

การวิเคราะห์ข้อมูลเครือข่ายเพื่อปรับแต่งเว็บแคชเซิร์ฟเวอร์

Traffic Analysis for Web Cache Server Configurations

จักรชัย โสอินทร์

ยืน ภู่วรรณ

สุรศักดิ์ สวงพงษ์

สำนักบริการคอมพิวเตอร์
มหาวิทยาลัยเกษตรศาสตร์
ถ.พหลโยธิน บางเขน
กรุงเทพมหานคร 10900
cpcccs@ku.ac.th

สำนักบริการคอมพิวเตอร์
มหาวิทยาลัยเกษตรศาสตร์
ถ.พหลโยธิน บางเขน
กรุงเทพมหานคร 10900
nguan@ku.ac.th

สำนักบริการคอมพิวเตอร์
มหาวิทยาลัยเกษตรศาสตร์
ถ.พหลโยธิน บางเขน
กรุงเทพมหานคร 10900
yuen@ku.ac.th

บทคัดย่อ

บทความนี้กล่าวถึงการวิเคราะห์ลักษณะการเรียกใช้ข้อมูลเว็บผ่านเว็บแคชเซิร์ฟเวอร์โดยพิจารณาจากสถิติในช่วงคาบเวลาหนึ่ง ค่าที่นำมาวิเคราะห์คือปริมาณการร้องขอใช้งานขนาดข้อมูลถ่ายโอน การใช้งานแบบครั้งเดียวและระยะเวลาการเปลี่ยนแปลงของข้อมูลโดยแยกแต่ละประเภทของไฟล์

ผลการวิเคราะห์พบว่าข้อมูลเท็กซ์และข้อมูลรูปภาพมีบทบาทสำคัญสูงกว่าข้อมูลประเภทอื่นคือมีการร้องขอใช้งานและขนาดข้อมูลถ่ายโอนรวมกันเป็นจำนวน 95% และ 79% ตามลำดับ การใช้งานแบบครั้งเดียวมีสัดส่วนครึ่งหนึ่งและเมื่อพิจารณาข้อมูลแยกตามขนาดของไฟล์พบว่าปริมาณการร้องขอไฟล์ข้อมูลเท็กซ์กับไฟล์ข้อมูลรูปภาพโดยส่วนใหญ่จะเกิดขึ้นกับไฟล์ที่มีขนาดน้อยกว่า 128 กิโลไบต์ และ 64 กิโลไบต์ตามลำดับ สำหรับขนาดข้อมูลถ่ายโอนพบว่าทั้งไฟล์ข้อมูลเท็กซ์และไฟล์ข้อมูลรูปภาพโดยส่วนใหญ่จะเกิดขึ้นกับไฟล์ที่มีขนาดน้อยกว่า 256 กิโลไบต์ และขนาดไฟล์ที่เหมาะสมสำหรับการแคชข้อมูลโดยไม่แยกประเภทของไฟล์คือ 64

กิโลไบต์ จากการทดลองยังพบว่าขนาดคิสก์ที่เหมาะสมของเว็บแคชเซิร์ฟเวอร์จะอยู่ที่ประมาณ 35-50% ของขนาดข้อมูลถ่ายโอน

Abstract

This article describes an analysis of data on web caches in a period of time. We emphasize the relationship of request rate, transfer size, one-time referencing, and changing of data separated by file types.

The results show that the texts and images have the highest request rate and transfer size of 95% and 79% respectively with 50% one-time referencing. We found that the most of requests occurs with the files size less than 128 and 64 kilobytes for texts and images respectively. While considering the transfer size, we found that users mostly request files at the size less than 256 kilobytes for both texts and images. From our experimental, the optimum file size for caching is equal to 64.

1. บทนำ

ปัจจุบันอัตราการใช้เว็บเพิ่มขึ้นอย่างรวดเร็วทั้งในส่วนของข้อมูลที่เป็นตัวอักษร รูปภาพหรือวิดีโอซึ่งมีขนาดและอัตราการใช้งานที่แตกต่างกัน การเพิ่มขึ้นนี้เป็นปัจจัยสำคัญที่ก่อให้เกิดความคับคั่งของข้อมูลบนเครือข่าย [6] ซึ่งส่งผลให้เกิดความล่าช้าในการร้องขอข้อมูลและมีระยะเวลารอเพื่อเข้าถึงข้อมูลยาวนานขึ้น

เว็บแคชเซิร์ฟเวอร์จึงได้รับการพัฒนาขึ้นเพื่อลดปัญหาข้างต้น หลักการทำงานของเว็บแคชเซิร์ฟเวอร์คือรอรับคำร้องขอข้อมูลจาก ไคลเอ็นต์ที่ประกอบด้วยข้อมูลประเภทต่างๆ หากเว็บแคชเซิร์ฟเวอร์พบข้อมูลที่ร้องขอในแคชก็จะนำข้อมูลส่งต่อไปยังไคลเอ็นต์ทันที แต่ถ้าไม่พบก็จะร้องขอข้อมูลไปยังเซิร์ฟเวอร์ต้นทางเพื่อถ่ายโอนข้อมูลและนำส่งต่อไปยังไคลเอ็นต์รวมทั้งจัดเก็บข้อมูลนั้นไว้เพื่อให้บริการสำหรับการร้องขอครั้งถัดไป

2. งานวิจัยที่เกี่ยวข้อง

การพัฒนาขีดความสามารถของเว็บแคชเซิร์ฟเวอร์มีเทคนิคต่างๆมากมาย [2] ทั้งนี้ล้วนแล้วแต่มีจุดประสงค์เพื่อเพิ่มสมรรถนะความเร็ว ความเชื่อถือได้และการขยายขนาดได้

อัตราการใช้งานซ้ำ (Hit Ratio) และค่าเวลาเฉลี่ยของข้อมูลตอบกลับ (Mean Response Time) เป็นค่าที่มักใช้วัดสมรรถนะการทำงานของเว็บแคชเซิร์ฟเวอร์ ค่านี้จะขึ้นอยู่กับหน่วยประมวลผลกลาง ขนาดหน่วยความจำ สมรรถนะของดิสก์และสมรรถนะของซอฟต์แวร์แคชที่ใช้งาน [7] นอกจากนี้ องค์ประกอบสำคัญอีกประการหนึ่งซึ่งจะมีผลกระทบต่อสมรรถนะของเว็บแคชเซิร์ฟเวอร์คือลักษณะการใช้งานของผู้ใช้บนเครือข่าย หากสามารถนำเอาข้อมูลที่ได้จากการเก็บบันทึกของเว็บแคชเซิร์ฟเวอร์มาวิเคราะห์หาค่าพารามิเตอร์ซึ่งเป็นองค์ประกอบสำคัญแล้วก็จะสามารถเพิ่มสมรรถนะของเว็บแคชเซิร์ฟเวอร์ได้โดยไม่จำเป็นต้องปรับเปลี่ยนองค์ประกอบการทำงานส่วนอื่น [5]

งานวิจัยนี้มุ่งเน้นการเพิ่มสมรรถนะเว็บแคชเซิร์ฟเวอร์โดยใช้เทคนิคการวิเคราะห์ปริมาณข้อมูลบนเครือข่ายเพื่อหาค่าพารามิเตอร์ที่สามารถปรับแต่งให้เว็บแคชเซิร์ฟเวอร์มีประสิทธิภาพเพิ่มมากขึ้น [1,3,4] ข้อมูลที่ได้จะผ่านการวิเคราะห์โดยพิจารณาถึงลักษณะการกระจายตัวของการใช้งาน การใช้งานแบบครั้งเดียว ขนาดข้อมูลถ่ายโอนและระยะเวลาการเปลี่ยนแปลงโดยแยกตามประเภทของไฟล์แต่ละประเภทเช่น ข้อมูลเท็กซ์ ข้อมูลรูปภาพ ข้อมูลเสียง ข้อมูลวิดีโอ และข้อมูลอื่นๆ (ข้อมูลไม่มีไทป์) เทียบกับขนาดไฟล์

บทความนี้จัดลำดับหัวข้อดังต่อไปนี้ หัวข้อถัดไปกล่าวถึงลักษณะการจัดเก็บข้อมูลแคชที่นำมาวิเคราะห์ หัวข้อที่ 4 กล่าวถึงขอบเขตของข้อมูลที่จัดเก็บ หัวข้อที่ 5 เป็นการวิเคราะห์ข้อมูลทางสถิติตามพารามิเตอร์ต่างๆ หัวข้อที่ 6 เสนอการทดลองปรับแต่งพารามิเตอร์เพื่อวัดสมรรถนะการทำงานและหัวข้อสุดท้ายเสนอผลสรุปการวิเคราะห์และแนวทางการวิจัยต่อไปในอนาคต

3. ลักษณะการจัดเก็บข้อมูลแคช

ข้อมูลที่นำมาวิเคราะห์ได้มาจากการจัดเก็บการร้องขอใช้งานจากเว็บแคชเซิร์ฟเวอร์ของมหาวิทยาลัยเกษตรศาสตร์ เว็บแคชเซิร์ฟเวอร์ใช้หน่วยประมวลผลกลางเพนเทียมทรีทำงานที่สัญญาณนาฬิกา 500 เมกะเฮิร์ตซ์ มีหน่วยความจำ 1 กิกะไบต์ และฮาร์ดดิสก์ชนิด SCSI-II ความจุ 9.1 กิกะไบต์ จำนวน 3 ตัว ต่อเชื่อมด้วย RAID-0 (ใช้เนื้อที่สำหรับเก็บแคชเพียง 19.5 กิกะไบต์) ระบบปฏิบัติการที่ใช้คือลินุกซ์ เรดแฮท 6.2 เคอร์เนล 2.2.16 ใช้โปรแกรมสควิดซ์ 2.2 Stable4 นโยบายการทดแทนข้อมูลที่ใช้คือ LRU (Least Recently Used) และจำกัดขนาดข้อมูลแคชต่อหน่วยสูงสุดที่ 64 เมกะไบต์

โครงสร้างการติดต่อภายนอกของเว็บแคชเซิร์ฟเวอร์มีช่องทางสื่อสารข้อมูลสองส่วนคือ ส่วนที่ออกภายนอกประเทศจะผ่านทางเครือข่ายของยูนิเน็ต โดยจำกัดช่องทางเครือข่ายที่ 2 เมกะบิตต่อวินาทีและส่วนของข้อมูลภายในประเทศ

จะผ่านเครือข่ายหลัก ของไทยสาร โดยมีจุดเชื่อมต่อเป็น 155 เมกะบิตต่อวินาที

4. ขอบเขตของข้อมูลที่จัดเก็บ

ข้อมูลที่จัดเก็บได้มาจากคาบเวลา 13 วันคือระหว่างวันที่ 4 ถึง 17 กุมภาพันธ์ 2544 (ยกเว้นวันที่ 15 ที่แคชปิดบริการ) ทั้งนี้ช่วงเวลาดังกล่าวจะครอบคลุมทั้งเวลางานปกติและวันหยุด

ปริมาณข้อมูลที่จัดเก็บจากค่าสถิติของโปรแกรมสควิดซ์โดยแยกในแต่ละวันเนื่องจากข้อมูลมีขนาดใหญ่ซึ่งมีขนาดรวมกันประมาณ 1.4 กิกะไบต์

ข้อมูลที่นำมาวิเคราะห์จะเป็นข้อมูลที่ผ่านเงื่อนไขสองประการคือ ประการที่หนึ่งเป็นข้อมูลการร้องขอใช้งานจากไคลเอ็นต์ผ่าน โพรโตคอลเอชทีทีพีเท่านั้น (ไม่รวมถึงบริการแบบอื่นเช่น เอฟทีพีหรือโกเฟอร์) ประการที่สองข้อมูลจะต้องผ่านการร้องขอใช้งานด้วยความถูกต้องโดยสมบูรณ์ (รหัส 200 OK) นอกจากนี้จะไม่รวมยูอาร์แอลแบบไดนามิก เช่น cgi, php, asp และยูอาร์แอลที่มีสัญลักษณ์พิเศษประกอบ เช่น ?, =, %, *, +, \$, &, @, [], () มาใช้วิเคราะห์ ยูอาร์แอลที่ผ่านเงื่อนไขทั้งสองประการข้างต้นมีสัดส่วนเป็นครึ่งหนึ่งของยูอาร์แอลรวมและมีขนาดข้อมูลเป็น 70% ของขนาดข้อมูลรวม

5. การวิเคราะห์ข้อมูล

ข้อมูลที่จัดเก็บตามข้อกำหนดในหัวข้อที่ 4 จะนำมาวิเคราะห์เพื่อหาพารามิเตอร์ที่สำคัญต่อสมรรถนะของเว็บแคชเซิร์ฟเวอร์ โดยจัดแบ่งการวิเคราะห์ออกเป็น 3 ส่วนซึ่งเป็นอิสระต่อกันดังต่อไปนี้คือ

- การวิเคราะห์การทำงาน โดยเปรียบเทียบ การวิเคราะห์ส่วนนี้จะเปรียบเทียบถึงค่าพารามิเตอร์ที่แตกต่างกันของการร้องขอใช้งาน ขนาดข้อมูลถ่ายโอนและการใช้งานแบบครั้งเดียวของข้อมูลรวม
- การวิเคราะห์เปรียบเทียบความสัมพันธ์ของค่าสี่ค่าต่อไปนี้ โดยแยกตามประเภทของไฟล์เทียบกับการเพิ่มขึ้นของ

ขนาดไฟล์คือ การร้องขอใช้งานโดยรวม ขนาดข้อมูลถ่ายโอนโดยรวม การร้องขอใช้งานแบบครั้งเดียว ขนาดข้อมูลถ่ายโอนแบบครั้งเดียว

- การวิเคราะห์การใช้ข้อมูลซ้ำซึ่งแสดงถึงจำนวนและระยะเวลาการเรียกใช้งานซ้ำโดยแยกตามประเภทของไฟล์

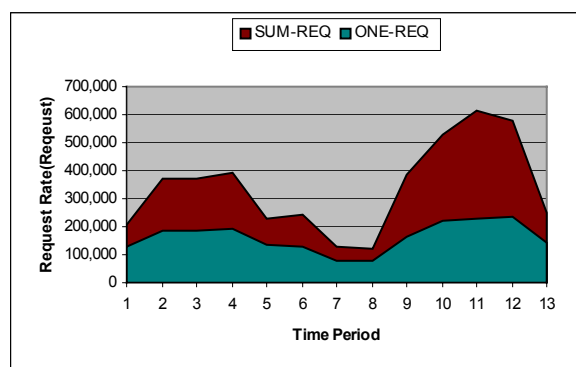
5.1 การวิเคราะห์การทำงานโดยเปรียบเทียบ

การวิเคราะห์ในหัวข้อนี้กล่าวถึงค่าพารามิเตอร์ที่ส่งผลกระทบต่อสมรรถนะของแคช ข้อมูลจะถูกวิเคราะห์ออกเป็นสามส่วนย่อยคือ ส่วนแรกคือค่าพารามิเตอร์โดยรวม ส่วนที่สองกล่าวถึงความสัมพันธ์ของการร้องขอใช้งานและขนาดข้อมูลถ่ายโอนรวมเทียบกับขนาดไฟล์และส่วนสุดท้ายจะกล่าวถึงความสัมพันธ์ของการใช้งานแบบครั้งเดียวเทียบกับขนาดไฟล์

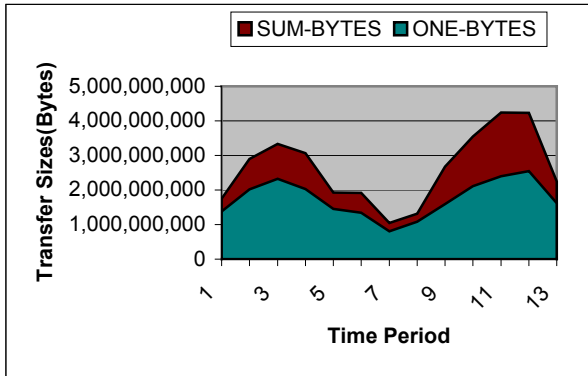
5.1.1 ค่าพารามิเตอร์โดยรวม

พารามิเตอร์ที่ได้จากการวิเคราะห์ในส่วนนี้แสดงได้ดังรูปที่ 1, 2 และตารางที่ 1, 2

รูปที่ 1 และ 2 แสดงถึงผลรวมการร้องขอใช้งานและขนาดข้อมูลถ่ายโอนโดยรวมและแบบครั้งเดียวในช่วงคาบเวลาซึ่งพบว่าการร้องขอใช้งานแบบครั้งเดียวเป็นสัดส่วนครึ่งหนึ่งของการร้องขอใช้งานทั้งหมด ($ONE-REQ \approx SUM-REQ/2$) ทั้งหมดในทำนองเดียวกันกับขนาดข้อมูลถ่ายโอน ($ONE-BYTES \approx SUM-BYTES/2$)



รูปที่ 1 : ปริมาณการร้องขอใช้งานโดยรวมและแบบครั้งเดียวในช่วงคาบเวลาที่จัดเก็บ



รูปที่ 2 : ปริมาณของขนาดข้อมูลถ่ายโอนโดยรวมและแบบครั้งเดียวในช่วงคาบเวลาที่จัดเก็บ

ตารางที่ 1 แสดงถึงค่าพารามิเตอร์เฉลี่ยเห็นได้ว่าการร้องขอใช้งานแบบครั้งเดียวมีปริมาณ 50 % ของการร้องขอใช้งานทั้งหมดและเป็น 85 % เมื่อเทียบกับจำนวนการอ้างอิงที่แตกต่างกัน (Distinct Request) แสดงว่ามีการร้องขอใช้งานแบบครั้งเดียวเป็นส่วนใหญ่

เมื่อสังเกตถึงความสัมพันธ์ระหว่างค่ากลางและค่าเฉลี่ยเห็นได้ถึงความแตกต่างกันค่อนข้างมากซึ่งแสดงว่าช่วง

ของไฟล์ที่มีขนาดใหญ่และเล็กจะมีความแตกต่างกันมากหรือไฟล์ขนาดเล็กเกิดขึ้นจำนวนมาก

ข้อมูลบางส่วนจากตารางที่ 1 จะแบ่งวิเคราะห์ตามประเภทไฟล์ข้อมูลออกเป็น 5 กลุ่มและแสดงผลการวิเคราะห์ที่ได้ดังตารางที่ 2 คือ AUDIO, IMAGE, APPLICATION, TEXT และ VIDEO

ตารางที่ 1 : ค่าพารามิเตอร์เฉลี่ยของการจัดเก็บค่าสถิติ

ITEM	KU
Total Request	4,415,543.00
Total Bytes Transfer (MBbytes)	32,599.56
Avg. Request/Day	339,657.15
Avg. Bytes/Day (MBbytes)	2,507.66
One-Time Request	2,098,985.00
One-Time Transfer (MBbytes)	21,665.03
One-Time Request/Day	161,460.38
One-Time Transfer/Day (Bytes)	1,666.54
Mean Transfer (Bytes)	8,086.15
Median Transfer (Bytes)	2,185.23
Distinct Request	2,488,404.00
Distinct Transfer (MBytes)	24,072.87
Max/Avg. Client Request	1367/860

ตารางที่ 2 : ค่าพารามิเตอร์แยกตามประเภทของข้อมูลไม่มีโทป์

ITEM	AUDIO	IMAGE	APPLICATION	TEXT	VIDEO	OTHER
% of Request	0.25	78.41	3.31	17.86	0.03	0.15
% of Bytes Transfer	1.55	52.57	17.11	27.37	1.05	0.34
%One-Time Request	0.16	37.65	0.97	8.70	0.02	0.03
%One-Time Bytes	1.32	31.62	13.05	17.41	0.98	2.07
Median (Bytes)	7,111.92	1,962.31	2,503.08	4,089.85	105,356.92	-
Mean (Bytes)	52,009.46	5,395.38	43,611.62	12,027.77	274,077.54	26,834.98

ผลจากตารางที่ 2 พบว่าประเภทของไฟล์ที่มีสัดส่วนสูงสุดของการร้องขอใช้งาน และขนาดข้อมูลถ่ายโอนได้แก่ข้อมูลรูปภาพและข้อมูลเท็กซ์ตามลำดับในขณะที่การร้องขอใช้งานแบบครั้งเดียวสำหรับข้อมูลรูปภาพและข้อมูลเท็กซ์มีประมาณครึ่งหนึ่ง (37.65:78.41 และ 8.70:17.86) ในทำนองเดียวกันกับขนาดข้อมูลถ่ายโอน (31.62:52.57 และ 17.41:27.37)

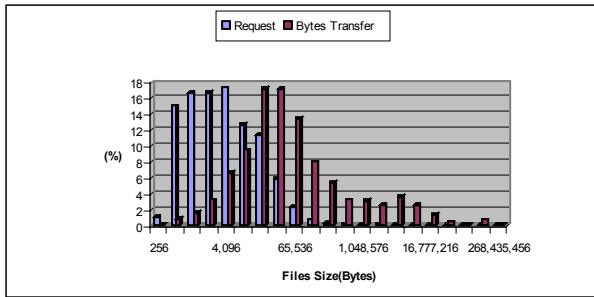
สำหรับข้อมูลวิดีโอพบว่าการร้องขอใช้งานและขนาดข้อมูลถ่ายโอนแบบครั้งเดียวใกล้เคียงกับการใช้งานทั้งหมด

(0.02:0.03 และ 0.98:1.05) ซึ่งแสดงให้เห็นว่าผู้ใช้งานมักร้องขอข้อมูลวิดีโอแบบครั้งเดียวทั้งสิ้นดังนั้นจึงไม่ควรแคชข้อมูลประเภทนี้เก็บไว้

เมื่อสังเกตถึงความสัมพันธ์ระหว่างค่ากลางและค่าเฉลี่ยเห็นได้ว่าไฟล์ข้อมูลเกือบทุกประเภทจะมีค่าแตกต่างกันค่อนข้างมาก (ค่ากลางน้อยกว่าค่าเฉลี่ยมาก) ซึ่งแสดงว่าช่วงของไฟล์ที่มีขนาดใหญ่และเล็กจะมีความแตกต่างกันมากหรือมีไฟล์ขนาดเล็กเกิดขึ้นจำนวนมากในทุกประเภทของไฟล์

5.1.2 ความสัมพันธ์ของการร้องขอใช้งานและขนาดข้อมูลถ่ายโอนเทียบกับขนาดไฟล์

รูปที่ 3 แสดงถึงช่วงขนาดไฟล์ข้อมูลที่มีเปอร์เซ็นต์การร้องขอใช้งานและขนาดข้อมูลถ่ายโอนโดยรวมมากที่สุด ซึ่งจะอยู่ในช่วง 512 ถึง 65,536 ไบต์ และ 2,048 ถึง 262,144 ไบต์ตามลำดับ

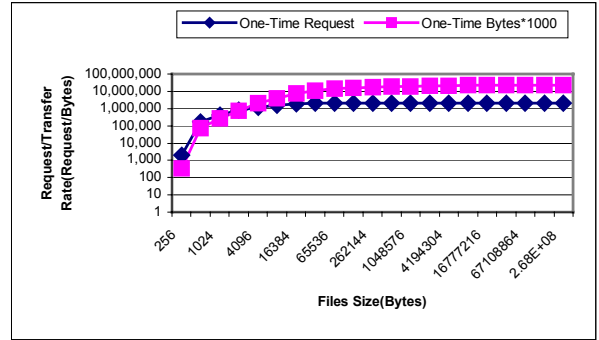


รูปที่ 3 : ความสัมพันธ์ของเปอร์เซ็นต์การร้องขอใช้งานและขนาดข้อมูลถ่ายโอนกับขนาดไฟล์

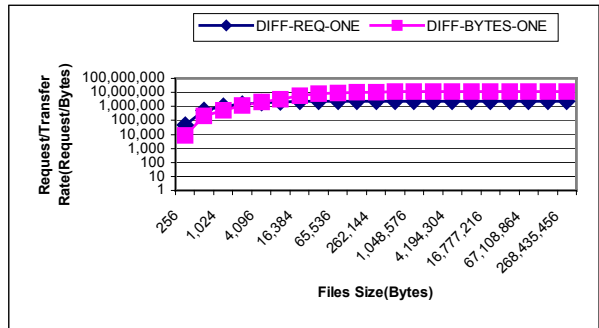
5.1.3 ความสัมพันธ์ของการใช้งานแบบครั้งเดียวกับขนาดไฟล์

การวิเคราะห์หัวข้อนี้พิจารณาการร้องขอใช้งานและขนาดข้อมูลถ่ายโอนแบบครั้งเดียวเท่านั้นเทียบกับขนาดไฟล์และนำข้อมูลที่ได้หาค่าผลต่างเปรียบเทียบกับกรร้องขอใช้งานและขนาดข้อมูลถ่ายโอนทั้งหมด

รูปที่ 4 แสดงถึงความสัมพันธ์ของการร้องขอใช้งานและขนาดข้อมูลถ่ายโอนแบบครั้งเดียวแบบสะสมพบว่าอัตราการเพิ่มขึ้นมีแนวโน้มคงที่เมื่อขนาดข้อมูลใหญ่ขึ้นและเมื่อนำค่าที่ได้จากกราฟเปรียบเทียบกับรูปที่ 1, 2 (แบบสะสม) หาผลต่างก็จะได้ผลลัพธ์ดังรูปที่ 5 ซึ่งแสดงถึงปริมาณการร้องขอใช้งานและขนาดข้อมูลถ่ายโอนที่แค่ทำให้บริการมากกว่าหนึ่งครั้งค่านี้เป็นค่าที่บ่งบอกถึงอัตราใช้งานซ้ำโดยอ้อมซึ่งพบว่าขนาดไฟล์ที่ประมาณ 64 และ 256 กิโลไบต์เป็นจุดที่มีอัตราการเพิ่มขึ้นลดลงหรือค่อนข้างคงที่ (เหมาะสมสำหรับขนาดแคชไฟล์)



รูปที่ 4 : การร้องขอใช้งานและขนาดข้อมูลถ่ายโอนแบบครั้งเดียวกับขนาดไฟล์สะสม



รูปที่ 5 : ผลต่างระหว่างการร้องขอใช้งาน ขนาดข้อมูลถ่ายโอนทั้งหมดและแบบครั้งเดียวกับขนาดไฟล์สะสม

5.1.4 ระยะเวลาการเปลี่ยนแปลงข้อมูล

การวิเคราะห์ในส่วนนี้ทำเพื่อตรวจหาว่าเว็บเพจที่แคชไว้มีการเปลี่ยนแปลงมากน้อยเพียงใด ตารางที่ 3 แสดงว่าข้อมูลที่กึ่งจะมีระยะเวลาการเปลี่ยนแปลงสั้นที่สุด (ต้นฉบับเว็บเพจเปลี่ยนแปลงบ่อย) ตามด้วยข้อมูลแอปพลิเคชัน ข้อมูลรูปภาพและข้อมูลเสียงตามลำดับส่วนข้อมูลวีดีโอพบว่าไม่มีการเปลี่ยนแปลงในคาบเวลา ผลที่ได้นี้สามารถนำไปใช้เป็นพารามิเตอร์เพื่อกำหนดนโยบายการกำหนดอายุของข้อมูลแต่ละประเภทของเว็บแคชเซิร์ฟเวอร์ได้

ตารางที่ 3 : ระยะเวลาการเปลี่ยนแปลงของข้อมูลแยกตามประเภทของข้อมูลไม่มีหน่วยโดยใช้ไรบอตเก็บค่าสถิติ

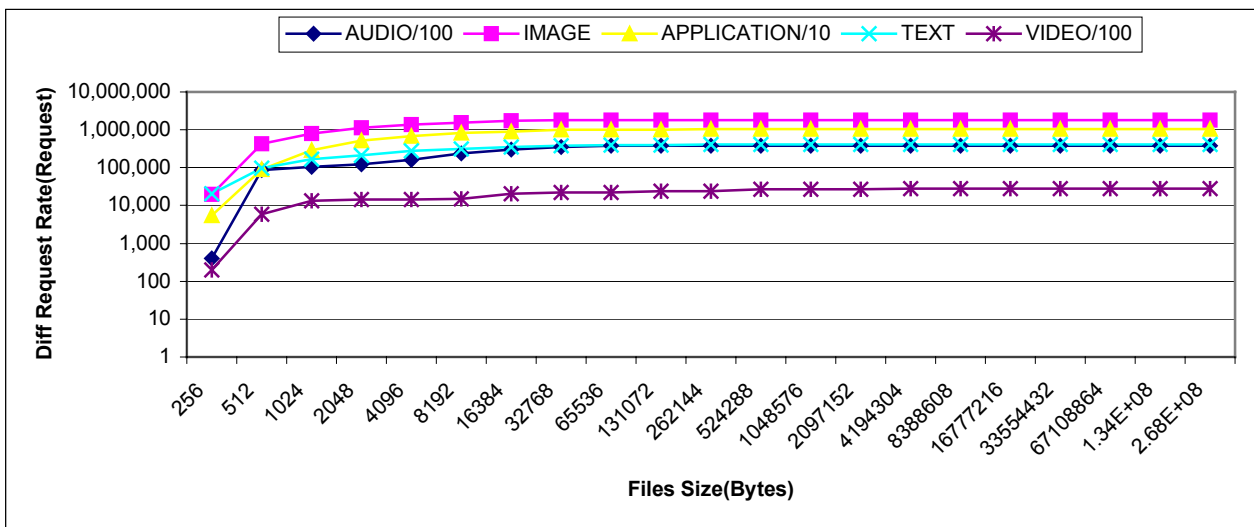
TYPE	AUDIO	IMAGE	APPLICATION	TEXT	VIDEO
MINUTES	930,190	309,032	41,276	30,811	-
DAY	645	214	28	24	-

5.2 การวิเคราะห์เปรียบเทียบความสัมพันธ์

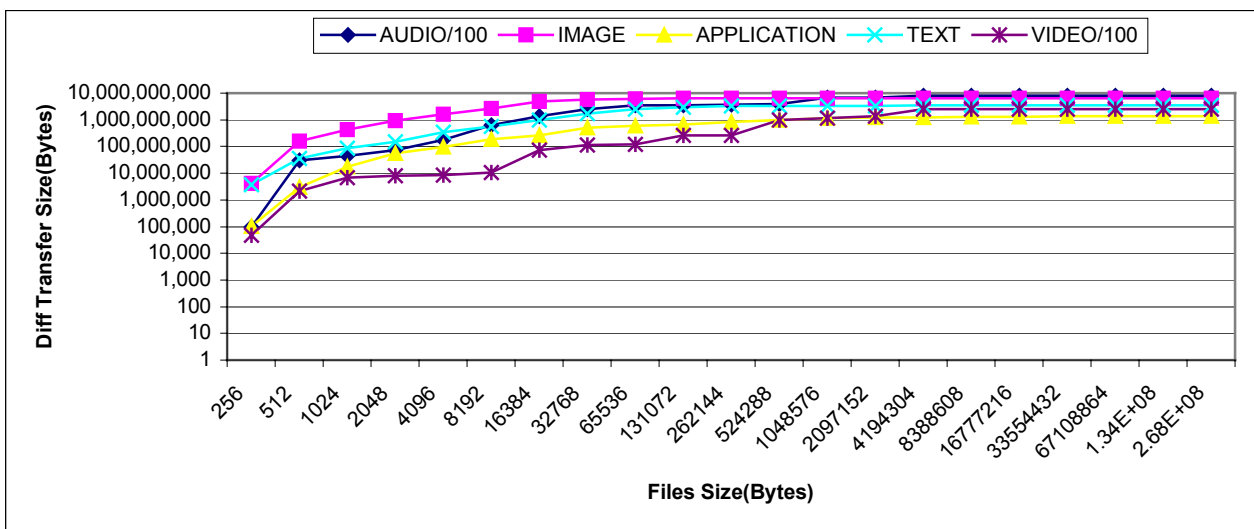
การวิเคราะห์หัวข้อนี้พิจารณาความสัมพันธ์ระหว่างปริมาณการร้องขอใช้งาน ขนาดข้อมูลถ่ายโอนโดยรวมและแบบครั้งเดียวกับขนาดไฟล์แบบสะสม โดยแยกตามประเภทของไฟล์และหาค่าผลต่างเพื่อหาช่วงขนาดไฟล์ที่มีผลกับพารามิเตอร์ที่วิเคราะห์

รูปที่ 6 และ 7 เมื่อพิจารณาผลต่างของการร้องขอใช้งานข้อมูลทั้งหมดและแบบครั้งเดียวพบว่าปริมาณการร้องขอ

ไฟล์ข้อมูลเท็กซ์กับไฟล์ข้อมูลรูปภาพโดยส่วนใหญ่จะเกิดขึ้นกับไฟล์ที่มีขนาดน้อยกว่า 128 กิโลไบต์ และ 64 กิโลไบต์ตามลำดับ สำหรับขนาดข้อมูลถ่ายโอนพบว่าทั้งไฟล์ข้อมูลเท็กซ์และไฟล์ข้อมูลรูปภาพโดยส่วนใหญ่จะเกิดขึ้นกับไฟล์ที่มีขนาดน้อยกว่า 256 กิโลไบต์ ซึ่งสามารถใช้เป็นพารามิเตอร์กำหนดอัตราการใช้งานซ้ำและอัตราขนาดข้อมูลถ่ายโอนการใช้งานซ้ำ (Hit Rate, Byte Hit Rate)



รูปที่ 6 : ผลต่างของการร้องขอใช้งานรวมและแบบครั้งเดียวแยกตามประเภทกับขนาดไฟล์แบบสะสม (สเกลลอการิทึม)



รูปที่ 7 : ผลต่างของขนาดข้อมูลถ่ายโอนรวมและแบบครั้งเดียวแยกตามประเภทกับขนาดไฟล์แบบสะสม (สเกลลอการิทึม)

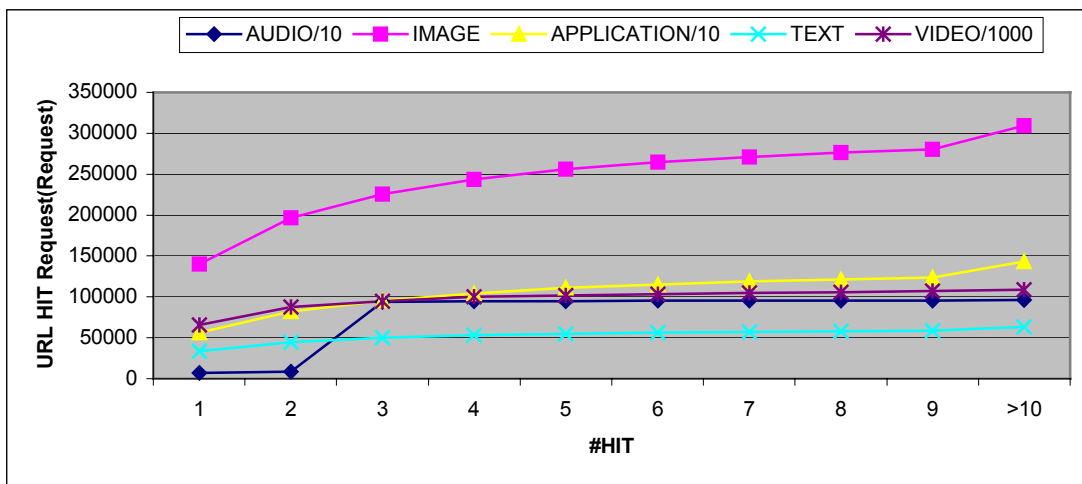
5.3 การวิเคราะห์การใช้งานซ้ำ

การวิเคราะห์หัวข้อนี้พิจารณาการใช้งานซ้ำของข้อมูล โดยจะวิเคราะห์การทำงานสองส่วน ส่วนแรกพิจารณาถึงจำนวนการใช้งานซ้ำของข้อมูลแบบสะสม ส่วนที่สองจะพิจารณาถึงค่าเฉลี่ยของระยะเวลาการใช้งานซ้ำของข้อมูลโดยแยกประเภทไฟล์ (ไม่วิเคราะห์ข้อมูลแยกตามขนาดเนื่องจากขนาดข้อมูลจะไม่คงที่เมื่อมีการเปลี่ยนแปลงหรือแก้ไขเพิ่มเติมข้อมูล)

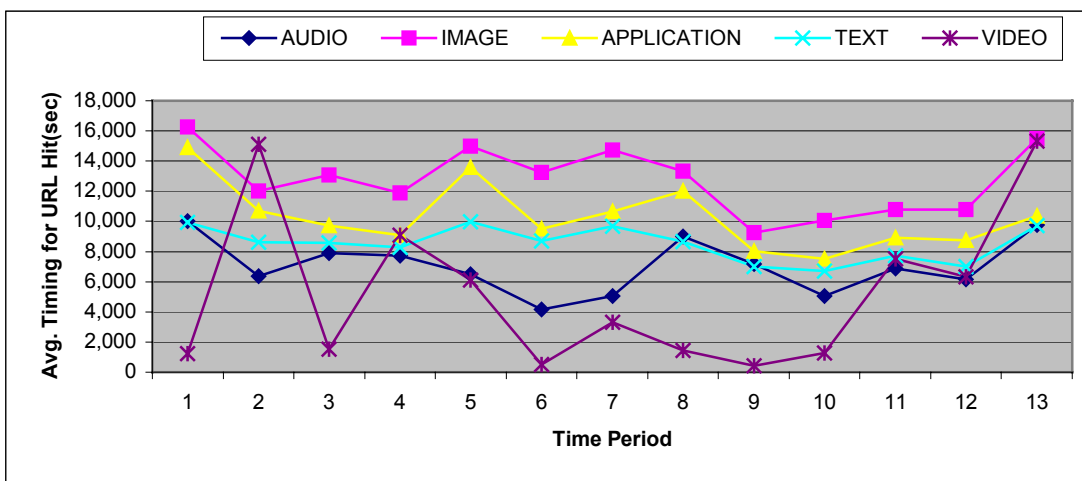
รูปที่ 8 แสดงถึงจำนวนการใช้งานซ้ำโดยเฉลี่ยพบว่า ข้อมูลรูปภาพมีการใช้งานซ้ำแบบสะสมเพิ่มขึ้นเป็นผลทำให้

การแคชข้อมูลประเภทนี้มีประสิทธิภาพสูงสุดและข้อมูลอื่นมีการใช้งานแบบซ้ำกันค่อนข้างน้อย

รูปที่ 9 และตารางที่ 4 แสดงถึงค่าเฉลี่ยของระยะเวลาการใช้งานซ้ำของข้อมูลพบว่าระยะเวลาการเปลี่ยนแปลงข้อมูลที่เท็กซ์และข้อมูลรูปภาพมีลักษณะเป็นคาบเวลาที่ส่วนข้อมูลอื่นมีการใช้งานที่ไม่แน่นอน (หมายเหตุ : เนื่องจากต้องการให้สามารถวาดกราฟได้ในรูปเดียวกันเพื่อแสดงถึงความแตกต่างของลักษณะไฟล์แต่ละประเภท ได้อย่างชัดเจน AUDIO/10 = ปริมาณไฟล์ประเภท AUDIO จะลดลง 10 เท่าของค่าที่ได้จากกราฟและไฟล์ประเภทอื่นในทำนองเดียวกัน)



รูปที่ 8 : จำนวนการใช้งานซ้ำเฉลี่ยของข้อมูลแบบสะสมแยกตามประเภทของไฟล์



รูปที่ 9 : ค่าเฉลี่ยของเวลาการร้องขอใช้งานข้อมูลซ้ำในคาบเวลาแยกตามประเภทของไฟล์

ตารางที่ 4 : ค่าเฉลี่ยรวมของเวลาการร้องขอใช้งานข้อมูลซ้ำแยกตามประเภทของข้อมูลไม่มีไฟป์

TYPE	AUDIO	IMAGE	APPLICATION	TEXT	VIDEO
MINUTES	117.6345	212.7092	166.8776	136.7763	88.7189

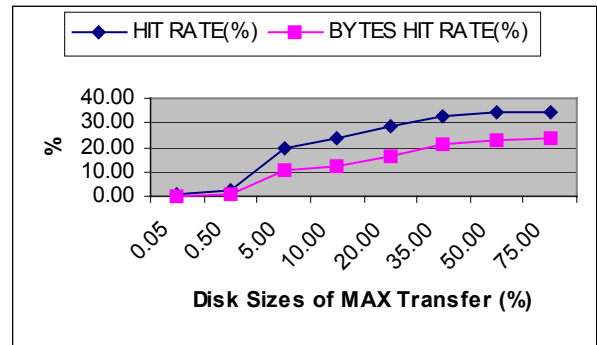
6. การทดลองเพื่อทดสอบสมรรถนะการทำงาน

เพื่อให้สามารถนำผลการวิเคราะห์ไปใช้ปรับแต่งเว็บแคชเซิร์ฟเวอร์ได้ จึงได้วางแผนการทดลองเป็นสองส่วนคือการหาขนาดดิสก์ที่เหมาะสมและขนาดแคชไฟล์ใหญ่สุดที่เหมาะสมสำหรับการใช้งานจริงโดยเปรียบเทียบกับขนาดไฟล์ที่ได้จากผลการวิเคราะห์ในหัวข้อที่ 5

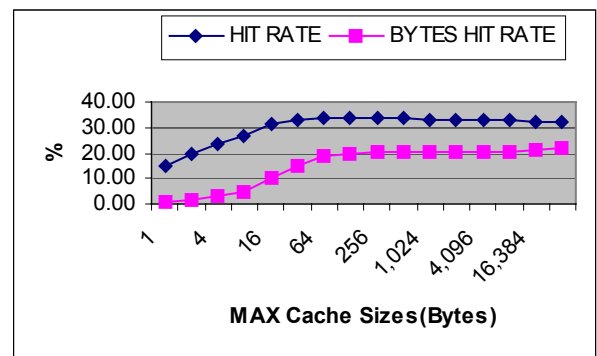
ข้อมูลทดสอบมีค่าผลรวมปริมาณการร้องขอใช้งานรวม 72,674 ครั้งและขนาดข้อมูลถ่ายโอนรวม 667,255,345 ไบต์ ทั้งนี้การทดลองจะใช้นโยบายการทดแทนข้อมูลแบบ LRU เนื่องจากเป็นนโยบายที่มีมาตรฐานและไม่มีความซับซ้อนของการทำงาน

การทดลองส่วนแรกใช้วิธีปรับขนาดดิสก์แคชจากน้อยไปหามากและวัดเปอร์เซ็นต์การใช้งานซ้ำ ผลการทดลองที่แสดงในรูปที่ 10 พบว่าเมื่อเพิ่มขนาดดิสก์ถึงที่ประมาณ 35-50% ส่งผลให้อัตราการใช้งานซ้ำมีแนวโน้มเพิ่มขึ้นอย่างลดลงหรือค่อนข้างคงที่ (ผลที่ได้เสมือนกับผลจากการวิเคราะห์การทำงานเฉลี่ยในช่วงคาบเวลาที่ใช้วิเคราะห์ 13 วัน) ดังนั้นจึงสรุปได้ว่าปริมาณดิสก์ที่สามารถเก็บข้อมูลโดยไม่ทำให้อัตราการใช้งานซ้ำลดลงจะอยู่ที่ค่าประมาณ 35-50% ของขนาดข้อมูลถ่ายโอนทั้งหมด

ส่วนที่สองจะใช้ผลลัพธ์จากการทดลองข้างต้นคือ ค่าขนาดดิสก์ที่ 35% ทดลองโดยปรับค่าขนาดแคชต่อหน่วยที่มากที่สุดและใช้นโยบายการทดแทนข้อมูลแบบ LRU ผลที่ได้ปรากฏว่าขนาดแคชที่ทำให้้อัตราการใช้งานซ้ำสูงคือ 64 กิโลไบต์ซึ่งตรงกับผลการวิเคราะห์การทำงานในช่วงคาบเวลาของการวิเคราะห์ดังรูปที่ 11



รูปที่ 10 : ความสัมพันธ์ระหว่างขนาดดิสก์กับอัตราการใช้งานซ้ำโดยใช้นโยบายการทดแทนข้อมูลแบบ LRU



รูปที่ 11 : ความสัมพันธ์ระหว่างขนาดไฟล์ที่แคชกับอัตราการใช้งานซ้ำที่ขนาดดิสก์ 35% โดยใช้นโยบายการทดแทนข้อมูลแบบ LRU

7. สรุป

ผลสรุปของงานวิจัยได้นำเสนอผลการวิเคราะห์ปริมาณและพฤติกรรมการใช้งานข้อมูลบนเครือข่ายที่ถูกใช้งานจริงของหน่วยงานซึ่งสามารถสรุปและสามารถนำไปเป็นแนวทางการปรับแต่งเว็บแคชเซิร์ฟเวอร์ทั้งด้านการจำกัดขนาดข้อมูล ลำดับความสำคัญของการแคชไฟล์ข้อมูล ขนาดแคชไฟล์และระยะเวลาการเปลี่ยนแปลงของไฟล์ทั้งนี้พารามิเตอร์ต่างๆสามารถปรับเปลี่ยนได้ตามความต้องการและความ

เหมาะสมของหน่วยงานนั้นๆ ทั้งนี้จะขึ้นอยู่กับนโยบายที่กำหนดไว้ซึ่งจากผลการวิเคราะห์สามารถสรุปได้ดังนี้

- ปริมาณการร้องขอใช้งานข้อมูลเท็กซ็มีจำนวน 17% ข้อมูลรูปภาพ 78% และขนาดข้อมูลถ่ายโอนข้อมูลเท็กซ็มีจำนวน 27% ข้อมูลรูปภาพ 52% เมื่อเทียบกับปริมาณการร้องขอใช้งานและขนาดข้อมูลถ่ายโอนทั้งหมด ดังนั้นจึงควรพิจารณาข้อมูลทั้งสองประเภทเป็นลำดับแรก
- ข้อมูลเท็กซ็มีการร้องขอใช้งานแบบครั้งเดียว 10% ข้อมูลรูปภาพ 37% เมื่อเทียบกับการร้องขอใช้งานข้อมูลรวมแต่ละประเภทแสดงว่าข้อมูลทั้งสองประเภทมีอัตราการร้องขอใช้งานแบบครั้งเดียวเป็นสัดส่วนครึ่งหนึ่ง
- การร้องขอใช้งานช่วงขนาดไฟล์ 512 ถึง 65,536 ไบต์มีจำนวน 97% และช่วงขนาดไฟล์ 2 ถึง 256 กิโลไบต์มี 77% ซึ่งจะส่งผลกระทบต่อโยบายการเลือกเก็บขนาดข้อมูลเมื่อขนาดดิสก์มีจำกัด
- จากรูปที่ 6 และ 10 พบว่าขนาดไฟล์ที่ทำให้อัตราการเพิ่มของการซ้ำของข้อมูลคงที่ของการแคชข้อมูลโดยไม่แยกประเภทของไฟล์คือ 64 กิโลไบต์
- เมื่อพิจารณาปริมาณข้อมูลแยกตามขนาดไฟล์พบว่าปริมาณการร้องขอไฟล์ข้อมูลเท็กซ็กับไฟล์ข้อมูลรูปภาพโดยส่วนใหญ่จะเกิดขึ้นกับไฟล์ที่มีขนาดน้อยกว่า 128 กิโลไบต์ และ 64 กิโลไบต์ตามลำดับ สำหรับขนาดข้อมูลถ่ายโอนพบว่าทั้งไฟล์ข้อมูลเท็กซ็และไฟล์ข้อมูลรูปภาพโดยส่วนใหญ่จะเกิดขึ้นกับไฟล์ที่มีขนาดน้อยกว่า 256 กิโลไบต์
- ขนาดดิสก์ที่เหมาะสมที่ทำให้อัตราการใช้งานซ้ำเพิ่มขึ้นค่อนข้างคงที่อยู่ที่ประมาณ 35-50% ของขนาดข้อมูลถ่ายโอนทั้งหมดแต่ทั้งนี้ขึ้นอยู่กับปริมาณดิสก์ที่มีอยู่เนื่องจากอัตราการใช้งานซ้ำสามารถเพิ่มขึ้นในอัตราลดลงหรือค่อนข้างคงที่กรณีมีดิสก์ไม่จำกัด
- ข้อมูลเท็กซ็และข้อมูลรูปภาพจะมีระยะเวลาการร้องขอใช้งานซ้ำค่อนข้างคงที่และระยะเวลาการเปลี่ยนแปลงของ

ข้อมูลบนเครือข่าย ข้อมูลเท็กซ็จะมีระยะเวลาการเปลี่ยนสั้นที่สุดเปรียบเทียบกับข้อมูลอื่นแต่ทั้งนี้ขึ้นอยู่กับประเภทของเว็บไซต์ด้วย

การวิเคราะห์เบื้องต้นเมื่อพิจารณาเฉพาะส่วนของปริมาณข้อมูลไม่มีไทป์เมื่อเปรียบเทียบกับช่วงเวลาการทำงานที่ต่างๆกันแต่มีนโยบายการทำงานเพื่อบริการของหน่วยงานที่เหมือนกัน (ในที่นี้เปรียบเทียบกับช่วงเวลาเปิดภาคการศึกษาทั้งภาคต้นเดือนกุมภาพันธ์และภาคปลายเดือนกรกฎาคม) พบว่าปริมาณข้อมูลจะมีสัดส่วนใกล้เคียงกัน ซึ่งผลที่ได้จากการวิเคราะห์ในช่วงเวลานี้เป็นผลทำให้ผลการวิเคราะห์ในคาบเวลาอื่นควรจะมีผลใกล้เคียงกันด้วย [8] ในทำนองเดียวเมื่อพิจารณาจากข้อมูลเบื้องต้นในหน่วยงานระดับเดียวกัน (ในที่นี้ใช้หน่วยงานการศึกษาคือ มหาวิทยาลัยเกษตรศาสตร์และสถาบันพระจอมเกล้าเจ้าคุณทหารลาดกระบัง) พบว่าปริมาณข้อมูลไม่มีไทป์จะมีสัดส่วนที่ใกล้เคียงกันเช่นกัน [9]

การวิเคราะห์จำกัดอยู่ที่การแยกแยะส่วนของข้อมูลไม่มีไทป์เท่านั้น นอกจากนี้ข้อมูลที่ใช้มีลักษณะคงที่ไม่มีลักษณะไดนามิกและยังไม่รวมถึงข้อกำหนดอื่นนอกเหนือขอบเขตการเลือกเก็บข้อมูลเช่น โครงสร้างการวางแคชตามลำดับชั้น การจัดสรรนโยบายการใช้งานเว็บแคชเซิร์ฟเวอร์ของหน่วยงานซึ่งมีความแตกต่างกัน ทำให้ผลการวิเคราะห์ไม่สามารถนำไปใช้ปรับแต่งโดยตรงกับเว็บแคชเซิร์ฟเวอร์ที่มีนโยบายการทำงานแตกต่างออกไปจากการวิเคราะห์นี้

การวิเคราะห์ข้อมูลแสดงถึงลักษณะการใช้งานเครือข่ายของหน่วยงานซึ่งจะมีผลต่อการกำหนดพารามิเตอร์สำหรับการใช้งานเว็บแคชเซิร์ฟเวอร์โดยไม่ต้องเพิ่มประสิทธิภาพระบบส่วนอื่นและสามารถนำไปประยุกต์ใช้กับระบบอัตโนมัติได้ งานวิจัยยังขยายผลศึกษาให้ครอบคลุมถึงส่วนของนโยบายการจัดตั้งแคชตามสภาพภูมิศาสตร์ การวางแคชตามลำดับชั้นและยังนำไปใช้สำหรับการวิเคราะห์การทำงานของข้อมูลทุกประเภทบนเครือข่ายเพื่อปรับปรุงพัฒนาระบบคุณภาพของข้อมูลบนเครือข่ายได้

8. เอกสารอ้างอิง

- [1] G.Abdulla, E.A.Fox, M.Abrams and S.Williams, *WWW Proxy Traffic Characterization with Application to Caching*, TR-97-03, Computer Science Dept, Virginia Tech, March 1997.
- [2] G.Barish and K.Obraczke, *World Wide Web Caching: Trends and Techniques*, IEEE Communication Magazine, 38(5):178--184, May 2000.
- [3] C.R.Carlos, A.Bestavros and Mark E.Crovella, *Characteristics of WWW Client based Traces*, BU-CS-95-010 Boston University, July 1995.
- [4] A.Mahanti, C.Williamson and D.Eager, *Traffic Analysis of a Web Proxy Caching Hierarchy*. IEEE Network Magazine, 14(3):16-23, May/June 2000.
- [5] I.Marshall and C.Roadknight, *Linking Cache Performance to User Behavior*, *Proceedings of the 3rd International WWW Caching Workshop*, Manchester England 1998.
- [6] K.Thompson, G.J.Miller and R.Wilder, *Wide-Area Internet Traffic Patterns and Characteristics*, IEEE/ACM Transactions on Networking, pp. 10--23, November 1997
- [7] D.Wessels, A.Rousskov and G.Chisholm, *The First IRCACHE Web Caching Bake-off*, *In Proceedings of 1999 The Fourth Web Caching Workshop*, San Diego, California , March/April 1999.
- [8] Department of Network and Communication Computer Center, *Web Cache Statistic at Kasetsart University Report*, February/July 2001.
Available
http://logs.ku.ac.th/cachelog/cache1_rpt/2001-Feb,
http://logs.ku.ac.th/cachelog/cache1_rpt/2001-Jul/index.html
- [9] Department of Network and Communication Computer Center, *Web Cache Statistic at Kasetsart University and King Mongkut's Institute of Technology Ladkrabang Report*, February 2001.
Available
<http://logs.ku.ac.th/cgi-bin/perl-app/cache1-stat.pl>
<http://www.kmitl.ac.th/~netcommu/proxystat/Chaokhun/index.html>