Virtualisation du réseau

M346 — Concevoir et réaliser des solutions cloud

Jérôme Frossard

EPAI

2 octobre 2024

Introduction

On peut dater la naissance du cloud à la mise sur le marché des premiers services web d'Amazon (AWS) au milieu des années 2000.

Mais aucune innovation ne surgit du néant. Ces services sont nés d'un besoin et ont été rendus possibles par un ensemble de technologies développées et théorisées au cours des décennies précédentes. Les technologies liées à l'Internet bien sûr, mais aussi celles liées à la virtualisation (calcul, stockage et réseau) que nous nous proposons d'explorer dans cette série de six présentations.

Après avoir abordé la virtualisation du calcul dans la deuxième présentation et la virtualisation du stockage dans la troisième, le but de cette quatrième présentation est d'aborder quelques aspects de la virtualisation du réseau.

Plan

Plan

Pile réseau

Une pile réseau (*network stack*) est un ensemble structuré de protocoles qui gère la communication entre les systèmes informatiques.

Le terme pile (*stack*) fait référence au fait que ces protocoles sont organisés en couches empilées les unes sur les autres.

Le **modèle TCP/IP** est un modèle descriptif qui reflète la structure réelle des protocoles utilisés dans l'Internet.

Vers la fin des années 1970, le **modèle OSI (*Open System Interconnection*)** a été proposé par l'ISO comme modèle de référence. Ce modèle plus détaillé que le modèle TCP/IP — et qui se voulait prescriptif — sert aujourd'hui de cadre théorique, mais n'est pas strictement implémenté dans les réseaux modernes.

Couches du modèle TCP/IP

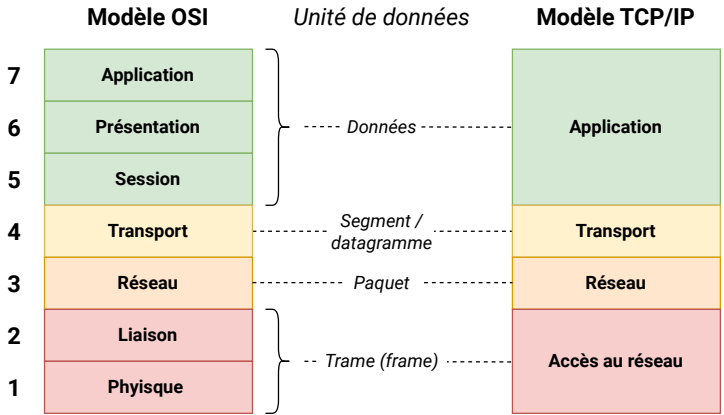
Dans cette présentation, nous nous référerons au modèle TCP/IP (*Internet Protocol*).

Le modèle TCP/IP est composé de quatre couches :

- **Couche Application**. Regroupe les protocoles qui fournissent des services aux utilisateurs (HTTP, SMTP, SSH), et des protocoles de support (DHCP, DNS, etc.).
- **Couche Transport**. Regroupe les protocoles qui permettent la transmission des données entre deux points, de manière fiable (TCP) ou non (UDP).
- **Couche Internet**. Regroupe les protocoles qui permettent l'acheminement des paquets de données à travers les différents réseaux. Le protocole IP (Internet Protocol) est central à cette couche.
- **Couche Accès réseau**. Regroupe les protocoles qui prennent en charge la transmission des données sur les supports physiques du réseau (Ethernet, WiFi, etc.).

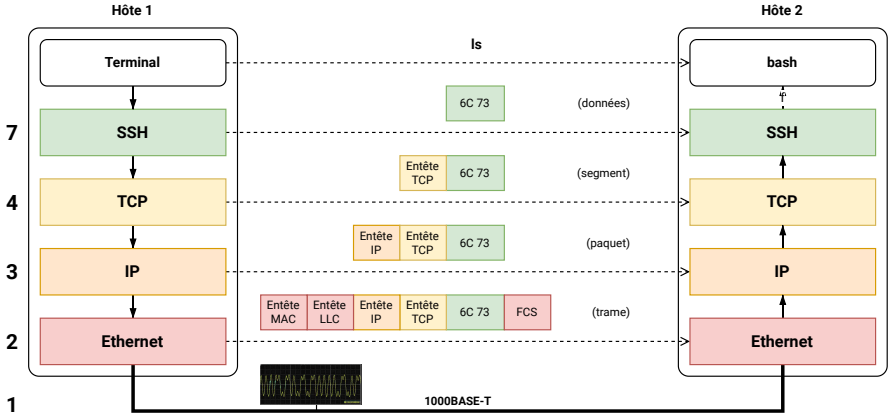
Modèle OSI et modèle TCP/IP

Correspondance entre modèle OSI et TCP/IP :



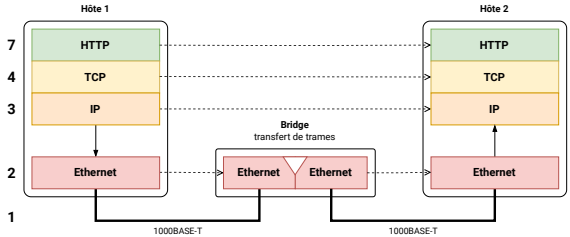
Communication directe entre deux hôtes

Le schéma ci-dessous illustre la manière dont une commande **ls** est transmise par le terminal client vers le shell distant avec le protocole SSH.



Plan

Pont (*bridge*)

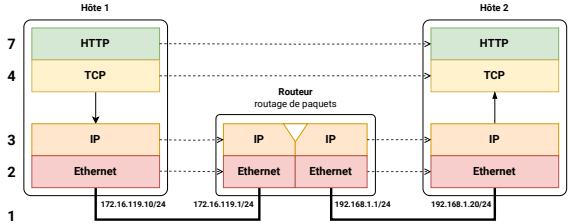


Dans un réseau, un pont (*bridge*) permet de relier plusieurs segments physiques pour former un seul segment de la couche d'accès (couche 2 de l'OSI) et un même domaine de diffusion (*broadcast domain*). Par exemple :

- deux segments *Gigabit Ethernet* (1000BASE-T),
- plusieurs segments *Gigabit Ethernet* et un segment 10GBASE-SR,
- un segment *Gigabit Ethernet* et un segment Wifi 6 (IEEE 802.11ax), etc.

La fonction d'un bridge est d'acheminer les **trames (*frame*)** entre les segments.

Routeur



Dans un réseau, la fonction d'un routeur est d'acheminer les **paquets** qui proviennent d'un hôte situé dans un réseau, vers un hôte situé dans un autre réseau.

Pour cela, le routeur examine l'en-tête de la couche réseau (couche 3 du modèle OSI), qui contient l'adresse IP de destination, afin de déterminer le réseau auquel elle appartient.

Il utilise alors une **table de routage** pour décider de l'interface à utiliser pour transférer les paquets vers l'hôte auquel ils sont destinés, ou vers le prochain routeur (*next hop*).

Commutateur (*switch*)

Un commutateur (*switch*) est un équipement réseau qui repose sur le principe d'un bridge multiport.

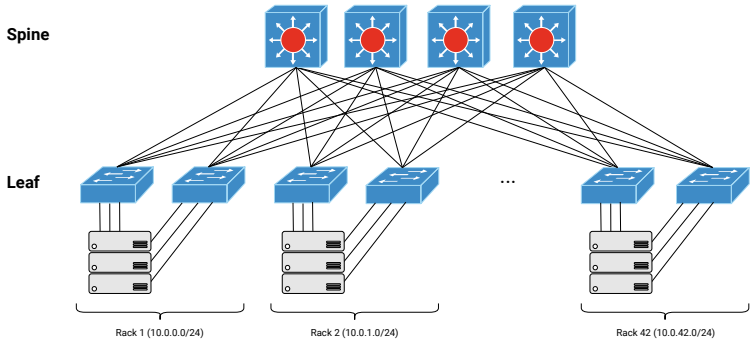
Cependant, dans une infrastructure d'entreprise, un switch de distribution ou d'accès offre des fonctionnalités qui vont bien au-delà de celles d'un bridge. Par exemple, il peut :

- Prendre en charge des protocoles de gestion des boucles comme STP (*spanning tree protocol*) ou TRILL (*Transparent Interconnection of Lots of Links*),
- Gérer des VLAN pour partitionner l'équipement en plusieurs switches virtuels,
- Utiliser des listes de contrôle d'accès (ACL) pour renforcer la sécurité,
- Prendre en charge l'agrégation de lien (LACP, MLAG, EVPN-MH, etc.) pour augmenter la bande passante et fournir de la tolérance de panne,
- Prendre en charge le routage des VLAN (switch L3), etc.

Plan

Architecture spine-leaf

L'architecture classique en trois couches est d'abord pensée pour le trafic entre les postes de travail et les serveurs (trafic nord-sud). La gestion du câblage prime sur la latence. En revanche, dans un centre de données, pour le trafic entre les serveurs hébergeant des applications fortement distribuées (trafic est-ouest), la priorité doit être accordée à la latence. Pour cela, on privilégie une autre approche : l'architecture **spine-leaf**.



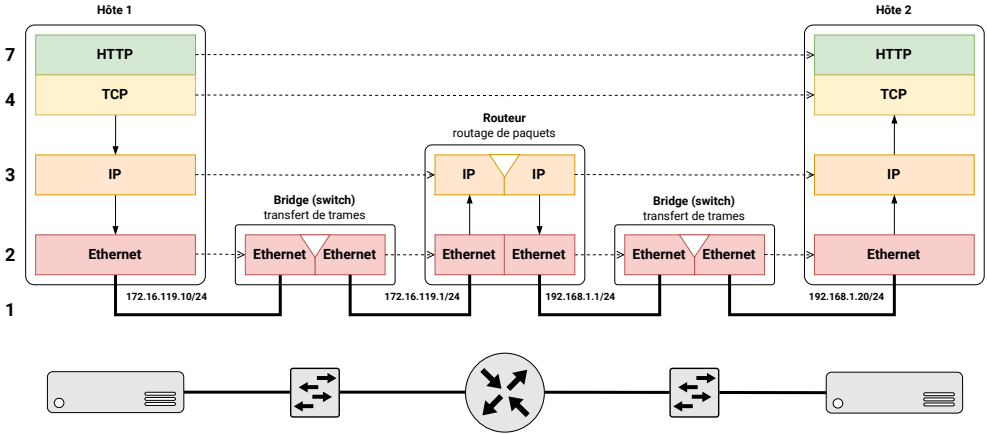
Pod, rack et sous-réseau

Dans un centre de données, les serveurs sont souvent organisés de la manière suivante :

- Un pod, souvent préfabriqué, regroupe jusqu'à quelques dizaines de racks.
- Un rack regroupe typiquement entre 20 et 50 serveurs et un ou deux ToR.
- Chaque switch ToR est connecté à chaque switch spine de l'architecture réseau, et chaque serveur est connecté à chaque ToR pour assurer la redondance.
- Les serveurs d'un même rack sont généralement placés dans un même sous-réseau, et chaque rack dispose de son propre sous-réseau distinct pour faciliter l'isolation du trafic et simplifier la gestion réseau.

Si deux serveurs se trouvent dans des racks différents, ils se trouvent dans des sous-réseaux distincts.

Communication entre deux serveurs



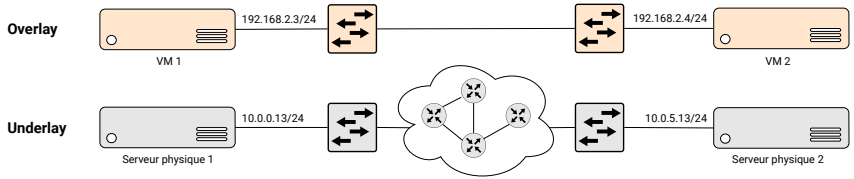
Plan

Réseau overlay et underlay

Dans un cloud, l'utilisateur n'a généralement pas de contrôle sur l'emplacement exact d'une VM ou d'un container. Par exemple :

- Si l'on crée deux VM, les serveurs physiques qui hébergent chacune d'elles peuvent se trouver dans des sous-réseaux et des domaines de diffusions distincts.
- Mais si ces VM ont été créées dans un même réseau virtuel, elles semblent se trouver dans un même sous-réseau et même domaine de diffusion.

Le réseau des serveurs est appelé **réseau underlay**, celui des VM, **réseau overlay**.



Réseau virtuel

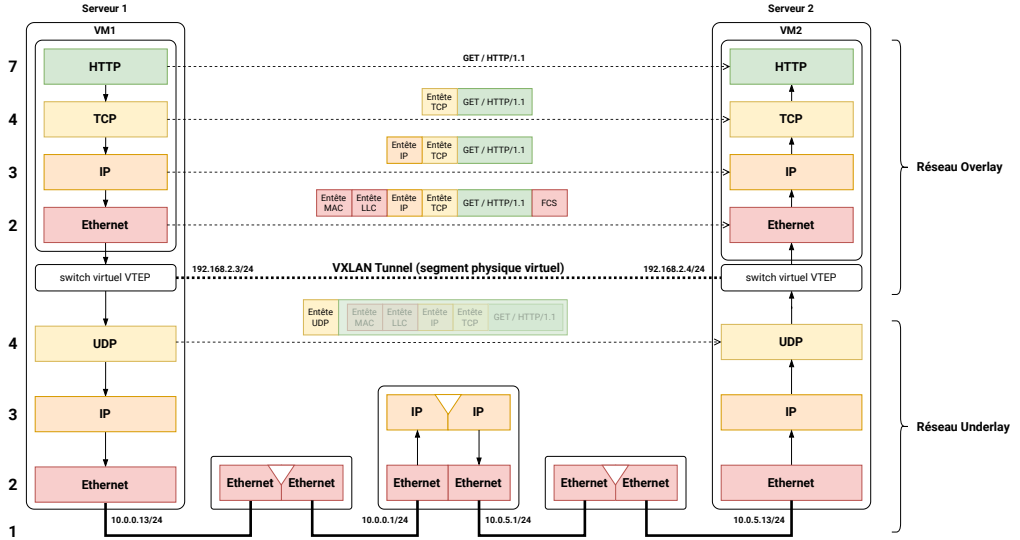
Il est possible de créer autant de réseau overlay que l'on veut au-dessus d'un réseau underlay. C'est pourquoi un réseau overlay est souvent appelé **réseau virtuel**.

Il existe différentes manières de créer un réseau overlay. L'une des plus courantes est l'utilisation de protocole d'encapsulation (ou de mise en tunnel).

Aujourd'hui, les protocoles d'encapsulation les plus utilisés sont :

- **GRE** (*General Routing Encapsulation*) encapsule différents types de paquets, y compris des paquets IP, dans des paquets IP.
- **OTV** (*Overlay Tunnel Virtualization*) encapsule des trames Ethernet dans des paquets IP.
- **VXLAN** (*Virtual Extensible LAN*) encapsule des trames Ethernet dans UDP.
- **GENEVE** (*Generic Network Virtualization Encapsulation*) encapsule n'importe quelles trames, y compris des trames Ethernet, dans UDP.

Protocole d'encapsulation



Plan

Acheminement des données et prise de décision (I/III)

Un équipement réseau physique ou virtuel permet l'acheminement de données (*data forwarding*) d'une source vers une destination.

Selon la couche à laquelle il travaille, l'équipement achemine des :

- Trames (*frame*), dans le cas d'un switch,
- Paquets, dans le cas d'un routeur,
- Segments de données ou des datagrammes dans le cas d'un pare-feu de type *packet filter*.
- Requête HTTP dans le cas d'un *reverse proxy*, etc.

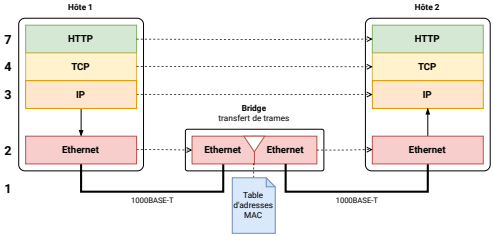
Acheminement des données et prise de décision (II/III)

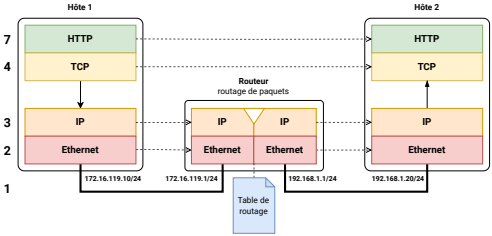
Pour acheminer correctement les données, un équipement prend des décisions basées sur ces données et des règles qui ont été préalablement définies.

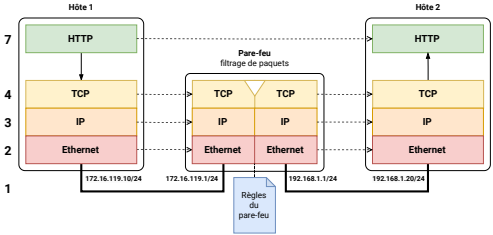
Ces règles peuvent, par exemple, prendre la forme de :

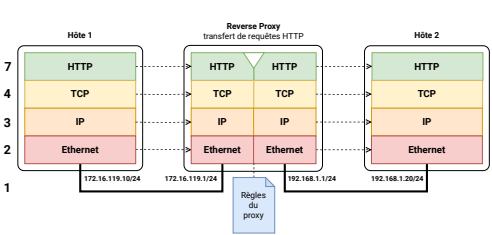
- Une table d'adresse MAC dans le cas d'un switch
- Une table de routage dans le cas d'un routeur
- Une table qui indique si un paquet est autorisé ou non en fonction de l'adresse source, l'adresse de destination, le port, etc., dans le cas d'un pare-feu de type *packet filter*
- Une table qui indique vers quel serveur doit être redirigé une requête HTTP en fonction de l'URI, de la méthode, etc., dans le cas d'un *reverse proxy*.

Acheminement des données et prise de décision (III/III)









Plan de données et plan de contrôle

Dans un équipement réseau, on peut distinguer deux sous-systèmes liés, mais bien distincts :

- Le **plan de données (*data plane*)** ou plan d'acheminement (*forwarding plane*), qui prend en charge l'acheminement des données d'un point A vers un point B, le plus rapidement possible.
- Le **plan de contrôle (*control plane*)**, qui permet de déterminer la bonne ou la meilleure manière d'acheminer les données sur la base de règles.

Réseau défini par logiciel (SDN)

Dans un cloud, le nombre d'équipements peut devenir très important, et leur gestion peut être grandement facilitée par la centralisation du plan de contrôle.

Pour cela, on utilise un système logiciel centralisé, appelé **contrôleur de SDN**, qui permet :

- De définir les règles d'acheminements pour l'ensemble du réseau de manière interactive à l'aide d'interfaces utilisateur·rices, ou automatisées avec des API
- De traduire ces règles sous des formes compréhensibles par les différents équipements.

Pour mettre en œuvre un **réseau défini par logiciel** (*software defined network* ou SDN), les équipements doivent supporter la séparation du plan de contrôle et la communication avec le contrôleur de SDN à l'aide d'un protocole standard comme OpenFlow.

Plan

Virtualisation des fonctions réseau (NFV)

La virtualisation des fonctions réseau (*network functions virtualization* ou NFV) permet de virtualiser les services réseau traditionnellement exécutés sur du matériel propriétaire.

Par exemple :

- Routeurs,
- Pare-feu,
- Répartiteur de charge, etc.

Ces services sont regroupés dans des machines virtuelles sur du matériel standard, ce qui permet aux opérateurs de faire fonctionner leur réseau sur des serveurs standard, plutôt que propriétaires.

NFV dans le cloud

Dans le cloud, il existe de nombreux services de type NFV tels que :

- **Répartiteur de charge :** Azure Load Balancer, Google Cloud Load Balancing, AWS Elastic Load Balancing, etc.
- **Passerelle d'API :** Azure Application Gateway, Google API Gateway, Amazon API Gateway, etc.
- **Pare-feu :** Azure Firewall, Google Cloud NGFW, AWS Network Firewall, etc.
- **Passerelle NAT :** Azure NAT Gateway, Google Cloud NAT, AWS NAT Gateway, etc.
- **VPN :** Azure VPN Gateway, Google Cloud VPN, AWS VPN, etc.
- **Routeur :** Azure Route Server, Google Cloud Router, AWS Transit Gateway, etc.

Dans un cloud privé virtuel, ces services permettent de gérer le trafic entrant et sortant de manière centralisée.

NFV et groupe de sécurité

À côté des services de type NFV, il est généralement possible de créer des groupes de sécurité (*security group*).

Un groupe de sécurité fonctionne comme un pare-feu à état (*stateful firewall*) dans lequel on peut définir des règles pour le trafic entrant et sortant.

En général, les règles par défaut sont :

- trafic sortant vers l'Internet autorisé
- trafic entrant bloqué

Les groupes de sécurité facilitent la gestion de la sécurité du trafic en appliquant un même ensemble de règles à toutes les ressources qui s'y trouvent.

Contrairement à un service de pare-feu, un groupe de sécurité n'est généralement pas une ressource en soi, mais un moyen simple de configurer le filtre de paquet de la pile TCP/IP des ressources qui s'y trouvent.