

KHẢO SÁT VÀ ĐÁNH GIÁ HỆ THỐNG PHỤC VỤ ĐÁM ĐÔNG TRỰC TUYẾN

LÊ QUYẾT THẮNG*

TÓM TẮT

Bài báo này sử dụng phương pháp Mô hình hóa một hệ thống phục vụ đám đông trực tuyến dưới dạng mạng các hàng chờ (Queueing Network) để khảo sát các tham số đặc trưng. Kết quả có thể hỗ trợ các quyết định về dự báo số lượng khách hàng trực tuyến hướng tới các giải pháp phòng ngừa tắc nghẽn. Phần ứng dụng trình bày kết quả khảo sát hệ thống Đăng kí học phần trực tuyến của Trường Đại học Cần Thơ.

Từ khóa: thời gian phục vụ, thời gian lưu trữ, chỉ số sử dụng, chỉ số tắc nghẽn, hệ số kinh nghiệm

ABSTRACT

Investigation and evaluation for an online crowd serving system

This paper uses the method of modeling online crowd serving system as an queueing network to survey and evaluate some characterized parameters. The results can support the decisions on forecasts of customers online in an optimal way toward solutions to prevent congestion. The application of method presents some results of the survey on online course registration system at Can Tho University.

Keywords: service time, residence time, utilisation, congestion, expert coefficient.

1. Giới thiệu

Ngày nay việc ứng dụng một dịch vụ trực tuyến trên Internet đã trở nên thông dụng và gần như không thể thiếu đối với các dịch vụ có uy tín. Mặc dù vậy các dịch vụ trực tuyến này luôn phải đối mặt với hai thách thức cơ bản:

- Nhu cầu tham gia trực tuyến tăng nhanh tạo ra đám đông khách hàng có thể làm tắc nghẽn dịch vụ;
- Cạnh tranh không lành mạnh dẫn đến việc sử dụng tin tặc tấn công dưới dạng đám đông “ảo” làm mất uy tín đối phương.

Để vượt qua các thách thức này người ta tập trung nghiên cứu các công nghệ mới về thiết kế và xử lý các hệ thống xử lý song song và đặc biệt thường chú ý nâng cấp các thiết bị phần cứng có cấu hình cao hơn nhiều lần. Việc lượng hóa bài toán để đánh giá và tối ưu hóa những hệ thống như vậy vẫn chưa được quan tâm nên có rất ít kết quả liên quan đến vấn đề này.

* TS, Trường Đại học Cần Thơ

Denning [4] đã đề cập đến phương pháp này nhưng ở giai đoạn đó người ta quan tâm chủ yếu đến tối ưu hóa năng lực tính toán của một hệ thống máy tính. Sau đó một loạt các công trình được công bố liên quan đến các bài toán tối ưu hóa mạng viễn thông.

Vấn đề đối với bài toán xử lý đám đông được đặt ra là hệ thống với một thiết kế cho trước có khả năng xử lý tối đa bao nhiêu khách hàng? Khi đám đông xảy ra làm sao xác định được thành phần nào trong hệ thống gây ra tắc nghẽn? Làm sao dự báo được khả năng tắc nghẽn hệ thống khi bắt đầu có đám đông truy cập và không ngừng tăng lên? Phương pháp mô hình hóa hệ thống như vậy dưới dạng mạng các hàng chờ và khảo sát các tham số đặc trưng của mô hình cho phép trả lời một phần các yêu cầu trên.

S. Athuraliya [2] đã đề xuất mô hình Quản trị hàng chờ chủ động (AQM, Active Queue Management) để tính các độ dài hàng chờ tương ứng với hiện tượng tắc nghẽn. Trên cơ sở này nhiều giải pháp khác nhau được xây dựng nhằm chủ động cắt bỏ các gói tin khi độ dài hàng chờ đã đạt tới một độ dài quy định, chẳng hạn như RED (Random Early Detection) của S. Floyd [5], DRED (Dynamic Random Early Drop) của nhóm J. Aweya [3], Discret Time DRED của AJ Hussein [6]. Cơ sở tính toán của các giải pháp dựa trên AQM là sử dụng bộ đếm các gói tin truy cập tại router để quản trị các hàng chờ. Như vậy, muốn ứng dụng AQM một cách tương tự cho các hệ thống trực tuyến trên Internet cần phải cài đặt bộ đếm các gói tin.

Vấn đề hiện nay là rất nhiều hệ thống xử lý trực tuyến không có các bộ đếm thời gian thực các truy cập hệ thống nên không thể sử dụng các mô hình AQM. Bài báo này vẫn tiếp cận mô hình mạng các hàng chờ để xây dựng một giải pháp đơn giản nhằm dự báo hiện tượng tắc nghẽn đối với các hệ thống phục vụ đám đông. Kết quả dự báo có thể hỗ trợ tốt cho các giải pháp phòng ngừa tắc nghẽn.

Xuất phát từ một đặc điểm rất đáng lưu ý là các tham số ổn định của hệ thống như: xác suất phát sinh một khách hàng và thời gian xử lý một khách hàng rất khó được xác định trên các hệ thống hiện có. Vì vậy bài báo đề xuất một khái niệm được gọi là Hệ số kinh nghiệm. Hệ số này lượng hóa các nhận xét theo kinh nghiệm của các nhà Quản trị hệ thống về “tính chậm chạp” của nó khi có đông khách hàng phát sinh. Từ cách lượng hóa Hệ số kinh nghiệm, bài báo đưa ra phương pháp khảo sát các tham số đặc trưng cho chỉ số tắc nghẽn. Kết quả khảo sát có thể tính được các ngưỡng kiểm soát số lượng đám đông và thời gian hoàn tất một yêu cầu của khách hàng. Khả năng kiểm soát này có thể hỗ trợ các phương pháp ra quyết định về phòng ngừa tắc nghẽn cũng như lọc ra các khách hàng có thời gian hoàn tất xử lý bất thường.

Phương pháp được ứng dụng vào khảo sát Hệ thống đăng kí học phần trực tuyến tại Trường Đại học Cần Thơ và đưa ra một số kết luận ban đầu.

2. Mô hình Mạng các hàng chờ

2.1. Các khái niệm cơ bản

Mạng các hàng chờ (Queueing Network) là một phương pháp mô hình hóa một hệ thống nối mạng các hàng chờ [7]. Kiến trúc tổng quát của một mạng các hàng chờ

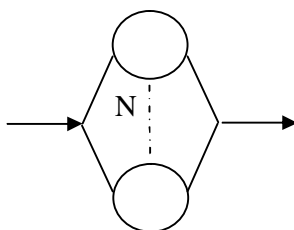
gồm các thành phần cơ bản: Trung tâm Phục vụ, Trung tâm Trì hoãn và Luồng khách hàng.

Trung tâm Phục vụ (Service Center) là một dịch vụ trong hệ thống xử lý một yêu cầu cụ thể của khách hàng và tạo ra một hàng chờ. Mô hình xếp hàng với kí pháp Kendall: $A/B/s/m/N/Z$ có thể áp dụng để tính toán các tham số tại mỗi trung tâm phục vụ như vậy. Một trung tâm Phục vụ được biểu diễn bởi khối cơ bản trong hình 1.



Hình 1. Khối cơ bản biểu diễn một trung tâm Phục vụ

Trung tâm Trì hoãn (Delay Center) là một khu vực dành riêng cho khách hàng đã đăng nhập và cần suy nghĩ trước khi quyết định tham gia một dịch vụ của hệ thống. Khối cơ bản biểu diễn một trung tâm Trì hoãn được trình bày trong hình 2.



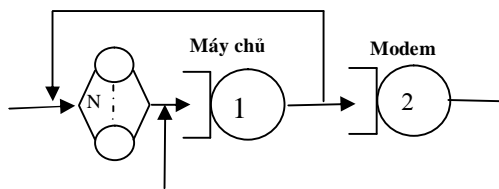
Hình 2. Khối cơ bản của một trung tâm Trì hoãn chứa N khách hàng

Luồng khách hàng là luồng nối mạng các hàng chờ với một lưu lượng nào đó. Các trung tâm Phục vụ hoặc Trì hoãn được kết nối với nhau bởi các luồng khách hàng tạo thành Mạng các hàng chờ.

Luồng khách hàng thường được phân loại theo Lưu lượng Giao dịch (Transaction Workload) hoặc theo Lưu lượng Lô (Batch Workload). Nếu luồng khách hàng là lưu lượng giao dịch thì Trung tâm Phục vụ sẽ phục vụ từng khách hàng một. Còn nếu luồng khách hàng là lưu lượng lô thì Trung tâm Phục vụ sẽ phục vụ từng lô khách hàng một.

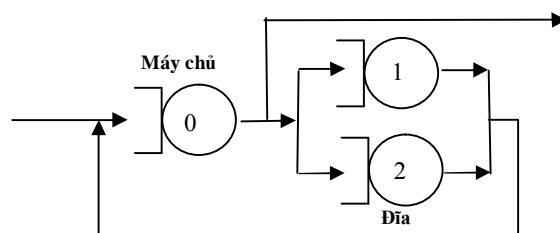
Các mạng hàng chờ cơ bản gồm: mạng Đóng (Close Queueing Network), mạng Mở (Open Queueing Network) và mạng tổng quát (General Queueing Network).

Mạng Đóng: khách hàng đi vào mạng và được xếp vào một trung tâm Trì hoãn sau đó đi tới một hoặc nhiều trung tâm phục vụ rồi lại quay về Trung tâm Trì hoãn nơi mà nó xuất phát. Mạng Đóng được minh họa trong hình 3.



Hình 3. Mạng đóng biểu diễn mô hình hóa một máy chủ quản lí N khách hàng truy cập đồng thời Internet qua Modem

Mạng Mở: khách hàng đi thẳng tới một hoặc nhiều trung tâm phục vụ, sau đó thoát khỏi mạng. Hình 4 mô tả sơ đồ mô hình hóa một hệ thống máy tính có 2 đĩa lưu trữ dữ liệu.



Hình 4. Mạng mở đối với một máy tính có 2 đĩa dữ liệu

Mạng hàng chờ (tổng quát): mô hình chung có thể có một trong các dạng: mạng Đóng, mạng Mở và mạng vừa Đóng vừa Mở

2.2. Khảo sát Hàm chỉ số tắc nghẽn mạng

2.2.1. Chỉ số tắc nghẽn

Mô hình xếp hàng áp dụng cho mỗi trung tâm phục vụ trong trường hợp xử lí đám đông có thể xếp xỉ bởi mô hình M/M/1. Giả thiết này là hợp lí.

Thứ nhất, những khách hàng có nhu cầu khác đám đông sẽ có mật độ thấp hơn rất nhiều, do đó tỉ lệ chia sẻ thời gian xử lí cho nhóm khách hàng này sẽ không đáng kể. Như vậy có thể sử dụng một lớp khách hàng có cùng nhu cầu và tạo ra đám đông cần được xử lí. Khi đó mô hình xếp hàng tại mỗi trung tâm phục vụ là M/M/s.

Thứ hai, hầu hết các hệ thống ứng dụng trực tuyến đều đơn giản hóa quá trình phân tải. Có nghĩa là mỗi hệ thống xử lí trực tuyến đều cài sẵn module phân tải cho s máy chủ xử lí song song khi có đám đông xảy ra. Với phương pháp phân tải như vậy, s máy chủ xử lí song song sẽ tạo ra s trung tâm phục vụ song song. Khi đó mô hình xếp hàng tại mỗi trung tâm phục vụ sẽ trở thành M/M/1.

Xét các tham số trên mạng các hàng chờ:

N – Tổng số khách hàng đăng nhập hệ thống và được lưu giữ tại một trung tâm trì hoãn trước khi tham gia một dịch vụ nào đó có trong hệ thống.

λ – Lưu lượng giao dịch trung bình của luồng đi tới một trung tâm phục vụ. Dưới mô hình xếp hàng còn được hiểu là mật độ trung bình khách hàng có nhu cầu được phục vụ.

p – Xác suất để một khách hàng đã đăng nhập đi tới một trung tâm phục vụ. Xác suất này ổn định đối với bài toán xử lý và không phụ thuộc λ . Quan hệ đầu tiên được xác định là: $p = \lambda/N$.

S – Thời gian phục vụ (Service Time) trung bình một yêu cầu của khách hàng tại một trung tâm phục vụ. S cũng là một tham số ổn định đối với bài toán xử lý và không phụ thuộc λ .

R – Thời gian lưu trú (Residence Time) trung bình của mỗi khách hàng tại một trung tâm phục vụ. Dưới mô hình xếp hàng thì R còn được gọi là thời gian đáp ứng (Response Time) một yêu cầu của Trung tâm.

R^* - Thời gian trung bình hoàn tất tất cả các xử lý tại tất cả các trung tâm phục vụ của một khách hàng và được tính theo công thức: $R^* = \sum_i R_i$, i chạy trên tất cả các trung tâm phục vụ [7].

Q – Độ dài hàng (Queue Length) trung bình trong một trung tâm phục vụ, hay còn là số khách hàng trung bình đang lưu trú tại Trung tâm. Mặt khác, theo luật Little [4], ta lại có quan hệ: $Q = \lambda R$.

φ – Chỉ số sử dụng (Utilisation) dịch vụ của một trung tâm phục vụ, hay còn được hiểu là mức tải của Trung tâm.

Từ mô hình M/M/1 của lý thuyết xếp hàng [1], ta có công thức tính chỉ số sử dụng: $\varphi = \lambda S = pNS$.

Để hệ thống không bị tắc nghẽn, về mặt lý thuyết [1]: $0 < \varphi < 1$. Nhưng trong thực tế, dấu hiệu tắc nghẽn xảy ra khi $\varphi > 0.7$, do thời gian lưu trú R trong một trung tâm phục vụ tăng lên rất nhanh hướng tới vô cùng.

Thật vậy, thời gian lưu trú tại một trung tâm phục vụ với mô hình M/M/1 có thể được tính theo chỉ số sử dụng [1] như sau:

Trước hết từ M/M/1 ta có công thức tính $Q = \varphi/(1-\varphi)$. Do $Q = \lambda R$ và $\varphi = \lambda S$, nên $R = S/(1-\varphi)$.

Như vậy có thể xem công thức tính R là một hàm số dạng Hyperbol biến thiên theo φ và có vận tốc biến thiên theo φ là đạo hàm bậc 1 của R , hay $R' = S/(1-\varphi)^2$.

Nếu khảo sát vận tốc tăng R' của thời gian lưu trú theo φ có thể thấy các đặc điểm cơ bản:

- Khởi động $\varphi = 0$, thì vận tốc tăng là S ;
- Khi φ đạt 0.5, thì vận tốc tăng lên: $4S$;
- Cho φ tăng lên tới 0.7 thì vận tốc tăng nhanh và vượt quá $10.5S$;

- Cho ρ tăng tới 0.8 vận tốc tăng rất nhanh và vượt 25S;
- Cho ρ vượt 0.9 vận tốc tăng hơn 100S và hướng tới vô cùng.

Kết quả khảo sát nhanh này cho thấy sự chậm chạp của hệ thống sẽ xảy ra khi có một trung tâm phục vụ trong hệ thống có $\rho > 0.7$. Vì vậy ρ có thể được gọi là *chỉ số tắc nghẽn* và “sự chậm chạp” còn có thể chấp nhận được của hệ thống trước khi bị tắc nghẽn hoàn toàn tương ứng với ρ nằm trong khoảng [0.7, 0.8].

2.2. Các bước khảo sát chỉ số tắc nghẽn mạng

Để có thể khảo sát được một hệ thống xử lý trực tuyến dưới mô hình mạng các hàng chờ cần thực hiện từng bước theo quy trình sau:

- 1- Xây dựng mô hình mạng các hàng chờ từ kết quả phân tích chức năng của hệ thống xử lý trực tuyến;
- 2- Tính các tham số ổn định không phụ thuộc lưu lượng luồng, gồm: p và S ;
- 3- Xây dựng các hàm khảo sát liên quan đến chỉ số tắc nghẽn mạng;
- 4- Khảo sát các hàm liên quan đến chỉ số tắc nghẽn mạng.

Bước 1: Xây dựng mô hình mạng các hàng chờ.

Ở bước đầu tiên này ta cần phải thực hiện một số phân tích để có thể xây dựng được mô hình mạng các hàng chờ:

- Phân tích xác định các trung tâm trì hoãn và trung tâm phục vụ;
- Phân tích các luồng kết nối các trung tâm và lưu lượng của các luồng;
- Xây dựng mô hình mạng các hàng chờ.

Bước 2: Tính các tham số ổn định không phụ thuộc vào lưu lượng luồng. Bước này yêu cầu phải thực hiện các khảo sát ban đầu từ kinh nghiệm chuyên gia hoặc Log File các sự kiện. Kết quả khảo sát cho phép phân tích để tính được: p và S .

Bước 3: Xây dựng hàm liên quan đến chỉ số tắc nghẽn.

Khảo sát các chỉ số tắc nghẽn, tức là khảo sát các hàm số liên quan đến “sự chậm chạp” của hệ thống khi xử lý nhu cầu của một đám đông.

Trước hết, quan hệ giữa N và ρ được biểu diễn bởi công thức: $\rho = \lambda S = pNS$. Do ρ thể hiện tính chậm chạp của một trung tâm phục vụ trong khoảng [0.7, 0.8] nên ta có thể xác định được vùng biến thiên tương ứng của N qua khảo sát quan hệ hàm:

$$N(\rho) = \rho/pS \quad (1).$$

Kết quả khảo sát sẽ giúp chúng ta xác định được các giá trị N_{min} và N_{max} tích hợp từ nhiều trung tâm phục vụ.

Quan hệ tiếp theo là quan hệ giữa thời gian lưu trú R với N .

$$\text{Từ các kết quả tính } Q \text{ và } R [1], \text{ ta có: } Q = \lambda R = \frac{\rho}{1-\rho}.$$

$$\text{Thay } \varphi = \lambda S = pNS, \text{ ta tính được: } R = R(N) = \frac{S}{1 - pSN} \quad (2).$$

Do khoảng biến động của N đã được xác định từ khảo sát hàm $N(\varphi)$, nên kết quả khảo sát vùng biến thiên của R theo N sẽ cho thấy mức độ “chậm chạp” của một trung tâm phục vụ.

Quan hệ quan trọng nhất là quan hệ giữa R^* và N và được khảo sát thông qua hàm

$$R^*(N) = \sum_i R_i(N), \text{ } i \text{ chạy trên tất cả các trung tâm phục vụ,} \quad (3).$$

Bước 4: Khảo sát các hàm liên quan đến chỉ số tắc nghẽn và tổng hợp kết quả.

Khảo sát hàm $N(\varphi)$ thấy được vùng biến thiên hợp lí của N . Đặc biệt có thể tìm được các giá trị N_{\min} và N_{\max} biểu diễn sự chậm chạp của hệ thống.

Khảo sát $R(N)$ và $R^*(N)$ sẽ chỉ ra sự tăng đột biến của R và R^* trong khoảng N_{\min} và N_{\max} . Kết quả khảo sát cũng chỉ ra R và R^* tăng đến lớn vô cùng khi N vượt quá N_{\max} và đó cũng là minh chứng cho sự tắc nghẽn hoàn toàn của hệ thống.

3. Khảo sát hệ thống Đăng kí học phần trực tuyến tại Trường Đại học Cần Thơ

3.1. Xây dựng mô hình mạng các hàng chờ

3.1.1. Mô tả tình huống

Trong chương trình đào tạo theo tín chỉ tại Trường Đại học Cần Thơ, sinh viên phải đăng kí môn học trực tuyến trên một hệ thống thông tin có bản quyền theo quy trình như sau:

- Tất cả sinh viên đều phải mở tài khoản của mình trong hệ thống để có thể đăng kí môn học trực tuyến.
- Khi đăng kí môn học, mỗi sinh viên đều phải đăng nhập. Để hệ thống có thể chấp nhận số lượng N sinh viên đăng kí môn học trực tuyến đồng thời đủ lớn, người ta bố trí hai máy chủ Đăng kí học phần và một máy chủ Phân tải. Máy chủ Phân tải có nhiệm vụ cân bằng tải cho hai máy chủ Đăng kí học phần, do đó nó sẽ kiểm soát và điều hướng các đăng nhập sang máy chủ nào có số lượng đăng kí trực tuyến ít hơn.
- Khi thực hiện đăng kí học phần, mỗi sinh viên phải suy nghĩ và điền vào mẫu đăng kí học phần (trực tuyến). Sau khi hoàn tất mẫu, sinh viên gửi kết quả cho máy chủ Đăng kí học phần. Máy chủ Đăng kí học phần xử lí mẫu và chuyển kết quả xử lí cho máy chủ Cơ sở dữ liệu (CSDL) để lưu lại các đăng kí mới của sinh viên.
- Sau mỗi lần đăng kí thành công sinh viên có thể xem lại kết quả và thoát khỏi hệ thống.

3.1.2. Phân tích các trung tâm trì hoãn và trung tâm phục vụ

Máy chủ đăng nhập xử lý các đăng nhập của sinh viên sau đó phân tải tạo thành Trung tâm phục vụ đầu tiên ở mức 0. Ta gọi Trung tâm này là Trung tâm 0 (Đăng nhập và Phân tải).

Sinh viên đăng nhập thành công và được phân tải sang máy chủ Đăng kí 1 hoặc 2. Trước khi đăng kí sinh viên phải cần thời gian suy nghĩ và điền đầy đủ mẫu đăng kí học phần trực tuyến. Như vậy, Máy chủ Đăng kí học phần 1 hoặc 2 phải tạo ra vùng nhớ riêng để lưu lại các sinh viên đăng nhập và cần thời gian suy nghĩ. Các vùng nhớ này tạo thành các trung tâm Trì hoãn. Tất cả sinh viên đăng nhập thành công đều được xếp vào một trong hai trung tâm Trì hoãn ở mức 1 để chuẩn bị đăng kí học phần. Máy chủ Đăng kí 1 tạo ra Trung tâm Trì hoãn 1.1, còn máy chủ Đăng kí 2 tạo ra Trung tâm Trì hoãn 1.2.

Sau khi hoàn thành mẫu đăng kí học phần, sinh viên yêu cầu được đăng kí chính thức. Máy chủ Đăng kí học phần sẽ thực hiện công đoạn xử lý mẫu đăng kí và chuyển cho máy chủ CSDL cập nhật: thêm, sửa hoặc xóa các đăng kí học phần của sinh viên. Với chức năng này, các máy chủ Đăng kí học phần sẽ tạo ra các trung tâm phục vụ ở mức 2. Ta gọi Trung tâm Phục vụ tương ứng với Máy chủ Đăng kí 1 là Trung tâm Phục vụ 2.1 (Đăng kí học phần 1), còn Trung tâm tương ứng với Máy chủ đăng kí 2 là Trung tâm Phục vụ 2.2 (Đăng kí học phần 2).

Máy chủ CSDL cập nhật các dữ liệu đăng kí bởi sinh viên và tạo ra Trung tâm Phục vụ ở mức 3 và được gọi là Trung tâm Phục vụ 3 (CSDL).

3.1.3. Phân tích luồng

Luồng sinh viên đăng nhập đi vào Trung tâm 0 là luồng khởi tạo sự hoạt động của hệ thống và có lưu lượng là λ .

Thoát khỏi Trung tâm 0 với luồng λ không đổi, sinh viên được gửi đến Trung tâm 1.1 với xác suất là q_1 hoặc đến Trung tâm 1.2 với xác suất tương ứng là q_2 . Do cân bằng tải nên $q_1 = q_2 = q = 1/2$.

Các mẫu đăng kí học phần hoàn chỉnh sẽ được chuyển đến Trung tâm Đăng kí học phần tương ứng (2.1 hoặc 2.2). Từ Trung tâm 1.1, mỗi sinh viên được chuyển đến Trung tâm 2.1 với xác suất p_1 . Nếu điền mẫu ở Trung tâm 1.2 thì sinh viên được chuyển đến Trung tâm 2.2 với xác suất p_2 . Do sinh viên có khả năng điền mẫu như nhau nên họ có cùng xác suất đi tới trung tâm Đăng kí học phần, vậy: $p_1 = p_2 = p$.

Thoát khỏi Trung tâm Đăng kí học phần, mỗi sinh viên sẽ có 3 trạng thái:

- Trạng thái 1: Đăng kí mới tạo luồng đi tới Trung tâm CSDL;
- Trạng thái 2: Kết thúc cập nhật tạo luồng quay về Trung tâm Trì hoãn;
- Trạng thái 3: Hoàn tất đăng kí học phần tạo luồng thoát khỏi hệ thống.

Do các sinh viên đều thực hiện quy trình đăng kí học phần tương tự nhau nên mỗi sinh viên dù thoát khỏi Trung tâm 2.1 hay 2.2 đều có phân phối xác suất giống nhau

theo 3 trạng thái trên. Như vậy mỗi sinh viên khi thoát khỏi trung tâm 2.1 hoặc 2.2 đều có phân phối xác suất:

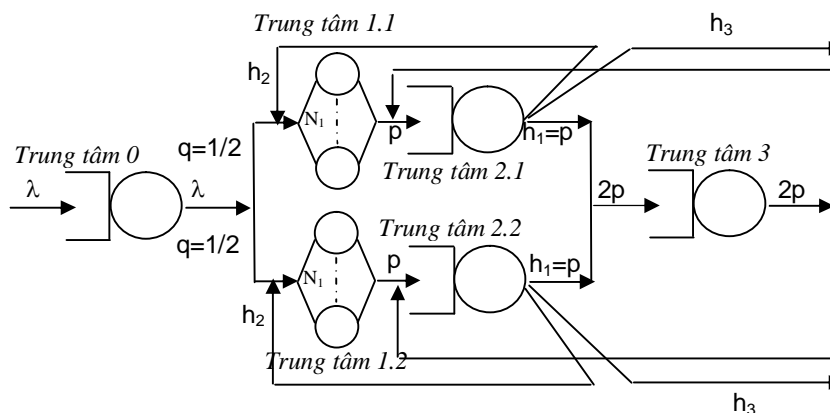
$$P(\text{Trạng thái 1}) = h_1, P(\text{Trạng thái 2}) = h_2 \text{ và } P(\text{Trạng thái 3}) = h_3.$$

Dễ dàng nhận thấy, mỗi sinh viên hoàn thành đăng kí học phần mới sẽ phải cập nhật CSDL, do đó: $h_1 = p$.

Ở mức 3, mỗi sinh viên hoàn thành cập nhật đăng kí học phần của mình sẽ tạo luồng với lưu lượng tương ứng ở đầu vào để quay về mức 2.

3.1.4. Mô hình mạng các hàng chờ

Kết quả phân tích trung tâm và luồng được tổng hợp thành mô hình mạng các hàng chờ của hệ thống Đăng kí học phần trực tuyến tại Trường Đại học Cần Thơ và được minh họa trong hình 5.



Hình 5. Mô hình mạng các hàng chờ của Hệ thống đăng kí học phần trực tuyến tại Trường Đại học Cần Thơ

3.2. Tính các tham số ổn định không phụ thuộc lưu lượng luồng

3.2.1. Hệ số kinh nghiệm

Do chỉ khảo sát chỉ số tải khi có đám đông đăng kí học phần nên ở đây ta chỉ quan tâm đến các tham số ổn định liên quan đến các trung tâm 2.1, 2.2 và 3.

Các tham số ổn định được quan tâm gồm:

- Xác suất p (xác suất để mỗi sinh viên hoàn thành điền mẫu đăng kí học phần và đi vào Trung tâm Đăng kí học phần);
- Thời gian S xử lí một mẫu đăng kí tại Trung tâm 2.1 hoặc 2.2 và Thời gian S_{CSDL} cập nhật Đăng kí học phần tại Trung tâm 3.

Bình thường nếu không có dữ liệu Log ghi lại các biến cố phát sinh tại các trung tâm người ta phải khảo sát kinh nghiệm của các nhà quản lí hệ thống và chấp nhận sai số nào đó. Đối với hệ thống đăng kí học phần tại Trường Đại học Cần Thơ, chúng tôi

khảo sát kinh nghiệm của các nhà quản lí hệ thống và chỉ nhận được 3 dữ liệu trung bình sau:

- Mỗi máy chủ Đăng kí học phần chạy chậm khi có khoảng 400 sinh viên đăng nhập đồng thời;
- Máy chủ CSDL chạy chậm khi có khoảng 500 sinh viên đăng nhập đồng thời;
- Thời gian trung bình mà hệ thống hoàn tất xử lí và cập nhật một mẫu đăng kí học phần dao động từ 2 - 4 giây. Như vậy thời gian trung bình là khoảng 4 giây khi toàn bộ hệ thống đạt ngưỡng đăng nhập tối đa $N = 500$ sinh viên.

Để tính các tham số ổn định p , S và S_{CSDL} cần xây dựng hệ ba phương trình cho các tham số này. Mặt khác như đã khảo sát ở phần trên: “sự chậm chạp” của hệ thống tương ứng với “sự chậm chạp” tại một trung tâm phục vụ nào đó và ở đó sẽ có chỉ số sử dụng $\varphi \in [0.7, 0.8]$. Nhưng ta lại cần một hệ số tương ứng với sự chậm chạp mà không phụ thuộc vào một trung tâm phục vụ nào. Ta đặt:

$$k \in [0.7, 0.8]$$

là một hệ số chỉ “sự chậm chạp” của một hệ thống và được gọi là hệ số kinh nghiệm. Như vậy, nếu xét “sự chậm chạp” xảy ra tại một trung tâm Đăng kí học phần thì $\varphi = k$. Tương tự, nếu xảy ra tại Trung tâm CSDL thì $\varphi_{CSDL} = k$.

3.2.2. Xây dựng Phương trình 1

Giả sử số sinh viên đăng nhập đồng thời tại một máy chủ đăng kí học phần đạt mức tối đa: $N_1 = 400$ sinh viên. Khi đó chỉ số tải tương ứng với thời gian xử lí có dấu hiệu chậm chạp tại máy chủ Đăng kí này là $\varphi = k$. Công thức (1) áp dụng cho Trung tâm 2.1 (hoặc 2.2) dẫn đến chỉ số tải: $\varphi = pSN_1$ và nhận được phương trình thứ nhất:

$$400pS = k \quad (4).$$

3.2.3. Xây dựng Phương trình 2

Tương tự đối với Trung tâm 3, ta sẽ xét trường hợp có $N = 500$ đăng kí học phần cần cập nhật và được phát sinh từ hai trung tâm 2.1 và 2.2. Khi đó có $N_1=250$ sinh viên đăng nhập đồng thời tại mỗi trung tâm 2.1 và 2.2. Mặt khác lưu lượng truy cập CSDL xuất phát từ hai trung tâm 2.1 và 2.2 và được bảo toàn theo luật Ép luồng (Forced Flow Law) trên mạng các hàng chờ [4], do đó lưu lượng này bằng: $\lambda_{CSDL} = pN_1 + pN_1 = 2pN_1 = pN$.

Từ công thức (1), ta có tiếp: $\varphi_{CSDL} = \lambda_{CSDL}S_{CSDL} = pNS_{CSDL}$ và suy ra phương trình thứ 2:

$$500pS_{CSDL} = k \quad (5).$$

3.2.4. Xây dựng Phương trình 3

Để tính R^* , ta sử dụng tổng số sinh viên đăng nhập là $N = 500$ kết nối các trung tâm 2.1, 2.1 và 3, khi đó sẽ có số $N_1 = 250$ sinh viên đăng nhập đồng thời lưu trú tại mỗi trung tâm Trì hoãn 1.1 và 1.2.

Bây giờ sử dụng công thức (3) cho Trung tâm 2.1 (hoặc 2.2) khi $N_1 = 250$ dẫn đến:

$$R = \frac{S}{1 - pSN_1} \text{ hay: } R = \frac{S}{1 - 250pS} \quad (6).$$

Còn áp dụng cho Trung tâm 3 với $N = 500$, ta có tiếp:

$$R_{CSDL} = \frac{S_{CSDL}}{1 - pS_{CSDL}N} \text{ hay: } R_{CSDL} = \frac{S_{CSDL}}{1 - 500pS_{CSDL}} \quad (7).$$

Suy ra:

$$R^* = 2R + R_{CSDL} = \frac{2S}{1 - 250pS} + \frac{S_{CSDL}}{1 - 500pS_{CSDL}} = 4(\text{giây})$$

Ta nhận được phương trình thứ 3:

$$\frac{2S}{1 - 250pS} + \frac{S_{CSDL}}{1 - 500pS_{CSDL}} = 4 \quad (8).$$

3.2.3. Giải hệ phương trình

Tổng hợp các kết quả ta có hệ ba phương trình (4), (5) và (8). Sử dụng phương pháp thế để giải hệ phương trình này.

Từ phương trình (4) ta tính: $p = \frac{k}{400S} \quad (9).$

Kết hợp (9), (5) và (8) nhận được phương trình tính

$$S: \frac{16S}{8 - 5k} + \frac{4S}{5(1 - k)} = 4 \quad (10),$$

suy ra:

$$S = \frac{5(8 - 5k)(1 - k)}{28 - 25k} \quad (11).$$

Sau khi tính được S từ (11), ta sẽ tính p và S_{CSDL} từ các công thức (9) và (10).

3.3. Xây dựng các hàm liên quan đến chỉ số tắc nghẽn biến thiên theo N

Do vai trò của N trong các hệ thức (1) và (2) xét tại Trung tâm 2.1 hoặc 2.2 là N_1 , nên:

$$N_1(\varphi) = \varphi/pS \text{ và } R(N_1) = \frac{S}{1 - pSN_1}. \text{ Nhưng tổng số } N = 2N_1, \text{ suy ra:}$$

$$N(\varphi) = 2\varphi/pS \text{ và } R(N) = \frac{2S}{2 - pSN}.$$

Thay (9) vào các hệ thức trên, ta nhận được:

$$N(\varphi) = \frac{800\varphi}{k} \quad (1^*)$$

và
$$R(N) = \frac{800S}{800 - kN} \tag{2*}$$

Áp dụng tiếp các hàm (1) và (2) cho Trung tâm 3, ta có tiếp:

$$N_{CSDL}(\varphi_{CSDL}) = \frac{\varphi_{CSDL}}{pS_{CSDL}} \text{ và } R_{CSDL}(N) = \frac{S_{CSDL}}{1 - pS_{CSDL}N}.$$

Sau khi thay thế (9) và (10) vào các hệ thức trên, ta có kết quả:

$$N_{CSDL}(\varphi_{CSDL}) = \frac{500\varphi_{CSDL}}{k} \tag{1**}$$

và
$$R_{CSDL}(N) = \frac{400S}{500 - kN} \tag{2**}$$

Cuối cùng hệ thức (3) trong trường hợp này có dạng:

$$R^*(N) = 2R(N) + R_{CSDL}(N) \tag{3*}.$$

3.4. Khảo sát các hàm liên quan đến chỉ số tắc nghẽn

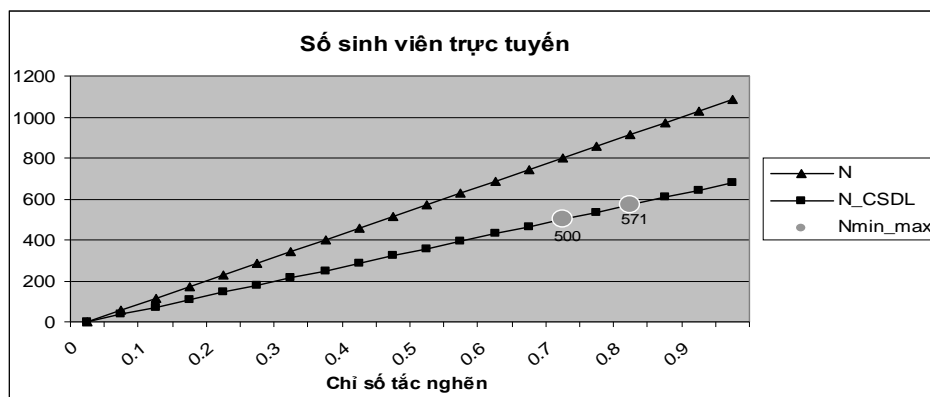
3.4.1. Khảo sát các hàm (1*) và (1**)

Mục tiêu của bài toán khảo sát này nhằm xác định khoảng biến động của số lượng sinh viên tham gia trực tuyến N tương ứng với chỉ số tắc nghẽn $\varphi \in [0.7, 0.8]$.

Sử dụng các hàm $N(\varphi)$ và $N_{CSDL}(\varphi_{CSDL})$ từ các hệ thức (1*) và (1**) để vẽ hai đồ thị biến thiên trên cùng khoảng (0, 1).

a. Trường hợp $k = 0.7$

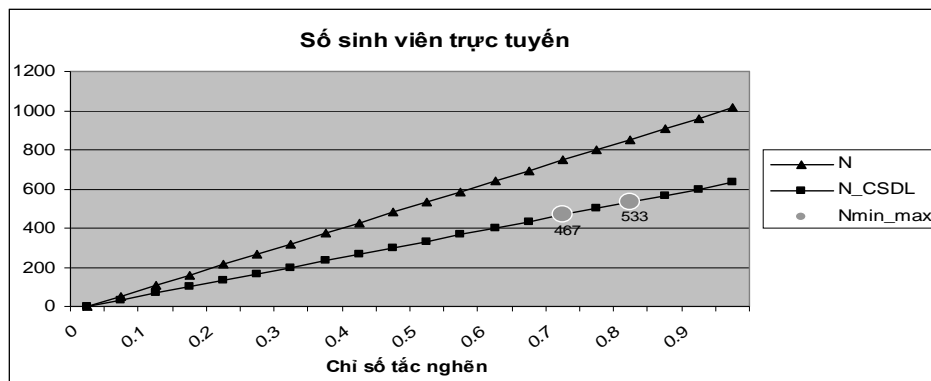
Trong đồ thị hình 6, các đường N và N_CSDL biểu diễn tương ứng các hàm $N(\varphi)$ và $N_{CSDL}(\varphi_{CSDL})$, còn các điểm Nmin_max biểu diễn các số lượng sinh viên trực tuyến tương ứng với khoảng biến thiên của $\varphi \in [0.7, 0.8]$. Các giá trị Nmin_max có thể sử dụng để dự báo nguy cơ tắc nghẽn hệ thống dưới giả thiết hệ số kinh nghiệm $k = 0.7$.



Hình 6. Đồ thị biểu diễn số sinh viên đăng nhập biến thiên theo chỉ số tắc nghẽn với $k = 0.7$

Các giá trị N_{min_max} không thể lấy các giá trị tương ứng trên đường N do các số lượng này lớn hơn các giá trị tương ứng trên đường N_CSDL , do đó chúng có thể gây tắc nghẽn Trung tâm CSDL dẫn đến tắc nghẽn toàn bộ hệ thống. Như vậy các giá trị đăng nhập tối đa có thể được phép trong trường hợp này là N nằm trong dãy giá trị từ 500 đến 571.

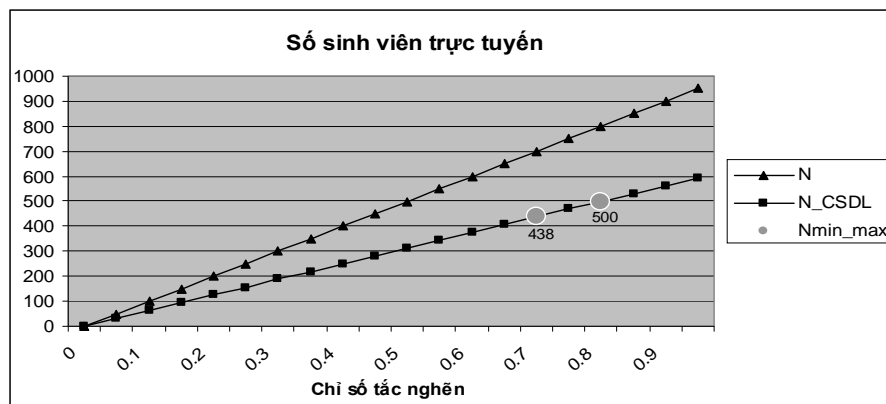
b. Trường hợp $k = 0.75$



Hình 7. Đồ thị biểu diễn số sinh viên đăng nhập biến thiên theo chỉ số tắc nghẽn với $k = 0.75$

Kết quả biến thiên trên đồ thị cho thấy số N đăng nhập tối đa có thể biến thiên từ 467 đến 533.

c. Trường hợp $k = 0.8$



Hình 8. Đồ thị biểu diễn số sinh viên đăng nhập biến thiên theo chỉ số tắc nghẽn với $k = 0.8$

Trường hợp này cho kết quả N tối đa có thể cho phép trực tuyến biến thiên trong khoảng từ 438 đến 500.

Ba kết quả trên cho thấy có thể sử dụng số lượng đăng kí trực tuyến tối đa biến thiên từ 438 đến 571. Tùy tốc độ tăng số lượng khách hàng trực tuyến mà chúng ta nên

sử dụng giá trị N tối đa trực tuyến phù hợp với thời gian chuẩn bị đối phó với trường hợp tắc nghẽn. Chẳng hạn nếu tốc độ khách hàng tăng số lượng nhiều nên sử dụng giá trị $N = 438$ tối đa cho phép trực tuyến, còn nếu tốc độ tăng ở mức trung bình nên lấy $N = 500$ tối đa cho phép trực tuyến.

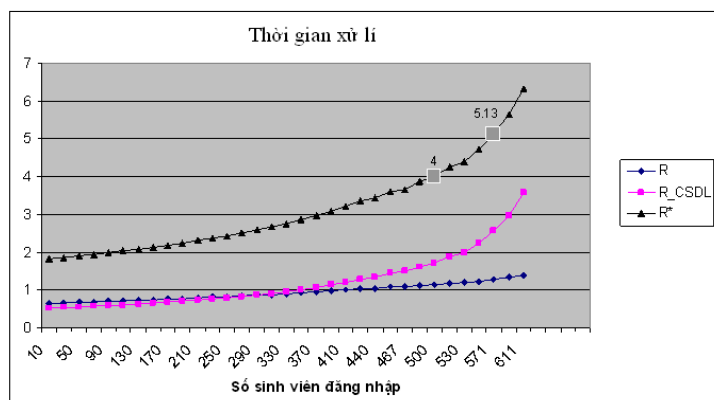
3.4.1. Khảo sát các hàm (2*), (2**) và (3*)

Mục tiêu của bài toán khảo sát các hàm $R(N)$, $R_{CSDL}(N)$ và $R^*(N)$ nhằm chỉ ra xu hướng tiến tới vô cùng của các thời gian xử lý tại từng trung tâm phục vụ và thời gian hoàn tất xử lý của toàn bộ hệ thống tương ứng với số lượng N khách hàng trực tuyến. Từ đây có thể thấy rõ hơn việc sử dụng các giá trị N tối đa cho phép trực tuyến từ kết quả khảo sát (1*) và (1**) là hợp lí.

a. Trường hợp $k = 0.7$

Trong đồ thị hình 9, các đường R , R_{CSDL} và R^* biểu diễn các hàm số tương ứng $R(N)$, $R_{CSDL}(N)$ và $R^*(N)$. Các giá trị đánh dấu trên đường R^* tương ứng với các số số lượng N khách hàng đăng nhập trực tuyến chạy từ 500 đến 571.

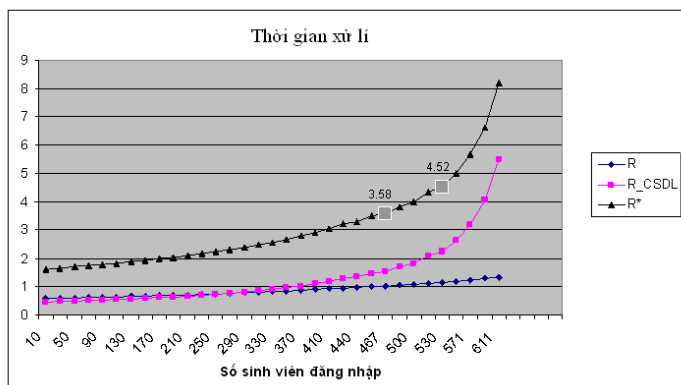
Kết quả khảo sát thời gian xử lý trong trường hợp $k = 0.7$ cho thấy thời gian R^* hoàn tất một đăng kí học phần tương ứng với các số lượng đăng nhập tối đa biến động từ 4 giây đến 5.13 giây. Quan sát các giá trị này cho thấy chúng đang nằm ở cung uốn chuyển sang tăng gấp đến vô cùng của đường Hyperbol. Hình ảnh tương tự của R_{CSDL} cho thông tin về nút tắc nghẽn của hệ thống nằm ở Trung tâm CSDL. Trong trường hợp này việc xác định N tối đa trong khoảng 500-571 là hợp lí.



Hình 9. Đồ thị biểu diễn Thời gian xử lý biến thiên theo số sinh viên đăng nhập với $k = 0.7$

b. Trường hợp $k = 0.75$

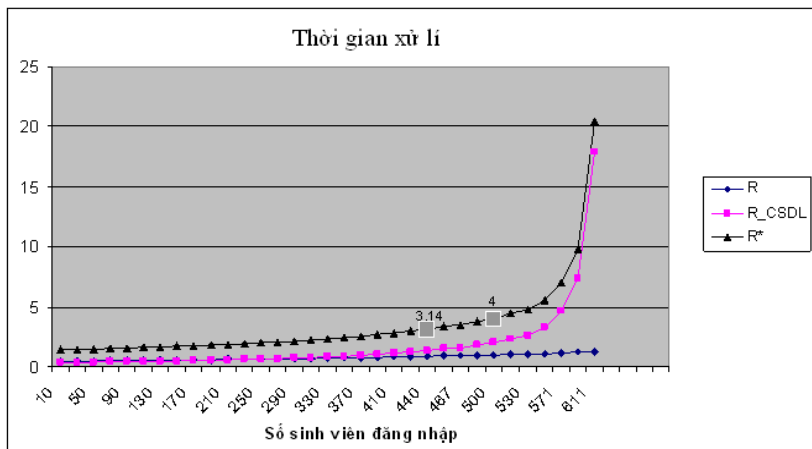
Kết quả được biểu diễn trong hình 10 cho thấy thời gian trung bình R^* mà hệ thống hoàn tất xử lý một đăng kí học phần nằm trong khoảng từ 3.58 giây đến 4.52 giây tương ứng với số lượng N đăng nhập tối đa cho phép trong khoảng 467-533. Trên đường R^* khoảng thời gian này cũng nằm trên cung uốn theo xu hướng tiến đến vô cùng của đường Hyperbol.



Hình 10. Đồ thị biểu diễn Thời gian xử lý biến thiên theo số sinh viên đăng nhập với $k = 0.75$

c. Trường hợp $k = 0.8$

Trường hợp này cho kết quả R^* biến động từ 3.14 giây đến 4 giây tương ứng với số sinh viên N đăng nhập tối đa trong khoảng [438, 500].



Hình 11. Đồ thị biểu diễn Thời gian xử lý biến thiên theo số sinh viên đăng nhập với $k = 0.8$

Kết quả khảo sát cả ba trường hợp cho thấy số lượng N sinh viên cho phép đăng nhập tối đa càng cao thì thời gian R^* hoàn tất ở các ngưỡng tắc nghẽn cũng cao theo. Ngưỡng thời gian trước khi tắc nghẽn cao nhất là $\text{Max}(R^*) = 5.13$ giây và thấp nhất là $\text{Min}(R^*) = 3.14$ giây.

Kết quả này cũng có thể hỗ trợ Nhà Quản trị hệ thống quan tâm xem xét sự không bình thường của các khách hàng có thời gian xử lý vượt quá 5 giây.

4. Kết luận

Các kết quả khảo sát cho thấy sự hữu hiệu của phương pháp khảo sát các tham số đặc trưng cho một hệ thống xử lý trực tuyến. Đặc biệt, phương pháp này đã sử dụng thông tin kinh nghiệm từ người quản trị và lượng hóa chúng thành hệ số kinh nghiệm k trượt trong khoảng $[0.7, 0.8]$.

Quy trình sử dụng phương pháp cần phải trải qua bốn bước:

1. Xây dựng chính xác mô hình mạng các hàng chờ từ kết quả phân tích chức năng của hệ thống xử lý trực tuyến.
2. Tính các tham số ổn định không phụ thuộc lưu lượng luồng, gồm: p và S .
3. Xây dựng các hàm khảo sát liên quan đến chỉ số tắc nghẽn mạng.
4. Khảo sát các hàm liên quan đến chỉ số tắc nghẽn mạng.

Các bước này đã được xử lý chi tiết từ các giải pháp lý thuyết đến thực nghiệm trên Hệ thống Đăng ký học phần trực tuyến của Trường Đại học Cần Thơ.

Giải pháp này có thể ứng dụng để khảo sát tất cả các hệ thống trực tuyến hiện nay. Kết quả khảo sát có thể hỗ trợ tìm ra các giải pháp ngăn chặn có hiệu quả các tình huống tắc nghẽn, tối ưu hóa các thiết kế mạng và đặc biệt có thể hỗ trợ các giải pháp phân tích và phòng ngừa các cuộc tấn công DDoS của tin tặc.

TÀI LIỆU THAM KHẢO

1. Lê Quyết Thắng, Phạm Nguyễn Khang (2013), *Lý thuyết xếp hàng và ứng dụng đánh giá hệ thống*, Nxb Đại học Cần Thơ.
2. S. Athuraliya, V.H. Li, S.H. Low, Q. Yin. (2001), REM: Active Queue Management, *IEEE Network*, 15(3), pp. 48-53.
3. J. Aweya, M. Ouellette, D.Y. Moutuno (2001), A Control Theoretic Approach to Active Queue Management, *Computer Network*, 36 (2-3), pp. 203-235.
4. P. J. Denning, J. P. Buzen (1978), The Operational Analysis of Queueing Network Models, *Computing Surveys*, Vol. 10, No 3, pp. 225-261.
5. S. Floyd, G. Ramakrishna, S. Shenker (2001), Adaptive RED: an Algorithm for Increasing the Robustness of RED's Active Queue Management, Technical Report, ICSI.

6. AJ Hussein, W. Mike, T. Fadi, AA. Amer (2007), Performance Evaluation for DRED Discret Time Queueing Network Analytical Model, Journal of Network and Computer Application, 31, pp. 750-770.
7. E. D. Lazowska, J. Zahorjan, G. S. Graham, K. C. Sevcik (1984), Quantitative System Performance, Computer System Analysis Using Queueing Network Models, Prentice-Hall, Inc.

*(Ngày Tòa soạn nhận được bài: 19-7-2013; ngày phản biện đánh giá: 21-11-2013;
ngày chấp nhận đăng: 13-12-2013)*