

CHƯƠNG BA

NÉN ẢNH, VIDEO VÀ AUDIO

Nội dung:

- Nén ảnh : JPEG
- Nén Video : MPEG -1
- Nén âm thanh

1. Nén ảnh

1.1. Quy trình nén ảnh JPEG

1.1.1. Mã hóa

Chuyển đổi không gian màu

Đầu tiên 1 ảnh cần được chuyển từ dạng RGB sang 1 không gian ảnh khác là YCbCr. Nó có 3 thành phần là Y, Cb và Cr : với Y là thành phần biểu diễn độ sáng và Cb,Cr là thành phần màu sắc. Không gian màu YCbCr cho phép việc nén tốt hơn với cùng 1 chất lượng ảnh.

Việc chuyển đổi này dùng chuẩn JFIF và được thực hiện tương thích hoàn toàn với file JPEG. Tuy nhiên, những ảnh JPEG “chất lượng cao” bỏ qua bước này và vẫn giữ không gian màu RGB và mỗi màu được nén và lượng tử hóa riêng biệt

Downsampling

Mắt người có thể nhìn thành phần Y chi tiết hơn thành phần Cb và Cr. Nhờ đó việc nén hiệu quả hơn. Bằng cách giảm các thành phần Cb và Cr (downsampling hay subsampling). Tỷ lệ giảm mẫu của ảnh JPEG thường là 4:2:0 (giảm 2 lần theo chiều ngang và dọc). Việc giảm mẫu giảm 35%-50% không gian lưu trữ ảnh. Việc xử lý Y, Cb và Cr được thực hiện riêng biệt và tương tự nhau.

Chia khối.

Sau khi giảm mẫu, mỗi kênh phải được chia thành những khối 8x8 điểm ảnh. Nếu dữ liệu cho 1 kênh không phải là các số nguyên thì bộ mã hóa phải chuyển đổi những khối có dữ liệu không phù hợp với những dữ liệu thừa.

Xử lý những điểm ảnh ở biên bằng những điểm ảnh đen

Việc thay thế những điểm ảnh ở biên phổ biến nhưng không phải là kỹ thuật tối ưu tránh viền nhìn thấy nhưng vẫn tạo ra những artifact với thiết bị đo màu của của những ô phủ đầy. Một chiến lược tốt hơn là phủ những điểm ảnh sử dụng những màu duy trì hệ số DCT của những điểm ảnh nhìn thấy, ít nhất là cho những thành phần tần số thấp. (ví dụ : phủ với những màu trung bình của phần nhìn thấy sẽ duy trì hệ số DC đầu tiên, nhưng việc khớp tốt nhất cho 2 hệ số AC tiếp theo sẽ cho ra những kết quả tốt hơn với ít nhìn thấy hơn những ô ở cạnh (theo đường viền).

Ảnh con 8x8 ở trên sau khi cân bằng histogram (vd : 154 → trắng, 55 → đen). Chú ý rằng điều này được hoàn thành chỉ vì mục đích thị giác và việc cân bằng không được thực hiện ở dữ liệu thử.

Tiếp đó, mỗi thành phần (Y,Cb,Cr) của mỗi khối 8x8 được chuyển đổi thành biểu diễn trên miền tần số, sử dụng biến đổi DCT 2 chiều, bình thường để biến đổi.

Mã hóa Entropy

Mã hóa entropy là 1 dạng đặc biệt của nén dữ liệu không mất mát thông tin. Nó là dạng nén dựa trên mã hóa độ dài (RLE) tạo nhóm những tần số tương tự với nhau, thêm độ dài những số 0 và sử dụng mã hóa Huffman cho phần khác 0.

Chuẩn JPEG thường sử dụng mã hóa Huffman thay vì mã hóa số học bởi vì thuật toán mã hóa số học chậm hơn rất nhiều thuật toán mã hóa Huffman kể cả mã hóa và giải mã dù cho phép nén tốt hơn khoảng 5%.

Nếu khối thứ I là Bi và vị trí mỗi khối là (i,j) với p,q=1..7 thì mỗi hệ số DCT biểu diễn bởi Bi(p,q). Vì vậy, với lược đồ trên, thứ tự mã hóa điểm ảnh là Bi(0,0), Bi(0,1), Bi(1,0), Bi(2,0), Bi(1,1), Bi(0,2), Bi(0,3), Bi(1,2) và tiếp tục như thế

Quá trình nén sẽ được thực hiện được gọi là mã hóa tuần tự (sequential encoding). Đường cơ sở của JPEG cũng hỗ trợ mã hóa tăng dần. Khi tiến hành mã hóa tuần tự, từng điểm ảnh theo thứ tự như trên được đưa vào và cứ thế tiếp tục cho đến khi còn lại toàn bộ các điểm ảnh còn lại có giá trị 0, kết thúc bằng ký tự “EOB”.

Việc mã hóa như trên sẽ chia các hệ số thành 2 phần

- Những hệ số có ý nghĩa có dấu.
- Những hệ số 0 đi theo sau cùng.

Như ví dụ các hệ số trên, hầu hết những hệ số đã lượng tử hóa là những số nhỏ. Những trường hợp này có nghĩa sẽ được mô tả bởi những từ mã ngắn hơn.

Chuẩn JPEG cung cấp những bảng Huffman khác nhau; bộ mã hóa có thể lựa chọn để tạo ra bảng Huffman tối ưu cho những tần số cụ thể cho những ảnh được mã hóa.

1.1.2. Giải mã

Quá trình giải mã sẽ đảo ngược các quá trình mã hóa ở trên.

1.2. Ví dụ về nén ảnh JPEG

1.2.1. Quá trình mã hóa

Ví dụ về 1 ảnh nhỏ 8x8 :

52	55	61	66	70	61	64	73
63	59	55	90	109	85	69	72
62	59	68	113	144	104	66	73
63	58	71	122	154	106	70	69
67	61	68	104	126	88	68	70
79	65	60	70	77	68	58	75
85	71	64	59	55	61	65	83
87	79	69	68	65	76	78	94

Chuyển những ảnh con sang 1 khoảng lấy 0 là điểm trung bình, với ảnh 8 bit có 256 giá trị thì trừ mỗi giá trị cho 128.

$$\frac{2^{bit}}{2} = \frac{2^8}{2} = 2^7 = 128$$

Kết quả nhận được là :

$$\begin{array}{c}
 x \\
 \longrightarrow \\
 \begin{bmatrix}
 -76 & -73 & -67 & -62 & -58 & -67 & -64 & -55 \\
 -65 & -69 & -73 & -38 & -19 & -43 & -59 & -56 \\
 -66 & -69 & -60 & -15 & 16 & -24 & -62 & -55 \\
 -65 & -70 & -57 & -6 & 26 & -22 & -58 & -59 \\
 -61 & -67 & -60 & -24 & -2 & -40 & -60 & -58 \\
 -49 & -63 & -68 & -58 & -51 & -60 & -70 & -53 \\
 -43 & -57 & -64 & -69 & -73 & -67 & -63 & -45 \\
 -41 & -49 & -59 & -60 & -63 & -52 & -50 & -34
 \end{bmatrix}
 \begin{array}{c}
 \downarrow \\
 y
 \end{array}
 \end{array}$$

Chuyển đổi DCT 64 điểm ảnh sang 1 kết hợp tuyến tính của 64 hình vuông. Chiều ngang là u , chiều dọc là v .

$$G_{u,v} = \alpha(u)\alpha(v) \sum_{x=0}^7 \sum_{y=0}^7 g_{x,y} \cos \left[\frac{\pi}{8} \left(x + \frac{1}{2} \right) u \right] \cos \left[\frac{\pi}{8} \left(y + \frac{1}{2} \right) v \right]$$

Với

u là thành phần ngang không gian tần số với $0 \leq u < 8$.

v là thành phần dọc không gian tần số với $0 \leq v < 8$.

$$\alpha_p(n) = \begin{cases} \sqrt{\frac{1}{8}}, & \text{if } n = 0 \\ \sqrt{\frac{2}{8}}, & \text{otherwise} \end{cases}$$

là 1 hàm

$g_{x,y}$ là giá trị điểm ảnh tại tọa độ (x, y)

$G_{u,v}$ là hệ số DCT tại coordinates (u, v)

Nếu thực hiện theo ví dụ trên ta có (làm tròn số)

$$\begin{array}{c}
 u \\
 \longrightarrow \\
 \begin{bmatrix}
 -415 & -30 & -61 & 27 & 56 & -20 & -2 & 0 \\
 4 & -22 & -61 & 10 & 13 & -7 & -9 & 5 \\
 -47 & 7 & 77 & -25 & -29 & 10 & 5 & -6 \\
 -49 & 12 & 34 & -15 & -10 & 6 & 2 & 2 \\
 12 & -7 & -13 & -4 & -2 & 2 & -3 & 3 \\
 -8 & 3 & 2 & -6 & -2 & 1 & 4 & 2 \\
 -1 & 0 & 0 & -2 & -1 & -3 & 4 & -1 \\
 0 & 0 & -1 & -4 & -1 & 0 & 1 & 2
 \end{bmatrix}
 \begin{array}{c}
 \downarrow \\
 v
 \end{array}
 \end{array}$$

Những thành phần vượt quá giới hạn 8 bit được gọi là hệ số DC, còn lại được gọi là hệ số AC. DCT có thể tạm thời tăng độ sâu của ảnh lên 11 hay 12 bit để lưu trữ. Khi lập trình có thể chuyển sang số 16 bit để biểu diễn tại thời điểm này. Ưu điểm của DCT là xu hướng tập trung hầu hết tín hiệu vào 1 góc như ở trên. Bước lượng tử hóa dưới đây tác động đến hiệu ứng này khi lập tức giảm kích thước của những hệ số DCT sang 8 bit hoặc thấp hơn, kết quả là một vùng lớn có giá trị là 0 và có thể được loại bỏ.

Việc tăng tạm thời kích thước tại bước này không ảnh hưởng tới việc thực hiện nén JPEG, bởi vì chỉ có 1 phần rất nhỏ của ảnh được lưu trữ trong dạng DCT đầy đủ tại thời điểm nào trong quá trình mã hóa hay giải mã.

Lượng tử hóa

Mắt người cảm nhận tốt khi nhìn sự thay đổi khác nhau nhỏ trong độ sáng qua 1 vùng tương đối lớn nhưng không tốt khi phân biệt chính xác độ mạnh của giá trị độ sáng tần số cao. Điều này cho phép giảm số lượng thông tin của thành phần tần số cao được thực hiện bằng cách chia mỗi thành phần trong miền tần số cho 1 số cố định với mỗi thành phần, rồi làm tròn số. Đây là quá trình gây mất mát thông tin chính trong cả quá trình. Kết quả quá trình này là hầu hết các thành phần tần số cao bị suy biến về 0, và nhiều thành phần khác là những số nhỏ hơn hoặc là số âm và cần ít bit hơn để lưu trữ.

Một ma trận lượng tử hóa quen thuộc là :

$$\begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}$$

Việc lượng tử hóa các hệ số DCT thực hiện bởi :

$$B_{j,k} = \text{round} \left(\frac{A_{j,k}}{Q_{j,k}} \right) \text{ for } j = 0, 1, 2, \dots, N_1 - 1; k = 0, 1, 2, \dots, N_2 - 1$$

Với A là hệ số DCT, Q là ma trận ở trên và B là hệ số được lượng tử hóa DCT

Kết quả nhận được sau khi tính sẽ là :

$$\begin{bmatrix} -26 & -3 & -6 & 2 & 2 & -1 & 0 & 0 \\ 0 & -2 & -4 & 1 & 1 & 0 & 0 & 0 \\ -3 & 1 & 5 & -1 & -1 & 0 & 0 & 0 \\ -4 & 1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Với hệ số DC -415 và làm tròn số ta có

$$\text{round} \left(\frac{-415}{16} \right) = \text{round}(-25.9375) = -26$$

Mã hóa Entropy

Thứ tự cho những hệ số ở trên có thể xem ở hình dưới

$$\begin{array}{cccccccc} -26 & & & & & & & \\ -3 & 0 & & & & & & \\ -3 & -2 & -6 & & & & & \\ 2 & -4 & 1 & -4 & & & & \\ 1 & 1 & 5 & 1 & 2 & & & \\ -1 & 1 & -1 & 2 & 0 & 0 & & \\ 0 & 0 & 0 & -1 & -1 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}$$

0	0	0	0	0	0	0
0	0	0	0	0	0	
0	0	0	0	0		
0	0	0	0			
0	0	0				
0	0					
0						
0						

Nếu khối thứ I là Bi và vị trí mỗi khối là (i,j) với p,q=1..7 thì mỗi hệ số DCT biểu diễn bởi Bi(p,q). Vì vậy, với lược đồ trên, thứ tự mã hóa điểm ảnh là Bi(0,0), Bi(0,1), Bi(1,0), Bi(2,0), Bi(1,1), Bi(0,2), Bi(0,3), Bi(1,2) và tiếp tục.

Quá trình nén sẽ được thực hiện được gọi là mã hóa tuần tự (sequential encoding). Đường cơ sở của JPEG cũng hỗ trợ mã hóa tăng dần. Khi tiến hành mã hóa tuần tự, từng điểm ảnh theo thứ tự như trên được đưa vào và cứ thế tiếp tục cho đến khi còn lại toàn bộ các điểm ảnh còn lại có giá trị 0, kết thúc bằng ký tự “EOB”. Khi áp dụng nguyên tắc trên, các hệ số còn lại phải nén sẽ là :

0
EOB

Phần tiếp theo sẽ thực hiện mã hóa Huffman.

1.2.2. Quá trình giải mã

Quá trình giải mã ảnh là quá trình đảo ngược những quá trình ở trên
Với hệ số ma trận DCT (sau khi thêm sự sai khác của hệ số DC)

$$\begin{bmatrix} -26 & -3 & -6 & 2 & 2 & -1 & 0 & 0 \\ 0 & -2 & -4 & 1 & 1 & 0 & 0 & 0 \\ -3 & 1 & 5 & -1 & -1 & 0 & 0 & 0 \\ -4 & 1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Và nhân các số trong ma trận với các hệ số trong ma trận lượng tử hóa ta có kết quả như sau :

$$\begin{bmatrix} -416 & -33 & -60 & 32 & 48 & -40 & 0 & 0 \\ 0 & -24 & -56 & 19 & 26 & 0 & 0 & 0 \\ -42 & 13 & 80 & -24 & -40 & 0 & 0 & 0 \\ -56 & 17 & 44 & -29 & 0 & 0 & 0 & 0 \\ 18 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Bằng cách đảo ngược DCT, kết quả trong ảnh với những giá trị là :

Có thể thấy sự sai khác chút ít giữa ảnh gốc và ảnh giải nén có thể thấy rõ nhất là ở góc dưới bên trái.

$$\begin{bmatrix} -68 & -65 & -73 & -70 & -58 & -67 & -70 & -48 \\ -70 & -72 & -72 & -45 & -20 & -40 & -65 & -57 \\ -68 & -76 & -66 & -15 & 22 & -12 & -58 & -61 \\ -62 & -72 & -60 & -6 & 28 & -12 & -59 & -56 \\ -59 & -66 & -63 & -28 & -8 & -42 & -69 & -52 \\ -60 & -60 & -67 & -60 & -50 & -68 & -75 & -50 \\ -54 & -46 & -61 & -74 & -65 & -64 & -63 & -45 \\ -45 & -32 & -51 & -72 & -58 & -45 & -45 & -39 \end{bmatrix}$$

Cộng 128 vào mỗi số ta có

$$\begin{bmatrix} 60 & 63 & 55 & 58 & 70 & 61 & 58 & 80 \\ 58 & 56 & 56 & 83 & 108 & 88 & 63 & 71 \\ 60 & 52 & 62 & 113 & 150 & 116 & 70 & 67 \\ 66 & 56 & 68 & 122 & 156 & 116 & 69 & 72 \\ 69 & 62 & 65 & 100 & 120 & 86 & 59 & 76 \\ 68 & 68 & 61 & 68 & 78 & 60 & 53 & 78 \\ 74 & 82 & 67 & 54 & 63 & 64 & 65 & 83 \\ 83 & 96 & 77 & 56 & 70 & 83 & 83 & 89 \end{bmatrix}$$

Và ta có ma trận sai khác giữa ảnh gốc và ảnh sau khi giải nén là:

$$\begin{bmatrix} -8 & -8 & 6 & 8 & 0 & 0 & 6 & -7 \\ 5 & 3 & -1 & 7 & 1 & -3 & 6 & 1 \\ 2 & 7 & 6 & 0 & -6 & -12 & -4 & 6 \\ -3 & 2 & 3 & 0 & -2 & -10 & 1 & -3 \\ -2 & -1 & 3 & 4 & 6 & 2 & 9 & -6 \\ 11 & -3 & -1 & 2 & -1 & 8 & 5 & -3 \\ 11 & -11 & -3 & 5 & -8 & -3 & 0 & 0 \\ 4 & -17 & -8 & 12 & -5 & -7 & -5 & 5 \end{bmatrix}$$

Với tỉ lệ lỗi trung bình là 5 giá trị trên 1 điểm ảnh

$$\frac{1}{64} \sum_{x=1}^8 \sum_{y=1}^8 |e(x, y)| = 4.8125$$

Lỗi thấy rõ nhất ở góc dưới bên trái khi điểm ảnh đó tối hơn so với ảnh gốc.

1.3. Giới thiệu chuẩn nén JPEG 2000

JPEG 2000 là 1 chuẩn nén ảnh dựa trên dạng sóng. Nó được tạo ra bởi ủy ban Joint Photographic Experts Group trong năm 2000 với mục đích thay thế chuẩn nén JPEG dựa trên biến đổi DCT (tạo ra vào năm 1992). Chuẩn nén này có phần mở rộng tên file là .jp2 cho những file hợp chuẩn ISO/IEC 15444-1 và .jpx cho phần mở rộng thứ 2, thuộc ISO/IEC 15444-2 khi kiểu MIME là ảnh/jp2.

JPEG 2000 đòi hỏi thời gian giải nén lớn hơn nhiều so với JPEG và cho phép việc tải xuống được cải tiến công phu. JPEG 2000 tăng cường hiệu ứng mờ với tỉ lệ nén cao hơn việc tạo “blocking và ringing” của JPEG, phức tạp việc nén trực tiếp của những tỉ lệ nén riêng của chúng. Những ảnh nén bằng chuẩn JPEG 2000 có chất lượng tốt hơn (với con người) ở tỉ lệ bit thấp.

JPEG 2000 được phổ biến là chuẩn ISO/IEC 15444-1:2000. Hiện nay, JPEG 2000 không được hỗ trợ rộng rãi bởi các trình duyệt và vì thế không dùng trên WWW. Dù cho những nhà cung cấp đã tích hợp mọi kiểu siêu dữ liệu bên ngoài, dữ liệu Exif vẫn không được nhúng.

1.4. Kỹ thuật sơ bộ

Mục đích của JPEG 2000 không chỉ là tăng cường hiệu năng nén mà còn thêm vào những tính năng như khả năng sửa đổi và mở rộng. Thực tế là JPEG 2000 tăng cường hiệu năng nén hơn so với chuẩn JPEG. Tỉ lệ nén rất cao và rất nhỏ được JPEG 2000 hỗ trợ. Các bước nén ảnh JPEG 2000 bao gồm.

Chuyển đổi thành phân màu

Ban đầu ảnh là dạng RGB được chuyển sang không gian ảnh khác, trái với JPEG, JPEG 2000 có 2 cách lựa chọn :

Chuyển đổi sang không gian ảnh YcbCr dùng trong JPEG, đây là chuyển đổi không thể đảo ngược vì lỗi do quá trình lượng tử hóa.

Sử dụng không gian ảnh YUV có thể đảo ngược vì không bị lỗi khi lượng tử hóa, cách chuyển đổi là :

$$\begin{bmatrix} Y_r \\ U_r \\ V_r \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}; \quad \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & \frac{3}{4} & -\frac{1}{4} \\ 1 & -\frac{1}{4} & -\frac{1}{4} \\ 1 & -\frac{1}{4} & \frac{3}{4} \end{bmatrix} \begin{bmatrix} Y_r \\ U_r \\ V_r \end{bmatrix}$$

Thành phần màu có thể , không phải là bắt buộc, giảm tỉ lệ độ phân giải. Trong trường hợp này, dù chuyển đổi sóng chia những ảnh thành những tỉ lệ, downsampling được giải quyết hiệu quả hơn bằng cách giảm tỉ lệ wavelet tốt nhất. Bước này được gọi chuyển đổi thành phần bội trong ngôn ngữ JPEG 2000.

Tiling

Sau khi chuyển đổi màu, ảnh được chia thành những phần được gọi là tile., những phần hình chữ nhật của ảnh được chuyển đổi và mã hóa riêng biệt. Những tile có kích thước bất kỳ, nó cũng có thể là toàn bộ ảnh như là 1 tile duy nhất. Khi 1 kích thước được chọn, tất cả tile có cùng kích thước (trừ những thành phần ở biên). Chia ảnh thành các tile là ý tưởng để bộ giải mã cần ít bộ nhớ hơn để giải mã ảnh và có thể giải mã chỉ 1 phần của bức ảnh, bỏ qua những phần khác. Nhược điểm của việc chia nhỏ là chất lượng

của ảnh sẽ giảm xuống. Sử dụng nhiều tile có thể tạo ra 1 hiệu ứng khối tương tự như chuẩn nén JPEG.

Chuyển đổi dạng sóng

Những tile này sau đó chuyển đổi sóng với độ lớn thay đổi, khác với ảnh JPEG dùng những khối kích thước 8×8 chuyển đổi DCT. JPEG 2000 dùng 2 chuyển đổi sóng riêng biệt :

- *không thể đảo ngược* : dạng chuyển đổi sóng CDF 9/7. kiểu chuyển đổi này là “không thể đảo ngược vì việc tập nhieu của lượng tử hóa phụ thuộc vào độ chính xác của bộ giải mã.
- *khả đảo ngược*: 1 phiên bản làm tròn của chuyển đổi sóng CDF 5/3. Nó chỉ dùng những hệ số nguyên, vì thế đầu ra không đòi hỏi làm tròn và do đó không gây ra nhiều lượng tử hóa. Cách này được dùng trong mã hóa không mất mát thông tin.

Lượng tử hóa

Sau khi chuyển đổi dạng sóng, những hệ số được tỉ lệ lượng tử hóa để giảm lượng bit biểu diễn chúng, đồng nghĩa với việc giảm chất lượng. Đầu ra là 1 tập các số nguyên cần được mã hóa bit-bit. Tham số có thể thay đổi để đặt chất lượng cuối cùng là bước lượng tử hóa : bước càng lớn, độ nén càng cao và chất lượng nén càng thấp. Với bước lượng tử hóa là 1 thì bước lượng tử hóa ko được thực hiện (dùng cho ko mất mát thông tin)

2. Nén Video

Quá trình nén video theo chuẩn MPEG-1

Các bác phệ cho em cái ảnh ở xờ nai 65 chương 3 vào đây

2.1. Sơ lược quá trình nén

- Quá trình nén frame I :
 - Áp dụng phép biến đổi DCT lên các khối 8×8 biểu diễn độ chói và sắc độ của frame.
 - Đầu ra của DCT (64 hệ số) được đưa qua bộ lượng tử hóa Q. Quá trình lượng tử hóa được điều khiển bởi hai tham số : bảng lượng tử hóa và các hệ số lượng tử.
 - Dòng bit sau khi lượng tử hóa được mã hóa entropy và đưa vào bộ đệm
- Mã hóa nén các frame P,B:
 - Frame tham chiếu cho quá trình nén frame P là frame I hoặc frame P trước đó.
 - 2 frame tham chiếu cho quá trình nén frame B là frame I trước đó và frame P sau đó.
 - Tính toán vec – tơ chuyển động bằng cách tìm kiếm một khối phù hợp với khối của frame hiện tại trong cửa sổ frame tham chiếu. Vec – tơ chuyển động sẽ được xử lý trên các khối DCT, Q ,mã hóa entropy và đưa vào bộ đệm
 - Tính sai số dự đoán (frame dự đoán bù chuyển động) bằng cách trừ từng điểm ảnh của Macroblock trên frame đang xử lý cho bản sao của nó trên frame tham chiếu.

- Tính được sự sai khác giữa frame hiện tại với frame dự đoán bù chuyển động. Kết quả được xử lý qua các khối DCT, Q, mã hóa entropy trước khi đưa vào bộ đệm
- Nội dung dữ liệu được nén cho chuỗi ảnh video $\{I(i,j,n)\}$:
 - Dữ liệu ảnh frame I
 - Các vec – tơ chuyển động đã ước lượng cho từng frame
 - Dữ liệu sai khác giữa frame dự đoán bù chuyển động so với frame ở thời điểm hiện tại.
- Bộ đệm và khối điều chỉnh đóng vai trò giữ tốc độ dòng bit là không đổi
- Chú ý, các frame sẽ được chuyển từ hệ màu RGB sang hệ YUV ($Y C_B C_R$)

2.2. Các khối chức năng

2.2.1. Biến đổi Cosine rời rạc (DCT)

- DCT hai chiều (có thể không cần ghi công thức)

$$F(u, v) = 1/4 C(u) C(v) \sum_{x=0}^7 \sum_{y=0}^7 f(x, y) \cos((2x+1)u\pi/16) \cos((2y+1)v\pi/16)$$

$$u, v, x, y = 0, 1, 2, \dots, 7$$

$$C(j) = 1/\sqrt{2}, j = 0$$

$$C(j) = 1, j > 0$$

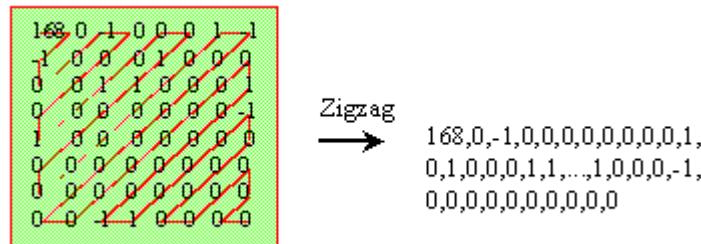
- DCT⁻¹ $f(x, y) = 1/4 \sum_{u=0}^7 \sum_{v=0}^7 C(u) C(v) F(u, v) \cos((2x+1)u\pi/16) \cos((2y+1)v\pi/16)$
- Đặc điểm:
 - Biến đổi từ miền không gian sang miền tần số
 - Là phép biến đổi giải tương quan (decorrelation) : loại bỏ các thành phần dư thừa ở các điểm ảnh lân cận → tối thiểu hóa biên độ dao động của các hàm tự tương quan.
 - Tập trung năng lượng ở miền tần số thấp (góc trên bên trái của ảnh) → khi lượng tử hóa cho phép đưa các giá trị đó về 0 mà không làm biến dạng ảnh
 - Là phép biến đổi trực giao và đối xứng → giảm chi phí tính toán bằng cách thay thế bởi phép biến đổi Cosine nhanh FCT
- Tối ưu bằng cách áp dụng lên các khối 8x8 vì:
 - Cho kết quả tốt nhất khi áp dụng FCT.
 - Tăng độ chính xác khi làm việc với dấu phẩy tĩnh, giảm sai số làm tròn
 - Các điểm ảnh lân cận có độ tương quan cao hơn, khi áp dụng DCT cho từng khối nhỏ sẽ tập trung năng lượng vào một số ít các hệ số biến đổi → chỉ gây ra mất mát thông tin cục bộ
- Vai trò trong sơ đồ : giảm độ bất định (entropy) của ảnh, tăng xác suất xuất hiện của một số giá trị nhất định → tăng tỉ số nén.

2.2.2. Lượng tử hóa

- Khái niệm : ánh xạ không tuyến tính có tồn hao từ khoảng biểu diễn liên tục biên độ tín hiệu vào khoảng biểu diễn bởi các giá trị rời rạc của các mức giá trị hay từ mã.
- Sử dụng bảng lượng tử $Q(m \times n)$ để lượng tử hóa:
Lượng tử hóa Q : $X_q[i,j] = \text{Round}(X[i,j]/Q[i,j])$
Lượng tử hóa ngược Q^{-1} : $X'[i,j] = X_q[i,j] * Q[i,j]$
Vấn đề : thiết kế bảng lượng tử để sai số lượng tử là nhỏ nhất, không làm biến dạng ảnh.
- Vai trò trong sơ đồ:
 - Do các giá trị sau phép biến đổi là số thực nên sử dụng bộ lượng tử hóa để chuyển về số nguyên \rightarrow biểu diễn đơn giản với số bit ít hơn
 - Thu hẹp dải giá trị, giảm số bit biểu diễn cho mỗi giá trị.

2.2.3. Mã hóa entropy (Entropy coding)

- Đặc điểm:
 - Mỗi mặt phẳng bit được mã hóa riêng bằng mã hóa số học thích nghi.
 - Mã hóa mỗi mặt phẳng bit được chia thành các khối
 - Mỗi mặt phẳng bit của mỗi khối mã hóa dựa theo nguyên tắc lan truyền bit
- Các bước thực hiện :
 - Tiền xử lý : sắp xếp lại theo chuỗi Zigzag để các giá trị 0 nằm cạnh nhau



- Mã hóa **DPCM** (Difference Pulse Coded Modulation) cho thành phần hệ số một chiều $C(0,0)$
- Mã hóa loạt dài (RLE) cho phần còn lại của chuỗi (các hệ số xoay chiều)
- Áp dụng các thuật toán mã hóa với độ dài thay đổi VLC (Huffman , Shannon, Shannon – Fano)

2.2.4. Ước lượng và bù chuyển động

- Các kỹ thuật dự đoán :
 - Dự đoán tiến nếu dự đoán chuyển động dựa trên ảnh đã được mã hóa
 - Dự đoán tiến nếu dự đoán chuyển động dựa trên ảnh sắp được mã hóa
 - Dự đoán 2 chiều (nội suy ảnh) nếu dự đoán chuyển động dựa trên cả dự đoán tiến và dự đoán lùi.
- Phương pháp dự đoán chuyển động gồm hai giai đoạn:
 - Đánh giá chuyển động để xác định vec – tơ chuyển động cho mỗi khối
 - Bù chuyển động để xác định ảnh bù sai số dự đoán

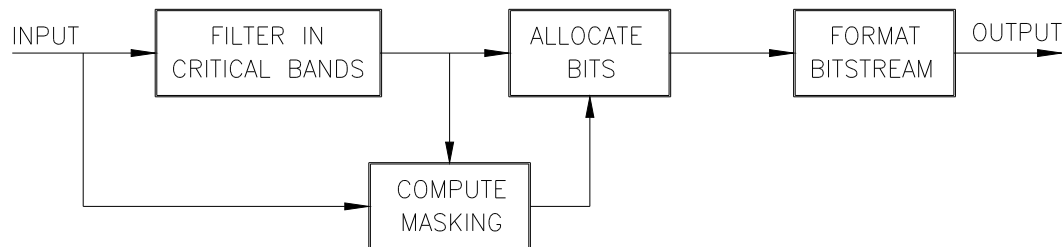
- Sử dụng phương pháp đối sánh khối để xác định vec – tơ chuyển động. Vec – tơ chuyển động chính là vec – tơ có tọa độ tính từ tâm khối hiện tại đến tâm khối tham chiếu.
- Thay vì phải mã hóa ảnh gốc với tỉ lệ lớn bit chuẩn MPEG mã hóa vec – tơ chuyển động và ảnh sai số dự đoán (ảnh bù chuyển động).

3. Nén audio theo chuẩn MPEG

3.1. Nén âm thanh MPEG.

Vài thông số:

- MPEG-1 : 1.5Mbits/s cho âm thanh và hình ảnh.
Khoảng 1.2 Mbits cho hình ảnh và 0.3Mbits/s cho âm thanh.
- Am thanh CD không nén dùng: $(44,100 \text{ mẫu/s} * 16\text{bit/mẫu} * 2 \text{ kênh}) > 1.4 \text{ Mbits/s}$
- Am thanh MPEG cung cấp các tần số lấy mẫu là 32, 44.1 và 48 kHz.
 - Giải thuật:
 1. Dùng bộ lọc thông để chia tín hiệu âm thanh thành các sub-band theo tần số, tương ứng với 32 băng giới hạn → lọc sub-band.
 2. Xác định số lượng che của mỗi band gây bởi các band lân cận bằng các kết quả bước 1 → mô hình âm - tâm lý.
 3. Nếu mức to của một băng mà nhỏ hơn ngưỡng che thì không mã hóa nó.
 4. Ngược lại, xác định số bit cần thiết để mã hóa sao cho nhiễu sinh ra bởi việc lượng tử hóa này thấp hơn đường cong che.
 - 5. Định dạng dòng dữ liệu bit :



Hình 5.4

Ví dụ:

Sau khi phân tích, 16 band đầu tiên trong số 32 band như sau:

Band	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Level(dB)	0	8	12	10	6	2	10	60	35	20	15	2	3	5	3	1

- Nếu mức to của âm thứ 8 là 60dB, nó sẽ che band thứ 7 ở mức 12dB và band thứ 9 ở mức 15dB.
Mức to ở band 7 là 10dB (<12dB) nên bị bỏ qua, không mã hóa.
Mức to band 9 là 35 (> 15dB) nên được tiếp tục xử lý.
- Layer I: bộ lọc loại DCT với 1 frame và độ rộng tần số như nhau trên mỗi sub-band. Mô hình âm-tâm lý chỉ sử dụng hiệu quả che tần số (Frequency masking).
- Layer II: sử dụng 3 frame trong bộ lọc (trước, hiện tại và kế tiếp, tổng cộng 1152 mẫu). Mô hình âm-tâm lý có sử dụng hiệu quả che nhất thời (Temporal masking).
- Layer III: dùng bộ lọc băng giới hạn tốt hơn, mô hình âm-tâm lý có sử dụng hiệu quả che nhất thời, và có dùng bộ mã hoá Huffman.

3.2. Các thông số dung trong MPEG.

Chuẩn MPEG cho phép ta chọn lựa các thông số cho việc nén âm thanh tốt nhất phù hợp với ứng dụng mà ta sử dụng. Lược đồ mã hóa cho các loại là tổng quát. Các thông số có thể chọn lựa trong bộ mã hóa MPEG bao gồm: Mode, Sampling frequency, bitrate, và Layer.

a. Mode.

Chuẩn MPEG có 4 chế độ:

- Mono.
- Dual channel.
- Stereo.
- Intensity Stereo (còn gọi là Joint Stereo).
tốc độ truyền bit thấp.

b. Sampling Frequency (tốc độ lấy mẫu).

Một số tốc độ lấy mẫu:

32kHz, 44.1kHz và 48kHz đối với MPEG1 (Tiêu chuẩn ISO/IEC 11172-3).

16kHz, 22.05kHz và 24kHz đối với MPEG2 (Tiêu chuẩn ISO/IEC 13818-3).

Khi chọn lựa tốc độ lấy mẫu cần xem xét các vấn đề:

- Tần số lấy mẫu càng lớn thì chất lượng âm thanh càng cao (độ dài frame nhỏ hơn).
- Băng thông tín hiệu giới hạn ở mức 15 kHz khi lấy mẫu ở tốc độ 32 kHz và 8 kHz ở tốc độ 16 kHz.
- Tần số lấy mẫu (kHz) và tốc độ của âm thanh mã hóa (kbps) có thể chọn độc lập.
- Tần số lấy mẫu 44.1 kHz hay 22.05 kHz là không thiết thực cho việc chọn lọc vì độ dài frame (byte) là thay đổi.
- Những file được lấy mẫu ở những tần số khác nhau thì rất khó khăn khi hòa trộn.
- Khi dùng đường nhập số AES/EBU, tần số lấy mẫu bị cố định bởi tín hiệu nhập. Nếu không bắt buộc, Digigram yêu cầu lấy mẫu ở 48 kHz hoặc 44.1 kHz cho phát thanh hay ứng dụng multimedia. Nếu ta phải sử dụng tốc độ bit thấp cho sự truyền có hiệu quả, tốc độ 24 kHz là thích hợp.

c. Bit Rate.

Mỗi Layer và chế độ có nhiều cách chọn lựa tốc độ bit (bit rate). Việc chọn tốc độ bit tùy thuộc trước tiên vào chất lượng âm yêu cầu. Băng thông tín hiệu là hẹp hơn nếu tốc độ bit thấp, khiến cho nó không thực tế đối với một số ứng dụng. Tốc độ bit được đo theo kilobits/sec(kbps).

Khi chọn lựa tốc độ bit cần xem xét các vấn đề:

d. Layers.

Chuẩn MPEG có ba layer.

Khi chọn lựa tốc độ bit cần xem xét các vấn đề:

- Ở cùng tốc độ bit, Layer II mang lại chất lượng âm thanh tốt hơn Layer I. Kết luận này là chủ quan, vì sự chênh lệch là rất khó phân biệt ở tốc độ bit 128 kbps và lớn hơn.
- Dùng Layer I thì việc chọn lọc chính xác hơn Layer II bởi vì độ phân giải của Layer I gấp ba lần Layer II.

Resolution Table

Sampling frequency	Layer I (384 samples)	Layer II (1152 samples)
32 kHz	12 ms	36 ms
44.1 kHz	# 8.71 ms	# 26.12 ms
48 kHz	8 ms	24 ms
16 kHz	24 ms	72 ms
22.05 kHz	# 17.42 ms	# 52.24 ms
24 kHz	16 ms	48 ms

Các điểm kỹ thuật mẫu chốt.

Chuẩn mã hóa âm thanh MPEG chỉ định việc ghi một số cố định các mẫu (384 cho Layer I và 1152 cho Layer II) để tạo ra một chuỗi các bytes gọi là “frame”. “Frame” là thực thể nhỏ nhất được điều khiển bởi một ứng dụng. Việc chọn tốc độ bit(kbps) thiết lập nên kích thước của frame đó theo byte.

Ví dụ:

Ở 48 kHz, 128 kbps, chế độ Mono, Layer II:

- 48,000 mẫu, tương ứng 1000 ms hay 1s, 1152 mẫu tương ứng 24 ms.
- Ở tốc độ bit 128,000 bits/s, 3072 bits (384 bytes) cần cho 24 ms.
- Chiều dài frame vì thế là 384 byte.
- Để lưu 1 phút, cần 960,000 byte (hay khoảng 1 Megabyte) đĩa trống.

4. Các phương pháp nén âm thanh khác

4.1. Nén âm thanh dùng mô hình âm - tâm lý.

a. Hệ thống nghe và phát âm của con người.

- Phạm vi nghe được từ 20 Hz đến 20 kHz, nhạy cảm ở 2 - 5kHz.
- Phạm vi phát âm bình thường từ 500 Hz đến 2 kHz.

b. Che tần số (Frequency masking)

“Ngưỡng che” (Threshold masking): sinh ra từ hiệu ứng che, mỗi âm với một tần số và mức to (dB) xác định sẽ có một “ngưỡng che” (xem hình 4.3 và 4.4)

c. Băng giới hạn.

- Thước đo tần số đồng bộ không tương xứng với độ rộng của đường cong che.
- Băng giới hạn có độ rộng là 100Hz đối với các tần số che $< 500\text{Hz}$, và càng tăng lên đối với các tần số $> 500\text{Hz}$.
- Định nghĩa một đơn vị mới cho tần số là bark (Barkhausen)
1 Bark = bề rộng của băng giới hạn:
 - Tần số $< 500\text{Hz}$: $1 \text{ bark} = \text{freq}/100$.
 - Tần số $> 500\text{Hz}$: $1 \text{ bark} = 9 + 4\log(\text{freq}/1000)$.

d. Che nhất thời (Temporal masking): che theo thời gian.

Tai người cũng có đặc tính lưu âm.

Nếu có một âm thanh lớn, rồi ngưng nó lại, mãi một lúc sau ta mới có thể nghe được một âm lân cận nhỏ hơn (xem hình 4.5 và 4.6).

4.2. Hiệu quả che.

Giả sử có một âm mạnh với tần số 1000Hz, và một âm kèm theo có tần số 1100Hz nhưng với cường độ âm nhỏ hơn 18dB. Ta sẽ không thể nghe thấy âm này vì nó đã bị che hoàn toàn bởi âm chủ 1000Hz. Nói một cách khác, một âm thanh yếu gần một âm thanh mạnh sẽ bị che. Hiệu quả che có ý nghĩa rằng ta có thể gia tăng mức ồn nền xung quanh một âm mạnh mà vẫn không nghe được tiếng ồn vì chúng sẽ bị che hoàn toàn. Tăng mức ồn nền còn có nghĩa là dùng ít bit để số hóa. Và điều này cũng giống như là ta đã nén âm thanh vậy.

Bây giờ hãy xem bộ mã hoá âm thanh MPEG hoạt động như thế nào. Bộ mã hóa chia phổ tần số (20Hz đến 20kHz) thành 32 dải nhỏ (sub-band). Mỗi sub-band giữ 1 phần nhỏ của phổ. Trong vùng trên của sub-band 8 ta phát một âm có tần số 1000Hz với mức to 60dB. Bộ mã hóa sẽ tính toán hiệu quả che của âm này và nhận ra rằng có một ngưỡng che cho toàn bộ sub-band thứ 8 (tất cả những âm có cùng tần số). Ngưỡng che này thấp hơn âm phát ra 35 dB. Tỉ số s/n có thể chấp nhận được là $60 - 35 = 25 \text{ dB}$, tương đương với 4 bit. Ngoài ra nó còn ảnh hưởng trên các sub-band 9-13 và 5-7 với hiệu quả che giảm dần từ sub-band 8. Hơn nữa, bộ mã hóa cũng xem xét mức độ nhạy cảm của tai đối với các tần số khác nhau. Tai người ít nhạy cảm với các tần số cao và thấp. Nó nhạy cảm nhất đối với tần số 2-4 kHz, cùng dải tần số với tiếng nói con người.

Các sub-band nên phù hợp với tai người., nghĩa là mỗi sub-band cần có các tần số có cùng các tính chất âm học tâm lý. Trong MPEG layerII, mỗi sub-band có độ rộng 625Hz, do đó cần phải có những bộ lọc băng thông phức tạp. Để các bộ lọc đỡ phức tạp, người ta thêm FFT (Fast Fourier Transform) vào song song với bộ lọc và sử dụng các thành phần phổ từ FFT như là các thông tin thêm vào bộ mã hóa. Bằng cách này ta sẽ lấy mật độ bit cao hơn đối với các tần số thấp mà tai người nhạy cảm hơn.

CHƯƠNG 4

ĐỒNG BỘ VIDEO VÀ AUDIO

1. Thuật toán đồng bộ theo dòng tại điểm tham chiếu.

• Giải pháp:

- Dòng dữ liệu audio có vai trò là chủ (principle jet), dòng Video là dòng phụ thuộc (slave jet) được đồng bộ theo dòng audio.

- Tại các điểm đồng bộ nhận thời gian của gói tin dòng Video được so sánh với nhận thời gian của gói tin dòng audio. Nếu một frame video bị trễ quá giới hạn sẽ bị loại bỏ.

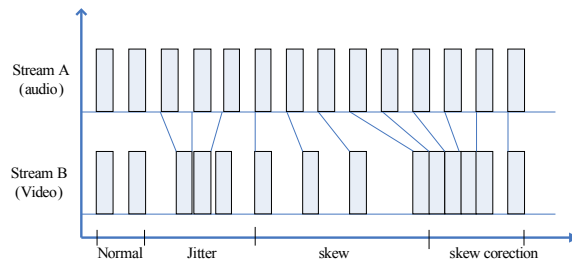
- Cảm thụ độ lệch giữa audio –video: khi tính độ lệch dòng audio–video (skew) thì:

+ Vùng đồng bộ (in sync): độ lệch cho phép từ -80 ms đến +80ms

+ Vùng mất đồng bộ (out of sync): độ lệch nằm ngoài hai giá trị giới hạn âm -160 ms và giới hạn dương +160ms.

+ Vùng trung gian (transient) độ lệch trong khoảng +80ms đến 160ms, và -160 ms đến -80ms.

• Nguyên tắc đồng bộ theo dòng audio:



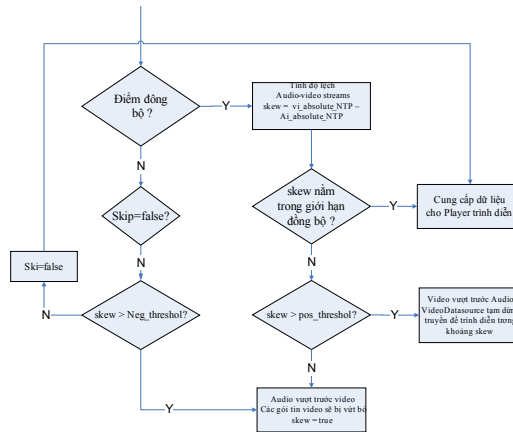
- Độ trễ rung “jitter”: sự khác nhau tức thời về thời gian trễ giữa các dòng video-audio.

- Độ lệch “skew”: độ lệch về thời gian giữa hai dòng audio và video

- Độ trễ điểm đầu cuối “end-to-end delay”

- Tại điểm đồng bộ nếu kết quả tính độ lệch “skew” vượt quá giới hạn thì xử lý loại bỏ gói video.

• Sơ đồ giải thuật đồng bộ theo dòng tại điểm tham chiếu:



2.1. Thuật toán đồng bộ thích nghi

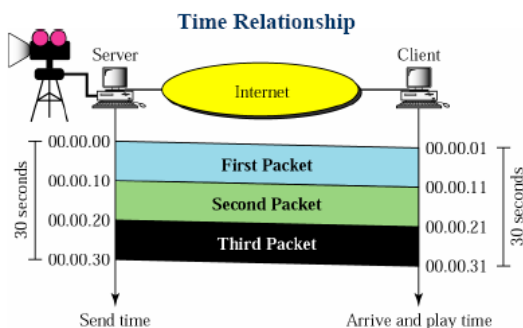
• Nguyên tắc:

Các khối dữ liệu (frame) đối tượng có độ trễ khác nhau được chuẩn hoá thời gian thích nghi khi trình diễn lại sao cho các đối tượng có độ trễ như nhau.

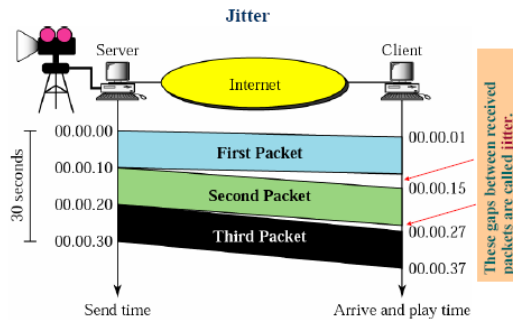
Ví dụ: Với hội nghị video một camera được kết nối với một máy server để truyền các thông tin video đã được quay được tới client. Như vậy thông qua camera và kết nối mạng mọi thông tin diễn ra trong hội nghị tại vị trí đặt server được hiển thị đầy đủ trên máy tính tại vị trí của client.

- Quá trình gửi các gói tin, cùng các liên quan thời gian được mô tả như hình bên dưới:

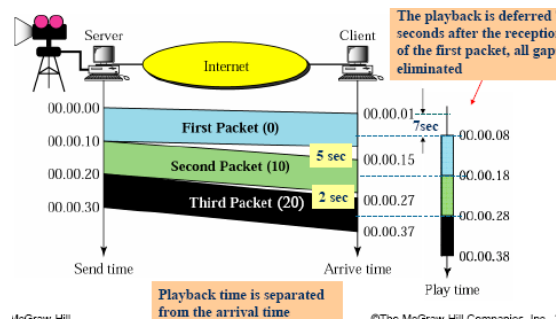
Nếu mỗi gói tin gửi 10 giây, sự trễ trên đường truyền các gói là như nhau và cùng bằng 1 giây thì tại nơi nhận có thể xem Video với tốc độ giống như khi nó được tạo ra.



Tuy nhiên độ trễ của các gói tin thường là không giống nhau:

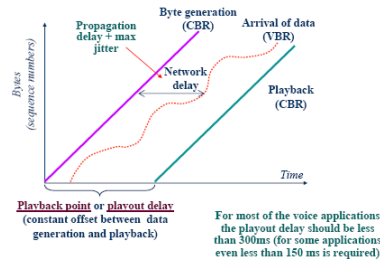


Giải pháp cho vấn đề độ trễ rung (jitter) là sử dụng nhãn thời gian *timestamp*. Mỗi gói có một nhãn thời gian liên quan tới gói đầu tiên (nhãn thời gian được đo tại nơi gửi, nhãn thời gian của gói đầu thường chọn ngẫu nhiên).



Để chia tách thời gian chơi lại (playback) theo thời gian đến ta sử dụng bộ đệm.

Tổng quát ta có sơ đồ mô tả độ trễ truyền khác nhau của các gói tin và thời điểm chuẩn hoá thời gian thích nghi.



Tính toán độ trễ trình diễn (playout delay):

$$D = ED + \beta EV$$

Trong đó:

D – Playout delay (set for each **talk spurt**)
 ED – Estimated average packet delay
 EV – Estimated average packet delay variation
 β – safety factor (usually $\beta = 4$)

Việc tính toán ED và EV được thực hiện cho mỗi gói tin và dựa vào nhãn thời gian:

$$ED_i := \alpha ED_{i-1} + (1-\alpha)(r_i - t_i)$$
$$EV_i := \alpha EV_{i-1} + (1-\alpha)|r_i - t_i - ED_i|$$

Với

α – weighting factor (typically $\alpha = 0.998$)
 r_i – time the packet i is received
 t_i – the timestamp of the i -th packet

3. Thuật toán đồng bộ thích nghi khôi phục đồng hồ thời gian.

Phát triển ý tưởng của thuật toán đồng bộ thích nghi là: các khối dữ liệu có độ trễ khác nhau được chuẩn hoá thời gian thích nghi khi trình diễn lại, sao cho các đối tượng có độ trễ như nhau. Tuy nhiên nhãn thời gian ở đây là sự kết hợp ngẫu nhiên của RTCP với RTP.

- Sau quá trình khởi động, khi nhận gói tin RTCP đầu tiên đồng hồ thời gian là: $t_0 = Ts_NTP$ (thời gian theo giao thức NTP)

- Độ lệch thời gian được tính: $\Delta ts_i = Ts_NTP - Ts_ref\ RTP$

- Cập nhật đồng hồ thời gian : $t_{i+1} = t_i + \text{delta_ts}_i$ với $i = 1, 2, 3 \dots N$

- Khi nhận được gói tin thứ N, đồng hồ được tính:

$$t_N = t_1 + \text{delta_ts}_i$$

- Trong đó: $\text{delta_ts}_i = \Delta ts_i / T_c$ và $S_delta_ts_i = \sum_{j=2}^N \frac{\Delta ts_j}{T_c}$

- Quá trình tiếp tục cho tới khi nhận được gói RTCP mới.

CHƯƠNG 5

TRUYỀN SỐ LIỆU ĐA PHƯƠNG TIỆN

Nội dung:

- Giao thức RTP
- Giao thức RTCP
- Giao thức SIP
- Chuẩn H323

5.1 Các giao thức giao vận

5.1.1. RTP

Các dịch vụ khác nhau đưa ra bởi RTP bao gồm

- Xác minh kiểu payload
- Đánh số trình tự : RTP không đảm bảo phân phát cá gói theo đúng thứ tự vì vậy số thứ tự được sử dụng để sắp xếp các gói theo trình tự tại phí nhận, xác định các gói bị mất
- Dấu bảo đảm thời gian : đảm bảo trình tự diễn ra đúng đắn theo thời gian thực mặc dù nó không cung cấp bất cứ kỹ thuật nào để đảm bảo sự phân phát đúng lúc hoặc cung cấp các sự đảm bảo chất lượng dịch vụ khác (QoS) mà nó nhờ vào các dịch vụ tầng thấp khác đảm bảo chuyển đó
- Giám sát phân phối : quá trình dữ liệu đến được giám sát liên tục bởi RTCP, là giao thức mà báo cho tầng RTP điều chỉnh các thông số truyền và mã hoá của nó nhằm để phân phát dữ liệu đúng đắn.

▪ Sau đây là các đặc trưng của RTP:

- RTP cung cấp các dịch vụ phân phát đầu cuối cho dữ liệu với đặc điểm thời gian thực, như audio và video tương tác.

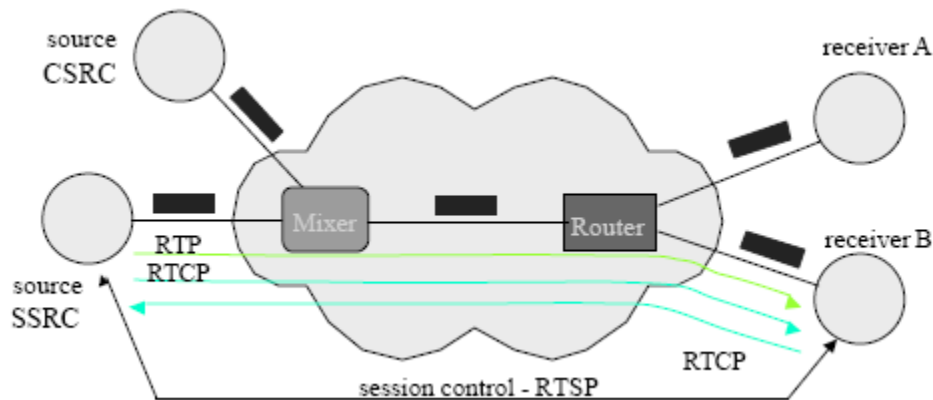
- Các ứng dụng thường chạy RTP ở bên trên UDP để sử dụng các dịch vụ dồn kênh và checksum. Nhưng giao vận RTP có thể độc lập và nó có thể sử dụng trên các giao thức khác.

- RTP tự bản thân không cung cấp các kỹ thuật để đảm bảo phân phát đúng thời gian, kịp thời hoặc cung cấp các đảm bảo chất lượng dịch vụ, nhưng nó nhờ vào các dịch vụ tầng dưới đảm bảo chuyển đó. RTP cho rằng mạng tầng dưới là tin cậy và phân phát các gói theo trình tự.

- RTP không thừa nhận bất cứ gì về mạng tầng dưới, ngoại trừ việc nó cung cấp khung. Mục đích thiết kế ban đầu của nó là Internet nhưng nó được dự định là độc lập giao thức. Ví dụ, việc thử nghiệm quá trình truyền RTP qua ATM AAL5 và IPv6 đang được xúc tiến.

- Trường PT (kiểu payload) của header RTP với 7 bit xác định kiểu nén/mã hoá và kiểu media của payload. Tại bất kỳ thời điểm nào, một bên gửi RTP được hỗ trợ để gửi chỉ duy nhất một kiểu payload đơn, mặc dù trong quá trình truyền kiểu payload có thể thay đổi.

- RTP cung cấp các chức năng phù hợp cho việc mang các nội dung thời gian thực, như là dấu thời gian (timestamp) và các kỹ thuật điều khiển để đồng bộ các dòng khác nhau với các thuộc tính thời gian. Vì RTP/RTCP có nghĩa vụ điều khiển luồng của một dòng media, nó sẽ không tự động đồng bộ các dòng khác nhau. Điều này xảy ra tại tầng ứng dụng.
- Cơ sở của luồng và kiểm soát nghẽn mạng được cung cấp bởi các báo cáo bên gửi và bên nhận RTCP. Chúng ta phân biệt tắc nghẽn đột biến và tắc nghẽn lâu dài. Bằng việc phân tích trường jitter của báo cáo bên gửi, chúng ta có thể đo được jitter vượt quá một khoảng thời gian xen giữa nhất định và chỉ ra nghẽn mạng trước khi nó xảy ra.
- RTP giúp cho xác minh kiểu payload, đánh trình tự các gói tin không theo thứ tự và đảm bảo đúng trình tự diễn ra của media bằng sử dụng dấu thời gian (timestamp).
- RTCP giám sát chất lượng của dịch vụ (QoS) và cũng truyền thông tin về những người tham dự trong một phiên làm việc.
- RTP là độc lập với các giao thức tầng dưới. Nó có thể làm việc trên bất cứ kiểu mạng nào như TCP/IP, ATM, frame relay..., CLNP(Connectionless Network Protocol), IPX (Internetwork Packet Exchange)...
- RTP hỗ trợ truyền dữ liệu tới nhiều đích sử dụng phân phối đa điểm nếu được cung cấp bởi tầng mạng dưới.
- RTP đánh trình tự các số cũng có thể được sử dụng để phát hiện chính xác các vị trí mất gói tin. Ví dụ các số trình tự có thể được sử dụng trong giải mã video, mà không cần phải giải mã các gói theo trình tự.



Hình 1 - Phiên RTP và các giao thức liên quan

5.1.2. RTCP

RTCP là giao thức điều khiển làm việc liên kết với RTP. Nó cung cấp hỗ trợ cho hội thảo thời gian thực cho các nhóm lớn trong internet, bao gồm xác minh nguồn và hỗ trợ cho các gateway và các bộ chuyển đổi multicast-to-unicast (translators). Nó được chuẩn hoá trong RFC 1889, RFC 1890. RTCP thực thi 4 chức năng sau.

- **Cung cấp thông tin cho ứng dụng:** Chức năng chính là cung cấp thông tin cho ứng dụng về chất lượng của việc phân phát dữ liệu. Mỗi gói RTCP chứa các báo cáo SR (sender report) và/hoặc RR (receiver report) báo cáo các số liệu thống kê hữu ích với ứng dụng. Các số liệu thống kê này bao gồm số gói đã được gửi, số gói mất, jitter,...

- **Nhận dạng nguồn RTP:** RTCP chứa một từ định danh tầng giao vận cho một nguồn RTP, gọi là CNAME. CNAME này được sử dụng để theo dõi những người tham gia trong một phiên làm việc RTP. Bên nhận sử dụng CNAME để liên kết nhiều dòng dữ liệu từ một người tham gia trong tập các phiên làm việc RTP có liên quan... ví dụ để đồng bộ video và audio.

- **Điều khiển khoảng thời gian truyền xen kẽ RCTP:** Để ngăn chặn lưu lượng điều khiển tràn ngập tài nguyên mạng và cho phép RTP có thể tăng số lượng người tham gia trong phiên làm việc, lưu lượng điều khiển được giới hạn đến 5% của toàn bộ lưu lượng phiên. Giới hạn này được đảm bảo bởi việc điều chỉnh tốc độ mà tại đó các gói tin RTCP được truyền như là một hàm của số những người tham gia hội thảo.

- **Vận chuyển thông tin điều khiển phiên làm việc tối thiểu:** Như một chức năng tùy chọn, RTCP có thể được sử dụng như một phương pháp thuận lợi để vận chuyển một lượng tối thiểu của thông tin tới mọi thành viên tham gia phiên làm việc..

5.2. Giao thức SIP

SIP cung cấp những khả năng sau:

- ✓ Xác định vị trí của điểm đích: SIP hỗ trợ phân giải địa chỉ, ánh xạ tên và điều hướng cuộc gọi.
- ✓ Xác định khả năng hỗ trợ đường truyền của điểm đích. Thông qua SDP (Session Description Protocol) SIP xác định mức thấp nhất của dịch vụ chung giữa 2 điểm cuối.
- ✓ Xác định tính sẵn sàng của điểm đích. Nếu một cuộc gọi không được hoàn thành do điểm đích không sẵn sàng, SIP sẽ xác định nguyên nhân và gửi lại 1 tin nhắn chỉ định tại sao điểm đích lại chưa sẵn sàng.
- ✓ Thiết lập 1 phiên giữa điểm nguồn và điểm đích nếu cuộc gọi được hoàn thành. SIP cũng hỗ trợ việc cho phép thay đổi giữa cuộc gọi, như là thêm một điểm cuối tham gia hội nghị, hoặc thay đổi những đặc điểm trên đường truyền, hoặc mã hóa...
- ✓ Điều khiển việc truyền và kết thúc cuộc gọi. SIP hỗ trợ truyền cuộc gọi từ điểm cuối này tới điểm cuối khác.

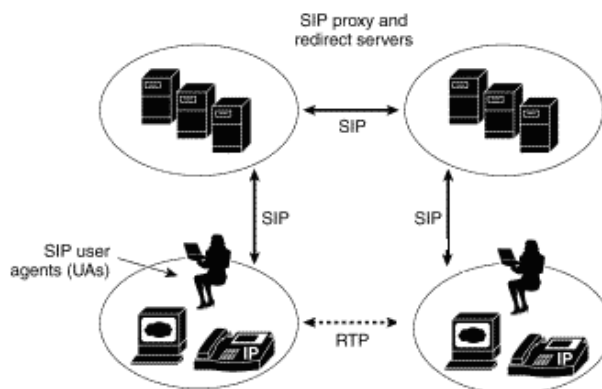
Thành phần của SIP

Kiến trúc của hệ thống SIP gồm 2 thành phần cơ bản : User Agent (UA) và Network Server (NA), được thể hiện rõ trong hình sau

1 User Agent có thể đóng 1 trong 2 hoặc cả 2 vai trò UAS(User Agent Server) và UAC (User Agent Client)

- ✓ UAC là ứng dụng khởi tạo yêu cầu SIP.
- ✓ UAS là ứng dụng liên hệ với người sử dụng khi 1 yêu cầu SIP được nhận và đáp trả với tư cách là người sử dụng.

User Agent thực hiện chức năng là UAS hay UAC phụ thuộc vào User Agent nào khởi tạo yêu cầu.



NS (Network Server) bao gồm

- ✓ PS (Proxy Server) :là thiết bị trung gian mà nhận yêu cầu SIP từ 1 client và chuyển tiếp yêu cầu với tư cách của người sử dụng..PS có thể cung cấp thêm dịch vụ xác thực(Authentication) , xác nhận quyền(authorization), điều khiển truy nhập mạng, định tuyến, truyền lại yêu cầu tin cậy, và an toàn(security).
- ✓ RS(Redirect Server) : cung cấp cho client thông tin về đích tiếp theo mà bản tin sẽ phải đến và sau đó client liên lạc với đích mới này, có thể hiểu đây chính là server đổi hướng truyền của bản tin.
- ✓ LS(Location Server):định vị , cung cấp cách xác định địa chỉ miền và người dùng.
- ✓ RS(Registrar server) giải quyết yêu cầu của UAC về việc đăng kí vị trí hiện thời của họ.Thường được đặt cùng với RS hoặc PS.

Cách thức SIP làm việc

SIP là một giao thức đơn giản mà sử dụng cách thức yêu cầu /đáp trả để thiết lập phiên truyền thông giữa nhiều thành phần trong mạng.

Một người sử dụng trong hệ thống SIP được xác định bởi 1 địa chỉ SIP duy nhất , có định dạng userID@domain.Người sử dụng đăng kí với RS sử dụng địa chỉ SIP đã được gán.RS cung cấp thông tin này cho LS khi có yêu cầu.

Khi một người sử dụng khởi tạo 1 cuộc gọi ,1 yêu cầu SIP được gửi tới SIP server (có thể là PS hoặc RS).Yêu cầu chưa địa chỉ của người gọi và người nhận.Ví trí của 1 người sử dụng có thể đăng kí động với SIP server.LS có thể sử dụng 1 hoặc nhiều giao thức (như là finger, rwhois, LDAP)để định vị người sử dụng.Bởi vì người sử dụng có thể đăng nhập vào nhiều hơn 1 máy tại cùng 1 thời điểm và bởi vì LS có thể có thông tin không chính xác , nó có thể trả về nhiều hơn 1 địa chỉ cho 1 người sử dụng.Trong trường hợp này PS sẽ cố gắng thử với mỗi địa chỉ được trả về cho đến khi định vị được người sử dụng.

Thiết lập cuộc gọi giữa 2 máy đầu cuối

Thiết lập cuộc gọi qua PS

Người gửi gửi yêu cầu INVITE tới PS, PS xác định đường dẫn và sau đó chuyển tiếp yêu cầu tới người nhận. Người nhận đáp trả lại PS, cái này sau đó lại được chuyển tiếp tới người gửi. PS chuyển tiếp bản tin báo nhận của cả 2 bên. Một phiên được thiết lập giữa người gửi và người nhận. RTP được sử dụng để truyền thông giữa người gửi và người nhận.

Nếu thiết lập cuộc gọi sử dụng RS

Người gửi gửi yêu cầu INVITE tới RS. RS liên hệ với LS để xác định đường dẫn tới người nhận, và RS gửi thông tin trả lại người gửi. Người gửi sau đó gửi bản tin báo nhận cho RS. Người gửi sau đó gửi yêu cầu INVITE tới thiết bị được chỉ định bởi RS (có thể là người nhận, cũng có thể là 1 server khác sẽ chuyển tiếp yêu cầu). Ngay khi yêu cầu tới đích, người nhận gửi đáp trả và người gửi gửi xác nhận RTP được sử dụng để truyền thông giữa người gửi và người nhận.

5.3. H323

1. Giới thiệu chuẩn H323 và các ưu điểm chuẩn H323

H323 là một chuẩn quốc tế về hội thoại trên mạng được đưa ra bởi hiệp hội viễn thông quốc tế ITU (International Telecommunication Union). Chuẩn H323 của ITU xác định các thành phần, các giao thức, các thủ tục cho phép cung cấp dịch vụ truyền dữ liệu đa phương tiện (multimedia) audio, video, data thời gian thực qua mạng chuyển mạch gói (bao gồm cả mạng IP) mà không quan tâm đến chất lượng dịch vụ. Đến nay, H323 đã phát triển thông qua hai phiên bản. Phiên bản thứ nhất được thông qua vào năm 1996 và phiên bản thứ hai được thông qua vào năm 1998. Ứng dụng vào chuẩn này rất rộng bao gồm cả các thiết bị hoạt động độc lập cũng như ứng dụng truyền thông nhúng trong môi trường máy tính cá nhân, có thể áp dụng cho đàm thoại điểm - điểm cũng như cho truyền thông hội nghị. H323 còn bao gồm cả chức năng điều khiển cuộc gọi, quản lý thông tin đa phương tiện và quản lý băng thông và đồng thời còn cung cấp giao diện giữa mạng LAN và các mạng khác.

* *Các ưu điểm của chuẩn H323:*

+ Cung cấp các bộ mã hoá đã được chuẩn hoá: H.323 thiết lập các chuẩn nén và giải nén cho các luồng dữ liệu audio và video, bảo đảm cho các thiết bị từ các nhà cung cấp khác nhau có sự hỗ trợ chung.

+ Tính tương thích cao: Người sử dụng có thể trao đổi dữ liệu mà không phải lo lắng về tính tương thích ở bên nhận.

+ Độc lập hệ thống mạng: H.323 được thiết kế để chạy ở tầng trên của kiến trúc mạng. Những giải pháp cơ bản của H.323 cho phép tận dụng được những cải tiến về kỹ thuật mạng và sự phát triển băng thông.

+ Độc lập với ứng dụng và hệ điều hành: H.323 không bị ràng buộc với phần cứng hay hệ điều hành.

+ Hỗ trợ đa điểm: Tuy H.323 có thể quản lý được những cuộc hội nghị có nhiều kết nối mà không cần sử dụng thêm một trình điều khiển đa điểm chuyên dụng nào, nhưng việc sử dụng MCU (Multipoint Control Unit – trình điều khiển đa điểm) sẽ cung cấp một

kiến trúc mạnh và linh hoạt hơn cho hội nghị kiểu nhiều kết nối.

+Quản lý băng thông: Việc truyền các dữ liệu truyền thông đa phương tiện đòi hỏi băng thông rất lớn và có thể làm nghẽn mạch. Để giải quyết vấn đề này, H.323 đưa ra trình quản lý băng thông, đảm bảo cho sự lưu thông trên mạng không bị tắc nghẽn.

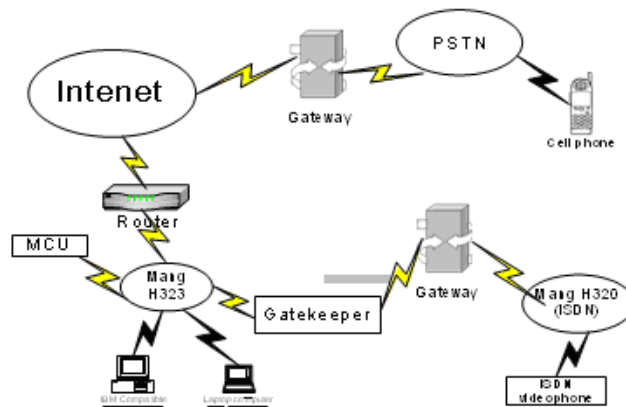
+ Hỗ trợ khả năng quản bá thông tin: Giúp cho việc sử dụng băng thông hiệu quả hơn.

+ Linh hoạt: Một hội nghị sử dụng chuẩn H.323 có khả năng tiếp nhận các thiết bị đầu cuối khác nhau.

+Khả năng hội nghị liên mạng: Nhiều người dùng muốn kết nối từ mạng LAN đến một đầu xa chẳng hạn như kết nối giữa hệ thống LAN với hệ thống ISDN. H.323 cũng hỗ trợ khả năng này và sử dụng kỹ thuật mã hoá chung từ các chuẩn hội nghị khác nhau để giảm thiểu thời gian chuyển đổi mã và tạo một hiệu suất tối ưu cho hội nghị.

2.Quy định cấu trúc các thành phần và các thiết bị vật lý trong chuẩn H323

Chuẩn H323 là chuẩn khuyến nghị quy định các thành phần chính như sau với 4 thành phần: Terminal, Gateway, Gatekeeper, MCU với cấu trúc như sau:



Hình 1.1 Mô hình mạng theo chuẩn H323

Trong đó có đưa ra các thành phần phân cứng tương ứng với mỗi thành phần như sau:

+Terminal:

+Gateway:

+Gatekeeper:

+MCU:

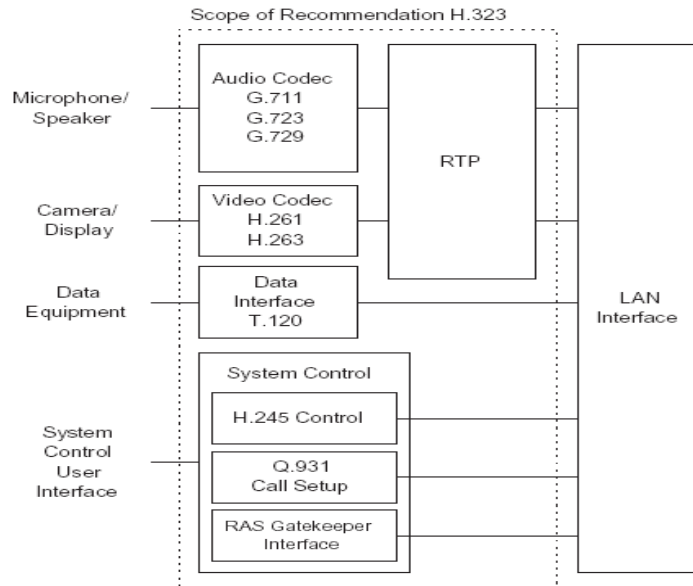
Sau đây là đặc điểm kiến trúc yêu cầu của từng thành phần:

2.1. Terminal:

H323 Terminal là một thiết bị đầu cuối trong mạng LAN có khả năng truyền thông hai chiều theo thời gian thực. Nó có thể là một máy PC hoặc một thiết bị độc lập. Tất cả các đầu cuối H323 đều phải được hỗ trợ khả năng truyền dữ liệu audio hai chiều, còn dữ liệu và video là lựa chọn. H323 chỉ ra những cách thức cho những hoạt động mà cần audio, video, dữ liệu làm việc chung với nhau được. Nó mở ra một thế hệ mới cho sử dụng điện thoại internet, hội nghị truyền thông. Các thiết bị đầu cuối H323 phải hỗ trợ chuẩn H245 được dùng để điều tiết các kênh truyền dữ liệu, và khả năng của thiết bị. Ngoài ra nó phải được hỗ trợ các thành phần sau:

+Giao thức báo hiệu H225 phục vụ trong quá trình thiết lập và hủy bỏ cuộc gọi.

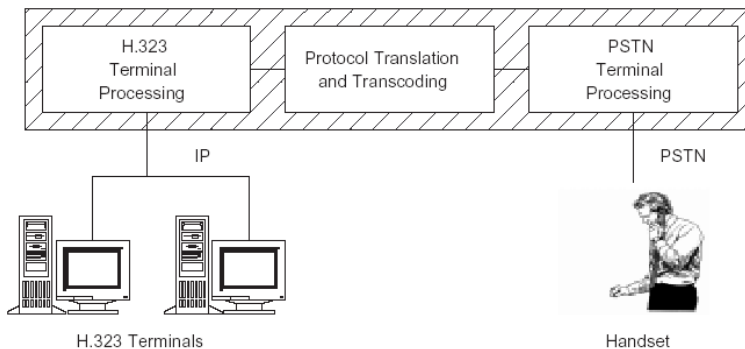
- +Giao thức H225 RAS (Registration/Admission/Status) thực hiện các chức năng đăng kí, thu nhận... với gatekeeper.
- +Giao thức Q.931 dùng cho báo hiệu và thiết lập cuộc gọi.
- +Giao thức RTP/RCTP để truyền và kết hợp các gói tin audio,video... Một đầu cuối H323 cũng có thể được trang bị thêm các tính năng như:
 - +Mã hoá và giải mã các tín hiệu audio, video.
 - +Hỗ trợ giao thức T120 phục vụ cho việc trao đổi thông tin số liệu (data).
 - +Tương thích với MCU để hỗ trợ các liên kết đa điểm



Hình 1.2: Các thành phần Terminal

2.2 Gateway:

Nhiệm vụ của gateway là thực hiện việc kết nối giữa 2 mạng khác nhau. H323 gateway cung cấp khả năng kết nối giữa 1 mạng H323 và một mạng khác (không phải H323). Việc kết nối này được thực hiện nhờ chức năng chuyển đổi giao thức trong quá trình thiết lập, giải phóng cuộc gọi và chức năng biến đổi khuôn dạng dữ liệu giữa hai mạng khác nhau của gateway. Như vậy đối với kết nối giữa hai thiết bị đầu cuối H323 thì không cần thiết phải có gateway, nhưng đối với cuộc gọi có sự tham gia của mạng chuyển mạch kênh thì gateway là bắt buộc phải có.

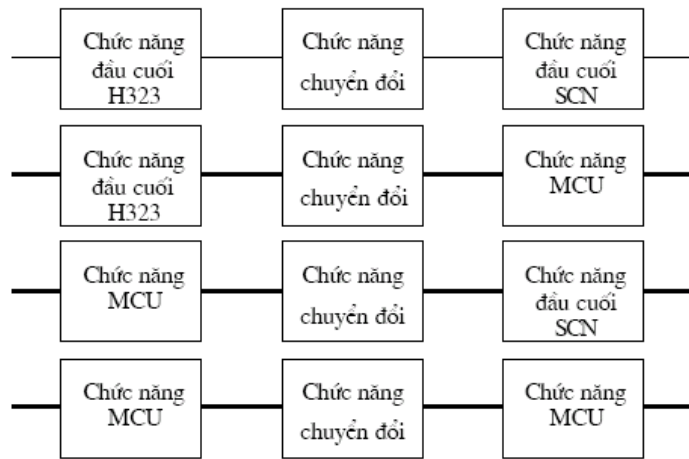


Hình 1.3: Thành phần Gateway

Gateway là một thành phần tùy chọn trong hội nghị H.323, thường là các máy tính có nhiều giao diện với các mạng khác nhau. Gateway cung cấp nhiều dịch vụ, tổng

quát nhất là chức năng biên dịch giữa các đầu cuối H.323 và các loại đầu cuối khác. Bằng những bộ chuyển mã thích hợp, Gateway H.323 có thể hỗ trợ những thiết bị đầu cuối tuân theo các chuẩn H.310, H.321, H.322 và V.70. Chức năng này bao gồm biên dịch giữa những khuôn dạng truyền (H.225.0 đến H.221) và giữa những thủ tục truyền thông (H.245 sang H.242). Ngoài ra, Gateway cũng biên dịch giữa các bộ mã hoá âm thanh và hình ảnh, thực hiện thiết lập và kết thúc cuộc gọi trên cả đầu mạng LAN và đầu mạng chuyển mạch điện tử SCN.

Gateway khi hoạt động sẽ có đặc điểm của một thiết bị đầu cuối H.323 hoặc một MCU trong mạng LAN và có đặc điểm của một thiết bị đầu cuối trong SCN hoặc một MCU trong SCN. Vì vậy ta có 4 cấu hình cơ sở của gateway. Mỗi gateway có thể có tổ hợp của các cấu hình cơ sở hoặc có thể gồm cả 4 cấu hình này.



Hình 1.4: Nội dung của Gateway

2.3 Gatekeeper:

Gatekeeper là một thành phần quan trọng trong mạng H323, nó được xem như bộ não của mạng.

Các chức năng bắt buộc của gatekeeper :

+Chức năng dịch địa chỉ: Gatekeeper sẽ thực hiện việc chuyển đổi từ một địa chỉ hình thức (dạng tên gọi) của các thiết bị đầu cuối và gateway sang địa chỉ truyền dẫn thực trong mạng (địa chỉ IP). Chuyển đổi này dựa trên bảng đối chiếu địa chỉ được cập nhật thường xuyên bằng bản tin đăng ký dịch vụ của các đầu cuối.

+Điều khiển truy nhập: Gatekeeper sẽ chấp nhận một truy nhập mạng LAN bằng cách sử dụng các bản tin H.225.0 là ARQ/ACF/ARJ . Việc điều khiển này dựa trên độ rộng băng tần và đăng ký dịch vụ hoặc các thông số khác do nhà sản xuất qui định.

+Điều khiển độ rộng băng tần: Gatekeeper hỗ trợ việc trao đổi các bản tin H.225.0 là BRQ/BCF/BRJ để điều khiển độ rộng băng tần của một cuộc gọi. Đây cũng có thể là một thủ tục rỗng có nghĩa là nó chấp nhận mọi yêu cầu về sự thay đổi độ rộng băng tần.

+Điều khiển miền: Một miền là một nhóm các đầu cuối H323, các gateway, MCU được quản lý bởi 1 gatekeeper. Trong một miền có tối thiểu một đầu cuối H323, mỗi miền chỉ có duy nhất một gatekeeper. Một miền hoàn toàn có thể độc lập với cấu trúc mạng, bao gồm nhiều mạng được kết nối với nhau. Thông qua các chức năng ở trên: dịch địa chỉ, điều khiển truy nhập, điều khiển độ rộng băng tần, gatekeeper cung

cấp khả năng quản lý miền.

Các chức năng không bắt buộc của Gatekeeper :

- +Điều khiển báo hiệu cuộc gọi.
- +Xử lý báo hiệu cuộc gọi.
- +Hạn chế truy nhập.
- +Giám sát độ rộng băng tần.
- +Giám sát cuộc gọi.

2.4 MCU (Multipoint Control Unit):

MCU là một điểm cuối (Endpoint) trong mạng, nó cung cấp khả năng nhiều thiết bị đầu cuối, gateway cùng tham gia vào một liên kết đa điểm (multipoint conference). Nó bao gồm một MC (Multipoint Controller) bắt buộc phải có và một MP (Multipoint Process) có thể có hoặc không.

Nhiệm vụ của MC là điều tiết khả năng audio, video, data giữa các thiết bị đầu cuối theo giao thức H245. Nó cũng điều khiển các tài nguyên của hội thoại bằng việc xác định dòng audio, video, data nào cần được gửi đến các đầu cuối. Tuy nhiên, MC không thao tác trực tiếp trên các dòng dữ liệu mà nhiệm vụ này được giao cho MP. MP sẽ thực hiện việc kết hợp, chuyển đổi, xử lý các bit dữ liệu.

2.5. Proxy

Proxy là một tác tử xử lý cuộc gọi, nó gởi hạn (terminate) các cuộc gọi từ một LAN hoặc một khu vực và tạo lập các phiên với các H.323 endpoints được định vị trong một LAN hoặc khu vực khác. Proxy cung cấp cho người quản trị sự linh hoạt để xác lập và đảm bảo chất lượng dịch vụ (quality of service). Đồng thời proxy cũng cung cấp một phương thức xác minh các kết nối "H.323 Videoconferencing" xuyên qua tường lửa, và cung cấp môi trường chuyển đổi địa chỉ NAT.

3. Các chuẩn mã hóa và giải mã dữ liệu trong H323:

3.1 Chuẩn CODEC tín hiệu audio:

Ở bên phát, tín hiệu Audio từ microphone trước khi được truyền tiếp phải được mã hoá. Còn ở bên nhận, chúng phải được giải mã trước khi đưa đến speaker. CODEC là dịch vụ tối thiểu mà đầu cuối H323 nào cũng phải có. Vì vậy một thiết bị đầu cuối H323 phải được hỗ trợ tối thiểu là một chuẩn CODEC. Hiện nay đang tồn tại một số chuẩn mã hoá như sau: G.711 (mã hoá tốc độ 64kbps); G.722 (64,56,48 kbps); G.723.1 (5.3 và 6.3 kbps); G.728 (16 kbps); G.729 (8 kbps).

Voice CODEC	Tốc độ	Độ phức tạp	Chất lượng	Độ trễ
G.711 PCM	64	Thấp	Rất tốt	Cực thấp
G.726 ADPCM	40/32/24	Thấp	Tốt (40K) Tối (16K)	Rất thấp
G.729 CS-ACELP	8	Cao	Tốt	Thấp
G.729 A CA-ACELP	8	Vừa phải	Khá tốt	Thấp
G.723 MP-MLQ	6,4/5,3	Cao vừa phải	Tốt (6,4 K) Tối (5,3 K)	Cao
G.723.1 MP-MLQ	6,4/5,3	nt	nt	nt
G.728 LD-CELP	16	Rất cao	Tốt	Thấp

Hình 1.6: Bảng so sánh các chuẩn CODEC

Việc lựa chọn thuật toán CODEC là một trong những yếu tố cơ bản để nâng cao chất lượng thoại Internet.

3.2 Chuẩn CODEC tín hiệu video:

Video CODEC mã hoá tín hiệu hình ảnh từ camera để truyền dẫn và giải mã các tín hiệu video nhận được (đã được mã hoá) để hiển thị hình ảnh. Trong H323, truyền hình ảnh có thể có hoặc không, vì vậy việc hỗ trợ video CODEC là tùy chọn. Tuy nhiên các đầu cuối cung cấp khả năng liên lạc hình ảnh phải được hỗ trợ giao thức mã hoá, giải mã tín hiệu video. Các giao thức hỗ trợ là H261, H263...

Mã hóa hình ảnh là khả năng tùy chọn. Mọi đầu cuối H.323 cung cấp truyền thông hình ảnh đều phải có khả năng mã hóa và giải mã hình ảnh theo chuẩn QCIF H.261 hoặc H.263.

Các bộ mã hóa hình ảnh khác và các dạng hình ảnh khác cũng có thể được dùng thông qua thỏa thuận trong H.245.

Các đầu cuối H.323 có thể hoạt động ở các tốc độ bit hình ảnh, tốc độ khung không cân đối và các giải pháp hình ảnh nếu có nhiều giải pháp hình ảnh hỗ trợ.

CHƯƠNG SÁU

MẠNG ĐA PHƯƠNG TIỆN

Nội dung

- VOIP
- Videoconferencing
- Tin nhắn MMS
- Các dịch vụ khác

1. VOIP

1.1. Khái niệm @iễn thoại IP :

, IP Telephony là một dịch vụ truyền thoại số dùng công nghệ mạng IP kết hợp với khả năng tính toán xử lý dữ liệu của các thiết bị @ụ cuối @ó thực hiện truyền tải các cuộc @um thoại.

Dịch vụ @iễn thoại IP bao gồm các dịch vụ truyền thoại, fax, multimedia, ... qua mạng IP @ic quản lý *vô chết linh dịch vô* (QoS). Các dịch vụ này cần có @ic thực hiện hoàn toàn trong phạm vi mạng IP hoặc cần có @ic thực hiện kết hợp giữa mạng IP và các mạng khác.

1.2. @éc @ióm của @iễn thoại IP.

@iễn thoại IP là một dịch vụ truyền thoại qua các mạng IP. Mạng IP @©y là các mạng dữ liệu số dùng bé giao thức TCP/IP cho các chức năng tầng giao vận (Transport Layer) và tầng mạng (Network Layer). Bên giao thức các tầng thấp hơn (các giao thức truy cập mạng: Network Access Protocols) cần có giao thức trong mạng LAN, X.25, Frame Relay, ATM, hay PPP... Bèi vậy so với @iễn thoại chuyển mạch kênh thông thường @iễn thoại IP cần nhiều khác biệt. @ó đây @ic nêu ra các khác biệt này, tríc hết ta xem xét @éc @ióm mạng PSTN, mạng chuyển mạch gói và mạng Internet.

PSTN, mạng chuyển mạch gói và mạng Internet:

Mạng @iễn thoại công cộng (Public Switched Telephone Network - PSTN) là mạng truyền thông dựa trên nền tảng kỹ thuật **chuyển mạch kênh (Circuit Switching)**. Chuyển mạch kênh là phương pháp truyền thông trong @ã một kênh truyền đến *dụng riêng* @ic thiết lập giữa hai thiết bị @ụ cuối thông qua một hay nhiều nút chuyển mạch trung gian. Dòng thông tin truyền trên kênh này là dòng bit truyền *liên tục* theo thời gian. Bềng thông của kênh dụng riêng @ic @m *bỏ* và cè *phn* trong quá trình liên lạc (64Kbps @èi với mạng @iễn thoại PSTN), và @é *trở* thông tin là rất nhỏ chø là thời gian truyền thông tin trên kênh (propagation time).

Khác với mạng chuyển mạch kênh, **mạng chuyển mạch gói (Packet Switching Network)** sử dụng *hệ thống lưu trữ - gửi (store-and-forward system)* tức là các nút mạng. Thông tin được chia thành các phần nhỏ (gửi lưu gói), mỗi gói được thêm các thông tin điều khiển cần thiết cho quá trình truyền như lưu địa chỉ nơi gửi, địa chỉ nơi nhận... Các gói thông tin được nút mạng xử lý và lưu trữ trong một thời gian nhất định rồi mới được truyền được nút tiếp theo sao cho việc sử dụng kênh của hiệu quả cao nhất. Trong mạng chuyển mạch gói không cần kênh định riêng cho một thiết lập, bằng thông của kênh logic giữa hai thiết bị. Các cuộc thông không cần định, và để trở thông tin lớn hơn mạng chuyển mạch kênh rất nhiều.

Khi các mạng sẽ liên kết với nhau, một mạng sẽ liên kết được thiết lập, để là **Internet**. Giao thức liên mạng sử dụng trong internet là giao thức IP (Internet Protocol). Cùng giêng như mạng điện thoại PSTN, mạng Internet cần quy mô toàn cầu, phí hội cho việc khai thác và phát triển các dịch vụ viễn thông trên thế giới. Tuy nhiên, giữa chúng cần sự khác biệt sâu sắc và mặt kỹ thuật:

- **Khác nhau về kỹ thuật chuyển mạch:** Mạng Internet sử dụng phương pháp định tuyến động (dynamic routing) dựa trên việc định địa chỉ không mang tính địa lý (non-geographic addressing). Ngược lại mạng PSTN, sử dụng phương pháp chuyển mạch tĩnh (static switching) dựa trên sẽ điện thoại phổ biến và vị trí của thuê bao (geographic telephone numbering).
- **Kiến trúc mạng:** Kiến trúc của Internet là phân tán. Trong khi để mạng PSTN tập trung các phần tử truyền đến và các chức năng truyền để điều khiển là với nhau tại một vị trí trung tâm trong mạng.

Kiến trúc phân tán của internet cần một ý nghĩa hết sức quan trọng. Nó tạo cho internet tính linh hoạt trong khai thác và ứng dụng các công nghệ mới. Điều này giúp thích cho sự phát triển mạng viễn thông trong một môi trường biến động.

1.3. Ưu điểm và nhược điểm của điện thoại IP:

+ **Giảm chi phí cuộc gọi:** Nếu dịch vụ điện thoại IP được triển khai, chi phí cho một cuộc gọi được giảm đáng kể so với chi phí truy cập internet. Nguyên nhân đến để chi phí thấp như vậy là do tính hiệu quả của truyền tải trong mạng IP cần khả năng sử dụng kênh hiệu quả cao. Chẳng hạn, kỹ thuật nén thoại tiên tiến giảm tốc độ bit từ 64 Kbps xuống thấp tới 8 Kbps (theo tiêu chuẩn nén thoại G.729A của ITU-T) kết hợp với tốc độ xử lý nhanh của các bộ vi xử lý ngày nay cho phép việc truyền tải theo thời gian thực mà cần ít thời gian để xử lý với lượng tại nguồn bằng thông thấp hơn nhiều so với kỹ thuật cũ.

So sánh mét cuộc gài trong m¹ng PSTN với mét cuộc gài qua m¹ng IP, ta thấy:

- Chi phí phải trả cho cước gọi trong mạng PSTN là chi phí phải trả trả ra để duy trì cho một kênh 64kbps suốt từ đầu cuối tới cuối kia thông qua một hệ thống các tầng mạng. Chi phí này sẽ với các cước gọi liên đới (liên đới, quốc tế) là khá lớn.
- Trong trường hợp cước gọi qua mạng IP, người sử dụng trong mạng PSTN chỉ phải duy trì kênh 64kbps ở cổng Gateway của mạng cung cấp dịch vụ thôi. Như cung cấp dịch vụ liên đới thoại IP sẽ làm nên nhiệm vụ nên, cũng gọi tên hiệu quả vụ gọi chóng lại qua mạng IP một cách cả hiệu quả nhất. Ở tất cả các Gateway này tất cả một mạng liên đới thoại khác cả người liên lạc đầu kia. Việc kết nối này như vậy làm giảm đáng kể chi phí cước gọi do phải liên lạc kênh truyền 64Kbps. Ngược lại thay thế bằng việc truyền thông tin qua mạng dữ liệu hiệu quả cao.

+ **TÝch hìp m¹ng tho¹i, m¹ng sè liÖu vµ m¹ng b_o hiÖu:**
 Trong ®iÖn tho¹i IP, tÝn hiÖu tho¹i, sè liÖu vµ ngay c¶ b_o hiÖu ®Öu
 cũ thÓ cïng ®i trªn cïng mét m¹ng IP. §iÖu nµy sÏ tiÖt kiÖm ®êc chi phÝ
 ®Çu t ®Ó x©y dùng nh÷ng m¹ng riªng rÏ.

+ **Khả năng mở rộng (Scalability):** Nếu nhu cầu hệ thống tăng lên thì việc mở rộng hệ thống sẽ rất dễ dàng. Có thể thêm vào hệ thống các thiết bị trong mạng internet cũng như các thiết bị khác để mở rộng hệ thống mà không cần phải thay đổi cấu trúc mạng. Chính việc mở rộng này đã mang lại cho dịch vụ Internet thời IP khả năng mở rộng dễ dàng hơn so với việc mở rộng truyền thống.

+ Kh«ng cÇn th«ng tin ®iÒu khiÓn ®Ó thiÕt lÛp k¸nh truyÒn vÛt lý: G¸i th«ng tin trong m¹ng IP truyÒn ®Õn ®Ých mụ kh«ng cÇn mét sù thiÕt lÛp k¸nh nµo. G¸i chØ cÇn mang ®¸p chØ cña n-í nhËn cuèi cïng lụ th«ng tin ®· c¸ thÓ ®Õn ®íc ®Ých. Do vÿy, viÖc ®iÒu khiÓn cuéc g¸i trong m¹ng IP chØ cÇn tÛp trung vµo chøc nng cuéc g¸i mụ kh«ng ph¶i tÛp trung vµo chøc nng thiÕt lÛp k¸nh.

+ **Quản lý băng thông:** Trong phiên giao dịch chuyển mạch kênh, tại nguồn băng thông cung cấp cho một cuộc liên lạc phụ thuộc vào phiên (một kênh 64Kbps) nhưng trong phiên giao dịch IP việc phân chia tại nguồn cho các cuộc thoại linh hoạt hơn nhiều. Khi một cuộc liên lạc diễn ra, nếu lưu lượng của mạng thấp, băng thông dành cho liên lạc sẽ cho chất lượng thoại tốt nhất cả hai; nhưng khi lưu lượng của mạng cao, mạng sẽ hạn chế băng thông của tổng cuộc gọi để mức duy trì chất lượng thoại chấp nhận được nhằm phục vụ cho tất cả các phiên nhiều người nhất.

+ **Nhiều t́nh n"ng đpch vô:** T́nh linh ho"t của m"ng IP cho phđp t"o ra nhiều t́nh n"ng mới trong đpch vô tho"i. V́ dụ cho bi"t th"ng tin v"o ng"i g"i tí hay mét thu" bao "i"n tho"i IP cũ th" cũ

nhiều sẽ liên lạc mọi chỗ cần một thiết bị (Ví dụ như một thiết bị IP Phone cần có một sẽ (điện thoại) dành cho công việc, một cho các cuộc gọi riêng tư).

+ **Khả năng multimedia:** Trong một “cuộc gọi” người sử dụng cần có khả năng chuyển đổi số đồng các dịch vụ khác như truyền file, chia sẻ dữ liệu, hay xem hình ảnh như cần người chuyển biến kia.

Điện thoại IP cũng cần những hạn chế:

+ **Kỹ thuật phức tạp:** Truyền tín hiệu theo thời gian thực trên mạng chuyển mạch gói rất khó khăn do một gói trong mạng lưu không theo trình tự và để tránh không có phân chia các gói thông tin khi truyền trên mạng. Số cần một dịch vụ thời gian thực, cần thiết phải cần một kỹ thuật nén tín hiệu và cần những yêu cầu khác: tốc độ nén (để giảm băng thông), cần khả năng suy rộng và tốc độ khôi phục thông tin của các gói mất đi... Tỉ lệ xử lý của các bé Codec (Coder and Decoder) phải rất nhanh để không làm việc chậm trễ thời gian xử lý. Sắp xếp cần sẽ có tập của mạng cũng cần các công việc liên quan như công nghệ mới như Frame Relay, ATM,... Để cần tỉ lệ cao hơn và hiệu quả phải cần một chỗ thực hiện chức năng QoS (Quality of Service).

+ **Vấn đề bảo mật (security):** Mạng Internet lưu một mạng cần tính riêng biệt và hỗn hợp (heterogenous network). Trong đó cần rất nhiều loại máy tính khác nhau cũng các dịch vụ khác nhau cũng sử dụng chung một cần sẽ có tập. Do vậy không cần quá (tính bảo mật) thông tin liên quan đến các nhân công như sẽ liên lạc truy cập số đồng dịch vụ của người dùng và các dữ liệu.

M« hình dịch vụ IP Telephony quy mô nhỏ cung cấp dịch vụ.

III.2.1. M« hình.

Số cần có lưu một dịch vụ để lập, cần phải cần một mô hình IP Telephony toàn diện hơn, cần khả năng (tính bảo mật) các yêu cầu:

- ♦ Hoạt động song song với mạng chuyển mạch kênh để cần có mang dịch vụ thời gian thực (điện thoại) cho (các) người sử dụng và cần (các) tài nguyên để chi phí thấp của việc truyền tín hiệu trên mạng chuyển mạch gói.

- ♦ Thực hiện các chức năng cần (các) tính năng của một dịch vụ viễn thông như tính cước, quản lý cuộc gọi...

- ♦ Cần có thực hiện những chức năng mới của (điện) thoại IP.

Số hoạt động cũng với mạng chuyển mạch kênh, mô hình bước phải bao gồm hỗ trợ của (các) (điện) thoại. Cần và sẽ cần cần thiết để cần một kết nối thông suốt giữa hai (các) cuối thông qua mạng IP. Trong mô hình, hỗ trợ của (các) số

dòng lụ SS7 nhng trong thùc tũ hũ thềng b, o hiũu cũ thũ lụ nh÷ng hũ thềng b, o hiũu kh, c nh R2...

C, c thụnh phÇn trong hũ thềng VoIP bao gảm:

➤ **Media Gateway:**

- ♦ ChuyỐn ®æi khu«n d¹ng th«ng tin: tũ d¹ng th«ng tin ghĐp kãnh theo thêi gian (TDM) trong m¹ng chuyỐn m¹ch kãnh sang d¹ng gãi trong m¹ng IP vµ ngíc l¹i.

- ♦ Thùc hiỐn c, c qu, trxn h xũ lý cÇn thiỐt kh, c nh: NĐn tÝn hiũu tho¹i (voice compression), nĐn kho¶ng lÆng (silence suppression), triỐt tiỐng vãng (echo cancellation)

- ♦ Cung cÊp nhiều giao điỐn vỄt lý cÇn thiỐt cho kỐt nêi: Giao điỐn vớ m¹ng chuyỐn m¹ch kãnh (E1/T1, PRI-ISDN, ...), giao điỐn vớ m¹ng IP (Ethernet, Fast Ethernet, Frame Relay, ...).

➤ **Signalling Gateway:** Phôc vô cho b, o hiũu gi÷a c, c terminal trong m¹ng chuyỐn m¹ch kãnh vµ c, c terminal trong m¹ng IP: §ăng gãi l¹i c, c th«ng ®iỐp SS7 thụnh c, c gãi phĩ hĩp vớ m¹ng IP, lăc c, c đĩng lu lĩng kh«ng phĩ hĩp... Thụnh phÇn nỳy kỐt hĩp ho¹t ®éng cũa m¹ng IP vµ m¹ng b, o hiũu SS7.

➤ **Call Control Center:**

- ♦ Hĩng đén Media Gateway c, c thiỐt lỄp, xũ lý vµ kỐt thóc đĩng th«ng tin media (th«ng tin thêi gian thùc) phôc vô cho cuéc gãi.

- ♦ Xũ lý th«ng tin b, o hiũu.

- ♦ Theo dâi tr¹ng th, i cũa tỄt c¶ c, c đĩng media ®ang truyền trong hũ thềng.

- ♦ Thùc hiỐn nhiều đpch vô cũa hũ thềng: TÝnh cíc, t¹o ra c, c b¶n ghi lu tr÷, c, c chøc n̄ng qu¶n lý m¹ng, qu¶n lý cuéc gãi...

➤ **C, c thụnh phÇn kh, c:** Bao gảm c, c terminal PSTN (m, y ®iỐn tho¹i, m, y fax), tæng ®µi PSTN PBX; thiỐt bP trong m¹ng IP (IP phone, IP PBX).

III.2.2. Hỗ trợ băng.

a) Mục đích của bài ra hai giải pháp VoIP:

Mét mặt, giải pháp cho các doanh nghiệp (Enterprise): Thông tin của doanh nghiệp sẽ được hai liên kết một cho các dịch vụ thoại qua mạng PSTN và một cho kết nối internet. Ở đó sẽ có các dịch vụ như e-mail, web... Với giải pháp VoIP cho doanh nghiệp, các doanh nghiệp sẽ có gateway của riêng họ. Ở kết nối các thiết bị ở cuối cuộc gọi truy cập thông (máy điện thoại, fax, PBX) vào Internet. Như vậy, thoại và các dịch vụ internet sẽ tích hợp vào một mạng truy cập chung thay vì phải dùng hai mạng truy cập như trước kia.

Hai mặt, giải pháp cho dịch vụ công cộng (Carrier): Như gateway của mạng lớn sẽ tích hợp cung cấp dịch vụ IP Telephony bên trên, kết nối thông xuyên vào mạng IP sẵn sàng cho người sử dụng truy cập tới ở đó sẽ có dịch vụ VoIP. Như người sử dụng với các máy điện thoại thông tin truy cập ở gateway thông qua dịch vụ điện thoại từ nhà mạng. Giải pháp này là dịch vụ điện thoại IP sẽ tích hợp với phần lớn người dùng, cung cấp cho người dùng như cuộc gọi băng dẹt gì, rồi.

b) Quay số truy cập dịch vụ.

Quay số truy cập dịch vụ liên quan đến vấn đề lộ lượng thông tin ở bên hính cuộc gọi đi qua mạng IP mà không đi qua mạng PSTN như dịch vụ điện thoại thông tin (trên hệ liên lạc phone to phone). Số truy cập dịch vụ IP Telephony người sử dụng cả số cả hai cách thức truy cập tùy thuộc vào khả năng của hệ thống.

- ♦ *Quay số hai giai đoạn:* Người sử dụng quay số ở liên lạc ở Gateway IP Telephony và quá trình giao tiếp với gateway sẽ quyết định việc hính tiếp cuộc gọi ở terminal nào ở bên kia.
- ♦ *Quay số một giai đoạn:* Người sử dụng quay một số truy cập để biết tìm theo số điện thoại để gọi và việc quay số chỉ phải thực hiện một lần. Còn về việc truy cập, hệ thống báo hiệu sẽ quyết định việc tùy chọn cuộc gọi thông qua mạng IP hay chỉ ở trong mạng PSTN như thông tin.

c) Thiết lập cuộc gọi:

Xét trên hệ người sử dụng truy cập dịch vụ IP Telephone tổ trong mạng PSTN sẽ có một người sử dụng khác cùng trong mạng PSTN. Số cuộc gọi cả số sẽ tạo ra, cần phải quan tâm ở ba giai đoạn báo hiệu thiết lập kết nối:

1. Giai đoạn thiết lập kết nối giữa người gọi và Gateway ở người sử dụng ở truy cập mạng IP (gateway 1).

2. Giai đoạn thiết lập cuộc gọi từ Gateway 1 vào cổng gateway đích (gateway 2).

3. Giai đoạn thiết lập liên lạc từ gateway 2 tới máy nhận thoại của người bắt gọi

c) Thông tin thoại trong hệ thống VoIP

Tín hiệu thoại từ người sử dụng được chuyển thành tín hiệu PCM ghép kênh theo thời gian (TDM) truyền qua một hệ thống mạng nội bộ cổng Media Gateway của dịch vụ IP Telephony. Media gateway thực hiện việc trích tiếng vọng, nén tín hiệu, và mã hóa tín hiệu thoại. Các xử lý thành phần này gọi là truyền tới gateway đích qua mạng IP. Media gateway đích thực hiện việc chuyển ngược tín hiệu gốc là truyền nhận người sử dụng ở cuối bên kia.

Mô hình (nền) tín hiệu thoại:

Tín hiệu thoại trong hệ thống VoIP được xử lý theo quy trình sau:

1. *Số hóa*: Tín hiệu thoại được lấy mẫu theo tần số 8 KHz và mã hóa theo chuẩn.

2. *Nén*: Dữ liệu số hóa được nén theo các tiêu chuẩn khác nhau như: G.711 (PCM 64kb/s), G.722 (Wideband Coder), G.723.1 (MPC-MLQ), G.726 (ADPCM), G.728 (LD-CELP), G.729/G.729A (CS-ACELP).

Trong trường hợp của Gateway giao tiếp với mạng chuyển mạch kênh (PSTN/ISDN), các dòng PCM 64Kbps luật A-law thì các giao diện mạng PSTN/ISDN được chuyển đổi thành mẫu nén, trích tiếng vọng và mã hóa theo một trong các chuẩn kể trên.

Mỗi phương pháp nén các đặc điểm riêng và các tham số trong thành phần điều khiển cốt lõi của mạng. Số lượng, các phương pháp nén, ta xem xét chúng theo 4 đặc điểm:

a) *Tốc độ bit (bit rate)*: Tốc độ bit là một đặc tính quan trọng khi nói về một phương pháp nén thoại, nó biểu hiện mức độ nén tín hiệu của phương pháp. Các chuẩn nén thoại trên cho các tốc độ bit từ 6,4Kbps/5,3Kbps (G.723.1) đến 64 Kbps (G.711).

b) *Độ trễ (Delay)*: Do vậy, độ trễ là một yếu tố phải được xem xét và kích thích khung thoại. Khung thoại càng lớn và tốc độ bit càng chậm thì độ trễ càng cao.

c) *Độ phức tạp (Complexity)*: Nén thoại được thực hiện bởi thành phần xử lý số (DSP) hay bởi thành phần CPU trong máy tính. Độ phức tạp của phương pháp nén được thể hiện ở mức độ phức tạp của DSP hoặc CPU cần thực hiện trong một đơn vị thời gian (MIPS - Millions of Instruction per Second) và sẽ liên quan đến chi phí thiết kế cho

thuật toán nén. Số phức tạp của phép toán liên quan đến giá trị của thời gian xử lý.

d) *Chất lượng tín hiệu (Quality)*: Chất lượng tín hiệu thời gian liên quan đến tỷ lệ tín hiệu trên nhiễu S/N của tín hiệu truyền tải hay số lỗi bit BER của dòng thời gian tùy chỉnh theo yêu cầu. Một hình thức khác của chất lượng thời gian liên quan với các loại tín hiệu khác nhau. Kết quả phụ thuộc vào thành phần là gì trong điều kiện nhiễu nền.

Dĩ vậy phụ thuộc vào các kết quả khác nhau của các phép toán nén thời gian theo số đồng trong các hệ thống VoIP.

Chuẩn nén	Tốc độ bit	Kích thước khung thời gian	Số phức tạp
G.711 PCM	64 Kb/s	125 μ s	
G.723 ADPCM	32 Kb/s	125 μ s	
G.722	48,56,64Kb/s	125 μ s	
G.728 LD-CELP	16 Kb/s	625 μ s	30 MIPS
G.729 CS-ACELP	8 Kb/s	10 ms	20 MIPS
G.729A	8 Kbps	10 ms	10,5 MIPS
G.723.1 MPC-MLQ	5,3 & 6,4Kb/s	30 ms	16 MIPS; 2200 tổ chức

2. Mô hình IP Videoconferencing.

1. Giới thiệu IP Videoconferencing.

a. Giới thiệu.

Videoconferencing là một công nghệ truyền thông tương tác. Nó kết hợp truyền video và Audio đồng thời, cho phép những người ở hai hoặc nhiều vị trí, khu vực khác nhau có thể nhìn và nghe những người ở vị trí khác như là khi họ tham dự các cuộc hội thoại mặt đối mặt (face-to-face).

Có nhiều loại dịch vụ Videoconferencing bao gồm: corporate communications, telemedicine, e-health, training, e-Learning, telecommuting và customer service.

Videoconferencing có thể ở dạng điểm điểm (point-to-point) hoặc multipoint. Với dạng điểm điểm thì chỉ có hai điểm cuối tương tác với nhau nhưng trong dạng multipoint thì có nhiều điểm cuối cùng tương tác.

Videoconferencing có thể triển khai trên nhiều hệ tầng mạng khác nhau như: ATM (Asynchronous Transfer Mode), ISDN (Integrated Services Digital Network) IP LANs với H.323, POTS (Plain Old Telephone Services), ... Tuy nhiên ở đây ta chỉ khảo sát Videoconferencing trên mạng IP sử dụng chuẩn H.323.

• Một số mô hình kịch bản của Videoconferencing:

- Các mô hình kịch bản Point-to-Point videoconferencing

- Person-To-Person
- Studio-To-Person
- Studio-To-Studio

- Mô hình kịch bản Multipoint videoconferencing:

b. Các chuẩn Videoconferencing.

Chuẩn H.323 cung cấp một nền tảng cho việc tra đổi Audio, video và các dữ liệu khác trên mạng IP. H.323 được giới thiệu và bảo hộ bởi International Telecommunication Union (ITU). Nó là một phần trong tập hợp chuẩn H.32x cho Videoconferencing. Bảng dưới đây liệt kê vài chuẩn trong H.32x cùng với các công nghệ mạng đi kèm.

H.320	ISDN	Circuit
H.321	ATM	Circuit
H.322	IsoEthernet	TDM
H.323	Ethernet	Packet
H.324	Analog	Circuit

• Các chuẩn nén dữ liệu trong IP videoconferencing:

Audio:

- **G.711** Là một giải thuật nén khá cũ dùng nén tiếng nói thoại trong giải tần cơ sở (3KHz analog Bandwidth), chất lượng mã hoá audio 48, 56, 64kbps.
- **G.722** Mã hoá cải tiến chất lượng Audio 48, 56, 64 kbps, 7KHz analog Bandwidth.
- **G.723.1**
- **G.728** : chất lượng mã hóa audio 16 kbps, 3KHz analog Bandwidth.
- **G.729** : chất lượng mã hóa audio 8 kbps, 3KHz analog Bandwidth

Video:

H.261, H.263

• Các chuẩn điều khiển trong IP Videoconferencing:

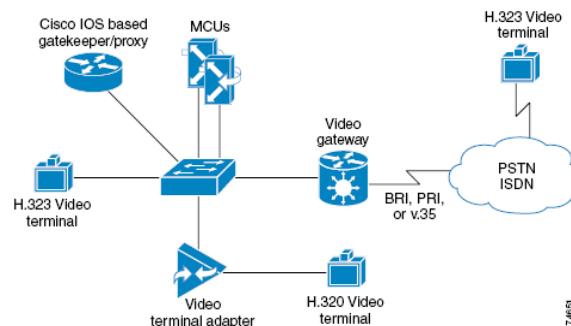
- **H.225:** Cung cấp sự kiện: Registration, admission, status (RAS) giữa các thiết bị H.323 và gatekeeper. Nó còn sử dụng trong khi cài đặt cuộc gọi để mở kênh truyền tín hiệu sử dụng chuẩn Q.931
- **H.245:** Cung cấp sự truyền tín hiệu điều khiển được sử dụng để trao đổi thông điệp điều khiển đầu cuối. Thông điệp điều khiển chứa các thông tin:
 - Khả năng trao đổi (Capabilities exchange)
 - mở, đóng các kênh logic được sử dụng để mang các dòng media
 - Thông điệp điều khiển luồng
 - Những lệnh vad chỉ thị chung.

• Chuẩn để chia sẻ dữ liệu: là **T.120**

• Kiến trúc của chuẩn H.323:

c. Mô hình các thành phần IP Videoconferencing.

Một hệ thống IP Videoconferencing với chuẩn H.323 thường có năm thành phần và được minh họa qua mô hình dưới đây. (**Xem chuẩn H323 chương 5**)



Hình B.1.1.5: Mô hình các thành phần của hệ thống IP Videoconferencing

e. Các chức năng của hệ thống Videoconferencing

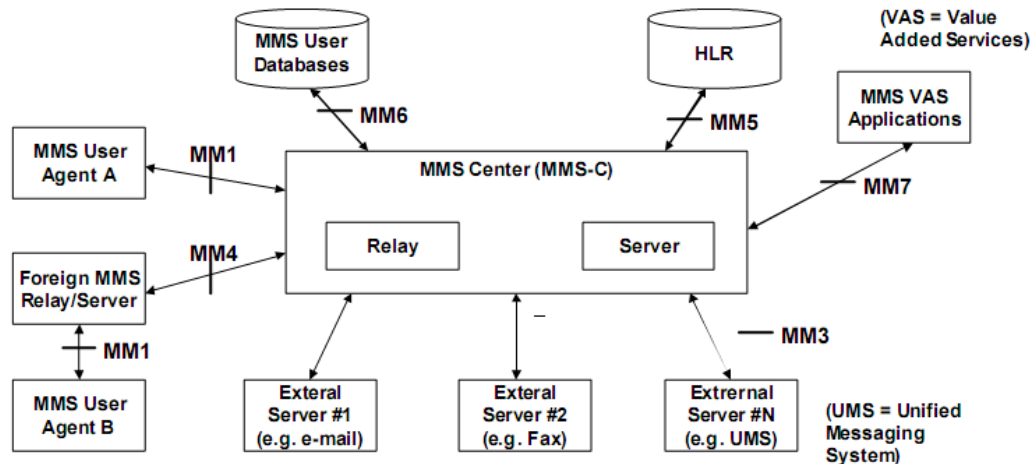
- Thiết lập kết nối đa điểm
- Điều khiển phiên làm việc động
- Điều độ hội nghị tự động
- Kết thúc

3. Tin nhắn MMS

I. Khái niệm:

Tin nhắn MMS là loại tin nhắn bao gồm cả văn bản, âm thanh, hình ảnh, video... (đa phương tiện). Người dùng với các thiết bị di động có hỗ trợ MMS của mình (như điện thoại, PDA, Pocket PC...) có khả năng tự soạn 1 tin nhắn MMS sinh động gửi cho bạn mình thông qua hệ thống mạng MMS hoặc nhận tin nhắn MMS từ một nhà cung cấp dịch vụ thứ 3 (chẳng hạn nhà cung cấp nhạc chuông, hình ảnh...).

II. Kiến trúc hệ thống mạng MMS:



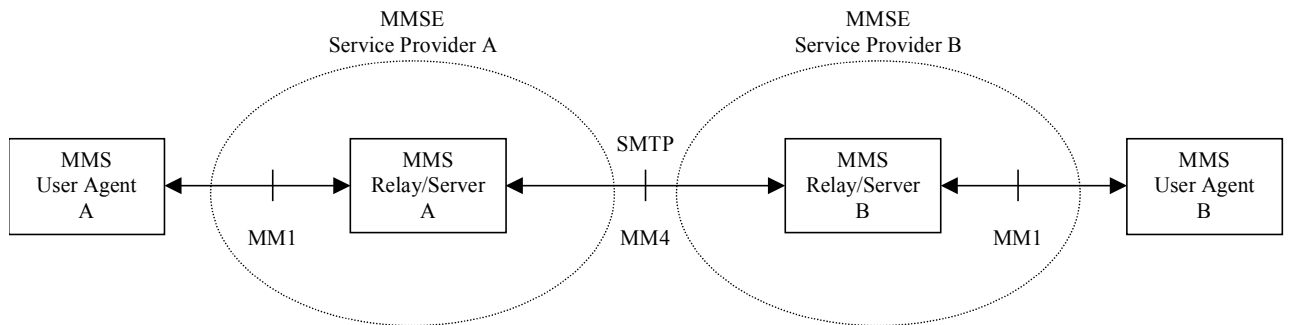
1. MMSE

Môi trường truyền của MMS (MMSE) gồm tất cả những phần tử tham gia gửi (phân phát), lưu trữ và nhận tin nhắn MMS, sau đây là hai phần tử em trình bày:

- **MMSC** là trung tâm của hệ thống thực hiện các chức năng xử lý tin nhắn MMS đến (Truyền phát vận chuyển tin nhắn) trong các mạng điện thoại di động. MMSC là sự kết hợp của 2 bộ phận MMS Relay và MMS Server
- **MMS VAS Server** - MMS Value Added Service: Máy chủ của nhà cung cấp dịch vụ gia tăng (thường gọi là dịch vụ thứ 3). Nhà cung cấp dịch vụ thứ 3 giao tiếp độc lập với User Agent để tạo nên cuộc giao dịch MMS và khi thực hiện nhắn tin MMS, VASP (Value Add Service Provider) buộc phải nhờ đến MMSC.

2. Các giao diện trong kiến trúc mạng MMS:

- **MM1**: giao diện giữa MMS Relay/Server và MMS Client, được sử dụng để:
 - + Submit tin nhắn MMS từ User Agent đến MMS Relay/Server
 - + User Agent Pull tin nhắn MMS từ MMS Relay/Server
 - + MMS Relay/Server Push thông tin về tin nhắn MMS tới User Agent (MM notification)
 - + Trao đổi các bản tin delivery report giữa MMS Relay/Server và User Agent
- **MM2**: giao diện giữa MMS Server và MMS Client
- **MM3**: giao diện giữa MMS Relay/Server và các server của hệ thống tin nhắn ngoài.
- **MM4**: giao diện giữa các MMS Relay/Server thuộc các MMSE khác nhau. Hoạt động dựa trên giao thức SMTP:



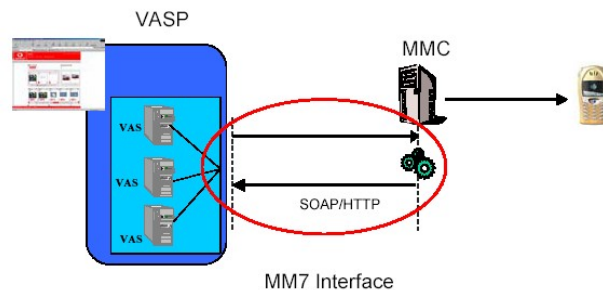
- MM5: giao diện giữa MMS Relay/Server với HLR, được sử dụng để cung cấp thông tin về các thuê bao.
- MM6: giao diện giữa MMS Relay/Server với MMS User Database
- MM7: giao diện giữa MMS Relay/Server với MMS VAS Server (Nhà cung cấp dịch vụ thứ 3). Hoạt động dựa trên nền tảng HTTP và SOAP (Sẽ được trình bày kĩ ở phần sau).

III. Giao diện MM7:

1. Khái niệm:

MM7 là giao diện trao đổi dữ liệu được xây dựng dựa trên nền tảng giao thức HTTP. Bao gồm những đặc điểm sau:

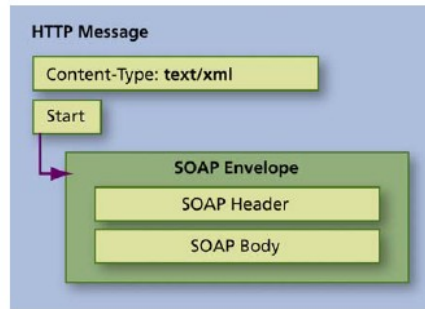
- Việc gửi và nhận được thực hiện thông qua HTTP.
- Cấu trúc gói tin của MM7 được xây dựng dựa trên HTML và SOAP. Đó là một loại định dạng dữ liệu XML - loại dữ liệu vạn năng cho phép đóng gói tất cả dữ liệu trong một và dễ dàng di chuyển trên Internet không gặp trở ngại nào (như Tường lửa, Proxy...). MM7 tồn tại trong quá trình VASP và MMSC (hay MMC) trao đổi thông tin với nhau:



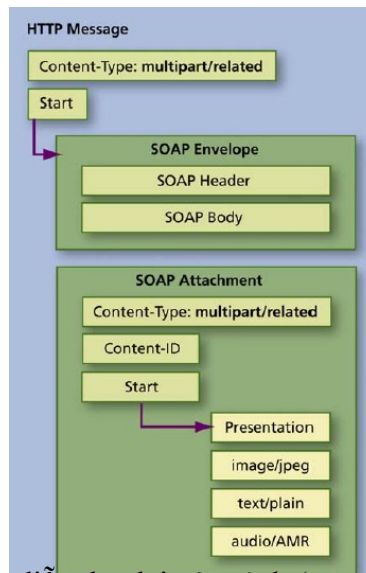
Một tin nhắn MMS với rất nhiều loại dữ liệu ở trong đã được đóng gói theo dạng SOAP – Simple Object Access Protocol – Giao thức truy cập đối tượng giản đơn (định dạng XML của Microsoft) theo các cách khác nhau được quy ước trong giao thức MM7 nhằm thực hiện nhiệm vụ truyền đạt thông tin.

2. Cấu trúc SOAP

Trong trường hợp tin nhắn MMS không có kèm các hình ảnh, âm thanh..., nó được tổ chức theo cấu trúc SOAP như sau:



Trong trường hợp tin nhắn MMS có kèm các hình ảnh, âm thanh..., nó được tổ chức theo cấu trúc SOAP như sau:



Các tin nhắn MMS có khả năng trình diễn, lặp lại các cảnh (tăng yếu tố sinh động của tin nhắn). Khởi đầu của Attachment được trỏ đến phần trình diễn.

4. Giới thiệu một số dịch vụ

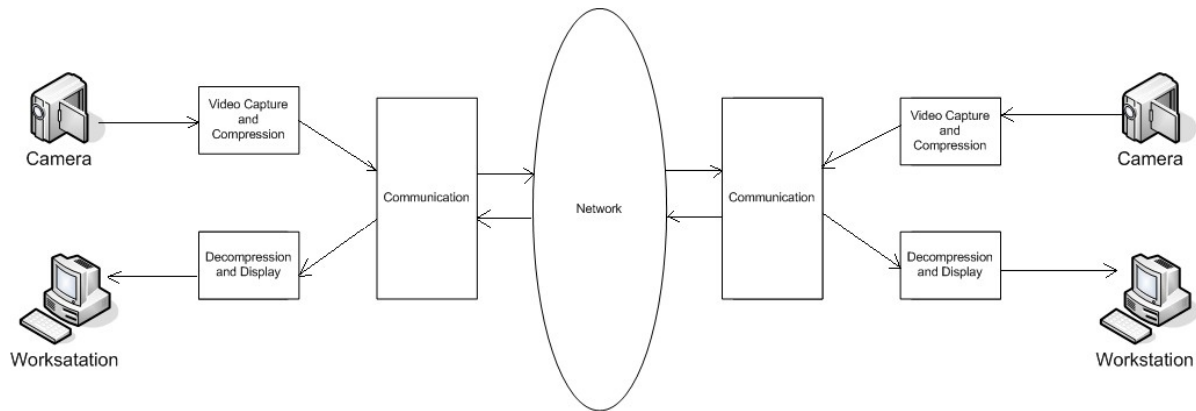
II. Các dịch vụ truyền video trên mạng Ip

2. Video chat

Chức năng:

Ứng dụng video chat cho phép 2 hoặc nhiều người đang ngồi tại những nơi khác nhau có thể vừa chat với nhau, vừa nhìn thấy được nhau. Về kĩ thuật hệ thống video chat cần có khả năng thiết lập kết nối ngang hàng, sau đó thực hiện truyền thông giữa các điểm này.

Các thành phần của hệ thống



Trong đó

Khối Video Capture and Compression thực hiện bắt hình ảnh từ camera sau đó thực hiện nén theo 1 chuẩn video nào đó (có thể là Mpeg hoặc H26X).

Khối Communication thực hiện chức năng thiết lập kết nối giữa 2 hoặc nhiều máy muốn sử dụng ứng dụng video chat, và truyền nhận dữ liệu video đã nén, hoặc dữ liệu text.

Khối Decompression và Display giải nén dữ liệu nhận được và thực hiện thị cho người sử dụng.

Mô hình hoạt động

Trong ứng dụng video chat không có yêu cầu về điều khiển tương tác giữa những người sử dụng dịch vụ.

Hạ tầng cơ sở và giao thức sử dụng

Dịch vụ videochat có thể triển khai trên hạ tầng mạng IP Lan, Wan và ngay cả trên Internet.

Video chat sử dụng chuẩn nén video mpeg, H26X và các giao thức truyền tải dữ liệu thời gian thực như RTP/RTCP.

II.3. Hội nghị video (video conference)

Chức năng

Hội nghị video cho phép người dùng ở các địa điểm khác nhau có thể tiến hành trao đổi thông tin đa phương tiện đảm bảo thời gian thực và có tính tương tác cao.

Về mặt xã hội một hội nghị bao giờ cũng phải đảm bảo được tổ chức tốt tức là định nghĩa ra các cơ chế, quy định dành cho các thành viên tham gia, quản trị các thành viên, có chức năng điều hành hội nghị về nội dung và tương tác giữa các thành viên, có chức năng giám sát toàn bộ hội nghị.

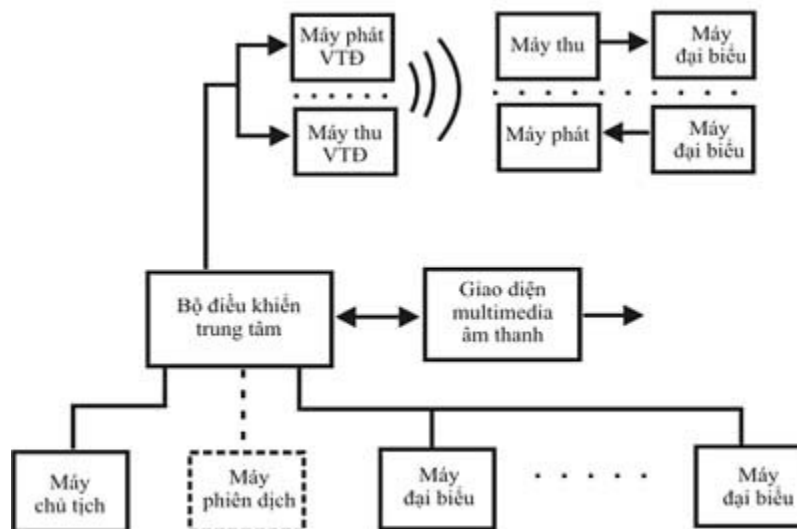
Về kỹ thuật hệ thống hội nghị video cần có các chức năng sau

- ✓ Thiết lập kết nối đa điểm
- ✓ Điều khiển phiên làm việc động
- ✓ Cung cấp dịch vụ thực mụ hội nghị (tức là dịch vụ cung cấp danh sách các tài nguyên chia sẻ, cho phép người sử dụng có thể dễ dàng lựa chọn và tải về những dữ liệu họ cần).
- ✓ Điều độ hội nghị tự động.
- ✓ Đồng bộ giữa âm thanh và hình ảnh.

Sơ đồ tổ chức chung của 1 hội nghị video

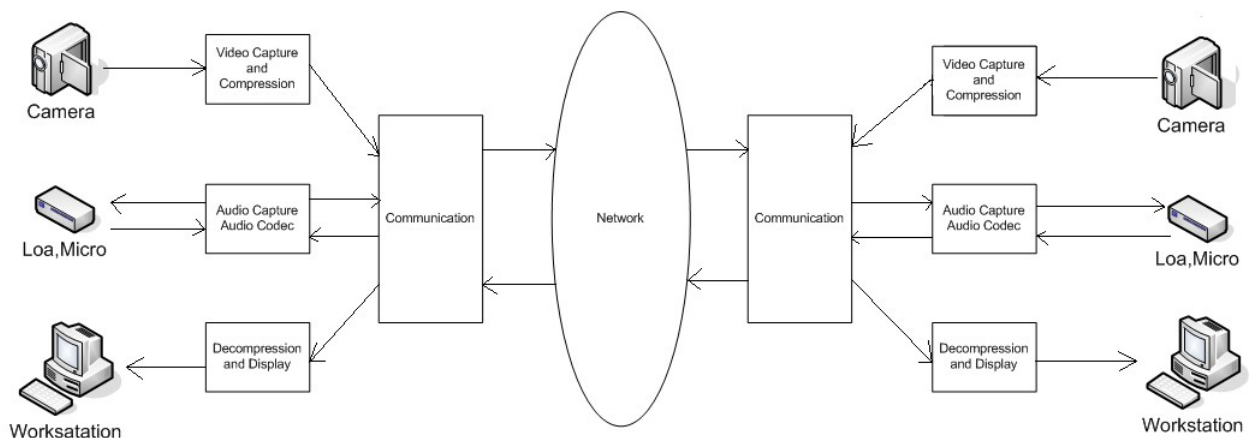
- Thiết bị điều hành trung ương (CCU) thực hiện việc quản lý và điều hành toàn bộ hệ thống hội nghị video. Chức năng của CCU gồm có: điều khiển các micro, cung cấp các tư liệu và hiển thị phương thức biểu quyết...

Ngoài ra hệ thống CCU còn có giao diện multimedia audio, video. Thông qua giao diện này, các tin tức và dữ liệu của hội nghị đưa tới các thiết bị analog như truyền hình quảng bá hoặc phát thanh quảng bá.



Trong phòng hội nghị, tín hiệu video được ghi bằng camera và hiển thị lên màn hình cỡ lớn. Thiết bị điều khiển CCU điều khiển các camera di chuyển đến các vị trí thích hợp để ghi toàn bộ hình ảnh của hội nghị. Sau đó đưa lên màn hiển thị. Trường hợp có một số ít đại biểu tại địa điểm xa thì có thiết bị thu phát vô tuyến điện liên lạc song công với các đại biểu đó.

Các thành phần của hệ thống hội nghị video



Trong đó

Khối Video Capture and Compression thực hiện bắt hình ảnh từ camera sau đó thực hiện nén theo 1 chuẩn video nào đó (có thể là Mpeg hoặc H26X).

Khối Decompression và Display giải nén dữ liệu nhận được và thực hiện thị cho người sử dụng.

Khối Audio Capture and Audio Codec thực hiện bắt âm thanh và nén, giải nén âm thanh trước khi truyền.

Khối Communication đảm nhận các chức năng chính của hệ thống hội nghị video, như đã trình bày ở trên.

Mô hình tương tác

Một đặc điểm quan trọng trong hội nghị video là tính tương tác giữa các thành viên tham gia hội nghị. Dựa trên cách thức điều khiển tương tác của hệ thống ta có thể chia ra làm 2 mô hình: điều khiển tập trung và điều khiển phân tán.

Trong điều khiển tập trung chỉ có 1 chủ tịch, cùng với 2 loại thành viên: thành viên đầy đủ và thành viên quan sát.

Trong điều khiển phân tán các thành viên tham gia hệ thống được tổ chức theo sơ đồ phân lớp, tức là một nhóm các thành viên mức thấp nhất được điều khiển bởi 1 chủ tịch, sau đó các chủ tịch này lại được điều khiển bởi 1 chủ tịch cấp cao hơn. Và cứ tiếp tục như vậy.

Hạ tầng cơ sở và giao thức sử dụng

Hội nghị video trên cơ sở IP có thể triển khai trên mạng LAN hoặc WAN, sử dụng chuẩn H323. Có thể nói chuẩn H323 chính là số hóa quy trình hội nghị viễn thông.

Trong chuẩn H323 có các chuẩn về dữ liệu: mã hóa và giải mã video: H261, H263; mã hóa và giải mã audio: G711, G722, G723.1, G728; mã hóa và giải mã văn bản: T.120; các giao thức điều khiển: H.225.0 điều khiển cuộc gọi, H.245 điều khiển hệ thống...

II.4. So sánh dịch vụ video chat và dịch vụ hội nghị video

	Video chat	Hội nghị video
Chức năng	Thiết lập kết nối ngang hàng	Thiết lập kết nối đa điểm Điều khiển phiên làm việc động Cung cấp dịch vụ thực mụ hội nghị (tức là dịch vụ cung cấp danh sách các tài nguyên chia sẻ, cho phép người sử dụng có thể dễ dàng lựa chọn và tải về những dữ liệu họ cần). Điều độ hội nghị tự động. Đồng bộ giữa âm thanh và hình ảnh.
Các thành phần hệ thống		So với video chat có thêm thành phần bắt âm thanh, mã hóa, giải mã âm thanh.
Mô hình hoạt động	Không có điều khiển tương tác	Có điều khiển tương tác.
Hạ tầng cơ	Ip Lan, Wan,	Ip Lan, Wan

sở	Internet	
Giao thức		So với video chat có thêm các giao thức điều khiển.

TÌM HIỂU GIAO THỨC TRUYỀN ÂM THANH CỦA YAHOO VOICE , SKYPE, ĐIỆN THOẠI 171

1. Dịch vụ Yahoo voice :

Yahoo Voice được tích hợp trên Yahoo messenger một phần mềm chat miễn phí được đông đảo người dùng trên thế giới chấp nhận nên so với các dịch vụ voice khác nó cũng có phần chiếm ưu thế trong quá trình quảng bá. Hiện tại Yahoo cũng như một số hãng khác miễn phí đàm thoại PC to PC.

+Các kịch bản có thể đàm thoại được của yahoo voice chat là:

PC-PC, PC - Phone hoặc là Phone – PC.

+Chuẩn Codec dữ liệu :iLBC (internet low bit rate), không sử dụng thuật encryption.

+Giao thức sử dụng : Sip (có TLS) và RTP (truyền dòng đa phương tiện).

Sip : Session Initiation Protocol là bộ giao thức điều khiển phiên đưa ra các phương thức để thực hiện Voip thuần túy bằng phần mềm .

+Yahoo Voice chỉ chạy được trên Window tức là không độc lập với hệ điều hành.

+Chất lượng âm thanh thoại của Yahoo tương đối tốt.

+Mức độ ổn định không cao.

+Giá thành của dịch vụ PC to Phone nếu cùng một chất lượng cao gấp 2 lần Skype.

+Vì xây dựng trên giao thức Sip và RTP nên gặp vấn đề khi vượt qua NAT và firewall.

2. Skype :

Skype hiện nay là một dịch vụ thoại có thị phần lớn nhất thế giới và nổi tiếng về độ ổn định khi thực hiện đàm thoại kể cả loại hình miễn phí PC- PC lẫn 2 loại hình có phí còn lại.

+Các kịch bản đàm thoại: Có 3 loại hình dịch vụ PC-PC (miễn phí), PC-Phone (2chiều), Phone – Phone (2chiều).

+Chuẩn CODECS và Encryption: CODECS: SVOPC (tần số lấy mẫu 16Khz tốc độ truyền tối thiểu 20kbps-chất lượng khá tốt). iLBC, iSAC (internet speech audio codec). G729 (8kbps).

+Giao thức dựa trên giao thức của mạng ngang hàng P2P (Peer to Peer) (nó sử dụng một mạng P2P che phủ-overlay).

+Tính tương thích cao có các phiên bản khác nhau dành cho Linux, Mac, Win.

+Chất lượng ổn định, được ưa dùng hơn yahoo voice.

+Giá thành thấp: cùng một chất lượng rẻ hơn 2 lần so với Yahoo voice.

+Không gặp trở ngại khi xuyên qua NAT và tường lửa.

+Tuy nhiên vấn đề bảo mật của Skype có vấn đề.

+Chất lượng thoại của Skype tốt và không yêu cầu máy người sử dụng có cấu hình cao. Khi máy người dùng cấu hình thấp thì nó vượt xa so với Yahoo voice về chất lượng thoại.

+Thị phần của Skype: Trên toàn thế giới rất lớn, PC to PC không bằng Yahoo Voice xong dịch vụ Phone call (Phone-PC và Phone-Phone), những dịch vụ trả tiền thì có thị phần lớn nhất thế giới.

3.Điện thoại 171 ở Việt Nam:

Dịch vụ điện thoại đường dài trong nước và quốc tế sử dụng giao thức IP gọi tắt là dịch vụ điện thoại "gọi 171". Đây là dịch vụ viễn thông (bao gồm điện thoại và fax) trên băng tần thoại cơ bản sử dụng công nghệ nén và giải nén tín hiệu để truyền đưa thông tin dưới dạng gói sử dụng giao thức Internet do Tổng Công ty Bưu chính - Viễn thông Việt Nam thiết lập, quản lý, khai thác, kinh doanh và phát triển đáp ứng nhu cầu ngày càng đa dạng của người sử dụng.

+Kịch bản đối thoại: Phone to Phone có cước phí nhỏ hơn điện thoại truyền thống.

+Chuẩn CODECS: G.723, G.729.

+Giao thức H323.

+Chất lượng tương đối tốt đảm bảo đàm thoại.