

Transmission, archivage et réseaux

HESAV - TRM BSc-S5



Mercredi 07.12.2016

Programme

- Représentation de l'information
- Milieu de transmission
- Technologies de communication
- Introduction à la notion de réseau
- Supports de stockage des données
- Sauvegarde
- Compression

Objectifs

- Comprendre les divers milieux de transmission et leur utilisation, voire leurs limitations
- Comprendre la notion de protocole de communication
- Appréhender la complexité d'un réseau et ses mécanismes d'adressage
- Savoir protéger vos données
- Notions de sécurité

Quelques notions de base

- **Bit**

Quantité minimale d'information pouvant être transmise par un message, peut prendre la valeur **0** ou **1**

- **Octet (byte)**

Plus petite unité adressable par un ordinateur moderne, est composé de 8 bits, valeur entre 0 et 255 (2^8-1)

- **Multiples usuels de l'octet**

1 ko
 10^3 octets



1 Mo
 10^6 octets



1 Go
 10^9 octets



1 To
 10^{12} octets



1 Po
 10^{15} octets



Représentation de l'Information

```
!"#$%&'()*+,-./
0123456789:;<=>?
@ABCDEFGHIJKLMNO
PQRSTUVWXYZ[\]^_
`abcdefghijklmno
pqrstuvwxyz{|}~
```

- Le texte
 - **ASCII** - American Standard Code for Information Interchange
Chaque lettre est représentée sur 7 bits → 127 possibilités
 - **ASCII étendu**
Utilisation d'un 8^{ème} bit permettant de passer à 256 caractères
Différentes versions régionales: ISO 8859-1, ISO 8859-15,...
Problèmes de compatibilité entre différents systèmes
 - **UTF-8** - Universal Character Set Transformation Format
Utilisation de 1 à 4 octets, rétro-compatible avec ASCII
Permet de représenter plus d'un million de caractères

Représentation de l'Information

- Les images

Les images numériques peuvent être soit:

- **matricielles**

tableau 2D de pixels

- **vectérielles**

chaque élément de l'image est représenté par une formule géométrique (courbe, cercle, polygone, etc.)



Représentation de l'Information

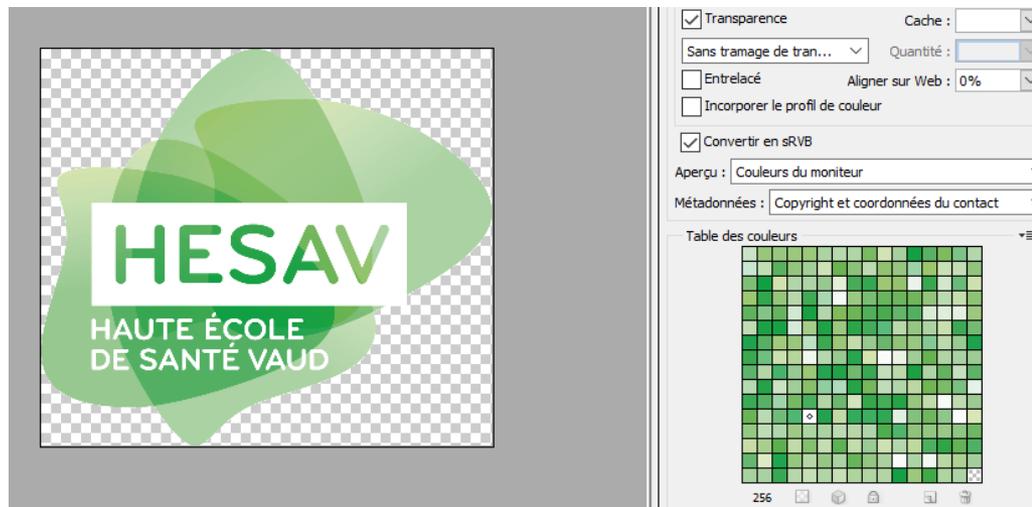
- Les images matricielles
 - **Noir/Blanc**
Chaque pixel est représenté par un bit (noir ou blanc)
 - **Niveau de gris**
Chaque pixel prend une valeur correspondant à l'intensité lumineuse, généralement entre 0 et 255, soit 1 octet



Représentation de l'Information

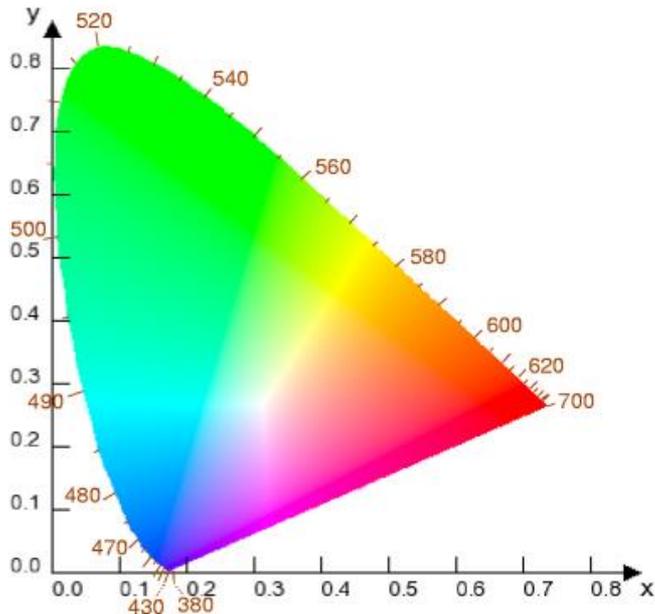
- Les images matricielles couleurs
 - **Image à palettes**

Une palette de couleurs est attachée à l'image, généralement composée de 255 couleurs, chaque pixel fait référence à l'index d'une couleur dans la palette



Représentation de l'Information

- Les images matricielles couleurs
 - Couleurs «vraies» - 24 bits - 16 millions de couleurs
Chaque pixel est représenté par 3 octets,
1 par couleur primaire (RVB – rouge, vert, bleu)



iPhone 7 plus – capteur de 12 mégapixels
 $12'000'000 \times 24 \text{ bits} = 36 \text{ Mo}$
→ compression

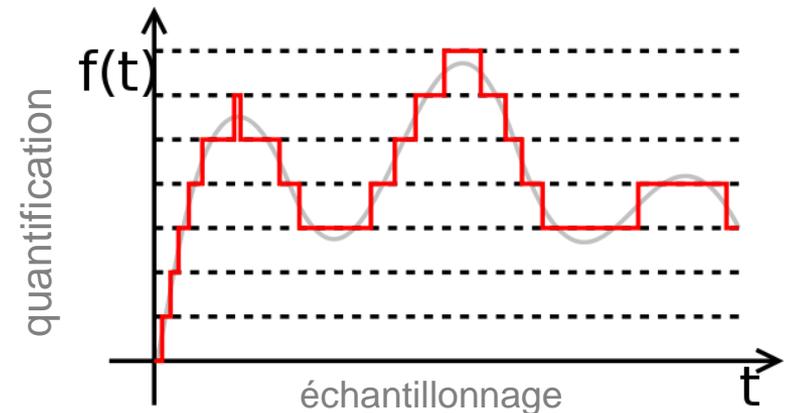


Représentation de l'Information

- **Audio**

- Echantillonnage + Quantification
- Dans le cadre du CD, le signal audio est découpé 44'100 fois par seconde (échantillonnage), chaque segment va prendre une valeurs (quantification) représentée par un nombre entre 0 et 65535 (16 bits – 2 octets)

CD: une musique stéréo de 5 minutes
 $2 \times 5 \times 60 \times 441000 \times 16 = \sim 53 \text{ Mo}$
→ compression



Théorie de l'Information

- 1948 - Claude Shannon - *A Mathematical Theory of Communications*
Théorie mathématique (probabiliste) permettant de quantifier la quantité d'information contenue dans un message échangé entre une source et un récepteur

*«Je l'ai vu, dis-je, vu, de mes propres yeux vu, ce qui s'appelle vu...»
Molière*

contient moins d'information que la phrase pourtant plus courte:

«J'ai vu Paul»

Théorie de l'Information

Domaines principaux de la théorie de l'information:

- la mesure quantitative de **redondance** d'un texte
 - **compression** de données si on réduit la redondance
 - **somme de contrôle** ou **code correcteur** si on ajoute de la redondance
- la **cryptographie**

Milieu de transmission

Un milieu de transmission est caractérisé par :

- Sa largeur de bande
fréquence maximale pouvant transiter dans le milieu sans affaiblissement excessif
- Son rapport signal-sur-perturbation
distance entre le niveau du signal et le niveau de *bruit* dans le canal de transmission

Milieu de transmission

- Une plus grande largeur de bande permet de placer plus d'informations dans le canal : l'*autoroute* est plus large
- Un meilleur rapport signal-sur-bruit permet de distinguer des nuances de signal plus fines
- Shannon a exprimé cette dépendance en définissant la capacité d'un canal:

$$C = W \cdot \log_2 \left(1 + \frac{S}{N} \right)$$

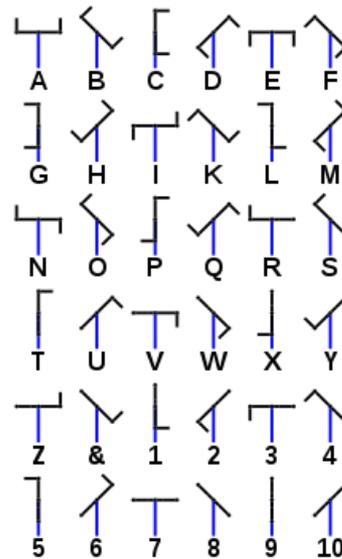
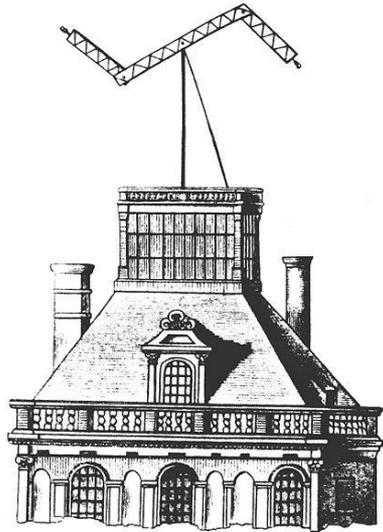
W = Largeur de bande [Hz]

S / N = rapport signal-sur-bruit

Les grandes dates

Transmissions point-à-point

- Tamtams, signaux de fumée, ...
- Télégraphe optique (Claude Chappe 1792)



Une tour tous les 10-15 km
Utilisable de jour uniquement
Premier message entre Paris et
Lille, 9 minutes pour transmettre un
symbole via 15 stations (200 km)

Les grandes dates

Transmissions point-à-point

- Télégraphe électromagnétique (Morse 1840)
- Câble transatlantique (1866) 4200 km, 7000 t
- Téléphone (1876) Alexander Bell et Elisha Gray
- Radio (1900) Nikola Tesla
- Guglielmo Marconi (1901)
Première liaison radio transatlantique

Les technologies plus récentes

- Transmissions
 - Multiplexage fréquentiel
 - Câble coaxial
 - Multiplexage temporel
 - Fibre optique (1970)
 - Transmission numérique par codage
 - Modulation par impulsions et codage MIC
 - Code Division Multiple Access CDMA
UMTS / LTE / WiFi

Acheminement d'information

- On distingue entre commutation de **circuits**, de **paquets**, et de **messages**
- Le **circuit** est une liaison fixe permanente pendant toute la durée de la liaison (ex. téléphone classique)
- Le **paquet** est un élément de données de longueur arbitraire (ex. IP)
- Le **message** est un élément de communication complet (ex. e-mail)

Acheminement d'information

- Dans une commutation de **circuits**, l'information est acheminée pratiquement en temps réel, au maximum du débit utilisable
- Dans une commutation de **paquets**, l'information à transmettre est divisée en paquets largement indépendants les uns des autres, chacun étant acheminé séparément
- Le **message** est quand à lui acheminé d'un bloc

La notion de réseau

- Toutes les technologies vues jusqu'à présent permettait la transmission:
 - À vue
 - D'une extrémité d'un fil à l'autre
 - De l'antenne de l'émetteur au récepteur
- Transmettre ne suffit pas : il faut aussi acheminer !
- L'ancêtre du réseau est le centraliste téléphonique



Situation actuelle

- Coexistence de nombreux réseaux; réseaux d'accès, réseaux mobiles, de transit, de diffusion, ...
- Distinction entre transmission numérique et analogique s'estompe
- L'information est de plus en plus représentée sous forme numérique
 - Facilité de transmission
 - Facilité d'acheminement

Le problème

- Une communication requiert :
 - Un (ou plusieurs) support physique
 - Une identification des partenaires
 - Un acheminement
 - Une garantie de qualité de transmission (indépendamment du support)
 - L'attention des partenaires
 - La confidentialité
 - La sécurité, l'authenticité

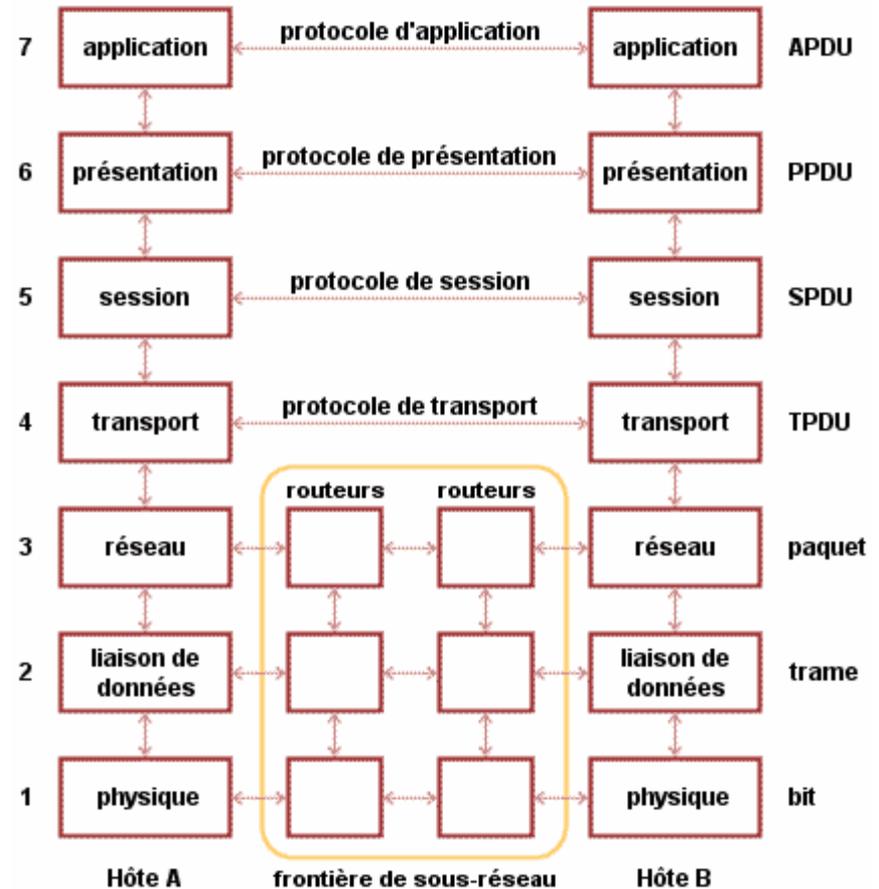
Protocole en couches

- La complexité des problèmes à résoudre les rend quasi impossibles à traiter en bloc
- On a donc imaginé une structuration des problèmes
- C'est le modèle à couches d'un protocole de communications
- TCP-IP, ou le modèle OSI, sont les principaux exemples de protocoles structurés en couches

Modèle OSI

Open Systems Interconnection

1. Physique
2. Liaison de Données
3. Réseau
4. Transport
5. Session
6. Présentation
7. Application

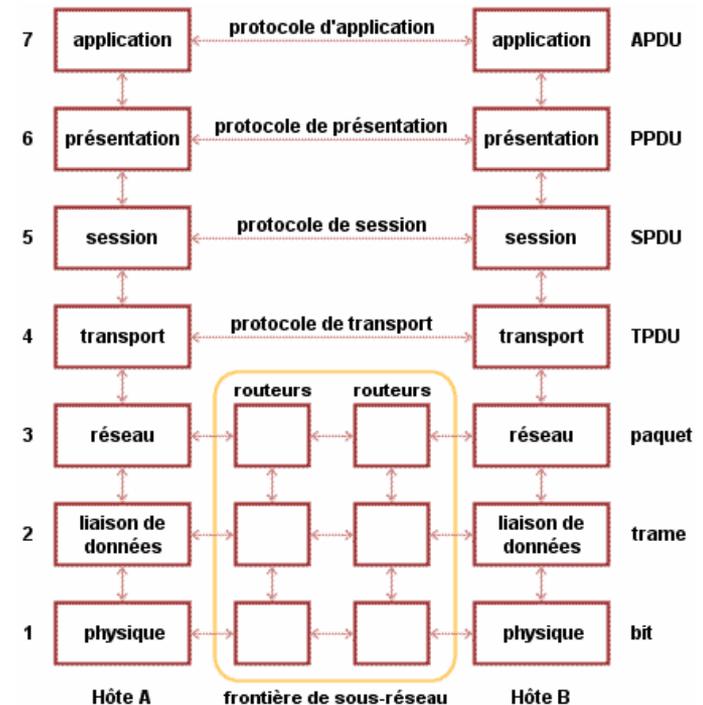


Modèle OSI

Open Systems Interconnection

1. Physique

La couche physique s'occupe de la transmission des bits de façon brute sur un canal de communication. Cette couche doit garantir la parfaite transmission des données (un bit 1 envoyé doit bien être reçu comme bit valant 1)



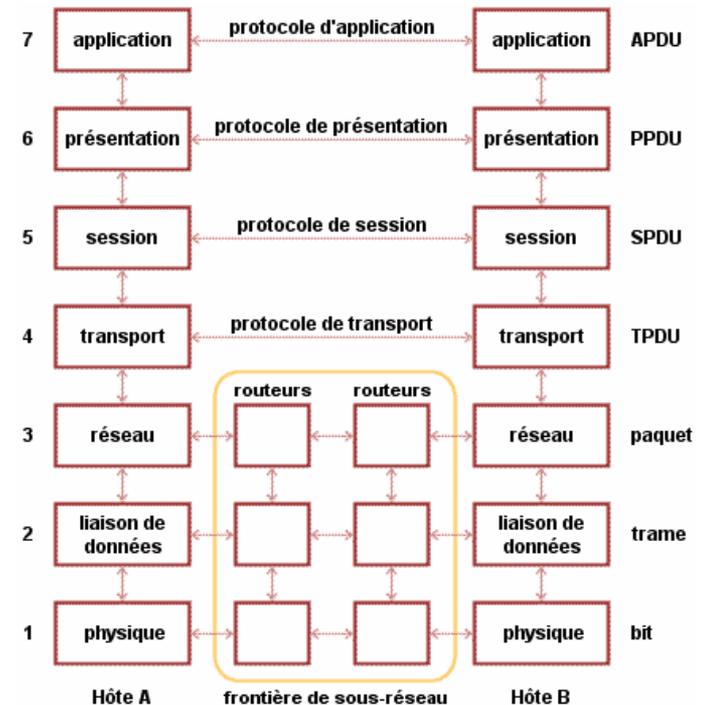
Modèle OSI

Open Systems Interconnection

2. Liaison de données

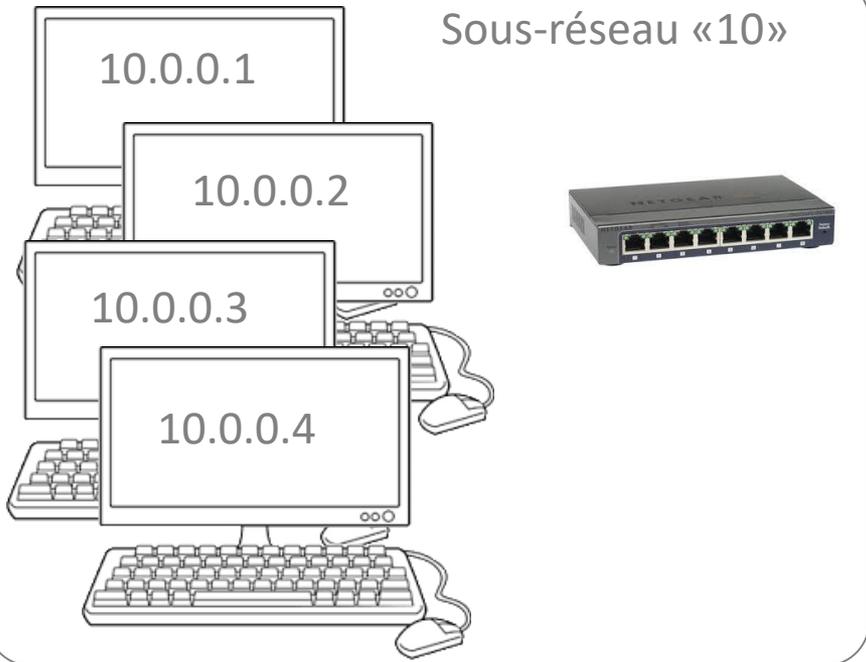
On regroupe plusieurs bits pour former une trame, un «bout» de message avec un émetteur et un récepteur.

Cette couche doit garantir que la trame reçue ne contient pas d'erreur de transmission

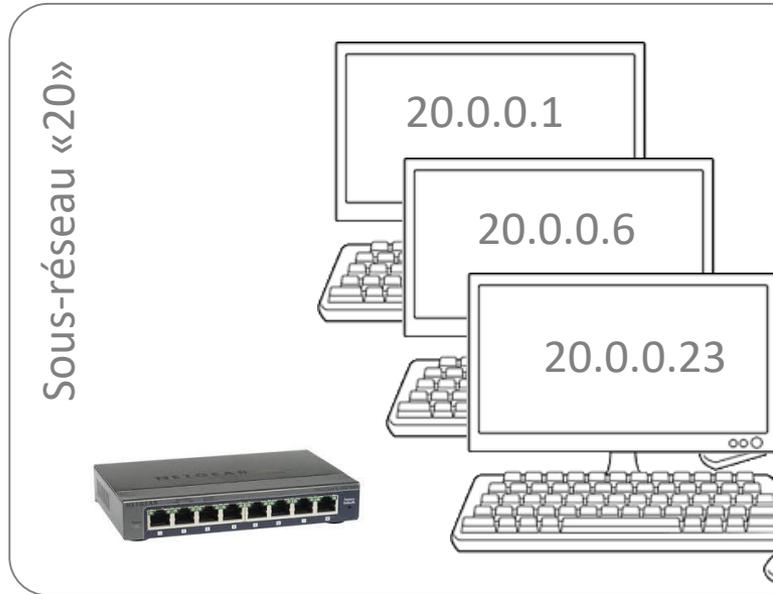


Routage de paquets - Exemple

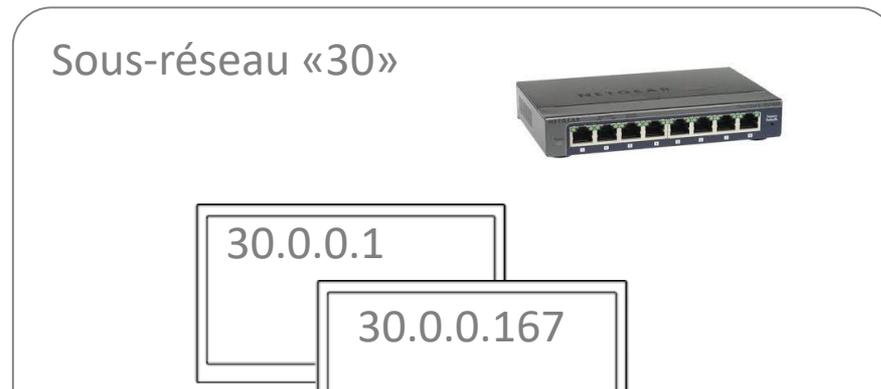
Sous-réseau «10»



Sous-réseau «20»

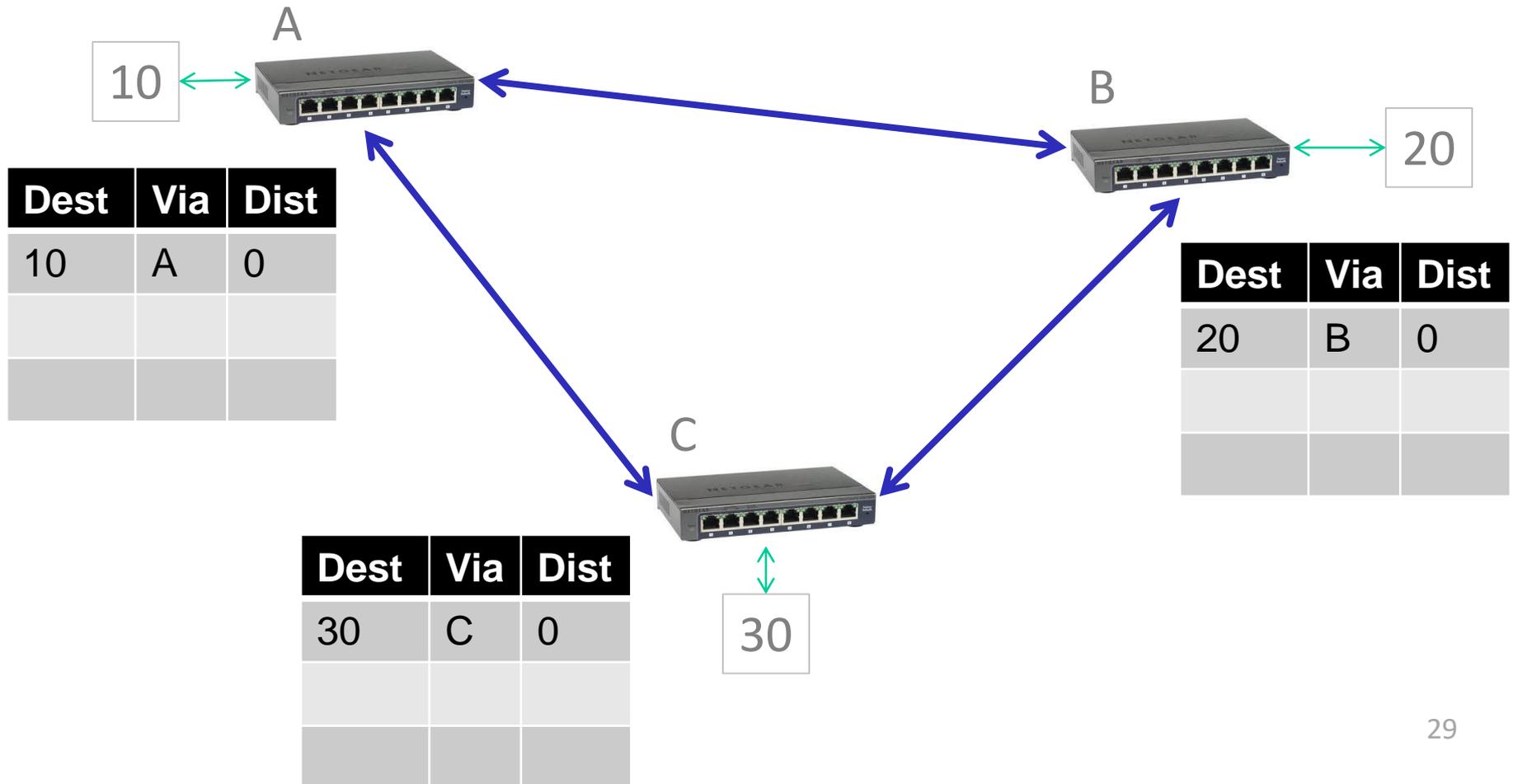


Sous-réseau «30»



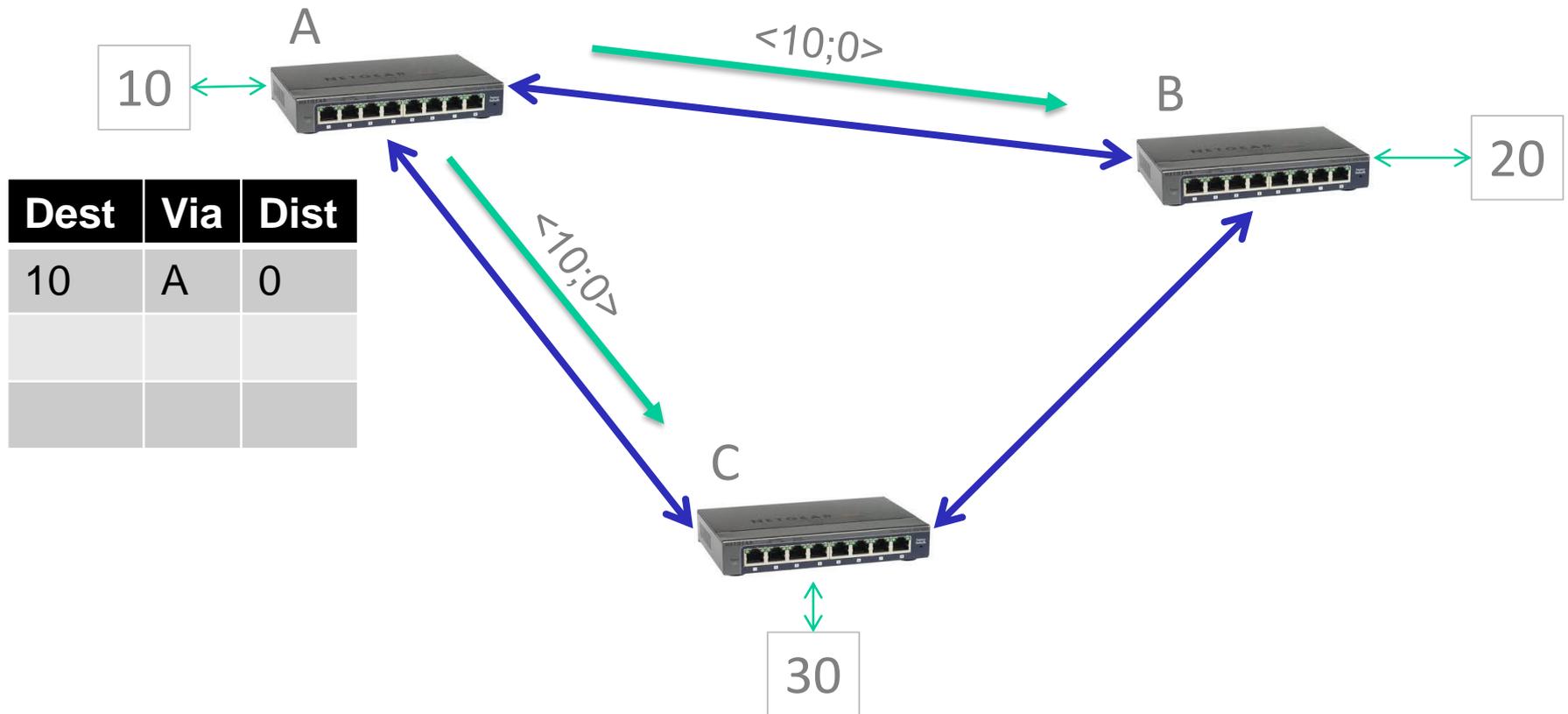
Routage de paquets

Etat initial

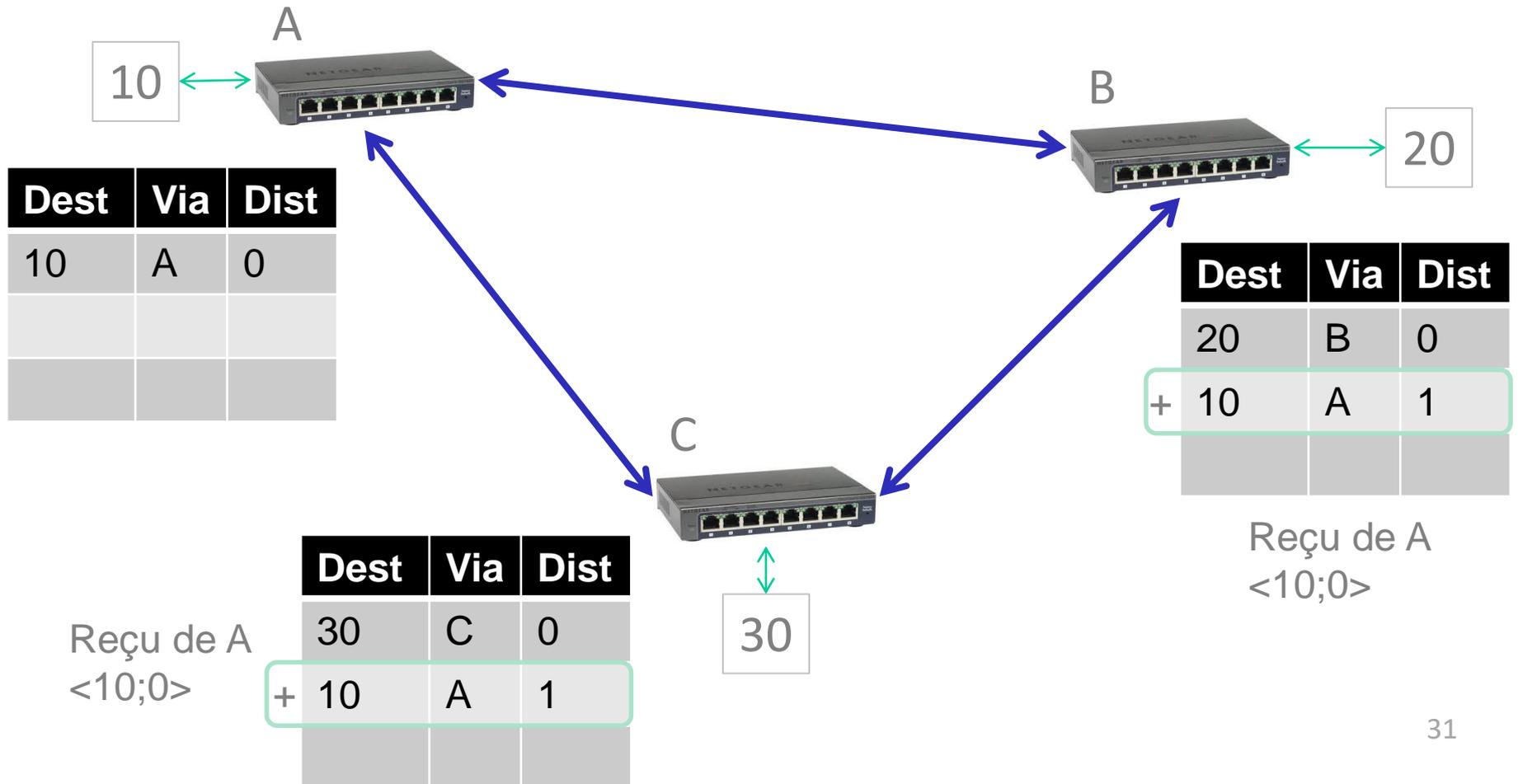


Routage de paquets

Annnonce de A à ses voisins

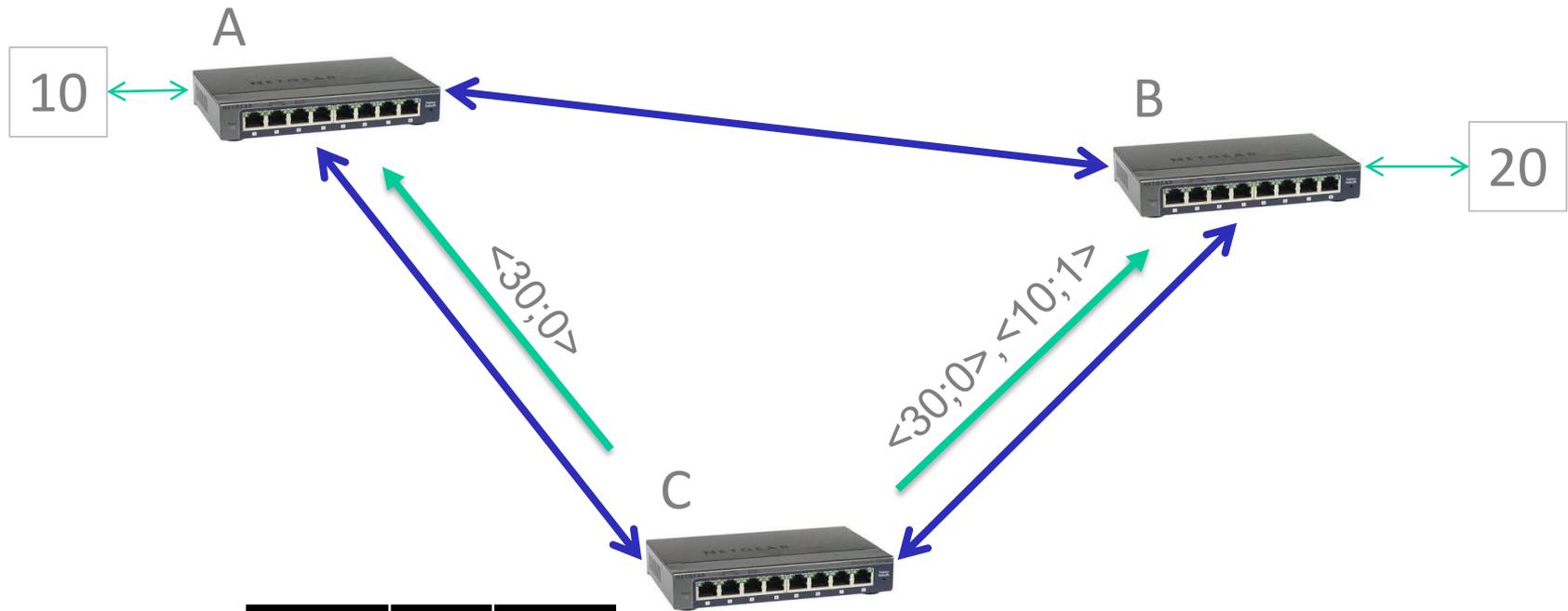


Routage de paquets



Routage de paquets

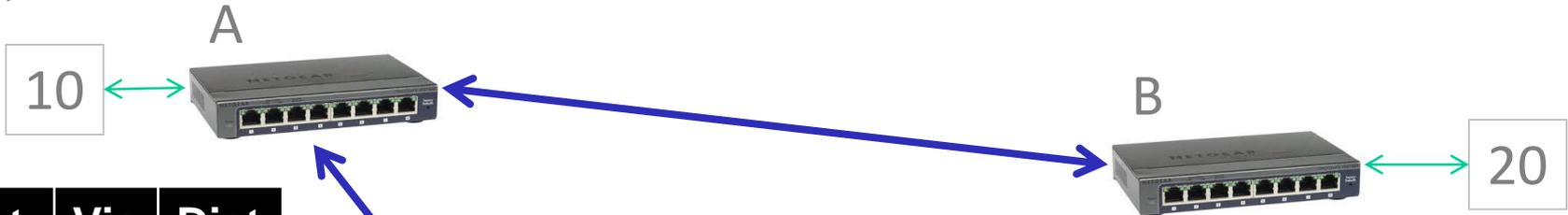
Annnonce de C à ses voisins



Dest	Via	Dist
30	C	0
10	A	1

Routage de paquets

Reçu de C
<30;0>



Dest	Via	Dist
10	A	0
+ 30	C	1

Dest	Via	Dist
20	B	0
10	A	1
+ 30	C	1
? 10	C	2

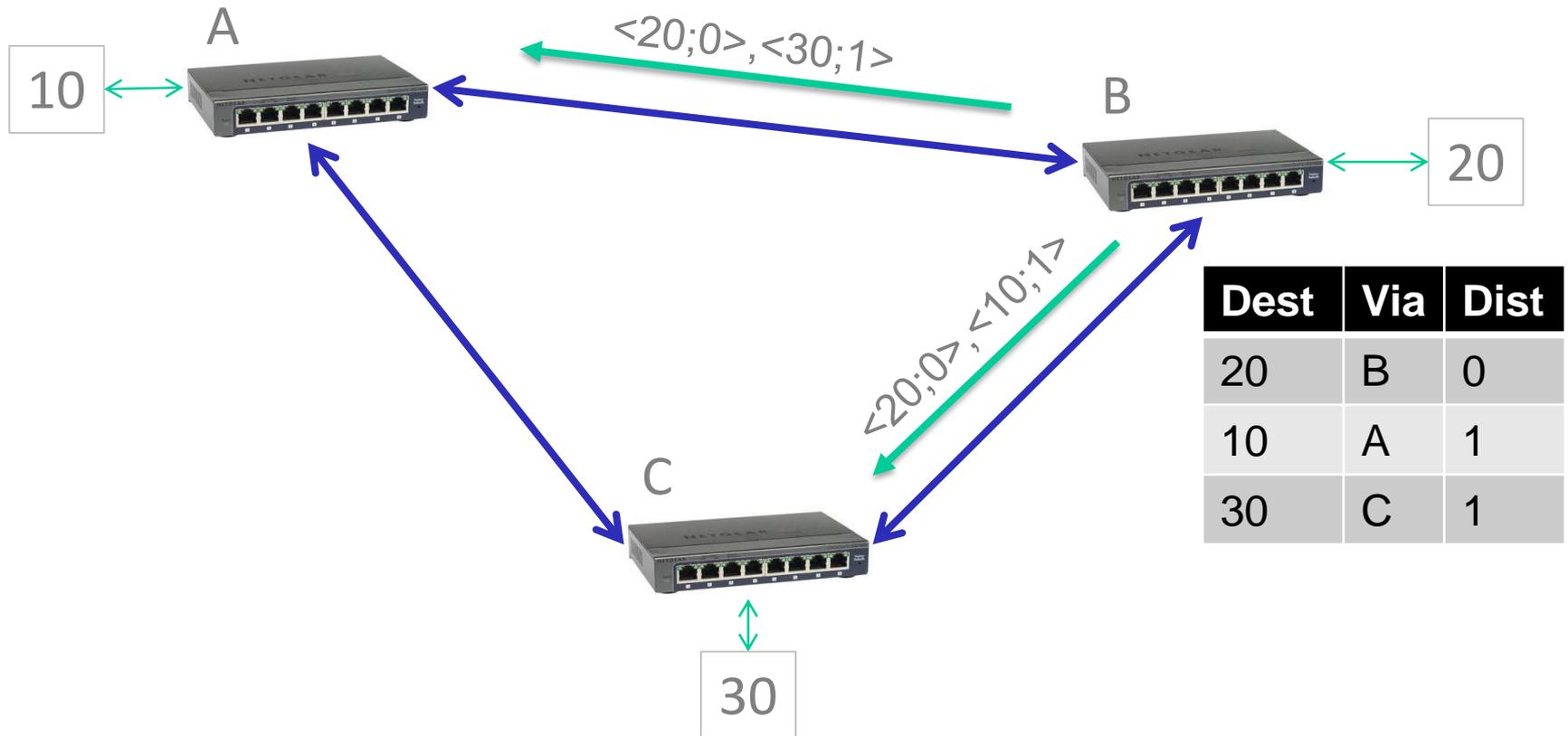
Dest	Via	Dist
30	C	0
10	A	1

30

Reçu de C
<30;0>, <10;1>

Routage de paquets

Annnonce de B à ses voisins



Dest	Via	Dist
20	B	0
10	A	1
30	C	1

Routage de paquets

Reçu de B
<20;0>, <30;1>

10



Dest	Via	Dist
10	A	0
30	C	1
+ 20	B	1
? 30	B	2

B



20

Dest	Via	Dist
20	B	0
10	A	1
30	C	1

C



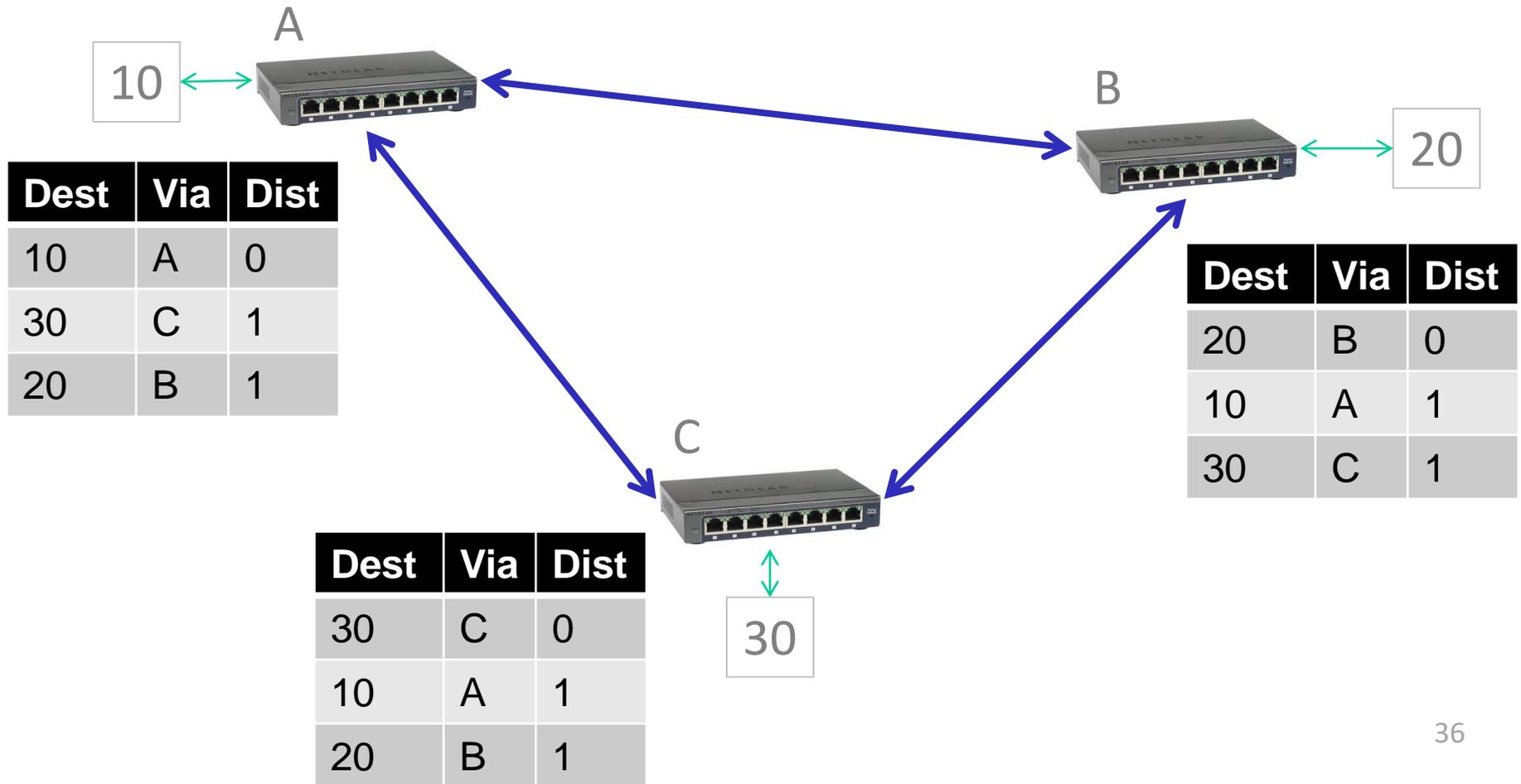
30

Reçu de B
<20;0>, <10;1>

Dest	Via	Dist
30	C	0
10	A	1
+ 20	B	1
? 10	B	2

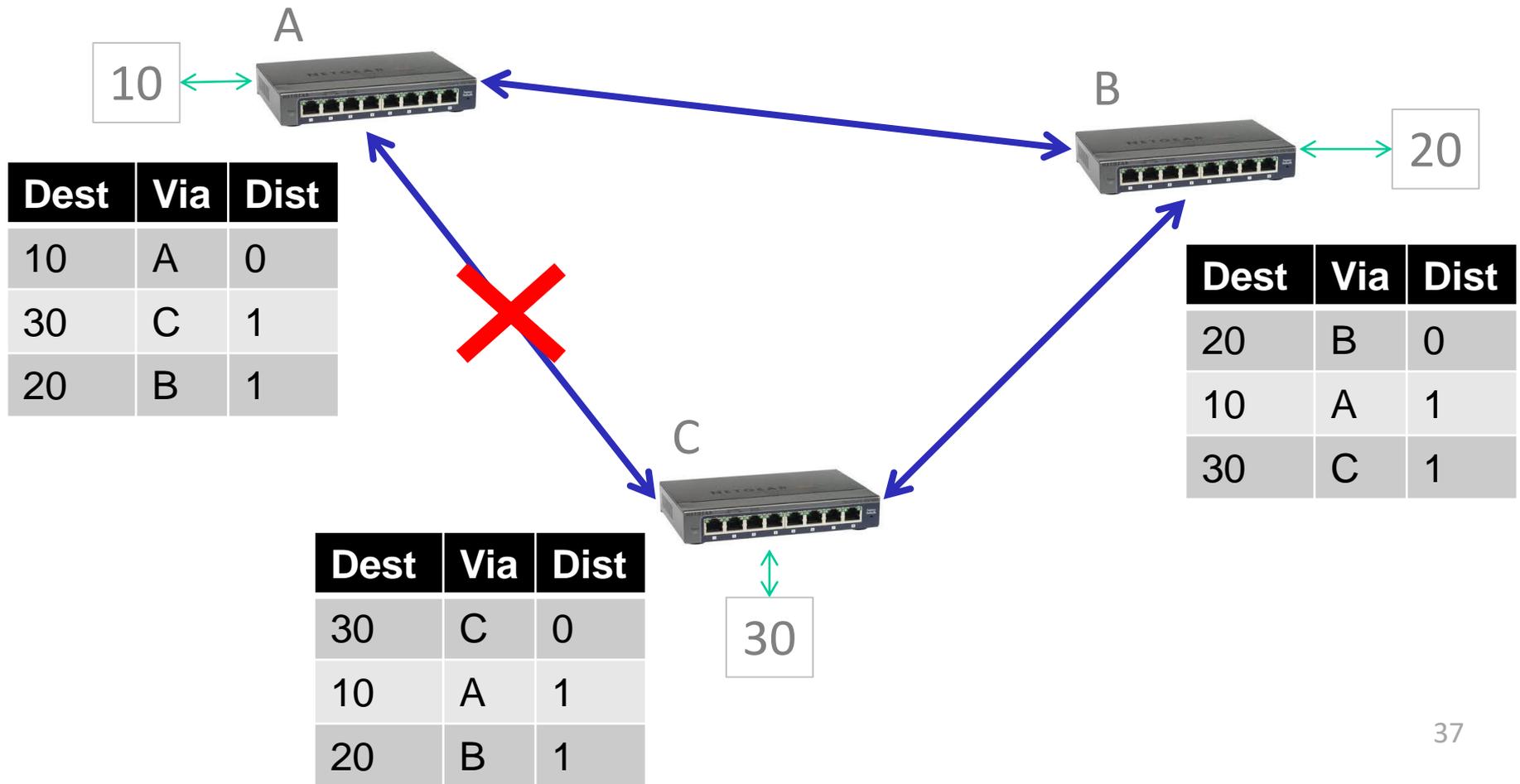
Routage de paquets

Convergence



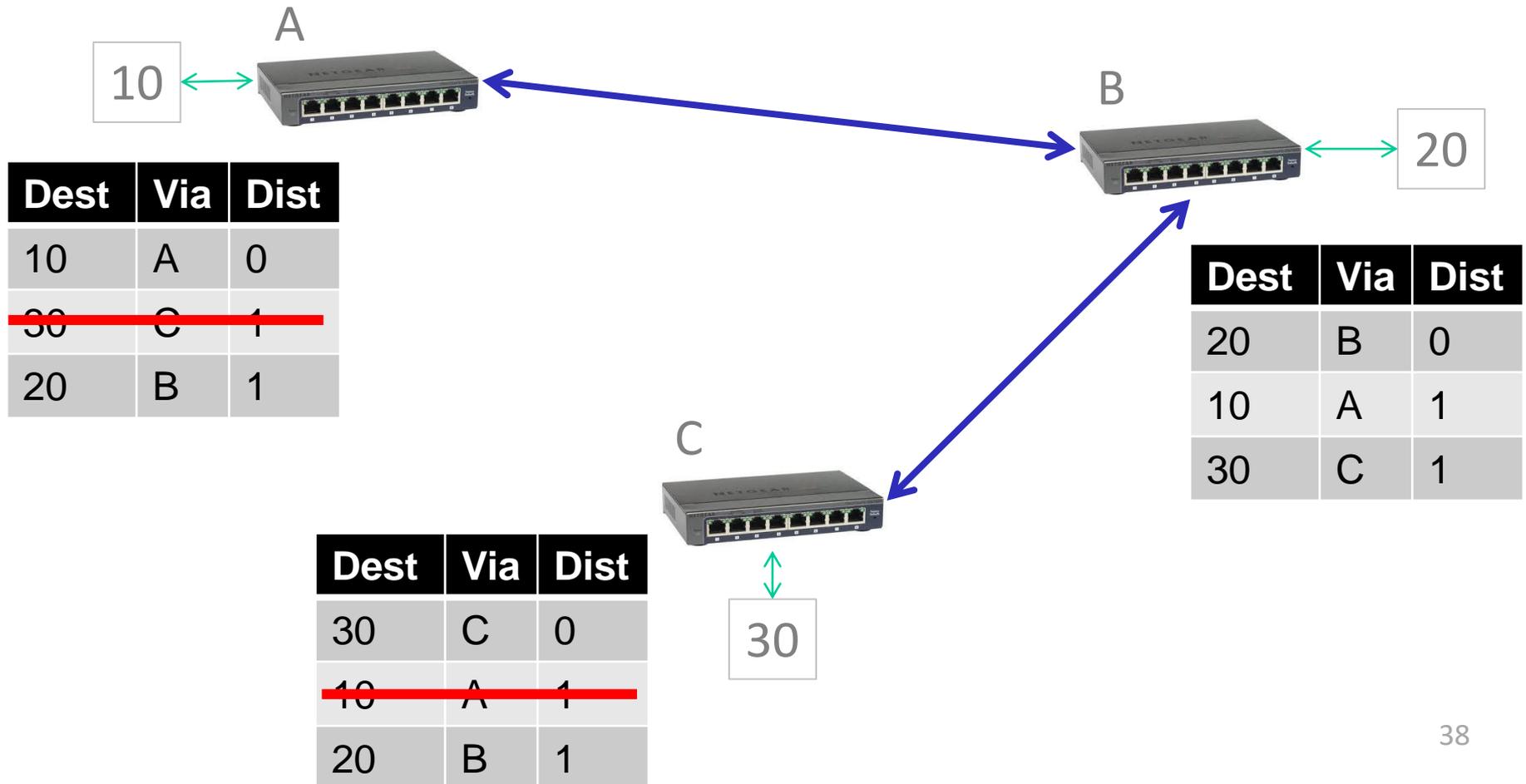
Routage de paquets

Etat stable ?



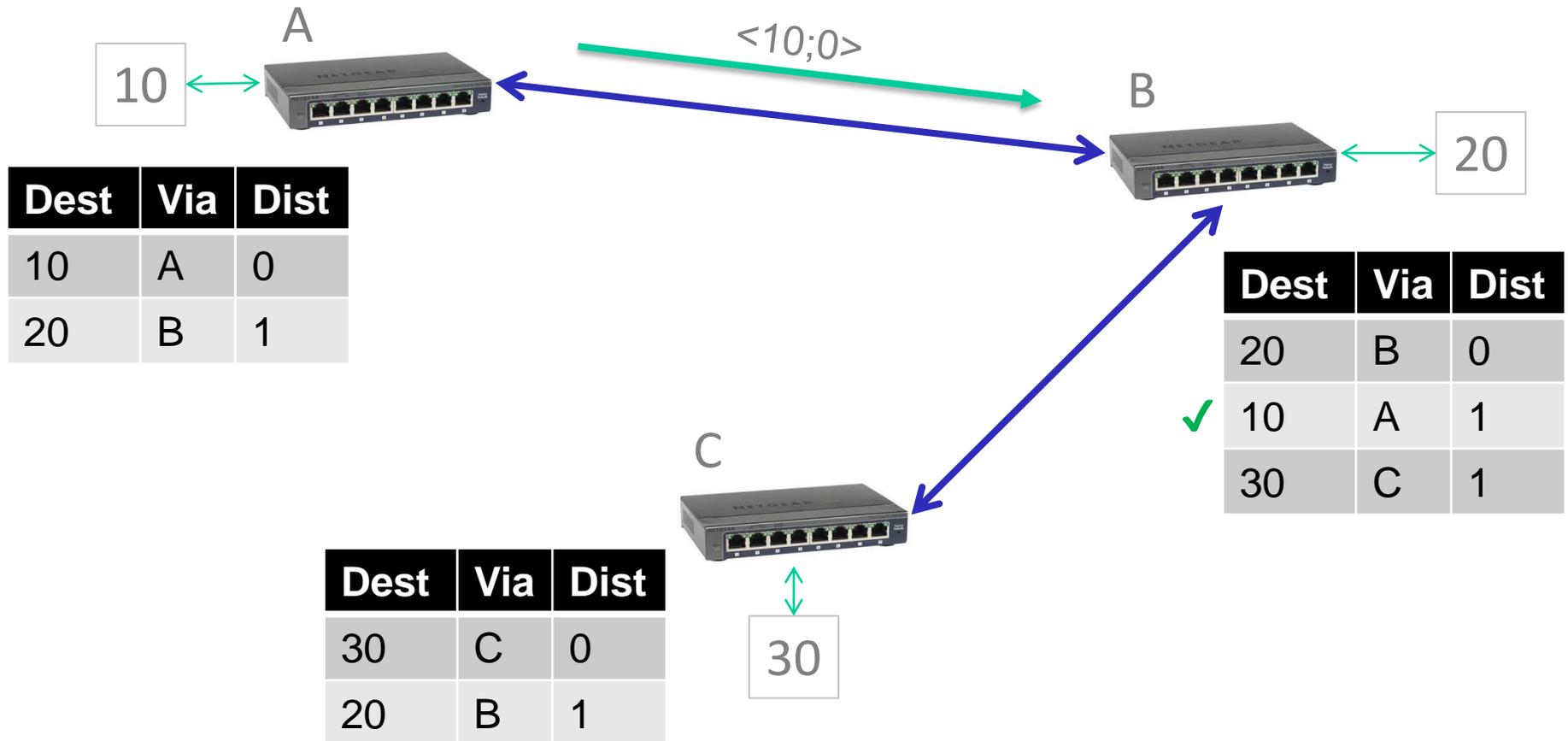
Routage de paquets

Timeout



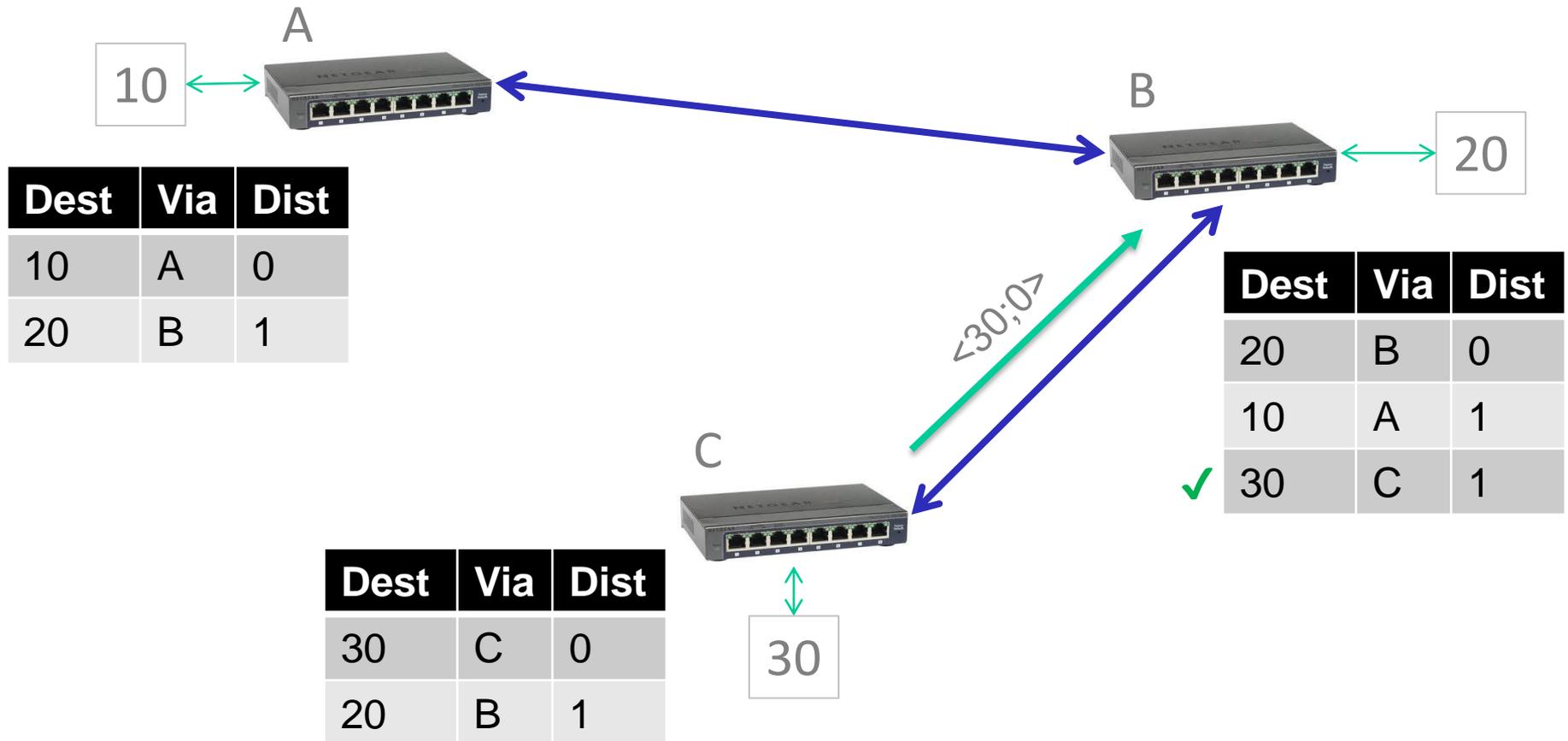
Routage de paquets

Annnonce de A à son voisin



Routage de paquets

Annnonce de C à son voisin



Routage de paquets

Annnonce de B à ses voisins

Reçu de B

<20;0>, <30;1>

10

A



<20;0>, <30;1>

B



20

Dest	Via	Dist
10	A	0
✓ 20	B	1
+ 30	B	2

<20;0>, <10;1>

C



30

Dest	Via	Dist
20	B	0
10	A	1
30	C	1

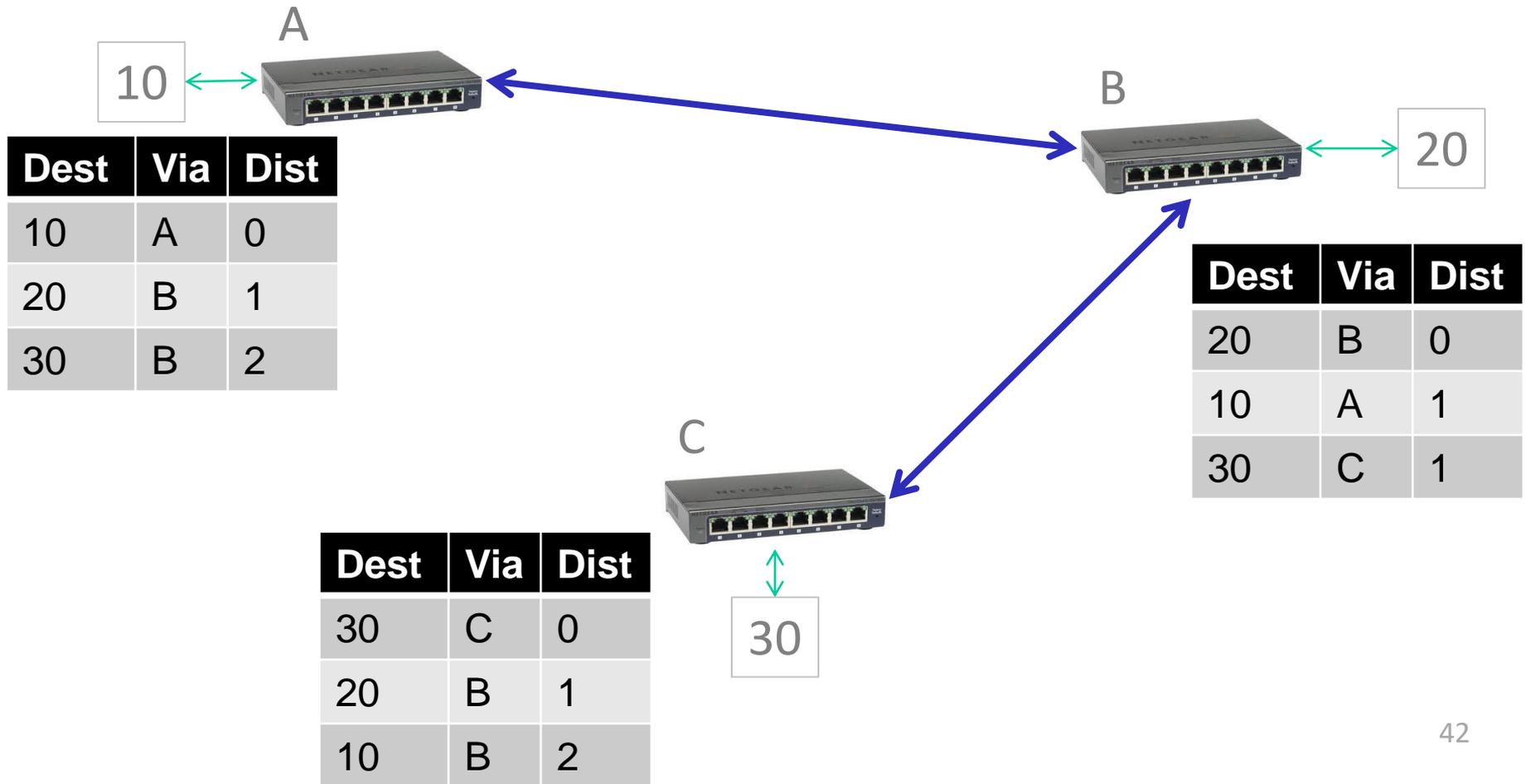
Reçu de B

<20;0>, <10;1>

Dest	Via	Dist
30	C	0
✓ 20	B	1
+ 10	B	2

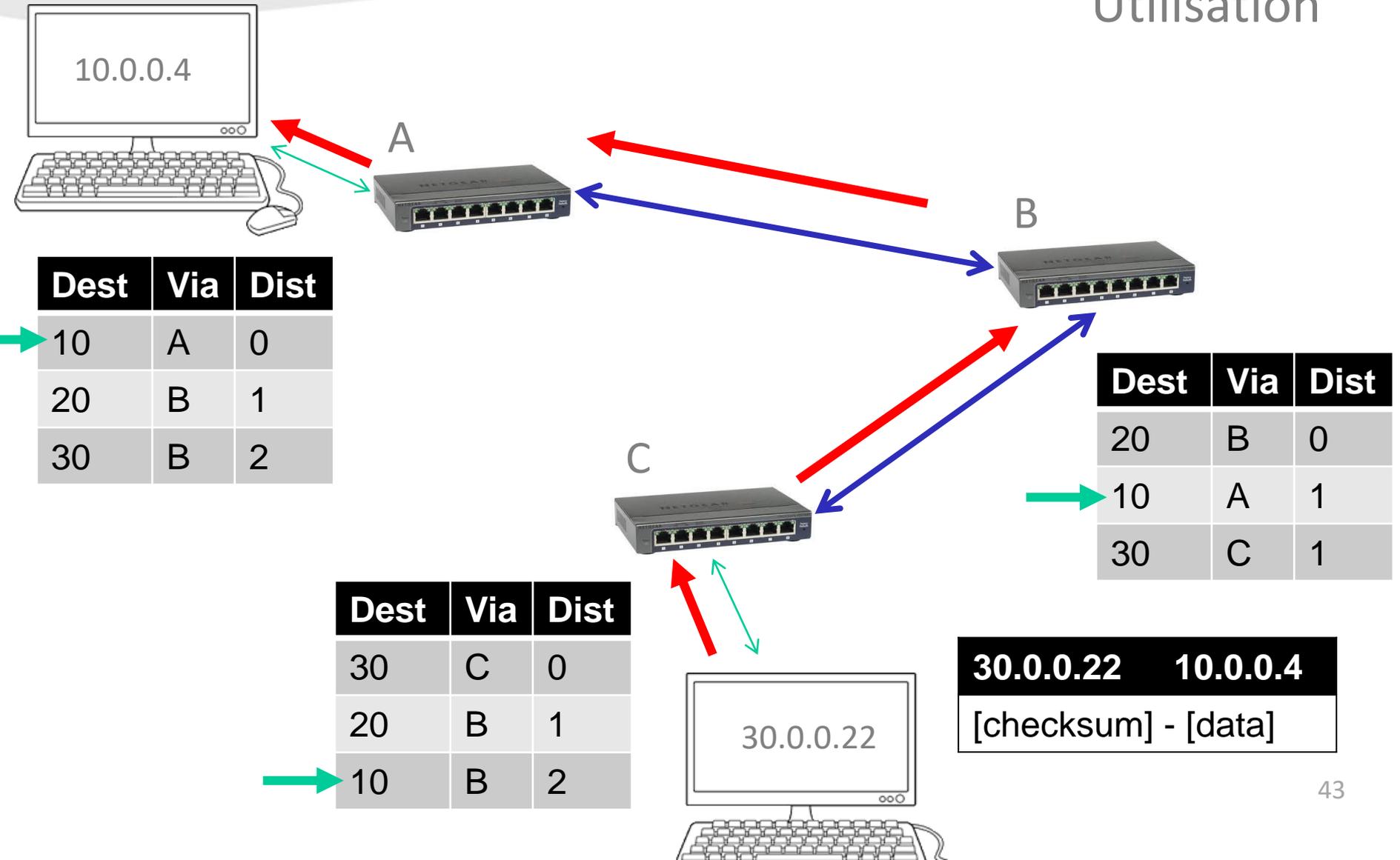
Routage de paquets

Convergence
Etat stable ?



Routage de paquets

Utilisation



Dest	Via	Dist
10	A	0
20	B	1
30	B	2

Dest	Via	Dist
20	B	0
10	A	1
30	C	1

Dest	Via	Dist
30	C	0
20	B	1
10	B	2

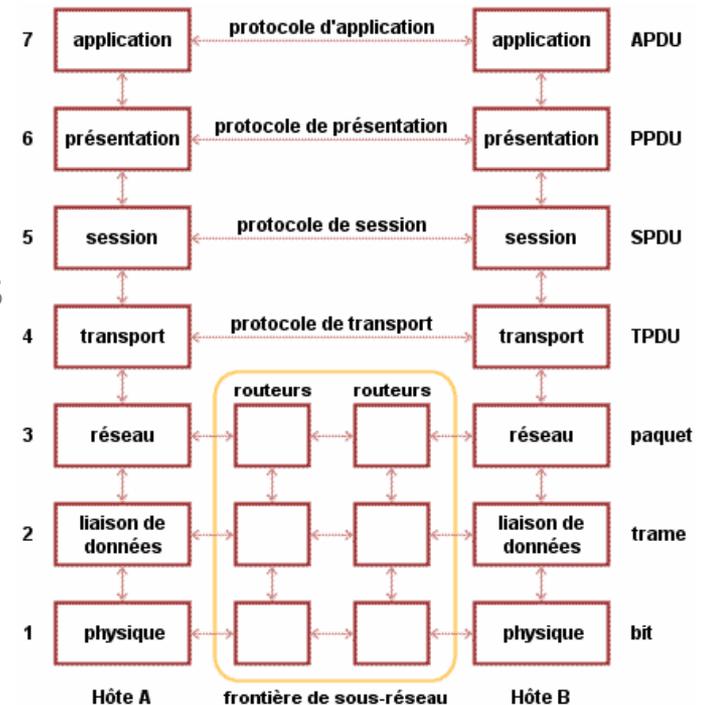
30.0.0.22 10.0.0.4
 [checksum] - [data]

Modèle OSI - Suite

Open Systems Interconnection

4. Transport

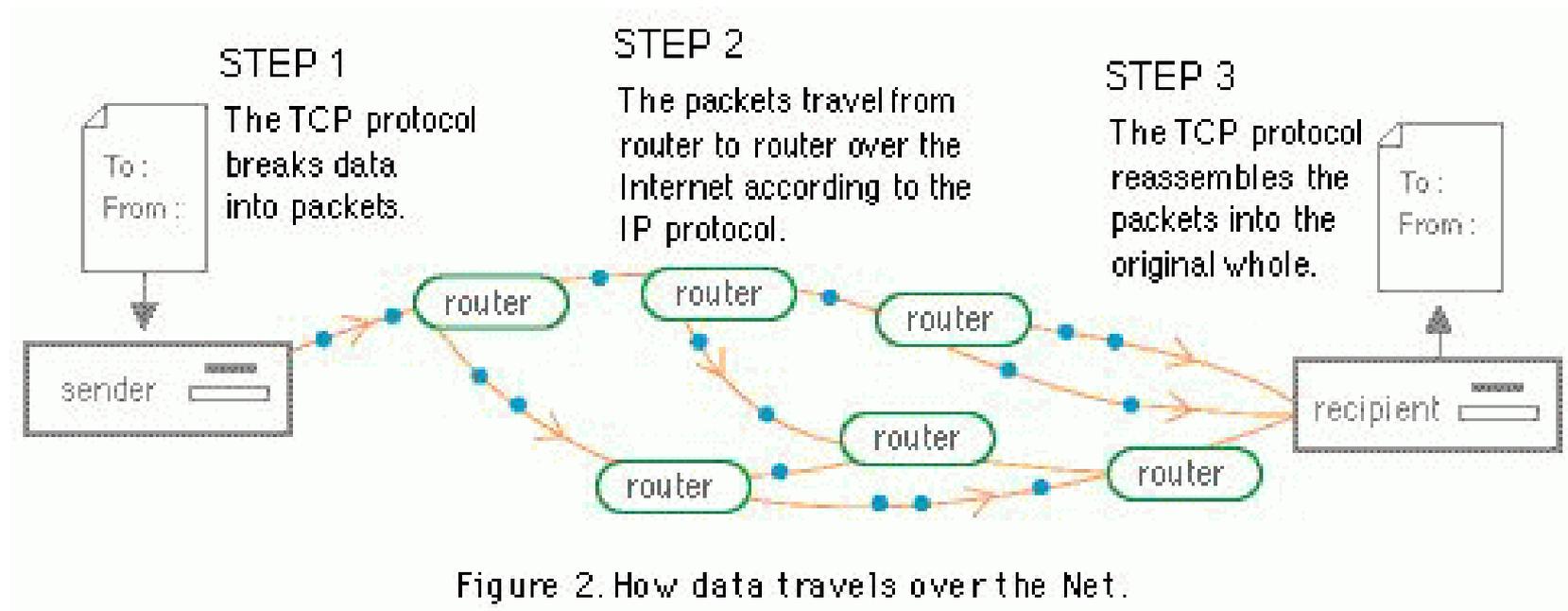
Cette couche est responsable du bon acheminement des messages complets au destinataire. Le rôle principal de la couche transport est de prendre les messages de la couche session, de les découper s'il le faut en unités plus petites et de les passer à la couche réseau, tout en s'assurant que les morceaux arrivent correctement de l'autre côté. Cette couche effectue donc aussi le ré-assemblage du message à la réception des morceaux



Protocole de transport – TCP/IP

- Les paquets IP échangés au niveau 3, sont indépendant les uns des autres
- Un paquet qui contient une erreur (checksum plus valide) sera simplement détruit
- La couche 4 permet d'envoyer des messages (divisés en plusieurs paquets IP) et de garantir la bonne réception
- Un mécanisme permet la retransmission des paquets perdus ou détruit

Protocole de transport – TCP/IP



Modèle OSI - Suite

Open Systems Interconnection

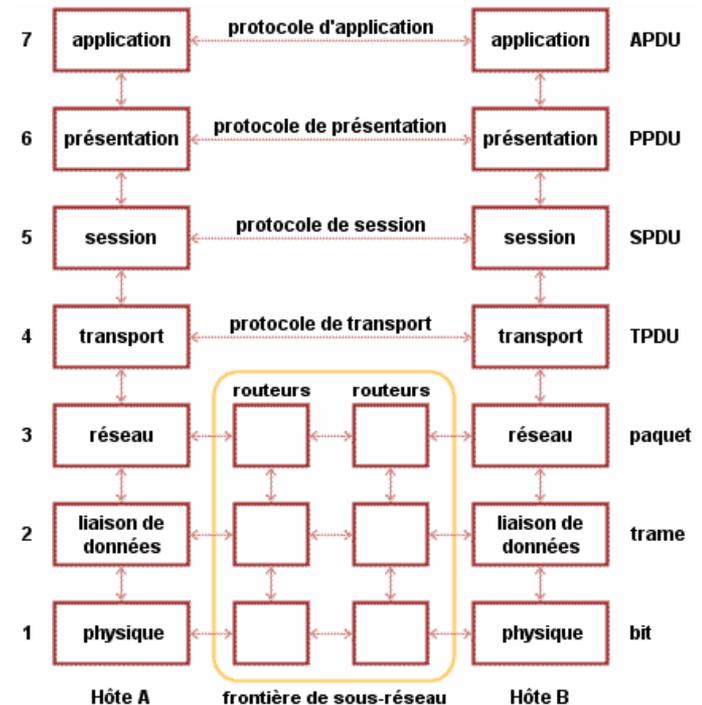
5. Session

6. Présentation

7. Application

Ces trois couches sont souvent intégrées dans une seule couche.

Le protocole HTTP, utilisé pour la consultation de pages web, intègre dans une même requête, un cookie de session et la page HTML (présentation, application)



OSI – Couches 5,6,7 – HTTP/HTML

Requête

```
GET /recherche?indexCatalogue=pagecontent&searchQuery=trm&wordsMode=0 HTTP/1.1
Host: hesav.ch
User-Agent: Mozilla/5.0 [...]
[...]
Referer: http://hesav.ch/[...]
Cookie: ga=GA1.2.47496119.1479602007; _gat=1
Connection: keep-alive
Upgrade-Insecure-Requests: 1
```

L'utilisateur effectue une recherche

Cookie de session
Permet de suivre
l'utilisateur d'une
page à l'autre

Réponse

```
HTTP/1.1 200 OK
Cache-Control: no-cache
Content-Type: text/html; charset=utf-8
[...]
Content-Length: 33169

<!DOCTYPE html>
<html lang="fr">
<head>
  <!--[if lt IE 9]>
<script src="/eigen/assets/js/html5.js" [...]
```

HTML: Présentation + contenu



FILIÈRES DE
FORMATION

FORMATION
CONTINUE ET
POSTGRADE

RELATIONS
INTERNATIONALES

RECHERCHE ET
DÉVELOPPEMENT

EXPERTISE ET
PRESTATIONS DE
SERVICE

Contact

fr en

71 search results for **trm**

Etudiante BSc TRM sur le ResearchBlog de l'University of Salford

<http://hesav.ch/autres/news/secteurs/rerelations-internationales/2016/09/26/etudiante-bsc-trm-sur-le-researchblog-de-l-university-of-salford>

Etudiante BSc **TRM** sur le ResearchBlog de l'University of Salford, Mathilde Widmer, étudiante BSc **TRM**, livre ses impressions sur sa participation à l'école d'été OPTIMAX 2016 - Manchester sur le ResearchBlog de l'University of Salford en août 2016. Consultez le blog ci-dessous et lisez le témoignage de Mathilde, en bas de page: <http://blogs.salford.ac.uk/research/2016/08/12/optimax-summer-school-review/>

Modèle OSI - TCP/IP

Open Systems Interconnection

Modèle OSI

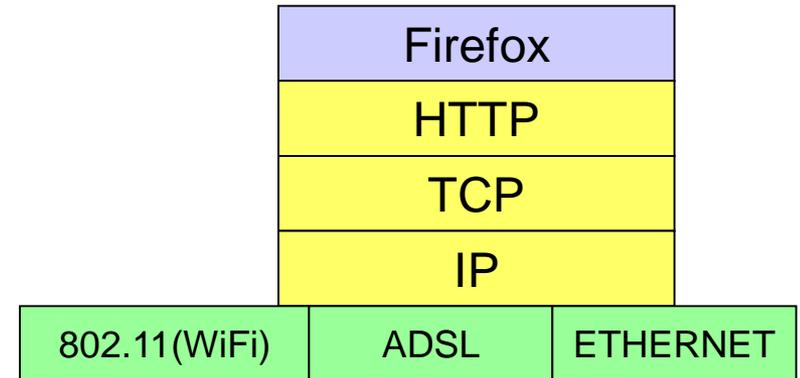
7	Application
6	Présentation
5	Session
4	Transport
3	Réseau
2	Liaison
1	Physique

TCP/IP

<i>Applications</i>
<i>Services Internet</i>
<i>Transport (TCP)</i>
<i>Internet (IP)</i>
<i>Accès au Réseau</i>

Protocole en couches

- Des couches homologues sont en principe interchangeables
- Une même application peut ainsi utiliser des milieux de transmission très différents



IP - Adressage

- Les adresses IPv4 sont constituées de 4 octets (32 bits), et sont donc de la forme x.y.z.w (ex. 192.168.1.24)
- Ce groupage permet aux routeurs du réseau IP d'aiguiller les « paquets » IP (messages IP élémentaires) pour les convoyer de la source vers la destination.

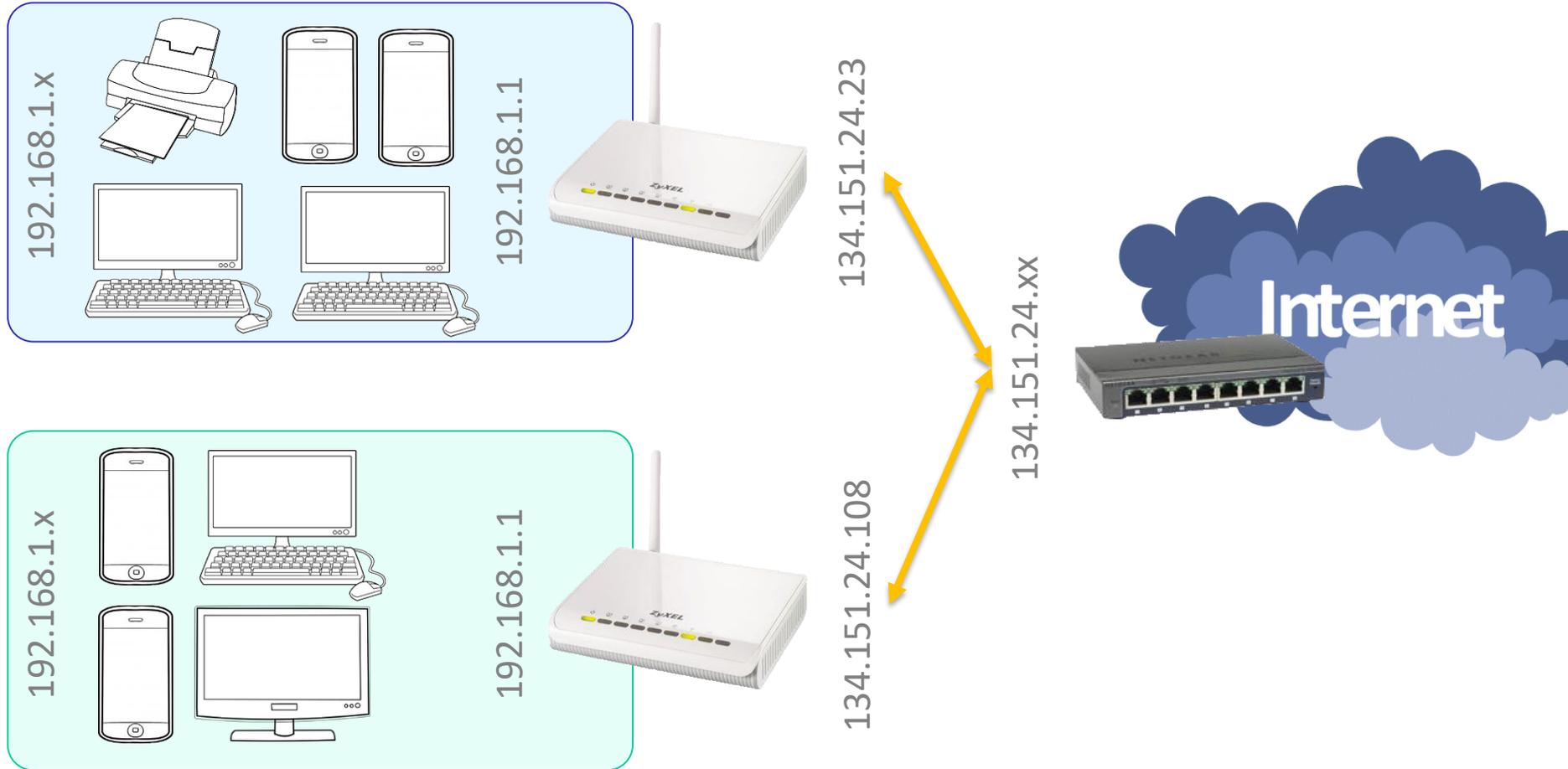
IP - Adressage

- En théorie, la capacité d'adressage (2^{32}) semble suffisante
- En réalité, elle est insuffisante, en raison de la structure rigide de l'adressage
- Les adresses IP doivent être uniques
- IPv6, prochaine évolution de IP étend le champ d'adressage à 128 bits

IP - Question

- Lorsque l'on achète un routeur ADSL dans le commerce, il est en principe configuré pour une adresse du type 192.168.1.xxx
- Rien ne s'oppose à ce que j'aie la même adresse que mon voisin qui a acheté le même routeur
- Où est l'erreur ?

IP - Question



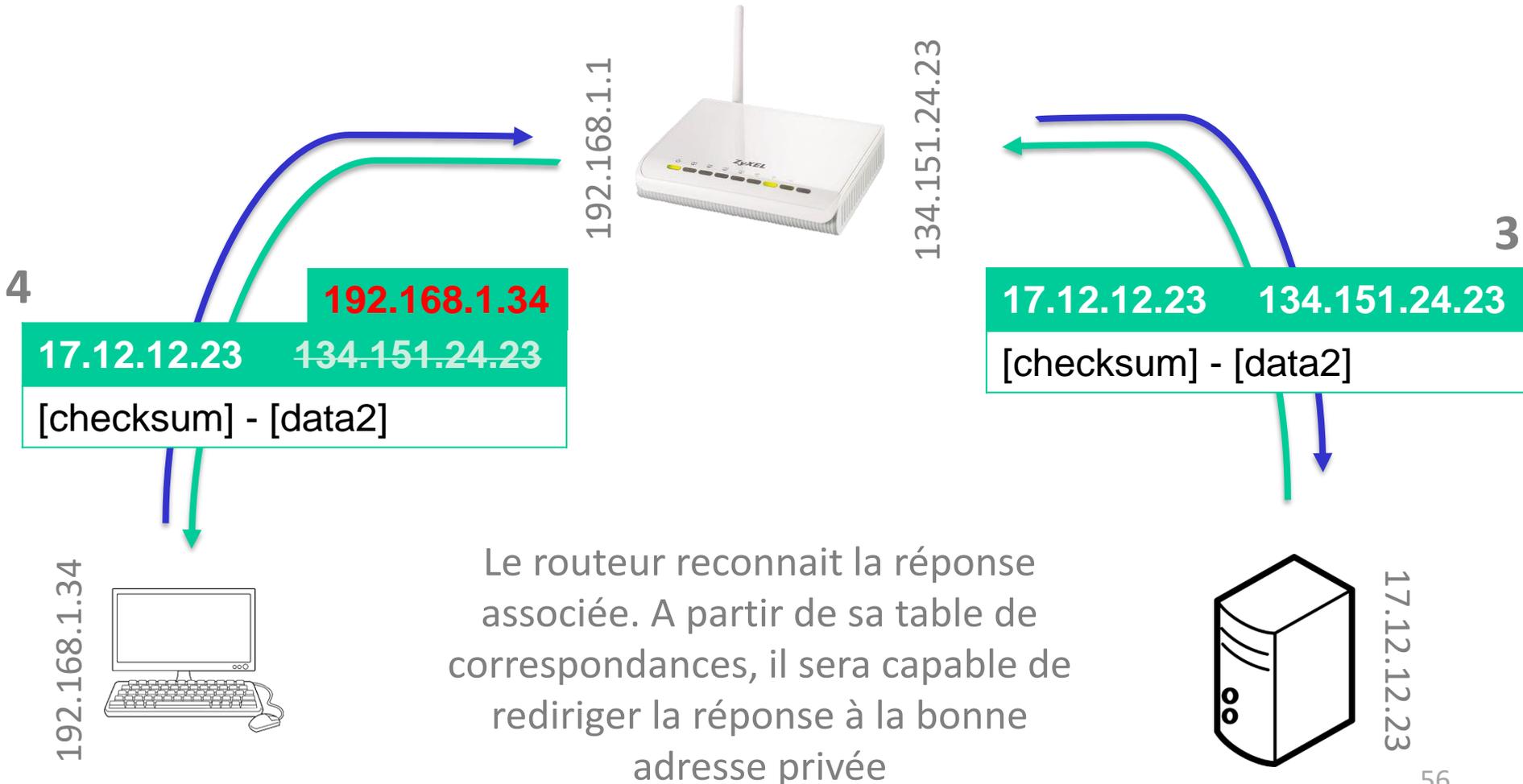
IP - NAT

Network address translation



IP - NAT

Network address translation



Adresses Internet

- L'adresse IP est rarement utilisée « telle quelle »
- On utilise plutôt une adresse dite DNS (Domain Name Service) par exemple *www.hesav.ch*
- Une base de données fait la correspondance entre adresse DNS et adresse IP

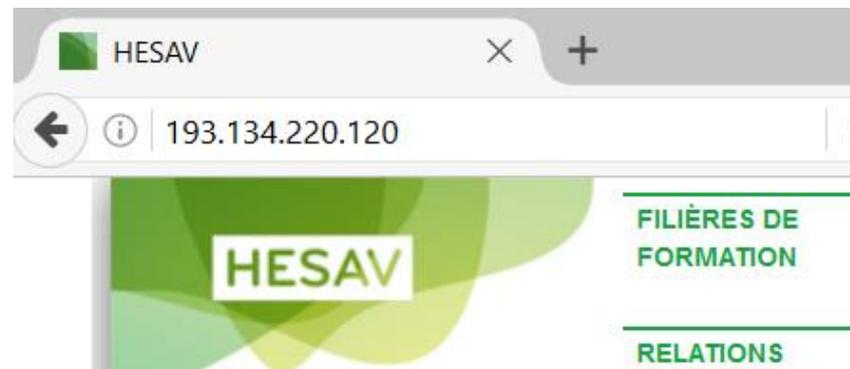
Adresses Internet

```
Windows PowerShell
Copyright (C) 2016 Microsoft Corporation. Tous droits réservés.

PS C:\Users\Fabien> nslookup hesav.ch
Serveur :      Unknown
Address:  192.168.123.1

Réponse ne faisant pas autorité :
Nom :      hesav.ch
Address:  193.134.220.120
```

Correspondance assurée par le DNS



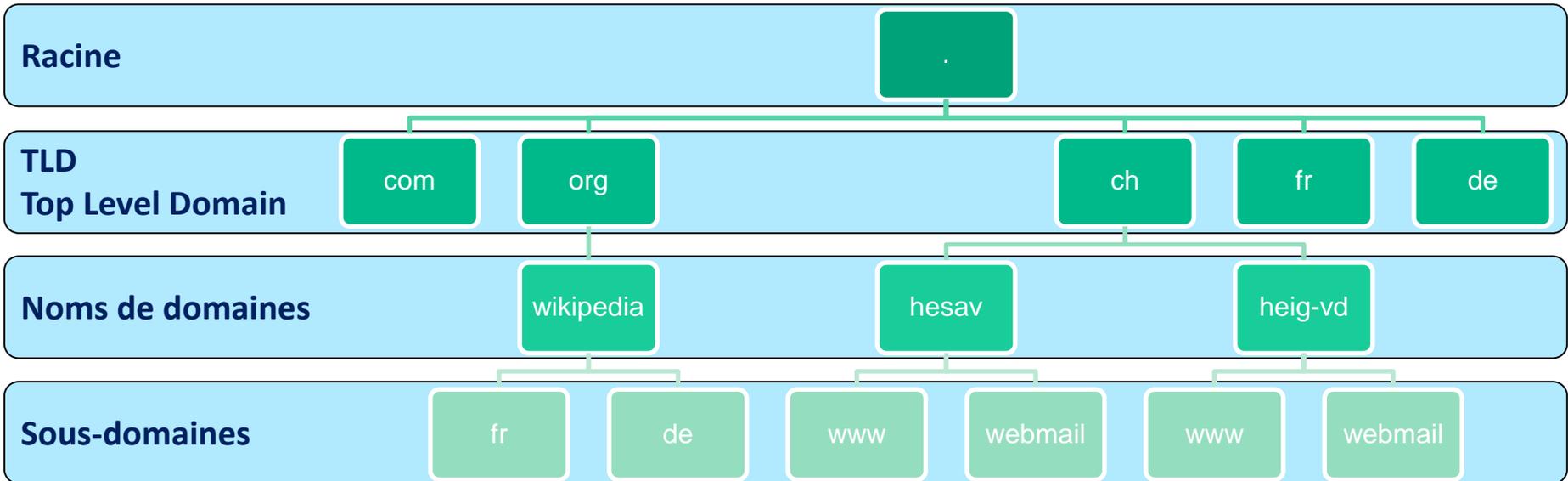
Adresses DNS

- Une adresse DNS est plus facile à retenir qu'une adresse IP
- Une adresse DNS devrait rester valable si on passe à IPv6
Par ex: 2001:0db8:0000:85a3:0000:0000:ac1f:8001
- L'adresse DNS se lit de droite à gauche
- Les champs à droite de l'adresse DNS permettent de désigner des ressources particulières sur le site adressé

Par ex: <http://hesav.ch/formation>

<http://hesav.ch/recherche-et-developpement>

Adresses DNS



Adresses DNS

- Chaque niveau doit pouvoir indiquer où trouver l'information pour les niveaux inférieurs
- Il existe 13 serveurs DNS « racines », leurs adresses IP sont connues « de tous »
- La résolution d'un nom de domaine s'effectue de droite à gauche, du plus général au plus précis:
webmail.hesav.ch



Adresses DNS - Exemple

- On contacte un des 13 serveurs racines:

```
PS C:\Users\Fabien> nslookup webmail.hesav.ch 193.0.14.129
```

```
Nom :      webmail.hesav.ch
Served by:
- a.nic.ch
      130.59.31.41
      2001:620:0:ff::56
      ch
```

Il nous répond que c'est nic.ch qui s'occupe du domaine **ch** et que l'on peut contacter 130.59.31.41

- On contacte donc 130.59.31.41:

```
PS C:\Users\Fabien> nslookup webmail.hesav.ch 130.59.31.41
```

```
Nom :      webmail.hesav.ch
Served by:
- ns02.heig-ud.ch
      193.134.216.115
      hesav.ch
```

Il nous répond que c'est heig-ud.ch qui s'occupe du domaine **hesav.ch** et que l'on peut contacter 193.134.216.115

- On contacte
193.134.216.115...

```
PS C:\Users\Fabien> nslookup webmail.hesav.ch 193.134.216.115
Serveur :  ns02.heig-ud.ch
Address:  193.134.216.115
Nom :      webmail.hesav.ch
Address:  193.134.218.198
```

Stockage de données

- Disques durs
 - Magnétiques - HDD (max 8 To, env. CHF 400)
 - Tête de lecture + disques magnétiques rotatifs pouvant être endommagés en cas de choc + bruit
 - Temps d'accès aux données pénalisé si multiples petits fichiers
 - Flash - SSD (max 4 To, env. CHF 1'500)
 - Solid State Drive
 - Meilleures performances en lecture et écriture
 - Pas de latence (pas de disques à mettre en rotation, ni de tête de lecture à déplacer)



Stockage de données

- Clés USB – Cartes mémoires
 - Mémoire flash (max 1 To, env. CHF 1'000)
 - Idéal pour le transport de documents
 - Certains modèles permettent le chiffrement des données stockées
- CD / DVD / Blu-Ray (max 50 Go, env. CHF 5)
 - Souvent à usage unique
 - Les portables récents ne possèdent plus de lecteurs...



Stockage de données – durée de vie

- CD / DVD / Blu-Ray
 - Généralement 4-5 ans
 - 10 ans pour les modèles à base d'or
- Clés USB – Cartes mémoires
 - Environ 10 ans
 - La technologie flash permet au maximum 10'000 – 100'000 cycles de lecture/écriture

Stockage de données – durée de vie

- Disques durs
 - Chiffres Google – sur 100'000 disques
 - 1,7% ont du être remplacés la première année
 - 8,0% la seconde année
 - 8,6% la troisième année
 - 6,0% la quatrième année
 - 7,0% la cinquième année
- Sous forte utilisation, 1/3 des disques sont défectueux les 5 premières années

Stockage de données – durée de vie

- Aucun support de données ne peut garantir un archivage à long terme
 - Il faudra envisager des copies sur de nouveaux supports régulièrement
- A plus court terme, il faut au minimum stocker les documents importants sur 2 supports
 - La probabilité est faible que les 2 soient défectueux en même temps

Stockage du futur ?

Microsoft voit le futur du stockage des données... dans notre ADN

MICROSOFT | STOCKAGE DE DONNÉES | ADN | BIOTECHS | NUMÉRIQUE & INFORMATIQUE
PAR JULIETTE RAYNAL PUBLIÉ LE 28/04/2016 À 16H33

- Et si l'ADN était le stockage du futur ?
- Microsoft est en train de tester le stockage de données sur des brins d'ADN synthétiques
- «un gramme d'ADN permettrait théoriquement de stocker mille milliards de Go de données»



Microsoft vient d'acheter 10 millions de brins d'ADN synthétiques à l'entreprise Twist Bioscience pour tester les capacités de stockage de cette macromolécule. A long terme, l'ADN pourrait être un moyen de réduire le coût du stockage des données. Une perspective qui intéresse particulièrement la firme de Redmond qui opère l'une des principales plates-formes de Cloud public.



$10^{12} \text{ Go} \gg 10^3 \text{ Go}$



- Des chercheurs de l'EPFZ sont parvenus à démontrer qu'une molécule d'ADN permettait de stocker des données numériques pendant plus de 2'000 ans !

Stockage de données – réplication

- Nous avons vu qu'un disque dur peut avoir une durée de vie relativement courte, la solution est de répliquer les données sur plusieurs disques



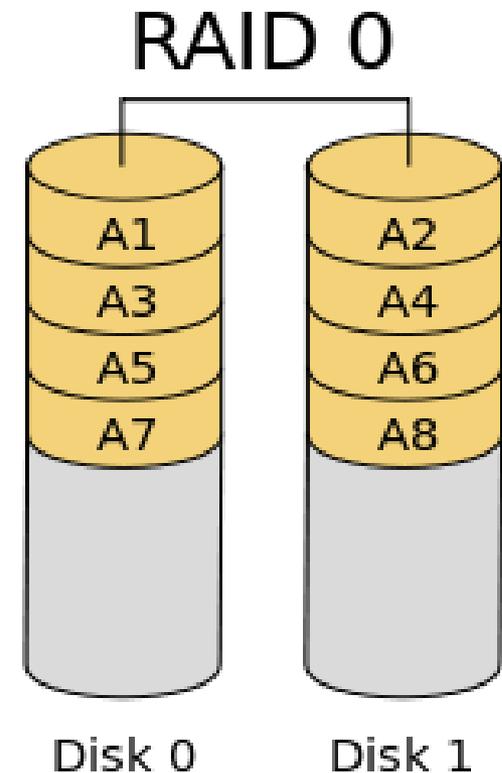
NAS personnel
2 disques



Stockage professionnel
25 disques

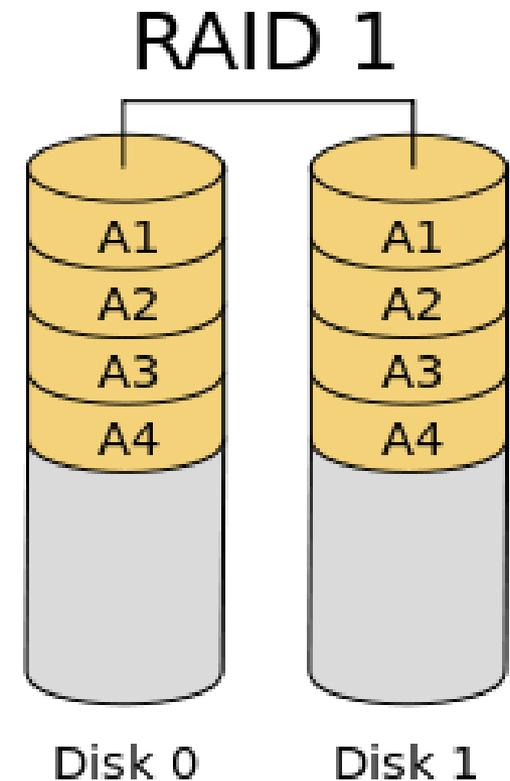
Réplication – RAID

- RAID: *Redundant Array of Independent Disks*
- **RAID 0** – Aucune redondance
- Ex: $2 \times 1 \text{ To} = 2 \text{ To}$
- Efficacité: 100 %



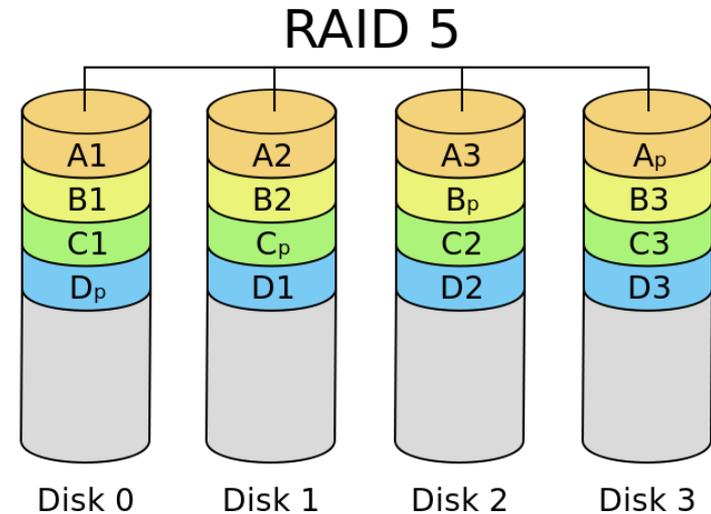
Réplication – RAID

- **RAID 1** – 2 copies identiques
- Ex: 2x 1 To = 1 To
- Efficacité: 50 %
- Si un disque est défectueux, on peut le remplacer et le contenu de l'autre est simplement recopié



Réplication – RAID

- **RAID 5** – Les données sont réparties sur N-1 disques, sur le dernier disque un bloc de parité est calculé
- Ex: $N \times 1 \text{ To} = (N-1) \times 1 \text{ To}$
- Efficacité: 75% (4 disques)
- Un disque défectueux peut être remplacé, les blocs de données et de parités manquants pourront être reconstruits à partir des autres disques

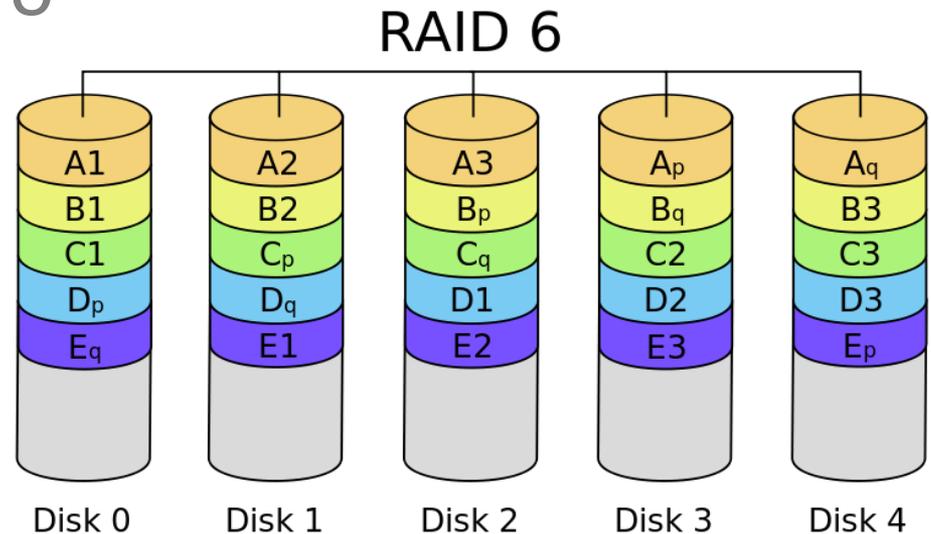


Réplication – RAID

- Les niveaux de RAID (1 et 5) vu jusqu'ici autorisent la perte d'un seul disque à la fois
- Le fait de changer un disque nécessite une lourde tâche de lecture sur les disques restants, ce qui peut achever un autre disque...
- Avec 2 disques défectueux, il y aura perte de données

Réplication – RAID

- **RAID 6** – évolution du RAID 5 qui accroît la sécurité en utilisant n informations redondantes au lieu d'une seule
- Ex: $N \times 1 \text{ To} = N-n \text{ To}$
- Efficacité: 60%
($N=5, n=2$)
- Jusqu'à n disques défectueux sont possibles



Stockage réseau

- Le stockage réseau se présente sous plusieurs variantes :
 - Disques partagés, disques virtuels, réseaux de stockage
 - Archivage
- Un disque partagé réclame un débit considérable pour être utilisable
- Un média d'archivage est moins gourmand en débit, mais souvent plus en taille

Stockage réseau

- Le stockage réseau dépend des performances du réseau...
- Sa fiabilité est celle du réseau
- La sécurité des données est conditionnée par la sécurité du réseau
- La connexion par un réseau d'accès peut dégrader les performances de manière significative, voire inadmissible

Stockage réseau

- Mettre de très gros fichiers sur un disque réseau peut être avantageux, mais parfois aussi désastreux
- Les protections d'accès à vos fichiers peuvent être aussi bonnes qu'on le souhaite. En pratique, elles sont rarement suffisantes
- Le temps de transfert peut s'avérer prohibitif

Stockage réseau

- L'imagerie médicale, de plus en plus sophistiquée, fait exploser les besoins en stockage
- Le besoin de stockage amplifie les besoins de bande passante
- Besoin de pose de fibres optiques, disques rapides, performances en entrées-sorties exceptionnelles

Exemples



max. 1 Mbit/s
Download

max. 0.2 Mbit/s
Vitesse
Upload

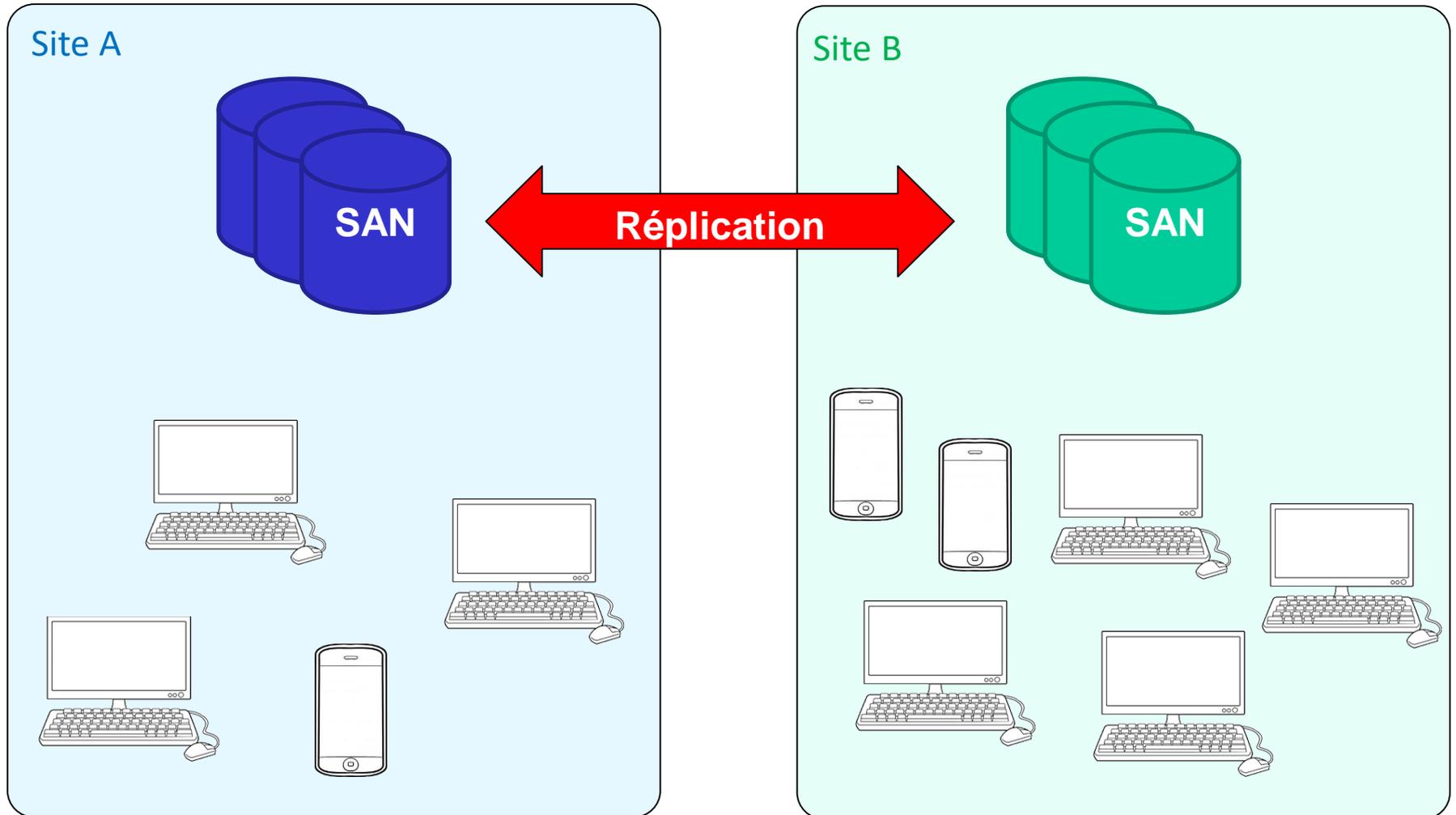


	Mo	Mb/s	Mb/s	Mb/s	Mb/s	Mb/s
Document	Taille	1000	100	10	2	0.1
Présentation Powerpoint	1.5	0.012 s	0.12 s	1.2 s	6 s	120 s
300 scan CT	157	1.256 s	12.56 s	125.6 s	628 s	12'560 s
500 JPEG fine 24 Mpxl	1'000	8 s	80 s	800 s	4000 s	80'000 s
1 film (LOTR part 2)	4'000	32 s	320 s	3'200 s	16'000 s	320'000 s
500 TIFF 8 Mpxl	96'000	768 s	7'680 s	76'800 s	384'000 s	7'680'000 s
500 TIFF 24 Mpxl	288'000	2'304 s	23'040 s	230'400 s	1'152'000 s	23'040'000 s

SAN

- La notion de Storage Area Network remplace la notion de disque partagé par la notion de réseau.
- Pour améliorer les performances, un même fichier est réparti sur plusieurs disques pouvant se trouver éloignés géographiquement
- Une redondance est nécessaire pour pallier aux défaillances d'un disque donné

SAN



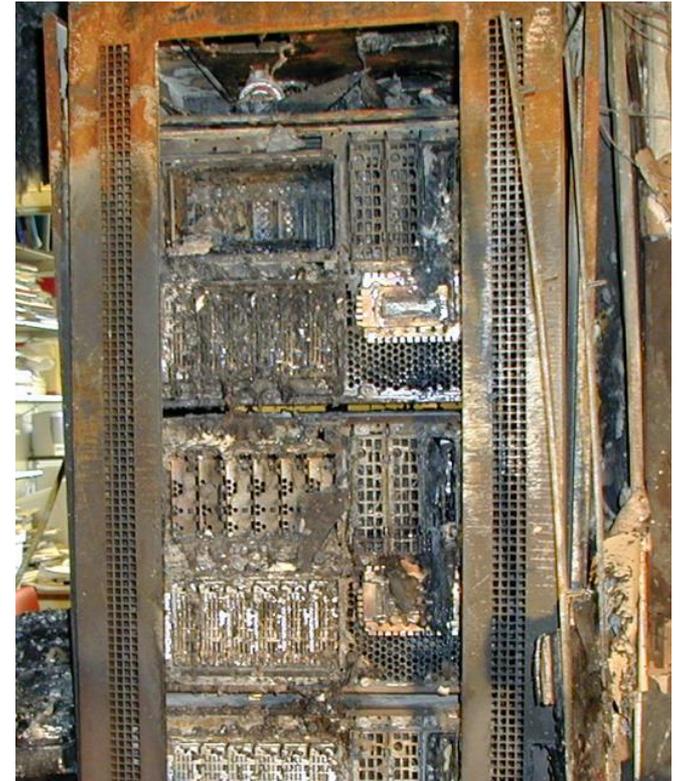
Archivage réseau

On distingue stockage et archivage :

- Le **stockage** concerne les documents actuels, sur lesquels on peut être amené à travailler régulièrement
- L'**archivage** concerne les documents ne présentant plus un intérêt immédiat
- l'archivage demande une opération spécialisée pour extraire l'information

Stockage de données – sauvegarde

- La répartition des données sur plusieurs disques permet de sécuriser les données contre un support défectueux
- Si l'intégralité du support peut être détruit le problème reste entier...



src: <http://www.haute-disponibilite.net>

Sauvegarde des données

- Le contenu d'un système de stockage doit être répliqué régulièrement sur un site distant
 - Les 2 sites doivent être éloignés au maximum (inondation, ouragan, incendie, guerre,...)
- La réplication à distance devra passer par un réseau (Internet), on va donc dépendre de la vitesse de transmission et devoir sécuriser (chiffrer) l'échange

Ransomware

- Ou «Rançongiciel» en français, est une famille de logiciels malveillants («virus informatique») ayant pour but principal de prendre en otage les données personnelles
- Généralement installé par l'ouverture d'un e-mail frauduleux, il va chiffrer toutes les données (*pdf, jpg, docx, xlsx,...*) auxquelles il a accès
- Il indiquera ensuite à l'utilisateur où verser l'argent pour recevoir la clé de déchiffrement

Ransomware

- Les dossiers partagés (NAS, Dropbox) seront aussi touchés.
- On peut espérer que les sauvegardes ne soient pas aussi accessibles comme dossier partagé...
- 17 février 2016, zdnet.fr:

Quand un ransomware paralyse un hôpital américain

Sécurité : *Un hôpital de Los Angeles a vu son système informatique touché par un ransomware, qui a progressivement bloqué la majorité de ses fonctions administratives pendant une semaine. Les attaquants réclameraient la coquette somme de 3,6 millions de dollars pour déverrouiller l'accès aux données chiffrées.*

Ransomware

- Le paiement de la rançon ne permet pas forcément de récupérer l'accès à ses données...
Difficile d'obtenir un service après-vente de la part de criminels

- 23 mai 2016, zdnet.fr:

Et ce sont les hôpitaux qui sont tout particulièrement à la pointe de l'actualité en ce domaine. Après celui de Los Angeles (Hollywood Presbyterian Medical Center), qui en février dernier a déboursé 17 000 dollars, c'est le Kansas Heart Hospital (Wichita, Kansas) qui vient de subir une attaque identique le 18 mai dernier.

Et pour rajouter du piquant à l'affaire, les kidnappeurs ont refusé de débloquer le système d'information pris en otage une fois la rançon versée sur leur compte [rapporte Networkworld](#). Les administrateurs s'attendaient à mettre à nouveau la main sur les données et les fichiers de l'hôpital. Ils ont reçu en lieu et place ...une nouvelle demande de paiement. Et cette fois l'hôpital a refusé de payer. Les sommes exigées par les pirates n'ont pas été divulguées.



Compression

- Nous avons vu tout à l'heure que la théorie de l'information introduit la notion de **redondance** de l'information
- Celle-ci peut être utilisée pour compresser les données, cela permet notamment:
 - Transmission plus rapide
 - Stockage plus efficace

Compression

- Il y a 2 catégories d'algorithmes de compression, avec et sans perte:
 - Sans perte, les données peuvent être reconstruites à l'identique
 - Avec perte, l'information reconstruite est plus ou moins similaire à l'original

Compression – Sans perte

- Le *RLE (run-length encoding, codage par plages)* utilisé par exemple pour les fax
- Un fax peut être vu comme une image composée de pixels noirs et blancs, on peut représenté une ligne par:

BBBBBBBBNNNNBBBBBBBBBBBBNNNNBBB...

- A la place d'indiquer chaque pixel indépendamment, nous allons les regrouper par suite de pixels de même couleurs

9B4N12B3N3B...

- On tire profit de la nature d'un fax, beaucoup de blanc, pixels noirs regroupés pour compresser les données



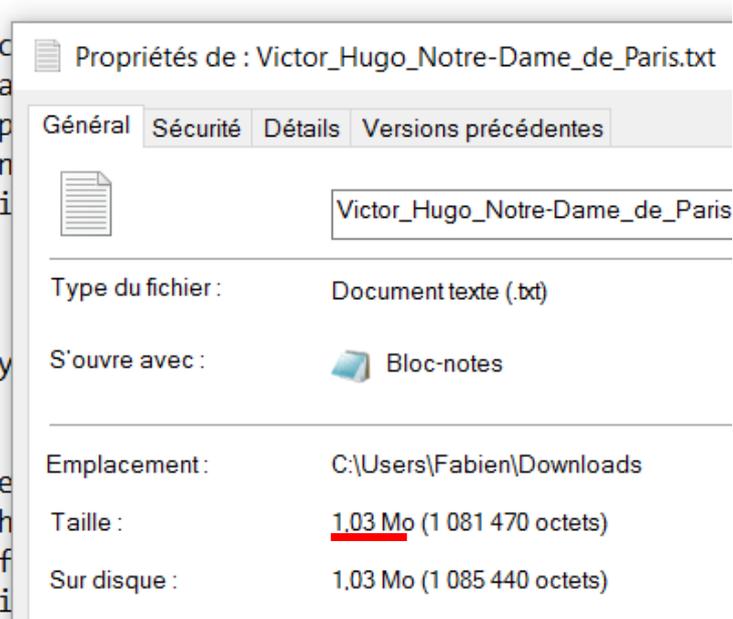
Compression – Sans perte

- Nous avons vu tout à l'heure que pour représenter du texte sur informatique, nous utilisons 1 byte par caractère
- Pour la version anglaise du livre «Notre dame de Paris», nous avons environ 184'000 mots et 1'000'000 de caractères, soit 10^6 octets, soit 1 Mo

ere was to be a bonfire on the Place de Braque, and a mystery at the Palace to the sound of the trumpet, the processions, by the provost's men, clad in robes of violet camelot, with large white

citizens, male and female, having converged from every direction, at early spots designated.

his choice; one, the bonfire; another mystery play. It must be stated, in history of Paris, that the greater part of the wards the bonfire, which was quite in



Compression – Sans perte

l : 167581
e : 106868
t : 71787
a : 64121
o : 60271
h : 56646
i : 54335
n : 51745
s : 50906
r : 48820
[...]
7 : 21
0 : 11
9 : 10
z : 8

- En regardant les statistiques caractère par caractère, nous nous rendons compte que le caractère *espace* représente 16.8% du contenu, le *e* 10.6% tandis que le *Z* représente seulement 0.0008%
- Le *codage de Huffman* se base sur cette constatation, on va représenter avec le minimum de bits possibles les caractères fréquents
- Par exemple en binaire:
 - *<espace>* 110 3 bits
 - *e* 010 3 bits
 - *Z* 101000100111111001 18 bits

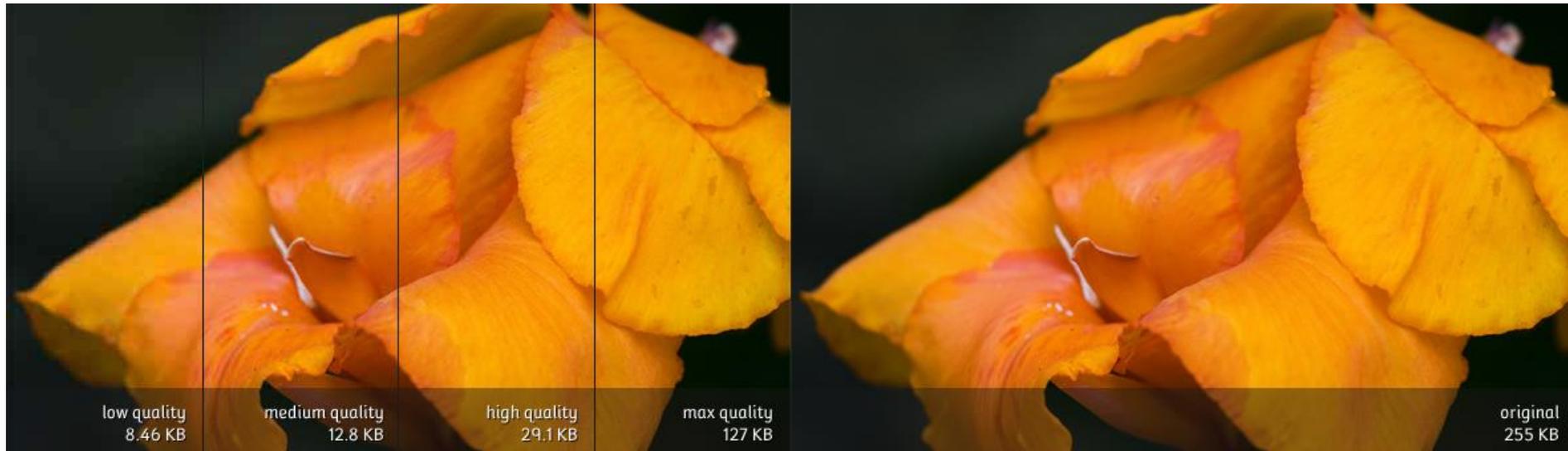
En moyenne il faut 4,49 bits par caractère, soit un gain d'environ 45 %

Compression – Avec perte

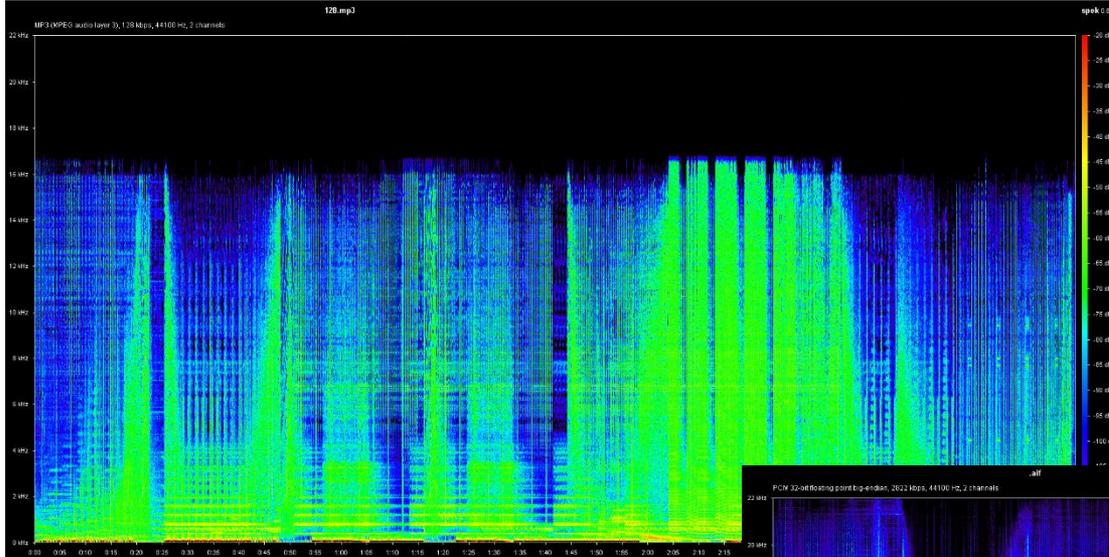
- La compression avec perte va tirer parti des limitations sensorielles de l'être humain pour garder uniquement la partie des données perceptibles
 - Son/Musique: L'oreille humaine peut percevoir les sons entre 16 Hz et 16'000 Hz environs
 - Image/Vidéo: L'œil humain est peu sensible aux hautes fréquences, rapides variations de couleur sur courte distance
- Le but est que la version compressée des données soit perçue à l'identique par une personne, bien qu'elle soit différente de l'original
- La compression avec perte n'est pas réversible

Compression – Avec perte

- JPEG



Compression – Avec perte



MP3 - 128 kbps

CD - 2822 kbps

