

Acheminement des données dans Internet : routage (présentation succincte)

Session de Remise à Niveau, Télécom Paris
Août-septembre 2021

Ken CHEN
ken.chen@univ-paris13.fr

Plan

+ Routage : généralités

- Circuit virtuel
- Routage des « *datagrams* »

+ Structure internet

- Système autonome

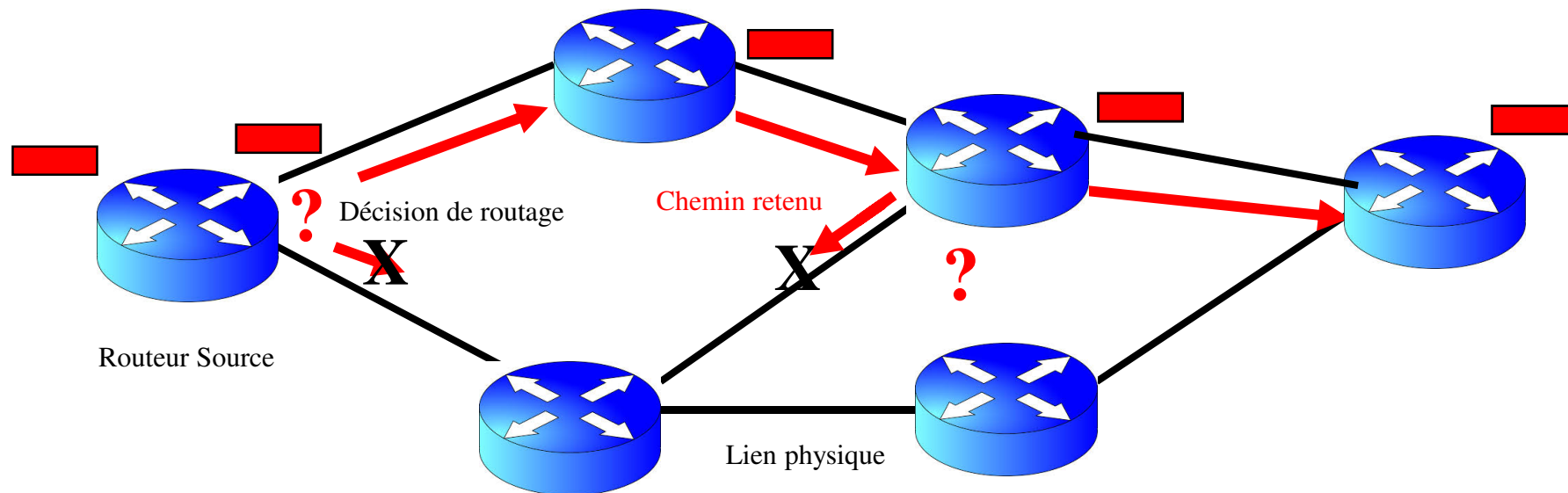
+ Routage et ses algorithmes

- Vecteur de distance (DV), Etat des liaisons (LS), connexité

+ Protocoles de routage

- *présentation succincte : RIP, OSPF*

Routage datagram (**Internet**)



Décisions prises par **chaque** routeur pour **chaque** paquet,

Décisions **locales concordantes** conduisant à **un** chemin **optimal global**

Routage

Routage

- + Contexte : Réseau maillé
- + Problème :
 - Trouver un chemin entre deux extrémités
- + Optimisation
 - Réseau maillé : chemins multiples
 - Critère pour choisir **un** chemin optimal
- + Deux problèmes
 - Etablissement de chemin proprement dit :
 - + **Routage**
 - Opération d'avancement de paquets
 - + **Forwarding, commutation**

Forwarding vs Routage

+ Acheminement (*forwarding*) : *Data plane*

- Opération « mécanique » et locale
- Dicté par une table de commutation (**FIB**)
 - + FIB=Forwarding Information Base (issue de RIB)
 - + **Point critique : nombre d'entrées dans FIB**
- Aiguillage vers une sortie selon un critère donné
 - + Couple <critère, sortie>
 - + Critère: Adresse de **destination** (ou référence à un **circuit virtuel [VC]**)

+ Routage (*routing*) : *Control Plane*

- Processus de construction de tableau de routage (**RIB**)
 - + RIB = Routing Information Base
- Dissémination des informations
 - + Via des échanges de messages (Protocoles de routage)
- *Base théorique : théorie des graphes, systèmes distribués*

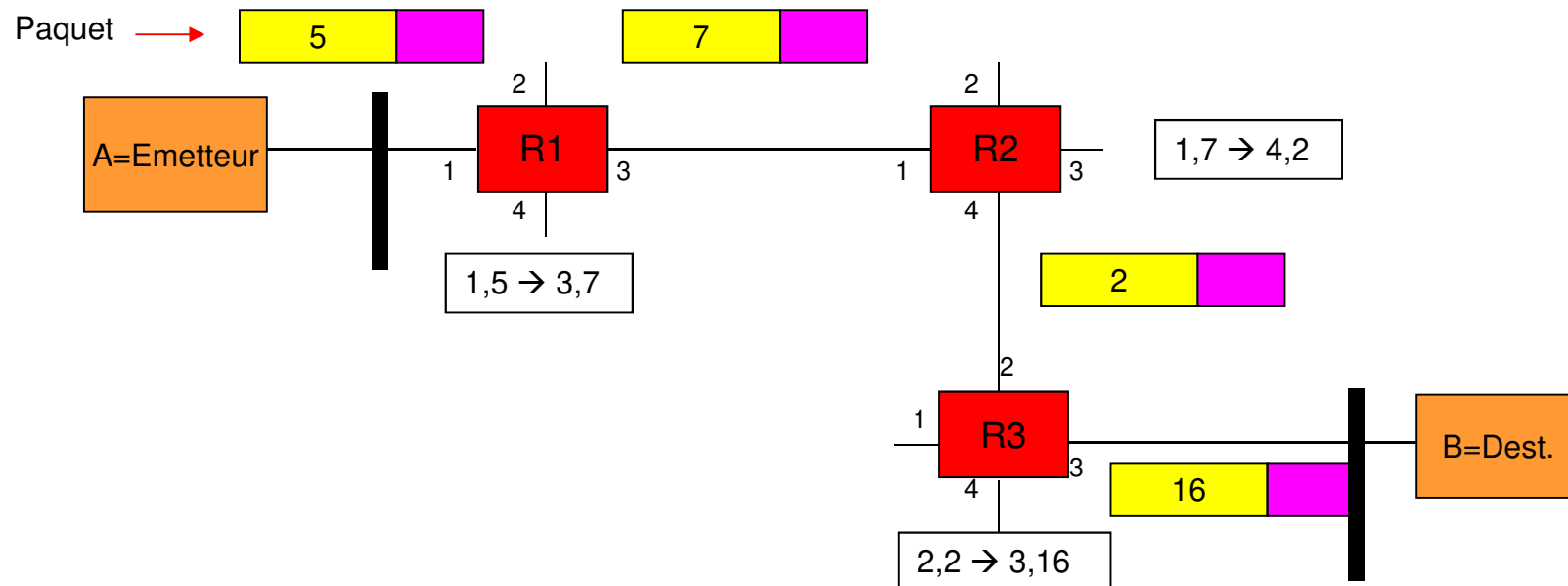
Forwarding : variantes

- + Gestion d'un circuit virtuel (VC) (*e.g. MPLS = support de IP*)
 - Routage établi à l'établissement
 - + Généralement **stable** dans le temps (réservation de ressource)
 - Liste de correspondance <label VC, sortie>
 - + Technique identique à la **commutation par label**
 - Souvent accompagné de la réservation des ressources
- + *Forwarding* des datagrammes (**mode pris par IP**)
 - Opération par paquet
 - + Pas intrinsèquement compatible avec réservation de resssources
 - Processus en parallèle d'établissement de Table de routage
 - Fluctuation du trafic (risque de congestion)
- + *Autres approches*
 - *Source Routing*
 - + *Exemple : IP avec indication du chemin dans le paquet*
 - + *Problème connexe : découverte de la route*
 - « *Hot potato* » : *aiguillage au hasard (réseau très dense avec tampon très limité)*

Commutation VC

- + Etablissement de connexion
 - Construction de la route
 - + Usuellement au fur et à mesure que le paquet Setup avance
 - + Réserve de ressource
 - + Un label est attribué localement <VC, lien> pour chaque segment
 - Accord entre deux routeurs voisins
 - Identification du VC à deux **interfaces d'accès** au réseau, **et de manière indépendante**, par une identification <flux, lien entrée/sortie>
- + Chaque paquet porte un label (identification du VC)
 - **Changement de label** d'un routeur à un autre
- + Opération routeur/commutateur
 - recherche la sortie dans la table de correspondance à l'aide du label
 - Changement de label (nouveau label reconnu le suivant)
 - Aiguillage vers la sortie

VC : schéma (e.g. *MPLS*)



Le même VC: pour A, c'est VC #5, pour B, c'est VC #16

Routage dans Internet

Internet et routage

- + Réseaux étendus et fortement maillés
- + Unité d'identification : réseau (prefix=network id)
 - un **seul** niveau d'agrégation <réseau, station>
 - Beaucoup de réseaux (de tailles très dispersées)
 - + 799213 réseaux (30/09/2019) (438725 agrégés)
- + Réseaux avec une topologie dynamique
 - Internet = un immense « parc » des routeurs
 - Pas de gestion centralisée
 - Les nœuds naissent et disparaissent
- + Communication en mode « Datagram »
- + Communications **unidirectionnelles** (SA -> DA)

Routage direct vs indirect (IP)

+ Routage directe

- Pour deux stations qui sont sur le même segment
- Avec ARP, le lien est établi directement

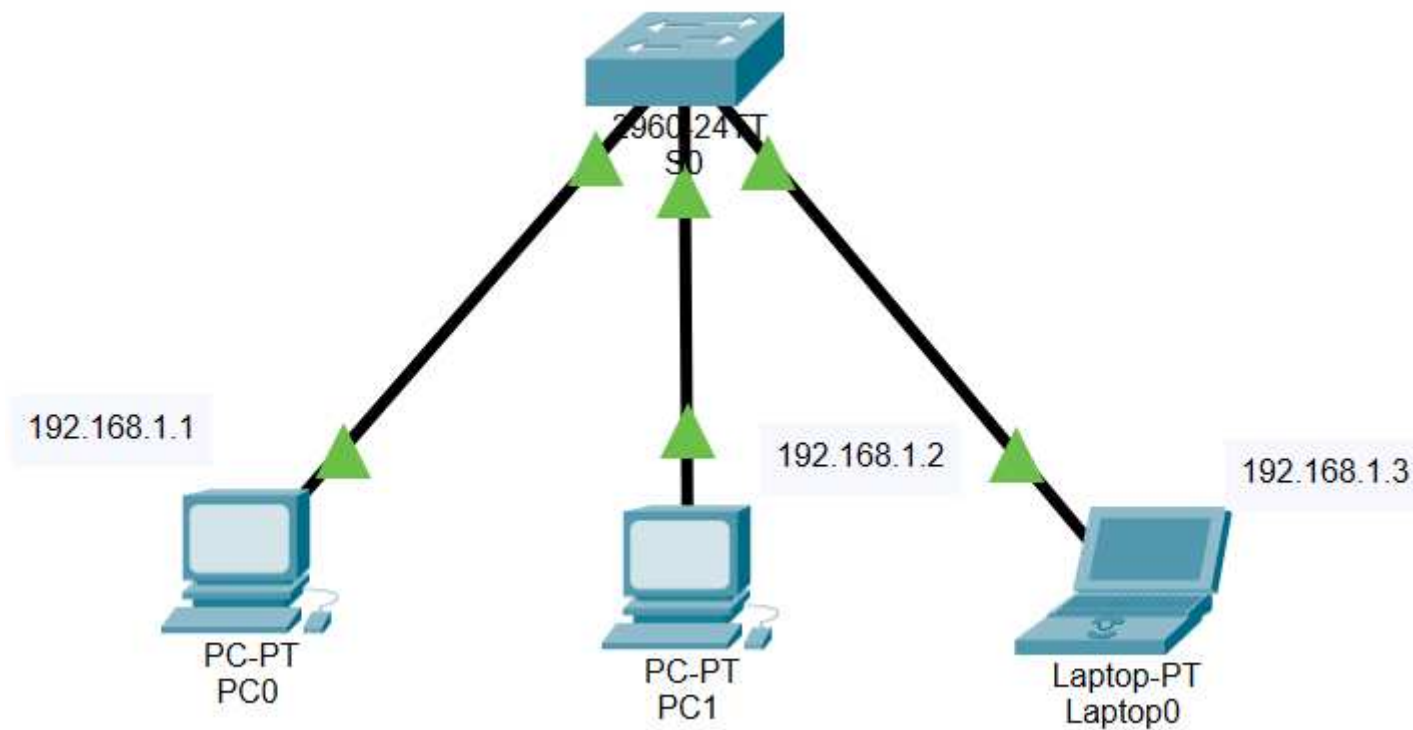
+ **Routage indirecte**

- La destination n'est pas sur le même segment
- Routage assuré par ***au moins un routeur***

+ Table de routage

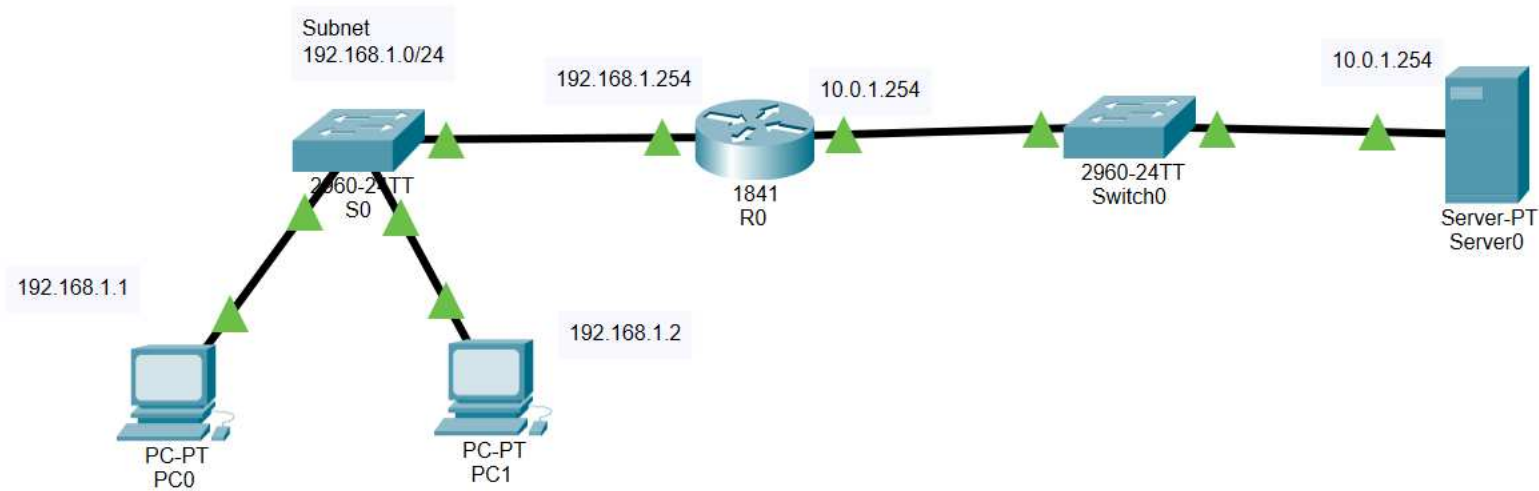
- Chaque ligne d'une table contient :
 - + Id. Réseau destination, Id. Réseau Prochain nœud, Id Lien de sortie
- Existence d'une ligne « *par défaut* » (***last resort***) : **0.0.0.0**
 - + *Similaire à « autres directions » des panneaux routiers*

Routage direct



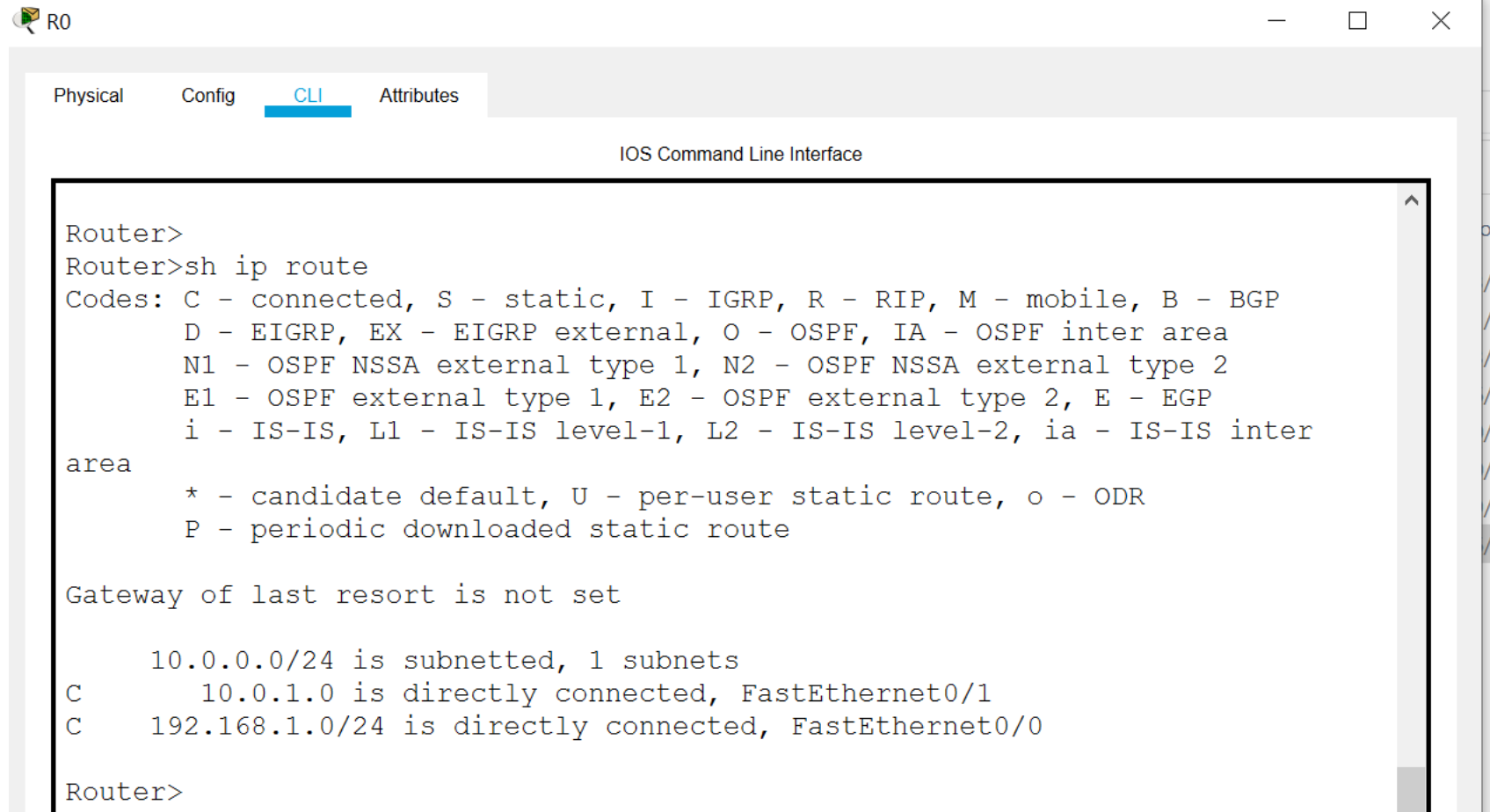
*Aucun routeur n'est nécessaire si communications internes
routage (table de routage / machine) établi grâce à ARP*

Routage indirect



Routeur : rôle capital

Routage indirect: routeur



```
Router>
Router>sh ip route
Codes: C - connected, S - static, I - IGRP, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2, E - EGP
       i - IS-IS, L1 - IS-IS level-1, L2 - IS-IS level-2, ia - IS-IS inter
area
       * - candidate default, U - per-user static route, o - ODR
       P - periodic downloaded static route

Gateway of last resort is not set

    10.0.0.0/24 is subnetted, 1 subnets
C       10.0.1.0 is directly connected, FastEthernet0/1
C     192.168.1.0/24 is directly connected, FastEthernet0/0

Router>
```

Table de routage

- Cas de celui d'un *host* (terminal)
 - Exemple sous windows (avec [netstat -r](#))
 - Destination
 - **<Prefix, Masque>** : équivalence opérationnelle de Adresse/L
 - Route par défaut (*last resort*) : **0.0.0.0**
 - Next Hop (Adr. passerelle)
 - Adresse du prochain routeur
 - Pas de prochain routeur si routage direct (*on-link* dans le jargon MS)
 - Interface
 - Identifier l'interface réseau (*ici via son adresse IP*)
 - Métrique
 - Distance (coût, etc.)
- Éléments utilisés par le **Forwarding**:
 - Conceptuel : **<destination, sortie>**
 - Pratique : **<prefix, masque>, Adr. interface**

Destination réseau	Masque réseau	Adr. passerelle	Adr. interface	Métrique
0.0.0.0	0.0.0.0	192.168.0.254	192.168.0.17	25
127.0.0.0	255.0.0.0	On-link	127.0.0.1	306
127.0.0.1	255.255.255.255	On-link	127.0.0.1	306
127.255.255.255	255.255.255.255	On-link	127.0.0.1	306
192.168.0.0	255.255.255.0	On-link	192.168.0.17	281

Structure Internet

Systèmes autonomes

Systèmes autonomes

- + Adresse IP :
 - un seul niveau d'agrégation <réseau, station>
- + Système autonome
 - agrégation des réseaux IP
 - Systèmes autonomes = *Autonomous System* (AS)
- + AS = Domaine sous la responsabilité d'une autorité unique
 - Indépendance vis-à-vis d'autres AS
 - Facilite l'organisation de grandes structures
 - + fournisseur d'accès, réseau des multinationales
- + AS = identification sur 16 bits (jusqu'à 2007) ou 32 bits
 - ASN=AS Number
 - + Exemple: 12568 (*asplain*), ou 234.13457 (*asdot*), 0.y pour y sur 16bits

AS : IGP et EGP

- + AS = domaine connexe
- + Tous les routeurs d'un AS s'informent via un IGP
 - IGP : *Interior Gateway Protocol*: **terme générique**
 - Exemple : RIP, **OSPF**
- + Quelques routeurs s'échangent des informations de routage avec leurs paires d'autres AS via des EGP
 - EGP : *Exterior Gateway Protocol*
 - Exemple : BGP
 - Remarque : EGP est un terme générique, mais aussi le nom d'un protocole de routage

Organisation Internet

- + AS interconnectés
 - Diminution taille des tables de routage
- + ASN affectés par IANA aux *cinq* RIR
 - IANA: Internet Assigned Number Authority
 - RIR : Regional Internet Registries
 - + RIPE NCC (Europe), ARIN (US-Canada), {AFRI,AP,LAC}NIC
- + Interconnexion ad hoc (*peering*)
- + Peering (*appairage*)
 - Raccordement entre AS (donc trafic, donc **coût BP**)
 - Echanges de trafic (destination finale)
 - Trafic en transit (souvent **payant**)
 - Mode d'échange : BGP

Protocoles de routage

Problème

Classification

DV – RIP

Link State – OSPF

Connexité - BGP

Problème

- + Réseau = modélisation par graphe
- + Réseau = système distribué
- + Routage
 - Recherche d'un chemin dans un graphe par un procédé distribué
 - Facile si tous les nœuds connaissait (de manière cohérente) le **graphe complet** (tous les nœuds)
- + Difficulté majeure
 - Chaque nœud ne connaît que ses **voisins immédiats**
 - Connaissance locale du réseau (graphe)
 - Connaissance globale : **difficile** et **coûteux**

Type d'information échangée

+ Connexité

- Une liste de réseaux
- Liste à une seule colonne

+ Vecteur de distance

- Une liste de réseaux avec un coût (distance) pour chacun
- Liste à deux colonnes

+ Etat des liens

- Une liste où chaque réseau est communiqué avec toutes les informations
- Une liste des objets (avec une structure de données)

Périmètre d'échanges

- + La propriété « **Atomicité** » est requise
- + Atomicité
 - information ou bien reçue par tous ou bien par personne
- + Périmètre naturel = les voisins directs
 - Avantage : identification naturelle, atomicité facile à assurer
 - Inconvénient : vitesses de convergence
 - + Le voisin de mon voisin de mon voisin m'a dit
- + Périmètre conceptuel : tous les routeurs
 - Avantage : convergence rapide
 - Difficulté : diffusion au groupe, atomicité moins facile à assurer

Trois familles (2+1)

+ **Protocole à vecteurs de distance (DV)**

- DV (Distance Vector)
- Périmètres : voisins directs
- RIP

+ **Protocole à états de liaison (LS)**

- LS (Link State)
- Périmètre : tous les routeurs
- OSPF

+ *Echanges entre paires sélectionnés*

- *BGP*

Procédé DV

- + Algorithme de **Bellman-Ford** : calcul **distribué** de routes.
- + Un routeur, par **diffusion** régulière, informe ses **voisins** des routeurs qu'il connaît.
- + Pour chaque destination connue, l'information est composée de:
 - Id. de destination (prefix et masque),
 - La distance (métrique) :
 - + Exemple de distance : **nombre de saut(s) : nombre de routeur(s)** à traverser pour atteindre le réseau de destination.
- + Le routeur mémorise aussi l'adresse du routeur suivant (next hop, le suivant sur le chemin mené au destinataire)
- + En réception, le routeur examine, pour chaque destination reçue, la nouvelle situation. Une **mise à jour** a lieu si :
 - Une route reçue est meilleure (**métrique inférieure**),
 - Une route reçue est **nouvelle**.
 - Dans les deux cas, l'annonceur devient le nouveau next-hop.
- + Exemple de réalisation : **RIP**

Procédé DV : exemple

+ Hypothèse :

- chaque saut a un coût unitaire
- R1 et R4 sont directement relié

+ Scenario: à l'instant t, un routeur R1 connaît, entre autres,

- **R2**, distance=**6**, next-hop=R3

+ A t+1, un message est reçu de R4 avec, entre autres

- **R2**, distance=**2**,
- **R5**, distance=**3**

+ Mise à jour

- **R2**: Distance plus courte : distance=3 (**2**+1)<**6**, next-hop=R4
- **R5** : Nouveau : R5, distance=4 (**3**+1) , next-hop=R4

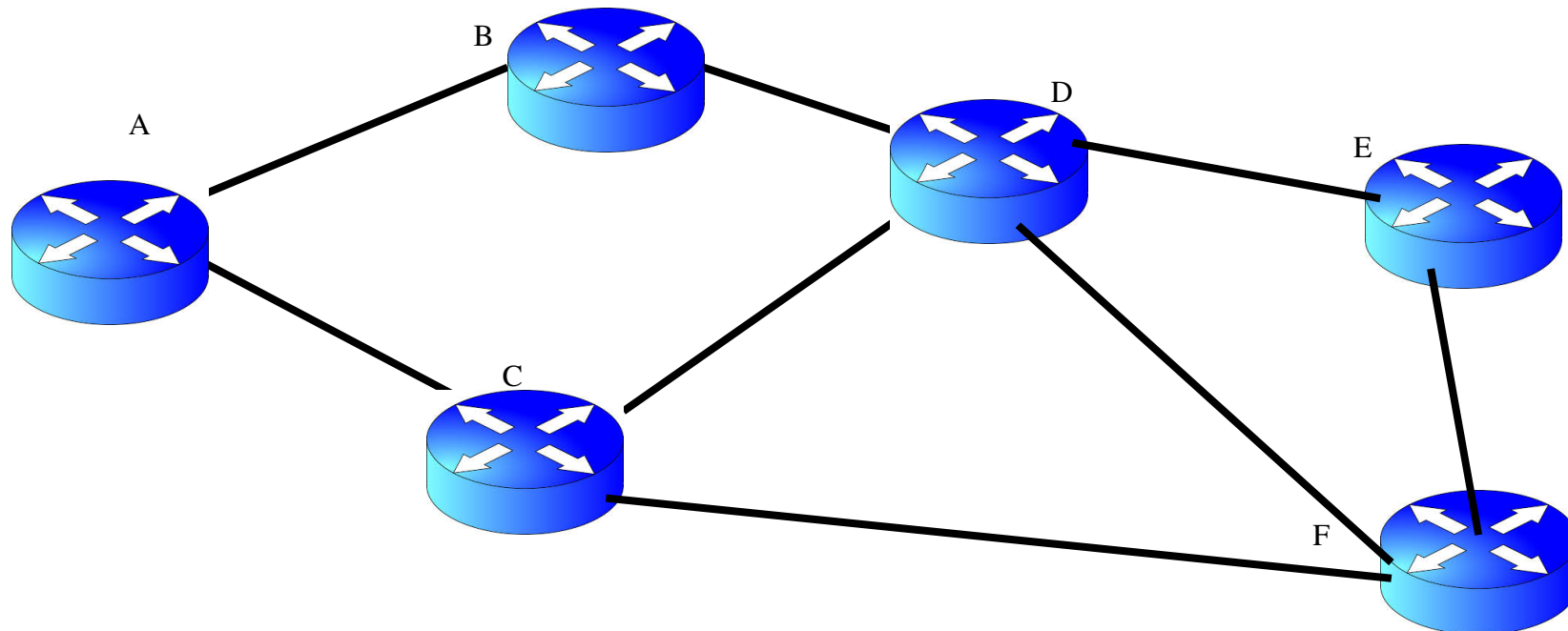
Procédé DV

Tableau routage in fine sur A

Destination	Distance	NextHop	remarque
A	0		
B	1	B	
C	1	C	
D	2	C	aurait pu être B
E	3	B	aurait pu être C
F	2	C	

Tableau routage in fine sur D

Destination	Distance	NextHop	remarque
A	2	B	aurait pu être c
B	1	B	
C	1	C	
D	0		
E	1	E	
F	1	F	



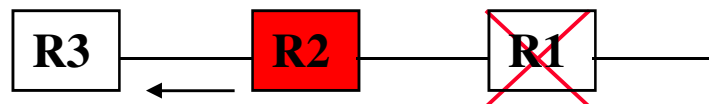
DV : Comptage à l'infini

+ Comptage à l'infini :

- R3 tient l'info de R1 via R2 :
- R1 tombe en panne
- R2 attend une mise à jour régulier
- Celle de R1 manque, $d(R1, R2)$ remet à l'infini par R2
- R3 informe R2 avec $(d(R1, R3)=2)$
- Mise à jour chez R2 : $d(R2, R1)=1+2=3$
- R2 informe $d(R2, R1)$ à R3
- $d(R1, R3)$ devient $3+1=4$,

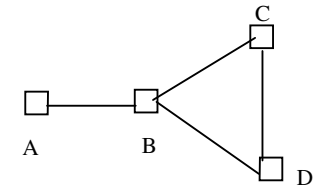
+ Remède

- Horizon coupé (split-horizon)
- Ne renvoie pas des informations apprises sur une interface, en utilisant la même interface



DV : comptage à l'infini

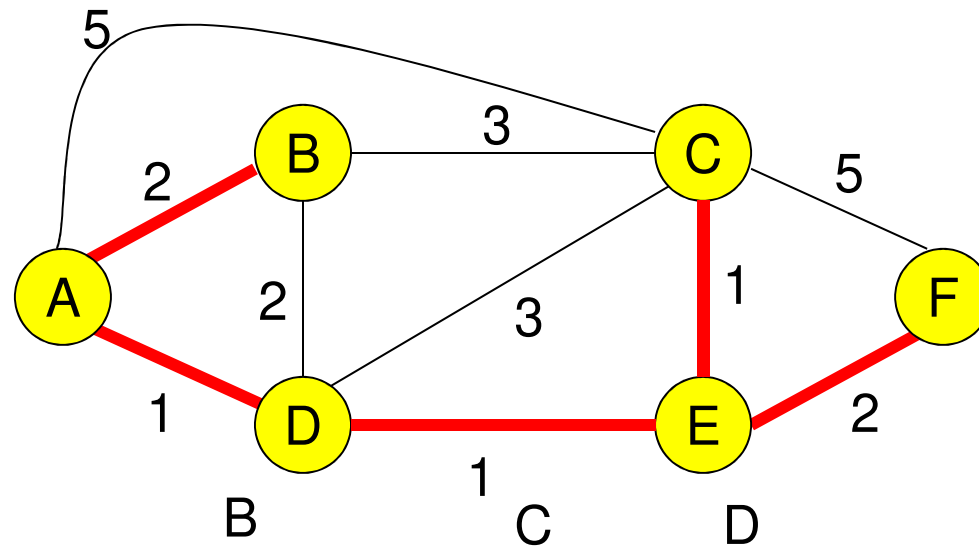
- + Même scénario que précédemment (A tombe en panne), avec une topologie plus complexe
 - Redondance de chemins avec l'ajout d'un quatrième routeur
 - L'horizon coupé ne peut pas éviter les boucles
- + Remède :
 - De demander au routeur de ne rien faire durant la période transitoire (**hold down**) de changement d'état
 - Dans ce cas précis (entre C et D) le routeur peut renvoyer sur l'interface d'apprentissage une route empoisonnée (= de métrique infinie) (**Poison Reverse**)



Procédé LS

- + Diffusion périodique par chaque routeur de l'état de tous ses liens (LS) à **tous les routeurs** du **domaine**
 - + Validité limitée dans le temps
 - + Information (LS) est envoyée généralement avec un numéro de séquence et un certain crédit (TTL):
 - + Garantie pour l'info la plus récente et évitement de boucles
 - Atomicité (cohérence) assurée avec acquittement
- + Mécanisme de dissémination plus complexe que DV
- + Tous les routeurs ont une **vision complète** et riche
 - Métriques multiples (débit, délai, coût, fiabilité, etc....).
- + l'algorithme **SPF (Dijkstra: Shortest Path First)** :
 - **Chaque routeur** détermine, **localement**, un plus court chemin vers **chaque destination**.
- + Exemple de réalisation : **OSPF**

Exemple



step	SPT	D(b), P(b)	D(c), P(c)	D(d), P(d)	D(e), P(e)	D(f), P(f)
0	A	2, A	5, A	1, A	~	~
1	AD	2, A	4, D		2, D	~
2	ADE	2, A	3, E			4, E
3	ADEB		3, E			4, E
4	ADEBC					4, E
5	ADEBCF					

LS vs DV

+ DV :

- Mécanisme de dissémination **simple**
- **Vision limitée** et indirecte du réseau
- Convergence **lente**
- Métrique **unique**

+ LS

- Mécanisme de dissémination **complexe**
- **Vision complète** du réseau (si *bonne dissémination!*)
- Convergence **rapide** (si bonne dissémination)
- Métriques **multiples**
 - + donc routes multiples **potentiels** suivants critère/métrique

Protocoles de routage RIP et OSPF

Principe
Protocole

Le protocole RIP (V1)

- + Défini dans RFC 1058
- + Très répandu sur des sites de tailles réduites
- + Utilise des datagrammes UDP (**port 520, système**)
 - Port système=mesure de sécurité
 - 1 datagramme contient 25 entrées au maximum
 - UDP : pas de fiabilité de communication
- + Métrique limitée à **15** (≥ 16 : route inaccessible)
- + Ne véhicule pas le masque du réseau
- + Ne conserve que la meilleure route

Fonctionnement de RIP (V1)

- + Période de diffusion : toutes les 30 s
 - Attention aux phénomènes de synchronisations entre routeurs (trafic instantané accrue!)
- + Une route (routeur) est invalidé ($D \geq 16$) au bout de 180 s si :
 - Aucun message n'a été reçu
 - Un message explicite a été reçu.
 - On lui communique toujours les info. de routage
- + La route est maintenue dans les tables, mais les voisins sont avertis du problème (**hold down de 180s**)
- + Au bout de **240 s**, suppression de la route (**flush**)

RIP V2

- + RIP Version 2 : RFC 1387 et 1388
- + Diffusion multicast (224.0.0.9)
- + Véhicule le masque de réseau
- + Premier protocole avec possibilité d'authentification

Open Shortest Path First

- + OSPF (version 2) est un protocole Link State :
 - Spécifications dans RFC 1583 (3/1994)
 - Créé pour remplacer RIP et les autres protocoles IGP propriétaires.
- + Utilise IP (protocole N°89) et le multicast
 - 224.0.0.5 (tous les routeurs OSPF de l'AS, et
 - 224.0.0.6 (entre les DR)
- + Routage basé sur SPF
- + OSPF utilise les types de service (ToS) d'IP
 - Permet la gestion de plusieurs routes pour une même destination,
 - + Selon critères du champ TOS : délai, débit, fiabilité, coût (1 à la fois).

OSPF : Caractéristiques (2)

- + OSPF supporte VLSM (masque véhiculé dans ces messages)
- + Découpage d'un AS en sous systèmes (aires)
 - Dissémination plus efficace des informations de routage.
- + Sécurité :
 - N° de séquence, Checksum,
 - Authentification possible pour les échanges entre routeurs.
- + Fréquence d'envoi des Informations (LSA)
 - LSA=Link State Announcement
 - quand l'état d'une ligne change
 - ou toutes les **30 minutes**.

Area, Backbone, ABR, ASBR

R1... R6=Internal
RA, RB, RC=ABR
RE = ASBR

