**Tổng quan về Internet và các khái niệm cơ bản**

* **Internet là gì?** Internet là một **mạng lưới của các mạng lưới**, kết nối hàng tỷ thiết bị trên toàn cầu. Ban đầu được phát triển bởi Bộ Quốc phòng Hoa Kỳ vào cuối những năm 1960, internet đã phát triển thành một hệ thống phức tạp, được sử dụng để truy cập thông tin, giao tiếp và kinh doanh. Để hiểu internet, trước tiên cần hiểu khái niệm **mạng** (network), đó là một nhóm các máy tính hoặc thiết bị được kết nối với nhau.
* **Cách internet hoạt động:** Ở cấp độ cao, internet hoạt động bằng cách kết nối các thiết bị và hệ thống máy tính với nhau thông qua một tập hợp các giao thức chuẩn. Dữ liệu được chia thành các **gói (packet)** nhỏ, được gửi qua các **bộ định tuyến (router)**, là các thiết bị chịu trách nhiệm định tuyến lưu lượng truy cập giữa các mạng.
* **Các khái niệm và thuật ngữ cơ bản:**
  + **Packet (Gói tin):** Một đơn vị dữ liệu nhỏ được truyền qua internet.
  + **Router (Bộ định tuyến):** Thiết bị định hướng các gói dữ liệu giữa các mạng khác nhau.
  + **IP Address (Địa chỉ IP):** Một định danh duy nhất được gán cho mỗi thiết bị trên mạng, dùng để định tuyến dữ liệu đến đúng đích.
  + **Domain Name (Tên miền):** Một tên dễ đọc do con người sử dụng để xác định một trang web, ví dụ: google.com.
  + **DNS (Hệ thống tên miền):** Hệ thống chịu trách nhiệm chuyển đổi tên miền thành địa chỉ IP.
  + **HTTP (Giao thức truyền siêu văn bản):** Giao thức được sử dụng để truyền dữ liệu giữa máy khách (ví dụ: trình duyệt web) và máy chủ (ví dụ: trang web).
  + **HTTPS (HTTP an toàn):** Một phiên bản mã hóa của HTTP, cung cấp giao tiếp an toàn giữa máy khách và máy chủ.
  + **SSL/TLS:** Các giao thức Secure Sockets Layer và Transport Layer Security được sử dụng để cung cấp giao tiếp an toàn trên internet.

**Vai trò của các giao thức và Địa chỉ IP/Tên miền**

* **Vai trò của các giao thức (protocols):** Các giao thức đóng vai trò quan trọng trong việc cho phép giao tiếp và trao đổi dữ liệu trên internet. Một giao thức là một tập hợp các quy tắc và tiêu chuẩn xác định cách thông tin được trao đổi giữa các thiết bị và hệ thống. Các giao thức phổ biến bao gồm **IP, TCP, UDP, DNS, HTTP**.
  + **IP (Internet Protocol):** Chịu trách nhiệm định tuyến các gói dữ liệu đến đúng đích.
  + **TCP (Transmission Control Protocol):** Đảm bảo các gói tin được truyền một cách đáng tin cậy và theo đúng thứ tự.
  + **UDP (User Datagram Protocol):** Một giao thức truyền tải dữ liệu không kết nối, thường dùng cho các ứng dụng thời gian thực như video call.
  + **DNS:** Được sử dụng để chuyển đổi tên miền thành địa chỉ IP.
  + **HTTP:** Được sử dụng để truyền dữ liệu giữa máy khách và máy chủ.
* **Địa chỉ IP và Tên miền:**
  + **Địa chỉ IP:** Là một định danh duy nhất cho mỗi thiết bị trên mạng, thường được biểu diễn bằng một chuỗi bốn số, phân tách bằng dấu chấm, ví dụ: "192.168.1.1".
  + **Tên miền:** Là các tên dễ đọc dùng để xác định các trang web và các tài nguyên internet khác, ví dụ: "google.com". Tên miền được dịch thành địa chỉ IP thông qua **DNS**.
  + **Hệ thống DNS:** Khi bạn nhập một tên miền vào trình duyệt web, máy tính của bạn gửi một truy vấn DNS đến một máy chủ DNS, máy chủ này sẽ trả lại địa chỉ IP tương ứng.

**HTTP, HTTPS và bảo mật kết nối internet**

* **HTTP và HTTPS:**
  + **HTTP (Hypertext Transfer Protocol):** Là giao thức được sử dụng để truyền dữ liệu giữa máy khách và máy chủ.
  + **HTTPS (HTTP Secure):** Là một phiên bản an toàn hơn của HTTP, mã hóa dữ liệu được truyền giữa máy khách và máy chủ bằng mã hóa SSL/TLS.
* **Bảo mật kết nối internet với SSL/TLS:**
  + **SSL/TLS:** Là các giao thức dùng để mã hóa dữ liệu được truyền qua internet, thường được sử dụng để cung cấp kết nối an toàn cho trình duyệt web, ứng dụng email và các chương trình truyền tệp.
  + **Chứng chỉ (Certificates):** Chứng chỉ SSL/TLS được sử dụng để thiết lập lòng tin giữa máy khách và máy chủ, chứa thông tin về danh tính của máy chủ và được ký bởi một bên thứ ba đáng tin cậy (Certificate Authority) để xác minh tính xác thực.
  + **Bắt tay (Handshake):** Trong quá trình bắt tay SSL/TLS, máy khách và máy chủ trao đổi thông tin để đàm phán thuật toán mã hóa và các tham số khác cho kết nối an toàn.
  + **Mã hóa (Encryption):** Sau khi kết nối an toàn được thiết lập, dữ liệu được mã hóa bằng thuật toán đã thỏa thuận và có thể được truyền một cách an toàn giữa máy khách và máy chủ.

**DNS Lookup và các thành phần liên quan**

* **DNS Lookup:** Quá trình DNS lookup (tra cứu DNS) là việc chuyển đổi một hostname (như www.example.com) thành một địa chỉ IP (ví dụ: 192.168.1.1). Quá trình này diễn ra "ở hậu trường" và không yêu cầu tương tác từ người dùng. Có 4 máy chủ DNS liên quan đến việc tải một trang web:
  + **DNS Recursor:** Có thể được coi như một người thủ thư, nhận truy vấn từ máy khách và chịu trách nhiệm đưa ra các yêu cầu để đáp ứng truy vấn DNS của khách.
  + **Root nameserver:** Là bước đầu tiên trong việc chuyển đổi tên máy chủ thành địa chỉ IP, tương tự như một chỉ mục trong thư viện.
  + **TLD nameserver (Top Level Domain):** Là bước tiếp theo trong tìm kiếm địa chỉ IP, lưu trữ phần cuối cùng của tên máy chủ (ví dụ: ".com").
  + **Authoritative nameserver:** Là bước cuối cùng, chứa bản ghi DNS chính thức và trả về địa chỉ IP cho tên máy chủ được yêu cầu.
* **Phân biệt giữa Authoritative DNS Server và Recursive DNS Resolver:**
  + **Recursive DNS resolver:** Nằm ở đầu quá trình truy vấn DNS, có nhiệm vụ theo dõi bản ghi DNS cho đến khi tìm thấy authoritative DNS nameserver.
  + **Authoritative DNS server:** Nằm ở cuối quá trình, thực sự chứa và chịu trách nhiệm về các bản ghi DNS.
* **Các bước trong DNS Lookup (khi không có cache):**

Người dùng nhập tên miền vào trình duyệt web và yêu cầu được gửi đến DNS recursive resolver.

Resolver truy vấn DNS root nameserver.

Root server trả lời với địa chỉ của TLD DNS server.

Resolver gửi yêu cầu đến TLD server.

TLD server trả lời với địa chỉ IP của nameserver của miền.

Resolver gửi truy vấn đến nameserver của miền.

Nameserver trả lời địa chỉ IP của miền cho resolver.

Resolver trả lời địa chỉ IP cho trình duyệt web.

Trình duyệt gửi yêu cầu HTTP đến địa chỉ IP.

Máy chủ trả lại trang web cho trình duyệt.

**DNS Caching, HTTP Request/Response và Xu hướng tương lai**

* **DNS Caching:** Để cải thiện hiệu suất và độ tin cậy, dữ liệu DNS được lưu trữ tạm thời ở nhiều vị trí, gần với máy khách yêu cầu nhất. Các vị trí lưu trữ DNS cache bao gồm:
  + **Browser DNS caching:** Trình duyệt web lưu trữ bản ghi DNS trong một khoảng thời gian nhất định.
  + **OS level DNS caching:** Hệ điều hành có một stub resolver lưu trữ bản ghi DNS cục bộ. Nếu không tìm thấy, stub resolver gửi yêu cầu đến recursive resolver của ISP.
  + **ISP recursive resolver caching:** Recursive resolver của ISP cũng lưu trữ bản ghi DNS. Nếu không có A record, nó sẽ truy vấn các NS record của authoritative nameserver. Nếu không có NS record, nó sẽ truy vấn TLD servers. Nếu không có record trỏ đến TLD server, nó sẽ truy vấn root server.
* **HTTP Request/Response:**
  + **HTTP Request:** Là cách các nền tảng giao tiếp internet, như trình duyệt web, yêu cầu thông tin cần thiết để tải một trang web. Một HTTP request bao gồm:
    - Phiên bản HTTP.
    - URL.
    - HTTP Method (ví dụ: GET, POST).
    - HTTP request headers.
    - HTTP body (tùy chọn).
  + **HTTP Response:** Là phản hồi của máy chủ cho một HTTP request. Một HTTP response bao gồm:
    - HTTP status code.
    - HTTP response headers.
    - HTTP body (tùy chọn).
    - **Status code:** Được sử dụng để chỉ ra xem một HTTP request có được hoàn thành thành công hay không.
* **Xu hướng và công nghệ mới nổi:**
  + **5G:** Công nghệ mạng di động thế hệ mới nhất, cung cấp tốc độ nhanh hơn, độ trễ thấp hơn và dung lượng lớn hơn.
  + **Internet of Things (IoT):** Mạng lưới các thiết bị vật lý, phương tiện, thiết bị gia dụng, được kết nối với internet và có thể trao đổi dữ liệu.
  + **Trí tuệ nhân tạo (AI):** Các công nghệ AI như học máy và xử lý ngôn ngữ tự nhiên đang được sử dụng để cung cấp các ứng dụng và dịch vụ.
  + **Blockchain:** Công nghệ sổ cái phân tán cho phép giao dịch an toàn, phi tập trung.
  + **Edge computing:** Xử lý và lưu trữ dữ liệu ở rìa mạng thay vì trong các trung tâm dữ liệu tập trung.