บทที่ 2

ทฤษฎีและหลักการ

ในบทนี้เป็นการกล่าวถึงทฤษฎีและหลักการพื้นฐานสำหรับการทำวิทยานิพนธ์ โดยเนื้อหาที่จะกล่าวถึงในบทนี้ประกอบด้วย หัวข้อ 2.1 การสื่อสารเสียง ซึ่งอธิบายถึงขั้นตอนในการสื่อสารเสียงบนเครือข่ายอินเตอร์เน็ต การบีบอัดเสียงโดยเน้นที่การบีบอัดเสียงด้วย G.723.1 หัวข้อ 2.2 การสื่อสารวีดิทัศน์ ซึ่งประกอบด้วยขั้นตอนในการสื่อสารวีดิทัศน์บนเครือข่ายอินเตอร์เน็ต การบีบอัดวีดิทัศน์ โดยเน้นที่การบีบอัดวีดิทัศน์ด้วย MPEG-4 หัวข้อที่ 2.3 กล่าวถึงโปรโตคอลที่ใช้สำหรับการสื่อสารเสียงและวีดิทัศน์บนเครือข่ายอินเตอร์เน็ตซึ่งก็คือ โปรโตคอลที่ชื่อว่า Real-time Transport Protocol (RTP) และสุดท้ายคือหัวข้อ 2.4 เป็นการสรุปเนื้อหาของบทนี้

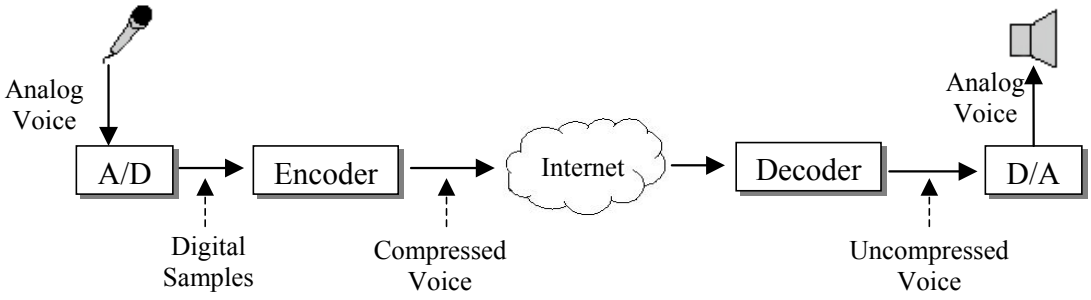
2.1 การสื่อสารเสียง

หลักการพื้นฐานของการสื่อสารเสียงที่จะกล่าวถึงในหัวข้อนี้หมายถึงการสื่อสารเสียงบนเครือข่ายอินเตอร์เน็ต ซึ่งจำเป็นจะต้องมีการแปลงสัญญาณเสียงจากอนาล็อก (Analog) ให้เป็นดิจิตอล (Digital) รวมทั้งต้องมีการลดขนาดของข้อมูลเสียงก่อนจะบรรจุลงในแพ็กเก็ตและส่งให้ผู้รับ และนอกจากนี้ผู้รับจะต้องเข้าใจข้อมูลที่ได้รับจากผู้ส่ง มิฉะนั้นผู้รับก็จะไม่สามารถนำข้อมูลไปประมวลผลได้ถูกต้อง ดังนั้นทั้งผู้รับและผู้ส่งจึงต้องใช้มาตรฐานในการสื่อสารเดียวกัน

2.1.1 ขั้นตอนในการสื่อสารเสียงบนเครือข่ายอินเตอร์เน็ต

ในการส่งเสียงพูดบนเครือข่ายอินเตอร์เน็ต ผู้ส่งจะต้องเปลี่ยนสัญญาณเสียงจากไมโครโฟนที่อยู่ในรูปแบบของสัญญาณอนาล็อกให้เป็นดิจิตอลเสียก่อน โดยใช้ตัวแปลงสัญญาณอนาล็อกเป็นดิจิตอล (Analog to Digital Converter, A/D) ดังแสดงในรูปที่ 2.1 โดยปกติแล้วเสียงพูดของมนุษย์จะมีความถี่ไม่เกิน 4 กิโลเฮิรตซ์ (kHz) [2] ดังนั้นอัตราการชักตัวอย่าง (Sampling Rate) ตามทฤษฎีของไนควิซต์ก็จะมีค่าเท่ากับ 8 กิโลเฮิรตซ์ (2 เท่าของความถี่สูงสุด) ซึ่งในระบบโทรศัพท์สาธารณะ (Public Switched Telephone Network, PSTN) ก็ใช้อัตราการชักตัวอย่างเท่ากับค่านี้เช่นกัน โดยในระบบโทรศัพท์ PSTN นั้นใช้วิธีการที่เรียกว่า Pulse Code Modulation (PCM) ในการแปลงระดับสัญญาณเสียงแบบอนาล็อกให้กลายเป็นค่าดิจิตอลที่เรียกว่า ค่าตัวอย่าง (Sample) โดยในกรณีที่ใช้คอมพิวเตอร์เป็นเครื่องมือในการสื่อสารเสียงนั้นหน้าที่นี้เป็นของการ์ดเสียง เมื่อได้ข้อมูลเสียงในรูปแบบของดิจิตอลแล้วก็สามารถส่งข้อมูลเสียงผ่านเครือข่ายอินเตอร์เน็ตไปให้กับผู้รับได้ แต่เพื่อไม่ให้ปริมาณข้อมูลเสียงที่ส่งผ่านเครือข่ายมี

มากเกินไปจึงต้องมีการบีบอัดข้อมูลเสียงเสียก่อน (Audio Compression) โดยผู้ส่งจะใช้ตัวเข้ารหัส (Encoder) ที่มีวิธีการเข้ารหัสตามมาตรฐานการบีบอัดที่เลือกใช้ ซึ่งรายละเอียดเกี่ยวกับการบีบอัดเสียงจะได้กล่าวถึงในหัวข้อถัดไป ส่วนทางฝั่งของผู้รับก็จะต้องนำแพ็กเก็ตเสียงที่ได้รับมาถอดรหัส จากนั้นก็แปลงสัญญาณเสียงจากดิจิตอลให้กลายเป็นอนาล็อกและส่งสัญญาณเสียงไปเล่น (Play) ออกทางลำโพง



รูปที่ 2.1 การส่งเสียงพูดในรูปแบบดิจิตอลบนเครือข่ายอินเตอร์เน็ต[2]

2.1.2 การบีบอัดเสียง (Audio Compression)

ในการส่งเสียงพูดผ่านเครือข่ายอินเตอร์เน็ตนั้น เพียงแต่การแปลงจากสัญญาณเสียงแบบอนาล็อกให้กลายเป็นข้อมูลเสียงแบบดิจิตอลนั้นยังไม่เพียงพอ สังเกตว่าในการแปลงสัญญาณเสียงจากอนาล็อกเป็นดิจิตอลซึ่งใช้อัตราการชักตัวอย่าง 8000 เฮิรตซ์ (ใน 1 วินาทีมีการชักตัวอย่าง 8000 ครั้ง) ถ้าหากใช้ข้อมูลดิจิตอล 8 บิตต่อหนึ่งค่าตัวอย่าง แบนด์วิดท์ที่ใช้ในการส่งข้อมูลก็จะเป็น 64 กิโลบิตต่อวินาที (kbps) และถ้าใช้ 16 บิตต่อค่าตัวอย่าง แบนด์วิดท์ก็เพิ่มเป็น 128 กิโลบิตต่อวินาที ซึ่งถือว่าเป็นค่าที่สูงทีเดียว ยิ่งไปกว่านั้นถ้าหากมีผู้ที่ต้องการสื่อสารเสียงพร้อมกันหลายคู่สนทนา เครือข่ายที่มีแบนด์วิดท์อย่างจำกัดก็จะไม่สามารถรองรับได้ ผลที่ตามมาก็คือผู้รับก็จะได้ยินเสียงไม่ชัดเจน เสียงไม่ต่อเนื่อง หรืออาจจะฟังไม่รู้เรื่องเลยก็ได้ ซึ่งวิธีแก้ไขก็คือต้องมีการลดขนาดของข้อมูลเสียงก่อนที่จะส่งข้อมูลผ่านเครือข่าย เทคนิคนี้เรียกว่าการบีบอัดเสียง

ในการบีบอัดเสียงนั้น วิธีการถอดรหัสจะต้องสอดคล้องกับวิธีการเข้ารหัส มิฉะนั้นจะไม่สามารถนำมาเสียงมาเล่นออกทางลำโพงได้ถูกต้อง สหภาพโทรคมนาคมนานาชาติ (International Telecommunication Union, ITU) ได้มีการกำหนดมาตรฐานสำหรับการบีบอัดเสียง (Audio Codec) เอาไว้หลายมาตรฐาน ตารางที่ 2.1 เป็นการเปรียบเทียบคุณสมบัติของมาตรฐานการบีบอัดเสียงแต่ประเภท โดยในคอลัมน์ของ Quality MOS (Mean Opinion Score)[25] เป็นค่าที่แสดงคุณภาพของเสียง ค่าสูงสุดมีค่าเท่ากับ 5 จะเห็นว่า G.711[26] นั้นมีคุณภาพสูงที่สุดก็จริง

แต่ก็มีอัตราบิต (Bit Rate) สูงถึง 64 kbps ทำให้ไม่สามารถนำมาใช้งานได้จริงถ้าเครือข่ายมีแบนด์วิดท์ต่ำกว่านี้ เช่น ในการเชื่อมต่ออินเตอร์เน็ตผ่านทางเครือข่ายโทรศัพท์ซึ่งมีความเร็วอยู่ที่ 28.8 ถึง 56 กิโลบิตต่อวินาทีเท่านั้น ส่วน G.729[27] และ G.723.1 ก็มีคุณภาพใกล้เคียงกับ G.711 แต่มีอัตราบิตน้อยกว่า มาตรฐานทั้งสองจึงน่าจะเป็นทางเลือกที่ดีมาตรฐานอื่น

ตารางที่ 2.1 เปรียบเทียบคุณสมบัติของมาตรฐานการบีบอัดข้อมูลเสียง[2]

Codec	Bit Rate (kbps)	Complexity compared with G.726	Algorithmic Delay	Quality MOS
G.711	64	very low	0.125	4.0
G.723.1	5.3*	8	37.5	3.9
G.723.1	6.3*	8	37.5	3.9
G.726	32	1	0.125	3.85
G.728	16	15	0.625	3.61
G.729	8	10	15	3.9
G.729A	8	6	15	3.7

*G.723.1 มี 2 อัลกอริทึม ซึ่งมีอัตราบิตต่างกันคือ 5.3 และ 6.3 กิโลบิตต่อวินาที

2.1.3 การบีบอัดเสียงด้วย G.723.1

เนื่องจากในวิทยานิพนธ์นี้ได้เลือกใช้ G.723.1[1] ในการบีบอัดเสียง ดังนั้นในหัวข้อนี้จึงได้อธิบายถึงรายละเอียดของการบีบอัดเสียงชนิดนี้ G.723.1 มีอัลกอริทึมในการบีบอัดเสียงอยู่ 2 อัลกอริทึม อัลกอริทึมแรกคือ Multi-Pulse Maximum Likelihood Quantization (MP-MLQ) ซึ่งมีการบีบอัดข้อมูลเสียงโดยแบ่งเป็นเฟรมซึ่งเทียบเท่ากับเสียงพูด 30 มิลลิวินาที (ms) หรือ 240 ค่าตัวอย่าง โดยใช้ค่าตัวอย่างขนาด 16 บิต ผลจากการบีบอัดจะได้ข้อมูลที่มีขนาดลดลงเหลือ 24 ไบต์ต่อเฟรมเท่านั้น อัลกอริทึมนี้มีอัตราบิตเท่ากับ 6.4 กิโลบิตต่อวินาที แต่ในมาตรฐาน G.723.1 ได้ระบุว่าการบีบอัดโดยใช้อัลกอริทึมนี้มีอัตราบิต 6.3 กิโลบิตต่อวินาที เนื่องจากไม่ได้คิดในส่วนของบิตที่เป็นเฮดเดอร์ (Header) ส่วนอัลกอริทึมที่สองมีชื่อว่า Algebraic Code-Excited Linear Prediction (ACELP) ซึ่งบีบอัดข้อมูลเสียงจาก 240 ค่าตัวอย่าง (ค่าตัวอย่างมีขนาด 16 บิต) ได้เป็นข้อมูลขนาด 20 ไบต์ ทำให้อัลกอริทึมนี้มีอัตราบิตเป็น 5.3 กิโลบิตต่อวินาที สำหรับอัลกอริทึมที่ใช้ในวิทยานิพนธ์นี้คือ MP-MLQ เนื่องจาก API ของการบีบอัดเสียงด้วย G.723.1 ที่ใช้ในการเขียนโปรแกรมของวิทยานิพนธ์นี้สามารถรองรับได้เฉพาะอัลกอริทึมนี้เท่านั้น

2.1.4 ปัจจัยที่มีผลกระทบต่อคุณภาพของเสียง

เนื่องจากเครือข่ายอินเตอร์เน็ตไม่ได้มีการรับประกันคุณภาพการบริการ ดังนั้นการสื่อสารเสียงบนเครือข่ายอินเตอร์เน็ตจึงมีโอกาสที่คุณภาพของเสียงที่ผู้รับได้รับจะต่ำกว่าคุณภาพของเสียงต้นฉบับได้ ซึ่งปัจจัยที่มีผลกระทบต่อคุณภาพของเสียงมีดังนี้

2.1.4.1 เวลาหน่วง

เวลาหน่วง คือ เวลาในการเดินทางของเสียงจากไมโครโฟนของผู้ส่งไปถึงลำโพงของผู้รับ ในระบบโทรศัพท์ PSTN จะไม่เกิดปัญหานี้เนื่องจากในการโทรแต่ละครั้ง จะมีการจองทรัพยากรสำหรับแต่ละคู่สายเอาไว้และค่าเวลาหน่วงจะน้อยมาก ประมาณ 50 – 70 มิลลิวินาที[28] ในขณะที่เครือข่ายอินเตอร์เน็ตไม่สามารถที่จะรับประกันได้ว่าค่าเวลาหน่วงเป็นเท่าใด โดยปกติคนเราจะรู้สึกถึงผลกระทบของเวลาหน่วงเมื่อค่าเวลาหน่วงสูงกว่า 250 มิลลิวินาที[28]

2.1.4.2 จิตเตอร์ (Jitter)

จิตเตอร์เกิดจากการที่เวลาหน่วงของแต่ละแพ็กเก็ตไม่เท่ากัน ทั้งนี้เนื่องจากในการเดินทางของแพ็กเก็ตบนเครือข่ายอินเตอร์เน็ตนั้น แต่ละแพ็กเก็ตอาจจะใช้เส้นทางที่ต่างกันได้แม้จะมีปลายทางที่เดียวกัน รวมถึงสภาพความคับคั่งของเครือข่ายที่เปลี่ยนแปลงได้ตลอดเวลาก็สามารถทำให้เกิดจิตเตอร์ได้เช่นกัน ปัญหานี้จึงทำให้สัญญาณเสียงที่ได้รับมีอาการกระตุกและฟังได้ไม่ชัดเจน ดังนั้นจึงต้องมีการใช้บัฟเฟอร์เพื่อรวบรวมข้อมูลก่อนเพื่อให้ข้อมูลมีความต่อเนื่องแล้วจึงส่งไปประมวลผลและเล่นออกทางลำโพงต่อไป การใช้บัฟเฟอร์ขนาดใหญ่จะสามารถกำจัดจิตเตอร์ได้ดีขึ้นแต่ก็มีผลทำให้มีเวลาหน่วงมากขึ้น แต่ถ้าใช้บัฟเฟอร์ขนาดเล็กเกินไปจะทำให้กำจัดจิตเตอร์ได้ไม่ดีนัก ดังนั้นจึงจำเป็นต้องหาวิธีที่ใช้ในการปรับขนาดของบัฟเฟอร์ ให้เหมาะสม

2.1.4.3 การสูญหายของแพ็กเก็ต (Packet Loss)

การที่แพ็กเก็ตสูญหายนั้นส่วนใหญ่มีสาเหตุมาจากความคับคั่งของเครือข่าย เพราะในขณะที่เราเตอร์ทำการประมวลผลแพ็กเกตหนึ่งอยู่ ถ้ามีอีกแพ็กเกตเข้ามาจะถูกเก็บไว้ในบัฟเฟอร์ของคิว และถ้าบัฟเฟอร์นั้นเต็มจะทำให้แพ็กเกตที่เข้ามาทีหลังถูกตัดทิ้ง ถึงแม้ว่าในการสื่อสารเสียงจะยอมให้มีการสูญหายของแพ็กเก็ตได้บ้าง แต่การที่มีแพ็กเก็ตสูญหายมากเกินไปก็อาจส่งผลกระทบต่อคุณภาพเสียงได้ โดยการสูญหายของแพ็กเก็ตจะมีผลกระทบอย่างชัดเจนเมื่อมีอัตราการสูญหายมากกว่า 5%[28][29]

2.1.4.4 แบนด์วิดท์ที่มีจำกัด (Limited bandwidth)

แบนด์วิดท์เป็นส่วนที่สำคัญมากในการสื่อสารเสียง ซึ่งปริมาณการใช้แบนด์วิดท์นั้นก็จะขึ้นอยู่กับอัลกอริทึมในการบีบอัดข้อมูลเสียง ถ้าแบนด์วิดท์ที่ใช้ในขณะนั้นไม่เพียงพอจะทำให้เกิดปัญหาต่าง ๆ ตามมาไม่ว่าจะเป็นเวลาหน่วง จิตเตอร์ และสูญหายของแพ็กเก็ต

2.1.5 โปรโตคอลที่ใช้ในการสื่อสารเสียงบนเครือข่ายอินเตอร์เน็ต

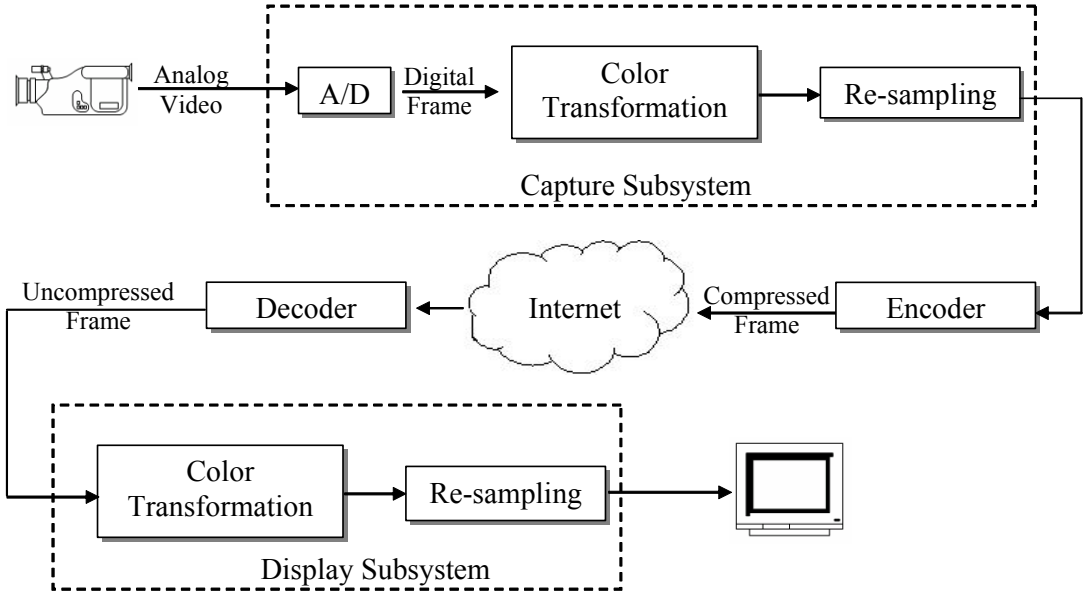
โปรโตคอลในระดับชั้นทรานสปอร์ต (Transport Layer) ของเครือข่ายอินเตอร์เน็ตที่มีการใช้งานกันอย่างแพร่หลายในปัจจุบันมีอยู่สองโปรโตคอลคือ TCP ซึ่งเป็นโปรโตคอลที่มีการรับประกันว่าสามารถส่งข้อมูลไปถึงปลายทางได้อย่างถูกต้องและครบถ้วน อีกโปรโตคอลหนึ่งก็คือ User Datagram Protocol (UDP) ซึ่งเป็นโปรโตคอลที่ไม่มีการรับประกันความน่าเชื่อถือของข้อมูล จะเห็นว่า TCP เป็นโปรโตคอลที่มีประสิทธิภาพมากกว่าและกลไกการทำงานซับซ้อนมากกว่า UDP แต่ TCP นั้นไม่เหมาะสำหรับการสื่อสารเสียง ทั้งนี้เนื่องจากความล่าช้าของกลไกการทำงานของ TCP นั่นเอง การสื่อสารเสียงนั้นต้องการความเร็ว โดยไม่จำเป็นว่าจะต้องส่งแพ็กเก็ตเสียงให้ไปถึงผู้รับอย่างครบถ้วน เพราะถึงแม้ว่าจะมีข้อมูลบางส่วนสูญหายไปบ้าง ก็ไม่ได้ส่งผลกระทบต่อคุณภาพของเสียงมากนัก ดังนั้น UDP จึงเป็นทางเลือกที่ดีกว่า แต่การที่จะใช้ UDP อย่างเดียวนั้นยังไม่เพียงพอ จึงได้มีการพัฒนาโปรโตคอลตัวหนึ่งที่ทำงานอยู่เหนือโปรโตคอล UDP โปรโตคอลตัวนี้มีชื่อว่า Real-time Transport Protocol (RTP) ซึ่งองค์กรที่เสนอโปรโตคอลตัวนี้ขึ้นมาคือ Internet Engineering Task Force (IETF) โปรโตคอล RTP ได้รับการพัฒนาขึ้นมาเพื่อใช้ในการสื่อสารแบบเวลาจริง ซึ่งรายละเอียดของของโปรโตคอลนี้จะกล่าวถึงอีกครั้งในหัวข้อที่ 2.3

2.2 การสื่อสารวีดิทัศน์

การสื่อสารวีดิทัศน์บนเครือข่ายอินเตอร์เน็ตมีขั้นตอนที่ใกล้เคียงกับการสื่อสารเสียง แต่แตกต่างกันตรงที่การจับวีดิทัศน์ (Video Capture) และการแสดงผลวีดิทัศน์มีรายละเอียดมากกว่า อีกทั้งข้อมูลของวีดิทัศน์มีขนาดใหญ่กว่าข้อมูลเสียงมาก ดังนั้นการบีบอัดวีดิทัศน์จึงถือเป็นสิ่งสำคัญมากสำหรับการสื่อสารประเภทนี้ นอกจากนี้ตัวเข้ารหัสที่ใช้ในการบีบอัดวีดิทัศน์ยังมีพารามิเตอร์ที่ต้องกำหนดมากกว่าการบีบอัดเสียง การปรับพารามิเตอร์ในการบีบอัดวีดิทัศน์บางค่าอาจทำให้ปริมาณการใช้แบนด์วิดท์ของการส่งวีดิทัศน์ลดลงได้โดยที่ไม่ส่งผลต่อคุณภาพของวีดิทัศน์มากนัก

2.2.1 ขั้นตอนในการรับและส่งวีดิทัศน์บนเครือข่ายอินเตอร์เน็ต

ในการสื่อสารวีดิทัศน์นั้น ภาพของผู้ร่วมสนทนาคนหนึ่งจะถูกส่งไปปรากฏยังอุปกรณ์แสดงผลของผู้ร่วมสนทนาอีกคน ซึ่งหลักการของการสื่อวิดีทัศน์เป็นไปในทำนองเดียวกับการสื่อสารเสียง เพียงแต่เปลี่ยนที่ตัวอุปกรณ์จากไมโครโฟนเป็นกล้องวีดิทัศน์ และเปลี่ยนจากลำโพงเป็นจอภาพเท่านั้น ในรูปที่ 2.2 เป็นขั้นตอนในการส่งวิดีทัศน์บนเครือข่ายอินเตอร์เน็ต โดยขั้นแรกจะต้องนำสัญญาณภาพจากกล้องวีดิทัศน์ซึ่งเป็นสัญญาณอนาล็อกมาเปลี่ยนเป็นสัญญาณดิจิตอลเสียก่อน โดยกล้องวีดิทัศน์จะจับภาพส่งมาทีละภาพหรือที่เรียกว่าเฟรม โดยความเร็วจะขึ้นอยู่กับอัตราเฟรม (Frame Rate) ที่กำหนด



รูปที่ 2.2 การส่งวีดิทัศน์บนเครือข่ายอินเตอร์เน็ต[2]

เมื่อได้ภาพในรูปแบบดิจิตอลแล้วก็ต้องประมวลผลภาพให้อยู่ในรูปแบบที่เหมาะสมกับวิธีการบีบอัดวีดิทัศน์ที่ใช้ รูปแบบดังกล่าวจะเกี่ยวกับระบบสี เช่น RGB หรือ YCbCr รวมไปถึงขนาดของภาพซึ่งจะต้องทำการชักตัวอย่างซ้ำ (Re-Sampling) เพื่อให้ได้ขนาดภาพตามที่ต้องการ เมื่อได้รูปแบบของภาพที่ถูกต้องแล้ว ข้อมูลของภาพดังกล่าวก็จะถูกส่งไปบีบอัดข้อมูลที่ตัวเข้ารหัสวีดิทัศน์ ก่อนที่จะส่งผ่านเครือข่ายอินเตอร์เน็ตไปยังฝั่งของผู้รับ เมื่อข้อมูลดังกล่าวเดินทางไปถึงผู้รับ ข้อมูลของภาพก็จะถูกนำไปถอดรหัสและเปลี่ยนรูปแบบของภาพให้เหมาะสมกับอุปกรณ์แสดงผล และเมื่อภาพถูกส่งมาแสดงติดต่อกันหลายภาพ ก็จะได้เป็นวีดิทัศน์ปรากฏแก่สายตาของผู้รับ

2.2.2 พารามิเตอร์ในการจับวีดิทัศน์ (Video Capture Parameters)

ในการจับวีดิทัศน์จากกล้องวีดิทัศน์หรืออุปกรณ์จับวีดิทัศน์ พารามิเตอร์ที่จะต้องกำหนดได้แก่ อัตราเฟรม ขนาดภาพ และระบบสี ซึ่งรายละเอียดของพารามิเตอร์แต่ละตัวมีดังนี้

2.2.2.1 อัตราเฟรม (Frame Rate)

วีดิทัศน์ก็คือการนำภาพหลาย ๆ ภาพมาแสดงต่อกันนั่นเอง อัตราเฟรมเป็นค่าที่บ่งบอกถึงความเร็วในการเปลี่ยนภาพดังกล่าว มีหน่วยเป็นเฟรมต่อวินาที (fps) วีดิทัศน์ที่มีอัตราเฟรมสูงจะมีความราบรื่นในการรับชมมากกว่าวีดิทัศน์ที่มีอัตราเฟรมต่ำ แต่ก็ทำให้อัตราบิตในการส่งวีดิทัศน์สูงขึ้นเช่นกัน โดยปกติแล้ววีดิทัศน์ของการฉายภาพยนตร์จะมีอัตราเฟรมอยู่ที่ 24 เฟรมต่อวินาที ส่วนการถ่ายทอดสัญญาณโทรทัศน์ในระบบ NTSC นั้นมีอัตราเฟรม 29.97 เฟรมต่อวินาที ส่วนระบบ PAL ใช้อัตราเฟรม 25 เฟรมต่อวินาที[2] สำหรับการถ่ายทอดวีดิทัศน์ผ่านเครือข่ายอินเตอร์เน็ตนั้นจะกำหนดอัตราเฟรมของวีดิทัศน์ให้ต่ำลงมา โดยใช้อัตราเฟรมเท่ากับ 15 เฟรมต่อวินาที [30][31][32]

2.2.2.2 ขนาดภาพ / ความละเอียด (Frame Size / Resolution)

ขนาดภาพของวีดิทัศน์เป็นค่าที่ระบุว่าความกว้างและความสูงของภาพมีจำนวนกี่จุดภาพ (Pixel) วีดิทัศน์ที่มีขนาดภาพใหญ่ย่อมเป็นที่พึงพอใจแก่ผู้ใช้ แต่ก็ต้องใช้อัตราบิตในการส่งวีดิทัศน์สูงกว่าวีดิทัศน์ที่มีขนาดภาพเล็กกว่า

2.2.2.3 ระบบสี (Color Space)

ระบบสีเป็นรูปแบบในการกำหนดค่าสำหรับแทนสี ระบบสีที่ใช้กันอย่างแพร่หลายในการสื่อสารวีดิทัศน์ได้แก่ ระบบ RGB และระบบ YCbCr โดยในระบบ RGB จุดภาพแต่ละจุดจะถูกนำเสนอด้วยค่า 3 ค่าคือ ค่าสีแดง (R), ค่าสีเขียว (G) และค่าสีน้ำเงิน (B) ส่วนระบบสี YCbCr จะแบ่งข้อมูลของจุดภาพตามการรับรู้ของประสาทตาของคน โดยจะแยกข้อมูลเกี่ยวกับความสว่างออกจากข้อมูลเกี่ยวกับระดับสี นั่นคือใช้องค์ประกอบ Y เก็บค่าความสว่างของจุดภาพ ส่วน Cb คือค่าความแตกต่างระหว่างค่าสีน้ำเงินกับค่าความสว่าง (B-Y) องค์ประกอบ Cr คือค่าความแตกต่างระหว่างค่าสีแดงกับค่าความสว่าง (R-Y) ส่วนค่าความแตกต่างระหว่างค่าสีเขียวกับค่าความสว่าง (Cg) นั้นไม่จำเป็นต้องเก็บ เนื่องจากว่าค่าของ Cb + Cr + Cg มีค่าคงที่ ดังนั้นถ้าหากทราบเพียงแค่ค่า Cb และ Cr ก็สามารถหาค่า Cg ได้[33] อย่างไรก็ตามไม่ว่าจะเก็บข้อมูลของ

จุดภาพในระบบ RGB หรือ YCbCr ก็สามารถแปลงเป็นอีกระบบหนึ่งได้เสมอ โดยหากต้องการแปลงจากระบบ RGB เป็นระบบ YCbCr ก็สามารถแปลงได้ดังนี้

$$Y = 0.299R + 0.587G + 0.114B$$
$$Cb = 0.564(B - Y)$$
$$Cr = 0.713(R - Y)$$

หรือในทางกลับกัน การแปลงจากระบบ YCbCr เป็นระบบ RGB ก็สามารถทำได้โดยใช้สมการต่อไปนี้

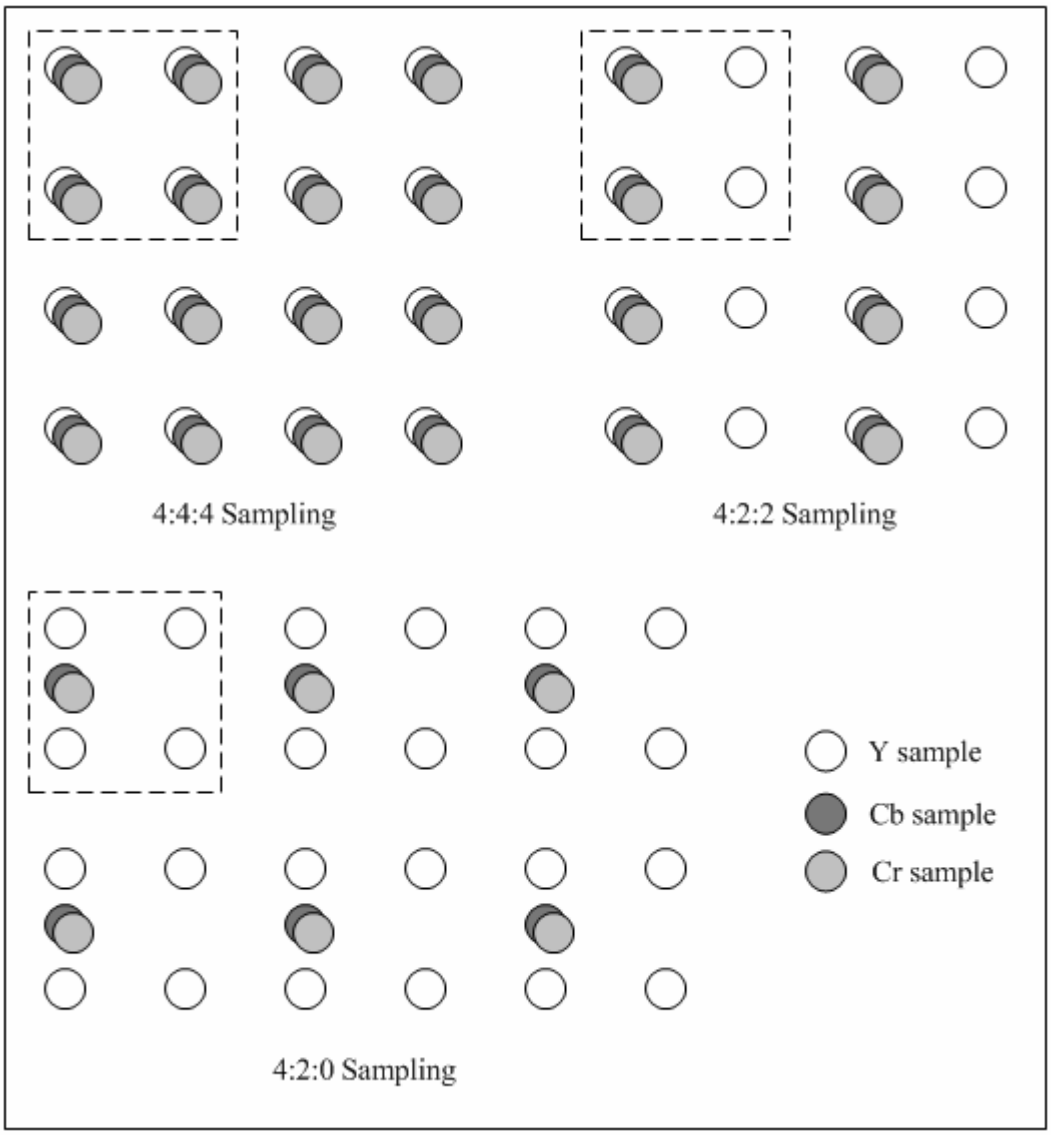
$$R = Y + 1.402Cr$$
$$G = Y - 0.344Cb - 0.714Cr$$
$$B = Y + 1.772Cb$$

นอกจากนี้ยังมีระบบสี YUV ซึ่งมีหลักการเดียวกับ YCbCr นั่นคือใช้องค์ประกอบ Y เก็บค่าความสว่างของจุดภาพ องค์ประกอบ U เก็บค่าความแตกต่างระหว่างค่าสีน้ำเงินและค่าความสว่าง ส่วนองค์ประกอบ V เก็บค่าความแตกต่างระหว่างค่าสีแดงและค่าความสว่าง หรืออาจจะกล่าวว่าใช้องค์ประกอบ U และ V แทนองค์ประกอบ Cb และ Cr ตามลำดับ มีหลายครั้งที่มีการเรียกชื่อระบบสีทั้งสองนี้สลับกัน ทั้งๆ ระบบสี YUV เป็นระบบสีที่ใช้ในระบบวีดิทัศน์อนาล็อก ส่วน YCbCr นั้นใช้กับวีดิทัศน์ในระบบดิจิตอล[34]

ระบบสี YCbCr มีรูปแบบการชักตัวอย่าง (Sampling Format) อยู่ 3 รูปแบบ คือ 4:4:4, 4:2:2 และ 4:2:0 ดังแสดงในรูปที่ 2.3 โดยตัวเลขทั้ง 3 ค่าเป็นอัตราการชักตัวอย่างสัมพัทธ์ (Relative Sampling Rate) ของแต่ละองค์ประกอบเมื่อนับในแนวนอน (Horizontal) ทั้งนี้เนื่องจากว่าสายตามนุษย์นั้นจะรับรู้การเปลี่ยนแปลงของสีได้น้อยกว่าการเปลี่ยนแปลงของความสว่าง ดังนั้นในการชักตัวอย่างบางรูปแบบจึงเก็บข้อมูลของระดับสีซึ่งอยู่ในองค์ประกอบ Cb และ Cr โดยใช้ความละเอียดที่ต่ำกว่าองค์ประกอบ Y

ในรูปแบบแรกคือ 4:4:4 อัตราการชักตัวอย่างของทุกองค์ประกอบเท่ากันหมด นั่นคือทุกจุดภาพจะมีข้อมูลของทั้งองค์ประกอบ Y, Cb และ Cr ส่วนในรูปแบบ 4:2:2 นั้นอัตราการชักตัวอย่างในแนวนอนขององค์ประกอบ Cb และ Cr เป็นครึ่งหนึ่งขององค์ประกอบ Y แต่รูปแบบสุดท้ายคือ 4:2:0 ตัวเลข 4:2:0 ไม่ตรงกับค่าอัตราการชักตัวอย่างสัมพัทธ์ของแต่ละองค์ประกอบ โดยอัตราการชักตัวอย่างขององค์ประกอบ Cb และ Cb เป็นครึ่งหนึ่งขององค์ประกอบ Y ไม่ว่าจะมองในแนวตั้งหรือแนวนอน การชักตัวอย่างในรูปแบบ 4:2:0 นี้เป็นที่นิยมใช้มากที่สุดไม่ว่าจะเป็นการประชุมวีดิทัศน์ โทรทัศน์ดิจิตอล และการเก็บวีดิทัศน์ลงแผ่น DVD[33] จะเห็นว่าการชักตัวอย่างโดยใช้รูปแบบนี้ใช้เนื้อที่ในการเก็บข้อมูลเพียงครึ่งหนึ่งของรูปแบบ 4:4:4 จากรูปที่ 2.3

หากพิจารณาการรวมกลุ่มของจุดภาพ 4 จุดที่อยู่ในเส้นประ และแต่ละองค์ประกอบใช้เนื้อที่ 8 บิต จะเห็นว่ารูปแบบ 4:2:0 ใช้เนื้อที่เพียง 48 บิต หรือ 12 บิตต่อจุดภาพ ในขณะที่รูปแบบ 4:4:4 ใช้เนื้อที่ 96 บิต หรือ 24 บิตต่อจุดภาพ



รูปที่ 2.3 รูปแบบการชักตัวอย่างของระบบสี YCbCr[32]

2.2.3 การบีบอัดข้อมูลวีดิทัศน์ (Video Compression)

การบีบอัดข้อมูลวีดิทัศน์มีจุดประสงค์เพื่อที่จะลดขนาดของข้อมูลก่อนที่จะส่งผ่านเครือข่าย เนื่องจากว่าหากใช้การส่งข้อมูลวีดิทัศน์โดยตรงโดยไม่มีการบีบอัดแล้วจะเป็นการสิ้นเปลืองทรัพยากรของเครือข่ายอย่างมาก ตัวอย่างเช่น การส่งวีดิทัศน์ที่มีอัตราเฟรม 15 เฟรม

ต่อวินาที ขนาดภาพ 320x240 จุดภาพ และใช้ข้อมูลขนาด 24 บิตต่อจุดภาพ จะเห็นว่าถ้าต้องการให้ส่งวีดีทัศน์ดังกล่าวได้อย่างราบรื่นแล้วจะต้องใช้แบนด์วิดท์ประมาณ 27.6 เมกะบิตต่อวินาที (Mbps)ในขณะที่การเชื่อมต่ออินเตอร์เน็ตผ่านทางเครือข่ายโทรศัพท์มีความเร็วเพียงแค่ 28.8 - 56 กิโลบิตต่อวินาทีเท่านั้น ดังนั้นการบีบอัดข้อมูลวีดิทัศน์จะถือว่าเป็นสิ่งที่มีความสำคัญเป็นอย่างมาก

มาตรฐานในการบีบอัดวีดิทัศน์ที่ใช้กันอยู่ในปัจจุบันนี้มีอยู่หลายมาตรฐาน แต่ละมาตรฐานก็ใช้วิธีการบีบอัดที่แตกต่างกัน ซึ่งสามารถแบ่งวิธีการเหล่านี้ได้เป็น 2 กลุ่ม โดยกลุ่มแรกนั้นจะบีบอัดวีดิทัศน์โดยใช้ข้อมูลภายในเฟรมเดียวกันเท่านั้น (Intra-Frame Compression) ตัวอย่างของมาตรฐานการบีบอัดวีดิทัศน์ที่จัดอยู่ในกลุ่มนี้คือ Motion JPEG (MJPEG) ส่วนกลุ่มที่สองนั้นนอกจากจะใช้การบีบอัดโดยใช้ข้อมูลภายในเฟรมเดียวกันแล้ว ยังใช้วิธีการบีบอัดแบบอ้างอิงกับเฟรมอื่น (Inter-Frame Compression) โดยการเปรียบเทียบว่าเฟรมปัจจุบันมีส่วนที่แตกต่างกับเฟรมที่อยู่ติดกันในส่วนใดบ้าง และเก็บข้อมูลเฉพาะส่วนที่แตกต่างกับเฟรมที่อยู่ติดกันเท่านั้น แทนที่จะเก็บข้อมูลของทั้งเฟรม ซึ่งวิธีการบีบอัดแบบนี้จะลดขนาดข้อมูลได้มากในกรณีของวีดิทัศน์ที่มีการเคลื่อนไหวน้อย

2.2.4 การบีบอัดวีดิทัศน์ด้วย MPEG-4

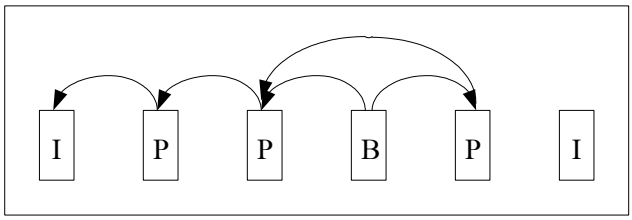
MPEG-4[3] เป็นมาตรฐานที่ได้รับการพัฒนาขึ้นโดย MPEG (Moving Picture Expert Group) เพื่อให้เป็นมาตรฐานเปิดทางด้านสื่อดิจิตอลสำหรับการเผยแพร่เนื้อหาที่มีอัตราบิตต่ำ ในมาตรฐาน MPEG-4 ประกอบด้วยข้อกำหนดหลายส่วน เช่น วิธีการบีบอัดวีดิทัศน์ วิธีการบีบอัดเสียง รูปแบบไฟล์พหุสื่อ เป็นต้น นอกจากจะเป็นข้อกำหนดสำหรับสื่อแต่ละประเภทแล้ว มาตรฐาน MPEG-4 ยังมีข้อกำหนดเกี่ยวกับระบบ (System) สำหรับที่จะนำวัตถุต่างชนิดกันมาแสดงผลพร้อมกันได้ ซึ่งจะมีลักษณะเช่นเดียวกับ AVI (Audio/Video Interleaved) แต่สิ่งที่ MPEG-4 แตกต่างจาก AVI ก็คือ นอกจากเสียงและวีดิทัศน์แล้ว MPEG-4 ยังสามารถรองรับวัตถุชนิดอื่นได้อีก ไม่ว่าจะเป็นข้อความ รูปภาพ ภาพเคลื่อนไหว รวมถึงวัตถุ 2 และ 3 มิติ ดังนั้นเมื่อกล่าวถึงการส่งสื่อ MPEG-4 จึงอาจจะไม่ใช่การส่งเฉพาะวีดิทัศน์เพียงอย่างเดียว แต่สำหรับในวิทยานิพนธ์นี้จะใช้มาตรฐาน MPEG-4 เฉพาะในส่วนของการบีบอัดวีดิทัศน์เท่านั้น โดยไม่ได้มีการใช้ข้อกำหนดเกี่ยวกับตัวระบบของ MPEG-4 และเนื้อหาของการบีบอัดวีดิทัศน์ด้วย MPEG-4 ที่จะกล่าวถึงในหัวข้อนี้ประกอบด้วย รูปแบบของวีดิทัศน์ต้นฉบับสำหรับการบีบอัดด้วย MPEG-4, ชนิดของเฟรมวีดิทัศน์ที่ผ่านการบีบอัดด้วย MPEG-4, พารามิเตอร์ในการบีบอัดวีดิทัศน์ด้วย MPEG-4, การบีบอัดในกรณีที่ไม่มีการควบคุมอัตราบิต และการบีบอัดวีดิทัศน์ในกรณีที่มีการควบคุมอัตราบิต

2.2.4.1 รูปแบบของวีดิทัศน์ต้นฉบับสำหรับการบีบอัดด้วย MPEG-4

ในการบีบอัดวีดิทัศน์จำเป็นจะต้องกำหนดรูปแบบของวีดิทัศน์ต้นฉบับให้เหมาะสมกับวิธีการบีบอัดที่ใช้ และในการบีบอัดวีดิทัศน์ด้วย MPEG-4 นั้น รูปแบบของวีดิทัศน์ต้นฉบับจะต้องใช้ระบบสี YCbCr โดยสามารถใช้ได้ทั้งแบบ 4:2:0, 4:2:2 และ 4:4:4 [33]

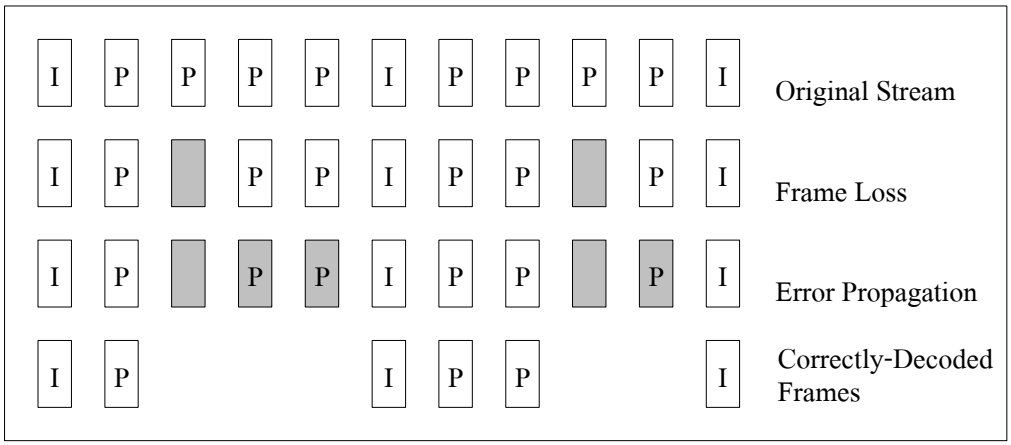
2.2.4.2 ชนิดของเฟรมวีดิทัศน์ที่ผ่านการบีบอัดด้วย MPEG-4

การบีบอัดวีดิทัศน์ตามมาตรฐาน MPEG-4 มีทั้งแบบที่ใช้ข้อมูลที่อยู่ภายในเฟรมเดียวกัน และการบีบอัดแบบที่ต้องอ้างอิงข้อมูลจากเฟรมอื่น ซึ่งนอกจะลดขนาดข้อมูลโดยใช้ความซ้ำซ้อนภายในเฟรมเดียวกันแล้ว ยังสามารถลดขนาดข้อมูลโดยใช้ความซ้ำซ้อนระหว่างเฟรมได้อีกด้วย ดังนั้นหากในสภาพแวดล้อมที่มีการเคลื่อนไหวต่ำ อย่างเช่น วีดิทัศน์ระหว่างการสนทนา เป็นต้น การบีบอัดวีดิทัศน์โดยใช้ MPEG-4 ก็สามารถลดขนาดข้อมูลได้มากทีเดียว โดยเฟรมวีดิทัศน์หลังการบีบอัดด้วย MPEG-4 แบ่งได้เป็น 3 ชนิดคือ เฟรม I (Intra Frame), เฟรม P (Predicted Frame) และเฟรม B (Bi-directional Frame) ในรูปที่ 2.4 ได้แสดงให้เห็นถึงความสัมพันธ์ของเฟรมแต่ละชนิด โดยเฟรม I เป็นเฟรมที่บีบอัดโดยใช้ข้อมูลจากเฟรมเดียวกันเท่านั้น ส่วนเฟรม P นั้นบีบอัดโดยการอ้างอิงกับเฟรม I หรือเฟรม P ที่อยู่ก่อนหน้า โดยเฟรม P นั้นจะมีขนาดน้อยกว่าเฟรม I เนื่องจากว่าเฟรม P เก็บเฉพาะข้อมูลที่ระบุว่ามีส่วนที่แตกต่างกับเฟรมก่อนหน้าอย่างไรบ้าง แต่เฟรม I นั้นเก็บข้อมูลที่ได้จากการบีบอัดวีดิทัศน์ทั้งเฟรม และชนิดสุดท้ายคือเฟรม B ใช้การบีบอัดโดยการอ้างอิงกับเฟรม P ที่อยู่ทั้งก่อนหน้าและถัดไป แต่ในการส่งวีดิทัศน์ที่ต้องการความเป็นเวลาจริง ไม่ควรใช้เฟรม B เนื่องจากเฟรม B จะต้องมีการรอวีดิทัศน์ในเฟรมถัดไปก่อนจึงจะบีบอัดได้ซึ่งจะทำให้เวลาหน่วงเพิ่มขึ้น



รูปที่ 2.4 ความสัมพันธ์ระหว่างเฟรมแต่ละชนิดที่ผ่านการบีบอัดวีดิทัศน์ด้วย MPEG-4

และเนื่องจากว่าการบีบอัดวีดิทัศน์ด้วย MPEG-4 นั้นมีการอ้างอิงข้อมูลระหว่างเฟรม ดังนั้นการสูญหายของเฟรมวีดิทัศน์อาจส่งผลให้การถอดรหัสเฟรมถัดไปเกิดข้อผิดพลาดได้ ปรากฏการณ์อย่างนี้เรียกว่า Error Propagation



รูปที่ 2.5 การเกิด Error Propagation เมื่อมีการสูญหายของเฟรมวีดิทัศน์

จากรูปที่ 2.5 จะเห็นว่าเมื่อมีเฟรมวีดิทัศน์สูญหาย เฟรม P หลังจากเฟรมที่สูญหายจะได้รับผลกระทบด้วย เนื่องจากในการบีบอัดเฟรม P นั้นต้องใช้ข้อมูลจากเฟรมที่อยู่ก่อนหน้า การถอดรหัสวีดิทัศน์จะกลับมาถูกต้องอีกครั้งเมื่อได้รับเฟรม I จากรูปจะเห็นว่าถึงแม้จะมีเฟรมสูญหายเพียง 2 เฟรม แต่สามารถส่งผลให้เกิดความผิดพลาดในการถอดรหัสวีดิทัศน์อีก 3 เฟรม รวมแล้วมีเฟรมวีดิทัศน์ที่เสียหายทั้งหมด 5 เฟรม ดังนั้นจะเห็นว่าถ้ามีจำนวนเฟรม P น้อยลงและเพิ่มจำนวนเฟรม I ให้มากขึ้น ก็จะสามารถลดผลกระทบจากการสูญหายของเฟรมวีดิทัศน์ได้ดีขึ้น แต่การเพิ่มจำนวนของเฟรม I ก็มีผลทำให้อัตราบิตของการส่งวีดิทัศน์เพิ่มขึ้นตามไปด้วย เนื่องจากว่าเฟรม I นั้นมีขนาดใหญ่กว่าเฟรม P มาก

2.2.4.3 พารามิเตอร์ในการบีบอัดวีดิทัศน์ด้วย MPEG-4

ในการบีบอัดวีดิทัศน์ด้วย MPEG-4 นั้นมีพารามิเตอร์ที่จะต้องกำหนดอยู่หลายค่า และพารามิเตอร์ที่จะกล่าวถึงต่อไปนี้ใช้ในกรณีที่มีการใช้เพียงเฟรม I และเฟรม P เท่านั้น ดังนั้นพารามิเตอร์เกี่ยวกับการบีบอัดเฟรม B จึงถูกตัดออกไป รวมทั้งพารามิเตอร์ที่มีความซับซ้อนสูง อย่างเช่น พารามิเตอร์เกี่ยวกับการประมาณความเคลื่อนไหว (Motion Estimation), เมตริกซ์การควอนไทซ์ (Quantization Matrix) ก็ไม่ได้กล่าวถึงเช่นกัน พารามิเตอร์ที่จะกล่าวถึงต่อไปนี้เป็นพารามิเตอร์ที่จะถูกนำไปใช้ในการออกแบบวิธีการควบคุมคุณภาพของวีดิทัศน์ซึ่งเป็นส่วนหนึ่งในวัตถุประสงค์ของวิทยานิพนธ์นี้

1) <u>ขนาดภาพ</u>
   ความกว้างและความสูงของวีดิทัศน์ต้นฉบับแต่ละเฟรม หน่วยเป็นพิกเซล
2) <u>อัตราเฟรม</u>
   ค่าอัตราเฟรมนี้ใช้ในการคำนวณในส่วนของการควบคุมอัตราบิต (Bit Rate Control) ดังนั้นควรกำหนดค่านี้ให้ตรงกับค่าอัตราเฟรมของวีดิทัศน์ต้นฉบับ

3) อัตราบิตเป้าหมาย (Target Bit Rate)
ในกรณีที่มีการควบคุมอัตราบิต ตัวเข้ารหัส MPEG-4 จะพยายามควบคุมให้อัตราบิตของการส่งวีดิทัศน์มีค่าใกล้เคียงกับค่าอัตราบิตเป้าหมายนี้

4) ระยะห่างระหว่างเฟรมหลัก (Key Frame Interval)
เฟรมหลักของการบีบอัดด้วย MPEG-4 ก็คือ เฟรม I เนื่องจากว่าเฟรม I ไม่ได้ขึ้นกับเฟรมอื่น ดังนั้นระยะห่างระหว่างเฟรมหลักจึงหมายถึงระยะห่างระหว่างเฟรม I สองเฟรมที่อยู่ติดกัน มีหน่วยเป็นเฟรม

5) ค่าระดับการควอนไทซ์ (Quantization Scale)
การควอนไทซ์เป็นขั้นตอนหนึ่งของการบีบอัดวีดิทัศน์ โดยกระบวนการนี้จะมีการตัดข้อมูลบางส่วนออกเพื่อช่วยให้สามารถลดข้อมูลได้ดีขึ้น การตัดข้อมูลจะมากหรือน้อยก็ขึ้นอยู่กับค่าระดับการควอนไทซ์ ถ้าค่าระดับการควอนไทซ์มากจะทำให้มีการตัดข้อมูลออกมาก คุณภาพของวีดิทัศน์ที่ส่งให้ผู้รับก็ต่ำลง แต่ขนาดข้อมูลที่จะส่งก็มีขนาดเล็กลงด้วย

2.2.4.4 การบีบอัดวีดิทัศน์ในกรณีที่ไม่มีการควบคุมอัตราบิต

ในการบีบอัดวีดิทัศน์แบบ MPEG-4 กรณีที่ไม่มีการควบคุมอัตราบิต พารามิเตอร์ที่ถือเป็นตัวแปรต้น ซึ่งต้องมีการกำหนดค่าก่อนที่จะเริ่มบีบอัดวีดิทัศน์ได้แก่ อัตราเฟรม ระยะห่างระหว่างเฟรมหลัก และค่าระดับการควอนไทซ์ ซึ่งหากกำหนดควอนไทซ์มากก็จะมีข้อมูลถูกตัดออกไปมากเช่นกัน นั่นคือคุณภาพของภาพในแต่ละเฟรมก็จะลดลงนั่นเอง ในกรณีนี้อัตราบิตของวีดิทัศน์หลังการบีบอัดจะไม่คงที่โดยจะขึ้นอยู่กับพารามิเตอร์ที่เป็นตัวแปรต้นทั้งสาม รวมทั้งลักษณะของภาพในแต่ละเฟรมของวีดิทัศน์ โดยความสัมพันธ์ระหว่างอัตราบิตและพารามิเตอร์ที่เป็นตัวแปรต้นทั้งสามเป็นดังนี้

$$B \propto \frac{F}{IQ}$$

โดย B คืออัตราบิตของการส่งวีดิทัศน์, F คืออัตราเฟรม, I คือค่าระยะห่างระหว่างเฟรมหลัก และ Q คือค่าระดับการควอนไทซ์ ซึ่งจากความสัมพันธ์ดังกล่าวจะเห็นว่าอัตราบิตแปรผันตามอัตราเฟรม แต่จะแปรผกผันกับค่าระยะห่างระหว่างเฟรมหลักและค่าระดับการควอนไทซ์ นั่นคือ หากมีการเพิ่มอัตราเฟรม อัตราบิตของการส่งวีดิทัศน์ก็จะเพิ่มตามไปด้วย แต่ถ้าเป็นการเพิ่มค่าระยะห่างระหว่างเฟรมหลักหรือเพิ่มค่าระดับการควอนไทซ์ อัตราบิตของการส่งวีดิทัศน์ก็จะลดลง

2.2.4.5 การบีบอัดวีดิทัศน์ในกรณีที่มีการควบคุมอัตราบิต

ในการบีบอัดวีดิทัศน์ด้วย MPEG-4 กรณีที่มีการควบคุมอัตราบิต หากกำหนดให้ขนาดภาพคงที่แล้ว พารามิเตอร์ที่เป็นตัวแปรต้นได้แก่ อัตราบิตเป้าหมาย อัตราเฟรม ระยะห่างระหว่างเฟรมหลัก ส่วนตัวแปรตามก็คือค่าระดับการควอนไทซ์ที่ใช้ในการบีบอัดวีดิทัศน์ในแต่ละเฟรม ซึ่งความสัมพันธ์ระหว่างค่าระดับการควอนไทซ์และตัวแปรต้นทั้งสามเป็นดังนี้

$$Q \propto \frac{F}{IB}$$

เมื่อ Q คือค่าระดับการควอนไทซ์, F คืออัตราเฟรม, I คือค่าระยะห่างระหว่างเฟรมหลัก และ B คืออัตราบิตเป้าหมาย ซึ่งจะเห็นว่าค่าระดับการควอนไทซ์แปรผกผันกับระยะห่างระหว่างเฟรมหลัก ดังนั้นหากกำหนดให้อัตราเฟรม และอัตราบิตเป้าหมายคงที่ การลดค่าระยะห่างระหว่างเฟรมหลักจะทำให้ค่าระดับการควอนไทซ์เพิ่มขึ้น นั่นคือมีการตัดข้อมูลของวีดิทัศน์ออกมากขึ้น ส่งผลให้เฟรมวีดิทัศน์หลังการบีบอัดมีขนาดลดลง นอกจากนี้การเพิ่มอัตราเฟรมหรือการลดอัตราบิตเป้าหมายก็ทำให้ค่าระดับการควอนไทซ์สูงขึ้นเช่นกัน

2.2.5 โปรโตคอลที่ใช้ในการสื่อสารวีดิทัศน์บนเครือข่ายอินเตอร์เน็ต

การสื่อสารวีดิทัศน์จัดเป็นการสื่อสารแบบเวลาจริงเช่นเดียวกับการสื่อสารเสียง ดังนั้นโปรแกรมประยุกต์สำหรับการสื่อสารวีดิทัศน์ในปัจจุบันนี้ส่วนใหญ่จึงเลือกใช้โปรโตคอล RTP เช่นเดียวกันกับการสื่อสารเสียง ซึ่งรายละเอียดของโปรโตคอลนี้จะได้กล่าวถึงในหัวข้อถัดไป

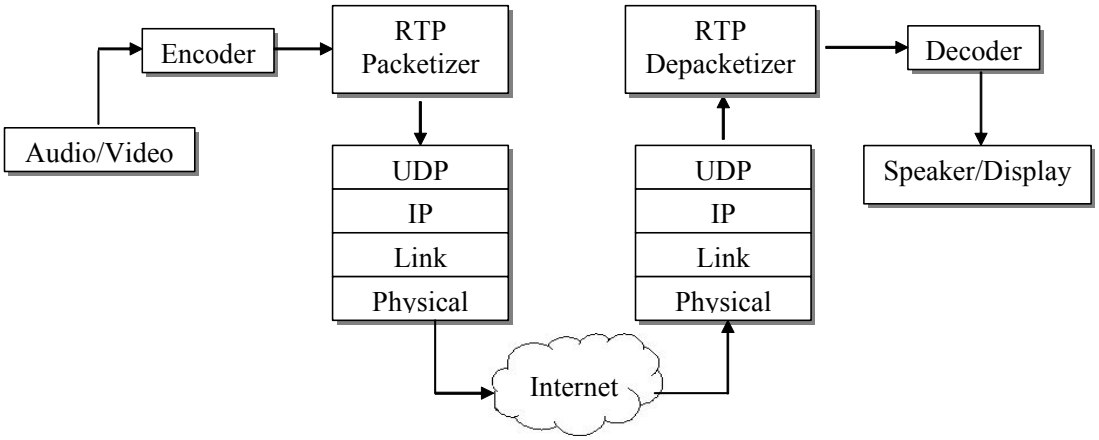
2.3 Real-time Transport Protocol (RTP)

RTP เป็นโปรโตคอลที่ได้รับการพัฒนาโดย IETF และมีข้อกำหนดอยู่ใน RFC 3550 [10] RTP ถูกออกแบบมาเพื่อใช้เป็นโปรโตคอลในการส่งข้อมูลของการสื่อสารที่ทำงานในลักษณะเวลาจริง ในวิทยานิพนธ์นี้จึงเลือกใช้โปรโตคอล RTP ทั้งในการสื่อสารเสียงและการสื่อสารวีดิทัศน์ หน้าที่สำคัญของ RTP คือ ทำให้ทางผู้รับสามารถนำข้อมูลที่ได้จากแพ็กเก็ตไปแสดงผลได้ถูกต้อง กลไกหลักที่ใช้ใน RTP ก็คือ การประทับเวลา (Time Stamping) และการกำหนดหมายเลขลำดับ (Sequence Numbering) ผู้ส่งจะกำหนดค่า Timestamp ซึ่งเป็นฟิลด์หนึ่งในเฮดเดอร์ของ RTP ตามเวลาที่ข้อมูลในไบต์แรกของแพ็กเกตถูกชักตัวอย่าง ทางด้านของผู้รับเมื่อได้รับแพ็กเก็ตจะใช้ค่า Timestamp เพื่อกำหนดจังหวะที่ถูกต้องในการแสดงข้อมูล และยังอาจจะใช้ในการประสานเวลาระหว่างสัญญาณเสียงและวีดิทัศน์ ส่วนกลไกการกำหนดหมายเลข

ลำดับนั้นใช้ฟิลด์หนึ่งในเฮดเดอร์ RTP ที่ชื่อว่า Sequence Number โดยค่าของหมายเลขนี้จะเพิ่มขึ้นตามจำนวนแพ็กเก็ตที่ส่ง ซึ่งผู้รับจะใช้ข้อมูลนี้ในการเรียงลำดับและตรวจสอบการสูญหายของแพ็กเก็ต นอกจากนี้ RTP ยังให้ข้อมูลเกี่ยววิธีการบีบอัดข้อมูลเสียงหรือวีดิทัศน์ เพื่อให้ผู้รับสามารถแสดงผลข้อมูลได้ถูกต้อง และนอกจากข้อกำหนดเกี่ยวกับการใช้โปรโตคอลในการขนส่งข้อมูลของสื่อแล้ว ใน RFC 3550 ยังมีข้อกำหนดของโปรโตคอลหนึ่งคือ RTP Control Protocol (RTCP) ซึ่งเป็นโปรโตคอลที่ถูกนำมาใช้งานร่วมกับ RTP โดยโปรโตคอล RTCP จะถูกนำไปใช้ในการรายงานเกี่ยวกับสถิติและคุณภาพการบริการ

2.3.1 บทบาทของ RTP

รูปที่ 2.6 แสดงให้เห็นถึงบทบาทของ RTP ในการสื่อสารเสียงและวีดิทัศน์ จะเห็นว่าเมื่อเสียงหรือวีดิทัศน์ได้ผ่านขั้นตอนการเข้ารหัสด้วยวิธีการบีบอัดข้อมูลแล้ว ข้อมูลที่ได้ก็จะมีขนาดลดลง หลังจากนั้นข้อมูลส่วนนี้ก็จะถูกนำมาบรรจุลงในแพ็กเก็ต RTP ซึ่งแต่ละแพ็กเก็ตก็จะมีการระบุด้วยว่าใช้วิธีการบีบอัดข้อมูลแบบใด เป็นแพ็กเก็ตในลำดับที่เท่าไร รวมไปถึงข้อมูลอื่น ๆ ที่จะกล่าวถึงในหัวข้อถัดไป ดังนั้นเมื่อข้อมูลนี้เดินทางผ่านเครือข่ายอินเตอร์เน็ตไปถึงฝั่งของผู้รับ ผู้รับก็สามารถใช้ข้อมูลนี้ในการแสดงผลได้ถูกต้อง



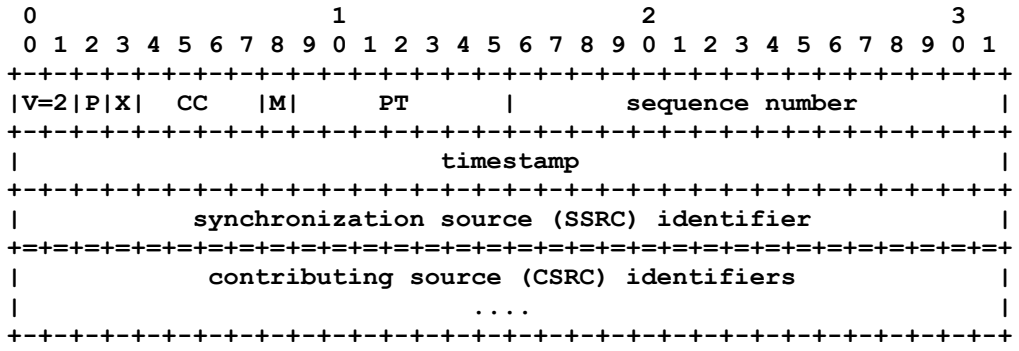
รูปที่ 2.6 บทบาทของ RTP ในการสื่อสารเสียงและวีดิทัศน์บนเครือข่ายอินเตอร์เน็ต

2.3.2 โครงสร้างแพ็กเก็ตของ RTP (RTP Packet Structure)

แพ็กเก็ต RTP แต่ละแพ็กเก็ตนั้นจะมีโครงสร้างดังที่แสดงในรูปที่ 2.7 โดยส่วนแรกของแพ็กเก็ตคือ ส่วนเฮดเดอร์ RTP (RTP Header) เป็นส่วนที่ระบุถึงคุณสมบัติของแต่ละแพ็กเก็ต โดยที่ข้อมูลเสียงหรือวีดิทัศน์ที่ผ่านการเข้ารหัสแล้วจะถูกนำไปวางในตำแหน่งของเพย์โหลด (Payload)

รูปที่ 2.7 โครงสร้างแพ็กเก็ตของ RTP

โดยทั่วไปเฮดเดอร์คือ โครงสร้างข้อมูลของแต่ละแพ็กเกตที่ถูกรวมเข้ากับข้อมูลที่จะส่ง ซึ่งในแต่ละโปรโตคอลจะมีโครงสร้างของเฮดเดอร์ที่แตกต่างกัน เพื่อกำหนดรูปแบบข้อมูลของโปรโตคอลตัวนั้น ๆ สำหรับเฮดเดอร์ของ RTP นั้นมีโครงสร้างดังรูปที่ 2.8



รูปที่ 2.8 โครงสร้างเฮดเดอร์ของ RTP[10]

จากโครงสร้างเฮดเดอร์ของ RTP ในรูปที่ 2.8 จะเห็นว่า RTP จะมีเฮดเดอร์อย่างน้อย 12 ไบต์ โดยมีรายละเอียดของฟิลด์ต่าง ๆ ดังนี้

- Version (V) กำหนดเวอร์ชันของโปรโตคอล RTP ที่ใช้ซึ่งปัจจุบันเป็นเวอร์ชัน 2
- Padding (P) เป็นฟิลด์ที่บอกว่าแพ็กเก็ตนั้นได้ถูกเติมด้วยข้อมูลที่เพิ่มให้พอดีกับขนาดของแพ็กเก็ตหรือไม่ และยังใช้ในกระบวนการการเข้ารหัสข้อมูล (Encryption) อีกด้วย
- Extension (X) เป็นฟิลด์ที่บอกว่าเฮดเดอร์ RTP มีการขยาย (Extension) หรือไม่
- CSRC Count (CC) บอกจำนวนของ Contribution Source Identifier ในแพ็กเกต โดยมี CSRC ได้ตั้งแต่ 0 – 15
- Marker (M) ค่าของบิตนี้จะถูกระบุโดย Profile และ Specification ซึ่งวิธีการใช้งานบิตนี้ได้กำหนดเอาไว้ใน RFC 3551[35]
- Payload Type (PT) บอกชนิดของข้อมูลภายในเพย์โหลด ซึ่งก็คือชนิดของการบีบอัดข้อมูลเสียงหรือวีดิทัศน์นั่นเอง ซึ่งใน RFC 3551 ได้มีการกำหนดเอาไว้ว่าใช้หมายเลขใดแทนวิธีการบีบอัดข้อมูลแต่ละแบบ

- Sequence Number บอกลำดับของแพ็กเก็ต และเพื่อให้ผู้รับตรวจสอบการสูญหายของแพ็กเก็ต และสามารถใช้ในการเรียงลำดับแพ็กเกตใหม่ได้
- Timestamp เป็นค่าที่บอกเวลาที่ใช้ในการชักตัวอย่างของข้อมูลในเพย์โหลด ซึ่งนำไปใช้ในกระบวนการคำนวณจิตเตอร์ และค่า Round Trip Time Delay
- SSRC เป็นตัวเลขประจำ Session นั่นคือแพ็กเก็ตที่มาจากผู้ส่งเดียวกัน และใช้ค่า SSRC เท่ากัน แสดงว่าเป็น Session เดียวกัน
- CSRC ฟิลด์นี้จะถูกใช้เมื่อมีการผสมสื่อ (Media Mixing) ซึ่งเป็นกระบวนการที่ใช้เมื่อมีการประชุมแบบหลายจุด (Multipoint Conference)

2.3.3 RTP Control Protocol (RTCP)

RTCP เป็นโปรโตคอลที่ถูกกำหนดอยู่ใน RFC 3550 เช่นเดียวกับ RTP โปรโตคอลตัวนี้ใช้ในการรายงานเกี่ยวกับสถิติและคุณภาพการบริการ RTCP ใช้ช่องสื่อสารแยกกันกับ RTP โดย RTCP นั้นใช้พอร์ตถัดจาก RTP ไปหนึ่งพอร์ต ช่องสื่อสารของ RTCP นั้นใช้สำหรับการรายงานเกี่ยวคุณภาพการบริการ เช่น จำนวนแพ็กเก็ตที่ได้รับ จำนวนแพ็กเก็ตที่สูญหาย จิตเตอร์ เป็นต้น แพ็กเก็ต RTCP นั้นแบ่งออกเป็น 5 ชนิดได้แก่

- SR(Sender Report) สำหรับบอกสถิติเกี่ยวกับการส่งข้อมูลของผู้ส่ง
- RR(Receive Report) สำหรับบอกสถิติเกี่ยวกับการรับข้อมูลของผู้รับ
- SDES (Source description items) เป็นรายละเอียดต่าง ๆ ของผู้ส่ง
- BYE สำหรับแจ้งถึงการจบการทำงาน
- APP (Application Specific Functions) สำหรับการกำหนดการทำงานบางอย่างในส่วนของโปรแกรมประยุกต์ ซึ่งไม่ได้ระบุเอาไว้เป็นมาตรฐาน

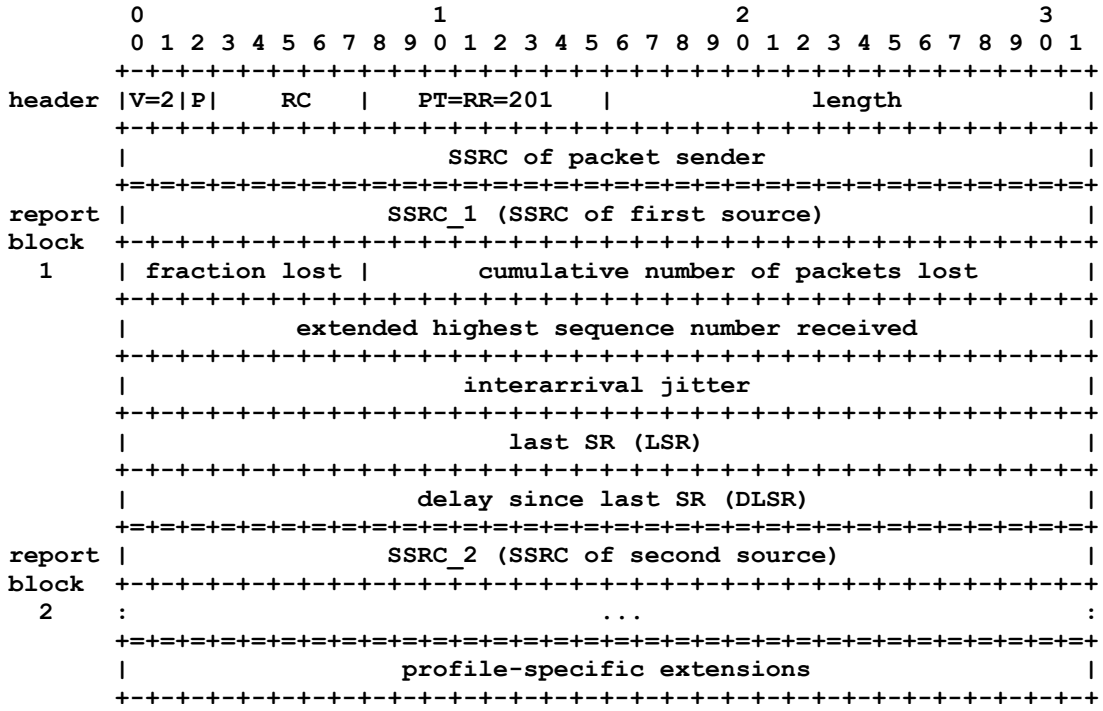
แพ็กเก็ต RTCP ที่น่าสนใจก็คือ RR เพราะว่าผู้รับจะรวบรวมสถิติต่าง ๆ ของข้อมูลเสียงหรือวีดิทัศน์ที่ได้รับจากแพ็กเก็ต RTP แล้วบรรจุลงในแพ็กเก็ต RR นี้และส่งให้กับผู้ส่ง ผู้ส่งสามารถใช้ข้อมูลในแพ็กเก็ตชนิดนี้ในการประเมินสถานะของเครือข่าย รวมถึงคุณภาพของสื่อที่ผู้รับได้รับ ในการออกแบบวิธีการควบคุมคุณภาพแบบปรับตัวของทั้งการสื่อสารเสียงและวีดิทัศน์ของวิทยานิพนธ์นี้ก็จะมีการใช้งานแพ็กเก็ตนี้เช่นกัน ในรูปที่ 2.9 เป็นโครงสร้างเฮดเดอร์ของแพ็กเก็ต RR ซึ่งฟิลด์ต่าง ๆ ที่กำหนดอยู่ในมาตรฐานแล้ว โครงสร้างเฮดเดอร์ของแพ็กเก็ต RTCP ยังอนุญาตให้มีการใส่ข้อมูลเพิ่มเติมลงในแพ็กเก็ตนี้ได้อีก โดยข้อมูลส่วนที่เพิ่มเติมนี้สามารถใส่ได้ในตำแหน่งที่เรียกว่า Profile-Specific Extension

เฮดเดอร์ของแพ็กเก็ต RR จะประกอบด้วย 3 ส่วน โดยส่วนแรกนั้นจะเป็นเฮดเดอร์ที่มีขนาดคงที่ มีขนาดรวมกันได้ 8 ไบต์ ซึ่งแต่ละฟิลด์มีความหมายดังนี้

- Version (V) กำหนดเวอร์ชันของ RTP ที่ใช้ซึ่งปัจจุบันเป็นเวอร์ชัน 2
- Padding (P) เป็นฟิลด์ที่บอกว่าแพ็กเก็ตนั้นได้ถูกเติมด้วยข้อมูลที่เพิ่มให้พอดีกับขนาดของแพ็กเก็ตหรือไม่ และยังใช้ในกระบวนการการเข้ารหัสข้อมูล (Encryption) อีกด้วย
- Payload Type (PT) ชนิดของแพ็กเก็ต ซึ่งในกรณีของแพ็กเก็ต RR ฟิลด์นี้มีค่าเท่ากับ 201
- Length ระบุความยาวของแพ็กเก็ต โดยค่านี้ในฟิลด์นี้มีที่มาดังนี้

  Length = (length of packet in byte / 4) − 1

  โดยค่าในฟิลด์ Length นี้จะมีค่าเป็นลบไม่ได้ ดังนั้นจะเห็นว่าความยาวของทั้งแพ็กเก็ตอย่างน้อยที่สุดต้องเป็น 4 ไบต์
- SSRC of packet sender หมายเลขประจำตัวที่ใช้ในการระบุผู้ที่จะส่งแพ็กเก็ต RR นี้



รูปที่ 2.9 โครงสร้างเฮดเดอร์ของแพ็กเก็ต Receiver Report (RR)[10]

เฮดเดอร์ของแพ็กเก็ต RR ในส่วนที่สองเรียกว่าบล็อกรายงาน (Report Block) ซึ่งในแพ็กเก็ต RR 1 แพ็กเก็ตอาจจะประกอบด้วยบล็อกรายงานหลายบล็อกก็ได้ถ้าหากว่าผู้รับได้รับสื่อจากผู้ส่งหลายคน โดยบล็อกรายงานแต่ละบล็อกจะมีเฮดเดอร์ดังนี้

- SSRC ค่า SSRC ที่ Session ที่ต้องการจะรายงานสถิติการรับ
- Fraction Lost อัตราการสูญหายของแพ็กเก็ตค่าในฟิลด์สามารถคำนวณได้ดังนี้

  Fraction Lost = (Number of lost packets / Number of expected packet) x 256

  และหากต้องการแปลงค่า Fraction Lost ให้กลายเป็นเปอร์เซ็นต์การสูญหายของแพ็กเก็ตก็สามารถทำได้ดังนี้

  Percent Lost = (Fraction Lost / 256) x 100

- Cumulative Number of Packets Lost ผลรวมของจำนวนแพ็กเก็ตที่สูญหายทั้งหมด
- Extended Highest Sequence Number Received หมายเลขลำดับของแพ็กเก็ตสุดท้ายที่ใช้ในการพิจารณาในการคำนวณค่าสถิติในแพ็กเก็ต RR นี้
- Interarrival Jitter ระบุค่าจิตเตอร์ของ Session โดยวิธีคำนวณค่าจิตเตอร์สามารถดูรายละเอียดได้ใน RFC 3550
- Last SR (LSR) เวลาที่ได้รับแพ็กเก็ต SR ครั้งสุดท้าย ซึ่งแพ็กเก็ตนี้ผู้ส่งสื่อจะส่งมาให้ผู้รับเป็นระยะ ๆ เพื่อรายงานสถิติการส่ง
- Delay Since Last SR (DLSR) ระยะห่างระหว่างเวลาที่ได้รับแพ็กเก็ต SR ครั้งสุดท้าย กับเวลาที่เริ่มส่งแพ็กเก็ต RR นี้

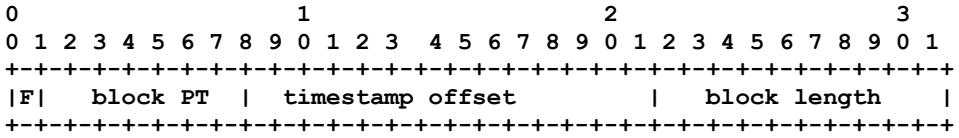
เฮดเดอร์ของแพ็กเก็ต RR ในส่วนที่สามซึ่งเป็นส่วนท้ายสุดของแพ็กเก็ตคือ ส่วน Profile-Specific Extension เป็นส่วนที่อนุญาตให้มีการเพิ่มเติมข้อมูลอื่นลงในแพ็กเก็ต RR ได้ ทั้งนี้จำเป็นจะต้องกำหนดค่าฟิลด์ Length ให้สัมพันธ์กับข้อมูลที่เพิ่มลงไปด้วย

2.3.4 การส่งข้อมูลซ้ำของเสียงโดยใช้โปรโตคอล RTP

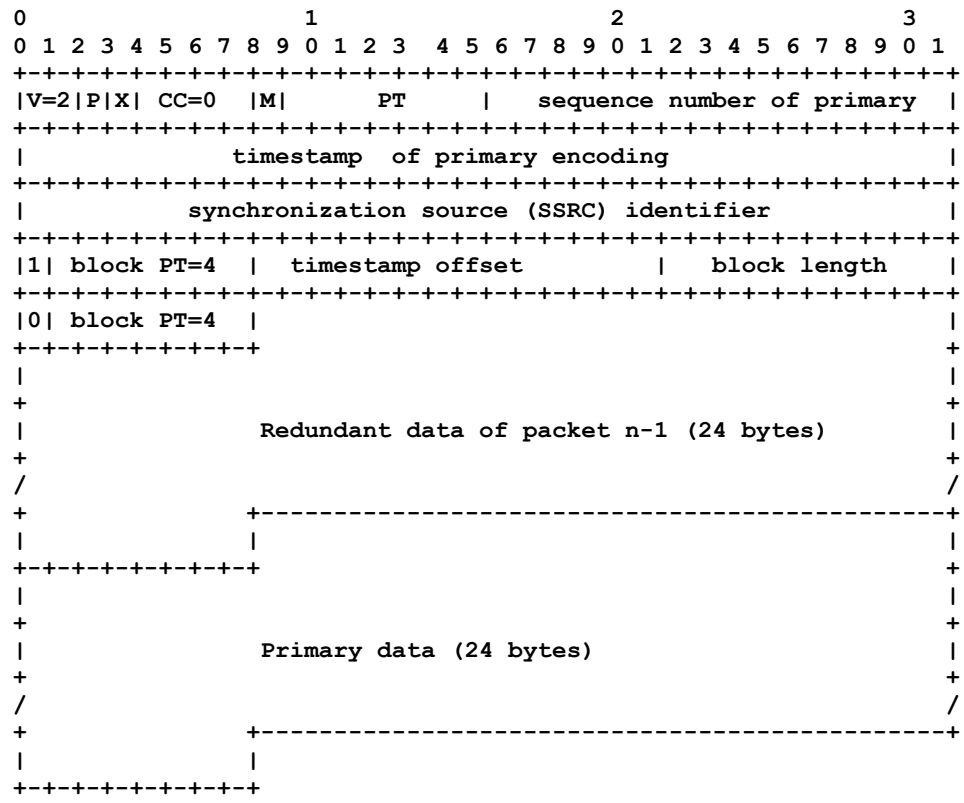
เนื่องจากการสื่อสารเสียงบนเครือข่ายอินเตอร์เน็ตนั้นไม่สามารถที่จะคาดหมายได้ว่าจะมีปริมาณแพ็กเก็ตที่สูญหายมากน้อยแค่ไหน ดังนั้นจึงได้มีการพัฒนาวิธีการที่จะลดผลกระทบจากการสูญหายของแพ็กเก็ต FEC เป็นวิธีการหนึ่งที่สามารถแก้ปัญหานี้ได้ โดยหลักการของ FEC คือในแต่ละแพ็กเก็ตเสียงจะมีการบรรจุข้อมูลซ้ำของแพ็กเก็ตอื่นอยู่ด้วย และในการสื่อสารเสียงโดยใช้โปรโตคอล RTP[10] นั้น ถ้าหากต้องการส่งข้อมูลซ้ำสามารถทำได้โดยการกำหนดรูปแบบของแพ็กเก็ตตามที่กำหนดใน RFC2198[17] จากรูปแบบของแพ็กเก็ตของ RTP นั้น หลังจากส่วนที่เป็นเฮดเดอร์แล้วก็จะตามด้วยส่วนที่เรียกว่า เพย์โหลด (Payload) ซึ่งก็คือข้อมูลเสียงที่ผ่านการบีบอัดแล้วนั่นเอง แต่เมื่อมีการส่งข้อมูลซ้ำ ส่วนของเพย์โหลดจะถูกแบ่งออกเป็นบล็อก ซึ่งมีทั้ง

บล็อกที่เป็นข้อมูลของแพ็กเก็ตปัจจุบันและบล็อกที่เป็นข้อมูลซ้ำของแพ็กเก็ตก่อนหน้า โดยแต่ละบล็อกอาจเป็นข้อมูลเสียงที่ผ่านการบีบอัดด้วยวิธีที่แตกต่างกันก็ได้ ซึ่งข้อมูลเสียงในแต่ละบล็อกจะมีเฮดเดอร์ดังรูปที่ 2.10 ซึ่งแต่ละฟิลด์มีรายละเอียดดังนี้

- F มีขนาด 1 บิต เป็นฟิลด์ที่ระบุว่ามีเฮดเดอร์ของบล็อกอื่นหลังจากนี้หรือไม่ ถ้า F เป็น 0 แสดงว่าเป็นเฮดเดอร์ของบล็อกสุดท้ายของแพ็กเก็ต
- block PT มีขนาด 7 บิต ใช้ในการระบุชนิดของเพย์โหลด (Payload Type) ซึ่งก็คือวิธีการการบีบอัดของข้อมูลเสียงในแต่ละบล็อก
- timestamp offset มีขนาด 14 บิต ค่านี้จะถูกนำไปลบกับ timestamp ที่อยู่ในเฮดเดอร์ RTP เพื่อใช้ในการระบุว่าเป็นข้อมูลซ้ำของแพ็กเก็ตใด
- block length มีขนาด 10 บิต ใช้ในการระบุความยาวของบล็อกในหน่วยไบต์



รูปที่ 2.10 เฮดเดอร์ของบล็อกในแพ็กเก็ต RTP ที่มีข้อมูลซ้ำของเสียง [17]



รูปที่ 2.11 ตัวอย่างแพ็กเก็ต RTP เมื่อมีการส่งข้อมูลซ้ำของเสียง 1 ชุด

สำหรับเฮดเดอร์ของบล็อกสุดท้ายจะมีแค่ 2 ฟิลด์คือ F และ block PT ทั้งนี้เนื่องจากข้อมูลในบล็อกสุดท้ายนั้นไม่ใช่ข้อมูลซ้ำ แต่เป็นข้อมูลเสียงของแพ็กเก็ตปัจจุบัน รูปที่ 2.11 เป็นตัวอย่างแพ็กเก็ต RTP เมื่อมีการส่งข้อมูลซ้ำของเสียงจำนวน 1 ชุด โดยทั้งข้อมูลหลักและข้อมูลซ้ำใช้การบีบอัดเสียงแบบ G.723.1 โดยทั้งบล็อกที่เป็นข้อมูลหลักและบล็อกที่เป็นข้อมูลซ้ำประกอบด้วยข้อมูลเสียงของ G.723.1 เพียง 1 เฟรม (24 ไบต์)

2.4 สรุป

ในบทนี้ได้กล่าวถึงหลักการพื้นฐานในการสื่อสารเสียงและวีดิทัศน์บนเครือข่ายอินเตอร์เน็ต โดยได้อธิบายถึงขั้นตอนในการรับและส่งสื่อทั้งสองชนิดตั้งแต่การติดต่อกับอุปกรณ์จนถึงการส่งข้อมูลของสื่อผ่านเครือข่าย และการบีบอัดโดยเน้นที่การบีบอัดเสียงด้วย G.723.1 และการบีบอัดวีดิทัศน์ด้วย MPEG-4 นอกจากนี้ยังได้อธิบายถึงโปรโตคอล RTP ซึ่งเป็นโปรโตคอลที่ใช้ในการส่งสื่อทั้งสองชนิดบนเครือข่ายอินเตอร์เน็ต โดยได้แสดงให้เห็นถึงรายละเอียดของโครงสร้างแพ็กเก็ต RTP และ RTCP รวมทั้งการส่งข้อมูลซ้ำของเสียงโดยใช้โปรโตคอล RTP ซึ่งเป็นหลักการพื้นฐานที่จะใช้การควบคุมความผิดพลาดของเสียงด้วยวิธี Forward Error Correction

จากหลักการพื้นฐานของการสื่อสารเสียงและวีดิทัศน์ที่ได้กล่าวไปแล้วนั้น จะเห็นว่ามีช่องทางสำหรับการควบคุมคุณภาพของสื่อทั้งสองชนิดได้ เช่น ในการโปรโตคอล RTP ผู้รับสามารถรายงานสถิติการรับให้ผู้ส่งได้รับทราบโดยใช้แพ็กเก็ต RTCP Receiver Report ในการสื่อสารเสียงก็สามารถที่จะควบคุมความผิดพลาดได้โดยใช้วิธีการส่งข้อมูลซ้ำ ซึ่งวิธีการนี้มีมาตรฐานรองรับแล้ว และในการสื่อสารวีดิทัศน์โดยใช้การบีบอัด MPEG-4 นั้นจะเห็นว่ามีพารามิเตอร์ของการบีบอัดอยู่หลายตัว ซึ่งการปรับเปลี่ยนค่าพารามิเตอร์เหล่านี้อาจจะช่วยในการควบคุมคุณภาพของวีดิทัศน์เมื่อสภาพของเครือข่ายมีการเปลี่ยนแปลง ซึ่งเนื้อหาเกี่ยวกับวิธีการควบคุมคุณภาพของการสื่อสารเสียงและวีดิทัศน์จะอยู่ในบทที่ 3 และ 4 ตามลำดับ