

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

เทคโนโลยีกริด (Grid) [1, 2] เป็นแนวคิดที่ยืดหยุ่นในการสร้างระบบการประมวลผลแบบกระจายด้วยกลุ่มทรัพยากรที่ร่วมใช้กันระหว่างองค์กรเพื่อบรรลุวัตถุประสงค์หนึ่งๆ ร่วมกัน โดยยังคงสอดคล้องกับนโยบายบริหารทรัพยากรของแต่ละองค์กรที่แตกต่างกัน ทรัพยากรอาจอยู่ในตำแหน่งที่กระจายไปตามสภาพภูมิศาสตร์และเขตการบริหาร กริดช่วยให้เราสามารถสร้างกลุ่มทรัพยากรด้วยทรัพยากรต่างๆ ที่เหมาะสมที่อยู่ในแต่ละองค์กรซึ่งตรงกับความต้องการของงานหรือวัตถุประสงค์ที่เกิดขึ้นจากความร่วมมือระหว่างองค์กรได้ เช่นงานที่ต้องการเครื่องมือทางวิทยาศาสตร์เฉพาะทางที่มีราคาแพงและมีอยู่จำนวนน้อย งานที่ประกอบด้วยปัญหาที่ยากต่อการหาคำตอบซึ่งต้องการพลังการคำนวณมาก งานที่ต้องการใช้ข้อมูลหรือสร้างข้อมูลขนาดใหญ่มากที่ต้องการการเข้าถึงข้อมูลอย่างปลอดภัย

ทรัพยากรบนกริดจึงมักเป็นทรัพยากรที่มีอยู่น้อยและจำเป็นต้องร่วมใช้กันระหว่างองค์กร เป็นทรัพยากรที่มีสมรรถนะสูง ทำงานเฉพาะทาง มีราคาสูง เช่นคอมพิวเตอร์เซิร์ฟเวอร์ ประสิทธิภาพสูงที่มีโปรแกรมคำนวณงานเฉพาะ คอมพิวเตอร์ให้บริการเพิ่มข้อมูลที่เก็บผลการทดลองจากอุปกรณ์วิทยาศาสตร์เป็นต้น ทรัพยากรบนกริดจึงมักวางตัวอยู่บริเวณศูนย์กลางของอินเทอร์เน็ตซึ่งอนุญาตให้ผู้ใช้กริดสามารถเข้าถึงได้ผ่านการควบคุมระยะไกล

ปัจจุบันเทคโนโลยีเพียร์ทูเพียร์ (Peer-to-Peer: P2P) [3, 4, 5, 6] เป็นที่นิยมและแพร่หลายมาก อันเนื่องมาจากลักษณะเฉพาะของระบบเพียร์ทูเพียร์และโปรแกรมประยุกต์ที่ใช้ร่วมกัน กล่าวคือระบบเพียร์ทูเพียร์มีลักษณะเฉพาะที่โดดเด่นใดๆ บนเครือข่ายที่เรียกว่าเพียร์สามารถปฏิบัติตนเป็นทั้งผู้ให้และผู้ใช้บริการแก่เพียร์อื่นๆ ได้ด้วยตัวเอง เพียร์สามารถเข้าร่วมและออกจากเครือข่ายได้อย่างยืดหยุ่นและพลวัต (Dynamic) ในขณะที่โปรแกรมประยุกต์ส่วนใหญ่ก็สร้างขึ้นเพื่อทำให้เพียร์ต่างๆ ร่วมบริการใช้ทรัพยากรที่มีอยู่บนแต่ละเพียร์ร่วมกัน บริการที่เป็นที่นิยมมากที่สุดคือการร่วมใช้แฟ้มข้อมูล (File sharing) ซึ่งมีอยู่หลากหลายตามประเภทแฟ้มข้อมูลและวิธีการร่วมใช้ เช่น Napster[7] Gnutella[8] KaZaa[9] และ BitTorrent[10] เป็นต้น ในขณะที่บริการการประมวลผลแบบเพียร์ทูเพียร์ก็เป็นที่สนใจในการสร้างการประมวลผลแบบกระจายขนาดใหญ่หลายๆ เช่นการวิเคราะห์สัญญาณวิทยุจากท้องฟ้าเพื่อค้นหาสิ่งมีชีวิตจากนอกโลกของ

SETI@home[11] การค้นหาจำนวนเฉพาะที่ใหญ่ที่สุดของ GIMPS (The great internet mersenne prime number search) [12] เป็นต้น

ทรัพยากรบนเครือข่ายเพียร์ทูเพียร์ส่วนใหญ่จึงมักเป็นคอมพิวเตอร์ส่วนบุคคลและอุปกรณ์พกพาที่สามารถเชื่อมต่อกับเครือข่ายอินเทอร์เน็ตได้ซึ่งมีใช้กันแพร่หลาย ทรัพยากรเหล่านี้มักวางตัวอยู่บริเวณขอบนอก (Edge) ของอินเทอร์เน็ต ซึ่งเห็นได้ชัดเจนว่าการเติบโตขึ้นทั้งด้านจำนวนและประสิทธิภาพของคอมพิวเตอร์ส่วนบุคคล การขยายตัวและการพัฒนาประสิทธิภาพการเชื่อมโยงเครือข่ายไปยังเครื่องคอมพิวเตอร์ส่วนบุคคลเหล่านี้มีประสิทธิภาพก้าวหน้าและเป็นไปอย่างรวดเร็วมาก

จะเห็นได้ว่ากริดและเพียร์ทูเพียร์ต่างก็มีเป้าหมายที่คล้ายคลึงกันคือการสร้างระบบประมวลผลแบบกระจายขนาดใหญ่มากด้วยทรัพยากรที่ได้จากการร่วมใช้กัน ทำให้สามารถนำเอาระบบที่ได้มาใช้ประมวลผลเพื่อบรรลุวัตถุประสงค์ใดๆ แต่กริดและเพียร์ทูเพียร์ต่างมีทรัพยากรที่กระจายอยู่ในบริเวณที่ต่างกันบนอินเทอร์เน็ต โดยเราไม่อาจมองข้ามแนวโน้มการเติบโตของปริมาณและประสิทธิภาพของคอมพิวเตอร์ส่วนบุคคลซึ่งเป็นทรัพยากรส่วนใหญ่ของเครือข่ายเพียร์ทูเพียร์

งานที่ถูกประมวลผลบนกริดส่วนใหญ่เป็นงานที่มีความละเอียดซับซ้อน ต้องการพลังคำนวณสูงและมักถูกประมวลผลเป็นเวลานานๆ โดยเข้าใช้ข้อมูลและสร้างข้อมูลที่มีขนาดใหญ่และมีจำนวนมาก บ่อยครั้งที่ผู้บริหารมักพบปัญหาเนื่องที่จัดเก็บข้อมูลของเครื่องคอมพิวเตอร์เซิร์ฟเวอร์บนกริดไม่เพียงพอ ในขณะที่ข้อมูลบางส่วนจำนวนไม่น้อยก็ใช้งานไม่บ่อย แต่ยังคงจำเป็นต้องเก็บไว้ การเพิ่มเติมดิสก์ข้อมูลบนเครื่องคอมพิวเตอร์เซิร์ฟเวอร์เหล่านี้ก็ไม่สะดวกและมีค่าใช้จ่ายสูง ในขณะที่คอมพิวเตอร์ส่วนบุคคลที่มีอยู่จำนวนมากภายในองค์กรที่บริหารอยู่และในเครือข่ายอินเทอร์เน็ตกลับมีเนื้อที่ว่างบนดิสก์ข้อมูลอยู่มาก การเพิ่มเติมดิสก์ข้อมูลบนเครื่องคอมพิวเตอร์ส่วนบุคคลก็สะดวกและมีค่าใช้จ่ายที่น้อยกว่า หากสามารถรวบรวมเนื้อที่ว่างบนเครื่องคอมพิวเตอร์ส่วนบุคคลเหล่านี้มาทำเป็นเสมือนดิสก์ข้อมูลแก่กริดได้ จะทำให้กริดมีดิสก์ข้อมูลเสมือนขนาดใหญ่มหาศาล อีกทั้งยังช่วยให้เกิดการใช้ทรัพยากรที่คุ้มค่าและลดค่าใช้จ่ายลง

งานวิจัยนี้ได้นำเสนอแนวคิดการรวบรวมทรัพยากรบนเครือข่ายเพียร์ทูเพียร์ ซึ่งเป็นทรัพยากรที่มีจำนวนมากและมีสมรรถนะดีมาให้แก่กริด โดยงานวิจัยนี้มุ่งเน้นที่การรวบรวมทรัพยากรข้อมูลคือเนื้อข้อมูลและเนื้อที่ข้อมูล ทำให้กริดเสมือนมีเนื้อที่จัดเก็บข้อมูลขนาดใหญ่มากที่สามารถเข้าใช้ จัดเก็บและเรียกคืนแฟ้มข้อมูลได้ ผู้วิจัยได้นำเสนอการรวบรวมทรัพยากรข้อมูลแก่ระบบกริดด้วยการใช้ระบบร่วมใช้แฟ้มข้อมูล โดยมีสมมุติฐานว่าแฟ้มข้อมูลที่จัดเก็บลงเครือข่ายเพียร์ทูเพียร์สามารถเข้าถึงได้จากทุกผู้ใช้กริด และมีส่วนเชื่อมต่อ (Gateway) ที่มี

ความสามารถของทั้งสองระบบที่เชื่อมต่อ ส่วนเชื่อมต่อเป็นเสมือนจุดเข้าถึงเนื้อที่จัดเก็บข้อมูลต่อผู้ใช้และโปรแกรมประยุกต์บนกริด การกำหนดตำแหน่งจัดเก็บและเรียกคืนแฟ้มข้อมูลบนทรัพยากรแบบกระจายบนระบบร่วมใช้แฟ้มข้อมูลนั้นใช้กลไกการทำดัชนีแบบกระจาย (Distributed indexing) เป็นเครื่องมือกำหนดความรับผิดชอบร่วมกันของทุกเพียร์บนเครือข่ายเพียร์ทูเพียร์

1.2 กรณีตัวอย่างปัญหา

กรณีตัวอย่างปัญหาได้แก่ระบบกริดประมวลผลพยากรณ์อากาศ ระบบกริดนี้มีแฟ้มข้อมูลทางอุตุนิยมวิทยาขนาดใหญ่ที่ถูกสร้างและถูกใช้งานเป็นจำนวนมากอันเนื่องจากการแปรผันกับเวลาและตำแหน่งทางภูมิศาสตร์ ความละเอียดของการแปรผัน และมิติที่ใช้ในการคำนวณ เช่น ข้อมูลความกดอากาศ ข้อมูลทิศทางและความเร็วลม ข้อมูลปริมาณน้ำฝนและความชื้นสัมพัทธ์ เป็นต้น ข้อมูลที่ถูกใช้หรือสร้างขึ้นเป็นเวลานานมักถูกใช้ไม่บ่อยนัก แต่ก็ยังคงมีความสำคัญและจำเป็นต้องคงเก็บไว้ กริดประมวลผลพยากรณ์อากาศจึงมีความต้องการดิสก์ข้อมูลขนาดใหญ่หลายๆ แต่ภายในหน่วยงานที่มีระบบกริดประมวลผลพยากรณ์อากาศก็มีเครื่องคอมพิวเตอร์ส่วนบุคคลที่เชื่อมต่อเครือข่ายภายในและเครือข่ายอินเทอร์เน็ตซึ่งต่างก็มีเนื้อที่ดิสก์ข้อมูลว่างไม่น้อย ซึ่งหากสามารถรวบรวมจากเครื่องคอมพิวเตอร์ส่วนบุคคลจำนวนมากๆ เหล่านี้แล้วก็จะได้ดิสก์ข้อมูลเสมือนขนาดใหญ่หลายๆ ดิสก์ข้อมูลเสมือนจึงเป็นคำตอบหนึ่งสำหรับการจัดเก็บข้อมูลที่กริดประมวลผลพยากรณ์อากาศสร้างขึ้นโดยเฉพาะอย่างยิ่งกับข้อมูลในช่วงเวลาในอดีตที่มีโอกาสใช้น้อยครั้ง

1.3 วัตถุประสงค์

1. เพื่อสร้างระบบรวบรวมทรัพยากรเนื้อข้อมูลและเนื้อที่จัดเก็บข้อมูลของระบบร่วมใช้แฟ้มข้อมูลบนเครือข่ายเพียร์ทูเพียร์มาให้แก่กริด
2. เพื่อศึกษาการจัดเก็บข้อมูลแบบกระจายบนเครือข่ายเพียร์ทูเพียร์ด้วยกลไกดัชนีแบบกระจาย ซึ่งจำเป็นต่อการกำหนดความรับผิดชอบของข้อมูล ความน่าเชื่อถือและความสามารถขยายตัวของระบบ

1.4 ขอบเขตงานวิจัย

ออกแบบและสร้างส่วนเชื่อมต่อกริดและระบบร่วมใช้แฟ้มข้อมูลบนเครือข่ายเพียร์ทูเพียร์ โดยใช้โกลบัสทูลคิท (Globus Toolkit) [13, 14] เป็นแพลตฟอร์มสำหรับกริด และใช้จังก์ซ์ตา (JXTA) [15] เป็นแพลตฟอร์มของเครือข่ายเพียร์ทูเพียร์ ทรัพยากรเนื้อที่จัดเก็บข้อมูลบนเครือข่าย

เพียร์ทูเพียร์ถูกใช้โดยกริดฝ่ายเดียว และเพิ่มข้อมูลที่ถูกสั่งให้จัดเก็บลงเครือข่ายเพียร์ทูเพียร์ สามารถเข้าถึงได้จากทุกผู้ใช้และโปรแกรมประยุกต์บนกริด

การทดลองทำบนกริดและเครือข่ายคอมพิวเตอร์ส่วนบุคคลภายในภาควิชาวิศวกรรมคอมพิวเตอร์ด้วยการถ่ายโอนรูปแบบต่างๆ

1.5 ขั้นตอนและวิธีดำเนินงานวิจัย

1. ศึกษากริด กลไกการทำงานและการจัดการข้อมูลบนกริด
2. ศึกษาเครือข่ายเพียร์ทูเพียร์ กลไกการทำงานและการจัดการข้อมูลบนเครือข่ายเพียร์ทูเพียร์
3. ศึกษากลไกการทำงานของระบบร่วมใช้เพิ่มข้อมูลบนเครือข่ายเพียร์ทูเพียร์ และกลไกดัชนีแบบกระจายบนเครือข่ายเพียร์ทูเพียร์
4. ออกแบบส่วนเชื่อมต่อกริดและระบบร่วมใช้เพิ่มข้อมูลบนเครือข่ายเพียร์ทูเพียร์
5. ทดลองระบบที่ได้บนกริดและเครือข่ายคอมพิวเตอร์ส่วนบุคคลภายในภาควิชาวิศวกรรมคอมพิวเตอร์สำหรับแต่ละรูปแบบการถ่ายโอน
6. สรุปผลการทดลองและเขียนวิทยานิพนธ์

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1. ทำให้สามารถสร้างระบบแบ่งร่วมเพิ่มข้อมูลด้วยเครือข่ายคอมพิวเตอร์ส่วนบุคคลในองค์กรที่มีอยู่เดิมแก่กริด ซึ่งเครือข่ายคอมพิวเตอร์ส่วนบุคคลเหล่านี้มักถูกใช้งานไม่เต็มประสิทธิภาพ
2. ลดค่าใช้จ่ายในการซื้อดีสก์เก็บข้อมูลที่มีราคาสูงสำหรับเครื่องคอมพิวเตอร์สมรรถนะสูงบนกริด ด้วยการใช้ดีสก์เก็บข้อมูลบนเครือข่ายคอมพิวเตอร์ส่วนบุคคลซึ่งมีจำนวนมากอยู่แล้ว

1.7 ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์

วิทยานิพนธ์นี้แบ่งเนื้อหาออกเป็น 7 บทดังนี้ บทที่ 1 เป็นบทนำซึ่งกล่าวถึงที่มาและความสำคัญของปัญหา รวมทั้งวัตถุประสงค์ของงานวิจัย บทที่ 2 สรุปกริดและโกลบัลสทูลคิท บทที่ 3 สรุปเพียร์ทูเพียร์และจังก์ชตา บทที่ 4 กล่าวถึงกลไกดัชนีแบบกระจายจีไอเอสพี (Global Information Sharing Protocol: GISP) [16, 17] ที่ใช้กับระบบร่วมใช้เพิ่มข้อมูล บทที่ 5 อธิบาย

การออกแบบและการสร้างระบบเชื่อมต่อ ผลการทดลองแสดงไว้ในบทที่ 6 และสุดท้ายข้อสรุปจากการวิจัยและข้อเสนอแนะถูกกล่าวไว้ในบทที่ 7

1.8 ผลงานที่ตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้รับการตีพิมพ์เป็นบทความทางวิชาการในหัวข้อเรื่อง “Integrating Peer-to-Peer File Sharing into the Grid” โดย กิตติศักดิ์ ถนอมธีระนันท์ และวีระเหมืองสิน ในงานประชุมวิชาการ “10th Annual National Symposium on Computational Science & Engineering (ANSCSE10)” ซึ่งจัดโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) ณ มหาวิทยาลัยเชียงใหม่ จังหวัดเชียงใหม่ ในระหว่างวันที่ 22-24 มีนาคม 2549



ศูนย์วิทยุทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย