Notes de cours

Comprendre le chemin de l'URL à la page web

Introduction

L'objectif de cette note est de comprendre ce qui se passe lorsque vous tapez une URL dans votre navigateur, en se concentrant sur le protocole HTTP.

Étape 1: Saisie de l'URL

L'utilisateur entre une URL dans la barre d'adresse du navigateur.
Exemple : `http://www.example.com`

Étape 2: Résolution DNS

La résolution DNS, pour "Domain Name System", est le processus par lequel un nom de domaine est traduit en son adresse IP correspondante. Lorsque vous tapez une URL dans votre navigateur, le système se tourne vers le serveur DNS pour trouver l'adresse IP associée au nom de domaine spécifié. Ce mécanisme permet aux utilisateurs de naviguer sur Internet en utilisant des noms de domaine faciles à retenir, tels que "www.example.com", au lieu d'adresses IP numériques, qui seraient plus difficiles à mémoriser. Le DNS agit donc comme un "annuaire téléphonique" d'Internet, fournissant un service de mappage entre les noms de domaine et les adresses IP. Ce système est fondamental pour le fonctionnement d'Internet tel que nous le connaissons.

Résolution du nom de domaine en adresse IP

Voici le processus de résolution :

1. **Cache Local**: Vérification du cache DNS local.
2. **Requête DNS**: Si l'IP n'est pas en cache, une requête DNS est envoyée.
3. **Réponse DNS**: Le serveur DNS renvoie l'adresse IP correspondante.

> **Qu'est-ce qu'une Adresse IP ?**
>
> L'adresse IP, pour "Internet Protocol", est un identifiant numérique unique attribué à chaque appareil connecté à un réseau IP. Cette adresse permet l'identification et la localisation des appareils sur le réseau.
>
> #### Format des Adresses IP
>
> ##### IPv4

L'IPv4 est la version la plus couramment utilisée des adresses IP. Elle est composée de quatre octets (de 0 à 255) séparés par des points. Par exemple, `192.168.0.1`.

IPv6

L'IPv6 est une version plus récente, conçue pour remplacer l'IPv4 en raison de l'épuisement des adresses disponibles. Les adresses IPv6 sont constituées de huit groupes de quatre chiffres hexadécimaux, séparés par des deux-points. Par exemple, `2001:0db8:85a3:0000:0000:8a2e:0370:7334`.

Types d'Adresses IP

Adresse IP Publique

Il s'agit d'une adresse unique sur tout l'Internet, utilisée pour les communications externes. Chaque site web a une adresse IP publique.
#####Adresse IP Privée
Ces adresses sont utilisées au sein d'un réseau local et ne sont pas routables sur l'Internet global. Les adresses IP privées permettent la communication entre les appareils au sein d'un même réseau local.

Classes d'Adresses IP (pour IPv4)

Les classes dans les adresses IP servent à catégoriser des blocs d'adresses IP en fonction de la taille du réseau qu'elles sont destinées à servir. Ce système de classement est spécifique aux adresses IPv4 et a été utilisé principalement pour des raisons de routage et d'attribution d'adresses. Voici un aperçu des différentes classes :

- **Classe A**: Les adresses de classe A étaient destinées à un petit nombre de réseaux qui avaient besoin d'un grand nombre de hôtes. Le premier octet est réservé pour l'identifiant réseau, et les trois octets restants sont utilisés pour les hôtes. Les adresses vont de `1.0.0.0` à `126.255.255.255`.
- **Classe B**: Cette classe était conçue pour les réseaux de taille moyenne. Les deux premiers octets sont utilisés pour l'identifiant réseau, et les deux derniers sont utilisés pour les hôtes. Les adresses vont de `128.0.0.0` à `191.255.255.255`.
- **Classe C**: Les adresses de classe C sont destinées à des réseaux plus petits. Trois des quatre octets sont utilisés pour l'identifiant réseau, laissant seulement un octet pour les hôtes. Les adresses vont de `192.0.0.0` à `223.255.255.255`.
- **Classe D**: Ces adresses sont utilisées pour le multicast et vont de `224.0.0.0` à `239.255.255.255`.
- **Classe E**: Réservée pour la recherche et l'expérimentation, elle englobe les adresses allant de `240.0.0.0` à `255.255.255.255`.
  Il est important de noter que le concept de classes d'adresses IP est maintenant en grande partie obsolète, remplacé par le CIDR (Classless Inter-Domain Routing), qui permet une allocation plus flexible des adresses IP.

Rôle dans la Communication

Lorsque vous voulez accéder à un site web, votre ordinateur utilise l'adresse IP du serveur web pour trouver sa localisation et établir une connexion. Les adresses IP sont essentielles pour le routage et la livraison de paquets de données à travers les réseaux.

Attribution des Adresses IP

- **Adresse IP Statique**: Cette adresse est fixe et attribuée manuellement. Elle ne change pas sauf si elle est modifiée manuellement.
- **Adresse IP Dynamique**: Cette adresse est attribuée automatiquement par un serveur DHCP et peut changer au fil du temps.

Types d'Adresses IP

Adresse IP Publique

Il s'agit d'une adresse unique sur tout l'Internet, utilisée pour les communications externes. Chaque site web a une adresse IP publique.

Adresse IP Privée

Ces adresses sont utilisées au sein d'un réseau local et ne sont pas routables sur l'Internet global. Les adresses IP privées permettent la communication entre les appareils au sein d'un même réseau local.

Le protocole DNS utilise généralement le protocole UDP pour ces opérations.

Processus de Requête DNS

1. **Cache Local:** Le navigateur vérifie d'abord son cache local pour voir si l'adresse IP correspondante au domaine est déjà connue.
2. **Système d'exploitation:** Si le cache local ne contient pas d'informations, le navigateur demande au système d'exploitation de résoudre le nom de domaine.
3. **Serveur DNS Configuré:** Le système d'exploitation envoie ensuite une requête au serveur DNS qui lui est configuré (souvent fourni par votre FAI).

Détails Techniques

- **Protocole:** DNS utilise le protocole UDP pour les requêtes.
- **Port:** La requête est généralement envoyée sur le port 53.
- **Format:** La requête contient le nom de domaine à résoudre et peut également spécifier le type de registre (A, AAAA, MX, etc.)

Types de Requêtes DNS

- **Requête Récurrente**: Le serveur DNS résout entièrement le nom de domaine.
- **Requête Itérative**: Le serveur DNS fournit des indications pour atteindre le serveur DNS qui peut résoudre la requête.

Le protocole UDP

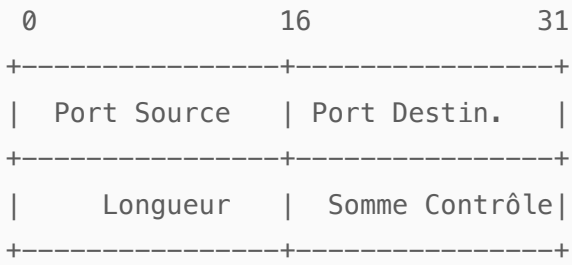
Structure d'un Datagramme UDP

- Un datagramme UDP est la plus petite unité d'information transmise sur le réseau via le protocole UDP.
- Il est composé d'un en-tête et des données du message.

Structure de l'En-tête

L'en-tête UDP est très simple et ne contient que quatre champs, chacun de 16 bits :

- **Port Source (16 bits)**: Le port de l'expéditeur.
- **Port de Destination (16 bits)**: Le port du récepteur.
- **Longueur (16 bits)**: La taille totale du datagramme, y compris l'en-tête et les données.
- **Somme de Contrôle (16 bits)**: Utilisé pour la détection d'erreurs.



Les Données

- Les données suivent immédiatement l'en-tête et constituent le "corps" du datagramme.
- La taille des données peut varier, mais la taille totale du datagramme (en-tête + données) doit être prise en compte dans le champ "Longueur".

Comprendre la perte de paquets dans le protocole UDP

- Le protocole UDP est rapide et efficace, mais il ne garantit pas la livraison des paquets.
- Cela conduit à des scénarios où des paquets peuvent être perdus.

Caractéristiques d'UDP liées à la Perte de Paquets

- **Sans Connexion:** UDP ne crée pas une connexion avant d'envoyer des paquets.
- **Sans Accusé de Réception:** UDP n'exige pas que le récepteur envoie un accusé de réception.
- **Sans Contrôle de Flux:** UDP ne gère pas le contrôle de flux, contrairement à TCP.

Causes Communes de Perte de Paquets

- **Congestion du Réseau:** Trop de paquets en circulation peuvent entraîner des pertes.
- **Buffer Overflow:** Si le tampon de réception est plein, les nouveaux paquets seront ignorés.

- **Latence Élevée:** Les paquets peuvent être supprimés s'ils ne sont pas livrés dans un délai spécifié.
- **Politiques de QoS:** Les routeurs peuvent avoir des politiques qui priorisent certains types de trafic.

Conséquences

- **Qualité de Service:** Dans des applications comme le streaming ou les jeux en ligne, la perte de quelques paquets n'est généralement pas critique.
- **Retransmission:** Les applications qui nécessitent une livraison fiable doivent gérer les retransmissions elles-mêmes.

Solutions

- **Retransmission par l'Application:** Les applications peuvent implémenter leur propre logique de retransmission.
- **Utiliser un Protocole Différent:** Si la livraison fiable est cruciale, TCP peut être une meilleure option

Résumé

- L'en-tête UDP est simple et ne comporte que quatre champs de 16 bits chacun.
- UDP est souvent utilisé pour des applications nécessitant un transfert rapide et où la perte de quelques paquets n'est pas critique.
- UDP ne garantit pas la livraison des paquets, ce qui peut entraîner leur perte pour diverses raisons.
- Les applications doivent être conçues en gardant à l'esprit ces limitations et doivent implémenter leurs propres mécanismes pour gérer la perte de paquets si nécessaire.

Types de Registres DNS

- **A**: Traduit un nom de domaine en une adresse IPv4.
- **AAAA**: Traduit un nom de domaine en une adresse IPv6.
- **CNAME**: Fournit un alias pour un autre nom de domaine.

Étape 3: Connexion TCP

Bien sûr. Après avoir obtenu l'adresse IP du serveur via une requête DNS, la prochaine étape consiste à établir une connexion TCP (Transmission Control Protocol) entre le navigateur web de l'utilisateur et le serveur web. Cette connexion est nécessaire pour garantir une communication fiable et ordonnée des données entre les deux parties.

Le processus d'établissement de cette connexion TCP est appelé "Three-Way Handshake" (poignée de main en trois étapes). Voici comment cela fonctionne :

1. **Étape 1 - SYN (Synchronize)**: Le client (en général, le navigateur web) envoie un paquet TCP avec le drapeau SYN (pour synchronisation) activé au serveur web, indiquant qu'il souhaite établir une connexion.
2. **Étape 2 - SYN-ACK (Synchronize-Acknowledgment)**: Le serveur reçoit le paquet SYN et y répond en envoyant un paquet SYN-ACK. Ce paquet signale que le serveur est ouvert à une connexion et a reçu la demande de synchronisation du client.
3. **Étape 3 - ACK (Acknowledgment)**: Le client reçoit le paquet SYN-ACK du serveur et envoie un dernier paquet avec le drapeau ACK (pour accusé de réception) activé. Ce paquet confirme que le client a bien reçu la réponse du serveur.

Une fois ces trois étapes complétées, la connexion TCP est établie, et le client et le serveur peuvent commencer à échanger des données de manière fiable et ordonnée. Cette connexion reste active jusqu'à ce qu'elle soit fermée par l'une des parties, généralement après que le navigateur ait reçu toutes les données nécessaires du serveur web.

Étape 4: Requête HTTP

Une fois la connexion TCP établie, le navigateur et le serveur peuvent commencer à échanger des données. Les types et formats de données dépendent de l'application en cours d'exécution sur le protocole TCP. Dans le cas de la navigation web, le protocole le plus couramment utilisé est le HTTP (HyperText Transfer Protocol) ou sa version sécurisée, HTTPS.

Types de Données Échangées

1. **Requêtes HTTP**: Le client (navigateur) envoie une requête HTTP au serveur pour demander un type de ressource spécifique. Cela peut être une page web, une image, un fichier CSS, un fichier JavaScript, etc.
2. **Réponses HTTP**: Le serveur renvoie une réponse HTTP qui contient la ressource demandée et des métadonnées comme le statut de la requête, le type de contenu, la longueur du contenu, etc.
3. **Données Binaires**: Dans le cas de fichiers plus gros comme des vidéos ou des documents, les données sont envoyées en format binaire.

Format des données

Requêtes HTTP

Une requête HTTP typique peut ressembler à ceci :

```
GET /index.html HTTP/1.1
Host: www.example.com
User-Agent: Mozilla/5.0
Accept: text/html
```

Réponses HTTP

Une réponse HTTP typique peut ressembler à ceci :

```
HTTP/1.1 200 OK
Content-Type: text/html; charset=UTF-8
Content-Length: 138

<!DOCTYPE html>
<html>
  <head>
    <title>Exemple</title>
  </head>
  <body>
    Bienvenue sur mon site !
  </body>
</html>
```

Données binaires

Les données binaires, telles que les images ou les vidéos, sont généralement envoyées dans un format spécifique à leur type, tel que JPEG pour les images ou MP4 pour les vidéos.

Qu'est-ce que HTTP ?

HTTP, ou HyperText Transfer Protocol, est un protocole de communication qui sert de fondement à toute communication de données sur le World Wide Web. Il définit la structure et la manière dont les requêtes et les réponses doivent être formulées entre les clients et les serveurs.

Communication sans état

HTTP est un protocole sans état, ce qui signifie que chaque requête est indépendante et qu'il n'y a pas de mémoire des interactions précédentes. Les sessions et les états doivent être gérés à un niveau supérieur, généralement via des cookies.

Éléments clés

1. **Méthodes HTTP**: Les méthodes telles que GET, POST, PUT, DELETE, etc., indiquent le type d'action que le client souhaite effectuer.
2. **Codes de statut**: Les codes de statut comme `200 OK` ou `404 Not Found` indiquent le résultat de la requête.
3. **Entêtes (Headers)**: Les informations supplémentaires concernant la requête ou la réponse sont contenues dans les entêtes.

Structure d'une Requête HTTP

Une requête HTTP typique se compose de :

- Une ligne de requête, qui contient la méthode HTTP et l'URL.
- Des entêtes de requête pour fournir des métadonnées.
- Un corps de requête éventuel.

```
GET /index.html HTTP/1.1
Host: www.example.com
User-Agent: Mozilla/5.0
```

Structure d'une Réponse HTTP

Une réponse HTTP se compose de :

- Une ligne de statut, qui contient le code de statut HTTP.
- Des entêtes de réponse pour fournir des métadonnées.
- Un corps de réponse contenant les données.

```
HTTP/1.1 200 OK
Content-Type: text/html
Content-Length: 1234

<!DOCTYPE html>...
```

Sécurité

HTTP ne chiffre pas les données en transit. Pour la sécurité, une version sécurisée appelée HTTPS (HTTP Secure) est utilisée, qui utilise le protocole SSL/TLS pour chiffrer la communication.

Les informations concernant le type d'ordinateur et le navigateur qui fait la requête sont souvent envoyées au serveur dans les en-têtes (headers) de la requête HTTP. L'en-tête le plus couramment utilisé pour transmettre ces informations est l'en-tête "User-Agent".

En-tête User-Agent

L'en-tête "User-Agent" contient une chaîne de caractères qui identifie le navigateur web, la version du navigateur, et souvent le système d'exploitation sur lequel le navigateur s'exécute. Par exemple, un en-tête User-Agent pourrait ressembler à :

```
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36
(KHTML, like Gecko) Chrome/58.0.3029.110
```

Dans cet exemple, l'en-tête nous indique que le navigateur est Chrome version 58.0, fonctionnant sur une machine Windows 10 64 bits, et utilise le moteur de rendu WebKit.

Moment de l'envoi

Ces informations sont envoyées au serveur chaque fois qu'une requête HTTP est faite. Cela inclut non seulement le chargement initial de la page, mais aussi toute requête subséquente faite par la page, comme le chargement d'images, de feuilles de style (CSS), de scripts (JavaScript), etc.

Autres En-têtes

D'autres en-têtes peuvent également être envoyés qui fournissent des informations sur les langues préférées (`Accept-Language`), les types de contenu acceptables (`Accept`), etc. Cependant, pour ce qui est du type de machine et du navigateur, l'en-tête principal à considérer est "User-Agent".

Utilisation des Informations

Les serveurs web peuvent utiliser ces informations pour diverses raisons, telles que le suivi des statistiques sur les types de navigateurs ou de systèmes d'exploitation utilisés, ou pour fournir du contenu spécifiquement formaté pour différents types de dispositifs ou de navigateurs.

En résumé, les informations sur le type de navigateur et le système d'exploitation sont généralement envoyées dans l'en-tête "User-Agent" de chaque requête HTTP, et elles peuvent être utilisées par les serveurs pour diverses tâches, y compris l'ajustement du contenu envoyé en réponse.

Étape 6: Rendu de la page

Le Rendu de la Page Web : Un Aperçu du Processus et des Outils

---

Introduction

Une fois que le navigateur a reçu la réponse HTTP contenant le contenu HTML, CSS et JavaScript, il commence le processus complexe du rendu de la page web. Ce processus met en jeu plusieurs moteurs et technologies pour transformer le code source en une page web visuellement interactive.

Étapes du Rendu

1. **Analyse du HTML**: Le moteur de rendu du navigateur décompose le document HTML en un arbre DOM (Document Object Model).
2. **Analyse du CSS**: Les règles CSS sont également analysées pour créer l'arbre CSSOM (CSS Object Model).
3. **Construction de l'Arbre de Rendu**: Le DOM et le CSSOM sont combinés pour créer un arbre de rendu qui décrit chaque élément visible sur la page.
4. **Mise en Page (Layout)**: Le navigateur calcule la géométrie de ces éléments visibles.
5. **Peinture (Painting)**: Les éléments sont ensuite dessinés sur l'écran dans un processus appelé "peinture".

Outils et Technologies

- **Moteur de Rendu**: Chaque navigateur utilise un moteur de rendu pour convertir le HTML, le CSS et le JavaScript en une page web. Exemples : Blink pour Chrome, Gecko pour Firefox.
- **JavaScript Engine**: Un moteur JavaScript, comme V8 pour Chrome ou SpiderMonkey pour Firefox, est utilisé pour exécuter le code JavaScript qui modifie le DOM et gère les interactions utilisateur.
- **GPU (Graphical Processing Unit)**: Pour des tâches de rendu graphique complexes, les navigateurs peuvent utiliser le GPU pour accélérer le processus de peinture.

Ressources Externes

- **Images, Vidéos, etc.**: Ces ressources sont souvent chargées séparément via des requêtes HTTP supplémentaires.
- **Web Fonts**: Des polices personnalisées peuvent également être chargées à partir de sources externes.

Gestion des Événements

- JavaScript écoute également divers événements, comme les clics de souris ou les entrées clavier, pour créer une page interactive.

Étape 7: Fermeture de la connexion TCP

Après le transfert des données, la connexion TCP est généralement fermée, sauf si des mécanismes comme "Keep-Alive" sont utilisés.