บทที่ 2
ทฤษฎี งานวิจัยที่เกี่ยวข้อง และผลการศึกษาการใช้งานเครือข่าย

เนื้อหาในบทนี้แบ่งออกเป็น 3 ส่วนหลัก โดยหัวข้อที่ 2.1 เกี่ยวข้องกับทฤษฎีทางด้านเครือข่ายที่สนับสนุนงานวิจัย หัวข้อที่ 2.2 เป็นข้อมูลที่ได้จากการศึกษารวบรวมความรู้จากงานวิจัยที่เกี่ยวข้อง และ หัวข้อที่ 2.3 เสนอโครงสร้างเครือข่ายของมหาวิทยาลัยสงขลานครินทร์ (SritrangNet) ผลที่ได้จากการศึกษาลักษณะการใช้งานเครือข่ายที่เกิดขึ้น

2.1 ทฤษฎีที่เกี่ยวข้องกับงานวิจัย

ผู้วิจัยได้แบ่งทฤษฎีที่เกี่ยวข้องกับงานวิจัยออกเป็น 3 หัวข้อย่อย ดังนี้คือ ความรู้พื้นฐานทางด้านเครือข่ายคอมพิวเตอร์ การอ่านข้อมูลในแพ็กเก็ต และคุณภาพของการให้บริการตามลำดับ สำหรับเนื้อหาในหัวข้อคุณภาพของการให้บริการจะกล่าวเฉพาะแนวคิดของคุณภาพของการให้บริการ เกณฑ์วัดที่จะทำให้เครือข่ายเกิดคุณภาพ กลไกการทำงานของคิว ซึ่งจัดวิธีการพื้นฐานที่ก่อให้เกิดคุณภาพของการให้บริการบนเครือข่าย และสถาปัตยกรรมอื่น ๆ ที่ได้รับการออกแบบเพื่อให้เครือข่ายมุ่งสู่คำว่าคุณภาพของการให้บริการ โดยรายละเอียดของแต่ละเรื่องได้กล่าวในหัวข้อที่ 2.1.1 ถึง 2.1.3 ตามลำดับ

2.1.1 ความรู้พื้นฐานทางด้านเครือข่ายคอมพิวเตอร์

เครือข่ายคอมพิวเตอร์ หมายถึง กลุ่มของอุปกรณ์เครือข่ายต่าง ๆ เช่น เครื่องคอมพิวเตอร์ ฮับ (Hub) สวิตซ์ (Switch) และเราเตอร์ (Router) ที่มีการเชื่อมต่อเพื่อใช้ในการแลกเปลี่ยนข้อมูลโดยไม่คำนึงถึงระยะทางระหว่างกัน ซึ่งการเชื่อมต่ออาจใช้สายเคเบิลธรรมดา สายเคเบิลใยแก้วนำแสง คลื่นไมโครเวฟ สัญญาณดาวเทียม หรือสื่อกลางอื่น ๆ ที่นำมาใช้ในการแลกเปลี่ยนข้อมูล

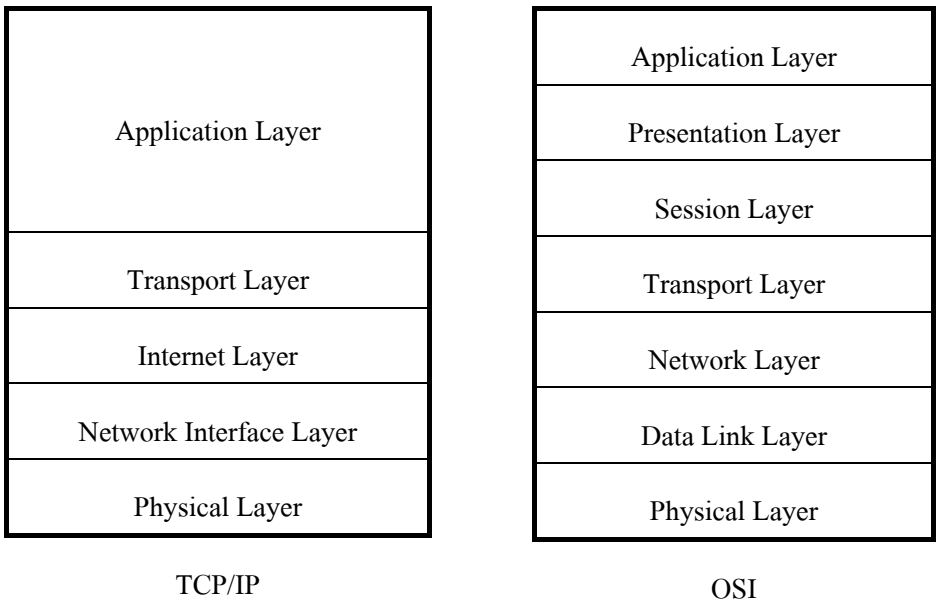
2.1.1.1 ประเภทของเครือข่าย

หากแยกประเภทของเครือข่ายคอมพิวเตอร์ตามขนาด หรือระยะทำการสามารถแบ่งออกได้ 2 ประเภท คือ เครือข่ายวงกว้าง (Wide Area Network: WAN) และเครือข่ายเฉพาะบริเวณ (Local Area Network: LAN) โดยที่ WAN เป็นเครือข่ายคอมพิวเตอร์ที่เกิดจากการเชื่อมโยงเครือข่ายย่อย หรือคอมพิวเตอร์ที่อยู่ห่างไกลกันเข้าด้วยกัน เช่น การเชื่อมต่อระหว่างจังหวัดโดยผ่านวงจรสื่อสารของผู้ให้บริการ หรือที่เรียกว่า Common carrier ส่วน LAN เป็นเครือข่าย

คอมพิวเตอร์ที่เกิดจากการเชื่อมโยงเครือข่ายขนาดเล็ก หรือคอมพิวเตอร์เฉพาะกลุ่มและมีระยะทำการเชื่อมโยงไม่ไกลมากนัก เช่น เครือข่ายภายในมหาวิทยาลัย และเครือข่ายภายในอาคารเดียวกัน เป็นต้น

2.1.1.2 โปรโตคอล

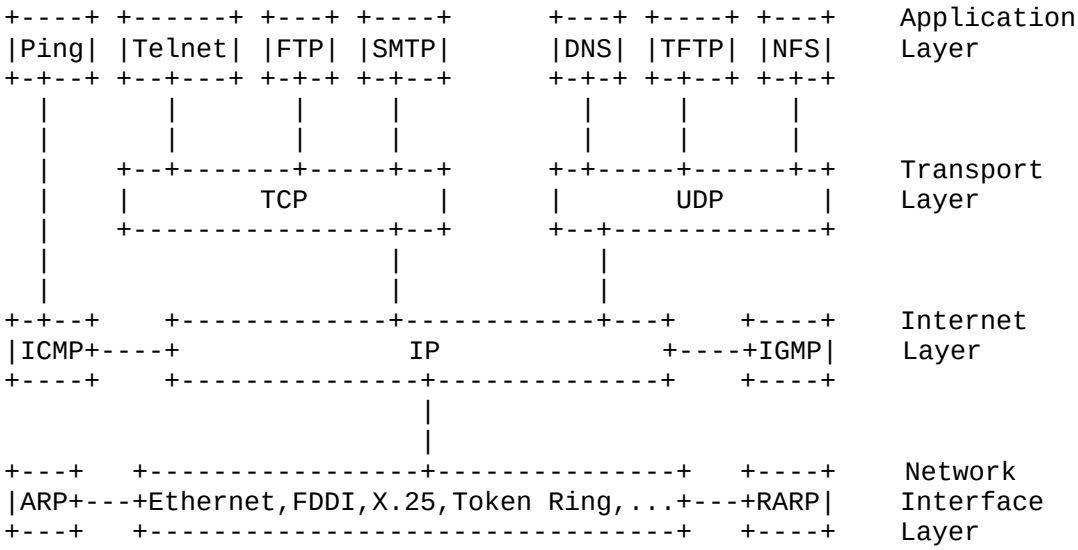
เพื่อให้อุปกรณ์เครือข่ายสามารถส่งผ่านข้อมูลถึงกันได้อย่างถูกต้อง จำเป็นต้องมีการกำหนดระเบียบ หรือมีการทำข้อตกลงในการสื่อสารระหว่างกัน หรือที่เรียกว่า โปรโตคอล (Protocol) ทั้งนี้โปรโตคอลมาตรฐานที่มักจะได้รับการอ้างอิงถึง คือ TCP/IP protocol suite และ OSI (Open System Interconnection) reference model มาตรฐานทั้งสองต่างได้รับการออกแบบให้แบ่งการทำงานออกเป็นชั้น ๆ หรือ เลเยอร์ (Layer) ดังรูปที่ 2.1 สำหรับรายละเอียดเกี่ยวกับหน้าที่ในแต่ละชั้นของ TCP/IP และ OSI สามารถศึกษาได้จาก [10, p.19-22] หรือ [8, chapter 1] สำหรับเอกสารชุดนี้ได้ให้ความสนใจเฉพาะ TCP/IP เนื่องจากเป็นโปรโตคอลที่มีการใช้งานบนเครือข่ายอินเทอร์เน็ต และอินทราเน็ตทั่วไปในปัจจุบัน



รูปที่ 2.1 แสดง TCP/IP protocol suite และ OSI reference model

2.1.1.3 โปรโตคอลในชุดของ TCP/IP

สาเหตุที่เรียกมาตรฐานของ TCP/IP เป็น Protocol suite เนื่องจากการทำงานของ TCP/IP เกิดจากการประสานงานกันระหว่างโปรโตคอลย่อยหลายตัวที่สัมพันธ์กันของแต่ละ Layer การซ้อนทับกันของโปรโตคอลจะเรียกว่า TCP/IP Stack ดังแสดงในรูปที่ 2.2



รูปที่ 2.2 แสดงตัวอย่างโปรโตคอลใน TCP/IP Stack

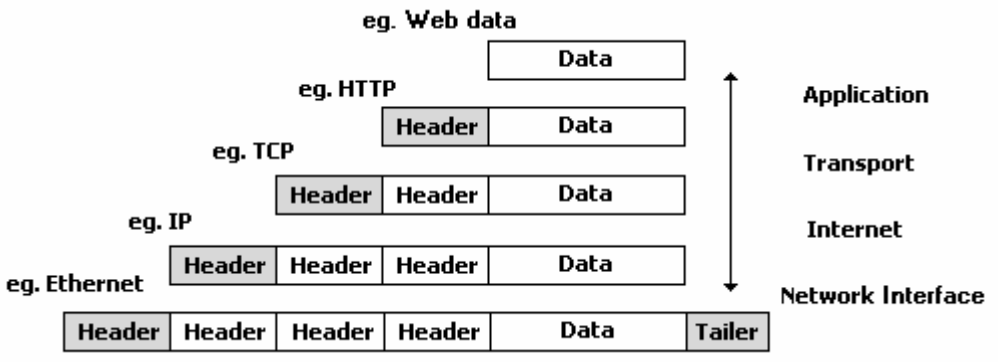
ในที่นี้จะกล่าวรายละเอียดพอสังเขปของโปรโตคอลสำคัญบางชนิดในชุดของ TCP/IP และเป็นโปรโตคอลที่เกี่ยวข้องในงานวิจัยนี้ คือ TCP UDP IP และ ICMP ดังนี้

- TCP (Transmission Control Protocol) เป็นโปรโตคอลให้บริการนำส่งข้อมูลที่ต้องการความน่าเชื่อถือในลักษณะต่าง ๆ เช่น ผู้รับปลายทางได้รับข้อมูลที่ถูกต้องตามลำดับและในเวลาที่กำหนดหรือในเวลาที่เหมาะสม โดย TCP มีกลไกเพื่อรับประกันการให้บริการเหล่านี้เช่น การส่งแพ็กเก็ตซ้ำใหม่เมื่อมีแพ็กเก็ตสูญหาย การกำจัดแพ็กเก็ตที่ซ้ำซ้อน การจัดลำดับของแพ็กเก็ต เป็นต้น
- UDP (User Datagram Protocol) เป็นโปรโตคอลใน Layer เดียวกับ TCP ให้บริการสำหรับการทำงานที่ต้องการความรวดเร็วมากกว่าความถูกต้องแม่นยำ
- IP (Internet Protocol) เป็นโปรโตคอลสำหรับนำส่งแพ็กเก็ตข้อมูลจากคอมพิวเตอร์เครื่องหนึ่ง (โฮสต์/Host) ไปยังคอมพิวเตอร์เครื่องอื่นในเครือข่ายอินเทอร์เน็ต

- ICMP (Internet Control Message Protocol) เป็นโปรโตคอลสำหรับรายงานความผิดพลาดที่เกิดขึ้นให้กับคอมพิวเตอร์ต้นทาง เพื่อให้ทราบว่ามีปัญหาอะไรขึ้นในกระบวนการนำส่งข้อมูล เช่น ไม่สามารถนำส่งแพ็กเก็ตไปยังเครื่องปลายทางได้ เป็นต้น หรือเรียกว่าทำ Error Reporting นอกจากนี้ ICMP ยังถูกใช้เพื่อสอบถามข้อมูลด้วย (Query) เช่น Echo request, Echo reply เป็นต้น

2.1.1.4 การส่งข้อมูลในเครือข่าย

การส่งข้อมูลจาก Application หนึ่งๆ เข้าสู่เครือข่าย กระบวนการส่งเริ่มจากการแบ่งย่อยข้อมูล และผนึกข้อมูลส่วนหัว (Header) ที่จำเป็นของโปรโตคอลในชั้นที่ต่ำกว่าลงไปเรื่อย ๆ หรือเรียกว่าการทำ Encapsulation เมื่อเครื่องปลายทางได้รับข้อมูล ฟังก์ชั่นที่เกี่ยวข้องในแต่ละ Layer จะถอดข้อมูลส่วนหัวออกเรื่อย ๆ จนกระทั่งเหลือเฉพาะข้อมูลต้นฉบับเมื่อถึง Layer บนสุด หรือเรียกว่าการทำ Decapsulation ดังแสดงในรูปที่ 2.3

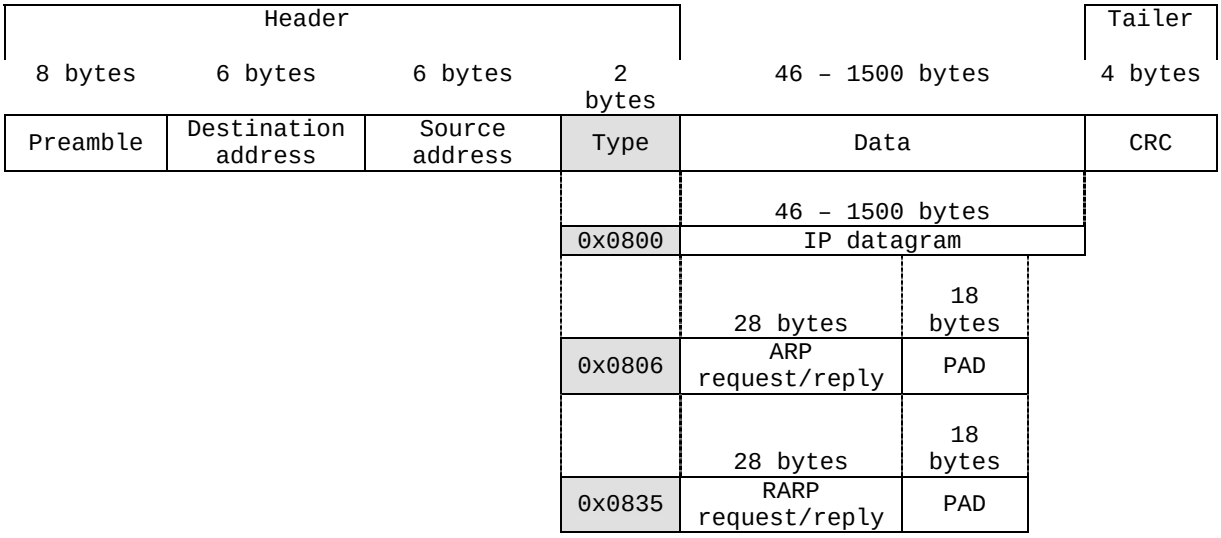


รูปที่ 2.3 แสดงการ Encapsulation (ลง) และ Decapsulation (ขึ้น)

2.1.2 การอ่านข้อมูลในแพ็กเก็ต

ในหัวข้อนี้จะอธิบายถึงวิธีการอ่านข้อมูลในแพ็กเก็ต โดยเฉพาะแพ็กเก็ตบนเครือข่าย LAN ที่ใช้โปรโตคอลอีเทอร์เน็ต (Ethernet) หรือ IEEE 802.3 CSMA/CD เป็นมาตรฐานในการสื่อสาร เมื่อแพ็กเก็ตเดินทางถึงการ์ดเครือข่าย/การ์ดแลน (Network Interface หรือ Network Card) ของเครื่องปลายทาง แพ็กเก็ตจะถูกนำส่งขึ้นไปยัง Network Interface Layer แพ็กเก็ตในชั้นนี้จะเรียกว่าเฟรม (Frame) มีโครงสร้างดังรูปที่ 2.4 ฟิลด์แอดเดรส (Address) หรือในที่นี้คือ MAC Address มีข้อมูลหมายเลขแอดเดรสของการ์ดเครือข่าย ฟิลด์ชนิด (Type) ใช้ระบุชนิดของ

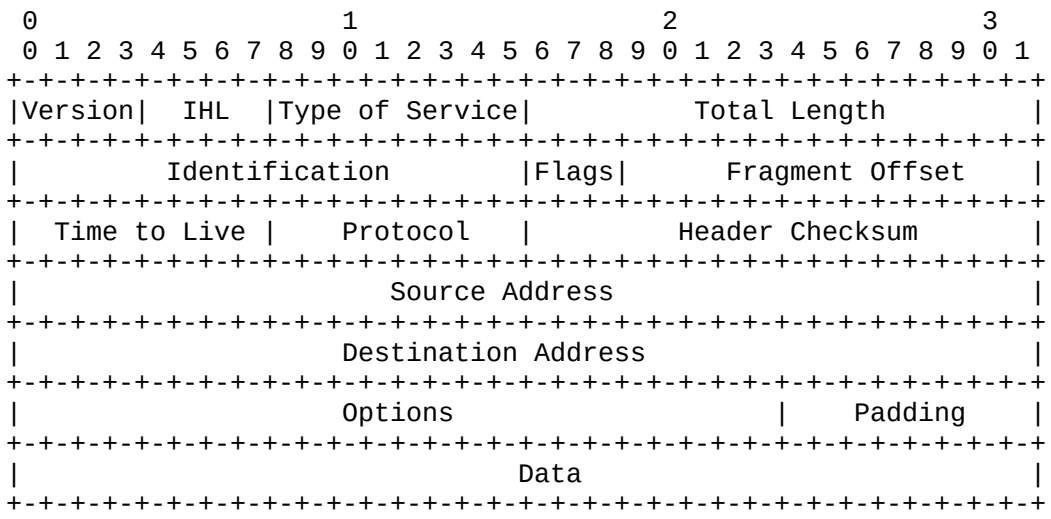
โปรโตคอลใน Internet Layer เช่น ค่า 0x0800 หมายถึง IP เป็นต้น ฟิลด์ข้อมูล (Data) ใช้บรรจุข้อมูลของโปรโตคอลในระดับชั้นที่สูงกว่าตามมาตรฐานอีเทอร์เน็ต โดยขนาดข้อมูลต้องมีค่าอยู่ระหว่าง 46 ถึง 1500 ไบต์ หากข้อมูลมีขนาดเล็กกว่า 46 ไบต์จำเป็นเติมข้อมูลเสริมให้เต็ม หรือที่เรียกว่า Padding



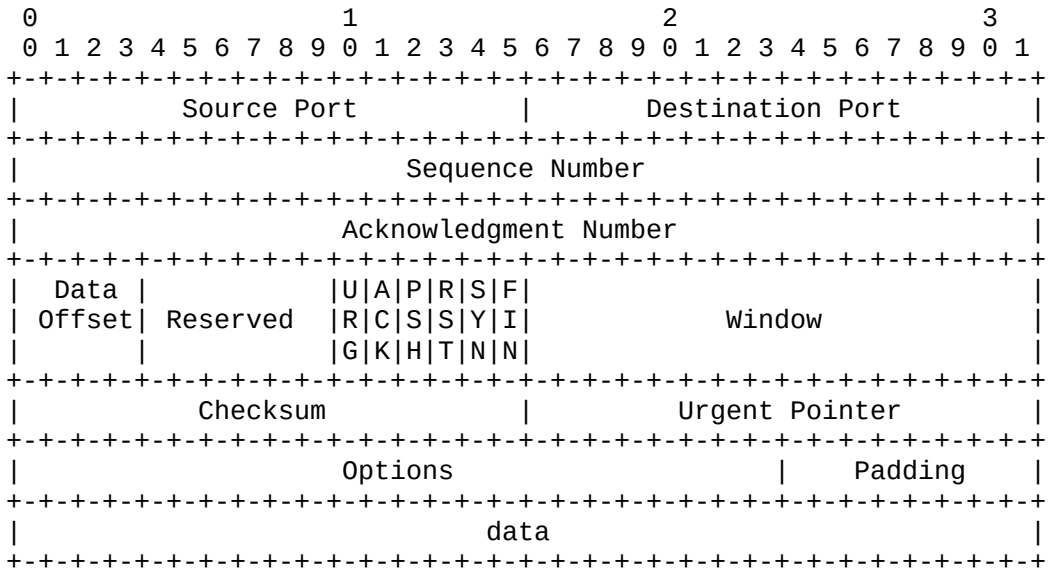
รูปที่ 2.4 โครงสร้างของเฟรมอีเทอร์เน็ต (Ethernet frame format)

เมื่อถอดข้อมูลส่วนหัวและส่วนท้ายของเฟรมที่มีค่าในฟิลด์ Type เป็น 0x0800 แพ็กเก็ตดังกล่าวจะถูกส่งไปใน Internet Layer จะเรียกแพ็กเก็ตในชั้นนี้ว่า ไอพีดาต้าแกรม (IP datagram) มีโครงสร้างของแพ็กเก็ตดังแสดงในรูปที่ 2.5 ประกอบด้วยส่วนหัวของ Internet Layer และส่วนของข้อมูล โดยปกติข้อมูลส่วนหัวมักมีขนาด 20 ไบต์ แสดงว่าไม่มีข้อมูลในฟิลด์ Options และ Padding สำหรับฟิลด์ที่ใช้ระบุขนาดส่วนหัวของ IP datagram คือ IHL (Internet Header Length) หากค่าในฟิลด์นี้เป็น 5 แสดงว่าข้อมูลส่วนหัวมีค่าเท่ากับ 5x4 หรือ 20 ไบต์ ฟิลด์ Total length ใช้ระบุขนาดจริงของ IP datagram มีหน่วยนับเป็นไบต์ เนื่องจากฟิลด์นี้มีขนาด 16 บิต ทำให้ IP datagram มีขนาดใหญ่สุดได้ไม่เกิน 2^{16}-1 หรือ 65,535 ไบต์ และค่าในฟิลด์นี้อาจมีค่าน้อยกว่า 46 ไบต์ตามขนาดของข้อมูลจริง ทั้งนี้ต้องมีการทำ Padding เพื่อให้ IP datagram มีขนาดต่ำสุดที่ 46 ไบต์ด้วย ฟิลด์สำคัญต่อมา คือ Protocol ใช้ระบุโปรโตคอลในระดับ Transport Layer เช่น 0x06 หมายถึง TCP (6) จะเรียกว่า TCP segment และ 0x11 หมายถึง UDP (17) จะเรียกว่า UDP message เป็นต้น โครงสร้างของแพ็กเก็ตใน Transport Layer ทั้ง 2 นี้ได้แสดงไว้ในรูปที่ 2.6 และ 2.7 ตามลำดับ ฟิลด์ Port มีไว้สำหรับแจ้งให้ทราบว่าแพ็กเก็ตนั้น ๆ เป็นของ Application ใด เช่น

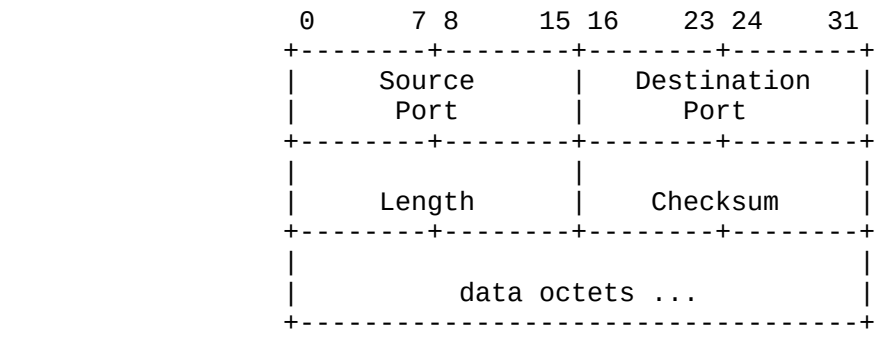
80/tcp หมายถึง www-http หรือ World Wide Web HTTP เป็นต้น สำหรับหมายเลขพอร์ตอื่น ๆ สามารถศึกษาเพิ่มเติมได้ที่เว็บไซต์ของ Internet Assigned Numbers Authority, http://www.iana.org/assignments/port-numbers นอกจากนี้รายละเอียดของฟิลด์อื่นที่มิได้กล่าวถึงในเอกสารฉบับนี้นี้สามารถอ่านเพิ่มเติมได้จาก RFC (http://www.ietf.org/rfc)



รูปที่ 2.5 แสดงโครงสร้างของ IP datagram [RFC 791]



รูปที่ 2.6 แสดงโครงสร้างของเซกเมนต์ทีซีพี (TCP segment) [RFC 793]



รูปที่ 2.7 แสดงโครงสร้างของเมสเสจยูดีพี (UDP Message) [RFC 768]

2.1.3 คุณภาพของการให้บริการ (Quality of Service: QoS)

จากการเติบโตของการใช้งานอินเทอร์เน็ตและอินทราเน็ต ทั้งการเพิ่มขึ้นของจำนวนผู้ใช้งานเครือข่าย และ Application หรือบริการที่มีบนเครือข่าย ได้ส่งผลให้ทรัพยากรของเครือข่ายที่มีอยู่จำกัด เช่น Bandwidth ไม่เพียงพอต่อความต้องการ หากเปรียบทรัพยากรของเครือข่ายเหมือนกับเงินรายได้ การเพิ่มขึ้นของบริการเปรียบได้กับความต้องการใช้จ่ายที่เพิ่มขึ้นเมื่อความต้องการเพิ่มมากขึ้น แต่รายรับคงที่ แสดงว่ามีเงินไม่เพียงพอต่อความต้องการ แต่การหาช่องทางเพิ่มรายได้ต้องเหนื่อยมากขึ้น อีกทางเลือกหนึ่งคือจัดวางระเบียบการใช้เงินด้วยการลดหรือตัดความต้องการที่ไม่จำเป็นทิ้งไป ในแง่ของการให้บริการเครือข่ายก็เช่นเดียวกัน หน่วยงานอาจเลือกแก้ปัญหาการขาดแคลนทรัพยากรด้วยการเพิ่ม Bandwidth หรือเปลี่ยนอุปกรณ์เครือข่ายที่ทำงานได้เร็วขึ้น แต่หน่วยงานต้องเสียค่าใช้จ่ายสูงมาก และเป็นการแก้ปัญหาเพียงบางจุดบนระบบเครือข่ายเท่านั้น หน่วยงานอาจเลือกควบคุมการใช้งานทรัพยากรเครือข่าย ด้วยการจัดลำดับความสำคัญของงาน เช่น ประเภทของ Application หรือกลุ่มของผู้ใช้งานเป็นต้น เพื่อให้การใช้ทรัพยากรของเครือข่ายเป็นไปอย่างมีประสิทธิภาพ

เนื้อหาในหัวข้อนี้ได้นำเสนอข้อมูลเกี่ยวกับ QoS ซึ่งเป็นแนวคิดในการแก้ปัญหาการขาดแคลนทรัพยากรเครือข่าย และเพิ่มประสิทธิภาพในการใช้งานเครือข่ายโดยที่มีค่าใช้จ่ายต่ำ ประกอบด้วยหัวข้อต่างๆ ต่อไปนี้ คือ คุณภาพของการให้บริการคืออะไร องค์ประกอบพื้นฐานของ QoS กลไกของ QoS เทคนิคการจัดการคิว และสถาปัตยกรรมของ QoS เป็นหัวข้อสุดท้าย

2.1.3.1 คุณภาพของการให้บริการคืออะไร [17] [26] [30]

ความหมาย หรือคำนิยามของ “คุณภาพของการให้บริการ” ไม่สามารถอธิบายให้แน่ชัดลงไปได้ แต่โดยทั่วไปการระบุว่าสิ่งใดมีคุณภาพต้องอาศัยเกณฑ์วัดบางอย่างที่เหมาะสมและเกี่ยวข้องกับสิ่งที่ต้องการวัด สมมติต้องการพิจารณาว่าภัตตาคาร หรือร้านอาหารใดมีคุณภาพหรือให้บริการได้ดีนั้น เกณฑ์วัดที่เหมาะที่จะนำมาใช้ในการวัด เช่น รสชาติของอาหารที่อร่อย การบริการและเอาใจใส่ที่ดีของพนักงาน หรือการเปิดเพลงที่ไพเราะ เป็นต้น แต่เกณฑ์เหล่านี้ใช่ว่าแต่ละคนจะมีระดับของการยอมรับ หรือให้คะแนนเหมือนกัน ในกรณีของเครือข่ายก็เช่นเดียวกัน การระบุว่าเครือข่ายใดมีคุณภาพจึงเป็นเรื่องที่ยาก

2.1.3.2 องค์ประกอบพื้นฐานของ QoS

จากที่กล่าวแล้วว่าขณะนี้ยังไม่มีการกำหนดเกณฑ์วัดที่แน่ชัดสำหรับพิจารณาคุณภาพของการให้บริการบนระบบเครือข่าย ได้มีกลุ่มวิจัยบางกลุ่มเช่น Light Reading [26] ได้เสนอเกณฑ์วัดหรือองค์ประกอบพื้นฐานที่ก่อให้เกิด QoS ไว้ 5 เรื่อง คือ Availability, Throughput, Latency/Delay, Jitter/Delay variation และ Loss โดยแต่ละองค์ประกอบมีรายละเอียดดังนี้

<u>เกณฑ์วัดที่ 1:</u> การมีให้ใช้งานได้ (Availability)

การมีให้ใช้งานได้เป็นการกำหนดสภาพความพร้อมของเครือข่าย อาจวัดในรูปของร้อยละ ในทางอุดมคติมักคาดหวังว่าเครือข่ายต้องสามารถใช้งานได้ร้อยละ 100 หรือกล่าวว่าเวลาของเวลาที่ใช้งานไม่ได้ (Downtime) มีค่าเป็นศูนย์ แต่ในความเป็นจริงไม่มีเครือข่ายใดสามารถให้บริการได้เช่นนั้นโดยมากผู้ให้บริการเครือข่ายจะตั้งค่าให้ต่ำกว่านี้ เช่นกำหนดค่าการมีให้ใช้งานได้คิดเป็นร้อยละ 99.999 แสดงว่าผู้ใช้จะไม่สามารถใช้งานเครือข่ายนั้นได้เพียง 5.256 นาทีต่อปีเท่านั้น องค์ประกอบที่ทำให้เกณฑ์นี้ประสบความสำเร็จ คือ ความน่าเชื่อถือและความทนทานของอุปกรณ์เครือข่าย ความมีเสถียรภาพของซอฟต์แวร์ และความสามารถในการพัฒนาหรือปรับปรุงเครือข่ายโดยไม่ต้องหยุดการทำงานของเครือข่าย

<u>เกณฑ์วัดที่ 2:</u> ช่องสัญญาณที่ส่งได้ (Throughput)

ในแง่ของการสื่อสารข้อมูล Throughput หมายถึง ความสำเร็จของการนำส่งข้อมูลจากจุดหนึ่งไปยังอีกจุดหนึ่งในช่วงเวลาที่กำหนด ซึ่งมักกำหนดในหน่วยบิตต่อวินาที (bps) สิ่งที่ต้องทำความเข้าใจคือ ค่านี้มิได้หมายถึงค่าสูงสุดของวงจรสื่อสารหรือช่องสัญญาณที่สามารถรับส่งได้ (Bandwidth) เพราะผู้ใช้รายอื่นสามารถใช้วงจรสื่อสารนี้ร่วมด้วยได้ ดังนั้นผู้ให้บริการอินเทอร์เน็ต (ISP) จึงใช้ค่า Throughput เป็นการรับประกันขนาดน้อยที่สุดของวงจรสื่อสารที่จะจัดสรรให้ใช้ได้ (Minimum throughput guarantee) เช่น เมื่อผู้เช่าเช่าวงจรขนาด 64 กิโลบิตต่อวินาที (Kbps) และ ISP ได้รับประกันว่าผู้เช่าจะใช้เครือข่ายได้ไม่ต่ำกว่า 32 Kbps ในกรณีนี้ค่า 32 เป็นค่า

Throughput ต่ำสุดที่ผู้ใช้งานต้องได้รับ สาเหตุที่ต้องรับประกันค่า Throughput เพราะบริการหรือ Application บางอย่างจำเป็นต้องมี Bandwidth เพียงพอจึงสามารถใช้งานได้อย่างมีประสิทธิภาพ เช่น การส่งสัญญาณภาพพร้อมเสียง

เกณฑ์วัดที่ 3: เวลาลาเทนซี่ (Latency)

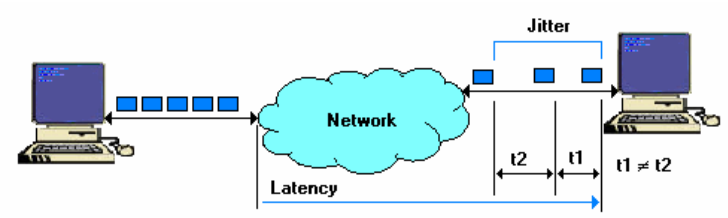
เวลาลาเทนซี่ มีความหมายเหมือนกับเวลาหน่วง (Delay) ซึ่งเป็นค่าเวลาการเดินทางของแพ็กเก็ตจากต้นทางไปยังปลายทาง สมมติแพ็กเก็ตเดินทางจากเครื่องต้นทางไปยังเครื่องให้บริการที่อยู่ห่างออกไป 5,000 กิโลเมตร ใช้เวลา 150 มิลลิวินาที ค่าเวลาลาเทนซี่คือ 150 ซึ่งเวลานี้อาจเกิดจาก

- Transmission delay: เป็นเวลาหน่วงของสื่อนำสัญญาณ
- Queueing delay: เป็นเวลาที่แพ็กเก็ตต้องรออยู่ในคิวก่อนการนำส่ง เมื่อเกิดความคับคั่งในเครือข่าย
- Packet reassembly delay: เป็นเวลาที่เกิดขึ้นเมื่อมีการรวมดาต้าแกรมชุดเดียวกันให้เป็นแพ็กเก็ต
- Router and other processing delay: เนื่องจากว่าโหนดที่เป็นทางเข้าออกของเครือข่ายหรือเกตเวย์ได้ใช้เวลาส่วนหนึ่งในการพิจารณา และเปลี่ยนข้อมูลบางอย่างในหัวของแพ็กเก็ต เช่น ค่า hop count ใน time-to-live เป็นต้น
- Other computer and storage delay: เป็นเวลาที่เกิดขึ้นที่จุดปลายของการเดินทางเมื่อมีการบันทึกข้อมูลลงสื่อ และเวลาในการเข้าถึงหน่วยบันทึกของอุปกรณ์จำพวกสวิตซ์และบริดจ์

เกณฑ์วัดที่ 4: เวลาจิตเตอร์ (Jitter)

เวลาจิตเตอร์เป็นความผันผวนของเวลาหน่วง (Delay variation) หรือเป็นความแตกต่างของเวลาลาเทนซี่ที่เกิดขึ้นกับแพ็กเก็ต หรืออธิบายด้วยสูตรคณิตศาสตร์ดังนี้ Jitter = $|Latency(P_n) - Latency(P_{n-1})|$ เมื่อ n คือลำดับที่ของแพ็กเก็ต สมมติฝ่ายส่งได้ส่งแพ็กเก็ตออกไปโดยที่แต่ละแพ็กเก็ตมีเวลาห่างกัน 30 ms หากเวลาที่แต่ละแพ็กเก็ตเดินทางไปถึงผู้รับห่างกัน 30 ms เท่ากันตลอด หมายความว่า Jitter มีค่าเป็นศูนย์ แต่ถ้าเวลาระหว่างสองแพ็กเก็ตที่ไปถึงผู้รับมีค่ามากกว่า 30 ms แสดงว่าแพ็กเก็ตที่ส่งออกครั้งหลังใช้เวลาในการไปถึงปลายทางมากกว่าแพ็กเก็ตก่อนหน้า สำหรับการส่งข้อมูลในเครือข่ายไอพีที่แพ็กเก็ตมีอิสระในการเลือกเส้นทาง หากระยะเวลาระหว่างแพ็กเก็ตมีค่าไม่เท่ากันจะส่งผลให้ข้อมูลไปถึงปลายทางผิดลำดับ ซึ่งจะกระทบโดยตรงกับ Application ที่ต้องการทำงานแบบเวลาจริง (Real time) เช่น ข้อมูลเสียงและ

ภาพเคลื่อนไหว หากแพ็กเก็ตใช้เวลาในการมาถึงปลายทางไม่เท่ากัน จะส่งผลให้การเล่นเสียงและภาพเกิดการขาดช่วง หรือกระตุกได้ เพื่อให้เข้าใจถึงเวลาลาเทนซี่ และเวลาจิตเตอร์ สามารถศึกษาจากรูปที่ 2.8



รูปที่ 2.8 แสดงเวลาลาเทนซี่ และเวลาจิตเตอร์

<u>เกณฑ์วัดที่ 5:</u> การสูญหาย (Loss)

การสูญหายในที่นี้ หมายรวมถึง ความผิดพลาดในระดับบิตของแพ็กเก็ตในขั้นตอนการนำส่งข้อมูล แพ็กเก็ตที่ได้สูญหายจริง และแพ็กเก็ตที่ถูกทิ้งเมื่อเกิดคับคั่งขึ้นในเครือข่าย

เกณฑ์วัดทั้ง 5 นี้ส่งผลต่อการใช้งานหรือการยอมรับได้ของการใช้งาน Application แต่ละประเภท เช่น Video Conference ต้องการบริการที่มีค่าหน่วงต่ำ ในขณะที่อาจจะมีการสูญหายของข้อมูลได้บ้าง แต่งานจำพวก E-Commerce, E-mail และ Web มีความต้องการความถูกต้องของข้อมูลสูง และหากมีการส่งข้อมูลช้าบ้างสามารถยอมรับได้ เป็นต้น การที่จะทำให้เกิด QoS บนเครือข่ายนั้นการเข้าใจในธรรมชาติของการสื่อสารของ Application เป็นเรื่องสำคัญอย่างหนึ่ง เพื่อให้ผู้ดูแลสามารถกำหนดวิธีการให้บริการแก่งานเหล่านั้นได้อย่างเหมาะสมต่อไป

2.1.3.3 กลไกของ QoS (QoS Mechanism) [14, p.16]

การพัฒนา QoS เพื่อนำมาใช้กับการจัดลำดับนำส่งข้อมูลบนเครือข่ายคอมพิวเตอร์ แบ่งออกได้เป็น 3 ส่วน คือ การจัดการกับคิว การจำแนกจัดกลุ่มแพ็กเก็ต และการจัดลำดับนำส่งข้อมูล โดยมีรายละเอียดในแต่ละส่วนดังนี้

1) การจัดการกับคิว (Queueing)

คิวเป็นเนื้อที่ในหน่วยความจำของอุปกรณ์เครือข่ายที่ได้รับการจัดสรรขึ้นเพื่อจัดเก็บแพ็กเก็ตที่มาถึง ซึ่งอาจมีเพียงคิวเดียว หรือมากกว่านั้นขึ้นอยู่กับเทคนิคที่ใช้ สำหรับงานในส่วนนี้จะเกี่ยวข้องกับ

- การนำแพ็กเก็ตที่ผ่านการจำแนกเข้าสู่คิวที่เหมาะสม
- การจัดการกับแพ็กเก็ตเมื่อคิวเต็ม หรือกำลังจะเต็ม
- การนำแพ็กเก็ตออกจากคิวเมื่อแพ็กเก็ตได้รับเลือกในขั้นตอนการจัดลำดับนำส่งข้อมูล

2) การจำแนกจัดกลุ่มแพ็กเก็ต (Classification)

เมื่อแพ็กเก็ตเดินทางมาถึงอุปกรณ์เครือข่าย แพ็กเก็ตจะได้รับการจัดกลุ่มเพื่อนำไปจัดเก็บในคิวใดที่เหมาะสม ซึ่งเกณฑ์การตัดสินขึ้นอยู่กับกฎ หรือเงื่อนไขที่ได้กำหนดไว้แล้ว เช่น การจัดกลุ่มตามวิธี Differentiated Service (DS) ซึ่งได้อธิบายในหัวข้อ 2.1.3.5 จะอาศัยฟิลด์ DS ในการจำแนกแพ็กเก็ต สำหรับวิธีอื่น ๆ อาจอาศัยฟิลด์ที่ส่วนหัวของไอพีมาใช้พิจารณาจำแนกแพ็กเก็ต เช่น หมายเลขไอพี และหมายเลขพอร์ต เป็นต้น

3) การจัดลำดับนำส่งข้อมูล (Scheduling)

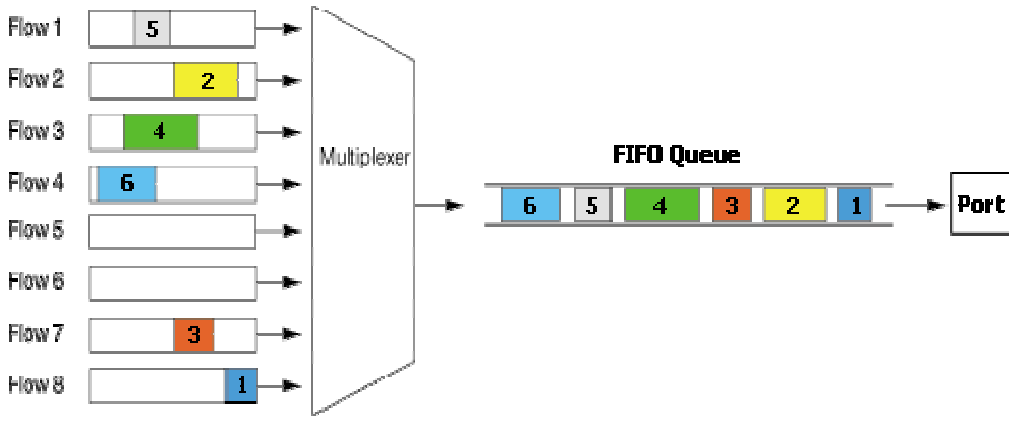
การเลือกแพ็กเก็ตออกจากคิว จำเป็นต้องพิจารณาว่าแพ็กเก็ตใดควรได้บริการก่อน แพ็กเก็ตใดจะได้รับบริการหลัง ตามข้อกำหนดที่ตั้งไว้ หรือเรียกว่าการจัดลำดับนำส่งแพ็กเก็ตที่ต้องการออกจากคิว

2.1.3.4 เทคนิคการจัดคิว (Queueing Disciplines) [22] [25] [31]

เทคนิคการจัดการคิว และวิธีการเลือกแพ็กเก็ตเพื่อการนำส่งที่มีการใช้งานอยู่ทั่วไป เช่น First-in First-out (FIFO) หรือ First-come First-served (FCFS), Priority Queue (PQ), Class-Based Queue (CBQ) และ Weighted Fair Queue (WFQ) เป็นต้น โดยแต่ละเทคนิคมีรายละเอียดดังนี้

1) First-in First-out (FIFO)

วิธีการจัดคิวแบบนี้ถือเป็นเทคนิคพื้นฐานที่สุดที่ใช้อยู่บนเราเตอร์ ในภาวะที่เกิดความคับคั่งในเครือข่าย เราเตอร์จะรับแพ็กเก็ตที่เข้ามาจากแต่ละช่องทางมาพักไว้ในคิวตามลำดับการมาถึง หลังจากนั้นจึงส่งแพ็กเก็ตออกไปตามลำดับ ดังแสดงในรูปที่ 2.9 สำหรับการจัดการหน่วยความจำเมื่อคิวเต็มใช้วิธีการทิ้งแพ็กเก็ตที่มาทีหลัง (Drop-tail)



รูปที่ 2.9 แสดงการจัดคิวแบบ First-In First-Out (FIFO) [25]

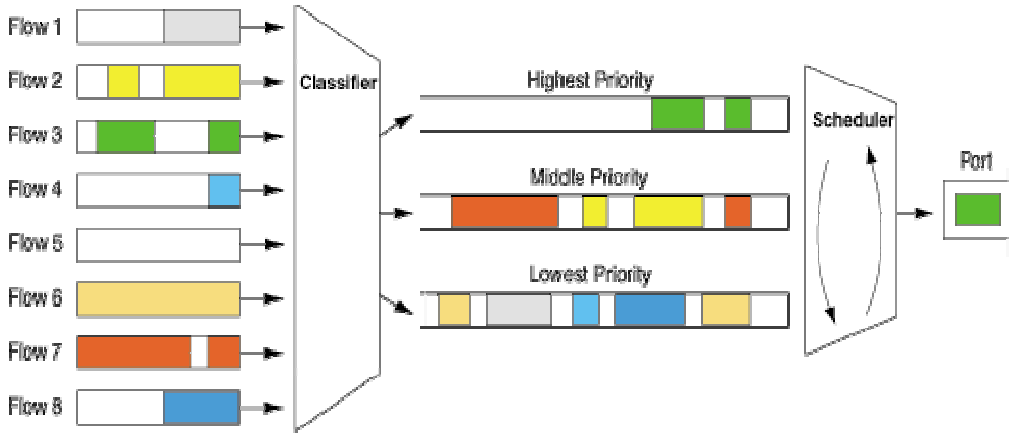
ข้อดีของการจัดคิวแบบ FIFO คือ ทำให้เราเตอร์สามารถส่งแพ็กเก็ตได้อย่างรวดเร็ว และลำดับของแพ็กเก็ตยังคงเดิม เนื่องจากไม่มีการเปรียบเทียบจัดกลุ่ม แต่วิธีการจัดคิวลักษณะนี้ได้ส่งผลกระทบโดยตรงกับ Application ที่จำเป็นต้องได้รับการโต้ตอบแบบทันทีทันใด นอกจากนี้ในช่วงที่เกิดความคับคั่ง แพ็กเก็ตใหม่ต้องรออยู่ในคิวนานจนกว่าจะถึงลำดับของตัวเองที่จะได้รับนำส่งเข้าสู่เครือข่าย เวลาหน่วงที่เกิดขึ้นได้ส่งผลให้เวลาจิตเตอร์สูงตามด้วย เมื่อมีแพ็กเก็ตในคิวมากขึ้น จนกระทั่งมีการทิ้งแพ็กเก็ต หากแพ็กเก็ตที่ถูกทิ้งใช้ TCP ในการนำส่งข้อมูล จะทำฝ่ายส่งต้องปรับข้อกำหนดที่เกี่ยวข้องกับการนำส่งแพ็กเก็ตตามสภาวะของเครือข่าย เช่น ลดขนาดของวินโดวส์ เป็นต้น ทำให้มีแพ็กเก็ต TCP เข้าสู่เครือข่ายน้อยลง ในขณะที่แพ็กเก็ตที่ใช้ UDP ยังคงใช้งานเครือข่ายตามข้อกำหนดเดิม นั่นคือ TCP เสียเปรียบในการใช้งานเครือข่ายแก่ UDP

2) Priority Queue (PQ)

เทคนิคการจัดคิวแบบ PQ ได้แบ่งหน่วยความจำของเราเตอร์ออกเป็นหลายคิว เพื่อรองรับแพ็กเก็ตที่มีระดับความสำคัญหรือมีความต้องการระดับบริการต่างกัน ดังรูปที่ 2.10 วิธีนี้ให้บริการแพ็กเก็ตที่มีระดับความสำคัญสูงก่อนเสมอ แพ็กเก็ตระดับความสำคัญต่ำกว่าจะไม่ได้รับบริการเลยหากยังคงมีแพ็กเก็ตในคิวที่มีระดับความสำคัญสูงกว่า ซึ่งการกำหนดความสำคัญของแพ็กเก็ตมีหลายวิธี เช่น การจำแนกด้วยประเภทของโปรโตคอลโดยกำหนดให้ TCP มีความสำคัญสูงกว่า UDP การจำแนกด้วย Application โดยกำหนดให้ Telnet มีความสำคัญสูงกว่า FTP หรือจำแนกจากแหล่งต้นทางของแพ็กเก็ตโดยกำหนดให้แพ็กเก็ตที่มาจากเราเตอร์สำคัญสูงกว่าแพ็กเก็ตของผู้ใช้งาน เป็นต้น

การจัดการกับหน่วยความจำเมื่อคิวเต็มตามวิธีของ PQ จะอาศัยการกำจัดแพ็กเก็ตใหม่ที่พึ่งมาถึงมี 2 วิธี คือ Preemptive discard algorithm และ Nonpreemptive discard algorithm โดยที่ Preemptive discard algorithm เลือกกำจัดแพ็กเก็ตที่มีระดับความสำคัญต่ำ เพื่อให้หน่วยความจำในเครื่องมีที่ว่างสำหรับแพ็กเก็ตที่มีระดับความสำคัญสูงกว่า ในขณะที่ Nonpreemptive discard algorithm เลือกทิ้งแพ็กเก็ตใหม่ที่ต้องการเข้าสู่คิวที่เต็มแล้ว โดยไม่ให้ความสนใจในเรื่องระดับความสำคัญของแพ็กเก็ต

ข้อดีของ PQ คือ สามารถจัดการเพื่อให้รับประกันได้ว่าแพ็กเก็ตของ Application หลักของหน่วยงานได้รับการนำส่งก่อน หรือ พร้อมที่จะได้รับบริการทันที แต่มีข้อเสียที่อาจทำให้แพ็กเก็ตที่มีความสำคัญต่ำกว่าต้องรออยู่ในคิวนาน หรืออาจไม่ได้รับบริการเลย เมื่อมีแพ็กเก็ตในระดับความสำคัญสูงกว่าเข้าสู่เครือข่ายเป็นจำนวนมาก และตลอดเวลา

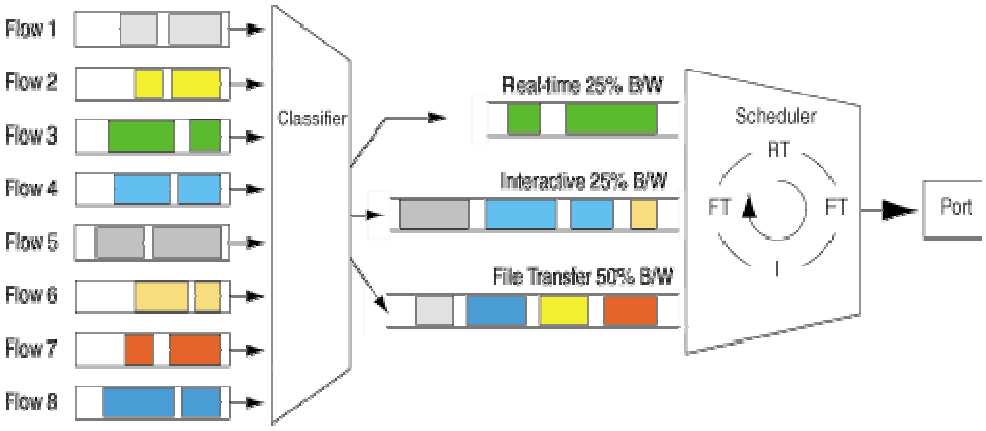


รูปที่ 2.10 แสดงการจัดคิวแบบ Priority Queue (PQ) [25]

3) Class-Based Queue (CBQ)

การจัดการคิวแบบ CBQ บางครั้งเรียกว่า Custom Queue (CQ) เป็นวิธีการจัดคิวที่มีการจำแนกแพ็กเก็ตออกเป็นกลุ่มเหมือนกับ PQ แต่วิธีการเลือกให้บริการแพ็กเก็ตต่างกัน โดย CBQ เลือกให้บริการแพ็กเก็ตในคิวต่าง ๆ แบบหมุนวน (Round-Robin) ตามข้อกำหนดที่ตั้งไว้ เพื่อแก้ปัญหาการรอโดยไม่มีกำหนดของแพ็กเก็ตที่มีระดับความสำคัญต่ำ ผู้ดูแลระบบสามารถกำหนดค่าสำหรับให้บริการแพ็กเก็ตในแต่ละคิวเพื่อประกันขนาดของช่องทางต่ำสุดที่แต่ละคิวจะได้รับ ซึ่งอาจอยู่ในรูปของขนาดไบต์ข้อมูล หรือจำนวนแพ็กเก็ต ดังนั้นในแต่ละรอบบริการแบบหมุนวน คิวต่าง ๆ จะได้รับบริการอย่างน้อยที่สุดตามค่าที่ได้ระบุไว้

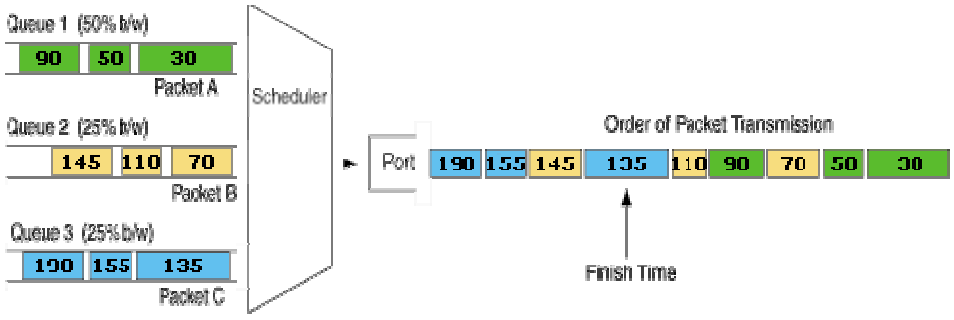
ในการกำหนดลักษณะการให้บริการแก่แพ็กเก็ตในคิวแต่ละระดับความสำคัญนั้น รูปที่ 2.11 เป็นตัวอย่างการกำหนดการให้บริการแบบถ่วงน้ำหนักให้กับ CBQ หรืออาจเรียกวิธีนี้ว่า Weigthed Round-Robin (WRR) นั่นคือกำหนดให้ข้อมูลแบบ Real-time (RT), Interactive (I) และ File transfer (FT) สามารถใช้ Bandwidth ได้ร้อยละ 25 25 และ 50 ตามลำดับ เนื่องจากมีระบบถ่วงน้ำหนักทำให้ลำดับของแพ็กเก็ตที่ออกไปจึงเป็นดังนี้ คือ RT FT I และ FT ตามลำดับ หากพิจารณาตามเทคนิคเดิมของ CBQ อาจกล่าวได้ว่าใน 1 รอบบริการ ระบบต้องให้บริการแก่ FT 2 ครั้ง หรือนำส่งออก 2 แพ็กเก็ต นำส่งแพ็กเก็ตประเภทอื่นรอบละ 1 แพ็กเก็ต ซึ่งวิธีการนี้จะได้ผลดีมากเมื่อแพ็กเก็ตในเครือข่ายมีขนาดเท่ากันทุกแพ็กเก็ต สำหรับเครือข่ายแบบไอพีที่แพ็กเก็ตมีขนาดหลากหลายวิธีนี้อาจไม่ใช่วิธีเหมาะสมมากนักโดยเฉพาะอย่างยิ่งกับเครือข่ายที่มีแพ็กเก็ตขนาดแตกต่างกันมาก



รูปที่ 2.11 แสดงการจัดคิวแบบ Class-Based Queue (CBQ) ที่มีการถ่วงน้ำหนัก [25]

4) Weighted Fair Queuing (WFQ)

WFQ หรือบางครั้งเรียกว่า Flow-based WFQ เป็นการจัดสรร Bandwidth ให้กับแต่ละคิวตามค่าถ่วงน้ำหนักเป็นค่าร้อยละของขนาด Bandwidth ของช่องทางออก WFQ จะคำนวณเวลาสิ้นสุดของการนำส่งแพ็กเก็ต (Finish time) ทุกแพ็กเก็ตที่อยู่ในคิวตามค่าตัวแปรต่อไปนี้ คือ ความเร็วของการสื่อสารข้อมูล จำนวนของคิวที่มีแพ็กเก็ตข้อมูล ค่าถ่วงน้ำหนัก และขนาดของทุกแพ็กเก็ตในคิว ในการเลือกนำส่งแพ็กเก็ตของ WFQ จะเลือกแพ็กเก็ตที่มีคาดว่าใช้เวลาในการนำส่งน้อยสุดก่อน ดูรูปที่ 2.12 ประกอบ แพ็กเก็ต A มีค่าเวลาสิ้นสุดที่ 30 ในขณะที่แพ็กเก็ต B มีค่าเวลาสิ้นสุดที่ 70 และ แพ็กเก็ต C มีค่าเวลาสิ้นสุดที่ 135 ดังนั้นลำดับการนำส่งจึงเป็นการนำส่งแพ็กเก็ตในคิวแรก 2 แพ็กเก็ต คิวที่สอง 1 แพ็กเก็ต คิวแรกอีก 1 แพ็กเก็ต คิวที่สองอีก 1 แพ็กเก็ต และคิวที่สาม 1 แพ็กเก็ต เรื่อยไปดังภาพ



รูปที่ 2.12 แสดงการจัดคิวแบบ Weighted Fair Queue (WFQ) [25]

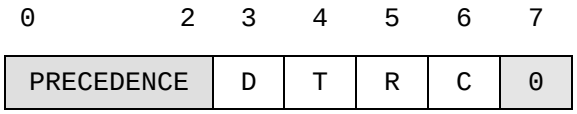
ข้อดีของวิธีการนี้ คือ เกิดการแบ่งบันการใช้ Bandwidth ตามค่าถ่วงน้ำหนัก และสามารถใช้งานได้ดีกับเครือข่ายที่มีขนาดของแพ็กเก็ตแตกต่างกันมาก แต่ข้อจำกัดหนึ่งที่สำคัญของวิธีนี้ คือ ใช้การคำนวณที่ซับซ้อน และต้องกระทำกับทุก ๆ แพ็กเก็ตที่ผ่านเข้ามา ซึ่งจะส่งผลกระทบต่อการให้บริการอย่างมากเมื่อมีการใช้งานเครือข่ายเพิ่มขึ้น หรือเครือข่ายที่มีจำนวนแพ็กเก็ตมาก

2.1.3.5 สถาปัตยกรรมของ QoS (QoS Architecture)

นอกจากเทคนิคที่ใช้ในการจัดการคิวแล้ว ระบบการรับส่งข้อมูลก็มีความสำคัญ ตัวอย่างของสถาปัตยกรรม หรือแนวทางของระบบการรับส่งข้อมูลที่ก่อให้เกิด QoS เช่น Type of Service Routing, Integrated Service (IntServ: IS), Differentiated Service (DiffServ: DS) และ Multiprotocol Label Switching (MPLS) โดยรายละเอียดของแต่ละวิธีการเป็นดังนี้

1) Type of Service Routing

สถาปัตยกรรมนี้ได้อาศัยฟิลด์ Type of Service ขนาด 1 ไบต์ ในส่วนหัวของ IP datagram ที่ได้ระบุถึงรูปแบบการให้บริการที่ IP datagram ควรได้รับมาใช้ในการเลือกวิธีการรับส่งข้อมูล เพื่อให้เราเตอร์สามารถเลือกส่ง IP datagram ไปในเส้นทางที่เหมาะสม และสอดคล้องกับความต้องการได้ ฟิลด์นี้แบ่งออกได้ 3 ส่วนดังรูปที่ 2.13 ส่วนแรกมีขนาด 3 บิต เรียกว่า Precedence ใช้สำหรับจัดลำดับความสำคัญของ IP datagram ซึ่งมีประโยชน์ในการแก้ปัญหาบางอย่างในระบบ เช่น เมื่อเกิดความคับคั่งในเครือข่าย IP datagram ที่ใช้ควบคุมการทำงานของเครือข่ายซึ่งมีระดับความสำคัญสูงจะยังคงได้รับการบริการให้ส่งผ่านไปได้ เป็นต้น ส่วนที่สองมีขนาด 4 บิต ซึ่งมีรหัสแทนแต่ละบิตคือ D T R และ C เพื่อใช้ในการกำหนดคุณภาพการให้บริการ ซึ่งมีความหมายดังนี้ Delay, Throughput, Reliability และ Cost ตามลำดับ โดยปกติเราเตอร์สามารถให้บริการได้ดีที่สุดเพียงลักษณะเดียวเท่านั้น จึงทำให้มีเพียงบิตเดียวเท่านั้นที่มี ค่าเป็น 1 นอกนั้นจะมีค่าเป็น 0 สำหรับบิตสุดท้ายยังมิได้นำไปใช้งานจึงกำหนดให้เป็นค่า 0 ไว้



รูปที่ 2.13 แสดงรูปแบบไบต์ของ Type of Service

รายละเอียดค่าข้อมูลในแต่ละบิต และความหมายของแต่ละค่ามีดังนี้

บิตที่ 0 - 2: Precedence Control ใช้กำหนดลำดับความสำคัญของ IP datagram มี 8 ระดับ จาก 7 (สูงสุด) ถึง 0 (ต่ำสุด) ดังรายละเอียดในตารางที่ 2.1 โดยเราเตอร์จะบริการ IP datagram ที่มีความสำคัญสูงกว่าก่อน

บิตที่ 3 - 6: ใช้กำหนดลักษณะของคุณภาพบริการ โดยมีค่าและความหมายตามตารางที่ 2.2

Precedence	ค่าแทนเลขฐานสิบ	ความหมาย
111	7	Network Control
110	6	Internetwork Control
101	5	Critical function
100	4	Flash Override
011	3	Flash
010	2	Immediate
001	1	Priority
000	0	Routine

ตารางที่ 2.1 แสดงค่าของ Precedence ใน Type of Service

บิตที่ 3 - 6	ความหมาย
1000	Minimize **D**elay หรือ เมื่อต้องการให้เราเตอร์ส่งผ่าน IP datagram บนเส้นทางที่มีเวลาหน่วงในการรับส่งต่ำสุด
0100	Maximize **T**hroughput หรือ เมื่อต้องการให้เราเตอร์ส่งผ่าน IP datagram บนเส้นทางที่มีความจุสูงสุด
0010	Maximize **R**eliability หรือ เมื่อต้องการให้เราเตอร์ส่งผ่าน IP datagram บนเส้นทางที่มีความน่าเชื่อถือสูง หรือ มีอัตราความผิดพลาดต่ำ
0001	Minimize monetary **C**ost หรือเมื่อต้องการส่งผ่านเส้นทางที่มีค่าใช้จ่ายน้อยสุด
0000	Normal best-effort service หรือเมื่อต้องการส่งแบบปกติ

ตารางที่ 2.2 แสดงค่าของบิตที่ 3 ถึงบิตที่ 6 ใน Type of Service

ตามเอกสาร RFC หมายเลข 1122 และ 1123 ได้ระบุว่าโดยทั่วไปหากกล่าวคำว่า ToS จะหมายถึงตำแหน่ง 5 บิตหลัง (Low-order 5 bits) ในฟิลด์ของ Type of Service สำหรับการใช้งานบิตกลุ่มนี้สามารถศึกษาได้จาก RFC 1349 ตัวอย่างการกำหนดค่าให้บิต D, T, R และ C ที่ได้รับการแนะนำเพื่อใช้งานกับ Application แต่ละอย่างได้แสดงในตารางที่ 2.3

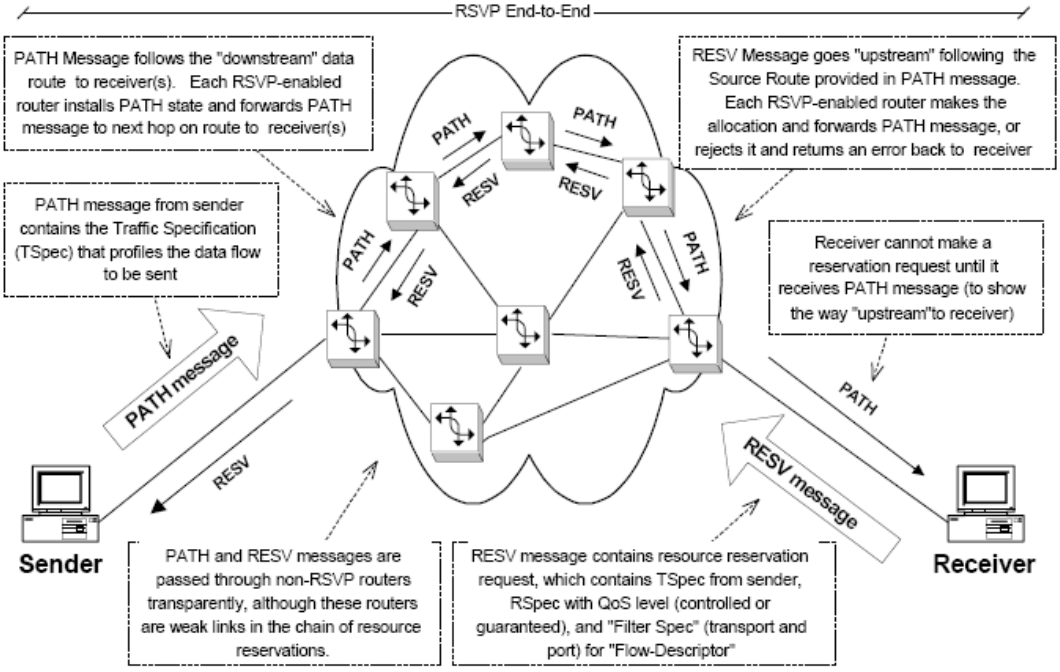
Application	D	T	R	C	ระดับบริการ
ICMP	0	0	0	0	Normal best-effort service
BOOTP	0	0	0	0	Normal best-effort service
NNTP	0	0	0	1	Maximize Reliability
IGP	0	0	1	0	Maximize Reliability
SNMP	0	0	1	0	Maximize Reliability
Telnet	1	0	0	0	Minimize Delay
FTP (data)	0	1	0	0	Maximize Throughput
FTP (control)	1	0	0	0	Minimize Delay
TFTP	1	0	0	0	Minimize Delay
SMTP (command)	1	0	0	0	Minimize Delay
SMTP (data)	0	1	0	0	Maximize Throughput
DNS (UDP query)	1	0	0	0	Minimize Delay
DNS (TCP query)	0	0	0	0	Normal best-effort service
DNS (zone)	0	1	0	0	Maximize Throughput

ตารางที่ 2.3 แสดงตัวอย่างการกำหนดบิต D, T, R และ C ของ Application ต่าง ๆ [3, น.63]

2) Integrated Service [RFC 1633] [31]

IntServ เป็นสถาปัตยกรรมที่ให้ความสนใจกับ Application ที่ต้องได้รับบริการตามเวลาจริง (Real-time service) เป็นหลัก เช่น เสียงและภาพเคลื่อนไหว สถาปัตยกรรมนี้ต้องการเราเตอร์ที่มีความสามารถในการจอง Bandwidth ของเครือข่ายด้วยโปรโตคอล RSVP (Resource Reservation Protocol) เพื่อให้ Application นั้นได้รับการบริการที่เหมาะสม และให้ผู้ใช้เกิดความพอใจ

RSVP เป็นการรับประกันช่องสื่อสารสำหรับการนำส่งข้อมูลจากต้นทางไปยังปลายทาง ทั้งแบบ Unicast หรือ Multicast โดยหลักการทำงานของ RSVP อธิบายพอสังเขป ควบคู่กับรูปที่ 2.14 ดังนี้



รูปที่ 2.14 แสดงหลักการทำงานของ Resource Reservation Protocol [31, p 7]

- เครื่องต้นทางที่เป็นฝ่ายส่ง (Sender) พิจารณาค่าต่ำสุด และสูงสุดของตัวแปรต่อไปนี้ เช่น Bandwidth เวลาหน่วง และเวลาจิตเตอร์ ซึ่งใช้เป็นรายละเอียดจราจร (Traffic Specification: Tspec) ของข้อมูลที่ต้องการนำส่ง หลังจากนั้น RSVP จะส่ง PATH message ที่ประกอบด้วย Tspec และ ข้อมูลอื่น ๆ ที่เกี่ยวข้องผ่านเราเตอร์ในเครือข่ายไปยังปลายทางที่เป็นฝ่ายรับ (Receiver)
- เมื่อฝ่ายรับได้รับ PATH message แล้ว ฝ่ายรับจะส่งคำร้องการจองทรัพยากร (Reservation request: RESV) ตามข้อกำหนดใน Tspec กลับตามเส้นทางเดิมไปยังฝ่ายส่ง ข้อมูลใน RESV จะมีรายละเอียดคำร้อง (Request specification: Rspec) ที่เกี่ยวข้องกับของ QoS และ รายละเอียดการกรอง (Filter specification: Filter spec) ที่ระบุรายละเอียดบางอย่างของแพ็กเก็ต เช่น โปรโตคอลที่ใช้ และหมายเลขพอร์ต เป็นต้น

- เมื่อเราเตอร์ในเส้นทางสื่อสารที่สนับสนุน RSVP ได้รับคำร้อง RESV เราเตอร์จะตรวจสอบว่าสามารถจองทรัพยากรตามที่ต้องการได้หรือไม่ หากไม่สามารถดำเนินการได้ เช่น มีทรัพยากรไม่เพียงพอ หรือไม่ได้รับสิทธิ เราเตอร์จะแจ้งให้ทางฝ่ายรับทราบถึงข้อผิดพลาดที่เกิดขึ้น แต่ถ้าเราเตอร์สามารถจองทรัพยากรได้เราเตอร์จะส่ง RESV ไปยังเราเตอร์ตัวถัดไป
- เมื่อเราเตอร์ตัวสุดท้ายในเส้นทางได้รับ RESV และยอมรับคำร้องขอที่ได้ระบุมา จะทำให้ได้ช่องทางที่ได้รับการประกันสำหรับการส่งข้อมูล

นอกจากการจองทรัพยากร IntServ ยังอาศัย Token-bucket เพื่อควบคุมให้ข้อมูลที่จะไหลเข้าสู่เครือข่ายมีความราบเรียบและสอดคล้องกับสภาพเครือข่ายในขณะนั้น โดยตัวแปรควบคุมของ Token-bucket เป็นข้อมูลส่วนหนึ่งของ Tspec และ Rspec

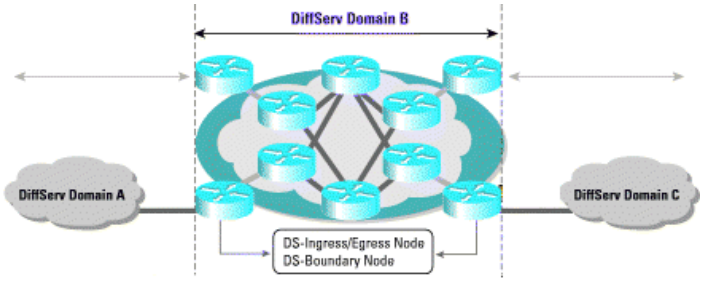
3) Differentiated Service [RFC 2475] [23] [31]

แม้ว่าการใช้ IntServ ได้ก่อให้เกิด QoS ในระดับหนึ่ง แต่ยังไม่ได้ผลดีเท่าที่ควรเมื่อนำไปใช้บนเครือข่ายอินเทอร์เน็ต เพราะการจองทรัพยากรของระบบเครือข่ายที่มีอยู่อย่างจำกัดออกมาอีก ทำให้ทรัพยากรของเครือข่ายอินเทอร์เน็ตในบางส่วนอาจถูกใช้อย่างไม่เต็มประสิทธิภาพ Differentiated Service หรือ DiffServ จึงเป็นแนวทางหนึ่งเพื่อเกิด QoS ในเครือข่ายซึ่งเรียกว่า โดยมีแนวคิดพื้นฐานที่ว่า Application แต่ละประเภทไม่ว่าจะเป็นภาพเคลื่อนไหว การถ่ายทอดเสียง การค้าแบบอิเล็กทรอนิกส์ และการบริการข้อมูลข่าวสาร เป็นต้น มีความต้องการ QoS ที่แตกต่างกัน ตัวอย่างเช่น การถ่ายทอดข้อมูลเสียงจะดีได้ต้องได้รับการบริการที่มีค่าจิตเตอร์ต่ำ และ Bandwidth ในช่วง 8 - 64 Kbps [31] ขึ้นอยู่กับวิธีการเข้ารหัสข้อมูลเสียง ในขณะที่การถ่ายโอนข้อมูลเอฟทีพีนั้นต้องการการรับประกันเรื่องของข้อมูลสูญหายมากกว่าการให้ความสำคัญเรื่องจิตเตอร์ เป็นต้น โครงสร้างของ DiffServ ประกอบด้วยกลุ่มของโหนดที่ต่อกันในเครือข่าย DS (DiffServ domain) ดังที่ได้แสดงในรูปที่ 2.15 โหนดที่สนับสนุน DiffServ เรียกว่า DS-compliant node หรือ DS node โหนดในเครือข่าย DS สามารถแบ่งได้ 2 กลุ่ม คือ DS Boundary Node (Edge) และ DS Interior Node (Core)

DS Boundary Node (Edge) คือ โหนดที่เชื่อมต่อ DS domain กับ DS domain อื่น หรือเครือข่ายอื่น โหนดกลุ่มนี้มี 2 ลักษณะ คือ

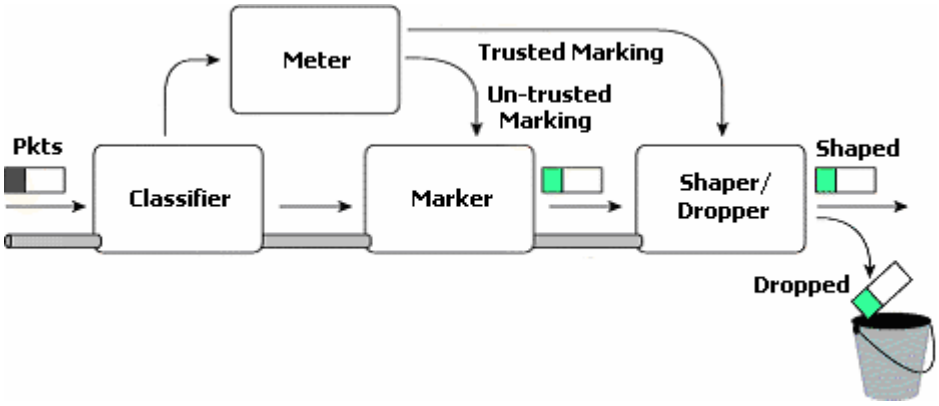
- DS Ingress Node เป็นจุดที่จราจรเข้าสู่ DS domain และ
- DS Egress Node เป็นจุดที่จราจรออกจาก DS domain

DS Interior Node (Core) คือ โหนดที่เชื่อมต่อกันอยู่ภายใน DS domain

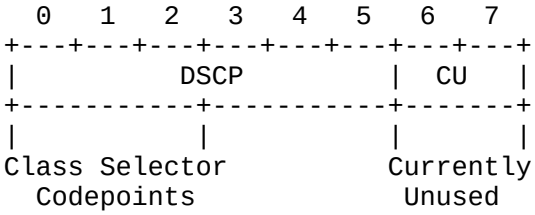


รูปที่ 2.15 แสดงสถาปัตยกรรมแบบ DiffServ [23]

เมื่อแพ็กเก็ตของ Application เดินทางมายัง DS Ingress Node โหนดนี้มีกระบวนการทำงานที่เกี่ยวข้องกัน 4 ขั้นตอนตามรูปที่ 2.16 หรือที่เรียกว่า Traffic Conditioner คือ Classifier, Meter, Marker และ Shaper/Dropper โดย **Classifier** จะอาศัยค่า Precedence ของ Type of Service มาสร้างแท็ก (Tag) ที่เรียกว่า Differentiated Service Field (DS-Field) ขนาด 8 บิต ดังรูปที่ 2.17 เพื่อกำหนดระดับของ QoS ที่เหมาะสมให้แพ็กเก็ต ซึ่งปัจจุบันใช้เพียง 6 บิตแรกที่เรียกว่า Differentiated Service Code Point (DSCP) เท่านั้น ทำให้ DiffServ สามารถแบ่งระดับของ QoS ได้ 2^6 หรือ 64 แบบ โดยค่า DSCP จะถูกนำมาใช้เพื่อหาวิธีการนำส่งที่เรียกว่า Per-Hop Behavior (PHB) ซึ่งมีความเกี่ยวกับขนาดของ Bandwidth ที่จะได้รับ วิธีการจัดคิว และวิธีการทิ้งแพ็กเก็ตเมื่อคิวเต็ม เป็นต้น หลังจากนั้นแพ็กเก็ตจะผ่าน **Meter** เพื่อตรวจสอบค่าตัวแปรต่าง ๆ ที่ใช้ในการส่งข้อมูลว่าสามารถยอมรับได้หรือไม่ เช่น DSCP ที่มีค่าเท่ากับ X จะใช้ Token-Bucket ด้วยความเร็ว R Mbps และขนาดข้อมูลสูงสุด B Bytes เป็นต้น หากค่าในตัวแปรผ่านการยอมรับแพ็กเก็ตจะถูกส่งต่อไปยัง Shaper มิฉะนั้นจะถูกส่งต่อไปยัง **Marker** ซึ่งมีหน้าที่ปรับค่า DSCP ให้เหมาะตามข้อกำหนดที่ได้จาก Meter ส่วน **Shaper** มีหน้าที่หน่วงเวลาเพื่อให้ส่งแพ็กเก็ตได้ตามข้อกำหนด หาก Shaper มีแพ็กเก็ตมากเกินกว่าหน่วยความจำที่สามารถรองรับได้ **Dropper** มีหน้าที่ทิ้งแพ็กเก็ตเหล่านั้น



รูปที่ 2.16 แสดง Traffic Conditioner Block (TCB) [23]



รูปที่ 2.17 แสดงโครงสร้างของ DS-Filed

ปัจจุบันค่ามาตรฐานของ PHB มีอยู่ 4 แบบ คือ Default PHB, Class-Selector PHB, Expedited Forwarding (EF) PHB และ Assured Forwarding (AF) PHB โดยแต่ละค่ามีคุณสมบัติดังนี้

Default PHB (000000)

เมื่อแพ็กเก็ตเข้ามายัง DS-compliant node และค่า DSCP ที่แนบมาไม่สามารถเลือก PHB ได้ DS-compliant node จะกำหนดค่า ‘000000’ ให้กับ DSCP ซึ่งเป็นค่าปริยาย ทำให้แพ็กเก็ตได้รับบริการแบบดั้งเดิม คือ Best-effort

Class-Selector PHB (xxx000)

ค่า DSCP จะอยู่ในรูปของ ‘xxx000’ โดย x มีค่าเป็น 0 หรือ 1 สำหรับ DSCP 3 บิตแรกเรียกว่า Class Selector Code points โดยแพ็กเก็ตที่มีค่า DSCP เป็น ‘110000’ (IP-Precedence 110) จะได้รับการบริการดีกว่าแพ็กเก็ตที่มีค่า DSCP เป็น ‘100000’ (IP-Precedence 100)

Expedited Forwarding (EF) PHB (101110)

ค่า DSCP ที่แนะนำให้ใช้สำหรับ EF คือ '101110' เหมาะกับแพ็กเก็ตที่ต้องการระดับการสูญหายต่ำ เวลาหน่วงต่ำ ความแปรปรวนของเวลาหน่วงต่ำ และต้องการรับประกัน Bandwidth เช่น แพ็กเก็ตจำพวก VoIP

Assured Forwarding (AF) PHB (xxxyyy)

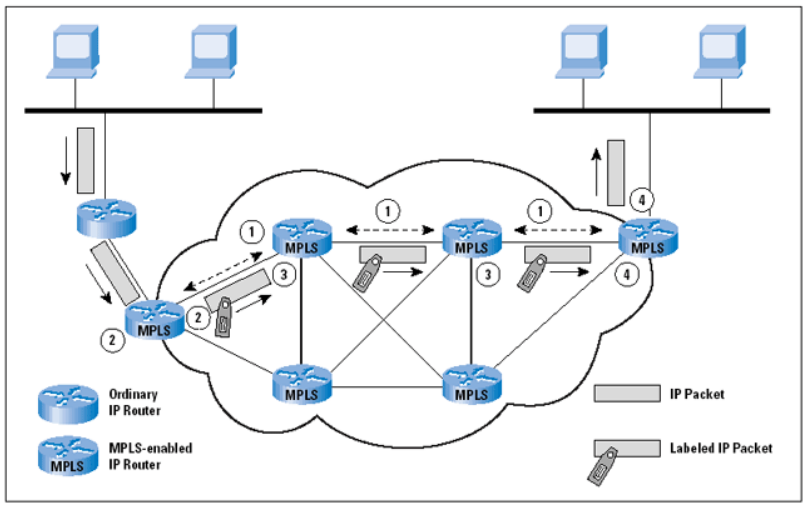
PHB แบบ AF จะแบ่ง Code point เป็น 2 ส่วน ซึ่งเขียนในรูปของ AFxy แทนค่า x ด้วยบิตจำนวน 3 บิต ใช้สำหรับแบ่งกลุ่มของแพ็กเก็ต โดยใช้เพียง 4 ค่า ดังที่ได้แสดงในตารางที่ 2.5 ค่านี้มีผลต่อการกำหนดขนาดของหน่วยความจำหรือบัฟเฟอร์ และขนาดของ Bandwidth ค่า y แทนด้วย 3 บิตหลัง ใช้กำหนดความสำคัญเมื่อจำเป็นต้องทิ้งแพ็กเก็ต มี 3 ระดับ ทำให้ PHB แบบ AF สามารถแบ่ง Code points ได้ 12 แบบ ตามตารางที่ 2.4 เมื่อเกิดความคับคั่งในเครือข่ายแพ็กเก็ตในกลุ่ม x ที่มีค่า y สูงกว่าจะถูกทิ้งก่อน

Drop Precedence	Class#1	Class#2	Class#3	Class#4
Low Drop Prec (1)	(AF11) 001010	(AF21) 010010	(AF31) 011010	(AF41) 100010
Medium Drop Prec (2)	(AF12) 001100	(AF22) 010100	(AF32) 011100	(AF42) 100100
High Drop Prec (3)	(AF13) 001110	(AF23) 010110	(AF33) 011110	(AF43) 100110

ตารางที่ 2.4 แสดง Codepoints ตามมาตรฐาน AF

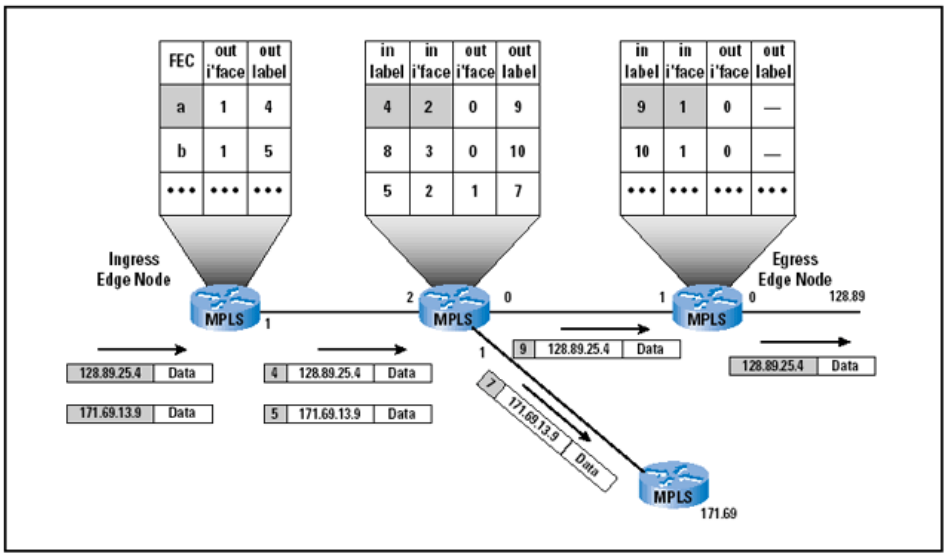
4) Multiprotocol Label Switching (MPLS) [RFC 3031] [6, น.71-73] [31] [32]

MPLS ได้รับการกำหนดให้เป็นมาตรฐานสำหรับควบคุมการจราจรบนเครือข่าย เมื่อปี ค.ศ. 2001 โดยหน่วยงานกำหนดมาตรฐาน IETF (Internet Engineering Task Force) เพื่อแก้ปัญหาความคับคั่งของข้อมูลในเครือข่ายที่นับวันยิ่งมีความรุนแรงมากขึ้น MPLS ได้นำข้อดีของการนำส่งข้อมูลบนเครือข่าย ATM ซึ่งเป็นเครือข่ายแบบ Virtual Circuit Switching กับกระบวนการกำหนดเส้นทางของเราเตอร์บนเครือข่าย TCP/IP มาผสมผสานกัน จึงทำให้ MPLS สามารถรับประกันค่า Throughput ให้กับผู้ใช้ได้ และรองรับ Application ที่ต้องการการโต้ตอบแบบทันทีทันได้ เช่น การถ่ายทอดภาพ และเสียงผ่านเครือข่ายอินเทอร์เน็ต โดยภาพและเสียงที่ได้นั้นมีคุณภาพใกล้เคียงกับการรับชมจากโทรทัศน์ หรือฟังวิทยุจริง



รูปที่ 2.18 แสดงเครือข่าย MPLS [32]

การใช้งาน MPLS เราเตอร์ในเครือข่ายต้องมีความสามารถรองรับการทำงานได้ดังรูปที่ 2.18 เราเตอร์ที่รองรับการทำงานแบบ MPLS จะเรียกว่า Label Switched Router (LSR) การทำงานของ MPLS จะมีการกำหนดเส้นทางของแพ็กเก็ตระหว่างจุด 2 จุด หรือระหว่างเราเตอร์ 2 ตัวไว้อย่างแน่นอน ในที่นี้เรียกว่า Forwarding Equivalence Class (FEC) ซึ่งเส้นทางนี้สามารถเปลี่ยนแปลงได้ตามสภาพของเครือข่าย เมื่อแพ็กเก็ตเดินทางมาถึง LSR ตัวแรกในเครือข่าย เราเตอร์จะพิจารณาว่าแพ็กเก็ตนั้นควรอยู่ใน FEC ใดตามข้อมูลส่วนหัวของแพ็กเก็ต เช่น ตำแหน่งของเครื่องต้นทาง ตำแหน่งของเครื่องปลายทาง หมายเลขเครือข่าย และหมายเลขพอร์ต เป็นต้น สำหรับกระบวนการนี้จะกระทำครั้งเพียงครั้งเดียวเท่านั้น ทำให้ LSR ที่อยู่ภายในเครือข่าย MPLS ไม่ต้องอ่านข้อมูลที่ส่วนหัวของไอพีอีก จากนั้น LSR ตัวแรกจะพิจารณาถึง QoS ของแพ็กเก็ต เช่น ทรัพยากรที่ต้องการ นโยบายในการจัดการคิว นโยบายในการทิ้งแพ็กเก็ต เป็นต้น ซึ่งรายละเอียดเหล่านี้เป็นตัวสร้างเส้นทางในการไหลของแพ็กเก็ตใน FEC เรียกว่า Label Switched Path (LSP) หลังจากนั้นจะระบุค่าที่ได้พิจารณาแล้วใน Label ของแพ็กเก็ต เมื่อ LSR ตัวถัดไปได้รับแพ็กเก็ตจะอ่านค่าใน Label และนำไปเปรียบเทียบกับข้อมูลในตารางว่าจะให้แพ็กเก็ตไหลไปทิศทางใด (Out interface) เราเตอร์จะเอา Label เดิมออก หลังจากนั้นจะใส่ Label ใหม่ตาม Out Label ดังรูปที่ 2.19 หลังจากนั้นจึงปล่อยแพ็กเก็ตออกไปยัง LSR ตัวถัดไป เมื่อแพ็กเก็ตมาถึง LSR ตัวสุดท้ายของเครือข่าย MPLS เราเตอร์จะนำ Label ออกจากแพ็กเก็ต แล้วปล่อยให้เป็นหน้าที่ส่วนหัวเดิมของแพ็กเก็ตในการส่งข้อมูลไปยังปลายทางที่แท้จริง



รูปที่ 2.19 แสดงการส่งต่อแพ็กเก็ตของ MPLS [32]

2.1.4 สรุปทฤษฎีที่เกี่ยวข้อง

เนื้อหาต่าง ๆ ที่ได้อธิบายในหัวข้อนี้ล้วนเป็นความรู้ที่จำเป็นและเกี่ยวข้องกับงานวิจัยนี้ได้เริ่มจากศึกษาความรู้ทั่วไปทางด้านเครือข่าย โปรโตคอลที่ใช้ในการสื่อสารโดยเฉพาะ TCP/IP ข้อมูลสำคัญที่ปรากฏในส่วนหัวของแพ็กเก็ตซึ่งใช้เป็นเงื่อนไขในการกำหนดระดับความสำคัญของแพ็กเก็ตเพื่อให้เกิด QoS ขึ้นบนเครือข่าย ตั้งแต่ Queuing Disciplines, Integrated Service, Differentiated Service และ Multiprotocol Label Switch โดยผู้วิจัยได้นำองค์ความรู้เหล่านี้มาใช้ในการออกแบบเครื่องมือต่าง ๆ ที่จำเป็นในการวิจัยครั้งนี้ ซึ่งจะมีกล่าวถึงในบทถัดไป

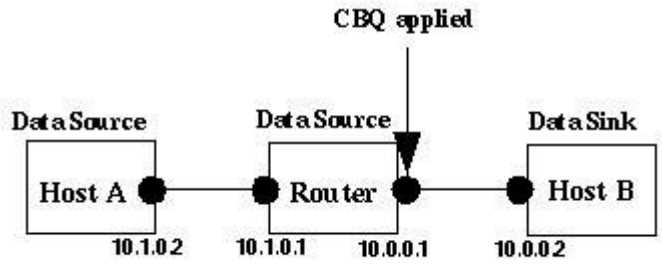
2.2 ผลงานที่เกี่ยวข้องและคล้ายคลึงกับงานที่ทำ

เนื้อหาในส่วนนี้เป็นการทบทวนเอกสารงานวิจัยที่เกี่ยวข้องกับงานวิจัยชิ้นนี้ ประกอบด้วยงานวิจัยเรื่อง “IP-007 IP QoS Testing – QoS services using Alternate Queueing” ของ George Uhl และ Raytheon STX [12], “Core-Stateless Fair Queueing: Acieving Approximately Fair Bandwidth Allocation in High Speed Networks” ของ I.Stoica และคณะ[13], “Design and Implementation of a Virtual Quality of Service MAC Layer (VQML) for Wireless LANs” ของ Mahbub Hassan และคณะ [15], “Quality of Service in IP Networks” ของ Karl Ahlin [14] และงานวิจัยเรื่อง “การประยุกต์ใช้ตัวจำลองระบบเครือข่ายสำหรับเครือข่ายคอมพิวเตอร์

มหาวิทยาลัยสงขลานครินทร์” ของหฤทัย สมบูรณ์รุ่งโรจน์ [11] และได้ศึกษารายละเอียดในส่วนที่เป็นผลิตภัณฑ์ โดยเลือกศึกษา NetScreen [29] ซึ่งผลิตภัณฑ์สำหรับบริหารการใช้งานเครือข่าย สำหรับรายละเอียดได้นำเสนอไว้ในหัวข้อที่ 2.2.1

2.2.1 งานวิจัยที่เกี่ยวข้อง

George Uhl และ Raytheon STX ได้นำเสนอรายงานการทดลองเพื่อทดสอบหน้าที่พื้นฐานของ QoS ในเรื่องการจัดลำดับนำส่งข้อมูลด้วยวิธีการจัดคิวแบบ CBQ และ WFQ เป็นการทดลองในห้องปฏิบัติการที่ประกอบด้วยเครื่องคอมพิวเตอร์จำนวน 3 เครื่องดังรูปที่ 2.20 เป็นเราเตอร์ 1 เครื่อง เป็นโฮสต์ 2 เครื่อง คือ A และ B โดยเครื่องเราเตอร์ใช้ระบบปฏิบัติการ FreeBSD 2.2.6-RELEASE ที่ได้รับการปรับปรุงเคอร์เนลด้วย ALTQ 1.0.1 release QoS OS modifications with ATM driver ของ Kenjiro Cho (http://www.csl.sony.co.jp/person/kjc.html) เพื่อใช้ในการจัดการกับการไหลของแพ็กเก็ต โฮสต์ A และ B ได้รับการเชื่อมต่อเข้ากับเราเตอร์ด้วยเทคโนโลยี ATM OC3 permanent virtual circuits (PVCs) แบบ unspecified bit rate กำหนดให้โฮสต์ A และเราเตอร์ทำหน้าที่เป็นตัวจำลองส่งแพ็กเก็ต (TCP และ/หรือ UDP) ส่วนโฮสต์ B ทำหน้าที่เป็นตัวรับแพ็กเก็ต



รูปที่ 2.20 แสดงเครือข่ายทดสอบงานวิจัยของ Geroge Uhl และคณะ

การทดลองได้ดำเนินตามสถานการณ์ที่ได้กำหนดไว้ที่มีข้อกำหนดแตกต่างกัน ผลการทดลองได้แสดงให้เห็นว่าหากไม่มีการควบคุมการใช้งาน Bandwidth ของโปรโตคอลจำพวก Connectionless หรือ UDP โปรโตคอลกลุ่มนี้จะใช้ Bandwidth ที่มีอยู่ให้มากเท่าที่ทำได้ แต่เมื่อกำหนดให้มีการควบคุมการใช้ Bandwidth ด้วยการจัดคิว พบว่า CBQ สามารถกำหนดให้โปรโตคอลต่าง ๆ ใช้ Bandwidth ได้ตามที่ค่าถ่วงน้ำหนักที่กำหนดไว้ ในขณะที่ WFQ ต้องกำหนดค่าถ่วงน้ำหนักให้กับ TCP มากกว่า UDP การแบ่งใช้งาน Bandwidth จึงเป็นไปอย่างยุติธรรม

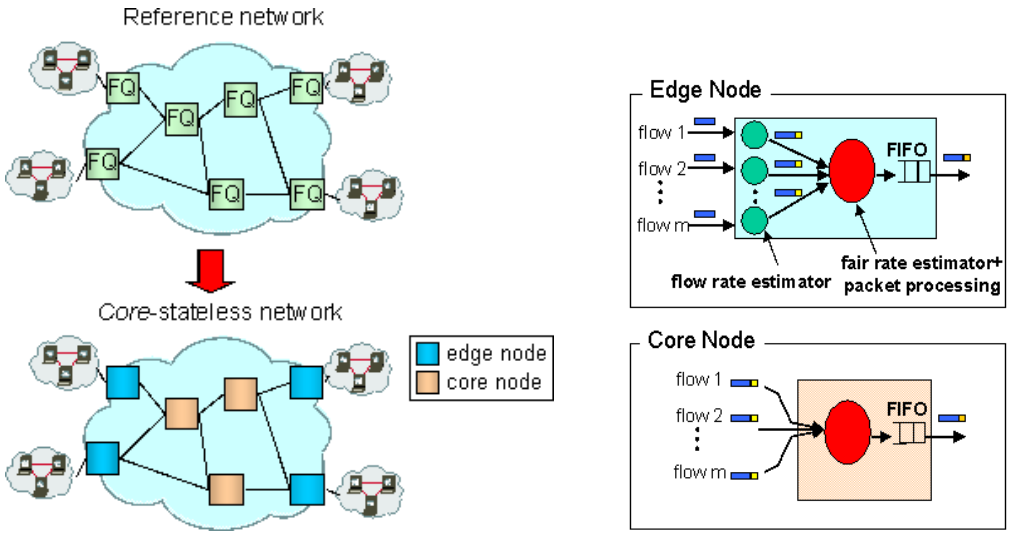
ในปี ค.ศ.1998 Ion Stoica และคณะได้นำเสนอวิธีการจัดสรร Bandwidth วิธีใหม่ในชื่อ CSFQ หรือ Core-Stateless Fair Queueing เพื่อให้อุปกรณ์เครือข่าย เช่น เราเตอร์สามารถจัดสรร Bandwidth ให้กับ Application ต่าง ๆ พร้อมทั้งทำงานได้ที่ความเร็วสูง วิธีการนี้ได้แบ่งเราเตอร์ออกเป็น 2 กลุ่ม คือ กลุ่มที่เป็น Edge node หรือโหนดที่อยู่บริเวณขอบของเครือข่ายที่เชื่อมต่อกับเราเตอร์ของเครือข่ายอื่น และ Core node หรือเราเตอร์ที่อยู่ภายในเครือข่าย CSFQ ดังรูปที่ 2.21 เมื่อแพ็กเก็ตเดินทางมาถึง Edge node ที่เป็น Ingress node หรือโหนดที่เป็นทางเข้า กิจกรรมหนึ่งที่จะเกิดขึ้นเฉพาะบน Edge node คือ หน่วยทำงาน Flow rate estimator จะประมาณค่าอัตราการไหลของจราจรที่เข้ามา (Flow arrival rate, r) และเพิ่มข้อมูลนี้เป็นส่วนหัวของแพ็กเก็ต (Label) ซึ่งส่วนหัวที่เพิ่มเข้าไปใหม่นี้จะถูกตัดทิ้งเมื่อแพ็กเก็ตแพ็กเก็ตเดินทางผ่าน Edge node ที่โหนดทางออกของเครือข่าย หลังจากผ่านกระบวนการประมาณค่าอัตราการไหลแล้วแพ็กเก็ตจะถูกส่งเข้าหน่วยทำงาน Fair rate estimator + Packet Processing ซึ่งกระบวนการนี้จะเกิดขึ้นกับเราเตอร์ทุกโหนดในเครือข่าย CSFQ หน่วยทำงานนี้มีหน้าที่ 3 อย่าง คือ

a) ประมาณอัตราความเป็นธรรม (Link fair rate, f) ตามสถานะของสายสื่อสารขาออก
b) นำส่งแพ็กเก็ตออกจากโหนดตามความน่าจะเป็น (Packet forwarding rate) ที่ได้จากการคำนวณโดยค่าอาศัยอัตราการไหลที่ได้ระบุที่หัวของแพ็กเก็ต

$$P = \min (1, f/r)$$

c) ก่อนแพ็กเก็ตจะถูกนำส่งออกไป หน่วยทำงานจะปรับปรุงค่าอัตราการไหล (r) ที่หัวของแพ็กเก็ตให้มีค่าเป็น

$$r' = \min(r, f)$$



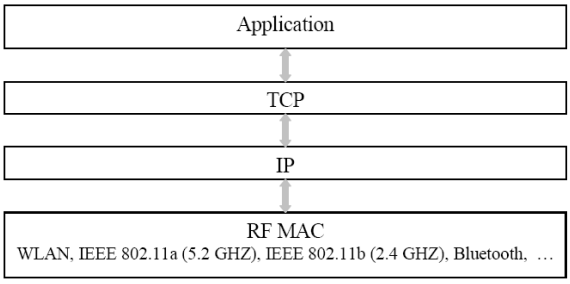
รูปที่ 2.21 แสดงโครงสร้างเครือข่ายแบบ Core-Stateless และหน้าที่ของโหนดในเครือข่าย

ในการทดลองได้เปรียบเทียบประสิทธิภาพของการจัดสรร Bandwidth ของ CSFQ กับกระบวนการวิธีอื่น ๆ ที่มีอยู่เดิม คือ FIFO (First In First Out) RED (Random Early Detection) FRED (Flow Random Early Drop) และ DRR (Deficit Round Robin) บนโปรแกรมจำลองเครือข่าย NS-2 โดยวิธีการทั้ง 4 นี้เป็นตัวแทนของระดับความซับซ้อนที่แตกต่างกัน DDR และ FRED จะให้ความสนใจในสถานะของแพ็กเก็ตในชุดเดียวกัน (Statefull) ในขณะที่ FIFO และ RED ไม่สนใจ เมื่อพิจารณาความซับซ้อนของ Edge nodes ใน CSFQ จะเทียบได้กับ FRED ในขณะที่ Core nodes มีระดับความซับซ้อนที่สามารถเปรียบเทียบได้กับ RED เท่านั้น การทดลองได้มีการกำหนดเงื่อนไขต่างกันในเรื่องชนิดของข้อมูล (TCP และ UDP) และ ลักษณะของเครือข่าย (Topology) พบว่า CSFQ บรรลุความเป็นธรรมของการใช้งาน Bandwidth ในทุกเงื่อนไขที่ใช้ทดลอง

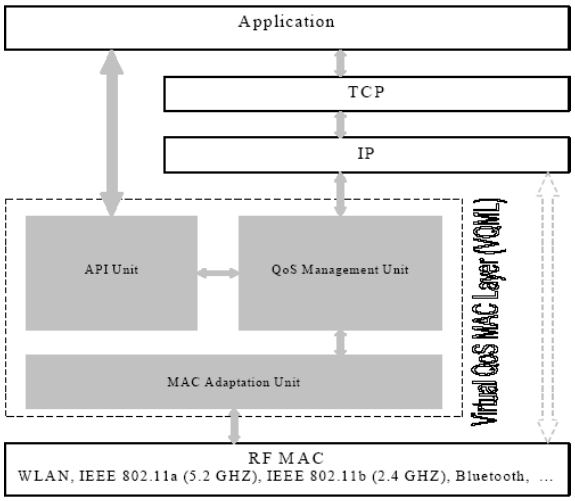
ผลการศึกษากระบวนการวิธีในการนำส่งข้อมูลในเครือข่ายที่ใช้ในอุปกรณ์เครือข่าย NetScreen ของบริษัท NetScreen Technologies ซึ่งเป็นอุปกรณ์ที่ใช้เชื่อมต่อเครือข่ายภายใน และเครือข่ายภายนอกเป็นหลัก เนื่องจาก Bandwidth ของเครือข่าย LAN ได้รับการพัฒนาจาก Ethernet เป็น Fast Ethernet จนกระทั่งเป็น Gigabit Ethernet ที่มีความเร็วในการรับส่งข้อมูลได้สูงถึง 10/100/1,000 Mbps ตามลำดับ ในขณะที่ Bandwidth ของ WAN ยังมีความเร็วในการรับส่งข้อมูลไม่ต่างจากเดิมมากนัก เช่น T1 ที่มีความเร็ว 1.54 Mbps หรือในกรณีที่เชื่อมต่อเครือข่ายในระยะทางไกลระหว่างองค์กรอาจติดต่อกันได้เพียง 128 Kbps หรือน้อยกว่านี้ ดังนั้นเมื่อใดก็ตามหากข้อมูลที่ไหลเวียน หรือจราจรมีขนาดใหญ่เกินกว่าขนาดของ Bandwidth ที่จะรองรับได้ ซึ่งอาจเกิดจากงานเพียงงานเดียว ก็อาจทำให้เครือข่าย WAN ไม่อยู่ในสภาพที่ให้บริการได้อีก การมีตัวจัดการจราจรที่จุดเชื่อมต่อระหว่างเครือข่ายภายใน และเครือข่ายภายนอกจึงมีความจำเป็นเพื่อจัดสรร Bandwidth ของ WAN ให้กับ Applications ต่าง ๆ ตามระดับความสำคัญ สำหรับวิธีการจัดสรร Bandwidth ของ NetScreen ไม่มีการเปิดเผยเนื่องจากเหตุผลทางธุรกิจ แต่ได้นำเสนอไว้ว่าวิธีการดังกล่าวสามารถจัดการกับจราจรแบบ Burst ได้ สามารถปรับเปลี่ยนการจัดสรร Bandwidth ให้กับ Application ที่มีระดับความสำคัญต่างกันได้ นอกจากนี้ยังอนุญาตให้แพ็กเก็ตของ Application ที่มีระดับความต่างกันสามารถใช้ Bandwidth ที่เหลืออยู่ร่วมกันได้ด้วย

Mahbub Hassan, Kenneth Lee และ Mohammad Rezvan ได้เสนอโปรโตคอล VQML (Virtual QoS MAC Layer) เพื่อทำให้การรับส่งข้อมูลมัลติมีเดียบนเครือข่ายไร้สาย (Wireless LANs) เกิด QoS โดยใช้เครือข่ายไร้สายของ Network Research Laboratory of UNSW เป็นกรณีทดลอง

งานชิ้นนี้ได้มีการปรับปรุงโครงสร้างของโปรโตคอลที่ใช้ในการติดต่อสื่อสารจากรูปที่ 2.22 เป็นรูปที่ 2.23 โดยการเพิ่มชั้น VQML แทรกระหว่างโปรโตคอลชั้นที่ 3 (IP) และโปรโตคอลชั้นที่ 2 (MAC) สำหรับการพัฒนา VQML ได้ใช้โปรแกรม Traffic control (TC) บนระบบปฏิบัติการลีนุกซ์ โดย QoS Management Unit มีหน้าที่รับแพ็กเก็ตจากชั้นที่สูงกว่า กำหนดระดับความสำคัญ (Access priority) ของการใช้งานอุปกรณ์ Wireless Station พร้อมทั้งจัดเก็บลงคิวที่เหมาะสม และหน่วยนำส่งจะเลือกแพ็กเก็ตจากคิวต่าง ๆ ตามวิธีการที่สามารถใช้งานได้ผ่าน System Call ที่ API ได้จัดเตรียมไว้ให้

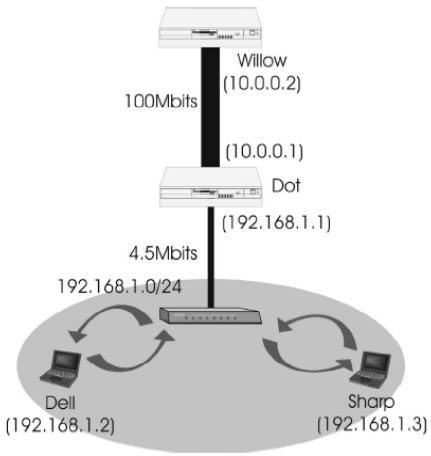


รูปที่ 2.22 แสดงโครงสร้างของโปรโตคอลที่ใช้บนเครือข่ายไร้สาย



รูปที่ 2.23 แสดงโครงสร้างของโปรโตคอลบนเครือข่ายไร้สายที่ได้เพิ่ม VQML ระหว่าง MAC และ Layer3 เพื่อให้เกิด QoS

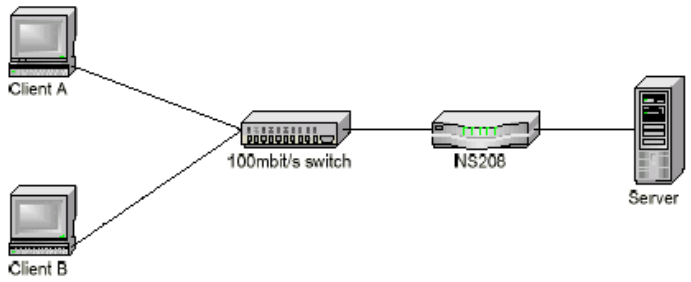
ในขั้นตอนการทดลองได้ทำการเชื่อมต่อเครือข่ายคอมพิวเตอร์ดังรูปที่ 2.24 ประกอบด้วยคอมพิวเตอร์ 4 เครื่องที่ติดตั้งลีนุกซ์เคอร์เนลรุ่น 2.4 และติดตั้ง VQML ที่เครื่อง Dell และ Sharp ซึ่งมีการเชื่อมต่อแบบไร้สายไปยัง Dot ให้ Dell เชื่อมต่อรับข้อมูลวิดีโอจาก Willow ในขณะที่ Sharp มีการจำลองส่งแพ็กเก็ตไปยัง Willow ด้วย ตามสภานะการณ์นี้ได้ก่อให้เกิดความคับคั่งในเส้นทางไร้สาย จากการทดลองพบว่าหากให้มีการติดต่อสื่อสารแบบปรกติ (ไม่ใช้ VQML) หรือ Best Effort ผลของการใช้ Application ประเภทนี้ ปรากฏว่าการแสดงภาพและเสียงไม่ตรงกัน แต่เมื่อเปิดใช้ VQML การแสดงผลวิดีโอเป็นที่น่าพอใจ ทำให้สรุปได้ว่า VQML ทำให้เกิด QoS กับ Application ประเภทวิดีโอ หรือมัลติมีเดีย



รูปที่ 2.24 แสดงเครือข่ายที่ใช้ในการทดสอบ VQML

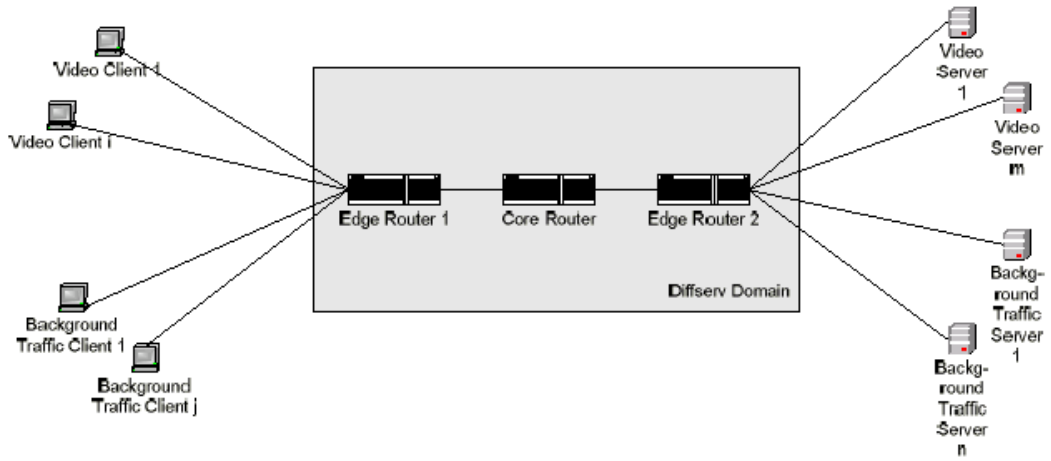
Karl Ahlin ได้ศึกษาทฤษฎีของ QoS โดยเน้นไปในส่วนที่เกี่ยวข้องกับเครือข่ายอินเทอร์เน็ตแบบแพ็กเก็ตสวิตชิง (Packet Switching) นอกจากนี้ได้มีการยกตัวอย่างการนำ QoS ไปประยุกต์อีกด้วย Karl ได้แบ่งการทดลองออกเป็น 2 ลักษณะคือ การทดลองบนเครือข่ายจริง (Real-World Testing) และการทดลองผ่านโปรแกรมจำลอง NS-2 รุ่น 2.1-b9

การทดลองบนเครือข่ายจริงได้เชื่อมต่ออุปกรณ์เครือข่ายดังรูปที่ 2.25 โดยอาศัยอุปกรณ์ Netscreen รุ่น 208 (NS208) ที่มีความสามารถในการกำหนดระดับความสำคัญให้กับแพ็กเก็ตได้ ส่วนงานที่ใช้ในการทดลองเกิดจากโปรแกรม Darwin Streaming Server 4.1, QuickTime และ Netload.pl โดย Darwin Streaming Server 4.1 ทำหน้าที่ให้บริการไฟล์วิดีโอในรูปแบบของ MPEG-4 และ QuickTime ติดตั้ง QuickTime Player 6 บนเครื่องลูกข่าย สำหรับ Netload.pl ซึ่งเป็นสคริปต์ของเพิลมีทั้งส่วนที่เป็นผู้ให้บริการ และผู้ร้องขอ เมื่อผู้ร้องขอได้ติดต่อไปยังผู้ให้บริการ สคริปต์ Netload.pl ที่อยู่ทางด้านผู้ให้บริการจะสร้างจราจรขึ้นมาจำนวนมากเท่าที่จะทำได้เพื่อส่งไปยังฝ่ายร้องขอ ในการทดลองกำหนดส่งแพ็กเก็ตจาก Netload.pl และ QuickTime ออกจากเครื่องให้บริการพร้อมกันผ่านอุปกรณ์ NS208 ที่ให้บริการแบบ Priority Queueing และถูกกำหนดให้เป็นคอขวดของเครือข่าย ไปยังเครื่องลูกข่าย A และ B ที่อยู่อีกด้านหนึ่ง และกำหนดให้จราจรจาก Netload.pl มีระดับความสำคัญสูงกว่าจราจรจาก QuickTime เมื่อเพิ่มจำนวนของเครื่องลูกข่าย พบว่าแพ็กเก็ตจาก QuickTime ไม่สามารถส่งผ่านไปยังเครื่องร้องขอได้ นั้นคือ NS208 ไม่ทำให้เกิด QoS บนเครือข่าย IP

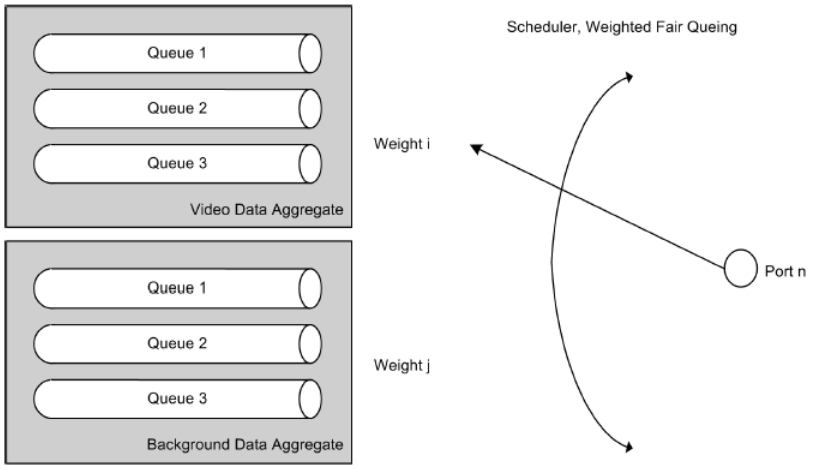


รูปที่ 2.25 แสดงเครือข่ายจริงสำหรับทดสอบ QoS ของ Karl

ในส่วนที่ทำการทดลองบนโปรแกรมจำลองเครือข่าย NS-2 นั้นได้เชื่อมต่อเครือข่ายตามรูปที่ 2.26 ประกอบด้วยเครื่องลูกข่ายสำหรับร้องขอข้อมูลวิดีโอ 1,000 เครื่อง ร้องขอข้อมูลอื่นอีก 1,000 เครื่อง และเป็นเครื่องสำหรับให้บริการวิดีโอ และข้อมูลทั่วไปอย่างละ 4 เครื่อง ข้อมูลจะถูกส่งจากเครื่องให้บริการไปยังเครื่องลูกข่ายโดยเว้นช่วงเวลาในการส่งระหว่างแพ็กเก็ตแบบสุ่ม (Random time intervals) เพื่อเลียนแบบลักษณะการส่งข้อมูลแบบ Burst บนเครือข่ายจริง ส่วนอุปกรณ์เครือข่ายในเส้นทางใช้วิธีการควบคุม Bandwidth แบบ DiffServ และอาศัยเทคนิคแบบ WRED ในการควบคุมความคับคั่งของเครือข่าย ซึ่งเป็นเทคนิคผสมระหว่างการจัดการคิวแบบ RED และ WFQ ดังรูปที่ 2.27



รูปที่ 2.26 แสดงเครือข่ายจำลองใน NS-2 สำหรับทดสอบ QoS ของ Karl



รูปที่ 2.27 แสดงวิธีการให้บริการแบบ WRED

จากการเปรียบเทียบผลที่ได้จากการทดลองเมื่อเราเตอร์ในเส้นทางเลือกให้บริการแบบ Best effort กับวิธีการให้บริการของ DiffServ พบว่า DiffServ ทำให้เวลาหน่วงของจราจรทั้งสองกลุ่มสูงขึ้น แต่ค่าความแปรปรวน และอัตราการถูกทิ้งของข้อมูลวิดีโอลดลง นอกจากนี้อัตราการส่งข้อมูลสูงขึ้นด้วย แต่ทำให้ประสิทธิภาพของการนำส่งจราจรอีกกลุ่มแย่ลง

Karl ได้สรุปว่า QoS มีความจำเป็นต่อเครือข่าย เพื่อเป็นการจัดเตรียมการให้บริการที่ดีกว่าแก่ลูกค้าตามระดับความสำคัญ และแนะนำให้มีการติดตั้ง QoS ในจุดที่เป็นคอขวดของระบบเครือข่าย ซึ่งเป็นการเพิ่มระดับของการให้บริการ โดยไม่ต้องปรับเปลี่ยนโครงสร้างเครือข่าย หรือเพิ่มขนาดของ Bandwidth ให้กับเครือข่าย

หฤทัย สมบูรณ์รุ่งโรจน์ได้เสนอผลจากการศึกษาวิธีการจัดคิวแบบ Weighted Fair Queueing และ Priority Queueing กับระบบเครือข่ายคอมพิวเตอร์มหาวิทยาลัยสงขลานครินทร์ โดยใช้ตัวจำลองระบบเครือข่าย NS-2 งานวิจัยนี้ได้แบ่งการทดลองออกเป็นสองส่วน คือ การทดสอบกับข้อมูลสมมติ และการทดสอบกับข้อมูลการใช้เครือข่ายจริงของมหาวิทยาลัยสงขลานครินทร์ ผลการทดสอบกับข้อมูลสมมติได้ว่าวิธีจัดการคิวสามารถแสดงความแตกต่างในการรับส่งข้อมูลได้ชัดเจนมากกว่าการทดสอบด้วยข้อมูลจริง และแตกต่างจากการรับส่งข้อมูลโดยไม่มีการจัดคิว ส่วนผลจากการทดสอบกับข้อมูลจริงพบว่าค่าอัตราการรับส่งข้อมูลก่อน และหลังการจัดคิวไม่แตกต่างกันมากนัก ทั้งนี้อาจมีสาเหตุเกิดจากปริมาณข้อมูลที่เข้าออกระหว่างเครือข่ายคอมพิวเตอร์มหาวิทยาลัยกับเครือข่ายอินเทอร์เน็ตมีปริมาณไม่มากพอที่จะทำให้ระบบเครือข่ายเกิดความแออัดของข้อมูล ถึงแม้ว่าการใช้วิธีจัดการคิวจะได้ผลไม่แตกต่างจากเดิม แต่ผู้วิจัยได้แนะนำว่าสามารถนำวิธีการจัดคิวมาปรับใช้ตามช่วงเวลาได้ โดยการแบ่งกลุ่มของผู้ใช้ในแต่ละคณะ/หน่วยงานออกเป็น 2 กลุ่ม คือ กลุ่มบุคลากร และกลุ่มนักศึกษา ในช่วงเวลาทำงานอาจเลือกใช้วิธีจัดคิวแบบ PQ โดยกำหนดให้กลุ่มบุคลากรมีค่าลำดับความสำคัญมากกว่ากลุ่มนักศึกษา และช่วงเวลาหลังเลิกงานอาจเลือกให้วิธีจัดคิวแบบ WFQ

2.2.2 สรุปงานวิจัยที่เกี่ยวข้อง

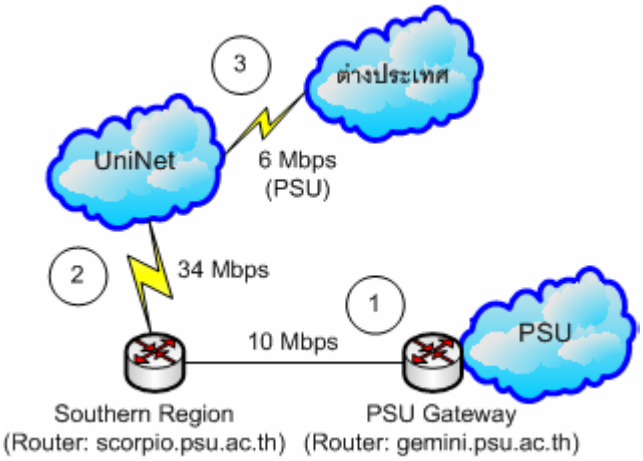
จากการศึกษางานวิจัยที่เกี่ยวข้องได้พบว่างานวิจัยส่วนใหญ่ได้ดำเนินการอยู่บนเครือข่ายจำลองที่สร้างขึ้นใน NS-2 เท่านั้น สำหรับงานวิจัยที่ดำเนินการทดสอบบนเครือข่ายจริงนั้นอยู่ภายใต้เครือข่าย ATM ดังนั้นงานวิจัยนี้จึงเป็นการทดสอบ QoS บนเครือข่ายจริงภายใต้สภาพแวดล้อมของมหาวิทยาลัยสงขลานครินทร์ที่ใช้ระบบสื่อสารแบบ Packet Switching

2.3 ผลการศึกษาโครงสร้างและการใช้งานเครือข่าย

ในหัวข้อนี้ผู้วิจัยได้นำเสนอความเป็นมาของเครือข่ายมหาวิทยาลัยสงขลา-นครินทร์ (SritrangNet) โครงสร้างของเครือข่าย และพฤติกรรมการใช้งานเครือข่ายที่เกิดขึ้นจริง โดยใช้ MRTG และ Traffic Tools (ttools, ภาคผนวก ข) เป็นเครื่องมือสำหรับเก็บรวบรวมข้อมูล

2.3.1 เครือข่ายของมหาวิทยาลัยสงขลานครินทร์ [18] [19] [20]

มหาวิทยาลัยสงขลานครินทร์เริ่มให้บริการอินเทอร์เน็ตตั้งแต่ปี พ.ศ.2539 ด้วยวงจรเช่าที่มี Bandwidth ขนาด 64 Kbps โดยเชื่อมต่อกับเครือข่ายไทยสาร[1]ที่กรุงเทพฯ และในปี พ.ศ. 2541 ได้เปลี่ยนมาใช้งานอินเทอร์เน็ตผ่านทางเครือข่าย UniNet[2] ด้วยวงจรภายในประเทศ Bandwidth 4 Mbps และวงจรออกต่างประเทศสำหรับมหาวิทยาลัยสงขลานครินทร์ที่มี Bandwidth 512 Kbps จากนั้นได้มีการปรับปรุงการเชื่อมต่อเรื่อยมาจนกระทั่งปี พ.ศ.2546 UniNet ได้ขยายวงจรระหว่างมหาวิทยาลัยสงขลานครินทร์กับเครือข่าย UniNet ที่ความเร็ว 34 Mbps วงจรออกต่างประเทศสำหรับมหาวิทยาลัยสงขลานครินทร์ที่มี Bandwidth 6 Mbps ดังที่ได้แสดงในรูปที่ 2.28



รูปที่ 2.28 แสดงการวงจรสื่อสารของมหาวิทยาลัยสงขลานครินทร์ไปยังเครือข่าย UniNet

[1] ดูรายละเอียดของเครือข่ายไทยสารได้จาก www.nectec.or.th

[2] ดูรายละเอียดของเครือข่าย UniNet ได้จาก www.uni.net.th

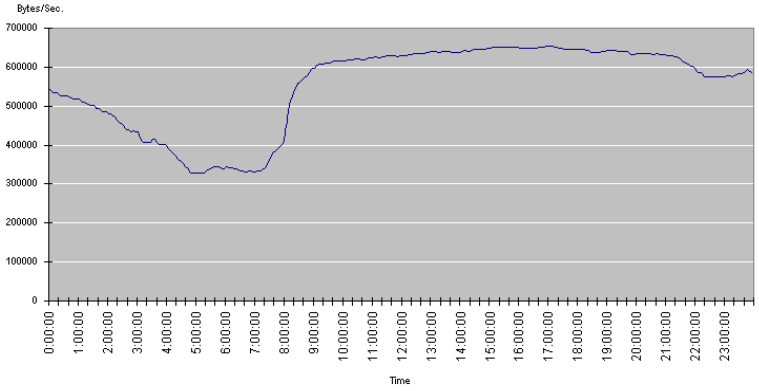
อุปกรณ์ที่ใช้ในการเชื่อมต่อระหว่างมหาวิทยาลัยสงขลานครินทร์ออกสู่อินเทอร์เน็ตผ่านวงจรสื่อสารของ UniNet เป็นเราเตอร์ Cisco 4500 ตรงกับตำแหน่ง 1 ในรูปที่ 2.28 ที่ทำหน้าที่เป็นอินเทอร์เน็ตเกตเวย์ของมหาวิทยาลัย ซึ่งใช้งานมาตั้งแต่ปี พ.ศ.2539 อุปกรณ์ตัวนี้มีพอร์ตเชื่อมต่อที่ออกแบบมาเพื่อใช้งานกับเครือข่ายที่มี Bandwidth 10 Mbps และมีการติดต่อสื่อสารระหว่าเราเตอร์แบบ Half duplex เท่านั้น จึงไม่สามารถรับและส่งข้อมูลได้พร้อมกันในขณะเดียวกันจากการตรวจสอบสมรรถนะของอุปกรณ์ พบว่าสภาวะการทำงานเฉลี่ยของ CPU มีค่าอยู่ร้อยละ 60 ถึง 70 ซึ่งสูงมาก และบางโอกาสมีค่าถึงร้อยละ 98 เป็นสาเหตุสำคัญอีกข้อหนึ่งที่ทำให้การใช้งานอินเทอร์เน็ตช้ามาก จากปัญหานี้ผู้วิจัยจึงได้เสนอแนวทางในการแก้ปัญหาด้วยการจัดระดับความสำคัญของ Application และลดจำนวนของแพ็กเก็ตของ Application ที่ไม่มีก่อประโยชน์ เพื่อให้ Application สำคัญมีโอกาสได้รับบริการที่ดีขึ้น และยังลดภาระงานให้กับอุปกรณ์เครือข่ายอีกประการหนึ่ง ดังนั้นผู้วิจัยจำเป็นต้องศึกษาถึงพฤติกรรมการใช้งานเครือข่ายที่เกิดขึ้นในขณะนั้น เพื่อให้ทราบถึงประเภทและข้อมูลอื่น ๆ ของ Application ที่มีการใช้งานจริงในเครือข่ายมหาวิทยาลัย

2.3.2 การศึกษาพฤติกรรมการใช้งานเครือข่ายมหาวิทยาลัยสงขลานครินทร์

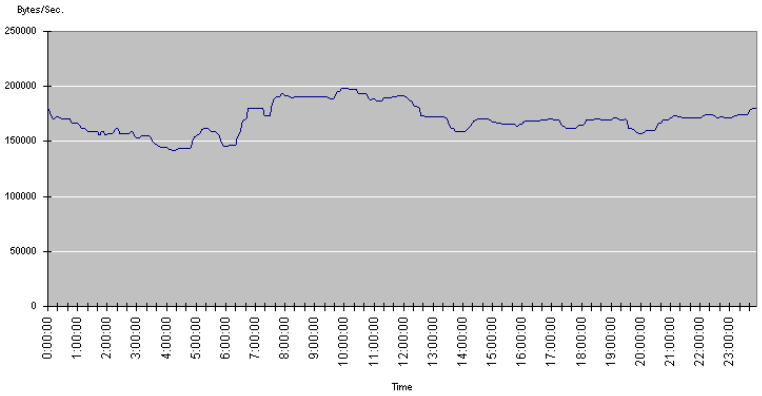
ในการศึกษาพฤติกรรมการใช้งานเครือข่ายของผู้ใช้ทั้งหมด ผู้วิจัยเลือกที่จะศึกษาถึงปริมาณและประเภทของข้อมูลที่ผ่านเข้าออกเครือข่ายมหาวิทยาลัย เช่น มีปริมาณข้อมูลผ่านเข้าออกในแต่ละเดือน แต่ละวัน ว่ามีปริมาณมากน้อยอย่างไร ช่วงเวลาใดมีการใช้งานเครือข่ายมากที่สุด เพื่อให้ทราบถึงพฤติกรรมการใช้งานเครือข่ายที่เกิดขึ้นจริง และทราบว่าที่ช่วงเวลาใดที่มีปัญหาความคับคั่งของข้อมูลในเครือข่าย ในการนี้ผู้วิจัยได้ทำการเก็บปริมาณข้อมูลที่เข้าออกเครือข่ายโดยใช้เครื่องมือชื่อ MRTG (Multi Router Traffic Grapher) ซึ่งเป็นโอเพนซอร์สที่นิยมใช้กันแพร่หลาย โดย MRTG จะติดต่อกับเราเตอร์หรืออุปกรณ์เป้าหมายในเครือข่ายผ่านโปรโตคอล SNMP (Simple Network Management Protocol) เพื่อขอค่าปริมาณข้อมูลที่ผ่านเข้าออกอุปกรณ์และนำมาแสดงในรูปแบบของกราฟ

จากการศึกษาพฤติกรรมการใช้งานเครือข่ายครั้งนี้ ผู้วิจัยได้ทำการเก็บสถิติการใช้งานเครือข่ายของผู้ใช้ทั้งหมดจากเราเตอร์ gemini.psu.ac.th (จากภาพประกอบ 2.28) ด้วยโปรแกรม MRTG ตลอดเดือนกันยายน พ.ศ. 2546 ได้เฉลี่ยค่าของการใช้งานโดยแบ่งตามช่วงเวลา และแยกเป็นปริมาณข้อมูลที่เข้าสู่เครือข่าย และปริมาณข้อมูลที่ออกจากเครือข่ายมหาวิทยาลัย ดังรูปที่ 2.29 และ 2.2 ตามลำดับ จากกราฟทำให้ทราบว่าในวันหนึ่ง ๆ ระดับการใช้งานเครือข่ายจะเพิ่มขึ้นอย่างรวดเร็วในช่วงเวลาประมาณ 7.00 – 10.00 น. และคงตัวจนกระทั่งเวลาประมาณ 22.00 น.ปริมาณ

การใช้งานเครือข่ายจึงจะค่อย ๆ ลดลง โดยปริมาณข้อมูลที่เข้าสู่เครือข่ายมีค่าสูงสุดประมาณ 657,327.5 Byte/s (Bps) หรือ 5.02 Mbps และปริมาณข้อมูลที่ออกจากเครือข่ายมีค่าสูงสุดประมาณ 198,689 Bps หรือ 1.52 Mbps

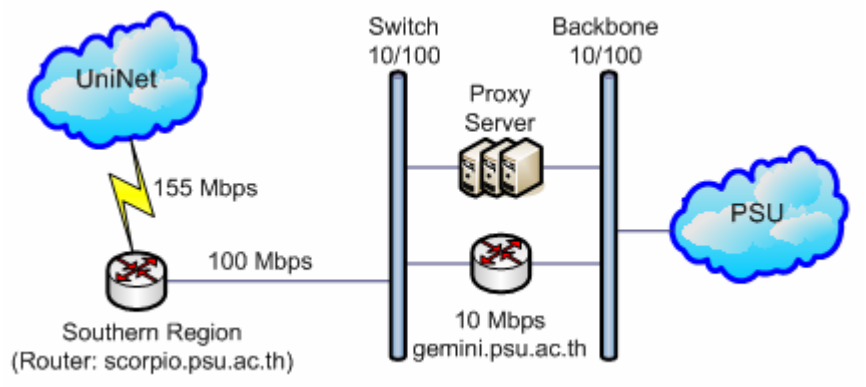


รูปที่ 2.29 แสดงปริมาณข้อมูลที่<u>เข้าสู่</u> SritrangNet เดือนกันยายน พ.ศ. 2546



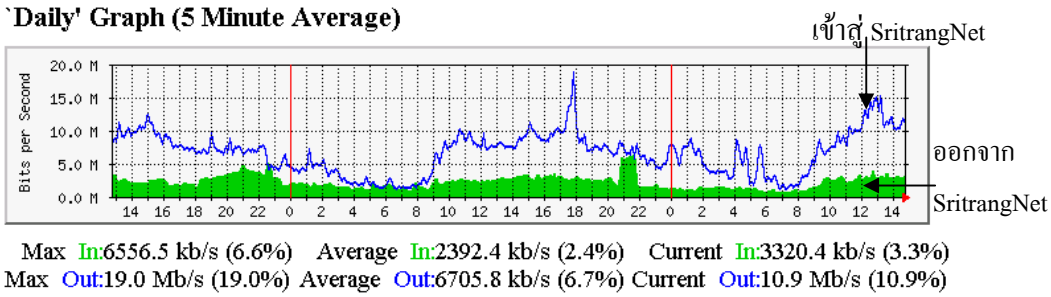
รูปที่ 2.30 แสดงปริมาณข้อมูลที่<u>ออกจาก</u> SritrangNet เดือนกันยายน พ.ศ. 2546

ต้นปี พ.ศ.2547 ศูนย์คอมพิวเตอร์ได้ปรับโครงสร้างการเชื่อมอุปกรณ์เครือข่ายภายในจากรูปที่ 2.28 เป็นรูปที่ 2.31 เพื่อให้ข้อมูล Web ที่ต้องการออกอินเทอร์เน็ตส่งผ่าน Proxy Servers โดยตรง ในขณะที่การใช้งานอินเทอร์เน็ตในด้านอื่นที่ไม่ใช่ Web ยังคงใช้งานผ่านเราเตอร์ gemini เช่นเดิม ทั้งนี้เพื่อให้สามารถรับส่งข้อมูลได้มากขึ้น และลดภาระงานให้กับเราเตอร์ gemini และในขณะนั้นทาง UniNet ได้ขยาย Bandwidth ของวงจร ATM จาก 34 Mbps เป็น 155 Mbps เพื่อเพิ่ม Bandwidth สำหรับใช้งานอินเทอร์เน็ตภายในประเทศด้วย



รูปที่ 2.31 แสดงการเชื่อมต่ออุปกรณ์เครือข่ายคอมพิวเตอร์ของมหาวิทยาลัยสงขลานครินทร์

หลังจากที่ได้มีการปรับเปลี่ยนโครงสร้างเครือข่ายแล้ว ผู้วิจัยได้เก็บข้อมูลการใช้งานเครือข่ายด้วย MRTG อีกครั้งในวันอังคารที่ 10 กุมภาพันธ์ พ.ศ.2547 บนเราเตอร์ scorpio ได้ข้อมูลดังแสดงในรูปที่ 2.32 พบว่าค่าเฉลี่ยปริมาณการใช้งานเครือข่ายที่เข้า (เส้น Out) และออกจาก (เส้น In) SritrangNet มีค่าเพิ่มขึ้น ค่าเฉลี่ยของปริมาณข้อมูลที่ขาเข้าเพิ่มขึ้นเป็น 6,705.8 Kbps (6.55 Mbps) และมีค่าสูงสุดเป็น 19.0 Mbps ในขณะที่ปริมาณข้อมูลขาเข้าจากรูปที่ 2.29 มีค่าสูงสุดเพียง 5.02 Mbps และค่าเฉลี่ยของปริมาณข้อมูลขาออกเพิ่มขึ้นเป็น 2,392.4 Kbps (2.34 Mbps) และมีค่าสูงสุดเป็น 6,556.5 Kbps (6.4 Mbps) ในขณะที่ปริมาณข้อมูลขาออกจากรูปที่ 2.30 มีค่าสูงสุดเพียง 1.52 Mbps เท่านั้น



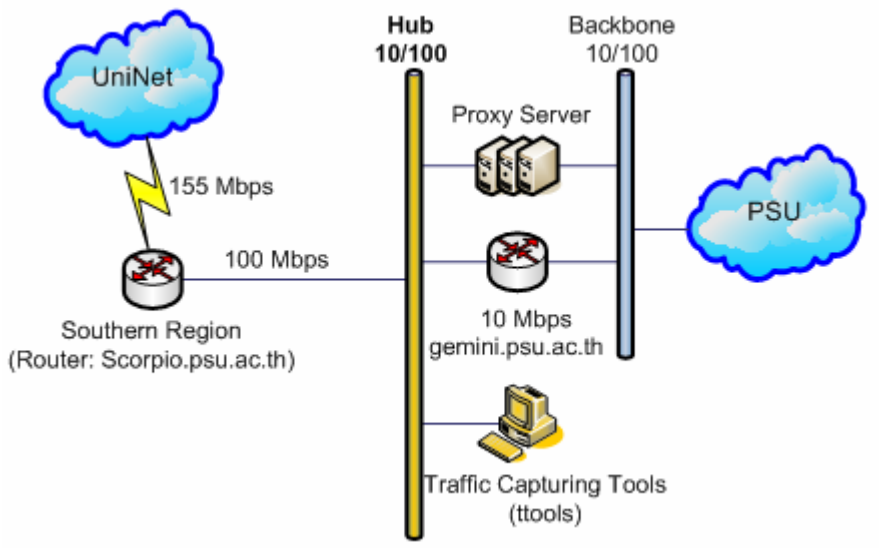
รูปที่ 2.32 แสดงปริมาณข้อมูลเข้าออกผ่าน Gateway Router (scorpio) ของ
มหาวิทยาลัยสงขลานครินทร์ เมื่อวันอังคารที่ 10 กุมภาพันธ์ พ.ศ. 2547

จากการศึกษาการใช้งาน SritrangNet ทั้งก่อนและหลังการปรับโครงสร้างเครือข่าย สรุปได้ว่า เราเตอร์ gemini (ดูรูปที่ 2.31 ประกอบ) ที่ทำหน้าที่เป็นทางออกของเครือข่ายและมีความสามารถในการส่งข้อมูลด้วยอัตราเร็ว 10 Mbps Half-duplex นั้นเป็นคอขวดจุดหนึ่งของ SritrangNet ไปยังเครือข่าย UniNet ถึงแม้ว่าข้อมูลจะได้รับการนำส่งมากขึ้น แต่ข้อมูลเหล่านี้อาจถูกส่งผ่านวงจรสื่อสารอื่นที่เป็นคอขวดได้อีก ดังนั้นหากมีการเลือกนำส่งแพ็กเก็ตตามความระดับสำคัญของ Application อาจช่วยเพิ่มประสิทธิภาพการใช้งานอินเทอร์เน็ตได้อีกระดับหนึ่ง

2.3.3 การเก็บบันทึกข้อมูลการใช้งานเครือข่าย

ผลการศึกษาพฤติกรรมการใช้งานเครือข่ายมหาวิทยาลัยสงขลานครินทร์ทำให้ทราบถึงปริมาณการใช้งานเครือข่ายของผู้ใช้ภายในมหาวิทยาลัย แต่ไม่สามารถระบุได้ว่าข้อมูลที่เข้าออกเครือข่ายเกิดจาก Application ใด

การเก็บข้อมูลการใช้งานเครือข่าย SritrangNet เมื่อวันอังคารที่ 23 มีนาคม 2547 ตั้งแต่เวลา ประมาณ 15.00 น. ถึง 16.20 น. (4865.954 วินาที) ด้วยเครื่องมือ Traffic Tools (tools) ในตำแหน่งดังแสดงในรูปที่ 2.33 พบว่ามี IP datagram เข้าออก SritrangNet ประมาณ 2,736 แพ็กเก็ตต่อวินาที (packet/second) คิดเป็น 11.095 Mbps ในขณะที่แพ็กเก็ตที่ใช้อีเทอร์เน็ตเฟรมอื่น เช่น ARP และ RARP เป็นต้น ได้เข้าและออก SritrangNet เพียง 1 packet/second คิดเป็น 0.4 Kpbs ถือว่าน้อยมาก ผู้วิจัยจึงเลือกเฉพาะ IP datagram ไปใช้ในสำหรับวิเคราะห์ในขั้นตอนต่อไป



รูปที่ 2.33 แสดงตำแหน่งติดตั้งเครื่องคอมพิวเตอร์สำหรับเก็บข้อมูลที่ไหลเข้าออก SritrangNet

เนื่องจากในขั้นตอนการเก็บข้อมูล ttools ได้บันทึกทั้งข้อมูลของแพ็กเก็ตที่เข้าสู่เครือข่าย และ แพ็กเก็ตที่ถูกส่งออกจากเครือข่าย ในการวิเคราะห์การใช้งานเครือข่ายต้องแยกข้อมูลทั้งสองส่วนออกจากกัน ผู้วิจัยอาศัยหมายเลขไอพีเพื่อแยกข้อมูลดังกล่าว แพ็กเก็ตที่มีหมายเลขไอพีต้นทางเป็น 192.100.77.0/24 202.12.73.0/24 202.12.74.0/24 และ 202.28.96.0/22 จัดเป็นแพ็กเก็ตที่ส่งออกจาก SritrangNet เนื่องจากเป็นหมายเลขที่ได้รับการจัดสรรสำหรับมหาวิทยาลัย-สงขลานครินทร์ แพ็กเก็ตที่มีหมายเลขไอพีต้นทางนอกจากนี้ถือเป็นแพ็กเก็ตที่มาจากภายนอก หรืออินเทอร์เน็ต โดยผลจากการศึกษาได้แสดงในตารางที่ 2.5 สามารถอธิบายว่าแพ็กเก็ตที่เข้าออกเครือข่ายเกือบทั้งหมดอาศัยโปรโตคอล TCP และ UDP (13,274,350 แพ็กเก็ตจากทั้งหมด 13,311,461 แพ็กเก็ต) เป็น TCP มากกว่า UDP และปริมาณข้อมูลที่เข้าสู่ SritrangNet มีมากกว่าปริมาณข้อมูลที่ส่งออกออกจากเครือข่าย ทั้งนี้เนื่องจากข้อมูลที่ออกจาก SritrangNet มักเป็นแพ็กเก็ตคำร้องขอข้อมูลมากกว่าข้อมูล และยังพบว่ามีการใช้งานโปรโตคอลอื่นในชุด TCP/IP นอกจาก TCP และ UDP เข้าและออกรวมกันน้อยมากเพียง 8 packet/second หรือ 0.007 Mbps

Protocol	ทิศทาง	จำนวนแพ็กเก็ต	Packet/s	MegaBytes	Mbps
TCP	เข้า	6,597,289	1,356	4,833.97	7.95
	ออก	6,105,754	1,255	1,749.45	2.88
UDP	เข้า	251,722	52	30.789	0.05
	ออก	319,585	66	30.6149	0.05
Other TCP/IP Protocol	เข้า + ออก	37,111	8	4.2566	0.007
รวม	เข้า + ออก	13,311,461	2,737	6,649.0805	10.937

ตารางที่ 2.5 แสดงจำนวน และปริมาณของ IP datagram ที่บันทึกเมื่อวันอังคารที่ 23 มีนาคม 2547

หมายเหตุ – ผลลัพธ์ที่แสดงในตารางนี้คลาดเคลื่อนจากตัวเลขที่ได้เสนอก่อนหน้าทั้งนี้เกิดจากการปัดเศษที่ได้จากการคำนวณ

ขั้นตอนต่อมาผู้วิจัยได้แยกวิเคราะห์แพ็กเก็ตตามประเภทของ Application หรือประเภทของบริการ โดยอาศัยหมายเลขพอร์ตที่ระบุอยู่ในแพ็กเก็ต ทั้งหมายเลขพอร์ตต้นทาง (Source Port/SPort) และหมายเลขพอร์ตปลายทาง (Destination Port/DPort) ผู้วิจัยได้เลือกวิเคราะห์บริการที่มักใช้งานอยู่ทั่วไป คือ Web (พอร์ต 80) FTP (พอร์ต 20 และ 21) และ Email (พอร์ต 25) ได้จำนวนและปริมาณของแพ็กเก็ตข้อมูลในกลุ่มที่สนใจนี้ดังตารางที่ 2.6 สำหรับแพ็กเก็ตที่เข้าสู่ SritrangNet และตารางที่ 2.7 สำหรับแพ็กเก็ตที่ออกจากเครือข่าย และรูปที่ 2.34 ถึง 2.37 ได้สรุปรวมข้อมูลตามประเภทของ Application (DPort + Sport)

	จำนวน TCP และ UDP แพ็กเก็ตที่เข้าสู่ มอ.		ปริมาณข้อมูลที่เข้าสู่ มอ.	
พอร์ต	แพ็กเก็ต	ร้อยละ	MB	ร้อยละ
DPort 20	61	0.0009	0.0024	0.0000
SPort 20	33,334	0.4867	30.1345	0.6194
DPort 21	971	0.0142	0.0488	0.0010
SPort 21	7,463	0.1090	0.8377	0.0172
DPort 25	54,595	0.7971	47.4733	0.9759
SPort 25	167,428	2.4446	8.6322	0.1774
DPort 80	488,822	7.1371	45.4291	0.9338
SPort 80	3,933,332	57.4292	3875.7478	79.6699
Other	2,163,005	31.5812	856.4499	17.6054
รวม	6,849,011	100	4,864.7556	100

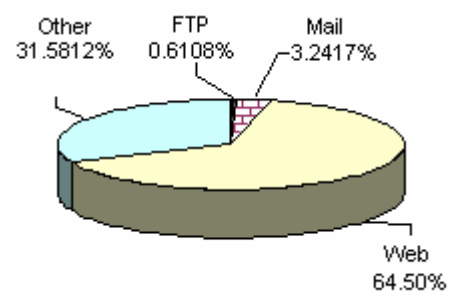
ตารางที่ 2.6 แสดงผลจำนวน และปริมาณของข้อมูล TCP และ UDP ที่เข้าสู่ SritrangNet แยกตามลักษณะการใช้งาน

	จำนวน TCP และ UDP แพ็กเก็ตที่ส่งออกจาก มอ.		ปริมาณข้อมูลที่ส่งออกจาก มอ.	
พอร์ต	แพ็กเก็ต	ร้อยละ	MB	เปอร์เซ็นต์
DPort 20	19,783	0.3079	0.8646	0.0486
Sport 20	57	0.0009	0.0259	0.0015
DPort 21	7,361	0.1146	0.3646	0.0205
Sport 21	891	0.0139	0.0596	0.0033
DPort 25	186,738	2.9063	97.9085	5.5003
SPort 25	41,861	0.6515	2.3228	0.1305
DPort 80	3,493,912	54.3771	339.2814	19.0601
SPort 80	571,690	8.8974	494.6255	27.7870
Other	2,103,046	32.7304	844.6101	47.4482
รวม	6,425,339	100	1,780.0631	100

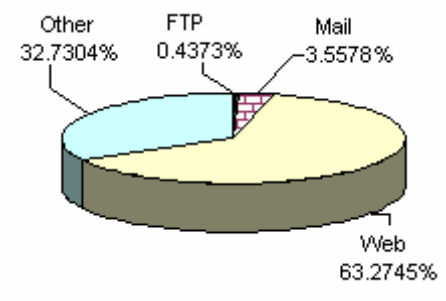
ตารางที่ 2.7 แสดงผลจำนวน และปริมาณของข้อมูล TCP และ UDP ที่ออกจาก SritrangNet แยกตามลักษณะการใช้งาน

หมายเหตุ - การตีความข้อมูลในตารางที่ 2.6 และ 2.7 สมมติใช้ข้อมูล Sport 80 ในตารางที่ 2.7 หมายถึง เครื่องให้บริการ Web ที่อยู่ใน SritrangNet ได้ส่งข้อมูลออกสู่อินเทอร์เน็ต หากแพ็กเก็ตที่ได้ส่งออกมไปเป็น Dport 80 หมายถึง ผู้ใช้ภายใน SritrangNet ได้ร้องขอข้อมูล Web จากเครื่องให้บริการที่อยู่ในอินเทอร์เน็ต

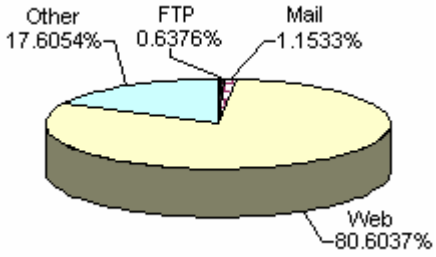
รูปที่ 2.34 แสดงจำนวนแพ็กเก็ตเข้าสู่ SritrangNet แยกตาม Application



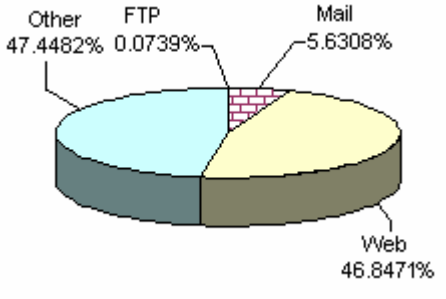
รูปที่ 2.35 แสดงจำนวนแพ็กเก็ตที่ส่งออกจาก SritrangNet แยกตาม Application



รูปที่ 2.36 แสดงปริมาณข้อมูลที่เข้าสู่ SritrangNet แยกตาม Application



รูปที่ 2.37 แสดงปริมาณข้อมูลที่ส่งออกจาก SritrangNet แยกตาม Application



รูปที่ 2.34 ได้แสดงให้เห็นว่าแพ็กเก็ตที่เข้าสู่ SritrangNet เป็นแพ็กเก็ต Web มากที่สุด รองลงมาเป็นแพ็กเก็ตอื่นที่ไม่อยู่ในความสนใจ Email และ FTP น้อยที่สุด เมื่อพิจารณารูปที่ 2.35 ประกอบจะเห็นว่าจำนวนแพ็กเก็ตของ Appliation แต่ละประเภทที่เข้าสู่เครือข่าย และจำนวนแพ็กเก็ตที่ออกจากเครือข่ายมีความสัมพันธ์ หรือมีจำนวนของแพ็กเก็ตใกล้เคียงกัน ข้อมูลนี้ทำให้ทราบว่าผู้ใช้งานอินเทอร์เน็ตภายในมหาวิทยาลัยได้ใช้งาน Application จำพวก Web มากที่สุด แต่ในมุมของปริมาณข้อมูลที่เข้าสู่เครือข่ายกับข้อมูลที่ออกจากเครือข่ายกลับมีค่าแตกต่างกันชัดเจนดังรูปที่ 2.36 และ 2.37 ตามลำดับ ข้อมูลกราฟได้แสดงให้เห็นว่าปริมาณข้อมูลที่เข้าสู่เครือข่ายเป็นข้อมูล Web มากที่สุด คือร้อยละ 80.59 ในขณะที่ปริมาณข้อมูล Web ที่ออกจากเครือข่ายมีเพียงร้อยละ 46.85 เท่านั้น ซึ่งมีค่าแตกต่างกันมากทั้งนี้เนื่องจากแพ็กเก็ต Web ที่ออกจาก

เครือข่ายส่วนใหญ่เป็นคำร้องที่มีขนาดของแพ็กเก็ตเล็ก พิจารณาข้อมูลในตารางที่ 2.7 ประกอบ (DPort80, แพ็กเก็ตมีขนาดเฉลี่ย 102 Bytes) ในขณะที่แพ็กเก็ตที่ตอบกลับมาจากเครื่องให้บริการในอินเทอร์เน็ตมีขนาดใหญ่กว่า พิจารณาข้อมูลในตารางที่ 2.6 ประกอบ (Sport 80, แพ็กเก็ตมีขนาดเฉลี่ย 1033 Bytes) รูปที่ 2.37 ได้แสดงให้เห็นอีกด้วยว่ามีปริมาณข้อมูลของ Application ที่อยู่นอกเหนือความสนใจ (Other) ได้ออกจากเครือข่ายในปริมาณมากพอสมควร โดยตัวอย่างของ Application ในกลุ่มนี้ เช่น

domain (Name-domain server: 53)
ntp (Network Time Protocol: 123)
rtsp (Real-time streaming protocol: 554)
hsrp (Router Protocol: 1985)
nsc-posa (NSC POSA; BGP daemon: 2605)
ircd (Internet Relay Chat: 6667)
pop3 (Post Office Protocol 3: email: 110)
https (HTTP over SSL: 443)
rsync (Rsync:873)
nsc-ccs (NSC CCS; OSPF daemon: 2604)
netmon (Dell Netmon; OSPF for IPv6 daemon: 2606)
เป็นต้น

2.3.4 สรุปผลการศึกษาโครงสร้างและการใช้งานเครือข่าย

เนื้อหาในบทนี้ทำให้ทราบถึงพฤติกรรมการใช้งานเครือข่ายของผู้ใช้ในมหาวิทยาลัยสงขลานครินทร์ ว่ามีการใช้เครือข่ายมากในช่วงเวลาประมาณ 10.00 น. ถึง 22.00 น. และมีการใช้ Application หรือบริการของ Web มากที่สุด นอกจากนี้ยังทำให้ทราบว่ามีบริการใดอีกบ้างที่ควรให้ความสนใจเพื่อเป็นแนวทางการกำหนดเงื่อนไขของการจัดกลุ่ม หรือการจัดลำดับความสำคัญให้กับแพ็กเก็ตของบริการนั้น ๆ โดยเนื้อหาที่เกี่ยวข้องกับการออกแบบเครื่องมือสำหรับจัดลำดับนำส่งข้อมูลจะอยู่ในบทถัดไป