บทที่ 4

อัลกอริทึมในการจัดคิว

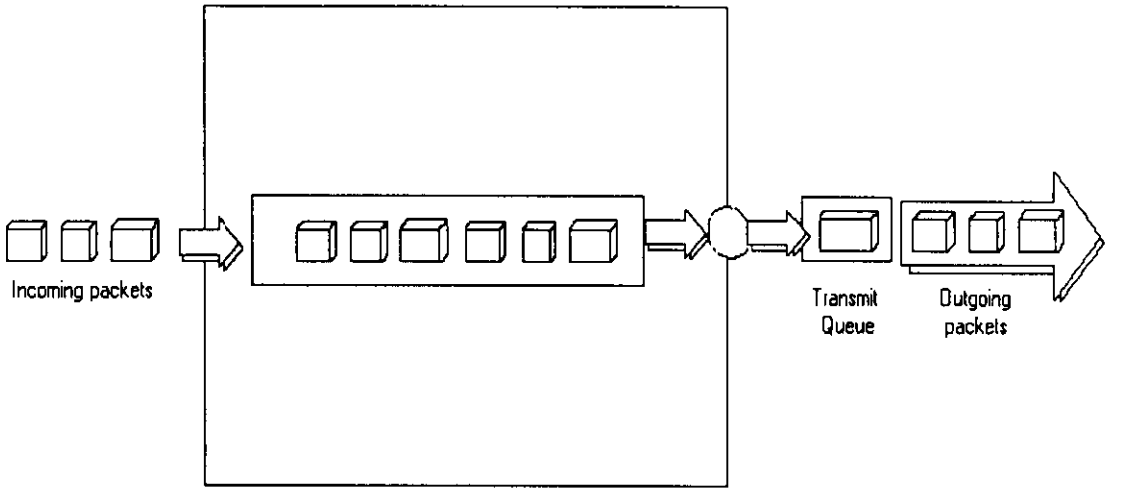
ความแออัด (Congestion) ของข้อมูลในระบบเครือข่ายเป็นสิ่งที่เกิดขึ้นได้เสมอในระบบเครือข่ายที่มีการใช้ทรัพยากรซึ่งมีอยู่อย่างจำกัดร่วมกัน อาทิเช่น แบนด์วิดท์ สาเหตุสำคัญของความแออัดเกิดจากผู้ใช้ส่งข้อมูลในขณะหนึ่ง ๆ ในปริมาณมากเกินกว่าที่ระบบเครือข่ายจะรองรับได้ ส่งผลให้ความสามารถในการรับส่งข้อมูลโดยรวมของระบบเครือข่ายมีประสิทธิภาพต่ำลง ดังนั้นเพื่อให้สามารถใช้แบนด์วิดท์ที่มีอยู่อย่างจำกัดร่วมกันได้อย่างมีประสิทธิภาพสูงสุด จึงต้องมีการจัดการข้อมูลที่เกิดขึ้นบนระบบเครือข่ายโดยใช้วิธีการจัดคิว ในบทนี้จะกล่าวถึงอัลกอริทึมในการจัดคิวแบบต่าง ๆ ไม่ว่าจะเป็นการจัดคิวแบบ First In First Out ซึ่งถือเป็นวิธีการจัดคิวที่ง่ายที่สุด การจัดคิวแบบ Fair Queuing โดยทุกแหล่งข้อมูลได้รับการบริการแบบเท่าเทียมกัน การจัดคิวแบบ Custom Queuing เป็นการจัดคิวแบบแบ่งช่องสื่อสารตามเปอร์เซ็นต์ที่ได้รับ การจัดคิวแบบ Random Early Detection เป็นวิธีการจัดคิวที่อาศัยหลักการหลีกเลี่ยงปัญหาก่อนที่ปัญหาจะเกิดขึ้น ส่วนการจัดคิวแบบ Weighted Fair Queuing คือ ทุกแหล่งข้อมูลได้รับสิทธิ์ในการใช้ช่องสื่อสารตามอัตราส่วนค่าน้ำหนักที่ได้รับ และวิธีการจัดคิวแบบ Priority Queuing โดยข้อมูลที่ได้รับค่าลำดับความสำคัญสูงกว่าจะได้สิทธิ์ในการใช้ช่องสื่อสารก่อน ส่วนข้อมูลที่ได้รับค่าลำดับความสำคัญต่ำกว่าจะถูกพักไว้ในคิวและรอจนกว่าช่องสื่อสารจะว่าง

4.1 การจัดคิวแบบ First In First Out (FIFO)

4.1.1 หลักการทำงาน

การจัดคิวแบบ FIFO หรือที่เรียกว่ามาก่อนได้รับบริการก่อน (First Come First Served หรือ FCFS) ถือว่าเป็นวิธีในการจัดคิวที่ง่ายที่สุด แสดงดังภาพประกอบ 4-1 โดยอาศัยหลักการที่เรียกว่า Store and Forward กล่าวคือ ข้อมูลชุดใดที่เข้ามาก่อนจะถูกจัดส่งออกไปก่อนตามลำดับ แต่เมื่อระบบเครือข่ายมีความแออัดเกิดขึ้น ข้อมูลจะถูกพักไว้ในคิวก่อน และจะถูกจัดส่งออกไปเมื่อช่องสื่อสารว่าง แต่วิธีนี้มีข้อเสียหลายประการ ได้แก่ ไม่มีการตัดสินใจในเรื่องของลำดับความสำคัญของข้อมูล และข้อมูลที่มาแบบไม่สม่ำเสมอ (Bursty Traffic) อาจยึดครองช่องสื่อสารตลอดเวลา ทำให้เกิดค่าหน่วงเวลา (Delay) ในคิวสูง ซึ่งส่งผลต่องานที่ต้องการความรวดเร็วในการส่ง เช่น ในการประชุมทางไกล (Video Conference) และเป็นสาเหตุให้ข้อมูลถูกทิ้งเนื่องจากข้อมูลล้น

คิว อีกทั้งการจัดคิวด้วยวิธีนี้ เมื่อเกิดความแออัดขึ้นในระบบเครือข่าย โปรโตคอลทีซีพีจะเข้าสู่กระบวนการหลีกเลี่ยงความแออัด (Congestion Avoidance) เพื่อบรรเทาปัญหา โดยการปรับลดอัตราการรับส่งข้อมูลให้อยู่ในปริมาณที่เหมาะสม แต่โปรโตคอลยูดีพีจะยังคงทำการส่งข้อมูลด้วยค่าอัตราการรับส่งข้อมูลที่คงเดิมโดยไม่ได้คำนึงถึงจำนวนข้อมูลที่ถูกทิ้ง ส่งผลให้อัตราการถูกทิ้งของข้อมูลสูงขึ้น ต่อมาได้มีการนำการจัดคิวด้วยวิธีนี้มาเป็นต้นแบบเพื่อพัฒนาไปสู่อัลกอริทึมที่ใช้ในการจัดคิวแบบอื่น ๆ โดยการกำจัดข้อเสียของการจัดคิวแบบเดิม



ภาพประกอบ 4-1 การจัดคิวแบบ First In First Out

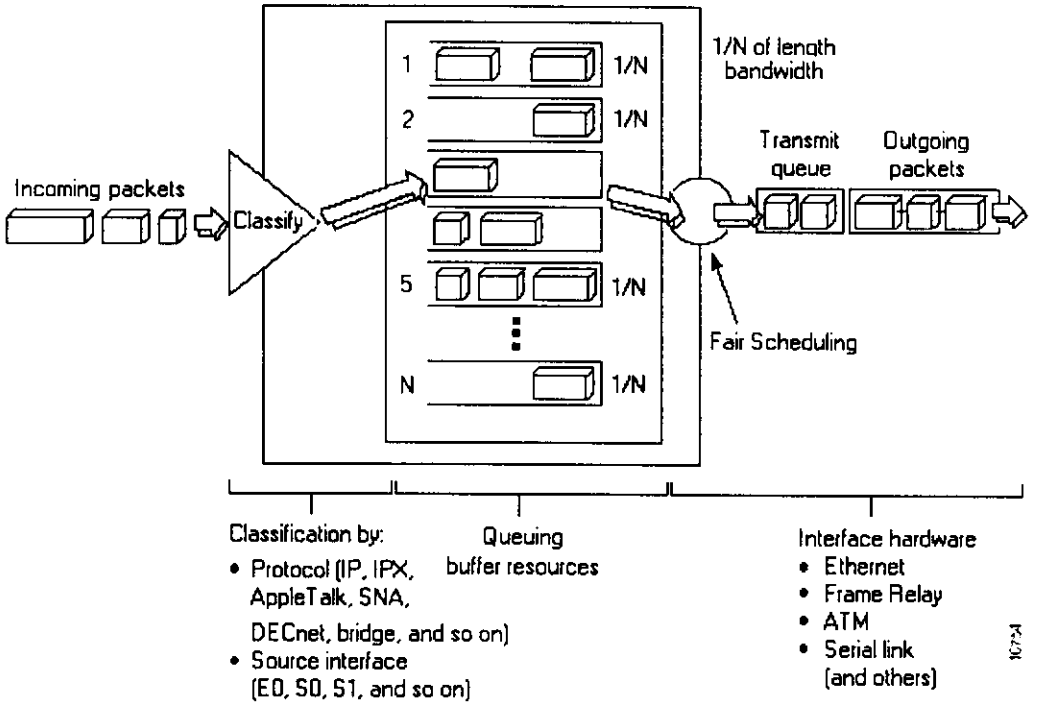
(ที่มา : http://www.cisco.com)

4.2 การจัดคิวแบบ Fair Queuing (FQ)

4.2.1 หลักการทำงาน

การจัดคิวด้วยวิธีการนี้เป็นการจัดคิวแบบยุติธรรม โดยรับประกันว่าทุกแหล่งข้อมูลจะได้รับการบริการที่เท่าเทียมกัน นั่นหมายถึงผู้ใช้จะได้รับสิทธิ์ในการใช้ช่องสื่อสารอย่างเท่าเทียมกัน และมีจุดประสงค์ที่จะกำจัดปัญหาของข้อมูลที่มาแบบไม่สม่ำเสมอที่อาจยึดครองช่องสื่อสารตลอดเวลา การจัดคิวด้วยวิธีนี้จะมีคิวหลายคิวเพื่อรองรับแพ็กเกตที่เข้ามาจากแหล่งข้อมูลแต่ละแหล่ง และแต่ละแพ็กเกตที่เข้ามาจะถูกนำไปเก็บไว้ในคิวอย่างเหมาะสม แล้วถูกส่งออกโดยจะสลับกันไปในแต่ละคิว ๆ ละ 1 แพ็กเกต โดยคิวที่ไม่มีข้อมูลจะถูกข้ามไป ดังภาพประกอบ 4-2 ข้อจำกัดของวิธีนี้ คือ ไม่สนใจเรื่องขนาดของแพ็กเกตแต่สนใจเพียงว่าใน 1 รอบแต่ละคิวต้องทำการส่งแพ็กเกตออกมา 1 แพ็กเกต ยกเว้นคิวที่ไม่มีข้อมูล หมายความว่าแพ็กเกตที่ส่งออกมาจาก

แต่ละคิวอาจมีขนาดไม่เท่ากัน ข้อจำกัดนี้สามารถกำจัดได้โดยใช้วิธีการจัดคิวที่เรียกว่า Bit-Around Fair Queuing (BRFQ) ซึ่งจะให้ความสำคัญกับขนาดของแพ็กเกตด้วย และพบว่าถ้าแพ็กเกตเข้ามาในคิวที่ซึ่งมีแพ็กเกตเพิ่งถูกส่งออกไป แพ็กเกตจะต้องรออยู่ในคิวเพื่อรอส่งในรอบถัดไป ทำให้เสียเวลาและส่งผลกับข้อมูลที่ต้องการความรวดเร็วและความต่อเนื่องของข้อมูล และเนื่องจากการจัดคิวด้วยวิธีนี้นิยมสร้างขึ้นจากซอฟต์แวร์มากกว่าฮาร์ดแวร์ทำให้มีข้อจำกัดในด้านความเร็ว จึงเหมาะสำหรับงานที่ไม่ต้องการความรวดเร็วในการส่ง แต่ต้องการความเท่าเทียมกันในการใช้งาน



ภาพประกอบ 4-2 การจัดคิวแบบ Fair Queuing

(ที่มา : http://www.cisco.com)

4.3 การจัดคิวแบบ Custom Queuing (CQ)

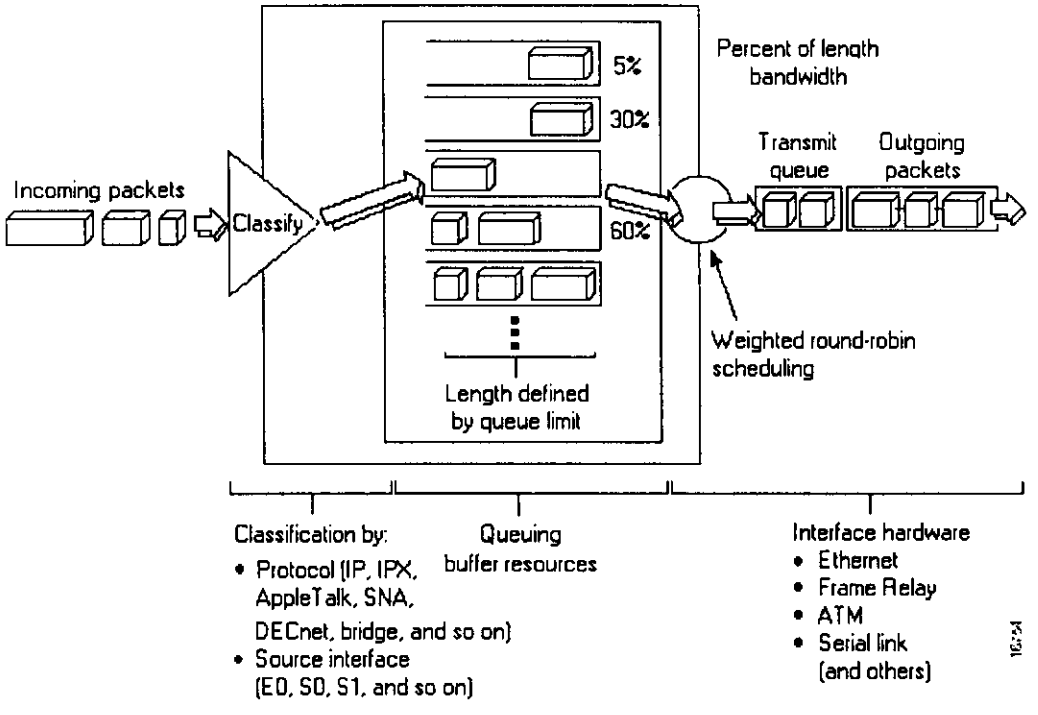
4.3.1 หลักการทำงาน

การจัดคิวด้วยวิธีนี้ถือเป็นวิธีที่รับประกันว่าอย่างน้อยทุกแหล่งข้อมูลต้องมีช่องสื่อสารให้อย่างแน่นอน และใช้วิธีการในการจัดแบ่งช่องสื่อสารโดยการกำหนดค่าแบนด์วิดท์ให้แต่ละคิวอย่างแน่นอนอยู่ในรูปเปอร์เซ็นต์ของแบนด์วิดท์ทั้งหมด ซึ่งทำให้สามารถคาดคะเนค่าอัตราการรับ

ส่งข้อมูลได้ล่วงหน้า และจะอนุญาตให้ทำการระบุจำนวนข้อมูลที่ต้องการส่งในหน่วยของจำนวนแพ็กเกตหรือในหน่วยของไบต์ได้ แบนด์วิดท์ของแต่ละคิวที่ควรได้รับมีความสัมพันธ์เป็นดังนี้

ขนาดแบนด์วิดท์ของจุดเชื่อมต่อ * จำนวนข้อมูล (หน่วยไบต์) / จำนวนข้อมูลทั้งหมด (หน่วยไบต์)

โดยแต่ละคิวจะทำการส่งข้อมูลสลับกันไป ถ้าคิวใดไม่มีข้อมูลที่จะส่งก็จะถูกข้ามไปยังคิวถัดไป และแบนด์วิดท์ในส่วนที่คิวนั้นได้รับแต่ไม่ได้ใช้งาน จะถูกนำมาใช้โดยคิวอื่น แสดงดังภาพประกอบ 4-3



ภาพประกอบ 4-3 การจัดคิวแบบ Custom Queuing

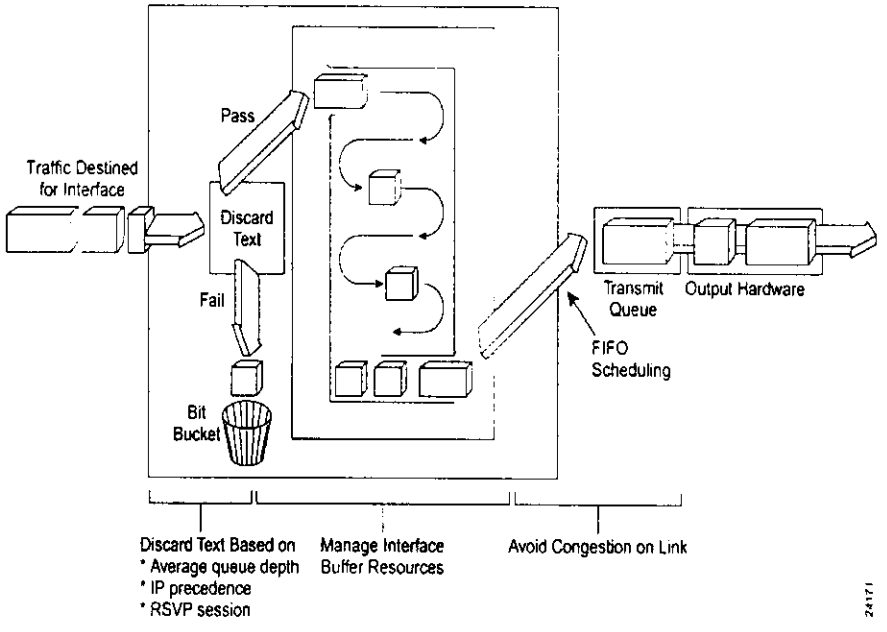
(ที่มา : http://www.cisco.com)

4.4 การจัดคิวแบบ Random Early Detection (RED)

4.4.1 หลักการทำงาน

การจัดคิวแบบ RED นี้เป็นอัลกอริทึมในการจัดคิวที่ถูกออกแบบมาเพื่อหลีกเลี่ยงความแออัดที่เกิดขึ้นบนระบบเครือข่ายก่อนที่จะเกิดปัญหาขึ้น จะใช้วิธีการทิ้งแพ็กเกตก่อนแทนที่จะรอจนกระทั่งคิวเต็ม โดยจะคอยตรวจสอบความแออัดจากการคำนวณค่าขนาดของคิวโดยเฉลี่ย (Average Queue Size) แล้วนำมาเปรียบเทียบกับค่าคงที่ ๆ ได้กำหนดไว้ 2 ค่า คือ ค่า Minimum Threshold และค่า Maximum Threshold ถ้าค่าขนาดของคิวโดยเฉลี่ยมีค่าน้อยกว่าค่า Minimum

Threshold แพ็กเกตที่เข้ามาจะถูกส่งไปยังปลายทางได้ตามปกติ แต่ถ้าค่าขนาดของคิวโดยเฉลี่ยมีค่าอยู่ระหว่างค่า Minimum Threshold และค่า Maximum Threshold แพ็กเกตที่เข้ามาจะถูกทำเครื่องหมาย (Mark) ไว้ที่ส่วนหัวของแพ็กเกต (Packet Header) ด้วยอัตราความน่าจะเป็นค่าหนึ่งซึ่งเป็นฟังก์ชันของค่าขนาดของคิวโดยเฉลี่ย แต่ถ้าค่าขนาดของคิวโดยเฉลี่ยมีค่ามากกว่าค่า Maximum Threshold แพ็กเกตที่ถูกทำเครื่องหมายไว้จะถูกทิ้ง และแพ็กเกตที่เข้ามาทุกแพ็กเกตจะถูกทำเครื่องหมายไว้ด้วย โดยแหล่งข้อมูลต้นทางจะทำการตรวจสอบการถูกทิ้งของข้อมูลพร้อมทั้งชะลอการส่งลงเมื่อขนาดของคิวโดยเฉลี่ยเกินกว่าค่าที่ได้กำหนดเอาไว้ การจัดคิวด้วยวิธีนี้ช่วยแก้ปัญหาในเรื่องของการทิ้งแพ็กเกตจากหลังทิ้งไป และช่องสื่อสารถูกใช้ตลอดเวลา



ภาพประกอบ 4-4 การจัดคิวแบบ Random Early Detection

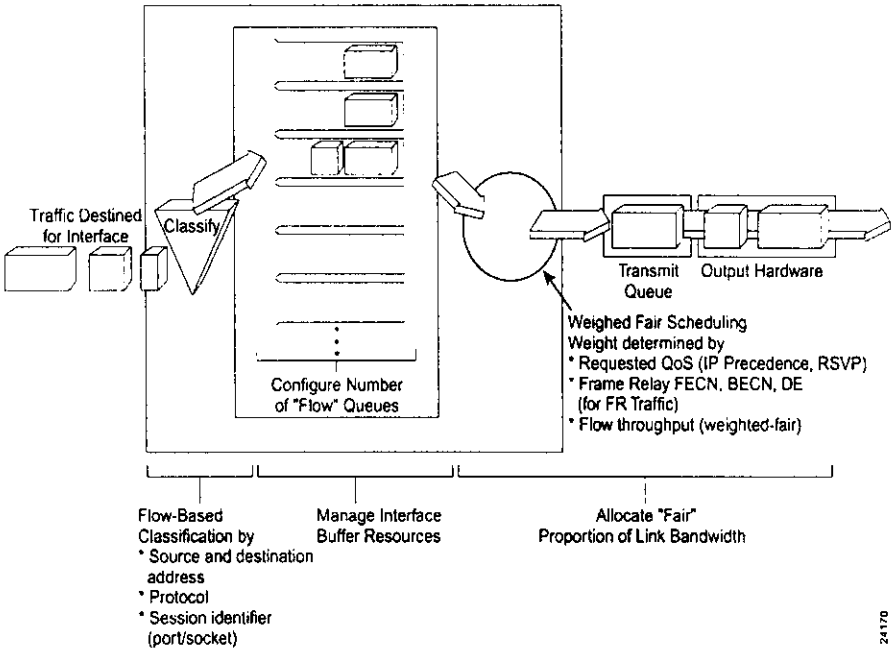
(ที่มา : http://www.cisco.com)

4.5 การจัดคิวแบบ Weighted Fair Queuing (WFQ)

4.5.1 หลักการทำงาน

การจัดคิวแบบจัดตามค่าน้ำหนักที่ได้รับ การจัดคิวด้วยวิธีนี้ถือเป็นอีกรูปแบบหนึ่งของการจัดคิวแบบ FQ โดยรับประกันว่าอย่างน้อยทุกแหล่งข้อมูลจะต้องได้รับสิทธิ์ในการใช้ช่องสื่อสารอย่างแน่นอน ด้วยวิธีการจัดแบ่งช่องสื่อสารตามอัตราส่วนของค่าน้ำหนักที่ได้รับ แสดงดังภาพประกอบ 4-5 การจัดคิวด้วยวิธีนี้เหมาะสำหรับใช้ในระบบเครือข่ายที่มีความต้องการในการใช้

แบนด์วิดท์ที่แตกต่างกัน แต่ต้องการให้ทุกแหล่งข้อมูลมีสิทธิ์ในการใช้ช่องสื่อสารนั้นได้โดยทำการจัดแบ่งช่องสื่อสารใหม่ ทำให้ช่องสื่อสารจะไม่ถูกปล่อยให้ว่างโดยไม่ได้ใช้งาน และการจัดคิวด้วยวิธีนี้ยังรองรับในเรื่องของขนาดแพ็กเกตที่ต่างกัน คือ แหล่งข้อมูลที่มีขนาดแพ็กเกตเล็กกว่าจะไม่เสียเปรียบแหล่งข้อมูลที่มีขนาดแพ็กเกตใหญ่กว่า และช่วยแก้ปัญหาข้อจำกัดของการจัดคิวแบบ FIFO ในเรื่องของการยึดครองช่องสื่อสารตลอดเวลาของกลุ่มข้อมูล หรือที่เรียกว่า Bursty Traffic



ภาพประกอบ 4-5 การจัดคิวแบบ Weighted Fair Queuing

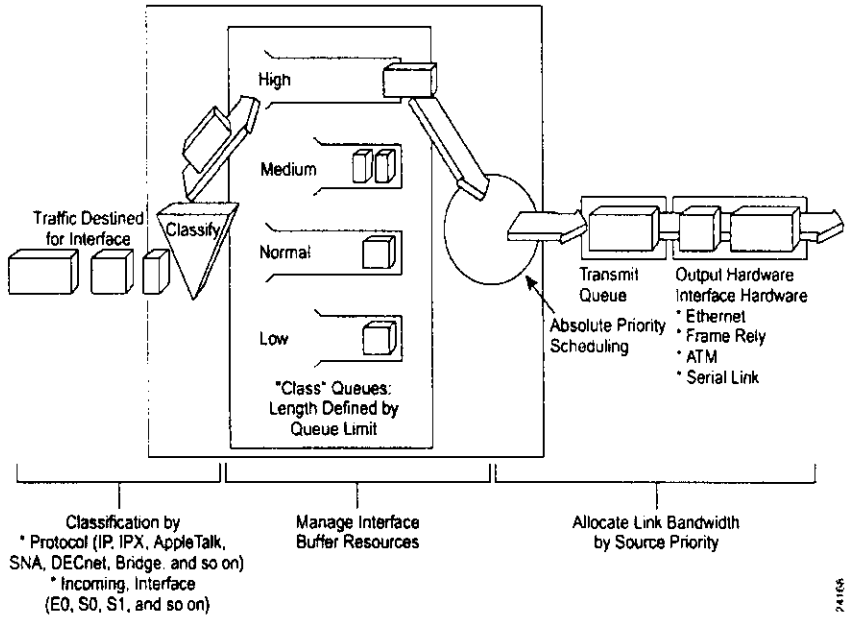
(ที่มา : http://www.cisco.com)

4.6 การจัดคิวแบบ Priority Queuing (PQ)

4.6.1 หลักการทำงาน

การจัดคิวด้วยวิธีการจัดตามลำดับความสำคัญถือเป็นอีกรูปแบบหนึ่งของการรับประกันคุณภาพในการให้บริการ QoS (Quality of Service) โดยเมื่อมีแพ็กเกตเข้ามายังตัวจัดการแบนด์วิดท์ แพ็กเกตนั้นจะถูกจัดเข้าไปยังคิวที่มีลำดับความสำคัญแตกต่างกันไปตามค่าลำดับความสำคัญที่ได้รับมา ซึ่งคิวที่มีลำดับความสำคัญสูงที่สุดจะได้รับสิทธิ์ในการส่งข้อมูลที่อยู่ภายในคิวออกไปก่อน เมื่อส่งออกไปจนหมด คิวที่มีลำดับความสำคัญรองลงมาจึงจะสามารถส่งข้อมูลออกไปได้ โดยในแต่ละคิวจะอาศัยหลักการทำงานของ FIFO กล่าวคือ แพ็กเกตใดเข้ามายังคิว

ก่อน จะถูกจัดส่งออกไปยังปลายทางก่อน แสดงดังภาพประกอบ 4-6 การจัดคิวด้วยวิธีการนี้ถือเป็นวิธีการในการช่วยแบ่งช่องสื่อสารให้ผู้ใช้อย่างเหมาะสมตามลำดับความสำคัญในการใช้งาน และข้อมูลที่มีลำดับความสำคัญมากจะไม่ถูกหน่วงเวลาโดยข้อมูลที่มีความสำคัญต่ำกว่า แต่มีข้อเสียคือ ถ้าทำการจัดลำดับความสำคัญไม่เหมาะสมอาจทำให้แพ็กเกตที่ได้รับลำดับความสำคัญต่ำไม่มีสิทธิ์ในการใช้ช่องสื่อสารเลย และส่งผลให้เกิดการหน่วงเวลาในคิวขึ้น



ภาพประกอบ 4-6 การจัดคิวแบบ Priority Queuing

(ที่มา : http://www.cisco.com)

4.7 สรุป

อัลกอริทึมในการจัดคิวมีด้วยกันหลากหลายแบบ ที่ได้กล่าวมาเป็นเพียงส่วนหนึ่งเท่านั้น และแต่ละอัลกอริทึมต่างก็มีข้อดีและข้อเสียแตกต่างกันไป ทั้งนี้ขึ้นกับวัตถุประสงค์และลักษณะการนำไปใช้งาน โดยในบทต่อไปจะได้กล่าวถึงการนำอัลกอริทึมในการจัดคิวมาปรับใช้เพื่อให้เหมาะกับระบบเครือข่ายคอมพิวเตอร์ของมหาวิทยาลัยสงขลานครินทร์