

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



Huỳnh Phi Long

**ĐỀ XUẤT THUẬT TOÁN CÂN BẰNG TẢI TRÊN
ĐIỆN TOÁN Đám MÂY THÔNG QUA HÀNH VI
NGƯỜI DÙNG CLOUD**

Chuyên ngành: Hệ thống thông tin

Mã số: 8.48.01.04

TÓM TẮT LUẬN VĂN THẠC SĨ

THÀNH PHỐ HỒ CHÍ MINH – NĂM 2022

Luận văn được hoàn thành tại:
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: PGS.TS. Trần Công Hùng
(Ghi rõ học hàm, học vị)

Phản biện 1:

Phản biện 2:

Luận văn sẽ được bảo vệ trước Hội đồng chấm luận văn thạc sĩ tại Học viện Công nghệ
Bưu chính Viễn thông

Vào lúc: giờ ngày tháng năm

Có thể tìm hiểu luận văn tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông.

MỞ ĐẦU

1. Tính cấp thiết của đề tài

Với sự phát triển của khoa học công nghệ, đặc biệt về lĩnh vực công nghệ thông tin – viễn thông, thông qua việc ứng dụng sự phát triển của công nghệ máy tính và dựa vào mạng Internet cho phép người dùng có thể lưu trữ, trích xuất dữ liệu từ nhiều nơi khác nhau và có thể truy cập qua mạng các dịch vụ công nghệ từ một nhà cung cấp nào đó, kết hợp sử dụng các công nghệ điện toán (song song, phân tán, ảo hóa...) gọi là “Điện toán đám mây” (ĐTĐM – Cloud Computing).

Trong những năm gần đây, sự tăng trưởng nhanh chóng của số lượng thiết bị đầu cuối di động làm phát sinh xu hướng không thể đảo ngược của việc ứng dụng rộng rãi điện toán đám mây. Sự chuyển dịch của điện toán đám mây từ thị trường máy tính để bàn sang thị trường di động trở thành hướng phát triển chính. Tuy nhiên, dường như có nhiều vấn đề phức tạp. Trong số đó, ba khía cạnh của “người dùng - môi trường - dịch vụ” là đặc biệt nổi bật. Việc phân loại rõ ràng hành vi của người dùng khi sử dụng các dịch vụ đám mây đã trở thành một vấn đề quan trọng cần được giải quyết khẩn cấp.

Hiện nay, việc thực hiện cân bằng tải trong điện toán đám mây thông qua hành vi người dùng Cloud là một thách thức lớn đối với các nhà nghiên cứu và nhà cung cấp dịch vụ đám mây dựa trên công nghiệp trên thế giới. Việc thiết lập một thuật toán cân bằng tải hiệu quả đáp ứng được hiệu năng hệ thống và làm thế nào sử dụng nguồn tài nguyên điện toán đám mây một cách có hiệu quả nhất là mục đích cuối cùng của điện toán đám mây muốn đạt đến. Ở nước ta hiện nay, các công trình nghiên cứu về cân bằng tải trong điện toán đám mây thông qua hành vi người dùng Cloud Computing cũng còn ít, hạn chế. Vì vậy, luận văn “Đề xuất thuật toán cân bằng tải trên điện toán đám mây thông qua hành vi người dùng Cloud” sẽ đi sâu nghiên cứu các kỹ thuật cân bằng tải đang được áp dụng hiện nay; đồng thời đề xuất cải tiến một kỹ thuật cân bằng tải và mô phỏng điện toán đám mây CloudSim. Đề cương luận văn bao gồm 03 phần:

1. Phần mở đầu
2. Phần nội dung gồm:

- Chương 1: Giới thiệu tổng quan về hệ thống cân bằng tải của điện toán đám mây.
- Chương 2: Các Công trình nghiên cứu liên quan.
- Chương 3: Đề xuất Thuật toán Cân bằng tải trên điện toán đám mây thông qua hành vi người dùng cloud.
- Chương 4: Mô phỏng, Thực nghiệm.

3. Phân Kết luận

2. Tổng quan về vấn đề nghiên cứu

Cân bằng tải là kỹ thuật phân phối khối lượng công việc đồng đều giữa hai hoặc nhiều máy tính, kết nối mạng, CPU, ổ cứng, hoặc các nguồn lực phân tán to lớn trên mạng, để có thể tận dụng có hiệu quả các nguồn lực, tối đa hóa thông lượng, cải thiện thời gian đáp ứng và thời gian xử lý dữ liệu; Đồng thời tránh tình trạng quá tải một số nút tính toán trong khi những nút khác được nạp tải nhẹ khi có nhiều yêu cầu xử lý cần được đáp ứng. Kỹ thuật cân bằng tải hiện nay chủ yếu tập trung vào hai kỹ thuật là cân bằng tải tĩnh và cân bằng tải động.

Kỹ thuật cân bằng tải tĩnh không thu thập thông tin trạng thái hiện tại hệ thống. Những yếu tố được đo lường trước khi gán công việc cho một nút tính toán như thời gian đến, qui mô nguồn tài nguyên, thời gian thực thi và giao tiếp các tiến trình.

Kỹ thuật cân bằng tải động trong tự nhiên không xem xét trạng thái trước đó hoặc hành vi của hệ thống, nó chỉ phụ thuộc vào hành vi hiện tại của hệ thống.

3. Mục đích nghiên cứu

Mục tiêu chính: “Đề xuất thuật toán cân bằng tải trên điện toán đám mây thông qua hành vi người dùng Cloud”.

Từ mục tiêu chính trên, Luận văn sẽ dự kiến các kết quả đạt được như sau:

- Tìm hiểu tổng quan về điện toán đám mây.
- Tìm hiểu về các thuật toán trên điện toán đám mây.
- Tìm hiểu về các thuật toán cân bằng tải trên điện toán đám mây.
- Tìm hiểu khả năng xảy ra quá tải, tài nguyên phân bổ không đồng đều, máy chủ quá tải và ngưng hoạt động.

- Nghiên cứu về hành vi người dùng Cloud.
- Đề xuất thuật toán có thể sử dụng tài nguyên hiệu quả hơn, tiết kiệm năng lượng.
- Trên cơ sở lý thuyết đã nghiên cứu, Luận văn đề xuất thuật toán nâng cao hiệu quả cân bằng tải của điện toán đám mây. Đánh giá hiệu quả của đề xuất cải tiến này trong môi trường mô hình và mô phỏng điện toán đám mây CloudSim; đồng thời nghiên cứu hướng tiếp cận mới về điện toán đám mây thông qua môi trường CloudSim...

4. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu

- Đối tượng nghiên cứu chính là thuật toán nâng cao hiệu quả cân bằng tải trên điện toán đám mây thông qua hành vi của người dùng Cloud (người dùng cá nhân – User, doanh nghiệp – Enterprise, tổ chức – Organization)
- Nghiên cứu các thuật toán cân bằng tải hiện đang sử dụng.

Phạm vi nghiên cứu

Phạm vi nghiên cứu trong Cloud:

- Xây dựng mô hình mô phỏng đám mây ở mức độ nhỏ: khoảng 10 – 30 máy ảo.
- Độ phức tạp trên mỗi máy ảo chỉ ở mức độ thấp: dưới 10 ứng dụng chạy trên trên các máy ảo.
- Yêu cầu (Request) gửi về máy chủ cũng đơn giản, đánh giá chính xác phải là hành động của người dùng Cloud.

5. Phương pháp nghiên cứu

Phương pháp luận

Dựa trên cơ sở là các lý thuyết về điện toán đám mây, các thuật toán cân bằng tải trên cloud.

Phương pháp đánh giá dựa trên cơ sở toán học

Trên cơ sở các lý thuyết về điện toán đám mây, khả năng xảy bị tắc nghẽn trên đám mây. Đề xuất ra thuật toán để nâng cao hiệu quả cân bằng tải trên đám

mây dựa trên các thuật toán đã nghiên cứu. Chứng minh thuật toán và đánh giá hiệu quả của thuật toán.

Phương pháp đánh giá bằng mô phỏng thực nghiệm

Xây dựng mô hình mô phỏng và thực nghiệm thuật toán đã đề xuất.

CHƯƠNG 1. GIỚI THIỆU TỔNG QUAN VỀ HỆ THỐNG CÂN BẰNG TẢI CỦA ĐIỆN TOÁN Đám MÂY

1.1. Tổng quan về điện toán đám mây

Lịch sử của điện toán đám mây bắt đầu từ năm 1983, khi Sun Microsystems đề xuất rằng “Web là máy tính”. Trong tháng 3 năm 2006, Amazon giới thiệu dịch vụ đám mây điện toán đàn hồi. Vào tháng 8 năm 2006, Eric Schmidt, Giám đốc điều hành của Google, lần đầu tiên đề xuất khái niệm “Điện toán đám mây” tại hội nghị công cụ tìm kiếm. Năm 2009, Nair M K. và Gopalakrishnan V. đã phát triển một khung hệ thống, sử dụng các dịch vụ web như SaaS và môi trường web để hiện thực hóa PaaS, thúc đẩy hiệu quả sự phát triển của điện toán đám mây. Takahiro Miyamoto và nhóm của ông đã nhận ra chức năng mạng của điện toán đám mây vào năm 2009, đặt nền tảng vững chắc cho sự phát triển của điện toán đám mây. Kể từ đó, điện toán đám mây đã bước vào thời kỳ phát triển nhanh chóng. Điện toán đám mây được phát triển từ điện toán song song, điện toán phân tán và điện toán lưới, như trong nó là một mô hình điện toán kinh doanh mới. Hiện tại, vẫn chưa có định nghĩa thống nhất về điện toán đám mây. Theo Wikipedia, định nghĩa điện toán đám mây là một phương thức tính toán mới dựa trên Internet, cung cấp tính toán theo yêu cầu cho người dùng cá nhân và doanh nghiệp thông qua các dịch vụ không đồng nhất và tự trị trên Internet. Eric Schmidt, Giám đốc điều hành của Google, cho rằng điện toán đám mây về cơ bản là một mô hình cung cấp dịch vụ, ảo hóa tài nguyên máy tính, tài nguyên lưu trữ và tài nguyên mạng bao gồm một số lượng lớn máy chủ, tạo thành một nhóm tài nguyên ảo bao gồm tài nguyên điện toán, lưu trữ và mạng. Quản lý và lên lịch thông qua một nền tảng điện toán đám mây thống nhất.

Điện toán đám mây là một ý tưởng đang phát triển trong thế giới CNTT, được sinh ra từ nhu cầu sử dụng máy tính khi đang di chuyển. Nó mang lại cho người dùng quyền truy cập vào dữ liệu, ứng dụng và bộ nhớ không được lưu trữ trên máy tính của họ. Để có một cái nhìn tổng quan về điện toán đám mây rất đơn giản, nó có thể được hiểu là một hệ thống phân phối cung cấp điện toán giống như cách một lưới điện cung cấp điện. Đối với người dùng máy tính bình thường, nó

mang lại lợi thế là cung cấp CNTT mà người dùng không cần phải có kiến thức chuyên sâu về công nghệ. Tương tự như cách một người tiêu dùng có thể tiếp cận điện năng mà không cần phải là một thợ điện. Cụ thể hơn, trong mô hình điện toán đám mây, tất cả các tài nguyên, thông tin và phần mềm đều được chia sẻ và cung cấp cho các máy tính, thiết bị, người dùng dưới dạng dịch vụ trên nền tảng một hạ tầng mạng công cộng (thường là mạng Internet). Các người dùng sử dụng dịch vụ như cơ sở dữ liệu, website, lưu trữ, ... trong mô hình điện toán đám mây không cần quan tâm đến vị trí địa lý cũng như các thông tin khác của hệ thống mạng đám. Người dùng cuối truy cập và sử dụng các ứng dụng đám mây thông qua các ứng dụng như trình duyệt web, các ứng dụng mobile (di động) hoặc máy tính cá nhân thông thường. Hiệu năng sử dụng phía người dùng cuối được cải thiện khi các phần mềm chuyên dụng, các cơ sở dữ liệu được lưu trữ và cài đặt trên hệ thống máy chủ ảo trong môi trường điện toán đám mây trên nền của trung tâm dữ liệu – hay còn gọi là “Data Center”. Đây là một thuật ngữ chỉ khu vực chứa server và các thiết bị lưu trữ, bao gồm nguồn điện và các thiết bị có khả năng sẵn sàng và độ ổn định cao. Bên cạnh đó còn bao gồm các tiêu chí khác như: tính module hóa cao, khả năng mở rộng dễ dàng, nguồn và làm mát, hỗ trợ hợp nhất server và lưu trữ mật độ cao.

Có 3 mô hình triển khai điện toán đám mây bao gồm: public (công cộng), private (riêng tư), và hybrid (“lai” giữa đám mây công cộng và riêng tư). Trong một đám mây công cộng, nhà cung cấp bên thứ ba cung cấp một loạt các dịch vụ cho công chúng qua internet. Dữ liệu từ một số khách hàng công ty hoặc cá nhân có thể dùng chung một máy chủ. Về nguyên tắc, đám mây riêng cũng tương tự như vậy, nhưng được thiết lập sau tường lửa và chỉ cung cấp các dịch vụ được lưu trữ cho một số lượng hạn chế người dùng được phê duyệt. Đám mây “lai” là môi trường đám mây mà kết hợp cung cấp các dịch vụ công cộng và riêng tư. Ngoài ra còn có “Community Cloud” là đám mây giữa các nhà cung cấp dịch vụ đám mây.

Điện toán đám mây bao gồm ba loại dịch vụ điện toán riêng biệt được phân phối từ xa tới khách hàng thông qua internet [1]. Khách hàng thường trả phí dịch vụ hàng tháng hoặc hàng năm cho nhà cung cấp để có quyền truy cập vào các hệ thống cung cấp phần mềm dưới dạng dịch vụ (SaaS), nền tảng dưới dạng dịch vụ (PaaS) và cơ sở hạ tầng như dịch vụ (IaaS) cho người đăng ký. Khách hàng đăng ký dịch vụ điện toán đám mây có thể gặt hái nhiều lợi ích, tùy thuộc vào nhu cầu kinh

doanh cụ thể của họ tại một thời điểm nhất định. Những ngày đầu tư vốn lớn vào phần mềm và cơ sở hạ tầng CNTT giờ đã trở thành dĩ vãng đối với bất kỳ doanh nghiệp nào chọn áp dụng mô hình điện toán đám mây để mua sắm các dịch vụ CNTT. Khả năng tiếp cận các nguồn lực CNTT mạnh mẽ trên cơ sở gia tăng đang san bằng sân chơi cho các tổ chức vừa và nhỏ, cung cấp cho họ các công cụ và công nghệ cần thiết để cạnh tranh trên thị trường toàn cầu mà không cần đầu tư trước đây vào các nguồn lực CNTT tiền đề. Những khách hàng đăng ký dịch vụ điện toán được cung cấp qua “cloud” có thể giảm đáng kể chi phí dịch vụ CNTT cho tổ chức của họ; và có quyền truy cập vào các dịch vụ tính toán cấp doanh nghiệp linh hoạt và nhanh nhẹn hơn, trong quá trình này.

Theo các loại hình dịch vụ, điện toán đám mây có thể được chia thành ba loại tương ứng như sau:

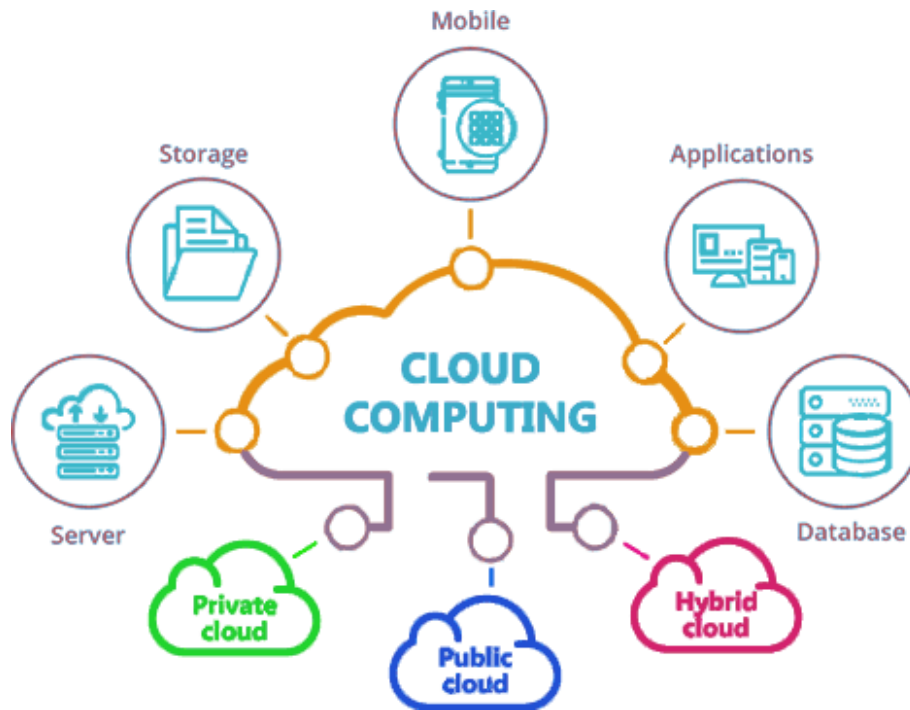
SaaS (Software as a Service) cung cấp cho khách hàng khả năng sử dụng các ứng dụng phần mềm từ xa thông qua trình duyệt web internet. Phần mềm như một dịch vụ còn được gọi là “phần mềm theo yêu cầu”. Khách hàng có thể truy cập các ứng dụng SaaS từ mọi nơi thông qua web vì các nhà cung cấp dịch vụ lưu trữ các ứng dụng và dữ liệu liên quan của họ tại vị trí của họ. Lợi ích chính của SaaS là chi phí sử dụng thấp hơn, vì phí thuê bao đòi hỏi một khoản đầu tư nhỏ hơn nhiều so với những gì thường gặp trong mô hình phân phối phần mềm truyền thống. Hầu như có thể loại bỏ phí cấp phép, chi phí cài đặt, phí bảo trì và phí hỗ trợ liên quan đến mô hình cung cấp phần mềm truyền thống bằng cách đăng ký mô hình phân phối phần mềm SaaS. Ví dụ về SaaS bao gồm: Ứng dụng Google và các ứng dụng email dựa trên internet như Yahoo! Mail, Hotmail và Gmail.

PaaS (Platform as a Service) cung cấp cho khách hàng khả năng phát triển và xuất bản các ứng dụng tùy chỉnh trong môi trường được lưu trữ thông qua web. Nó đại diện cho một mô hình mới để phát triển phần mềm đang tăng nhanh chóng về mức độ phổ biến của nó. Một ví dụ về PaaS là Salesforce.com. PaaS cung cấp một khuôn khổ để phát triển, thử nghiệm, triển khai và bảo trì phần mềm nhanh trong một môi trường tích hợp. Giống như SaaS, lợi ích chính của PaaS là chi phí sử dụng thấp hơn, vì phí thuê bao yêu cầu đầu tư nhỏ hơn nhiều so với những gì thường gặp phải khi triển khai các công cụ truyền thống để phát triển, thử nghiệm và triển khai phần mềm. Các nhà cung cấp PaaS xử lý việc bảo trì nền tảng và nâng cấp hệ thống,

dẫn đến giải pháp hiệu quả hơn và tiết kiệm chi phí cho việc phát triển phần mềm doanh nghiệp.

IaaS (Infrastructure as a Service) hoặc cơ sở hạ tầng như một dịch vụ, cho phép người dùng truy cập trực tiếp vào tài nguyên lưu trữ, tài nguyên mạng và tài nguyên máy tính bên dưới. IaaS sử dụng công nghệ ảo hóa để ảo hóa và đóng gói tài nguyên máy tính, tài nguyên lưu trữ và tài nguyên mạng của máy chủ, đồng thời cung cấp các tài nguyên này dưới dạng API. Khi cần sử dụng các tài nguyên này, người dùng không cần mua các thiết bị phần cứng như máy chủ mà chỉ cần mua các tài nguyên này từ các nhà sản xuất cung cấp dịch vụ IaaS. Nền tảng điện toán đám mây IaaS cung cấp quản lý và lập kế hoạch của các tài nguyên này. Ví dụ điển hình bao gồm: Đám mây tính toán đàn hồi (EC2) và Dịch vụ lưu trữ đơn giản (S3) của Amazon.

Bằng cách cung cấp điện toán, lưu trữ và các ứng dụng này dưới dạng dịch vụ mà không phải là sản phẩm, đám mây mang lại lợi thế kinh doanh và chi phí. Đám mây di chuyển tất cả các dịch vụ này ra bên ngoài trang web đến một nhà thầu hoặc một cơ sở tập trung. Tập trung dữ liệu cho phép chia sẻ chi phí giữa tất cả người dùng. Đám mây hoàn thành những gì CNTT luôn tìm kiếm; tăng khả năng tính toán mà không cần phải cung cấp cơ sở hạ tầng mới. Các ứng dụng có thể có của điện toán đám mây là theo cấp số nhân. Người dùng giao diện với đám mây thông qua trình duyệt web của họ, loại bỏ nhu cầu cài đặt nhiều ứng dụng phần mềm.



Hình 0.1: Mô hình điện toán đám mây [2]

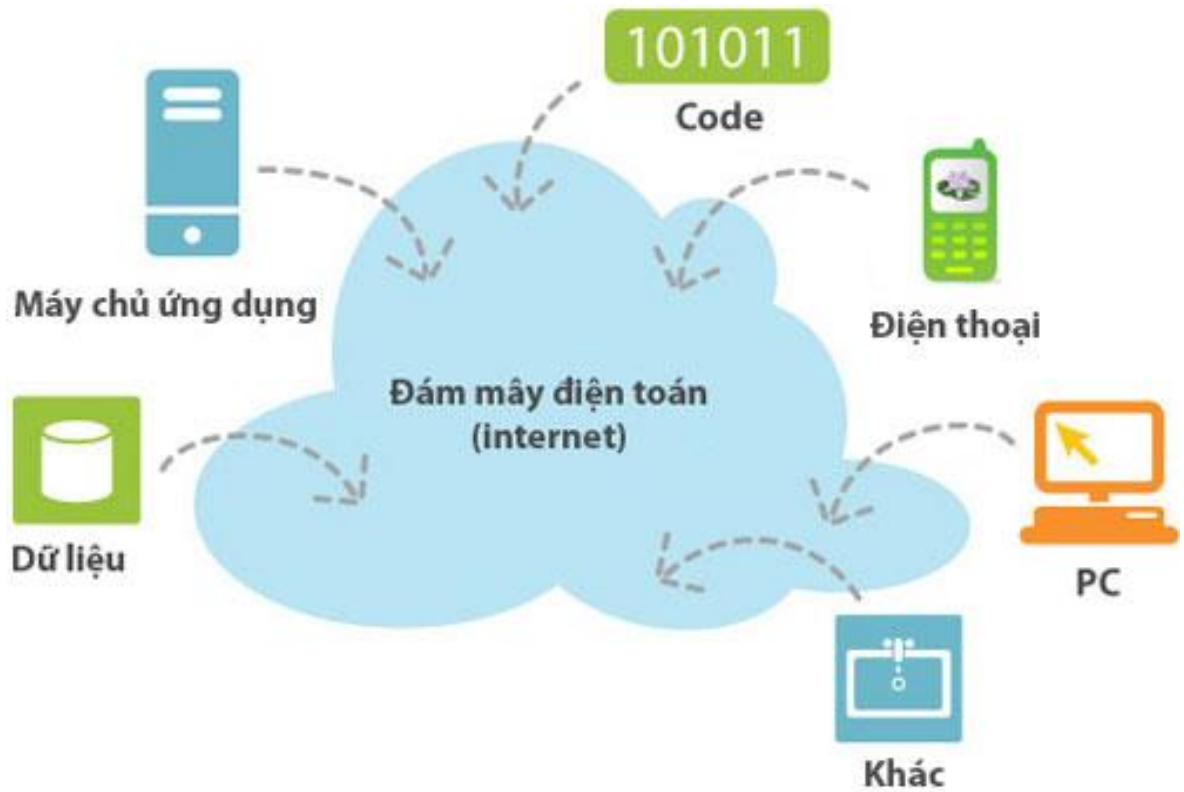
Điện toán đám mây là một mô hình điện toán nơi mà mọi giải pháp liên quan đến CNTT đều được cung cấp dưới dạng các dịch vụ qua mạng Internet, giúp người dùng thoát khỏi việc phải đầu tư không ít nhân lực, công nghệ và hạ tầng để triển khai hệ thống. Qua đó tối giản chi phí cũng như thời gian triển khai, tạo điều kiện cho người sử dụng nền tảng điện toán đám mây tập trung được tối đa nguồn lực vào công việc chuyên môn của họ. Chính vì vậy, lợi ích mà điện toán đám mây mang lại không chỉ gói gọn trong phạm vi người sử dụng nền tảng điện toán đám mây mà còn từ phía các nhà cung cấp dịch vụ điện toán. Bên cạnh đó, điện toán đám mây [2], [3] cũng đang là xu hướng được phát triển mạnh mẽ hiện nay. Kế thừa các mạng lưới trước đây cùng với các khái niệm máy tính phân tán để tích hợp các tài nguyên máy tính, lưu trữ, nền tảng và các dịch vụ khác theo nhu cầu một cách thuận tiện và nhanh chóng, đồng thời cho phép kết thúc sử dụng dịch vụ, giải phóng tài nguyên dễ dàng, giảm thiểu các giao tiếp với nhà cung cấp. Theo đó, mô hình chính là cho phép sử dụng dịch vụ theo yêu cầu; cung cấp khả năng truy cập dịch vụ qua mạng rộng rãi từ máy tính để bàn, máy tính xách tay tới thiết bị di động; với tài nguyên tính toán động, phục vụ nhiều người năng lực tính toán phần mềm dẻo, đáp ứng nhanh với nhu cầu từ thấp đến cao.

Điện toán đám mây được dựa trên công nghệ ảo hóa [4], thông qua các dịch vụ mạng để cung cấp cho người dùng với các nguồn lực cơ bản, nền tảng ứng dụng, phần mềm và các dịch vụ khác. Trong trường hợp IaaS (cơ sở hạ tầng như một dịch vụ), các nhà phát triển cung cấp một môi trường ứng dụng phần mềm hoàn chỉnh bằng cách tập hợp các phần cứng, phần mềm và các thiết bị có liên quan lại với nhau để đáp ứng thỏa thuận chất lượng dịch vụ với người dùng. Công nghệ máy ảo (Virtual Machine) thường được sử dụng trong các trung tâm dữ liệu, máy tính cụm và các ứng dụng khác. Công nghệ này cho phép nhiều hệ điều hành có thể chạy trên cùng một máy tính và cung cấp các dịch vụ độc lập đáng tin cậy, cải tiến rất nhiều khả năng sử dụng lại các tài nguyên vật lý. Ngoài ra, điện toán đám mây [5] còn là một hướng nghiên cứu rộng, sẽ đem lại giá trị lớn về các chi phí cho các doanh nghiệp trên toàn thế giới. Điện toán đám mây sẽ giúp giải quyết được việc lưu trữ dữ liệu trên hệ thống nhanh, gọn, nhẹ. Cung cấp các dịch vụ về cơ sở hạ tầng, nền tảng phần mềm, các dịch vụ theo yêu cầu người dùng thông qua Internet.

Điện toán đám mây [6] là một mô hình dịch vụ công nghệ thông tin, kế thừa các mạng lưới trước đây trên thế giới giúp người dùng truy cập tài nguyên dữ liệu, lưu trữ đến hệ thống quản lý, xử lý dữ liệu phức tạp của các hệ thống như Google, Facebook... Người dùng chỉ truy cập vào thiết bị đầu cuối để truy xuất vào các tài nguyên trên điện toán và bên trong hệ thống điện toán sẽ lập lịch xử lý các yêu cầu trên bao gồm xử lý thời gian chờ, thời gian xử lý tín hiệu đến thời gian hoàn thành nhiệm vụ. Chính vì vậy mà điện toán đám mây [7] đang chuyển đổi ngành CNTT, thay đổi cách thức sử dụng và cung cấp phần mềm và phần cứng. Làm cho việc sử dụng các tài nguyên máy tính theo yêu cầu như băng thông, lưu trữ hoặc các ứng dụng phần mềm và điện toán có sẵn. Nó che giấu sự phức tạp của cơ sở hạ tầng cơ bản, cho phép người dùng cuối tập trung vào sản phẩm của chính họ mà không cần nhiều khoản đầu tư vào phần cứng. Theo hợp đồng dịch vụ đã được thiết lập giữa nhà cung cấp điện toán và khách hàng, các ràng buộc về chất lượng dịch vụ (QoS) nhất định được xác định thông qua các thỏa thuận theo mức dịch vụ (SLA). Tuân thủ với các SLA này, nhà cung cấp đảm bảo cung cấp một chất lượng nhất định cho dịch vụ đã thỏa thuận. Việc sử dụng các máy ảo cho phép sử dụng tốt hơn các tài nguyên phần cứng hiện tại trong khi vẫn duy trì QoS yêu cầu. Để tránh sự xuống cấp của hiệu suất, máy ảo được di chuyển từ quá tải đến các máy không sử dụng

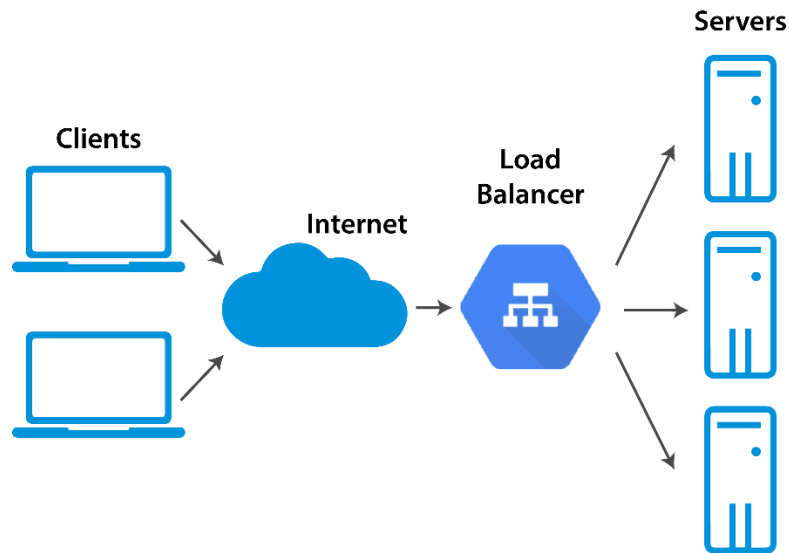
được. Vì vậy, các thuật toán phát hiện là cần thiết để chủ động phân loại quá tải và không quá tải. Các thuật toán chủ động xác định một kế hoạch tối ưu cho việc di chuyển và phân bổ các máy ảo trong thời gian chạy.

Điện toán đám mây là một kiểu mẫu mới [8], [9] và đang không ngừng tiến hóa trong tính toán. Cơ chế cân bằng tải được chia thành các nguồn lực và cung cấp các nguồn lực cùng với nhiệm vụ lập kế hoạch giữa các hệ thống phân phối. Trong cân bằng tải truyền thống phải đối mặt với một số vấn đề khác nhau của các giai đoạn cung cấp tài nguyên trong môi trường đám mây. Nó cũng có tác động to lớn trong các hệ thống đám mây về hiệu suất và về vấn đề đo lường do sự tham gia của các thông số cân bằng tải khác nhau và bản chất của môi trường đám mây. Ngày nay [10], điện toán đám mây là một cách để giữ phần cứng cũng như phần mềm ở một nơi và sử dụng nó từ mọi nơi trong thế giới này. Nó đã làm cho phần cứng yêu cầu linh hoạt hơn nhiều. Do đó, mọi người có cơ hội sử dụng nhiều tài nguyên khi cần và phải trả số tiền chỉ cho khoảng thời gian họ đã sử dụng nguồn dung lượng cụ thể, được gọi là dịch vụ trả tiền cho mỗi lần sử dụng, làm cho ngành công nghiệp công nghệ thông tin hướng tới việc kinh doanh điện toán đám mây. Giống như một CPU với nhiều lõi, những doanh nghiệp sở hữu một cụm của các CPU/Máy vật lý đó được gọi là đám mây. Các cụm có một số lượng hữu hạn không gian và bộ nhớ.



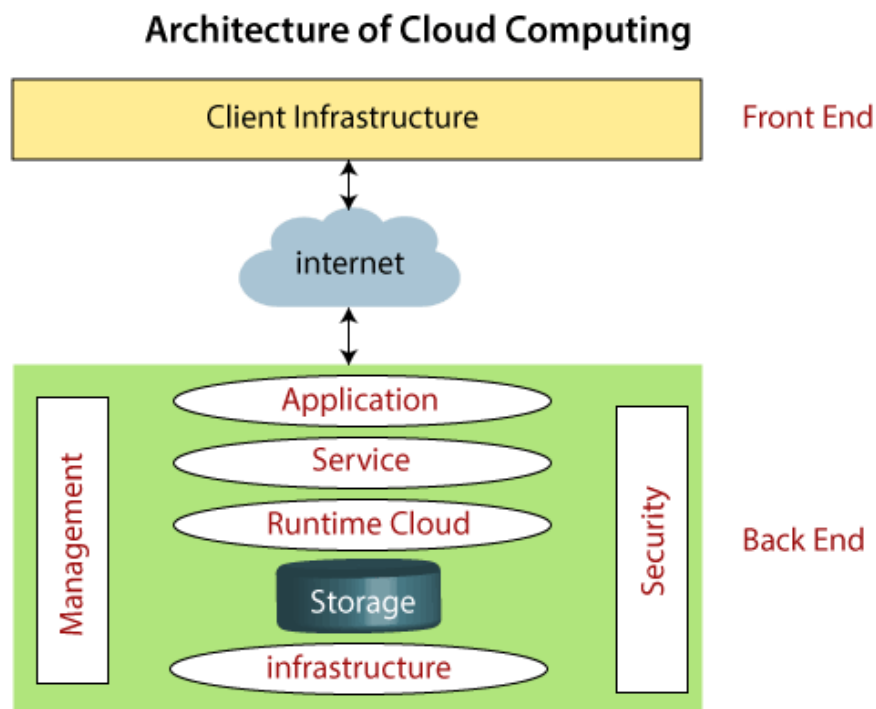
Hình 0.2: Cung cấp tài nguyên đám mây [5]

Người dùng sẽ nhận được không gian và bộ nhớ trong một khoảng thời gian từ cụm được phân bổ khi trả tiền cho dịch vụ này. Khi người dùng đòi hỏi các nguồn lực bao gồm bộ nhớ, không gian và băng thông được thực hiện bởi các công ty thông qua phân bổ các máy chủ đến nền tảng nhu cầu khách hàng. Cung cấp tài nguyên trên đám mây là quá trình cung cấp không gian bộ nhớ ảo từ các nguồn lực bằng cách tổng hợp máy vật lý (PM) hay còn được gọi là máy ảo (VM). Lúc này, bộ cân bằng tải sẽ quản lý ghép kênh các tài nguyên theo đúng với yêu cầu từ người dùng.



Hình 0.3: Cân bằng tải trong điện toán đám mây [6]

Những biện pháp cân bằng trước đây có hiệu quả trong việc cải thiện thời gian phản hồi và thời gian phục vụ của đám mây tuy nhiên nó lại không cung cấp đúng chất lượng dịch vụ. Các QoS có thể cung cấp hiệu quả bằng cách thêm tham số của nó vào tham số cân bằng tải. Xem xét băng thông như tham số, mà phải đối mặt với các vấn đề suy giảm và những vấn đề khác sẽ làm cho ngưỡng giá trị chính xác hơn, do đó QoS sẽ được coi là có hiệu quả. Vì vậy, giảm thiểu yêu cầu được cấp phát cho các máy vật lý với đúng khả năng cung cấp của các máy ảo và duy trì trạng thái ổn định trong suốt thời gian cung cấp dịch vụ.



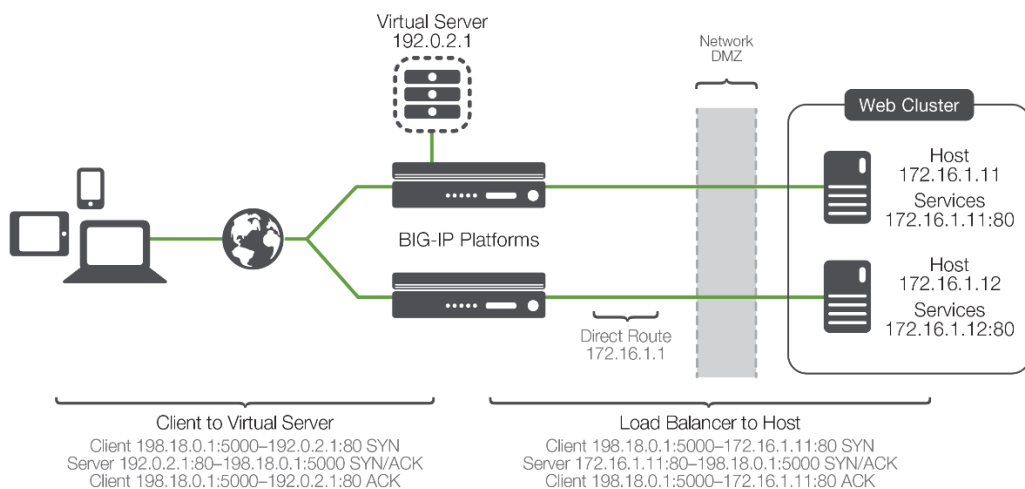
Hình 0.4: Kiến trúc của điện toán đám mây [8]

1.2. Tổng quan về cân bằng tải trong điện toán đám mây

1.2.1 Giới thiệu về cân bằng tải

Hiện nay, lĩnh vực CNTT đang phát triển ngày càng mạnh mẽ và nhu cầu về tài nguyên lưu trữ và tính toán tăng trưởng không kém. Với lượng lớn dữ liệu được tạo và trao đổi qua mạng ngày một tăng lên thì sự đòi hỏi thêm về nhu cầu tài nguyên máy tính ngày càng nhiều. Đám mây đã giúp các doanh nghiệp tận dụng lợi ích của tài nguyên điện toán được chia sẻ trên môi trường ảo hóa, cùng với đó rất nhiều doanh nghiệp đã sử dụng các dịch vụ dựa trên đám mây ở dạng này hoặc dạng khác. Điều này đưa chúng ta đến khái niệm cân bằng tải trong điện toán đám mây.

Kéo theo sự phát triển rộng khắp của Internet, các website hay các ứng dụng trực tuyến đang được rất nhiều người truy cập và sử dụng. Lượng truy cập này quá lớn thường dẫn đến các vấn đề về hạ tầng mạng, ngoài ra khả năng xử lý của Server sẽ bị tắc nghẽn cục bộ. Chính vì vậy, cân bằng tải luôn luôn là một trong những tính năng công nghệ rất quan trọng giúp các máy chủ ảo hoạt động đồng bộ và hiệu quả hơn thông qua việc phân phối đồng đều các tài nguyên. Giải pháp của cân bằng tải chính là việc phân bố đồng đều lưu lượng truy cập giữa hai hay nhiều các máy chủ có cùng chức năng trong cùng một hệ thống. Với phương pháp này, hệ thống sẽ giảm thiểu tối đa tình trạng một máy chủ bị quá tải và ngưng hoạt động hoặc khi một máy chủ gặp sự cố, cân bằng tải sẽ chỉ đạo phân phối công việc của máy chủ đó cho các máy chủ còn lại, đẩy thời gian uptime (thời gian hoạt động) của hệ thống lên cao nhất và cải thiện năng suất hoạt động tổng thể.



Hình 0.5: Mô hình Cân bằng tải trong điện toán đám mây [9]

Trong môi trường phân tán, cân bằng tải là một trong những chủ đề hết sức quan trọng. Vì điện toán đám mây được coi là một trong những nền tảng tốt nhất giúp lưu trữ dữ liệu với chi phí tối thiểu và có thể truy cập mọi lúc qua internet, cân bằng tải cho điện toán đám mây đã trở thành một lĩnh vực nghiên cứu rất thú vị và quan trọng. Cân bằng tải nhằm mục đích thỏa mãn người dùng và sử dụng tỷ lệ tài nguyên cao bằng cách đảm bảo phân bổ hợp lý. Có rất nhiều khó khăn trong các kỹ thuật cân bằng tải như bảo mật, khả năng chịu lỗi, v.v ... vốn phổ biến trong môi trường điện toán đám mây hiện đại. Nhiều nhà nghiên cứu đã đề xuất một số kỹ thuật thuật toán để tăng cường nhằm tìm ra những phương án tốt nhất cho cân bằng tải.

Trong cân bằng tải thời gian gần đây đã nổi lên chủ đề phân tán dự đoán quá tải [11] như một giải pháp đầy hứa hẹn, trong đó chuyển sang cấp độ giám sát tình trạng tắc nghẽn của mỗi con đường và phân tán dòng chảy trực tiếp đến con đường ít tắc nghẽn. Cách tiếp cận này có nhiều lợi thế thực tiễn. Là một lược đồ phân phối, nó có thể mở rộng hơn và có thể đối phó với lưu lượng truy cập nhanh hơn cách lịch trình tập trung. Là một phương pháp tiếp cận dữ liệu, nó không phụ thuộc vào ngăn xếp mạng của máy chủ lưu trữ và ngay lập tức mang lại lợi ích cho tất cả lưu lượng truy cập khi triển khai. Khả năng hiển thị tắc nghẽn cuối cùng của nó cũng làm cho nó trở nên mạnh mẽ hơn mà không cần cấu hình lại máy điều khiển. Mấu chốt của việc thiết kế một giao thức cân bằng tải tắc nghẽn là chúng ta cần phải biết thông tin về tắc nghẽn thời gian thực từ tất cả các đường đi giữa nguồn dòng chảy và điểm đến. Một cách tiếp cận đơn giản là sử dụng thông tin định hướng đường đi cuối: một switch ToR duy trì các chỉ số tắc nghẽn đầu cuối cho tất cả các đường dẫn từ chính nó đến các thiết bị chuyển mạch ToR khác trong mạng. Các chỉ số tắc nghẽn có thể được thu thập bằng các gói dữ liệu. Thông thường, có hàng trăm đường dẫn tồn tại giữa hai ToR thiết bị chuyển mạch và công tắc ToR có thể giao tiếp với hàng trăm các thiết bị chuyển mạch ToR khác. Quan trọng hơn, không thể để thu thập thông tin tắc nghẽn thời gian thực cho tất cả các đường dẫn này, vì sẽ không có đủ dòng chảy đồng thời xảy ra đi cùng với tất cả chúng cùng một lúc. Trong giai đoạn đầu, chỉ có nguồn và thiết bị chuyển mạch ToR đích tham gia để lựa chọn tốt nhất đường dẫn từ ToR đến tầng tổng hợp. Chuyển đổi nguồn ToR sẽ gửi số liệu tắc nghẽn của nó đến đích ToR, chúng sẽ kết hợp với các chỉ số tắc nghẽn để chọn con

đường tốt nhất cho lớp tổng hợp. Trong giai đoạn thứ hai, tập hợp đã chọn sau đó chọn công tắc lỗi tốt nhất theo một cách tương tự về tình trạng tắc nghẽn của bước nhảy thứ hai và thứ ba. Con đường quyết định lựa chọn sau đó được duy trì tại ToR và tập hợp thiết bị chuyển mạch. Về cơ bản hai giai đoạn lựa chọn đường dẫn sử dụng thông tin một phần đường dẫn để tìm đường tốt nhất cho dòng chảy. Bằng cách khai thác các tính chất cấu trúc của 3 tầng, lựa chọn đường dẫn hai giai đoạn làm giảm đáng kể sự phức tạp mà không có nhiều hiệu suất. Trên thực tế, đánh giá cho thấy rằng thực hiện lựa chọn đường dẫn trên mỗi cơ sở lưu lượng trong TCP là tốt nhất và không gây ra việc sắp xếp lại gói tin cũng như không gây bất kỳ độ trễ nào.

Cân bằng tải luôn là chủ đề nghiên cứu nóng của các trung tâm dữ liệu đám mây, và mục tiêu của nó là đảm bảo rằng mọi tài nguyên máy tính có thể xử lý các nhiệm vụ một cách hiệu quả và nhanh chóng. Cuối cùng, việc sử dụng nguồn lực được cải thiện. Các nhà nghiên cứu đã đề xuất một loạt cân bằng tĩnh, động và chiến lược lập kế hoạch cân bằng tải. Ngoài ra, cũng có một số nghiên cứu sử dụng công nghệ di chuyển trực tiếp của máy ảo để đáp ứng các yêu cầu đám mây, nhiệm vụ của trung tâm dữ liệu là yêu cầu hiệu suất và giới hạn tải. Các chiến lược cân bằng tải hiện được chia thành hai loại: cân bằng tải tĩnh và cân bằng tải năng động [12]. Thuật toán lập lịch cân bằng tải tĩnh thường bao gồm round robin, weighted rounded robin [13]. Các thuật toán tĩnh chỉ sử dụng một số thông tin tĩnh mà không thể phản ánh tải động. Hiện nay, hầu hết các nền tảng mã nguồn mở cả IaaS đã sử dụng các thuật toán tĩnh để tiến hành lập kế hoạch tài nguyên. Lợi thế của thuật toán lập kế hoạch cân bằng tải tĩnh là nó rất đơn giản để sử dụng. Nhưng trong các trung tâm dữ liệu đám mây quy mô lớn có tính không đồng nhất của tài nguyên và nhu cầu người sử dụng là không nhất quán, hiệu quả cân bằng tải tĩnh không được lý tưởng. Cân bằng tải động (DLB - Dynamic Load Balancing), nó chủ yếu được sử dụng trong lĩnh vực phân phối máy tính song song, và mục tiêu chính của nó là làm thế nào để phân phối tải hợp lý hơn giữa nhiều máy chủ để tránh một số hiện tượng mà một số các nút máy tính bị quá tải và một số nút có tải nhẹ và do đó để cải thiện toàn bộ hiệu suất của hệ thống. Chi phí truyền thông bổ sung được tạo ra trong quá trình DLB sẽ làm suy giảm hiệu năng hệ thống của cân bằng tải động. Vì vậy, làm thế nào để giảm truyền gói tin trên cao nhất giữa các nút trong quá trình DLB trở

thành một vấn đề quan trọng sẽ ảnh hưởng đến hiệu suất của DLB. Tuy nhiên, một số thuật toán ở trên không thể đáp ứng được sự lựa chọn và bản chất của cơ cấu cân bằng tải tối ưu cùng một lúc. Vì vậy, những cách phân phối tiếp cận thường có được sự tối ưu cục bộ của các giải pháp. Và hiệu quả của việc giải quyết vấn đề phân phối tải trong một số trường hợp đặc biệt không phải là lý tưởng. Vì vậy, nó có thể đảm bảo cân bằng tải và sử dụng hiệu quả tài nguyên vật lý của toàn bộ cụm. Tuy nhiên, cân bằng tải là vấn đề và chi phí chung của đám mây trong các trung tâm dữ liệu không được xem xét. Nó chỉ tập trung vào quản lý máy ảo để tăng cường quản lý các trung tâm dữ liệu đám mây và nâng cao hiệu quả hoạt động của các trung tâm dữ liệu điện toán đám mây.

Theo tài liệu [14], Cân bằng tải có thể được chia thành 2 loại là: Cân bằng tải cục bộ và cân bằng tải toàn cục. Cân bằng tải cục bộ được sử dụng để cân bằng dự báo tải trong một trung tâm. Nó phân phối yêu cầu từ phía máy khách sang máy chủ đáp ứng nhu cầu. Về cân bằng tải toàn cục, nó quản lý và kiểm soát yêu cầu từ phía khách hàng tự động đến máy chủ qua nhiều trung tâm dữ liệu. Nó xử lý lưu lượng trên cả hai mặt gói truyền tải. Xử lý cân bằng tải toàn cầu cho sự phức tạp, nhưng đồng thời điều này là rất hữu ích cho truyền tải gói tin trên trung tâm dữ liệu mạng. Tính khả dụng đảm bảo rằng, trong trường hợp thất bại, hệ thống tiếp tục hoạt động như mong đợi.

1.2.2. Mục đích cân bằng tải

Cân bằng tải có một số mục đích chính sau đây:

- Tăng khả năng đáp ứng, tránh tình trạng quá tải trên máy chủ, đảm bảo tính linh hoạt và mở rộng cho hệ thống.
- Tăng độ tin cậy và khả năng dự phòng cho hệ thống: Sử dụng cân bằng tải giúp tăng tính khả dụng cao (HA - High Availability) cho hệ thống, đồng thời đảm bảo cho người dùng không bị gián đoạn dịch vụ khi xảy ra lỗi sự cố lỗi tại một điểm cung cấp dịch vụ.
- Tăng tính bảo mật cho hệ thống: Thông thường khi người dùng gửi yêu cầu dịch vụ đến hệ thống, yêu cầu đó sẽ được xử lý trên bộ Cân bằng tải, sau đó thành phần Cân bằng tải mới chuyển tiếp các yêu cầu cho các máy chủ bên trong. Quá trình trả lời cho khách hàng cũng thông qua thành phần Cân bằng

tải, vì vậy mà người dùng không thể biết được chính xác các máy chủ bên trong cũng như phương pháp phân tải được sử dụng. Bằng cách này có thể ngăn chặn người dùng giao tiếp trực tiếp với các máy chủ, ẩn các thông tin và cấu trúc mạng nội bộ, ngăn ngừa các cuộc tấn công trên mạng hoặc các dịch vụ không liên quan đang hoạt động trên các công cụ khác.

1.3. Tổng quan về trí tuệ nhân tạo (AI)

Trí tuệ nhân tạo là một lĩnh vực liên quan đến chuyên ngành khoa học máy tính và công nghệ thông tin, bản chất của trí tuệ nhân tạo vẫn do con người làm ra, họ xây dựng các thuật toán, lập trình bằng các công cụ phần mềm công nghệ thông tin, giúp các máy tính có thể tự động xử lý các hành vi thông minh như con người. Trí tuệ nhân tạo có khả năng tự thích nghi, tự học và tự phát triển, tự đưa ra các lập luận để giải quyết vấn đề, có thể giao tiếp như người...tất cả là do AI được cài một cơ sở dữ liệu lớn, được lập trình trên cơ sở dữ liệu đó và tái lập trình trên cơ sở dữ liệu mới sinh ra. Cứ như vậy cấu trúc của AI luôn luôn thay đổi và thích nghi trong điều kiện và hoàn cảnh mới

1.4. Tổng quan về học máy (ML)

Học máy (Machine Learning/ML) là một phương pháp để tạo ra AI. ML liên quan đến các chương trình máy tính viết lập trình của riêng chúng để hoàn thành một nhiệm vụ định trước. Quá trình này có thể được giám sát, bán giám sát, hoặc không giám sát. Trong học tập có giám sát, máy được cung cấp dữ liệu trong đó mỗi ví dụ trong tập dữ liệu được gắn nhãn với câu trả lời. Các sau đó máy học thông qua thử và sai để dự đoán câu trả lời từ tập dữ liệu đã nhập. Học tập không giám sát liên quan đến việc phân tích dữ liệu đầu vào mà không có câu trả lời xác định. Điều này thường được sử dụng để mô hình hóa cấu trúc và phân phối dữ liệu. Cuối cùng, học tập bán giám sát là một phương pháp kết hợp liên quan đến việc kết hợp dữ liệu được gắn nhãn và không được gắn nhãn. Điều này có thể giúp giảm bớt gánh nặng của nhiệm vụ ghi nhãn. Sử dụng các thuật toán phân lớp của ML để tiến hành phân lớp người dùng dựa trên các đặc trưng của họ để thực hiện việc cân bằng tải.

1.5. Người dùng cloud và hành vi người dùng cloud

Việc xác định hành vi người dùng có vai trò rất quan trọng trong nghiên cứu này. Việc xác định hành vi người dùng thông qua đoạn request họ gửi lên cloud, với

từng loại người dùng với những mục đích sử dụng khác nhau thì cấu trúc của những đoạn request này cũng sẽ khác nhau.

Vậy nên dựa vào sự khác nhau trong cấu trúc, thông tin của các request được gửi trên Cloud ta có thể dự đoán những tác vụ mà người dùng sẽ yêu cầu thực hiện khi kết nối với Cloud. Từ đó có sự phân bổ tài nguyên cho các người dùng hợp lý, tránh việc lãng phí tài nguyên.

1.6. Kết luận chương 1

Hiểu biết được những khái niệm tổng quan về điện toán đám mây, hiểu biết thuật toán điện toán đám mây giải quyết những vấn đề tắc nghẽn, gói tin mất mát khi truyền dữ liệu qua môi trường điện toán, mục đích tăng hiệu quả cân bằng của hệ thống.

CHƯƠNG 2. CÁC CÔNG TRÌNH LIÊN QUAN

2.1. Giới thiệu chung

Chương này sẽ giới thiệu các công trình nghiên cứu của Việt Nam và trên thế giới có liên quan, ứng dụng vào đề tài. Những công trình này đã đề xuất nhiều kỹ thuật phân bổ tài nguyên, phát hiện hành vi người dùng, cân bằng tải,... trong cloud. Qua việc nghiên cứu các công trình này cũng góp phần củng cố thêm cơ sở lý thuyết và định hình hướng nghiên cứu cho đề tài.

2.2. Các công trình nghiên cứu tại Việt Nam

Trong bài báo [12] của Trần Công Hùng và các cộng sự đăng trên tạp chí Khoa học công nghệ Thông tin và truyền thông số 04(CS.01) 2018 của Học viện Công nghệ Bưu chính viễn thông, đã đề xuất một thuật toán cân bằng tải nhằm giảm thời gian đáp ứng trên điện toán đám mây, ý tưởng chính là sử dụng thuật toán dự báo ARIMA để dự báo thời gian đáp ứng, từ đó đưa ra cách giải quyết phân phối tài nguyên hiệu quả dựa vào giá trị ngưỡng thời gian. Bài báo đã đưa ra thuật toán, thử nghiệm mô phỏng với mô hình nhỏ và đã đạt được một số kết quả mô phỏng khá tích cực, và tiềm năng trong dự báo tương lai gần.

Trong bài báo [13] của tác giả Nguyễn Thanh Thủy và các cộng sự đăng trên tạp chí “International Journal of Computer Science and Network, Volume 4, Issue 2, April 2015”, đã trình bày một cách tiếp cận để cải thiện thuật toán ngăn chặn bế tắc, đề lên lịch cho các chính sách cung cấp tài nguyên để phân bổ tài nguyên không đồng nhất. Thuật toán ngăn chặn bế tắc có độ phức tạp thời gian chạy là $O(m \cdot n)$ (min (m, n)), trong đó m là số lượng tài nguyên và n là số lượng quy trình. Họ đề xuất thuật toán phân bổ nhiều tài nguyên cho các dịch vụ cạnh tranh đang chạy trong các máy ảo trên nền tảng phân tán không đồng nhất. Các thí nghiệm cũng so sánh hiệu suất của phương pháp đề xuất với các công việc liên quan khác.

2.3. Một số công trình nghiên cứu trên thế giới

Trên tạp chí Journal of Internet Technology Volume năm 2019 [15], nhóm tác giả Ruoshui Liu, Xin Wang, Juan Du, Ping Xie, cũng đã công bố quốc tế “A Cloud User Behavior Authentication Model Based on Multi-label Hyper-network”. Với sự ra đời của kỷ nguyên Dữ liệu lớn, việc bảo mật thông tin người dùng là đặc

biệt quan trọng. Làm thế nào để xây dựng lòng tin giữa người dùng và đám mây là một vấn đề quan trọng. Để giải quyết vấn đề này, bài báo này đề xuất một mô hình xác thực hành vi người dùng đám mây dựa trên mạng đa cấp nhãn, thực hiện phân chia chi tiết hành vi của người dùng và cải thiện độ chính xác của phát hiện bất thường. Phương pháp này huấn luyện cơ sở dữ liệu hành vi bình thường của người dùng thành một siêu mạng, thêm hành vi của người dùng hiện tại như một thể hiện vào siêu mạng để phân loại. Nếu một nhãn được tìm thấy thành công trong phân loại này, nhãn đó được xác định là người dùng bình thường. Nếu không, mô hình cập nhật trọng số của siêu mạng, thay thế siêu cạnh và tìm kiếm lại nhãn. Nếu nhãn được tìm thấy, nó được xác định là người dùng rủi ro, nếu không nó được xác định là người dùng độc hại. Kết quả mô phỏng cho thấy có sự cải thiện đáng kể về độ chính xác của phân loại. Áp dụng phương pháp trong bài báo này để phát hiện hành vi của người dùng có thể cải thiện hiệu quả tỷ lệ phát hiện hành vi của người dùng, thực hiện phân tích chi tiết về hành vi của người dùng và cải thiện khả năng xử lý hành vi của người dùng.

Năm 2021, Xun Xu; Shuo Zeng; Yuanjie He [19], đã công bố nghiên cứu về “The impact of information disclosure on consumer purchase behavior on sharing economy platform Airbnb”, thông qua các bằng chứng thực nghiệm từ Airbnb từ tám thành phố lớn ở Hoa Kỳ, nhóm tác giả xem xét vai trò của thông tin tiết lộ trong việc tác động đến hành vi mua hàng của người tiêu dùng trên nền tảng kinh tế chia sẻ này. Chúng tôi phân tích việc công bố thông tin từ bốn khía cạnh - đó là thông tin gì (tức là nội dung thông tin), từ đâu (tức là, nguồn thông tin), ở định dạng nào (hình thức trình bày thông tin), và bao nhiêu (số lượng thông tin). Chúng tôi tìm thấy cả ba nguồn thông tin - nhà cung cấp, nền tảng và người tiêu dùng ngang hàng - ảnh hưởng đến người tiêu dùng hành vi mua hàng. Liên quan đến thông tin do các nhà cung cấp đăng tải, chúng tôi nhận thấy mối quan hệ lồi giữa thông tin (tức là số lượng ảnh và độ dài của mô tả) và hành vi mua hàng của người tiêu dùng. Tuy nhiên, không có mối quan hệ đáng kể nào giữa phần tự mô tả của nhà cung cấp (văn bản và ảnh) và hành vi mua hàng của người tiêu dùng hành vi được tìm thấy. Về thông tin được đăng bởi nền tảng, cả khuyến nghị từ nền tảng và thông tin xác minh nhà cung cấp ảnh hưởng tích cực đến hành vi mua hàng của người tiêu dùng. Đối

với thông tin về tương tác giữa nhà cung cấp và người tiêu dùng, tỷ lệ phản hồi cao và tốc độ phản hồi nhanh của nhà cung cấp sẽ nâng cao khả năng mua hàng của người tiêu dùng hành vi. Tuy nhiên, việc cung cấp kết nối với hồ sơ mạng xã hội của nhà cung cấp ảnh hưởng tiêu cực đến người tiêu dùng hành vi mua hàng. Về thông tin từ người tiêu dùng ngang hàng, chúng tôi nhận thấy mặc dù về tổng thể, người tiêu dùng xếp hạng ảnh hưởng tích cực đến hành vi mua hàng của người tiêu dùng, ảnh hưởng đó giảm đi khi xếp hạng vượt quá mức nhất định các ngưỡng. Nghiên cứu của chúng tôi cung cấp các gợi ý cho chủ sở hữu nền tảng để tối ưu hóa bố cục trình bày thông tin trực tiếp thông qua thiết kế nền tảng hoặc gián tiếp thông qua hướng dẫn tiết lộ thông tin của nhà cung cấp để tạo điều kiện thuận lợi cho việc tìm kiếm và thu thập thông tin của người tiêu dùng nhằm giảm rủi ro được nhận thức, nâng cao lòng tin đối với các nhà cung cấp và nền tảng, đồng thời nâng cao ý định và hành vi mua hàng của họ.

2.4. Kết luận Chương 2

Chương này đã trình bày một số công trình nghiên cứu liên quan trong nước và quốc tế. Các nghiên cứu này giúp ta hiểu rõ hơn về cân bằng tải trên môi trường điện toán đám mây, nắm bắt được ưu nhược điểm của các thuật toán cân bằng tải cũng như hành vi của người dùng hiện nay, từ đó đưa ra các thuật toán cải tiến với mục tiêu nâng cao hiệu suất. Chương tiếp theo sẽ trình bày đề xuất thuật toán cân bằng tải trên điện toán đám mây thông qua hành vi người dùng cloud.

CHƯƠNG 3. ĐỀ XUẤT THUẬT TOÁN CÂN BẰNG TẢI TRÊN ĐIỆN TOÁN ĐÁM MÂY THÔNG QUA HÀNH VI NGƯỜI DÙNG CLOUD

3.1. Giới thiệu chung

Trong vài năm trở lại đây, nhiều bài báo đã trình bày các thuật toán trong cân bằng tải [20], [21], [22] và đề xuất cải tiến chúng với mục đích nâng cao hiệu suất cân bằng tải trong môi trường điện toán đám mây. Sau quá trình nghiên cứu và tham khảo những công trình nghiên cứu trên, đề tài này quyết định sẽ sử dụng một số thuật toán phân lớp (classification) trong AI, cụ thể là sử dụng thuật toán Cây quyết định J48 kết hợp với việc nghiên cứu hành vi người dùng cloud nhằm đề xuất một ý tưởng cân bằng tải mới, từ đó có thể giúp cho các nhà cung cấp dịch vụ cũng như người dùng cloud hoạt động hiệu quả hơn. Hướng đến mục tiêu này, chương này sẽ trình bày tổng quát về ý tưởng bên cạnh đó cũng đề xuất việc xây dựng thuật toán cân bằng tải trên điện toán đám mây thông qua hành vi của người dùng cloud nhằm nâng cao hiệu suất cân bằng tải trên môi trường điện toán đám mây bao gồm hệ thống host/datacenter và các máy ảo.

3.2. Mô hình nghiên cứu

Thuật toán phân lớp cây quyết định J48 (Decision Tree J48 Algorithm) sẽ được áp dụng vào mô hình nghiên cứu với mục tiêu phân bổ tài nguyên tương ứng với các Request thông qua hành vi của người dùng cloud mà ta đã dự đoán. Hành vi người dùng ở đây được phân chia trên request mà họ mang tới dựa theo các chỉ như kích thước của đầu vào (Filesize), đầu ra (OutputSize), độ dài (Length), thời gian thực hiện (execStartTime), thời gian kết thúc (finishTime),... Sau quá trình phân loại các tác vụ theo hành vi người dùng, bộ cân bằng tải sẽ tiến hành phân bổ request chứa tác vụ với hành vi cần nhiều tài nguyên hơn vào những máy ảo có khả năng xử lý tốt hoặc ngược lại, vào những máy vừa và thấp. Thông qua cách tiếp cận này, thuật toán mà luận văn đề xuất sẽ cải thiện được thời gian xử lý cân bằng tải trên đám mây, và có thể ứng dụng trên môi trường đám mây theo thời gian thực. Luận văn cũng xin tạm đặt tên cho thuật toán là J-TUBA (J48 classifier of Task through User Behavior).

Mục tiêu:

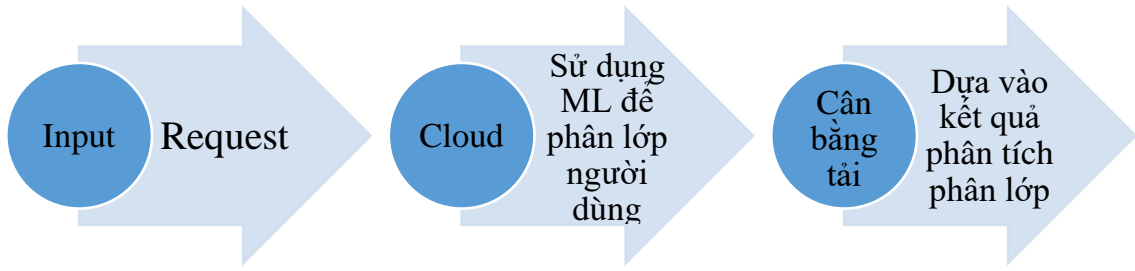
- Hạn chế và giảm thiểu các rủi ro cho datacenter.
- Làm giảm thời gian sống của các request trong đám mây.
- Ngăn chặn mất cân bằng tải và hạn chế tối đa sự mất cân bằng tải giữa các máy ảo.
- Giải quyết các request nhanh hơn, phân loại các tác vụ với các hành vi người dùng khác nhau tương ứng với các request, sử dụng có hiệu quả nguồn tài nguyên đám mây, đáp ứng một cách tốt nhất cho người dùng.
- Có thể phân lớp được các request tiếp theo tương ứng với hành vi người dùng đã được phân lớp ở trên, từ đó có kế hoạch đưa các request này sang những máy ảo/host có khả năng xử lý tải tương ứng.
- Sắp xếp các máy ảo/host/tài nguyên sao cho mức độ sử dụng từ cao đến thấp để phân bổ các tác vụ hợp lý.

Giả định:

- Bộ cân bằng tải sẽ biết trước các dịch vụ nào đang chạy trên các máy ảo vào bất cứ thời điểm nào.
- Đề xuất tập trung vào mô phỏng dịch vụ Web (Web Service), các web server (máy ảo) sẽ biết trước thời gian xử lý của từng dịch vụ chạy trên web và trên từng máy ảo.
- Hai máy ảo có cấu hình tương đương nhau về RAM, vi xử lý, và I/O thì thời gian thực thi của các dịch vụ sẽ không mấy là khác nhau.

Mô hình nghiên cứu:

- Quá trình cân bằng tải được thực hiện gồm các bước như sau:
 - + Bước 1: Nhận thông tin input (các request nhận được)
 - + Bước 2: Sử dụng các thuật toán phân lớp của máy học để tiến hành phân lớp các request dựa trên các đặc trưng của các request. Các đặc trưng này dựa trên các hành vi của người dùng trên internet.
 - + Bước 3: Dựa vào kết quả phân lớp các request tiến hành cân bằng tải.



Hình 0.1 Mô hình cân bằng tải

- Với đầu vào là các yêu cầu từ người dùng (request) thuật toán đề xuất sẽ xử lý và đưa vào các máy ảo phù hợp để tiến hành quá trình cân bằng tải.

- Mô hình này áp dụng K-means (dựa vào tính chất của các request) mà phân loại các đầu vào này, sau đó từ các tác vụ mà request mang tới (CPU Usage, RAM Usage, Power) ta dự báo thông số cloud sao cho phù hợp. Để có thể tiến hành phân lớp với kỹ thuật này, thuật toán sẽ sử dụng bộ dữ liệu trong lịch sử cloud đã được lưu lại (dữ liệu gần đây nhất).

3.3 Thuật toán Cây quyết định J48 và ứng dụng vào phân lớp request dựa trên hành vi người dùng

- Thuật toán Cây quyết định J48

J48 là một mã nguồn mở của Java triển khai thuật toán cây quyết định C4.5. J48 là một phần mở rộng của ID3, các tính năng bổ sung của J48 là tính toán các giá trị bị thiếu, cắt tĩa cây quyết định, dẫn xuất các quy tắc, v.v.. Là một bộ phân loại cây quyết định (decision tree classifier), J48 sử dụng mô hình học máy dự đoán để tính toán giá trị kết quả của một mẫu mới dựa trên các giá trị thuộc tính khác nhau của dữ liệu có sẵn. Các nút bên trong của cây quyết định biểu thị các thuộc tính khác nhau; các nhánh giữa các nút cho chúng ta biết các giá trị có thể có mà các thuộc tính này có thể có trong các mẫu quan sát, trong khi các nút đầu cuối cho chúng ta biết giá trị cuối cùng (phân loại) của biến phụ thuộc.

- Phân lớp tác vụ dựa theo hành vi người dùng

Dựa vào bộ dữ liệu trước đây khi xử lý những tác vụ đầu tiên, ta sử dụng J48 để phân lớp dựa trên hành vi người dùng cho các task tiếp theo. Tương ứng với mỗi yêu cầu (Request) sẽ có một tác vụ mà máy tính cần phải thực hiện để phục vụ

người dùng. Chính vì thế, bất kỳ một request được gửi đến cloud đều có thể được phân lớp dựa trên tác vụ tương ứng của nó.

3.4 Thuật toán đề xuất J-TUBA

Thuật toán đề xuất J-TUBA (J48 classifier of Task through User Behavior), dựa vào hành vi người dùng (User Behavior được mô tả ở trên) tương ứng với các request. Để có thể phân bổ tài nguyên cho các request một cách tối ưu và hiệu quả nhất ta sử dụng thuật toán J48 để phân lớp các request này. Bên cạnh đó, các tài nguyên (máy ảo/ host) sẽ được sắp xếp theo mức độ sử dụng tăng dần. Kết hợp với đánh giá số lần sai cùng với sai số, ta cải thiện thuật toán bằng cách áp dụng máy học vào, tuy nhiên, việc áp dụng này sẽ ít diễn ra vì có sai số cho phép.

Dựa vào tham khảo từ tài liệu [20], luận văn này xin đề xuất thuật toán gồm 3 nhóm module chính:

- (1) Module tính toán ra các thông số của request bằng thuật toán K-means
- (2) Module phân lớp tác vụ theo hành vi người dùng
- (3) Module phân bổ các dịch vụ (chọn máy ảo)

Thuật toán J-TUBA

1. **For each** Request in CloudRequests
2. isLocated = false;
3. UserBehavior = {Po, CPU, RAM}_{new} = K-means(T₁, T₂, ..., T_n); // Module 1
4. Request.UserBehaviorClass = J48(UserBehavior); //J48 là mô hình phân lớp tác vụ
5. **For each** VM in VMList
6. **If** isFitSituation(Request. UserBehaviorClass, VM)
7. AllocateRequestToVM(VM, Request); // Module 3
8. isLocated = true;
9. **End If**
10. **End For**
11. **If**(!isLocated)
12. VM = VMList.getSelectedVM(); // Module 2
13. AllocateRequestToVM(VM, Request);
14. **End If**

15. End For

Phương pháp đánh giá thuật toán J-TUBA

Kết quả đạt được từ thuật toán mà luận văn đề xuất đã đáp ứng các mục tiêu, chẳng hạn như giới hạn số lượng yêu cầu xếp hàng để phân phối, cải thiện thời gian xử lý và phản hồi của đám mây trung tâm so với các thuật toán cũ. Điều này cũng có nghĩa là với thuật toán được đề xuất, hiệu năng của điện toán đám mây được cải thiện so với bốn thuật toán FCFS, MaxMin, MinMin và Round Robin.

3.5. Kết luận chương 3

Để phục vụ cho thuật toán cân bằng tải, chương này đã trình bày lý do tác giả chọn phương pháp phân lớp tác vụ [17], [21], [22] theo hành vi của người dùng cloud. Để cloud có thể giữ được trạng thái an toàn và hoạt động liên tục, thuật toán đề tài thực hiện đã cải tiến các thuật toán cân bằng tải trong điện toán đám mây – J-TUBA (J48 classifier of Task through User Behavior) sẽ giải quyết được sự cân bằng tải dựa trên việc cải thiện thời gian thực thi tác vụ từ đó lượng số lượng thất bại trong việc triển khai sẽ giảm đi, số điểm chết nút sẽ ít hơn các thuật toán hiện tại.

CHƯƠNG 4. MÔ PHỎNG, THỰC NGHIỆM

4.1. Giới thiệu chung

Chương này sẽ trình bày việc cài đặt và mô phỏng thuật toán J-TUBA (J48 classifier of Task through User Behavior), cụ thể là sử dụng thuật toán phân lớp Cây quyết định J48 (Decision Tree J48 Algorithm) với mục tiêu là loại các tác vụ tương ứng với các Request dựa trên đặc trưng của tác vụ đó. Đặc trưng này tính toán dựa trên hành vi của người dùng cloud (loại người dùng và mục đích sử dụng). Sau khi phân loại các tác vụ theo đặc trưng, bộ cân bằng tải sẽ tiến hành phân phối các request có tác vụ đó vào những máy ảo/host phù hợp. Từ đó, phân bổ request có nhu cầu xử lý nhiều vào máy ảo/host có mức độ hoạt động thấp nhất. Với cách tiếp cận này, thuật toán đề xuất J-TUBA sẽ cải thiện thời gian xử lý cân bằng tải trên cloud, và ứng dụng trên môi trường cloud theo thời gian thực. Sau khi tiến hành các bước như trên ta thu được các kết quả từ đó phân tích tính hiệu quả của thuật toán đề ra.

4.2. Xây dựng mô hình mô phỏng – thực nghiệm

Dựa vào bộ dữ liệu chứa các request, luận văn này sử dụng thuật toán K-Means để phân loại request dựa trên các đặc trưng của request (hành vi người dùng cloud)

Bước 1: Tiếp nhận giá trị input

Bước 2: Phân tích các input để rút trích các đặc trưng của các input

Bước 3: Dựa vào các đặc trưng trên chúng ta sử dụng machine learning để phân lớp các đầu vào.

Bước 4: Dựa vào kết quả phân lớp ta tiến hành cân bằng tải cho hệ thống.

Dựa vào dữ liệu của các request mà ta có thể biết, ta sử dụng thuật toán K-means để phân loại request bằng cách tính toán ra bộ Priority = {Power, CPU, RAM}, từ đó ta biết cách phân bổ tài nguyên cho cái request vào các máy ảo đã phân cụm. Kết hợp với đánh giá số lần sai, và sai số, ta cải thiện thuật toán bằng cách áp dụng máy học vào, tuy nhiên, việc áp dụng này sẽ ít diễn ra vì có sai số cho phép.

Ta tiến hành giả lập môi trường cloud bằng cách lập trình trên ngôn ngữ JAVA và sử dụng bộ thư viện CloudSim. Môi trường này sẽ bao gồm 5 đến 15 máy

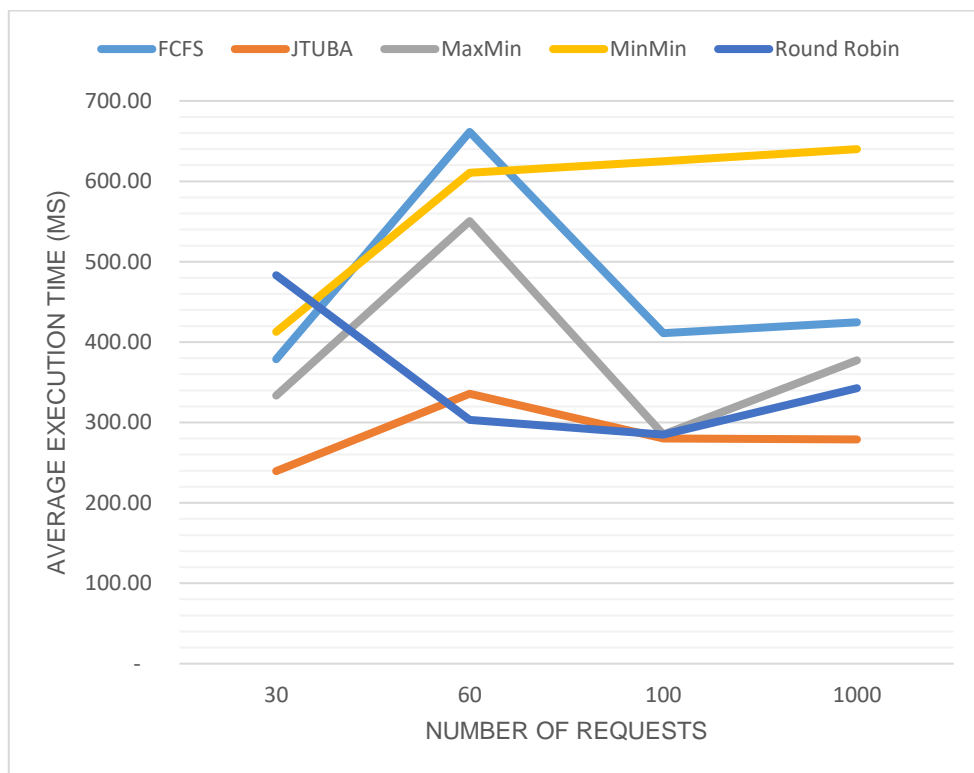
ảo, và đồng thời tạo môi trường request ngẫu nhiên tới các dịch vụ trên cloud này (bao gồm dịch vụ cung cấp máy ảo, dịch vụ cung cấp và đáp ứng người dùng của cloudSim) để tiến hành thử nghiệm.

Cài đặt thuật toán K-means, J48 trên môi trường mô phỏng được phát triển bởi bộ thư viện Weka và kiểm nghiệm kết quả.

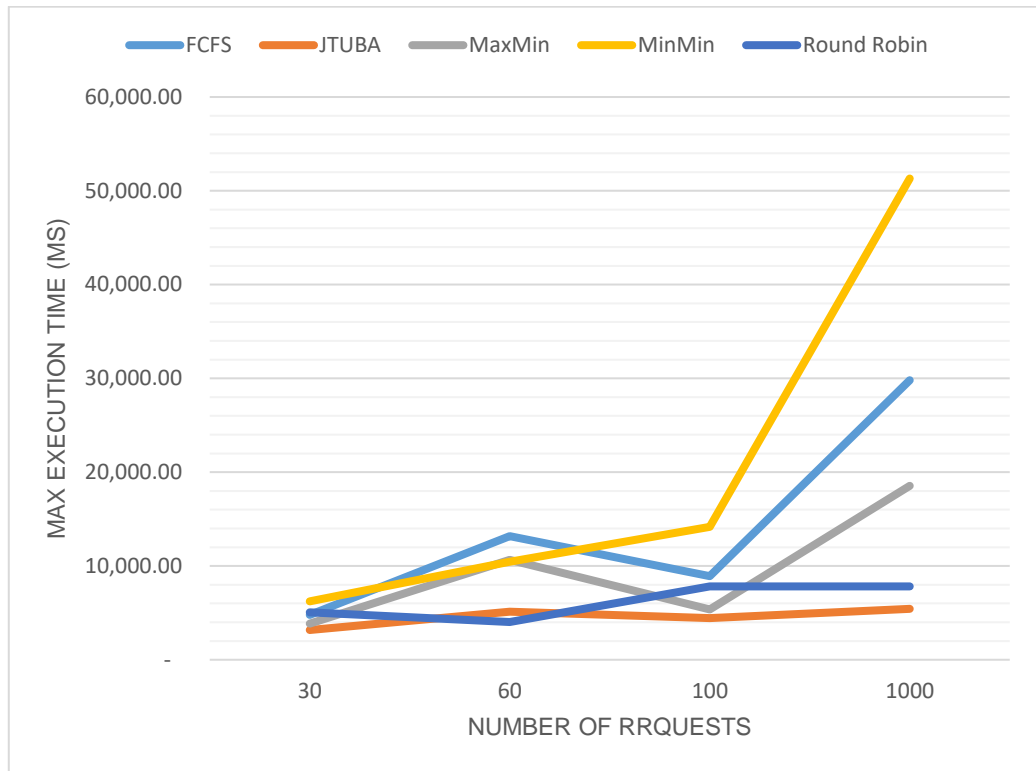
4.3. Kết quả thực nghiệm của mô hình

Tiến hành chạy thực nghiệm mô phỏng trên CloudSim với 5 máy ảo đã được dựng sẵn để đáp ứng các request, với các request được khởi tạo có chiều dài và kích thước ngẫu nhiên và số lượng Request lần lượt là 30, 60, 100 và 1000

Với 4 trường hợp 30, 60, 100 và 1000 request, sau khi so sánh thời gian xử lý của các thuật toán với cùng điều kiện ta đã thấy được sự phân bố khá ổn định và hợp lý của thuật toán đề xuất J-TUBA, thời gian xử lý của các máy ảo không quá khác biệt so với thời gian xử lý của các thuật toán khác trên cloud (ở trường hợp ít và nhiều request). Hình 4.5 và 4.6 cho thấy J-TUBA luôn thấp nhất, kể cả giá trị trung bình lẫn giá trị max.



Hình 4.5 Thời gian thực hiện trung bình của 5 thuật toán từ 30-1000 Request



Hình 0.2 Thời gian thực hiện lớn nhất của 5 thuật toán từ 30-1000 Request

Thuật toán đề xuất đã cho thấy hiệu quả khi máy ảo có số lượng cao lên thì J-TUBA đảm bảo thời gian phản hồi và thời gian xử lý tốt, giảm chi phí của các trung tâm dữ liệu đám mây. Tuy nhiên thuật toán vẫn còn 1 số nhược điểm như:

- Nếu số lượng máy ảo nhiều thì việc tìm ra máy có Usage nhỏ nhất là khó khăn hơn, nên có thể tìm một máy có Usage phù hợp là đạt.
- Chưa sắp xếp các máy ảo theo danh sách tăng dần.
- Nếu số lượng máy ảo nhiều thì việc tìm ra máy có Usage nhỏ nhất là khó khăn hơn, nên có thể tìm một máy có Usage phù hợp là đạt.
- Chưa sắp xếp các máy ảo theo danh sách tăng dần.

4.4 Phân tích, đánh giá hiệu quả của mô hình

Chương này đã trình bày mô hình thực nghiệm mô phỏng, các thông số cũng như kịch bản đưa ra là dựa vào quá trình request của các browser trên môi trường cloud. Từ đó, ghi nhận các thông số về thời gian xử lý của các máy ảo, và của cloud. Việc chạy thực nghiệm mô phỏng với thông số 5 máy ảo, chịu tải từ 30 tới 1000 request đã cho thấy kết quả tương đối tốt, việc phân bổ các request tới các máy ảo xử lý khá đồng đều, và tính khả thi cao.

KẾT LUẬN

Luận văn “**Đề xuất thuật toán cân bằng tải trên điện toán đám mây thông qua hành vi người dùng cloud**” nghiên cứu các thuật toán phân lớp từ các đặc trưng của request – hành vi người dùng. Từ đó phân bổ các tác vụ sao cho hợp lý và nâng cao cân bằng tải trong môi trường điện toán đám mây, sử dụng hiệu quả nguồn tài nguyên đám mây. Từ các thuật toán [23] đã có, tiến hành phân tích làm rõ chúng sau đó đánh giá đưa ra lợi thế cũng như nhược điểm của từng thuật toán, xem xét các nhược điểm đã phân tích để đề xuất thuật toán mới nhằm cải tiến và nâng cao khả năng cân bằng tải so với thuật toán cũ. Quá trình nghiên cứu của tác giả đã đạt được những mục tiêu như sau:

- Nghiên cứu tổng quan về đám mây và các kỹ thuật cân bằng tải được dùng trong môi trường điện toán đám mây.

- Nghiên cứu cách tiếp cận điện toán đám mây thông qua mô phỏng sử dụng bộ thư viện CloudSim. Cài đặt, mô phỏng các kỹ thuật cân bằng tải, các thuật toán MaxMin, MinMin, Round Robin và thuật toán tự nhiên FCFS. Các giá trị thu được khi mô phỏng đưa ra để phân tích so sánh với nhau để tổng hợp các ưu nhược điểm của các thuật toán từ đó có hướng đề xuất một thuật toán mới để khắc phục những mặt hạn chế đó.

- Kết quả đạt được từ thuật toán đề xuất đáp ứng được các mục tiêu như: cải thiện thời gian đáp ứng, hạn chế tài nguyên rảnh rỗi, máy ảo có năng lực xử lý mạnh sẽ được xử lý nhiều yêu cầu hơn. Giúp cân bằng tải hiệu quả hơn thuật toán được so sánh là MaxMin, MinMin, Round Robin và thuật toán tự nhiên FCFS.

- Thuật toán đề xuất J-TUBA có thể dùng để áp dụng trên thực tế.

Hạn chế của luận văn

- Chưa được ứng dụng vào môi trường thực tế.
- Thời gian đáp ứng và xử lý chưa cải thiện được nhiều.

Hướng phát triển đề tài

- Đưa thuật toán đề xuất ứng dụng vào môi trường đám mây cụ thể, hệ thống thực.

- Tiếp tục nghiên cứu, tối ưu các tham số đầu vào nhằm nâng cao hơn hiệu năng cân bằng tải.