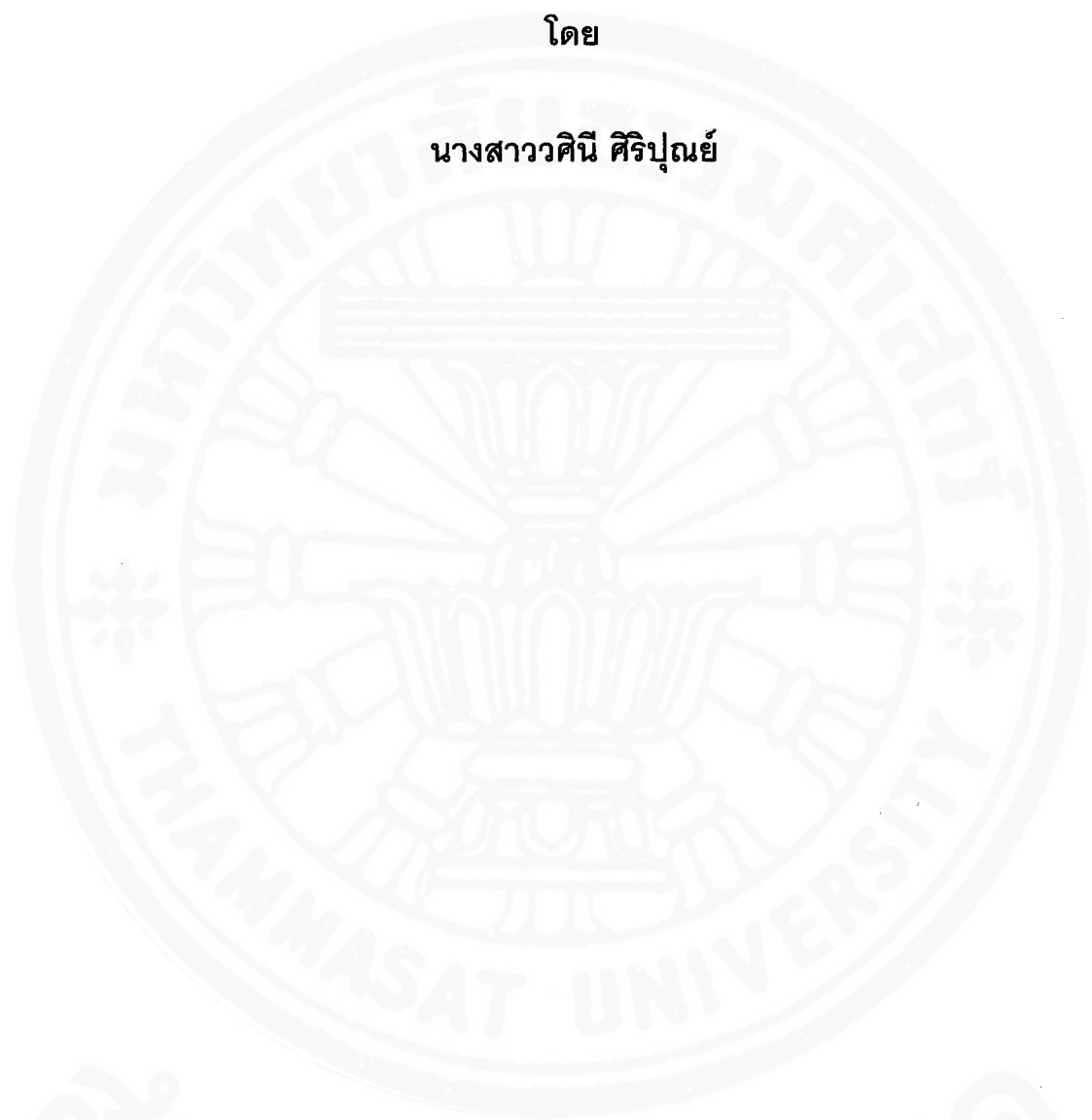


ไลฟ์เช็คพอยน์แบบเทรดบนเวอร์ชวลแมชชีน

โดย

นางสาวศินี ศิริบุญย์



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

พ.ศ. 2552

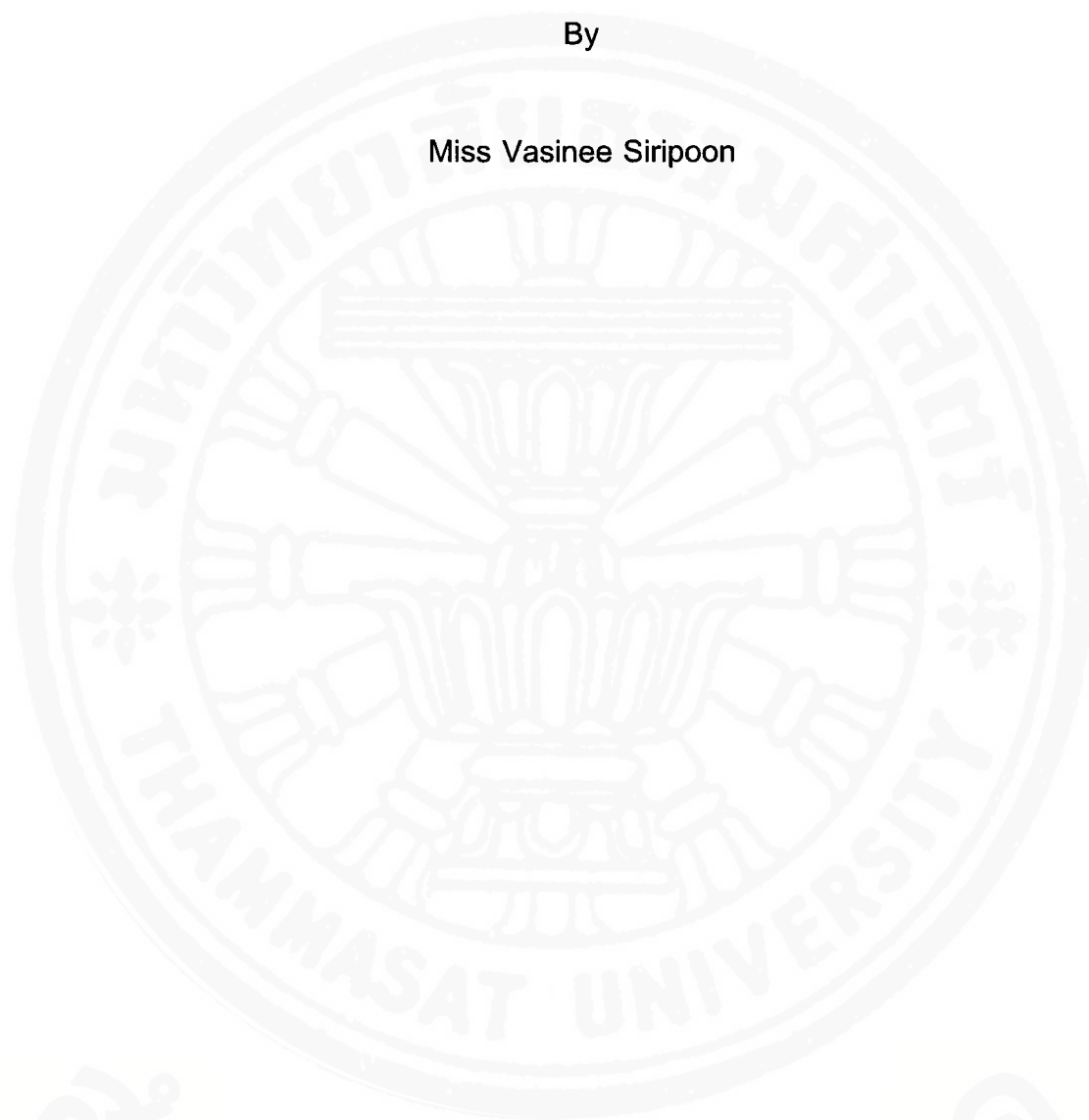
อธิปัทนการ

จาก

Thread-based Live Checkpointing of Virtual Machines

By

Miss Vasinee Siripoon



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science

Department of Computer Science

Faculty of Science and Technology

Thammasat University

2009

มหาวิทยาลัยธรรมศาสตร์
คณะวิทยาศาสตร์และเทคโนโลยี

วิทยานิพนธ์

ของ

นางสาวศินี ศิริบุญ

เรื่อง

ไลฟ์เชื้อพอยน์แบบเทอร์คบนเวอร์ชวลแมชชีน

ได้รับการตรวจสอบและอนุมัติ ให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต

เมื่อ วันที่ 1 พฤษภาคม พ.ศ. 2553

ประธานกรรมการสอบวิทยานิพนธ์

เด่นดวง ประดับสุภาณ

(ดร.เด่นดวง ประดับสุวรรณ)

กรรมการและอาจารย์ที่ปรึกษาวิทยานิพนธ์

กษิติศ ชาญชัย

(ดร.กษิติศ ชาญชัย)

กรรมการสอบวิทยานิพนธ์

รองศาสตราจารย์ ดร.เยาวดี เต็มธนาภัทร์

กรรมการสอบวิทยานิพนธ์

ดร.ศรเทพ วรรณรัตน์

คณบดี

รองศาสตราจารย์ สายทอง อมรวิเศษฐ์

บทคัดย่อ

ซิสเต็มส์เวอร์ชวลแมชชีนช่วยให้ฮาร์ดแวร์แพลตฟอร์มของโฮสเครื่องหนึ่งสามารถสนับสนุนเกสโอเอสหลายระบบได้ในเวลาเดียวกัน ด้วยเทคโนโลยีเวอร์ชวลแมชชีน ผู้ใช้สามารถรันระบบปฏิบัติการที่แตกต่างกันได้บนฮาร์ดแวร์เดียวกัน เครื่องมือที่สำคัญอย่างหนึ่งของเวอร์ชวลแมชชีนมอนิเตอร์คือไลฟ์ไมเกรชัน ซึ่งเป็นความสามารถในการย้ายเวอร์ชวลแมชชีนจากเครื่องหนึ่งไปอีกเครื่องหนึ่งได้โดยเกือบจะไม่ต้องหยุดเวอร์ชวลแมชชีน ไลฟ์ไมเกรชันในเควีเอ็ม (KVM) สามารถใช้ในการทำเช็คพอยน์เวอร์ชวลแมชชีนได้โดยการเก็บสถานะของเวอร์ชวลแมชชีนลงในไฟล์ปกติ โปรโตคอลในการทำเช็คพอยน์นี้จำเป็นที่จะต้องหยุดเวอร์ชวลแมชชีนในระหว่างที่ทำเช็คพอยน์ ซึ่งอาจมีผลต่อประสิทธิภาพการทำงานของแอปพลิเคชัน ในการทดลองเบื้องต้นพบว่าการแก้ไขโปรโตคอลการเช็คพอยน์เดิมให้ทำงานโดยที่ไม่ต้องหยุดเวอร์ชวลแมชชีนสามารถช่วยปรับปรุงประสิทธิภาพการทำงานของแอปพลิเคชันได้ อย่างไรก็ตาม ข้อเสียที่สำคัญของวิธีการนี้คือในกรณีที่แอปพลิเคชันมีความต้องการใช้หน่วยความจำและการเขียนหน่วยความจำมาก จะทำให้การเช็คพอยน์ใช้เวลามากขึ้นและประสิทธิภาพการทำงานของแอปพลิเคชันลดลง งานวิจัยนี้จึงเสนอวิธีการไลฟ์เช็คพอยน์แบบเทรด ซึ่งหมายถึงการเช็คพอยน์เวอร์ชวลแมชชีนโดยใช้เทรดในเทคนิคไลฟ์ไมเกรชัน ด้วยวิธีการนี้ เทรดจะทำเช็คพอยน์ในขณะที่เวอร์ชวลแมชชีนกำลังทำงาน ทำให้เวอร์ชวลแมชชีนโปรเซสสามารถทำงานต่อไปได้โดยไม่ต้องทำเช็คพอยน์จนกว่าจะถึงเวลาที่ต้องหยุดเพื่อบันทึกสถานะ การอิมพลีเมนต์วิธีการนี้ทำในเควีเอ็ม และการประเมินประสิทธิภาพใช้โปรแกรมจาก NPB ผลการทดลองพบว่าโปรโตคอลไลฟ์เช็คพอยน์แบบเทรดสามารถแก้ปัญหาของกลไกการเช็คพอยน์ที่ใช้ไลฟ์ไมเกรชันโดยตรงได้ ซึ่งวิธีการนี้จะทำให้โอเวอร์เฮดลดลง แอปพลิเคชันทำงานเสร็จเร็วขึ้น ในขณะที่เวลาที่เครื่องหยุดทำงานยังคงน้อยมากเช่นเดิม

คำสำคัญ: เช็คพอยน์, เวอร์ชวลแมชชีน, ไลฟ์เช็คพอยน์, ไลฟ์ไมเกรชัน

Abstract

A system virtual machine (VM) provides a single-host hardware platform that supports multiple guest operating system environments concurrently. By using VM technology, users could run different operating systems on the same hardware. A key feature of virtual machine monitor is live migration, which is the ability to transport a VM from one host to another without noticeable guest downtime. KVM live migration supports the VM checkpointing capability by storing the VM state in a regular file. Its checkpointing protocol requires a VM to be stopped during the entire process of checkpointing that may affect the application performance. In the preliminary study, a slightly modified version of live-migration checkpointing protocol that allows a VM to continue its execution during the checkpoint time can improve the application performance. However, an important drawback to this implementation is that the delayed checkpoint latency induced by large memory requirement and high memory access can degrade performance. In this study, a novel thread-based live checkpointing is proposed. This model leverages live migration mechanism introducing a thread, which is responsible for the majority of the checkpoint activities. This enables a VM process to continue normal execution without performing checkpoint operations until it reaches the final stage of checkpointing. This approach is implemented in KVM and its performance evaluations are conducted using the NAS parallel benchmarks. Experiments show that thread-based live checkpoint protocol can lift out the restrictions inherited in the live-migration-based checkpointing mechanism. A notable improvement of execution time and lower overhead can be found whereas its downtime is negligible.

Keywords: Checkpoint, Virtual Machine, Live Checkpointing, Live Migration

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณ ดร.กษิตศ ชาญเขียว อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้ความรู้ และคำแนะนำเป็นอย่างดี ทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี

ขอขอบคุณ ดร.เด่นดวง ประดับสุวรรณ, รศ.ดร. เยวดี เต็มธนาภักดิ์ และ ดร. ศรเทพ วรรณรัตน์ กรรมการสอบวิทยานิพนธ์ ที่ได้ให้คำแนะนำ และขอขอบคุณภาควิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยธรรมศาสตร์ ที่เชื้อเพื่อสถานที่และวัสดุอุปกรณ์สำหรับการทำวิจัย

วคินี ศิริบุญ

มหาวิทยาลัยธรรมศาสตร์

พ.ศ. 2553

สำนักหอสมุด

สารบัญ

	หน้า
บทคัดย่อ.....	(2)
กิตติกรรมประกาศ.....	(4)
สารบัญตาราง.....	(7)
สารบัญภาพประกอบ.....	(10)
บทที่	
1. บทนำ.....	1
ความเป็นมาและความสำคัญของปัญหา.....	1
วัตถุประสงค์การวิจัย.....	4
สมมุติฐานการวิจัย.....	4
ขอบเขตของการวิจัย.....	4
ประโยชน์ที่คาดว่าจะได้รับ.....	5
นิยามศัพท์เฉพาะ.....	5
2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	6
ประเภทของเวอร์ชวลไลเซชัน.....	6
เวอร์ชวลแมชชีน (Virtual Machine).....	9
สถาปัตยกรรมของเวอร์ชวลแมชชีน.....	10
เวอร์ชวลแมชชีนมอนิเตอร์ (Virtual Machine Monitor).....	13
การอิมพลีเมนต์ VMM.....	14
เวอร์ชวลไลเซชันซอฟต์แวร์และการทำไลฟ์ไมเกรชัน.....	16

งานวิจัยที่เกี่ยวข้องกับการเช็คพอยน์เวอร์ซอลแมชชีนด้วยวิธีไลฟ์ไมเกรซัน.....	29
3. วิธีการวิจัย.....	33
การเช็คพอยน์เวอร์ซอลแมชชีนในเควีเอ็ม.....	35
ไลฟ์เช็คพอยน์โปรโตคอล (Live checkpoint protocol).....	37
เทรดเบสไลฟ์เช็คพอยน์โปรโตคอล (Thread-based live checkpoint protocol).....	38
การสร้างไฟล์สถานะของเวอร์ซอลแมชชีน.....	49
ดิสก์อิมเมจ.....	50
การรีสตาร์ทเวอร์ซอลแมชชีน.....	50
วิธีการทดลอง.....	51
การวิเคราะห์ข้อมูล.....	56
การทดลองเบื้องต้น.....	56
4. ผลการวิจัย.....	66
5. สรุปผลการวิจัยและข้อเสนอแนะ.....	79
รายการอ้างอิง.....	89
ภาคผนวก.....	93
ก. ข้อมูลเกี่ยวกับการคัดลอกหน่วยความจำ.....	94
ข. ข้อมูลเกี่ยวกับเวลาการทำงานของเวอร์ซอลแมชชีนและเทรด.....	96
ประวัติการศึกษา.....	98

สารบัญตาราง

ตารางที่	หน้า
3.1 NAS Parallel Benchmarks คลาส A.....	53
3.2 NAS Parallel Benchmarks คลาส B.....	53
3.3 ขนาดข้อมูล, ขนาด ADC, ขนาดผลลัพธ์ และรายละเอียดอื่นๆของเบนช์มาร์ก DC	54
3.4 การใช้หน่วยความจำของเบนช์มาร์กที่ใช้ในการทดลองและขนาด active memory ของเวอร์ชวลแมชชีน.....	54
3.5 การใช้หน่วยความจำของเบนช์มาร์กที่ใช้ในการทดลองเบื้องต้นและขนาด active memory ของเวอร์ชวลแมชชีน.....	57
3.6 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของเวลาการทำงาน เวลาเช็คพอยน์ และ เวลาที่เวอร์ชวลแมชชีนหยุดทำงาน ของเบนช์มาร์ก CG คลาส B สำหรับ การทดลองเบื้องต้น.....	57
3.7 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของเวลาการทำงาน เวลาเช็คพอยน์ และ เวลาที่เวอร์ชวลแมชชีนหยุดทำงาน ของเบนช์มาร์ก LU คลาส A สำหรับ การทดลองเบื้องต้น.....	58
3.8 การทดสอบความเท่ากันของความแปรปรวนของเวลาการทำงานและ เวลาเช็คพอยน์ของเบนช์มาร์ก CG คลาส B และ LU คลาส A ระหว่างวิธีการ เช็คพอยน์ที่แตกต่างกัน.....	59
3.9 การทดสอบความแตกต่างของเวลาการทำงานและเวลาเช็คพอยน์ของเบนช์มาร์ก CG คลาส B ระหว่างวิธีการเช็คพอยน์ที่แตกต่างกันด้วย F-test.....	59
3.10 การทดสอบความแตกต่างของเวลาการทำงานและเวลาเช็คพอยน์ของเบนช์มาร์ก CG คลาส B ระหว่างวิธีการเช็คพอยน์ที่แตกต่างกันเป็นรายคู่ด้วย Scheffe.....	60
3.11 การทดสอบความแตกต่างของเวลาการทำงานและเวลาเช็คพอยน์ของเบนช์มาร์ก LU คลาส A ระหว่างวิธีการเช็คพอยน์ที่แตกต่างกันด้วย F-test.....	61
3.12 การทดสอบความแตกต่างของเวลาการทำงานและเวลาเช็คพอยน์ของเบนช์มาร์ก LU คลาส A ระหว่างวิธีการเช็คพอยน์ที่แตกต่างกันเป็นรายคู่ด้วย Scheffe.....	61
3.13 ค่าเฉลี่ยของขนาดของไฟล์สถานะเวอร์ชวลแมชชีนที่รันเบนช์มาร์ก CG คลาส B และ LU คลาส A.....	62

3.14	โอเวอร์เฮดของวิธีการเช็คพอยน์ต์แต่ละแบบในการทดลองเบื้องต้น.....	63
4.1	ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของเวลาการทำงาน เวลาเช็คพอยน์ต์ และเวลาที่เวอร์ชวลแมชชีนหยุดทำงาน ของเบนซ์มาร์ก EP คลาส B.....	66
4.2	ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของเวลาการทำงาน เวลาเช็คพอยน์ต์ และเวลาที่เวอร์ชวลแมชชีนหยุดทำงาน ของเบนซ์มาร์ก CG คลาส B.....	67
4.3	ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของเวลาการทำงาน เวลาเช็คพอยน์ต์ และเวลาที่เวอร์ชวลแมชชีนหยุดทำงาน ของเบนซ์มาร์ก BT คลาส A.....	67
4.4	ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของเวลาการทำงาน เวลาเช็คพอยน์ต์ และเวลาที่เวอร์ชวลแมชชีนหยุดทำงาน ของเบนซ์มาร์ก DC คลาส W.....	68
4.5	การทดสอบความเท่ากันของความแปรปรวนของเวลาการทำงานและเวลาเช็คพอยน์ต์ของเบนซ์มาร์กที่ใช้ในการทดลอง ระหว่างวิธีการเช็คพอยน์ต์ที่แตกต่างกัน.....	69
4.6	การทดสอบความแตกต่างของเวลาการทำงานและเวลาเช็คพอยน์ต์ของเบนซ์มาร์ก EP คลาส B ระหว่างวิธีการเช็คพอยน์ต์ที่แตกต่างกันด้วย F-test.....	69
4.7	การทดสอบความแตกต่างของเวลาการทำงานและเวลาเช็คพอยน์ต์ของเบนซ์มาร์ก EP คลาส B ระหว่างวิธีการเช็คพอยน์ต์ที่แตกต่างกันเป็นรายคู่ด้วย Scheffe.....	70
4.8	การทดสอบความแตกต่างของเวลาการทำงานและเวลาเช็คพอยน์ต์ของเบนซ์มาร์ก CG คลาส B ระหว่างวิธีการเช็คพอยน์ต์ที่แตกต่างกันด้วย F-test.....	71
4.9	การทดสอบความแตกต่างของเวลาการทำงานและเวลาเช็คพอยน์ต์ของเบนซ์มาร์ก CG คลาส B ระหว่างวิธีการเช็คพอยน์ต์ที่แตกต่างกันเป็นรายคู่ด้วย Scheffe.....	71
4.10	การทดสอบความแตกต่างของเวลาการทำงานและเวลาเช็คพอยน์ต์ของเบนซ์มาร์ก BT คลาส A ระหว่างวิธีการเช็คพอยน์ต์ที่แตกต่างกันด้วย F-test.....	72
4.11	การทดสอบความแตกต่างของเวลาการทำงานและเวลาเช็คพอยน์ต์ของเบนซ์มาร์ก BT คลาส A ระหว่างวิธีการเช็คพอยน์ต์ที่แตกต่างกันเป็นรายคู่ด้วย Scheffe.....	73
4.12	การทดสอบความแตกต่างของเวลาการทำงานและเวลาเช็คพอยน์ต์ของเบนซ์มาร์ก DC คลาส W ระหว่างวิธีการเช็คพอยน์ต์ที่แตกต่างกันด้วย F-test.....	74
4.13	การทดสอบความแตกต่างของเวลาการทำงานและเวลาเช็คพอยน์ต์ของเบนซ์มาร์ก DC คลาส W ระหว่างวิธีการเช็คพอยน์ต์ที่แตกต่างกันเป็นรายคู่ด้วย Scheffe.....	74
4.14	ขนาดของไฟล์สถานะและเวลาที่ใช้ในการรีสตาร์ทเวอร์ชวลแมชชีน.....	75

4.15	ขนาดโอเวอร์เฮดของแต่ละวิธีการเช็คพอยน์ต์.....	76
ก.1	ข้อมูลการคัดลอกหน่วยความจำของแต่ละวิธีการเช็คพอยน์บนเบนซ์มาร์ก EP คลาส B.....	94
ก.2	ข้อมูลการคัดลอกหน่วยความจำของแต่ละวิธีการเช็คพอยน์บนเบนซ์มาร์ก CG คลาส B.....	95
ก.3	ข้อมูลการคัดลอกหน่วยความจำของแต่ละวิธีการเช็คพอยน์บนเบนซ์มาร์ก BT คลาส A.....	95
ก.4	ข้อมูลการคัดลอกหน่วยความจำของแต่ละวิธีการเช็คพอยน์บนเบนซ์มาร์ก DC คลาส W.....	95
ข.1	เวลาการทำงานของเวอร์ชวลแมชีน.....	96
ข.2	เวลาการทำงานของเช็คพอยน์ท์ของเทรค.....	97

สำนักหอสมุด

สารบัญภาพประกอบ

ภาพที่		หน้า
1.1	เวอร์ชวลแมชชีนมอนิเตอร์ (Virtual Machine Monitor หรือ VMM).....	2
2.1	ฮาร์ดแวร์อิมูเลชัน.....	7
2.2	พูลเวอร์ชวลไลเซชัน.....	8
2.3	พาราเวอร์ชวลไลเซชัน.....	9
2.4	การแปลง ISA ของซอฟต์แวร์เวอร์ชวลไลซ์ในซิสเต็มส์เวอร์ชวลแมชชีน.....	11
2.5	ความสัมพันธ์ระหว่างเวอร์ชวลแมชชีนมอนิเตอร์และเวอร์ชวลแมชชีน.....	14
2.6	วิธีการแปลงไบนารีในการเวอร์ชวลไลซ์สถาปัตยกรรม x86 ของวีเอ็มแวร์.....	17
2.7	โครงสร้างของเครื่องที่รันเซนไฮเปอร์ไวเซอร์.....	19
2.8	ไลฟ์ไมเกรชันในเซนไฮเปอร์ไวเซอร์.....	20
2.9	ไมโครเคอร์เนลไฮเปอร์ไวเซอร์ (Microkernel Hypervisor).....	22
2.10	สถาปัตยกรรมของเควีเอ็ม.....	25
2.11	การเช็คพอยน์โดยใช้โลคอลโฮสไลฟ์ไมเกรชัน.....	30
2.12	สถาปัตยกรรมระดับสูงของ Remus.....	31
3.1	ขั้นตอนการเช็คพอยน์เวอร์ชวลแมชชีนตามปกติในเควีเอ็ม.....	35
3.2	ขั้นตอนการเช็คพอยน์เวอร์ชวลแมชชีนของไลฟ์เช็คพอยน์โปรโตคอล.....	37
3.3	ขั้นตอนการเช็คพอยน์เวอร์ชวลแมชชีนของเทรดเบสไลฟ์เช็คพอยน์โปรโตคอล.....	39
3.4	อัลกอริทึมของเทรดเบสไลฟ์เช็คพอยน์โปรโตคอล.....	40
3.5	การทำงานเช็คพอยน์ของเทรดขณะที่เวอร์ชวลแมชชีนกำลังทำงานตามปกติ.....	41
3.6	สถานการณ์การเขียนหน่วยความจำของเวอร์ชวลแมชชีนและการคัดลอกหน่วยความจำของเทรด.....	43
3.7	การอิมพลีเมนต์ do_migrate_state().....	46
3.8	การอิมพลีเมนต์ migrate_fd_put_ready() ในเวอร์ชันที่ใช้เทรด.....	46
3.9	การคำนวณเงื่อนไขการหยุดคัดลอกหน่วยความจำใน ram_save_live().....	47
3.10	การชิงโครโนซ์เวอร์ชวลแมชชีนและเทรดด้วย mutex.....	48
3.11	การป้องกันการสลับลำดับคำสั่งด้วย memory barrier.....	48
3.12	การป้องกันการสลับลำดับคำสั่งในการคัดลอกหน่วยความจำของเทรดและ	

การเขียนหน่วยความจำของเวอร์ชวลแมชชีน.....	49
3.13 การรีเซ็ตาร์ทเวอร์ชวลแมชชีนจากไฟล์สถานะ.....	50
3.14 เช็คพอยน์โอเวอร์เฮดและดาวน์โหลดใหม่ของวิธีการเช็คพอยน์แต่ละแบบในการทดลอง เบื้องต้น.....	63
4.1 เช็คพอยน์โอเวอร์เฮดและดาวน์โหลดใหม่ของแต่ละวิธีการเช็คพอยน์.....	77
5.1 เวลาในการทำงานของเบนช์มาร์กที่มีการเช็คพอยน์และไม่มีการเช็คพอยน์.....	80
5.2 โอเวอร์เฮดของแต่ละวิธีการเช็คพอยน์.....	82
5.3 เวลาในการเช็คพอยน์เวอร์ชวลแมชชีน.....	84
5.4 ขนาดของไฟล์สถานะเวอร์ชวลแมชชีน.....	86
ข.1 เวลาในการทำงานเช็คพอยน์ของเทรคและเวอร์ชวลแมชชีน.....	97

สำนักหอสมุด