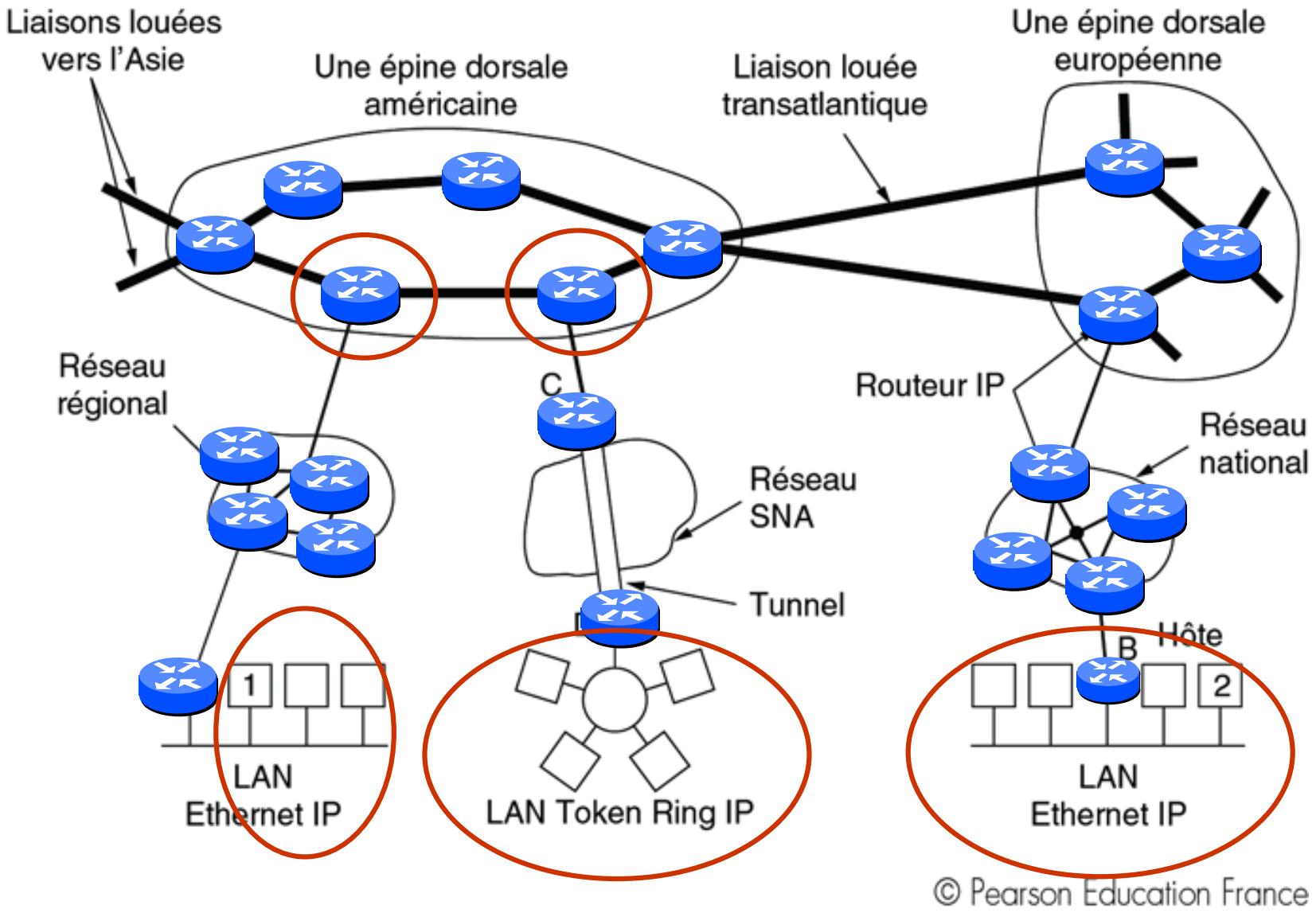
Réseaux IP, routage avancé

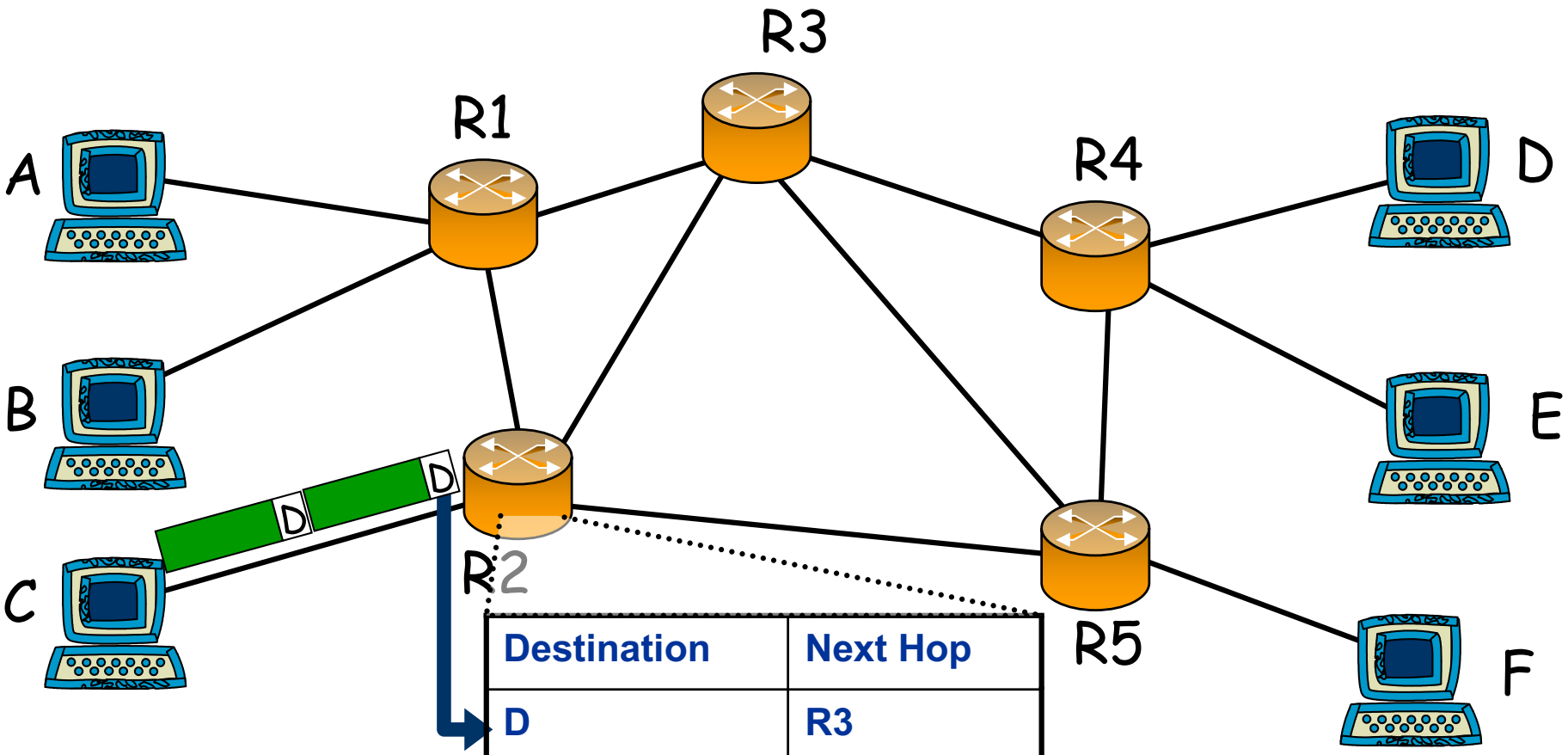
C. Pham
Université de Pau et des Pays de l'Adour
Département Informatique
http://www.univ-pau.fr/~cpham
Congduc.Pham@univ-pau.fr

Copyright

La diversité des réseaux

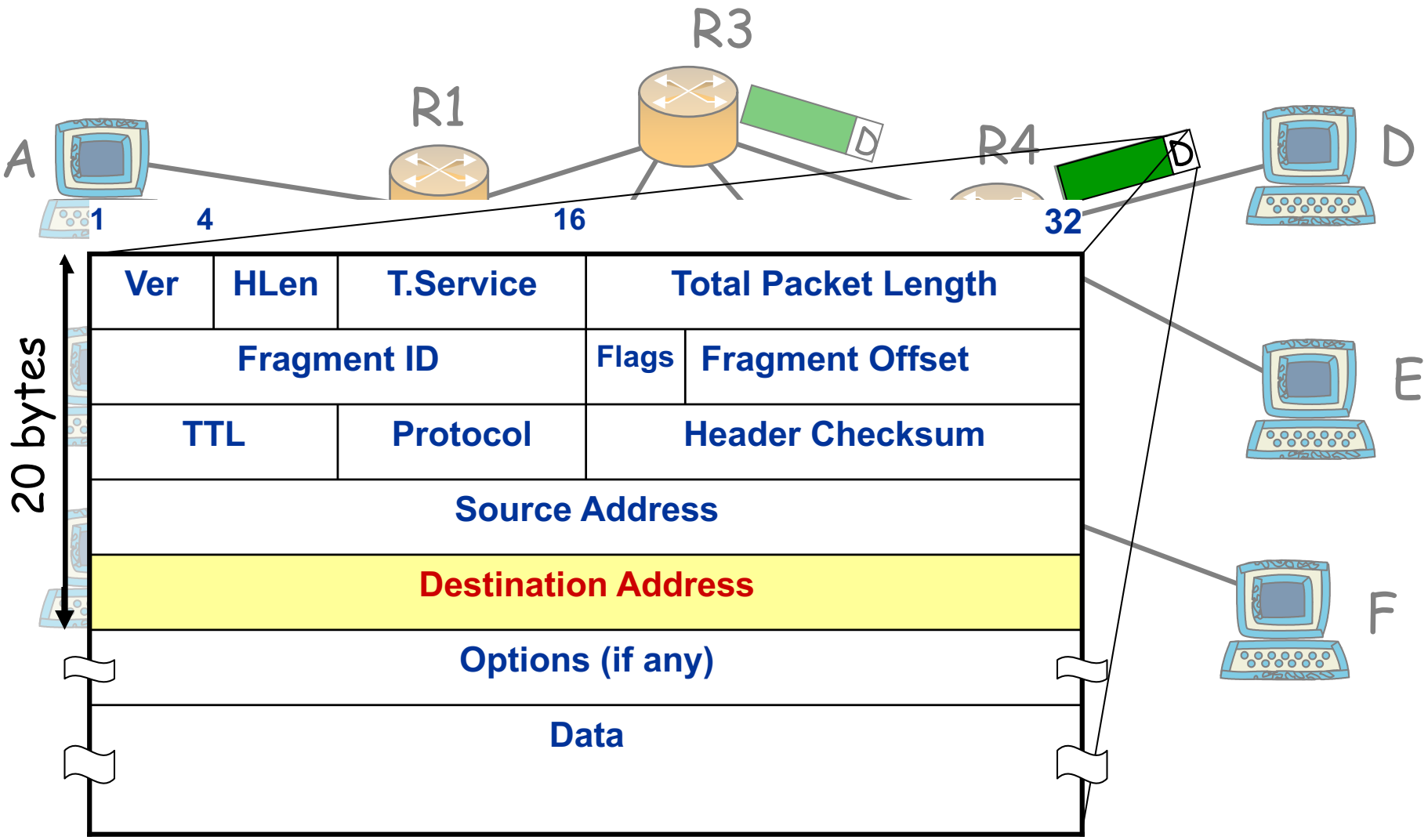


Le routage de proche en proche illustré



Destination	Next Hop
D	R3
E	R3
F	R5

Le routage IP



High Performance Routers


©Juniper

©Procket Networks

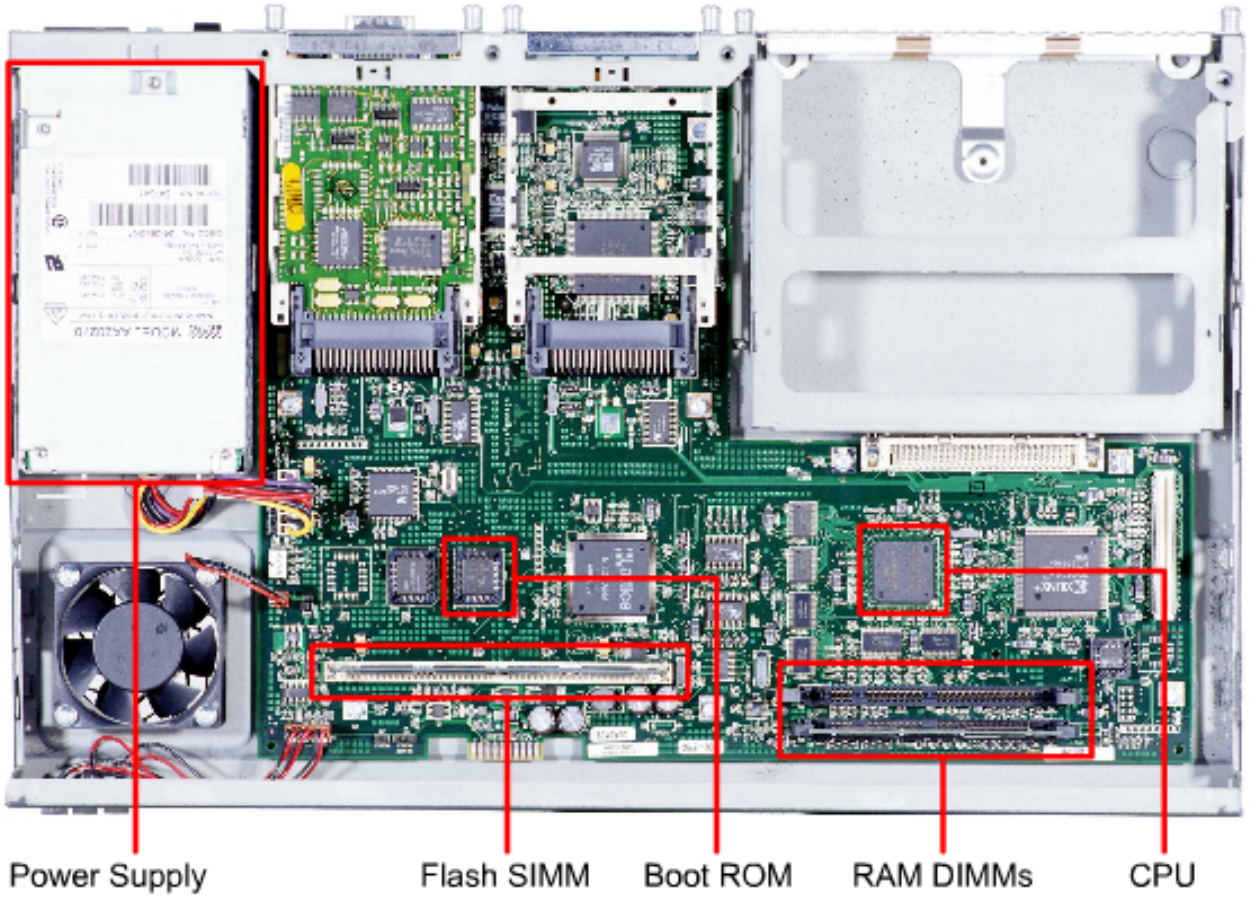

©Alcatel


©Nortel Networks

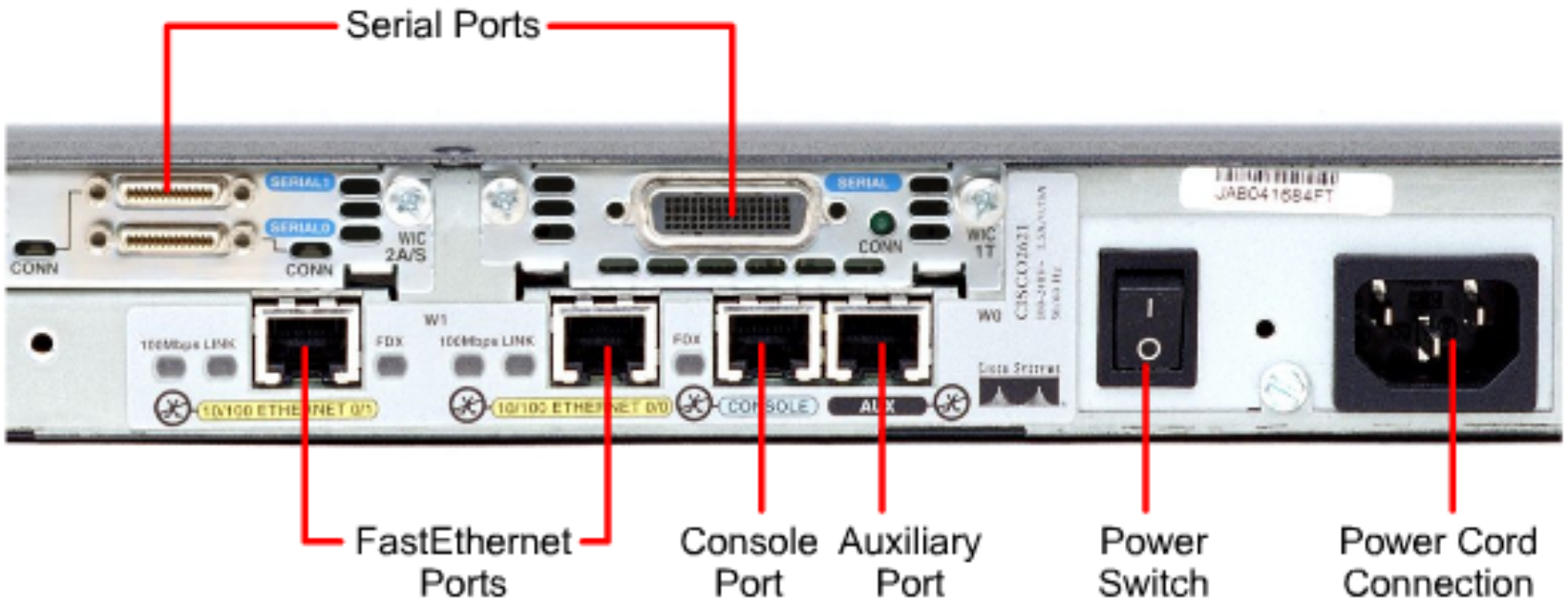

©Lucent

and more…

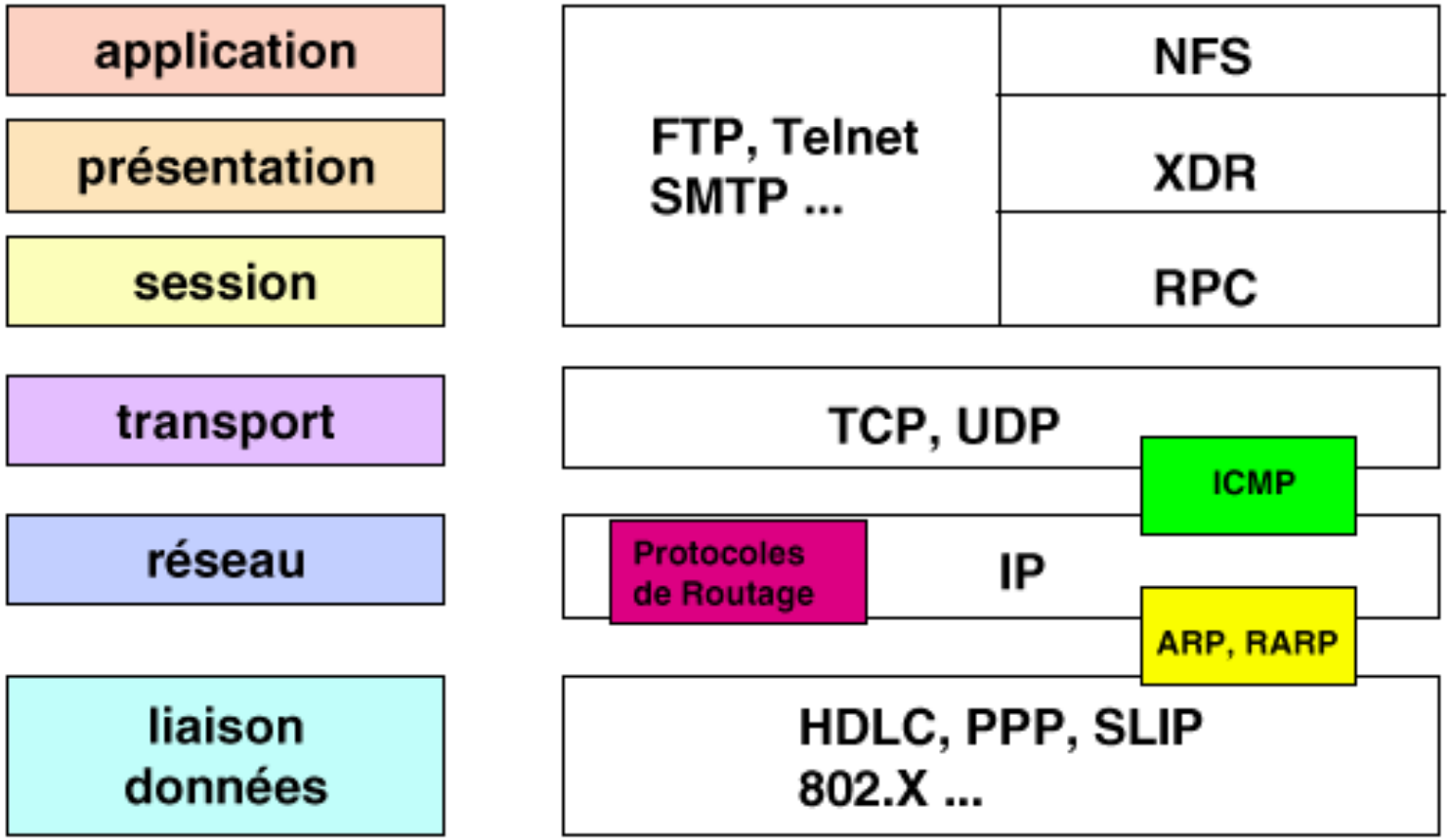
Internal Components of a 2600 Router



External Connections on a 2600 Router



Les protocoles de l ’Internet



- **Couche transport dans l'Internet (light, détaillé)**
- TCP/IP sur les réseaux locaux
- Couche application

Les protocoles de routage

- **Fonctions de base, protocole de routage**
  - déterminer et mise à jour des tables de routages,
  - répartition des charges pour éviter les congestions,
  - critères pour déterminer le coût d'une liaison (nombre de noeuds, temps de traversée, taille des files d'attente etc.)

- **Fonctions avancées, liées à la qualité de service**
  - définir des classes de trafic, ordonnancement,
  - instaurer la sécurité,
  - contrôle de flux et contrôle de congestion,
  - qualité de service: temps-réel, multimédia etc.

Hiérarchie dans l'Internet

- **Trois niveaux de hiérarchie dans les adresses**
  - adresse réseaux,
  - adresse sous-réseaux,
  - adresse de la machine.
- **Le réseaux de backbone ne publient les routes qu'aux réseaux, et pas aux sous-réseaux.**
  - e.g. 135.104.*, 192.20.225.*
- **Malgré cela, il y a environ 118,000 adresses de réseaux dans les routeurs de backbones (2003)**
- **Les gateways communiquent avec le backbone pour trouver le meilleur noeud suivant pour chaque réseau dans l'Internet.**

Les protocoles de routages pour réseaux paquets

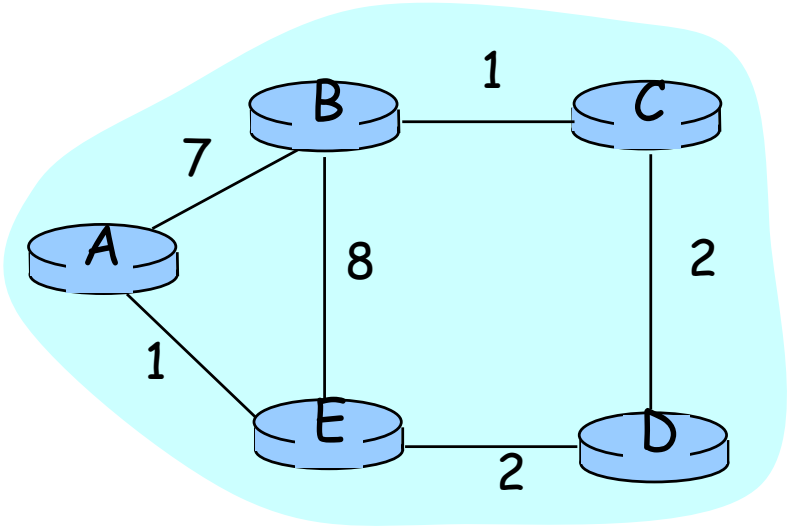
- **Vecteur de distance (Distance-Vector, DV)**
  - chaque routeur ne connaît initialement que le coût de ses propres liaisons, les routeurs échangent entre-eux des informations de coûts,
  - chaque routeur n'a qu'une vision partielle du réseau: coût vers chaque destination,
  - fonctionne bien sur des systèmes de petite taille.

- **Etat des liens (Link-State, LS)**
  - chaque routeur construit une vision complète de la topologie du réseau à partir d'informations distribuées,
  - ne pas confondre connaître la topologie et connaître tous les noeuds terminaux,
  - fonctionne sur des grands réseaux.

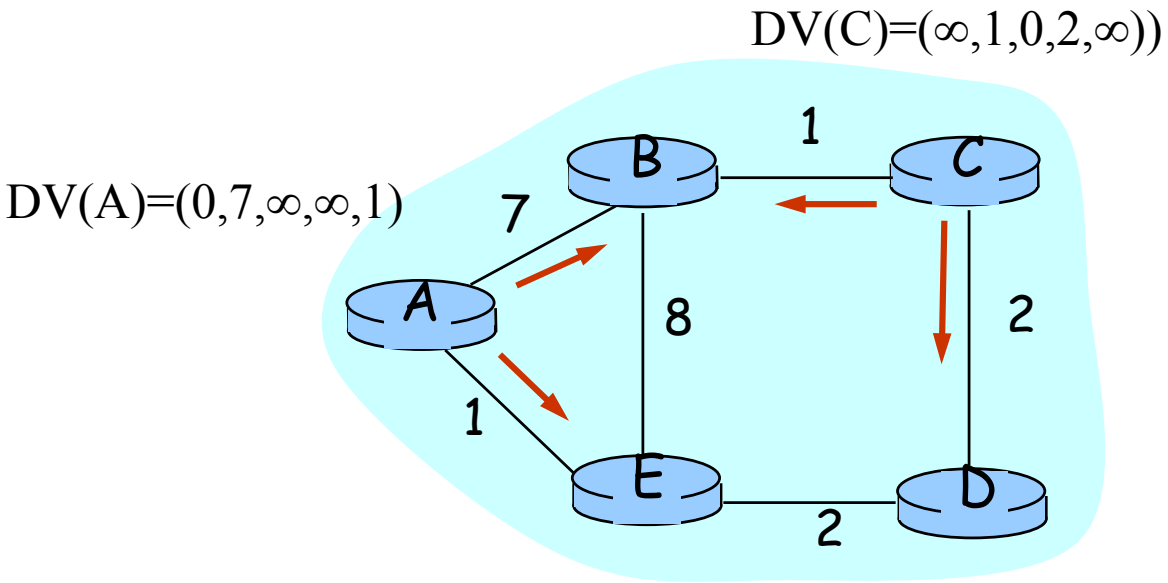
L'approche vecteur de distance (1)

- **Vecteur de distance (Distance-Vector, DV)**
  - chaque routeur ne connaît initialement que le coût de ses propres liaisons vers ses voisins direct. C'est le vecteur initial
  - chaque routeur va échanger son vecteur initial avec tous ses voisins
  - après un certain nombre d'itérations, chaque routeur va connaître le coût vers chaque destination,
  - fonctionne bien sur des systèmes de petite taille.

$$DV(A)=(0,7,\infty,\infty,1)$$



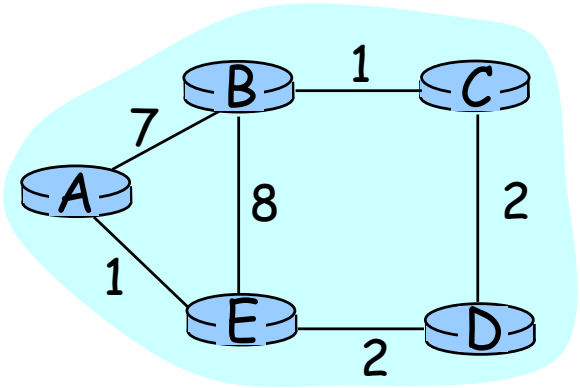
L’approche vecteur de distance (2)



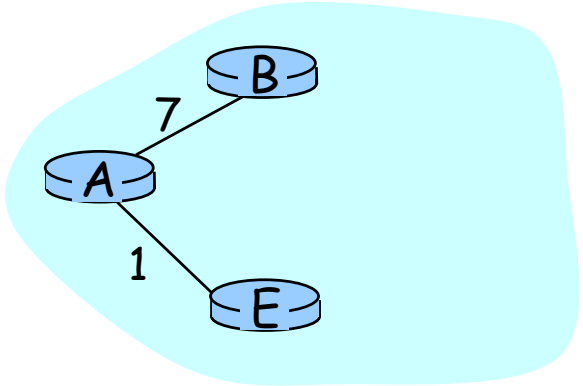
Pas obligatoirement de synchronisation dans les envois de messages

L'approche vecteur de distance (2)

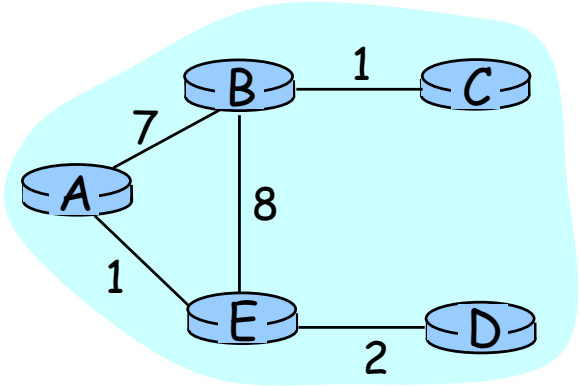
- **Condition de consistence: D(i,j) = c(i,k) + D(k,j)**
- **L'algorithme DV (Bellman-Ford) évalue cette condition de manière récursive**
  - À la m-iéme itération, le critère de consistence est vérifié, en supposant que chaque nœud N "voit" les noeuds et les liens à m-sauts (ou moins) de lui (i.e. on a une vision à m-sauts)



Réseau d'étude

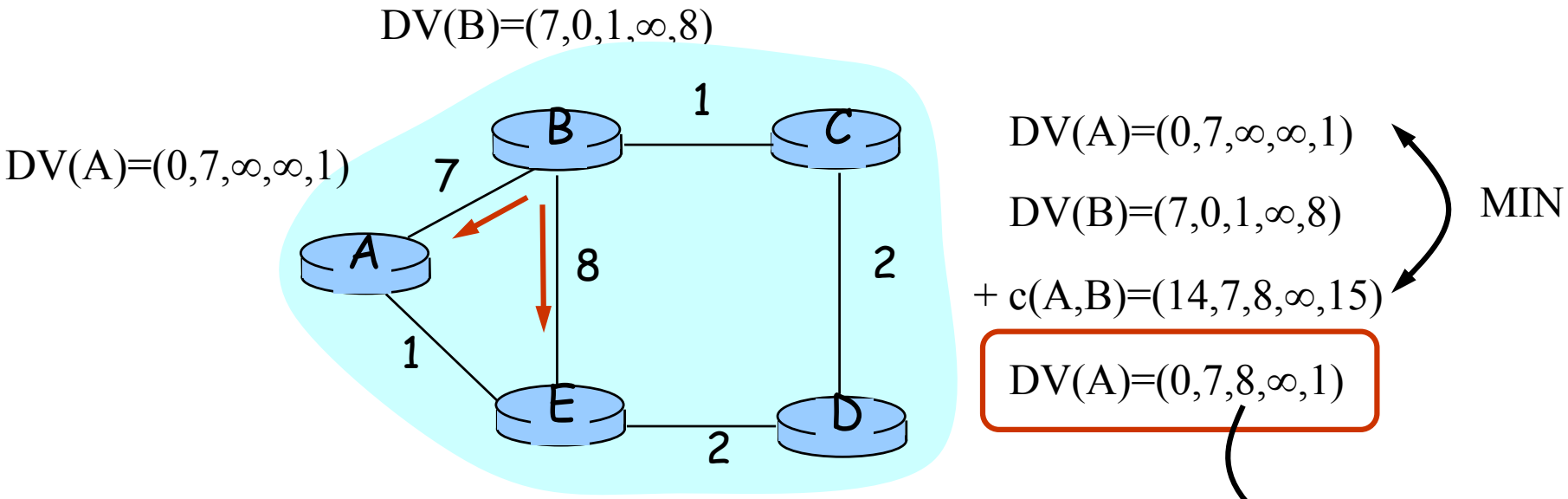


Vision de A à 1-saut (après la 1ère itération)



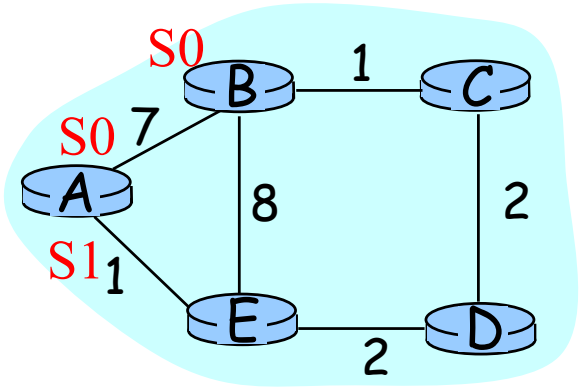
Vision de A à 2-sauts (après la 2nd itération)

Algorithme DV (3)



- **A reçoit de B: DV(B,*)=(7,0,1, ∞,8)**
- **Pour tout voisin k, si c(i,k) + D(k,j) < D(i,j), alors:**
  - D(i,j) = c(i,k) + D(k,j)
  - prochain-saut(j) = k
- **Pour voisin B, si c(A,B)+D(B,C) < D(A,C), alors:**
  - D(A,C) = c(A,B) + D(B,C)
  - prochain-saut(C) = B
  - Plus précisément: prochain-saut(C)=interface menant vers B

Exemple

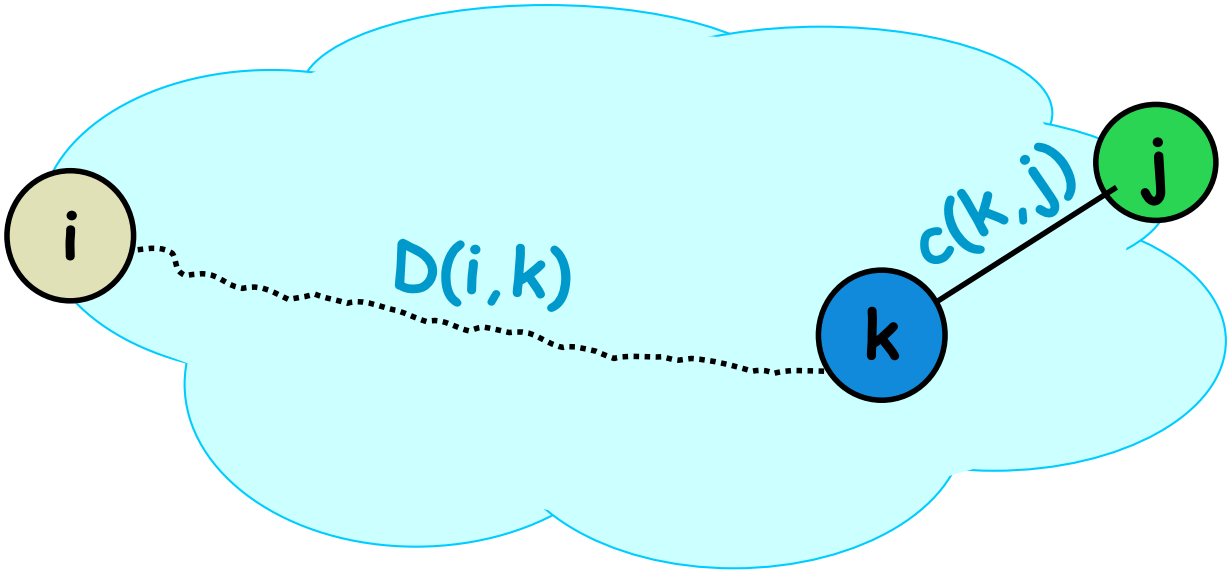


Réseau d'étude

	A	B	C	D	E
	(0,7,-,-,1)	(7,0,1,-,8)	(-,1,0,2,-)	(-,-,2,0,2)	(1,8,-,2,0)
S0, +7	(14,7,8,-,15)				
min	(0,7,8,-,1)	➜Pour aller vers C, A envoie sur l'interface S0			
S1, +1	(2,9,-,3,1)				
	(0,7,8,3,1)	➜Pour aller vers D, A envoie sur l'interface S1			
		(7,14,15,10,8)			
		(7,0,1,10,8)	➜Pour aller vers D, B envoie sur l'interface S0		

Approche "état des liens" (2)

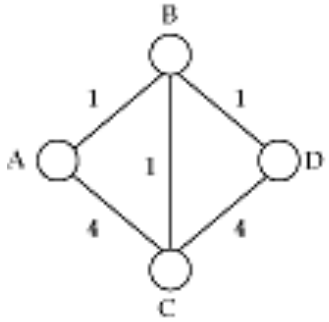
- **L'approche état des liens est itérative, et pivote autour des destinations j, and leur prédécesseurs k = p(j)**
  - Une autre vue du critère de consistence est utilisée:
  - D(i,j) = D(i,k) + c(k,j)

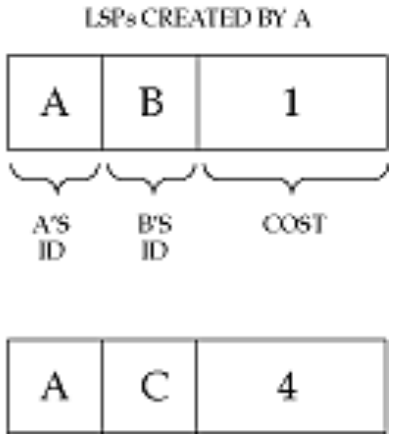


- **Chaque noeud i collecte tous les états c(*,*) d'abord puis exécute localement l'algorithme de plus court chemin (Dijkstra).**

Diffusion de la topologie

- **Un routeur décrit son voisinage avec un *link state packet (LSP)***





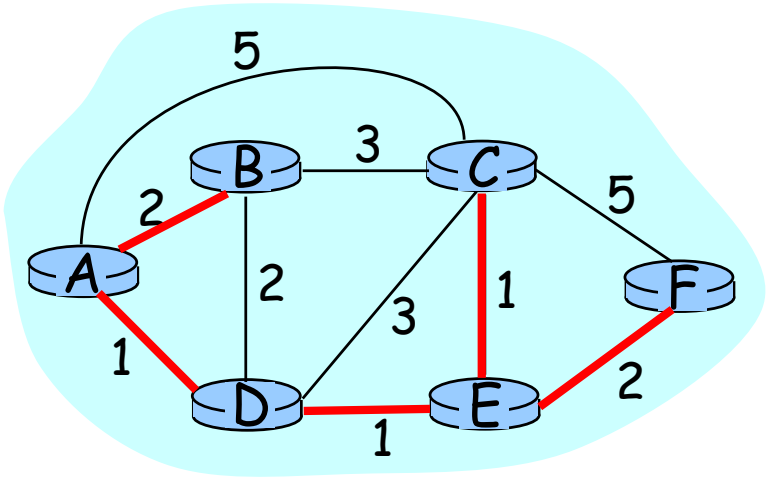
- **Utilise une diffusion contrôlée pour distribuer l'information dans le réseau**
  - Garde le LSP dans une base de données de LSP
  - Si nouvelle, transmet sur chaque interface, sauf l'interface entrante
  - Un réseau avec E sommets transmettra au plus 2E fois

Link State (LS) Approach…

- **After each iteration, the algorithm finds a new destination node j and a shortest path to it.**
- **After m iterations the algorithm has explored paths, which are m hops or smaller from node i.**
  - It has an m-hop view of the network just like the distance-vector approach
- **The Dijkstra algorithm at node i maintains two sets:**
  - set N that contains nodes to which the shortest paths have been found so far, and
  - set M that contains all other nodes.
  - For all nodes k, two values are maintained:
    - D(i,k): current value of distance from i to k.
    - p(k): the predecessor node to k on the shortest known path from i

Dijkstra's algorithm: *example*

Step	set N	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	infinity	infinity
1	AD	2,A	4,D		2,D	infinity
2	ADE	2,A	3,E			4,E
3	ADEB		3,E			4,E
4	ADEBC					4,E
5	ADEBCF					

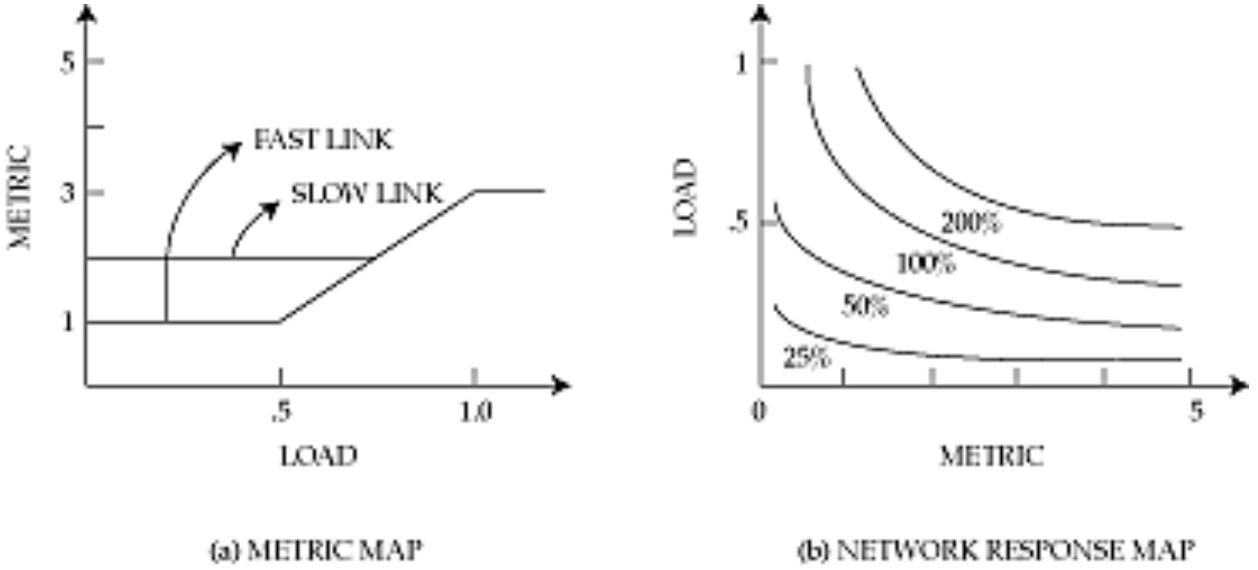


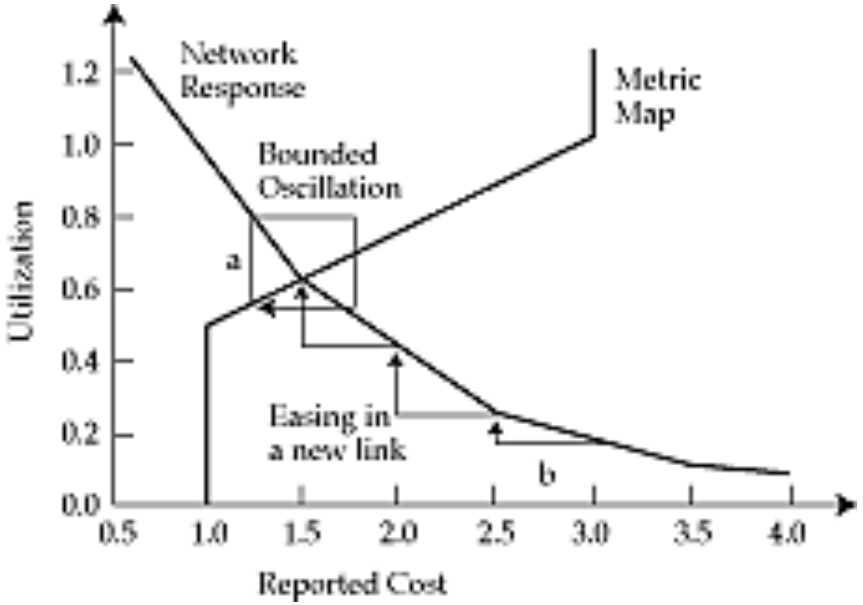
The shortest-paths spanning tree rooted at A is called an SPF-tree

Misc: How to assign the Cost Metric?

- **Choice of link cost defines traffic load**
  - Low cost = high probability link belongs to SPT and will attract traffic
- **Tradeoff: convergence vs load distribution**
  - Avoid oscillations
  - Achieve good network utilization
- **Static metrics (weighted hop count)**
  - Does not take traffic load (demand) into account.
- **Dynamic metrics (cost based upon queue or delay etc)**
  - Highly oscillatory, very hard to dampen (DARPAnet experience)
- **Quasi-static metric:**
  - Reassign static metrics based upon overall network load (demand matrix), assumed to be quasi-stationary

Routing dynamics





Added by C. Pham

Les protocoles de routage (interne)

- **RIP (v1 et v2)**
  - Routing Information Protocol, v2 supporte le VLSM
  - Nombre de saut comme métrique
  - Nombre de saut maximum = 15
  - Mise à jour des tables de routage toutes les 30s
- **IGRP**
  - Interior Gateway Routing Protocol (Cisco)
  - Bande passante et délai comme métrique
  - Mise à jour des tables de routage toutes les 30s
- **OSPF**
  - Open Shortest Path First, supporte le VLSM
  - Notion de zones administratives
  - Utilise SPF (Dijkstra) pour calculer le plus court chemin
  - Le coût d'un lien dépend de la capacité (10^8/capacité)
  - Paquet HELLO toutes les 10s ou 30s
- **EIGRP**
  - Enhanced IGRP (Cisco), supporte le VLSM
  - Utilise l'équilibrage
  - Utilise DUAL (Diffused Update Algorithm) pour calculer le + court chemin

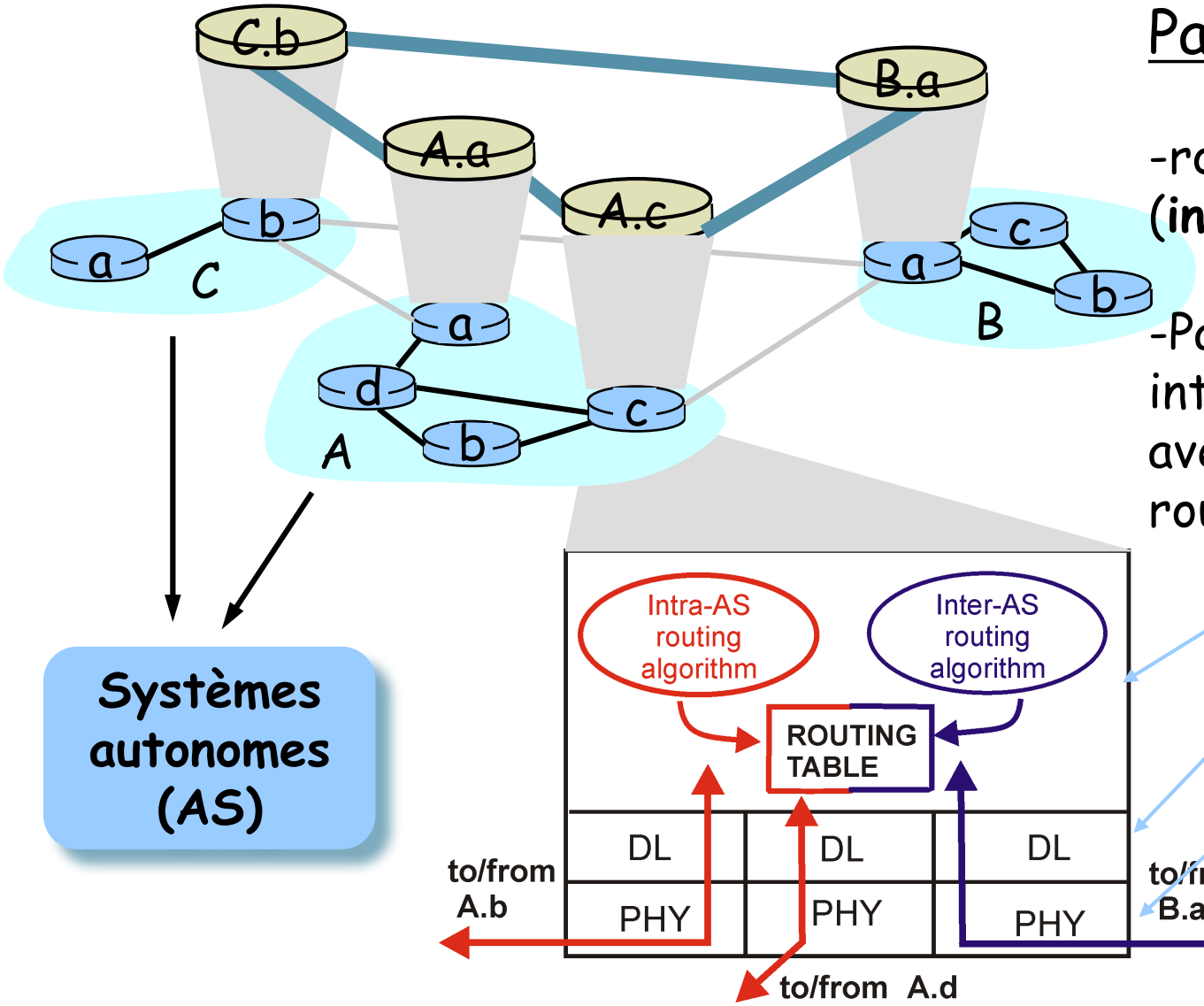
Le routage dans l'Internet

- **Interior Routing**
  - utilise RIP (Routing Information Protocol, DV), IGRP/EIGRP (cisco, DV), IS-IS (LS) et OSPF (Open Shortest Path First, LS). Ce dernier est celui qui est préféré
  - protocole d'échange de données de routage périodiques entre routeurs adjacents.

- **Exterior Routing**
  - utilise EGP (Exterior Gateway Protocol, DV), BGP (Border Gateway Protocol, DV). Ce dernier est celui qui est préféré.
  - connexion TCP entre les routeurs pour les échanges d'informations,
  - routage politique.

- **Notion de peering et d'accords entre AS**

Le vrai routage dans l'Internet



Passerelles:

-routage externe (**inter-AS**) entres eux

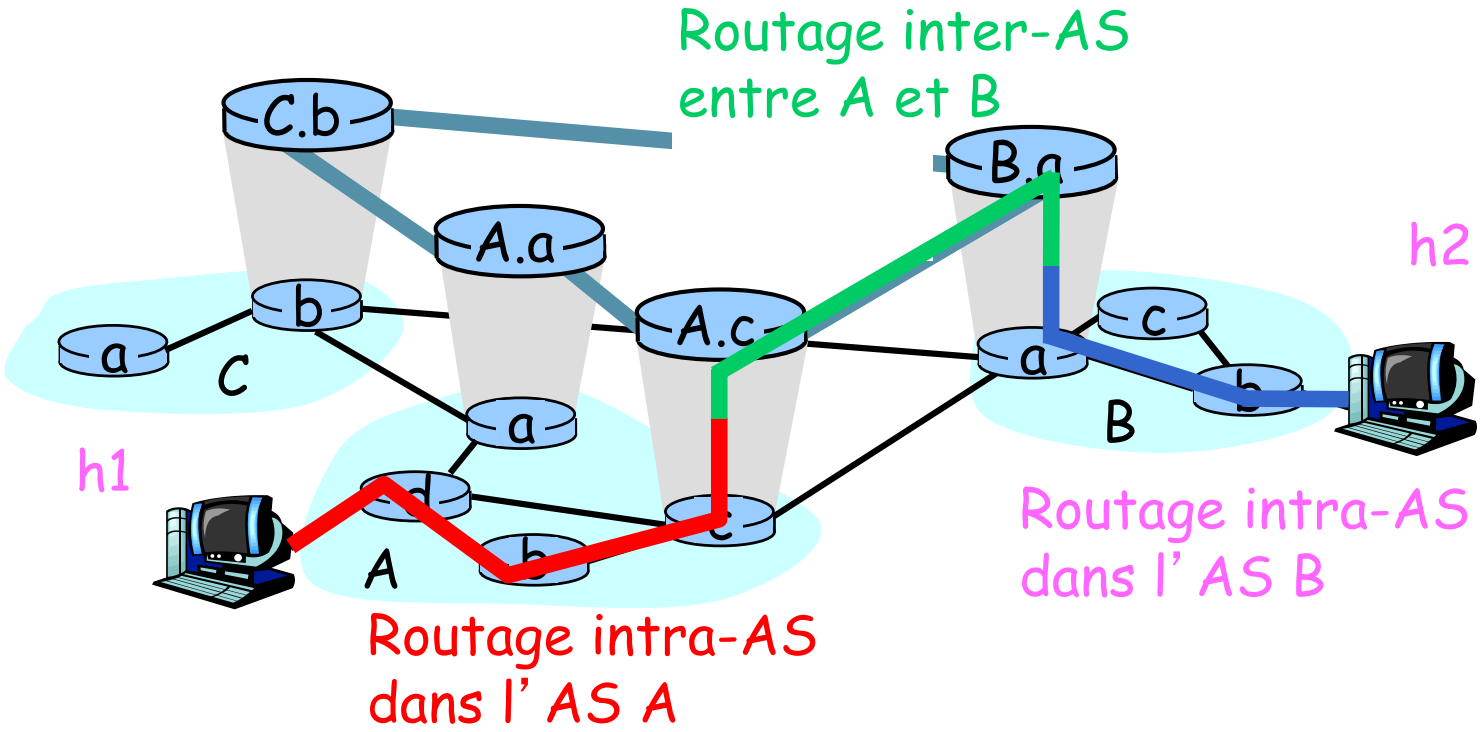
-Participent au routage interne **(intra-AS)** avec les autres routeurs de l'AS

Couche réseau

Liaison

Couche physique

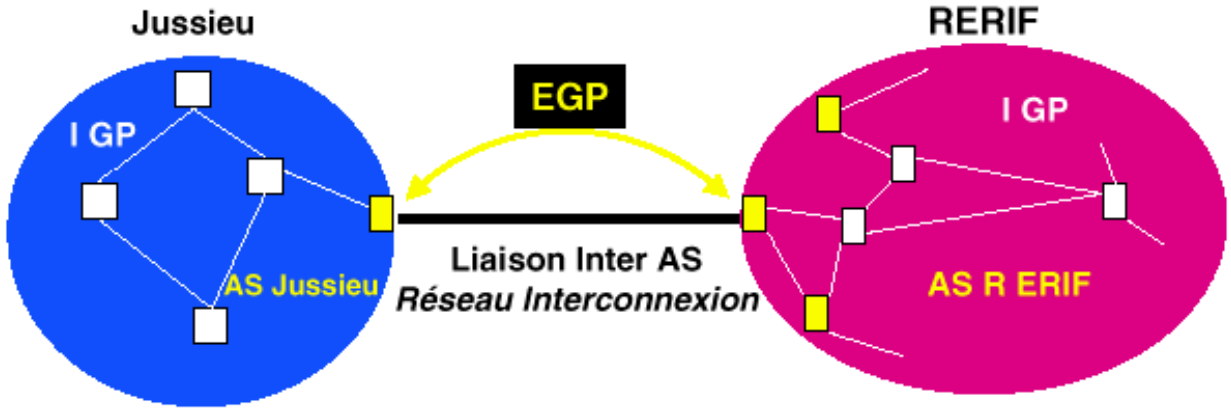
Exemple de routage interne et externe



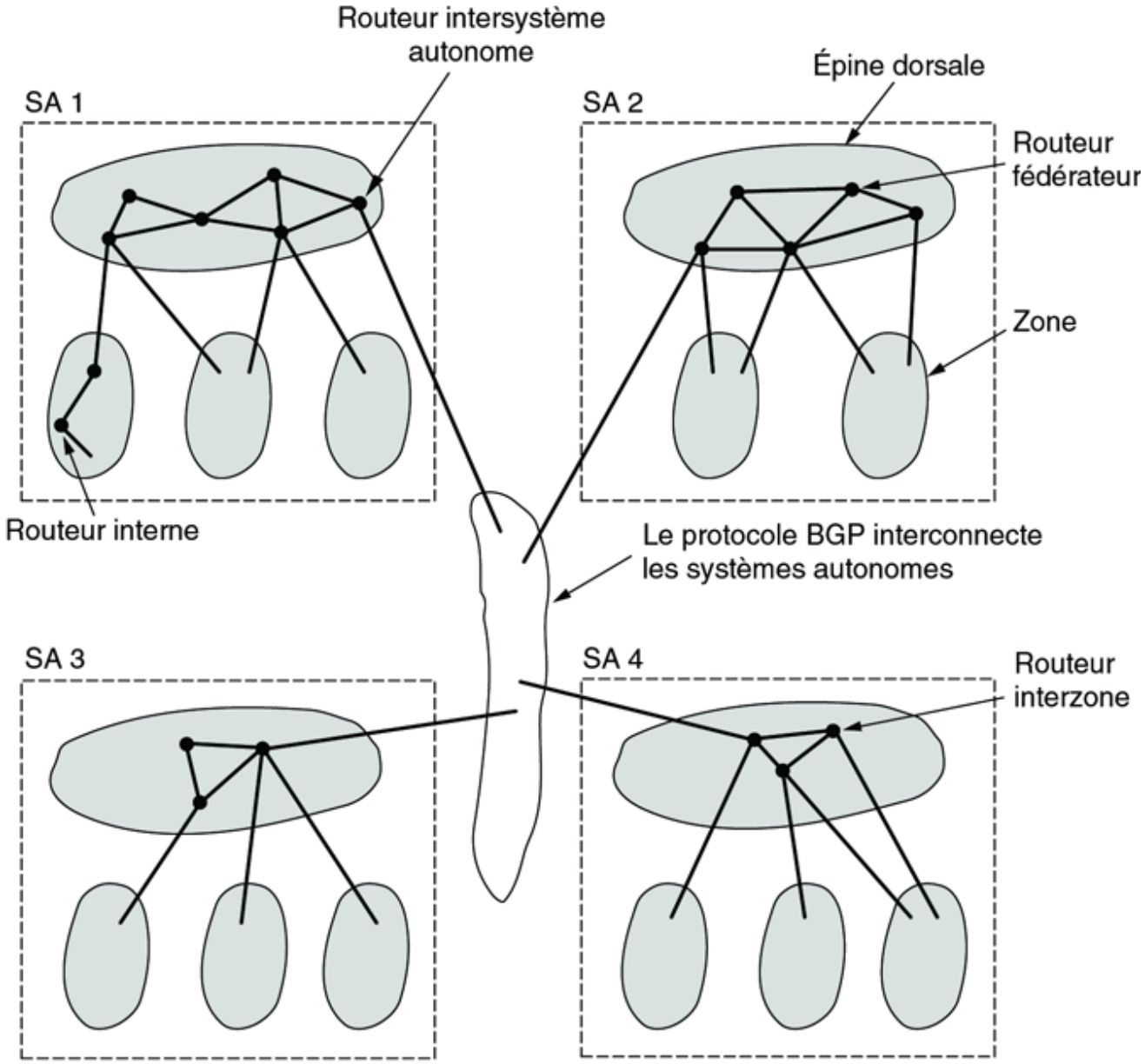
Seulement quelques routeurs (de 0 à 50) dans chaque AS

Organisation en systèmes autonomes

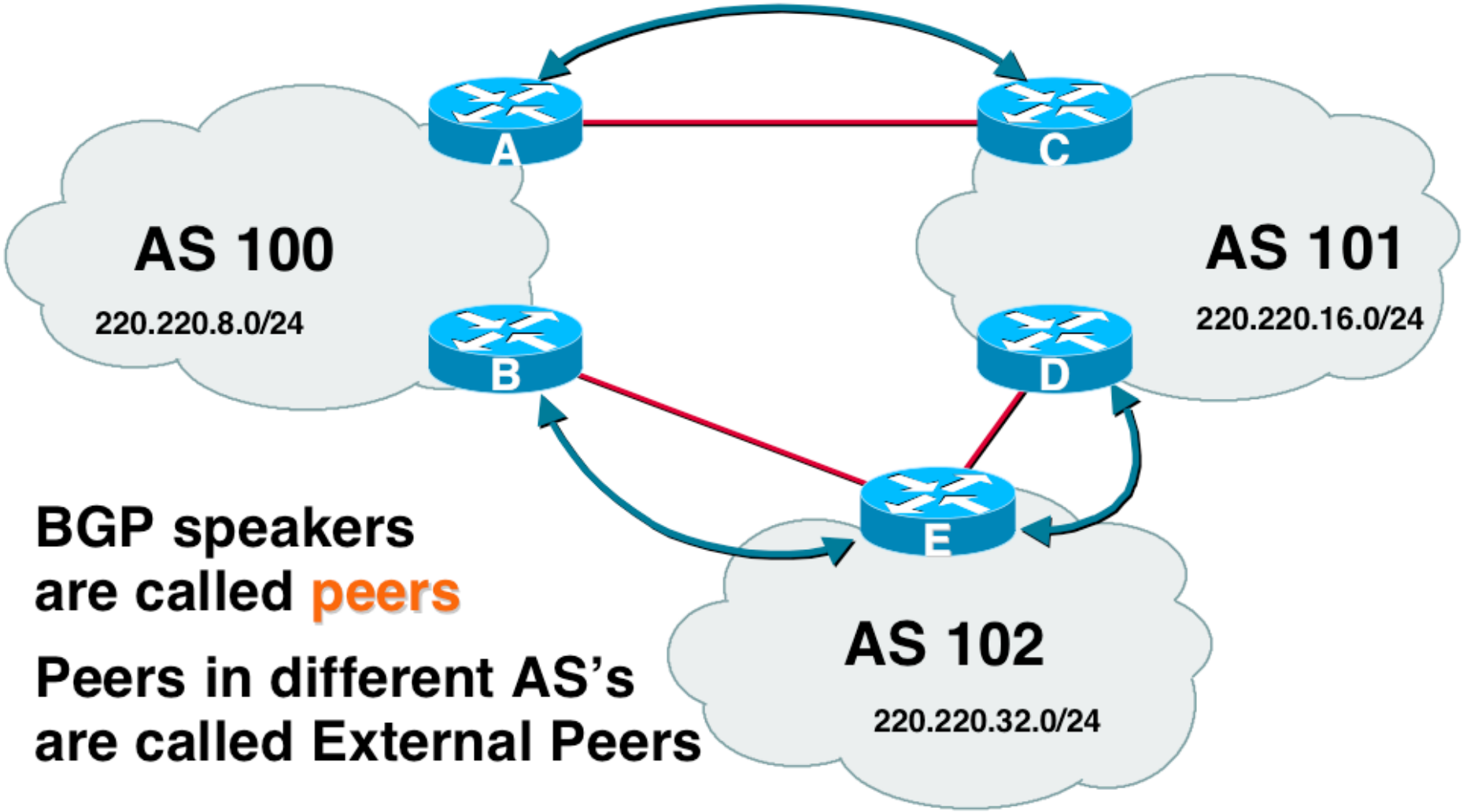
- **L ’Internet est organisée en un ensemble de systèmes autonomes (Autonomous System)**
- **Chaque AS est un ensemble de réseaux et de routeurs sous une administration communes**
  - entreprise, campus, réseau régional…
  - toutes les parties d ’un AS doivent être connexes
- **Les numéros d’AS sont délivrés par le NIC-France**
  - un numero = 16 bits (ex: Renater = AS 1717)
- **Le routage entre AS est appelé routage externe**



Systèmes autonomes et routage externe

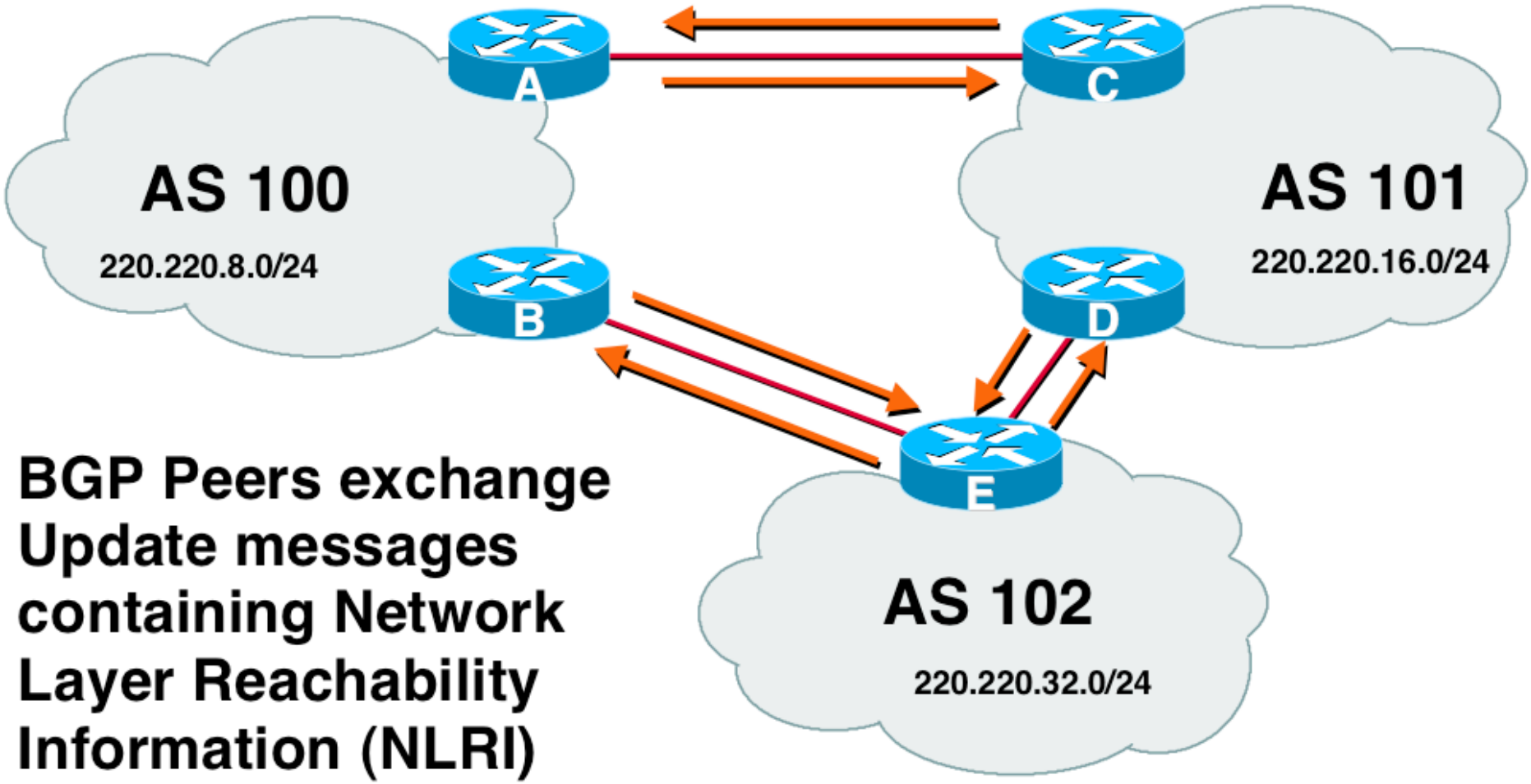


BGP background (1)



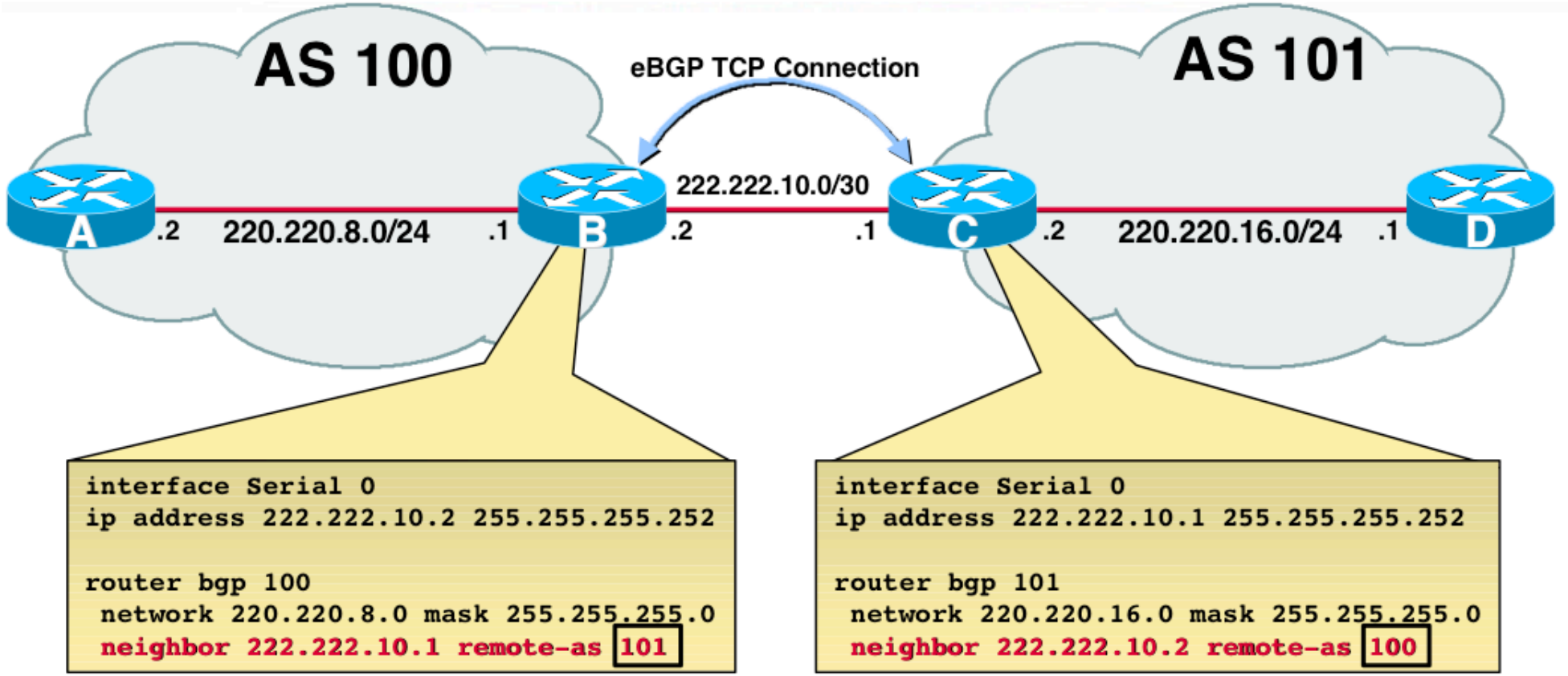
From CISCO

BGP background (2)



From CISCO

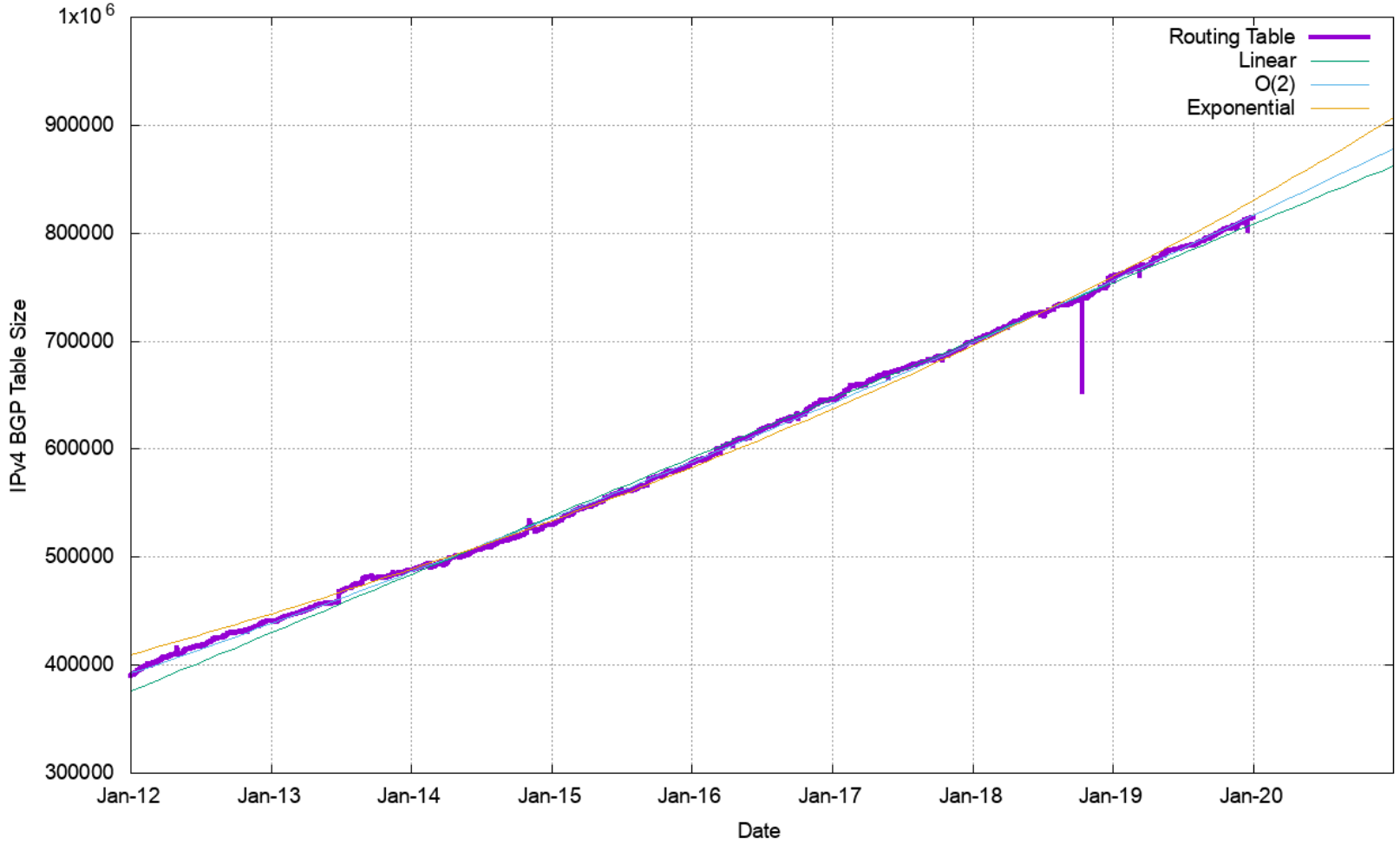
BGP background (3)



- **BGP Peering sessions are established using the BGP "neighbor" configuration command**
  - External (eBGP) is configured when AS numbers are different

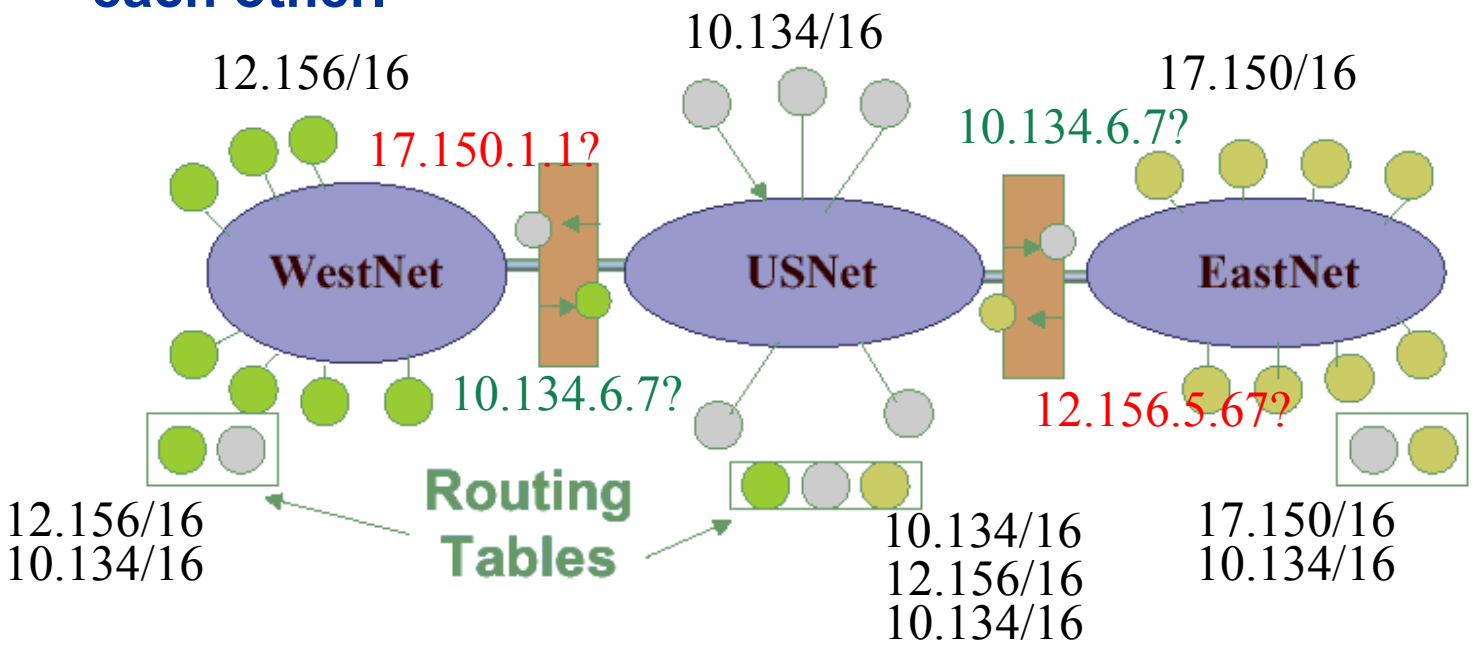
From CISCO

Evolution du nombre d'entrée dans un routeur inter-domaine BGP



Peering is not transit…

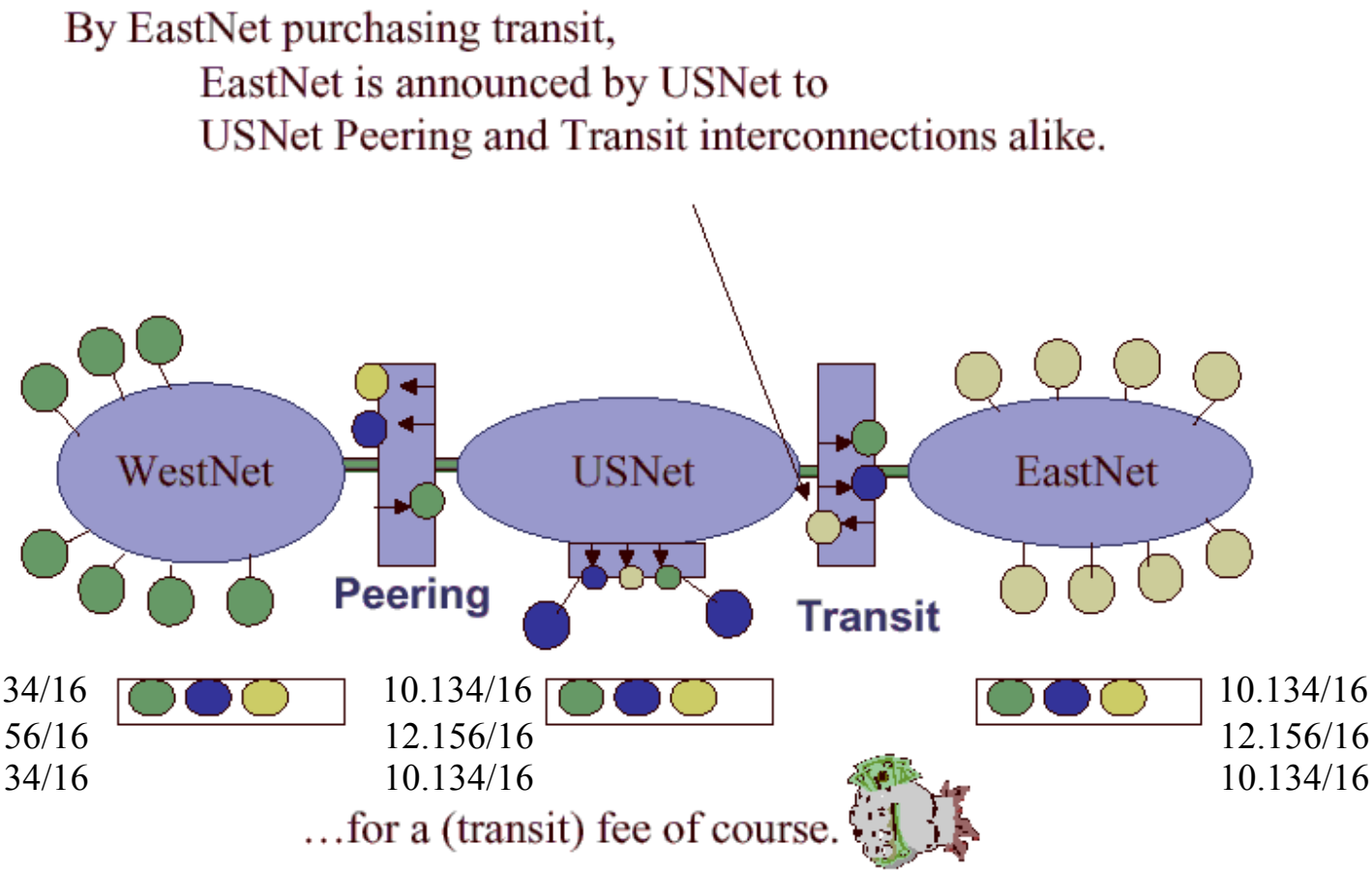
- **Peering consists in establishing a commercial relationship to give subscribers of both providers full connectivity to each other.**



- **No transitivity! WestNet has not access to EastNet, and vice-versa**

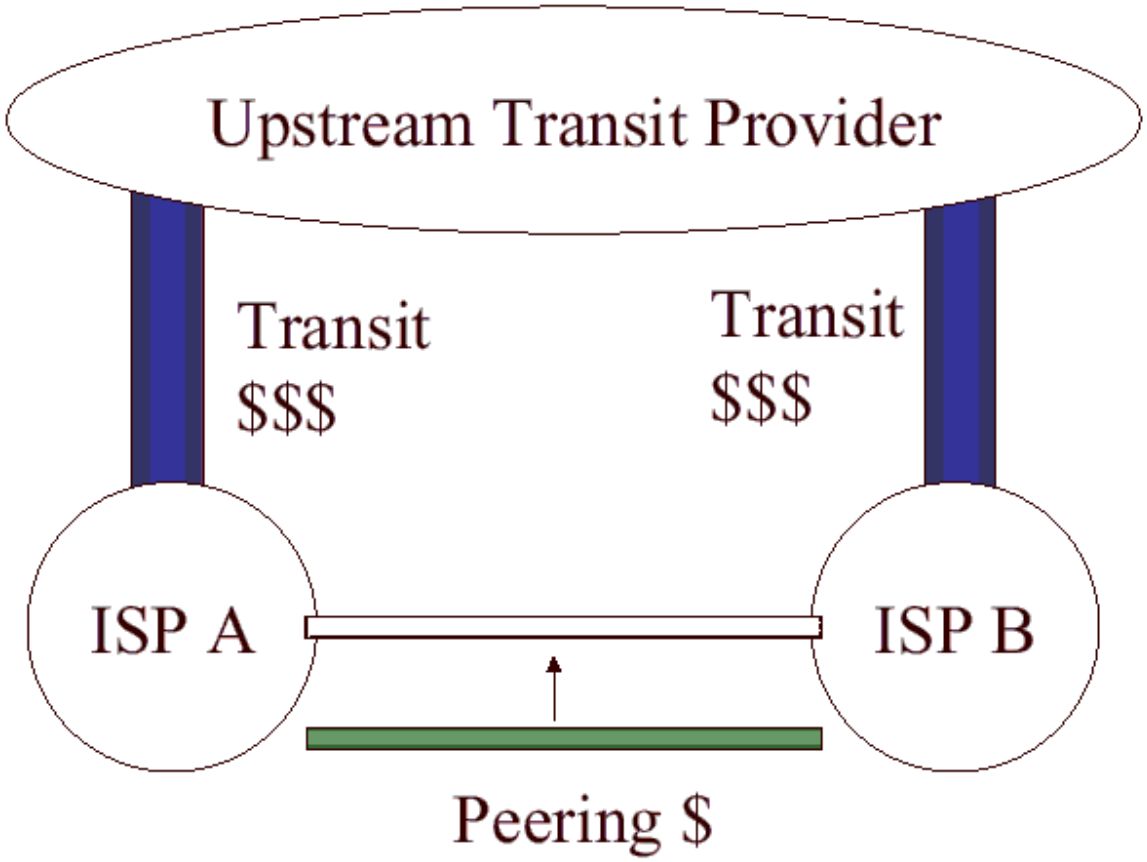
Transit

- **Transit consists in establishing a commercial relationship in which an ISP give (sell) the access of all (or in part, Europe) destinations in its routing table**



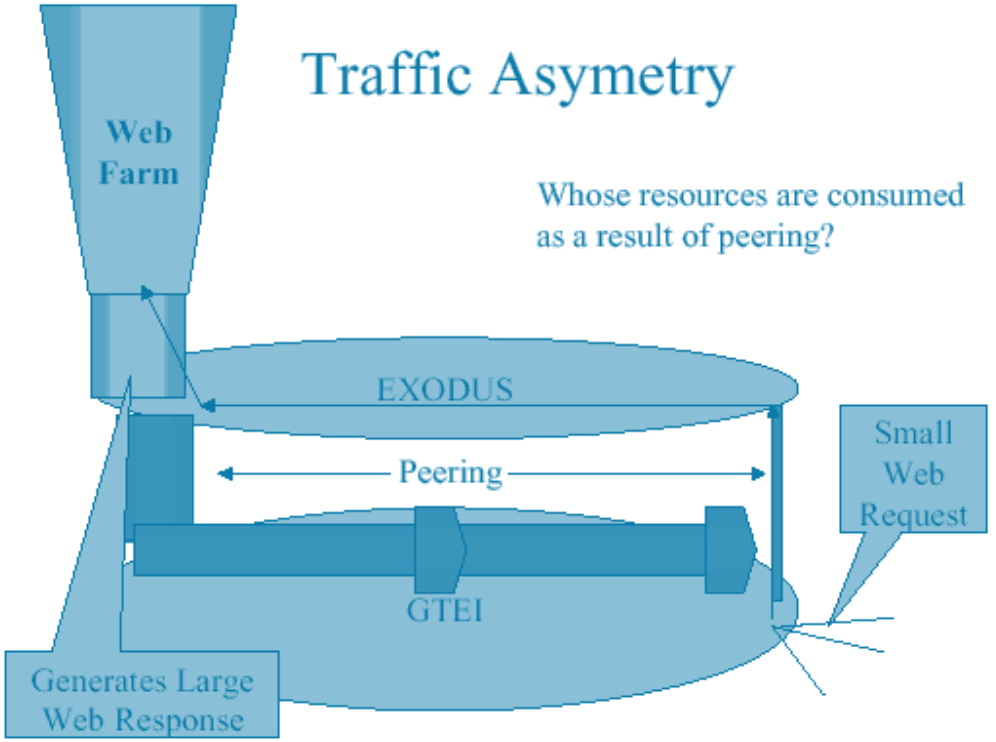
Peering or transit?

- Transit deals can be very expensive ($150000/month for an OC-3 capacity)
- Private peering is often very beneficial



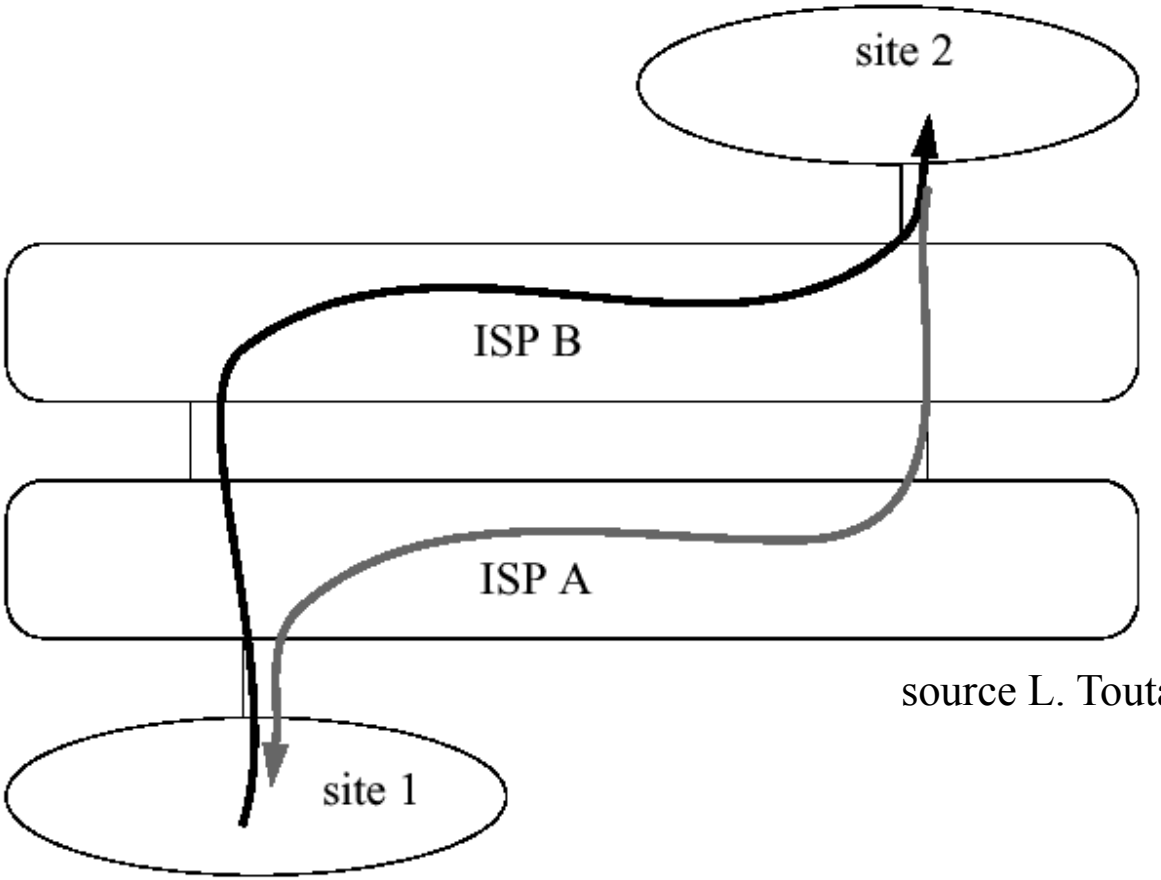
But beware of asymetrical traffic flows

- **An ISP with a lot of interesting contents can consume a lot of resources in its neighboring ISP! A ratio of (4:1) is oftenly put in the transit deal.**



Asymétrie du routage

- ou comment se débarrasser des packets le plus vite possible…



source L. Toutain