Architecture d'un

Routeur

Source: Cours de Jean-Patrick Gelas, Lyon, France

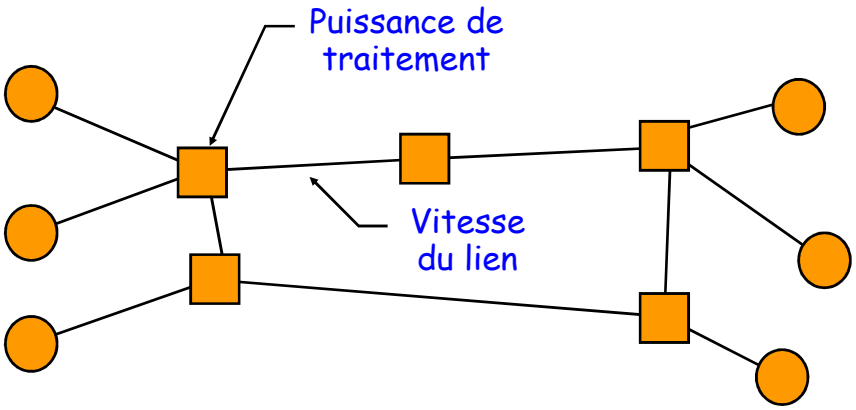
A quoi ressemble un routeur ?

- Routeurs d'accès (ex: xDSL, ISDN)
- Routeurs de coeur
- Commutateurs ATM

2

Performance de routage

Qui décide des performances sur Internet ?

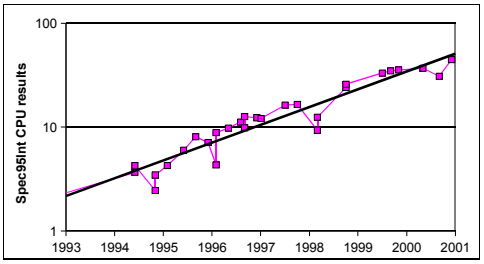


4

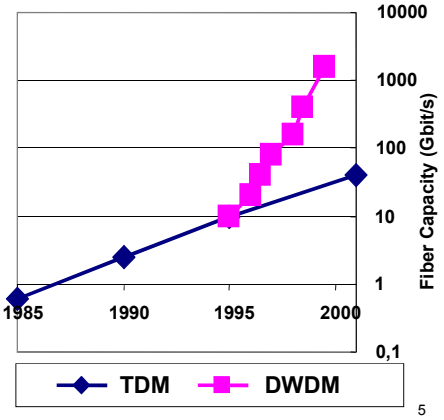
Pourquoi avons nous besoin de routeurs plus rapide ?

Pour éviter qu'ils ne deviennent les goulots d'étranglement.

Capacité de traitement
x2 / 1 an



Vitesse des liens
x2 / 7 mois

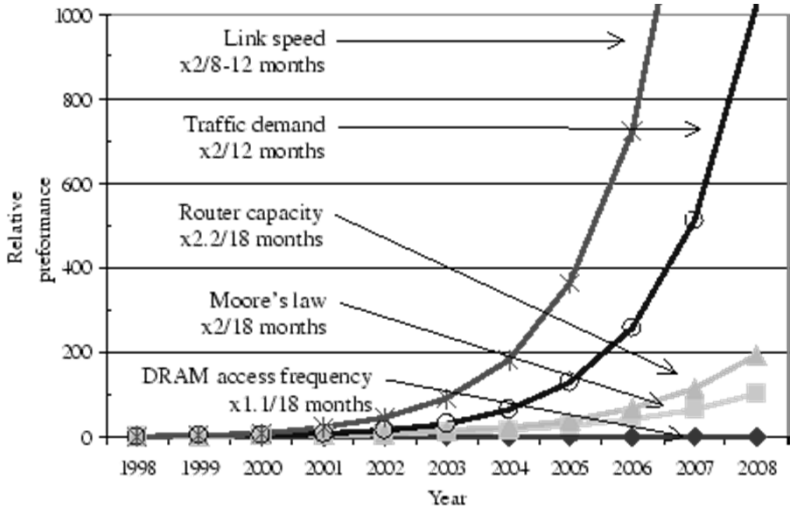


Source: SPEC95Int & David Miller, Stanford.

5

Bande passante mémoire

Le goulot d'étranglement est la vitesse des mémoires... et donc, potentiellement les routeurs.



6

Contraintes de performance

- Avec les débits actuellement rencontrés, un routeur doit effectuer des millions d'opérations par seconde.

Year	Line	Line-rate (Gbps)	40B packets (Mpps)
1998-99	OC12	0.622	1.94
1999-00	OC48	2.5	7.81
2000-01	OC192	10.0	31.25
2002-03	OC768	40.0	125

31.25 Mpps □ 32 ns

… sachant que : DRAM: 50-80 ns ; SRAM: 5-10 ns

7

Traitement par paquet dans un routeur IP

1. Réception d'un paquet arrivant sur une interface d'entrée.
2. Identification du port de sortie à partir de l'adresse de destination du paquet (« *Lookup* »).
3. Manipulation de l'en-tête du paquet : décrement du TTL, mise à jour du *checksum* de l'en-tête (et traitement des options IP).
4. Commutation du paquet vers le port de sortie.
5. Ordonnancement et mise en attente du paquet dans un tampon.
6. Transmission du paquet sur une interface de sortie.

8

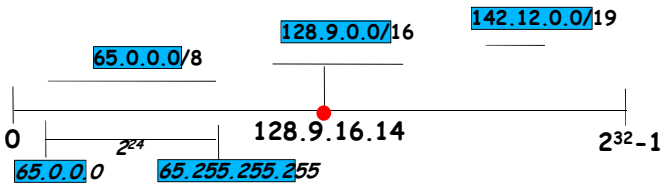
Rappel sur la phase de lookup

- Lookup :
  - Le choix de l'interface de sortie dépend de l'adresse de destination du paquet et du contenu de la table de routage,
  - la recherche dans la table de routage se fait selon le préfixe le plus long : *Best Matching Prefix* (BMP).
  - la rapidité est primordiale.
- Best Matching Prefix :
  - table de routage = paires d'entrées (Préfixe, Interface),
  - pour une @IP donnée, l'entrée avec le plus long préfixe est choisi.

9

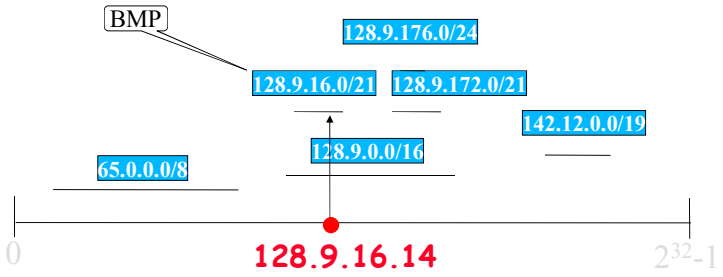
Exemple du « BMP »

Destination IP Prefix	Outgoing Port
65.0.0.0/8 (Taille du préfixe)	3
128.9.0.0/16	1
142.12.0.0/19	7



10

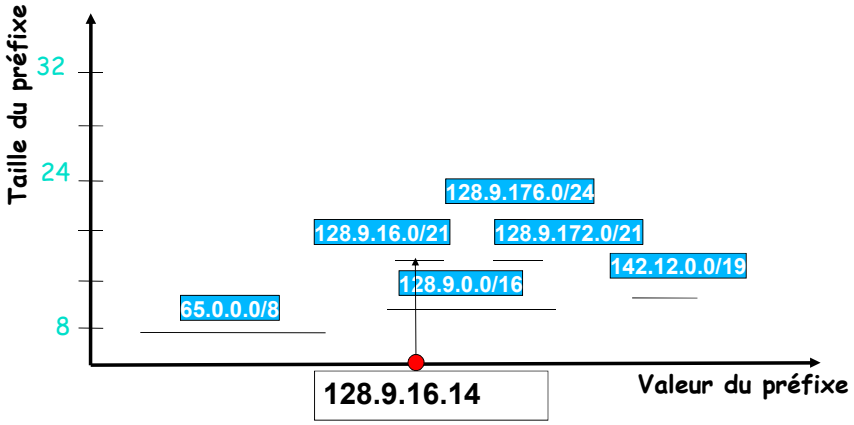
Prefixes can Overlap



11

Difficultés du BMP

Recherche dans un espace à 2 dimensions



Distribution de la taille des paquets et difficultés du *lookup*

- Sur un lien de réseau dorsale (*backbone*) les paquets ne sont pas de taille égale. Difficulté supplémentaire pour automatiser.
  - 75% des paquets < 552 octets,
  - environ 50% des paquets < 44 octets (paquets d'acquittements),
  - 10% des paquets > 1500 octets.
- Difficultés du *lookup*
  - Les tables de routage peuvent avoir des milliers d'entrées,
  - Le préfixe des adresses de destination est de longueur variables (ex: 100101* ou 1* ou 11001100 00110001 01010001)
  - L'adresse de destination peut correspondre à plusieurs préfixes, il faut prendre le plus long.

13

Et dans 10 ans...

- Si on considère qu'il n'y aura pas plus d'opérations par paquet
- Les routeurs seront 200 fois plus rapide...
- mais le trafic aura multiplié par 1000 !
- La solution est peut être de multiplié par 5 le nombre de routeurs => 5 fois plus de consommation (énergie + espace) ?

14

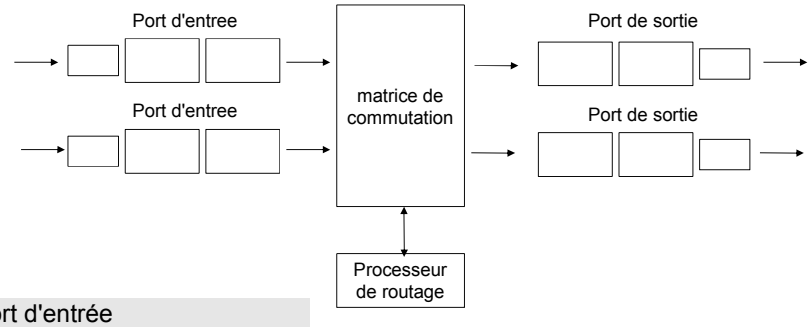
Métriques de performances

- Temps de « lookup »
- Espace de stockage
- Temps d'une mise à jour
- Temps de pré- et post- traitement

15

Architecture

Architecture simplifiée d'un routeur



- Port d'entrée
- Matrice de commutation
- Port de sortie
- Processeur de routage

17

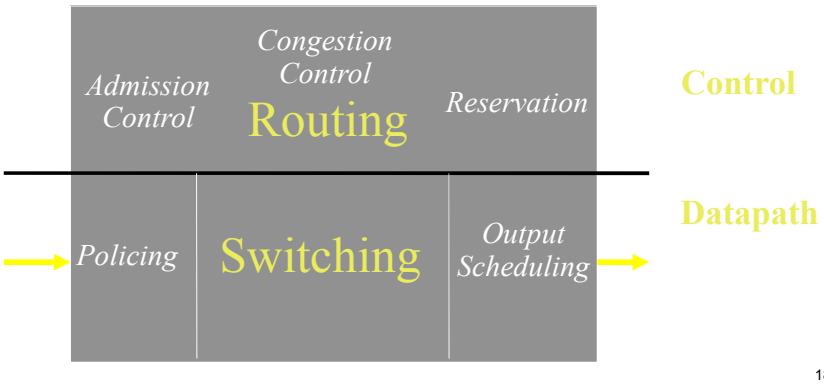
Architecture d'un routeur

Contrôle et Acheminement

Deux niveaux:

Contrôle : Admission, Routage, Réservation.

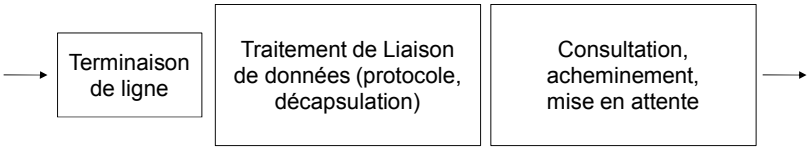
Acheminement : Commutation, Ordonnancement des files.



18

Port d'entrée

- Le port d'entrée est en charge :
  - des fonctionnalités de la **couche physique** qui consiste à relier une liaison d'entrée au routeur,
  - des fonctionnalités de la **couche liaison** de données nécessaires à l'interfonctionnement avec celle de la couche liaison de données à l'autre extremité de la liaison d'entrée.
  - de la **consultation** et **l'acheminement** des paquets entrants dans la matrice de commutation vers le port de sortie approprié.



19

Port d'entrée (suite)

- Dans la pratique, un routeur regroupe plusieurs ports sur une même **carte de communication**.
- Les paquets de contrôle (par exemple porteur d'information pour les protocoles de routage RIP, OSPF ou BGP) sont transmis vers le processeur de routage.
- Le choix du port de sortie se fait en fonction d'informations contenues dans la table de **routage** (élaborée par le processeur de routage).

20

Port d'entrée (suite)

- Dans la pratique un routeur doit pouvoir réaliser des millions de consultations par seconde. L'objectif est que le traitement au port d'entrée se fasse au débit **de la liaison.**
  Exemple : Combien de consultations par seconde pour un lien à 2.5 Gbps avec des paquets de 256 octets ? Réponse : 1 million de consultations par seconde.
- Différentes techniques ont été développées pour augmenter la vitesse des consultations.
  - Les mémoires à contenu adressable (CAM, *Content Addressable Memory*) peuvent prendre en charge des adresses 32bits et retourner l'entrée correspondante de la table en un laps de temps constant.
  - ….
- Dès que le port de sortie d'un paquet à été identifié, celui-ci peut entrer dans la matrice de commutation.

21

Matrice de commutation

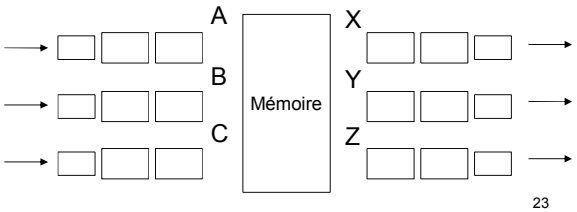
Logée au coeur du routeur, la matrice de commutation assure le transfert des paquets entre le port d'entrée et le port de sortie approprié.
L'opération de commutation peut se faire de trois façons :

- Commutation par action sur les mémoires
- Commutation par bus
- Commutation via un réseau d'interconnexion (commutation *crossbar*)

22

Commutation par action sur les mémoires

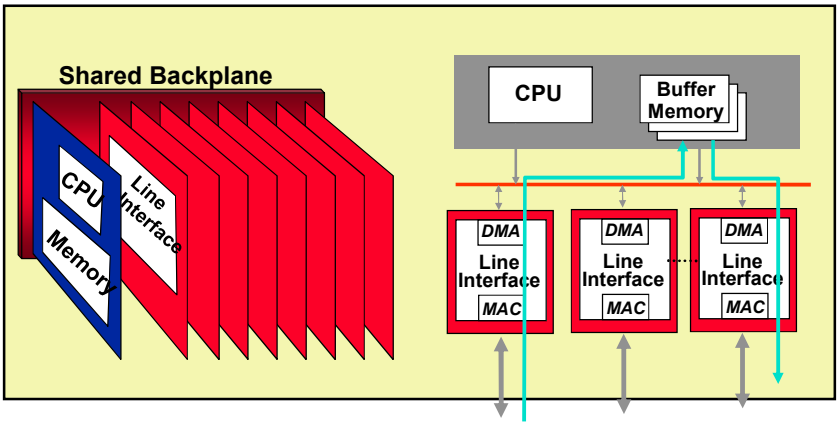
- Les premiers routeurs étaient de simple ordinateurs.
- Le port d'entrée signale l'arrivée d'un paquet. Une copie du paquet est placée dans la mémoire du processeur qui extrait l'adresse de destination, consulte la table et copie le paquet dans le tampon du port de sortie approprié.
- La consultation de l'adresse de destination et la commutation du paquet peut etre placés dans les cartes de communication



o Catalyst 8500 Cisco
o 1200 Bay Networks Accelar

23

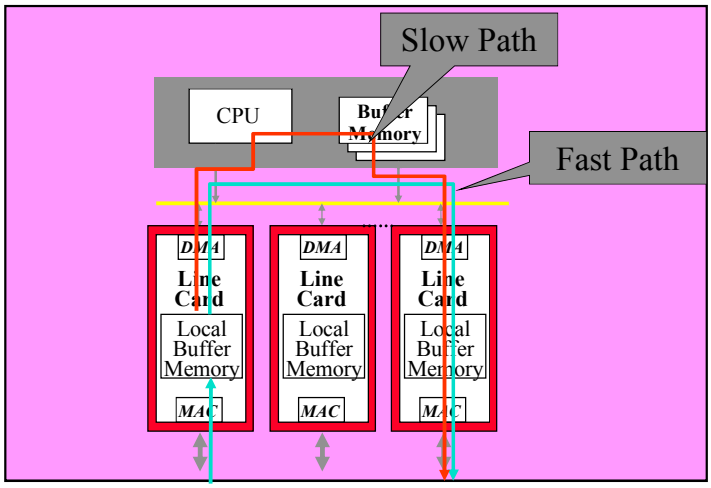
Routeur IP (1ère génération)



Le goulot d'étranglement peut être le CPU, l'adaptateur, ou le bus d'E/S.
Qu'est ce qui coute ? (bus, mémoire, l'interface, cpu)

24

2nd génération de routeur

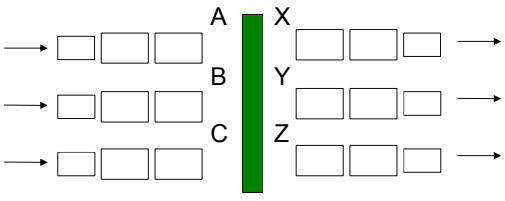


Intelligence embarqué sur les cartes d'interface.

25

Commutation par bus

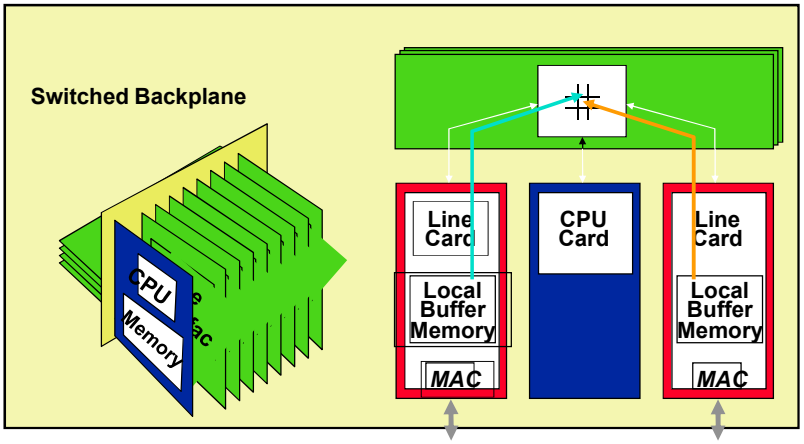
- Le port d'entrée transfère les paquets directement au port de sortie grâce à un bus partagé, sans passer par le processeur de routage.
- Puisqu'il s'agit d'un bus partagé, un paquet seulement peut être transféré à la fois.
- C'est le débit de ce bus qui détermine la vitesse de commutation du routeur.



o 1900 Cisco (Packet Exchange Bus, 1Gbps)
o CoreBuilder 5000, 3com (Packet Channel, 2 Gbps)

26

3ème génération de routeur (switch/router)

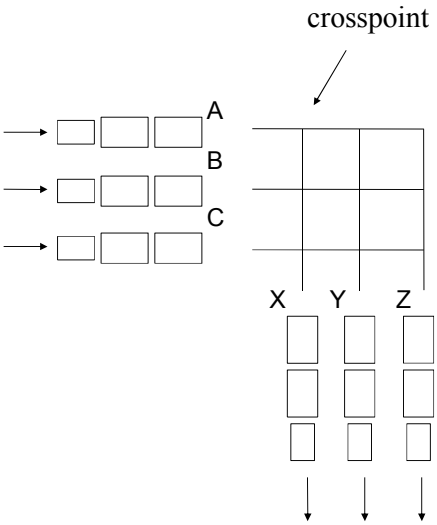


Fournit des chemin parallèles.

27

Commutation crossbar

- Pour s'affranchir des limitations de débit associés au partage d'un bus commun, on peut utiliser un réseau d'interconnexion de type ***crossbar***.
- Composé de 2*N* bus reliant *N* ports d'entrés à *N* ports de sorties.
- Autorise de nombreux ports (densité) et de haut débits.
- Le plus simple des *space division switch*.
- Dans une matrice crossbar, pendant chaque instant de commutation, une et une seule entrée est connectée à une sortie (un seul *crosspoint* actif).
- Pas de bloquage interne.

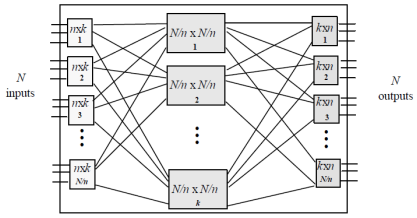


o 12000 Cisco (60Gbps)

28

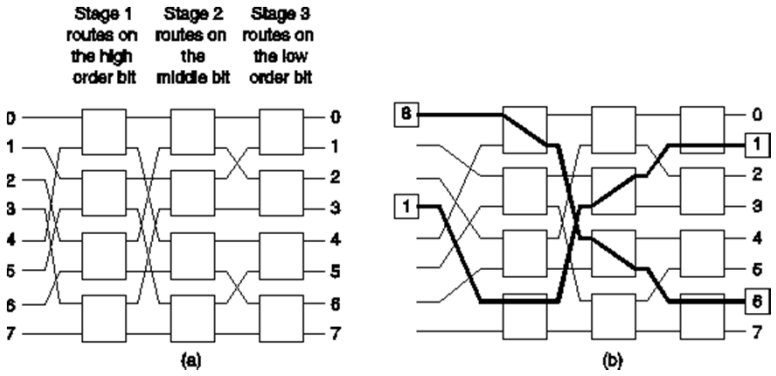
Commutation spatiale multi-étages : *Multistage crossbar*

- On peut économiser des *crosspoints* si on peut y attacher plus d'une ligne d'entrée.
- Supporte mieux la passage à l'échelle que le crossbar...
- Peut générer des bloquages internes (à moins d'avoir suffisamment d'étage intermédiaires).



29

Commutateur Banyan

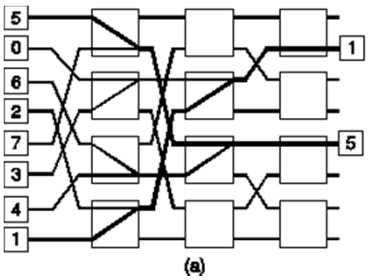


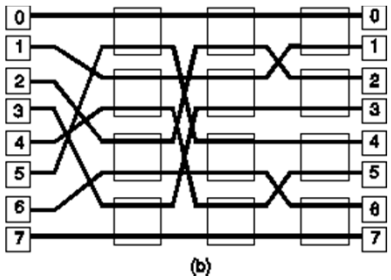
Banyan 8x8 à trois étages

(6 = 110b 1 = 001b)

30

Banyan : Conflits





Lorsque deux cellules se dirigent vers la même sortie simultanément... collisions !

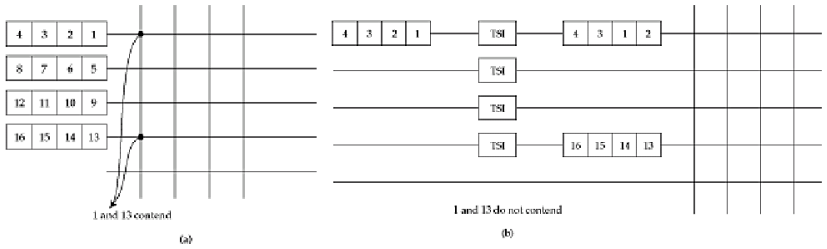
Ici aucun conflit.

31

Techniques alternatives pour économiser des crosspoints

Time space switching

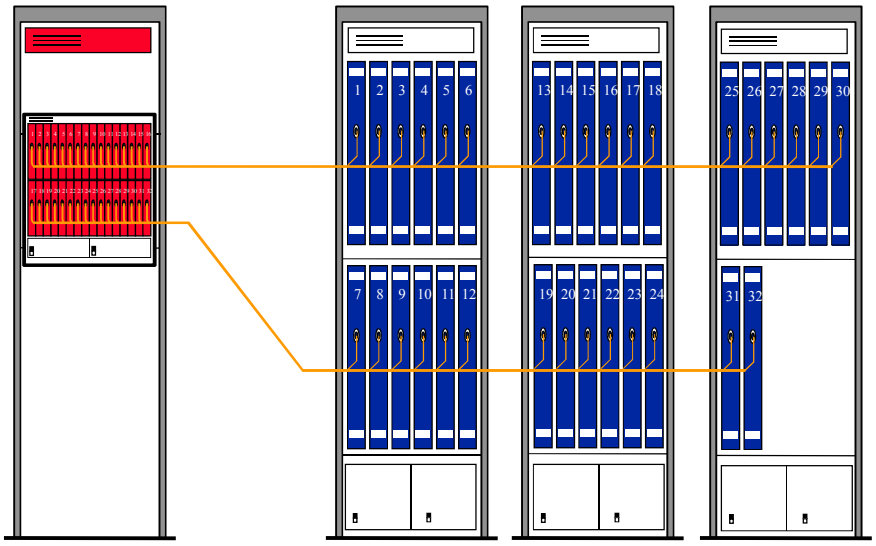
- On place à chaque entrée d'une matrice crossbar un TSI.



32

Fourth-Generation Switches/Routers

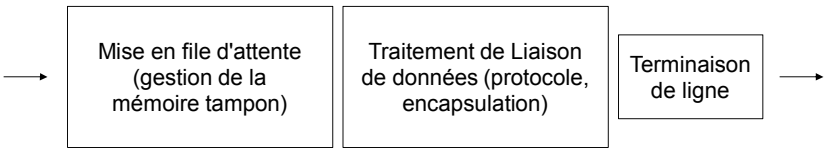
Clustering and Multistage



33

Port de sortie

- Le port de sortie prend les datagrammes stockés dans sa mémoire et les transmet à la liaison de sortie.
- Mise en file d'attente si la matrice de commutation livre les paquets au port de sortie à une vitesse supérieure au débit de la liaison de sortie.



34

Processeur de routage

- Le processeur de routage est en charge de :
  - l'exécution des protocoles de routage (RIP, OSPF, BGP),
  - de la mise à jour des informations et des tables de re-acheminement (*forwarding table*),
  - ainsi que de certaines fonctions d'administration de réseau.

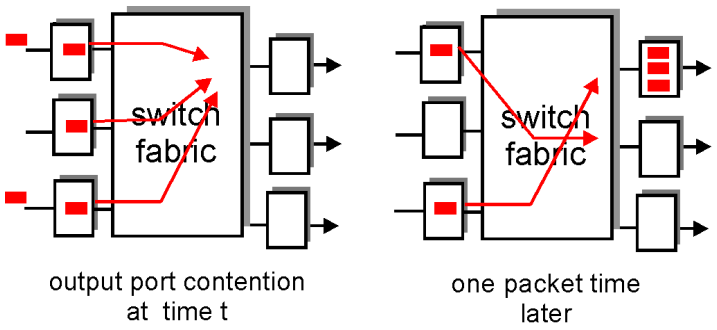
35

Files d'attentes

- Des files de paquets peuvent se former au niveau des ports d'entrée et de sortie.
- Si la taille d'une file augmente trop, la disponibilité du tampon mémoire se réduit, ce qui risque de conduire à la perte de certains paquets.
- L'endroit exact où a lieu la perte dépend aussi bien de la densité du trafic que de la vitesse relative de la matrice de commutation et du débit de la liaison.

36

File d'attente au niveau du port de sortie



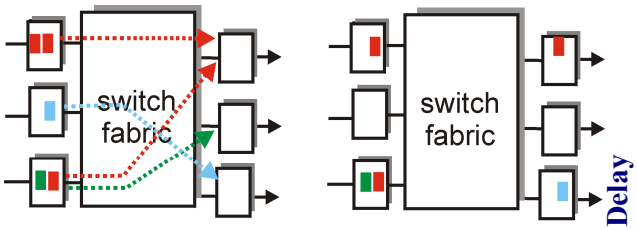
output port contention at time t

one packet time later

- Un **gestionnaire de paquets** doit déterminer quel paquet de la file doit être transmis :
  - FIFO
  - WFQ (Weighted Fair Queing) partage la liaison de sortie de façon équitable entre les différentes connexions ayant des paquets en attente.
- L'ordonnancement des paquets joue un rôle crucial dans les **garanties de qualité de services**.

37

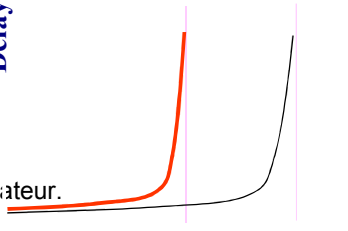
Blocage de tête de ligne (HOL, *Head of the Line*)



output port contention at time t - only one red packet can be transferred

green packet experiences HOL blocking

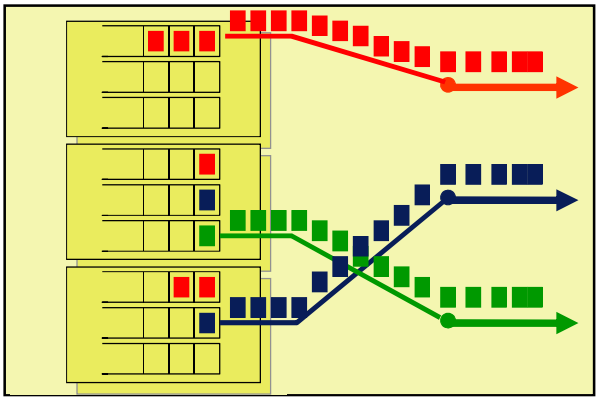
Blocage de tête de ligne sur une file d'entrée du commutateur.

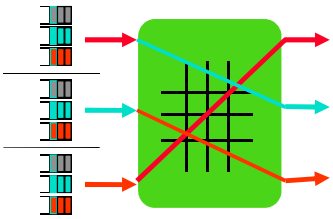


38

Solution : Virtual Output Queues (VOQ)

- Chaque entrée maintient autant de files d'attente que de port de sorties.
- Associé à un algorithme efficace d'ordonnancement on peut atteindre 100% de la bande passante.





39

Conclusion

- Nous avons décrit l'intérieur d'un routeur dans lequel nous avons identifié 4 éléments :
  - Port d'entrée : couche physique et liaison, consultation et acheminement des paquets entrants dans la matrice de commutation.
  - Matrice de commutation : un réseau à l'intérieur du réseau (mémoire, bus, crossbar).
  - Port de sortie : assure les fonctions inverse du port d'entrée.
  - Processeur de routage : en charge de l'exécution des protocoles de routage, de la mise à jour des tables et d'administration.

40

Conclusion

- La différence de rapidité d'évolution des différents composants implique un goulot d'étranglement pour le futur,
- Nécessite de trouver des solutions alternatives pour répondre au décalage entre la demande croissante de trafic et la capacité des routeurs
- Le traitement de l'en-tête est de plus en plus confié à des composants dédiés (ASICs, FPGAs ou network processors).

41