

Approche technique de l'espace numérique

Chapitre 2

Chargement d'une page web



Pablo Rauzy <pr@up8.edu>
pablo.rauzy.name/teaching/aten

Chargement d'une page web

- ▶ Que se passe-t-il quand on demande à son navigateur de charger une page web ?

- ▶ Que se passe-t-il quand on demande à son navigateur de charger une page web ?
 - requête DNS (*Domain Name System*),

- ▶ Que se passe-t-il quand on demande à son navigateur de charger une page web ?
 - requête DNS (*Domain Name System*),
 - requête HTTP (*HyperText Transfer Protocol*).

- ▶ Que se passe-t-il quand on demande à son navigateur de charger une page web ?
 - requête DNS (*Domain Name System*),
 - requête HTTP (*HyperText Transfer Protocol*).

- ▶ On va prendre l'exemple de `www.univ-paris8.fr`.

- ▶ Le but de la requête DNS est de trouver l'adresse IP associée au nom de domaine du site web qu'on veut consulter.
- ▶ Cette requête va être adressée à un *serveur de nom*, sous forme d'un *paquet*.
- ▶ Un paquet est constitué d'entêtes et d'un corps.

Création d'un paquet DNS

- ▶ Votre système crée donc un paquet DNS.
- ▶ Les entêtes de ce paquet disent (entre autres) :
 - je suis une requête (et pas une réponse),
 - si le message arrive en une seule fois ou si il est tronqué,
 - j'autorise (ou non) les requêtes récursives.
- ▶ Le corps de ce paquet dit :
 - pour le nom de domaine `www.univ-paris8.fr`,
 - je veux l'adresse IPv4 (**A**),
 - sur internet (**IN**).

Anatomie d'un paquet DNS

Entêtes :

Field	Description	Length (bits)
QR	Indicates if the message is a query (0) or a reply (1)	1
OPCODE	The type can be QUERY (standard query, 0), IQUERY (inverse query, 1), or STATUS (server status request, 2)	4
AA	Authoritative Answer, in a response, indicates if the DNS server is authoritative for the queried hostname	1
TC	TrunCation, indicates that this message was truncated due to excessive length	1
RD	Recursion Desired, indicates if the client means a recursive query	1
RA	Recursion Available, in a response, indicates if the replying DNS server supports recursion	1
Z	Zero, reserved for future use	3
RCODE	Response code, can be NOERROR (0), FORMERR(1, Format error), SERVFAIL (2), NXDOMAIN (3, Non existent domain), etc. ^[31]	4

Anatomie d'un paquet DNS

Requête :

Field	Description	Length (octets)
NAME	Name of the requested resource	Variable
TYPE	Type of RR (A, AAAA, MX, TXT, etc.)	2
CLASS	Class code	2

Réponse :

Field	Description	Length (octets)
NAME	Name of the node to which this record pertains	Variable
TYPE	Type of RR in numeric form (e.g., 15 for MX RRs)	2
CLASS	Class code	2
TTL	Count of seconds that the RR stays valid (The maximum is $2^{31}-1$, which is about 68 years)	4
RDLENGTH	Length of RDATA field (specified in octets)	2
RDATA	Additional RR-specific data	Variable, as per RDLENGTH

Encapsulation dans un datagramme UDP

- ▶ La requête ne part pas encore de votre système, avant ça elle est *encapsulée* dans un paquet UDP.
- ▶ UDP signifie *User Datagram Protocol*.
- ▶ Le rôle de UDP est celui d'une enveloppe de *transport* des données (ici notre paquet DNS) d'un programme source à un programme destination.

Encapsulation dans un datagramme UDP

- ▶ La requête ne part pas encore de votre système, avant ça elle est *encapsulée* dans un paquet UDP.
- ▶ UDP signifie *User Datagram Protocol*.
- ▶ Le rôle de UDP est celui d'une enveloppe de *transport* des données (ici notre paquet DNS) d'un programme source à un programme destination.
- ▶ Le système crée donc un paquet UDP.
- ▶ Les entêtes de ce paquet UDP contiennent (entre autres) :
 - le port source (par exemple **43234**),
 - le port destination (53, le port assigné aux serveurs de noms).
- ▶ Le corps de ce paquet contient le paquet DNS.

Port Source (16 bits)	Port Destination (16 bits)
Longueur (16 bits)	Somme de contrôle (16 bits)
Données (longueur variable)	

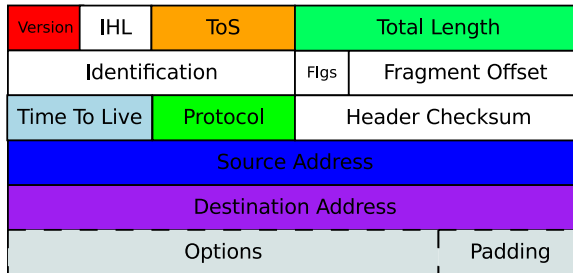
Encapsulation dans un paquet IP

- ▶ La requête ne part toujours pas de votre système ! Avant ça, elle est encapsulée dans un paquet IP.
- ▶ IP signifie *Internet Protocol*.
- ▶ Le rôle de IP est de servir de transporteur sur le *réseau*.

Encapsulation dans un paquet IP

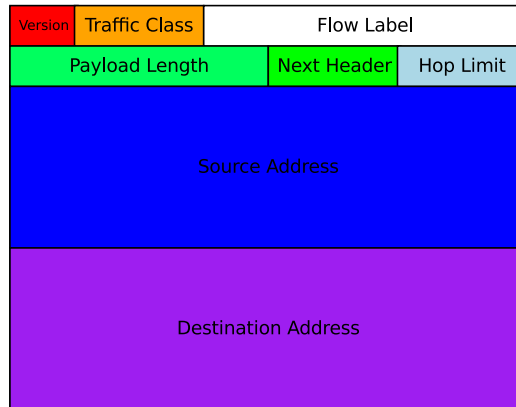
- ▶ La requête ne part toujours pas de votre système ! Avant ça, elle est encapsulée dans un paquet IP.
- ▶ IP signifie *Internet Protocol*.
- ▶ Le rôle de IP est de servir de transporteur sur le *réseau*.
- ▶ Le système crée donc un paquet IP.
- ▶ Les entêtes de ce paquet IP contiennent (entre autres) :
 - la version du protocole utilisé (4 ou 6),
 - le protocole utilisé au niveau supérieur (pour nous 17 pour UDP),
 - l'adresse source (par exemple 192.168.1.51),
 - l'adresse destination (par exemple 192.168.1.254).

IPv4 :



Anatomie d'un paquet IP

IPv6 :



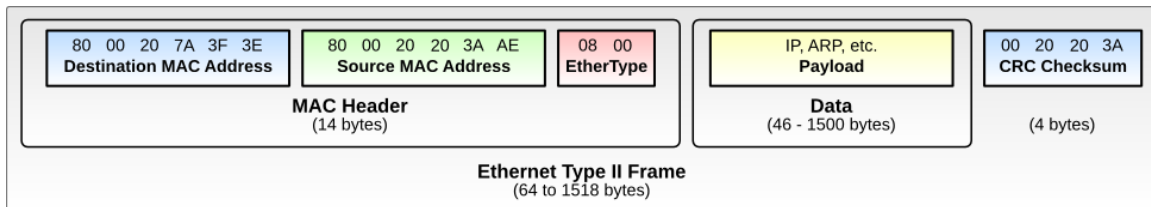
Encapsulation dans une trame Ethernet

- ▶ La requête ne part toujours pas de votre système !
Avant ça, elle est encapsulée dans un paquet Ethernet.

Encapsulation dans une trame Ethernet

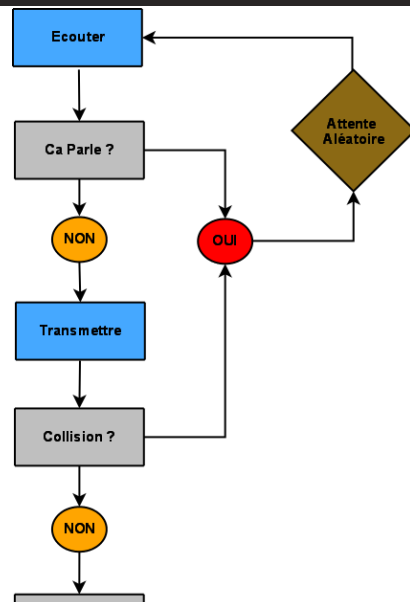
- ▶ La requête ne part toujours pas de votre système !
Avant ça, elle est encapsulée dans un paquet Ethernet.
- ▶ Le système crée donc un paquet Ethernet.
- ▶ Les entêtes de ce paquet Ethernet contiennent (entre autres) :
 - le protocole utilisé au niveau supérieur (pour nous `0x0800` pour IPv4),
 - l'adresse mac source (par exemple `3c:a9:f4:78:78:78`),
 - l'adresse mac destination (par exemple `e8:f1:b0:78:78:78`).

Anatomie d'une trame Ethernet



Ethernet et la politesse

- ▶ Le principe d'Ethernet est le même que celui des ondes radio : on communique sur un canal commun et chacun possède une clef unique (adresse MAC).
- ▶ L'orchestration se fait avec le protocole CSMA/CD (Carrier Sense Multiple Access with Collision Detection) expliqué ci-contre.



Transfert des données en WiFi

- ▶ Arrive enfin le déplacement “physique” des données.
- ▶ Dans mon cas, elles sont transmises suivant le protocole 802.11n (WiFi) jusqu'à ma box ADSL.

- ▶ La réponse à ma requête DNS est transmise par ma box ADSL à mon ordinateur de la même manière.
- ▶ Si la réponse n'était pas en cache au niveau de la box, celle-ci fait alors récursivement la requête à un autre serveur de noms (celui de mon fournisseur d'accès) :
 - Cette fois-ci, le déplacement physique s'est fait en ADSL (au moins au début),
 - la liaison avec le protocole ATM (par exemple, sur la partie ADSL).

Réponse DNS reçue

- ▶ Une fois la réponse DNS reçue, on connaît l'adresse IP de la machine qui héberge `www.univ-paris8.fr` : `193.54.174.19`.
- ▶ On peut donc lui faire notre requête HTTP.

- ▶ Le but de la requête HTTP est de récupérer le contenu de la page web qu'on veut visiter.
- ▶ Cette requête va être adressée à un *serveur web*.

Création de la requête HTTP

- ▶ Votre système crée la requête HTTP.
- ▶ HTTP est un protocole haut niveau (application).
- ▶ La requête commence par un verbe (**GET**), suivie du chemin de la ressource demandée (/), suivie de la version du protocole (**HTTP/1.1**).
- ▶ Viennent ensuite les entêtes, dont :
 - le serveur (**www.univ-paris8.fr**),
 - l'agent utilisateur (**Mozilla (...)**),
 - potentiellement des cookies, etc.
- ▶ Vient ensuite le corps de la requête (mais celui-ci est vide pour une requête **GET**).

Encapsulation dans des paquets TCP

- ▶ Comme pour la requête DNS, on a besoin d'une enveloppe.
- ▶ Cette fois-ci cependant :
 - la requête peut être trop grande pour rentrer dans une seule enveloppe,
 - on veut être sûr d'avoir transmis correctement toutes les enveloppes,
 - et aussi être capable de les remettre dans l'ordre.

Encapsulation dans des paquets TCP

- ▶ Comme pour la requête DNS, on a besoin d'une enveloppe.
- ▶ Cette fois-ci cependant :
 - la requête peut être trop grande pour rentrer dans une seule enveloppe,
 - on veut être sûr d'avoir transmis correctement toutes les enveloppes,
 - et aussi être capable de les remettre dans l'ordre.
- ▶ On utilise donc TCP (*Transmission Control Protocol*) à la place de UDP.
- ▶ TCP permet des transmissions par *sessions* :
 - on commence par établir une connexion,
 - on transfère les données,
 - puis on ferme la connexion.
- ▶ Chaque paquet TCP contient en entête (entre autres) :
 - le port source (par exemple **57743**),
 - le port destination (**80**, le port assigné aux serveurs web).
 - un numéro de séquence,
 - un numéro d'acquittement.
- ▶ Le corps de chaque paquet comprend un morceau de la requête HTTP.

Anatomie d'un paquet TCP

Offsets	Octet	0								1								2								3							
Octet	Bit	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0	0	Source port																Destination port															
4	32	Sequence number																															
8	64	Acknowledgment number (if ACK set)																															
12	96	Data offset	Reserved 0 0 0				N S	C W R	E C E	U R G	A C K	P R H	R S T	S S N	Y I N	Window Size																	
16	128	Checksum																Urgent pointer (if URG set)															
20	160	Options (if <i>data offset</i> > 5. Padded at the end with "0" bytes if necessary.)																															
...																															

- ▶ Comme les paquets UDP, chaque paquet TCP est encapsulé dans un paquet IP.
- ▶ Qui sont eux-mêmes encapsulés dans des paquets Ethernet (dans un premier temps).
- ▶ Qui eux transitent au niveau de la couche physique (WiFi, puis ADSL, etc.)

- ▶ Le *routage* des paquets est assuré par le protocole IP.
- ▶ Le paquet est transmis de réseaux en réseaux via des *routeurs*, jusqu'à arriver à destination.
- ▶ Un routeurs est un dispositif relié à au moins deux réseaux, avec une table de routage qui lui indique quoi faire en fonction de l'adresse de destination d'un paquet.
- ▶ Dans chaque *routeur* sur le trajet il se passe ceci :
 - Si l'adresse de destination est directement accessible, y envoyer le paquet.
 - Si il y a une entrée qui correspond à l'adresse de destination, y router le paquet.
 - Si il existe une route par défaut, y envoyer le paquet.
 - Sinon renvoyer un message d'erreur.

- ▶ Il existe différents protocoles de routage, en charge aussi du maintien à jour des routes existantes.
- ▶ À l'intérieur d'un système autonome, c'est le protocole OSPF (*Open Shortest Path First*) qui est le plus courant.
- ▶ Entre systèmes autonomes, c'est le protocole BGP (*Border Gateway Protocol*) qui est utilisé.

- ▶ La réponse HTTP subit le même traitement que la requête, mais depuis le serveur web.
- ▶ Une fois les différents paquets TCP arrivés et remis dans l'ordre, la réponse est recomposée et le contenu de la page web récupéré.
- ▶ Viennent ensuite les requêtes des autres composants de la page web (images, feuilles de styles, scripts, etc.).

Couches réseaux : le modèle OSI

