NGHIÊN CỨU CÁC YẾU TỐ TÁC ĐỘNG VÀO TẢI CỦA HỆ THỐNG WEB ĐỂ GIẢI QUYẾT BÀI TOÁN TẮC NGHẼN TRONG MẠNG WEB LỚN

RESEARCHING THE FACTORS THAT IMPACT ON THE LOAD OF THE WEB SITE SYSTEM TO SOLVE CONGESTION IN THE NETWORK OF LARGE WEB

Vũ Văn Đốc

Khoa Công nghệ thông tin, Trường Đại học Kinh tế - Kỹ thuật Công nghiệp

Đến Tòa soạn ngày 24/4/2015, chấp nhận đăng ngày 20/5/2015

Tóm tắt: Trong những năm gần đây đã có rất nhiều công nghệ cải tiến cơ sở hạ tầng mạng internet, tuy nhiên, đối với các hệ thống Web lớn trên thế giới việc tìm ra một giải pháp chống tắc nghẽn khi số lượng người truy cập lớn vẫn luôn là một trong những thách thức đối với nhà cung cấp dịch vụ web. Vì vậy việc tìm kiếm, đánh giá các giải pháp hiện có để có thể áp dụng vào giải quyết vấn đề tắc nghẽn trong mạng web lớn luôn là vấn đề được quan tâm hàng đầu nhằm đáp ứng lượng người truy nhập ngày càng tăng.

Bài báo đề cập đến các yếu tố tác động vào tải của hệ thống web nhằm tìm kiếm, đánh giá các giải pháp hiện có để giải quyết bài toán tắc nghẽn trong mạng web lớn.

Từ khóa: Cân bằng tải, mạng web lớn.

Abstract: There have been many technologies that improve the infrastructure of internet in the recent years. However, for the large web systems in the world, finding out a solution for preventing the obstruction of the internet when having a large numbers of people accessing is the challenge for the web suppliers. So seaching and evaluating the current solutions in order to be able to apply for solving the obstruction of the big network is always the first concerve to meet the assessers increasing daily.

In this article, I study the factors effecting on loading the web system on order to finding, evaluating the curent sollusions to solve the obstruction of the big network.

Keywords: Load balancing, large web.

1. TỔNG QUAN VỀ CÂN BẰNG TẢI

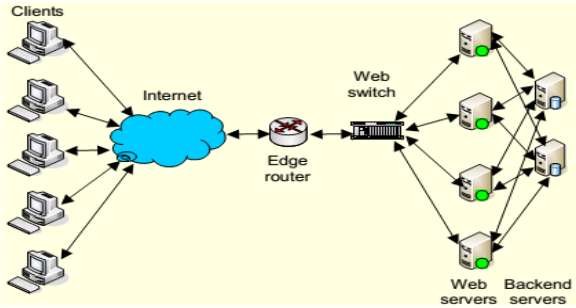
1.1. Giới thiệu

Với số lượng người sử dụng web ngày càng cao đòi hỏi các ứng dụng web phải có khả năng chạy trên nhiều máy chủ (web server). Cân bằng tải cải thiện khả năng mở rộng của một ứng dụng hoặc cụm máy chủ bằng cách phân phối tải trên nhiều máy chủ, đồng thời cũng có thể hướng lưu lượng truy nhập đến các máy chủ khác nếu một máy chủ hoặc ứng dụng gặp sự cố.

Cân bằng tải cung cấp các giải pháp cải thiện an toàn bằng cách bảo vệ các cụm máy chủ với nhiều hình thức của tấn công từ chối dịch vụ (DoS).

Có nhiều cách phân loại cân bằng tải khác nhau, tuy nhiên khi phân loại cân bằng tải theo lớp giao thức trong mô hình OSI thì cân

bằng tải được chia làm 2 loại đó là cân bằng tải không nhận biết nội dung và cân bằng tải nhận biết nội dung.



Hình 1. Mô hình cân bằng tải lưu lượng

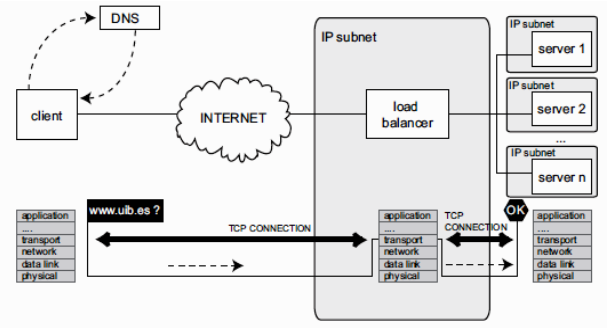
1.2. Cân bằng tải không nhận biết nội dung

Cân bằng tải không nhận biết nội dung là bộ cân bằng tải không nhận biết thông tin của ứng dụng chứa trong các yêu cầu gửi đến và thường được nhắc tới như là các bộ cân bằng tải ở lớp 4 (lớp transport).

Cân bằng tải không nhận biết nội dung cho phép thiết bị front-end nhận biết về các kết nối TCP giữa các client và server. Do đó, bộ cân bằng tải gửi đi các yêu cầu theo địa chỉ IP và cổng TCP.

1.3. Cân bằng tải nhận biết nội dung

Bộ cân bằng tải nhận biết nội dung làm việc tại lớp ứng dụng. Bộ cân bằng tải nhận biết các nội dung ứng dụng của các yêu cầu gửi tới. Điều này làm cho việc định tuyến nhận biết nội dung cụ thể hơn đối với các ứng dụng có thể cung cấp các dịch vụ khác biệt, nhưng cũng phức tạp hơn so với phương pháp tiếp cận không nhận biết nội dung [2].



Hình 2. Cân bằng tải theo kiến trúc 2 chiều

Một số giải pháp nhận biết địa phương đã được đề xuất như các chính sách phân phối trong thiết kế cân bằng tải nhận biết nội dung.

2. PHÁT HIỆN VÀ GIÁM SÁT BURSTINES TRONG HỆ THỐNG WEB

2.1. Tính truyền loạt và kiểm soát truy nhập

Khi các yêu cầu truy nhập tới hệ thống máy chủ web tăng đột biến, đó là hiện tượng truy nhập đồng loạt hay tính truyền loạt (burstiness). Đây là một trong những nguyên nhân quan trọng gây tắc nghẽn hệ thống web. Do đó việc phát hiện burstiness được xem là vấn đề quan trọng trong việc duy trì và cải thiện hiệu năng của hệ thống. Burstiness được phát hiện trong một số trường hợp:

- Dựa trên lưu lượng mạng.
- Dựa trên giao thức TCP.
- Trong cơ sở dữ liệu.
- Dựa trên số lần phục vụ.

2.2. Các yếu tố Burstiness

Việc giám sát hệ thống được thực hiện bằng cách sử dụng lập khe thời gian thích ứng dựa trên yếu tố burstiness. Ta sẽ xem xét sáu phương pháp tiếp cận khác nhau nhằm xác định các yếu tố burstiness để so sánh với tác động của chúng và phát hiện những ưu nhược điểm của chúng theo các trường hợp tương ứng.

2.2.1. Yếu tố burstiness 1 (BF1)

Yếu tố Burstiness đầu tiên được đề xuất bởi Menasc'e và Almeida. Yếu tố này đòi hỏi phải biết tỷ lệ đến trung bình của các giao dịch HTTP cho một máy chủ web, được đo trong khoảng thời gian 0, 1, ..., k - 1, ký hiệu là μ_λ. Đối với một khe k và một máy chủ web, $\lambda(k)$ đại diện cho tỷ lệ xuất hiện tương ứng của nó. Nếu $\lambda(k) > \mu_\lambda$, khe đó được coi là một khe bursty.

Yếu tố burstiness này thể hiện được tỷ lệ lưu

lượng đến nhưng không thể hiện được các thay đổi nhanh chóng của nó đối với hệ thống.

2.2.2. Yếu tố burstiness 2 (BF2)

Yếu tố burstiness thứ hai được coi là một trong những sửa đổi BF1, bằng cách tạo sự khác biệt tương đối của tỷ lệ xuất hiện của hai khe trước đó. Do đó, việc sửa đổi yếu tố burstiness cũng phụ thuộc vào tỷ lệ đến tăng hoặc giảm. Trong trường hợp này, yếu tố burstiness cũng khác nhau với các tỉ lệ đến khác nhau.

2.2.3. Yếu tố burstiness 3 (BF3)

Yếu tố burstiness biểu diễn cho đỉnh tăng lưu lượng truy nhập đến tới một máy chủ web. Do đó, chúng ta xem xét mức tối đa của j khe bursty liên tiếp sẽ gây ra sự gia tăng tỷ lệ trong các yếu tố burstiness. Đây là lý do cản trở trong yếu tố burstiness, phụ thuộc vào một bản ghi của các khe bursty trước đó.

2.2.4. Yếu tố burstiness 4 (BF4)

Đề xuất này bao gồm sự khác biệt lớn nhất của mức tương đối giữa hai khe, do đó sẽ cho phép các yếu tố burstiness làm theo tất cả các biến thể tỷ lệ đến và sự cản trở. Trong trường hợp này, yếu tố burstiness cũng nhạy cảm với những thay đổi trong tỷ lệ đến.

2.2.5. Yếu tố burstiness 5 (BF5)

Sử dụng phép ngoại suy tuyến tính của tỷ lệ đến để phát hiện các khe bursty thay vì mức trung bình của các tỷ lệ đến, như Baryshnikov mô tả [4]. Tính toán dự báo tỷ lệ đến trong các khe tiếp theo cho một máy chủ web với t(k) là thời gian cuối cùng của khe k.

2.2.6. Yếu tố burstiness 6 (BF6)

Đề xuất cuối cùng, được giới thiệu bởi Wang [3], đã được xem xét để phân tích tác động của nó và so sánh với các đề xuất khác. Sự khác biệt chính giữa các đề xuất này với tất cả những đề xuất trước đây là yếu tố này được tính cho từng yêu cầu HTTP gửi đến một máy chủ web.

2.3. Lập khe thời gian thích nghi

Sau khi các yếu tố burstiness được xác định, chúng được sử dụng để thiết lập khe thời gian thích nghi. Khoảng thời gian của khe tiếp theo được xác định bởi giá trị của yếu tố burstiness trên khe hiện thời, khi truyền loạt tăng lên được phát hiện, khoảng thời gian của khe tiếp theo được giảm để kiểm tra tỷ lệ đến đến sớm và sau đó điều chỉnh lại các tham số của thuật toán. Nếu sự truyền loạt giảm đi, khoảng thời gian của khe tiếp theo được mở rộng để giảm chi phí. Bằng cách kiểm soát burstiness ở tỉ lệ đến, và khoảng thời gian các khe thử nghiệm, việc giảm đột ngột hiệu năng trong tương lai của các máy chủ web có thể được dự báo.

3. THUẬT TOÁN CÂN BẰNG TẢI VÀ ĐIỀU KHIỂN TRUY NHẬP

3.1. Tối thiểu hóa chi phí

Phương pháp điều khiển truy nhập và thuật toán cân bằng tải dựa trên dự báo thông lượng đối với một hệ thống web. Để thuật toán có độ phức tạp thấp, thời gian gọi xử lý phải dựa trên tỷ lệ lưu lượng truy cập đến. Đồng thời định nghĩa việc lập khe thời gian thích ứng thông qua tập các yêu cầu truy nhập thường xuyên đến máy chủ web. Sau đó, sử dụng việc lập khe thời gian thích ứng để đặt thời gian cho việc gọi thuật toán kiểm soát truy nhập và cân bằng tải. Tính truyền loạt được phát hiện trong hệ thống ảnh hưởng đến quá trình thực hiện thuật toán theo cách sau: khi các yếu tố truyền loạt tăng lên, thì thuật toán được gọi thường xuyên hơn và ngược lại.

Có hai phương pháp để gọi thuật toán, đó là thực hiện gọi định kỳ và không định kỳ.

3.2. Tổng quan về thuật toán cân bằng tải

Có năm yếu tố dự đoán lưu lượng truy cập, năm yếu tố này cung cấp cho chúng ta các xu hướng tác động đến hệ thống, từ đó quyết định đến quá trình duy trì việc thực hiện thuật toán cân bằng tải trong hệ thống. Nghiên cứu này xem xét đến tính ưu tiên khác nhau và các loại yêu cầu khác nhau. Các yêu cầu Service Level Agreement (SLA) được xác định trong khái niệm mức độ sử dụng CPU của các máy chủ web.

Các tham số sử dụng trong thuật toán

CPU là nguyên nhân chính gây ra "nghẽn cổ chai" của hệ thống Web khi nội dung động được yêu cầu. Do đó, mức độ sử dụng CPU được xem như là thước đo chính để ước tính và kiểm soát hiệu suất của hệ thống web.

Các số liệu được sử dụng nhằm thực thi thuật toán cân bằng tải:

- Tỷ lệ lưu lượng truy cập đến hệ thống cho một khe *k*, *λ(k)*, được tính bằng cách chia tổng số yêu cầu gửi đến khe *k* theo thời gian, *d(k)*.
- Số lần phục vụ cần thiết để xử lý các yêu cầu tĩnh và động, thu được từ máy chủ web. Số lần phục vụ trung bình tại mỗi khe là *δ(k)*, và sử dụng đại lượng này để ước tính mức độ sử dụng các máy chủ web.
- Thông lượng trung bình tại một khe *k* là *x(k)*, cũng được xác định từ máy chủ web, và được sử dụng để ước lượng thông lượng của các máy chủ web trong khe tiếp theo.
- Mức sử dụng trung bình CPU của máy chủ web *u(k)*, được sử dụng như một yếu tố trong biểu thức thực hiện dự báo thông lượng, và cũng được sử dụng để kiểm soát lỗi cho dự báo thông lượng.

Số lần phục vụ và điều chỉnh dự báo của mức độ sử dụng trong các máy chủ dựa trên thông lượng đã được dự báo bởi vì, số lần phục vụ sẽ tăng cao khi máy chủ bắt đầu quá tải.

3.3. Dự báo thông lượng

Trước hết, thuật toán đề xuất được xây dựng dựa trên quá trình dự báo lưu lượng truy nhập đến hệ thống web. Theo các nghiên cứu trước đây, quá trình thực hiện dự báo lưu lượng truy cập được chia thành 3 loại và được trình bày trong tài liệu [7].

3.3.1. P1: based on filtering - phương pháp dự báo dựa trên quá trình lọc

Phương pháp dự báo này là trung bình di chuyển giữa giá trị thông lượng đã được thiết lập trong khe cuối cùng và giá trị trung bình của thông lượng thực được đo trong hai khe cuối cùng.

$$\hat{x}_1(k+1) = (1 - a(k+1)).\hat{x}_1(k) + a(k+1).\frac{2}{\frac{1}{x(k)} + \frac{1}{x(k-1)}} \quad (1)$$

a(k+1) là đại lượng thể hiện tần suất xuất hiện trọng số cân bằng cho các biểu thức trong hàm tính toán khoảng thời gian của khe kế tiếp *d(k+1)*, vì thế, đại lượng này gián tiếp phụ thuộc vào các yếu tố burstiness: $a(k+1) = \frac{2.T - d(k+1)}{2.T + d(k+1)}$

Vì vậy, quá trình ước lượng thông lượng phụ thuộc vào hai điều kiện: dự đoán thông lượng cuối cùng được dự báo và giá trị trung bình của thông lượng trước đó. Kết quả của việc tính toán này cho phép lọc các giá trị thông lượng truy cập dựa trên xác suất xuất hiện tương ứng với *a(k+1)*.

Lợi ích chính của dự báo này là làm mịn đỉnh lưu lượng để đảm bảo hiệu suất xử lý của các máy chủ trong thời gian dài.

3.3.2. P2: based on burstiness - Dựa trên burstiness

Quá trình dự báo này thực hiện dựa trên các yếu tố burstiness được mô tả ở trên. Quá trình được thực hiện thông qua việc xác định thừa số xu hướng (factorise the tendency) của các

burstiness trong hai giai đoạn cuối với các biến đổi thông lượng là $\beta(k+1)$:

$$\beta(k+1+=(b(k)-b(k-1)).|x(k)-x(k-1)| \quad (2)$$

Yếu tố này sẽ xét thông lượng khác nhau trong hai khe cuối. Yếu tố này được tính toán dựa trên việc xác định giá trị sai khác của yếu tố burstiness hiện tại và yếu tố burstiness trước đó. Máy chủ sử dụng yếu tố này để xác định khả năng tăng hoặc giảm thông lượng qua khe trước đó:

$$\hat{x}_2(k+1)=x(k)-(\beta(k+1).u(k)) \quad (3)$$

Mục tiêu của quá trình ước tính này là giảm dự báo thông lượng truy cập khi tính truyền loạt của các yêu cầu được phát hiện, quá trình này phụ thuộc vào hiệu suất sử dụng hiện hành của máy chủ.

3.3.3. P3: based on filtering and burstiness - dựa trên bộ lọc và tính truyền loạt

Phương pháp này được thự hiện dựa trên sự kết hợp cả hai phương pháp P1 và P2. Phương pháp này được đề xuất để thực hiện cân bằng các yếu tố dự báo khả năng thực hiện các phương pháp **P1** và **P2** có thể có trong hệ thống.

$$\hat{x}_3(k+1)=\frac{2}{\frac{1}{\hat{x}_1(k+1)}+\frac{1}{\hat{x}_2(k+1)}} \quad (4)$$

Để đạt được kết quả tốt hơn, tôi đã mở rộng nghiên cứu nhằm cải thiện các phương pháp vốn có và đề xuất thêm hai phương pháp thực hiện dự báo lưu lượng truy cập mới.

3.3.4. P4: based on Least Mean Square (LMS) - dựa trên bình phương trung bình nhỏ nhất

Thuật toán LMS [5] được sử dụng để dự báo lưu lượng truy cập. LMS giới thiệu một thủ tục lặp cho phép chỉnh liên tiếp đối với vector trọng số, qua đó giảm thiểu giá trị sai khác trung bình.

Đặt $\underline{w}(k+1)$ là đại lượng đặc trưng cho các vector trọng số của bộ lọc LMS đối với khe *k*. Toán tử $\underline{w}(k+1)$ có thể được biểu diễn thông qua biểu thức đệ quy sau:

$$\underline{w}(k+1)=\underline{w}(k)+\mu.\left[x(k)-\hat{x}_4(k)\right].\underline{x}(k) \quad (5)$$

trong đó *M* là số các đại lượng trọng số được sử dụng trong bộ lọc thích nghi, μ là thông số kích thước bước. Vectơ $w(k)$ và $x(k)$ được xác định như sau:

$$\underline{w}(k)=\left[w_0,w_1,...,w_{M-1}(k)\right]^T$$

$$\underline{x}(k)=\left[x(k),x(k-1),....,x(k-M+1)\right]^T$$

Giá trị dự đoán thông lượng truy cập thu được bằng cách dự báo tuyến tính theo biểu thức:

$$\hat{x}_4(k+1)=\sum_{x=0}^{M-1}w(x).x(k-x) \quad (6)$$

3.3.5. P5: based on Normalised Least Mean Square (NLMS) - dựa trên bình phương trung bình nhỏ nhất thông thường

Phiên bản chuẩn hóa của bộ lọc LMS [6] đã được đề xuất nhằm tránh sự nhạy cảm (sensitivity) của thuật toán LMS đối với các loại dữ liệu đầu vào $x(k)$ khác nhau. Giải pháp thực hiện chuẩn hóa biểu thức trước đó bằng cách chia véc tơ $x(k)$ cho bình phương Euclide của nó.

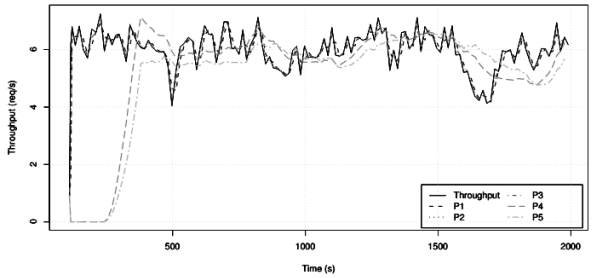
$$\underline{w}(k+1)=\underline{w}(k)+\mu.\left[x(k)-\hat{x}_5(k)\right].\frac{\underline{x}(k)}{\|\underline{x}(k)\|^2} \quad (7)$$

3.3.6. Các kết quả dự đoán thông lượng

Thử nghiệm mô phỏng trong OPNET Modeler để mô tả các tác động của những yếu tố dự báo. Các kết quả được thể hiện trong hình 3 bao gồm lưu lượng truy cập của

một máy chủ kèm theo năm quá trình dự báo lưu lượng truy cập được giới thiệu ở trên. Qua đó, ta có thể thấy được cách thức các phương pháp thực hiện dự đoán P1 - P3 tương ứng với các lưu lượng truy cập kể từ khi bắt đầu quá trình mô phỏng. Sự tham gia của phương pháp P4-P5 thực hiện dự báo dựa trên lý thuyết lọc. Do đó, các phương pháp này cần nhiều khe hơn để có được một ước lượng tốt nhất. Trong trường hợp của các kết quả mô phỏng trình bày trong hình 3, toàn bộ quá trình dự báo thông lượng truy cập cần gần 400 giây.

Điều này là do số lượng các đại lượng trọng số (*M*) đã được thiết lập ở giá trị 20, vì thế trong 20 slots đầu tiên không có dự báo thông lượng truy cập.



Hình 3. Các dự đoán thông lượng

Kể từ khe 21, quá trình dự báo thông lượng truy cập xảy ra khá chậm trong suốt thời gian khe 21 đạt giá trị chấp nhận được. Tuy nhiên, rất khó để kiểm tra hiệu quả của các dự báo bằng thị giác. Để đảm bảo tính tin cậy của các kết quả dự báo, chúng ta cần tính toán các giá trị sai khác trung bình, các giá trị này thể hiện tính đúng đắn qua quá trình thực hiện thuật toán kiểm soát truy nhập và cân bằng tải.

4. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

4.1. Kết luận

Qua phần trình bày trên, nội dung bài báo đã đề cập đến các vấn đề cơ bản sau:

- Giới thiệu các giải pháp cân bằng tải trang web, tách các kiến trúc mạng và chính sách phân phối. Đồng thời, giới thiệu kỹ thuật phân tải dựa trên các giao thức OSI, bao gồm các kỹ thuật nhận biết nội dung. Qua đó, chứng minh kỹ thuật cân bằng tải nhận biết nội dung hữu dụng hơn kỹ thuật cân bằng tải không nhận biết nội dung. Tuy nhiên, hai kỹ thuật này không phải lúc nào cũng tách biệt nhau, trong một số trường hợp, kỹ thuật cân bằng tải nhận biết nội dung chuyển nhiệm vụ phân phối cho kỹ thuật cân bằng tải không nhận biết nội dung để đạt được sự cân bằng tải tối ưu nhất.

- Các đề xuất kiểm soát truy nhập cũng được nghiên cứu. Chúng được phân loại theo cách gọi định kỳ và không định kỳ. Cách gọi phi định kỳ được thực hiện khi có một yêu cầu mới truy nhập hoặc phiên kết nối tới cho hệ thống. Tỷ lệ đến được xem là tham số quan trọng, có thể làm thay đổi tần số gọi kiểm soát truy nhập để tránh tình trạng tắc nghẽn có thể có trong hệ thống web.

- Xác định và nghiên cứu sáu yếu tố burstiness khác nhau. Dựa trên các yếu tố burstiness để lập khe thời gian thích nghi. Nó được dùng để đặt tần số giám sát, được sử dụng trong thuật toán cân bằng tải và kiểm soát truy nhập và cũng để đặt thời hạn khe thích hợp dựa trên tính bursty của các yêu cầu tới hệ thống.

Trong số các yếu tố burstiness đã trình bày, kết luận BF1, BF5 và BF6 rất cứng nhắc và không thay đổi đáng kể giá trị của chúng với những thay đổi của tỉ lệ yêu cầu tới. Tuy nhiên, các yếu tố bao gồm sự ngăn cản làm cho BF3 và BF4 đạt giá trị tối đa một cách dễ dàng khi phát hiện các khe bursty liên tiếp, BF2 - yếu tố burstiness được áp dụng trong thuật toán, vì cho phép nó nhận biết của mức độ tăng tỉ lệ yêu cầu đến độc lập với khối lượng tải thực tế trong hệ thống.

- Xây dựng thuật toán cân bằng tải và điều khiển truy nhập, dựa trên lập khe thời gian thích nghi. Khe thời gian đó được sử dụng để gọi thực thi thuật toán. Do đó, thuật toán cân bằng tải đưa các yếu tố burstiness cho lần gọi thực thi của nó. Với năm yếu tố tiên đoán thông qua được và kết luận những dự báo phù

hợp tốt hơn thuật toán cân bằng tải được dựa trên LMS vì nó là một trong những cách có thời gian phản ứng ít nhất và tác động ổn định nhất. Điều này có nghĩa là các dịch vụ vẫn được đảm bảo mặc dù tỷ lệ đạt đến ngưỡng của hệ thống web.

- Khi so sánh chiến lược lập khe thời gian thích nghi và khe thời gian cố định, qua phân tích và đi đến kết luận: việc lập khe thời gian thích nghi đạt hiệu quả tốt hơn khi đánh giá mức độ sử dụng CPU và thời gian phản hồi. Việc cố định khe thời gian không đảm bảo việc lưu trữ thông tin mức độ sử dụng CPU một cách chính xác.

4.2. Hướng phát triển

Vấn đề thiết kế các yếu tố burstiness phải tiếp tục nghiên cứu để tìm kiếm hoặc cải thiện yếu tố burstiness mang lại lợi ích tốt nhất. Đặc biệt trong trường hợp phát hiện bursty trong các khe liên tiếp. Yếu tố burstiness có thể phù hợp cho tất cả các tỉ lệ yêu cầu đến, một hệ thống web có thể mong đợi và đã đáp ứng các tỉ lệ yêu cầu đến khác nhau trong phạm vi của các giá trị được định nghĩa cho nó.

Việc phát triển các yếu tố burstiness nhận biết tỷ lệ đến trong số các dịch vụ khác nhau theo các yêu cầu đòi hỏi truy nhập gửi đến. Do đó, một vấn đề nữa sẽ là sự khác biệt về lưu lượng ở các yếu tố burstiness. Vấn đề đó có thể cải thiện hiệu suất của hệ thống web khi tất cả các loại dịch vụ có thể không được yêu cầu với cường độ tương tự trong các hệ thống web.

Khi nghiên cứu thuật toán kiểm soát truy nhập, việc dự phòng tài nguyên trong các máy chủ web, được thực hiện mà không quan tâm tới số lượng dịch vụ được yêu cầu.

TÀI LIỆU THAM KHẢO

[1] K.Gilly, C.Juiz, R.Puigjaner, "*An up-to-date survey in web load balancing*", World Wide Web (2011), Vol.14(2), pp.105-131.

[2] M.Aron, P.Druschel, W.Zwaenepoel, "*Ecient support for P-HTTP incluster-based web servers*". Proc. The Annual Conference on USENIX Annual Technical Conference (1999).

[3] Zheng Wang and Jon Crowcroft. "*Analysis of burstiness and jitter in real-time communications*". In proc.of SIGCOMM (1993).

[4] Yuliy Baryshnikov, Ed Coffman, Dan Rubenstein, and Teddy Yimwadsana. "*Traffic prediction on the internet*". Technical report ee200514-1, ComputerNetworking Research Center Columbia Univeristy, (May 2002).

[5] Simon Haykin. "*Adaptive Filter Theory*". Prentice-Hall (1991).

[6] Graham C. Goodwin and Kwai Sang Sin. "*Adaptive Filtering:Predictionand Control. Prentice-Hall*" (1984).

[7] Katja Gilly, Salvador Alcaraz, Carlos Juiz, and Ramon Puigjaner. "*Comparison of predictive techniques in cluster-based network servers with resource allocation*".In proc. Of the12th Annual Meeting of the IEEE/ACM International Symposium on Modeling, Analysis, and Simulation of Computerand.

Thông tin liên hệ: **Vũ Văn Đốc**

Điện thoại: 0912648561 - Email: vvdoc@uneti.edu.vn

Khoa Công nghệ thông tin, Trường Đại học Kinh tế - Kỹ thuật Công nghiệp