(12) **BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ**

(19) **Cộng hòa xã hội chủ nghĩa Việt Nam (VN)**
CỤC SỞ HỮU TRÍ TUỆ

(11) **1-0021696**

(51)[7] **H04L 29/08**, 29/06

(13) **B**

(21) 1-2015-05010
(22) 12.06.2014
(86) PCT/EP2014/062220 12.06.2014
(87) WO2014/206749 31.12.2014
(30) 13305910.5 28.06.2013 EP
(45) 25.09.2019 378
(43) 25.04.2016 337
(73) THOMSON LICENSING (FR)
1-5 rue Jeanne d'Arc, F-92130 Issy-les-Moulineaux, France
(72) TAIBI, Charline (FR), HOUDAILLE, Remi (FR), GOUACHE, Stephane (FR)
(74) Công ty Luật TNHH T&G (TGVN)

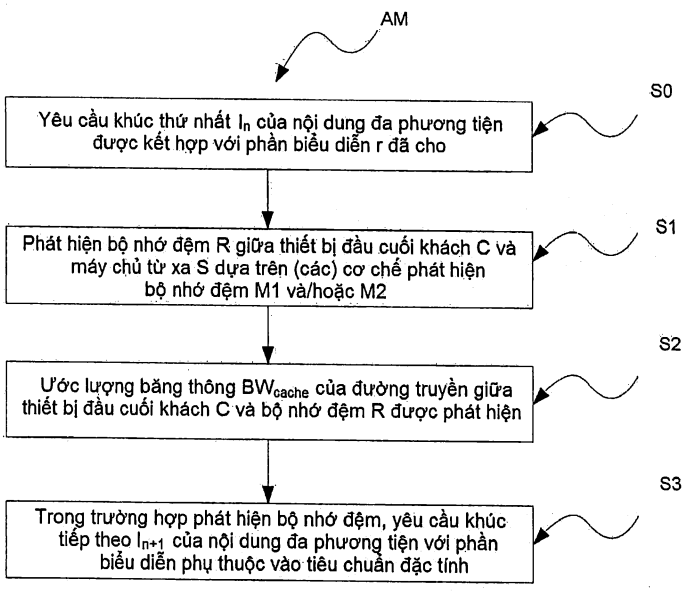
(54) PHƯƠNG PHÁP LÀM THÍCH ỨNG HÀNH VI TẢI XUỐNG CỦA THIẾT BỊ ĐẦU CUỐI KHÁCH ĐƯỢC CẤU HÌNH ĐỂ NHẬN NỘI DUNG ĐA PHƯƠNG TIỆN, VÀ THIẾT BỊ ĐẦU CUỐI

(57) Sáng chế đề xuất phương pháp làm thích ứng hành vi tải xuống của thiết bị đầu cuối khách được cấu hình để nhận nội dung đa phương tiện từ ít nhất một máy chủ, ít nhất một phần biểu diễn của nội dung đa phương tiện này là sẵn có, trong đó phương pháp này bao gồm các bước:

yêu cầu (S0) phần thứ nhất của nội dung đa phương tiện với phần biểu diễn đã cho;

phát hiện (S1) nếu bộ nhớ đệm được đặt dọc theo đường truyền giữa thiết bị đầu cuối khách và máy chủ, dựa trên yêu cầu của phần thứ nhất;

trong trường hợp (S3) bộ nhớ đệm được phát hiện, yêu cầu phần thứ hai của nội dung đa phương tiện với phần biểu diễn phụ thuộc vào ít nhất một tiêu chuẩn đặc tính.



LĨNH VỰC KỸ THUẬT ĐƯỢC ĐỀ CẬP

Sáng chế đề cập chung đến lĩnh vực kỹ thuật tạo dòng thích ứng qua, ví dụ nhưng không chỉ là HTTP (HyperText Transfer Protocol - Giao thức truyền siêu văn bản) và, cụ thể, sáng chế đề cập đến phương pháp làm thích ứng hành vi tải xuống của thiết bị đầu cuối khách được cấu hình để nhận nội dung đa phương tiện từ một hoặc một số máy chủ.

TÌNH TRẠNG KỸ THUẬT CỦA SÁNG CHẾ

Phần này nhằm giới thiệu với người đọc các khía cạnh khác nhau của tình trạng kỹ thuật, mà có thể liên quan đến các khía cạnh khác nhau của sáng chế mà được mô tả và/hoặc được yêu cầu bảo hộ sau đây. Sự đề cập này được tin là hữu ích trong việc cung cấp thông tin tình trạng kỹ thuật cho người đọc để giúp hiểu rõ hơn về các khía cạnh khác nhau của sáng chế. Do đó, cần hiểu là sự trình bày này cần được đọc với ý nghĩa như vậy, và không phải là sự thừa nhận tình trạng kỹ thuật.

Việc tạo dòng thích ứng qua HTTP đang nhanh chóng trở thành kỹ thuật chính để phân phối nội dung đa phương tiện. Trong số các giao thức tạo dòng thích ứng HTTP đã được sử dụng, các giao thức nổi bật nhất là HLS (HTTP Live Streaming - Tạo dòng trực tiếp HTTP) của Apple, SSS (Silverlight Smooth Streaming - Tạo dòng êm ánh sáng bạc) của Microsoft, ADS (Adobe Dynamic Streaming - Tạo dòng động Adobe) của Adobe và DASH (Dynamic Adaptive Streaming over HTTP - Tạo dòng thích ứng động qua HTTP) được phát triển bởi 3GPP trong nhóm SA4.

Khi thiết bị đầu cuối khách muốn phát nội dung nghe nhìn (hay nội dung A/V) trong khi tạo dòng thích ứng, trước tiên nó phải lấy tệp mô tả cách thức có thể thu được nội dung A/V này. Điều này thường được thực hiện thông qua giao thức HTTP nhờ lấy tệp mô tả, gọi là bản kê, từ URL (Uniform Resource Locator - Phần tử định vị tài nguyên đồng nhất), nhưng cũng có thể thu được bởi phương tiện khác (ví dụ truyền phát rộng, thư điện tử, SMS và dạng tương tự). Về cơ bản, bản kê liệt kê các phần biểu

diễn sẵn có của nội dung A/V này (về tốc độ bit, độ phân giải và các thuộc tính khác). Bản kê này được tạo ra trước và được phân phối cho thiết bị đầu cuối khách bởi, ví dụ, máy chủ từ xa.

Thực vậy, dòng dữ liệu tương ứng với nội dung A/V với các chất lượng khác nhau sẵn có trên máy chủ HTTP. Chất lượng cao nhất được kết hợp với tốc độ bit cao, chất lượng thấp nhất được kết hợp với tốc độ bit thấp. Điều này cho phép phân phối với nhiều thiết bị đầu cuối khác nhau mà có thể chịu các điều kiện mạng biến đổi cao.

Toàn bộ dòng dữ liệu này được phân chia thành các khúc mà được tạo ra sao cho thiết bị đầu cuối khách có thể chuyển êm từ một mức chất lượng thành một mức khác giữa hai khúc. Kết quả là, chất lượng viđeo có thể biến đổi trong khi phát nhưng ít khi bị gián đoạn (còn được gọi là treo).

Tùy thuộc vào giao thức, bản kê có thể có các định dạng khác nhau. Đối với giao thức HLS của Apple, nó là danh sách phát M3U8, được gọi là "danh sách chủ". Mỗi phần tử của danh sách phát này là một danh sách phát khác, một cho mỗi phần biểu diễn. Theo các giao thức khác (ví dụ DASH), bản kê được tạo ra từ một hoặc nhiều tệp XML mô tả tất cả các phần biểu diễn một cách lần lượt. Trong trường hợp bất kỳ, việc tạo ra bản kê đơn giản như tạo ra tệp văn bản và ghi văn bản theo ngữ pháp tất định.

Đã biết rõ là, theo băng thông sẵn có của nó, thiết bị đầu cuối khách chọn phần biểu diễn tốt nhất ở thời điểm đã cho để tối ưu hóa sự cân bằng giữa chất lượng (ví dụ chất lượng viđeo) và độ vững chắc đối với các sự biến đổi mạng. Băng thông sẵn có được xác định một cách động, ở mọi khúc nhận được. Thực vậy, RTT (Round Trip Time - Thời gian đi về) được xác định giữa lúc phát yêu cầu HTTP cho khúc đã cho và lúc nhận sự đáp ứng HTTP tương ứng (sau đây được gọi là HTTP RTT) thường được đo và được sử dụng để ước lượng băng thông sẵn có dọc theo đường truyền.

Tốc độ nhận ở phía máy khách biến đổi theo thời gian khi tải khúc xuống. Ở thời điểm bắt đầu, thiết bị đầu cuối khách phát yêu cầu HTTP cho khúc. Trước tiên có khoảng thời gian "nghỉ" tương ứng với HTTP RTT của yêu cầu HTTP đó. Sau đó, nhận được các gói của khúc. Các gói này đạt tới tốc độ đỉnh của kết nối. Cuối cùng, tốc độ nhận lại hạ thành zêrô khi hoàn thành việc tải xuống của khúc.

Thiết bị đầu cuối khách theo đó có thể ước lượng cả HTTP RTT của yêu cầu HTTP và băng thông đỉnh sẵn có, và sau đó sử dụng các trị số được ước lượng này để xác định kích thước khúc tối đa có thể được yêu cầu với xác suất cao là được nhận trong khoảng thời gian của một khúc.

Như đã được biết đến, các thiết bị đầu cuối khách thường lấy trung bình ước lượng băng thông nhờ công thức sau đây:

$$BW_n = \alpha BW_{n-1} + (1-\alpha)D_n$$

trong đó:

– BWn là băng thông trung bình cho khúc n, được sử dụng cho yêu cầu khúc n+1 tiếp theo,

– Dn là tốc độ dữ liệu nhận được tức thời của khúc n (giữa lúc bắt đầu và kết thúc của việc truyền khúc n),

– α sao cho $0 \le \alpha \le 1$.

Hơn nữa, các thiết bị đầu cuối khách còn sử dụng một số vùng đệm để bảo vệ đối với việc đột ngột thiếu băng thông. Để điền vùng đệm, các thiết bị đầu cuối này yêu cầu các khúc đủ nhỏ để được nhận trong thời gian ngắn hơn so với khoảng thời gian của khúc, hỏi khúc tiếp theo ngay khi đã nhận được khúc trước. Khi vùng đệm ở kích thước bình thường của nó, thiết bị đầu cuối khách thử tải các khúc mà vừa với khoảng thời gian của khúc. Nếu một số khúc tải quá chậm, thì vùng đệm được dùng và thiết bị đầu cuối khách sẽ thử điền nó một lần nữa bởi các khúc tiếp theo.

Khi bộ nhớ đệm (cache) được đặt dọc theo đường truyền giữa thiết bị đầu cuối khách và máy chủ từ xa mà thường xuất hiện, một khúc có thể đã được lưu trữ trong bộ nhớ đệm này, trong trường hợp một máy khách khác trước đó đã yêu cầu cùng một khúc với cùng phần biểu diễn hoặc trong trường hợp CDN (Content Delivery Network - Mạng phân phối nội dung) đã cung cấp khúc trong bộ nhớ đệm.

Theo đó, sự đáp ứng đối với yêu cầu HTTP cho khúc đã cho này nhanh hơn so với nếu khúc đến từ máy chủ từ xa. HTTP RTT của yêu cầu HTTP giữa thiết bị đầu cuối khách và bộ nhớ đệm có thể nhỏ hơn nhiều so với HTTP RTT giữa thiết bị đầu cuối khách và máy chủ từ xa, do đường truyền ngắn hơn.

Ngoài ra, trong trường hợp có mặt bộ nhớ đệm dọc theo đường truyền (khúc được yêu cầu đang được lưu trữ trong bộ nhớ đệm), tốc độ đỉnh có thể tốt hơn, đặc biệt là khi có sự tắc nghẽn trên đường truyền đó, nằm giữa bộ nhớ đệm và máy chủ từ xa.

Do thiết bị đầu cuối khách thường không phân biệt các sự đáp lại được gửi bởi máy chủ từ xa hoặc bởi bộ nhớ đệm trung gian, nên nó dịch sai sự biến đổi băng thông là sự biến đổi của các điều kiện mạng từ đầu cuối đến đầu cuối, trong khi thực ra nó đang quan sát sự chuyển đường truyền từ đường "thiết bị đầu cuối khách đến máy chủ" thành đường "thiết bị đầu cuối khách đến bộ nhớ đệm".

Theo đó, sự ước lượng băng thông được thực hiện bởi thiết bị đầu cuối khách được ước lượng quá cao và không phản ánh một cách chính xác các đặc trưng đường truyền từ đầu cuối đến đầu cuối như dự tính.

Sự ước lượng quá cao này thường dẫn đến trải nghiệm kém tốt đối với người dùng đầu cuối. Thực vậy, nếu băng thông được ước lượng cao hơn so với dự tính, thiết bị đầu cuối khách tạo dòng thích ứng thường yêu cầu khúc từ phần biểu diễn có chất lượng cao hơn (ví dụ tốc độ bit cao hơn). Theo đó, khúc được yêu cầu này có xác suất thấp hơn để nằm trong bộ nhớ đệm (do giả sử là bộ nhớ đệm đã được điền bởi thiết bị đầu cuối khách trước đó đang phát cùng nội dung đa phương tiện ở tốc độ bit không đổi) khi phần biểu diễn thay đổi. Thời gian tải xuống được kết hợp với khúc được yêu cầu này sẽ dài hơn nhiều so với dự tính, dẫn đến khúc được yêu cầu đến quá muộn. Sau đó, thiết bị đầu cuối khách sẽ chuyển lại phần biểu diễn có chất lượng thấp hơn, phần này có khả năng được tìm thấy trong bộ nhớ đệm một lần nữa.

Kết quả là, thiết bị đầu cuối khách đang chuyển qua lại giữa các khúc chất lượng cao và thấp – không ngừng bị gián đoạn do không tìm thấy trong bộ nhớ đệm – điều này triệt tiêu hoàn toàn các lợi ích của việc đệm.

Sáng chế nhằm khắc phục ít nhất là một số trong các vấn đề nêu trên để cải thiện chất lượng trải nghiệm của người dùng đầu cuối.

BẢN CHẤT KỸ THUẬT CỦA SÁNG CHẾ

Sáng chế đề xuất phương pháp làm thích ứng hành vi tải xuống của thiết bị đầu cuối khách được cấu hình để nhận nội dung đa phương tiện từ ít nhất một máy chủ, ít nhất một phần biểu diễn của nội dung đa phương tiện này là sẵn có,

khác biệt ở chỗ phương pháp này bao gồm, tốt hơn nếu ở phía máy khách, các bước:

- yêu cầu phần thứ nhất của nội dung đa phương tiện với phần biểu diễn đã cho;
- phát hiện nếu bộ nhớ đệm được đặt dọc theo đường truyền giữa thiết bị đầu cuối khách và máy chủ, dựa trên yêu cầu của phần thứ nhất;
- trong trường hợp bộ nhớ đệm được phát hiện, yêu cầu phần thứ hai của nội dung đa phương tiện với phần biểu diễn phụ thuộc vào ít nhất một tiêu chuẩn đặc tính.

Theo đó, nhờ sáng chế, việc làm thích ứng hành vi tải xuống của thiết bị đầu cuối khách khi có mặt (các) bộ nhớ đệm trung gian có thể tránh được sự phát lại thất thường và có thể khôi phục khả năng tận dụng số lượng lớn bộ nhớ đệm được triển khai. Thực tế, hầu hết các máy chủ Tạo dòng thích ứng HTTP gửi dữ liệu khúc với các phần đầu "không-đệm" để ngăn việc đệm, để tránh cho các thiết bị đầu cuối khách bị rối trong việc ước lượng băng thông của chúng khi một số khúc được đệm và các khúc khác không được đệm. Sáng chế có thể đem các lợi ích của việc đệm trở lại cho toàn bộ kiến trúc máy khách máy chủ, cải thiện toàn bộ đặc tính mạng.

Tốt hơn nếu, phương pháp này còn bao gồm bước:

- ước lượng băng thông của đường truyền giữa thiết bị đầu cuối khách và bộ nhớ đệm được phát hiện.

Theo tiêu chuẩn đặc tính này, phần thứ hai được yêu cầu của nội dung đa phương tiện có thể được xác định với:

- hoặc cùng phần biểu diễn như phần biểu diễn của phần thứ nhất được lưu trữ trong bộ nhớ đệm được phát hiện này, bất kể kết quả của việc ước lượng băng thông;
- hoặc phần biểu diễn thay thế có tính đến băng thông được ước lượng, phần biểu diễn mới này khác với phần biểu diễn của phần thứ nhất.

Hơn nữa, thuận lợi nếu yêu cầu của phần thứ hai này có thể bao gồm đoạn thông tin mà bộ nhớ đệm được phát hiện có thể hiểu được, sao cho, trong trường hợp phần thứ hai không được lưu trữ trong bộ nhớ đệm được phát hiện, thiết bị đầu cuối khách nhận tin nhắn chỉ định là phần thứ hai không sẵn có từ bộ nhớ đệm này. Ví dụ, đoạn thông tin này có thể là chỉ dẫn "chỉ khi nào được đệm" trong phần đầu điều khiển của yêu cầu HTTP.

Theo một khía cạnh khác, phương pháp này còn có thể bao gồm bước, trong trường hợp việc tải xuống của nội dung đa phương tiện từ bộ nhớ đệm được phát hiện thỏa mãn ít nhất một tiêu chuẩn tải xuống, yêu cầu phần thêm nữa của nội dung đa phương tiện này với phần biểu diễn mới, mà khác với phần biểu diễn của phần thứ nhất.

Theo phương án ưu tiên của sáng chế, bước phát hiện bộ nhớ đệm còn bao gồm xác định thời gian đi về cho yêu cầu thiết lập kết nối từ thiết bị đầu cuối khách đến máy chủ.

Ngoài ra, bước phát hiện bộ nhớ đệm còn có thể bao gồm đo độ trễ nhận giữa lúc phát yêu cầu cho phần thứ nhất của nội dung đa phương tiện và lúc bắt đầu nhận phần thứ nhất được yêu cầu này.

Hơn nữa, bước phát hiện bộ nhớ đệm còn có thể bao gồm so sánh thời gian đi về được xác định của yêu cầu thiết lập kết nối và độ trễ nhận được đo.

Theo một khía cạnh khác của phương án ưu tiên, trong trường hợp độ chênh lệch giữa độ trễ nhận được đo và thời gian đi về được xác định của yêu cầu thiết lập kết nối:

- ít nhất là bằng ngưỡng phát hiện thứ nhất, thì bộ nhớ đệm được phát hiện dọc theo đường truyền giữa thiết bị đầu cuối khách và máy chủ, mà không lưu trữ phần thứ nhất được yêu cầu vào trong bộ nhớ đệm được phát hiện này;
- nếu không thì:
  - hoặc là bộ nhớ đệm có mặt dọc theo đường truyền giữa thiết bị đầu cuối khách và máy chủ, và phần thứ nhất được yêu cầu đến từ bộ nhớ đệm;

- hoặc không có bộ nhớ đệm dọc theo đường truyền giữa thiết bị đầu cuối khách và máy chủ, và phần thứ nhất được yêu cầu đến từ máy chủ.

Theo một biến thể hoặc bổ sung của phương án ưu tiên, bước phát hiện bộ nhớ đệm còn có thể bao gồm:

– đo thời gian đáp ứng giữa lúc phát yêu cầu phản hồi từ thiết bị đầu cuối khách đến máy chủ và lúc nhận sự đáp ứng đối với yêu cầu phản hồi này;

– so sánh thời gian đi về được xác định của yêu cầu thiết lập kết nối với thời gian đáp ứng.

Ngoài ra, bước phát hiện bộ nhớ đệm còn có thể bao gồm so sánh thời gian đáp ứng được đo với độ trễ nhận được đo.

Theo một khía cạnh bổ sung của biến thể hoặc bổ sung này, trong trường hợp độ chênh lệch giữa thời gian đáp ứng được đo và thời gian đi về được xác định của yêu cầu thiết lập kết nối:

– lớn nhất là bằng ngưỡng phát hiện thứ hai, thì không có bộ nhớ đệm nào được phát hiện dọc theo đường truyền giữa thiết bị đầu cuối khách và máy chủ, khúc đến từ máy chủ;

– ít nhất là bằng ngưỡng phát hiện thứ ba, thì bộ nhớ đệm được phát hiện dọc theo đường truyền giữa thiết bị đầu cuối khách và máy chủ, và khúc:

  - được tải vào trong bộ nhớ đệm được phát hiện, trong trường hợp độ chênh lệch giữa thời gian đáp ứng được đo với độ trễ nhận được đo ít nhất là bằng ngưỡng phát hiện thứ tư; hoặc
  - đến từ máy chủ mà xuất hiện ở phía đối diện.

Theo một khía cạnh khác, tiêu chuẩn đặc tính này có thể thuộc về nhóm gồm các tiêu chuẩn bao gồm ít nhất là:

– tiêu chuẩn liên quan đến chất lượng của nội dung đa phương tiện;

– tiêu chuẩn liên quan đến tốc độ tải xuống của nội dung đa phương tiện.

Sáng chế còn đề xuất thiết bị đầu cuối khách được cấu hình để làm thích ứng hành vi tải xuống của nó để nhận nội dung đa phương tiện từ ít nhất một máy chủ, ít

nhất một phần biểu diễn của nội dung đa phương tiện này là sẵn có. Theo sáng chế, thiết bị đầu cuối này bao gồm:

- môđun truyền thông để yêu cầu phần thứ nhất của nội dung đa phương tiện với phần biểu diễn đã cho;
- bộ phát hiện bộ nhớ đệm để phát hiện nếu bộ nhớ đệm được đặt dọc theo đường truyền giữa thiết bị đầu cuối khách và máy chủ, dựa trên yêu cầu của phần thứ nhất;
- môđun quyết định để yêu cầu, trong trường hợp bộ nhớ đệm được phát hiện, phần thứ hai của nội dung đa phương tiện với phần biểu diễn phụ thuộc vào ít nhất một tiêu chuẩn đặc tính.

Ngoài ra, thiết bị đầu cuối này còn có thể bao gồm bộ ước lượng băng thông để ước lượng băng thông của đường truyền giữa thiết bị đầu cuối này và bộ nhớ đệm được phát hiện.

Theo tiêu chuẩn đặc tính này, phần thứ hai được yêu cầu (I_{n+1}) của nội dung đa phương tiện có thể được xác định với:

- hoặc cùng phần biểu diễn như phần biểu diễn của phần thứ nhất được lưu trữ trong bộ nhớ đệm được phát hiện này, bất kể kết quả của việc ước lượng băng thông;
- hoặc phần biểu diễn thay thế có tính đến băng thông được ước lượng, phần biểu diễn mới này khác với phần biểu diễn của phần thứ nhất.

Ngoài ra, yêu cầu của phần thứ hai này có thể bao gồm đoạn thông tin mà bộ nhớ đệm được phát hiện có thể hiểu được, sao cho, trong trường hợp phần thứ hai không được lưu trữ trong bộ nhớ đệm được phát hiện, thiết bị đầu cuối khách nhận tin nhắn chỉ định là phần thứ hai không sẵn có từ bộ nhớ đệm này.

Hơn nữa, thiết bị đầu cuối còn có thể được cấu hình để yêu cầu phần thêm nữa của nội dung đa phương tiện này với phần biểu diễn mới, mà khác với phần biểu diễn của phần thứ nhất, trong trường hợp việc tải xuống của nội dung đa phương tiện từ bộ nhớ đệm được phát hiện thỏa mãn ít nhất một tiêu chuẩn tải xuống. Các khía cạnh nhất định tương ứng về phạm vi với các phương án được đề xuất sẽ được đưa ra sau đây.

Cần hiểu rằng các khía cạnh này được đưa ra chỉ nhằm cung cấp cho người đọc thông tin tóm tắt ngắn gọn về các dạng nhất định mà sáng chế có thể thực hiện và cần hiểu rằng các khía cạnh này không nhằm giới hạn phạm vi bảo hộ của sáng chế. Thực vậy, sáng chế có thể bao trùm nhiều khía cạnh khác nhau mà có thể không được đưa ra sau đây.

MÔ TẢ VẮN TẮT CÁC HÌNH VẼ

Sáng chế sẽ được hiểu rõ hơn và được minh họa bởi phương án sau đây và các ví dụ thực hiện, theo cách không giới hạn, có dựa vào các hình vẽ kèm theo, trong đó:

Fig.1 là sơ đồ về kiến trúc mạng Máy khách-Máy chủ trong đó có thể thực hiện sáng chế;

Fig.2 là sơ đồ khối của một ví dụ về thiết bị đầu cuối khách theo phương án ưu tiên của sáng chế;

Fig.3 là sơ đồ tiến trình mô tả cơ chế phát hiện bộ nhớ đệm thứ nhất được thực hiện bởi thiết bị đầu cuối khách trên Fig.2;

Fig.4A và Fig.4B là các hình vẽ minh họa TCP-RTT không có bộ nhớ đệm (Fig.4A) và có bộ nhớ đệm đặt dọc theo đường truyền (Fig.4B);

Fig.4C, Fig.4D và Fig.4E là các hình vẽ thể hiện HTTP-RTT không có bộ nhớ đệm (Fig.4C) và có bộ nhớ đệm (Fig.4D và Fig.4E) dọc theo đường truyền, khúc đã cho hoặc không được đệm (Fig.4D) hoặc được đệm (Fig.4E);

Fig.5 là sơ đồ tiến trình mô tả cơ chế phát hiện bộ nhớ đệm thứ hai được thực hiện bởi thiết bị đầu cuối khách trên Fig.2;

Fig.6 là sơ đồ tiến trình minh họa phương pháp làm thích ứng hành vi tải xuống được thực hiện bởi thiết bị đầu cuối khách trên Fig.2.

Trên Fig.1 và Fig.2, các khối được biểu diễn là các thực thể chức năng thuần túy, chúng không nhất thiết tương ứng với các thực thể riêng rẽ về vật lý. Cụ thể, chúng có thể được phát triển dưới dạng phần mềm, phần cứng, hoặc được thực hiện trong một hoặc một số mạch tích hợp, bao gồm một hoặc nhiều bộ xử lý.

Ở mọi vị trí có thể, các số chỉ dẫn giống nhau sẽ được sử dụng trên toàn bộ các hình vẽ để chỉ các phần giống nhau hoặc các phần tương tự.

MÔ TẢ CHI TIẾT SÁNG CHẾ

Cần hiểu rằng các hình vẽ và các phần mô tả của sáng chế đã được đơn giản hóa để minh họa các phần tử có liên quan nhằm hiểu rõ sáng chế, trong khi bỏ qua, nhằm mục đích rõ ràng, nhiều phần tử khác được thấy trong các hệ thống và phương pháp phân phối nội dung đa phương tiện số đặc trưng. Tuy nhiên, vì các phần tử này đã được biết rõ trong lĩnh vực, nên đề cập chi tiết đến các phần tử này không được đưa ra ở đây. Phần mô tả này nhắm đến tất cả các thay đổi và biến đổi đã được biết đến đối với các chuyên gia trong lĩnh vực.

Theo phương án ưu tiên, sáng chế được mô tả liên quan đến giao thức tạo dòng thích ứng HTTP. Dĩ nhiên, sáng chế không bị giới hạn vào môi trường cụ thể này và đương nhiên là giao thức tạo dòng khác có thể được xem xét và được thực hiện.

Như được mô tả trên Fig.1, kiến trúc mạng Máy khách-Máy chủ, trong đó có thể thực hiện sáng chế, bao gồm thiết bị đầu cuối khách C, cổng GW và một hoặc nhiều máy chủ HTTP S (chỉ có một máy chủ được biểu diễn trên Fig.1).

Thiết bị đầu cuối khách C – được kết nối với cổng GW thông qua mạng thứ nhất N1 (như mạng gia đình hoặc mạng doanh nghiệp) – muốn kết nối với máy chủ HTTP S thông qua mạng thứ hai N2 (như mạng Internet). Mạng thứ nhất N1 được kết nối với mạng thứ hai N2 nhờ cổng GW.

Máy chủ HTTP S tạo dòng cho các khúc đến thiết bị đầu cuối khách C, theo yêu cầu máy khách, sử dụng giao thức tạo dòng thích ứng HTTP qua một hoặc nhiều kết nối TCP/IP.

Theo phương án ưu tiên như được mô tả trên Fig.2, thiết bị đầu cuối khách C bao gồm ít nhất là:

- giao diện của kết nối 1 (nối dây và/hoặc không dây, ví dụ như Wi-Fi, Ethernet, v.v.) đến mạng thứ nhất N1;
- môđun truyền thông 2 chứa các ngăn xếp giao thức để truyền thông với máy chủ HTTP S. Cụ thể, môđun truyền thông 2 bao gồm ngăn xếp TCP/IP đã được

biết rõ trong lĩnh vực. Đương nhiên, nó có thể là loại mạng và/hoặc phương tiện truyền thông bất kỳ mà cho phép thiết bị đầu cuối khách C truyền thông với máy chủ HTTP S;

- môđun tạo dòng thích ứng 3 mà nhận nội dung đa phương tiện tạo dòng HTTP từ máy chủ HTTP S. Nó liên tục lựa chọn khúc ở tốc độ bit mà khớp tốt hơn với các ràng buộc mạng và các ràng buộc của chính nó;
- máy phát viđeo 4 được làm thích ứng để giải mã và kết xuất nội dung đa phương tiện;
- một hoặc nhiều bộ xử lý 5 để chạy các ứng dụng và các chương trình được lưu trữ trong bộ nhớ bất khả biến của thiết bị đầu cuối khách C;
- phương tiện lưu trữ 6, như bộ nhớ khả biến, để tạo vùng đệm cho các khúc nhận được từ máy chủ HTTP S trước khi truyền chúng đến máy phát viđeo 4;
- buýt trong B1 để kết nối các môđun khác nhau và tất cả các phương tiện đã được biết rõ đối với chuyên gia trong lĩnh vực để thực hiện các chức năng chung của thiết bị đầu cuối khách.

Theo phương án ưu tiên, thiết bị đầu cuối khách C là thiết bị phương tiện di chuyển được, điện thoại di động, máy tính bảng hoặc máy tính xách tay. Dĩ nhiên, thiết bị đầu cuối khách C có thể không bao gồm máy phát viđeo hoàn chỉnh, mà chỉ một số phần tử con như các phần tử để giải dồn kênh và giải mã nội dung phương tiện và có thể dựa vào phương tiện ngoài để hiển thị nội dung được giải mã cho người dùng đầu cuối. Trong trường hợp này, thiết bị đầu cuối khách C là bộ giải mã viđeo có khả năng HAS (HTTP Adaptive Streaming - Tạo dòng thích ứng HTTP), như bộ giải mã tín hiệu truyền hình (set-top box).

Theo sáng chế, thiết bị đầu cuối khách C được cấu hình để làm thích ứng hành vi tải xuống của nó để nhận nội dung đa phương tiện từ máy chủ S.

Để thực hiện điều này, thiết bị đầu cuối khách C còn bao gồm:

- bộ phát hiện bộ nhớ đệm 7 được làm thích ứng để phát hiện bộ nhớ đệm dọc theo đường truyền giữa thiết bị đầu cuối khách C và máy chủ S;

- bộ ước lượng băng thông 8 được cấu hình để xác định băng thông của đường truyền giữa thiết bị đầu cuối khách C và bộ nhớ đệm R được phát hiện và/hoặc băng thông của đường truyền giữa thiết bị đầu cuối khách C và máy chủ S;
- môđun quyết định 9 được cấu hình để yêu cầu các khúc của nội dung đa phương tiện này theo tiêu chuẩn đặc tính. Theo một biến thể, môđun quyết định 9 có thể được tích hợp hoặc là vào trong môđun truyền thông 2 hoặc vào trong môđun tạo dòng thích ứng 3.

Cụ thể, theo phương án ưu tiên này, bộ phát hiện bộ nhớ đệm 7 của thiết bị đầu cuối khách C thực hiện cơ chế thứ nhất M1 để phát hiện bộ nhớ đệm R giữa thiết bị đầu cuối khách C và máy chủ S, như được biểu diễn trên Fig.3.

Cơ chế phát hiện bộ nhớ đệm thứ nhất M1 bao gồm các bước sau:

- xác định (bước E0) thời gian đi về cho yêu cầu thiết lập kết nối từ thiết bị đầu cuối khách C đến máy chủ S (được gọi là TCP-RTT như được minh họa trên Fig.4A và Fig.4B);
- đo (bước E1) độ trễ nhận $Delay_{Rx}$ giữa lúc phát yêu cầu cho khúc I_n đã cho của nội dung đa phương tiện, với phần biểu diễn r đã cho, và lúc bắt đầu nhận khúc I_n được yêu cầu này;
- so sánh (bước E2) thời gian đi về được xác định của yêu cầu thiết lập kết nối TCP-RTT với độ trễ nhận $Delay_{Rx}$ được đo.

Khi khúc I_n được yêu cầu nằm trong bộ nhớ đệm R (để được phát hiện), độ trễ nhận $Delay_{Rx}$ có thể bao gồm:

- thời gian đi về HTTP-RTT để yêu cầu khúc I_n đã cho giữa thiết bị đầu cuối khách C và bộ nhớ đệm R (như được mô tả trên Fig.4E);
- thời gian cho bộ nhớ đệm R kiểm tra tính sẵn có của khúc I_n được yêu cầu;
- thời gian để truy xuất dữ liệu từ bộ nhớ đệm R;
- thời gian di chuyển của gói dữ liệu thứ nhất của khúc I_n.

Tất cả các thời gian này được dự tính là ngắn khi bộ nhớ đệm được đặt một cách thích hợp.

Khi khúc I_n được yêu cầu không được tải vào trong bộ nhớ đệm R, độ trễ nhận $Delay_{Rx}$ bao gồm độ trễ bổ sung được tạo ra từ:

- thời gian đi về HTTP-RTT để yêu cầu khúc I_n đã cho giữa bộ nhớ đệm R và máy chủ S (như được thể hiện trên Fig.4D);
- thời gian cho máy chủ S truy xuất khúc I_n;
- thời gian di chuyển của gói dữ liệu thứ nhất từ máy chủ S đến bộ nhớ đệm R.

Độ trễ bổ sung này thường dài hơn nhiều so với độ trễ trước vì:

- số lượng bước nhảy đến máy chủ S có thể lớn, dẫn đến RTT lớn, mỗi nút định tuyến có khả năng bị tắc nghẽn;
- băng thông TCP có thể giảm đột ngột nếu xảy ra tắc nghẽn ở vị trí nào đó trên đường;
- máy chủ S có nhiều máy khách và có thể bị quá tải.

Trong trường hợp độ chênh lệch giữa độ trễ nhận $Delay_{Rx}$ được đo và thời gian đi về được xác định của yêu cầu thiết lập kết nối TCP-RTT ít nhất là bằng ngưỡng phát hiện thứ nhất Th1 khác 0 (cụ thể là $Delay_{Rx} - TCPRTT \geq Th1 \neq 0$), thì bộ nhớ đệm R được phát hiện dọc theo đường truyền giữa thiết bị đầu cuối khách C và máy chủ S. Khúc I_n đã cho được yêu cầu không được lưu trữ trong bộ nhớ đệm R được phát hiện (như được thể hiện trên Fig.4D).

Trong trường hợp độ chênh lệch giữa độ trễ nhận $Delay_{Rx}$ được đo và thời gian đi về được xác định của yêu cầu thiết lập kết nối TCP-RTT nhỏ hơn ngưỡng phát hiện thứ nhất (cụ thể là $Delay_{Rx} - TCPRTT < Th1$):

- hoặc là bộ nhớ đệm R có mặt dọc theo đường truyền giữa thiết bị đầu cuối khách C và máy chủ S, và khúc I_n đã cho được yêu cầu đến từ bộ nhớ đệm R (như được thể hiện trên Fig.4E);
- hoặc không có bộ nhớ đệm dọc theo đường truyền giữa thiết bị đầu cuối khách C và máy chủ S, và khúc I_n đã cho được yêu cầu đến từ máy chủ S (như được thể hiện trên Fig.4C).

Nhờ so sánh $Delays_{Rx}$ tiếp theo, bộ phát hiện bộ nhớ đệm 7 có thể biết nếu khúc được yêu cầu đến từ cùng một nguồn như khúc nhận được trước đó, bất kể nó từ máy chủ S hoặc từ bộ nhớ đệm R.

Cũng cần chú ý là, đối với khúc đã cho, khi đã phát hiện được bộ nhớ đệm dọc theo đường truyền, thông tin liên quan đến sự có mặt của bộ nhớ đệm này có thể được tái sử dụng sau này.

Hơn nữa, theo một biến thể, bộ phát hiện bộ nhớ đệm 7 của thiết bị đầu cuối khách C có thể thực hiện cơ chế thứ hai M2 để phát hiện bộ nhớ đệm R giữa thiết bị đầu cuối khách C và máy chủ S.

Như được thể hiện trên Fig.5, cơ chế thứ hai M2 này bao gồm các bước E0 và E1 của cơ chế thứ nhất M1 và các bước khác là:

- đo (bước E3) thời gian đáp ứng T_{Rx} giữa lúc phát yêu cầu phản hồi Giao thức tin nhắn điều khiển Internet (còn gọi là phản hồi ICMP (Internet Control Message Protocol – Giao thức tin nhắn điều khiển Internet)) từ thiết bị đầu cuối khách C đến máy chủ S và lúc nhận sự đáp ứng đối với yêu cầu phản hồi ICMP này;
- so sánh (bước E4) thời gian đi về TCP-RTT được xác định của yêu cầu thiết lập kết nối với thời gian đáp ứng T_{Rx};
- so sánh (bước E5) thời gian đáp ứng T_{Rx} được đo với độ trễ nhận $Delay_{Rx}$ được đo.

Khi độ chênh lệch giữa thời gian đáp ứng T_{Rx} được đo và thời gian đi về TCP-RTT được xác định của yêu cầu thiết lập kết nối lớn nhất là bằng ngưỡng phát hiện thứ hai Th2 (ví dụ với trị số gần với zêrô), cụ thể là $|T_{RX} - TCPRTT| \le Th2 \approx 0$, thì không có bộ nhớ đệm nào được phát hiện dọc theo đường truyền giữa thiết bị đầu cuối khách C và máy chủ S, khúc I_n đến từ máy chủ S.

Ngược lại, khi độ chênh lệch giữa thời gian đáp ứng T_{Rx} được đo và thời gian đi về TCP-RTT được xác định của yêu cầu thiết lập kết nối ít nhất là bằng ngưỡng phát hiện thứ ba Th3 khác 0, cụ thể là $T_{Rx} - TCPRTT \ge Th3 \ne 0$, thì bộ nhớ đệm R được phát hiện dọc theo đường truyền giữa thiết bị đầu cuối khách C và máy chủ S, và khúc I_n:

– được tải vào trong bộ nhớ đệm R được phát hiện, trong trường hợp độ chênh lệch giữa thời gian đáp ứng T_{Rx} được đo với độ trễ nhận $Delay_{Rx}$ được đo ít nhất là bằng ngưỡng phát hiện thứ tư (ví dụ là một nửa thời gian đáp ứng T_{Rx}), cụ thể là $T_{Rx} - Delay_{Rx} \geq Th4 = \frac{T_{Rx}}{2}$; hoặc

– đến từ máy chủ S mà xuất hiện ở phía đối diện.

Cơ chế thứ hai M2 liên quan đến phép đo rõ ràng của "khoảng cách" đến máy chủ từ xa S nhờ gửi yêu cầu phản hồi ICMP đến máy chủ S này và nhờ đo thời gian đáp ứng T_{Rx}, do bộ nhớ đệm R sẽ không làm gì với yêu cầu ICMP này.

Nếu TCP-RTT và T_{Rx} là gần nhau, thì nó có nghĩa là chỉ có một kết nối TCP trực tiếp đến máy chủ S (xem Fig.4A). Nếu bộ nhớ đệm R thiết lập kết nối TCP kép (xem Fig.4B), thì RTT đầy đủ của phản hồi ICMP (cụ thể là T_{Rx}) lớn hơn nhiều.

Ngoài ra, có thể hiểu rõ rằng có thể được kết hợp cả hai cơ chế để đạt được độ tin cậy tốt hơn trong việc phát hiện bộ nhớ đệm. Một số phép đo của các thông số nêu trên có thể được thực hiện để giới hạn rủi ro về các lỗi do, ví dụ, sự tắc nghẽn mạng tạm thời.

Ngoài ra, bộ ước lượng băng thông 8 của thiết bị đầu cuối khách C được cấu hình để ước lượng băng thông sẵn có của đường truyền nhờ công thức nêu trên:

$$BW_n = \alpha BW_{n-1} + (1-\alpha)D_n$$

Khi thiết bị đầu cuối khách C biết liệu khúc I_n được nhận thông qua bộ nhớ đệm R hoặc từ máy chủ từ xa S nhờ bộ phát hiện bộ nhớ đệm 7, bộ ước lượng băng thông 8 được cấu hình để giữ các trị số riêng rẽ của độ trễ nhận $Delay_{Rx}$ và tốc độ đỉnh cho cả hai tình huống (khúc I_n được đệm, khúc I_n đến từ máy chủ S) và để duy trì hai trị số khác nhau để ước lượng băng thông:

– một trị số BW_{server} cho đường truyền "thiết bị đầu cuối khách C đến máy chủ S";

– một trị số BW_{cache} cho đường truyền "thiết bị đầu cuối khách C đến bộ nhớ đệm R được phát hiện".

Việc ước lượng băng thông có thể được cập nhật khi nhận được thông tin phát hiện từ bộ phát hiện bộ nhớ đệm 7 của thiết bị đầu cuối khách C.

Theo một khía cạnh khác, do tốc độ đỉnh chịu ảnh hưởng chủ yếu bởi liên kết truy nhập "dặm cuối" (last mile) và/hoặc bởi việc sử dụng liên kết thường trú không dây, nếu thiết bị đầu cuối khách C quan sát được độ giảm tốc độ đỉnh, thì nó có thể giả sử độ giảm tương tự cho cả hai tình huống, sao cho cả hai việc ước lượng có thể được cập nhật để tính đến độ giảm tốc độ đỉnh.

Theo phương án ưu tiên, trong trường hợp bộ nhớ đệm R đã được phát hiện, môđun quyết định 9 thực hiện quy trình quyết định sau đây để quyết định – phụ thuộc vào tiêu chuẩn đặc tính – yêu cầu khúc tiếp theo I_{n+1} với:

- cùng phần biểu diễn r như phần của khúc I_n được lưu trữ trong bộ nhớ đệm R, để tối đa hóa xác suất ở lại trong bộ nhớ đệm R này, bất kể kết quả của việc ước lượng băng thông. Trong trường hợp này, môđun quyết định 9 có thể yêu cầu phần biểu diễn với tốc độ bit được kết hợp mà không phải là gần nhất dưới ước lượng băng thông. Theo đó độ vững chắc được ưu tiên hơn – thông qua đệm – so với chất lượng tối đa; hoặc
- phần biểu diễn thay thế r' (khác với phần của khúc I_n) tính đến ước lượng băng thông (ví dụ BW_{cache} trong trường hợp khúc I_n được đệm, BW_{server} trong trường hợp khúc I_n đến từ máy chủ S). Thuận lợi nếu phần biểu diễn thay thế r' này có thể phụ thuộc vào ước lượng băng thông. Trong trường hợp phần biểu diễn thay thế r' được xác định với tốc độ bit cao hơn so với phần biểu diễn r hiện thời của khúc I_n, thì chất lượng được ưu tiên hơn so với độ vững chắc.

Theo một khía cạnh khác của phương án ưu tiên, yêu cầu của khúc tiếp theo I_{n+1} với cùng phần biểu diễn r hoặc với phần biểu diễn thay thế r' thuận lợi nếu có thể bao gồm đoạn thông tin mà có thể hiểu được bởi bộ nhớ đệm R được phát hiện (ví dụ phần đầu HTTP "chỉ khi nào được đệm"), sao cho, trong trường hợp khúc tiếp theo I_{n+1} không được lưu trữ trong bộ nhớ đệm R được phát hiện, thiết bị đầu cuối khách C nhận tin nhắn chỉ định là khúc tiếp theo I_{n+1} này không sẵn có trong bộ nhớ đệm R. Điều này cho phép phạt độ trễ giới hạn để kiểm tra tính sẵn có trong bộ nhớ đệm R của

khúc thêm nữa. Nếu phần biểu diễn thay thế r' không sẵn có trong bộ nhớ đệm R, thì thiết bị đầu cuối khách C có thể một lần nữa yêu cầu khúc I_{n+1} với phần biểu diễn r.

Theo một ví dụ, tiêu chuẩn đặc tính có thể được lựa chọn bởi người dùng đầu cuối của thiết bị đầu cuối khách C, để lựa chọn các sự ưu tiên của họ, ví dụ chất lượng của nội dung đa phương tiện, tốc độ tải xuống, v.v.. Theo một biến thể, tiêu chuẩn đặc tính này có thể được xác định một cách tự động liên quan đến loại nội dung đa phương tiện (ví dụ sự kiện thể thao, phim ảnh, phim tài liệu, v.v.). Rõ ràng là, theo một biến thể khác, một hoặc nhiều tiêu chuẩn bổ sung có thể được sử dụng.

Theo một khía cạnh khác của phương án ưu tiên, khi việc tải xuống của nội dung đa phương tiện từ bộ nhớ đệm R được phát hiện thỏa mãn tiêu chuẩn tải xuống (ví dụ việc tải xuống từ bộ nhớ đệm là đủ nhanh để vùng đệm 6 của thiết bị đầu cuối khách C bị đầy), môđun quyết định 9 có thể yêu cầu khúc I_k thêm nữa của nội dung đa phương tiện với phần biểu diễn mới r'' – phần này sẽ khác với phần biểu diễn r của khúc I_n – để cố gắng thay đổi phần biểu diễn (ví dụ chất lượng cao hơn) và xác định liệu phần biểu diễn mới r'' này có thể chịu được (có nghĩa là khúc I_k tiếp theo sẽ đến đúng lúc và ước lượng băng thông tiếp theo cho phép tiếp tục với phần biểu diễn mới r'' này) hay không. Trong trường hợp phần biểu diễn mới r'' không thể chịu được, môđun phát hiện 9 chuyển trở lại về phần biểu diễn r được đệm trước đó.

Theo một khía cạnh bổ sung của phương án ưu tiên, danh sách cờ cho tất cả các phần biểu diễn của nội dung đa phương tiện có thể được lưu trữ trong thiết bị đầu cuối khách C, danh sách cờ này chỉ thị liệu phần biểu diễn đã được phát hiện trước đó trong bộ nhớ đệm R hay không. Khi môđun phát hiện 9 muốn thay đổi phần biểu diễn r hiện thời, danh sách cờ này có thể cung cấp phần chỉ thị về khả năng tìm thấy nó trong bộ nhớ đệm R được phát hiện.

Có thể hiểu rõ rằng phụ thuộc vào tỷ lệ giữa ước lượng băng thông có (BW_{cache}) hoặc không có bộ nhớ đệm (BW_{server}), việc lựa chọn tốc độ bit ngay dưới mức hiện thời có thể dẫn đến các điều kiện nhận kém hơn so với các mức hiện thời khi khúc I_n hiện thời nằm trong bộ nhớ đệm R và khúc tiếp theo I_{n+1} thì không. Môđun quyết định 9 theo đó có thể quyết định lựa chọn tốc độ bit thấp hơn nhiều để bù cho sự bất lợi của việc không có bộ đệm.

Theo sáng chế, và như được thể hiện trên Fig.6, thiết bị đầu cuối khách C được cấu hình để làm thích ứng hành vi tải xuống của nó để nhận nội dung đa phương tiện đã cho từ mạng N2. Cụ thể, thiết bị đầu cuối khách C có thể thực hiện phương pháp thích ứng AM sau đây bao gồm các bước:

- yêu cầu (bước S0) khúc thứ nhất I_n của nội dung đa phương tiện với phần biểu diễn r đã cho;
- phát hiện (bước S1) bộ nhớ đệm R giữa thiết bị đầu cuối khách C và máy chủ từ xa S bao gồm nội dung đa phương tiện đã cho dựa trên (các) cơ chế phát hiện bộ nhớ đệm M1 và/hoặc M2;
- ước lượng (bước S2) băng thông BW_{cache} của đường truyền giữa thiết bị đầu cuối khách C và bộ nhớ đệm R được phát hiện;
- trong trường hợp bộ nhớ đệm R đã được phát hiện trong bước S1, yêu cầu (bước S3) khúc tiếp theo I_{n+1} của nội dung đa phương tiện với phần biểu diễn r', r' phụ thuộc vào tiêu chuẩn đặc tính, theo quy trình quyết định như được mô tả ở trên.

Nhờ sáng chế, hành vi của thiết bị đầu cuối khách Tạo dòng thích ứng HTTP có thể được làm thích ứng để phát hiện sự có mặt của bộ nhớ đệm dọc theo đường truyền và để điều chỉnh quy trình quyết định của nó theo đó. QoE (Quality of Experience - Chất lượng trải nghiệm) của người dùng đầu cuối có thể được cải thiện.

Thật vậy, khi khúc I_n được yêu cầu (khúc n của phần biểu diễn r) đã có mặt trong bộ nhớ đệm R được phát hiện, có xác suất cao là có khúc tiếp theo ở cùng tốc độ (khúc I_{n+1} của phần biểu diễn r) cũng được tải vào trong bộ nhớ đệm R. Trong một số trường hợp, có thể thuận lợi nếu giữ cùng phần biểu diễn r và yêu cầu khúc cụ thể này (khúc I_{n+1} của phần biểu diễn r). Lợi ích rõ rệt trong trường hợp này có thể là việc tái sử dụng của các khúc được đệm, điều này dẫn đến trải nghiệm phát lại mượt và giảm đáng kể lưu lượng giữa bộ nhớ đệm R và máy chủ S. Các quyết định khác có thể được thực hiện dựa trên việc nhận biết về sự có mặt của bộ nhớ đệm R.

Ngoài ra, cũng cần hiểu rằng bộ nhớ đệm được phát hiện có thể tương ứng với bộ nhớ đệm trong của cổng GW cho phép:

- các khúc được yêu cầu trước đó của nội dung đa phương tiện được tải xuống nhanh hơn;
- các nhà điều hành quá độ cung cấp cho bộ nhớ đệm trong của cổng với nội dung của chúng để cải thiện trải nghiệm của người dùng.

Các tham khảo được bộc lộ trong phần mô tả, các điểm yêu cầu bảo hộ và các hình vẽ có thể được đưa ra một cách độc lập hoặc theo kết hợp thích hợp bất kỳ. Các dấu hiệu có thể, khi thích hợp, được thực hiện trong phần cứng, phần mềm, hoặc kết hợp của phần cứng và phần mềm.

Các số tham chiếu xuất hiện trong các điểm yêu cầu bảo hộ chỉ nhằm minh họa và sẽ không có tác dụng giới hạn phạm vi bảo hộ của các điểm yêu cầu bảo hộ này.

Sáng chế đã được mô tả theo phương án ưu tiên của nó, nhưng rõ ràng rằng sáng chế có thể có nhiều biến đổi và phương án nằm trong khả năng của người có hiểu biết trung bình trong lĩnh vực kỹ thuật và không cần nỗ lực sáng tạo thêm. Do đó, phạm vi bảo hộ sáng chế được xác định bởi phạm vi bảo hộ của các điểm yêu cầu bảo hộ sau đây.

Trong các điểm yêu cầu bảo hộ của sáng chế, phần tử bất kỳ được biểu diễn là phương tiện để thực hiện chức năng quy định (ví dụ bộ phát hiện bộ nhớ đệm 7, bộ ước lượng băng thông 8, môđun quyết định 9, v.v.) nhằm bao trùm cách thức bất kỳ để thực hiện chức năng đó bao gồm, ví dụ, a) kết hợp của các phần tử mạch (ví dụ một hoặc nhiều bộ xử lý) mà thực hiện chức năng đó hoặc b) phần mềm dưới dạng bất kỳ, bao gồm theo đó là, phần sụn, vi mã hoặc dạng tương tự, được kết hợp với hệ mạch thích hợp để chạy phần mềm đó nhằm thực hiện chức năng. Sáng chế như được xác định bởi các điểm yêu cầu bảo hộ này dựa trên thực tế là các chức năng được cung cấp bởi các phương tiện đã đề cập khác nhau được kết hợp và đưa ra cùng nhau theo cách thức mà các điểm yêu cầu bảo hộ này cần phải có. Theo đó, nó được coi là phương tiện bất kỳ, phương tiện này có thể cung cấp các chức năng đó, mà tương đương với các phương tiện được thể hiện trong bản mô tả này.

YÊU CẦU BẢO HỘ

1. Phương pháp làm thích ứng hành vi tải xuống của thiết bị đầu cuối khách (C) được cấu hình để nhận nội dung đa phương tiện từ ít nhất một máy chủ (S), ít nhất một phần biểu diễn của nội dung đa phương tiện này là sẵn có,

phương pháp này bao gồm các bước:

yêu cầu (S0) phần thứ nhất (I_n) của nội dung đa phương tiện với phần biểu diễn đã cho;

phát hiện (S1) nếu bộ nhớ đệm (R) được đặt dọc theo đường truyền giữa thiết bị đầu cuối khách (C) và máy chủ (S), dựa trên yêu cầu của phần thứ nhất (I_n);

trong trường hợp bộ nhớ đệm (R) được phát hiện, yêu cầu (S3) phần thứ hai (I_{n+1}) của nội dung đa phương tiện với phần biểu diễn phụ thuộc vào ít nhất một tiêu chuẩn đặc tính.

2. Phương pháp theo điểm 1, phương pháp này còn bao gồm bước:

ước lượng băng thông của đường truyền giữa thiết bị đầu cuối khách (C) và bộ nhớ đệm (R) được phát hiện.

3. Phương pháp theo điểm 2, trong đó, theo tiêu chuẩn đặc tính này, phần thứ hai (I_{n+1}) được yêu cầu của nội dung đa phương tiện được xác định với:

hoặc cùng phần biểu diễn như phần biểu diễn của phần thứ nhất (I_n) được lưu trữ trong bộ nhớ đệm (R) được phát hiện, bất kể kết quả của việc ước lượng băng thông;

hoặc phần biểu diễn thay thế có tính đến băng thông được ước lượng, phần biểu diễn mới này khác với phần biểu diễn của phần thứ nhất (I_n).

4. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 3, trong đó yêu cầu của phần thứ hai (I_{n+1}) bao gồm đoạn thông tin mà bộ nhớ đệm (R) được phát hiện có thể hiểu được, sao cho, trong trường hợp phần thứ hai (I_{n+1}) không được lưu trữ trong bộ nhớ đệm (R) được phát hiện, thiết bị đầu cuối khách (C) nhận tin nhắn chỉ định là phần thứ hai (I_{n+1}) không sẵn có từ bộ nhớ đệm (R) này.

5. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 4, trong đó phương pháp này còn bao gồm bước:

trong trường hợp việc tải xuống của nội dung đa phương tiện từ bộ nhớ đệm (R) được phát hiện thỏa mãn ít nhất một tiêu chuẩn tải xuống, yêu cầu phần (I_k) thêm nữa của nội dung đa phương tiện với phần biểu diễn mới, mà khác với phần biểu diễn của phần thứ nhất (I_n).

6. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 5, trong đó bước phát hiện bộ nhớ đệm (R) còn bao gồm:

xác định thời gian đi về cho yêu cầu thiết lập kết nối (TCP-RTT) từ thiết bị đầu cuối khách (C) đến máy chủ (S).

7. Phương pháp theo điểm 6, trong đó bước phát hiện bộ nhớ đệm (R) còn bao gồm:

đo độ trễ nhận giữa lúc phát yêu cầu cho phần thứ nhất (I_n) của nội dung đa phương tiện và lúc bắt đầu nhận phần thứ nhất (I_n) được yêu cầu này.

8. Phương pháp theo điểm 7, trong đó bước phát hiện bộ nhớ đệm (R) còn bao gồm:

so sánh thời gian đi về được xác định của yêu cầu thiết lập kết nối (TCP-RTT) và độ trễ nhận được đo.

9. Phương pháp theo điểm 7 hoặc 8, trong đó bước phát hiện bộ nhớ đệm (R) còn bao gồm:

đo thời gian đáp ứng giữa lúc phát yêu cầu phản hồi từ thiết bị đầu cuối khách (C) đến máy chủ (S) và lúc nhận sự đáp ứng đối với yêu cầu phản hồi này;

so sánh thời gian đi về được xác định của yêu cầu thiết lập kết nối (TCP-RTT) với thời gian đáp ứng.

10. Thiết bị đầu cuối được cấu hình để làm thích ứng hành vi tải xuống của nó để nhận nội dung đa phương tiện từ ít nhất một máy chủ (S), ít nhất một phần biểu diễn của nội dung đa phương tiện này là sẵn có, thiết bị đầu cuối này bao gồm:

môđun truyền thông (3) để yêu cầu phần thứ nhất (I_n) của nội dung đa phương tiện với phần biểu diễn đã cho;

bộ phát hiện bộ nhớ đệm (7) để phát hiện nếu bộ nhớ đệm (R) được đặt dọc theo đường truyền giữa thiết bị đầu cuối khách (C) và máy chủ (S), dựa trên yêu cầu của phần thứ nhất (I_n);

môđun quyết định (9) để yêu cầu, trong trường hợp bộ nhớ đệm (R) được phát hiện, phần thứ hai (I_{n+1}) của nội dung đa phương tiện với phần biểu diễn phụ thuộc vào ít nhất một tiêu chuẩn đặc tính.

11. Thiết bị đầu cuối theo điểm 10, thiết bị đầu cuối này còn bao gồm bộ ước lượng băng thông để ước lượng băng thông của đường truyền giữa thiết bị đầu cuối này và bộ nhớ đệm được phát hiện.

12. Thiết bị đầu cuối theo điểm 11, trong đó, theo tiêu chuẩn đặc tính này, phần thứ hai (I_{n+1}) được yêu cầu của nội dung đa phương tiện được xác định với:

hoặc cùng phần biểu diễn như phần biểu diễn của phần thứ nhất (I_n) được lưu trữ trong bộ nhớ đệm (R) được phát hiện, bất kể kết quả của việc ước lượng băng thông;

hoặc phần biểu diễn thay thế có tính đến băng thông được ước lượng, phần biểu diễn mới này khác với phần biểu diễn của phần thứ nhất (I_n).

13. Thiết bị đầu cuối theo điểm bất kỳ trong số các điểm từ 10 đến 12, trong đó yêu cầu của phần thứ hai (I_{n+1}) bao gồm đoạn thông tin mà bộ nhớ đệm (R) được phát hiện có thể hiểu được, sao cho, trong trường hợp phần thứ hai (I_{n+1}) không được lưu trữ trong bộ nhớ đệm được phát hiện (R), thiết bị đầu cuối khách (C) nhận tin nhắn chỉ định là phần thứ hai (I_{n+1}) không sẵn có từ bộ nhớ đệm (R) này.

14. Thiết bị đầu cuối theo điểm bất kỳ trong số các điểm từ 10 đến 13, trong đó thiết bị đầu cuối còn được cấu hình để yêu cầu phần (I_k) thêm nữa của nội dung đa phương tiện với phần biểu diễn mới, mà khác với phần biểu diễn của phần thứ nhất (I_n), trong trường hợp việc tải xuống của nội dung đa phương tiện từ bộ nhớ đệm (R) được phát hiện thỏa mãn ít nhất một tiêu chuẩn tải xuống.

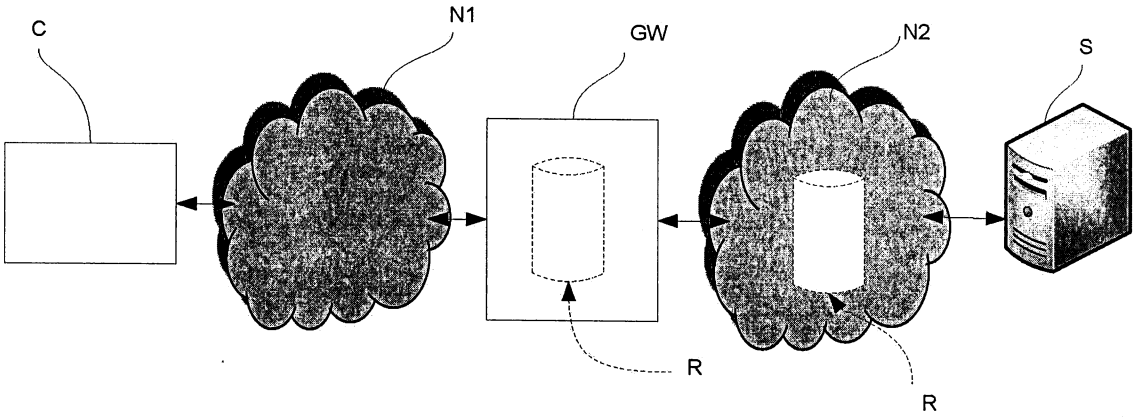

C
N1
GW
N2
S
R
R

Fig.1

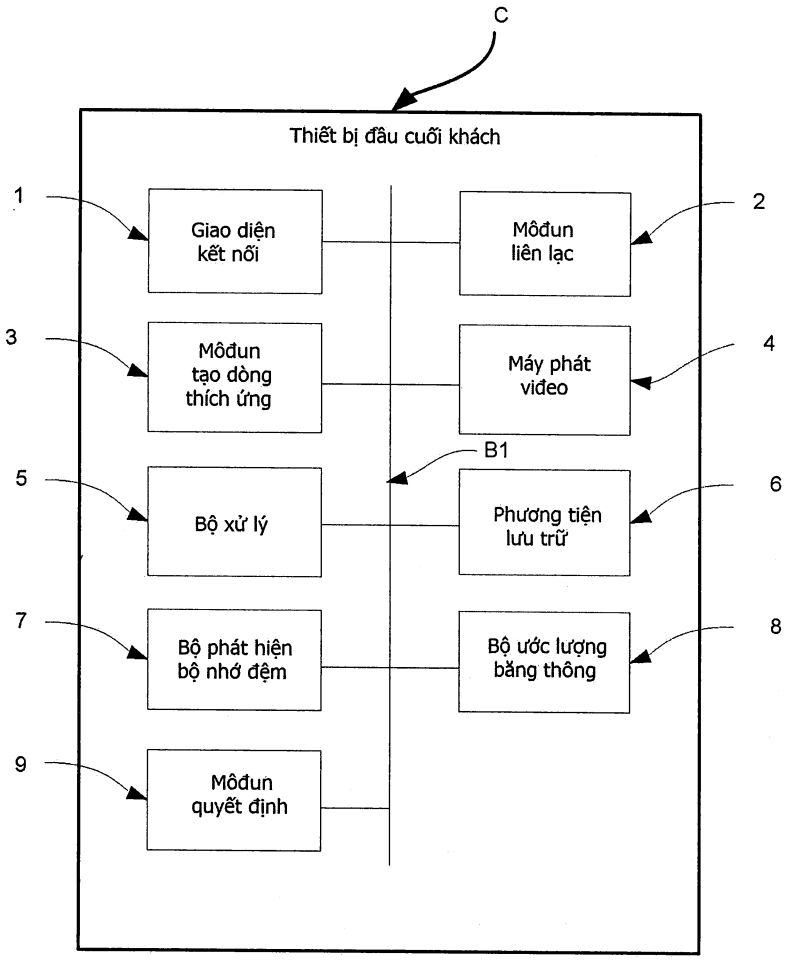

C
Thiết bị đầu cuối khách
1
Giao diện kết nối
2
Môđun liên lạc
3
Môđun tạo dòng thích ứng
4
Máy phát video
B1
5
Bộ xử lý
6
Phương tiện lưu trữ
7
Bộ phát hiện bộ nhớ đệm
8
Bộ ước lượng băng thông
9
Môđun quyết định

Fig. 2

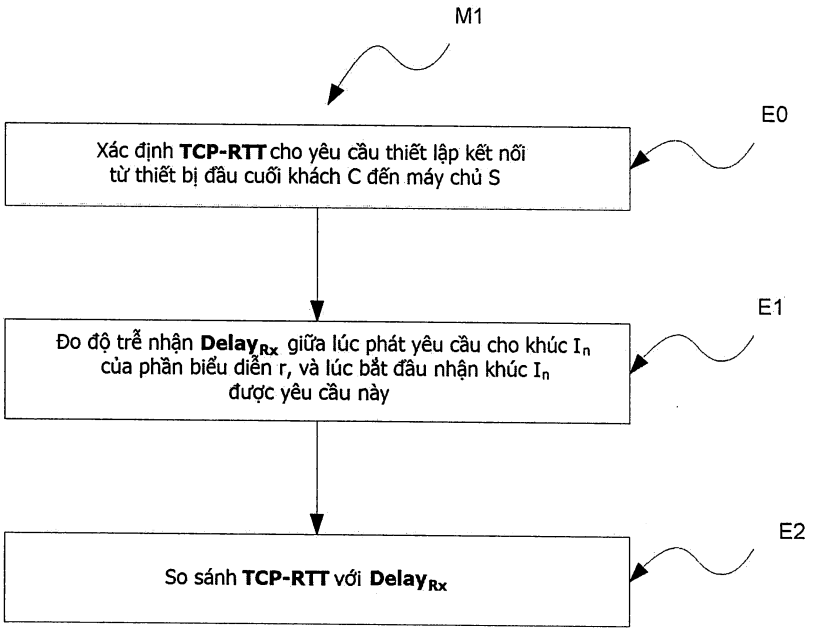

M1
E0
Xác định **TCP-RTT** cho yêu cầu thiết lập kết nối từ thiết bị đầu cuối khách C đến máy chủ S
E1
Đo độ trễ nhận **Delay$_{Rx}$** giữa lúc phát yêu cầu cho khúc I_n của phần biểu diễn r, và lúc bắt đầu nhận khúc I_n được yêu cầu này
E2
So sánh **TCP-RTT** với **Delay$_{Rx}$**

Fig.3

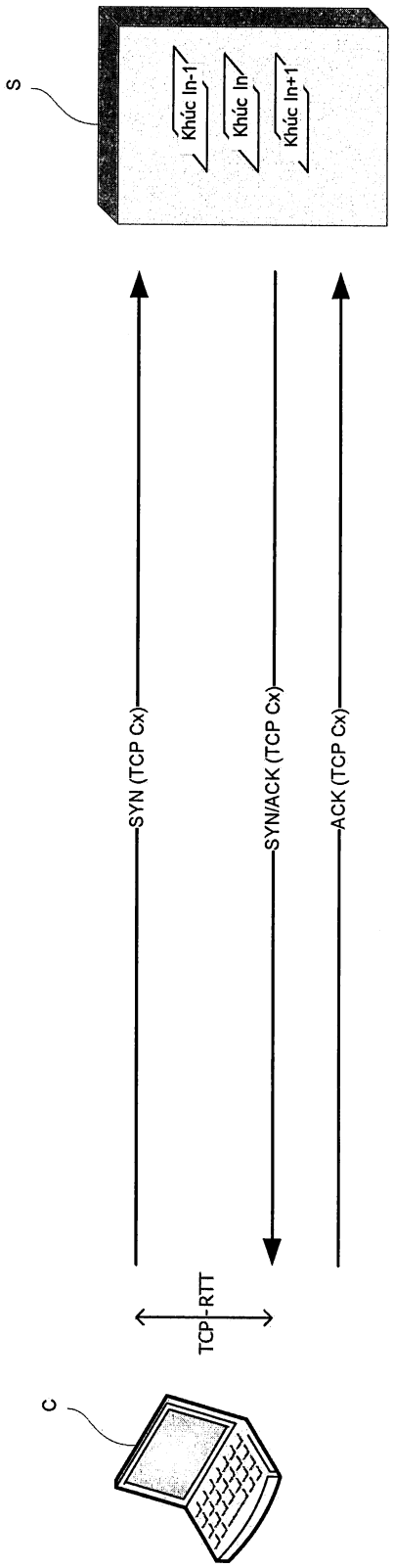

S
Khúc In-1
Khúc In
Khúc In+1
SYN (TCP Cx)
SYN/ACK (TCP Cx)
ACK (TCP Cx)
TCP-RTT
C

Fig.4A

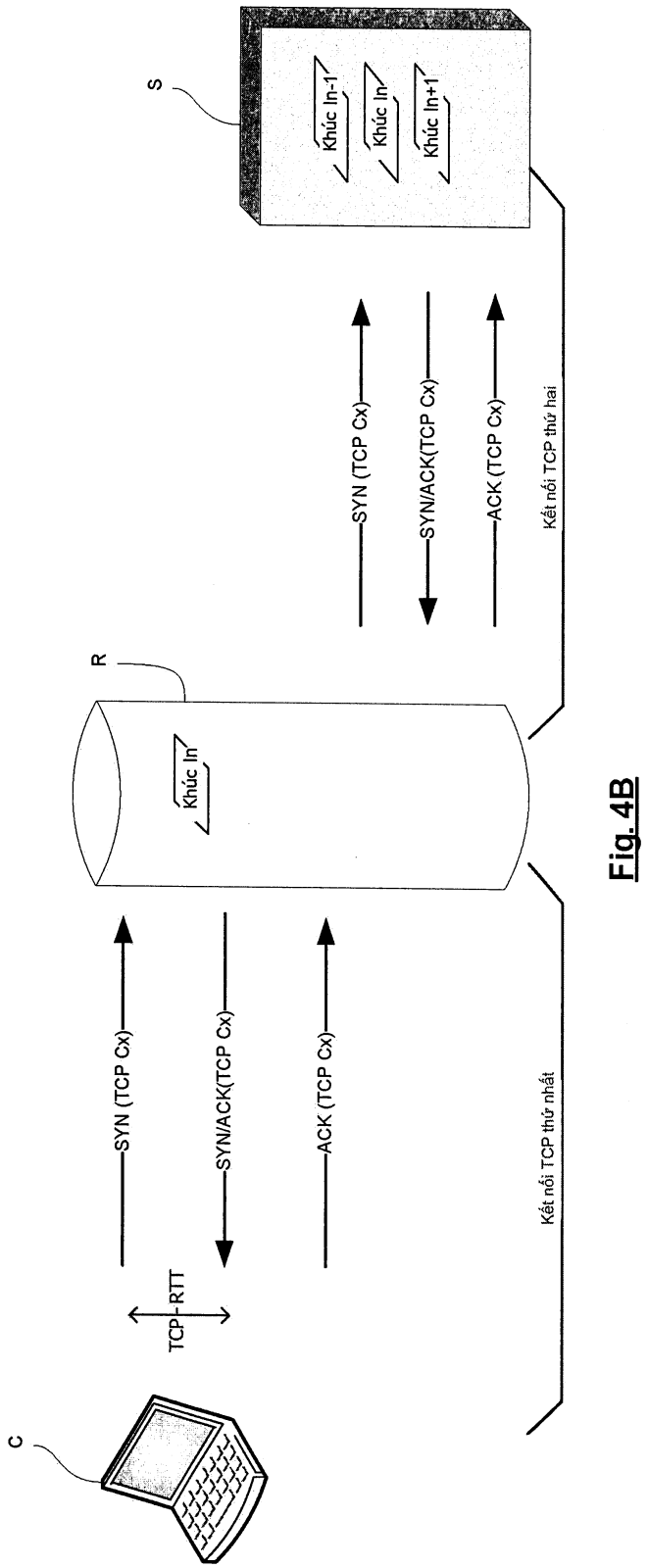

S
Khúc In-1
Khúc In
Khúc In+1
SYN (TCP Cx)
SYN/ACK(TCP Cx)
ACK (TCP Cx)
Kết nối TCP thứ hai
R
Khúc In
SYN (TCP Cx)
SYN/ACK(TCP Cx)
ACK (TCP Cx)
Kết nối TCP thứ nhất
TCP-RTT
C

Fig.4B

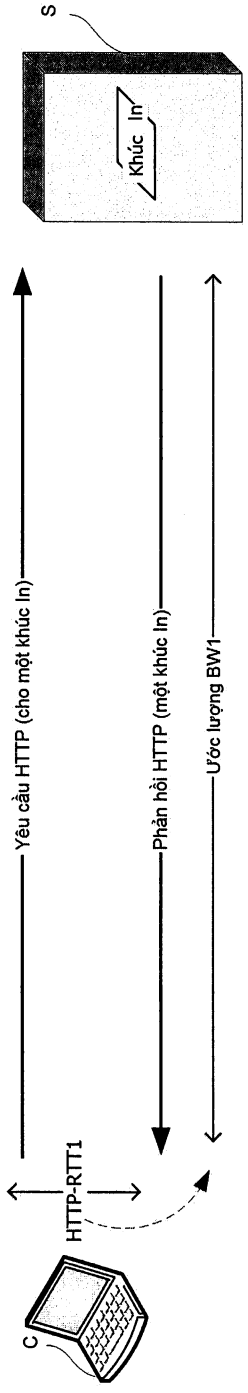


Fig.4C

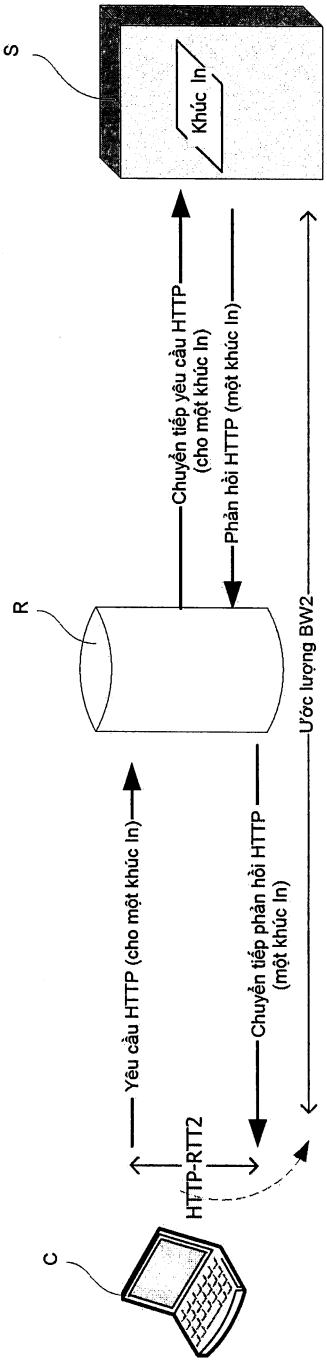


Fig.4D

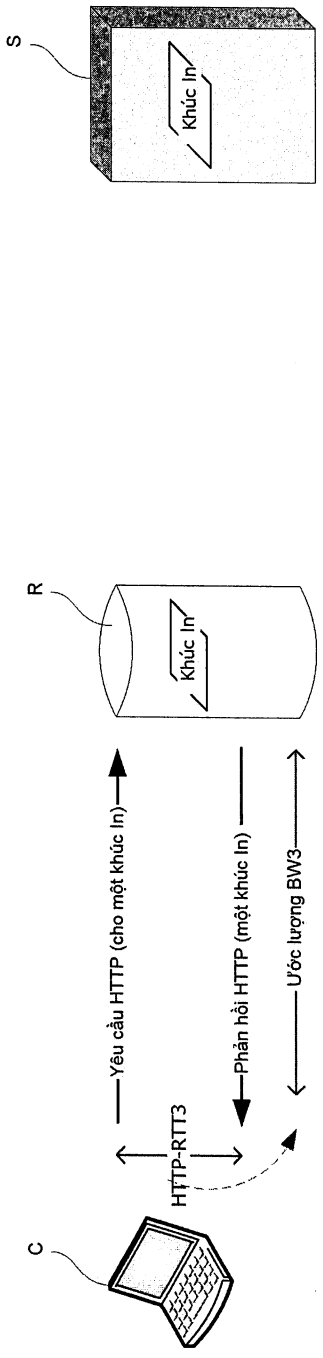


Fig.4E

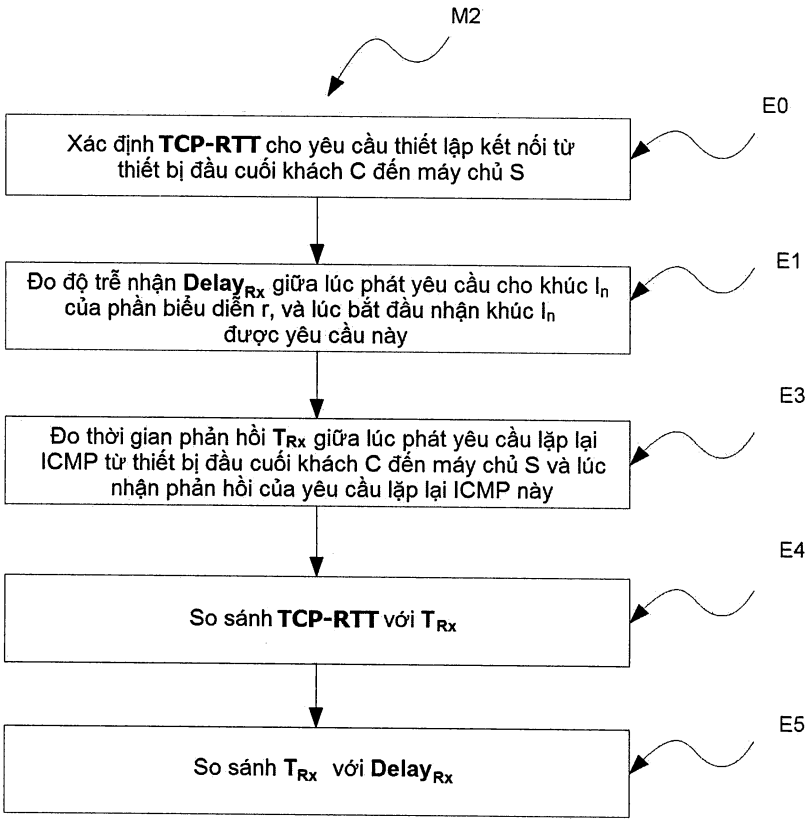

M2
E0
Xác định **TCP-RTT** cho yêu cầu thiết lập kết nối từ thiết bị đầu cuối khách C đến máy chủ S
E1
Đo độ trễ nhận $\mathbf{Delay_{Rx}}$ giữa lúc phát yêu cầu cho khúc I_n của phần biểu diễn r, và lúc bắt đầu nhận khúc I_n được yêu cầu này
E3
Đo thời gian phản hồi $\mathbf{T_{Rx}}$ giữa lúc phát yêu cầu lặp lại ICMP từ thiết bị đầu cuối khách C đến máy chủ S và lúc nhận phản hồi của yêu cầu lặp lại ICMP này
E4
So sánh **TCP-RTT** với $\mathbf{T_{Rx}}$
E5
So sánh $\mathbf{T_{Rx}}$ với $\mathbf{Delay_{Rx}}$

Fig. 5

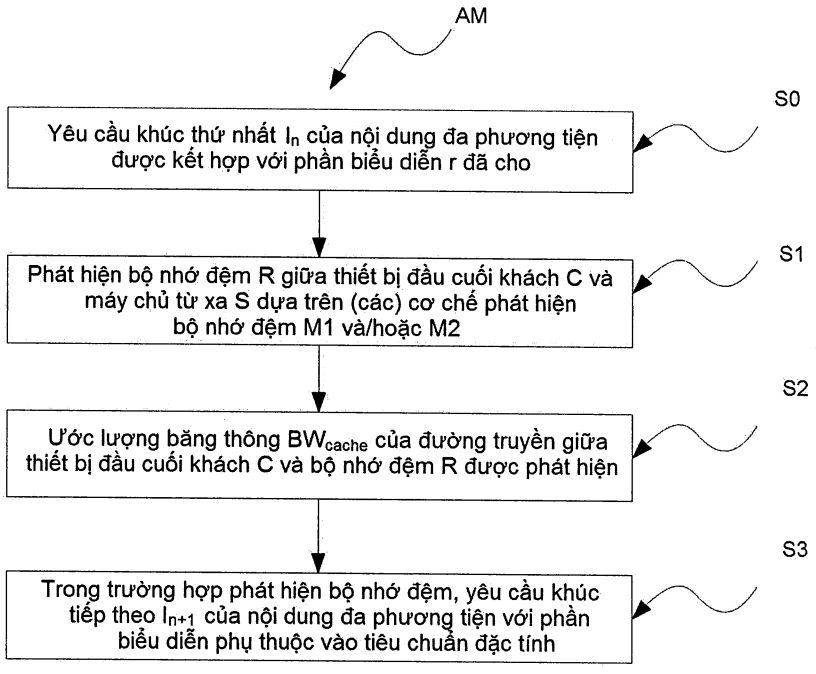

AM
S0
Yêu cầu khúc thứ nhất I_n của nội dung đa phương tiện được kết hợp với phần biểu diễn r đã cho
S1
Phát hiện bộ nhớ đệm R giữa thiết bị đầu cuối khách C và máy chủ từ xa S dựa trên (các) cơ chế phát hiện bộ nhớ đệm M1 và/hoặc M2
S2
Ước lượng băng thông BW_{cache} của đường truyền giữa thiết bị đầu cuối khách C và bộ nhớ đệm R được phát hiện
S3
Trong trường hợp phát hiện bộ nhớ đệm, yêu cầu khúc tiếp theo I_{n+1} của nội dung đa phương tiện với phần biểu diễn phụ thuộc vào tiêu chuẩn đặc tính

Fig. 6