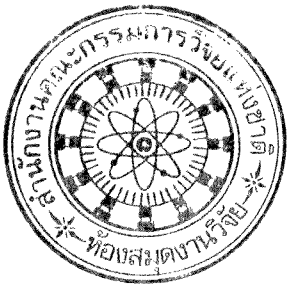


บทที่ 4

ผลการวิเคราะห์ข้อมูล

การเสนอผลการวิเคราะห์ข้อมูลประสิทธิภาพของนโยบายแทนที่ข้อมูลทั้งสี่จะเสนอโดยใช้กราฟประกอบการวิเคราะห์เพื่อตอบคำถามการวิจัยของงานวิจัยนี้ทั้งสองข้อตามลำดับ ส่วนตารางข้อมูลเชิงตัวเลขจะนำเสนอในภาคผนวก ข และเพื่อให้สามารถตีความวิเคราะห์ข้อมูลได้อย่างถูกต้องจำเป็นที่จะต้องนำเสนอผลการวิเคราะห์คุณลักษณะของข้อมูลนำเข้าแต่ละชุด และผลการวิเคราะห์ค่าที่ดีที่สุดของตัวแปรต้น Window size ของนโยบายการแทนที่ข้อมูลที่เสนอก่อนที่จะนำเสนอผลการวิเคราะห์เชิงเปรียบเทียบผลลัพธ์ประสิทธิภาพของนโยบายแทนที่ข้อมูลทั้งสี่

คุณลักษณะของข้อมูลนำเข้า

คุณลักษณะเบื้องต้นของข้อมูลนำเข้าที่ใช้ในการจำลองการทำงานของนโยบายการแทนที่ข้อมูลทั้งสี่ที่วิเคราะห์ด้วยโปรแกรมสนับสนุนการจำลองการทำงาน แสดงในตารางที่ 1 ทั้งนี้ขนาดข้อมูลรวมสูงสุดของข้อมูลที่ไม่ซ้ำกันถูกใช้เป็นขนาดแคช (Cache size) ขั้นต่ำที่ไม่ทำให้เกิด Cache miss เพราะพื้นที่จัดเก็บข้อมูลในแคชเพียงพอสำหรับการเก็บข้อมูลที่ไม่ซ้ำกันที่ถูกร้องขอทั้งหมด

ตาราง 1 คุณลักษณะเบื้องต้นของข้อมูลนำเข้าแต่ละชุด

คุณลักษณะ	ชุดข้อมูล		
	BO	SV	NY
จำนวนคำร้องขอข้อมูล (เรคคอร์ด)	205,226	441,084	599,097
ขนาดข้อมูลที่ถูกร้องขอทั้งหมด (ไบต์)	2,401,517,003	7,113,486,583	4,712,041,132
จำนวนชิ้นข้อมูลที่ไม่ซ้ำกัน	70,944	248,508	158,552
ขนาดข้อมูลรวมสูงสุดของข้อมูลที่ไม่ซ้ำกัน	694,759,006	1,065,863,067	1,323,954,264

ค่าตัวแปรต้น Window sizes ที่ดีที่สุด

ประสิทธิภาพการทำงานของนโยบายแทนที่ข้อมูล Cloud ที่เสนอส่วนหนึ่งขึ้นอยู่กับตัวแปรต้น Window size ซึ่งถูกกำหนดค่าที่ดีที่สุดซึ่งได้จากการทดลองจำลองการทำงานที่ค่า Window sizes ต่างๆ กันด้วยชุดข้อมูลจริง พบว่าค่าที่ดีที่สุดเป็นไปตามตารางที่ 2 ค่าในวงเล็บคือค่า Window sizes สัมพัทธ์ในแนวดิ่งและแนวนอนของตารางตามลำดับ

ตาราง 2 ค่า Window size ที่ดีที่สุดที่ใช้ในการจำลองการทำงานของนโยบาย Cloud ที่เสนอ

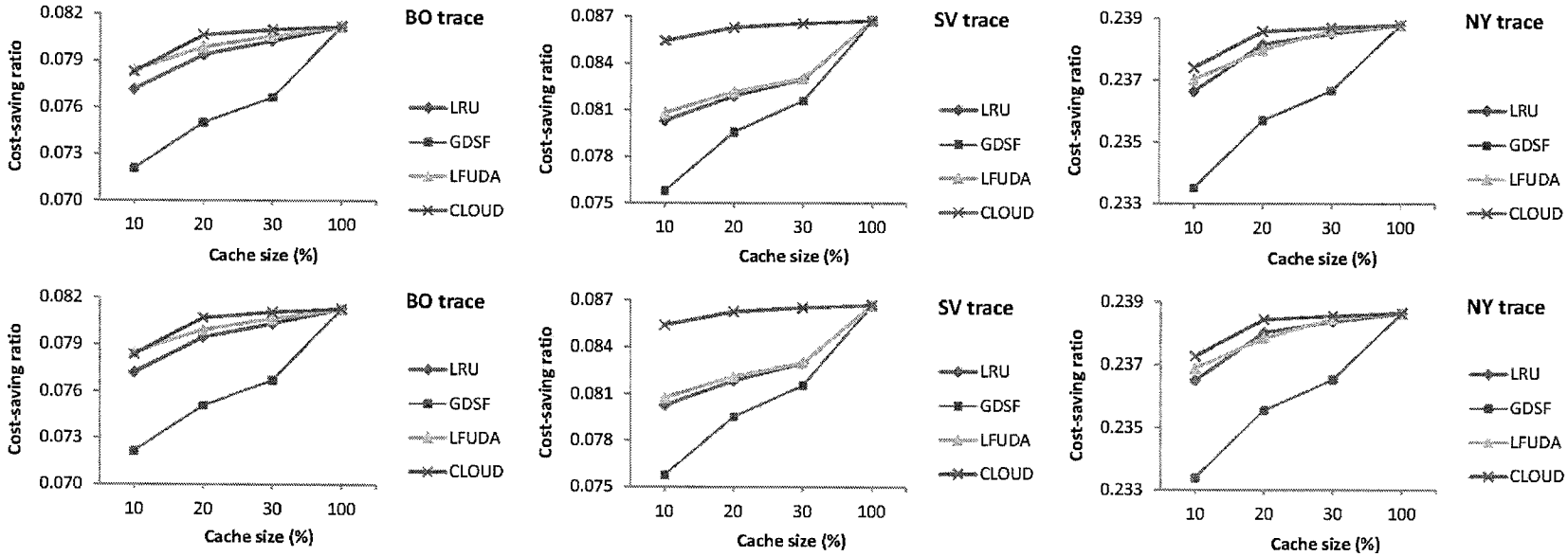
ชุดข้อมูล	ขนาดแคช (ร้อยละของขนาดแคชขั้นต่ำที่ไม่ทำให้เกิด Cache miss)			ขนาดแคชสัมพัทธ์
	10	20	30	
BO	215 (1.00,1.00)	625 (1.00,2.91)	675 (1.00,3.14)	1.00
SV	425 (1.98,1.00)	1175 (1.88,2.76)	1200 (1.78,2.82)	1.53
NY	700 (3.26,1.00)	1550 (2.48,2.21)	1575 (2.33,2.25)	1.91

ประสิทธิภาพเชิงเปรียบเทียบ

ผลลัพธ์ประสิทธิภาพของนโยบายแทนที่ข้อมูลต่างๆ ถูกแสดงเป็นกลุ่มตามลำดับหน่วยวัดสมรรถนะทางเศรษฐศาสตร์และทางเทคนิค คือ อัตราประหยัดค่าโหลดข้อมูล (ภาพประกอบ 1), อัตราประหยัดเวลาโหลดข้อมูล (ภาพประกอบ 2), อัตราขนาดรวมของข้อมูลที่พบในแคช (ภาพประกอบ 3) และอัตราการพบข้อมูลในแคช (ภาพประกอบ 4) โดยในแต่ละกลุ่มจะแบ่งแสดงเป็นสองแตามชุดข้อมูลที่ใช้ในการจำลองการทำงานและแบ่งตามแบบจำลองอัตราค่าโหลดข้อมูล

จากภาพประกอบ 1 จะพบว่า

1. นโยบาย Cloud ที่เสนอมีอัตราประหยัดค่าโหลดข้อมูลสูงสุด ยกเว้นในชุดข้อมูล BO ที่ขนาดแคช 10% พบว่า LFU-DA มีอัตราประหยัดค่าโหลดข้อมูลสูงกว่า Cloud เล็กน้อยคือ 0.14% แสดงว่านโยบายการแทนที่ข้อมูลในเว็บแคชที่พิจารณาปัจจัยขนาดของข้อมูลที่ร้องขอ, อัตราค่าโหลดข้อมูล, สถิติจำนวนครั้งที่ข้อมูลถูกร้องขอ, สถิติระยะเวลาที่ข้อมูลถูกร้องขอครั้งล่าสุด, อายุ



ภาพประกอบ 1 ผลการเปรียบเทียบอัตราประหยัดค่าโหลดข้อมูลแบบอัตราค่าโหลดข้อมูลอัตราเดียวในกราฟแถวบน และแบบอัตราค่าโหลดข้อมูลหลายอัตราในกราฟแถวล่าง

ใช้งานคงเหลือของข้อมูล และสถิติระยะเวลาของการโหลดข้อมูล สามารถประหยัดค่าใช้จ่ายในการโหลดข้อมูลจากกลุ่มเมฆได้มากกว่านโยบาย LRU, GDSF และ LFU-DA ในเกือบทุกกรณี

2. อัตราประหยัดค่าโหลดข้อมูลของนโยบาย LRU อยู่ในระดับใกล้เคียงกับของนโยบาย LFU-DA ส่วนนโยบาย GDSF ให้อัตราประหยัดค่าโหลดข้อมูลต่ำสุด ทั้งนี้เพราะ GDSF เลือกแทนที่ข้อมูลขนาดใหญ่ก่อน

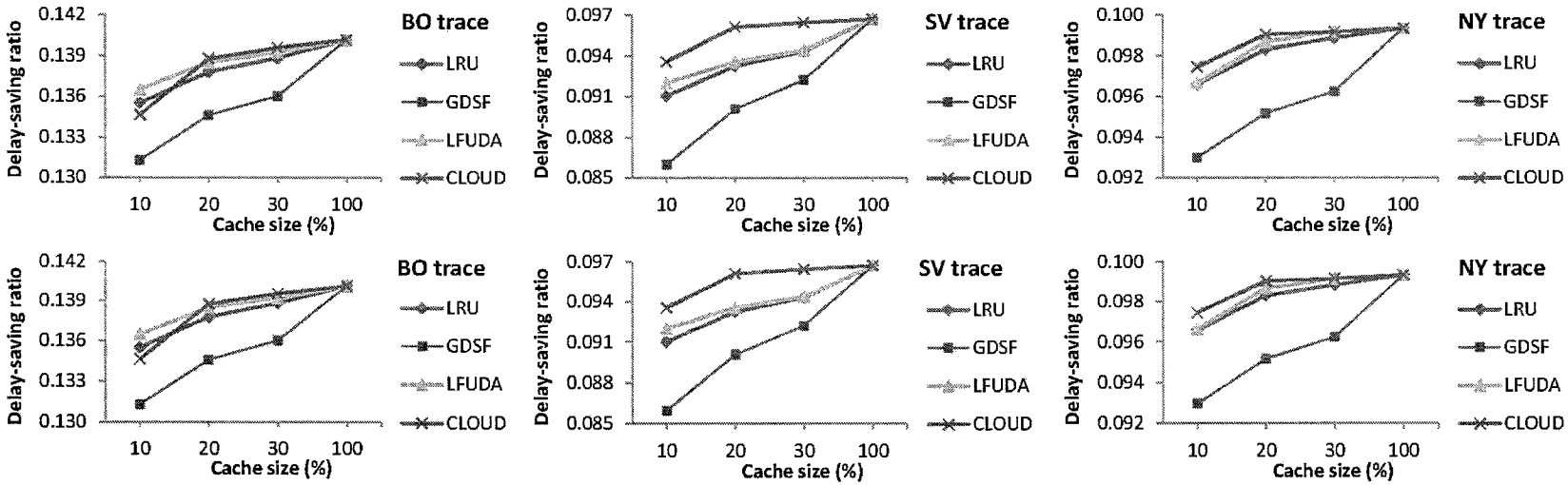
3. ทั้งในแบบจำลองอัตราค่าโหลดข้อมูลแบบอัตราเดียวและหลายอัตรา เมื่อขนาดแคชมีค่าเท่ากับหรือมากกว่า 20% นโยบาย Cloud จะให้อัตราประหยัดค่าโหลดข้อมูลค่อนข้างคงที่ที่สุดเมื่อเปรียบเทียบกับอีกสามนโยบาย

4. อัตราประหยัดค่าโหลดข้อมูลของแบบจำลองอัตราค่าโหลดข้อมูลทั้งสองแบบจำลองแตกต่างกันน้อยมาก ซึ่งจากข้อมูลเชิงตัวเลขทำให้ทราบว่าอัตราประหยัดค่าโหลดข้อมูลในแบบจำลองอัตราค่าโหลดข้อมูลแบบอัตราเดียวสามารถสูง (ในกรณีชุดข้อมูล SV และ NY) หรือต่ำกว่าในแบบจำลองอัตราค่าโหลดข้อมูลแบบหลายอัตรา (ในกรณีชุดข้อมูล BO) ดังแสดงในภาคผนวก ข

จากภาพประกอบ 2 จะพบว่า

1. ที่ขนาดแคชตั้งแต่ 20% ขึ้นไป นโยบาย Cloud ลดระยะเวลาตอบสนองโดยรวมต่อผู้ใช้ได้มากที่สุด แสดงว่านโยบายการแทนที่ข้อมูลในเว็บแคชที่พิจารณาปัจจัยขนาดของข้อมูลที่ร้องขอ, อัตราค่าโหลดข้อมูล, สถิติจำนวนครั้งที่ข้อมูลถูกร้องขอ, สถิติระยะเวลาที่ข้อมูลถูกร้องขอครั้งล่าสุด, อายุใช้งานคงเหลือของข้อมูล และสถิติระยะเวลาของการโหลดข้อมูล จะสามารถลดระยะเวลาการตอบสนองผู้ใช้บริการซอฟต์แวร์กลุ่มเมฆได้มากกว่านโยบาย LRU, GDSF และ LFU-DA

2. อัตราประหยัดเวลาโหลดข้อมูลของนโยบาย LRU อยู่ในระดับใกล้เคียงกับของนโยบาย LFU-DA ส่วนนโยบาย GDSF ให้อัตราประหยัดเวลาโหลดข้อมูลต่ำสุด ส่วนหนึ่งเป็นเพราะ GDSF เลือกแทนที่ข้อมูลขนาดใหญ่ก่อนซึ่งปกติข้อมูลขนาดใหญ่จะใช้เวลาโหลดนานกว่าข้อมูลขนาดเล็ก



ภาพประกอบ 2 ผลการเปรียบเทียบอัตราประหยัดเวลาโหลดข้อมูลแบบอัตราค่าโหลดข้อมูลอัตราเดียวในกราฟแถวบน และแบบอัตราค่าโหลดข้อมูลหลายอัตราในกราฟแถวล่าง

3. ทั้งในแบบจำลองอัตราค่าโหลดข้อมูลแบบอัตราเดียวและหลายอัตรา เมื่อขนาดแคชมีค่าเท่ากับหรือมากกว่า 20% นโยบาย Cloud จะให้อัตราประหยัดเวลาโหลดข้อมูลค่อนข้างคงที่ที่สุดเมื่อเปรียบเทียบกับอีกสามนโยบาย

4. อัตราประหยัดเวลาโหลดข้อมูลของทั้งสองแบบจำลองอัตราค่าโหลดข้อมูลแตกต่างกันน้อยมากดังแสดงในภาคผนวก ข อย่างไรก็ตาม การทดลองเพิ่มเติมทำให้ทราบว่าการเพิ่มอัตราค่าโหลดข้อมูลสูงสุดของแบบจำลองอัตราค่าโหลดข้อมูลแบบหลายอัตรา จะทำให้อัตราประหยัดเวลาโหลดข้อมูลในแบบจำลองอัตราค่าโหลดข้อมูลทั้งสองจะแตกต่างกันมากยิ่งขึ้น ข้อมูลเชิงตัวเลขของการทดลองนำเสนออยู่ในตอนท้ายของบทนี้

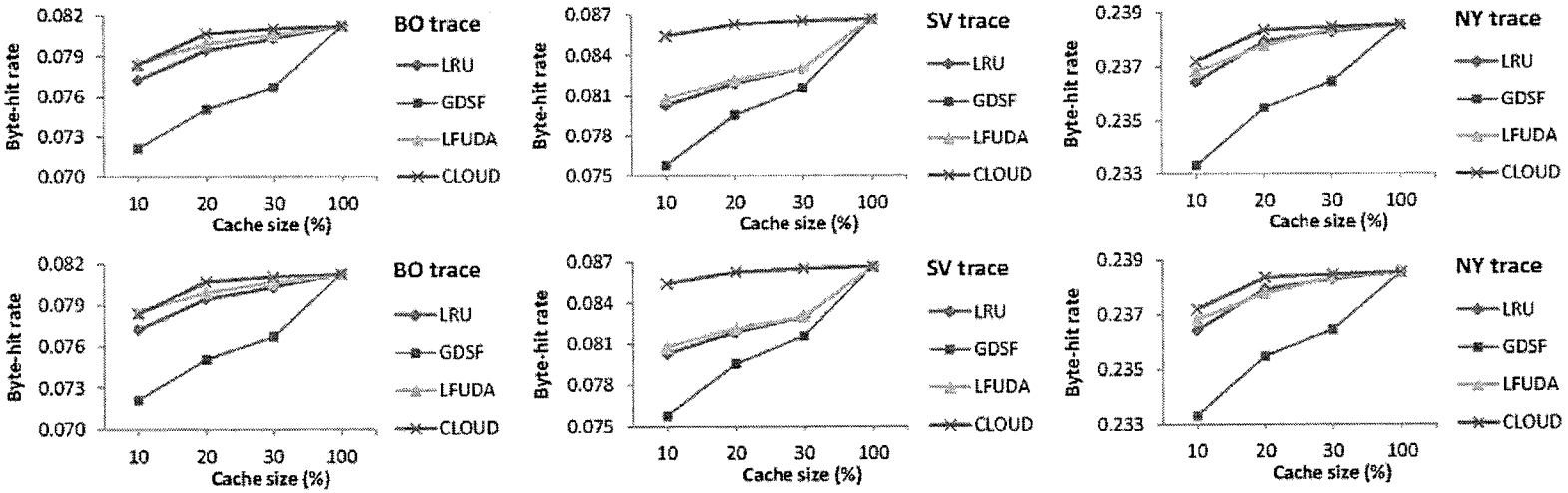
นอกเหนือจากผลการวิเคราะห์ตามวัตถุประสงค์ทั้งสอง งานวิจัยในด้านนี้ยังนิยมวิเคราะห์ประสิทธิภาพตามหน่วยวัดสมรรถนะอีกสองหน่วยดังนำเสนอในสองภาพประกอบต่อไปนี้

จากภาพประกอบ 3 จะพบว่า

1. ที่ขนาดแคชตั้งแต่ 20% ขึ้นไป นโยบาย Cloud ลดขนาดรวมของข้อมูลที่ต้องโหลดใหม่จากกลุ่มเมฆได้มากที่สุด ยกเว้นที่ขนาดแคช 10% ที่นโยบาย LFU-DA ลดขนาดรวมข้อมูลที่ต้องโหลดใหม่ได้มากกว่า Cloud ประมาณ 0.14% เหตุผลคือนโยบาย Cloud พยายามไม่แทนที่ข้อมูลขนาดใหญ่ก่อน

2. อัตราขนาดรวมของข้อมูลที่พบในแคชของนโยบาย LRU อยู่ในระดับใกล้เคียงกับของนโยบาย LFU-DA ส่วนนโยบาย GDSF ให้อัตราขนาดรวมของข้อมูลที่พบในแคชต่ำสุด เพราะ GDSF เลือกแทนที่ข้อมูลขนาดใหญ่ก่อน

3. ทั้งในแบบจำลองอัตราค่าโหลดข้อมูลแบบอัตราเดียวและหลายอัตรา เมื่อขนาดแคชมีค่าเท่ากับหรือมากกว่า 20% นโยบาย Cloud จะให้อัตราขนาดรวมของข้อมูลที่พบในแคชค่อนข้างคงที่ที่สุดเมื่อเปรียบเทียบกับอีกสามนโยบาย



ภาพประกอบ 3 ผลการเปรียบเทียบอัตราขนาดรวมของข้อมูลที่พบในแคชแบบอัตราค่าโหลดข้อมูลอัตราเดียวในกราฟแถวบน และแบบอัตราค่าโหลดข้อมูลหลายอัตราในกราฟแถวล่าง

4. ไม่สามารถเห็นความแตกต่างระหว่างอัตราขนาดรวมของข้อมูลที่พบในแคชของทั้งสองแบบจำลองอัตราค่าโหลดข้อมูลแตกต่างกันได้ดังแสดงในชุดตัวเลขผลลัพธ์ประสิทธิภาพในภาคผนวก ข ทั้งนี้การทดลองเพิ่มเติมทำให้ทราบว่าการเพิ่มอัตราค่าโหลดข้อมูลสูงสุดของแบบจำลองอัตราค่าโหลดข้อมูลแบบหลายอัตรา จะทำให้อัตราขนาดรวมของข้อมูลที่พบในแคชในแบบจำลองอัตราค่าโหลดข้อมูลทั้งสองจะแตกต่างกันมากยิ่งขึ้น ข้อมูลเชิงตัวเลขของการทดลองนำเสนออยู่ในตอนท้ายของบทนี้

จากภาพประกอบ 4 จะพบว่า

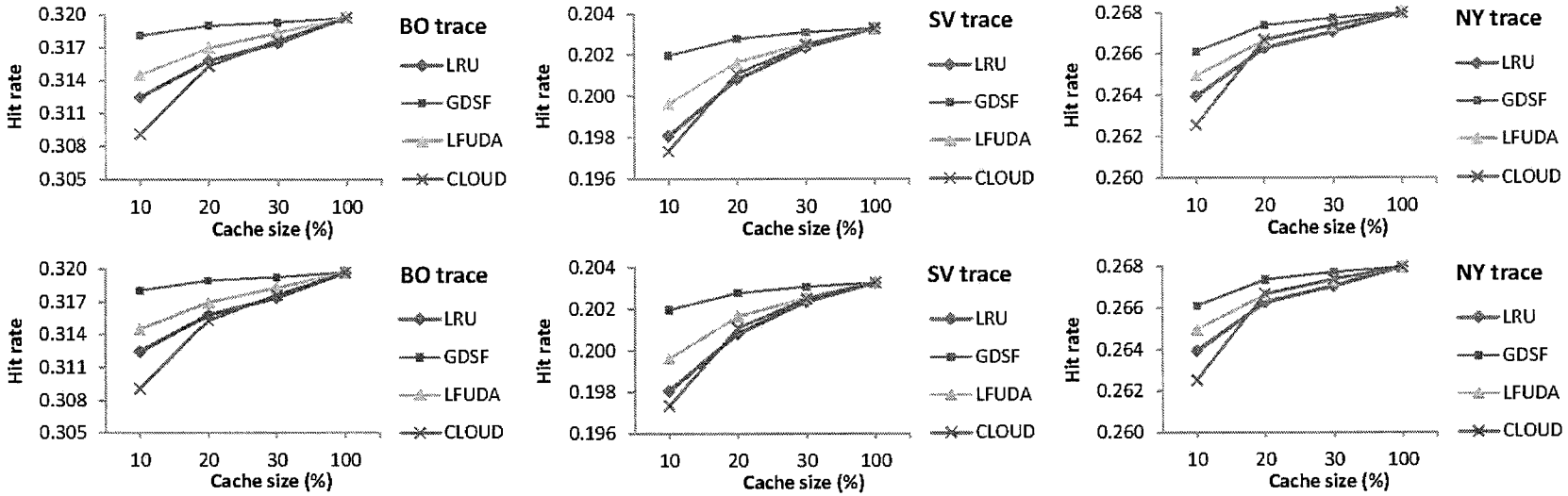
1. ที่ขนาดแคช 10% นโยบาย Cloud มีอัตราการพบข้อมูลในแคชต่ำสุด ที่ขนาดแคชอื่นนโยบาย Cloud มีอัตราการพบข้อมูลในแคชดีกว่า LRU เป็นส่วนใหญ่และดีกว่า LFU-DA ในบางกรณี ปรากฏการณ์นี้สอดคล้องกับข้อเท็จจริงที่ค้นพบโดย Podlipnig and Böszörmenyi (2003), Williams et al. (1996) และ Balamash and Krunz (2004) ว่าการเลือกแทนที่ข้อมูลขนาดใหญ่ก่อนจะทำให้อัตราการพบข้อมูลในแคชสูง แต่นโยบาย Cloud ที่เสนอมีอัตราการพบข้อมูลในแคชต่ำเพราะเลือกแทนที่ข้อมูลขนาดเล็กก่อน

2. นโยบาย GDSF ให้อัตราการพบข้อมูลในแคชสูงสุดแม้ว่าจะมีประสิทธิภาพภายใต้ตัววัดสมรรถนะอื่นต่ำสุด เหตุผลคือนโยบาย GDSF เลือกแทนที่ข้อมูลขนาดใหญ่เพื่อให้มีพื้นที่จัดเก็บข้อมูลขนาดเล็กจำนวนมากจึงเพิ่มโอกาสการพบข้อมูลในแคช

3. ไม่สามารถพบความแตกต่างของอัตราการพบข้อมูลในแคชของทั้งสองแบบจำลองได้ดังแสดงในภาคผนวก ข ทั้งนี้การทดลองเพิ่มเติมทำให้ทราบว่าการเพิ่มอัตราค่าโหลดข้อมูลสูงสุดของแบบจำลองอัตราค่าโหลดข้อมูลแบบหลายอัตรา จะทำให้อัตราการพบข้อมูลในแคชในแบบจำลองอัตราค่าโหลดข้อมูลทั้งสองจะแตกต่างกันมากยิ่งขึ้น ข้อมูลเชิงตัวเลขของการทดลองถูกนำเสนอในย่อหน้าถัดไป

ผลการวิเคราะห์เชิงบูรณาการของหน่วยวัดสมรรถนะทั้งสี่ในภาพประกอบ 1 ถึง 4 เป็นดังต่อไปนี้

1. ประสิทธิภาพโดยรวมของแต่ละนโยบายแปรเปลี่ยนไปขึ้นอยู่กับชุดข้อมูลหรือพฤติกรรมการบริโภคข้อมูลของผู้ใช้



ภาพประกอบ 4 ผลการเปรียบเทียบอัตราการพบข้อมูลในแคชแบบอัตราค่าโหลดข้อมูลอัตราเดียวในกราฟแถวบน และแบบอัตราค่าโหลดข้อมูลหลายอัตราในกราฟแถวล่าง

2. เนื่องจากค่าโหลดข้อมูลแปรผันตรงกับขนาดข้อมูล ดังนั้นอัตราประหยัดค่าโหลดข้อมูลจึงปรากฏในการทดลองว่ามีอัตราเติบโตเท่ากับอัตราขนาดรวมของข้อมูลที่พบในแคช ดังนั้นจึงสามารถประหยัดค่าโหลดข้อมูลและแบนด์วิดท์ของระบบเครือข่ายได้พร้อมกันที่อัตราเดียวกันไม่ว่าจะใช้นโยบายแทนที่ข้อมูลใดๆ

3. ภาพประกอบที่ 1, 3 และ 4 พบว่านโยบายที่ให้อัตราประหยัดค่าโหลดข้อมูล และอัตราขนาดรวมของข้อมูลที่พบในแคชสูง จะให้อัตราการพบข้อมูลในแคชต่ำ ปรากฏการณ์นี้เป็นหลักฐานอีกชิ้นหนึ่งที่สนับสนุนข้อค้นพบที่ว่านโยบายแทนที่ข้อมูลที่เลือกแทนที่ข้อมูลขนาดใหญ่ก่อนจะมีประสิทธิภาพอัตราการพบข้อมูลในแคชสูง แต่อัตราขนาดรวมของข้อมูลที่พบในแคชรวมถึงอัตราประหยัดค่าโหลดข้อมูลจะต่ำ

4. ความแตกต่างของประสิทธิภาพทั้งสี่ของตัววัดสมรรถนะระหว่างแบบจำลองอัตราค่าโหลดข้อมูลทั้งสองแบบจำลองจะเห็นได้ชัดเจนขึ้นเมื่ออัตราค่าโหลดข้อมูลทั้งสองที่ใช้ในแบบจำลองอัตราค่าโหลดข้อมูลแบบหลายอัตรามีความแตกต่างกันมากขึ้น เช่น การทดลองเพิ่มเติมโดยใช้อัตราค่าโหลดข้อมูล 0.117997 และ 1.17997 ดอลล่าร์ (แทน 0.117997 และ 0.15 ดอลล่าร์) ที่ขนาดแคช 20% ด้วยชุดข้อมูล NY ให้ผลลัพธ์อัตราประหยัดค่าโหลดข้อมูล, อัตราประหยัดเวลาโหลดข้อมูล, อัตราขนาดรวมของข้อมูลที่พบในแคช และอัตราการพบข้อมูลในแคชของแบบจำลองอัตราค่าโหลดข้อมูลแบบหลายอัตราต่ำกว่าของแบบจำลองอัตราค่าโหลดข้อมูลแบบอัตราเดียว 0.00026097, 0.00000348, -0.00000057 และ 0.00000334 ตามลำดับ

5. ทั้งในแบบจำลองอัตราค่าโหลดข้อมูลแบบอัตราเดียวและหลายอัตรา เมื่อขนาดแคชมีค่าเท่ากับหรือมากกว่า 20% นโยบาย Cloud จะให้อัตราประหยัดค่าโหลดข้อมูล, อัตราประหยัดเวลาโหลดข้อมูล และอัตราขนาดรวมของข้อมูลที่พบในแคชค่อนข้างคงที่ที่สุดเมื่อเปรียบเทียบกับอีกสามนโยบาย