได้

ข้อเสนอแนะ

1. การเลือกใช้ขั้นตอนวิธีการแทนที่โดยการใช้ค่าฮิตเรโช และค่าไบท์ฮิตเรโชเป็นเครื่องวัดประสิทธิภาพของพร็อกซีแคชอาจไม่สามารถลดเวลาตอบสนองทางด้านผู้ใช้เว็บได้เนื่องจากเครื่องวัดทั้งสองค่านี้ไม่สามารถบอกถึงภาระที่เกิดบนพร็อกซีเซิร์ฟเวอร์และเวลาที่ใช้ในการส่งข้อมูลไปยังผู้ใช้เว็บที่เปลี่ยนไปจากเดิมได้ ดังนั้นเพื่อให้การเลือกใช้ขั้นตอนวิธีการแทนที่เกิดประโยชน์โดยตรงต่อผู้ใช้เว็บจึงควรจะพิจารณาขั้นตอนวิธีการแทนที่ที่พิจารณาเวลาที่ใช้ในการส่งข้อมูลไปยังผู้ใช้เว็บด้วย

2. การจำลองการทำงานของพร็อกซีแคชนี้ไม่ได้พิจารณาเอกสารประเภทไดนามิกเพจ (dynamic page) ซึ่งสภาพการใช้เว็บในปัจจุบันมีแนวโน้มของปริมาณการใช้งานที่สูงขึ้น และเอกสารประเภทนี้ไม่สามารถจัดเก็บโดยพร็อกซีแคชได้ ดังนั้นการปรับปรุงประสิทธิภาพของพร็อกซีแคชเพียงที่เดียวอาจไม่ช่วยให้เวลาตอบสนองลดลงได้ถ้าสัดส่วนของไดนามิกเพจมากกว่าสแตติกเพจ (static page)

รายการอ้างอิง

[1] Abrams, M. , Standridge , C. R. , Abdulla , G. , Williams , S. , and Fox , E. A. Caching Proxies:Limitations and Potentials. Proceedings of the 4th International WWW Conference (December 1995).

[2] Cao, P. , and Irani , S. Cost-Aware WWW Proxy Caching Algorithms. Proceedings of the 1997 USENIX Symposium on Internet Technologies and Systems (USITS-97) (December 1997).

[3] Glassman, S. A caching relay for the world wide web. First International Conference on the World Wide Web (May 1994).

[4] Cunha, C. , Bestavros, A., and Crovella, M. Characteristics of WWW client-based traces. Technical Report TR-95-010, Boston University, Computer Science Dept., Boston, MA 02215, USA (April 1995).

[5] Nishikawa, N., Hosokawa, T., Mori Y., Yoshidab, K., and Tsujia, H. Memory-based architecture for distributed WWW caching proxy. 7th WWW Conference (April 1998).

[6] Barford, P., Bestavros, A., Bradley, A., and Crovella, M. Changes in Web Client Access Patterns: Characteristics and Caching Implications. Special Issue on Characterization and Performance Evaluation.

[7] Breslau, L., Cao, P., Fan, L., Phillips, G., and Shenker, S. Web caching and zipf-like distributions: Evidence and Implications. Technical report, University of Wisconsin Madison, Department of Computer Science, 1210 West Dayton Street (July 1998).

[8] Shi, Y., Watson, E., and Chen, Y. Model-driven simulation of world-wide-web cache policies In Proceedings of the 1997 Winter Simulation Conference.

[9] Bharat, K., and, Broder, A. Measuring the Web. In the 7th International World Wide Web Conference

[10] Lorenzetti, P., Rizzo, L., and Vicisano, L. Replacement policies for a proxy cache.<http://www.iet.unipi.it/luigi/research.html>

[11] Duska, B. M., Marwood, D., and Feeley, M. J. The measured access characteristics of world-wide-web client proxy caches. In Proceedings of the USENIX Symposium on Internet Technologies and Systems (December 1997).

[12] Arlitt, M. F., and Williamson, C. L. Internet web servers : Workload characterization and Performance Implications. IEEE/ACM Transactions on Networking Vol. 5, No. 5 (October 1997).

[13] Breslau, L., Cao, P., Fan, L., Phillips, G. and Shenker, S. On the implications of Zipf's Law for web caching. Technical report, University of Wisconsin Madison, Department of Computer Science, 1210 West Dayton Street (April 1998).

[14] Pitkow, J. E., and Recker, M. M. A simple yet robust caching algorithm based on dynamic access patterns. In the 2th International World Wide Web Conference (October 1994)

[15] Williams, S., Abrams, M., Standridge, C. R., Abdulla, G., and Fox, E. A. Removal policies in network caches for World-Wide Web documents. Proceedings of Sigcomm'96 (August 1996).

[16] ภาณุพันธ์ สุวรรณมาตร์. การวิเคราะห์สถิติการใช้อินเทอร์เน็ตในระบบเครือข่ายระดับสถาบันอุดมศึกษา. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2541

[17] Hines, W. W., Montgomery, D. C. Probability and Statistics in Engineering and Management Science. John Wiley & Sons.

[18] Jain, R. The art of computer systems performance analysis. John Wiley & Sons.

ภาคผนวก

ภาคผนวก ก

รูปแบบของข้อมูลการเรียกขอที่พร็อกซีแคชบันทึกไว้

time elapsed remotehost code/status bytes method URL rfc931 peerstatus/peerhost

time	เวลาที่บันทึกการเรียกขอ
elasped	เวลาที่ใช้ในการรับข้อมูลจากเว็บเซิร์ฟเวอร์มีหน่วยเป็นมิลลิวินาที
remotehost	เลขไอพีของเว็บเซิร์ฟเวอร์
code/status	รหัสการตอบกลับจากเว็บเซิร์ฟเวอร์
bytes	ปริมาณข้อมูลที่รับมาจากเว็บเซิร์ฟเวอร์มีหน่วยเป็นไบต์
method	วิธีการเรียกขอข้อมูล
URL	ที่อยู่ของข้อมูลที่เรียกขอ
rfc931	ชื่อของผู้เรียกขอแต่โดยปกติจะไม่มีการเก็บข้อมูลส่วนนี้ไว้
peerstatus/peerhost	วิธีการส่งต่อการเรียกขอ/ไอพีปลายทางที่การเรียกขอถูกส่งต่อไป

ตัวอย่างการเรียกขอ

939519239.071 131 161.200.88.211 TCP_MISS/200 6092 GET
http://www.chula.ac.th/home/images/prakaew.gif - DIRECT/www.chula.ac.th image/gif

ความหมายของรหัสการตอบกลับ (response code)

1XX - Informational - แสดงสถานะของเว็บเซิร์ฟเวอร์ว่าพร้อมที่จะตอบรับการเรียกขอที่มีรูปแบบตามเวอร์ชันที่ไคลเอนต์ได้แสดงความต้องการมา

2XX - Successful - การเรียกขอมายังเว็บเซิร์ฟเวอร์ได้รับและพร้อมที่จะส่งข้อมูลกลับตามการเรียกขอ

3XX - Redirection - การตอบกลับการเรียกขอเพื่อแสดงสถานะของเอกสารจากการเรียกขอ

4XX - Client Error - ข้อมูลที่ไคลเอ็นต์ส่งมาไม่ครบหรือเกิดความผิดพลาด

5XX - Server Error - มีความผิดพลาดเกิดที่เซิร์ฟเวอร์ขณะกำลังประมวลผลการเรียกขอ

ภาคผนวก ข

โปรแกรมที่พัฒนาขึ้นเพื่อใช้ในการจำลองการทำงานของพร็อกซีแคช

รูปแบบของคำสั่งการเรียกใช้โปรแกรม

perl <ชื่อโปรแกรม> <ชื่อแฟ้มข้อมูลการเรียกขอ> <ชื่อแฟ้มข้อมูลที่ใช้บันทึกผลลัพธ์>

ตัวอย่างแสดงการใช้โปรแกรมในการจำลองการทำงานของพร็อกซีแคช

$ perl lru_min.pl cu-11-peak.cut cu-11-peak.min

แฟ้มข้อมูลการเรียกขอ

มีรูปแบบการบันทำการเรียกขอตามรูปแบบในภาคผนวก ก

แฟ้มข้อมูลที่ใช้บันทึกผลลัพธ์

การบันทึกผลการจำลองการทำงานของพร็อกซีแคชแบ่งออกเป็น 2 ส่วนดังนี้

ข้อมูลของแฟ้มที่บันทึกการเรียกขอ ได้แก่

ก. "Total Request" -ปริมาณการเรียกขอทั้งหมด

ข. "Unique Request"-ปริมาณเอกสารที่แตกต่างกันจากการเรียกขอทั้งหมด

ค. "Total Byte" -ขนาดของเอกสารรวมจากการเรียกขอทั้งหมดมีหน่วยเป็นไบต์

ง. "Unique Byte" -ขนาดของเอกสารรวมจากเอกสารที่แตกต่างกันจากการเรียกขอทั้งหมด

จ. "____ Algorithm"-ขั้นตอนวิธีการแทนที่ที่ใช้ในการจำลองการทำงานของพร็อกซีแคช

2 ผลการจำลองการทำงานของพร็อกซีแคช มีรูปแบบดังนี้

ก. ขนาดของพร็อกซีแคช(เปอร์เซ็นต์ของ “Unique Byte ”)

ข. ขนาดของพร็อกซีแคช(ไบท์)

ค. ปริมาณของการเรียกขอที่พบในแคช(ครั้ง)

ง. ปริมาณของการเรียกขอที่ไม่พบในแคช(ครั้ง)

จ. ค่าฮิตเรโช

ฉ. ปริมาณของข้อมูลที่พบในแคช(ไบท์)

ช. ปริมาณของข้อมูลที่ไม่พบในแคช(ไบท์)

ซ. ค่าไบท์ฮิตเรโช

ตัวอย่างของผลลัพธ์ที่ได้มีลักษณะดังรูปที่ ข.1

```
Total Request  : 1240852
Unique Request : 452043
Total Byte     : 9757468531
Unique Byte    : 4267472109
FIFO Algorithm
1 42674721 300715 940137 0.242 1773466380 7984002151 0.182
2 85349442 383522 857330 0.309 2266335821 7491132710 0.232
5 213373605 495470 745382 0.399 3056906503 6700562028 0.313
10 426747210 576816 664036 0.465 3803079594 5954388937 0.390
```

รูปที่ ข.1 ตัวอย่างผลลัพธ์ที่ได้จากโปรแกรมจำลองการทำงานของพร็อกซีแคช

ภาคผนวก ค

1 ผลการจำลองการทำงานของพร็อกซีแคชเพื่อหาค่าขีดแบ่งที่เหมาะสมสำหรับขั้นตอนวิธีการแทนที่ LRU_TH

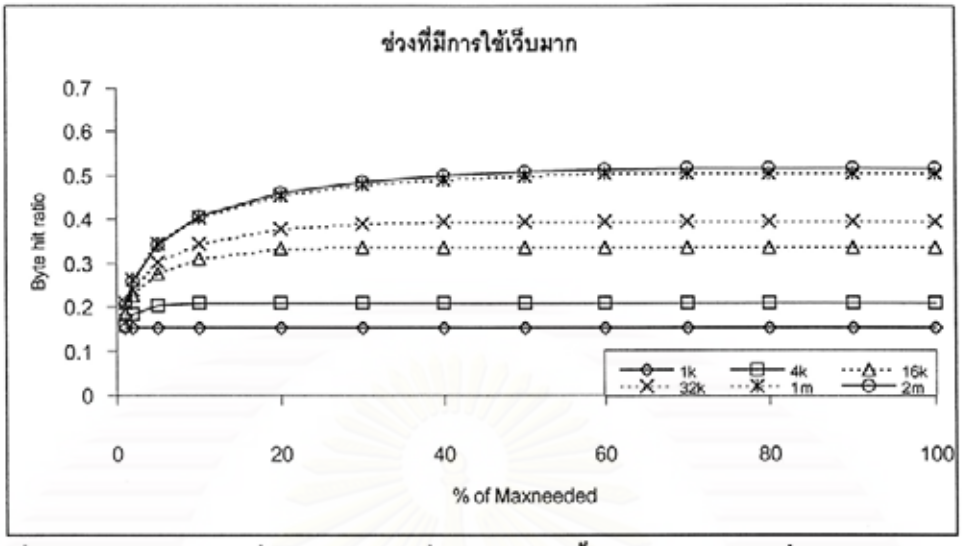
การทดสอบเปรียบเทียบระหว่างขั้นตอนวิธีการแทนที่แบบ LRU-TH กับขั้นตอนวิธีการแทนที่แบบอื่นๆ และการพิจารณาเลือกใช้ขั้นตอนวิธีการแทนที่แบบ LRU-TH กับพร็อกซีแคชจำเป็นที่จะต้องจำลองการทำงานของพร็อกซีแคชเพื่อหาค่าขีดแบ่งที่เหมาะสมเพื่อให้ขั้นตอนวิธีการแทนที่แบบ LRU-TH ให้ประสิทธิภาพสูงสุดโดยพิจารณาจากค่าฮิตเรโชและไบท์ฮิตเรโชที่ได้

การทดสอบเพื่อหาค่าขีดที่เหมาะสมสำหรับขั้นตอนวิธีการแทนที่แบบ LRU-TH สำหรับสภาพการใช้เว็บแต่ละแห่งแบ่งออกเป็น 4 การทดสอบ ได้แก่ การทดสอบกับข้อมูลการใช้เว็บช่วงที่มีการใช้เว็บมากโดยมีค่าฮิตเรโชและค่าไบท์ฮิตเรโชเป็นเครื่องวัด และการทดสอบกับข้อมูลการใช้เว็บช่วงที่มีการใช้เว็บน้อยโดยมีค่าฮิตเรโชและค่าไบท์ฮิตเรโชเป็นเครื่องวัด แต่ละการทดสอบจะมีจำนวนการจำลองการทำงานของพร็อกซีแคชตามจำนวนของชุดข้อมูลที่มีอยู่ เช่น ข้อมูลการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย มีข้อมูลการใช้เว็บในช่วงที่มีการใช้เว็บมาก 8 ชุด ดังนั้นการจำลองการทำงานของพร็อกซีแคชในช่วงที่มีการใช้เว็บมากมีทั้งหมด 8 ชุด และผลการทดสอบสำหรับข้อมูลการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย ในช่วงที่มีการใช้เว็บมาก คือค่าเฉลี่ยของค่าฮิตเรโชและค่าไบท์ฮิตเรโชที่ได้จากการจำลองการทำงานของพร็อกซีแคชทั้งหมด 8 ชุด เป็นต้น

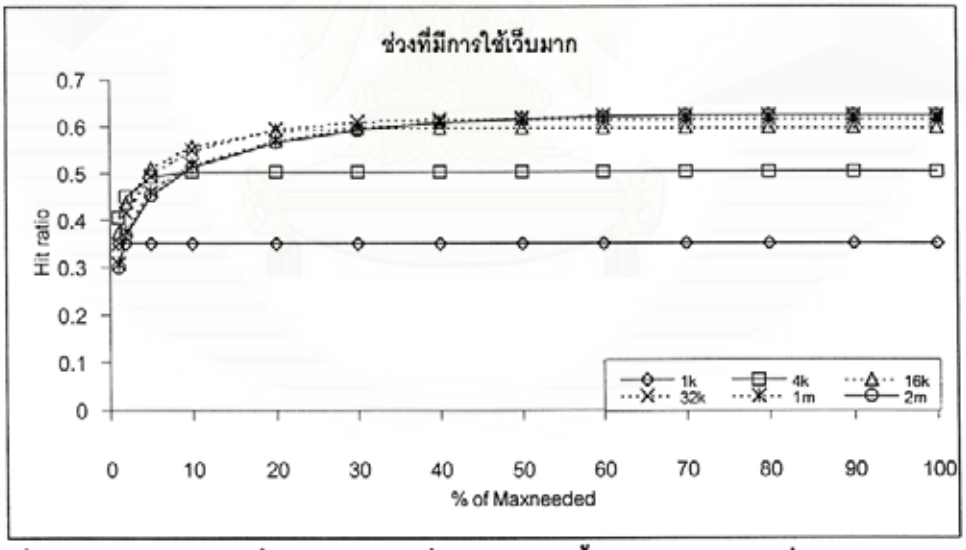
ผลการทดสอบแสดงให้เห็นว่าค่าขีดแบ่งที่เหมาะสมของขั้นตอนวิธีการแทนที่แบบ LRU-TH สำหรับสภาพการใช้เว็บที่จุฬาลงกรณ์มหาวิทยาลัยและเอ็นแอลเอเอ็นอาร์ทั้งช่วงที่มีการใช้เว็บมากและช่วงที่มีการใช้เว็บน้อย คือ 1 เมกกะไบต์ เนื่องจากประสิทธิภาพของขั้นตอนวิธีการแทนที่ที่ค่าขีดแบ่ง 1 เมกกะไบต์ใกล้เคียงกับค่าขีดแบ่ง 2 เมกกะไบต์มาก

การทดสอบแบ่งตามข้อมูลที่ใช้ในการทดสอบได้ผลดังนี้

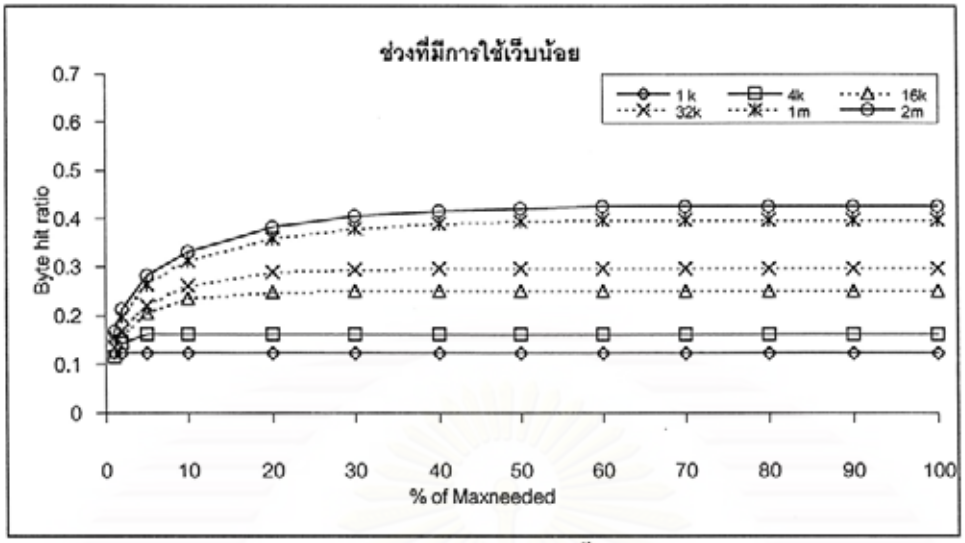
1.1 ข้อมูลการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศจุฬาลงกรณ์มหาวิทยาลัย



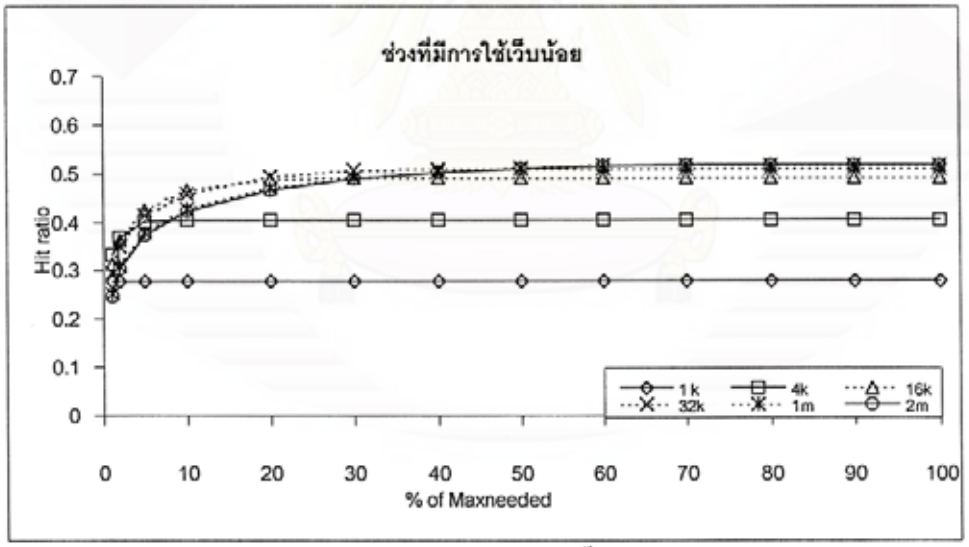
รูปที่ ค.1 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศของจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าไบท์ฮิตเรโช



รูปที่ ค.2 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศของจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าฮิตเรโช

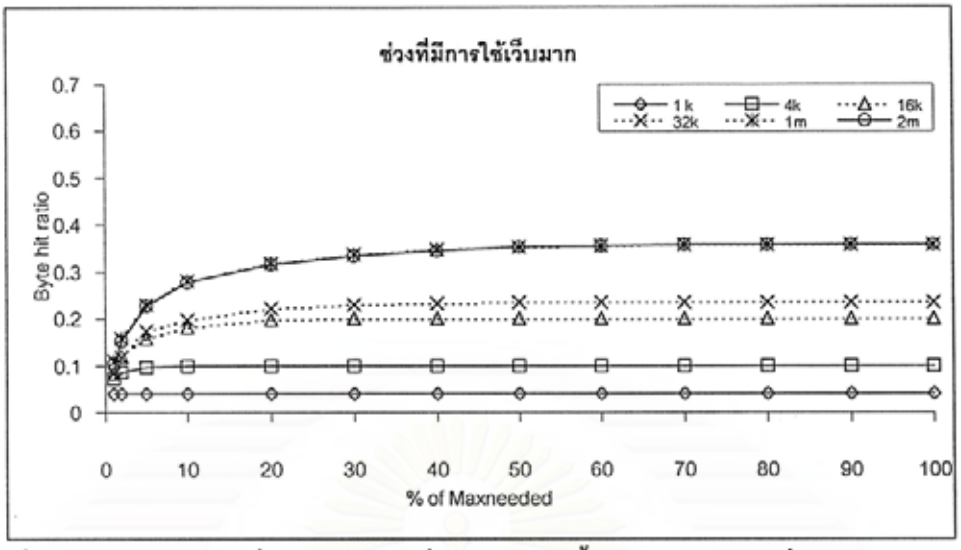


รูปที่ ค.3 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศของจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าไบท์ฮิตเรโช

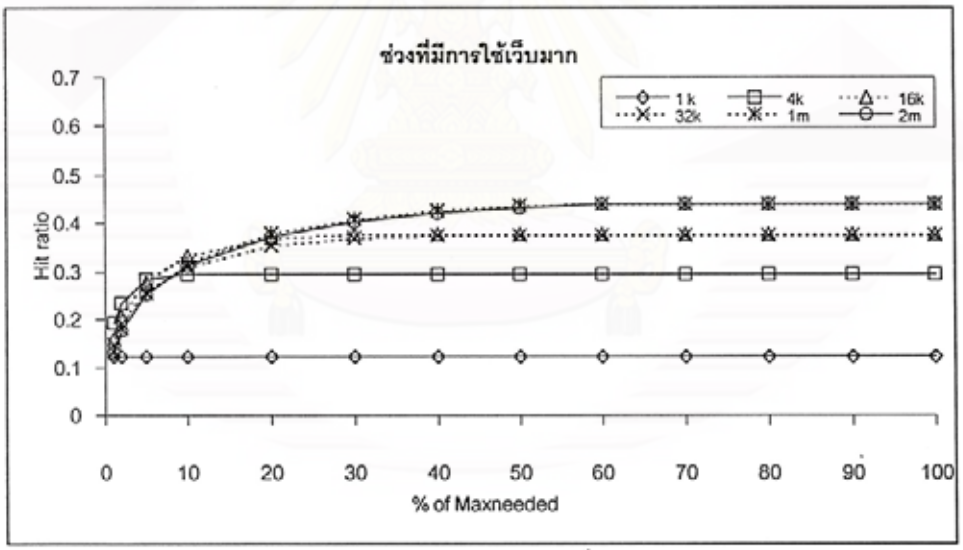


รูปที่ ค.4 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศของจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าฮิตเรโช

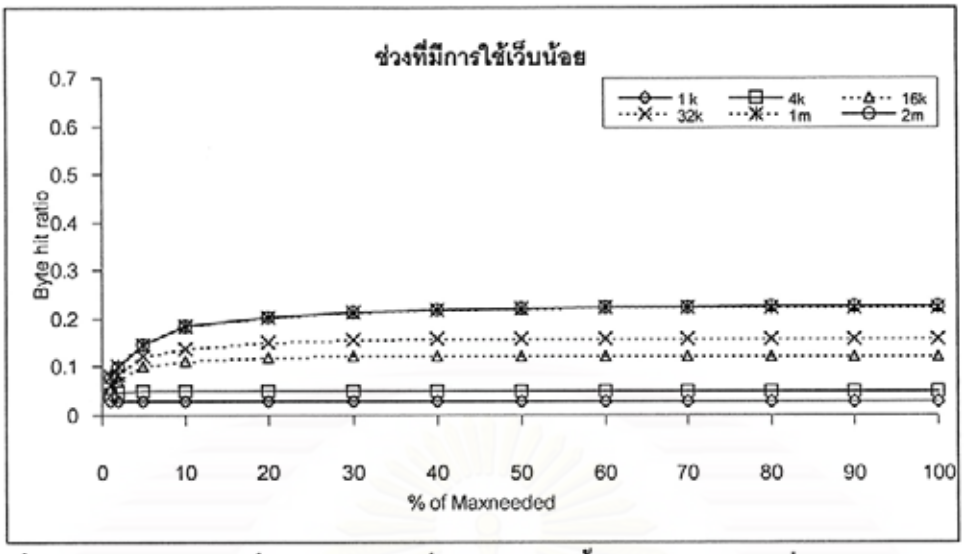
1.2 ข้อมูลการใช้เว็บจากพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย



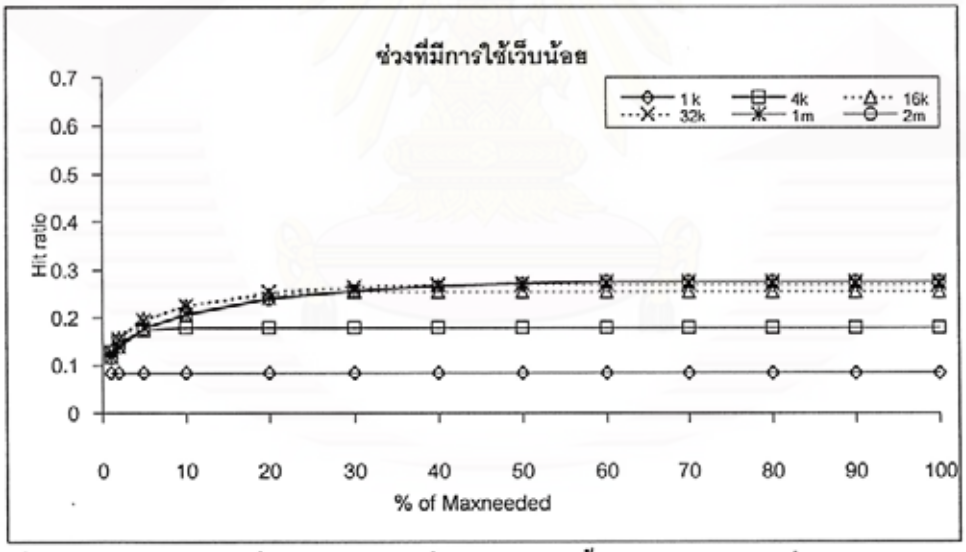
รูปที่ ค.5 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย โดยพิจารณาจากค่าไบท์ฮิตเรโช



รูปที่ ค.6 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย โดยพิจารณาจากค่าฮิตเรโช

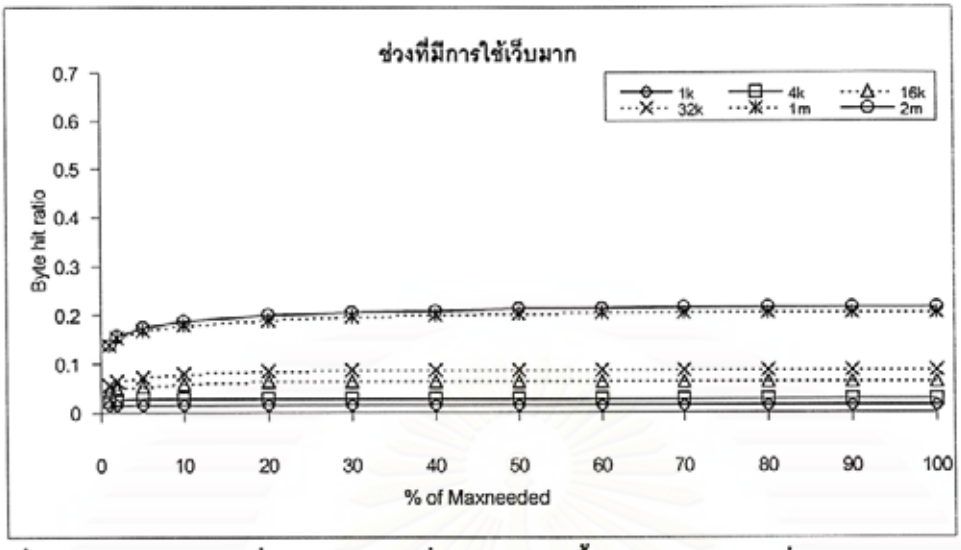


รูปที่ ค.7 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย โดยพิจารณาจากค่าไบท์ฮิตเรโช

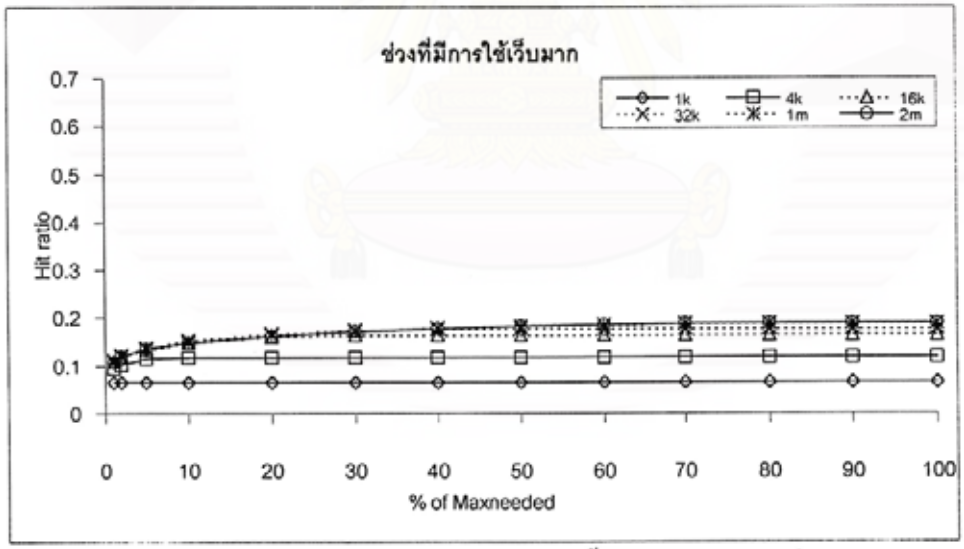


รูปที่ ค.8 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย โดยพิจารณาจากค่าฮิตเรโช

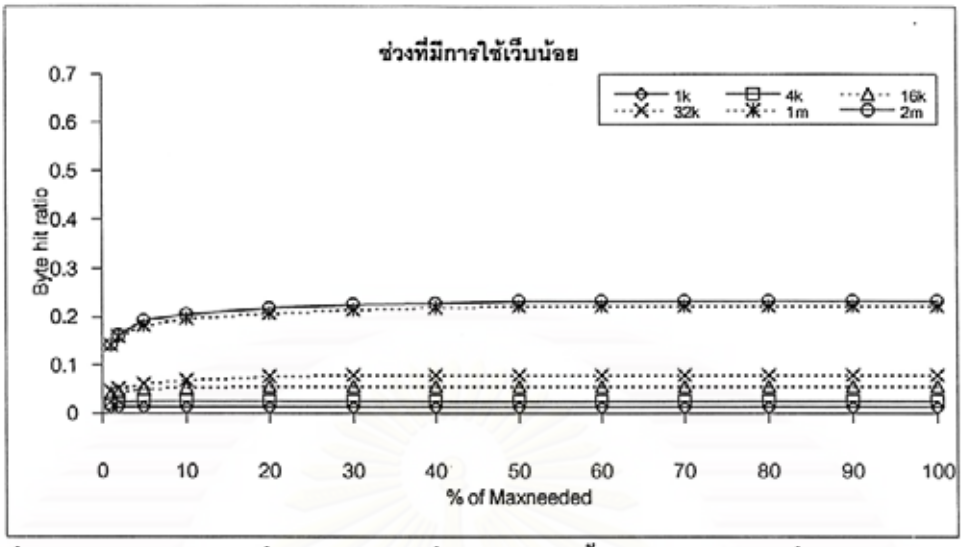
1.3 ข้อมูลการใช้เว็บจากเอ็นแอลเอเอ็นอาร์



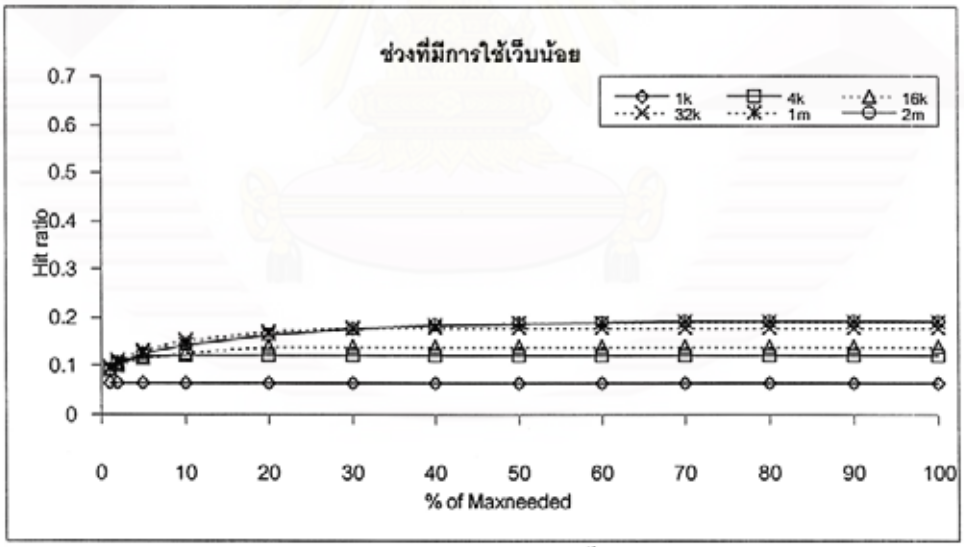
รูปที่ ค.9 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของเอ็นแอลเอเอ็นอาร์โดยพิจารณาจากค่าไบท์ฮิตเรโช



รูปที่ ค.10 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของเอ็นแอลเอเอ็นอาร์โดยพิจารณาจากค่าฮิตเรโช



รูปที่ ค.11 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของเอ็นแอลเอเอ็นอาร์โดยพิจารณาจากค่าไบท์ฮิตเรโช



รูปที่ ค.12 ผลการทดสอบเพื่อหาค่าขีดแบ่งที่เหมาะสมกับขั้นตอนวิธีการแทนที่ LRU-TH สำหรับสภาพการใช้เว็บของเอ็นแอลเอเอ็นอาร์โดยพิจารณาจากค่าฮิตเรโช

2 ผลการจำลองการทำงานของพร็อกซีแคชที่มีขั้นตอนวิธีการแทนที่แบบต่างๆ

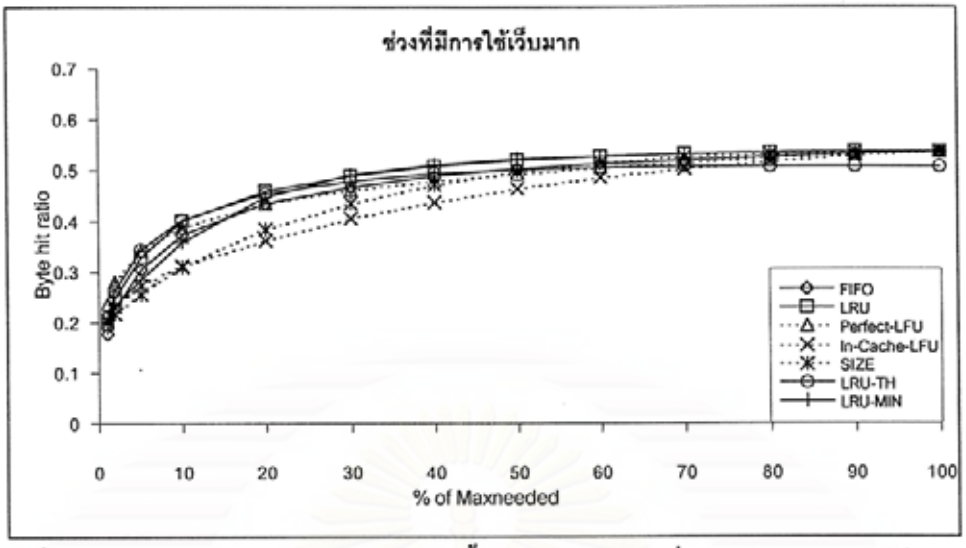
การทดสอบเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่ทั้งหมด 7 แบบ ได้แก่ FIFO LRU Perfect-LFU In-Cache-LFU SIZE LRU-TH และ LRU-MIN ใช้การจำลองการทำงานของพร็อกซีแคช โดยเลือกใช้ค่าขีดแบ่งที่เหมาะสมได้จากผลการทดสอบหาค่าขีดแบ่งในข้อ 1 สำหรับขั้นตอนวิธีการแทนที่แบบ LRU-TH

การทดสอบเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บแต่ละแห่งจะทดสอบกับชุดข้อมูลช่วงที่มีการใช้เว็บมากและช่วงที่มีการใช้เว็บน้อยของการใช้เว็บแต่ละวัน โดยมีค่าฮิตเรโชและค่าไบท์ฮิตเรโชเป็นเครื่องวัด และใช้ค่าเฉลี่ยของผลการทดสอบแต่ละช่วงเวลาและเครื่องวัดเป็นผลการทดสอบเปรียบเทียบ

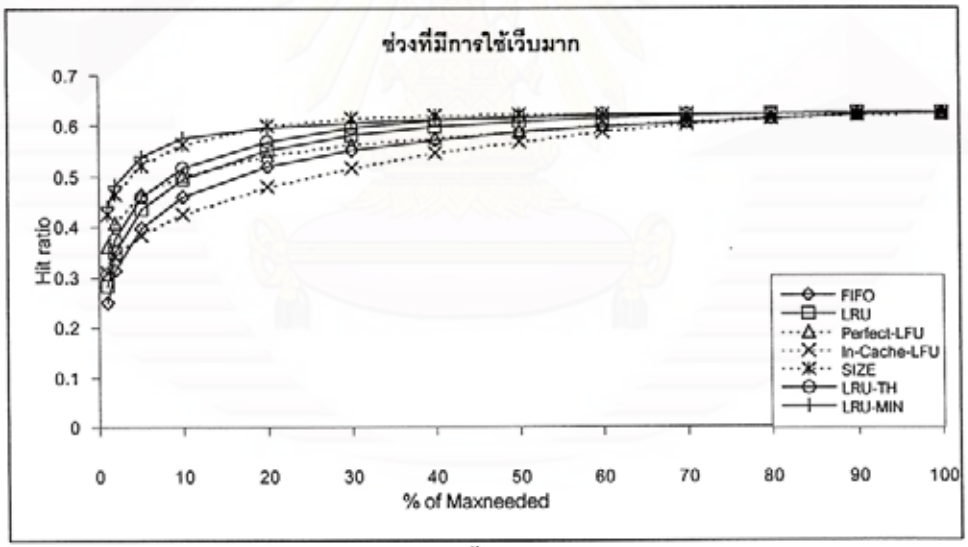
ผลการทดสอบที่ได้พบว่าขั้นตอนวิธีการแทนที่ที่ให้ประสิทธิภาพสูงเมื่อพิจารณาจากค่าไบท์ฮิตเรโชคือขั้นตอนวิธีการแทนที่แบบ Perfect-LFU และ LRU-TH ในกรณีที่พร็อกซีแคชมีความจุน้อยกว่า 10 เปอร์เซ็นต์ของความจุที่ไม่มีการแทนที่ และขั้นตอนวิธีการแทนที่แบบ LRU-MIN ในกรณีที่พร็อกซีแคชมีความจุมากกว่า 10 เปอร์เซ็นต์ของความจุที่ไม่มีการแทนที่ ส่วนการพิจารณาจากค่าฮิตเรโชจะได้ขั้นตอนวิธีการแทนที่ที่ให้ประสิทธิภาพสูงคือขั้นตอนวิธีการแทนที่แบบ LRU-MIN ในกรณีที่พร็อกซีแคชมีความจุน้อยกว่า 10 เปอร์เซ็นต์ของความจุที่ไม่มีการแทนที่ และขั้นตอนวิธีการแทนที่แบบ SIZE ในกรณีที่พร็อกซีแคชมีความจุมากกว่า 10 เปอร์เซ็นต์ของความจุที่ไม่มีการแทนที่

การทดสอบแบ่งตามข้อมูลที่ใช้ในการทดสอบได้ผลดังนี้

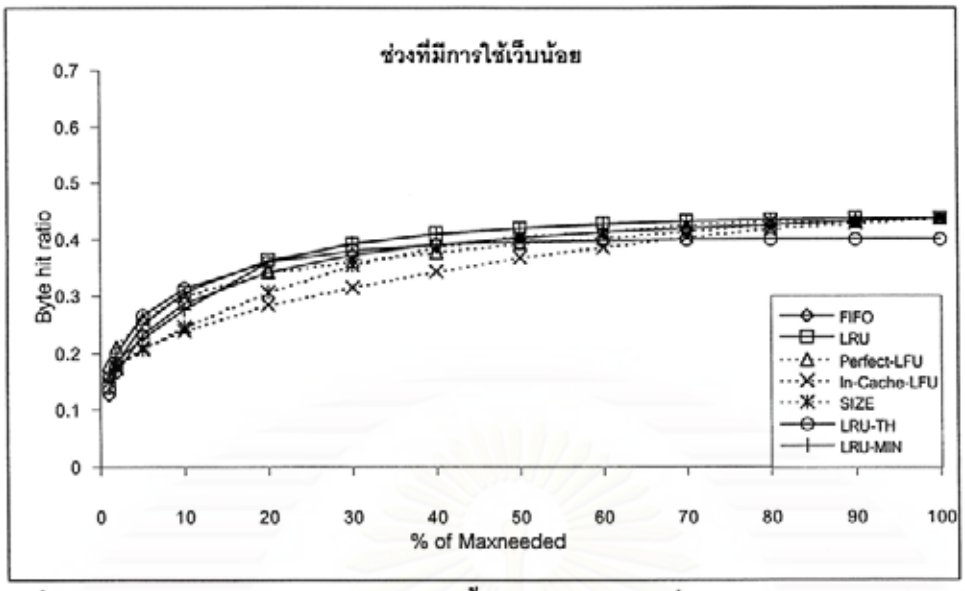
2.1 ข้อมูลการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศจุฬาลงกรณ์มหาวิทยาลัย



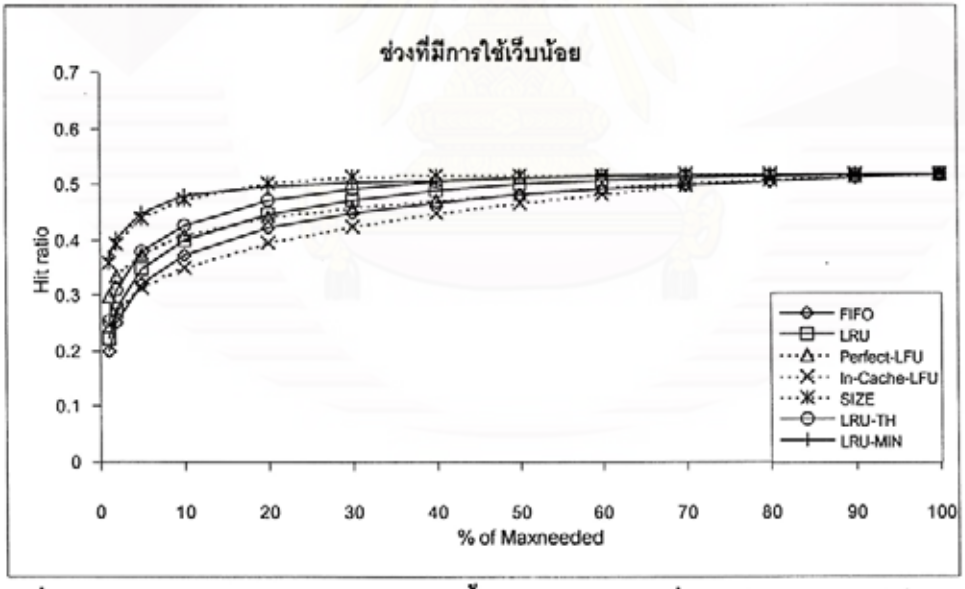
รูปที่ ค.13 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าไบท์ฮิตเรโช



รูปที่ ค.14 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าฮิตเรโช

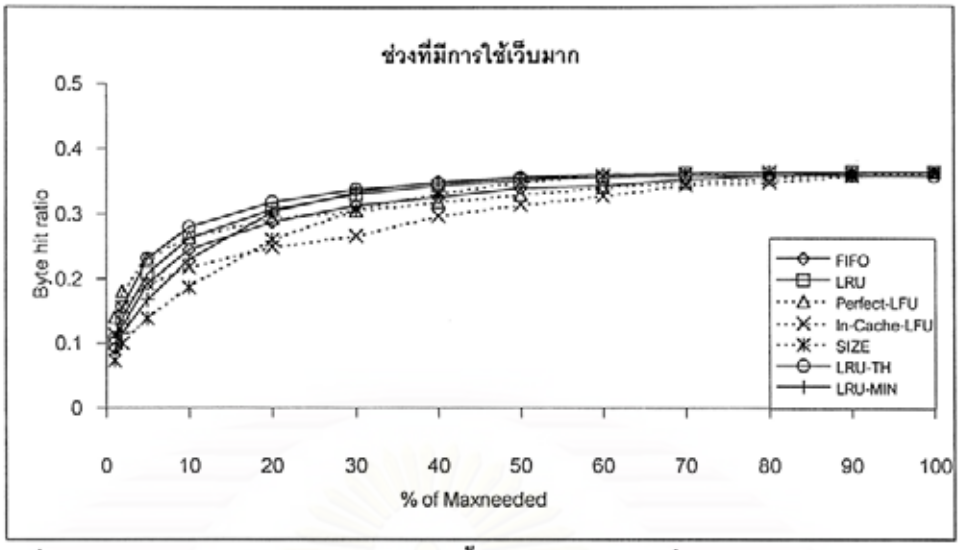


รูปที่ ค.15 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าไบท์ฮิตเรโช

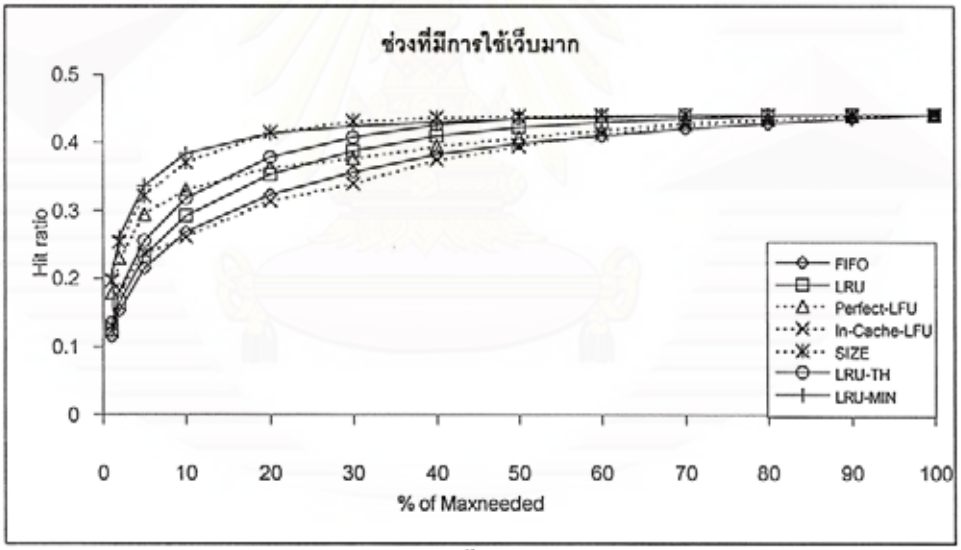


รูปที่ ค.16 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศจุฬาลงกรณ์มหาวิทยาลัยโดยพิจารณาจากค่าฮิตเรโช

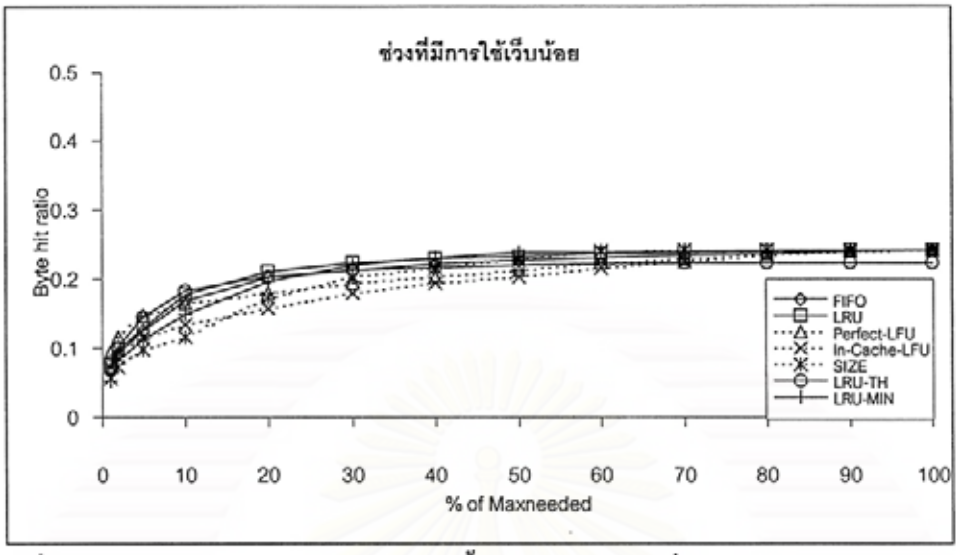
2.2 ข้อมูลการใช้เว็บจากพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย



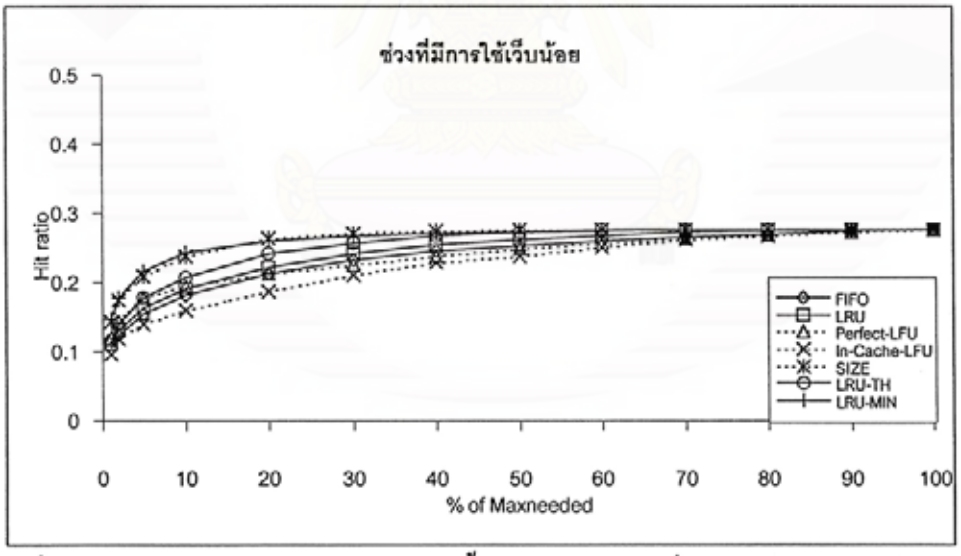
รูปที่ ค.17 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย โดยพิจารณาจากค่าไบท์ฮิตเรโช



รูปที่ ค.18 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย โดยพิจารณาจากค่าฮิตเรโช

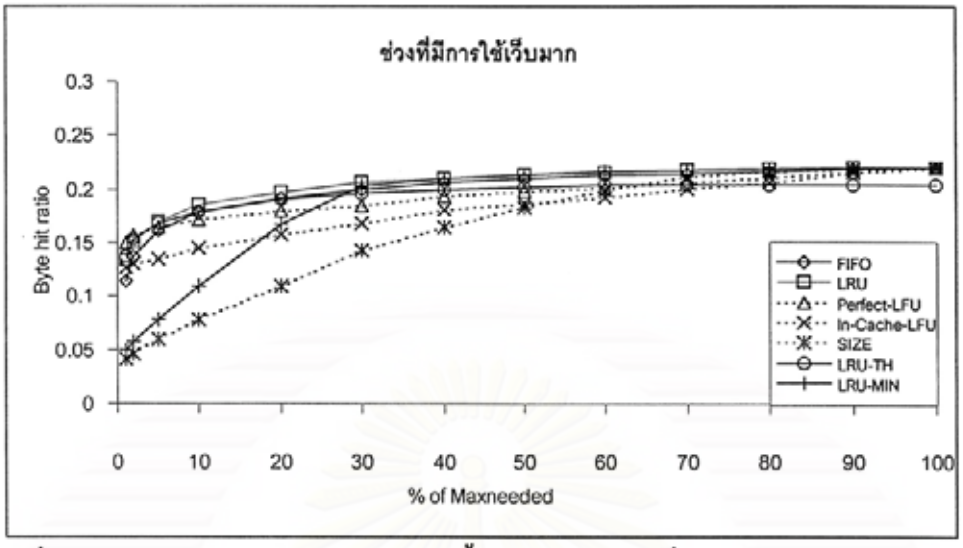


รูปที่ ค.19 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย โดยพิจารณาจากค่าไบท์ฮิตเรโช

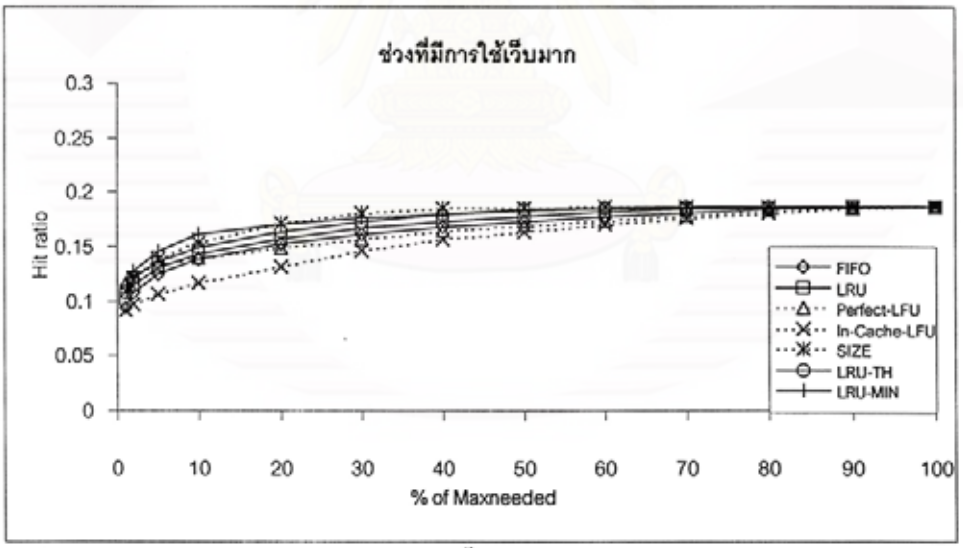


รูปที่ ค.20 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของพร็อกซีเซิร์ฟเวอร์ระดับคณะของจุฬาลงกรณ์มหาวิทยาลัย โดยพิจารณาจากค่าฮิตเรโช

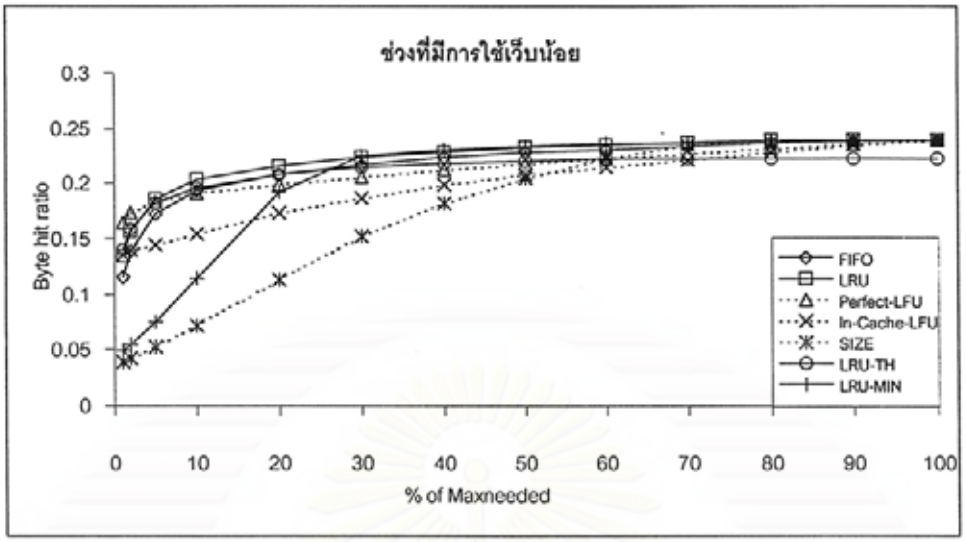
2.3 ข้อมูลการใช้เว็บจากเอ็นแอลเอเอ็นอาร์



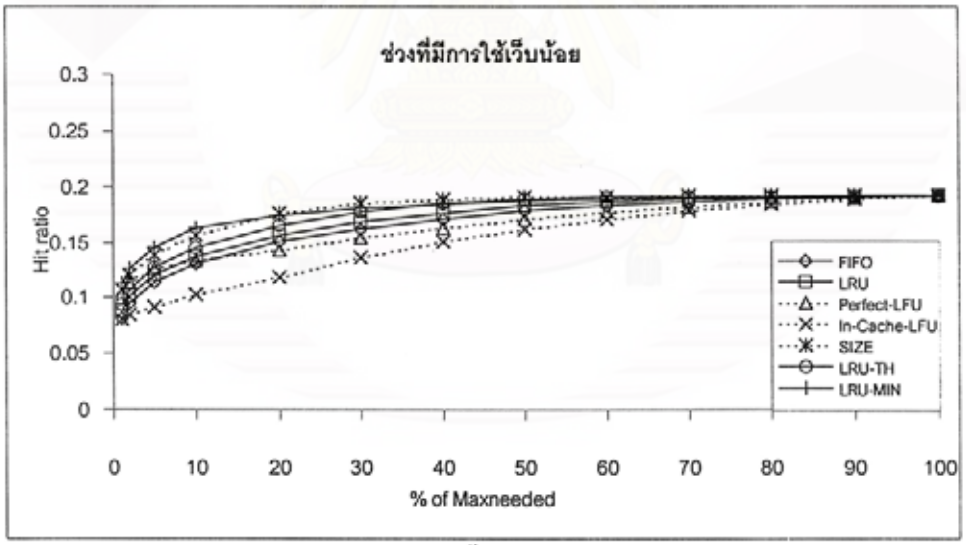
รูปที่ ค.21 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของเอ็นแอลเอเอ็นอาร์โดยพิจารณาจากค่าไบท์ฮิตเรโช



รูปที่ ค.22 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของเอ็นแอลเอเอ็นอาร์โดยพิจารณาจากค่าฮิตเรโช



รูปที่ ค.23 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของเอ็นแอลเอเอ็นอาร์โดยพิจารณาจากค่าไบท์ฮิตเรโช



รูปที่ ค.24 ผลการทดสอบประสิทธิภาพของขั้นตอนวิธีการแทนที่ภายใต้สภาพการใช้เว็บของเอ็นแอลเอเอ็นอาร์โดยพิจารณาจากค่าฮิตเรโช

ภาคผนวก ง

การประมาณค่า maxneeded สำหรับสภาพการใช้เว็บของสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย

การประมาณค่า maxneeded ใช้การเปรียบเทียบค่าฮิตเรโชและค่าไบท์ฮิตเรโชในช่วงที่มีการใช้เว็บมากและน้อยจากสภาพการใช้เว็บจริงกับค่าฮิตเรโชและไบท์ฮิตเรโชที่ได้จากการจำลองการทำงานของพร็อกซีแคชและใช้ขนาดของพร็อกซีแคชที่ใช้ในการจำลองการทำงานของพร็อกซีแคชเป็นค่าประมาณของค่า maxneeded

ค่าฮิตเรโชและไบท์ฮิตเรโชที่ได้จากการนับการเรียกขอที่พบเอกสารในแคชจากแฟ้มที่บันทึกการเรียกขอมีค่าดังตารางที่ ง.1

	วันที่	ค่าฮิตเรโช	ค่าไบท์ฮิตเรโช
ช่วงที่มีการใช้เว็บมาก (8:00 น. - 20:00 น.)	9 ต.ค. 2542	0.434	0.351
	10 ต.ค. 2542	0.448	0.344
	11 ต.ค. 2542	0.472	0.411
	12 ต.ค. 2542	0.450	0.390
	13 ต.ค. 2542	0.468	0.405
	14 ต.ค. 2542	0.449	0.390
	15 ต.ค. 2542	0.519	0.484
	16 ต.ค. 2542	0.427	0.330
ช่วงที่มีการใช้เว็บน้อย(20:00 น.- 8:00 น.)	9-10 ต.ค. 2542	0.402	0.322
	10-11 ต.ค. 2542	0.423	0.315
	11-12 ต.ค. 2542	0.398	0.309
	12-13 ต.ค. 2542	0.407	0.311
	13-14 ต.ค. 2542	0.411	0.278
	14-15 ต.ค. 2542	0.397	0.272
	15-16 ต.ค. 2542	0.395	0.280

ตารางที่ ง.1 ค่าฮิตเรโชและค่าไบท์ฮิตเรโชที่ได้จากสภาพการใช้เว็บจริง

ค่าเฉลี่ยของค่าฮิตเรโชและค่าไบท์ฮิตเรโชสำหรับการเรียกขอในช่วงการใช้เว็บมากและช่วงใช้เว็บน้อยมีค่าดังตารางที่ ง.2

	ค่าฮิตเรโช	ค่าไบท์ฮิตเรโช
ช่วงที่มีการใช้เว็บมาก	0.458	0.388
ช่วงที่มีการใช้เว็บน้อย	0.405	0.298

ตารางที่ ง.2 ค่าเฉลี่ยของค่าฮิตเรโช และค่าไบท์ฮิตเรโชจากแฟ้มที่บันทึกการเรียกขอ

การประมาณค่า maxneeded สำหรับสภาพการใช้เว็บใช้ค่าฮิตเรโชและค่าไบท์ฮิตเรโชที่ได้จากตารางที่ ง.2 ไปเปรียบเทียบกับผลการจำลองการทำงานของพร็อกซีแคชที่มีขั้นตอนวิธีการแทนที่แบบ LRU ตามแผนภูมิในรูปที่ ค.13 ค.14 ค.15 และ ค.16 จะได้ค่า maxneeded สำหรับการใช้เว็บดังตารางที่ ง.3

	ใช้ค่าฮิตเรโชเปรียบเทียบ	ใช้ค่าไบท์ฮิตเรโชเปรียบเทียบ
ช่วงที่มีการใช้เว็บมาก	6.89 %	9.20 %
ช่วงที่มีการใช้เว็บน้อย	11.28 %	9.27 %

ตารางที่ ง.3 ค่า maxneeded ที่ได้จากการเปรียบเทียบผลการทดลองกับสภาพการใช้เว็บจริง

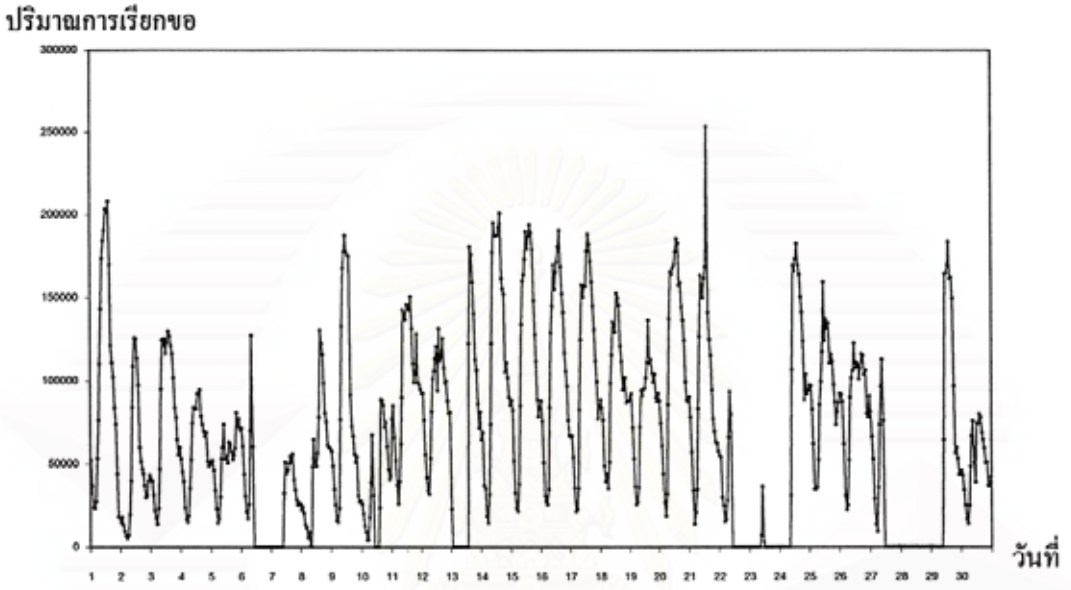
ค่า maxneeded สำหรับสภาพการใช้เว็บจากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย ได้จากการหาค่าเฉลี่ยของค่า maxneeded จากตารางที่ ง. 3 ได้ค่าเท่ากับ 9.16 เปอร์เซ็นต์

ภาคผนวก จ

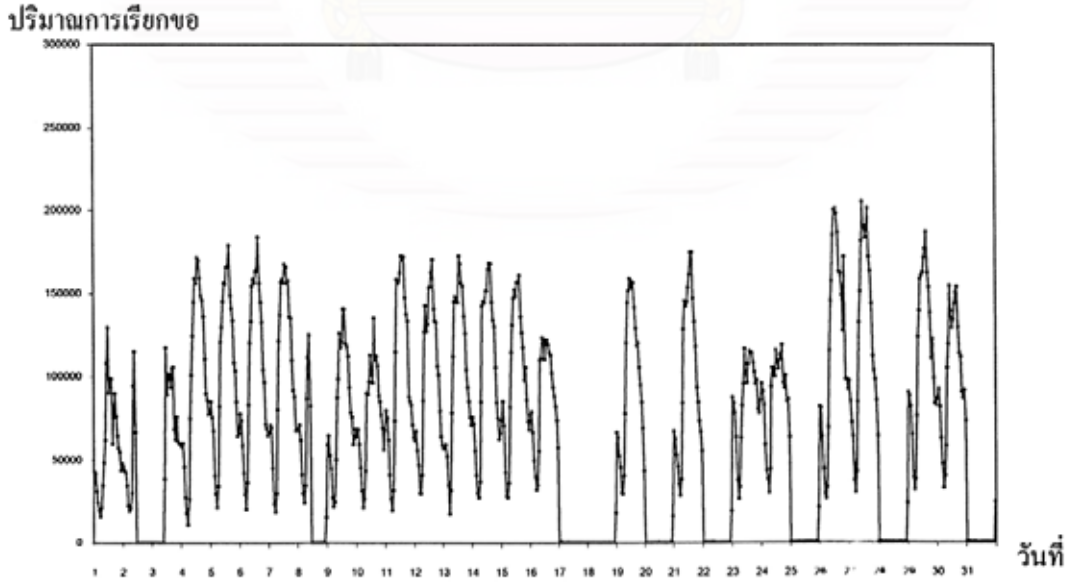
การวิเคราะห์ข้อมูลการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัย

การใช้เว็บในช่วงเดือนกันยายน พ.ศ. 2542 ถึง เดือนตุลาคม พ.ศ. 2542

ข้อมูลการใช้เว็บที่ใช้ในการศึกษาแบ่งออกเป็น 2 ส่วน คือ ข้อมูลการใช้เว็บในช่วงเปิดภาคการศึกษาและข้อมูลการใช้เว็บในช่วงปิดภาคการศึกษา ข้อมูลที่ใช้คือข้อมูลการใช้เว็บเดือนกันยายน พ.ศ. 2542 และข้อมูลการใช้เว็บเดือนตุลาคม พ.ศ. 2542 โดยมีปริมาณการเรียกขอเฉลี่ยในแต่ละชั่วโมงดังแผนภูมิในรูปที่ จ.1 และ จ.2

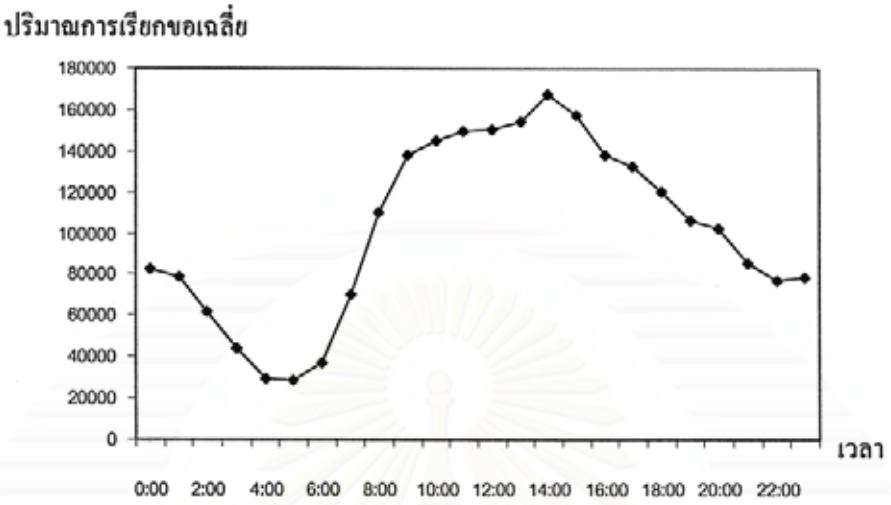


รูปที่ จ.1 ปริมาณการเรียกขอต่อชั่วโมงช่วงเดือนกันยายน พ.ศ. 2542

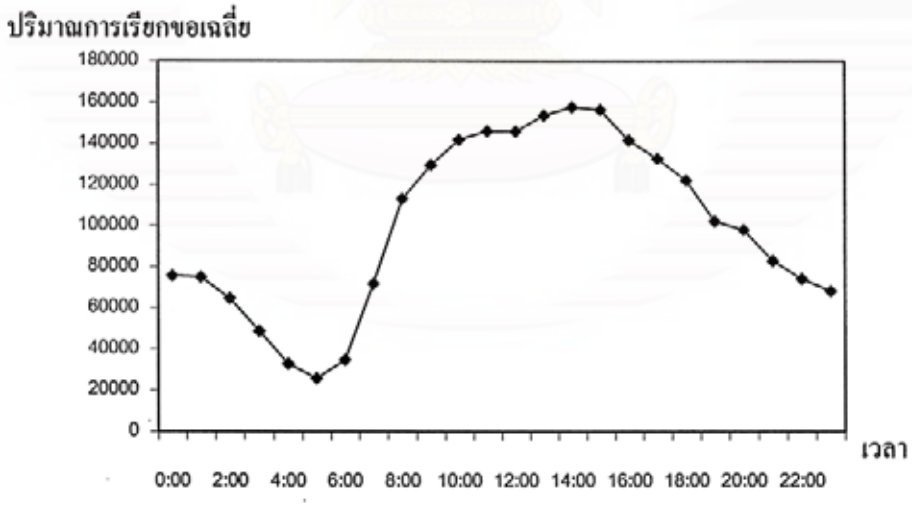


รูปที่ จ.2 ปริมาณการเรียกขอต่อชั่วโมงช่วงเดือนตุลาคม พ.ศ. 2542

การใช้เว็บบางชั่วโมงมีปริมาณการเรียกขอน้อยมากหรือมีค่าเป็นศูนย์เนื่องจากบางช่วงเวลามีการรีบูตแคชเอ็นจิ้นเนื่องจากมีการปรับปรุงซอฟต์แวร์ที่ใช้ในการจัดการระบบแคชหรือแคชเอ็นจิ้นมีการทำงานขัดข้อง ดังนั้นในการพิจารณาช่วงเวลาที่มีการใช้เว็บมากและช่วงที่มีการใช้เว็บน้อยจะใช้ข้อมูลการใช้เว็บวันที่มีข้อมูลการใช้เว็บครบทุกชั่วโมง โดยปริมาณการเรียกขอเฉลี่ยในแต่ละชั่วโมงสำหรับข้อมูลการใช้เว็บในเดือนกันยายนและเดือนตุลาคมสามารถแสดงได้ดังแผนภูมิในรูปที่ จ.3 และ จ.4



รูปที่ จ.3 ปริมาณการเรียกขอเฉลี่ยแต่ละชั่วโมงของข้อมูลการใช้เว็บ
จากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย เดือนกันยายน พ.ศ. 2542



รูปที่ จ.4 ปริมาณการเรียกขอเฉลี่ยแต่ละชั่วโมงของข้อมูลการใช้เว็บ
จากสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย เดือนตุลาคม พ.ศ. 2542

ข้อมูลการใช้เว็บในช่วงเดือนกันยายนและเดือนตุลาคมแบ่งออกได้เป็น 2 ช่วงคือช่วงเวลาที่มีการใช้เว็บมากและช่วงเวลาที่ใช้เว็บน้อยโดยพิจารณา 12 ชั่วโมงที่มีการใช้เว็บมากจากแผนภูมิในรูปที่ จ.3

และ จ.4 เป็นช่วงที่มีการใช้เว็บมากและอีก 12 ชั่วโมงเป็นช่วงที่มีการใช้เว็บน้อย ได้ช่วงที่มีการใช้เว็บมากและช่วงที่มีการใช้เว็บน้อยของเดือนกันยายนและเดือนตุลาคมเป็นช่วงเวลาเดียวกันดังตารางที่ จ.1

	ตั้งแต่	ถึง
ช่วงที่มีการใช้เว็บมาก	8:00 น.	8:00 น.
ช่วงที่มีการใช้เว็บน้อย	20:00 น.	20:00 น.

ตารางที่ จ.1 ช่วงเวลาที่มีการใช้เว็บมากและน้อยของข้อมูลการใช้เว็บเดือนกันยายนและเดือนตุลาคม

การทดสอบค่าเฉลี่ยของปริมาณการใช้เว็บที่ใช้ในการจำลองการทำงานของพร็อกซีแคช

ข้อมูลทางสถิติของข้อมูลการใช้เว็บที่ใช้ในการจำลองการทำงานของพร็อกซีแคชและข้อมูลการใช้เว็บในช่วงเดือนกันยายนและเดือนตุลาคมมีค่าดังตารางที่ จ.2

		ค่าเฉลี่ย ของปริมาณการเรียกขอ	ค่าเบี่ยงเบนมาตรฐาน ของปริมาณการเรียกขอ
ข้อมูลการใช้เว็บที่ใช้ในการจำลองการทำงานของพร็อกซีแคช	**ช่วงที่มีการใช้เว็บมาก**	1563677	214196
	ช่วงที่มีการใช้เว็บน้อย	711908	47514
ข้อมูลการใช้เว็บเดือนกันยายน	**ช่วงที่มีการใช้เว็บมาก**	1671262	298563
	ช่วงที่มีการใช้เว็บน้อย	776760	83972
ข้อมูลการใช้เว็บเดือนตุลาคม	**ช่วงที่มีการใช้เว็บมาก**	1643276	233904
	ช่วงที่มีการใช้เว็บน้อย	752499	83451

ตารางที่ จ.2 ข้อมูลทางสถิติของข้อมูลการใช้เว็บในช่วงเดือนกันยายนและเดือนตุลาคม

การทดสอบการทดสอบค่าเฉลี่ยของปริมาณการใช้เว็บที่ใช้ในการจำลองการทำงานของพร็อกซีแคชใช้การทดสอบแบบ t-test โดยการทดสอบแบ่งออกเป็น 2 ส่วน คือ

1. การทดสอบความแตกต่างระหว่างค่าเฉลี่ยของปริมาณการใช้เว็บที่ใช้ในการจำลองการทำงานของพร็อกซีแคชกับค่าเฉลี่ยของปริมาณการใช้เว็บเดือนกันยายน
2. การทดสอบความแตกต่างระหว่างค่าเฉลี่ยของปริมาณการใช้เว็บที่ใช้ในการจำลองการทำงานของพร็อกซีแคชกับค่าเฉลี่ยของปริมาณการใช้เว็บเดือนตุลาคม

ผลการทดสอบแบบ t-test โดยมีระดับนัยสำคัญเท่ากับ 0.05 (confidence level = 95%)ได้ผลดังตารางที่ จ.3

	ข้อมูลการใช้เว็บที่ใช้ทดสอบกับข้อมูลการใช้เว็บเดือนกันยายน	ข้อมูลการใช้เว็บที่ใช้ทดสอบกับข้อมูลการใช้เว็บเดือนตุลาคม
ช่วงที่มีการใช้เว็บมาก	ไม่แตกต่าง	ไม่แตกต่าง
ช่วงที่มีการใช้เว็บน้อย	แตกต่าง	ไม่แตกต่าง

ตารางที่ จ.3 ผลการทดสอบแบบ t-test ที่ระดับนัยสำคัญเท่ากับ 0.05

ผลการทดสอบแบบ t-test เพื่อพิจารณาความแตกต่างระหว่างค่าเฉลี่ยของปริมาณการเรียกขอของข้อมูลการใช้เว็บที่ใช้ในการจำลองการทำงานของพร็อกซีแคชและข้อมูลการใช้เว็บในช่วงเดือนกันยายนและเดือนตุลาคม สรุปได้ว่าข้อมูลการใช้เว็บที่นำมาใช้ในการจำลองการทำงานของพร็อกซีแคชมีปริมาณการเรียกขอที่ไม่แตกต่างกับปริมาณการเรียกขอในช่วงเดือนตุลาคมและข้อมูลการใช้เว็บที่นำมาใช้ในการจำลองการทำงานของพร็อกซีแคชมีปริมาณการเรียกขอในช่วงที่มีการใช้เว็บมากไม่แตกต่างกับปริมาณการเรียกขอในช่วงเดือนกันยายนหรือในช่วงเปิดภาคการศึกษา สำหรับปริมาณการเรียกขอในช่วงที่มีการใช้เว็บน้อยแตกต่างกับปริมาณการเรียกขอในช่วงเดือนกันยายน

ข้อมูลการใช้เว็บที่ใช้ในการจำลองการทำงานของพร็อกซีแคชเลือกใช้ข้อมูลในช่วงวันที่ 9 ถึง 16 ตุลาคม พ.ศ. 2542 เนื่องจากเป็นช่วงเวลาที่ใกล้เคียงกับข้อมูลการใช้เว็บที่รวบรวมมากจากพร็อกซีเซิร์ฟเวอร์ระดับคณะและข้อมูลการใช้เว็บจากเอ็นแอลเอเอ็นอาร์ที่นำมาใช้ในการเปรียบเทียบผลการจำลองการทำงานของพร็อกซีแคช และสามารถใช้แทนข้อมูลการใช้เว็บในช่วงปิดภาคการศึกษาและในช่วงที่มีการใช้เว็บมากของช่วงเปิดภาคการศึกษาซึ่งเป็นปริมาณการเรียกขอส่วนใหญ่ได้

ประวัติผู้วิจัย

นายพรทวี วัฒนวิทูกูร เกิดเมื่อวันอังคารที่ 2 พฤศจิกายน พ.ศ. 2519 ที่จังหวัดกรุงเทพมหานคร สำเร็จการศึกษาปริญญาตรีวิศวกรรมศาสตรบัณฑิต ภาควิชาเครื่องกล คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2540 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต ที่ภาควิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อ พ.ศ. 2540