

## **NÂNG CAO HIỆU QUẢ CỦA NGUYÊN TẮC HẠN CHẾ VÀ BÙ BẰNG THÔNG TRONG MẠNG IP**

PHẠM VĂN THƯƠNG

*Tập đoàn Bưu chính Viễn thông Việt Nam*

**Tóm tắt.** Tại mỗi thời điểm, bộ định trình tại mỗi nút mạng lựa chọn một gói tin đưa ra khỏi nút mạng để chuyển tiếp đi. Việc lựa chọn gói tin chuyển đi quyết định tốc độ của luồng gói tin. Do tính biến thiên khó lường trước của lưu lượng và khả năng bùng phát dữ liệu của các ứng dụng trong mạng IP hiện nay, việc hạn chế băng thông của luồng tin để tránh ảnh hưởng tới các luồng tin khác, bù băng thông cho các luồng tin phục hồi sau khi bị lỗi vẫn còn là vấn đề nan giải. Trên cơ sở mô hình MPWPS [5], bài báo đề xuất giải pháp nâng cao hiệu quả của nguyên tắc hạn chế, bù băng thông với những điểm mới như sau: 1) Phương pháp ước lượng tốc độ mới nâng cao hiệu quả hạn của cơ chế hạn chế tốc độ; 2) Sử dụng mô hình điều khiển có phản hồi để lựa chọn tham số hiệu chỉnh trọng số, nâng cao hiệu quả bù QoS.

**Abstract.** At each time, the scheduler of each network node selects a packet of packet flows to forward. Principle of selecting packets to transmit decides the rate of each flow. Because of the unpredictable traffic and possible outbreaks of the application in current IP networks, the bandwidth limitations of the packet flow to avoid affecting other flows and compensating bandwidth for the flow after error recovery remains problematic. On the basis MPWPS model [5], this paper proposes a solution to improve the effectiveness of the principle to limit and compensate bandwidth two features: 1) New periodical average rate estimating method for raising the efficiency of the bandwidth limitation mechanism; 2) Use the feedback control model to choose the weight adjustment parameter and to improve QoS compensation mechanism.

**Keyword.** Bandwidth Scheduler, Rate Limitation, QoS Compensation.

### **1. MỞ ĐẦU**

Mạng chuyển mạch gói sử dụng giao thức Internet (mạng IP) đã được triển khai rộng khắp, cho phép truyền tải lưu lượng các dịch vụ đa dạng (thoại, truyền dữ liệu, đa phương tiện,...) trên cùng một hạ tầng mạng. Mạng IP thực hiện phân tách lưu lượng các dịch vụ, đảm bảo yêu cầu chất lượng khác nhau cho các dịch vụ. Đảm bảo chất lượng dịch vụ (QoS – Quality of Service) liên quan đến nhiều cơ chế ở các lớp, các thành phần mạng khác nhau (các nút mạng trung gian, và các nút mạng đầu cuối): 1) Các cơ chế: Chấp nhận kết nối, định dạng lưu lượng đầu vào (shaping), định trình (scheduling), điều khiển luồng tránh tắc nghẽn, quản lý bộ đệm. 2) Các lớp: Lớp vật lý, lớp IP, lớp truyền tải, lớp ứng dụng. 3) Các thiết bị: Thiết bị trung gian, thiết bị đầu cuối.

Tại mỗi thời điểm, bộ định trình tại mỗi nút mạng lựa chọn một gói tin đưa ra khỏi nút mạng để chuyển tiếp đi. Việc lựa chọn gói tin chuyển đi quyết định tốc độ của luồng gói tin.

Do tính biến thiên khó lường trước của lưu lượng và khả năng bùng phát dữ liệu của các ứng dụng trong mạng IP hiện nay, việc hạn chế băng thông của luồng tin để tránh ảnh hưởng tới các luồng tin khác, bù băng thông cho các luồng tin phục hồi sau khi bị lỗi vẫn còn là vấn đề nan giải.

Đối với các luồng tin có yêu cầu tốc độ tối đa, tại mỗi thời điểm cần ước lượng tốc độ hiện thời để thực hiện điều chỉnh tốc độ phù hợp. Thực hiện bù băng thông cho các luồng tin bị lỗi sau đó phục hồi để đảm bảo sự công bằng giữa các luồng trong thời gian dài hạn. Những công trình nghiên cứu về vấn đề này tới nay còn những hạn chế như sau:

- Hạn chế tốc độ: Trong các cơ chế định trình đến nay (điển hình như trong các tài liệu [1,2,3]), băng thông cấp phát cho các ứng dụng chưa hạn chế tốc độ tối đa, do đó có thể gây ra biến thiên trễ và bùng nổ lưu lượng ở nút mạng kế tiếp. Điều này cũng làm nảy sinh nguy cơ nút mạng bị tấn công từ chối dịch vụ (DoS). Công trình trong [4] đã thực hiện hạn chế tốc độ tuy nhiên còn hạn chế là khi phát hiện luồng vượt quá tốc độ tối đa thì dừng không xử lý gói tin dẫn tới giảm hiệu suất của bộ định trình. Công trình trong [5] thực hiện hạn chế tốc độ tức thời của luồng tin, điều này dẫn tới băng thông đạt được của luồng chưa sát với giá trị giới hạn tối đa.

- Giám sát các tham số QoS để hiệu chỉnh trọng số, bù QoS căn cứ trên thông tin giám sát được. Vấn đề này đã được đề cập trong [1,6,7], song chưa cụ thể và vẫn còn những tồn tại về tham số giám sát, phương thức bù, điều khiển tốc độ... Công trình [5] chọn giá trị tham số hiệu chỉnh trọng số cố định, do đó chưa đảm bảo độ ổn định trọng số của các luồng trong trường hợp số luồng thay đổi.

Trên cơ sở mô hình MPWPS [5], bài báo đề xuất giải pháp nâng cao hiệu quả của nguyên tắc hạn chế, bù băng thông với những điểm mới như sau: 1) Phương pháp ước lượng tốc độ nâng cao hiệu quả hạn chế tốc độ; 2) Nguyên tắc lựa chọn tham số hiệu chỉnh trọng số đảm bảo ổn định trọng số của các luồng khi thực hiện bù QoS trong trường hợp số luồng gói tin biến động.

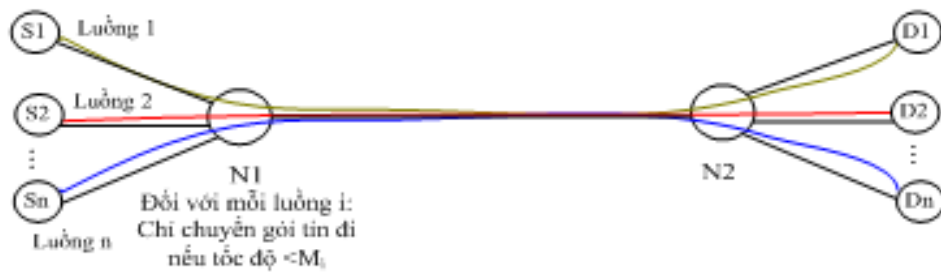
Bố cục của bài báo như sau: Phần 2 trình bày một nguyên tắc nâng cao hiệu quả hạn chế tốc độ. Phần 3 trình bày nguyên tắc lựa chọn tham số hiệu chỉnh bù QoS. Phần 4 trình bày một số kết quả mô phỏng. Cuối cùng là phần kết luận.

## 2. NGUYÊN TẮC ƯỚC LƯỢNG TỐC ĐỘ

Đã có các cơ chế ước lượng băng thông trong mạng IP được đề xuất như: Ước lượng băng thông từ băng thông trung bình đo được như trong EWMA [8], dùng kích thước gói thay đổi, mô hình một gói, mô hình cặp gói, mô hình chuỗi gói [9]... Các nghiên cứu đã chỉ ra phương pháp chung về việc ước lượng tốc độ, tuy nhiên việc áp dụng một dịch vụ, một luồng gói tin cụ thể tại một nút mạng chưa được chỉ rõ. Trong [5] đề xuất cơ chế ước lượng tốc độ tức thời dựa trên kích thước gói tin và khoảng thời gian tính từ khi gửi gói tin trước đó đến thời gian gửi gói tin tiếp theo của một luồng tin. Tốc độ tức thời biến động mạnh, do đó sử dụng tốc độ tức thời để kiểm soát tốc độ tối đa chưa chính xác. Trong phần tiếp theo, bài báo đề xuất cơ chế ước lượng tốc độ mới phục vụ cho việc hạn chế tốc độ tối đa của luồng gói tin.

Trong mạng IP, trên một kết nối cho phép số luồng truy nhập và lưu lượng sử dụng của mỗi luồng biến động theo thời gian. Xét trường hợp khách hàng sử dụng một kết nối giữa 2 điểm, trên đó truyền các luồng lưu lượng của các dịch vụ khác nhau như thoại, hội nghị truyền hình, web, tải file. Tổng băng thông của kết nối được cung cấp theo SLA trong hợp

đồng giữa khách hàng và nhà cung cấp dịch vụ. Các cơ chế như DSCP, CoS, MPLS EXP thực hiện bằng phần cứng cho phép gán nhãn, phân lớp lưu lượng, thực hiện chính sách ưu tiên cho các dịch vụ với số lượng lớp hữu hạn (8-64 lớp). Tuy nhiên, các cơ chế này chưa có khả năng phân biệt, đảm bảo chất lượng cho từng luồng tin, nghĩa là với cùng một dịch vụ, luồng tin của các người sử dụng được ưu tiên như nhau. Do đó có thể xảy ra trường hợp, ví dụ khi một nhân viên sử dụng dịch vụ tải file có thể ảnh hưởng tới chất lượng tải file của lãnh đạo một doanh nghiệp. Phát biểu bài toán: Trên một kết nối ra tại nút mạng cần phục vụ  $n$  luồng kết nối (1.. $n$ ) như trên Hình 1. Mỗi luồng kết nối phục vụ yêu cầu kết nối cho một dịch vụ của một người sử dụng, với giá trị băng thông tối đa là  $M_i$ . Vấn đề cần giải quyết là: đưa ra cơ chế xử lý tại nút mạng để đảm bảo băng thông cung cấp cho luồng  $i$  luôn nhỏ hơn hoặc bằng  $M_i$ . và đảm bảo QoS cần thiết của các luồng tin khác.



Hình 1. Cơ chế hạn chế tốc độ luồng  $i$  tại nút mạng  $N1$

Trong [5], phương pháp xử lý là xác định tốc độ của một luồng trước khi chuyển gói tin của luồng đó đi:

$$rate_i^{k+1} = L_i^{k+1} / delta\_t \quad (1)$$

trong đó,  $L_i^k$  là kích thước gói tin thứ  $k$  của luồng  $i$ ,  $delta\_t$  là khoảng thời gian từ khi truyền gói tin thứ  $k$  tới thời gian truyền gói tin  $k + 1$ . Gói tin của luồng chỉ được chuyển đi nếu tốc độ của luồng không vượt quá tốc độ tối đa  $M_i$  cho phép [5]:

$$rate_i^{k+1} / \delta < M_i, \quad (2)$$

trong đó,  $\delta$  là hằng số. Tốc độ tức thời của một luồng gói tin biến động mạnh, nếu hạn chế tốc độ tức thời theo tốc độ tối đa thì tốc độ trung bình đạt được sẽ nhỏ hơn nhiều so với tốc độ tối đa cần đạt. Để giải quyết vấn đề này, ta cần xem xét việc hạn chế tốc độ trung bình của luồng  $i$  sao cho không vượt quá tốc độ tối đa  $M_i$ . Phương pháp EWMA đo bằng thông của các luồng tin ở liên kết ra như sau [8]:

$$\overline{C_n^{K+1}} = \beta \times \overline{C_n^K} + (1 - \beta) \times C_n^{K+1}, \quad (3)$$

trong đó,  $C_n^{K+1}$  là băng thông đo được tại thời điểm  $K + 1$ ,  $\overline{C_n^K}$  là băng thông đo được ở thời điểm  $K$ ,  $\overline{C_n^{K+1}}$  là băng thông ước lượng được tại thời điểm  $K + 1$ ,  $\beta = 0,9$  là hệ số ước lượng,  $n$  là số các luồng tin được cấp phát băng thông.

Ta áp dụng công thức (3) cho trường hợp tính tốc độ trung bình của luồng  $i$  khi truyền gói tin  $k + 1$  như sau:

$$\overline{rate_n^{k+1}} = \beta \times \overline{rate_n^k} + (1 - \beta) \times rate_n^{k+1}. \tag{4}$$

Tuy nhiên tốc độ trung bình tính theo (4) phụ thuộc nhiều vào lịch sử truyền các gói tin của luồng  $i$ , chưa sát với tốc độ tức thời. Ví dụ, khi một luồng bị lỗi, tốc độ tức thời của luồng đó bằng 0 trong suốt thời gian lỗi. Trong thời gian đó, tốc độ trung bình của luồng giảm dần và trở thành rất nhỏ khi lỗi xảy ra trong thời gian dài. Khi luồng được phục hồi, tốc độ tức thời lớn, nhưng tốc độ trung bình của nó vẫn rất nhỏ. Để tốc độ trung bình tính trong công thức (4) sát với tốc độ tức thời của luồng  $i$ , bài báo đề xuất nguyên tắc PAE (Periodical Average QoS parameter Estimating) tính tốc độ trung bình của luồng  $i$  trong mỗi khoảng thời gian ngắn  $Tgs$  như sau:

Gọi  $l$  là số gói tin của luồng  $i$  được truyền đi trong khoảng thời gian  $Tgs$ , ta có:

$$\overline{rate_i^{Tgs}} = \sum_{j=1}^l L_i^j / Tgs, \tag{5}$$

trong đó,  $\overline{rate_i^{Tgs}}$  là tốc độ trung bình của luồng  $i$  trong khoảng thời gian  $Tgs$ ;  $L_i^j$  là kích thước gói tin thứ  $j$  của luồng  $i$  được gửi đi trong khoảng thời gian  $Tgs$ .

Thay  $rate_i^k$  trong công thức (4) bằng  $rate_i^{Tgs}$  ta có:

$$\overline{rate_i^{k+1}} = \beta \times \overline{rate_i^{Tgs}} + (1 - \beta) \times rate_i^{k+1}. \tag{6}$$

Để hạn chế tốc độ của luồng  $i$ , trước khi chuyển mỗi gói tin của nó đi, ta kiểm tra tốc độ trung bình, và gói tin chỉ được chuyển đi nếu tốc độ trung bình  $\leq M_i$ :

$$\overline{rate_i^{k+1}} \leq M_i. \tag{7}$$

Bảng 1. Tốc độ trung bình ước lượng theo phương pháp EWMA và PAE

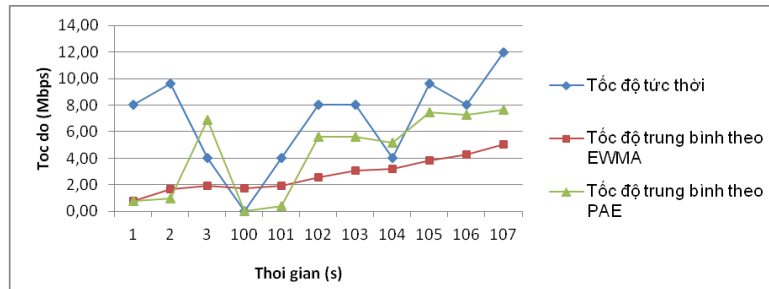
STT	Nội dung	Đơn vị	Các gói tin tới										
			1	2	3	100	101	102	103	104	105	106	107
1	Thời điểm t	ms											
2	Độ lớn gói tin chuyển đi ( $P_i^l$ )	Byte	1000	1200	500	500	500	1000	1000	500	1200	1000	1500
3	delta t	ms	1	1	1	97	1	1	1	1	1	1	1
4	Tốc độ tức thời	Mbps	8,00	9,60	4,00	0,04	4,00	8,00	8,00	4,00	9,60	8,00	12,00
5	Tốc độ trung bình theo EWMA	Mbps	0,80	1,68	1,91	1,72	1,95	2,56	3,10	3,19	3,83	4,25	5,02
6	Tốc độ trung bình theo PAE	Mbps	0,80	0,96	6,88	0,00	0,40	5,60	5,60	5,20	7,44	7,28	7,68

Để đánh giá hiệu quả của nguyên tắc hạn chế tốc độ dựa trên việc giám sát tốc độ trung bình trong khoảng thời gian  $Tgs$  nêu trên, ta xét ví dụ như trong bảng 1.

Giả sử trong khoảng thời gian từ  $t = 4$  đến  $t = 99ms$ , luồng  $i$  bị lỗi và phục hồi tại thời điểm  $t = 100ms$ . Bảng 1 là kết quả so sánh tốc độ xác định theo các phương pháp:

- 1) Tốc độ tức thời được tính theo công thức (1).
- 2) Tốc độ trung bình ( $\overline{rate_i^{EWMA}}$ ) được tính theo công thức (4).
- 3) Tốc độ trung bình ( $\overline{rate_i^{PAE}}$ ) được tính theo công thức (6).

Hình 2 là kết quả so sánh tốc độ tức thời, tốc độ trung bình tính theo *EWMA* và *PAE* sau khi luồng  $i$  phục hồi, từ thời điểm  $t = 101ms$  tốc độ trung bình tính bằng nguyên tắc *PAE* phù hợp với biến động tốc độ tức thời và ở mức trung bình so với đường biểu diễn tốc độ tức thời và tốc độ trung bình tính theo *EWMA*. Như vậy ta có thể nhận xét rằng, tốc độ ước lượng theo nguyên tắc *PAE* phù hợp hơn so với *EWMA* [8] và tốc độ tức thời [5].



Hình 2. Tốc độ tức thời, tốc độ trung bình của luồng gói tin

Độ phức tạp tính toán của PAE: Với mỗi gói tin của luồng cần hạn chế tốc độ, cần thực hiện tính tốc độ trung bình theo công thức (5), (6) và so sánh với tốc độ tối đa theo công thức (7). Như vậy độ phức tạp tính toán của PAE là  $O(m)$ , với  $m$  là số lượng gói tin của các luồng cần hạn chế băng thông.

### 3. NGUYÊN TẮC LỰA CHỌN THAM SỐ HIỆU CHỈNH TRỌNG SỐ BÙ QoS

Trong [5], tác giả sử dụng hàm định giờ để tính lại các tham số QoS sau mỗi khoảng thời gian  $Tgs$  (short-term) và trong toàn thời gian truyền (long-term); xác định được luồng có QoS cần bù, luồng có QoS tốt nhất có thể giảm trọng số để bù cho luồng khác. Sau mỗi chu kỳ giám sát  $Tgs$ , MPWPS thực hiện bù cho luồng bị suy giảm QoS để đảm bảo sự bình đẳng trong việc phục vụ các luồng tin. Để bù băng thông cho luồng bị suy giảm QoS được khôi phục từ trạng thái lỗi, cần băng thông để bù. Vậy băng thông để bù được lấy từ đâu, có ảnh hưởng đến các luồng tin khác hay không? Các công trình trước đây [1,5,6,7] sử dụng nguyên tắc bù như sau: 1) Nguyên tắc trong [7] là dành sẵn một phần băng thông cho mục đích bù; 2) Nguyên tắc trong [1,6] là ép buộc các luồng tin không yêu cầu thời gian thực nhường lại một phần băng thông để bù. 3) Nguyên tắc trong [5] là giảm trọng số của luồng có QoS tốt nhất trong chu kỳ giám sát trước đó, tăng trọng số của luồng cần bù. Nếu chọn phương án 1, khi không có luồng nào cần bù, băng thông dành sẵn không được sử dụng gây lãng phí tài nguyên. Nếu chọn phương án 2 mà không có giám sát cụ thể, việc bù có thể ảnh hưởng nhiều đến QoS của các luồng tin khác. Nguyên tắc bù trong [5] sử dụng phương án 3 khắc phục được nhược điểm của phương án 1, 2.

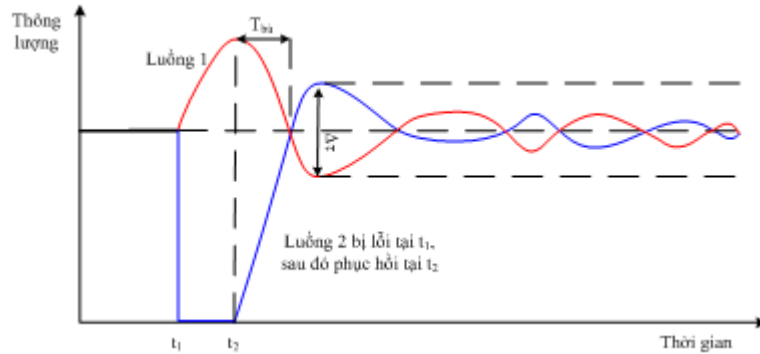
Nguyên tắc bù QoS trong [5] như sau: Giám sát để biết luồng có QoS tốt nhất và kém nhất; giảm QoS của luồng tốt nhất để bù cho luồng có QoS kém nhất tại thời điểm đó. Điều chỉnh giá trị trọng số  $w_{ci}$  dựa vào giá trị các tham số QoS giám sát được trong khoảng thời gian  $Tgs$  ngay trước đó. Cụ thể như sau [5]:

$$1) \text{ Đối với luồng được giảm QoS để bù: } w_{ci} = w_i - r_i. \quad (8)$$

$$2) \text{ Đối với luồng được bù: } w_{ci} = w_i + r_i. \quad (9)$$

3) Đối với các luồng khác:  $w_{ci} = w_i$ . (10)

Trong đó  $r_i$  là hệ số điều chỉnh trọng số. Vấn đề là cần chọn  $r_i$  bằng bao nhiêu để việc điều chỉnh thực hiện nhanh, thông qua thực nghiệm, [5] chọn  $r_i = 0,1$  (10% băng thông của kết nối ra). Giá trị cho  $r_i$  ảnh hưởng tới thời gian bù, độ ổn định trọng số, hiệu suất sử dụng băng thông của các luồng. Việc chọn giá trị  $r_i = 0,1$  trong [5] có thể cho kết quả tốt trong trường hợp số luồng nhỏ, tuy nhiên khi nút mạng xử lý hàng triệu luồng thì việc chọn  $r_i$  như vậy dẫn tới trọng số của các luồng liên tục biến động mạnh.



Hình 3. Biến đổi thông lượng của các luồng trong khi bù

Xét trường hợp hai luồng 1, 2 có trọng số bằng nhau, luồng 2 bị lỗi trong khoảng thời gian  $[t_1, t_2]$  sau đó phục hồi. Hình 3 mô tả sự biến đổi thông lượng của các luồng trong quá trình bù. Thời gian bù  $T_{bù}$  được tính từ khi luồng 2 phục hồi đến khi thông lượng của nó trở về ngang bằng với luồng 1.

Yêu cầu đối với cơ chế bù như sau:

- Thời gian bù ( $T_{bù}$ ) ngắn.
- Biến thiên thông lượng của các luồng ( $\Delta\tau$ ) nhỏ.

Việc bù được thực hiện bằng cách hiệu chỉnh trọng số một giá trị là  $r_i$  đối với luồng có thể bù và luồng cần bù lần lượt theo công thức (9) và (10). Tuy nhiên giá trị  $r_i$  cần lựa chọn phù hợp với số luồng, trọng số của các luồng và sự mất công bằng về thông lượng của luồng cần bù so với các luồng khác tại mỗi thời điểm.

Công trình [10] đã sử dụng quy luật điều chỉnh tỷ lệ vi tích phân (PID) để giải quyết bài toán cấp phát tài nguyên theo trọng số tĩnh  $w_i$  (với  $i = 1..n$ ) cho  $n$  tác vụ ứng dụng tích cực, biểu diễn trong phương trình như sau [10]:

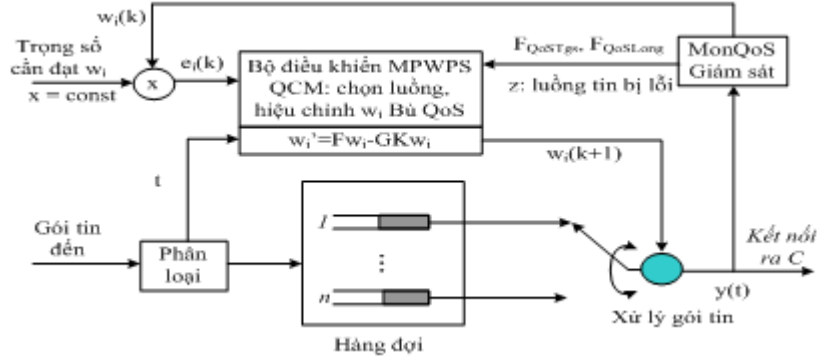
$$u(k) = u(k - 1) + \alpha e(k) + \beta[e(k) - e(k - 1)], \tag{11}$$

trong đó,  $k$  là hằng số thời gian,  $u(k)$  là đầu ra của bộ điều khiển ở thời điểm  $k$ ,  $e(k)$  là sai lệch trạng thái của hệ thống so với trạng thái tham chiếu;  $\alpha, \beta$  là các hệ số điều chỉnh. Tác giả trong [10] đã chứng minh được rằng hệ thống hệ thống điều khiển hoạt động ổn định khi các tham số  $\alpha, \beta$  được lựa chọn thỏa mãn công thức (12):

$$\alpha + 2\beta < 4w_{\min} / \sum_{\forall i} w_i, \tag{12}$$

trong đó,  $w_{\min} = \min\{w_i\}$  với mọi  $i = 1..n$ .

Bài báo sử dụng mô hình điều khiển trạng thái có phản hồi để xây dựng nguyên tắc lựa chọn tham số  $r_i$  để hiệu chỉnh trọng số bù QoS (Hình 4), ý nghĩa các ký hiệu như sau:



Hình 4. Nguyên tắc điều chỉnh trọng số bù QoS

$x$  là giá trị trọng số cần đạt của các luồng  $x = \{w_i\}$  (với  $i = 1..n$ ), như vậy  $x = \text{const}$ ;

$w(k) = \{w_i(k)\}$  (với  $i = 1..n$ ) là giá trị trọng số của các luồng do *MonQoS* đo được sau chu kỳ giám sát  $k$ . Giá trị của  $w(k)$  được cập nhật rồi rạc sau mỗi chu kỳ  $Tgs$ ;

$F_{QoS}^i(t) = T_i(t)/w_i$  [5] là giá trị hàm QoS của luồng  $i$  trong khoảng  $[0, t]$ ;

$F_{QoSTgs}^i(k) = (T_i(t_2) - T_i(t_1))/w_i$  [5] là giá trị hàm QoS của luồng  $i$  đo được sau chu kỳ giám sát thứ  $k$ ;

$z$ : tập các luồng tin bị lỗi được phục hồi do *MonQoS* xác định được sau chu kỳ giám sát  $k$ ;

$e(k) = \{w_i(k) - w_i\}$  (với  $i = 1..n$ ) là sai lệch giữa trọng số đo được so với trọng số cần đạt của luồng  $i$ , được cập nhật sau chu kỳ giám sát  $k$ ;

$w_i(k+1)$  (với  $i = 1..n$ ) là giá trị trọng số đầu ra của bộ điều khiển bù QoS. Bộ định trình cần xử lý cho luồng  $i$  tại chu kỳ  $k+1$  với trọng số  $w_i(k+1)$  nhằm đạt được trọng số cho luồng  $i$  trong thời gian dài hạn là  $w_i$ ;

$y(t)$  là các gói tin được truyền đi ở kết nối đầu ra tại thời điểm  $t$ .

Chức năng *MonQoS* đo các tham số QoS theo chu kỳ giám sát  $Tgs$ , khi xác định được luồng bị lỗi được phục hồi ( $z$ ), dựa trên các tham số QoS xác định được ( $w_i(k)$ ,  $F_{QoSTgs}^i$ ,  $F_{QoSLong}^i$ ); bộ điều khiển MPWPS hiệu chỉnh trọng số của các luồng để bù QoS cho luồng tin bị lỗi, đảm bảo phục vụ công bằng cho các luồng theo trọng số ( $w_i$ ).

Áp dụng luật điều khiển PID dạng rời rạc trong công thức (11) [10] để điều chỉnh trọng số của luồng  $i$  và luồng  $j$ . Giá trị trọng số của luồng  $i$  được tăng lên và trọng số của luồng  $j$  giảm đi một lượng bằng nhau về giá trị tuyệt đối ( $r_i$ ) như sau [10]:

$$r_i(k+1) = \alpha e(k+1) + \beta [e(k+1) - e(k)], \quad (13)$$

trong đó  $\alpha, \beta$  được chọn sao cho thỏa mãn điều kiện như công thức (12) [10]:

$$\alpha + 2\beta < 4w_{\min} / \sum_{\forall i} w_i. \quad (14)$$

Giá trị trọng số của luồng  $i$  được tăng lên theo công thức sau:

$$w_i(k + 1) = w_i(k) + r_i(k + 1). \tag{15}$$

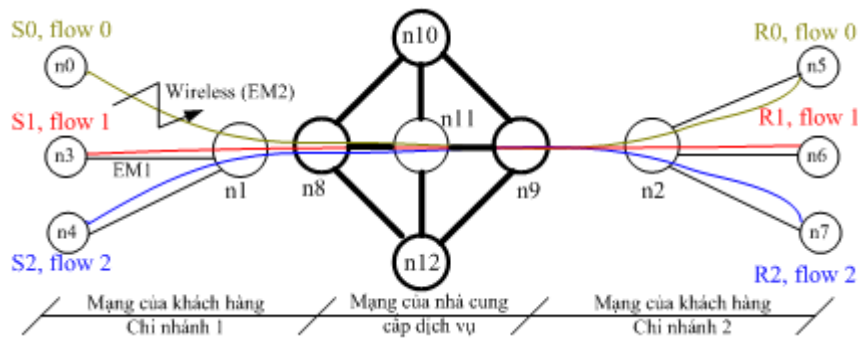
Giá trị trọng số của luồng  $j$  được giảm đi theo công thức sau:

$$w_j(k + 1) = w_j(k) - r_i(k + 1). \tag{16}$$

Nguyên tắc lựa chọn  $r_i$  theo công thức (13) đảm bảo sự ổn định của thông lượng của các luồng ngay cả trong trường hợp số luồng lớn.

#### 4. MÔ PHỎNG

Bài báo sử dụng NS-2 [11] để mô phỏng và so sánh kết quả trong các trường hợp: 1) Sử dụng MPWPS [5]. 2) Sử dụng MPWPS có áp dụng nguyên tắc PAE. Trong khuôn khổ bài báo, tác giả chưa thực hiện mô phỏng việc lựa chọn tham số  $r_i$  theo mô hình điều khiển phản hồi nêu trong mục 3, mà tập trung mô phỏng nguyên tắc PAE, với giá trị  $r_i=0,1$ .



Hình 5. Sơ đồ mô phỏng

Sơ đồ (Hình 5) mô tả kết nối mạng giữa 2 địa điểm. Hai nút mạng n1, n2 của khách hàng được kết nối lần lượt với nút n8 và n9 của nhà cung cấp dịch vụ với băng thông 2Mbps. Kết nối n0, n3, n4 tới n1 và kết nối n2 tới n5, n6, n7 đều bằng 1Mbps. Kết nối giữa các nút mạng của nhà cung cấp dịch vụ (n8, n9, n10, n11, n12) là 100Mbps. Bảng 2 mô tả nguồn lưu lượng mô phỏng.

#### Kết quả mô phỏng

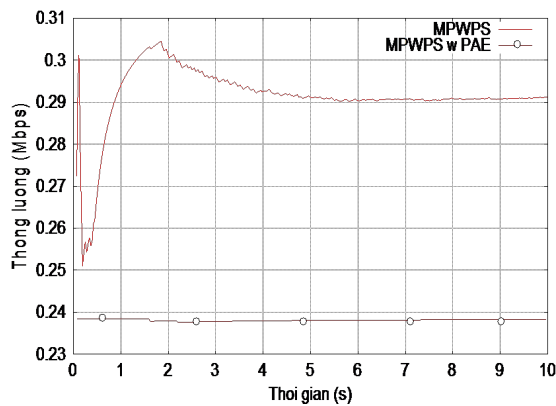
- Kết quả mô phỏng thông lượng được nêu trong Hình 6a, trong cả hai trường hợp (không sử dụng PEA và có sử dụng PEA), thông lượng của luồng 2 đều được hạn chế < 0,3Mbps. Khi sử dụng PEA, thông lượng của luồng 2 ổn định hơn so với khi không dùng PEA.
- Kết quả mô phỏng băng thông nêu trong Hình 6b: Trường hợp không sử dụng PEA, băng thông của luồng 2 biến thiên mạnh. Băng thông của luồng 2 ổn định hơn khi dùng PEA.

Kết quả mô phỏng cho thấy khi sử dụng phương pháp PEA ước lượng tốc độ nhằm mục đích hạn chế tốc độ trong MPWPS, vẫn đảm bảo khả năng hạn chế tốc độ của luồng có yêu cầu tốc độ tối đa, đồng thời tốc độ đạt được ổn định hơn.

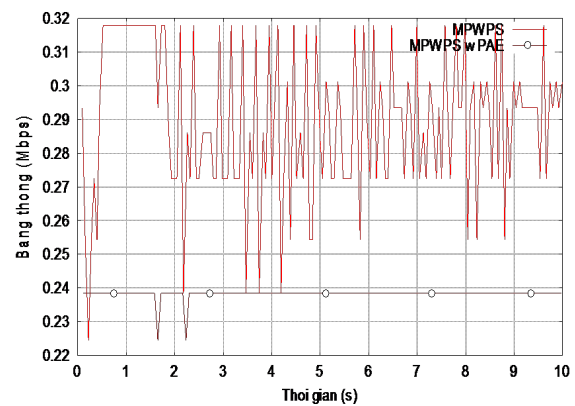


Bảng 2. Nguồn lưu lượng sử dụng để mô phỏng

Luồng	Luồng 0	Luồng 1	Luồng 2
Kích thước hàng đợi (byte)	15000	15000	15000
Trọng số	0,3	0,3	0,3
Nguồn lưu lượng	UDP-CBR	UDP-CBR	UDP-CBR
Độ lớn gói tin (byte)	1000	1000	1000
Khoảng thời gian phát gói tin (s)	0,008	0,008	0,008
Tốc độ phát (Mbps)	0,954	0,954	0,954
Mô hình lỗi	$= 0$ với $t \in [0, 6s; 1, 6s]$	-	-
Băng thông được chia xẻ theo trọng số (Mbps)	0,67	0,67	0,67
Tốc độ tối đa (Mbps)	-	-	0,3
Thời gian mô phỏng (s)	0-10	0-10	0-10



Hình 6a. Thông lượng của luồng 2 khi dùng MPWPS và MPWPS+PEA



Hình 6b. Băng thông của luồng 2 khi dùng MPWPS và MPWPS+PEA

Hình 6. Kết quả mô phỏng MPWPS và MPWPS+PEA

## 5. KẾT LUẬN

Bài báo đã nghiên cứu các vấn đề giám sát, hạn chế tốc độ tối đa; bù  $QoS$  cho luồng tin có  $QoS$  thấp và được phục hồi sau lỗi và đề xuất các giải pháp mới như sau:

- Nguyên tắc PAE sử dụng để xác định tốc độ của luồng gói tin, phục vụ việc hạn chế băng thông của luồng tin. PAE đáp ứng yêu cầu hạn chế tốc độ và sự ổn định của tốc độ đối với luồng tin hạn chế tốc độ tối đa. Nhược điểm của nguyên tắc này là độ phức tạp tính toán là  $O(m)$ , với  $m$  là số gói tin của các luồng cần hạn chế tốc độ.
- Đề xuất áp dụng mô hình điều khiển có phản hồi trong việc xác định giá trị tham số hiệu chỉnh trọng số bù  $QoS$  đảm bảo hệ thống hoạt động ổn định ngay cả trong trường hợp nút mạng phục vụ nhiều luồng gói tin đồng thời tại một kết nối đầu ra.

Kết quả mô phỏng nguyên tắc PAE bằng NS-2 [11] phù hợp với lý thuyết được trình bày đã chứng tỏ tính khả thi của nguyên tắc này. Trong khuôn khổ bài báo này, tác giả chưa thực hiện được mô phỏng việc lựa chọn tham số hiệu chỉnh trọng số bằng mô hình điều khiển có phản hồi nêu trong mục 3.

Nghiên cứu tiếp theo sẽ là những ảnh hưởng khi số luồng lớn và biến động đối với PAE;

thực hiện mô phỏng để đánh giá hiệu quả của mô hình điều khiển có phản hồi trong việc lựa chọn tham số hiệu chỉnh trọng số.

### TÀI LIỆU THAM KHẢO

- [1] Van Thuong Pham, Dang Hai Hoang, A compensation mechanism for bandwidth allocation in IP wireless networks, *Proc. of the 12th Int. Conference on Advanced Communication Technology (ICACT)*, Vol. 2, 2010 (p.1189–1193).
- [2] C. Hu, Y.Tang, K.Chen, B.Liu, Dynamic queueing sharing mechanism for per-flow quality of service control, *IET Commun* **4** (4) (2010) 472–483.
- [3] Dwekat A. Ziyad, “Practical Fair Queueing Schedulers: Simplification through Quantization”, PhD Thesis, North Carolina State University, US, (2009).
- [4] J.F.Lee, M.C.Chen, Y.Sun, WF2Q-M: worst-case fair weighted fair queueing with maximum rate control, *Computer Networks* **51** (2007) 1403–1420.
- [5] Pham Van Thuong, Hoang Dang Hai, A QoS monitoring and compensation mechanism with rate limitation in IP network, *Journal of Science and Technology* (accepted) Hanoi University of Technology (90) (2012).
- [6] Phạm Văn Thương, Hoàng Đăng Hải, Cơ chế định trình giám sát và bù chất lượng dịch vụ, *Tạp chí Khoa học và Công nghệ*, Viện Khoa học Việt Nam, ISSN 0866 708X, (2011) 87–100.
- [7] D. H. Hoang, *Quality of Service Control in the Mobile Wireless Environment*, Peter Lang Publishers, Frankfurt/M-Berlin-Bern-Bruxelles-New York-Oxford-Wien, 2003.
- [8] W. Stallings, *High-Speed Networks: TCP/IP and ATM Design Principles*, ISBN 0-13-525965-7, Prentice Hall, 1998.
- [9] Van Jacobson, *Congestion Avoidance and Control*, 1998.
- [10] Vũ Hoàng Hiếu, “Mô hình tương thích QoS theo ứng dụng trong môi trường đa phương tiện phân tán”, Luận án Tiến sĩ kỹ thuật, 2008.
- [11] K. Fall, K. Varadhan, “The ns Manual”, The VINT Project, (2008).

Ngày nhận bài 06 - 9 - 2012

Nhận lại sau sửa ngày 07 - 12 - 2012