BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC TÂY BẮC

GIÁO TRÌNH

LẬP TRÌNH MẠNG

NGUYỄN DUY HIẾU
MAI VĂN TÁM

SƠN LA, NĂM 2019

LỜI NÓI ĐẦU

Trong những năm gần đây, lập trình mạng luôn là một trong những nội dung quan trọng trong lĩnh vực công nghệ phần mềm. Nhờ sự phát triển vượt bậc trong lĩnh vực mạng máy tính, những phần mềm máy tính dùng cho doanh nghiệp hiện nay sử dụng rất nhiều trên môi trường mạng đặc biệt là Internet. Mạng máy tính là nơi các kỹ thuật liên quan tới mô hình khách/chủ, mô hình phân tán hay mô hình hợp tác được triển khai. Các ứng dụng mạng xử lý tập trung hoặc phân tán, tận dụng tối đa sức mạnh của các hệ thống phần cứng để mang lại hiệu quả cao.

Giáo trình này phục vụ giảng dạy và học tập học phần *Lập trình mạng* tại Trường Đại học Tây Bắc. Trong nội dung, ngoài những kiến thức cơ bản về mạng máy tính có liên quan, chúng tôi sẽ hướng dẫn cách thức làm việc với kỹ thuật lập trình socket với giao thức TCP, UDP và kỹ thuật lập trình phân tán RMI. Trong từng phần kiến thức, giáo trình cũng cung cấp các ví dụ minh họa. Đây là những ví dụ hết sức cơ bản, giúp bạn đọc hiểu được cách thức mà các ứng dụng mạng hoạt động. Từ những ví dụ này, độc giả có thể tự phát triển để tạo ra các ứng dụng phức tạp hơn, mạnh mẽ hơn.

Chúng tôi lựa chọn ngôn ngữ lập trình Java để trình bày về các kỹ thuật liên quan tới lập trình mạng cũng như các ví dụ minh họa. Việc lựa chọn ngôn ngữ lập trình Java không chỉ bởi Java hiện nay luôn là một trong những ngôn ngữ hàng đầu để phát triển phần mềm mà còn bởi Java là ngôn ngữ vốn sinh ra để giải quyết các vấn đề liên quan tới các ứng dụng mạng.

Giáo trình không tránh khỏi những sơ suất. Chúng tôi mong nhận được các ý kiến đóng góp quý báu của quý thầy cô và các bạn sinh viên để hoàn thiện giáo trình hơn nữa.

Chúng tôi xin chân thành cảm ơn.

NHÓM TÁC GIẢ

MỤC LỤC

DANH MỤC HÌNH ẢNH

CHƯƠNG 1. CÁC KHÁI NIỆM CƠ BẢN VỀ MẠNG MÁY TÍNH

1.1 Mạng máy tính

Một mạng máy tính là một tập hợp các máy tính và các thiết bị. Các máy tính và các thiết bị trên mạng có thể gửi và nhận dữ liệu với các máy tính và các thiết bị khác. Với mạng được kết nối bằng dây dẫn, các bit dữ liệu sẽ được biến đổi thành các sóng điện từ di chuyển dọc theo dây dẫn. Các mạng không dây truyền dữ liệu bằng cách sử dụng sóng vô tuyến. Và với những đường truyền có khoảng cách lớn, dữ liệu được truyền đi bằng cách sử dụng cáp quang. Cáp quang sử dụng các sóng ánh sáng với độ dài bước sóng khác nhau để truyền dữ liệu.

Mỗi một thiết bị trên một mạng được gọi là một *nút mạng* (node). Hầu hết các nút là các máy tính, tuy nhiên các nút cũng có thể là các máy in, router, cầu nối, gateway (thiết bị nối ghép hai mạng cục bộ không cùng họ với nhau, hoặc mạng cục bộ với một mạng diện rộng, với một máy tính mini hay máy tính lớn…), các thiết bị cuối câm (thiết bị cuối không có bộ xử lý trung tâm và các ổ đĩa - dumb terminal). Ta sẽ dùng thuật ngữ *nút* để nói đến bất kỳ thiết bị nào trên mạng và dùng thuật ngữ *host* để nói đến một nút là một máy tính đa năng. Mỗi một nút mạng có một địa chỉ. Một địa chỉ là một chuỗi các byte xác định duy nhất một nút.

Địa chỉ được gán cho từng nút mạng sẽ khác nhau đối với các mạng khác nhau. Các địa chỉ Ethernet được gắn vào phần cứng vật lý Ethernet. Các nhà sản xuất phần cứng Ethernet sử dụng mã nhà sản xuất được chỉ định trước để đảm bảo không có xung đột giữa địa chỉ trong phần cứng của họ và địa chỉ của phần cứng của nhà sản xuất khác. Mỗi nhà sản xuất chịu trách nhiệm đảm bảo rằng không có bất kỳ hai card mạng Ethernet nào có cùng địa chỉ. Các địa chỉ Internet thường được gán cho một máy tính bởi Nhà cung cấp dịch vụ Internet (Internet Service Provider - ISP).

Các mạng máy tính hiện đại là các mạng chuyển mạch gói (packet-switched): dữ liệu truyền trên mạng được chia nhỏ thành các đoạn được gọi là các gói (packet). Mỗi một gói chứa thông tin về bên gửi và bên nhận. Một số ưu điểm của việc chia nhỏ dữ liệu thành các gói được gán địa chỉ là:

+ Các gói tin từ nhiều nguồn trao đổi thông tin khác nhau có thể được truyền đi trên cùng một đường dây và sẽ làm giảm giá thành cho việc xây dựng các mạng. Các máy tính có thể chia sẻ chung một đường truyền mà không sợ bị can nhiễu lẫn nhau.

+ Các mã kiểm tra tổng (checksum) có thể được sử dụng để phát hiện gói tin có bị hư hại trong quá trình truyền tin hay không.

Để các mạng máy tính hoạt động được ta cần phải có những quy tắc hoạt động, đó chính là *các giao thức* (protocol). Một giao thức là một tập hợp chính xác các luật định nghĩa cách thức các máy tính giao tiếp với nhau: khuôn dạng của các địa chỉ, cách chia dữ liệu được thành các gói… Ví dụ:

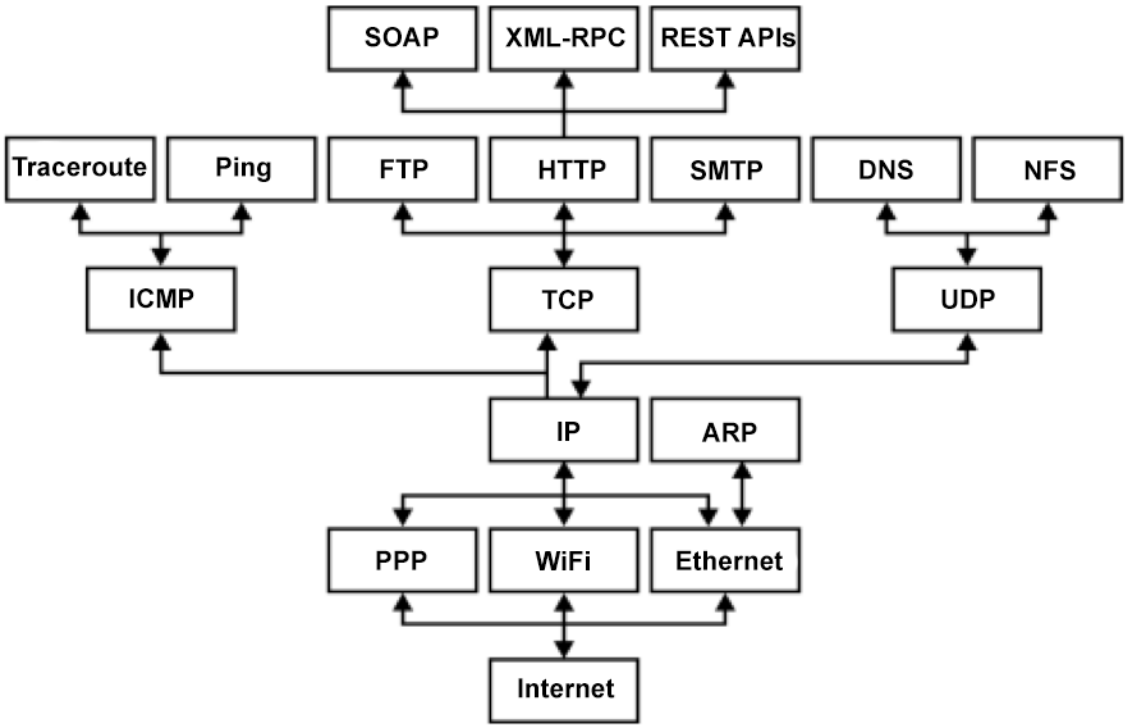
+ Giao thức HTTP (Hypertext Transfer Protocol) quy định cách thức trao đổi thông tin giữa các trình duyệt web (web browsers) với máy chủ dịch vụ web (webserver).

+ Chuẩn IEEE 802.3 quy định cách mã hóa các bit thành các tín hiệu điện trên từng loại dây dẫn cụ thể.

Các chuẩn giao thức mở, đã được công bố cho phép phần mềm và thiết bị từ các nhà cung cấp khác nhau giao tiếp với nhau. Ví dụ một máy chủ web không cần phải quan tâm người dùng sẽ sử dụng hệ điều hành nào trên thiết bị của họ, chẳng hạnh như Windows, Unix, Android, iOS… vì tất cả các thiết bị này đều sử dụng chung giao thức HTTP.

1.2. Các lớp của một mạng

Truyền dữ liệu trên một mạng là một quá trình phức tạp. Để người phát triển ứng dụng và để người sử dụng không cần nhìn thấy sự phức tạp của quá trình truyền dữ liệu, các thành phần khác nhau của một mạng truyền thông được chia thành nhiều lớp. Mỗi một lớp biểu diễn một mức khác nhau của việc trừu tượng hóa giữa phần cứng vật lý (dây dẫn, điện…) và thông tin được truyền.

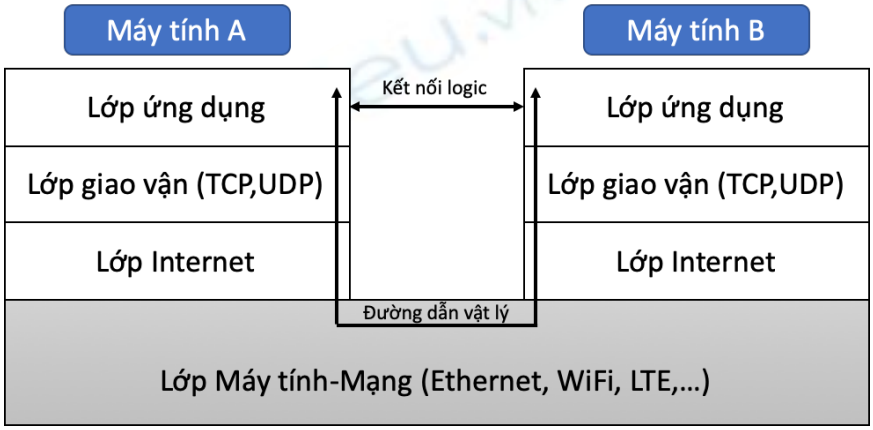


Hình 1.1: Các giao thức trong các lớp khác nhau của một mạng

Về mặt lý thuyết, mỗi một lớp chỉ trao đổi với các lớp kề ngay trên và kề ngay dưới. Việc tách một mạng thành các lớp cho phép ta sửa đổi thậm chí là thay thế phần mềm trong một lớp mà không ảnh hưởng đến các lớp khác miễn sao cho các giao diện giữa các lớp không thay đổi.

Sơ đồ trong *Hình 1.1* trình bày một mô hình phân tầng của các giao thức có thể có trong một mạng. Trong khi các giao thức các lớp ở giữa tương đối ổn định trong mạng Internet thì các giao thức trên đỉnh và dưới đáy thay đổi rất nhiều.

Tồn tại nhiều mô hình chia lớp khác nhau như mô hình OSI, mô hình TCP/IP. Mỗi mô hình phù hợp với các yêu cầu của một kiểu mạng cụ thể. Trong tài liệu này ta sẽ sử dụng mô hình TCP/IP, đây là mô hình chuẩn bốn lớp phù hợp với mạng Internet. Sơ đồ sau sẽ trình bày cấu trúc của mô hình TCP/IP.



Hình 1.2: Các lớp trong mô hình TCP/IP

Các lớp trong mô hình TCP/IP bao gồm:

+ Lớp ứng dụng (Application Layer).

+ Lớp giao vận (Transport Layer) sử dụng các giao thức TCP hoặc UDP.

+ Lớp Internet (Internet Layer) sử dụng giao thức IP.

+ Lớp máy tính-mạng (Host-To-Network Layer) sử dụng các giao thức Ethernet, WiFi, LTE…

Một ứng dụng được chạy trong lớp ứng dụng và chỉ trao đổi thông tin với lớp giao vận. Lớp giao vận chỉ trao đổi thông tin với lớp ứng dụng và lớp Internet. Lớp Internet, đến lượt mình, chỉ trao đổi thông tin với lớp giao vận và lớp máy tính-mạng và không trao đổi trực tiếp với lớp ứng dụng. Lớp máy tính-mạng chuyển dữ liệu thông qua các dây dẫn, cáp quang hoặc các đường truyền vật lý khác tới lớp máy tính-mạng trên hệ thống từ xa khác trong mạng. Lớp máy tính-mạng trên hệ thống từ xa sau đó sẽ chuyển dữ liệu lên các lớp bên trên tới lớp ứng dụng trong hệ thống này. Ta sẽ tìm hiểu chức năng của mỗi lớp ở các phần dưới đây.

1.2.1 Lớp máy tính-mạng

Lớp máy tính-mạng (lớp vật lý) quy định cách thức một giao diện mạng cụ thể. Chẳng hạn như một card Ethernet hoặc một ăng-ten WiFi gửi các IP datagram qua kết nối vật lý của nó tới mạng cục bộ và tới mạng diện rộng.

Lớp máy tính-mạng được xây dựng bởi phần cứng kết nối các máy tính với nhau (dây dẫn, cáp quang, sóng vô tuyến) và đôi khi còn được gọi là lớp vật lý của một mạng. Các lập trình viên sử dụng ngôn ngữ Java không cần phải quan tâm tới lớp này trừ khi có các vấn đề về kỹ thuật cần phải khắc phục. Các lập trình viên chỉ cần quan tâm đến hiệu suất của mạng, ví dụ khách hành của chúng ta sử dụng một mạng cáp quang tốc độ cao ta sẽ cần phải thiết kế một giao thức và các ứng dụng phù hợp với tốc đô của mạng bên phía khách hàng.

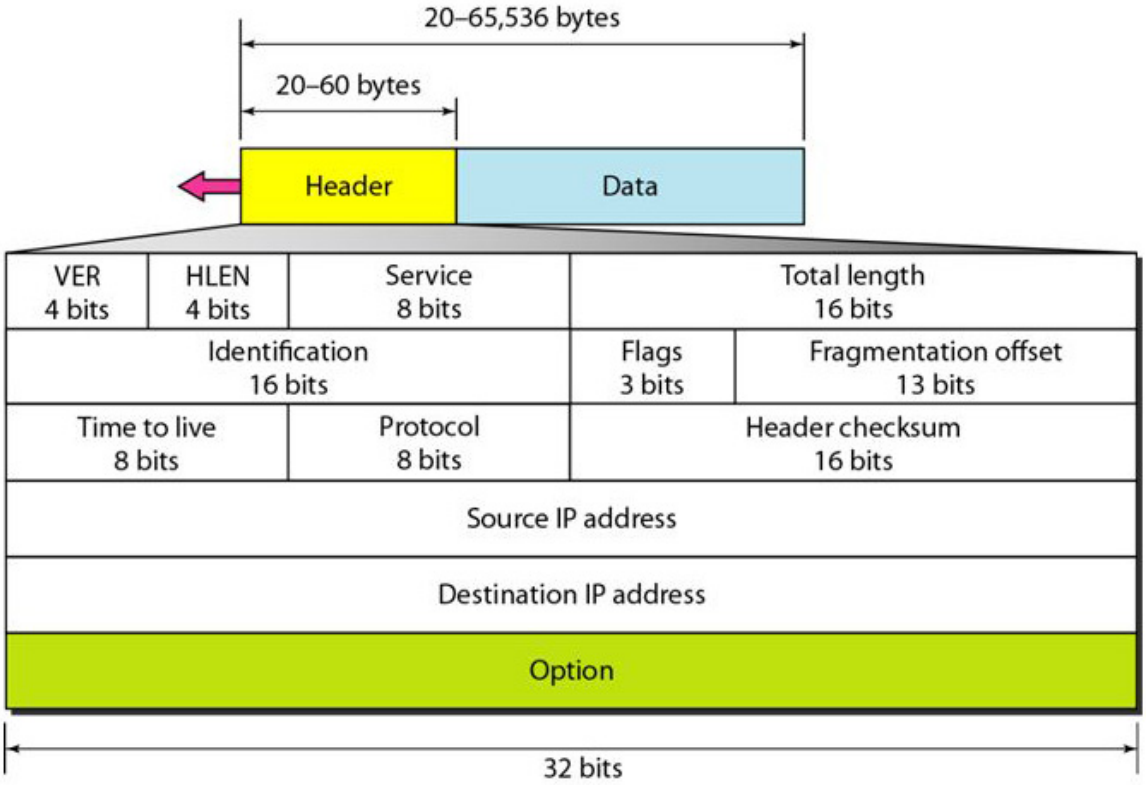
1.2.2 Lớp Internet

Lớp tiếp theo trong một mạng và là lớp đầu tiên ta cần phải quan tâm đến đó là *lớp Internet*. Trong mô hình OSI, lớp Internet có tên gọi là lớp mạng. Một giao thức cho lớp mạng quy định cách các bit và các byte dữ liệu được tổ chức thành một nhóm lớn hơn được gọi là các packet và mô hình đánh địa chỉ để các máy tính có thể tìm thấy nhau trong mạng. *Giao thức Internet* (Internet Protocol - IP) là giao thức được sử dụng rộng rãi nhất trên thế giới và Java hiểu được các giao thức của lớp này.

Giao thức IP có hai phiên bản là IPv4 và IPv6. IPv4 sử dụng các địa chỉ 32 bit và IPv6 sử dụng các địa chỉ 128 bit. IPv6 bổ sung thêm một số tính năng kỹ thuật để trợ giúp quá trình định tuyến. Hiện nay IPv6 đã được sử dụng rộng rãi và có khả năng sẽ vượt qua IPv4 về số lượng người dùng. Cần phải có các gateway đặc biệt và các giao thức đường ống (tunnel) để hai giao thức IPv4 và IPv6 có thể cùng hoạt động trên một mạng.

Cả hai giao thức IPv4 và IPv6 gửi dữ liệu qua lớp Internet trong các packet được gọi là *datagram*. Mỗi một datagram trong IPv4 chứa một phần đầu (header) có độ dài từ 20 byte đến 60 byte và một khối lượng dữ liệu (payload) lên đến 65.515 byte. Trong thực tế thì payload trong IPv4 nhỏ hơn nhiều, từ khoảng vài chục byte đến khoảng 8 kilobyte. Một datagram của IPv6 chứa một header lớn hơn và payload có thể lên đến 4 gigabyte.

Sơ đồ trong *Hình 1.3* trình bày cấu trúc của một IPv4 datagram. Trong sơ đồ này tất cả các bit và các byte được biểu diễn dưới dạng big-endian: MSB đến LSB tính từ trái qua phải.



Hình 1.3: Cấu trúc của một IPv4 datagram

Bên cạnh chức năng định tuyến và đánh địa chỉ, mục đích thứ hai của lớp Internet là cho phép các lớp máy tính đến mạng với nhiều kiểu khác nhau có thể trao đổi được với nhau. Các router Internet chuyển đổi các giao thức giữa WiFi và Ethernet, Ethernet và DSL, DSL và cáp quang. Nếu không có lớp Internet, mỗi máy tính chỉ có thể trao đổi với các máy tính khác trên cùng một kiểu mạng. Lớp Internet có nhiệm vụ kết nối các mạng không đồng nhất sử dụng các giao thức không đồng nhất với nhau.

1.2.3 Lớp giao vận

Các datagram chưa được xử lý có nhiều hạn chế. Đáng chú ý nhất là không có một sự bảo đảm nào cho các datagram sẽ được chuyển giao tới bên nhận, thậm chí nếu được chuyển đi thì các datagram cũng có thể bị hư hỏng trong quá trình truyền. Phần kiểm tra tổng của phần header (header checksum) cũng chỉ có thể phát hiện những hư hỏng trong phần header và không thể phát hiện những hư hỏng trong phần dữ liệu của một datagram. Cuối cùng là các datagram có thể không bị hư hỏng trong quá trình gửi thì cũng có thể sẽ đến đích không theo đúng thứ tự mà các datagram đã được gửi đi vì các datagram có thể đi theo những tuyến đường khác nhau từ nguồn đến đích. Ví dụ: datagram A được gửi trước datagram B nhưng không có nghĩa là datagram A sẽ đến đích trước datagram B.

Lớp giao vận có nhiệm vụ để đảm bảo các gói tin được nhận theo đúng thứ tự đã được gửi đi và dữ liệu không bị mất hay bị hư hỏng. Nếu một gói tin bị mất, lớp giao vận có thể yêu cầu bên gửi gửi lại gói tin đó. Các mạng IP cài đặt cơ chế này bằng cách bổ sung thêm một phần header vào mỗi datagram. Có hai giao thức quan trọng nhất tại lớp giao vận đó là giao thức TCP (Transmission Control Protocol) và giao thức UDP (User Datagram Protocol).

Giao thức TCP là giao thức cho phép truyền lại các dữ liệu bị mất hay bị hư hỏng và chuyển giao các byte theo đúng thứ tự đã được gửi đi. TCP được gọi là giao thức đáng tin cậy. Giao thức UDP cho phép bên nhận phát hiện các gói tin bị hư hỏng nhưng không đảm bảo các gói tin được chuyển giao theo đúng thứ tự. UDP thường nhanh hơn rất nhiều so với TCP. Giao thức UDP là giao thức không tin cậy, tuy nhiên UDP vẫn được sử dụng nhiều trong các mạng Internet.

1.2.4 Lớp ứng dụng

Lớp ứng dụng có trách nhiệm chuyển dữ liệu đến người sử dụng. Ba lớp bên dưới làm việc cùng với nhau để quy định cách dữ liệu được truyền từ một máy tính đến một máy tính khác. Lớp ứng dụng quyết định cần phải làm gì với dữ liệu sau khi dữ liệu đã được truyền tới lớp này. Tồn tại nhiều giao thức khác nhau trong lớp ứng dụng, ví dụ: các giao thức HTTP, HTTPS cho World Wide Web; các giao thức SMTP, POP, IMAP cho email; các giao thức FTP, FSP, TFTP cho truyền file; giao thức NFS cho truy nhập file; các giao thức Gnutella and BitTorrent cho chia sẻ file; các giao thức Session Initiation Protocol (SIP) and Skype cho truyền tiếng nói…

1.3 Giao thức IP, TCP và UDP

1.3.1 Khái quát về giao thức IP, TCP và UDP

Giao thức Internet - IP được phát triển trong thời chiến tranh lạnh giữa Liên Xô và Mỹ và được đỡ đầu bởi quân đội Mỹ. IP cần phải đáp ứng được các yêu cầu:

Đầu tiên, giao thức IP phải đủ mạnh để toàn bộ hệ thống mạng không bị ngừng hoạt động nếu Liên Xô tấn công hạt nhân vào một vị trí nào đó, chẳng hạn ở Cleveland. Tất cả các thông báo vẫn được chuyển đến đích (ngoại trừ đến Cleveland). Do đó, giao thức IP được thiết kế để cho phép có nhiều tuyến đường giữa hai điểm bất kỳ và để định tuyến các gói tin không đi qua các router đã bị hỏng.

Thứ hai, quân đội Mỹ có nhiều loại máy tính khác nhau và tất cả các máy tính này phải trao đổi được với nhau. Do đó, giao thức IP cần phải là một giao thức mở và không phụ thuộc vào hệ điều hành. Do có nhiều tuyến đường giữa hai điểm bất kỳ và đường đi nhanh nhất giữa hai điểm có thể thay đổi liên tục nên các gói tin của

một luồng dữ liệu có thể không đi cùng một tuyến đường đến đích và có thể đến đích không theo đúng thứ tự đã được gửi.

Để cải thiện mục tiêu trên, giao thức TCP đã được đặt bên trên IP để cho phép điểm cuối của một kết nối có khả năng phản hồi các gói tin đã nhận được và yêu cầu gửi lại các gói tin bị mất hoặc bị hư hỏng. TCP còn cho phép các gói tin được sắp xếp lại theo đúng thứ tự đã được gửi đi. Tuy nhiên, TCP lại chứa một phần không nhỏ thông tin phụ được gọi là overhead, phần này sẽ ảnh hưởng đến tốc độ truyền các gói tin. Do đó, nếu thứ tự của các gói tin không thật sự quan trọng và việc một vài gói tin bị hư hỏng hay bị mất không ảnh hưởng đến luồng dữ liệu thì các gói tin đôi khi được gửi đi mà không cần được đảm bảo và TCP cung cấp việc sử dụng giao thức UDP.

Như đã trình bày ở phần *Lớp giao vận*, UDP là giao thức không tin cậy nên giao thức này có thể ảnh hưởng đến việc truyền file. Tuy nhiên giao thức này lại thích hợp cho các ứng dụng khi việc mất một phần dữ liệu sẽ không bị người dùng cuối phát hiện, ví dụ việc mất một vài bit từ một tín hiệu video hay audio sẽ không làm giảm đến chất lượng của dòng bit dữ liệu. Các mã sửa lỗi có thể được đưa vào trong các dòng dữ liệu sử dụng UDP tại mức ứng dụng để xử lý các thông tin bị mất.

Một số lượng lớn các giao thức có thể chạy trên đỉnh của giao thức IP, phổ biến nhất là giao thức Internet Control Message Protocol - ICMP. ICMP sử dụng các datagram chưa được xử lý để chuyển tiếp các thông báo lỗi giữa các máy tính. Ping là chương trình được sử dụng phổ biến nhất trong giao thức ICMP. Java chỉ hỗ trợ TCP, UDP và các giao thức trong lớp ứng dụng, Java không hỗ trợ ICMP. Tất cả các giao thức khác của các lớp giao vận, lớp Internet và các lớp bên dưới chỉ có thể cài đặt trong Java bằng cách triển khai các mã gốc (native code).

1.3.2 Địa chỉ IP và tên miền

Mỗi máy tính trên một mạng IPv4 được xác định bởi một địa chỉ IP duy nhất. Địa chỉ IP này gồm 4 byte, thường được viết dưới dạng bốn chữ số thập phân ngăn cách nhau bởi dấu chấm, ví dụ 192.168.1.1. Mỗi một chữ số thập phân được biểu diễn bằng một byte, do đó phạm vi của mỗi số thập phân sẽ từ 0 đến 255. Khi các gói tin được truyền đi trên mạng, địa chỉ của máy gửi và địa chỉ của máy nhận sẽ được chứa trong phần header của gói tin. Các router dọc theo đường truyền sẽ dựa vào địa chỉ IP của máy nhận để lựa chọn tuyến đường tốt nhất gửi các gói tin đến đích. Máy nhận sẽ sử dụng địa chỉ IP của máy gửi để biết ai đã gửi các gói tin đến mình.

Có khoảng hơn bốn tỷ địa chỉ IPv4, tuy nhiên các địa chỉ này không được phân phối phù hợp giữa các vùng trên thế giới. Tháng 4 năm 2011, châu Á và Australia

sử dụng hết các địa chỉ IP được phân phối. Tháng 9 năm 2012, châu Âu cũng sử dụng hết các địa chỉ IP được phân phối. Bắc Mỹ, Mỹ latinh và châu Phi vẫn còn một số khối địa chỉ IP chưa sử dụng, tuy nhiên các địa chỉ này cũng sẽ nhanh chóng được sử dụng hết.

Có một sự dịch chuyển từ sử dụng IPv4 sang IPv6. IPv6 là các địa chỉ có độ dài 16 byte, đủ để cung cấp địa chỉ IP cho toàn bộ con người, toàn bộ máy tính và trong thực tế toàn bộ các thiết bị trên thế giới. Cách biểu diễn địa chỉ IPv6:

+ Thông thường Các địa chỉ IPv6 được viết thành tám khối, mỗi khối gồm bốn chữ số viết trong hệ thập lục phân, các khối được ngăn cách nhau bởi dấu **:**, ví dụ FEDC:BA98:7654:3210:FEDC:BA98:7654:3210.

+ Các chữ số không (0) ở đầu không cần phải viết. Nếu trong một địa chỉ IPv6 có nhiều khối gồm toàn bộ số 0 thì những khối này có thể được thay thế bằng cặp dấu ::, ví dụ, địa chỉ FEDC:0000:0000:0000:00DC:0000:7076:0010 có thể được viết ngắn gọn thành FEDC::DC:0:7076:10. Mỗi cặp :: chỉ được phép xuất hiện nhiều nhất một lần trong một địa chỉ IPv6

+ Trong các mạng kết hợp cả IPv4 và IPv6, bốn byte cuối cùng của một địa chỉ IPv6 đôi khi được viết như là một địa chỉ IPv4, ví dụ địa chỉ FEDC:BA98:7654:3210:FEDC:BA98:7654:3210 có thể được viết thành FEDC:BA98:7654:3210:FEDC:BA98:118.84.50.16.

Rất khó để con người nhớ được các địa chỉ IP viết dưới dạng các con số, đặc biệt là với địa chỉ IPv6. Do đó, hệ thống tên miền *Domain Name System* (DNS) đã được phát triển để chuyển các địa chỉ IP thành các tên máy (hostname) cho dễ nhớ, ví dụ địa chỉ 208.201.239.101 được chuyển thành *www.oreilly.com*. Khi các chương trình viết bằng Java truy nhập mạng, các chương trình này cần phải xử lý cả các địa chỉ viết bằng các số và cả các hostname tương ứng. Các phương thức để thực hiện công việc này được cung cấp bởi lớp `java.net.InetAddress`.

Một vài máy tính, đặc biệt là các máy chủ, có địa chỉ cố định. Các máy tính khác, đặc biệt là các máy khách trên các mạng cục bộ và các kết nối không dây sẽ nhận các địa chỉ khác nhau mỗi khi các máy tính được khởi động và kết nối vào mạng. Các địa chỉ này được cung cấp bởi DHCP Server. DHCP là viết tắt của Dynamic Host Configuration Protocol - Giao thức cấu hình host động.

Trong địa chỉ IPv4, một vài khối địa chỉ và khuân mẫu có dạng đặc biệt. Tất cả các khối địa chỉ bắt đầu bằng 10., bằng 172.16. đến 172.31. và bằng 192.168. là chưa được cấp cho bất kỳ máy tính nào trên mạng. Các khối địa chỉ này được gọi là các khối địa chỉ không thể định tuyến được (non-routable) và có thể được sử dụng trên

các mạng cục bộ. Không một máy tính nào sử dụng địa chỉ trong các khối trên có thể truy nhập được vào Internet. Các địa chỉ không thể định tuyến được rất thích hợp cho việc xây dựng các mạng riêng, các mạng này sẽ không xuất hiện trên mạng Internet. Các địa chỉ IPv4 bắt đầu bằng 127 luôn có nghĩa là các địa chỉ loopback cục bộ (local loopback address), nghĩa là các địa chỉ này luôn chỉ tới máy tính cục bộ. Hostname cho địa chỉ này thường là localhost. Trong IPv6, địa chỉ 0:0:0:0:0:0:0:1 (cũng được viết tắt là ::1) là địa chỉ loopback. Địa chỉ 0.0.0.0 luôn được chỉ đến máy gửi nhưng cũng có thể được dùng như là địa chỉ nguồn mà không phải địa chỉ đích. Tương tự, trong IPv4, bất kỳ địa chỉ nào bắt đầu bằng 0. (tám bit không) là chỉ đến một máy tính trong cùng một mạng cục bộ.

Trong IPv4 địa chỉ 255.255.255.255 được sử dụng làm địa chỉ quảng bá (broadcast address). Các gói tin được gửi đến địa chỉ này sẽ được tất cả các máy trong mạng cục bộ nhận và sẽ không được gửi ra bên ngoài mạng cục bộ. Địa chỉ quảng bá thường được sử dụng để phát hiện các máy trong cùng một mạng cục bộ, ví dụ: Khi một máy tính trong mạng cục bộ được khởi động, máy tính này sẽ gửi một thông báo đặc biệt đến địa chỉ 255.255.255.255 để tìm DHCP Server. Tất cả các máy tính trong mạng sẽ nhận được thông báo này nhưng chỉ có DHCP Server sẽ trả lời cho máy gửi. Thực tế, DHCP Server gửi sẽ các thông tin về cấu hình mạng cục bộ, bao gồm cả địa chỉ IP mà máy tính vừa khởi động sẽ sử dụng trong phiên làm việc và địa chỉ của DNS Server để máy tính này có thể sử dụng để phân giải tên miền.

1.3.3 Các cổng

Các máy tính hiện đại thực hiện nhiều việc khác nhau cùng một lúc. Email cần phải được tách ra khỏi các yêu cầu FTP, các yêu cầu về FTP cần được tách biệt với lưu lượng truy nhập web. Điều này được thực hiện thông qua cổng (port). Mỗi máy tính với một địa chỉ IP có hàng nghìn cổng logic, chính xác là 65.535 cổng trên giao thức tầng giao vận. Các cổng này hoàn toàn là trừu tượng trong bộ nhớ của máy tính và không phải là các cổng vật lý như cổng USB. Mỗi một cổng được xác định bằng một con số nằm giữa 1 và 65.535. Mỗi một cổng có thể được cấp cho một dịch vụ cụ thể. Ví dụ, giao thức HTTP thường sử dụng cổng 80, chúng ta thường nói máy chủ web lắng nghe (listen) các kết nối đang đến trên cổng 80. Khi một máy tính với một địa chỉ IP cụ thể gửi dữ liệu (thông thường là các yêu cầu) đến máy chủ web thì máy tính cũng sẽ gửi dữ liệu tới cổng 80 trên máy này. Máy chủ web sẽ kiểm tra mỗi gói tin nhận được để phát hiện ra số hiệu của cổng bên máy gửi và sau đó sẽ gửi dữ liệu đến bất kỳ chương trình nào đang nghe trên cổng đó. Bảng sau sẽ liệt kê danh sách các cổng thường được sử dụng trong các ứng dụng.

Giao thức	Cổng	Giao thức	Mục đích
echo	7	TCP/UDP	Giao thức kiểm tra dùng để xác minh hai máy tính có thể kết nối với nhau bằng cách: một máy tính sẽ gửi lại thông tin được gửi từ máy tính phía bên kia.
daytime	13	TCP/UDP	Cung cấp biểu diễn ASCII về thời gian hiện tại trên Server.
ftp data	20	TCP	Được sử dụng để truyền các file.
ftp data	21	TCP	Được sử dụng để gửi các lệnh của FTP như *put* và *get.*
ssh	22	TCP	Được sử dụng để mã hóa đăng nhập an toàn từ xa
telnet	23	TCP	Được sử dụng cho tương tác, các phiên làm việc từ xa.
smtp	25	TCP	Simple Mail Transfer Protocol - Giao thức gửi thư giữa hai máy.
whois	43	TCP	Một dịch vụ về thư mục đơn giản cho các quản trị mạng Internet.
finger	79	TCP	Một dịch vụ trả về thông tin người dùng hay một nhóm người dùng trên hệ thống cục bộ.
http	80	TCP	Giao thức truyền siêu văn bản.
pop3	110	TCP	Port Office Protocol Version 3 - Là giao thức truyền các email được tích lũy cho một khách hàng thỉnh thoảng mới truy nhập mạng .

1.4 Mạng Internet

1.4.1 Các khối địa chỉ Internet

Mỗi nhà cung cấp dịch vụ Internet (ISP) được cấp một số khối địa chỉ IPv4. Khi một tổ chức muốn thiết lập một mạng máy tính kết nối với Internet, ISP sẽ cấp cho tổ chức này một khối địa chỉ, mỗi một khối sẽ được cố định phần đầu của địa chỉ, ví dụ nếu phần đầu của địa chỉ là 216.245.85 thì tổ chức này có thể sử dụng các địa chỉ từ 216.245.85.0 đến 216.245.85.255 (lưu ý các địa chỉ 216.245.85.0 và 216.245.85.255 không được dùng để gán địa chỉ cho các máy tính). Vì phần cố định gồm 24 bit đầu tiên nên sẽ đực ký hiệu là /24. Nếu một khối có phần cố định là /23 thì sẽ còn 9 bit để gán địa chỉ cho các máy tính, do đó sẽ có 2^9-2 = 510 địa chỉ được gán cho các máy tính. Nếu một mạng có có phần cố định là /30 thì chỉ còn 2 bit để gán địa chỉ cho các máy tính và như vậy sẽ chỉ có tối đa 2 địa chỉ được dùng để gán cho các máy tính trong mạng.