IPv6

IP (Internet Protocol)

C’est un protocole de niveau 3 du modèle OSI (Open System Interconnection)

Une des tâches primordiale du protocole IP est le **routage des données** dans le réseau, les stations sont identifiées par une **adresse unique**.

C’est un protocole *sans connexion*, chaque datagramme transmis contient l’adresse du destinataire et de l’émetteur.

Le protocole IP n’assure pas que le paquet transmis arrive à destination, si un datagramme est perdu le protocole IP ne prévoit rien (c’est la responsabilité de TCP – niveau 4).

Adresses IPv6

Adresses IPv6

Les adresses IPv6 ont 128 bits (un multiple de 64), on les note par paquets de 16 bits en hexadécimal (hextet)

xxxx : xxxx : xxxx : xxxx : xxxx : xxxx : xxxx : xxxx

Les premiers bits forment le préfixe qui est le numéro du réseau. La longueur du préfixe est variable, on le note

xxxx : xxxx : xxxx : xxxx : xxxx : xxxx : xxxx : xxxx / prefix

ou prefix est 0…128.

Les bits restant forment l' *Interface ID*.

Adresses IPv6

L'association IANA (Internet Assigned Numbers Authority) alloue certaines plages d'adresses à des fonctions particulières

Les adresses de 2000::/3 à 3fff::/3 Global unicast

Les adresses de fc00::/7 à fdff::/7 Unique local unicast

Les adresses de fe80::/10 à febf::/3 Link-local unicast

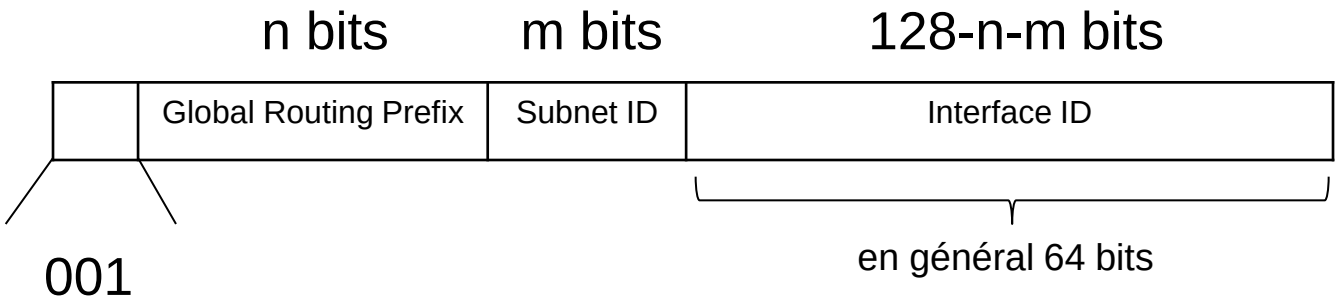
Les adresses de ff00::/8 à ffff::/8 Multicast

Adresses Unicast

Global Unicast Address (GUA)
Link Local Unicast Address (L-LUA)

Global Unicast Address

Les adresses GUA sont utilisées pour router les datagrammes. Ces adresses sont divisées en 3 parties

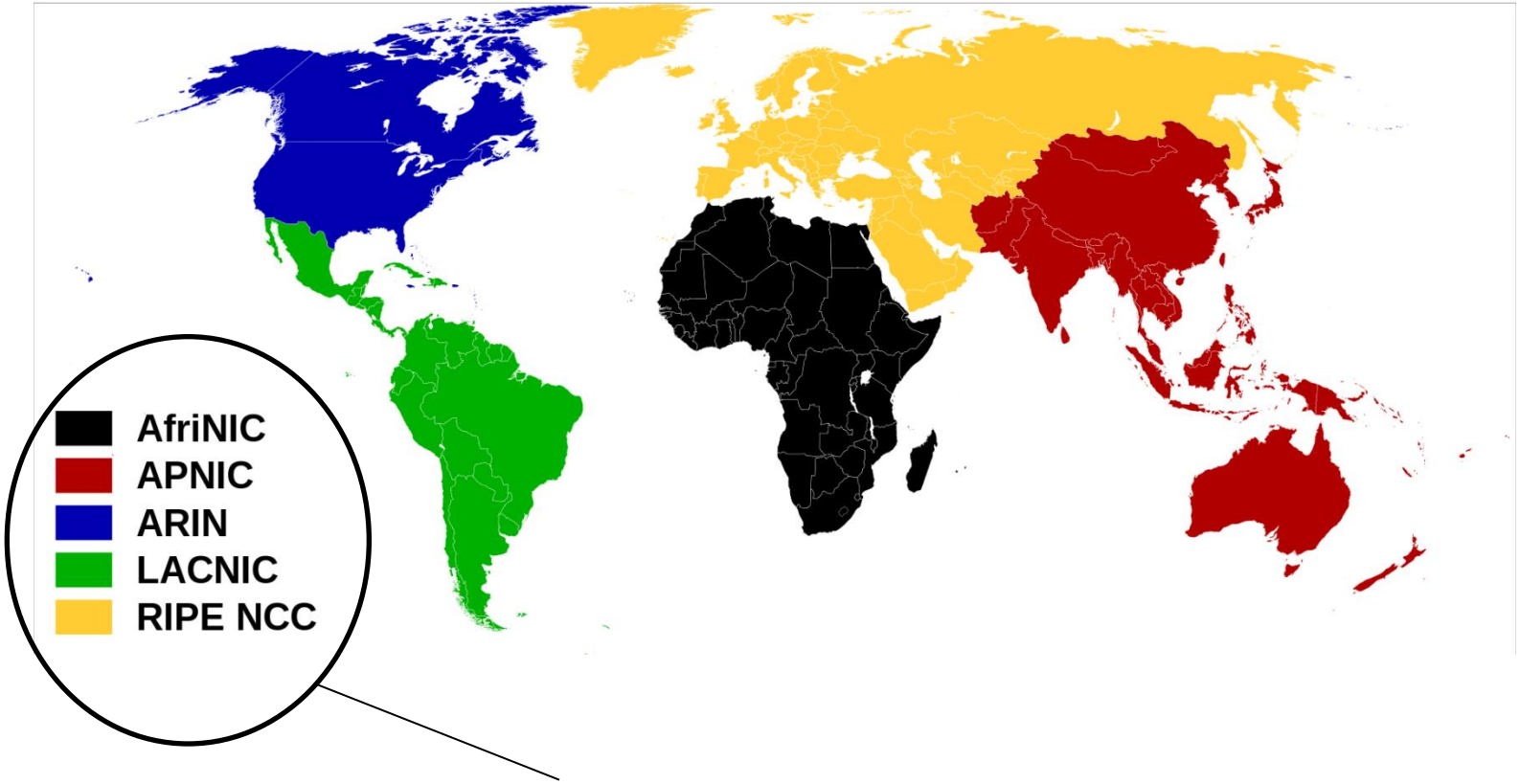


Le premier **hextet** varie de 2000::/3 à 3fff::/3

L'association IANA alloue des blocs d'adresses GUA aux 5 registres Internet régionaux (Regional Internet Registries).

Les RIRs redistribuent des blocs d'adresses à des RIRs locaux (FAI, opérateurs de réseaux) qui les redistribuent aux utilisateurs.

Global Unicast Address



Les 5 RIRs, NIC = Network Information Center

Source: www.iana.org

Allocation adresses GUA

Les adresses GUA sont attribuées aux interfaces

manuellement (statique)

dynamiquement

Pour l'allocation dynamique il y a plusieurs possibilités.

L'interface génère un message Router Sollicitation (RS) sur le réseaux local (link). Un routeur qui reçoit ce message répond avec un message Router Advertisement (RA)

Allocation adresses GUA

Dans le message RA se trouvent

- Le préfixe et sa longueur, les infos sur le link (sous réseau)
- L'adresse de la passerelle par défaut (pour accéder Internet)

La station peut:

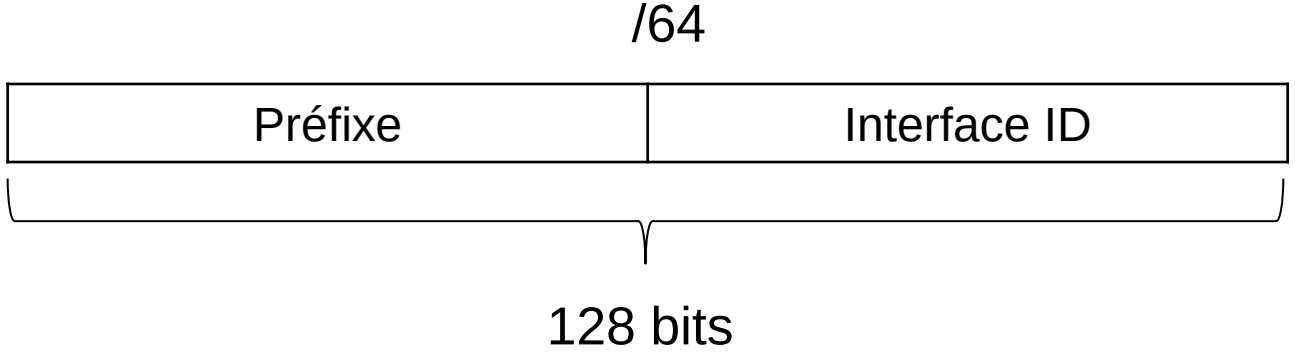
1. Stateless address configuration (SLAAC) elle crée une adresse en utilisant le préfixe du RA et l'adresse source du RA come passerelle par défaut.
2. Idem qu'au point 1. avec en plus contact à un serveur DHCP pour obtenir des informations complémentaires, p. ex. adresse d'un serveur DNS.
3. Le RA suggère de contacter un serveur DHCP, l'adresse de la passerelle par défaut est la source du RA.

Allocation adresses GUA

Dans les cas 1. et 2., le préfixe de l'adresse est celui des adresses du RA et l'Interface ID peut-être

1. Une valeur aléatoire 64 bits
2. L'identificateur EUI-64 (48 bits) de l'interface avec 16 bits (fffe) insérés au milieu.

Dans les deux cas la station crée sa propre adresse (128 bits, /64) sans serveur DHCP.



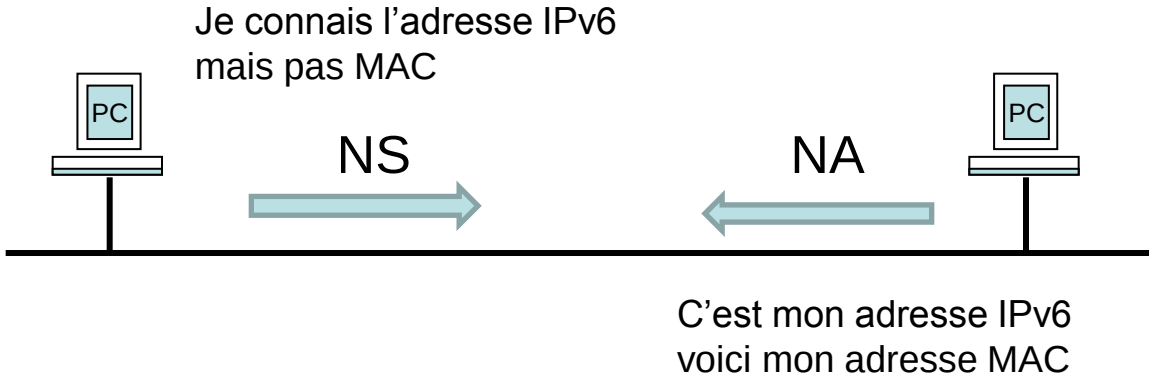
Remarque

Les routeurs émettent périodiquement des Router Advertisement RA, typiquement toutes les 200 secondes.

Remarque

Les message RS et RA sont des messages du protocole Neighbour Discovery Protocol (NDP). C'est un protocole qui a été développé pour IPv6, en fait un sous-protocole de ICMPv6 (Internet Control Message Protocol).

NDP propose aussi des messages Neighbour Solicitation NS et Neighbour Advertisment NA. Ces messages sont utilisés pour remplacer le protocole ARP de IPv4



Link-Local Unicast Address

C’est une adresse qui utilisée uniquement sur le link local (LAN, subnet).

Une station possède obligatoirement une adresse L-LUA.

L’adresse est générée automatique par l’OS de la station, elle commence par fe80::/10.

Les routeurs ne propagent jamais au réseau Internet des messages si la source à une adresse L-LUA.

Pour transmettre un message Router Solicitation, l’adresse L-LUA est indispensable.

Unique Local Address

Les adresses Unique Local Address (ULA), ne permettent pas d'accéder à Internet, ce sont des adresses **privées**.

Elles sont néanmoins **routées** par les routeurs, ce qui permet d'interconnecter des sites (privés) différents.

Les machines qui ont une adresse ULA n'accède jamais internet. Ces adresses sont aussi **globalement unique** pour assurer que si on connecte deux sites alors il n'y a pas de conflits avec les adresses ULA.

Les adresses sont **générées localement** par un algorithme pseudo-random.

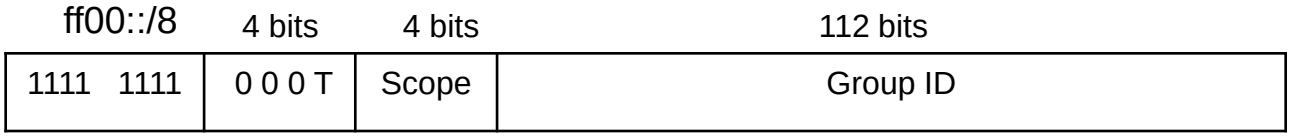
Adresses Multicast

Well-Known
Solicited Node Multicast

Adresses Multicast

Les adresses multicast déterminent des *groupes multicasts*.

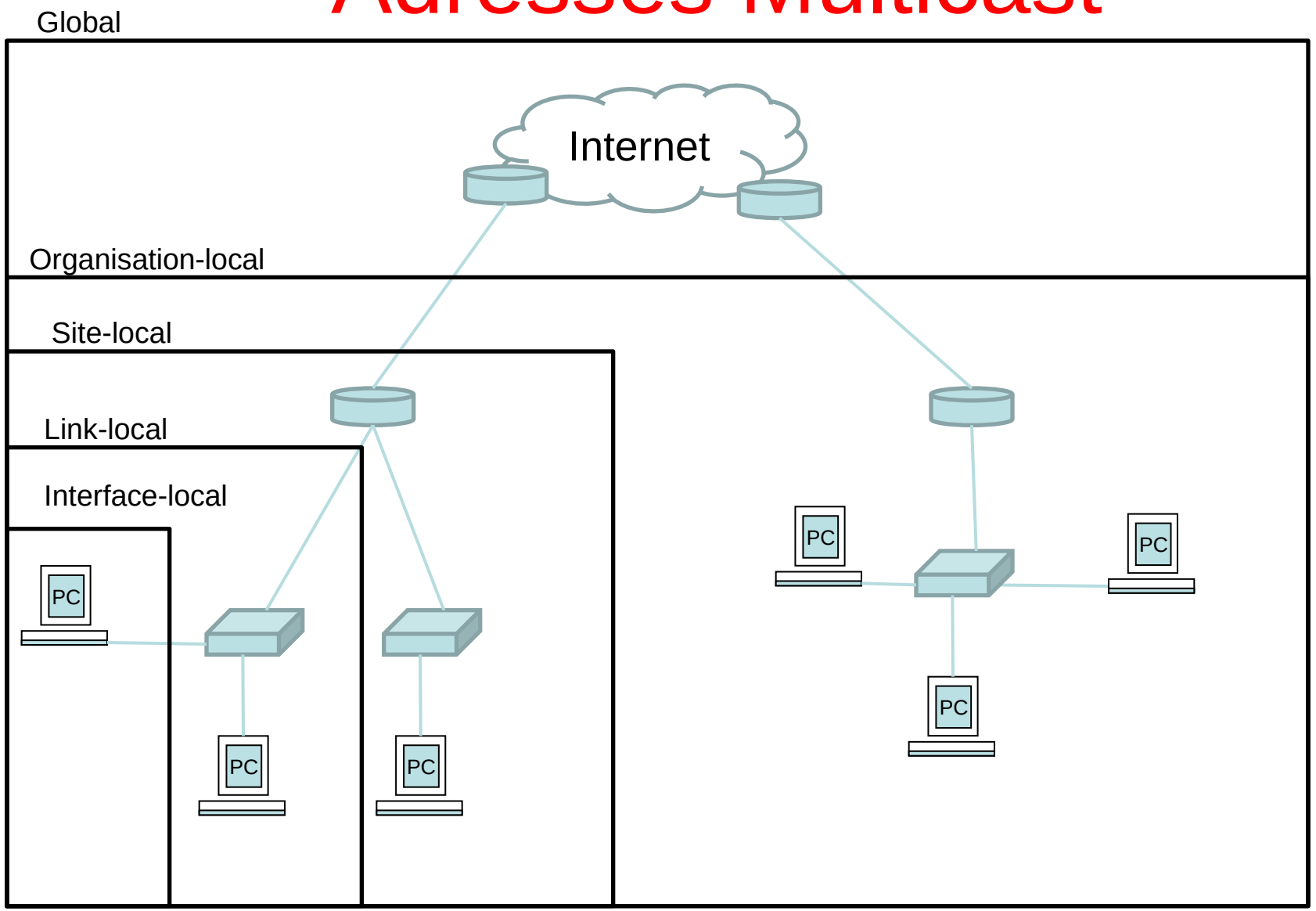
L'adresse de l'expéditeur d'un message à un groupe multicast doit être unicast (one-to-many).



T: **0** adresse permanente (well-known) assignée par l'IANA, **1** adresse non-permanente

Scope: **0** inutilisé, **1** Interface local, **2** Link-Local, **3** Unicast Prefix based **4** Admin local **5** Site local, etc.

Adresses Multicast



Filtrage des datagrammes

Les routeurs sont configurés manuellement pour filtrer les datagrammes en fonction de leur *scope.*

Les datagrammes link-local sont filtrés automatiquement.

Essayer les commandes

- netsh interface ipv6 show joins (window)
- netstat –h (linux)

pour voir les adresses multicast associées à votre machine.

Well-known ff00::/12 multicast

ff01::1 : tous les noeuds
ff01::2 : tous les routeurs

ff02::1 : tous les noeuds
ff02::2 : tous les routeurs
ff02::5 : tous les routeurs OSPF (Open Shortest Path First)
ff02::6 : OSPF designated routers
ff02::9 : tous les routers RIP (Routing Information Protocol)
ff02::a : tous les routeurs EIGRP (Enhanced Interior Gateway Routing Protocol)
ff02::1:2 : serveurs DHCP (Dynamic Host Configuration Protocol)

ff05::2 : tous les routeurs
ff05::1:3 : serveurs DHCP

Solicited-node Multicast (SNM)

A chaque adresse globale GUA, ULA, L-LUA est associé une adresse SNM.

Le préfixe de l'adresse est ff02:0:0:0:0:1:ff00::/104.

Pour chaque adresse Unicast l'adresse correspondante SNM est crée en *complétant le préfixe avec les 24 bits de poids faibles de l'adresse unicast*.

Chaque adresse SNM correspond à un groupe multicast. L'interface accepte un message et le passe à la couche réseau de la station si elle fait partie du groupe. Sinon, elle est ignorée (filtrée)

Exercice

Trois PC se trouvent sur le même lien avec les adresses:

PCA: 2001:db8:cafe:1:aaaa::200

PCB: 2001:db8:cafe:1:bbbb::200

PCC: 2001:db8:cafe:1:aaaa::201

Un quatrième PC transmet un message Neighbor solicitation (NS) avec l'adresse de PCA.

Indiquez les adresses utilisées pour former le message NS et comment il est traité par PCA-PCB et PCC.

Adresses SNM - usages

1. Dans certaines situation, une machine connaît l'adresse IP du destinataire **sur le même lien** mais pas l'adresse MAC. La station transmet un message Neighbor solicitation protocole NDP, en utilisant pour adresse destination l'adresse **Solicitated-Node multicast** qui correspond à l'adresse ipv6.

L'intérêt de cette technique par rapport à utiliser une adresse broadcast est que les interfaces peuvent filtrer les messages multicast.

Adresses SNM - usages

2. Les adresses Unicast sont générées par les stations en utilisant des mécanismes aléatoires (préfixe connu complété par des bits obtenus aléatoirement). Pour vérifier que l'adresse est bien unique la station peut envoyer une message NS (Neighbor solicitation – NDP) à l'adresse multicast correspondant sur le lien (DAD: Duplicate Addresse Detection).

Mapping des adresses multicast layer 3 -> layer 2

Les adresses SNM sont converties en adresses MAC pour être routées au niveau 2.

L'adresse MAC utilisées est l'adresse 33-33-xx-xx-xx-xx (48 bits)

et les xx-xx-xx-xx sont les 24 bits de poids faibles de l'adresse multicast.

Les trames avec des adresses MAC 33-33-xx-xx-xx-xx sont transmises par les commutateurs sur tous les ports (sauf le port entrant - broadcast).

Les interfaces Ethernet des stations acceptent les adresses 33-33-xx-xx-xx-xx qui correspondent à une adresses SNM (et donc Unicast). Les autres sont filtrées déjà au niveau 2.

Remarques

IPv4 utilise des adresses de broadcast, les adresses 33-33-xx-xx-xx-xx n'existent pas. L'intérêt du mécanisme de ipv6 est que les datagrammes peuvent être filtrés au niveau 2 (si les 24 bits de poids faibles ne correspondent pas).

Les adresses *Well-Known Multicast* ont aussi des équivalent 33-33-xx-xx-xx-xx.

Routage des adresses multicast niveau 3

Un routeur va transmettre un datagramme uniquement si l'adresse de la destination se trouve dans sa table de routage.

Pour permettre aux stations d'interagir avec les routeurs ils utilisent le **protocole MLD** (Multicast Listener Discovery).

Il y a 3 types de messages:

1. **Multicast Listener Query**: périodiquement le routeur transmet un message qui peut-être:
    1. **General Query**: Demande aux stations qui appartiennent à des groupes de s'anoncer, envoyé à l'adresse ff02::1.
    2. **Multicast Address Specific Query**: Envoyé à l'adresse multicast et attend une réponse pour vérifier que l'adresse est toujours actuelle.

Routage des adresses multicast niveau 3

2. **Multicast Listener Report**: Une station transmet au routeur pour indiquer son appartenance à un groupe. Elle utilise l'adresse ff02::16 qui correspond aux routeurs MLDv2.

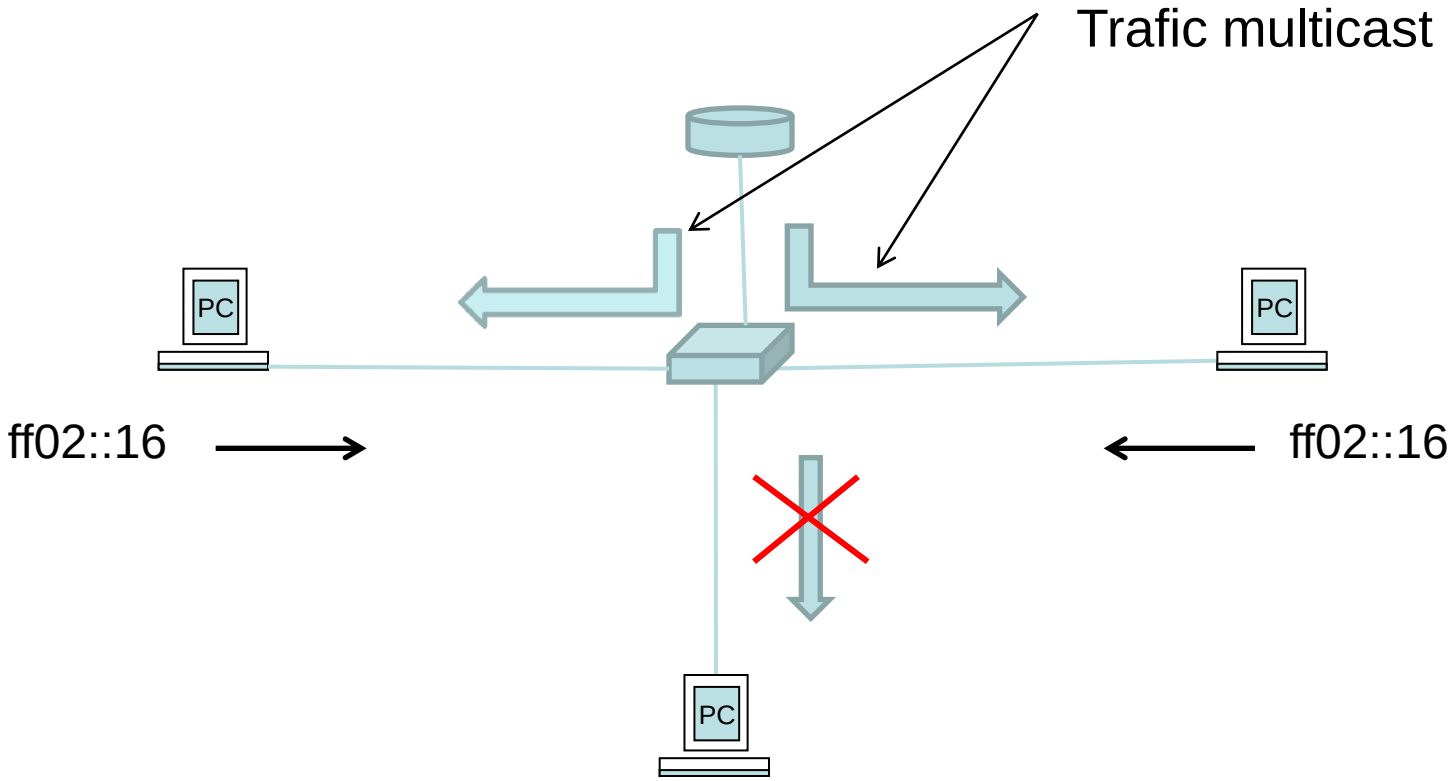
3. **Multicast Listener Done**: Une station transmet au routeur pour indiquer quelle quitte le groupe multicast. Le message est transmis à l'adresse ff02:2.

Filtrage multicast niveau 2

Les stations doivent s'annoncer auprès d'un routeur pour indiquer quelles sont destination d'une adresse multicast.

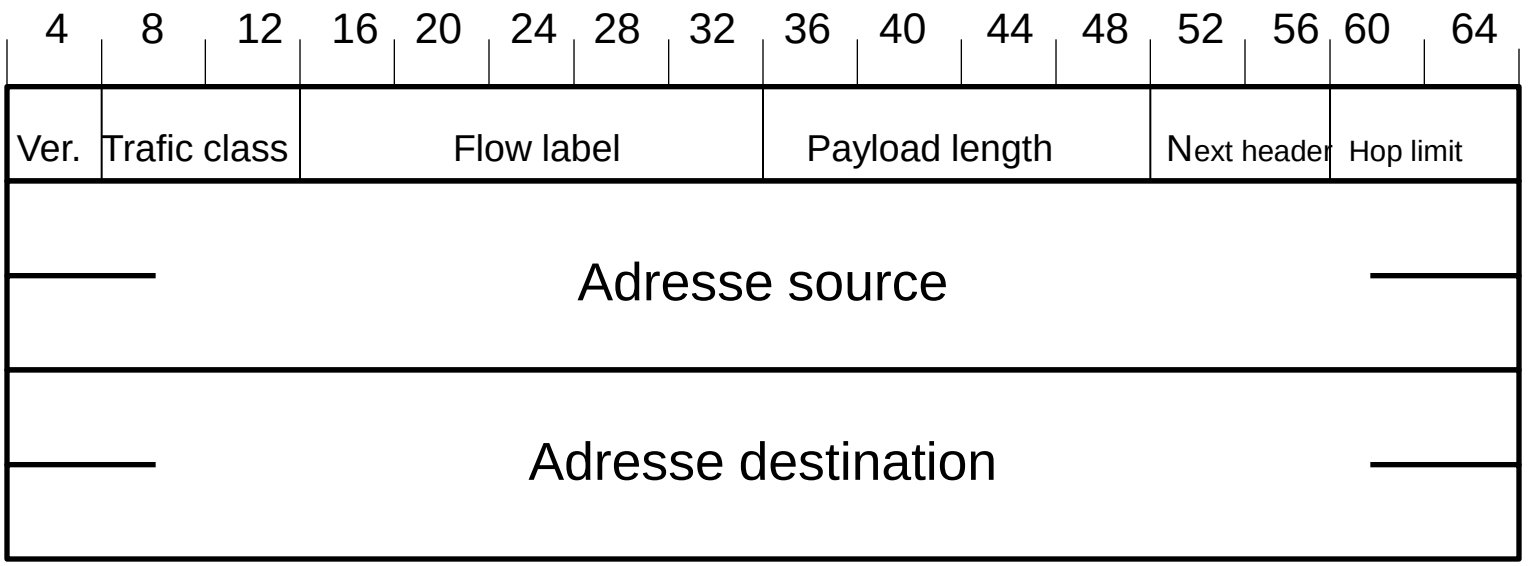
L'adresse utilisée au niveau 3 est ff02::16, au niveau 2 l'adresse est 33-33-00-00-00-16.

MLD Snooping est utilisé par les commutateurs pour détecté les stations qui appartiennent à des groupes multicast et transmettre les trames uniquement sur les ports qui mènent à une telle station.


Trafic multicast
PC
PC
PC
ff02::16
ff02::16

Format des datagrammes IP

Datagrammes IP



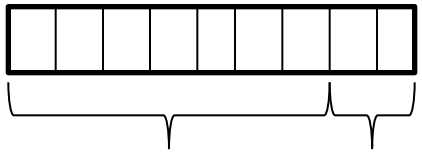
Format optimisé pour les architectures 64 bits.

En-tête de taille fixe (= 40 bytes).

Datagrammes IP

Ver. : la version du protocole (6).

Trafic class :



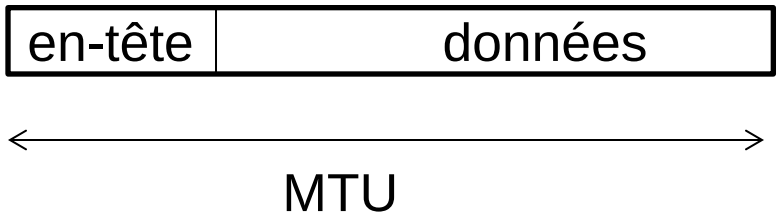
Le champ DSCP (DiffServ Code Point) code sur 6 bits le service offert aux datagrammes par le routeur. Ces bits permettent d'assurer une **qualité de service** (QoS) qui varie en fonction des datagrammes. Par exemple, implémenter des mécanismes de priorités

Datagrammes IP

Flow label : Ce champ permet d'associer un même identifiant à tous les datagrammes qui composent un même flow de données. Ca permet aux routeurs de traiter tous les datagrammes d'une même flot de manière identique.

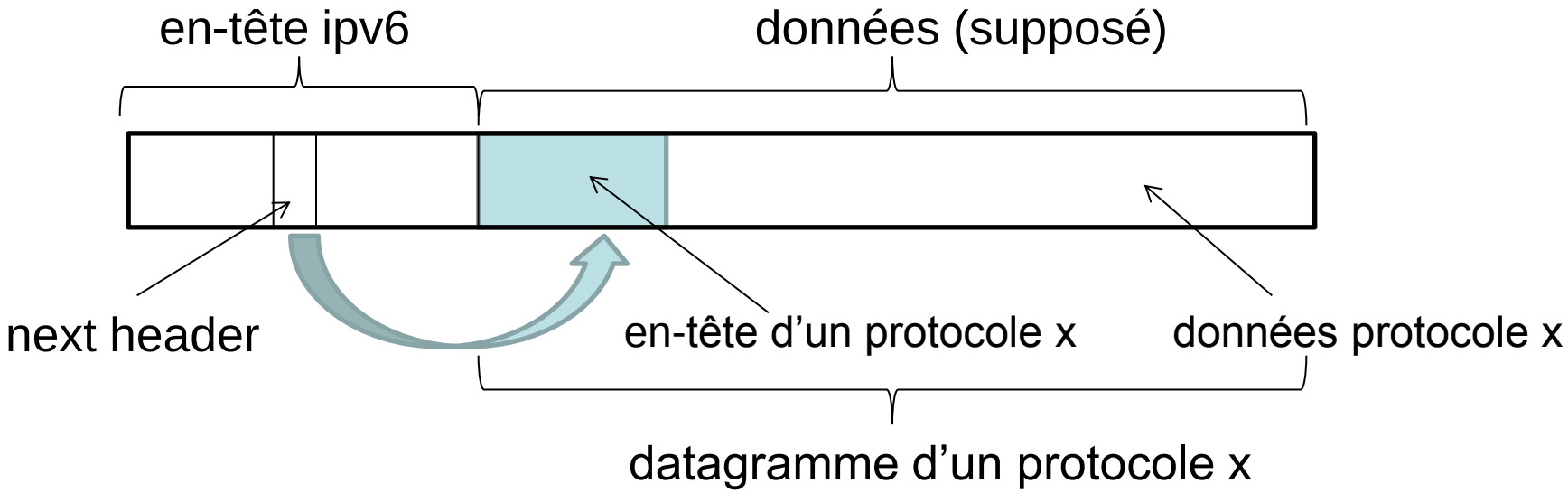
Payload length : Indique la taille des données du datagramme sans l'entête. Pour un datagramme IPv6 le Maximum Transmission Unit (**MTU**) la taille maximum d'un datagramme (payload + en-tête) doit être supérieure à 1280 bytes (typiquement 1500 bytes).

Les 16 bits du champ payload length limitent à 65Kbytes la taille maximale du champ données.



Datagrammes IP

Next header : Ce champ permet de spécifier que les bits qui suivent l'en-tête du datagramme ne sont pas des données mais l'en-tête d'un autre protocole **encapsulé** dans la datagramme ipv6.



Next header – quelques exemples

0 : Hop-by-hop extension header
1 : Internet Control Message protocol (ICMPv4)
2 : Internet Group Management Protocol (IGMPv4)
4 : IPv4 encapsulation
5 : Internet Stream Protocol (ST)
6 : Transmission Control Protocol (TCP)
7 : User Datagram Protocol (UDP)
46 : Resource Reservation Protocol (RSVP)
50 : Encapsulating Security Payload (ESP)
51 : Authentication Header (AH)
88 : Enhanced Interior Gateway Routing Protocol (EIGRP)
89 : Open Shortest Path First (OPSF)

Datagrammes IP

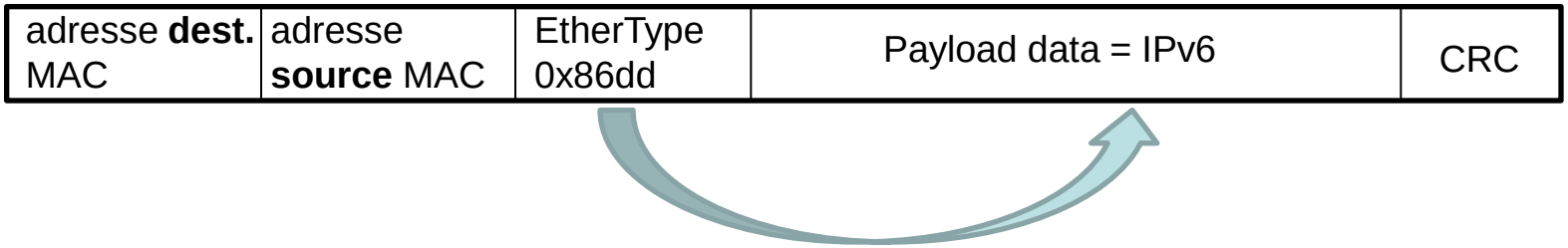
Hop limit : Ce champ est décrémenté par chaque routeur. S'il atteint 0 alors le datagramme est détruit et un paquet ICMP Time Exceeded est envoyé à la source.

Remarques

Il n'y a pas de **checksum** dans l'en-tête IPv6 pour vérifier l'intégrité du datagramme car un tel checksum existe au niveau 2 et au niveau 4 (TCP et UDP). Normalement, le checksum est facultatif pour UDP (avec IPv4) avec IPv6 il devient obligatoire.

L'adresse **source** doit être un adresse **unicast**.

IPv6 over Ethernet



Exemple
Next header et IPsec

IPsec

IPsec est une suite de protocoles utilisées pour garantir la sécurité des transmissions. En particulier, il s'agit d'assurer

- L'authentification des communicants
- L'intégrité des données

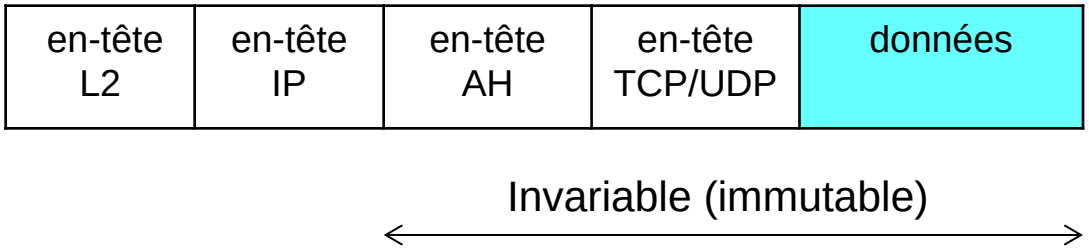
Deux mécanismes principaux sont proposés

- Authentication Header (AH) – authentification et intégrité des données.
- Encapsulating Security Payload (ESP) – authentification, intégrité et cryptage des données.

Mode Transport/Tunnel

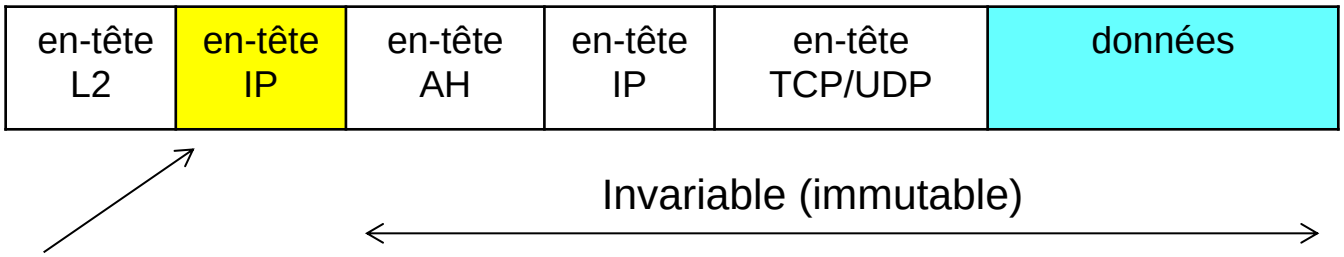
Trame à transmettre

Protocol AH en mode transport



Mode Transport/Tunnel

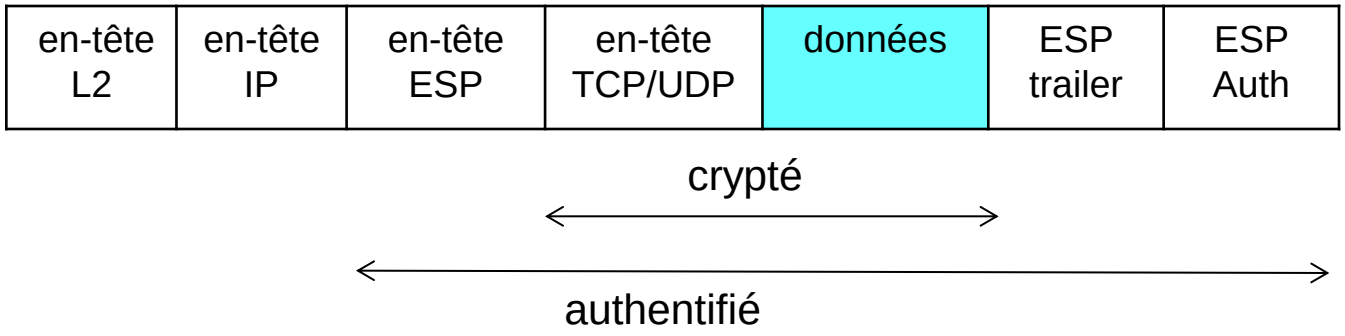
Protocol AH en mode tunnel



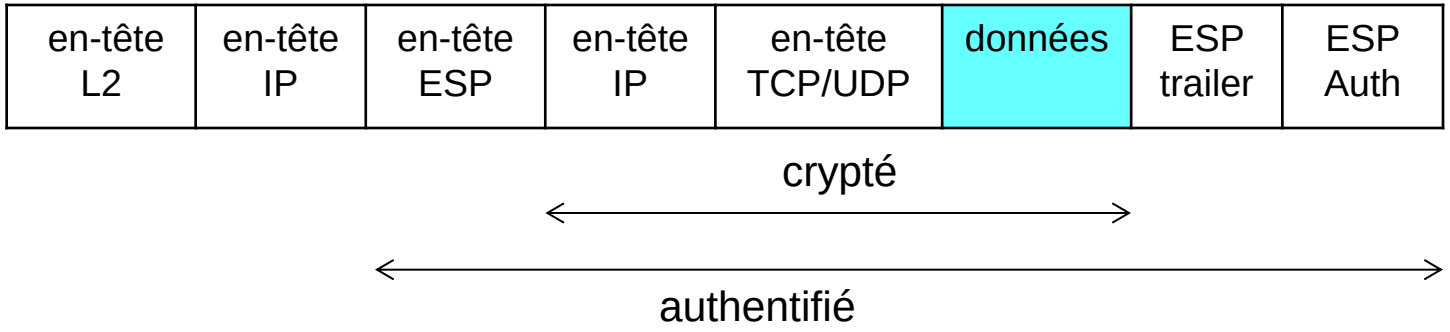
nouvelle en-tête IP

Mode Transport/Tunnel

Protocol ESP en mode transport



Protocol ESP en mode tunnel



Algorithme de routage pour Internet

Organisation

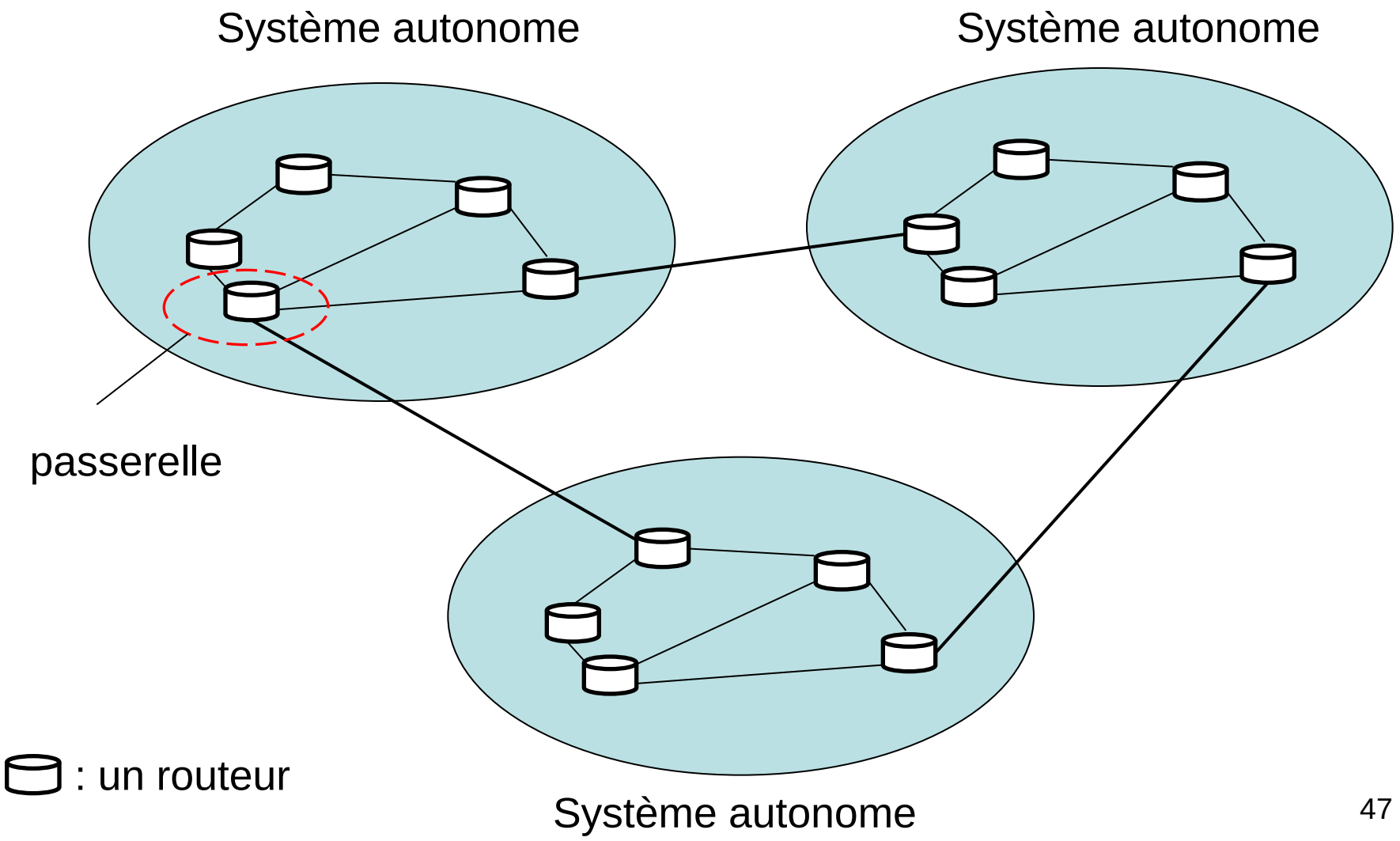
Il n’existe pas un unique algorithme de routage pour tout Internet.

Le réseau est composé de **systèmes autonomes (AS)**, qui sont en général administré par une seule entité et chacun utilise sont propre algorithme de routage.

Un protocole de routage utilisé dans un AS s’appelle **Interior Gateway Protocol (IGP).**

Pour communiquer entre systèmes autonomes les protocoles utilisés s’appellent **Exterior Gateway Protocol (EGP)**.

Organisation



Routage par vecteur de distance

Généralités

Le routage par vecteur de distance s'appelle aussi quelque fois algorithme de Bellman-Ford ou de Ford-Fulkerson.

C'est un protocole prévu come IGP (interior gateway protocol)

- En général la taille du réseau est limitée, par exemple le diamètre du graphe de communication est limité à 15 (15 hops).
- Il utilise une métrique fixe pour déterminer les distances entre routeurs (administration).
- La vitesse de convergence peut-être lente pour des réseaux de (trop) grandes tailles.

Généralités

Le protocole est un protocole de routage par le **plus court chemin**.

Il existe d'autres protocoles qui utilisent les chemine les plus courts, se qui change c'est la manière de calculer ces chemins.

On note d(x,y) la distance entre les routeurs x et y. Comme c'est une distance elle doit satisfaire

- d(x,y) = d(y,x) symétrie
- d(x,y)=0 ⇔ x=y
- d(x,y) >= 0
- d(x,z) <= d(x,y) + d(y,z) inégalité du triangle

Métrique (distance)

Pour la définition de la métrique (fonction distance), on suppose qu'elle est connue pour deux routeurs x et y qui sont voisins (communiquent directement). On note x~y deux routeurs voisins et D(x,y) la distance.

Un chemin est une suite de routeurs x1, x2, … , xn tels que xi~x(i+1). La **longueur du chemin** est la somme des longueurs D(xi,x(i+1)).

On note P(x,y) l'ensemble des chemins de x à y et w(p) la longueur du chemin p.

On définit **la distance** entre deux routeurs x et y comme étant

$$d(x,y) = \min\{ w(p) , p \text{ appartient à } P(x,y)\}$$

Calcul de la distance

La définition de la distance suppose qu'on est capable d'énumérer tous les chemins et de calculer leur distance.

Une définition équivalente qui conduit à un algorithme distribué est de noter que la fonction d(x,y) satisfait

$$d(x,y) = \begin{cases} 0 & \text{si } x=y \\ \min\{ D(x,z) + d(z,y), x\sim z\} & \text{sinon} \end{cases}$$

Le chemin le plus court de x à y passe par z un voisin de x (x~z) et depuis z on suit le chemin le plus court vers y.

Algorithme

On exécute l'algorithme suivant pour calculer la fonction d(x,y).

Initialisation: D(x,y) = D(y,x) est donnée

d(x,y)= D(x,y) si x~y (x et y sont voisins)

d(x,y) = ∞ si x ≠ y et ne sont pas voisins

d(x,x) = 0

repeat

choisir x et y au hasard, x ≠ y

calculer d(x,y) = min{ D(x,z) + d(z,y), x~z }

AlgorithmeV2

On exécute l'algorithme suivant pour calculer la fonction d(x,y).

Initialisation: D(x,y) = D(y,x) est donnée

d(x,y)= D(x,y) si x~y (x et y sont voisins)

d(x,y) = ∞ si x ≠ y et ne sont pas voisins

d(x,x) = 0

repeat

choisir x et z au hasard, x ~ z

pour tout y

calculer d(x,y) = min{ d(x,y), D(x,z) + d(z,y)}

Exercices

a. Montrez l'équivalence entre les deux définitions.

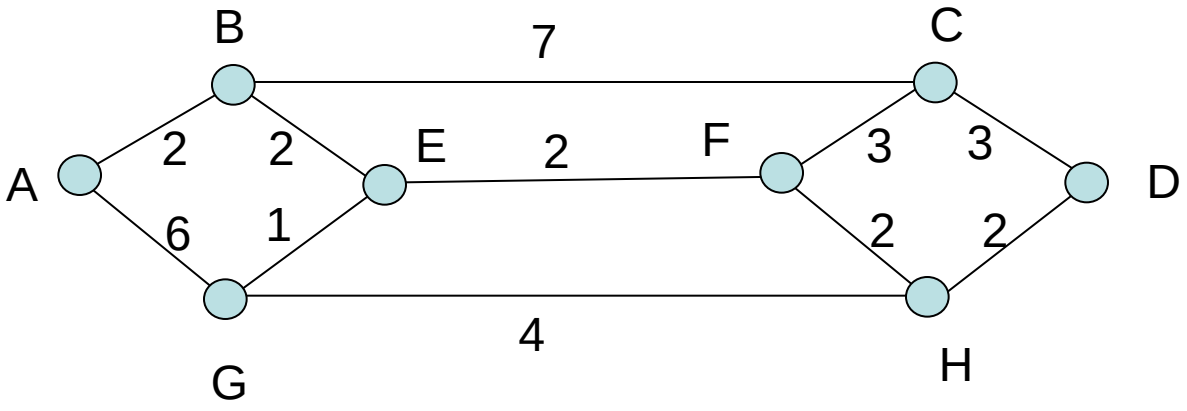
b. Montrez qu'on a bien définit une distance. Notez de la fonction D(x,y) doit être symétrique.

c. Montrez que l'algorithme pour le calcul des distances converge.

Routage par vecteur de distance

Pour le routage chaque routeur doit mettre à jour **sa table de routage**, qui contient (au minimum) pour chaque routeur du réseau, la longueur du plus court chemin ainsi que le routeur **voisin** par lequel ce chemin passe.

Exemple: les routeurs sont notés A, B, …



Routage par vecteur de distance

On peut vérifier que pour ce réseau de communication, la table de routage finale en A doit être

destination	distance	next hop (via)
B	2	B
C	9	B
D	10	B
E	4	B
F	6	B
G	5	B
H	8	B

Routage par vecteur de distance

Initialement, les routeurs connaissent uniquement les routeurs voisins.
La table de routage en A vaut:

Table de routage en A		
destination	distance	next hop (via)
B	2	B
C	∞	-
D	∞	-
E	∞	-
F	∞	-
G	6	G
H	∞	-

Routage par vecteur de distance

En fait, A peut ne pas avoir connaissance des routeurs non voisins, il n'y a pas d'entrée dans la table.

Table de routage en A		
destination	distance	next hop (via)
B	2	B
G	6	G

Les routeurs vont tous appliquer l'algorithme de calcul des chemins les plus courts. Mais pour aller plus vite, ils transmettent les informations à tous leurs voisins.

Routage par vecteur de distance

Périodiquement, ils vont transmettre à leurs voisins leur vecteur de distance. Pour A c'est au départ

Vecteur distance A	
destination	distance
B	2
G	6

Routage par vecteur de distance

La table de routage originale de B est

Table de routage en B		
destination	distance	next hop (via)
A	2	A
C	7	C
E	2	E

A la réception du vecteur de A, B apprend l'existence d'un routeur G, qui se trouve à distance 6 + 2 = distance de B à A + distance de A à G

Table de routage en B		
destination	distance	next hop (via)
A	2	A
C	7	C
E	2	E
G	8	A

Routage par vecteur de distance

Le routeur B va à son tour transmettre son vecteur distance à A, C et E.

Vecteur distance de B	
destination	distance
A	2
C	7
E	2
G	8

Le routeur A apprend l'existence des routeurs C et E qui se trouvent à distances 9 et 4 en passant par B

Table de routage en A		
destination	distance	next hop (via)
B	2	B
C	9	B
E	4	B
G	6	G

Routage par vecteur de distance

Pour la dernière mise-à-jour discutée de la table de routage en A, il faut remarquer que A à connaissance de deux chemins pour aller à G.

- Le chemin original de sa table de routage avec longueur 6.
- Le chemin annoncé par B. La longueur de ce chemin est 8+2=10, la longueur du chemin de B vers G + la longueur du chemin vers B.

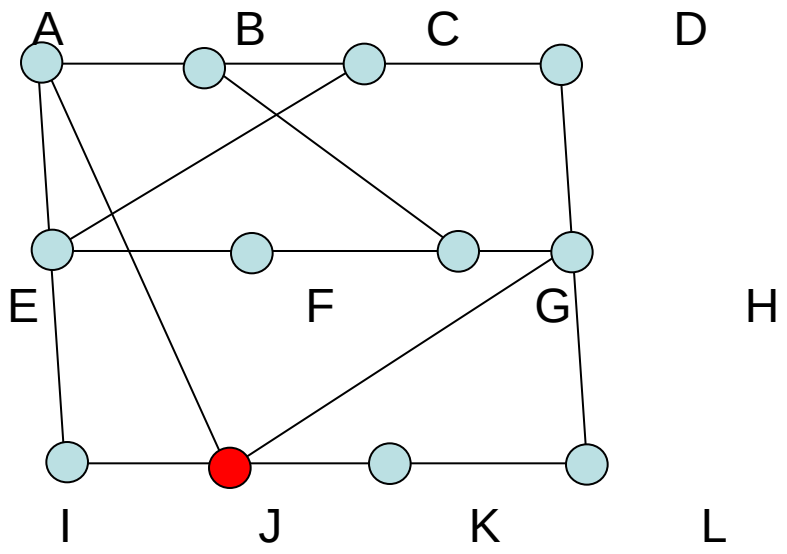
A compare ces deux chemins et conserve uniquement le plus court.

Exercices

- Continuez à décrire les mise-à-jour des tables de routage (de A et B) en supposant que E puis B transmettent leur vecteur distance.

Exercices

On considère le réseau et les tables de routages ci-dessous



tables des routeurs A I H K

vers	A	I	H	K
A	0	24	21	21
B	12	36	31	28
C	25	18	19	36
D	40	27	8	24
E	14	7	30	22
F	23	20	19	40
G	18	31	6	31
H	17	20	0	19
I	21	0	14	22
J	9	11	7	10
K	24	22	22	0
L	29	33	9	9

Exercices

A un instant donnée le routeur J reçoit tous les vecteurs de distance de ces voisins A, I, H, et K.

Supposons que J ait mesuré des délais de 8 ms pour JA, 10 ms pour JI, 12 ms pour JH et 6 ms pour JK.

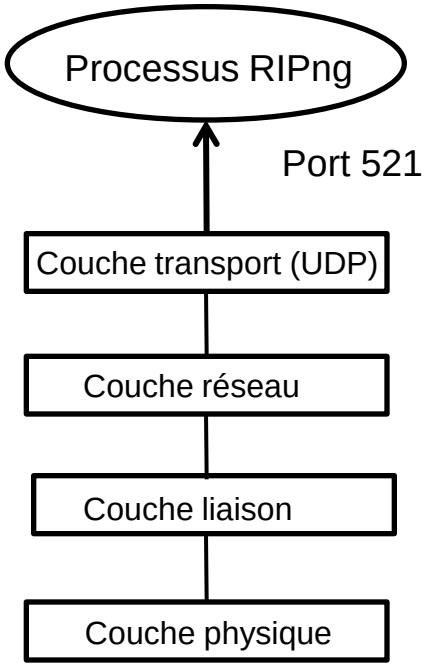
Construisez la table de routage de J.

RIPng

Routing Information Protocol next generation

RIPng

- C’est un protocole à vecteur de distance
- La métrique utilisée est le nombre de sauts (hops)
- La taille des réseaux est limitée à 15 sauts (16 = infini)
- Les routeurs échangent des **datagrammes UDP** sur le port 521



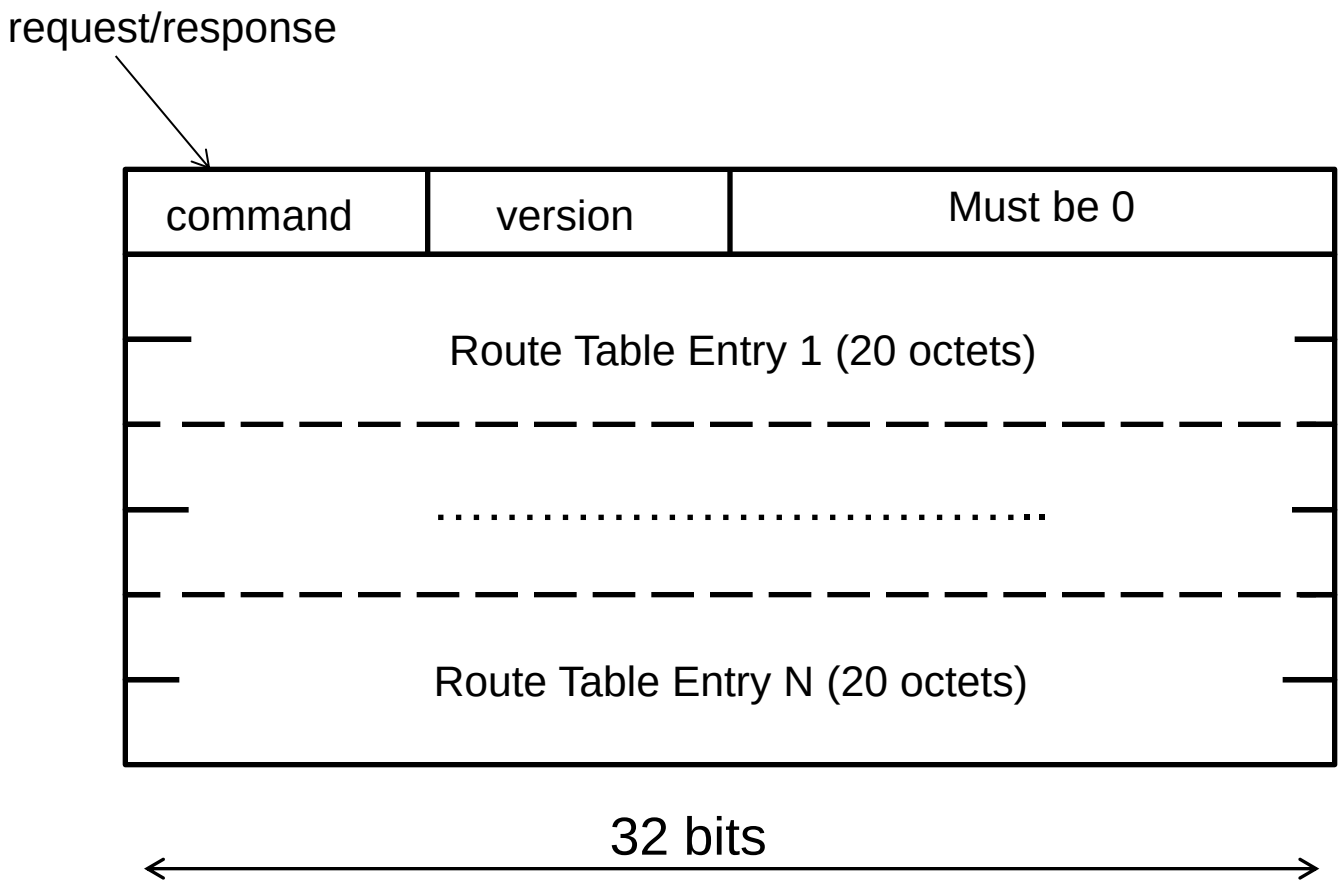
RIPng

Les routeurs échangent deux types de messages

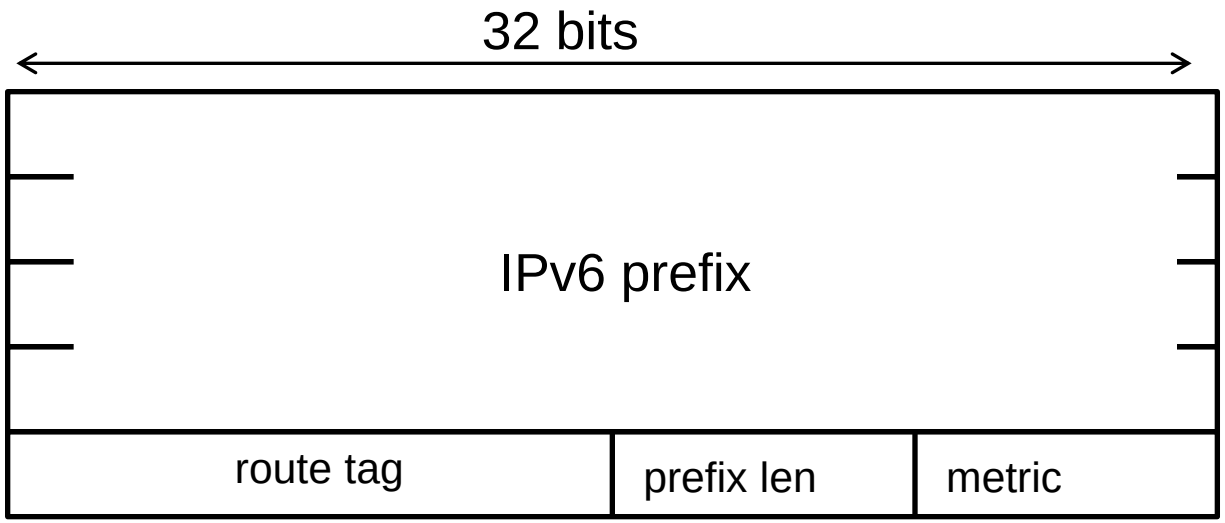
- **request:** c'est une requête pour que le routeur qui reçoit le message transmette sa table de routage.
- **response:** Un message qui contient des informations de routage (table de routage complète ou partielle).

Les message de type **response** sont aussi transmis périodiquement, toutes les 30 secondes. La rfc parle de *Unsolicitated Response Message.*

RIPng datagrammes



RIPng RTE



Le champ *route tag* est inclut pour permettre de distinguer des informations qui proviennent d'autres protocoles de routages.
Le champ metric appartient à [1,15], la valeur 16 = infini

Pour la mise-à-jour de la table de routage, l'adresse de l'émetteur du message est utilisée pour remplir le champ **next hop** de la table de routage.

RIPng RTE

Si le champ **metric = 0xFF**, alors le RTE s'appelle un **next hop RTE**.

Un routeur utilise un next hop RTE pour indiquer que le next hop à inclure dans la table de routage n'est pas l'adresse de l'expéditeur du datagramme UDP mais l'adresse indiquée dans l'entrée RTE.

L'information est valable pour toutes les entrées RTE qui suivent.

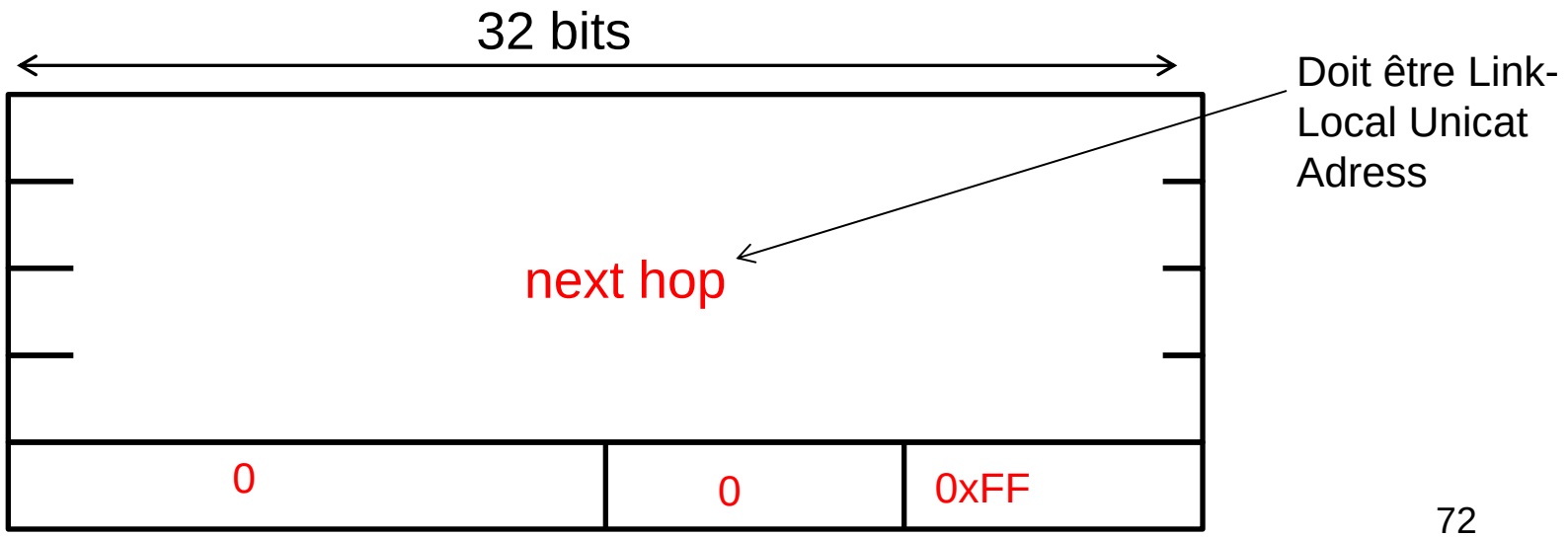


Table de routage

La table de routage gérée par le processus RIPng doit contenir

- Une entrée avec le préfixe ipv6 de la destination.
- Une métrique qui indique le nombre de sauts depuis le routeur vers la destination.
- Un drapeau pour indiquer que la route à été modifiée récemment.
- Des timers pour gérer la validité de l'entrée de la table.

Processus

Le processus RIPng fait appel à deux processus complémentaires:

- Processus d'entrée
- Processus de sortie

Le processus d'entrée est activé à chaque réception de message *response.*

Lorsque le processus d'entrée **met à jour une entrée de la table de routage**, le drapeau de l'entrée correspondante est positionné pour indiquer un **changement récent**.

Le drapeau sera désactivé lorsque l'entrée de la table de routage à été transmise dans un message de type *response* aux autres routeurs.

Processus entrée

Lorsque la table de routage est modifiée avec une entrée *valide* (nouvelle route ou mise-à-jour d'une route valide), l'entrée est associée à un timer, initialisé à 180 secondes. C'est la durée de validité de l'entrée.

Lorsque'une entrée n'est plus valide, soit parce que la métrique = 16 ou que le timer à expiré, le processus entrée initialise un timer ***garbage-collection timer*** à 120 secondes. L'entrée sera supprimée de la table à expiration du timer. Ce timer est utile pour laisser du temps à un autre processus RIPng de communiquer des informations sur cette destination et aussi pour laisser du temps au processus de sortie de transmettre l'information modifiée.

Après chaque modification de la table de routage le processus d'entrée sollicite le processus de sortie (triggered updates) pour qu'il transmette les modifications aux autres routeurs.

Processus sortie

Le processus de sortie est activé

- Par le processus d'entrée sil reçoit un message *request* d'un autre routeur.
- Par le processus d'entrée s'il reçoit une *triggered updates.*
- Régulièrement, toutes les 30 secondes.

Le drapeau qui indique qu'une entrée de la table vient d'être modifié permet au processus de sortie de limiter la tailles des données transmises. En particulier, après une triggered updates, seules les entrées nouvellement modifiées sont transmises (et le drapeau désactivé).

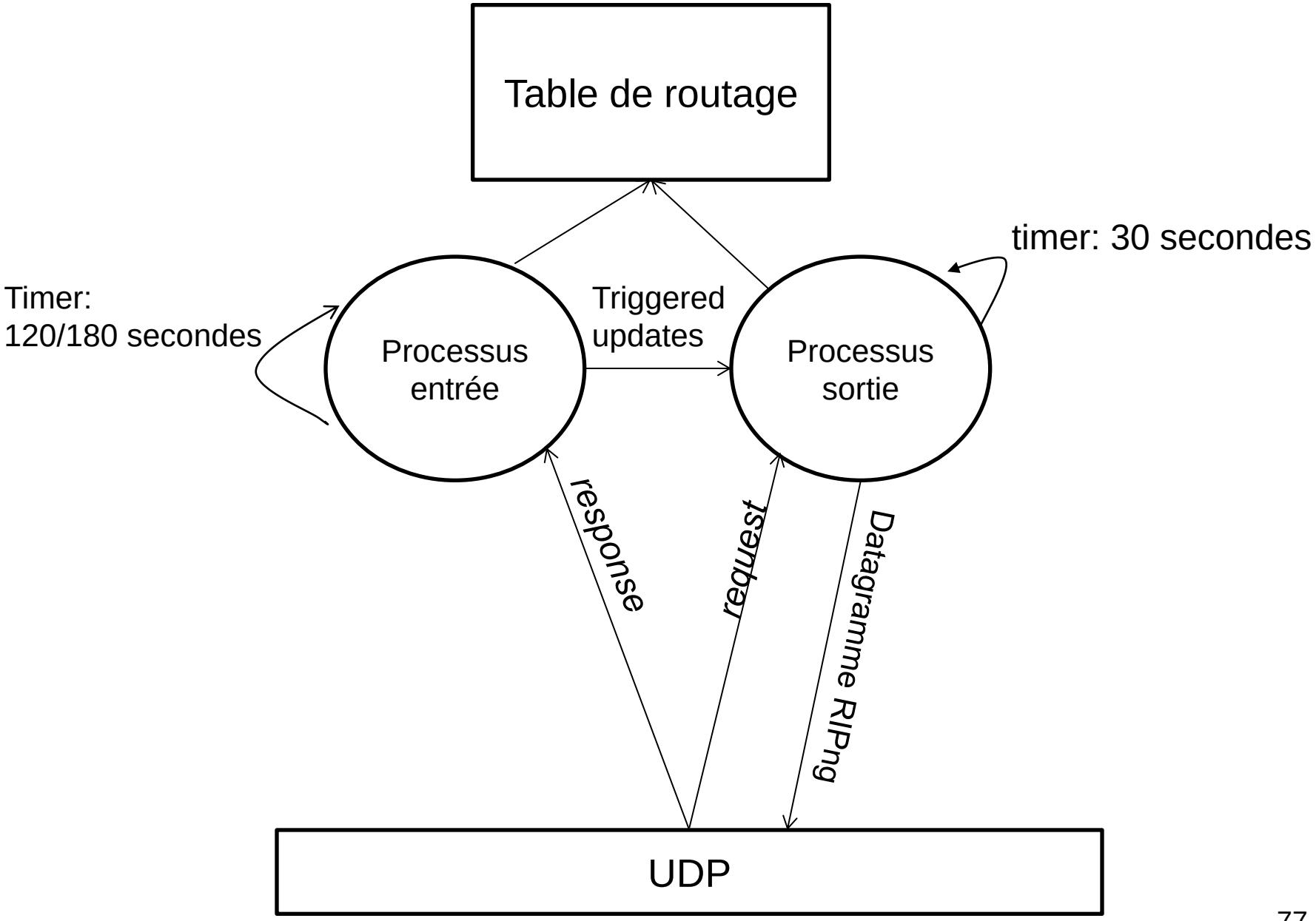


Table de routage
timer: 30 secondes
Timer:
120/180 secondes
Processus
entrée
Triggered
updates
Processus
sortie
response
request
Datagramme RIPng
UDP

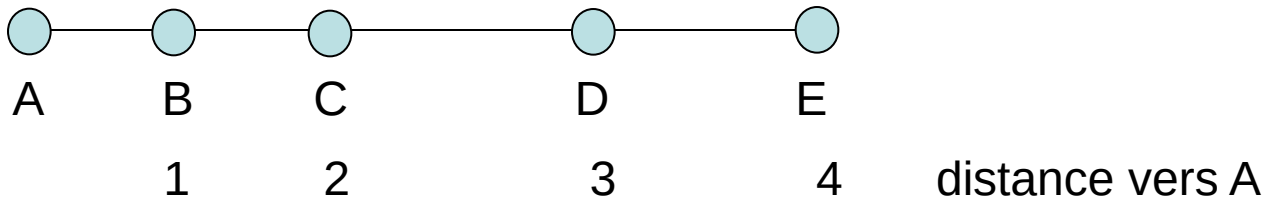
Remarques

Le processus d'entrée est susceptible de modifier plusieurs entrée de la table de routage successivement. Après chaque modification il transmet un signal *triggered updates* au processus de sortie pour qu'il transmette un datagramme RIPng.

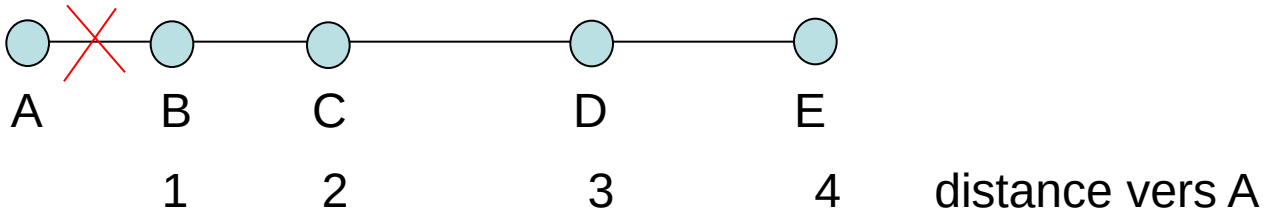
Pour éviter que le processus de sortie transmette plusieurs datagrammes dans un trop court laps de temps, il est recommandé qu'il attende 5 secondes avant de transmettre le datagramme. Si pendant les 5 secondes une nouvelle notification survient alors il modifie le datagramme et attend 5 secondes a nouveau.

Remarques

Dans certaines situations les algorithmes à vecteur de distance convergent lentement. Un exemple est :



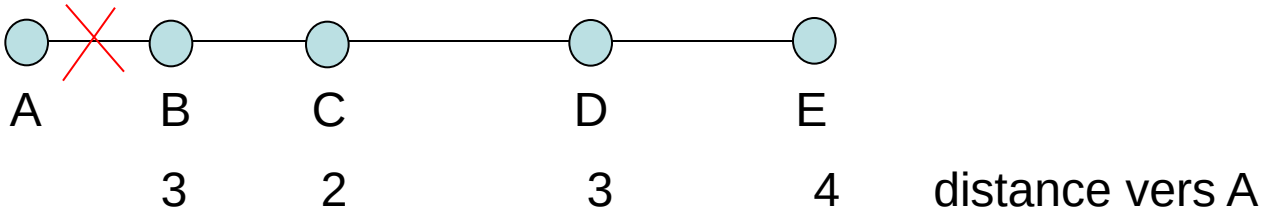
Les routeurs sont en ligne, et les distances sont les distance des plus courts chemins vers A. La métrique est le nombre de sauts. Supposons que le lien entre les routeurs A et B n'existe plus.



Remarques

Le routeur B ne reçoit plus d'information de A. Il perd le chemin de longueur 1 vers A.

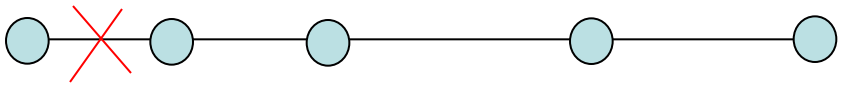
Quand il reçoit le datagramme de C, il apprend qu'il y a un chemin vers A qui passe par C. La longueur de ce chemin est 2 + 1 = 3



Remarques

Ensuite, les routeurs continuent à s'échanger des messages.

C apprend qu'il y a un chemin vers A qui peut-être soit via B soit via D, dans les deux cas il est de longueur 3 + 1 = 4, etc.



A	B	C	D	E
	3	4	3	4
	5	4	5	4
	5	6	5	6
	7	6	7	6
	…	…	…	…
	∞	∞	∞	∞

Dans la littérature ce phénomène porte le nom de *couting to infinity.*

Remarques

Les distances vers A vont converger vers l'infini, **ce qui est correcte.**

Le problème c'est la vitesse de convergence (observer que RIPng limite à 15 la distance).

Pour pallier à ce problème, les routeurs utilisent le mécanisme **d'horizon partagé ou éclaté.** L'idée est qu'on ne transmet pas une route à un routeur par lequel passe la route (l'adresse du routeur est celle du next hop). Dans notre exemple, C ne transmet pas la route vers A à B puisque B apparait comme le next hop dans sa table de routage.

Remarques

Le mécanisme **d'horizon partagé ou éclaté avec empoisonnement** consiste à transmettre des valeurs infinies aux routeurs next hops.

Dans notre exemple, C transmet la valeur infinie à B et la convergence est immédiate.

RIPng utilise le mécanisme d'horizon partagé avec empoisonnement, la valeur 16 étant l'infini.

Routage par état des liens

Routage par information d'état de lien

Le **routage par information d'état de lien** ou **routage par état de lien** a été introduit pour palier à la lente convergence de l'algorithme par vecteur de distance.

Les idées de bases de l'algorithme sont:

1. découvrir les routeurs voisins et leur adresse réseau
2. calculer le délai (coût) pour atteindre chaque voisin
3. construire un paquet résumant les informations découvertes
4. transmettre le paquet aux autres routeurs
5. calculer le plus court chemin vers chaque routeur

Découverte des voisins

Contrainte: chaque routeur possède un identificateur unique.

Lorsqu'un routeur démarre, il s'annonce aux routeurs voisins en transmettant un paquet de contrôle HELLO (sur chaque ligne point-à-point ou réseau broadcast - Ethernet). Les routeurs répondent en s'identifiant.

Mesure du coût de la ligne exemple

Pour mesurer le délai d'acheminement vers chacun de ces voisins un routeur peut transmet un paquet de contrôle ECHO qui doit être retourné le plus rapidement par le routeur destinataire. Le routeur mesure le délai aller-retour et obtient ainsi une estimation du délai de transmission. Cette méthode suppose les délais symétriques.

Pour transmettre le paquet ECHO, le routeur le place dans la file d'attente des paquets à transmettre. Le routeur peut donc comptabiliser le temps d'attente dans la file d'attente et ainsi prendre en compte la charge du réseau. Il peut aussi ignorer la charge du réseau et mesurer seulement le délai à partir du moment ou le paquet est effectivement transmis.

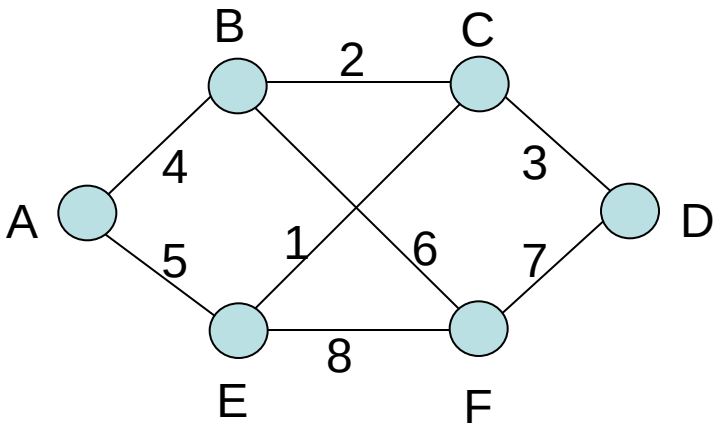
Paquet d'état des liens

Une fois les informations recueillies par le routeur elles sont transmises aux autres routeurs. Ce paquet contient:

1. l'identité du routeur émetteur
2. un numéro de séquence
3. l'âge du paquet
4. la liste des voisins directes ainsi que le délai (coût) estimé pour l'atteindre.

Ces paquets peuvent être transmis périodiquement et/ou après les événements importants dans le réseau tels que détection d'une panne, établissement d'un nouveau routeur, changement significatif dans la mesure du coût de la ligne, …

Paquet d'état des liens



A	
Séqu.	
Âge	
B	4
E	5

B	
Séqu.	
Âge	
A	4
C	2
F	6

C	
Séqu.	
Âge	
B	2
D	3
E	1

D	
Séqu.	
Âge	
C	3
F	7

E	
Séqu.	
Âge	
A	5
C	1
F	8

F	
Séqu.	
Âge	
B	6
D	7
E	8

Distribution des paquets d'état de lien

L'idée de base est distribuer les paquets par **inondation**.

Chaque routeur mémorise le couple paquet + numéro de séquence.

Un paquet est retransmis (inondation) seulement si c'est un nouveau paquet.

Si le numéro de séquence est inférieur au numéro le plus grand reçu, il est considéré obsolète et ignoré.

Distribution des paquets d'état de lien

L'utilisation du numéro de séquence pose quelque problèmes:

1. Lorsque le numéros de séquence passe de la valeur maximale à 0, tous les paquets vont être ignorés
2. Si un routeur tombe en panne, après réinitialisation il retransmet des paquets avec des numéros de séquence 0, 1, 2, … qui seront ignorés
3. si une erreur de transmission modifie le numéro de séquence de 4 à 1028 (1 bit faux) tous les paquets 5, 6, …, 1028 vont être ignorés.

Le champ âge indique en secondes la durée de vie du paquet (dans le routeur récepteur). Ce champ est régulièrement décrémenté et lorsque sa valeur est nulle le paquet est détruit.

Ainsi l'effet d'une erreur dans un paquet d'état de lien à un effet fini dans le temps.

Calcul des routes

Une fois qu'un routeur dispose d'une image globale du sous-réseau de communication, c'est-à-dire l'information relative à tous les liens, il peut appliquer l'algorithme de Dijkstra pour déterminer le plus court chemin vers tous les routeurs.

OSPF

Open Shortest Path First

OSPF

Le protocole OSPF est le protocole standard retenu par l'IETF pour le routage par état des liens.

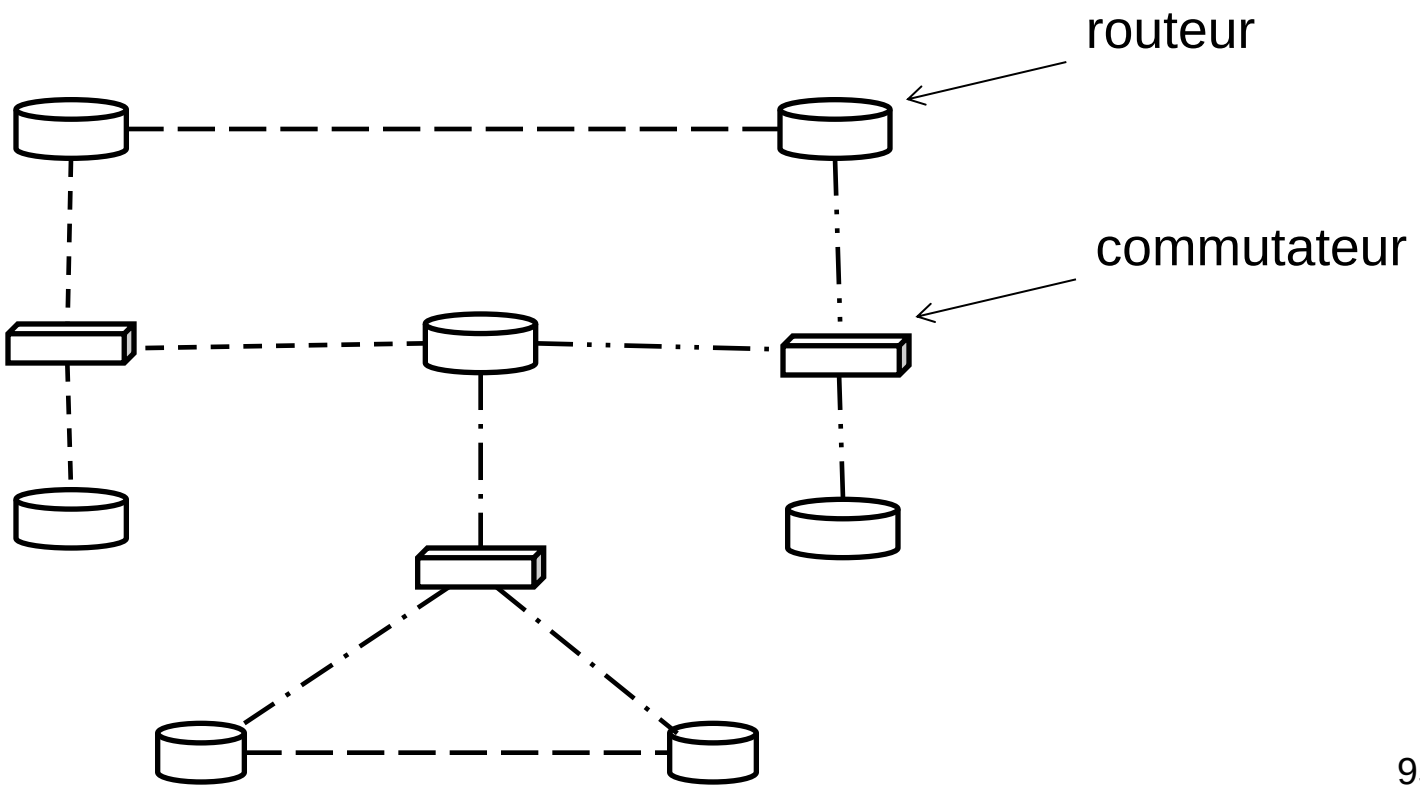
Les datagrammes OSPF sont transmis comme des datagrammes IPv6 en utilisant le champ *next header = 89.*

Les adresses multicast (Well-Known) ff02::5 et ff02::6 permettent d'envoyer un datagrammes à tous les routeurs sur le même lien.

Topologie

Les routeurs sont interconnectés (incomplet)

- Par des liaisons point à point (protocole PPP, HDLC, …)
- Via un réseau broadcast (Ethernet,…)



Hello

Les routeurs transmettent périodiquement des **paquets Hello** pour connaître leurs voisins.

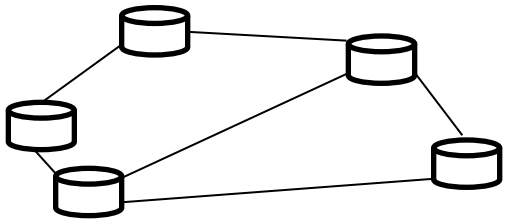
Ces paquets sont transmis avec une adresse destination multicast ff02::5.

Ils contiennent les informations sur le routeurs

- Identificateurs du routeur, de l'aire, de l'interface
- La priorité du routeur
- Le designated routeur

Graphe de communication

OSPF détermine un graphe de communication dont les sommets sont les routeurs. Il y a une arête entre deux sommets si les routeurs échangent des informations de routage.

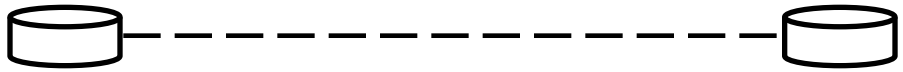


Les routeurs sont connectés via une interface à un autre routeur. Cette interface est associée à une adresse Link-Local Unicast Address (fe80::/10).

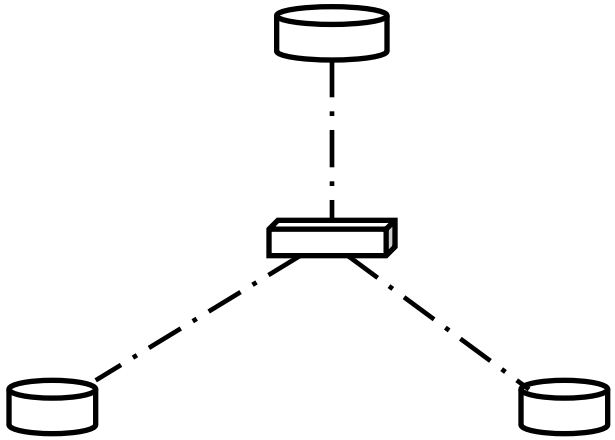
Graphe de communication

L'idée est que deux routeurs doivent être adjacents dans le graphe de communication s'ils communiquent directement.

Quand on a une liaison point-à-point les graphes de communication sont le même que les communications réelles.

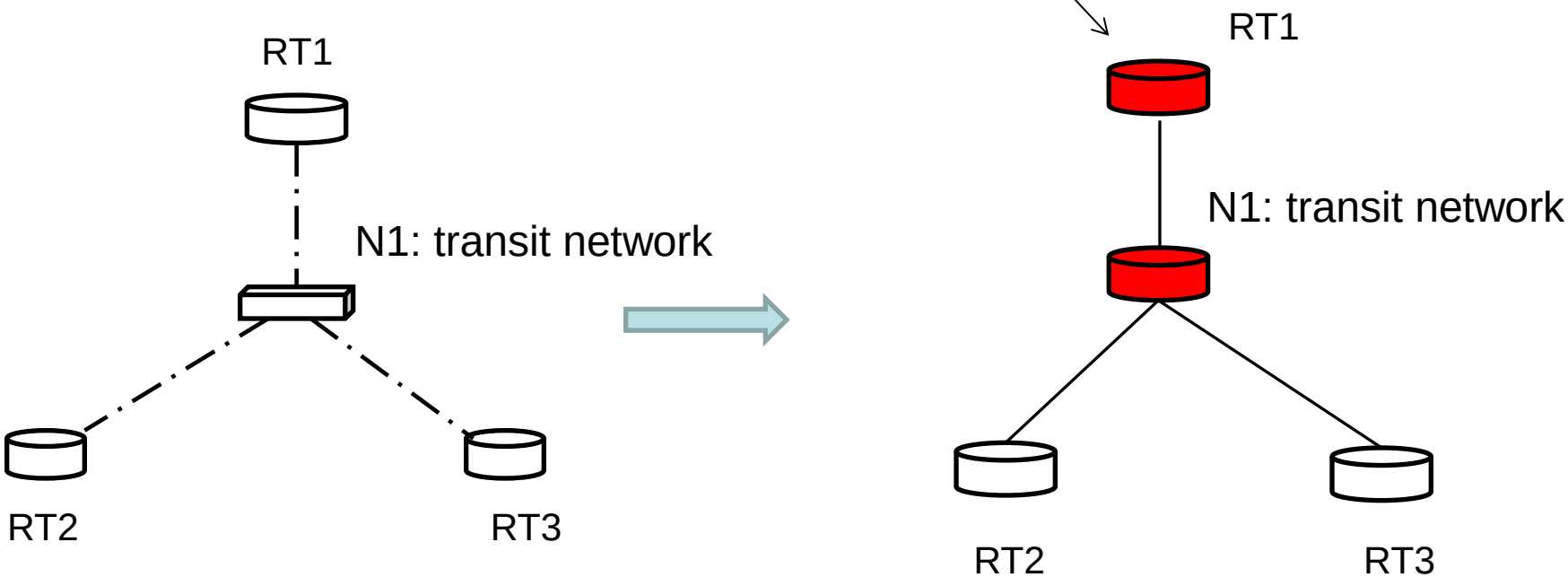


Par contre si les routeurs sont connectés à un même réseau de type broadcast (Ethernet) il faut clarifier le graphe de communication.



Designated Router

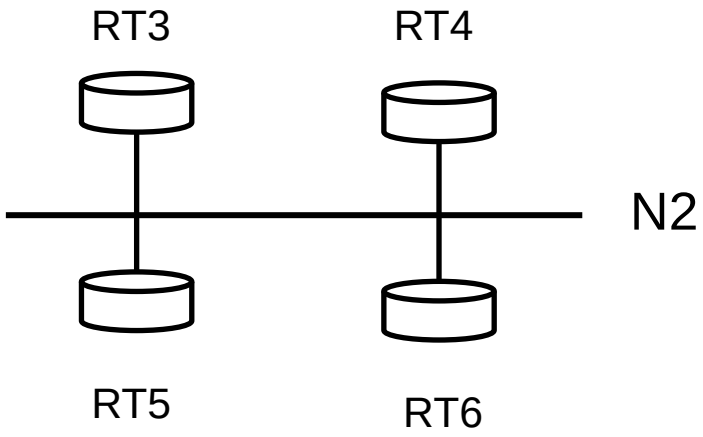
OSPF propose d'élire un routeur ***Designated Router (DR)***, c'est ce routeurs qui sera connecté à tous les autres.



Designated router

Un réseau est un **transit network** si plusieurs routeurs y sont connectés. Pour le graphes de communication, un transit network est vu comme un nœud (actif). Il émet des paquets LSA (Link State Advertisement). En fait c'est le Designated Router qui jour le rôle du réseau.

Le coût du chemin d'un réseau de transit à un routeur est **nul**.

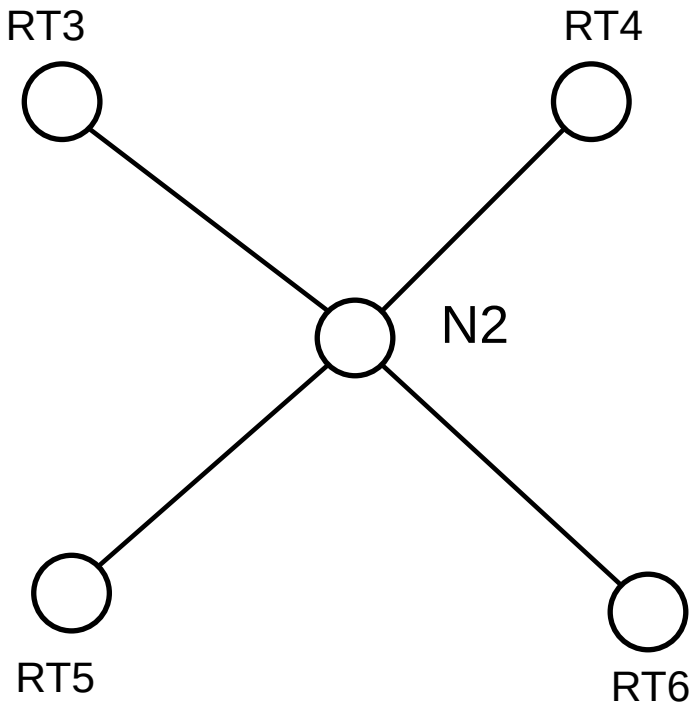


de

vers	RT3	RT4	RT5	RT6	N
RT3					X
RT4					X
RT5					X
RT6					X
N	X	X	X	X	

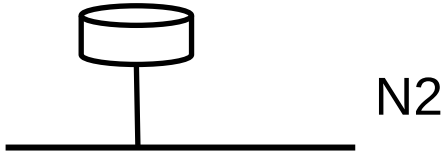
Graphe d'adjacence

Designated Router



Réseau Stub

Un réseau connecté à un seul routeur est un **Stub Network.** Il apparait comme une feuille du graphe,



vers \ de		RT3	N2
	RT3		
	N2	X	

Graphe d’adjacence

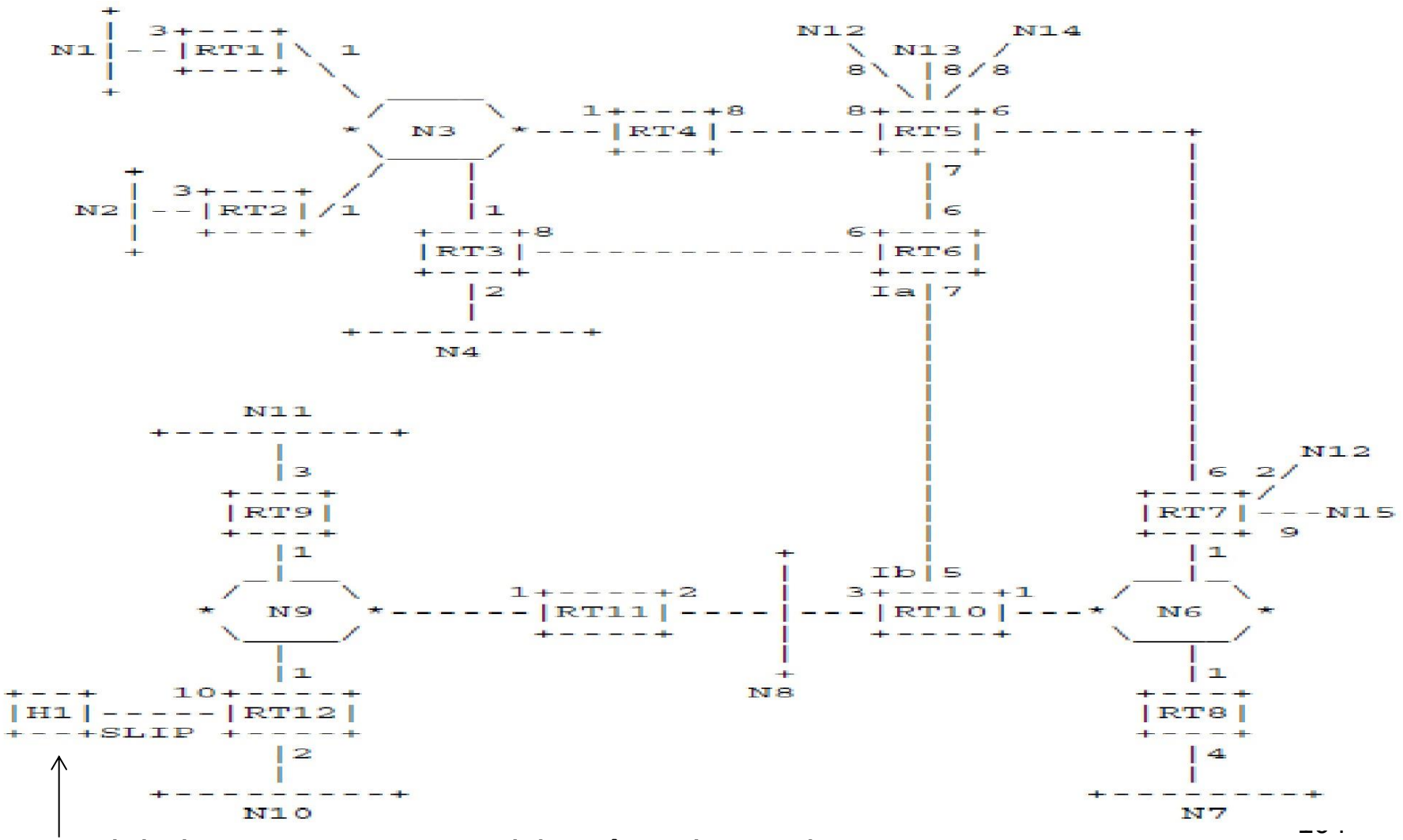
Exercices

Le réseau sur la slide suivante vient de la rfc 2328.

- Quels sont les réseaux Stubs, transits?
- Faites la liste des émetteurs de LSAs (Link State Advertisements).
- Représentez le graphe d'adjacence sur la forme d'une table avec les colonnes qui correspondent au LSA émis et le lignes les destinations. A l'intersection des lignes/colonnes vous notez le coût du lien.
- Dessinez l'arbre SPF pour le routeur RT6.

Remarque oubliez Ia et Ib ce sont des interfaces. La liaison entre RT6 et RT10 se fait par une connexion *numbered,* c'est-à-dire que la connexion possède une adresse IP différente du routeur (voir p. 19 rfc 2328). Elles n'émettent pas de LSA mais peuvent être destinations.

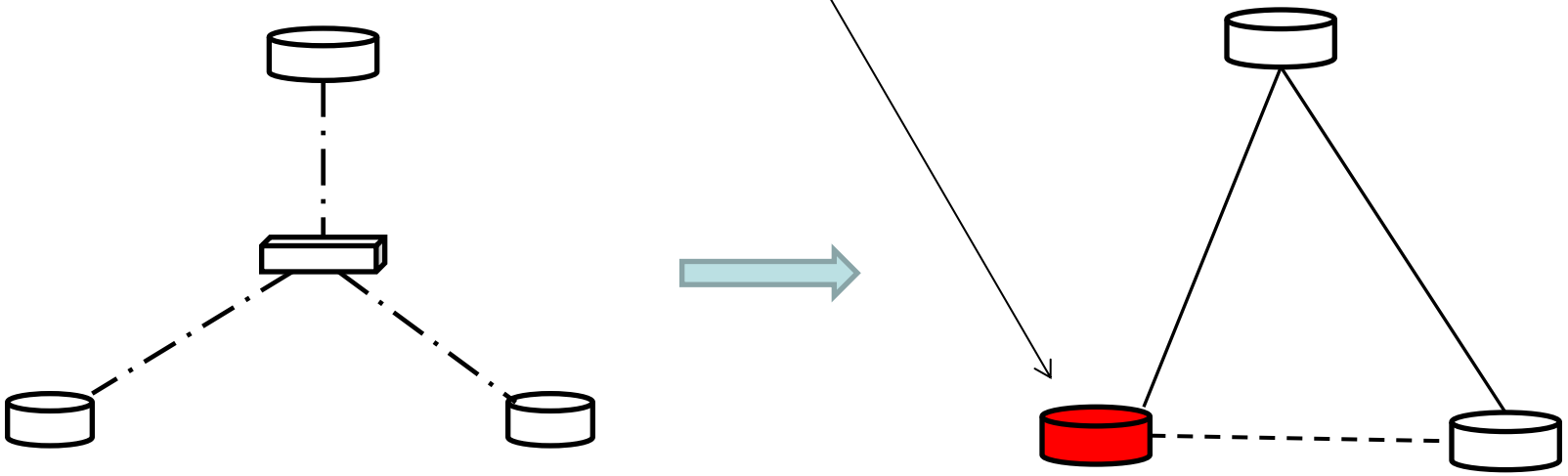
Exercices



Serial Line Internet Protocol (ancêtre de PPP)

Backup Designated Router

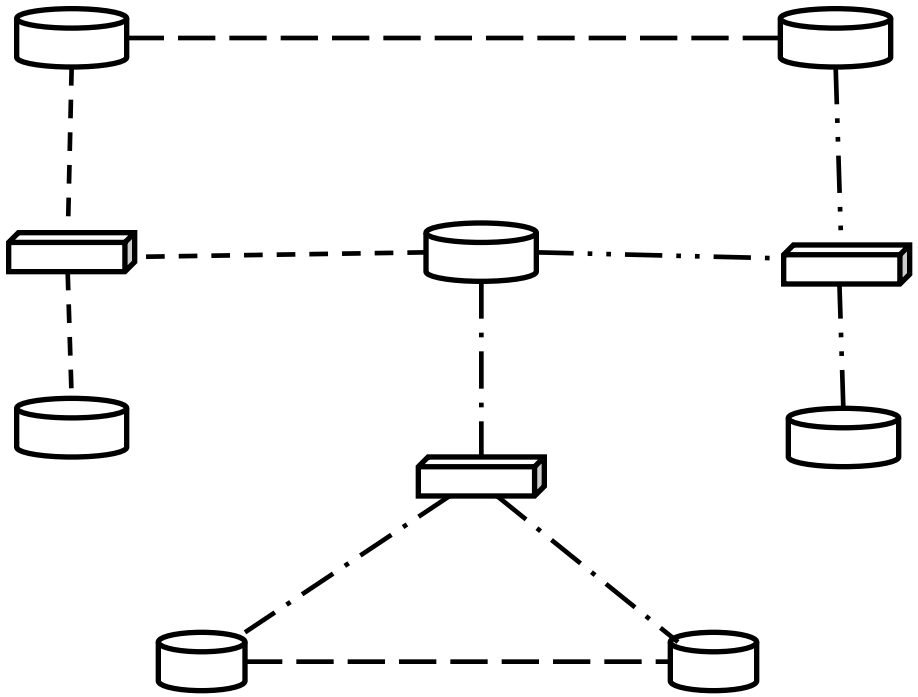
Un ***Backup Designated Router (BDR)*** est aussi élu. L'idée est qu'il va collecter toutes les informations pour être DR en cas de panne.



Le BDR va transmettre beaucoup moins d'information sur ces liens que le DR. Le but est de ne pas pénaliser le trafic.

Exercice

En choisissant par hasard les DR et BDR, dessinez le graphe de communication du réseaux ci-dessous.



Types de paquets

Les routeurs échangent 5 types de paquets:

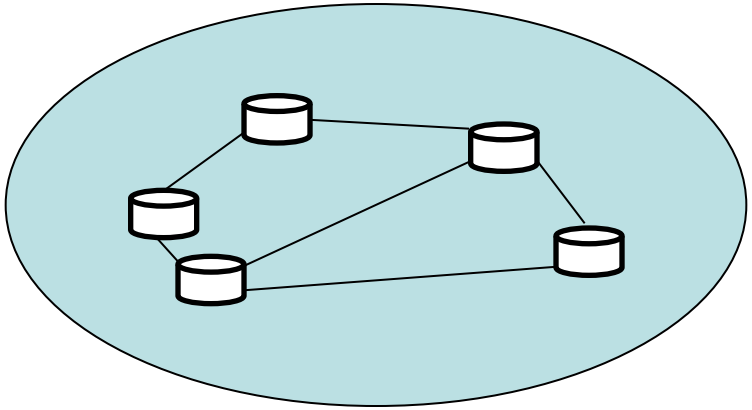
- Les paquets HELLO, pour découvrir les voisins et construire le graphe de communication.
- Les paquets DBD synchronisation des Bases de Données des routeurs (Database Decriptor packet), pour échanger les liens connus.
- Les paquets LSR, Link State Request packet, après réception d'un DBD un LSR permet de demander des informations sur un lien pas à jour.

Types de paquets

- Les paquets LSU (Link State Update) sont utilisé pour le flooding des LSA (Link State Advertisement)
- Les paquets LSAck (Link State Aknowledment) pour acquitter la réception des LSAs.

Zones de routage

Système autonome



Le protocole OSPF peut s'appliquer à tous les routeurs du système autonome. Dans ce cas, tous les routeurs ont leurs base d'état des liens identiques.

Si le nombre de routeurs est très grand alors il est possible de diviser le système autonome en zone (**Area)**. On a alors 2 algorithmes de routage **intra-area** et **inter-area routing.**

Zone de routage

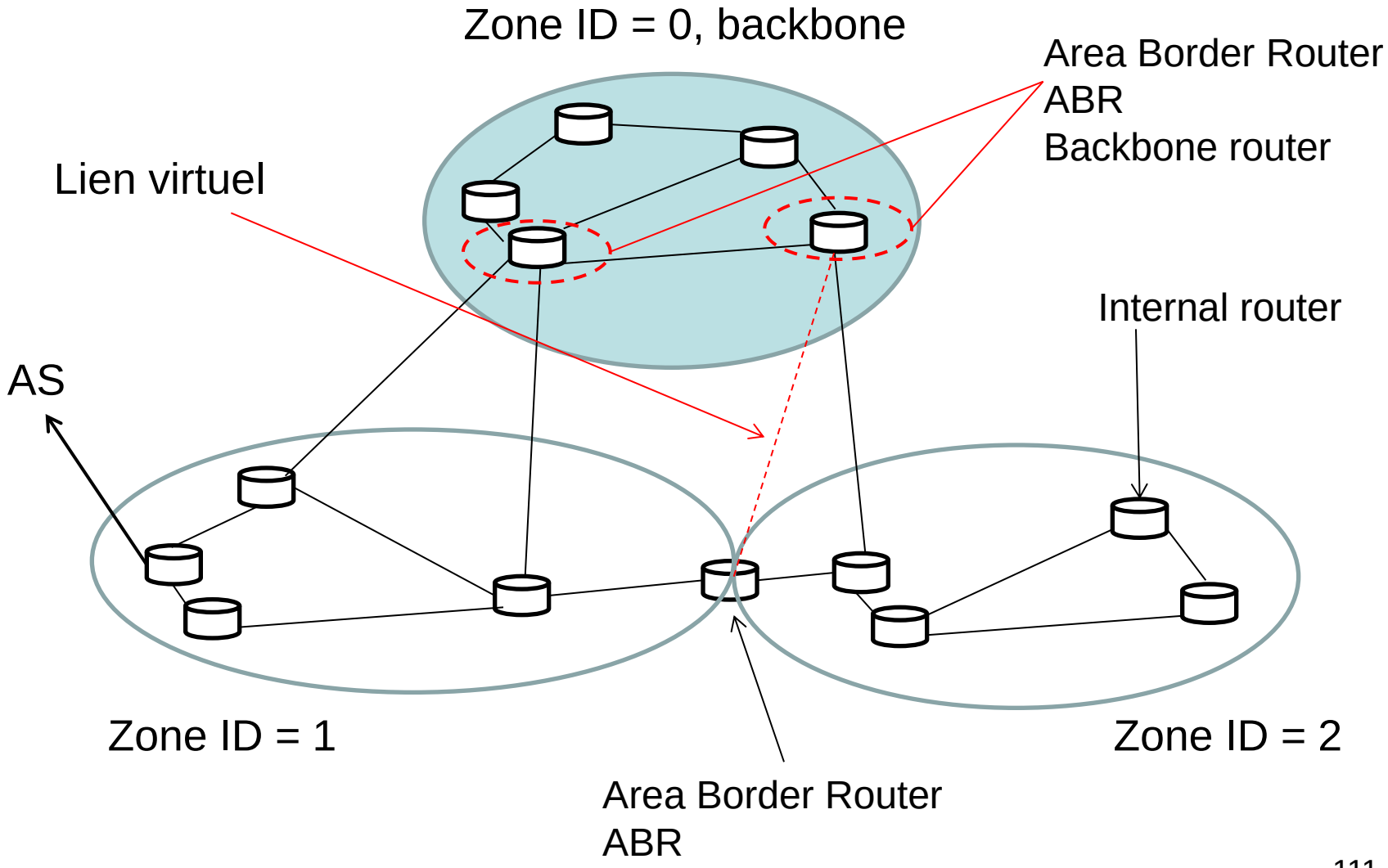
Une zone contient des réseaux contigus et les routeurs qui ont des interfaces qui les connectent.

Par contigus on veut dire que les hôtes du même zone communiquent sans sortir de la zone.

Pour chaque zone les routeurs exécutent une version de l'algorithme de routage par état des liens. Ils n'utilisent pas d'informations provenant d'autres zones. La topologie d'une zone est donc opaque pour les autres zones.

Un routeur qui a des interfaces dans différentes zones possède une base de donnée d'état des liens par zone à laquelle il est connecté.

Topologie


Zone ID = 0, backbone
Area Border Router
ABR
Backbone router
Lien virtuel
Internal router
AS
Zone ID = 1
Zone ID = 2
Area Border Router
ABR

Types de routeurs

On a 4 types de routeurs

- Backbone router : au moins une interface dans la zone 0
- Internal router : toutes ses interfaces dans la même zone
- Area border router : relié à plusieurs zones, exécute une copie de l'algorithme de routage par état des liens par zone.
- Autonomous system border router : connect la zone à un AS.

Backbone

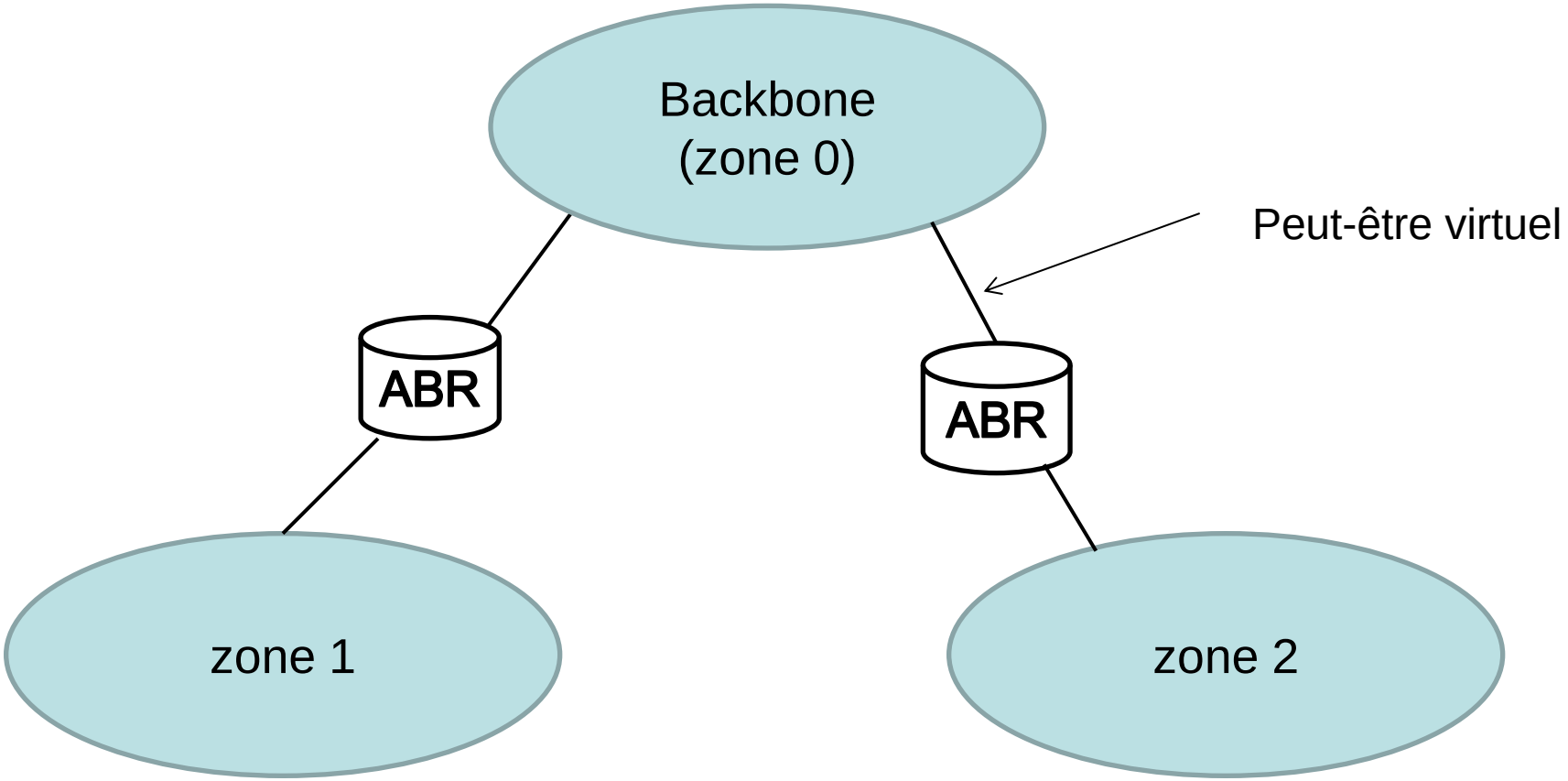
Les zones sont déterminées par l'administrateur du système.

La zone 0 est le backbone du réseau, les informations de routage d'une zone A vers une zone B transitent toujours par le backbone.

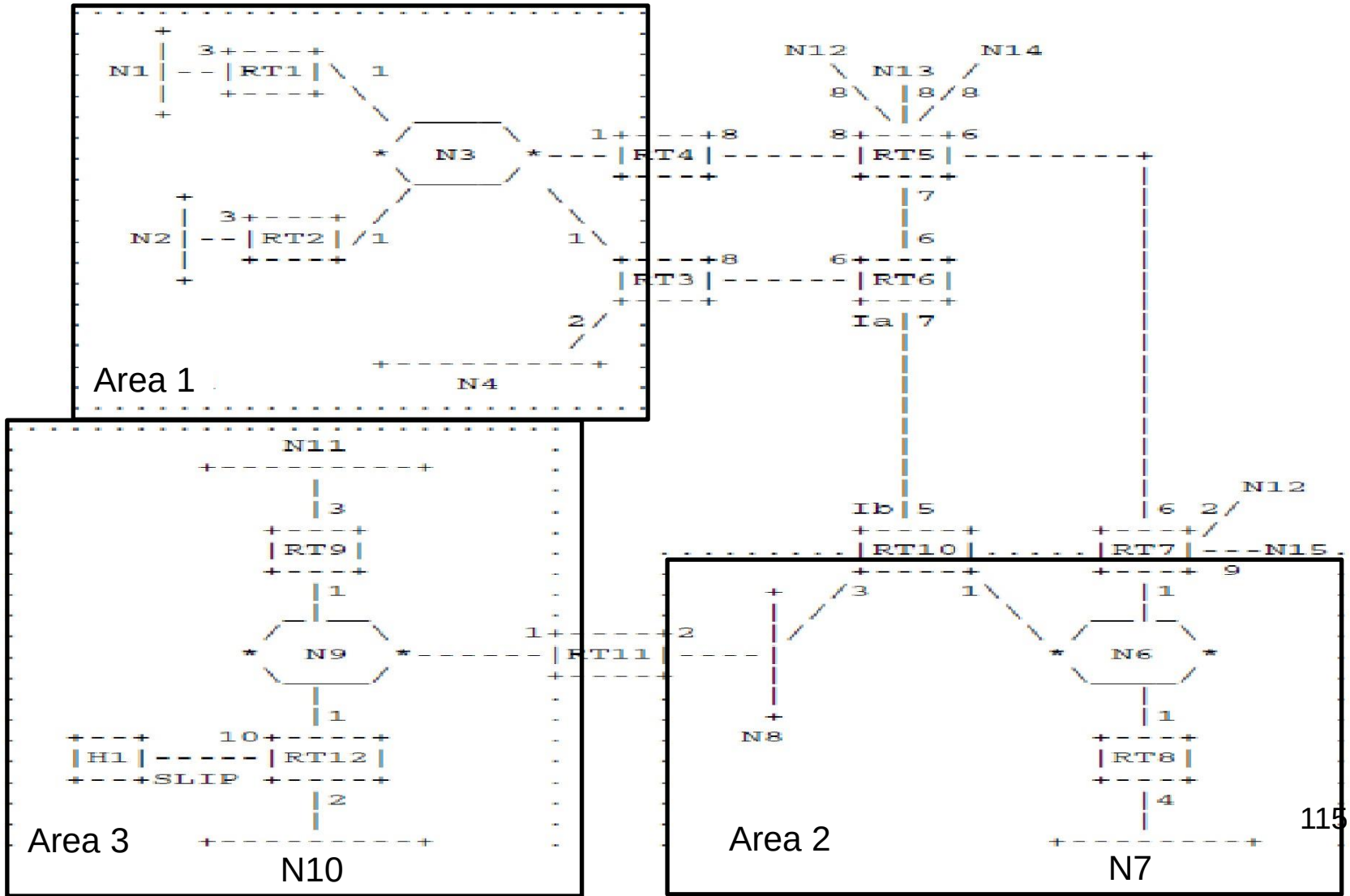
Ce sont les Area Border Router qui diffuse l'information à l'extérieur de la zone au backbone. Tous les ABR appartiennent au backbone.

Si l'ABR n'est pas physiquement relié au backbone alors l'administrateur introduit un lien virtuel (virtual link). On peut définir un lien virtuel entre deux routeurs qui ont une interface dans une même zone. Le poids du lien virtuel est le poids du chemin intra-zone.

Topologie en Hub



Exemple


Area 1
N1
RT1
N2
RT2
N3
N4
RT4
RT3
N12
N13
N14
RT5
RT6
Ia
Ib
RT10
RT7
N15
N11
RT9
N9
RT11
H1
SLIP
RT12
N10
Area 3
N8
N6
RT8
N7
Area 2

Zone 1

de

	RT1	RT2	RT3	RT4	RT5	RT7	N3
RT1							0
RT2							0
RT3							0
RT4							0
RT5			14	8			
RT7			20	14			
N1	3						
N2		3					
N3	1	1	1	1			
N4			2				
N6			16	15			
N7			20	19			
N8			18	18			
N9-N11-H1			29	36			
N12					8	2	
N13					8		
N14					8		
N15						9	

v
e
r
s

Remarques

RT11 appartient aux zones 2 et 3. C'est donc un Backbone router. Pour le connecter aux autres routeurs du backbone un **lien virtuel** est ajouté entre RT11 et RT10 (entre RT11 et un autre routeur qui appartient à une même zone que lui).

RT11 a été configuré pour annoncer une unique route vers les destinations à l'intérieur de la zone 3.

ABR

Les informations de routage transmises par l'ABR à l'extérieur de la zone (non backbone) à laquelle il appartient ne donne pas le détail de la topologie interne de la zone.

Les noms de réseaux appartenant à la zone sont publiés ainsi que la distance intra-zone du routeur ABR au réseau. Quelque fois une distance peut-être publiée pour tous les réseaux (voir RT11 dans l'exemple).

Stratégie pour le routage

Les routeurs qui appartiennent à une même zone se partagent les informations et connaissent la topologie de la zone à laquelle ils appartiennent.

Pour router un datagrammes en-dehors de la zone, il est d'abord transmis à un ABR de la zone (c'est-à-dire au backbone) qui connait la topologie du backbone et envoie le datagrammes à un ABR qui fait partie de la même zone que la destination.

Finalement, le dernier ABR transmet le datagramme vers la destination en utilisant le routage intra zone.

Stratégie pour le routage

Chaque routeur connait les routeurs qui connectent le système autonome à un autre système autonome et les routeurs ABR qui y mène (ainsi que le coût).