

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN
THÔNG

LÊ QUANG HÙNG

**KỸ THUẬT ĐIỀU KHIỂN LƯU LƯỢNG TRONG
MẠNG MÁY TÍNH CỤM**

Chuyên ngành: Khoa học máy tính

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên, năm 2015

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn “Kỹ thuật điều khiển lưu lượng trong mạng máy tính cụm” là công trình nghiên cứu do tôi thực hiện dưới sự hướng dẫn của PGS.TS. NGUYỄN VĂN TAM. Các nội dung được trình bày trong luận văn là những kết quả đạt được trong thời tôi gian thực đề tài dưới sự hướng của tập thể giáo viên hướng dẫn, tôi không sao chép nguyên bản lại kết quả của các nghiên cứu đã từng được công bố và đây cũng là kết quả của quá trình nghiên cứu, học tập và làm việc nghiêm túc của tôi trong quá trình học cao học. Bên cạnh đó, trong một số nội dung luận văn là kết quả phân tích, nghiên cứu, tổng hợp từ nhiều nguồn tài liệu khác. Các thông tin tổng hợp hay các kết quả lấy từ nhiều nguồn tài liệu khác đã được tôi trích dẫn một cách đầy đủ và hợp lý. Nguồn tài liệu tham khảo có xuất xứ rõ ràng và được trích dẫn hợp pháp.

Các số liệu và thông tin sử dụng trong luận văn này là trung thực.

Thái Nguyên, ngày 28 tháng 08 năm 2015

Người cam đoan

MỤC LỤC

LỜI CẢM ƠN	Error! Bookmark not defined.
LỜI CAM ĐOAN.....	1
MỤC LỤC.....	3
DANH SÁCH CÁC TỪ VIẾT TẮT.....	5
DANH MỤC HÌNH ẢNH.....	6
MỞ ĐẦU	8
CHƯƠNG I: TỔNG QUAN MÁY TÍNH CỤM VÀ VẤN ĐỀ ĐIỀU KHIỂN LƯU LƯỢNG.....	10
1.1 Tổng quan về hệ thống máy tính cụm.....	10
1.1.1 Khái niệm máy tính cụm.....	10
1.1.2 Các loại máy tính cụm	11
1.1.3 Kiến trúc của một Cluster	11
1.1.4 Chế độ hoạt động của Cluster	13
1.1.5 Linux Cluster.....	14
1.2 Lý thuyết lưu lượng.....	16
1.2.1 Khái niệm về lưu lượng và đơn vị Erlang.....	16
1.2.2 Hệ thống tổn thất (Loss System) và công thức Erlang B.....	20
1.2.3 Hệ thống trễ (Delay) và công thức Erlang C	22
1.3 Khái niệm điều khiển lưu lượng	23
1.4 Nhiệm vụ điều khiển lưu lượng	23
1.5 Các cơ chế điều khiển lưu lượng điển hình.....	25
1.5.1 Cơ chế cấp lại ARQ	25
1.5.2 Cơ chế cửa sổ	27
1.5.3 Cơ chế điều khiển truy cập mạng (hạn chế băng thông).....	28
CHƯƠNG 2: KỸ THUẬT ĐIỀU KHIỂN LƯU LƯỢNG TRONG MẠNG MÁY TÍNH CỤM	30
2.1 Hệ thống hàng đợi	30
2.1.1 Các đặc trưng của hệ thống hàng đợi []	30
2.1.2 Phân tích một số mô hình hàng đợi.....	31

2.1.3 Kỹ thuật hàng đợi.....	36
2.2 Điều khiển lưu lượng theo thuật toán gáo rò (leaky bucket)	43
2.2.1 Nguyên lý của thuật toán gáo rò	43
2.2.2 Mô hình giải tích	44
2.2.3 Thuật toán gáo rò (Leaky bucket)	45
2.2.4 Thuật toán gáo rò trong điều khiển lưu lượng	47
CHƯƠNG 3: CHƯƠNG TRÌNH THỰC NGHIỆM	51
3.1 Nhiệm vụ của luận văn.....	51
3.1.1 Bài toán đặt ra	51
3.1.2 Mô hình hệ thống	52
3.2 Xây dựng một Web Cluster.....	52
3.2.1 Mô hình	52
3.2.2 Cài đặt hai server chạy hệ điều hành Linux trên Vmware Workstation.....	54
3.2.3 Cài đặt Apache và PHP trên Linux CentOS	55
3.2.4 Cài đặt và cấu hình heartbeat trên các Node	56
3.3 Chương trình thuật toán gáo rò	58
3.4 Một số kịch bản thử nghiệm.....	61
3.5 Đánh giá kết quả.....	63
KẾT LUẬN.....	64
TÀI LIỆU THAM KHẢO.....	65
PHỤ LỤC : KẾT QUẢ HIỂN THỊ SAU KHI CÀI ĐẶT XONG HEARTBEAT ...	66

DANH SÁCH CÁC TỪ VIẾT TẮT

Từ viết tắt	Viết đầy đủ	Ý nghĩa
ARQ	Automatic Repeat Request	yêu cầu lặp lại tự động
TCP	Transmission Control Protocol	Giao thức điều khiển truyền vận
FBA-TCP	Fair Bandwidth Allocation for TCP	Phân bổ băng thông công bằng cho TCP
LCC	Lost Calls Cleared	Mô hình LCC (Mô hình tổn thất)
PCT -I	Pure Chance Traffic Type I	
PASTA	Poisson Arrival See Time Average	
FEC	Forward Error Correction	Sửa lỗi trực tiếp bên thu
WFQ	Weighted Fair Queue	Hàng đợi cân bằng có trọng số
HA Cluster	High-availability clusters/HA	Hệ thống máy tính cụm có độ sẵn sàng cao

DANH MỤC HÌNH ẢNH

Hình 1.1 Hệ thống máy tính cụm của trung tâm vật lý lý thuyết, viện Vật lý – Viện Hàn Lâm Khoa học và công nghệ Việt Nam	11
Hình 1.2 Kiến trúc một Hadoop Cluster	13
Hình 1.3 Nguyên lý hoạt động của một Cluster.....	13
Hình 1.4 Sơ đồ nguyên lý của một Linux cluster lớn	15
Hình 1.5: Lưu lượng mạng (mật độ)(bằng số thiết bị bận) là một hàm thời gian (đường cong C). Lưu lượng trung bình trong khoảng thời gian T (đường cong D)....	17
Hình 1.6 Hoạt động của mạng khi không có sự kiểm soát	23
Hình 1.7 Phát lại theo cơ chế dừng và đợi	25
Hình 1.8 Nguyên tắc hoạt động của cơ chế cửa sổ trượt	27
Hình 1.9: (a) thuật toán gáo rò với nước, (b) thuật toán gáo rò với các gói tin ...	28
Hình 2.1 Mô hình chung của hệ thống hàng đợi.....	30
Hình 2.2 Chuỗi Markov của một quá trình sinh-tử.....	33
Hình 2.3 Chuỗi Markov của hàng đợi M/M/1	34
Hình 2.4 Chuỗi Markov của hàng đợi M/M/1	35
Hình 2.5 Leaky bucket	37
Hình 2.6 Token Bucket	38
Hình 2.8 Hàng đợi ưu tiên.....	41
Hình 2.9 Custom Queue.....	42
Hình 2.10 Weighted Fair Queue	43
Hình 2.11 Mô hình gáo rò	43
Hình 2.12 Mô hình gáo rò bằng kí hiệu toán học	44
Hình 2.13 Mô hình chuyển đổi sang hàng đợi	45
Hình 2.14 Lưu đồ thuật toán gáo rò	46
Hình 2.15 Thuật toán gáo rò	47
Hình 2.16 Điều khiển lưu lượng đưa vào mạng bằng thuật toán cái gáo rò	47
Hình 2.17 Sử dụng thuật toán cái gáo rò để giới hạn trễ tối đa	48
Hình 2.18 Ví dụ chức năng định dạng lưu lượng của thuật toán gáo rò	49

Hình 3.1 Mô hình mạng của công ty cổ phần Sách giáo dục điện tử EDC	51
Hình 3.2 Mô hình Web Cluster	53
Hình 3.3a Nội dung file ifcfg-eth0 cho máy Node 1	54
Hình 3.3b Nội dung file ifcfg-eth0 cho máy Node 2	55
Hình 3.4 Cấu hình httpd	58
Hình 3.5 : Trước khi điều khiển lưu lượng	62
Hình 3.6 Sau khi điều khiển lưu lượng	63

MỞ ĐẦU

Ngày nay, nền kinh tế của đất nước đang ngày một phát triển và đang hoà nhập với nền kinh tế của khu vực cũng như của thế giới. Cùng với sự phát triển đó mạng máy tính đã và đang trở nên rất quan trọng đối với chúng ta trong mọi lĩnh vực như: Khoa học, quốc phòng, thương mại, giáo dục...hiện nay ở nhiều nơi, mạng đã trở thành một nhu cầu không thể thiếu được.

Việc truyền dữ liệu trong mạng phụ thuộc vào nhiều yếu tố, trong đó có chiến lược cung cấp tài nguyên của mạng (đường truyền, bộ nhớ đệm...). Nếu khả năng tài nguyên có hạn, không có chiến lược cấp phát phù hợp và thích nghi với trạng thái luôn thay đổi của mạng thì dẫn đến tình trạng tắc nghẽn do khả năng tài nguyên trên các thiết bị mạng đáp ứng không nổi. Trong khi đó, có một số thiết bị mạng nào đó có ít dữ liệu truyền qua lại không được tận dụng. Tắc nghẽn có thể xuất hiện ở nhiều vị trí khác nhau trong mạng và đó là kết quả của một số nguyên nhân sau:

Thời gian chờ xử lý, các gói tin xếp hàng trong hàng đợi quá lớn. Nếu các luồng gói tin đột ngột đi vào từ nhiều Interface và tất cả đều muốn đi ra cùng một đường nên hàng đợi sẽ bị đầy (do phải lưu gói tin và chuyển tiếp gói tin ...). Nếu khả năng xử lý của các nút mạng yếu sẽ dẫn đến tắc nghẽn. Kích thước bộ đệm của hàng đợi quá nhỏ. Nếu bộ đệm không đủ dung lượng để lưu các luồng gói tin thì một số gói tin sẽ bị mất. Độ trễ lớn, tần suất lỗi mạng cao và sự chênh lệch về băng thông giữa các liên kết: Làm tăng số lượng gói tin tại các interface đầu vào của các router biên trong mạng làm cho khả năng tắc nghẽn trong mạng tăng lên, và cũng đồng nghĩa với việc số lượng gói tin bị loại bỏ cũng tăng lên nếu các Router biên không có cơ chế hành xử hợp lý.

Trong quá trình học tập và nghiên cứu tôi đã tìm hiểu được một số kiến thức về mạng và tìm hiểu các thuật toán để đi đều khiến lưu lượng mạng để đảm bảo dữ liệu trong mạng không bị tắc nghẽn. Chính vì thế tôi đã lựa chọn đề tài “Kỹ thuật điều khiển lưu lượng trong mạng máy tính cụm”.

✓ **Đối tượng và phạm vi nghiên cứu**

Đối tượng:

- Kiến trúc hệ thống máy tính cụm
- Thuật toán gáo rò

Phạm vi nghiên cứu:

- Điều chỉnh lưu lượng trong mạng máy tính cụm bằng thuật toán gáo rò

✓ **Hướng nghiên cứu của đề tài**

- Tìm hiểu thành phần, chức năng của hệ thống máy tính cụm, và các vấn đề nảy sinh trong quá trình khởi động mạng máy tính cụm

- Tìm hiểu thuật toán gáo rò.
- Giải quyết bài toán điều khiển lưu lượng dựa trên thuật toán gáo rò

✓ **Những nội dung nghiên cứu chính:**

Luận văn được trình bày trong 3 chương:

Chương 1: Giới thiệu tổng quan về máy tính cụm và các vấn đề điều khiển lưu lượng trong thực tế

Chương 2: Kỹ thuật điều khiển lưu lượng trong mạng máy tính cụm

Chương 3: Đây là chương demo việc triển khai điều khiển lưu lượng trên máy tính cụm bằng thuật toán gáo rò. Trong chương này sẽ hướng dẫn tạo ra một cluster, dùng thuật toán gáo rò để điều khiển lưu lượng.

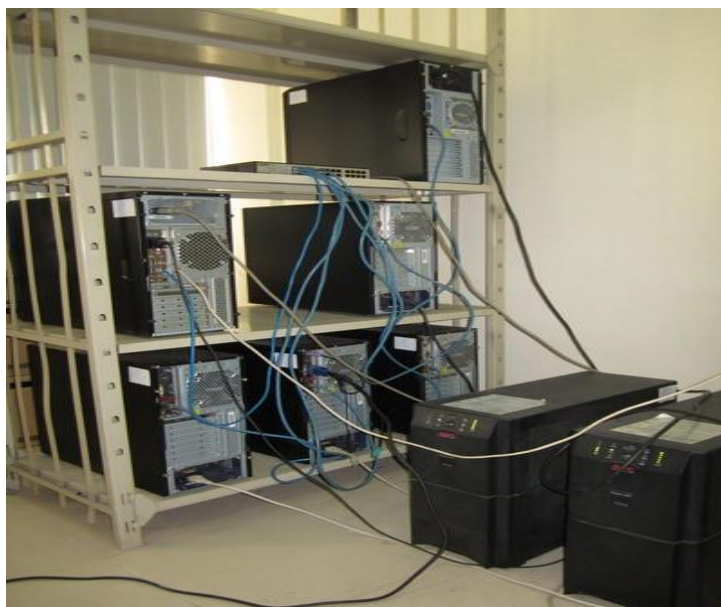
CHƯƠNG I: TỔNG QUAN MÁY TÍNH CỤM VÀ VẤN VỀ ĐIỀU KHIỂN LƯU LƯỢNG

1.1 Tổng quan về hệ thống máy tính cụm

1.1.1 Khái niệm máy tính cụm

Máy tính cụm (Cluster computers) là một cụm gồm nhiều máy tính hoạt động cùng nhau, xử lý các công việc như 1 máy tính logic hay còn gọi là cluster (cụm máy tính). Chúng thường (nhưng không nhất thiết) có chung cấu hình phần cứng và phần mềm. Để đảm bảo việc toàn bộ cụm máy tính (cluster computer) có chung một hệ thống điều hành cũng như các tùy chọn về phần mềm và cấu hình, bạn có thể dùng chung file ảo hóa để khởi động từ đó (ví dụ file .iso). Việc này có thể hoàn thành nhờ sử dụng mạng như một phương thức khởi động, thay vì sử dụng ổ đĩa thông thường. [1]

Các bộ phận của hệ thống máy tính cụm được liên kết với nhau thông qua các loại cáp dùng cho mạng LAN có thể là cáp thường, tốt nhất là cáp quang, để đảm bảo việc truyền dữ liệu được nhanh hơn. Máy tính cụm cùng chạy một hệ điều hành giống nhau. Hệ thống máy tính cụm là 1 giải pháp để nâng cao khả năng của các máy tính không mạnh để chúng trở nên mạnh hơn khi kết hợp với nhau và đặc biệt là tiết kiệm nguồn tài chính.



Hình 1.1 Hệ thống máy tính cụm của trung tâm vật lý lý thuyết,