

Les réseaux Informatiques

Evolution des réseaux Informatiques.....	10
Cas d'école dans les années 80	11
Types de réseaux.....	12
En fonction de la méthode d'accès :	13
Maintenant.....	14
Ethernet / IEEE 802.3.....	15
Carte d'identité du protocole	16
Etat de la normalisation	17
Architecture	18
Equipements.....	19
10 base 2 (Thin Ethernet)	20
Hub : Emulation d'un bus	21
matériel	22
Sur un PC	23
Algorithme du CSMA/CD	24
Algorithme du CSMA/CD	25
Algorithme du CSMA/CD	26
Les collisions.....	27
Topologie d'un réseau Ethernet	28
Résolution des collisions	29
Quand s'applique le CSMA/CD ?	30
Les commutateurs (switchs)	31
Exemple de trafic	32

Les commutateurs (switchs)	33
Format des trames.....	34
Influence de la vitesse sur le format des trames	35
Auto-négociation.....	36
Questions	37
Les ponts transparents.....	39
ponts filtrants	40
Cas complexe	41
Cas complexe (suite)	42
Algorithme du Spanning Tree.....	43
Exemple de déroulement de l'algorithme du Spanning Tree	44
Les réseaux virtuels.....	45
Problème sur un commutateur	46
Réseau virtuel = réduire l'espace de diffusion	47
Sécurité	48
Comment définir l'appartenance à un VLAN ?	49
Comment définir l'appartenance à un VLAN ? (suite)	50
Comment définir des VLAN sur plusieurs équipements ?	51
Appartenance à un VLAN	52
GARP	53
GVRP	54
Réseau Fédérateur.....	55
Format Propriétaire: ISL (cisco)	56
Format propriétaire : IEEE 802.10	57
IEEE 802.1p + IEEE 802.1Q	58
La couche LLC (Logical Link Control).....	59
Carte d'identité du protocole	60
Ethernet vs IEEE 802.3	61
Format des trames	62

Quelques valeurs de SAP	63
Champ contrôle	64
Questions	65
SNAP.....	67
Carte d'identité du protocole	68
Ethernet vs LLC vs SNAP	69
Format des trames	70
Questions	71
IP - UDP - TCP.....	73
Carte d'identité du protocole IP.....	74
Organisation de la standardisation.....	75
Processus de création des RFC.....	76
Accès libre aux documents	77
Architecture d'un système UNIX	78
Architecture de la couche 3	79
L'adressage IP.....	80
Avant 1994 :	81
Plan d'adressage	82
Notion de sous-réseau / Netmask	83
La table de routage.	84
Questions	85
Faits Historiques	86
Evolution du nombre d'adresses IP	87
Principe	88
Exemple (www.iana.org) Août 2001	89
Allocation administrative	90
Délégation aux Opérateurs	91
Adressage CIDR	92
Multidomiciliation (multi-homing) difficile	93
efficace au niveau mondial ?	94
Evolution des tables de routage du cœur du réseau In-	
ternet	95

Mesures d'urgence	96
Relais applicatifs	97
Traduction d'adresses NAT	98
Pour et contre des NAT	99
La taille de l'internet (source MP Brig SPAWAR, NGI)	100
Les bases de données des RIR.....	101
Les centres de coordination de l'Internet	102
Les autres bases de données	103
Les objets répertoriés	104
La base RIPE.....	105
La base RIPE	106
Aux US, la base ARIN	108
Voir la délégation de bloc correspondante	109
IPv6.....	110
Espace d'Adressage	111
Construction des Adresses	112
Construction d'une adresse IPv6	113
Adresses Globales	114
La Théorie : RFC 2374 (Aggregatable Global Unicast)	115
La pratique : Le 6bone	116
La pratique: les autorités régionales	117
Format des paquets IPv4.....	118
Options	119
Questions	120
Le protocole ARP / Carte d'identité	122
Format d'une trame ARP	123
Questions	124
Questions	125
Le protocole ICMP / Carte d'identité.....	126
Quelques types de paquets ICMP	127

L'outil Traceroute	128	Congestion	162
Exemple de routage avec traceroute (très ancien)	129	Algorithme du slow start	163
Outils : traceroute	131	Exemple	164
Patate Chaude	133	Régime Stationnaire	165
Format des paquets IPv6.....	134	Exercices	167
En-tête IPv6	135	Exemple par simulation	168
Extensions: différence avec les options	136	Exemple par simulation	169
Extension: avec IPv6	137	Exemple par mesure	170
L'ordre est important	138	Plusieurs versions de TCP	171
Neighbor Discovery.....	139	Reno	172
Autoconfiguration	140	Reno	173
Autoconfiguration	141	New Reno	174
Les protocoles de Niveau 4	142	Modélisation de TCP.....	175
Le protocole UDP / Carte d'identité	143	Hypothèses	176
UDP.....	144	Modélisation	177
pseudo-en-tête	145	TCP Equitable ?.....	178
Pseudo en-ête	146	TCP Equitable ?	179
RTP.....	147	Problèmes de TCP.....	180
RTP = Application Layer Framing	148	Exemple	181
RTCP	149	Exemple FIFO	182
Le protocole TCP.....	150	Exemple FIFO	183
Format des messages	151	Exemple RED	184
Ouverture de la connexion	152	Exemple RED	185
transfert de données	153	Explicit Congestion Notification.....	186
transfert avec erreurs	154	Problèmes de TCP.....	187
transfert avec erreur	155	Acquittements Sélectifs	188
fermeture de connexion	156	Problèmes de TCP	189
Adaptation à l'environnement	157	Autres problèmes de TCP	190
Exemple (mesuré dans le noyau SunOs 4.1.3)	158	Nouveaux besoins	191
Limitation du trafic	159	Mécanismes de contrôle de flux par Van Jacobson	192
Contrôle de flux	160	TCP	192
Fenêtre d'anticipation	161	UDP	192

TCP Friendly.....	193	Les extensions de RIP	229
Configuration d'une station Unix.....	196	RIP-II	230
Tables de routage	197	Authentification	231
Questions	198	Authentification par Secret commun	232
Tables de routage (suite)	200	Interior Gateway Routing Protocol (IGRP)	233
Modification des tables de routage	202	IGRP (suite)	234
Modification des tables de routage (ICMP redirect)	203	OSPF (Open Shortest Path First).....	235
Administration des équipements avec SNMP.	204	Exemple	236
Type des variables	205	Tables de routage initiales	237
MIB sous forme d'arbre	206	Base de données OSPF	238
L'arbre des OID plus complet... ..	207	Inondation	239
Exemple de MIB pour IP.	208	algorithme du Short Path First	240
Primitives et Objets	211	Exemple pour le routeur A	241
format des messages	212	Exemple pour le routeur A	242
format des messages (suite)	213	Exemple pour le routeur A	243
format des messages (suite)	214	Exemple pour le routeur A	244
Exemple sur un tableau.	215	Exemple pour le routeur A	245
Exemple suite	216	Exemple pour le routeur A	246
Remarques	217	Exemple pour le routeur A	247
Openview	218	Exemple pour le routeur A	248
Openview (suite)	219	Exemple pour le routeur A	249
Fonctions d'un routeur	220	Arbre des plus courts chemins pour A	250
Les algorithmes de routage.....	221	Mise à jour de la table de routage de A	251
classification des algorithmes de routage.	222	Struturation du réseau	252
Algorithme du Distant Vector.....	223	Division en aire	253
Exemple de déroulement de l'algorithme du Distant Vec-	224	Liaison Virtuelle	254
tor	224	Liaison Virtuelle	255
Problème de convergence	225	Aire pas si terminale	256
Horizon coupé	226	Routes entre les aires	257
RIP.....	227	Exemple	258
format des paquets	228	Diffusion des routes exterieures à l'aire	259
		Mise à jour de la table de routage de A	260
		Mise à jour de la table de routage de B	261
		Routes Externes	262
		Récapitulatif sur les Bases de Données d'OSPF .	263

Le protocole OSPF.....	264	Protocoles	297
Format des En-tête OSPF	265	Exemple	298
Paquet d'Hello	266	Exemple	299
Algorithme du Link State	267	MPLS.....	300
Paquet de description de la base de données	268	Manque de scalabilité de VP/DLCI.	301
paquet de demande d'état de lien	269	Limite des IGP (suite)	302
mise à jour de l'état de lien	270	Améliorer l'efficacité du routage	303
acquiescement.	271	Principe de MPLS (Multi Protocol Label Switching)	304
Exemple.....	272	Avantages de MPLS (suite)	305
Exemple : Hello (équipement isolé)	273	Avantages de MPLS	306
Exemple Hello (équipement isolé)	274	Avantages de MPLS	307
Exemple : Hello (équipement isolé)	275	Principe de fonctionnement	308
Exemple : Echange de Hello	276	Exemple	309
Exemple : Echange de Hello	277	Exemple : fusion de circuit (VC merging)	310
Exemple hello	278	Problème avec ATM	311
Exemple : Description de la base de données	279	Exemple : Ingénierie de Trafic	312
Exemple : Description de la base de données	280	Optimisation du penultieme routeur MPLS	313
Exemple : Description de la base de données	281	Distribution des etiquettes.....	314
Exemple : Description de la base de données	282	techniques de distribution des étiquettes	315
Exemple : Demande d'information	283	techniques de distribution des étiquettes	316
Exemple : Mise à jour de la base de données	284	techniques de distribution des étiquettes	317
Exemple : Mise à jour de la base de données	285	techniques de distribution des étiquettes	318
Exemple : Mise à jour de la base de données	286	Deux modes de distribution	319
Exemple : Mise à jour de la base de données	287	Exemple de mode ordonné	320
Exemple : Mise à jour de la base de données	288	Exemple de mode ordonné	321
Exemple	289	Exemple de mode ordonné	322
Exemple de configuration OSPF.....	290	Protocoles de distribution des étiquettes.....	323
exemple passage de RIP à OSPF	291	LDP : Echanges protocolaires	324
exemple passage de RIP à OSPF	292	Systèmes Autonomes.....	325
exemple passage de RIP à OSPF	293	traceroute 1.2.3.4	326
IS-IS.....	294	La commande whois	327
Adressage	295	Traceroute avec systeme autonome	330
Niveaux	296	Router Advisery	332

Cas du multihoming	333	Les outils de contrôle et de mesure.....	366
Annonces	334	Exemple sur route-server.cerf.net	367
BGP.....	335	Exemple sur route-server.cerf.net	368
Principes	336	Exemple sur route-server.cerf.net	369
iBGP	337	Exemple sur route-server.cerf.net	370
Attributs.	338	Exemple sur route-server.cerf.net	372
Utilisation d'une Adresse LoopBack	339	Exemple sur route-server.cerf.net	373
AS_PATH	340	Exemple sur route-server.cerf.net	374
Next_HOP	341	Exemple sur route-server.cerf.net	375
MULTI_EXIT_DISC	342	Exemple sur route-server.cerf.net	376
LOCAL_PREF	343	Exemple sur route-server.cerf.net	377
Synchronisation	344	Stabilité des routes	378
Synchronisation	345	Dampering	379
Attributs pour iBGP	346	Exemple sur route-server.cerf.net	380
Attributs pour iBGP	347	Exemple sur route-server.cerf.net	381
Règles de selection	348	BGP Multiprotocole.....	382
Règles de selection (suite)	349	BGP & MPLS.....	383
Exemple	350	Cas sans IGP	384
Points d'échange.....	351	Cas avec IGP	385
Point d'interconnexion : France Sfinx	352		
Echange d'Annonces sur le Sfinx	353		
Négociation entre opérateur	354		
Serveur de routes	355		
Le centre du Monde Universitaire	356		
Agrégation	357		
Agrégation très contrôlée	358		
Configuration de BGP sur un Cisco.....	359		
Exemple simple	360		
Traitement des Attributs	361		
Filtrage des NRLI	362		
Filtrage des NRLI	363		
Filtrage des NRLI	364		
Les expressions régulières	365		
		Le multimédia et les réseaux	
		Rappel sur les architectures de réseau	387
		Internet : réseau "best-effort"	388
		Deux fenêtres de Congestion	389
		Fenêtre d'anticipation	390
		Régime Stationnaire	391
		Fenêtre de congestion (algorithme du slow start)	392

Exemple	393	Modification des files d'attente.....	423
Exemple par simulation	394	Equité pondérée	424
Exemple par simulation	395	Exemple	425
Exemple par mesure	396	Remarques	426
Plusieurs versions de TCP	397	Weighted Fair Queueing.....	427
Reno	398	Preuve	428
Reno	399	Preuve (suite)	429
New Reno	400	Equité	430
TCP Equitable ?	401	La réservation de ressources dans l'Internet	
TCP Equitable ?	402	431	
Problèmes de TCP.....	403	Réservation de ressources	432
Exemple	404	Exemple	433
Exemple FIFO	405	Network Calculus	434
Exemple FIFO	406	Network Calculus	435
Exemple RED	407	Network calculus	436
Exemple RED	408	RSVP	437
Problèmes de TCP	409	RSVP.....	438
Exemple	410	RSVP	439
Problèmes de TCP	411	RSVP	440
Autres problèmes de TCP	412	RSVP : Message Path	441
Nouveaux besoins	413	RSVP : Message RESV	442
La qualité de service dans Internet	414	Exemple de service garanti	443
Mécanismes de contrôle de flux par Van Jacobson	415	Message Path reçu par le client	444
TCP	415	Calcul par le client de la réservation	445
UDP	415	Message RESV	446
Peut-on modifier TCP ?	416	Calcul de C et D	447
Différencier les traitements : Oublier la FIFO		Service contrôlé	448
417		RSVP (suite).....	449
Problèmes	418	Sur le niveau 2.....	450
ordonnanceur, pas une question de vitesse	419	ISSLOW	451
résultat de l'ordonnancement	420	IS802	452
Trois types de systèmes	421	La différenciation de services.....	453
Mêmes problème dans les machines.....	422		

Internet à 3 niveaux	454
Service Level Specification	455
Comportement des routeurs	456
Octet DiffServ (DSCP)	457
Classes de services EF	458
Classes de services AF	459
Conclusion.....	460
Conclusion	461
Conclusion.....	463

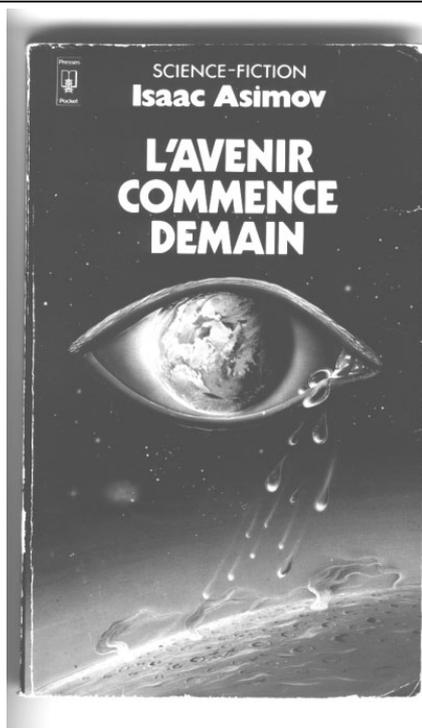
Les réseaux Informatiques

Laurent Toutain

17 janvier 2003 - C:\Documents and Settings\laurent\Mes documents\Support_ENST\Intro.fm

9

ÉVOLUTION DES RÉSEAUX INFORMATIQUES



2

TOUS LES ENNUIS DU MONDE

L'INDUSTRIE la plus vaste de la Terre était centrée sur Multivac... Multivac, l'ordinateur géant qui en cinquante ans avait grandi jusqu'à remplir de ses diverses branches Washington et ses faubourgs, puis avait étendu ses tentacules à toutes les villes et bourgs du globe.

Une armée de fonctionnaires l'approvisionnaient sans cesse de renseignements alors qu'une autre armée interprétait ses réponses et établissait les corrélations entre elles. Un véritable corps du génie patrouillait à l'intérieur, tandis que mines et usines s'acharnaient à maintenir continuellement au complet, avec précision, pour donner pleine et entière satisfaction, les stocks de réserve de pièces de rechange.

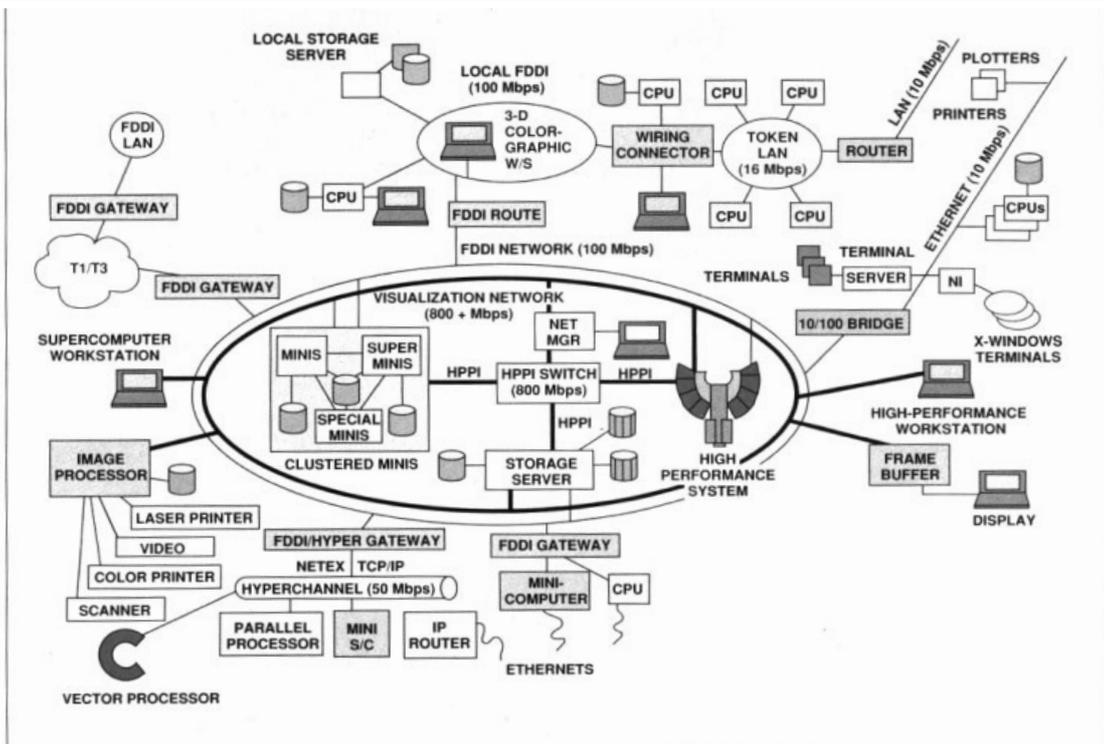
Multivac dirigeait l'économie de la Terre et venait en aide à sa science. Mais il était avant tout le centre de recueil et de classement de tous les faits connus relatifs à chaque individu de la Terre.

Et tous les jours, une partie des devoirs de Multivac consistait à examiner les quatre milliards d'ensembles de faits portant sur chacun des êtres humains qui garnissaient ses entrailles pour extrapoler une nouvelle journée à partir de cette base. Tous les Services de Rectification de la Terre recevaient

11

17 janvier 2003 - C:\Documents and Settings\laurent\Mes documents\Support_ENST\Intro.fm

CAS D'ÉCOLE DANS LES ANNÉES 80



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Intro.fm

TYPES DE RÉSEAUX

EN FONCTION DE LA DISTANCE

Caractéristiques de distance	LAN Local Area Network Réseau Local	MAN Metropolitan Area Network Réseau de campus	WAN Wide Area Network Réseau public
Taille géographique	1 mètre à 2 kilomètres	1 mètre à 200 kilomètres	quelques milliers de kilomètres
Nombre d'abonnés	2 à 200	2 à 1000	plusieurs milliers
Opérateur	l'utilisateur ou son service	regroupement d'utilisateurs	différent des utilisateurs
Facturation	gratuit	forfait	volume et durée
Débits	1 à 100 Mbits/s	1 à 100 Mbits/s	de 50 b/s à 2 Mbits/s
Taux d'erreur	inférieur à 10 ⁻⁹	inférieur à 10 ⁻⁹	10 ⁻³ à 10 ⁻⁶
Délai	1 à 100 ms	10 à 100 ms	inférieur à 0.5 s

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Intro.fm

EN FONCTION DE LA MÉTHODE D'ACCÈS :

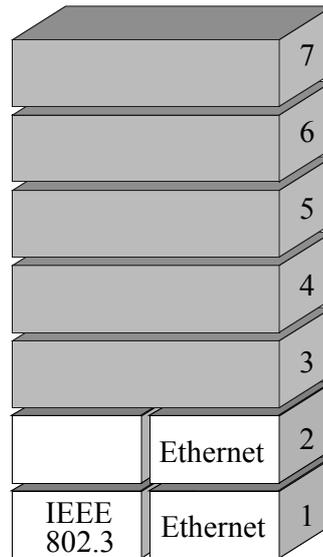
- Diffusion : Ethernet, Token Ring, FDDI.
- **Non Broadcast Multiple Access** : réseau téléphonique, X.25, ATM
- Point-à-point : liaison spécialisée, téléphone rouge, Circuit Virtuel ou Permanent
- Un équipement maître peut diffuser, les autres ne peuvent dialoguer qu'avec le maître : GSM, RNIS.

- ⇒ Pour quel type de réseau, un adressage est-il nécessaire ?
- ⇒ Quels sont les types de réseaux qui résistent au facteur d'échelle ?
- ⇒ Où se trouve Internet dans ce modèle ?

MAINTENANT

- Prédominance au niveau 2 d'Ethernet (LAN et MAN) : IEEE 802.3
 - Augmentation des débits => 100 Mbit/s
 - Réseaux Virtuels
- Disparition des supports partagés
 - câblage en étoile autour d'un équipement actif
 - émulation de la diffusion : Hub
 - Suppression de la diffusion (quand c'est possible) : switches, commutateurs
- Réseaux sans fils : IEEE 802.11
- Technologies WAN :
 - ATM,
 - Frame Relay,
 - LL
- Forte Prédominance au niveau 3 d'IP :
 - Inter-dépendance (par exemple multicast)

ETHERNET / IEEE 802.3



CARTE D'IDENTITÉ DU PROTOCOLE

- Ethernet proposé par Digital, Intel et Xerox (DIX).
- Normalisé par IEEE : IEEE 802.3.
- Ethernet vs IEEE 802.3 :
 - même format de trame
 - affectation des champs différente => Deux protocoles INCOMPATIBLES
- Protocole évolutif :
 - vitesse de transmission (10, 100, 1 000, projets à 10 000 Mbit/s)
 - support (coaxial, paire torsadée, fibre optique)
 - méthode d'accès (CDMA/CD, half duplex, full duplex)

ÉTAT DE LA NORMALISATION

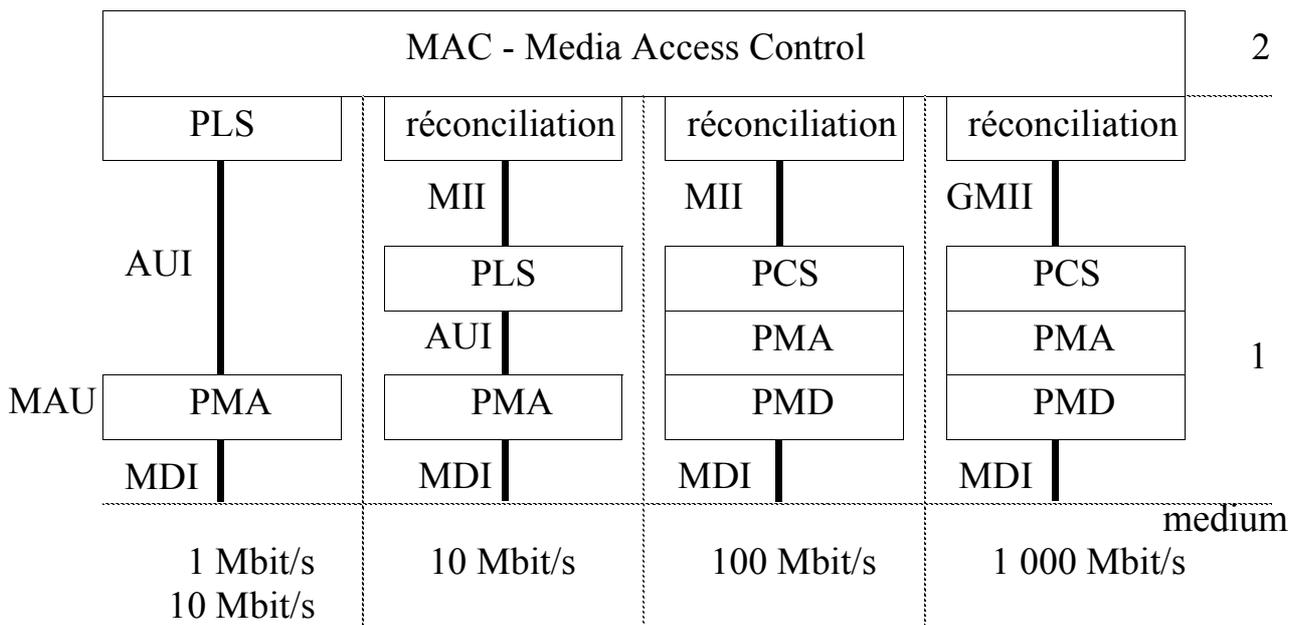
1 Mbit/s	Paire Torsadée	1BASE5		HD
10 Mbit/s	coaxial	10BASE5 <i>Thick Ethernet</i>		HD
		10BASE2 <i>Thin Ethernet</i>		HD
	TV	10BROAD36		HD
	Fibre Optique	10BASE-F	10BASE-FB	HD / FD
			10BASE-FL	HD / FD
			10BASE-FP	HD / FD
Paire Torsadée	10BASE-T		HD / FD	
100 MBit/s	PT	100BASE-T	100BASE-T2	HD / FD
			100BASE-T4	HD
		100BASE-X	100BASE-TX	HD / FD
	Fibre Optique		100BASE-FX	HD / FD
1000 Mbit/s	Fibre Optique	1000BASE-X	1000BASE-LX	HD / FD
			1000BASE-SX	HD / FD
	Paire Torsadée		1000BASE-CX	HD / FD
			1000BASE-TX	HD / FD

HD : Half Duplex - FD : Full Duplex

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Ethernet_plus.fm

ARCHITECTURE

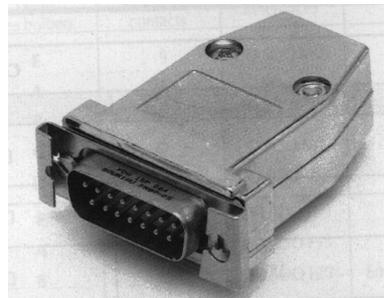
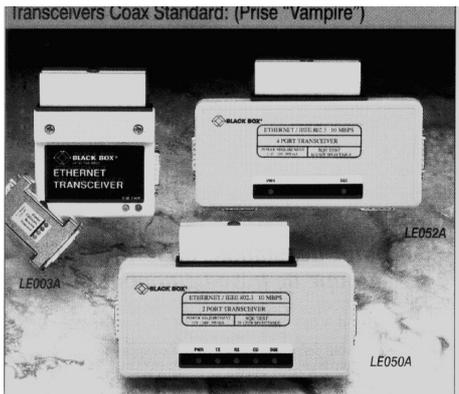
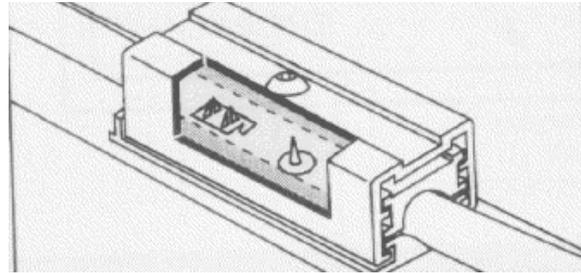
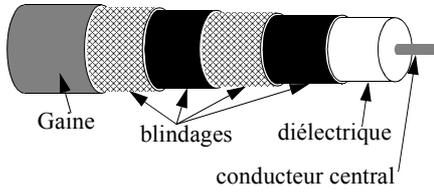
couche ISO



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Ethernet_plus.fm

EQUIPEMENTS

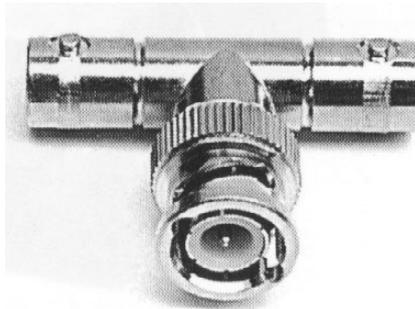
10 BASE 5 (THICK ETHERNET)



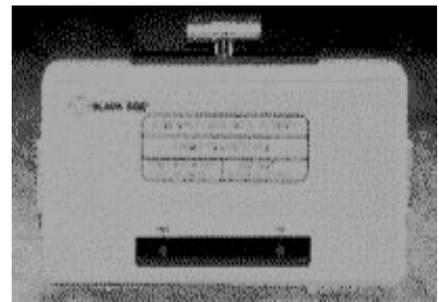
10 BASE 2 (THIN ETHERNET)



câble thin ethernet

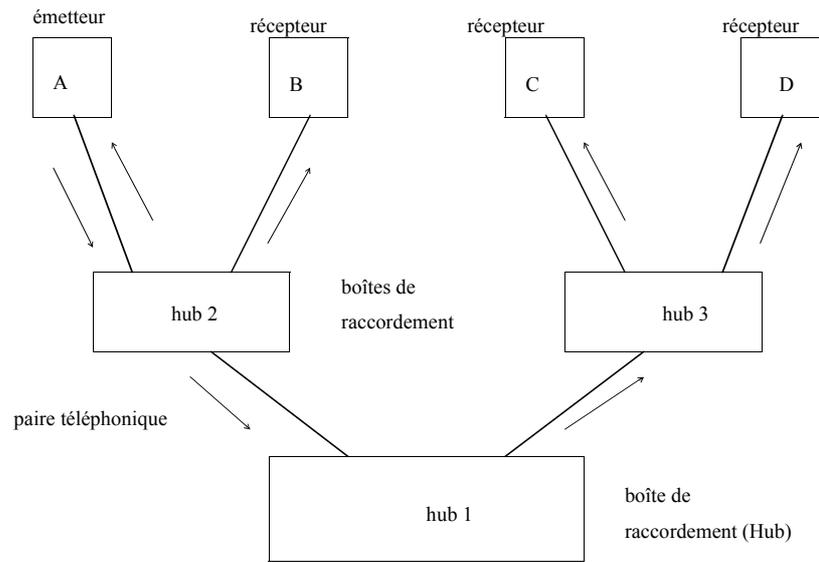


prise BNC en T



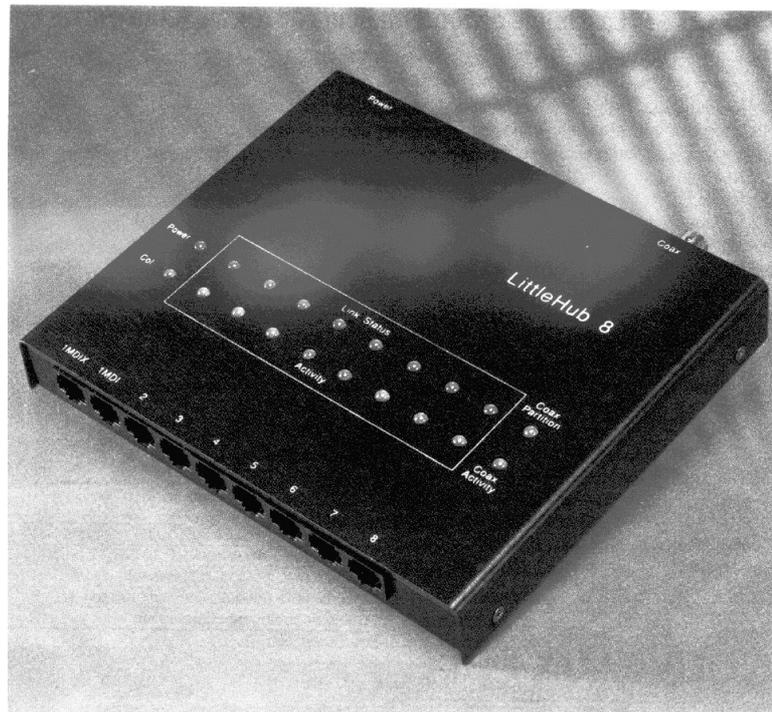
Transceiver

HUB : EMULATION D'UN BUS



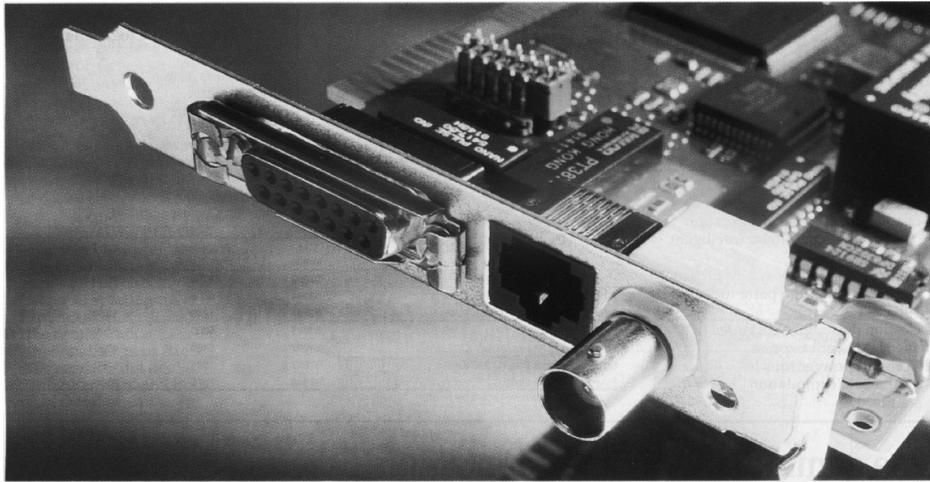
17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\Ethernet_plus.fm

MATÉRIEL



17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\Ethernet_plus.fm

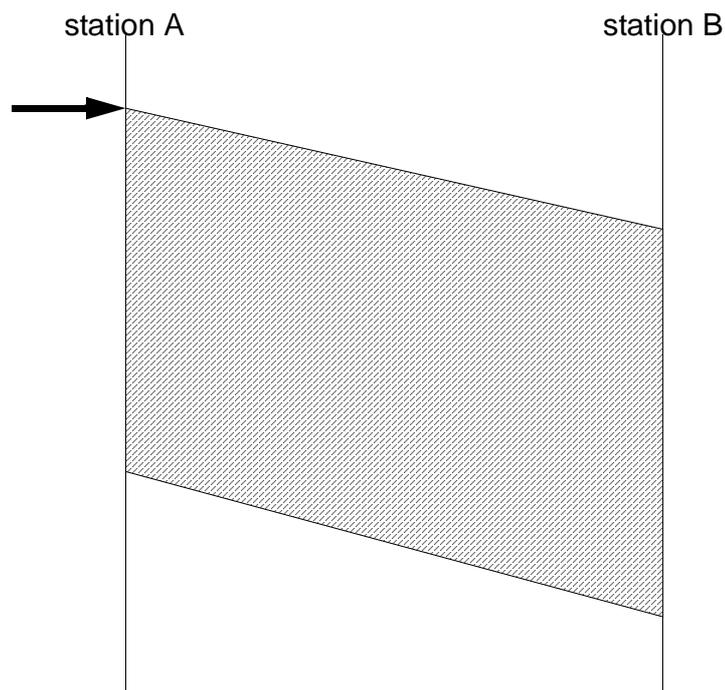
SUR UN PC



+ A quel type de réseau Ethernet correspondent les différentes prises de cette carte ?

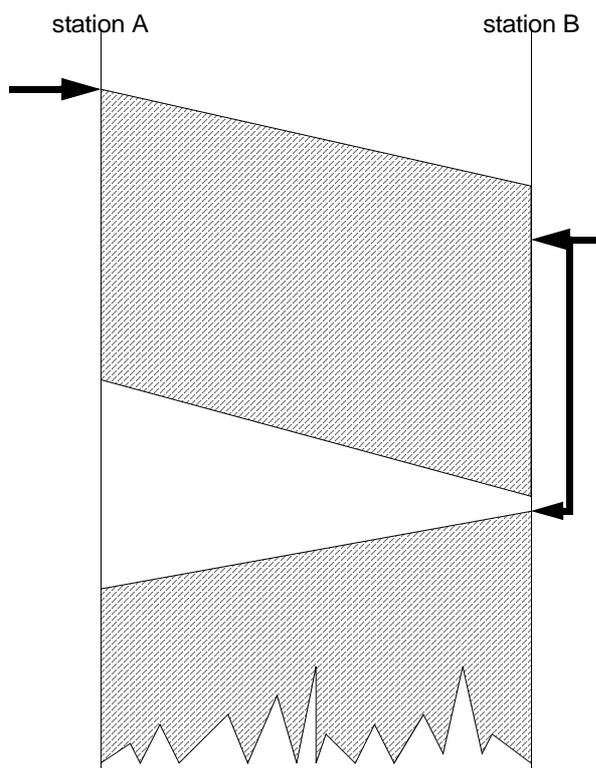
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Ethernet_plus.fm

ALGORITHME DU CSMA/CD

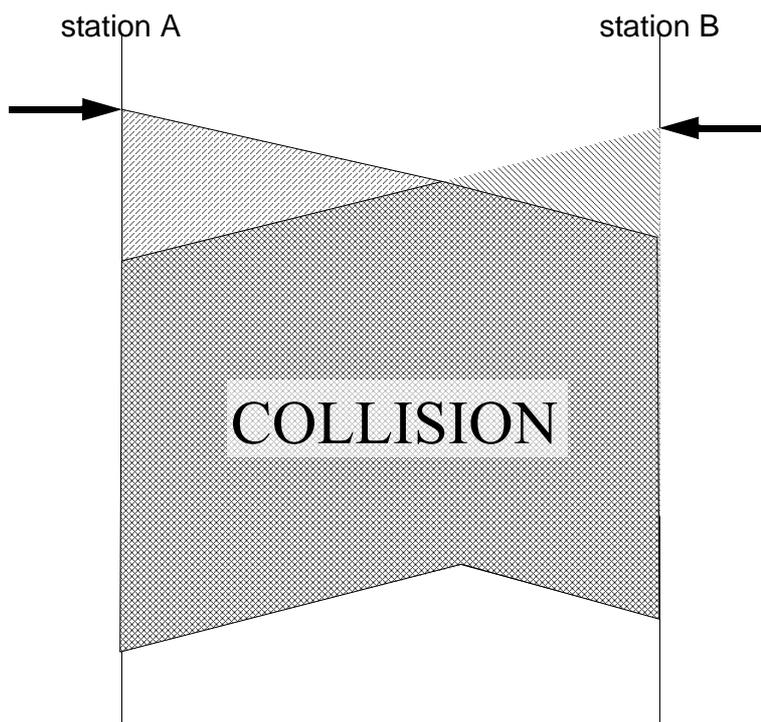


17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Ethernet_plus.fm

ALGORITHME DU CSMA/CD



ALGORITHME DU CSMA/CD



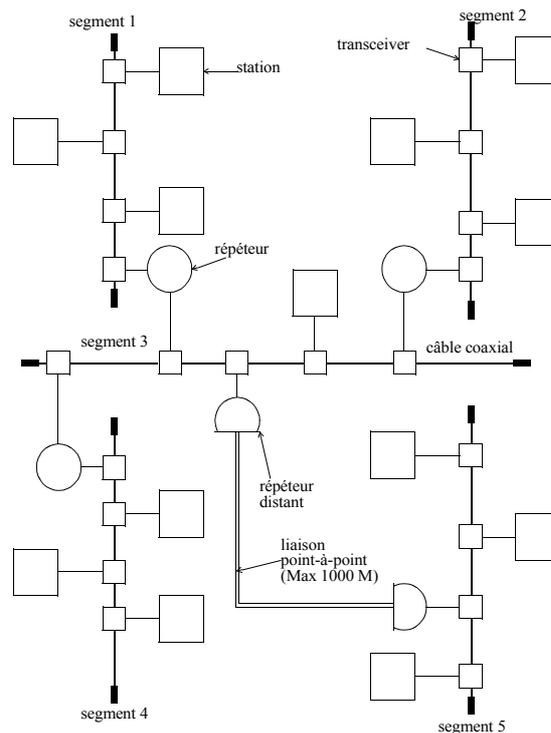
LES COLLISIONS

- Il faut que TOUTES les stations soient dans le même état
 - ⇒ La durée d'émission doit être d'au moins 2 fois la durée de propagation du signal

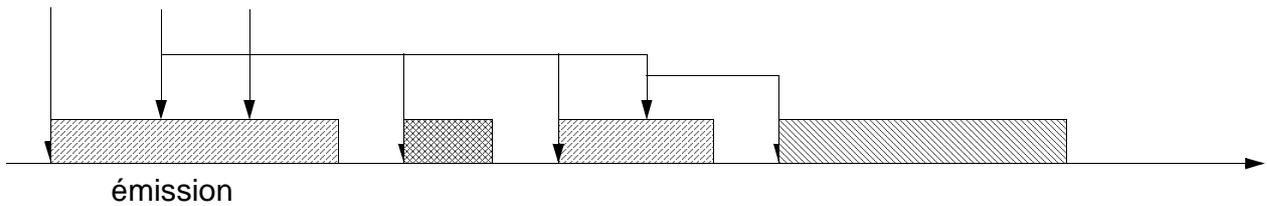
- Si la trame est trop courte, il faut ajouter des bits de bourrage.
 - ⇒ La topologie doit être limitée pour éviter des durées de propagation qui forceraient à allonger la longueur des trames

- La durée minimale d'émission est de 51.2 μ s (soit 64 octets pour Ethernet à 10 Mbits/s).
 - ⇒ Les règles de câblage sont (pour le 10 base 5):
 - des segments de 500 mètres maximum,
 - traversée de 4 répéteurs.

TOPOLOGIE D'UN RÉSEAU ETHERNET



RÉSOLUTION DES COLLISIONS



- si collision attende de 0 ou 1 Tranche Canal
- si de nouveau collision : attente de 0, 1 2 ou 3 Tranche Canal
- si > 10 collisions : attente de 0 à 2^{10} Tranche Canal
- si > 16 collisions : arrêt de la tentative

+ Pas de garantie d'émission (ni de borne)

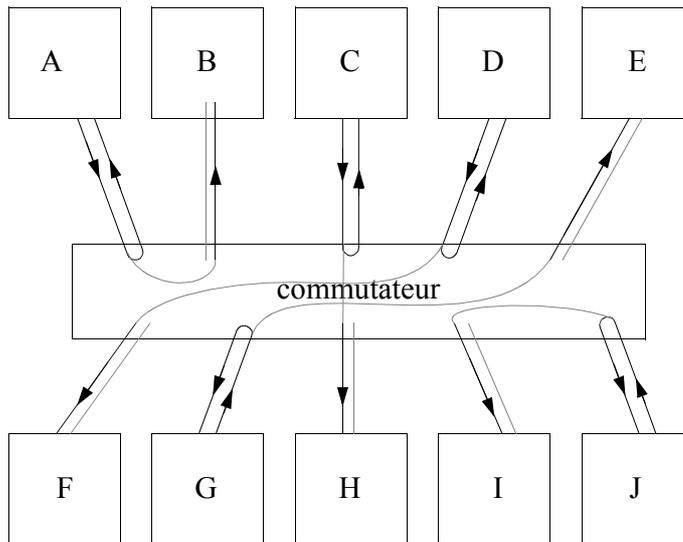
QUAND S'APPLIQUE LE CSMA/CD ?

- topologie en bus
- Emulation d'un bus avec un Hub
- Avec un commutateur et un équipement Half-duplex (utilisé pour faire du contrôle de flux)

+ Le CSMA/CD n'est pas mis en œuvre dans le cas d'un commutateur et d'équipements terminaux full-duplex :

- suppression des contraintes de câblage
- augmentation des débits
- problème de contrôle de flux

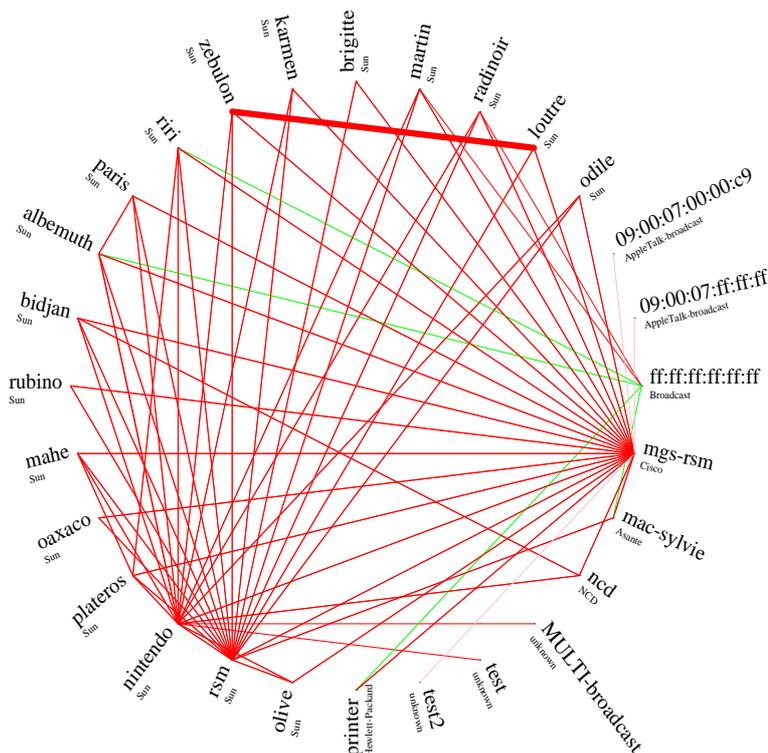
LES COMMUTATEURS (SWITCHS)



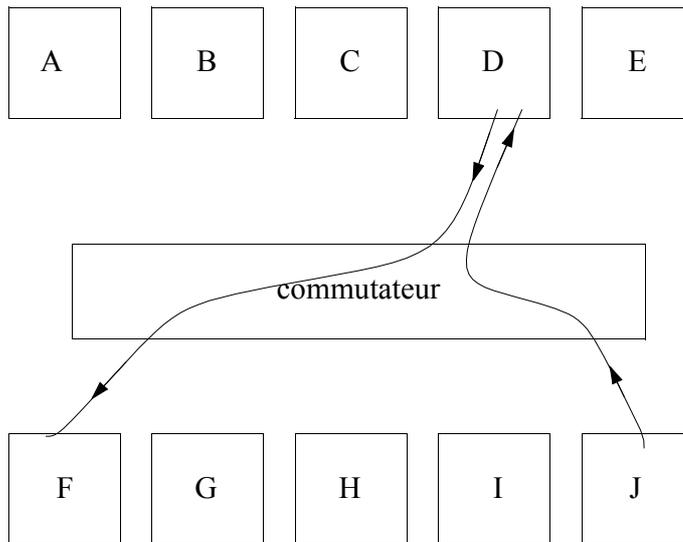
- à la volée
- stockage et retransmission

+ Switch n'implique pas full duplex (les équipements doivent l'être)

EXEMPLE DE TRAFIC

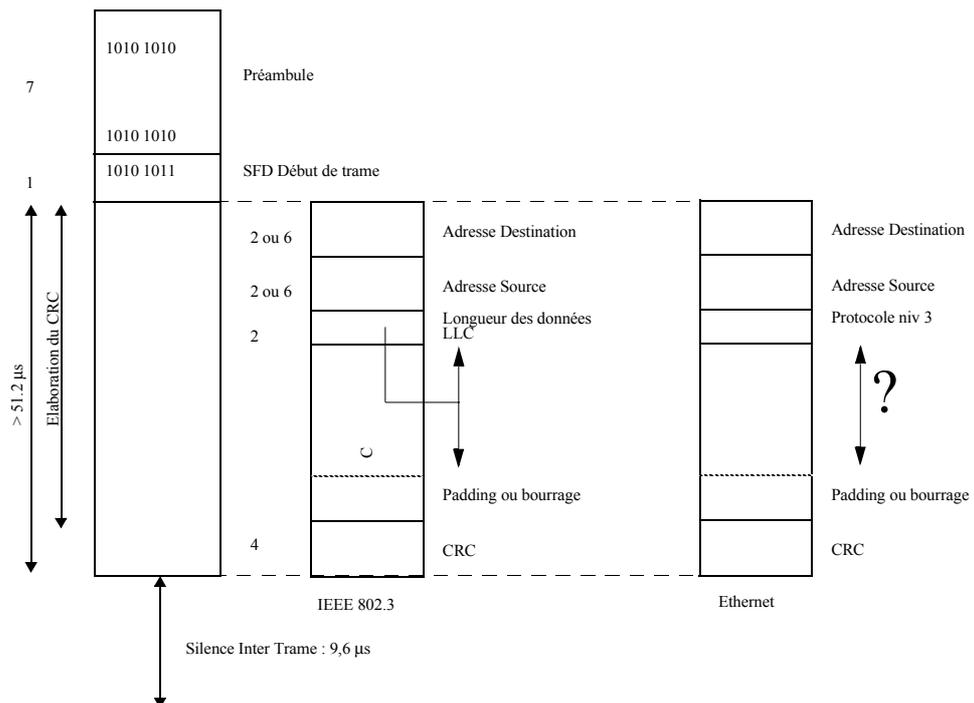


LES COMMUTATEURS (SWITCHS)

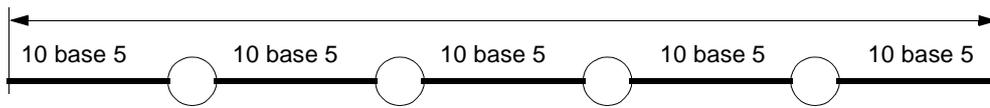


- Problème de contrôle de flux (le CSMA/CD peut être vu comme un algorithme de contrôle de flux)
- message PAUSE

FORMAT DES TRAMES



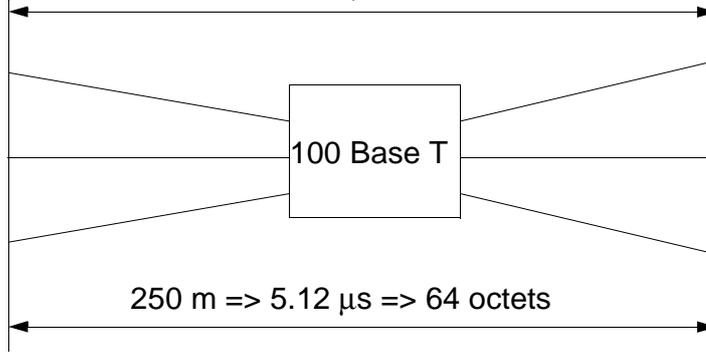
INFLUENCE DE LA VITESSE SUR LE FORMAT DES TRAMES



Solution 1

2.5 Km => 51.2 μ s => 640 octets

Interconnexion ???

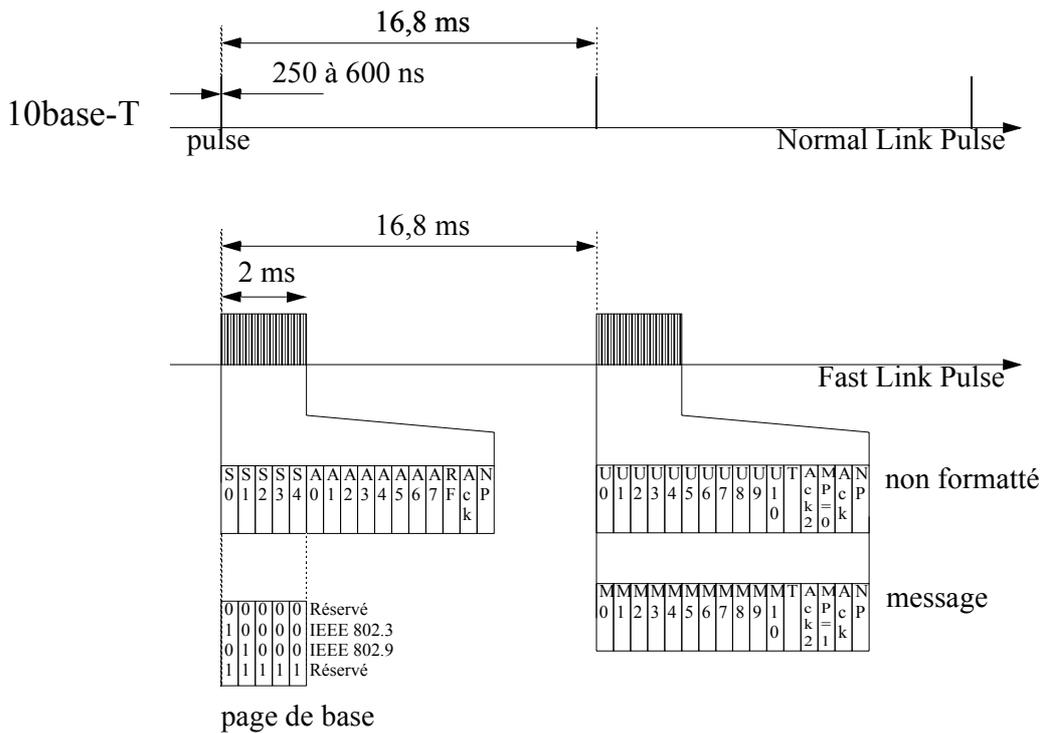


Solution 2

250 m => 5.12 μ s => 64 octets

+ Quelle distance pour Ethernet 1 Giga ?

AUTO-NÉGOCIATION



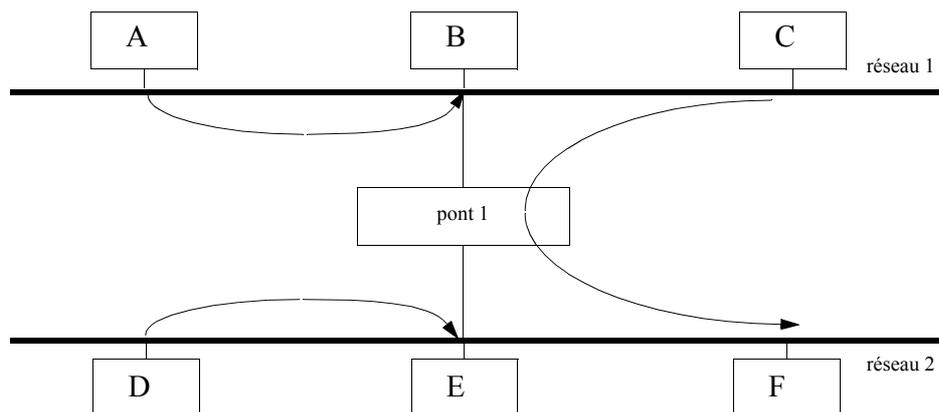
LES PONTS TRANSPARENTS

- Principe de fonctionnement :
 - Le pont écoute toute l'activité en mode *Promiscuous*.
 - Il stocke dans sa mémoire les trames.
 - Il retransmet vers le (ou les) autre(s) sous-réseau(x) les messages stockés .
- Un pont n'a pas besoin d'adresse MAC pour fonctionner
- Un pont n'a pas besoin d'être configuré.

QUESTIONS

- ⇒ Quelle adresse est mise dans le champ source des trames recopiées ?
- ⇒ Quels protocoles de niveau 3 peuvent être utilisés ?

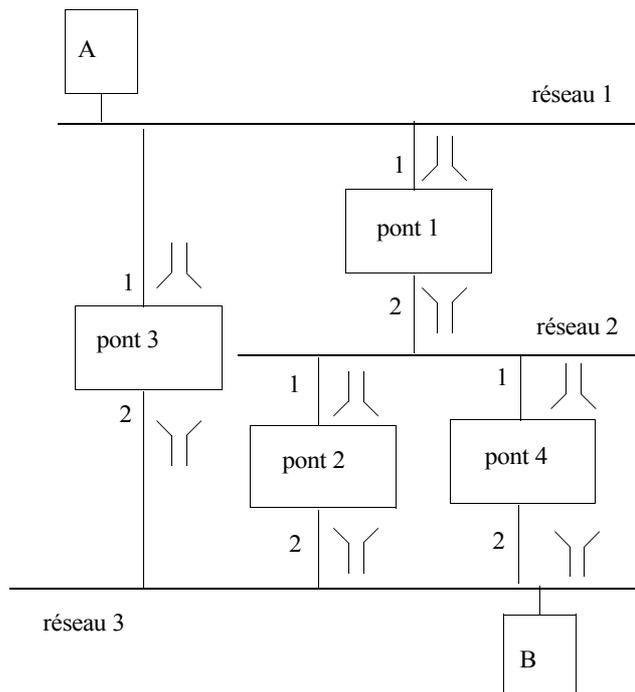
PONTS FILTRANTS



- Les ponts permettent de réduire le trafic global du réseau
- Toujours pas besoin d'adresse MAC ou d'administration

- ⇒ Est-ce toujours valable avec 3 réseaux et 2 ponts ?

CAS COMPLEXE



Que se passe-t-il quand A émet un message vers B ?

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Pontage.fm

CAS COMPLEXE (SUITE)

Quand plusieurs routes sont possibles, il y a duplication des messages et inondation du réseau

- + IL faut absolument éviter les boucles lors de l'interconnexion des réseaux.
- + Mais :
 - la redondance doit toujours être possible
 - possibilité d'erreur
- + Mettre en œuvre dans les ponts un protocole qui permettra de trouver un arbre couvrant (*Spanning Tree*).
- + Les ponts doivent échanger des messages

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Pontage.fm

ALGORITHME DU SPANNING TREE

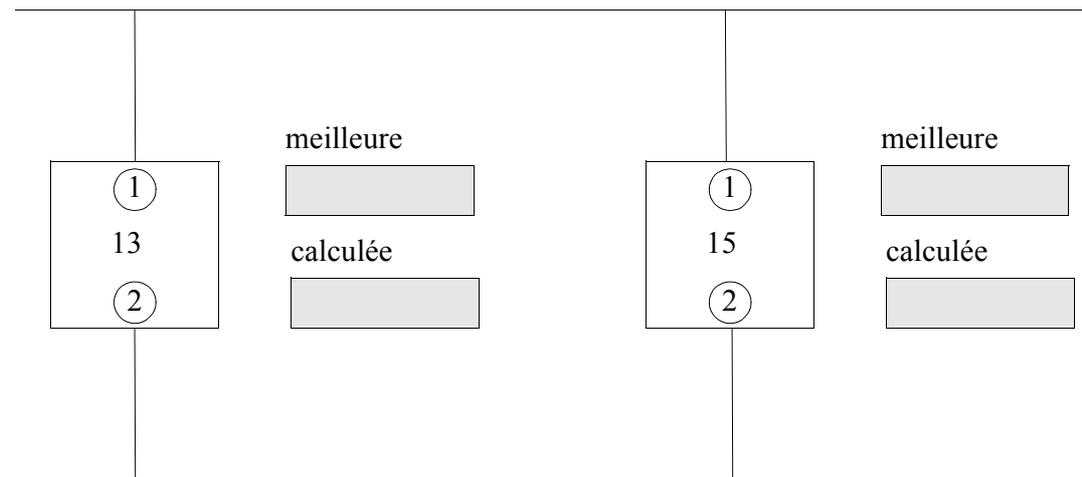
Les ponts vont échanger des messages contenant :

- L'identité supposée de la racine. A l'initialisation, ils se supposent racine.
- Le coût supposé de la liaison. Pour un pont racine, ce coût est nul.
- L'identité de l'émetteur.
- Le numéro du port sur lequel le message est émis.

L'algorithme pour chaque pont est le suivant :

- Recherche du meilleur message sur ses ports.
- Si une des configurations est meilleure que la configuration :
- Cette voie devient le chemin pour la racine.
Une nouvelle configuration est calculée. Le coût est augmenté de 1.
 - Les ports qui ont une configuration meilleure que celle nouvellement calculées sont désactivés Les autres ports font partis du *Spanning Tree*.
 - Cette configuration sera émise sur les ports autres que celui qui mène vers la racine
- Si aucun message n'est meilleur que celui émis par le pont, celui-ci se considère comme racine.

EXEMPLE DE DÉROULEMENT DE L'ALGORITHME DU SPANNING TREE



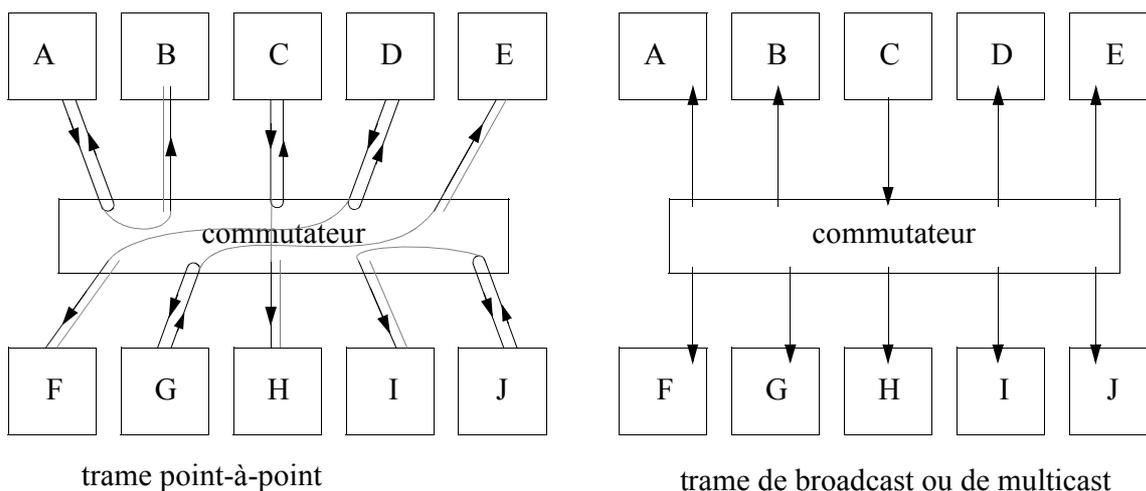
- Garder les avantages des réseaux de niveau 2 :
 - autoconfiguration (adresses, pontage (Spanning Tree), ...)
 - rapidité de transmission grâce à la commutation

- Offrir des fonctionnalités de réseaux de niveau 3 :
 - extensibilité
 - utilisation sur de grandes distances
 - filtrage du trafic

+ Réduire l'espace de diffusion

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Pontage.fm

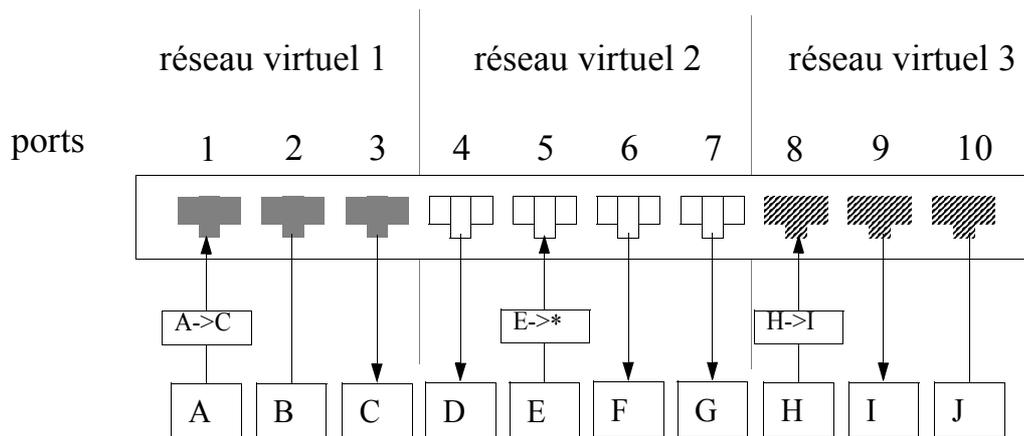
PROBLÈME SUR UN COMMUTEUR



+ Même problème avec un réseau construit avec des demi-ponts et des LL.

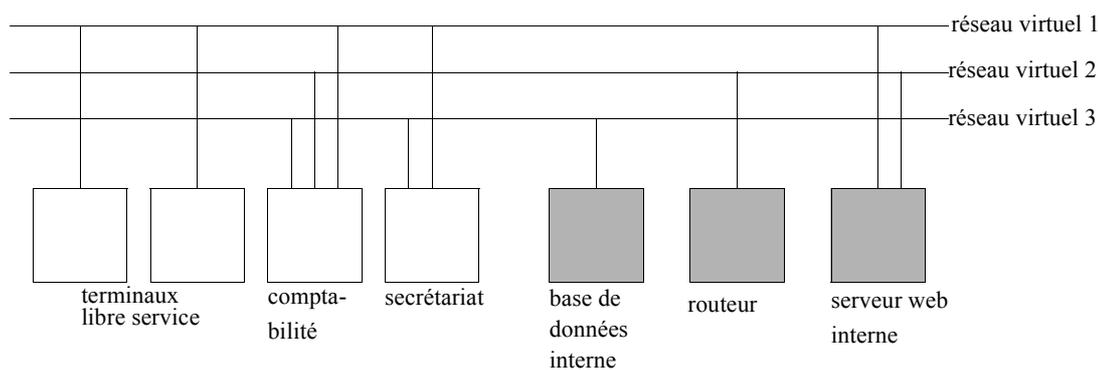
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Pontage.fm

RÉSEAU VIRTUEL = RÉDUIRE L'ESPACE DE DIFFUSION



- La communication entre les réseaux virtuels est impossible
=>Utilisation d'un routeur
- Les VLAN fonctionnent sur les ponts et les commutateurs

SÉCURITÉ



+ Un seul réseau physique

⇒ De combien d'adresses IP doivent disposer les équipements ?

COMMENT DÉFINIR L'APPARTENANCE À UN VLAN ?

- Par port :
 - simplicité d'administration
 - très sécurisé
 - manque de souplesse pour les utilisateurs nomades

- Par adresse MAC :
 - souplesse pour les nomades
 - manque de sécurité
 - difficulté d'administration (il faut manipuler les adresses MAC)
 - problèmes avec les protocoles de niveau 3 (par exemple : sous-réseau IP)
Doit être lié avec des protocoles de configuration automatique (DHCP)

COMMENT DÉFINIR L'APPARTENANCE À UN VLAN ? (SUITE)

- Par protocole de niveau 3 :
 - Permet de limiter les trames en diffusion (Novell : RIP; SAP; IP : ARP)

- Par adresse de niveau 3 :
 - Plus de souplesse pour les nomades
 - Un réseau virtuel par réseau IP
 - Plus lent car analyse des trames/paquets plus complexe

- Combinaison de ces différents modes

COMMENT DÉFINIR DES VLAN SUR PLUSIEURS ÉQUIPEMENTS ?

- Enorme problème de normalisation, solutions constructeurs

- Arrivée prochaine de normes IEEE :
 - IEEE 802.1p (supplément à IEEE 802.1D) + IEEE 802.1Q :
numérotation des VLAN, partage d'information entre les équipement.

- Utilisation intensive de plateformes d'administration.

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Pontage.fm

APPARTENANCE À UN VLAN

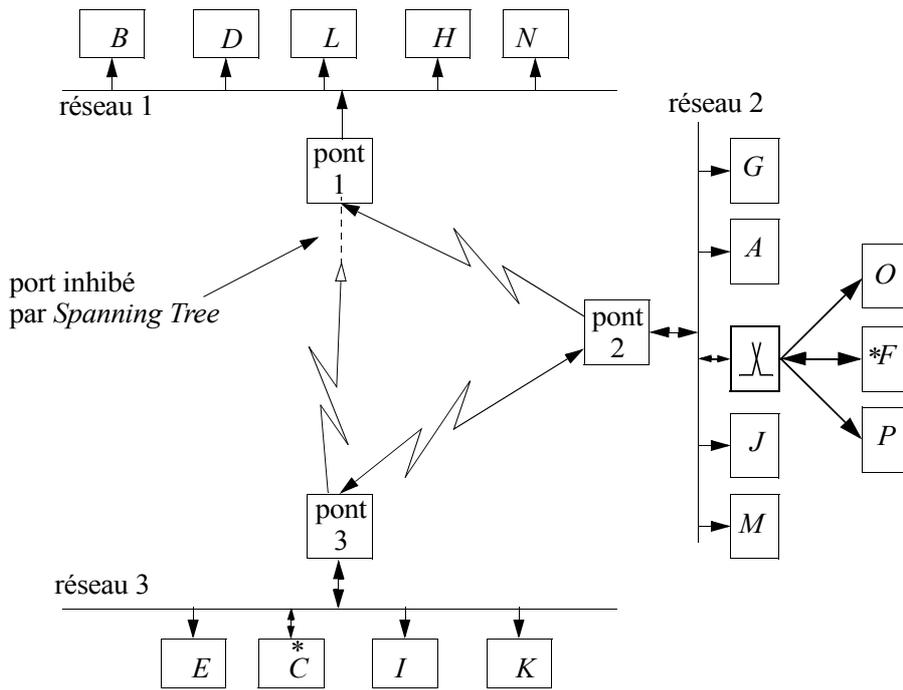
- La machine 12-34-45-67-89 appartient au VLAN 1

- Comment faire pour partager cette information à tous les VLAN ?
 - plateforme d'administration :
 - découvertes des équipements d'interconnexion du réseau
 - découvertes des stations (adresses MAC) utilisées sur le réseau
 - attribution d'un VLAN à une adresse MAC (outil graphique)
 - envoi de cette information à tous les équipements d'interconnexion

 - Protocole d'inondation des équipements d'interconnexion :
 - GVRP : GARP VLAN Registration Protocol
 - GARP : Generic Attribute Registration Protocol

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Pontage.fm

GARP



3 états des stations :

- très anxieuse
- anxieuse
- tranquille

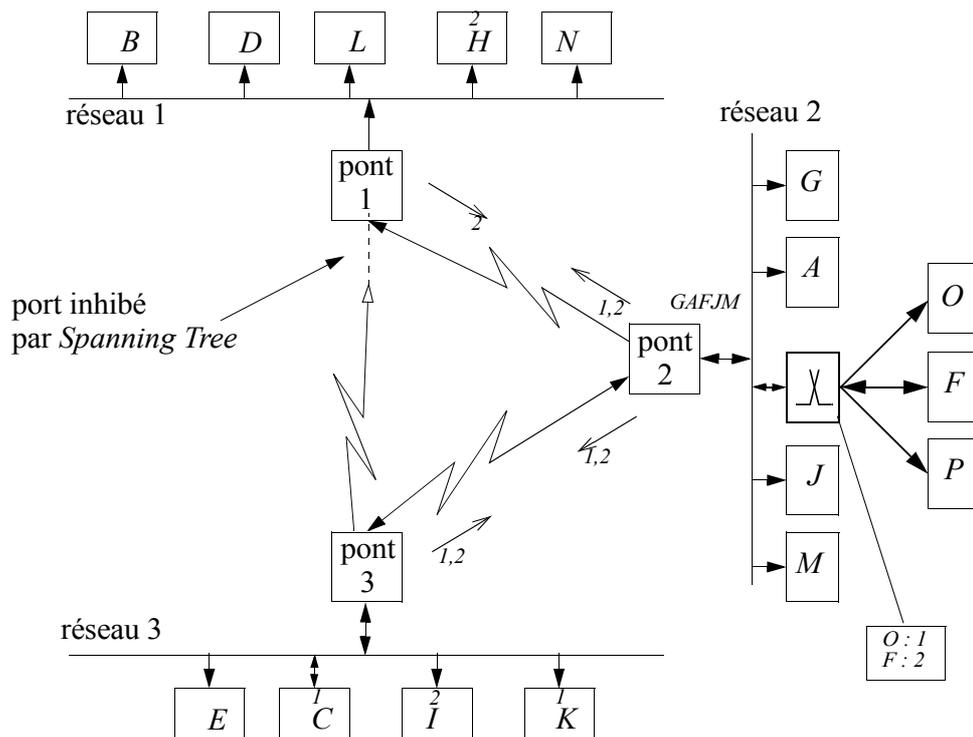
Plusieurs rôles :

- Observateur
- Membre :
- actif
- passif

Périodiquement reconstitution du groupe.

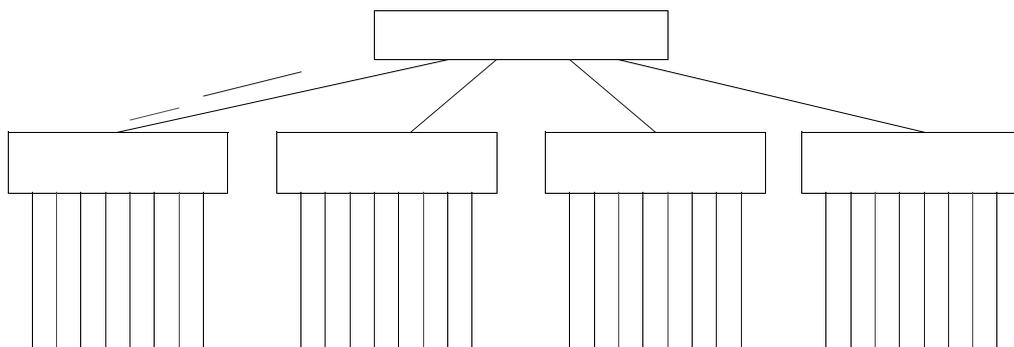
17/ janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Pontage.fm

GVRP



17/ janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Pontage.fm

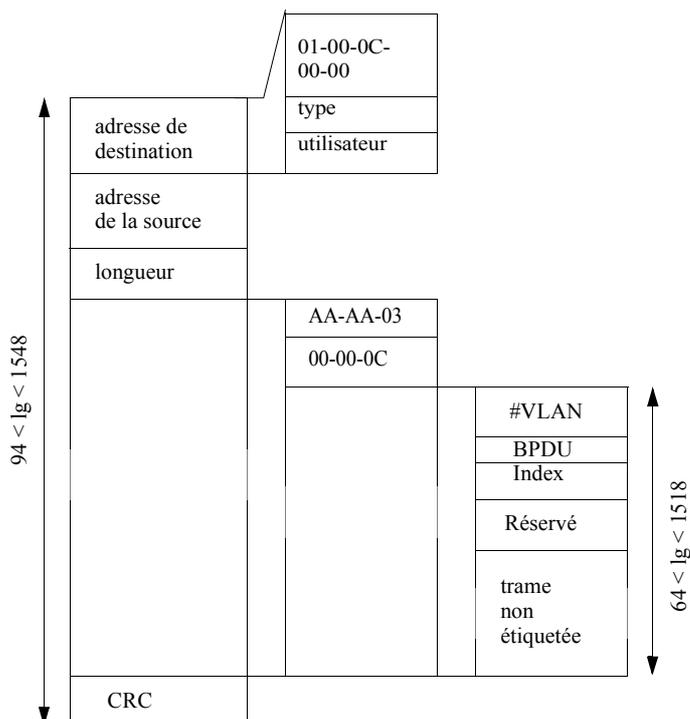
RÉSEAU FÉDÉRATEUR



- Marquage des trames sur le réseau fédérateur
- Trames normales sur le réseau entre le dernier équipement d'interconnexion et la station.

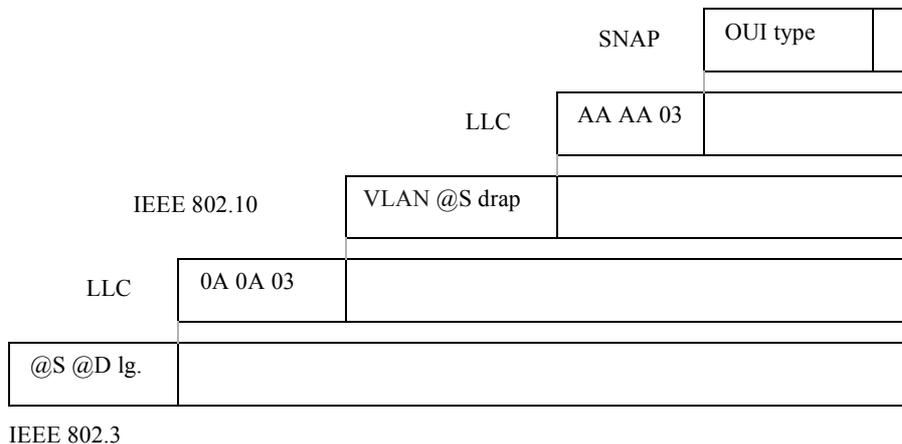
+ Solution constructeurs (Cisco : ISL, ...)

FORMAT PROPRIÉTAIRE: ISL (CISCO)



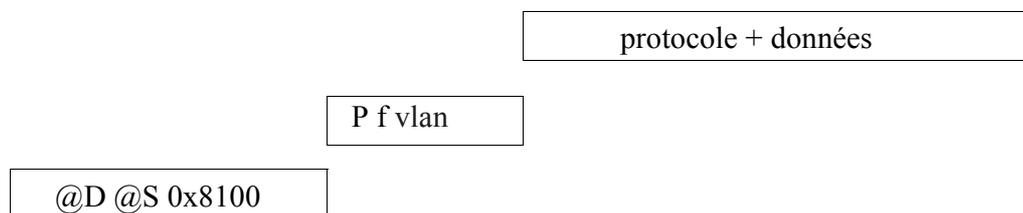
FORMAT PROPRIÉTAIRE : IEEE 802.10

- Norme pour le chiffrement des données
- Réutilisation pour la numérotation des VLAN
- Dans le cas de matériel hétérogène sur un anneau FDDI
- Encapsulation coûteuse.



IEEE 802.1P + IEEE 802.1Q

- Nouvelle norme, peu de produit
- Encapsulation réduite



- 12 bits pour numérotter le VLAN
- Taille des trames peut dépasser 1518 octets => problème de compatibilité avec les anciens équipements
- Projet de norme IEEE 802.3ac pour autoriser des longueurs de trames plus grandes

LA COUCHE LLC (LOGICAL LINK CONTROL)



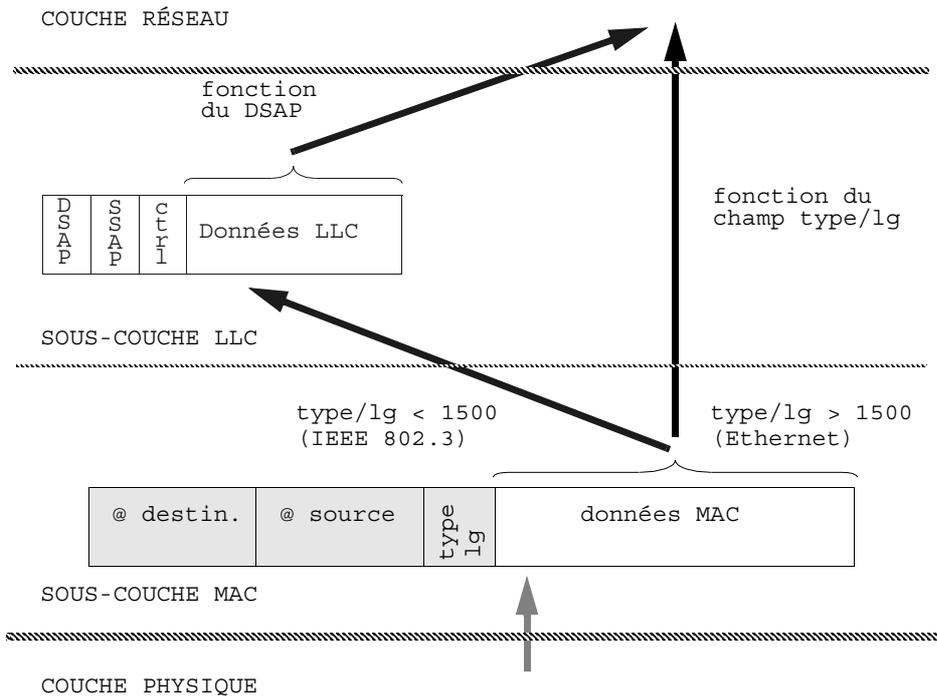
CARTE D'IDENTITÉ DU PROTOCOLE

- Permet de contrôler la communication entre les entités distantes

- Trois modes de fonctionnement :
 - LLC type 1 ou mode datagramme.
 - LLC type 2 ou mode transfert sur connexion.
 - LLC type 3 ou mode datagramme acquitté.

- N'est pas utilisé par Ethernet, uniquement par IEEE 802.3 et IEEE 802.5

ETHERNET VS IEEE 802.3



FORMAT DES TRAMES

DSAP 8 bits	SSAP 8 bits	Control 8 ou 16 bits	information 8*M bits
----------------	----------------	-------------------------	-------------------------

- Le SAP (Service Access Point) permet de désigner la couche supérieure. Le format du SAP est :

AAAAAAUI avec

- AAAAAA : Numéro du SAP.
- U : adresse universelle ou locale,
- I : adresse individuelle ou de groupe.

QUELQUES VALEURS DE SAP

SAP hexa.	SAP décimal	SAP binaire	signification	Equivalent Ethernet
0x00	0	0000 0000	SAP Nul	
0x02	2	0000 0010	Gestion de la couche LLC	
0x06	6	0000 0110	Réseau IP	0x800
0x42	66	0100 0010	Gestion du <i>Spanning Tree</i>	
0x7E	126	0111 1110	X.25 niveau 3	0x805
0xAA	170	1010 1010	SNAP	
0xE0	224	1110 0000	IPX: (protocole du réseau NetWare de Novell)	0x8137

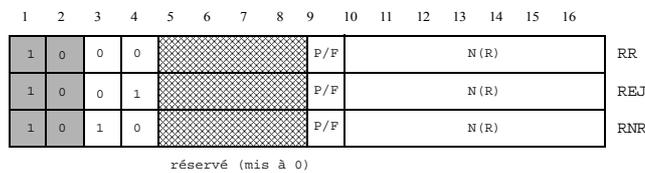
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\LLC_SNAP.fm

CHAMP CONTRÔLE

Trame de type I (Information)



Trames de type S (Supervision)



Trames de type U (Non Numérotées)



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\LLC_SNAP.fm

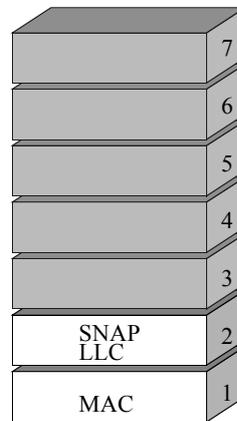
QUESTIONS

- + Combien de SAP peuvent être attribués ?
- + Quelle est la longueur de l'en-tête d'un datagramme de type 1 ?
- + Comment faire la différence entre une trame Ethernet et IEEE 802.3 ?
- + IEEE 802.3 et Ethernet peuvent-ils cohabiter sur un même réseau ?
- + Un répéteur peut-il être utilisé pour interconnecter un réseau Ethernet à un réseau IEEE 802.3 ?

+ Poursuivre le désassemblage de la trame IEEE 802.5 :

```
0 - 10 40 00 00 1D A0 CE 51    00 00 1D A0 D1 33 AA AA
16 - 03 00 00 00 08 00 45 00    00 82 01 01 40 00 FF 11
32 - 60 0B C0 2C 4D 01 C0 2C    4D 04 00 A1 00 A1 00 6E
48 - 00 00 30 64 02 01 00 04    06 70 75 62 6C 69 63 A0
64 - .....
```

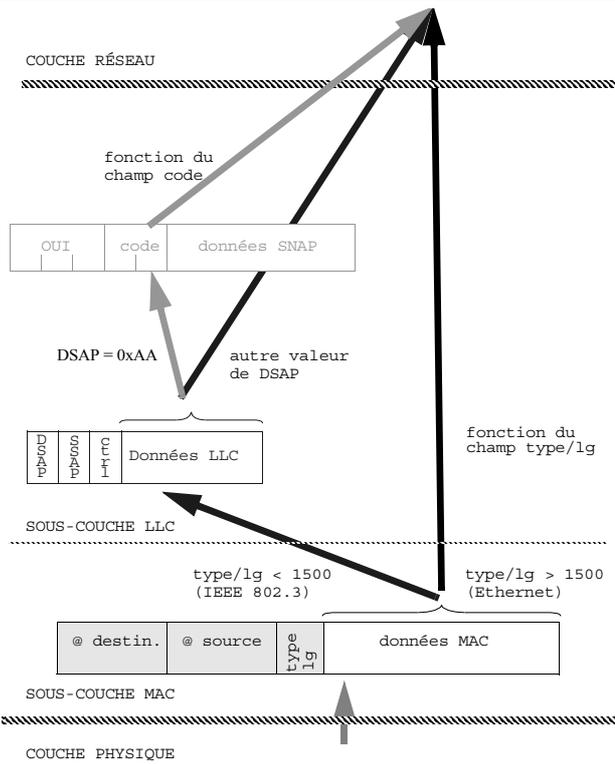
SNAP



Carte d'identité du protocole

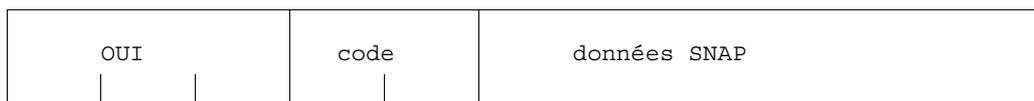
- Permet de combler les lacunes de LLC :
 - Adresse sur un nombre impair de bits : peu performant,
 - Espace de valeurs des SAP limités,
 - Compatibilité avec l'adressage d'Ethernet.
- SNAP propose une encapsulation supplémentaire au dessus de LLC
- Valeur du SAP : 0xAA
- SNAP ne met pas en œuvre de protocole supplémentaire

ETHERNET VS LLC VS SNAP



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\LLC_SNAP\m

FORMAT DES TRAMES



-
-
- OUI : Organizational Unit Identifier (3 octets) désigne un vendeur. Identique aux 3 premiers octets d'une adresse MAC
- Généralement le OUI est à 0 dans les trames SNAP
- Code : identique aux codes rencontrés dans le champ `proto` des trames Ethernet.

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\LLC_SNAP\m

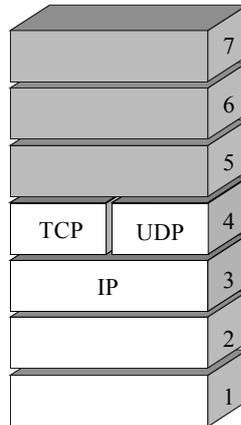
QUESTIONS

+ Poursuivre le désassemblage de la trame IEEE 802.5 :

```
0 - 10 40 00 00 1D AO CE 51    00 00 1D AO D1 33 AA AA
16 - 03 00 00 00 08 00 45 00    00 82 01 01 40 00 FF 11
32 - 60 0B CO 2C 4D 01 CO 2C    4D 04 00 A1 00 A1 00 6E
48 - 00 00 30 64 02 01 00 04    06 70 75 62 6C 69 63 AO
64 - .....
```

+ Quel serait le format pour des trames Ethernet et IEEE 802.3

IP - UDP - TCP



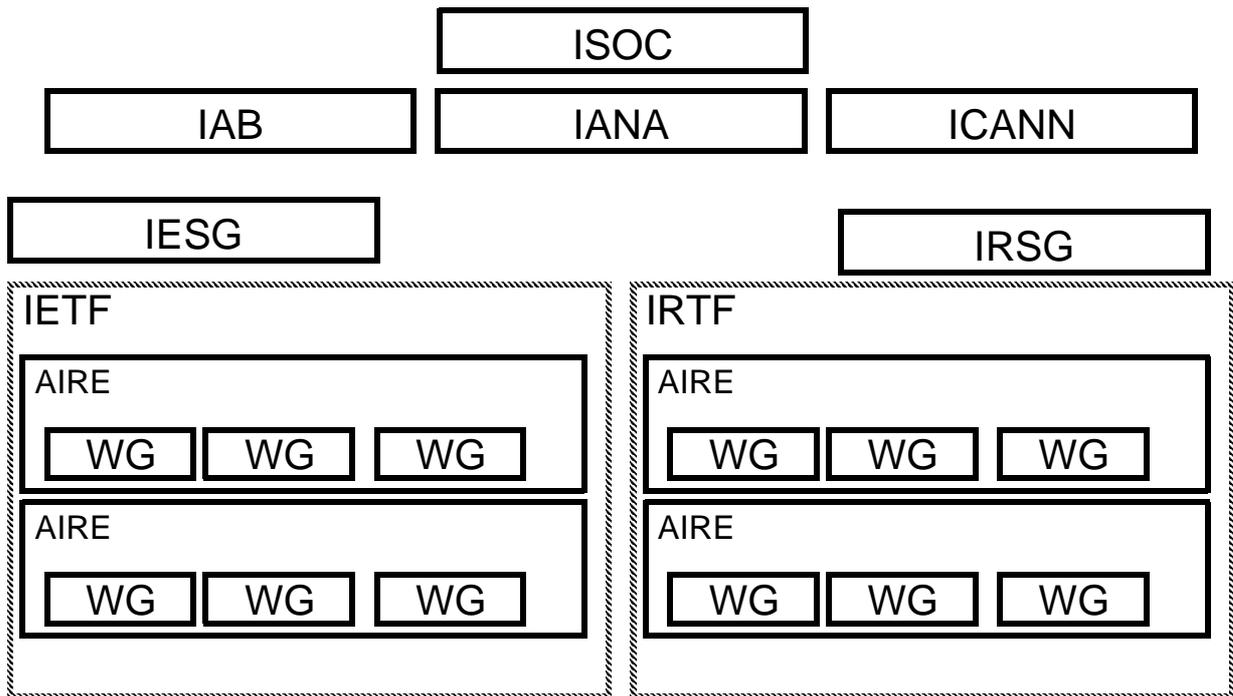
17 janvier 2003 - C:\Documents and Settings\loutain\Mes documents\Support_ENST\IP.fm

CARTE D'IDENTITÉ DU PROTOCOLE IP

- Protocole en mode datagramme
- La portée des paquets IP est mondiale => Importance de l'adressage
- À la base du plus grand réseau informatique mondial : Internet

17 janvier 2003 - C:\Documents and Settings\loutain\Mes documents\Support_ENST\IP.fm

ORGANISATION DE LA STANDARDISATION



PROCESSUS DE CRÉATION DES RFC

Standard track :

- Soumission d'un draft personnel
 - mail à draft-editor@ietf.org
 - draft-tell-mon-sujet-favori-00.txt
- via les mailing list et les meeting IETF (3 fois par an)
 - faire adopter le draft comme working group item
 - draft-ietf-wg-mon-sujet-favori-00.txt
 - atteindre un consensus sur la mailing list (last call)
 - donner le document à un directeur d'aire
 - last call dans tous les groupes
 - si acceptation : envoi au rfc editeur et à l'IANA si nécessité d'allouer des valeurs protocolaires.
- RFC : proposed standard, puis draft standard puis standard

ACCÈS LIBRE AUX DOCUMENTS

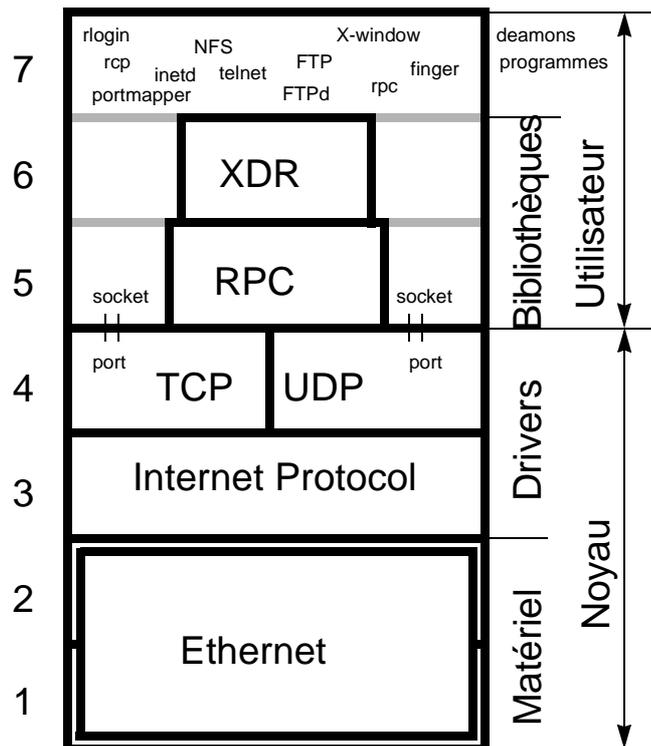
- www.ietf.org
- accès aux working groups (adresse des mailing list, wg items,...)
- accès aux RFC, aux drafts
- dates de réunions IETF

CRÉATION D'UN WORKING GROUP

- BOF (Bird Of a Feather, Birth Of a Feature),
- consensus, trouver une aire
- rédiger une charte

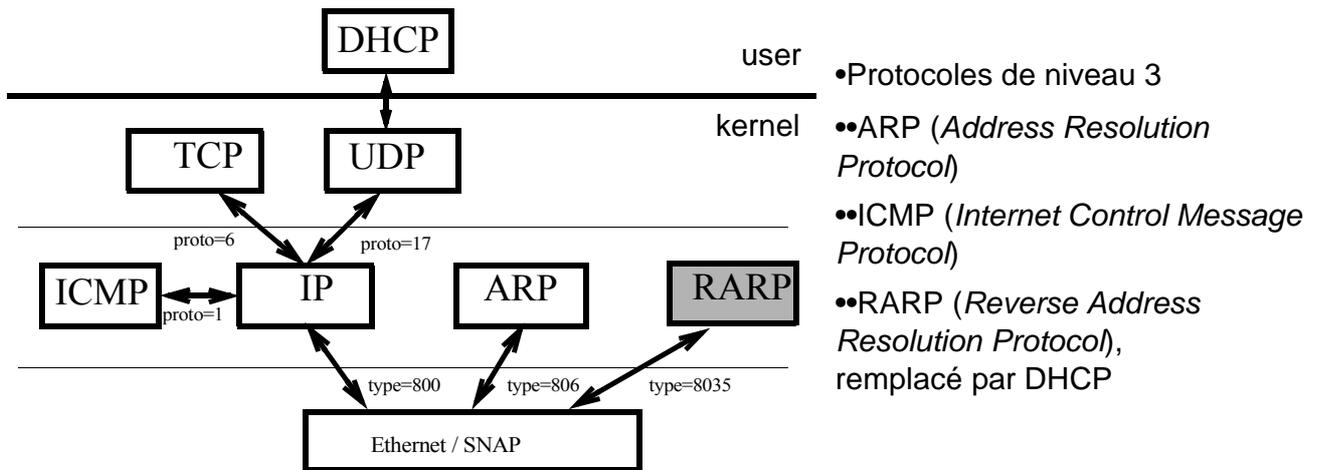
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP.fm

ARCHITECTURE D'UN SYSTÈME UNIX



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP.fm

ARCHITECTURE DE LA COUCHE 3



- Protocoles de niveau 3
- ARP (*Address Resolution Protocol*)
- ICMP (*Internet Control Message Protocol*)
- RARP (*Reverse Address Resolution Protocol*), remplacé par DHCP

- Protocoles de niveau 4
 - UDP (*User Datagram Protocol*)
 - TCP (*Transmission Control Protocol*)

L'ADRESSAGE IP

Propriétés des adresses :

- Unique dans le réseau => IP : sur terre
- Facile à attribuer
 - pour le site demandeur
 - pour l'administration qui gère l'espace d'adresses
- Simplifier la localisation de réseau => scalabilité
 - ce n'est pas le cas des adresses MAC
- Permettre la traduction entre espace d'adressage
 - nom de machine <=> adresse IP
 - adresse IP <=> adresse MAC

AVANT 1994 :

Classe A	0	net_id (7 bits)			host_id (24 bits)									
Classe B	1	0	net_id (14 bits)						host_id (16 bits)					
Classe C	1	1	0	net_id (21 bits)							host_id (8 bits)			
Classe D	1	1	1	0	Adresses Multicast (28 bits)									
Classe E	1	1	1	1	Adresses réservées pour un futur usage (28 bits)									

- Les classes A, B et C désignent des stations
- La classe D définit des adresses de groupe
- La classe E est réservée (non utilisée)

Règles d'allocation :

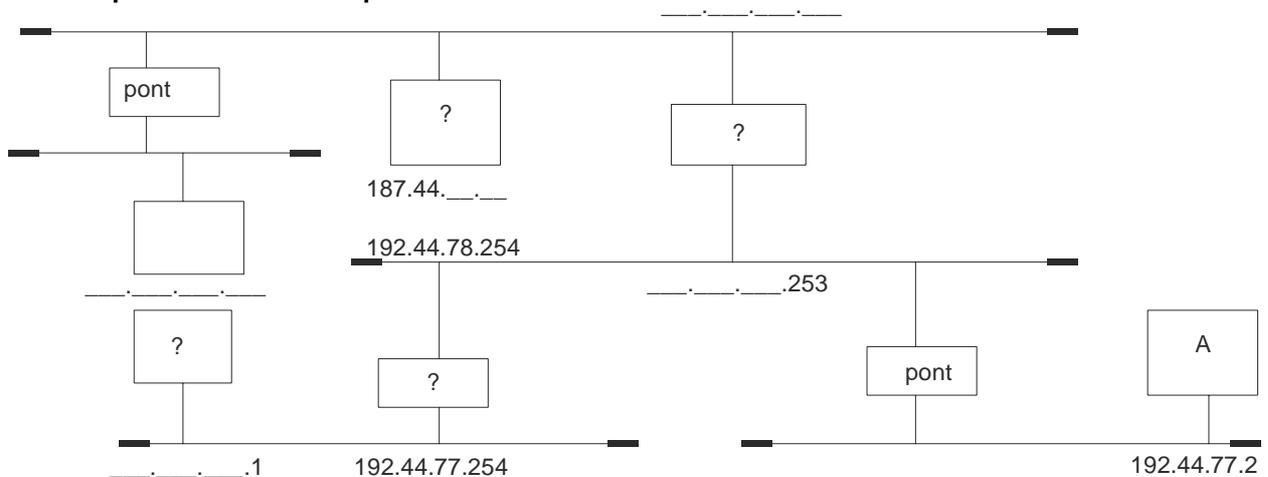
- adressage dit à *plat* (flat address space)
- gestion par le NIC (Network Information Center)

+ Plan d'adressage surtout fait pour faciliter l'administration

PLAN D'ADRESSAGE

+ Un routeur a une adresse IP par réseau.

+ Le pont est transparent.

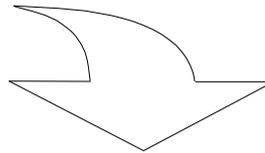


⇒ Est ce que la station A peut avoir cette adresse ?

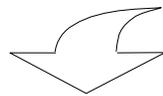
- + Les numéros attribués (classes B ou C) par le NIC désignent un réseau. La partie restante est à administrer par l'ingénieur système. Elle peut être divisée en réseaux et sous-réseaux.

Adresses IP : 192.44.77.79

NetMask : 255.255.255.192

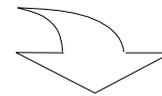


1100 0000.0010 1100.0100 1101.0100 1111
1111 1111.1111 1111.1111 1111.1100 0000



réseau

machine



1100 0000.0010 1100.0100 1101.0100 0000
192.44.77.64

1111
15

- Permet d'indiquer la direction que devront suivre les paquets.
- Contient deux informations :
 - la destination (machine ou réseau IP)
 - le chemin directement accessible (attachement local ou routeur sur le même sous-réseau).

Exemple :

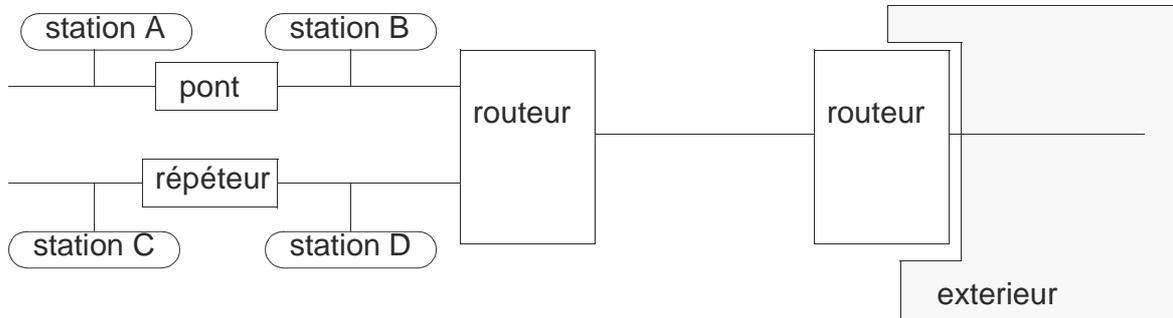
Destination	Gateway
default	mgs-rsm
192.44.77.0	bloodmoney
192.108.119.0	nintendo



Cette table provient de la machine bloodmoney :
 - Proposez une adresse IP pour cette machine,
 - Proposez une table de routage pour les machines nintendo et msg-rsm.

QUESTIONS

Soit le réseau suivant :



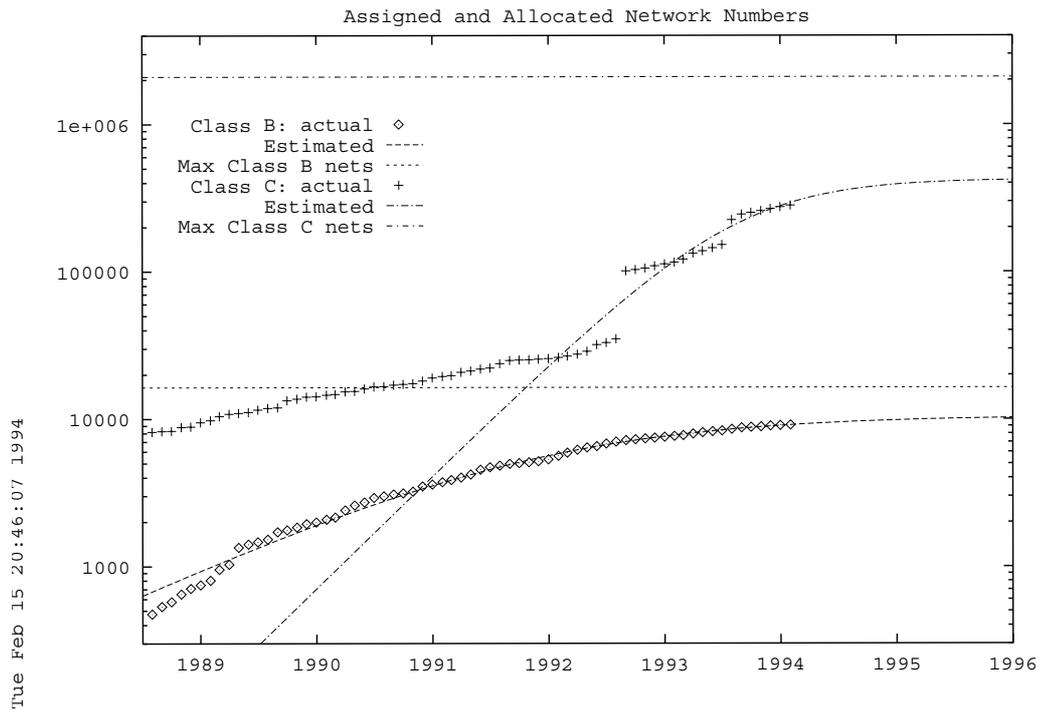
Le NIC vous attribue le numéro de réseau 192.45.67.

⇒ Quelle est la classe de cette adresse ?
 Définir un netmask pour chacun des réseaux IP.
 Donner une adresse pour une station sur chacun des sous-réseaux.
 Combien d'adresses IP ont les routeurs ?

FAITS HISTORIQUES

- 1er Janvier 1983 : Réseau de recherche ~ 100 équipements connectés
- 1992 : Ouverture aux activités commerciales
 - croissance exponentielle
- 1993 : Epuisement des adresses de classe B
 - Prédiction d'un effondrement du réseau pour 1994!
 - l'utilisation des classes C provoque la saturation des tables de routage des routeurs du cœur du réseau.
- Mesures d'urgence :
 - N'accorder qu'exceptionnellement des adresses de classe B
 - Réformer les principes d'allocation des adresses, la notion de classe disparaît
 - CIDR (*Classless Internet Domain Routing*) RFC 1519
 - adresse réseau = préfixe + longueur de préfixe
 - moins de pertes : allocation proche des besoins de la société
 - permet l'agrégation (réduit la taille des tables de routage)

EVOLUTION DU NOMBRE D'ADRESSES IP



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP.fm

PRINCIPE

CIDR peut être vue comme une extension du netmask.

Au lieu de noter les bits indiquant la partie fixe, on indique la longueur de cette partie

Exemple :

- ex classe A : 3.0.0.0 \Rightarrow 3.0.0.0/8 ou 3/8
- ex classe B : 128.93.0.0 \Rightarrow 128.93.0.0/16 ou 128.93/16
- ex classe C : 192.44.77.0 \Rightarrow 192.44.77.0/24 ou 192.44.77/24
- default : 0.0.0.0/0 ou 0/0

Si les bits après le préfixe \neq 0 \Rightarrow adresse d'équipement

Si les bits après le préfixe = 1 \Rightarrow adresse de diffusion

† Une fois le plan d'adressage CIDR mis en place, récupération des anciennes classe A et B pour les allouer suivant ces règles.

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP.fm

EXEMPLE (WWW.IANA.ORG) AOÛT 2001

000/8 IANA - Reserved	Sep 81	028/8 DSI-North	Jul 92	053/8 Cap Debis CCS	Oct 93	196/8 Various Registries	May 93
001/8 IANA - Reserved	Sep 81	029/8 DISA	Jul 91	054/8 Merck and Co., Inc.	Mar 92	197/8 IANA - Reserved	May 93
002/8 IANA - Reserved	Sep 81	030/8 DISA	Jul 91	055/8 Boeing	Apr 95	198/8 Various Registries	May 93
003/8 General Electric	May 94	031/8 IANA - Reserved	Apr 99	056/8 U.S. Postal Service	Jun 94	199/8 ARIN - North America	May 93
004/8 BBN Inc.	Dec 92	032/8 Norsk Infor.	Jun 94	057/8 SITA	May 95	200/8 ARIN - Central and South America	May 93
005/8 IANA - Reserved	Jul 95	033/8 DLA	Jan 91	058/8 IANA - Reserved	Sep 81	201/8 Reserved - Central and South America	May 93
007/8 IANA - Reserved	Apr 95	034/8 Halliburton Company	Mar 93	059/8 IANA - Reserved	Sep 81	202/8 APNIC - Pacific Rim	May 93
008/8 BBN Inc.	Dec 92	035/8 MERIT Network	Apr 94	060/8 IANA - Reserved	Sep 81	203/8 APNIC - Pacific Rim	May 93
009/8 IBM	Aug 92	036/8 IANA - Reserved	Jul 00	061/8 APNIC - Pacific Rim	Apr 97	204/8 ARIN - North America	Mar 94
010/8 IANA - Private Use	Jun 95	(Formerly Stanford University)	Apr 93	062/8 RIPE NCC - Europe	Apr 97	205/8 ARIN - North America	Mar 94
011/8 DoD	May 93	037/8 IANA - Reserved	Apr 95	063/8 ARIN	Apr 97	206/8 ARIN - North America	Apr 95
012/8 AT&T Bell Laboratories	Jun 95	038/8 PSII	Sep 94	064/8 ARIN	Jul 99	207/8 ARIN - North America	Nov 95
013/8 Xerox Corporation	Sep 91	039/8 IANA - Reserved	Apr 95	065/8 ARIN	Jul 00	208/8 ARIN - North America	Apr 96
014/8 IANA - PDN	Jun 91	040/8 Eli Lilly and Company	Jun 94	066/8 ARIN	Jul 00	209/8 ARIN - North America	Jun 96
015/8 Hewlett-Packard Company	Jul 94	041/8 IANA - Reserved	May 95	067/8 ARIN	May 01	210/8 APNIC - Pacific Rim	Jun 96
016/8 DEC	Nov 94	042/8 IANA - Reserved	Jul 95	068/8 ARIN	Jun 01	211/8 APNIC - Pacific Rim	Jun 96
017/8 Apple Computer Inc.	Jul 92	043/8 Japan Inet	Jan 91	069-079/8 IANA - Reserved	Sep 81	212/8 RIPE NCC - Europe	Oct 97
018/8 MIT	Jan 94	044/8 Amateur Radio Digital	Jul 92	080/8 RIPE NCC	Apr 01	213/8 RIPE NCC - Europe	Mar 99
019/8 Ford Motor Company	May 95	045/8 Interop Show Network	Jan 95	081/8 RIPE NCC	Apr 01	214/8 US-DOD	Mar 98
020/8 CSC	Oct 94	046/8 BBN Inc.	Dec 92	082-095/8 IANA - Reserved	Sep 81	215/8 US-DOD	Mar 98
021/8 DDN-RVN	Jul 91	047/8 Bell-Northern Research	Jan 91	096-126/8 IANA - Reserved	Sep 81	216/8 ARIN - North America	Apr 98
022/8 DISA	May 93	048/8 Prudential Securities Inc.	May 95	127/8 IANA - Reserved	Sep 81	217/8 RIPE NCC - Europe	Jun 00
023/8 IANA - Reserved	Jul 95	049/8 Joint Technical Command		128-191/8 Various Registries	May 93	218/8 APNIC - Pacific Rim	Dec 00
024/8 ARIN - Cable Block	May 01	Returned to IANA		192/8 Various Registries - MultiRegional	May 93	219-223/8 IANA - Reserved	Sep 81
025/8 RSRE	Jan 95	050/8 Joint Technical Command		193/8 RIPE NCC - Europe	May 93	224-239/8 IANA - Multicast	Sep 81
026/8 DISA Agency	May 95	Returned to IANA		194/8 RIPE NCC - Europe	May 93	240-255/8 IANA - Reserved	Sep 81
027/8 IANA - Reserved	Apr 95	051/8 DSS UK UK	Aug 94	195/8 RIPE NCC - Europe	May 93		
		052/8 duPont de Nemours, Inc.	Dec 91				

17 janvier 2003 - C:\Documents and Settings\oulain\Mes documents\Support_ENST\IP\m

L'adressage IP

89

ALLOCATION ADMINISTRATIVE

L'IANA alloue des blocs d'adresses à des autorités régionales :

- RIPE (*Réseaux IP Européens*) - NCC (*Network Coordination Center*) :
 - Europe + Afrique supra équatoriale+ Moyen Orient + Russie
- ARIN (*American Registration Internet Number*) :
 - Amériques + Afrique sub équatoriale
- APNIC (*Asia Pacific Network Information Center*)
 - Asie Pacifique
- En création :
 - Afrique
 - Amérique Latine

17 janvier 2003 - C:\Documents and Settings\oulain\Mes documents\Support_ENST\IP\m

L'adressage IP

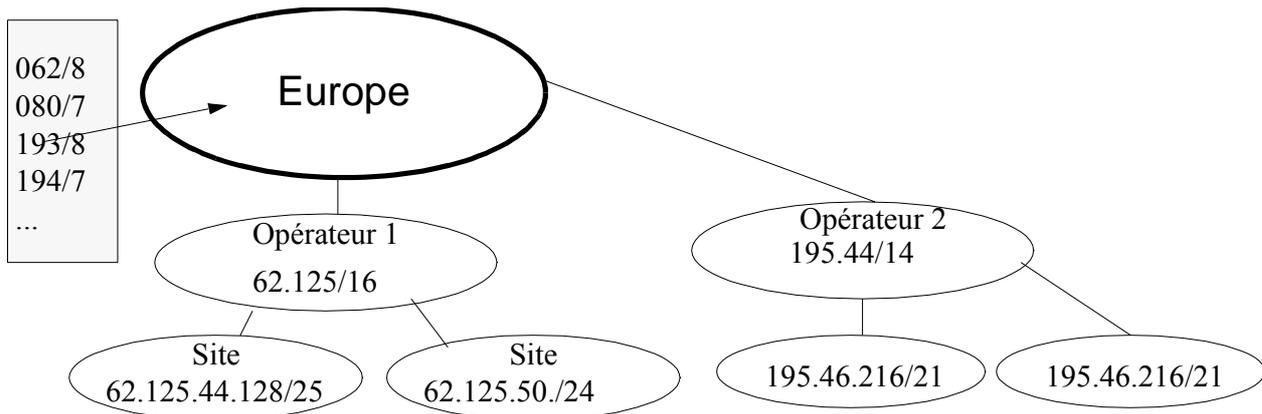
90

DÉLÉGATION AUX OPÉRATEURS

- Chaque autorité régionale délègue les préfixes qui lui ont été alloués.
- Les opérateurs redèlèguent à leur clients (site ou ISP)
- ...

+ CIDR : plan d'adressage hierarchique

+ Administrativement :

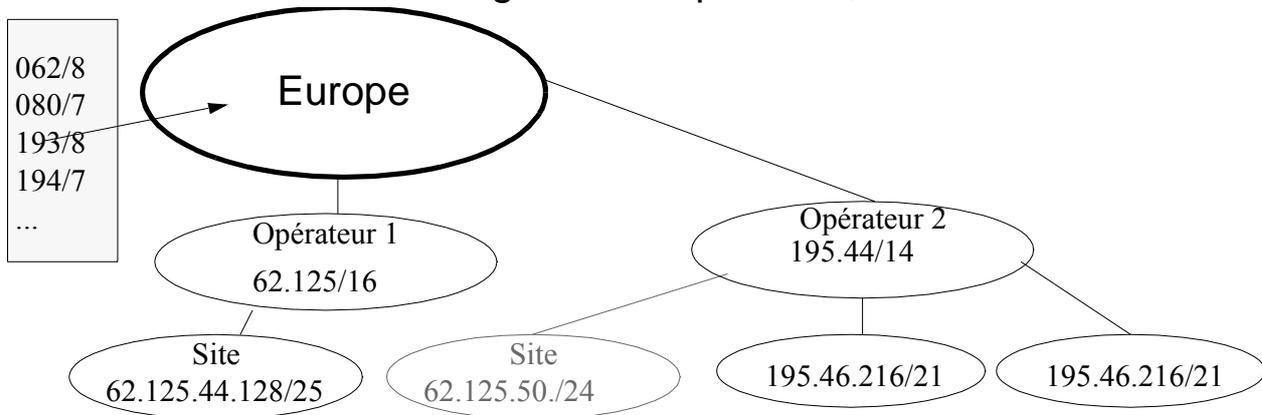


17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP\m

ADRESSAGE CIDR

+ L'adresse appartient à l'opérateur, plus au site,

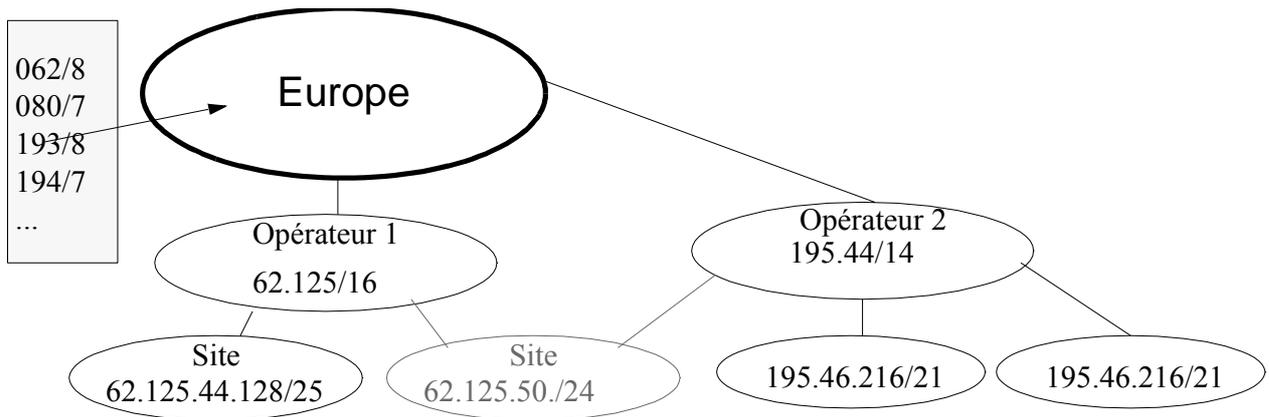
+ Renumérotation si changement d'opérateur,



- Bloque la concurrence

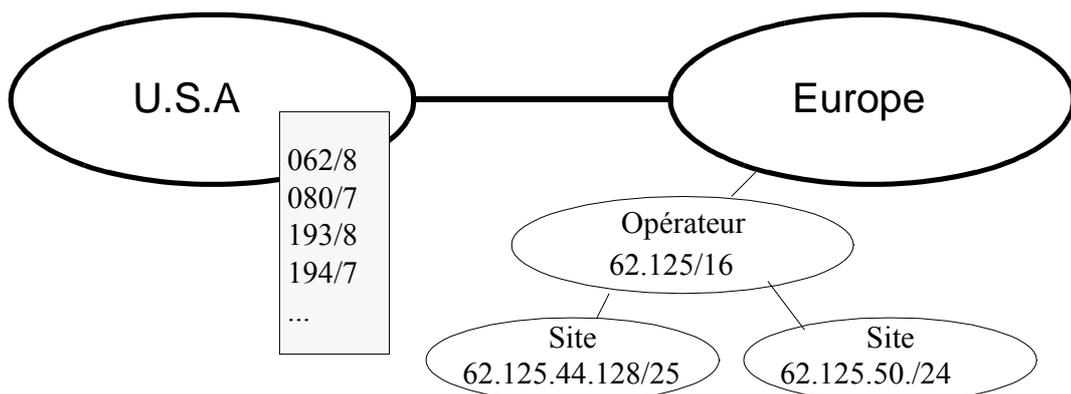
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP\m

MULTIDOMICILIATION (MULTI-HOMING) DIFFICILE



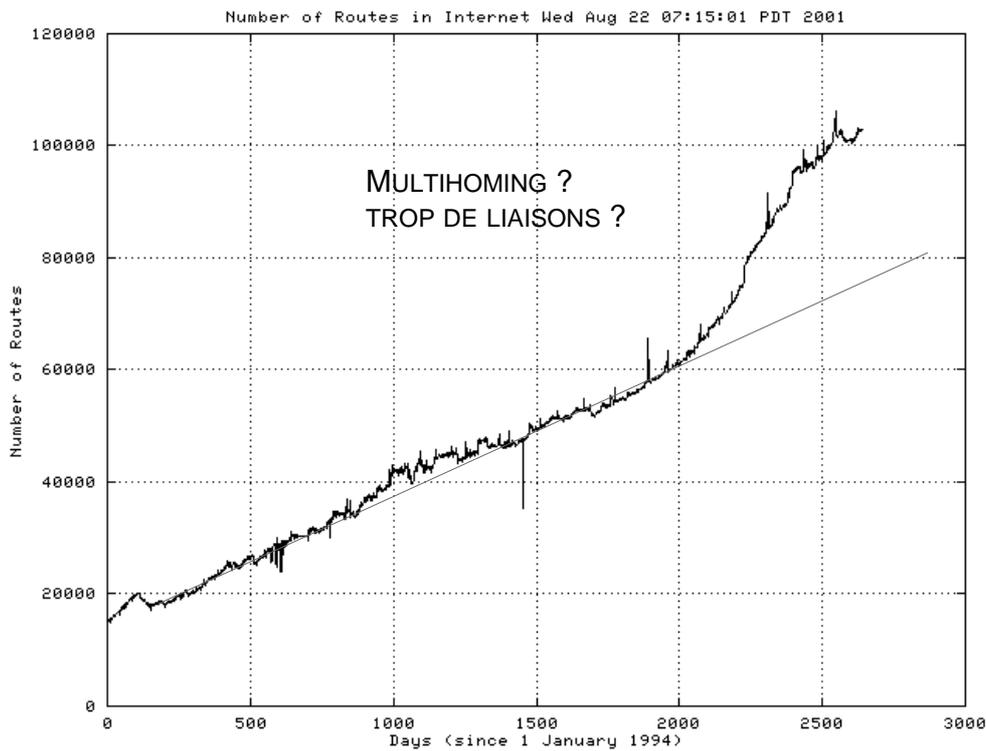
- Service critique => fiabilisation en cas de panne d'un provider

EFFICACE AU NIVEAU MONDIAL ?



- Des US, l'Europe devrait être réduit à quelques préfixes
- Evolution des tables de routage du cœur (core) du réseau.
 - connaissance de tous les préfixes (avec agrégation),
 - pas de route par défaut.

EVOLUTION DES TABLES DE ROUTAGE DU CŒUR DU RÉSEAU INTERNET



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP.fm

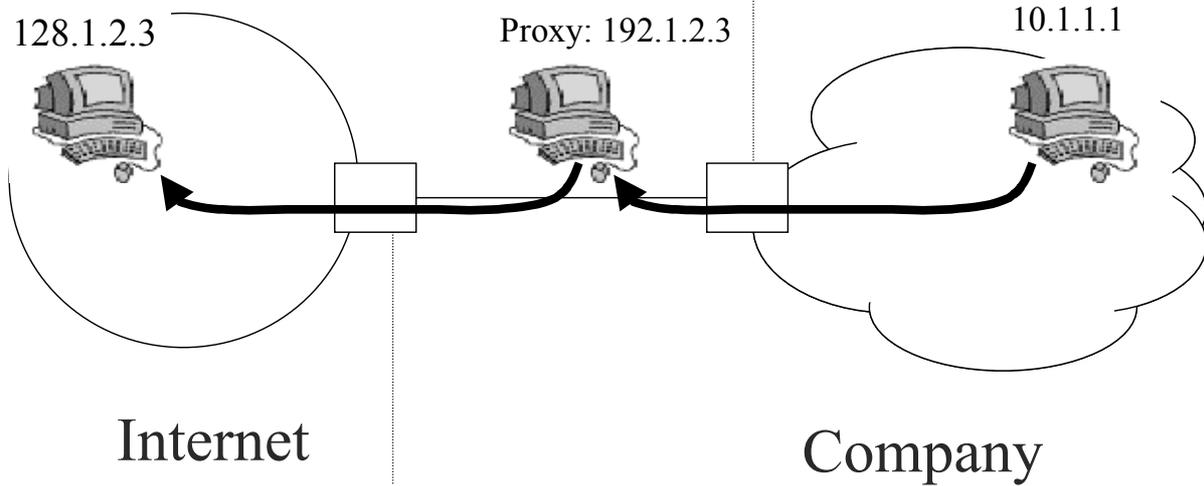
MESURES D'URGENCE

- Permettre les plan d'adressage privés :
 - 10/8
 - 172.16/16
 - 192.168/16
- Ces préfixes ne seront jamais attribuée officiellement à un site, ils peuvent donc être utilisés en interne sans conflits avec des préfixes existants.
- Que faire des ces adresses privées :
 - les utiliser pour des équipements qui ne sont pas accessible de l'extérieur (imprimante,...)
- mettre des mécanismes permettant de sortir (proxy ou NAT) RFC 1631, 2663 et 2993
-
- Similaire aux architectures par la sécurité (Firewall)

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP.fm

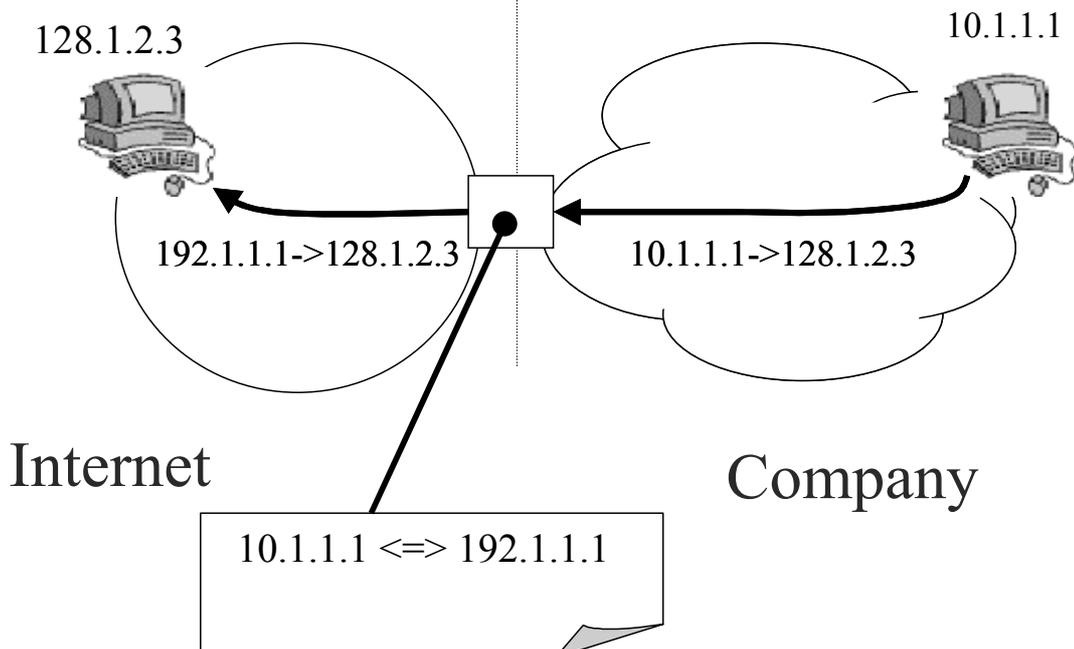
Public address space

Private address space



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP.fm

TRADUCTION D'ADRESSES NAT



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP.fm

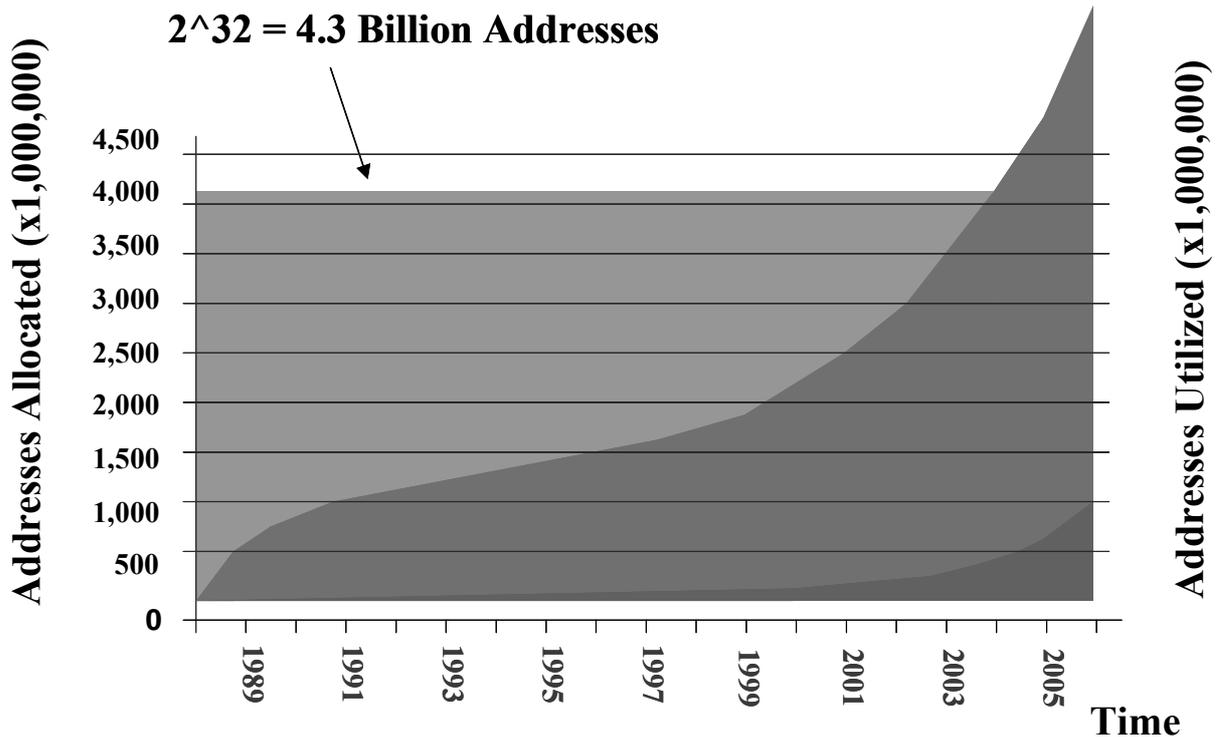
Avanantages :

- Réduit le besoin d'adresses officielles
- Facilite la renumérotation lors d'un changement de fournisseur
- transparent à certaines applications
 - pas de configuration d'un proxy
- Architecture proche de celle utilisée pour la sécurité

Inconvénients :

- Traduction parfois complexe
 - adresses dans les données
- Résiste mal au facteur d'échelle
- Introduit des états dans le réseau
- Casse le modèle de bout-en-bout
- Pas de sécurité si IPsec.

Doit être utilisé dans des petits sites pour le mode Client/Serveur.



LES BASES DE DONNÉES DES RIR

- RIR : Regional Internet Registries
- Maintient une base de données de préfixes attribuées avec les contacts (techniques et administratifs)
- interrogeable avec l'outil `whois`.
- exemple

```
>whois -h whois.ripe.net 193.52.74.0
% This is the RIPE Whois server.
% The objects are in RPSL format.
% Rights restricted by copyright.

inetnum: 193.52.74.0 - 193.52.74.255
netname: FR-OR-ENSTRE
descr: ENST Bretagne - antenne de Rennes
descr: BP 78, 35512 Cesson Sevigne CEDEX,
France
country: FR
admin-c: BA2459-RIPE
tech-c: MF32-RIPE
tech-c: VB2111-RIPE
status: ASSIGNED PA
mnt-by: RENATER-MNT
changed: <Annie.Renard@inria.fr> 19930908
changed: rensvp@renater.fr 20000531
source: RIPE

route: 193.52.74.0/24
descr: FR-OR-ENSTRE
origin: AS2200
mnt-by: RENATER-MNT
changed: RenSVP@Renater.fr 19991008
source: RIPE

person: Bernard AYRAULT
address: ENST Bretagne - antenne de Rennes
address: BP 78
address: 35512 Cesson Sevigne CEDEX
address: FRANCE
phone: +33 2 98 00 11 00
fax-no: +33 2 98 00 10 02
e-mail: Bernard.Ayrault@enst-bretagne.fr
nic-hdl: BA2459-RIPE
changed: rensvp@renater.fr 20000531
source: RIPE

person: Marc FRADIN
address: ENST Bretagne - antenne de Rennes
address: BP 78, F-35512 Cesson Sevigne CE-
DEX, France
phone: +33 2 99 12 70 24
fax-no: +33 2 99 12 70 30
e-mail: Marc.Fradin@enst-bretagne.fr
nic-hdl: MF32-RIPE
mnt-by: RENATER-MNT
changed: Annie.Renard@inria.fr 19940810
changed: rensvp@renater.fr 20000531
source: RIPE

person: Vincent BRARD
address: ENST Bretagne - antenne de Rennes
address: BP 78, F-35512 Cesson Sevigne CE-
DEX, France
phone: +33 2 99 12 70 11
fax-no: +33 2 99 12 70 30
```

Les bases de données des RIR

101

LES CENTRES DE COORDINATION DE L'INTERNET

- IANA (ou ICANN)
- InterNIC (USA) : Monde
- RIPE-NCC : Europe
- APNIC : Asie Pacifique

LES BASES DE DONNÉES DE RÉFÉRENCE

- InterNIC (USA) : rs.internic.net
- ARIN (Amérique du Nord) : whois.arin.net
- RIPE-NCC (Pays Bas) : whois.ripe.net
- APNIC (Japon) : whois.nic.ad.jp

LES AUTRES BASES DE DONNÉES

- Merit (USA) : <http://www.ra.net>
- MCI (USA) : <http://www.mci.net>
- Ca'net (Canada) : <http://www.canet.ca>
- ...

BASES RÉGIONALES

- AFNIC (France) : <http://www.nic.fr>
- Allemagne : <http://www.nic.de>
- ...

LES OBJETS RÉPÉTORIÉS

+ Pour RIPE la définition est donnée dans des documents RIPE-xxx

- Adresse de réseau : appartenance (RIPE 141 et 142)
- Systèmes Autonomes : politiques de routage (RIPE 147, 140 et 181)
- Noms de domaine : serveur de noms, réseaux (RIPE 49)
- Personnes : coordonnées des responsables (RIPE 119)
- Routes : ASN, agrégats (RIPE 181)
- ...

+ Pour d'autres organismes, la structuration est plus libre

LA BASE RIPE

La base RIPE n'est pas directement utilisée pour le routage, mais :

- permet d'avoir une vision centralisée du réseau
- permet de trouver les incohérences dans les règles de routage distribué

La base RIPE est maintenue par les gestionnaires du système autonome.

La base RIPE est gérée par les réseaux qui ont plusieurs attachements à plusieurs systèmes autonomes

Par exemple, RENATER gère les informations des systèmes autonomes qui lui sont connectés.

✦ La base contient la liste des AS avec lesquels des informations de routage sont échangées.

LA BASE RIPE

>whois 192.44.77.0

```
inetnum: 192.44.77.0
netname: FNET-ENSTB-3
descr: ENST Bretagne - antenne de Rennes
descr: BP 78, 35512 Cesson Sevigne CEDEX, France
country: FR
admin-c: Pierre Rolin
tech-c: Marc Fradin
remarks: OR
changed: <Annie.Renard@inria.fr> 930908
source: RIPE
```

```
route: 192.44.77.0/24
descr: FNET-ENSTB-3
origin: AS1717
remarks: OR
mnt-by: AS1717-MNT
changed: ripe-dbm@ripe.net 941121
source: RIPE
```

```
person: Pierre Rolin
address: ENST Bretagne - antenne de Rennes
```



Campus de Rennes

address: BP 78, F-35512 Cesson Sevigne CEDEX, France
phone: +33 99 12 70 21
e-mail: Pierre.Rolin@rennes.enst-bretagne.fr
changed: Annie.Renard@inria.fr 940810
source: RIPE

person: Marc Fradin
address: ENST Bretagne - antenne de Rennes
address: BP 78, F-35512 Cesson Sevigne CEDEX, France
phone: +33 99 12 70 24
e-mail: Marc.Fradin@rennes.enst-bretagne.fr
changed: Annie.Renard@inria.fr 940810
source: RIPE

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP\m



Campus de Rennes

AUX US, LA BASE ARIN

>whois -h whois.arin.net 18.0.0.0

Massachusetts Institute of Technology (NET-MIT-TEMP)
1 Amherst Street Cambridge, MA 02139-1986

Netname: MIT
Netblock: 18.0.0.0 - 18.255.255.255

Coordinator:
Schiller, Jeffrey I (JIS-ARIN) jis@MIT.EDU
+1 617 253-8400 (FAX) +1 617 258-8736

Domain System inverse mapping provided by:

STRAWB.MIT.EDU	18.71.0.151
W20NS.MIT.EDU	18.70.0.160
BITSY.MIT.EDU	18.72.0.3

Record last updated on 26-Sep-98.
Database last updated on 12-Feb-99 16:12:44 EDT.

The ARIN Registration Services Host contains ONLY Internet Network Information: Networks, ASN's, and related POC's. Please use the whois server at rs.internic.net for DOMAIN related Information and nic.mil for NIPRNET Information.

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP\m

VOIR LA DÉLÉGATION DE BLOC CORRESPONDANTE

```

>whois -h whois.ripe.net 193.52.0.0/16
% This is the RIPE Whois server.
% The objects are in RPSL format.
% Rights restricted by copyright.

inetnum: 193.48.0.0 - 193.52.255.255
netname: FR-RENATER-193-48-193-52
descr: DELEGATED BLOCK
descr: Provider Local Registry
descr: RENATER
country: FR
admin-c: ML36-RIPE
tech-c: IM3-RIPE
tech-c: MG21-RIPE
status: ALLOCATED UNSPECIFIED
mnt-by: RIPE-NCC-HM-MNT
changed: marten@ripe.net 19930901
changed: auto-dbm@ripe.net 19950420
changed: auto-dbm@ripe.net 19950421
changed: hostmaster@ripe.net 19990413
changed: hostmaster@ripe.net 19990603
source: RIPE

route: 193.52.0.0/16

descr: RENATER
descr: ENSAM - 151, Boulevard de l'hopital,
descr: 75013 Paris
descr: FRANCE
origin: AS2200
mnt-by: RENATER-MNT
changed: RenSVP@Renater.fr 19991008
source: RIPE

person: Michel LARTAIL
address: GIP RENATER
address: ENSAM
address: 151, Boulevard de l'hopital, 75013 Paris
phone: +33 1 53 94 20 30
fax-no: +33 1 53 94 20 31
e-mail: Michel.Lartail@renater.fr
nic-hdl: ML36-RIPE
changed: rensvp@renater.fr 19980312
source: RIPE

person: Isabelle MOREL
address: Ministère de l'Education Nationale,
address: de la Recherche et de la Technologie

address: 110, rue de Grenelle, 75007 Paris
address: France
phone: +33 1 55 55 15 51
fax-no: +33 1 55 55 39 19
e-mail: isabelle.morel@education.gouv.fr
nic-hdl: IM3-RIPE
mnt-by: RENATER-MNT
changed: rensvp@renater.fr 19980403
changed: rensvp@renater.fr 20000830
source: RIPE

person: Marie-Helene GUILMIN
address: GIP RENATER
address: ENSAM
address: 151, Boulevard de l'hopital, 75013 Paris
address: France
phone: +33 1 53 94 20 38
fax-no: +33 1 53 94 20 31
e-mail: marie-helene.guilmin@renater.fr
nic-hdl: MG21-RIPE
changed: rensvp@renater.fr 19971121
source: RIPE

```

⇒ Quel préfixe pourrait annoncer Renater aux autres opérateurs ?

IPv6

L'adresse passe de 32 bits à 128 bits :

- Entre 1 564 et 3 911 873 538 269 506 102 adresses par m²
 - Justification des adresses de longueur fixe.
- Règles d'attribution identiques à celle d'IPv4 :
 - Utilise les principes de CIDR:
 - Préfixe / longueur de préfixe
- Notation Haxadécimale
 - 3FFE:302:12::/48
 - 3FFE:302:12:2:a00:20ff:fe18:964c/64
- L'agrégation devrait réduire la taille des tables de routage
- Les interfaces ont plusieurs adresses IPv6

ESPACE D'ADRESSAGE

Reserved	0000 0000	1/256	
Unassigned	0000 0001	1/256	
Reserved for NSAP Allocation	0000 001	1/128	
Reserved for IPX Allocation	0000 010	1/128	
Unassigned	0000 011	1/128	
Unassigned	0000 1	1/32	
Unassigned	0001	1/16	
Aggregatable Global Unicast Addresses	001	1/8	[RFC2374]
Unassigned	010	1/8	
Unassigned	011	1/8	
Unassigned	100	1/8	
Unassigned	101	1/8	
Unassigned	110	1/8	
Unassigned	1110	1/16	
Unassigned	1111 0	1/32	
Unassigned	1111 10	1/64	
Unassigned	1111 110	1/128	
Unassigned	1111 1110 0	1/512	
Link-Local Unicast Addresses	1111 1110 10	1/1024	
Site-Local Unicast Addresses	1111 1110 11	1/1024	
Multicast Addresses	1111 1111	1/256	

+ 1/8 de l'espace d'adressage est utilisé par le plan actuel

CONSTRUCTION DES ADRESSES

Types d'adresses :

- Unicast : point à point
- Multicast : idem aux classe D d'IPv4
- Anycast : un équipement (par exemple, le plus proche au sens du routage)

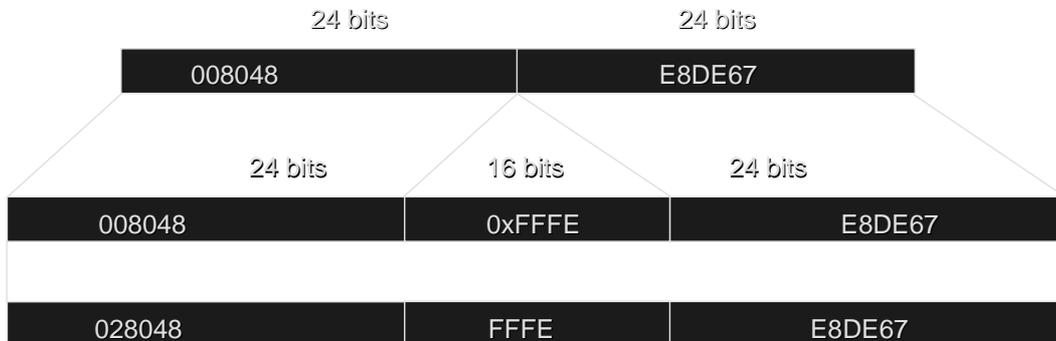
Format d'adresses :

- Loopback : ::1 idem à 127.0.0.1 en IPv4
- Link local : utilisé lors du démarrage, construit en utilisant l'adresse MAC
- Site local : équivalent à une adresse privée pour IPv4
 - utilisé quand le site n'est raccordé à aucun opérateur
 - *pourrait* être utilisé pour les connexions dans un site pour résister aux renumérotations
- Global : lié à un opérateur
 - 6bone : 3FFE...
 - Official : 200x...

CONSTRUCTION D'UNE ADRESSE IPV6

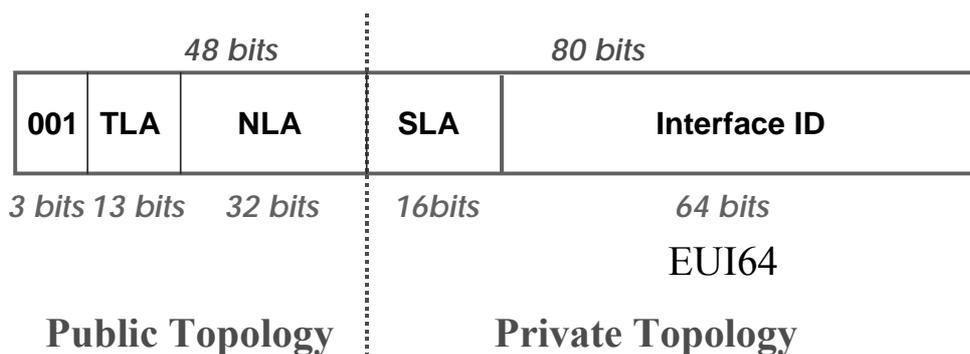
Identifiant d'interface:

- 64 bits en fin d'adresses



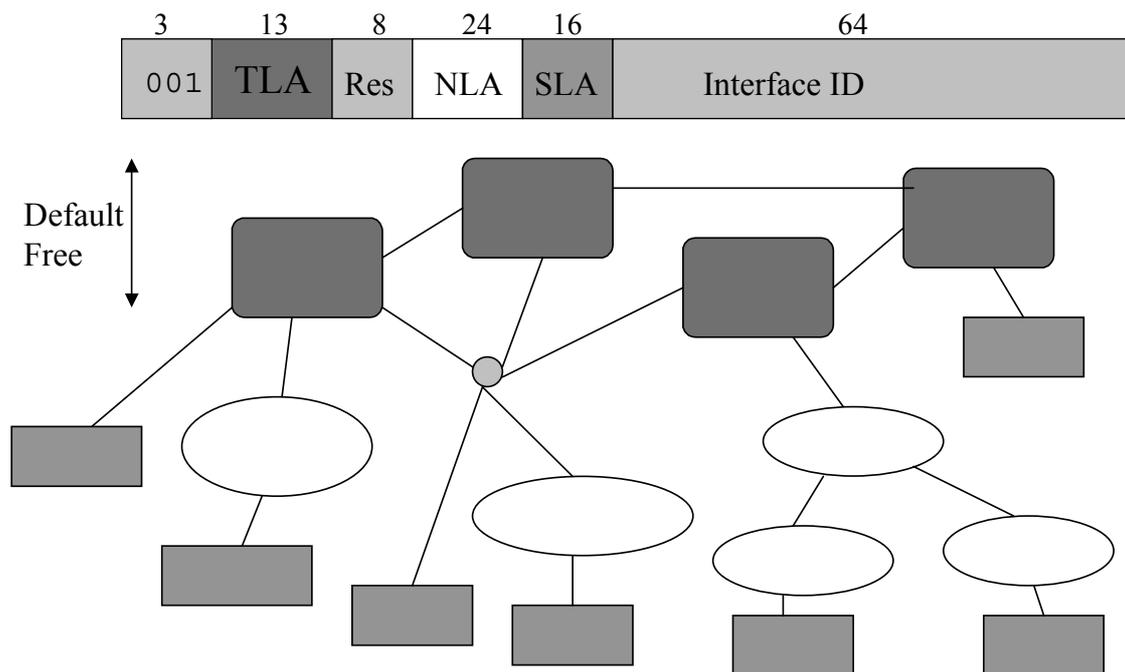
- peut aussi être tiré aléatoirement pour augmenter la confidentialité
- Adresse lien-local : FE80::identifiant d'interface

ADRESSES GLOBALES



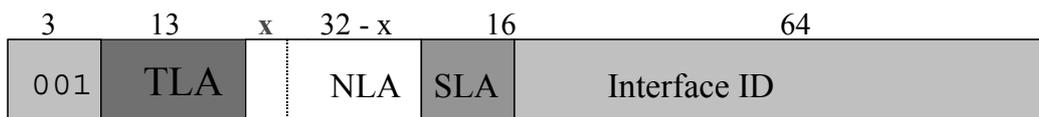
- TLA : Top Level Aggregator => /16
- NLA : Next Level Aggregator => /48
 - peut être récursif
- SLA : Site Level Aggregator => /64
 - 16 bits pour numérotter les réseaux dans la topologie privée

LA THÉORIE : RFC 2374 (AGGREGATABLE GLOBAL UNICAST)



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP\m

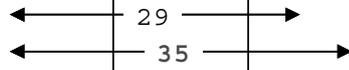
LA PRATIQUE : LE 6BONE



- TLA : 1FFE => 3FFE::/16
- Plan d'adressage géré par le groupe *ngtrans* de l'IETF, accès libre :
 - INNER/US-VA 3FFE:0000::/24
 - TELEBIT/DK 3FFE:0100::/24
 - SICS/SE 3FFE:0200::/24
 - G6/FR 3FFE:0300::/24
 - JOIN/DE 3FFE:0400::/24
 - ...
 - TRUMPET/AU 3FFE:8000::/28
 - ICM-PL/PL 3FFE:8010::/28
 - IJJ/JP 3FFE:8020::/28
 - QTPVSIX/EU 3FFE:8030::/28
 - APAN-KR 3FFE:8040::/28

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP\m

LA PRATIQUE: LES AUTORITÉS RÉGIONALES



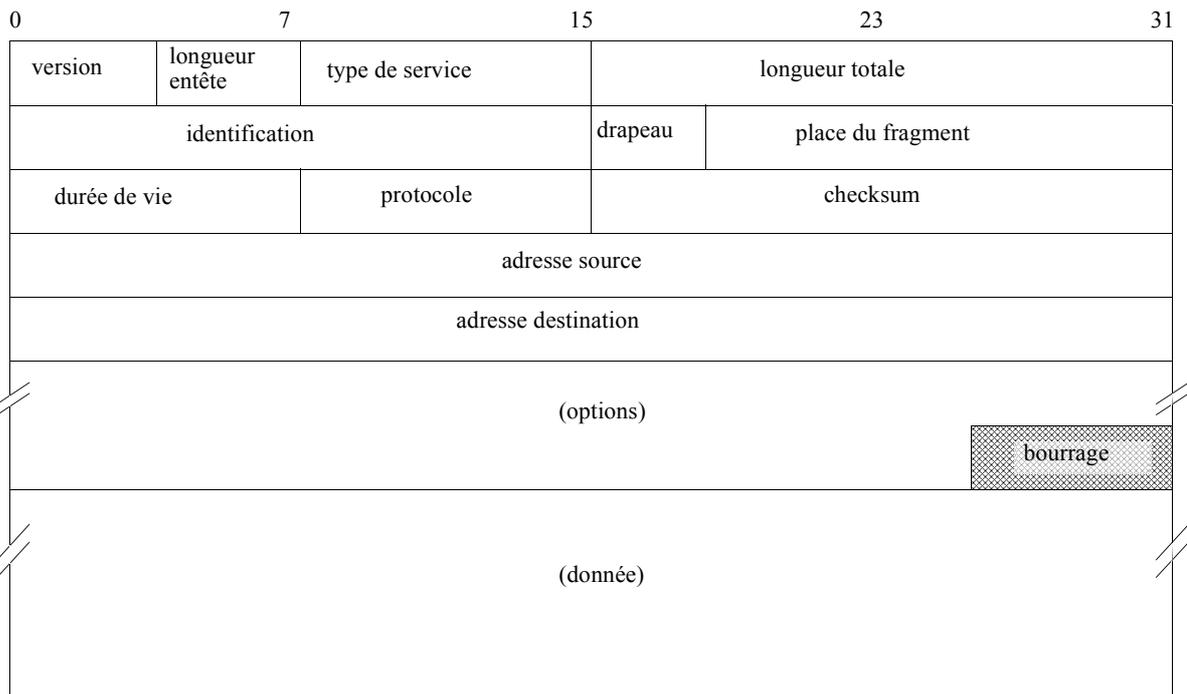
0000 000X XXXX X	2001:0000::/29 - 2001:01F8::/29	IANA
0000 001X XXXX X	2001:0200::/29 - 2001:03F8::/29	APNIC
0000 010X XXXX X	2001:0400::/29 - 2001:05F8::/29	ARIN
0000 011X XXXX X	2001:0600::/29 - 2001:07F8::/29	RIPE-

NCC

0 0000 0000 0000	0x0000	2000::/16 Reserved
0 0000 0000 0001	0x0001	2001::/16 Sub-TLA Assignments
0 0000 0000 0010	0x0002	2002::/16 "6to4"
1 1111 1111 1110	0x1FFE	3FFE::/16 6bone Testing
1 1111 1111 1111	0x1FFF	3FFF::/16 Reserved

17 janvier 2003 - C:\Documents and Settings\louain\mes documents\Support_ENST\IPv6

FORMAT DES PAQUETS IPV4.



17 janvier 2003 - C:\Documents and Settings\louain\mes documents\Support_ENST\IPv4

OPTIONS

- EOL (End Of Option List). Cette option permet de terminer la liste des options.
- NOP (No Operation). Option nulle qui peut être utilisée pour aligner les options sur des mots de 32 bits.
- SEC (Security) . Cette option permet de marquer le niveau de sécurité du datagramme.
- LSR (Loose Source Route).
 - peu performante, les tunnels sont préférés.
 - trou de sécurité: interdit dans les sites
 - utilisé dans le réseau pour contrôler le routage
- RR (Record Route). Si cette option est présente, les routeurs ajoutent leur adresse dans le champ paramètre.
- SSR (Strict Source Route).
- Traceroute => Attaque par denie de service (DoS attack)
- RTRALT (Router Alert)

QUESTIONS

⇒ Poursuivre le désassemblage de la trame IEEE 802.5 :

0 - 10 40 00 00 1D AO CE 51 00 00 1D AO D1 33 AA AA
16 - 03 00 00 00 08 00 45 00 00 82 01 01 40 00 FF 11
32 - 60 OB CO 2C 4D 01 CO 2C 4D 04 00 A1 00 A1 00 6E
48 - 00 00 30 64 02 01 00 04 06 70 75 62 6C 69 63 AO
64 -

⇒ Que vaut le champ option ?

⇒ De quelle classe sont les adresses IP ?

LE PROTOCOLE ARP / CARTE D'IDENTITÉ

- Identifié par le champ type = 0x806 dans la trame Ethernet ou SNAP.
- Permet de faire la correspondance entre une adresse IP et une adresse MAC.
- Les applications ne manipulent que des adresses IP. Le système doit retrouver l'adresse MAC correspondante.
- Les systèmes construisent une table de correspondance.
- Quand un système ne trouve pas l'adresse MAC, il DIFFUSE une requête ARP et attend une réponse.
- La station qui reconnaît dans la requête ARP son adresse IP répond en émettant une réponse ARP avec son adresse MAC.

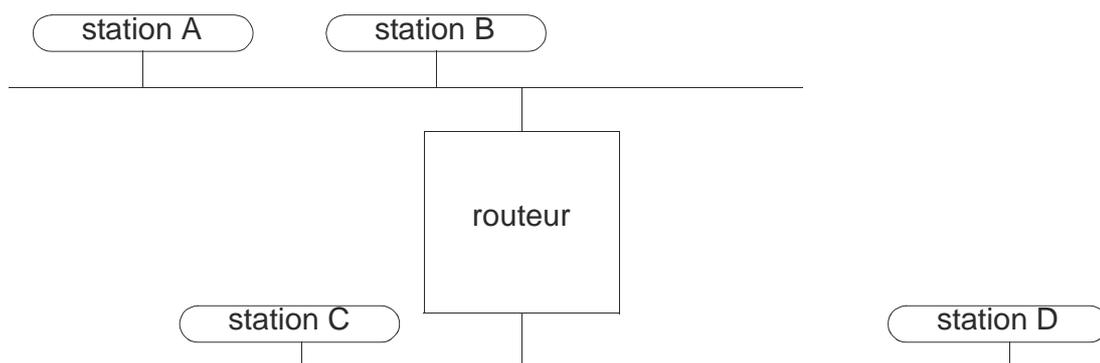
FORMAT D'UNE TRAME ARP

0	7	15	23	31
espace d'adressage physique		espace d'adressage logique		
lg @ physique	lg @ protocole	code		
adresse physique de l'émetteur de la trame...				
adresse physique (suite)		adresse du protocole de ...		
... l'émetteur de la trame		adresse physique du récepteur...		
... de la trame (inconnue)				
adresse du protocole récepteur du paquet				

17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\IP\m

QUESTIONS

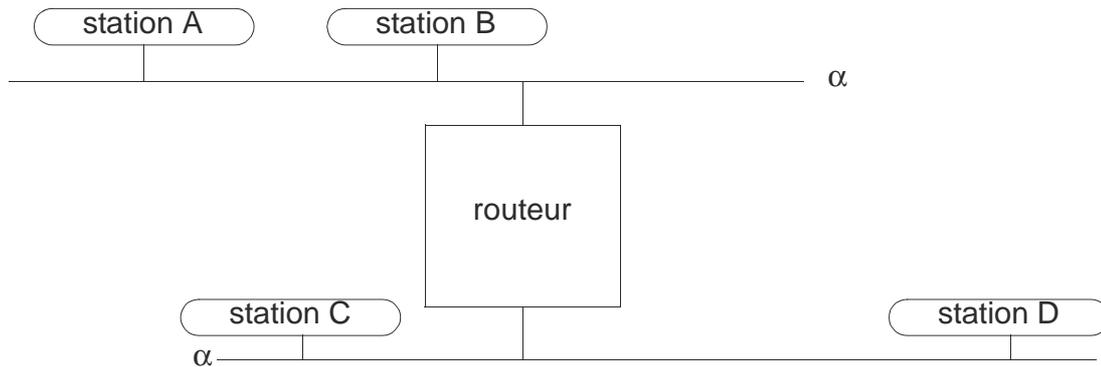
Soit le réseau suivant :



⇒ Que se passe-t-il quand :
A veut émettre vers B ?
A veut émettre vers C ?

17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\IP\m

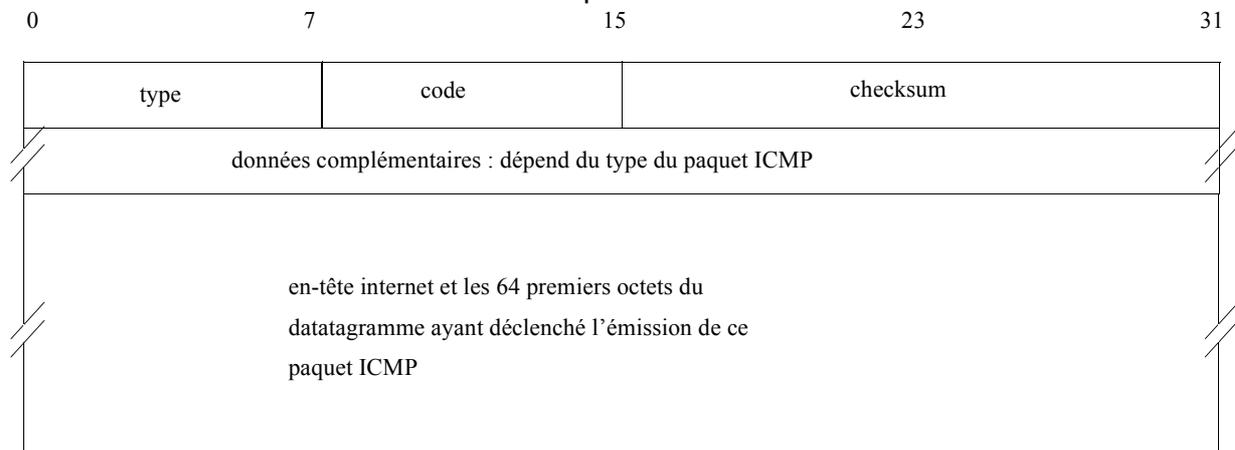
QUESTIONS



- ⇒ Le routeur peut-il répondre aux requetes pour C venat de A ?
- ⇒ Quel intérêt (cas d'une liaison point-à-point par modem)
- ⇒ Quand une interface IP est configurée, elle émet un ARP demandant la résolution de son adresse IP. Pourquoi ?

LE PROTOCOLE ICMP / CARTE D'IDENTITÉ

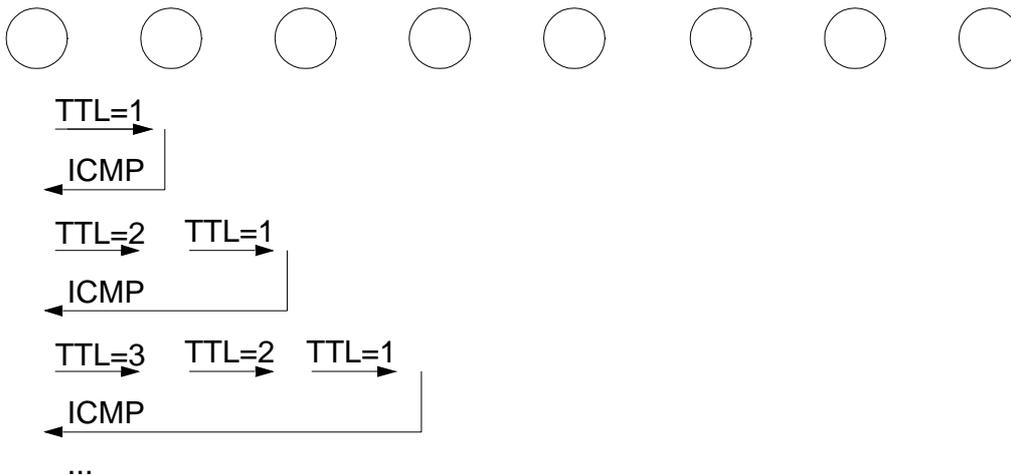
- ICMP (*Internet Control Message Protocol*)
- Encapsulé dans IP (champ proto = 1) mais de niveau 3.
- Sert à contrôler le bon déroulement du protocole IP.



QUELQUES TYPES DE PAQUETS ICMP

- Le message ne peut atteindre sa destination (type = 3)
 - 0 : le réseau ne peut être atteint
 - 1 : la station ne peut être atteinte
 - 2 : le protocole ne peut être atteint
 - 3 : la fragmentation est nécessaire et bit DF=1
 - 4 : port inaccessible
 - 5 : route proposée en option n'est pas valable
- Durée de vie expirée (type = 11)
 - 0. La durée de vie a expiré pendant le transit
 - 1 : La durée de vie a expiré pendant le réassemblage
- Source Quench (type = 4)
- Indication de redirection (type = 5)
- Echo / La commande Ping (type = 8 : demande et type = 0 : réponse)

L'OUTIL TRACEROUTE



- Le message ICMP (TTL expired) permet d'obtenir l'adresse IP du routeur
- Une interrogation du DNS permet de trouver le nom de l'équipement.

EXEMPLE DE ROUTAGE AVEC TRACEROUTE (TRÈS ANCIEN)

```
#traceroute 133.11.11.11
traceroute to 133.11.11.11 (133.11.11.11), 30 hops max, 40 byte packets
 1 mgs-rsm (192.44.77.2) 2 ms 4 ms 2 ms
 2 ft-renater (193.52.72.1) 3 ms 2 ms 2 ms
 3 rennes3.or-br.ft.net (193.48.78.41) 11 ms 7 ms 3 ms
 4 rennes1.or-br.ft.net (193.48.78.25) 4 ms 4 ms 4 ms
 5 192.93.43.210 (192.93.43.210) 12 ms 18 ms 12 ms
 6 stamand2.renater.ft.net (192.93.43.138) 13 ms 12 ms 14 ms
 7 stamand1.renater.ft.net (192.93.43.34) 18 ms 14 ms 12 ms
 8 stamand3.renater.ft.net (192.93.43.17) 17 ms 13 ms 18 ms
 9 rbs1.renater.ft.net (192.93.43.121) 32 ms 21 ms 22 ms
10 (192.121.156.226) 25 ms 28 ms 30 ms
11 icm-dc-2b-S40-1984k.icp.net (192.157.65.129) 182 ms 201 ms 181 ms
12 icm-dc-1-F00.icp.net (144.228.20.101) 193 ms * 299 ms
13 * icm-fix-e-H20-T3.icp.net (192.157.65.122) 139 ms 182 ms
14 * mf0.enss145.t3.ans.net (192.203.229.246) 184 ms *
15 t3-2.crss56.Washington-DC.t3.ans.net (140.222.56.3) 251 ms 179 ms 183 ms
16 t3-1.crss72.Greensboro.t3.ans.net (140.222.72.2) 199 ms 197 ms 192 ms
17 t3-0.crss104.Atlanta.t3.ans.net (140.222.104.1) 203 ms 203 ms 187 ms
18 t3-2.crss64.Houston.t3.ans.net (140.222.64.3) 206 ms * 204 ms
19 t3-0.crss112.Albuquerque.t3.ans.net (140.222.112.1) 269 ms 238 ms *
20 t3-1.crss16.Los-Angeles.t3.ans.net (140.222.16.2) 286 ms * 280 ms
21 t3-2.crss8.San-Francisco.t3.ans.net (140.222.8.3) 289 ms * 278 ms
```

17 janvier 2003 - C:\Documents and Settings\ouain\Mes documents\Support_ENST\IP\m

```
22 * * t3-0.enss144.t3.ans.net (140.222.144.1) 256 ms
23 192.203.230.6 (192.203.230.6) 322 ms 300 ms 294 ms
24 132.160.252.2 (132.160.252.2) 389 ms 405 ms *
25 * 133.11.208.101 (133.11.208.101) 388 ms *
26 133.11.210.2 (133.11.210.2) 422 ms 431 ms 429 ms
27 utsun.su-tokyo.ac.jp (133.11.11.11) 411 ms 479 ms 433 ms
```

17 janvier 2003 - C:\Documents and Settings\ouain\Mes documents\Support_ENST\IP\m

⇒ Est ce que les datagramme IP peuvent prendre des chemins différents?

⇒ Que peut-on déduire du routage dans l'Internet?

OUTILS : TRACEROUTE

```
>./traceroute rki.kbs.co.kr
traceroute to rki.kbs.co.kr (210.115.193.23): 1-30 hops, 38 byte packets
 1 mgs-in.rennes.enst-bretagne.fr (193.52.74.2) 3.11 ms 2.74 ms 2.31 ms
 2 ft-renater.rennes.enst-bretagne.fr (193.51.128.81) 3.21 ms 3.16 ms 3.2 ms
 3 193.48.78.41 (193.48.78.41) 4.21 ms 93.6 ms 38.2 ms
 4 rennes.or-br.ft.net (193.48.78.25) 4.47 ms 4.50 ms 4.60 ms
 5 rennes.renater.ft.net (193.55.253.170) 4.77 ms 9.50 ms 4.90 ms
 6 stamand2.renater.ft.net (195.220.180.153) 10.3 ms 12.9 ms 10.6 ms
 7 rbs2.renater.ft.net (195.220.180.34) 10.3 ms 10.4 ms 10.3 ms
 8 paii.renater.ft.net (195.220.180.29) 13.1 ms 12.1 ms 12.1 ms
 9 relay-pos-6.opentransit.net (193.55.152.70) 87.1 ms 87.9 ms 86.9 ms
10 sl-bb11-rly-0-1.sprintlink.net (144.232.8.209) 127 ms 96.3 ms 88.0 ms
11 sl-bb2-dc-4-0-0.sprintlink.net (144.232.7.142) 101 ms 90.2 ms 88.5 ms
12 core7-hssi0-0-0.Washington.cw.net (206.157.77.33) 89.9 ms 89.3 ms 93.2 ms
13 bordercore2.SanFrancisco.cw.net (166.48.14.1) 160 ms (ttl=240!) 172 ms (ttl=240!) 159 ms (ttl=240!)
14 dacom.SanFrancisco.cw.net (166.48.15.246) 192 ms (ttl=239!) 164 ms (ttl=239!) 161 ms (ttl=239!)
15 gateway.bora.net (203.233.35.249) 298 ms (ttl=238!) 300 ms (ttl=238!) 321 ms (ttl=238!)
16 210.120.128.4 (210.120.128.4) 312 ms (ttl=237!) 311 ms (ttl=237!) 324 ms
17 203.233.37.146 (203.233.37.146) 317 ms (ttl=236!) 336 ms (ttl=236!) 324 ms (ttl=236!)
```

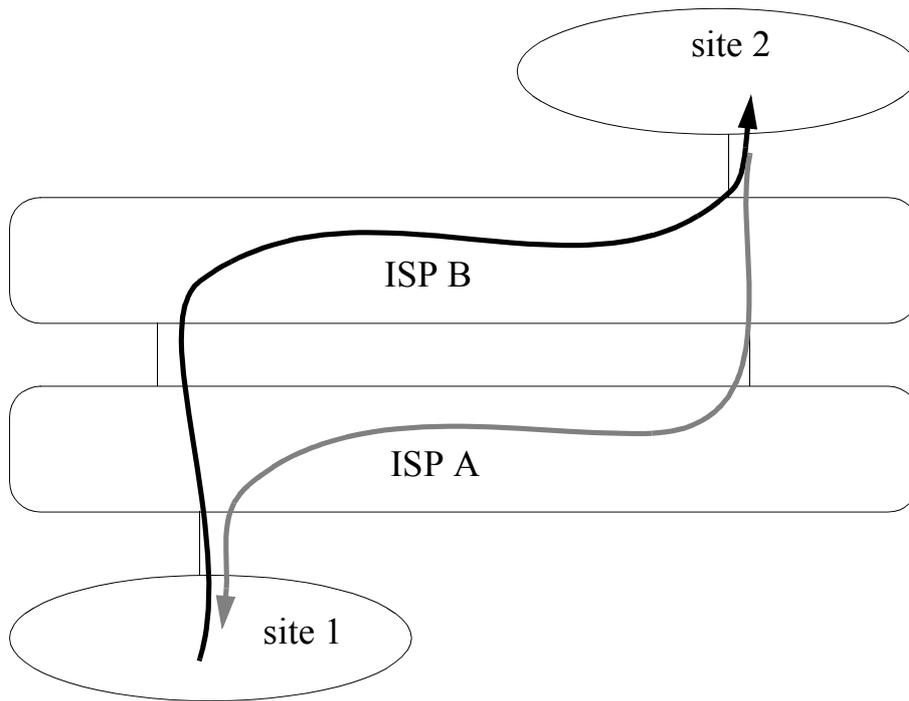
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\FI\m

```
18 rki.kbs.co.kr (210.115.193.23) 346 ms (ttl=108!) 375 ms (ttl=108!) 329 ms (ttl=108!)
```

⇒ Expliquer les valeurs entre parenthèses

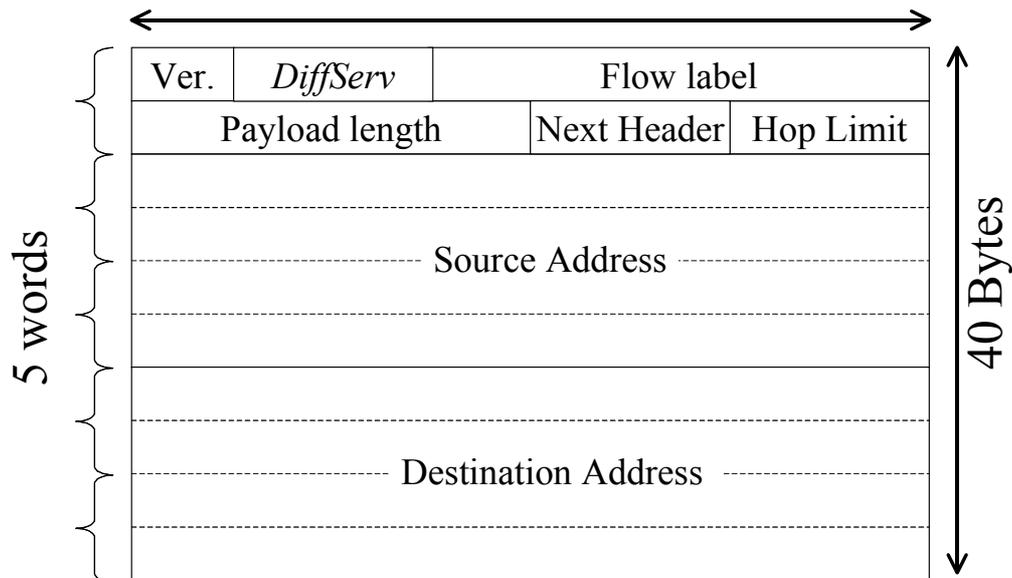
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\FI\m

PATATE CHAUDE



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP\m

FORMAT DES PAQUETS IPV6



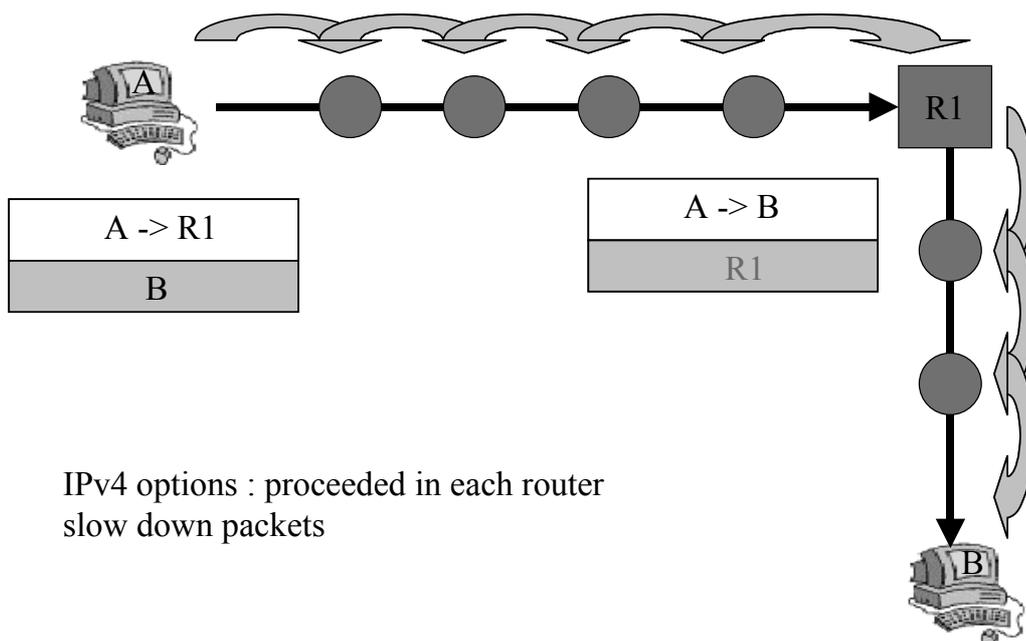
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP\m

+ Simplification de l'en-tête => fondamentaux du routage

- Suppression:
 - checksum : recalcul après chaque changement du champ TTL
 - options : remplacées par des extentions
 - fragmentation : idem + généralisation MTU discovery
- Limites ?
 - taille des paquets : 64 Ko par défaut, sinon extention jumbogramme
 - Hop Limit : semble être constant

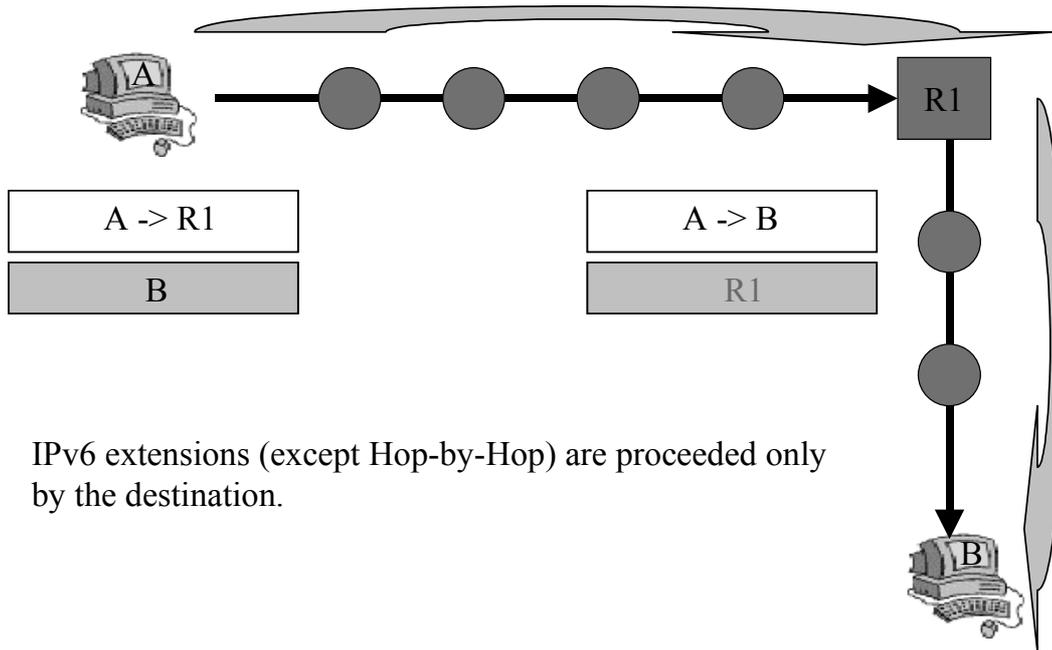
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP\im

EXTENTIONS: DIFFÉRENCE AVEC LES OPTIONS



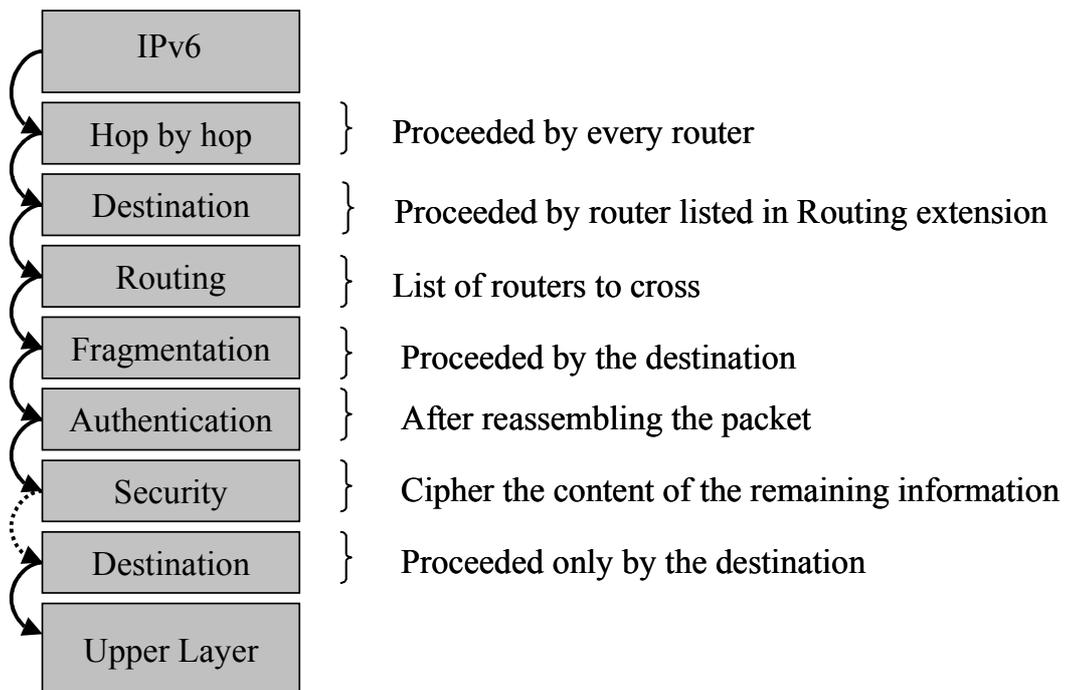
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP\im

EXTENSION: AVEC IPV6



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP\fm

L'ORDRE EST IMPORTANT

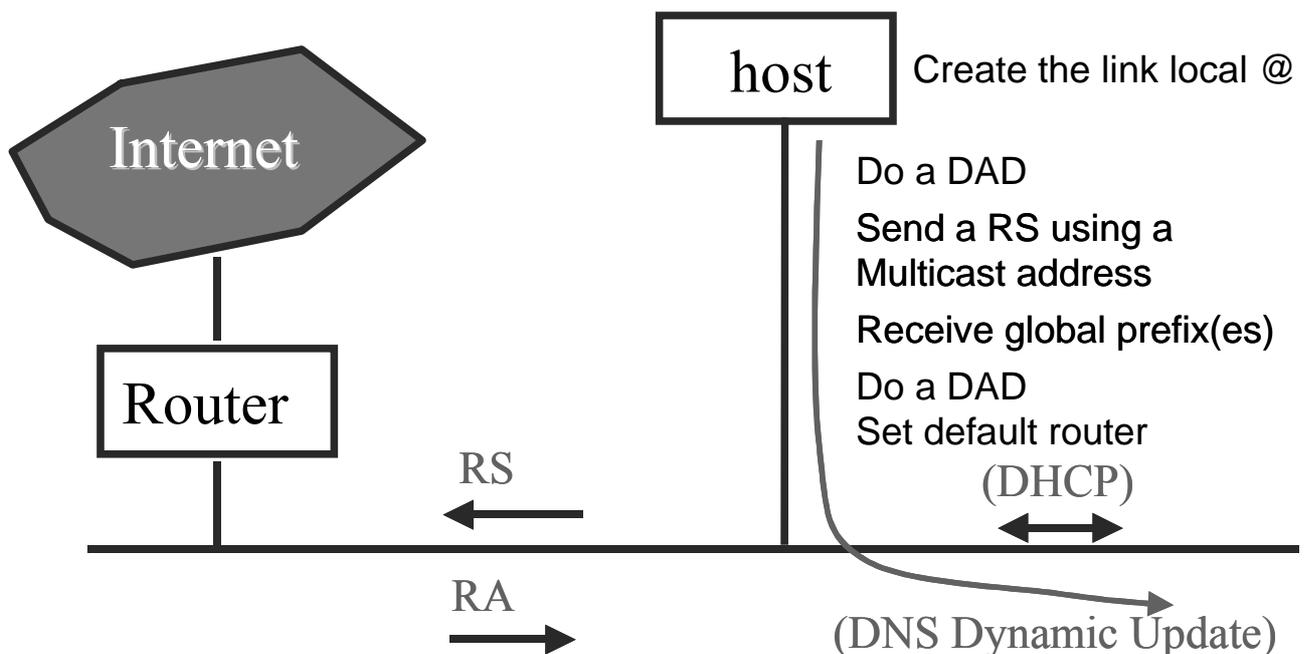


17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP\fm

NEIGHBOR DISCOVERY

- Synthèse des fonctions d'ARP et d'ICMP
- Utilisation de messages ICMPv6 : Neighbor Discovery
 - Utilisation de l'adresse lien-local
 - Utilisation d'adresses de multicast
 - Suppression du broadcast => meilleure scalabilité (surtout si le niveau 2 gère les groupes multicast : IEEE 802.1 p/Q).
- Neighbor Discovery sert à :
 - autoconfiguration
 - localiser les équipements de routage (routeur par défaut) (équivalent d'ICMP routeur discovery)
 - résolution d'adresses (équivalent d'ARP pour IPv4)
 - détecter les adresses dupliquées (équivalent de l'ARP gratuit en IPv4)
 - maintenir l'information d'accessibilité vers les voisins (NUD)
 - est utilisable pour les réseaux NBMA (*Non Broadcast Multi Access networks*)

AUTOCONFIGURATION



AUTOCONFIGURATION

- Seuls les routeurs continuent à être configurés manuellement
- Les équipements peuvent obtenir dynamiquement une adresse IPv6 :
 - La configuration sans état est très utilisée
 - La configuration avec état DHCPv6 est controversée (uniquement pour les paramètres supplémentaire (DNS, imprimante,...))
- Les équipements doivent pouvoir s'enregistrer automatiquement dans le DNS
 - Changer une carte Ethernet d'une machine la fait "disparaître" du réseau
 - Dynamic Update pour IPv6
 - Problème de sécurité ?

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP\m

LES PROTOCOLES DE NIVEAU 4

Notion de port

- Présent avec UDP et TCP
- Les adresses IP permettent de désigner une machine unique dans le monde
- Le port permet de désigner un programme sur cette machine
- Mode client/serveur
 - Serveur : attends des requêtes sur un port bien déterminé
 - Client : envoie des requêtes vers un serveur, le n° de port peut être dynamique
- Pour des raisons de sécurité "authentifier" les serveurs :
 - 0..1023 : en mode protégé
 - 1024...65535 : attribuable à n'importe quelle application.

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP\m

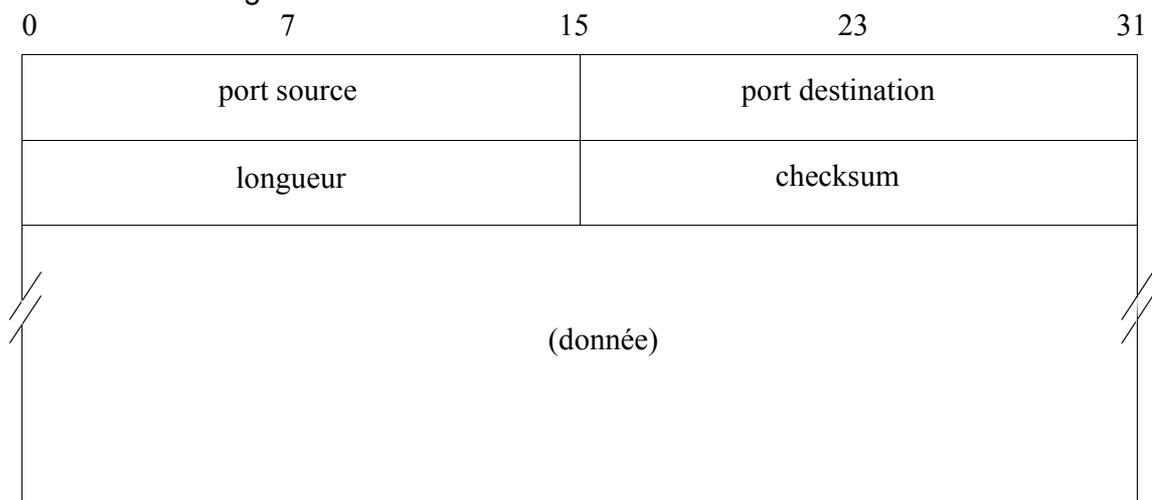
LE PROTOCOLE UDP / CARTE D'IDENTITÉ

- Identifié par le champ proto = 17 dans le paquet IP
- Protocole en mode datagramme :
 - Aucune sécurité de bout-en-bout
 - Les adresses de classe D sont utilisables uniquement avec ce protocole
- Aucun traitement protocolaire:
 - aucun contrôle d'erreur, de séquençement
 - aucun contrôle de flux.
- Utilisé uniquement pour contenir l'information sur les numéros de port
- Utilisé par :
 - flux multimédia (pour éviter les contrôles et autoriser le multicast)
 - les requêtes DNS

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP\m

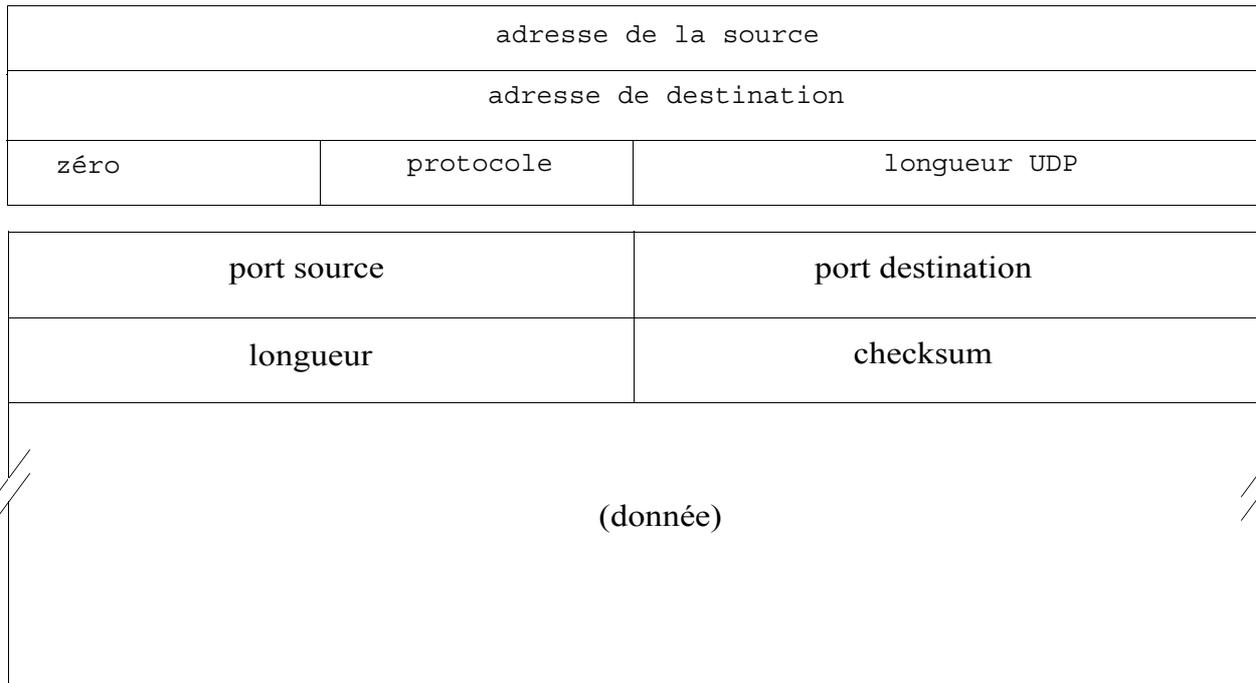
UDP

- Format du message :



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\IP\m

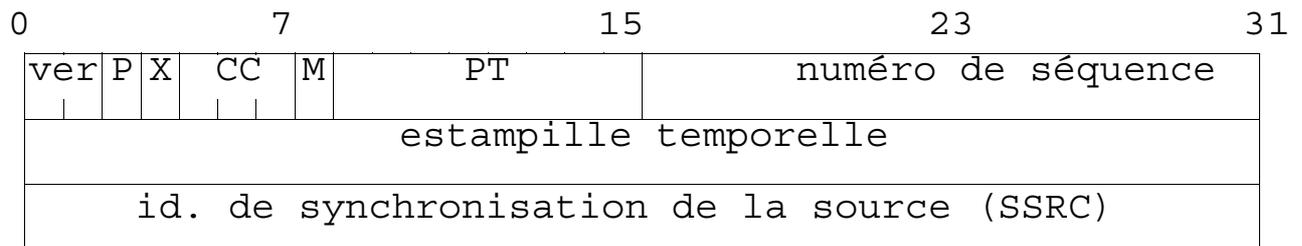
PSEUDO-EN-TÊTE



PSEUDO EN-ÊTE

- Violation du modèle OSI :
 - des données de niveau 3 servent à calculer un champ de niveau 4
 - IP et UDP (ou TCP) sont mis en œuvre au même niveau dans le système d'exploitation,
 - "niveau 4" : remplit certains champs du paquet IP (@s, @d, protocole, longueur,...) et tous ceux de niveau 4.
 - les autres champs sont mis à 0
 - pseudo en-tête = propriété de l'algorithme de calcul : l'addition est commutative
- Dans IPv6, plus de checksum au niveau 3 :
 - la détection des erreurs est reportée sur le pseudo en-tête,
 - détection uniquement par le destinataire,
 - simplification des équipements d'interconnexion.

RTP



- version = 2;
- P = 1 bits de bourrage (Padding), le dernier octet du paquet contient la taille du bourrage ;
- X = 1 : l'en-tête contient d'autres champs que ceux obligatoires ;
- CC = nombre de champ CSRC de l'en-tête ;
- M = 0 : fin d'une information séquentée au niveau RTP;
- PT = nature de l'information multimédia

RTP = APPLICATION LAYER FRAMING

- L'application choisi la manière d'encapsuler les données
 - Permet de mieux respecter la sémantique des infomations multimédia
 - Contrôle d'erreur aussi à la charge de l'application
- Le réseau ignore tout de l'encapsulation RTP
 - convention entre des applications
 - problèmes pour la compression des en-têtes

RTCP

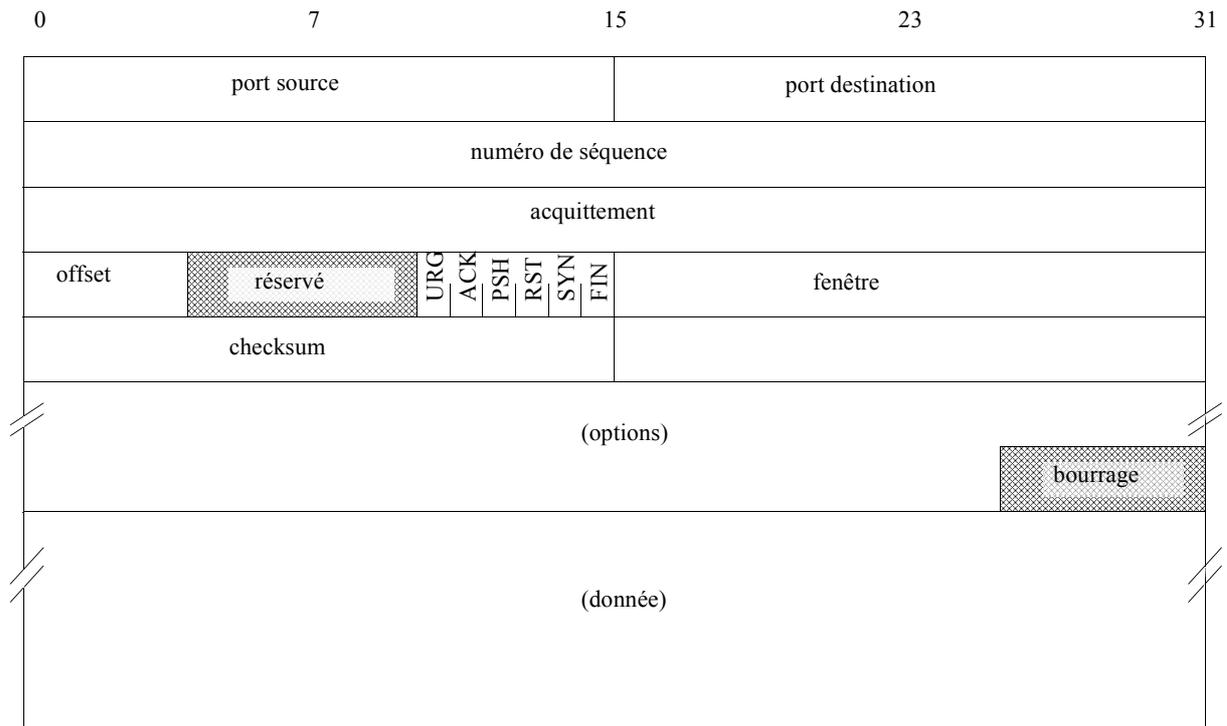
- Mécanisme de contrôle des paquets échangés entre une source et les récepteurs (multicast ou unicast) :
 - Donne des statistiques sur le taux d'erreur de la liaison : perte de paquets, temps de traversé,...
 - peut être utilisé pour un contrôle de flux
 - Manque de réactivité => a long terme:
 - adapter le codage au débit du client

LE PROTOCOLE TCP

Carte d'identité :

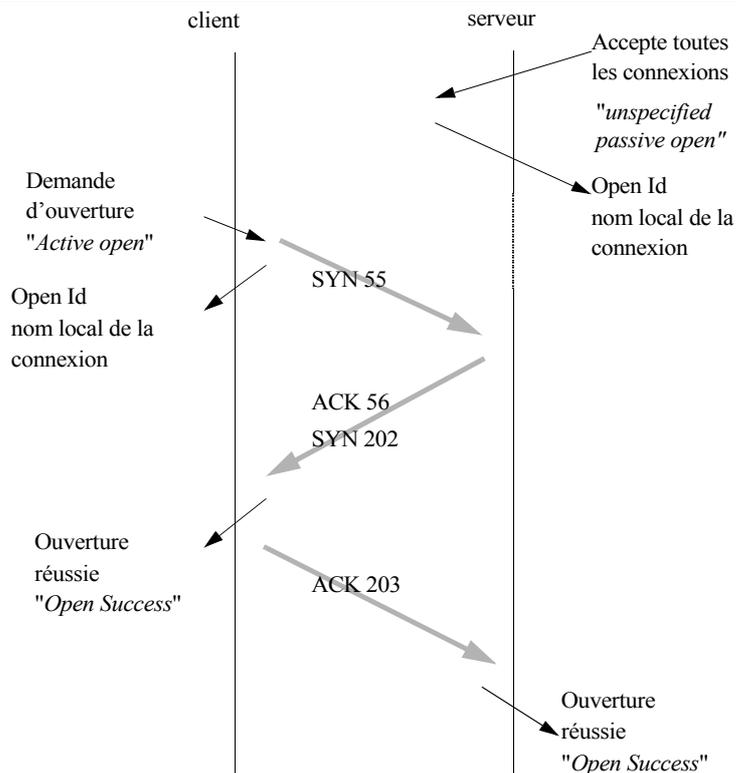
- Identifié par le champ proto = 6 dans le paquet IP
- Protocole en mode connecté :
 - Contrôle de séquençement, de flux
 - Possibilité de données express
 - Uniquement en point-à-point
- Protocole complexe
- Trois phases :
 - Ouverture de la connexion
 - Transfert de données
 - Fermeture de la connexion
- Numérotation des octets transmis

FORMAT DES MESSAGES



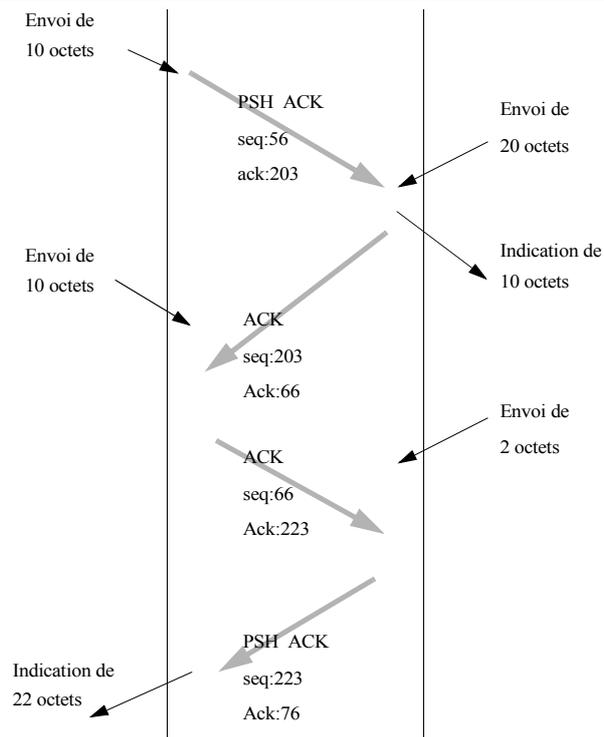
17 janvier 2003 - C:\Documents and Settings\outain\Mes documents\Support_ENST\IP.fm

OUVERTURE DE LA CONNEXION

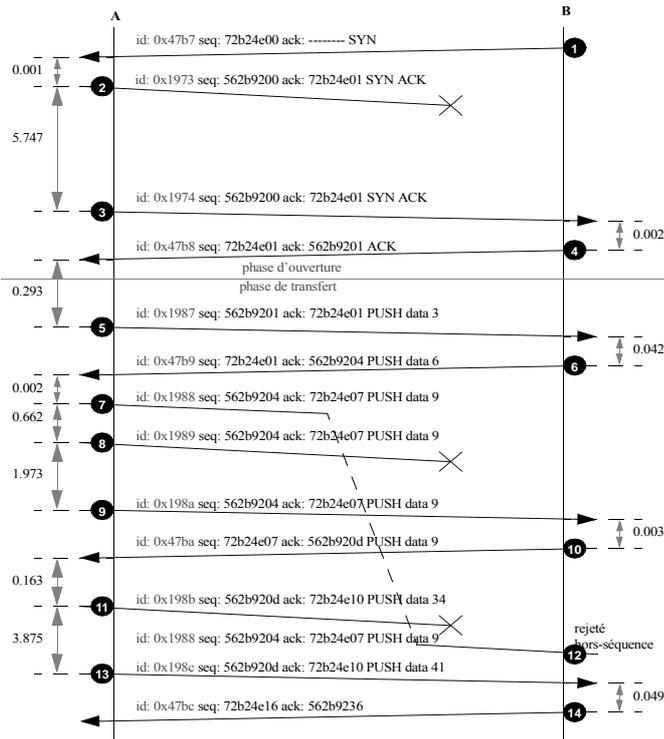


17 janvier 2003 - C:\Documents and Settings\outain\Mes documents\Support_ENST\IP.fm

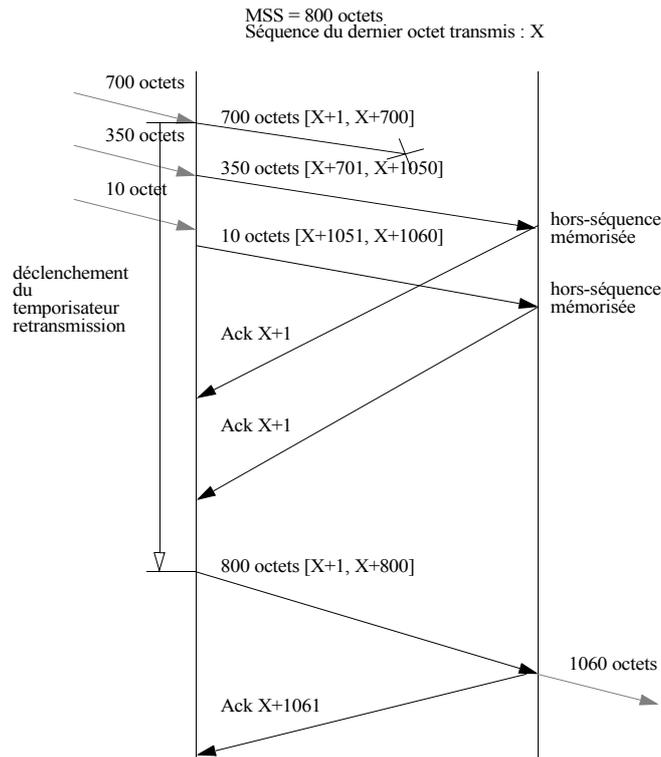
TRANSFERT DE DONNÉES



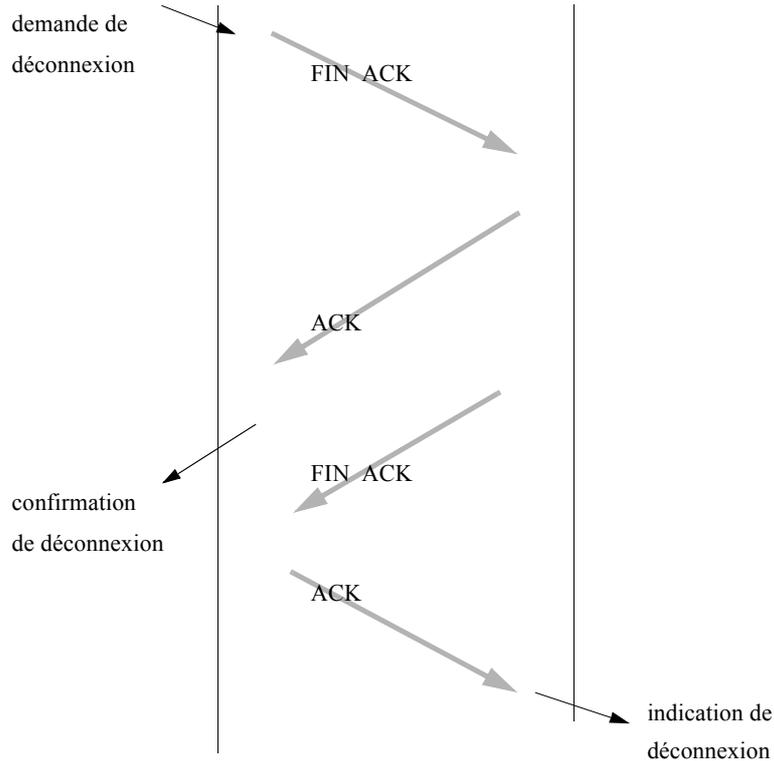
TRANSFERT AVEC ERREURS



TRANSFERT AVEC ERREUR



FERMETURE DE CONNEXION



156

ADAPTATION À L'ENVIRONNEMENT

- Gestion dynamique des valeurs temporisation
- RTT : *Round Trip Time*. Temps mesuré pour un aller et retour.
- L'estimation du RTT va déterminer le temporisateur de *Retransmission Time Out* (RTO).
- Ne pas prendre en compte le RTT lors de retransmissions (algorithme de Karn)

On a :

$$\begin{aligned}
 \text{erreur} &= \text{mesure} - \text{moyenne} \\
 \text{moyenne} &= \text{moyenne} - \alpha \times \text{erreur} \\
 \text{deviation} &= \text{deviation} + \beta \times (|\text{erreur}| - \text{deviation})
 \end{aligned}$$

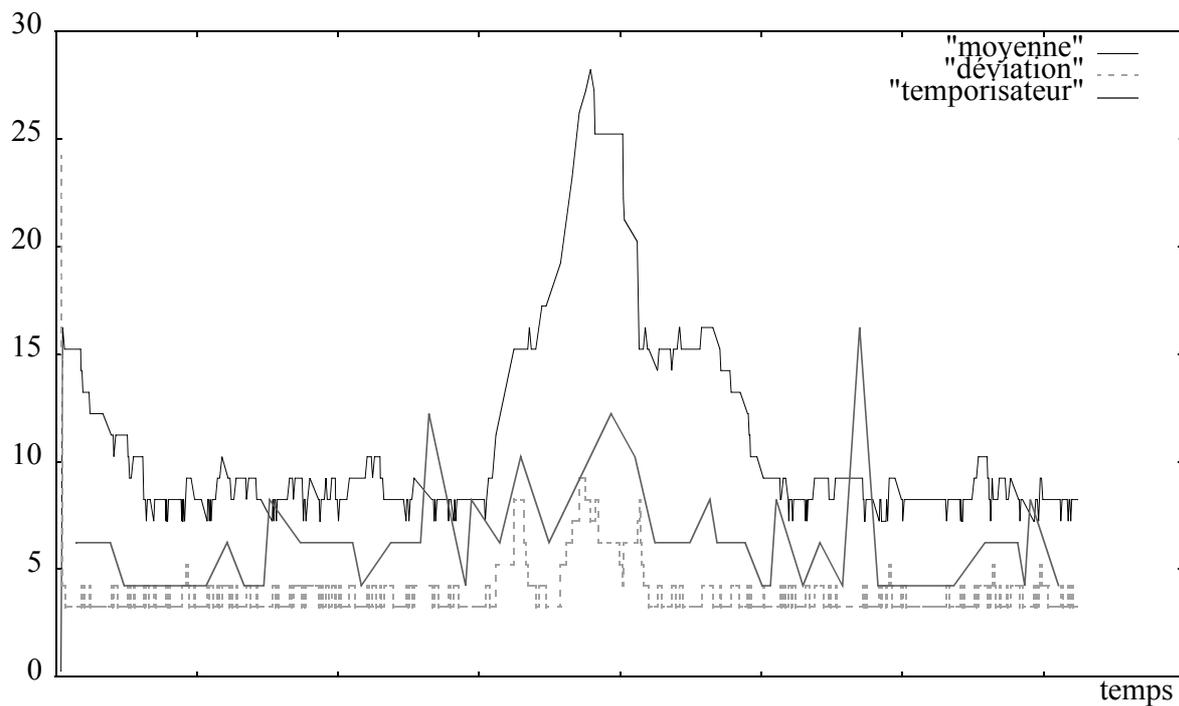
et

$$RTO = \text{moyenne} + 4 \times \text{deviation}$$

+ L'utilisation d'une bonne estimation améliore les performances

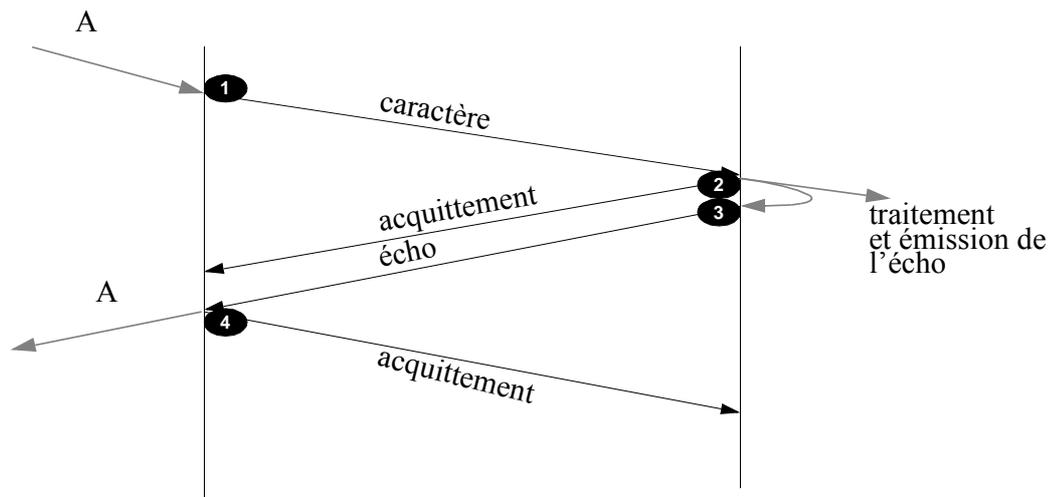
157

EXEMPLE (MESURÉ DANS LE NOYAU SUNOS 4.1.3)



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_suite.fm

LIMITATION DU TRAFIC

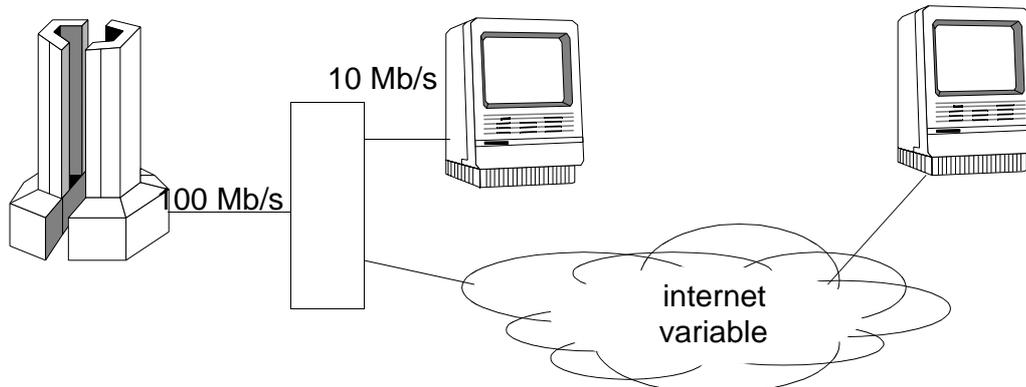


- Retard à l'acquittement de 200 ms pour permettre le "Piggy Backing"
- Algorithme de Naggle :
"on ne peut émettre sur le réseau qu'un paquet de petite taille non acquité"

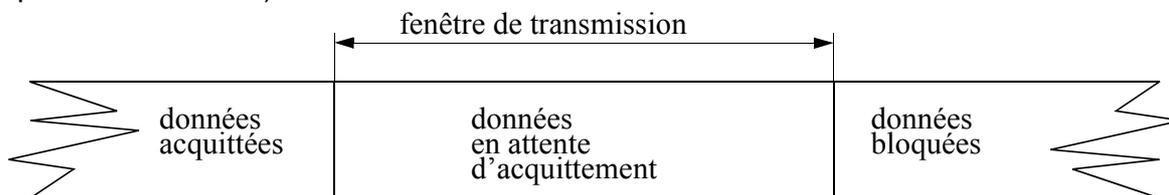
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_suite.fm

CONTRÔLE DE FLUX

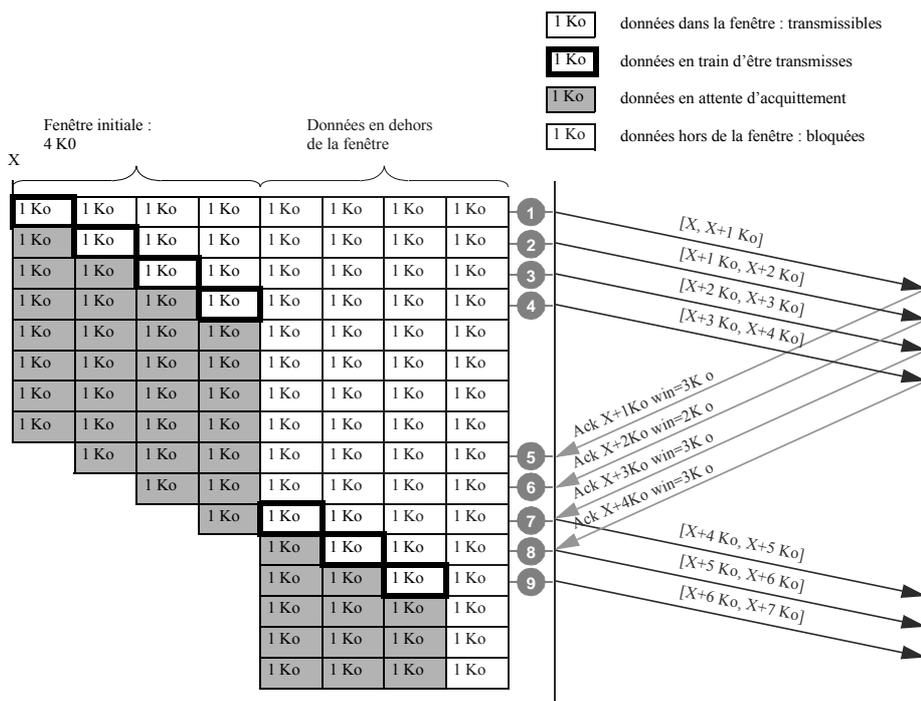
- Adapter dynamiquement le trafic émis à la configuration du réseau et au récepteur.



Basé sur la taille d'une fenêtre de retransmission (nombre d'octets non acquittés qui peuvent être émis)



FENÊTRE D'ANTICIPATION



CONGESTION

Si le débit du réseau varie (dépend des autres utilisateurs), il peut y avoir congestion des routeurs. Règles :

- Les congestions font perdre des paquets, qui seront retransmis, et qui à leur tour feront des congestions, qui feront perdre des paquets et qui ...
- La perte d'un paquet provient plus souvent d'une congestion que d'une erreur de transmission.
- La perte d'un paquet peut se détecter quand l'émetteur reçoit plusieurs acquittements identiques.
- La perte de plusieurs paquets se détecte quand la temporisation de retransmission se déclenche.

+ Pour s'adapter au débit, faire varier la taille de la fenêtre d'anticipation.

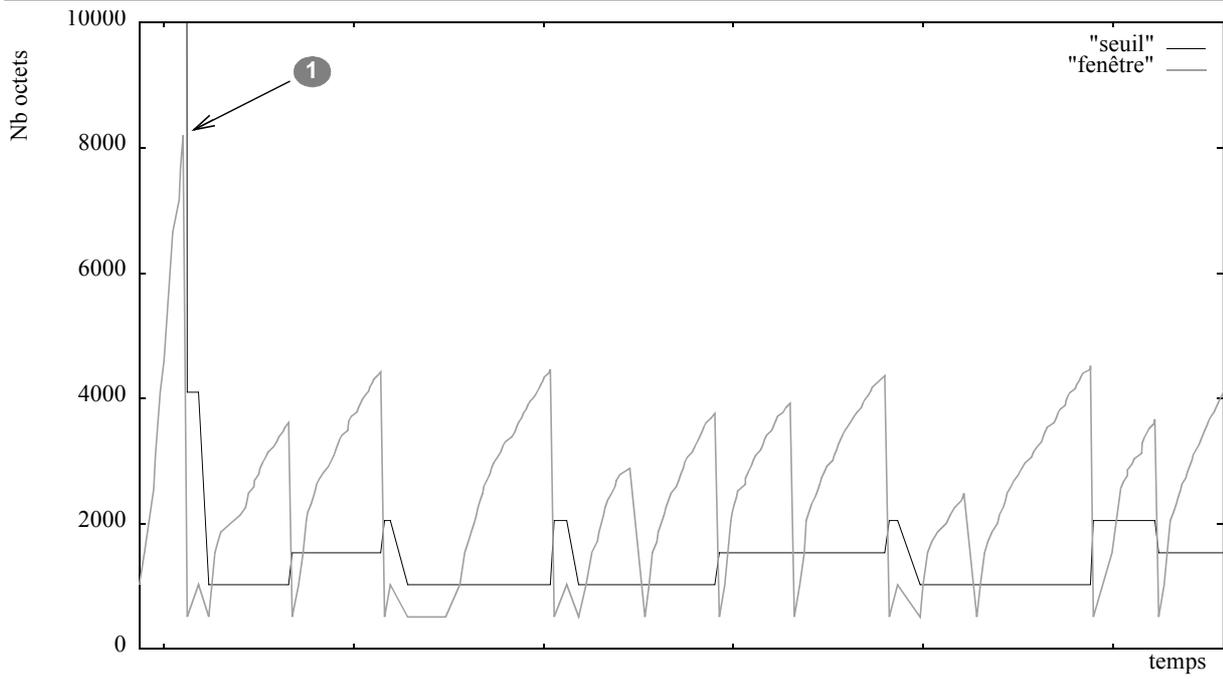
162

ALGORITHME DU SLOW START

- Soit la variable seuil initialement à 65 535 et la taille de la fenêtre initialement à 1.
- A chaque réception d'un acquittement, ajouter 1 à la taille de la fenêtre.
La taille de la fenêtre croît exponentiellement.
- Si la taille de la fenêtre atteint le seuil, incrémenter linéairement sa taille.
- Si un paquet est perdu (réception de trois acquittements identiques) alors :
taille de la fenêtre = taille de la fenêtre / 2
- Si pertes de plusieurs paquets (déclenchement du timer) alors :
seuil = taille de la fenêtre / 2 (le seuil ne peut pas être inférieur à 2 segments)
et
taille de la fenêtre = 1

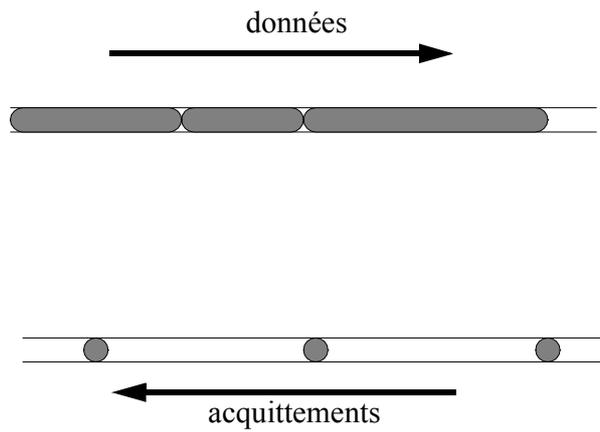
163

EXEMPLE



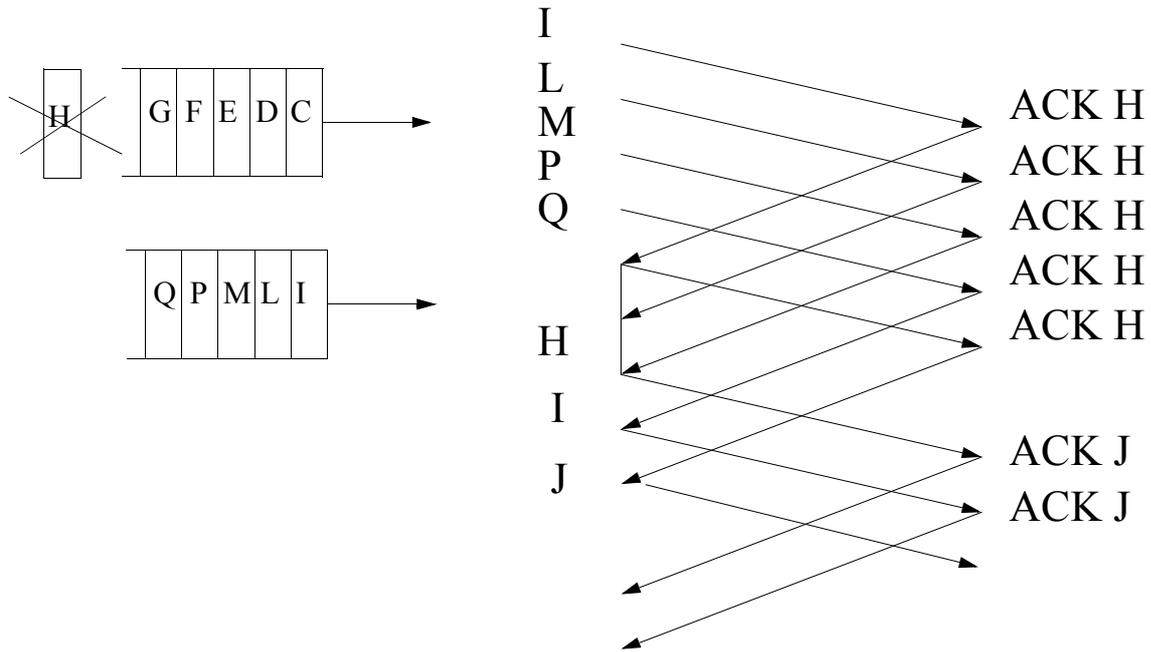
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_suite.fm

RÉGIME STATIONNAIRE



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_suite.fm

EXEMPLE

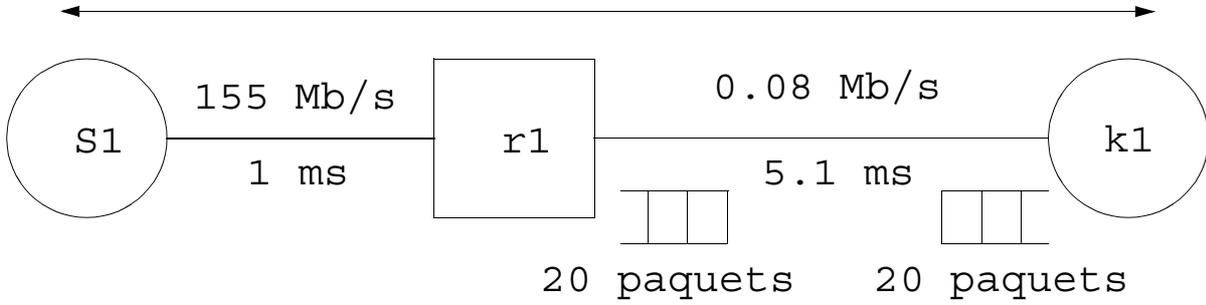


EXERCICES

⇒ Est ce que TCP supporte un déséquencelement des paquets, comme un réseau datagramme pourrait en introduire ?

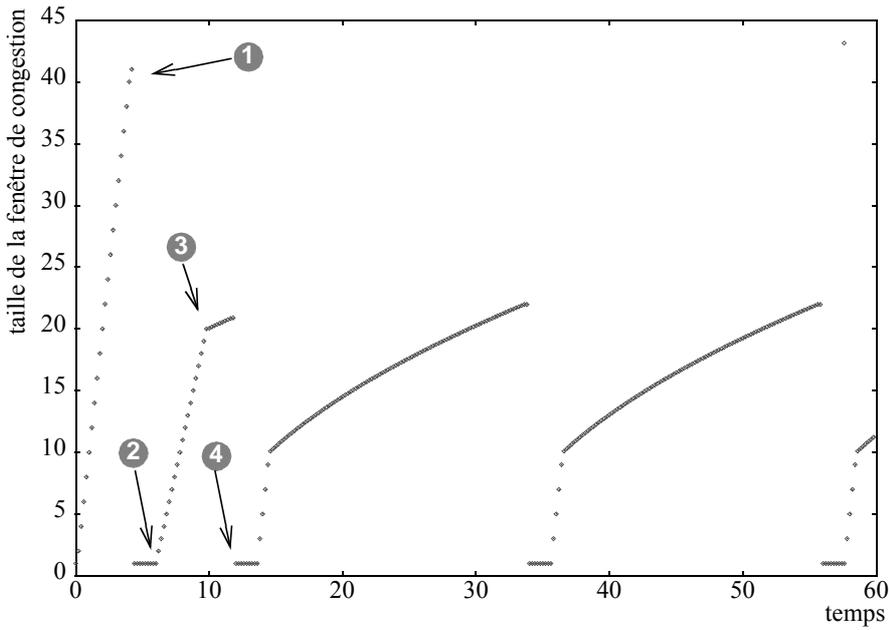
EXEMPLE PAR SIMULATION

tcp Tahoe



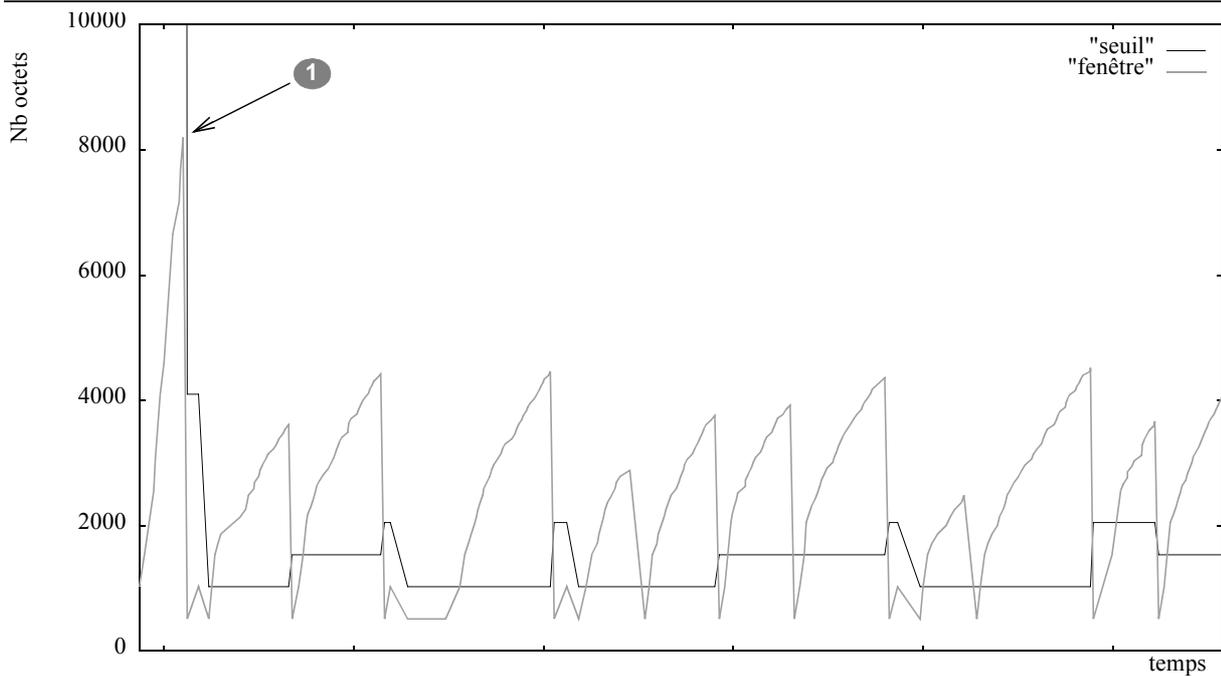
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_sulle.fm

EXEMPLE PAR SIMULATION



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_sulle.fm

EXEMPLE PAR MESURE



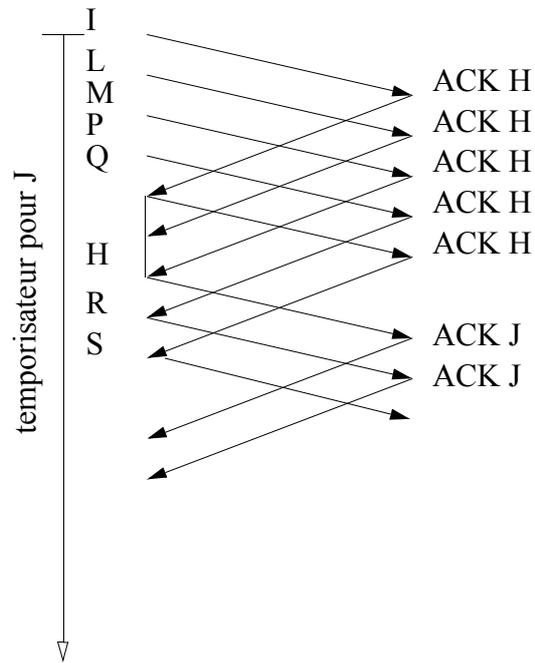
170

PLUSIEURS VERSIONS DE TCP

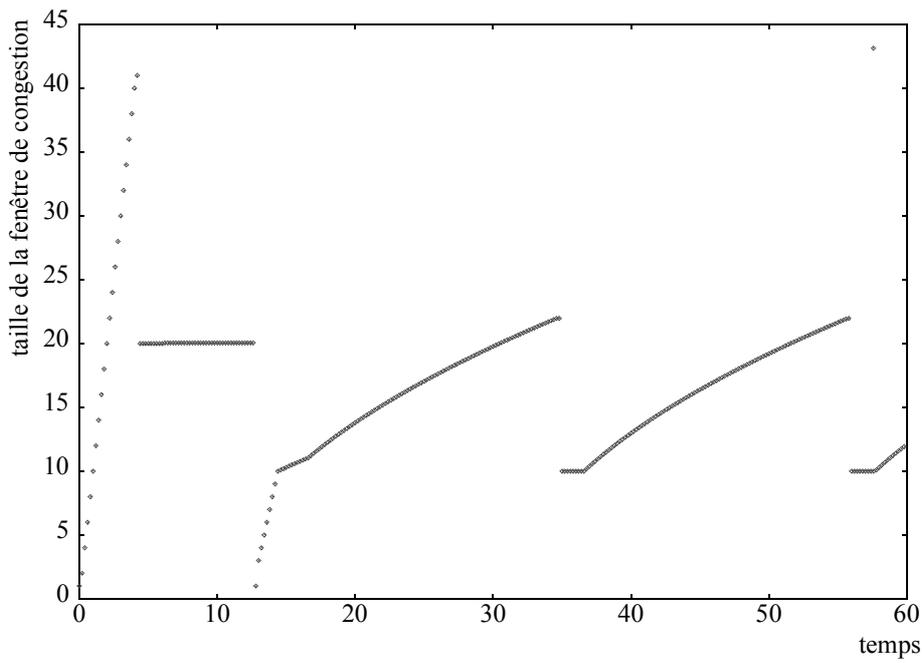
- Nom lié à la version de BSD dans laquelle elle est parue
 - Tahoe : initiale de Van Jacobson
 - Reno : Une perte unique (i.e. un acquittement dupliqué), n'est pas une congestion grave.
 - Pas de phase de Slow Start,
 - Division de la fenêtre par 2 et phase de Congestion Avoidance,
 - Retransmission uniquement du paquet perdu.
 - new Reno :
 - Correction des bugs de Reno.

171

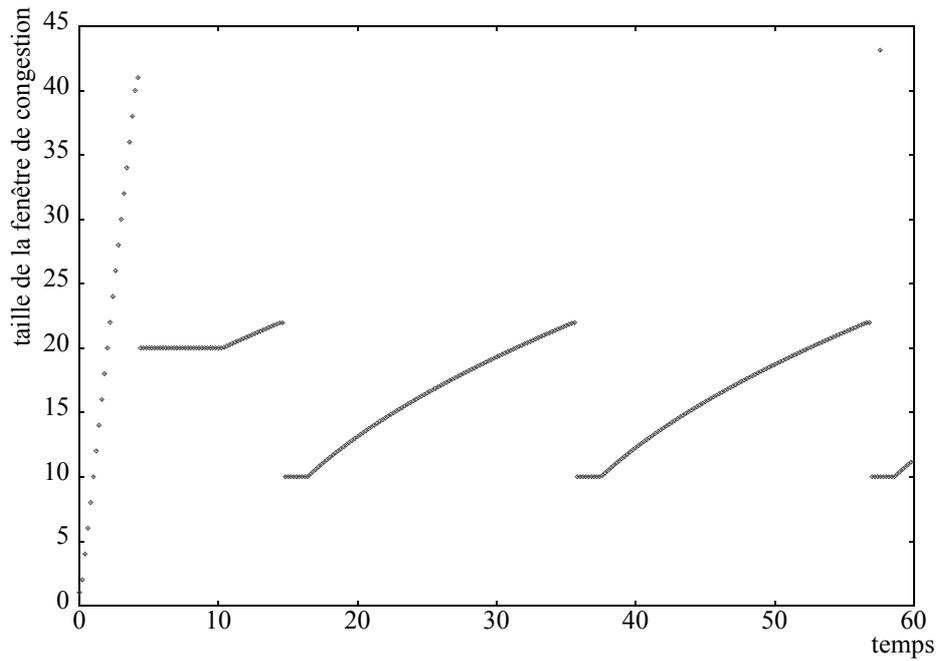
RENO



RENO



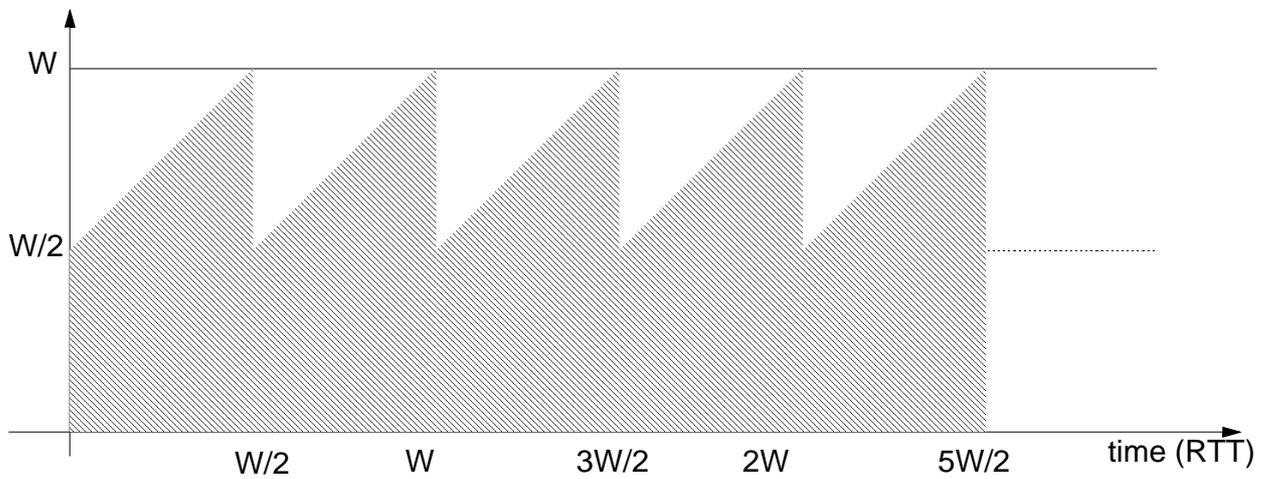
NEW RENO



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_suite.fm

MODÉLISATION DE TCP

- Simulation de petits agrégats de flux
- Modélisation analytique relativement difficile
- Approximation de TCP new Reno



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_suite.fm

 HYPOTHÈSES

Regime stationnaire:

- étude dans la phase de *congestion avoidance*
- un seul flux TCP
- fenêtre de congestion de W
 - perte d'un paquet quand la fenêtre de congestion atteint W
 - la fenêtre est dans ce cas divisée par 2
 - la fenêtre augmente d'un segment par RTT
 - au bout de $\frac{W}{2}RTT$, la fenêtre atteint à nouveau son maximum
 - il est délivré $\frac{3}{8}W^2$ paquets avant une perte.
- Soit p la probabilité de perte, on émet donc $1/p$ paquets avant de perdre un paquet.
 - $\left(\frac{1}{p} = \frac{3}{8}W^2\right) \Rightarrow W = \sqrt{\frac{8}{3p}}$

 MODÉLISATION

- la bande passante est le rapport entre le nombre d'octets transmis et la durée d'un cycle

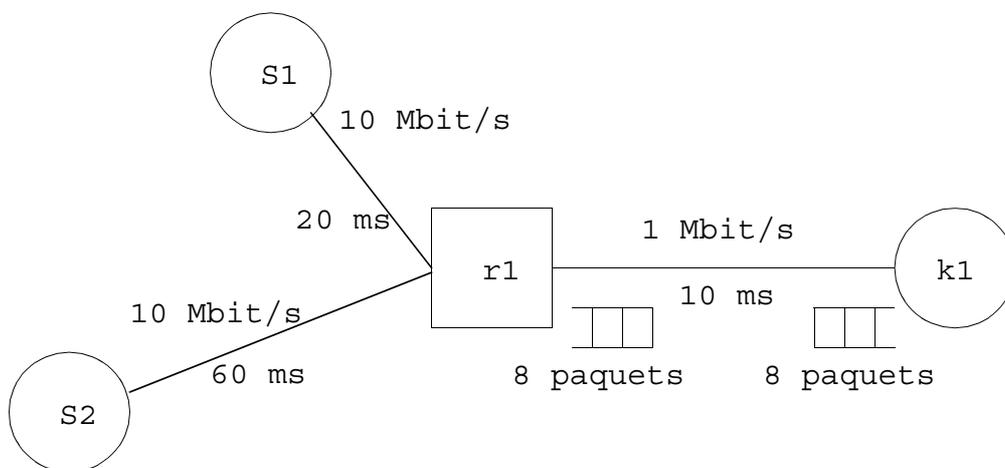
$$BW = \frac{\text{donnees}}{\text{duree}} = \frac{MSS \times \frac{3}{8} \times W^2}{RTT \times \frac{W}{2}} = \frac{MSS \times C}{RTT \times \sqrt{p}}, C = \sqrt{\frac{3}{2}}$$

- d'autres constantes C peuvent être calculées en fonction d'autres hypothèses sur les flux TCP^[1]

pertes périodiques	C	pertes aléatoires	C
Acquittement pour chaque paquet	$1.22 = \sqrt{\frac{3}{2}}$	Acquittement pour chaque paquet	1.31
Acquittement retardé	$1.22 = \sqrt{\frac{3}{2}}$	Acquittement retardé	0.93

1. *The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm*, M. Mathis, J. Semke, J. Mahdavi, ACM SIGCOMM, Vol 27, N° 3, July 1997, ISSN#0146-4833

TCP EQUITABLE ?

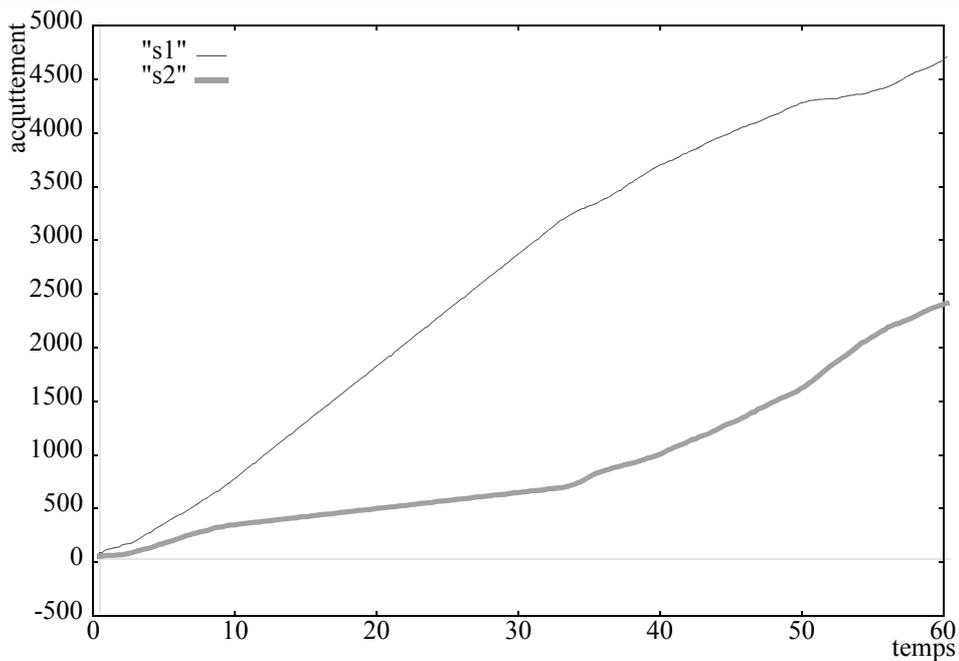


17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_suite.fm

TCP Equitable ?

178

TCP EQUITABLE ?



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_suite.fm

TCP Equitable ?

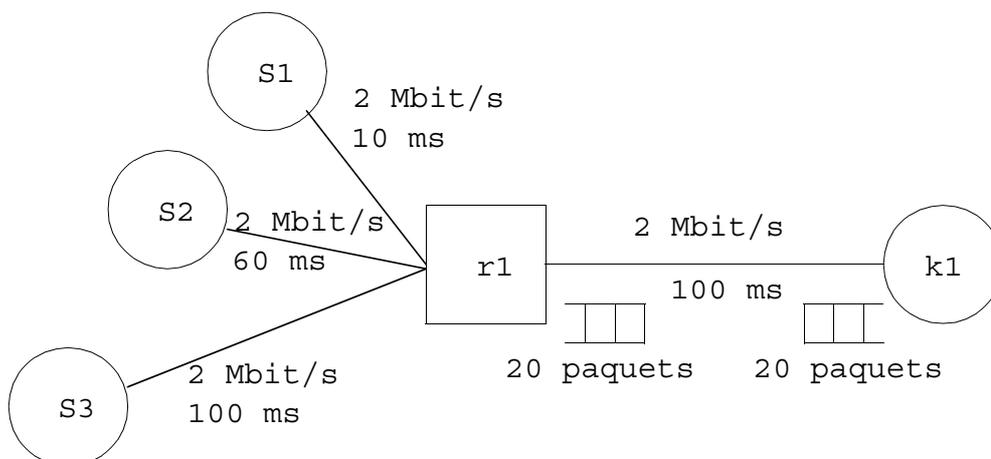
179

+ Synchronisation :

- Goulot d'étranglement :
 - saturation de la file d'attente du routeur
 - toutes les connexions perdent des paquets
 - toutes les connexions réduisent en même temps leur débit
- Solution :
 - RED
 - probabilité de perte proportionnelle à l'occupation de la file d'attente
 - les plus gros flux sont pénalisés

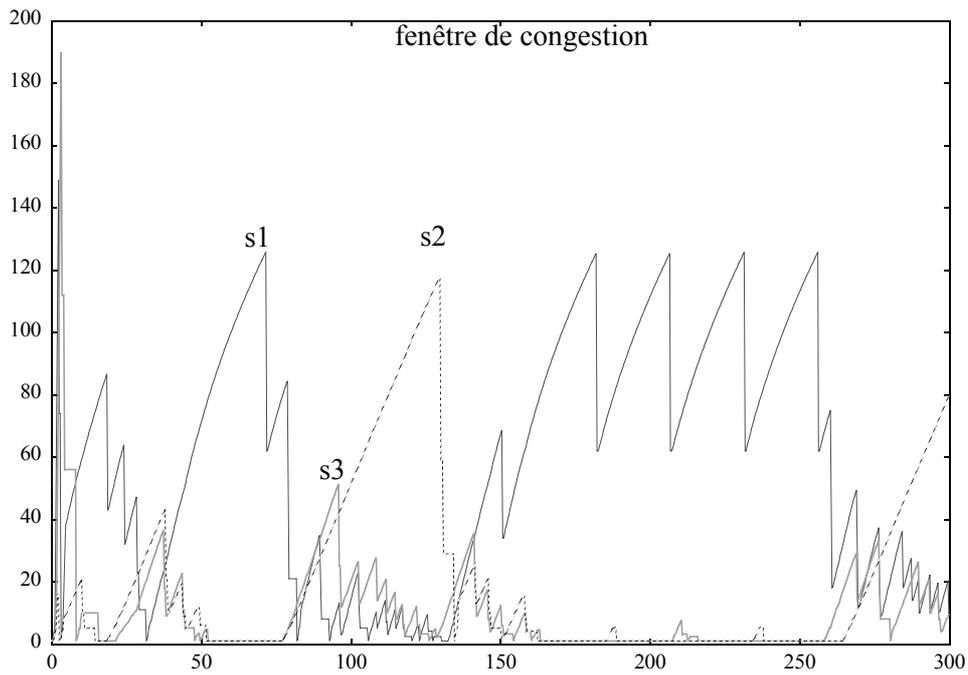
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_suite.fm

EXEMPLE



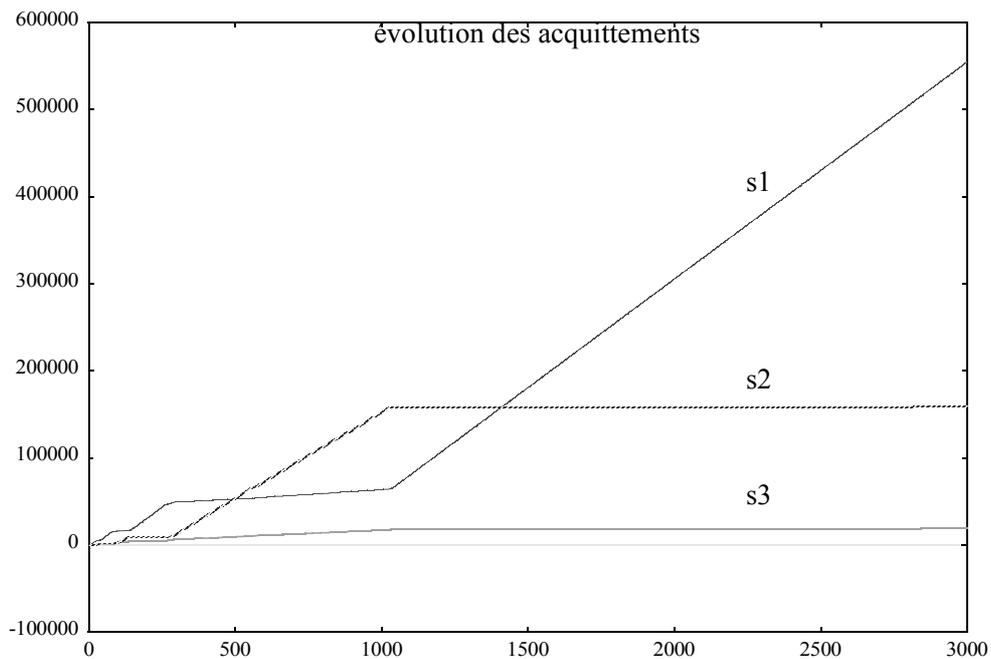
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_suite.fm

EXEMPLE FIFO



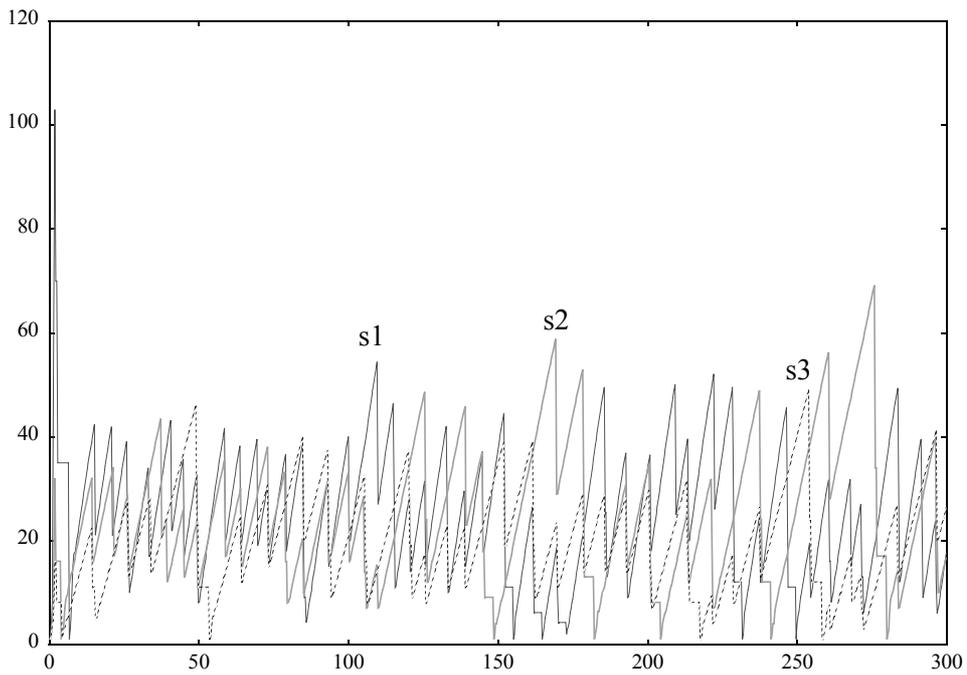
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_suite.fm

EXEMPLE FIFO



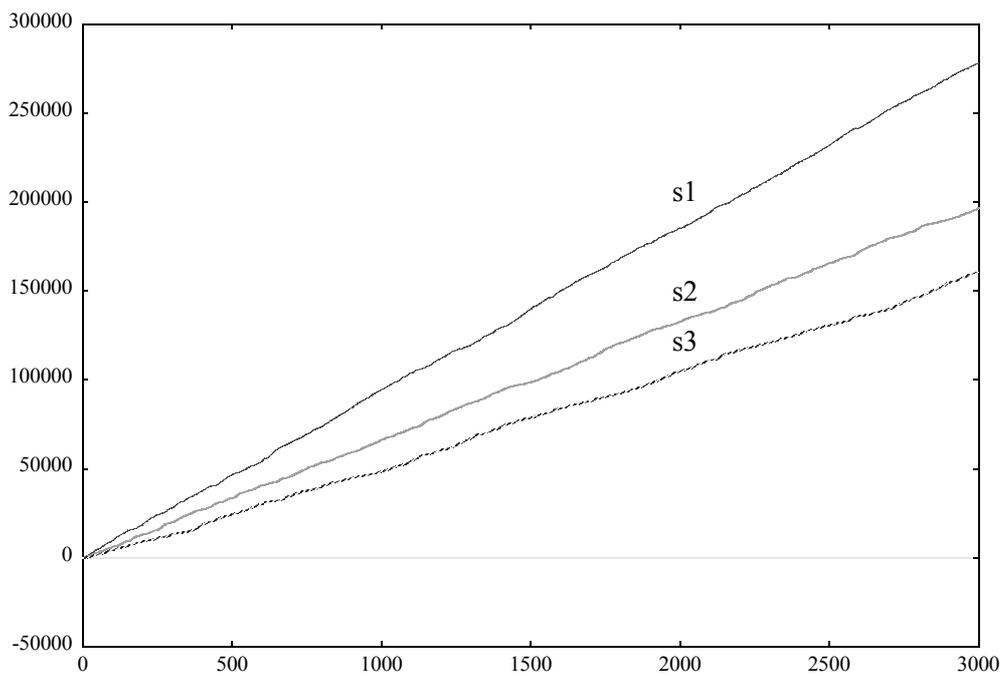
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_suite.fm

EXEMPLE RED



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_suite.fm

EXEMPLE RED



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_suite.fm

EXPLICIT CONGESTION NOTIFICATION

RED indique que les paquets en "excès" doivent être marqués

- Dans TCP le marquage signifie suppression
- Il faut donc 3 acquittements dupliqués pour que l'émetteur détecte la perte
- et retransmettre des informations perdus

ECN utilise deux bits non utilisés par DiffServ (CU : *Currently Unused*) pour marquer les paquets

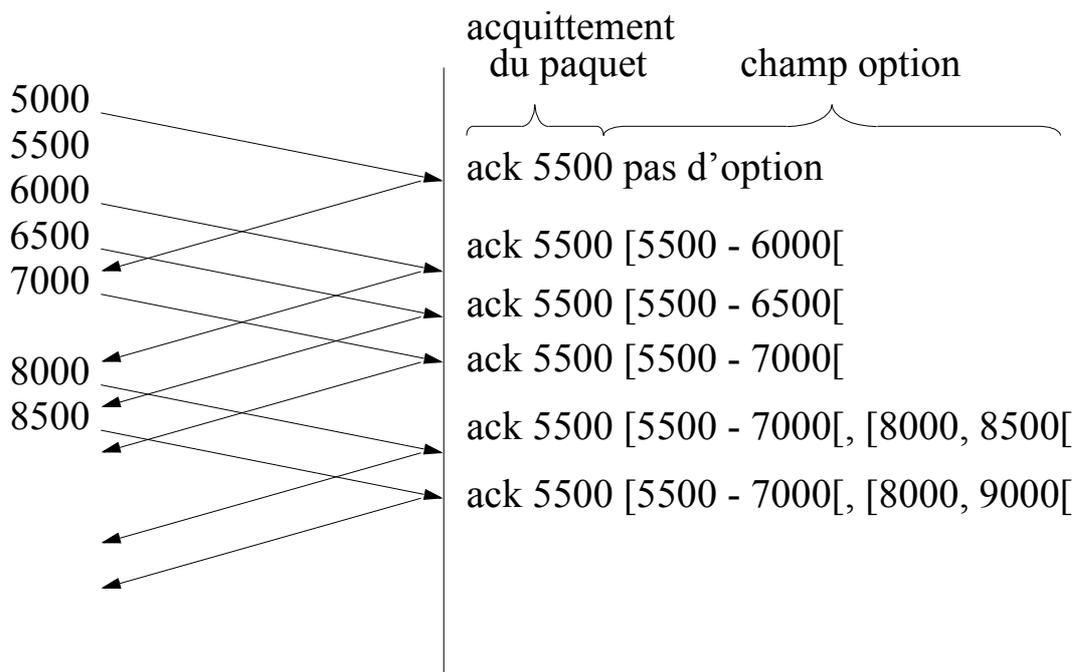
- un bit pour indiquer que le flux supporte ECN
- un bit positionné par les routeurs pour indiquer la congestion
 - Quand le destinataire reçoit un paquet avec ce bit positionné, il marque un champ dans le paquet TCP pour indiquer à la source qu'une congestion a lieu sur le chemin
- Equivalent au FECN de Frame Relay

PROBLÈMES DE TCP

+ Liens satellites ou radio :

- Beaucoup d'erreurs de transmission
- Algorithme de gestion des congestions prend les pertes de paquets pour des congestions
 - Réduction du débit d'émission
- Solution :
 - Acquittements sélectifs
 - Dans le champ option des messages TCP
 - Evite de tomber dans le déclenchement du temporisateur de retransmission

ACQUITTEMENTS SÉLECTIFS



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_suite.fm

PROBLÈMES DE TCP

+ Liens asymétriques

- ADSL, réseaux câblés, satellites,...
- La perte de paquet a lieu sur les acquittements pas sur les données
- Réduction du débit d'émission
- Solution ?
 - retarder les acquittements
 - enlever des acquittements dans la file d'attente
 - faire une autre connexion TCP (~proxy)

+ HTTP 1.1

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_suite.fm

AUTRES PROBLÈMES DE TCP

Problème d'équité

- ouvrir plusieurs connexions (ex Netscape)
- mauvaises mises en œuvre de TCP
- petits transferts
- multimédia

Solutions

- Modifier les applications (ex. p-HTTP)
- Mettre un contrôle de flux au niveau applicatif (ex : RTCP)
- Problèmes de cohérence
- Mettre un contrôle dans le réseau
 - dans chaque routeur.
 - à l'entrée.
 - pour chaque flux ?

NOUVEAUX BESOINS

Transports d'autres types d'informations :

- vidéo
- téléphonie
- jeux interactifs distribués

Contraintes de :

- débit
 - mécanismes de contrôle de flux de TCP inappropriés
- pertes
- délai
 - interactivité (téléphonie)

MÉCANISMES DE CONTRÔLE DE FLUX PAR VAN JACOBSON

TCP

- Estimation du délai :
 - $erreur = mesure - moyenne$
 - $moyenne = moyenne - \alpha \times erreur$
 - $dev = dev + \beta \times (|erreur| - dev)$
 - $RTO = moyenne + 4 \times dev$
- Slow start :
 - si pertes de plusieurs paquets (declenchement du timer) alors :*
 - seuil = taille de la fenêtre / 2 (le seuil ne peut pas être inférieur à 2 segments)*
 - et*
 - taille de la fenêtre = 1*

UDP

From list-mgr@ISI.EDU Tue Jan 10 22:35:19 1995
 Received: by venera.isi.edu (5.65c/5.61+local-20)
 id <AA25275>; Wed, 11 Jan 1995 06:31:42 -0800
 Received by rx7.ee.lbl.gov for mbonera@isi.edu (5.65/1.44)
 id AA01872; Wed, 11 Jan 95 06:35:21 -0800
 Message-Id: <9501111435.AA01872@rx7.ee.lbl.gov>
 To: mbone-na
 Subject: **jpuckett@168.18.130.249 sendig 600kb/s video**
 Date: Wed, 11 Jan 95 06:35:19 PST
 From: **Van Jacobson** <van@ee.lbl.gov>

Someone named jpuckett in Georgia (domain peachnet.edu) is sending ridiculous amounts of video (400-600kb/s) to the "Live from Antarctica" session (and most other mbone video sessions).
 The reverse dns mapping for this address doesn't work so I haven't been able to send email to this individual. **Could someone get in touch with them & explain to them that this behavior is antisocial?** Thanks.
 - Van

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_sulle.fm

TCP FRIENDLY

- Mettre un contrôle de flux aux applications utilisant UDP pour que celles-ci ne conduisent pas à une famine des connexions TCP
- Le controle de flux conduit à réduire le nombre de paquets émis :
 - augmentation du taux de compression
- Reprendre la formule

$$BW = \frac{MSS \times C}{RTT \times \sqrt{p}}, C = 1,22$$

- Le taux de perte permet de calculer la bande passante que devrait produire l'application.

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\TCP_sulle.fm

CONFIGURATION D'UNE STATION UNIX

```

/usr/etc/ifconfig interface [ address_family ]
    [ address [ dest_address ] ] [ netmask mask ]
    [ broadcast address ] [ up ] [ down ] [ trailers ]
    [ -trailers ] [ arp ] [ -arp ] [ private ]
    [ -private ] [ metric n ] [ auto-revarp ]
  
```

EXEMPLE

```

root@bloodmoney[27]# ifconfig le0 bloodmoney netmask +
Setting netmask of le0 to 255.255.255.128
root@bloodmoney[28]# ifconfig -a
le0: flags=863<UP,BROADCAST,NOTRAILERS,RUNNING>
    inet 192.44.77.81 netmask ffffff80 broadcast 192.44.77.0
    ether 8:0:20:1c:74:84
lo0: flags=849<UP,LOOPBACK,RUNNING>
    inet 127.0.0.1 netmask ff000000
  
```

TABLES DE ROUTAGE

bloodmoney (192.44.77.81) :

```

>netstat -n -r
Routing tables
Destination      Gateway          Flags  Refcnt Use  Interface
192.108.119.16   192.44.77.77    UGHD   1    1683  le0
127.0.0.1        127.0.0.1       UH     2    12971  lo0
default          192.44.77.2     UG     3    16977  le0
192.44.77.0      192.44.77.81   U      13   5780   le0
  
```

nintendo (192.44.77.77) :

```

>rsh nintendo netstat -n -r
Routing tables
Destination      Gateway          Flags  Refcnt Use  Interface
127.0.0.1        127.0.0.1       UH     3    351344  lo0
default          192.44.77.2     UG     3    17388997  le0
192.44.77.128    192.44.77.252   U      26   504768  le2
192.44.77.0      192.44.77.77    U      24   10702069  le0
192.108.119.0    192.108.119.1   U      2    249777  le1
  
```

QUESTIONS

⇒ Combien d'interfaces possède nintendo ?

⇒ Quelle route est prise pour un datagramme émis de bloodmoney vers la machine 192.108.119.16 ?

⇒ Quelle route est prise pour un datagramme émis de bloodmoney vers la machine 192.108.119.17 ?

TABLES DE ROUTAGE (SUITE)

sur mgs (192.44.77.2) routeur CISCO

>telnet mgs

Trying 192.44.77.2 ...

Connected to mgs-rsm.

Escape character is '^'].

User Access Verification

Password:

mgs>enable

Password:

mgs#show ip ro

Codes: I - IGRP derived, R - RIP derived, O - OSPF derived

C - connected, S - static, E - EGP derived, B - BGP derived

i - IS-IS derived, D - EIGRP derived

* - candidate default route, IA - OSPF inter area route

E1 - OSPF external type 1 route, E2 - OSPF external type 2 route

L1 - IS-IS level-1 route, L2 - IS-IS level-2 route

EX - EIGRP external route

Gateway of last resort is 193.52.72.1 to network 0.0.0.0

17 janvier 2003 - C:\Documents and Settings\loutain\Mes documents\Support_ENST\configuration_fm

Configuration d'une station Unix

200

192.44.77.0 is subnetted (mask is 255.255.255.128), 2 subnets

C 192.44.77.128 is directly connected, Ethernet5

C 192.44.77.0 is directly connected, Ethernet0

C 193.52.72.0 is directly connected, Ethernet2

193.52.74.0 is subnetted (mask is 255.255.255.192), 3 subnets

S 193.52.74.128 [1/0] via 193.52.74.4

C 193.52.74.0 is directly connected, Ethernet3

C 193.52.74.64 is directly connected, Ethernet1

192.108.119.0 is subnetted (mask is 255.255.255.192), 2 subnets

C 192.108.119.64 is directly connected, Ethernet4

17 janvier 2003 - C:\Documents and Settings\loutain\Mes documents\Support_ENST\configuration_fm

Configuration d'une station Unix

201

MODIFICATION DES TABLES DE ROUTAGE

```
/usr/etc/route [ -fn ] add|delete [ host|net ] destination [gateway [ metric ] ]
```

exemple :

```
root@bloodmoney[30]# netstat -r
Routing tables
Destination      Gateway          Flags  Refcnt Use  Interface
localhost        localhost        UH     2    13569  lo0
192.44.77.0      bloodmoney       U      18   13272  le0
root@bloodmoney[31]# ping 133.11.11.11
sendto: Network is unreachable
root@bloodmoney[32]# route add 0.0.0.0 mgs-rsm 1
add net 0.0.0.0: gateway mgs-rsm
root@bloodmoney[33]# netstat -r
Routing tables
Destination      Gateway          Flags  Refcnt Use  Interface
localhost        localhost        UH     2    13591  lo0
default          mgs-rsm          UG     0     0      le0
192.44.77.0      bloodmoney       U      16   13566  le0
root@bloodmoney[34]# ping 133.11.11.11
133.11.11.11 is alive
```

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\configuration.fm

MODIFICATION DES TABLES DE ROUTAGE (ICMP REDIRECT)

```
root@bloodmoney[35]# ping atm2
ICMP Host redirect from gateway mgs-rsm (192.44.77.2)
to nintendo (192.44.77.77) for atm2 (192.108.119.16)
atm2 is alive
root@bloodmoney[36]# netstat -r
Routing tables
Destination      Gateway          Flags  Refcnt Use  Interface
atm2             nintendo         UGHD   0     0      le0
localhost        localhost        UH     2    13591  lo0
default          mgs-rsm          UG     0     13     le0
192.44.77.0      bloodmoney       U      17   13669  le0
```

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\configuration.fm

ADMINISTRATION DES ÉQUIPEMENTS AVEC SNMP

Un standard :

- snmp (Simple Network Management Protocol)

basé sur le principe de centres de gestion qui gèrent des agents snmp placés dans les équipements.

Concept ouvert :

- coté équipements : chaque constructeur fournit la description de ses équipements via la MIB (Management Information Base)
- coté centre de gestion snmp : ils savent accepter tout type d'équipement, sous réserve de fourniture de la MIB.

De nombreux produits centre de gestion :

- Spectrum (Cabletron),
- Sun NetManager,
- HP Openview ...
- fournis par les constructeurs d'équipements réseau et les constructeurs Unix.

TYPE DES VARIABLES

- INTEGER
- OCTET STRING
- IpAddress est composé de 4 entiers. Ce type code une adresse IP.
- NetworkAddress : type générique pour plusieurs adresses de niveau 3
Seules les adresses IP sont autorisées.
- Counter : valeurs positives entre 0 et $2^{32}-1$. L peut jamais décroître sauf en cas de débordement.
- Gauge : valeurs positives entre 0 et $2^{32}-1$. peut croître ou décroître.
- TimeTicks : durée en centièmes de secondes
- Opaque : n'est pas interprété par SNMP.

EXEMPLE DE MIB POUR IP.

objet	type	ac st	description
ipForwarding 1	INTEGER	rw m	indique si l'équipement sert de routeur (1) ou s'il rejette les paquets qui ne lui sont pas adressés (2).
ipDefaultTTL 2	INTEGER	rw m	valeur par défaut de la durée de vie placée dans les paquets à l'émission.
ipInReceives 3	Counter	ro m	nombre de paquets reçus des interfaces.
ipInHdrErrors 4	Counter	ro m	nombre de paquets rejetés à cause d'une erreur dans l'entête (checksum, durée de vie à 0, ...).
ipInAddrErrors 5	Counter	ro m	nombre de paquet reçus rejetés à cause d'une fausse adresse.
ipForwDatagrams 6	Counter	ro m	nombre de paquets routés (i.e. réémis sur une autre interface)
ipInUnknownProtos 7	Counter	ro m	nombre de paquets rejetés à cause d'un numéro de protocole de niveau supérieur non mis en œuvre.
ipInDiscards 8	Counter	ro m	nombre de paquets rejetés (par exemple à cause d'un manque de place en mémoire).
ipInDelivers 9	Counter	ro m	nombre de paquets remis à la couche supérieure (incluant ICMP).
ipOutRequests 10	Counter	ro m	nombre de paquets soumis pas les couches supérieures (incluant ICMP).
ipOutDiscards 11	Counter	ro m	nombre total de paquets rejetés
ipOutNoRoutes 12	Counter	ro m	nombre de paquets rejetés car ils ne pouvaient pas être routés.
ipReasmTimeout 13	Counter	ro m	nombre de paquets rejetés à la suite de la perte d'un fragment.
ipReasmReqds 14	Counter	ro m	nombre de paquets qui ont nécessités un réassemblage.
ipReasmOKs 15	Counter	ro m	nombre de paquets ayant réussis à être réassemblés.

17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\Administration.fm

objet	type	ac st	description
ipReasmFails 16	Counter	ro m	nombre de paquets qui n'ont pas pus être réassemblés à cause d'une erreur.
ipFragOKs 17	Counter	ro m	nombre de paquets qui ont été fragmentés sans erreurs.
ipFragFails 18	Counter	ro m	nombre de paquets qui n'ont pu être fragmentés (par exemple bit DF=1).
ipFragCreates 19	Counter	ro m	nombre de fragments créés.
ipAddrTable 20	ipAddrEntry : 20.1	na m	configuration des interfaces.
ipAdEntAddr 20.1.1	IpAddress	ro m	adresse IP de cette entrée.
ipAdEntIfIndex 20.1.2	INTEGER	ro m	numéro de l'interface correspondante.
ipAdEntNetMask 20.1.3	IpAddress	ro m	netmask associée à l'adresse IP.
ipAdEntBcastAddr 20.1.4	INTEGER	ro m	adresse IP de diffusion
ipAdEntReasm MaxSize 20.1.5	INTEGER (0..65535)	ro m	longueur du plus grand datagramme IP qui peut être réassemblé.
ipRouteTable 21	ipRouteEntry : 21.1	na m	définition des tables de routage.
ipRouteDest 21.1.1	IpAddress	rw m	adresse du réseau ou de la station destination pour lequel la table de routage est définie.
ipRouteIfIndex 21.1.2	INTEGER	rw m	numéro de l'interface.
ipRouteMetric1 21.1.3	INTEGER	rw m	coût ou métrique principale : -1 la route n'est pas accessible.
ipRouteMetric2 ipRouteMetric3 ipRouteMetric4 21.1.4 - 6	INTEGER	rw m	métriques complémentaires.

17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\Administration.fm

objet	type	acst	description
ipRouteNextHop 21.1.7	IpAddress	rw m	adresse du routeur qui permet d'atteindre la destination annoncée.
ipRouteType 21.1.8	INTEGER	rw m	type de la route : 1 : autres 2 : invalide 3 : directe (même sous-réseau IP) 4 : indirecte (à travers un routeur intermédiaire).
ipRouteProto 21.1.9	INTEGER	ro m	comment la route a été apprise (1 autre méthode, 2 configuration manuelle, 3 par la plate-forme d'administration, 4 ICMP, 5 EGP, 6 GGP, 7 Hello, 8 RIP, 9 IS-IS, 10 ES-IS, 11 Cisco IGRP, 12 bbnSpfIgp, 13 OSPF, 14 BGP).
ipRouteAge 21.1.10	INTEGER	rw m	durée en seconde depuis la dernière modification.
ipRouteMask 21.1.11	IpAddress	rw m	netmask associé à la route.
ipRouteMetric5 21.1.12	INTEGER		métrique complémentaire.
ipRouteInfo 21.1.13	OBJECT IDENTIFIE R	ro m	identificateur vers le protocole de routage qui a mis cette route
ipNetToMediaTable 22	IpNetToMe dia Entry :	na m	table de correspondance entre les adresse IP et physique. Remplace l'objet at.
ipNetToMedia IfIndex 22.1.1	INTEGER	rw m	numéro de l'interface.
ipNetToMedia PhysAddress 22.1.2	PhysAddre ss	rw m	adresse physique.
ipNetToMedia NetAddress 22.1.3	IpAddress	rw m	adresse IP.
ipNetToMedia Type 22.1.4	INTEGER	rw m	type de média (1 autre, 2 invalide, 3 dynamique, 4 statique)
ipRoutingDiscards 23	Counter	ro w	nombre d'entrées (même valides) supprimées de la table de routage.

PRIMITIVES ET OBJETS

Le gestionnaire et les agents SNMP s'échangent des primitives simples sur UDP/IP pour manipuler des objets. Les primitives sont:

- SET positionner un objet (ex : config)
- GET obtenir un objet (ex : compteur)
- GET-NEXT obtenir le suivant
- TRAP génération d'une alerte (ex : ligne HS)

les échanges sont protégés par le "nom de communauté".

M.I.B. (Management Information Base) les objets sont mémorisés dans une base d'informations de gestion de réseau M.I.B. Les objets sont des éléments d'information tels que :

- des compteurs statistiques,
- des états de liaisons,
- des alertes en cas d'incident,
- des paramètres de configuration ...

RFC1157-SNMP DEFINITIONS ::= BEGIN

IMPORTS

ObjectName, ObjectSyntax, NetworkAddress, IpAddress, TimeTicks **FROM** RFC1155-SMI;

-- top-level message

```
Message ::= SEQUENCE {  
    version      INTEGER {  
        version-1(0)           -- version-1 for this RFC  
    },  
    community   OCTET STRING, -- community name  
    data        ANY  
}
```

```
PDU ::= CHOICE {  
    get-request  GetRequest-PDU,  
    get-next-request  GetNextRequest-PDU,  
    get-response  GetResponse-PDU,  
    set-request  SetRequest-PDU,  
    trap         Trap-PDU  
}
```

END

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Administration.fm

```
GetRequest-PDU ::= [0] IMPLICIT SEQUENCE {  
    request-id RequestID,  
    error-status ErrorStatus, -- always 0  
    error-index ErrorIndex, -- always 0  
    variable-bindings VarBindList  
}
```

```
GetNextRequest-PDU ::= [1] IMPLICIT SEQUENCE {  
    request-id RequestID,  
    error-status ErrorStatus, -- always 0  
    error-index ErrorIndex -- always 0  
    variable-bindings VarBindList  
}
```

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Administration.fm

```

GetResponse-PDU ::= [2] IMPLICIT SEQUENCE {
    request-id RequestID,
    error-status ErrorStatus,
    error-index ErrorIndex,
    variable-bindings VarBindList
}

SetRequest-PDU ::= [3] IMPLICIT SEQUENCE {
    request-id RequestID,
    error-status ErrorStatus, -- always 0
    error-index ErrorIndex, -- always 0
    variable-bindings VarBindList
}
    
```

EXEMPLE SUR UN TABLEAU.

```

SNMP: len: 38 version: int(1) 0x00 comm: string(6) "public" type: GET-NEXT
req-id: int(2) 0x5e31 error: int(1) 0x00 error-index: int(1) 0x00
var: obj(7) 1 3 6 1 2 1 2 1 val: empty(0)          "Combien d'interfaces ? "

SNMP: len: 40 version: int(1) 0x00 comm: string(6) "public" type: RESPONSE
req-id: int(2) 0x5e31 error: int(1) 0x00 error-index: int(1) 0x00
var: obj(8) 1 3 6 1 2 1 2 1 0 val: int(1) 0x06    " il y a 6 interfaces"

SNMP: len: 178 version: int(1) 0x00 comm: string(6) "public" type: GET-NEXT
req-id: int(3) 0x00a2a2 error: int(1) 0x00 error-index: int(1) 0x00
var: obj(9) 1 3 6 1 2 1 2 2 1 1 val: empty(0)
....
var: obj(9) 1 3 6 1 2 1 2 2 1 10 val: empty(0)

SNMP: len: 219 version: int(1) 0x00 comm: string(6) "public" type: RESPONSE
req-id: int(3) 0x00a2a2 error: int(1) 0x00 error-index: int(1) 0x00
var: obj(10) 1 3 6 1 2 1 2 2 1 1 1 val: int(1) 0x01
var: obj(10) 1 3 6 1 2 1 2 2 1 2 1 val: string(9) "Ethernet0"
...
var: obj(10) 1 3 6 1 2 1 2 2 1 10 1 val: counter(4) 0x6b055aa0
    
```

Exemple suite

SNMP: len: 188 version: int(1) 0x00 comm: string(6) "public" type: GET-NEXT
req-id: int(3) 0x00a422 error: int(1) 0x00 error-index: int(1) 0x00
var: obj(10) 1 3 6 1 2 1 2 2 1 1 1 1 val: empty(0)

.....

var: obj(10) 1 3 6 1 2 1 2 2 1 9 1 val: empty(0)
var: obj(10) 1 3 6 1 2 1 2 2 1 10 1 val: empty(0)

SNMP: len: 220 version: int(1) 0x00 comm: string(6) "public" type: RESPONSE
req-id: int(3) 0x00a422 error: int(1) 0x00 error-index: int(1) 0x00
var: obj(10) 1 3 6 1 2 1 2 2 1 1 2 val: int(1) 0x02
var: obj(10) 1 3 6 1 2 1 2 2 1 2 2 val: string(9) "Ethernet1"
var: obj(10) 1 3 6 1 2 1 2 2 1 3 2 val: int(1) 0x06

....

var: obj(10) 1 3 6 1 2 1 2 2 1 9 2 val: time(2) 0x0420
var: obj(10) 1 3 6 1 2 1 2 2 1 10 2 val: counter(5) 0x009f209925



Que se passerait-il si GET avait été utilisé à la place de GET-NEXT ?



Est-ce que la numérotation des interfaces doit être contigue ?

REMARQUES

Le protocole est standard mais pas l'offre commerciale:

- il n'est pas toujours facile de monter une solution fédérant des équipements hétérogènes.
- Le temps pour rendre l'administration opérationnelle n'est pas négligeable./

SNMP n'a pas été conçu pour gérer les sites distants: trafic, sécurité. De bonnes solutions propriétaires.

SNMPv2 améliore la gestion des sites distants (secours et hiérarchisation) et la sécurité.

RMON = la sonde distante.

ADMINISTRATION OSI - CMIS/CMIP

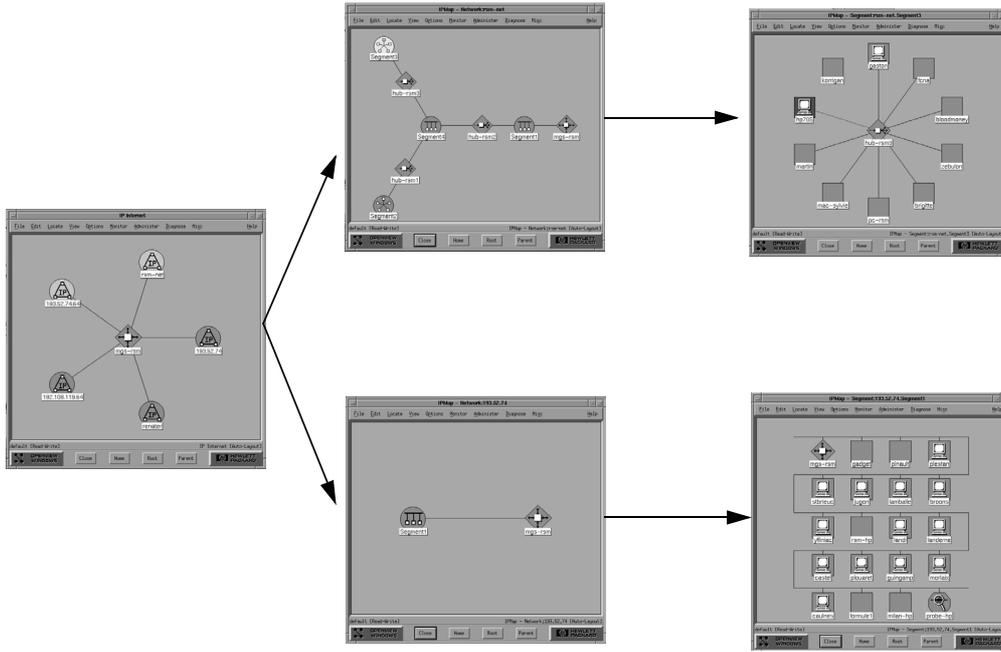
Mêmes principes que snmp, mais respecte l'architecture OSI

Plus lourd à développer, encore peu de solutions.

De gros constructeurs se sont engagés :

- Network Management Forum (AT&T, DEC, HP...)

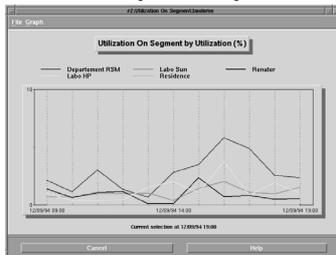
Openview



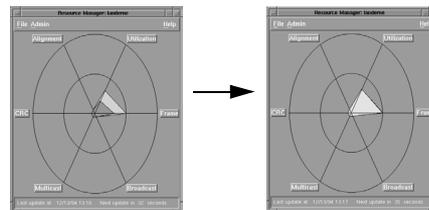
17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\Administration.fm

Openview (suite)

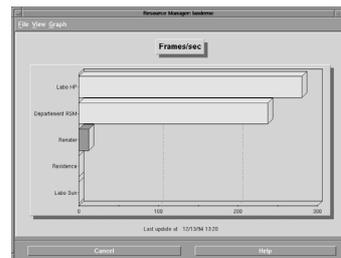
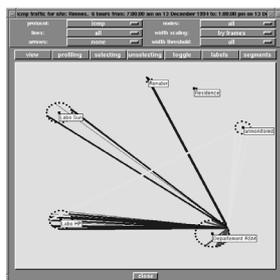
History Analyser



Resource Manager



Traffic Expert



17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\Administration.fm



FONCTIONS D'UN ROUTEUR

- Contrairement aux ponts, les routeurs doivent être configurés.
- Ils doivent connaître les adresses des routeurs ou des stations vers lesquels ils envoient les paquets.
- Ces informations peuvent être entrées manuellement (routage statique) ou acquises par dialogue entre routeurs.
- Dans ce dernier cas, les routeurs peuvent détecter la panne ou l'ajout d'un routeur sur le réseau et modifier leur plan de routage.
- Deux grandes familles d'algorithmes existent :
 - Le distance vector et
 - Le Link State.

220

LES ALGORITHMES DE ROUTAGE

+ Permettent l'échange automatique de tables de routage.

Table de routage :

- pour aller :
 - réseau,
 - sous-réseau,
 - machine.
- passer par :
 - attachement local,
 - routeur.
- avec un coût de (*métrique*) :
 - nombre de sauts,
 - fonction du débit, délai...

CLASSIFICATION DES ALGORITHMES DE ROUTAGE.

	Monde Internet	Monde ISO
routage intra-domaines (IGP)	distant vector : RIP, RIP-II, IGRP link state :OSPF, DUAL : EIGRP	IS-IS
routage inter-domaines (EGP)	EGP (obsolète) BGP	IDRP
Entre routeur et équipement	ICMP redirect	IS-ES

17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\ROUTAGE_Interne.fm

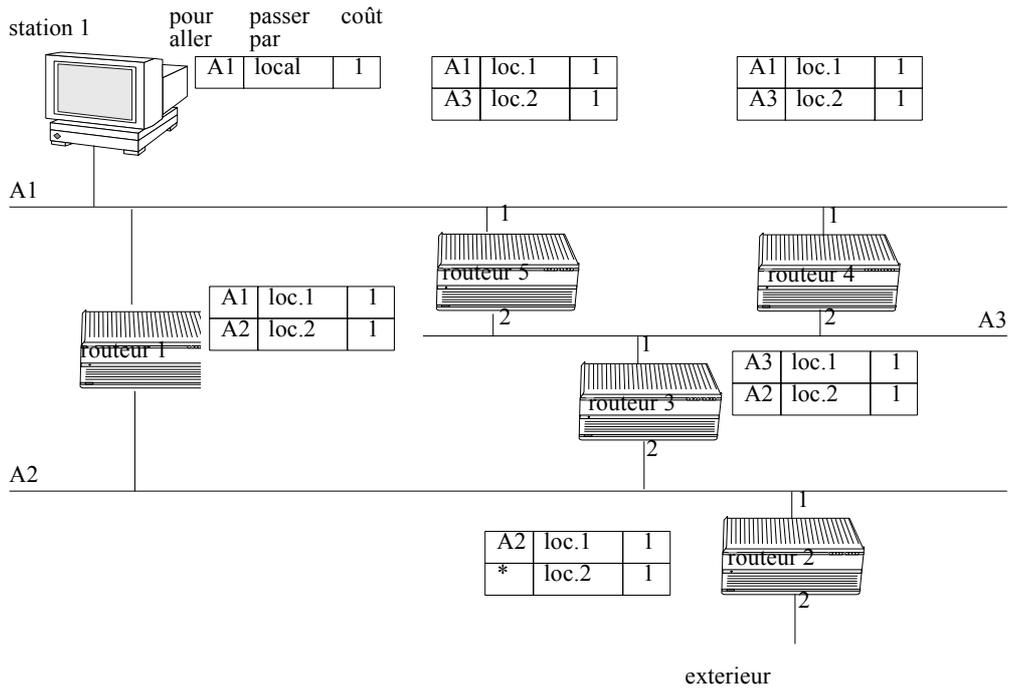
ALGORITHME DU DISTANT VECTOR

- Echange d'information entre routeurs adjacents.
- Les routeurs diffusent vers les nœuds adjacents leur table de routage rudimentaire constituée de ses différents voisins accessibles et du coût de la liaison.
- Quand un routeur reçoit une nouvelle table, il effectue les traitements suivants pour chaque entrée de la table reçue :
 - Si l'entrée n'est pas dans sa table, il la rajoute.
 - Si le coût de la route proposée par la table plus le coût de la route pour aller jusqu'au routeur (émetteur de la table) est supérieur au coût indiqué dans sa table, sa table de routage est modifiée pour prendre en compte cette nouvelle route.
 - Sinon, il n'y a pas de changement.
- La modification d'une entrée dans la table d'un routeur engendre l'émission de la nouvelle table sur tous les ports du routeur.

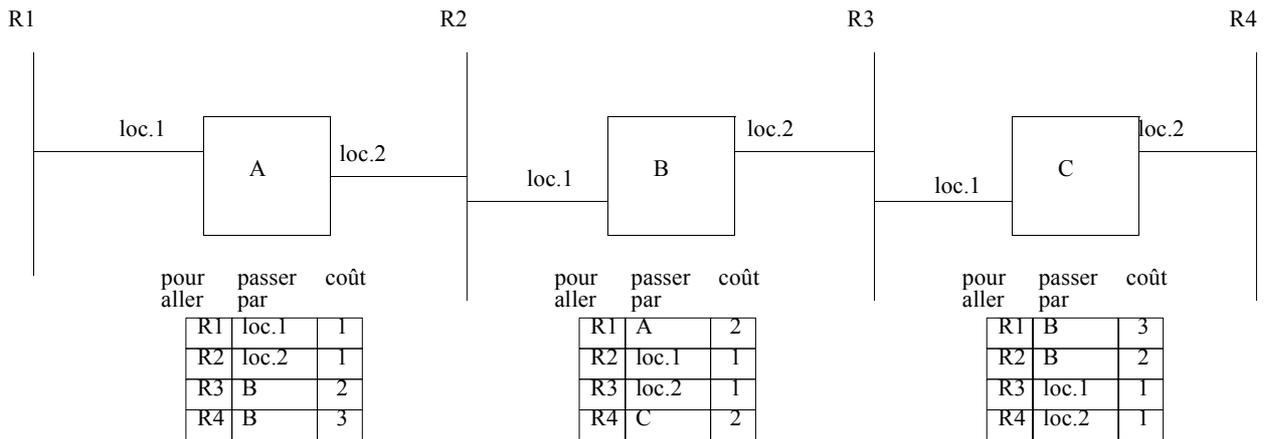
† Les échanges entre les routeurs continuent jusqu'à ce que l'algorithme converge

17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\ROUTAGE_Interne.fm

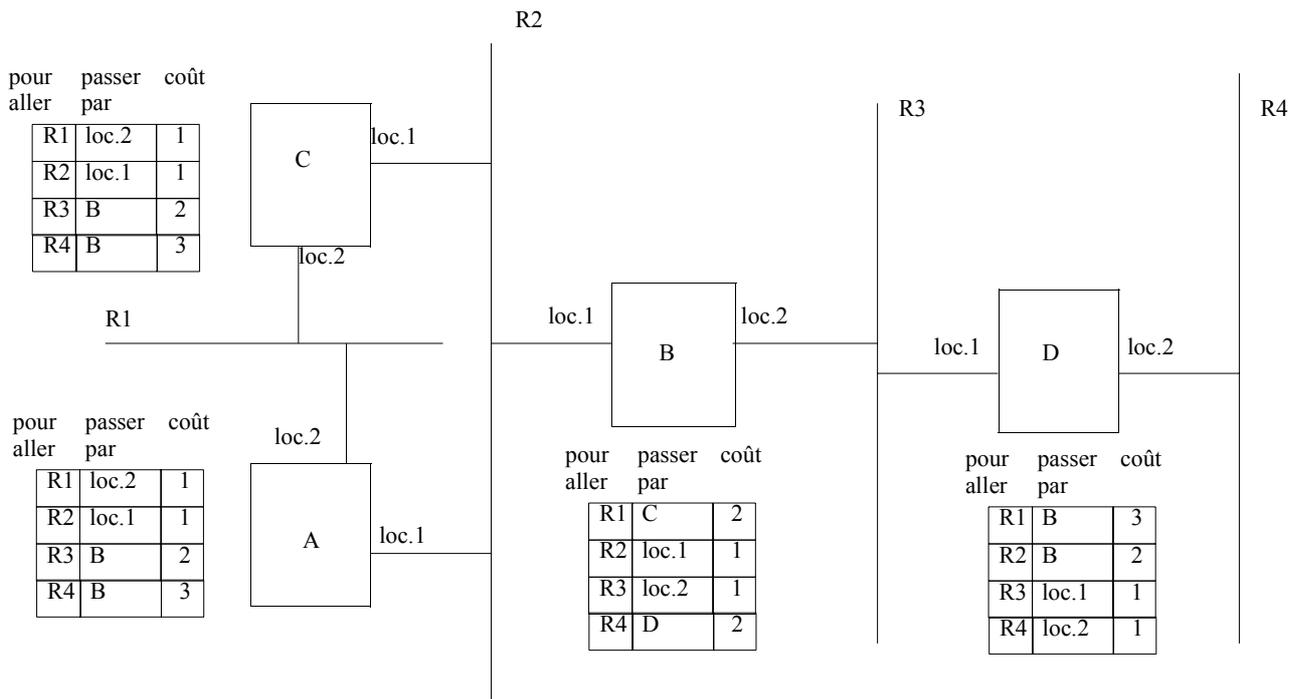
EXEMPLE DE DÉROULEMENT DE L'ALGORITHME DU DISTANT VECTOR



PROBLÈME DE CONVERGENCE



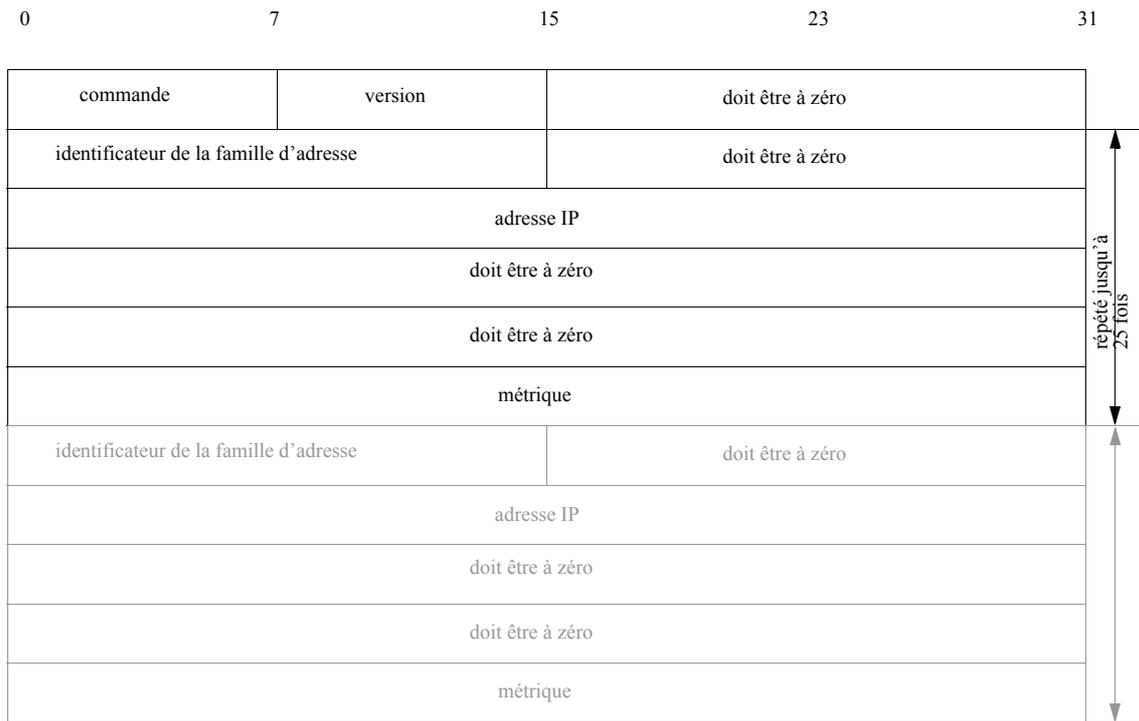
HORIZON COUPÉ



RIP

- Algorithme de type Distant Vector
- Implémenté sur la plupart des systèmes (sur BSD : routed)
- Limité à 15 routeurs
- Au dessus de UDP, port 520
- A servi de base pour de nombreux autres protocoles
 - AppleTalk, Novel / IPX, Banyan Vines ...
- Broadcasts toutes les 30 secondes

FORMAT DES PAQUETS

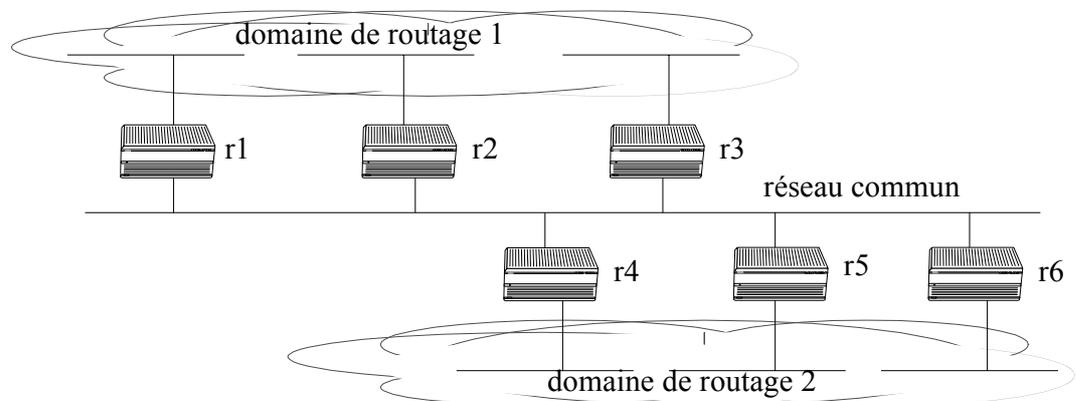


RIP

228

LES EXTENSIONS DE RIP

- RIP-2 (Janvier 1993)
 - Utilisation d'une adresse de Multicast (224.0.0.9)
 - Prise en compte des *netmask*
 - Plusieurs plans de routage (*route tag*)



- Le champ prochain routeur indique la destination qui sera insérée dans la table de routage du destinataire.

RIP

229

RIP-II

0	7	15	23	31
commande	version	domaine de routage		
identificateur de la famille d'adresse		route Tag		
adresse IP				
masque de sous-réseau				
prochain routeur				
métrique				

répété jusqu'à
25 fois

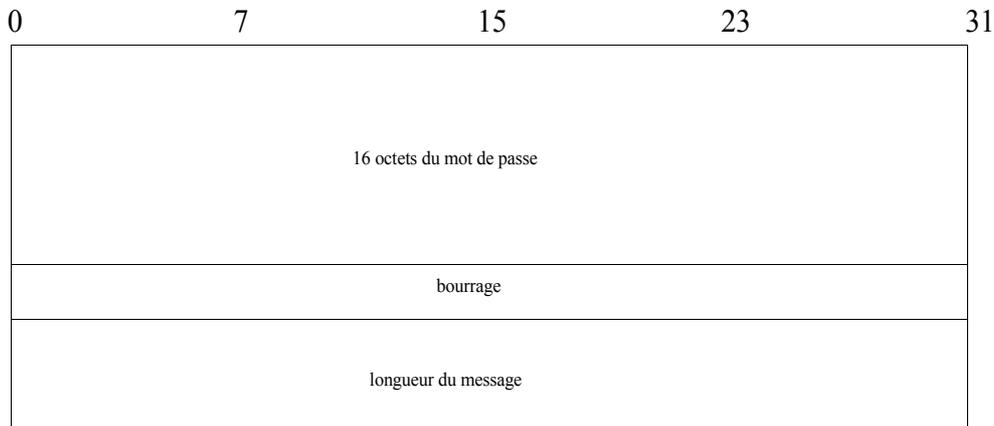
AUTHENTIFICATION

- Authentification simple : mot de passe
 - Peu fiable (découverte par écoute du réseau)
 - Permet surtout de se prémunir contre les équipements mal configurés

0	7	15	23	31
commande	version	domaine de routage		
0xFF		type d'authentification		
longueur du mot de passe		identifiant	longueur de l'identifiant	
numéro de séquence				
réservé				
réservé				
.....				
0xFF		0x1		
domaine de routage				

AUTHENTIFICATION PAR SECRET COMMUN

- Authentification pas sceau MD5
 - Contre le rejeu
 - Le mot de passe (secret) n'est jamais transmis sur le réseau
 - Les champs suivants sont ajoutés lors du calcul du sceau.



INTERIOR GATEWAY ROUTING PROTOCOL (IGRP)

- Propriétaire : Cisco
- Broadcasts de mise à jour toutes les 90 secondes
- Mesures de protection contre les boucles de routage
- Routage des sous-réseaux
- Routage multi-chemins (secours + load balancing)
 - La table de routage peut avoir plus d'un chemin pour atteindre une même destination
 - Chaque chemin = adresse_prochain_routeur, N°_interface
 - Avantages :
 - secours automatique si un chemin devient indisponible
 - partage du trafic entre les chemins de métriques voisines
- Gestion des routes par défaut (candidates)

IGRP (SUITE)

- La métrique composite d'IGRP
 - la métrique tient compte de :
 - la bande passante (B)
 - le délai de propagation (D)
 - la charge de la liaison (C)
 - et sa fiabilité (F)
 - => B et D sont des caractéristiques statiques de la liaison.
 - => C et F sont des variables dynamiques calculées par le routeur.
- Avantages :
 - affiner le routage en fonction des besoins, des coûts des ressources, de la redondance ...
 - en faisant varier les composantes de la métrique.
- Evolution :
 - Extended IGRP (E-IGRP)

OSPF (OPEN SHORTEST PATH FIRST)

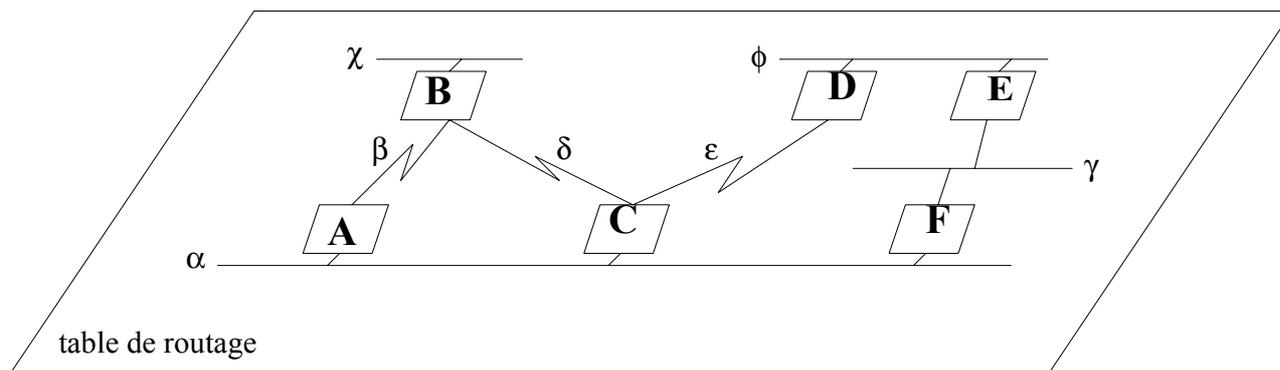
- + Pour de grands réseaux,
- + Force à structurer le réseau,
- + Poids attribuable aux liens, généralement fonction du débit :

$$\text{coût} = \frac{10^8}{\text{bande passante en b/s}}$$

- Pour un Ethernet à 10 Mbit/s le coût est de 10

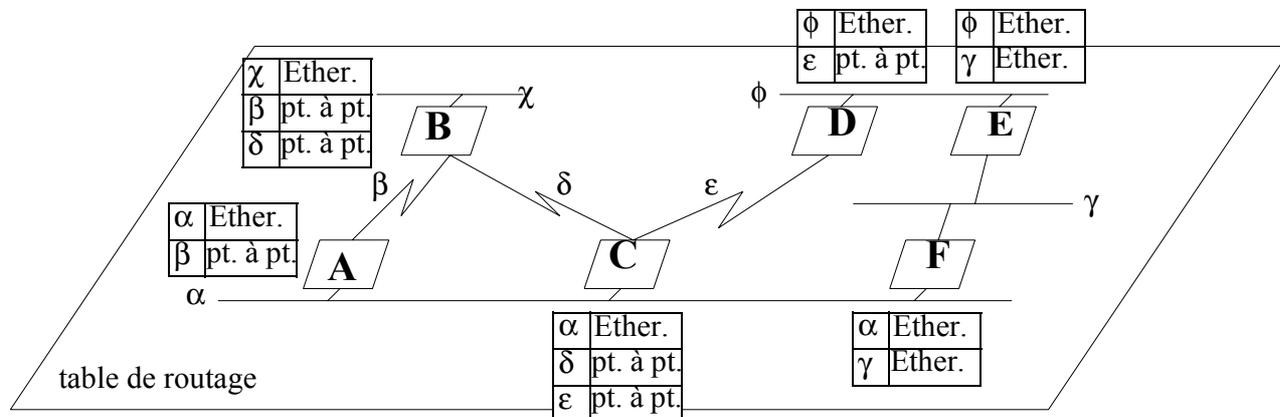
- + Utilise l'algorithme du Link State

EXEMPLE



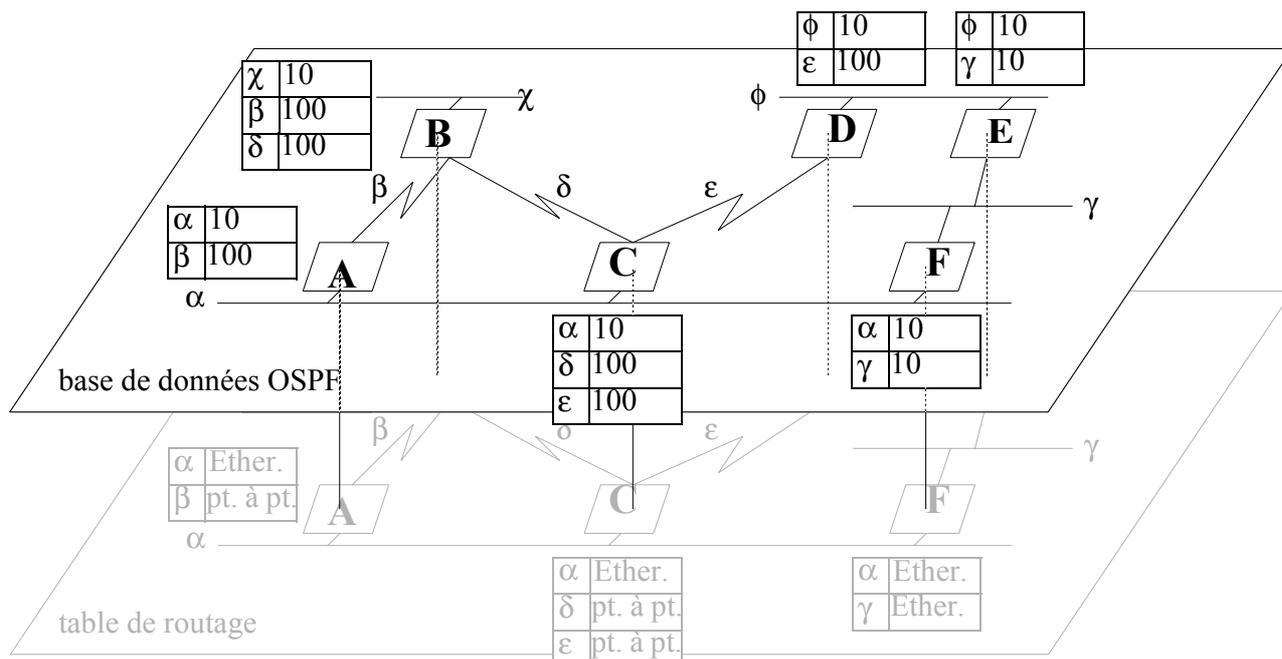
- Chaque routeur connaît uniquement les réseaux auquel il est connecté

TABLES DE ROUTAGE INITIALES



- Ces informations sont recopiées dans la base de donnée OSPF du routeur avec le coût choisi par l'administrateur

BASE DE DONNÉES OSPF



- Ces informations sont recopiées dans la base de donnée OSPF du routeur avec un coût.

INONDATION

- Un mécanisme d'inondation fiable permet de faire partager à tous les routeurs les informations de chaque routeur
- Après un temps de convergence, chaque routeur dispose des mêmes informations :

A	B	C	D	E	F
α 10	χ 10	α 10	ϕ 10	ϕ 10	α 10
β 100	β 100	δ 100	ϵ 100	γ 10	γ 10
	δ 100	ϵ 100			

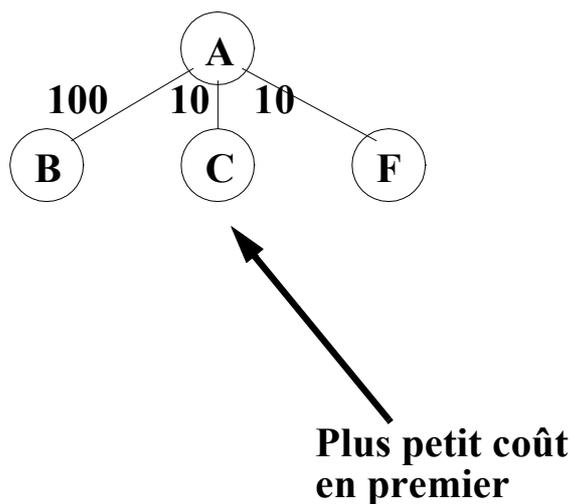
- Chaque routeur exécute l'algorithme du plus court chemin (SPF : Short Path First) pour trouver le plus court chemin pour aller des réseaux de son routeur vers les autres réseaux.

ALGORITHME DU SHORT PATH FIRST

Chaque routeur calcule sa table de routage à la réception de paquets LSP. A l'initialisation mettre le routeur dans la variable PATH puis :

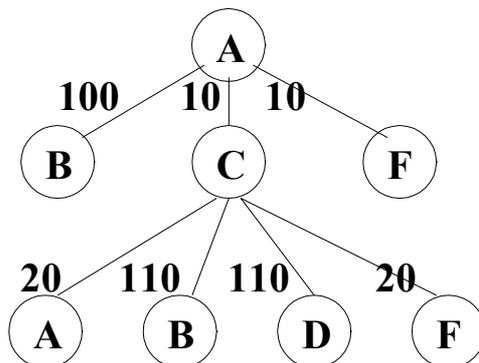
- Pour chaque routeur N contenu dans la structure PATH, examiner les LSP (c'est-à-dire les voisins immédiats) de N.
 - Pour chaque voisin M de N ajouter le coût de la liaison de la racine jusqu'à N au coût de la liaison de N à M.
 - Si M n'est ni dans la structure PATH ni dans la structure TENT avec un meilleur coût, insérer M avec le coût calculé et la direction N dans TENT.
- Si TENT est vide, l'algorithme est terminé, sinon prendre dans TENT l'entrée qui a le coût minimum, la mettre dans PATH et redérouler l'algorithme au début.

EXEMPLE POUR LE ROUTEUR A



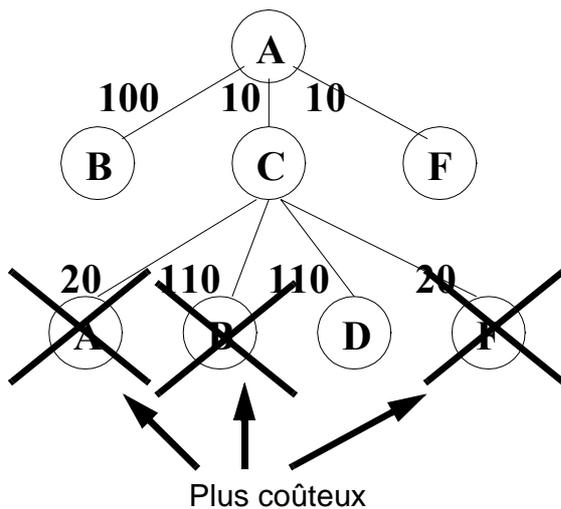
- A regarde les coûts pour ses voisins immédiats

EXEMPLE POUR LE ROUTEUR A

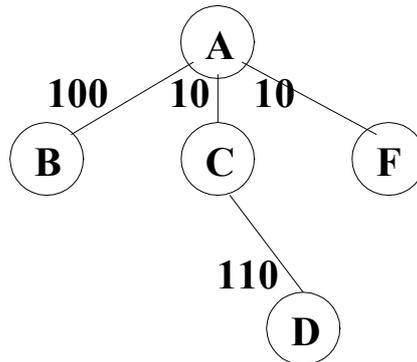


- A regarde les coûts cumulés pour les voisins immédiats de C

EXEMPLE POUR LE ROUTEUR A

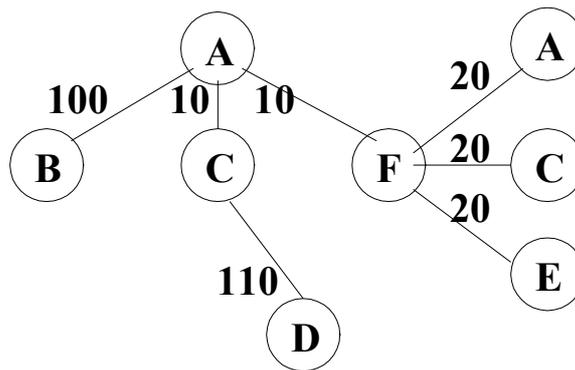


EXEMPLE POUR LE ROUTEUR A

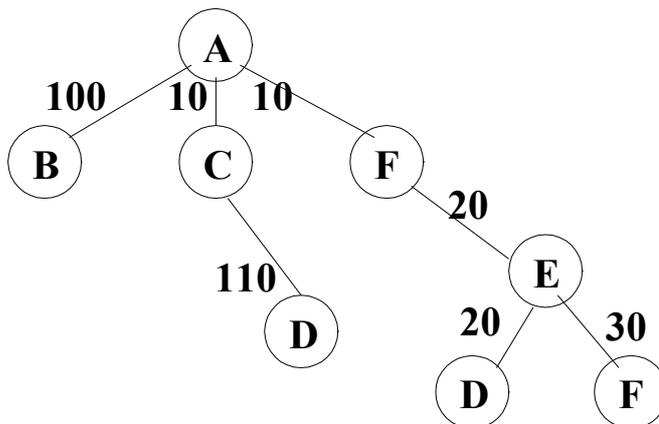


- A s'intéresse maintenant au plus petit chemin de l'arbre non encore exploré
- A calcule des coûts cumulés pour les voisins immédiats de F

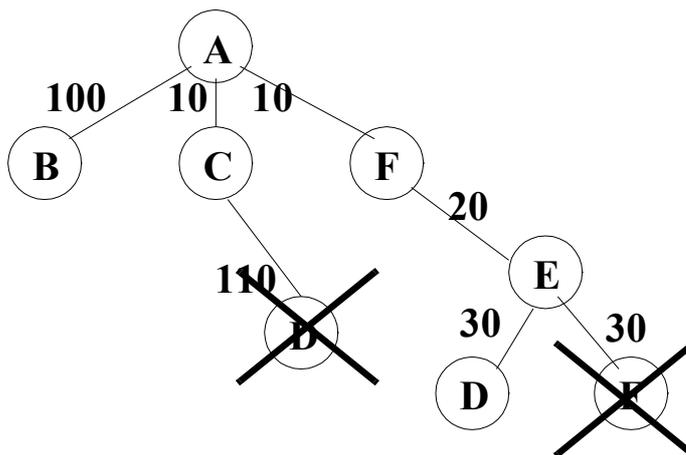
EXEMPLE POUR LE ROUTEUR A



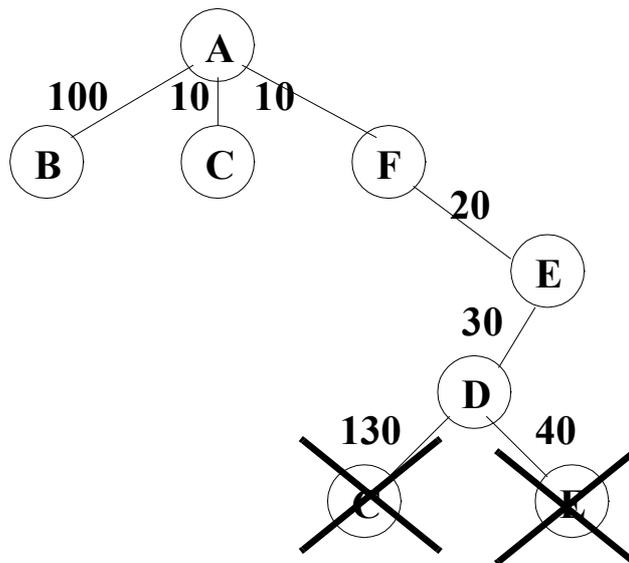
EXEMPLE POUR LE ROUTEUR A



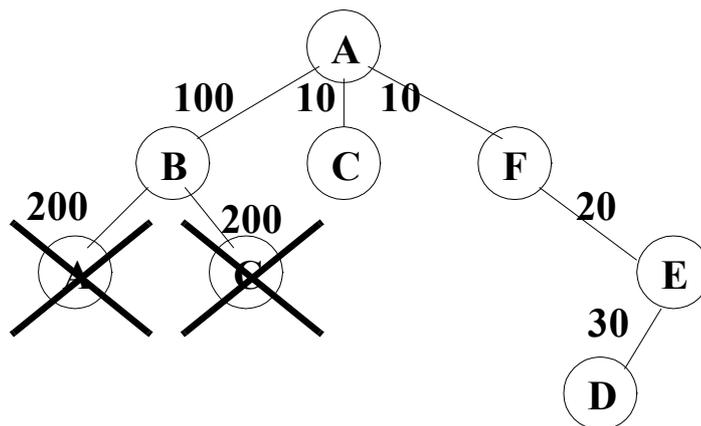
EXEMPLE POUR LE ROUTEUR A



EXEMPLE POUR LE ROUTEUR A

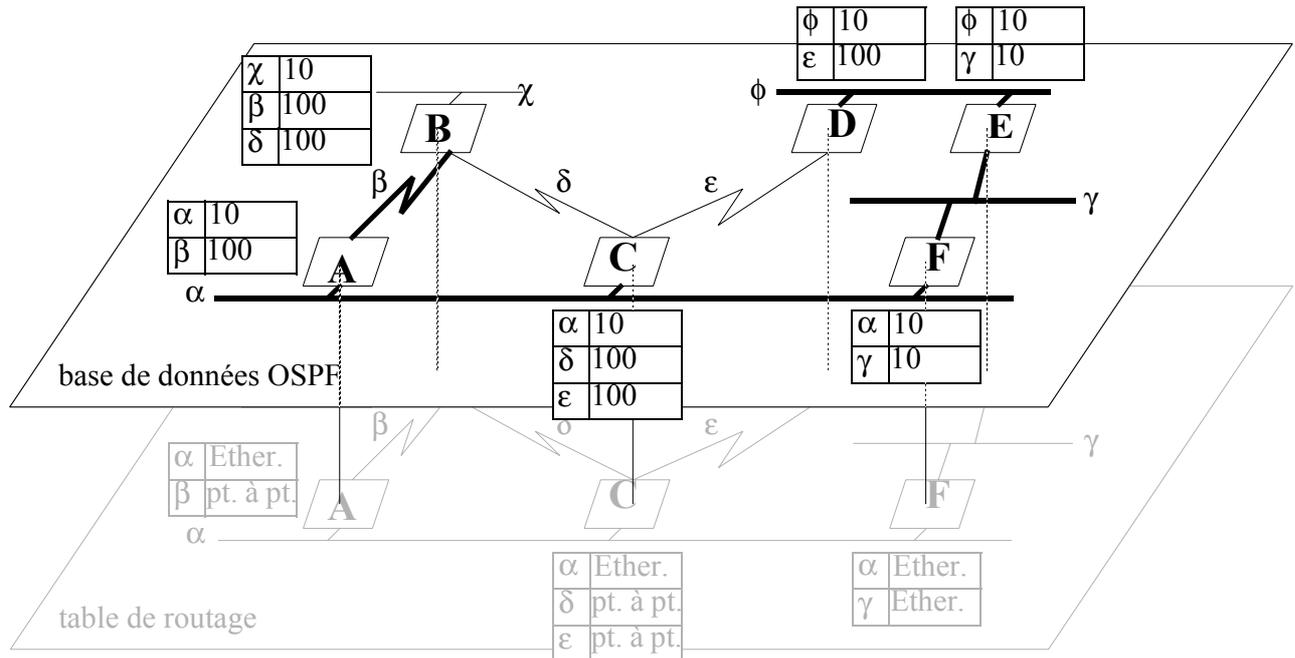


EXEMPLE POUR LE ROUTEUR A



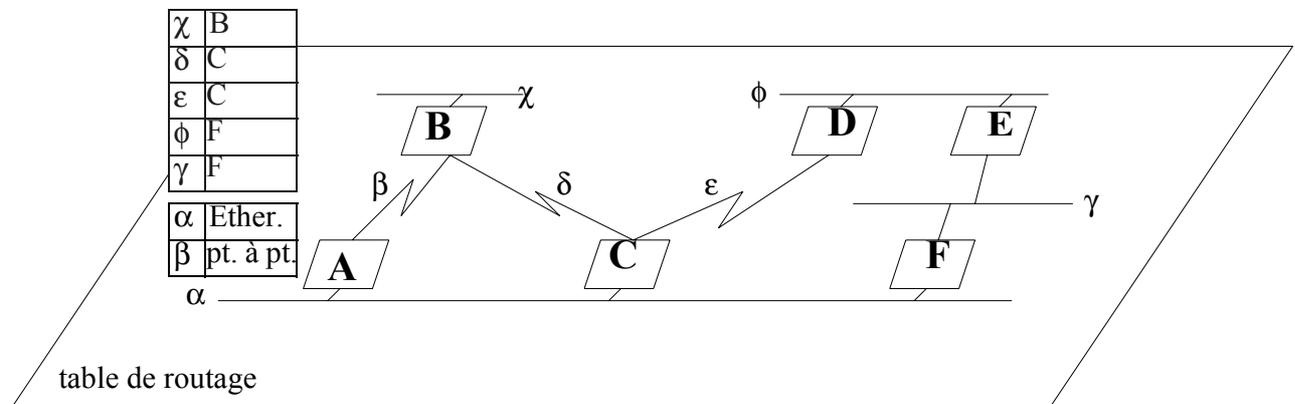
- A en déduit l'arbre des plus courts chemins

ARBRE DES PLUS COURTS CHEMINS POUR A



- A peut donc calculer sa table de routage

MISE À JOUR DE LA TABLE DE ROUTAGE DE A



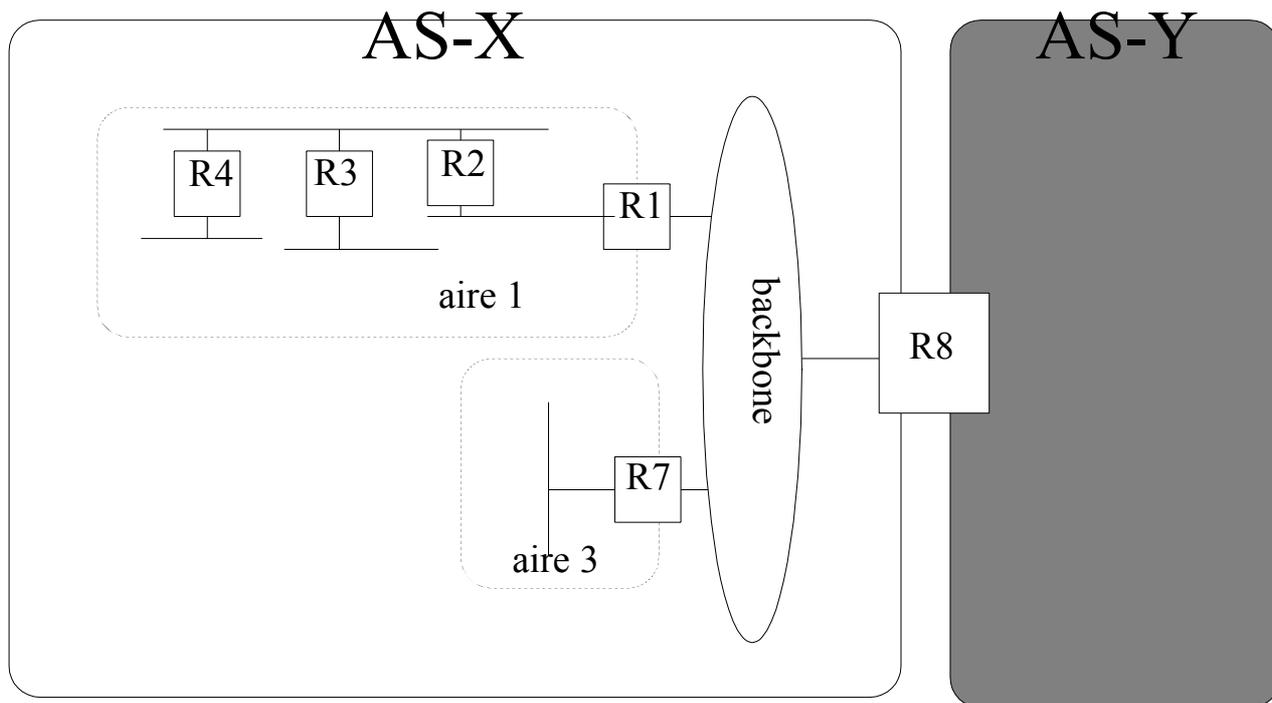
? Faites la même chose pour le routeur F

Principes :

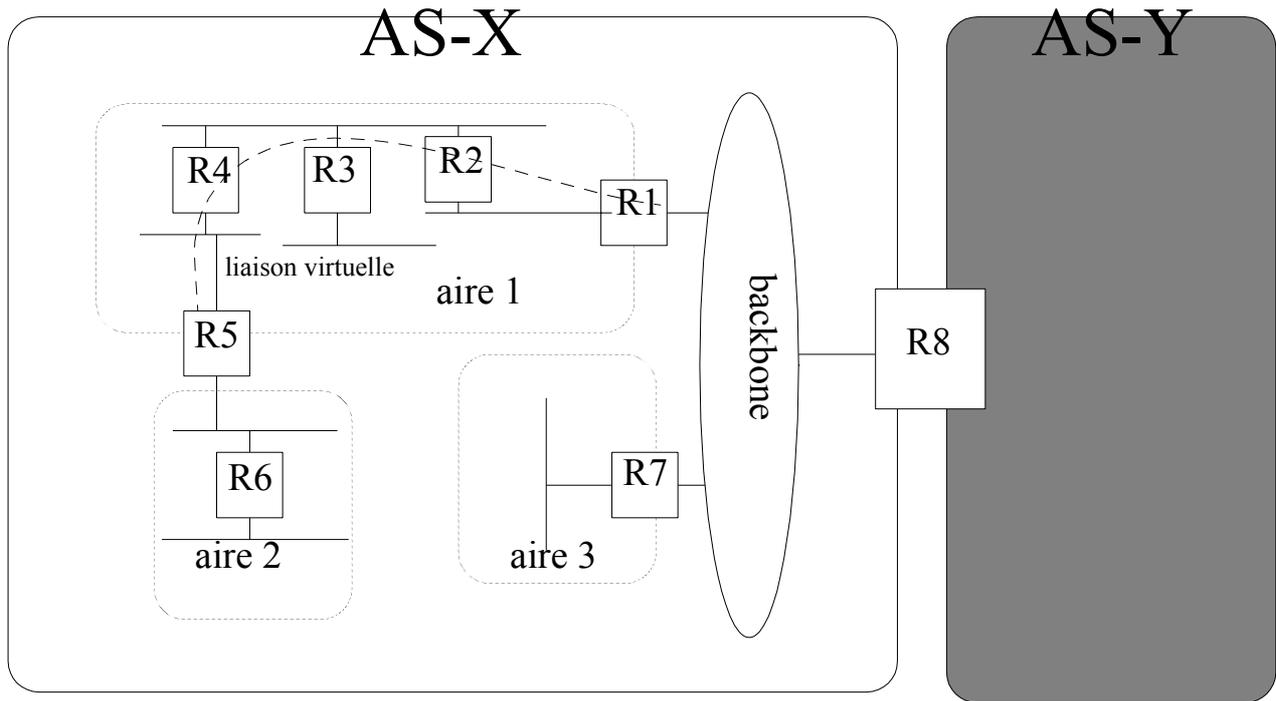
- Faire connaître à tous les routeurs les visions locales
 - Protocole d'inondation évitant les boucles,
 - Permettre les mises à jours quand l'état d'un routeur évolue.
- Calculer une route en fonction de cette vue générale
 - Algorithme du link-state
 - Algorithme coûteux en temps de calcul
 - Doit être relancé après chaque modification d'un routeur

+ Pour éviter des instabilités et un trop grand échange d'informations

- Division en aires : moins d'instabilités, moins d'échange de routes
- Définition d'un plan d'adressage conforme : possibilité d'agrégation

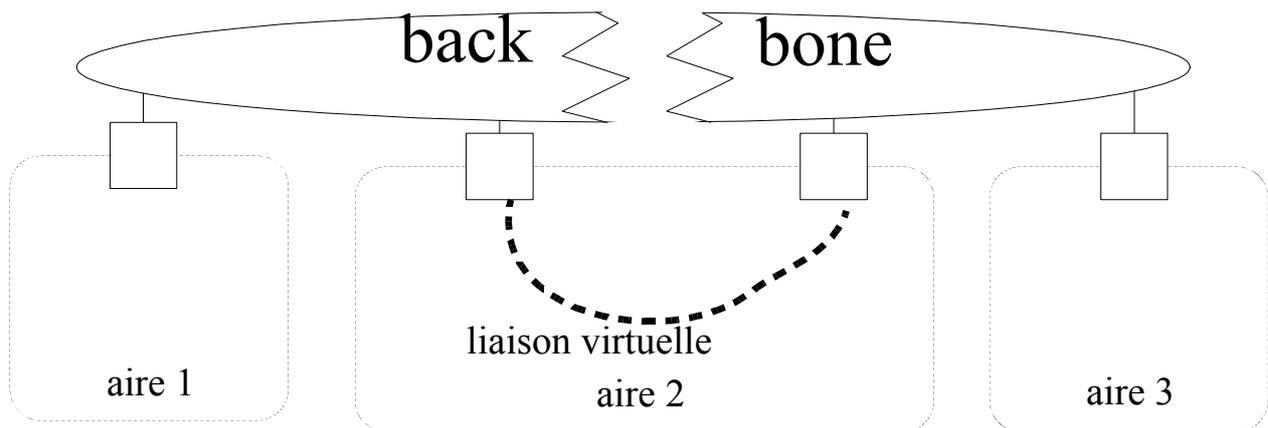


LIAISON VIRTUELLE



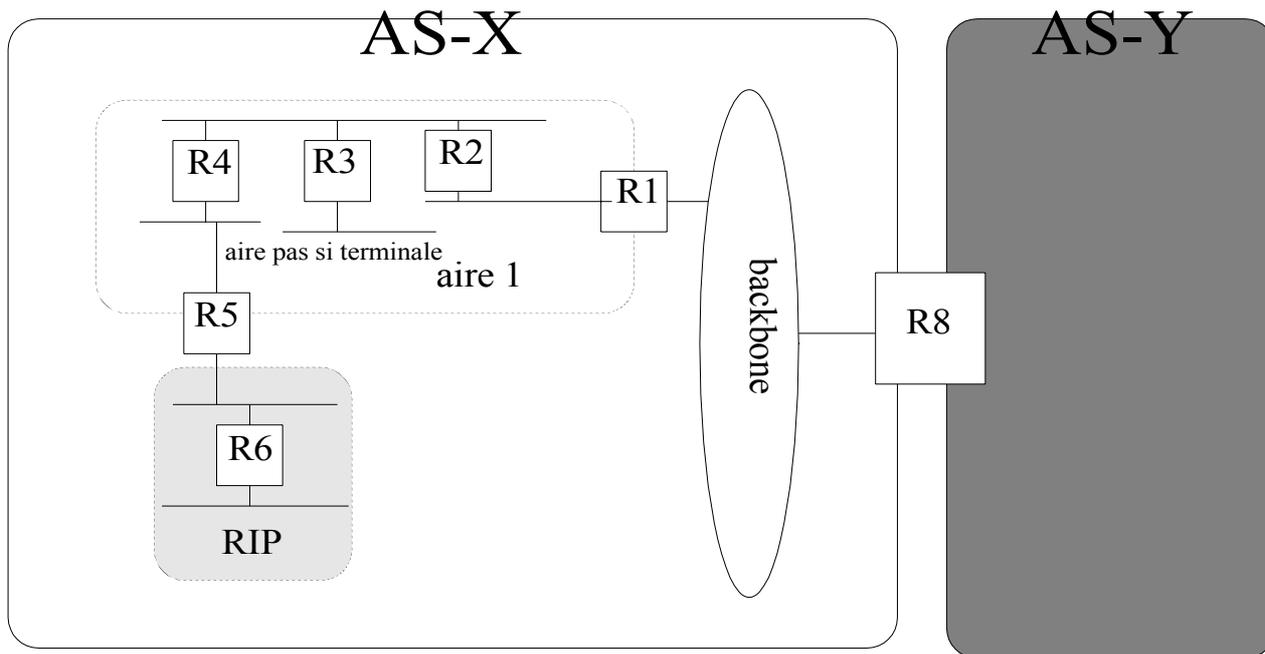
- A déconseiller à cause du manque de stabilité des routes

LIAISON VIRTUELLE



- Le backbone doit toujours être connexe
- Le lien virtuel peut servir à le réparer

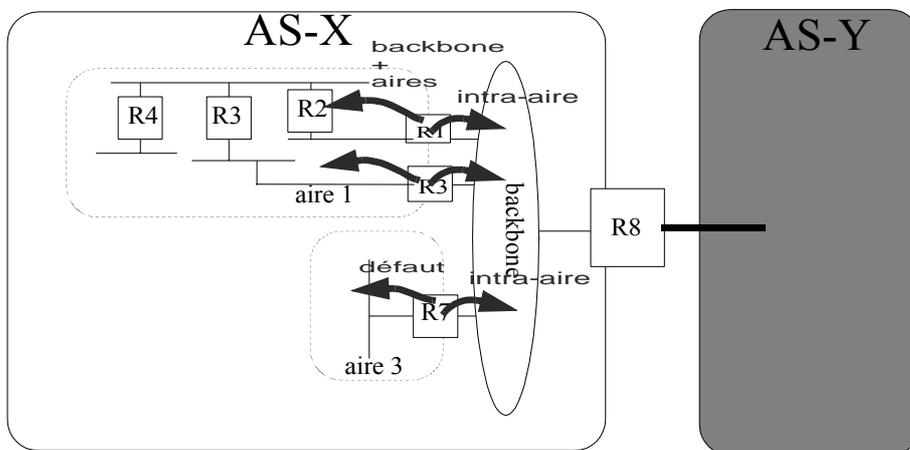
AIRE PAS SI TERMINALE



- Mais qui utilise encore RIP dans ce genre de configuration ?

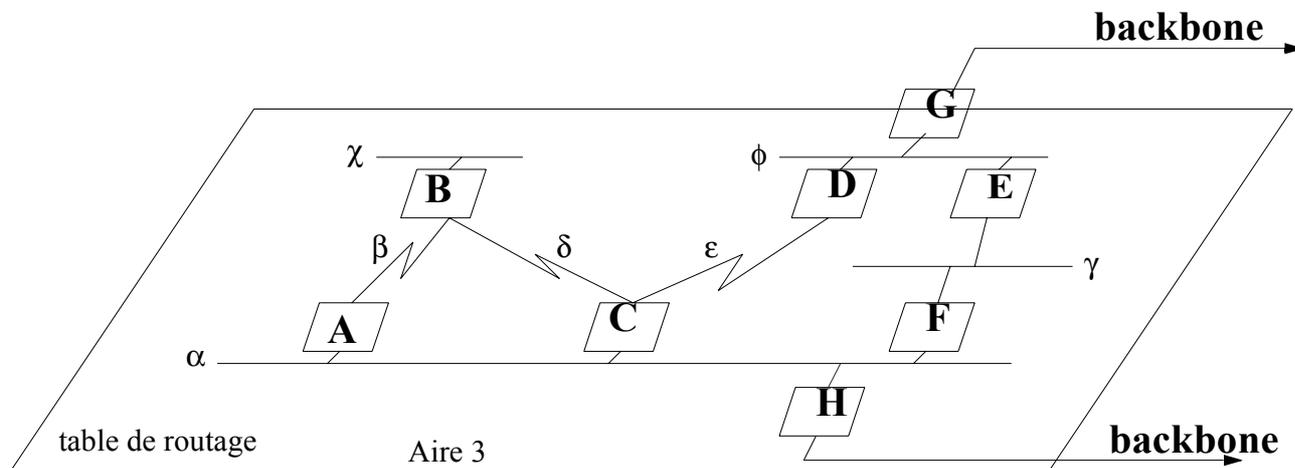
ROUTES ENTRE LES AIRES

- Des réseaux dans les autres aires :
 - Contient la liste de tous les réseaux contenus dans une aire.
 - Produit uniquement par le routeur en frontière d'Aire
 - Les routeurs à l'intérieur de l'aire ajoutent ces informations à leur table de routage



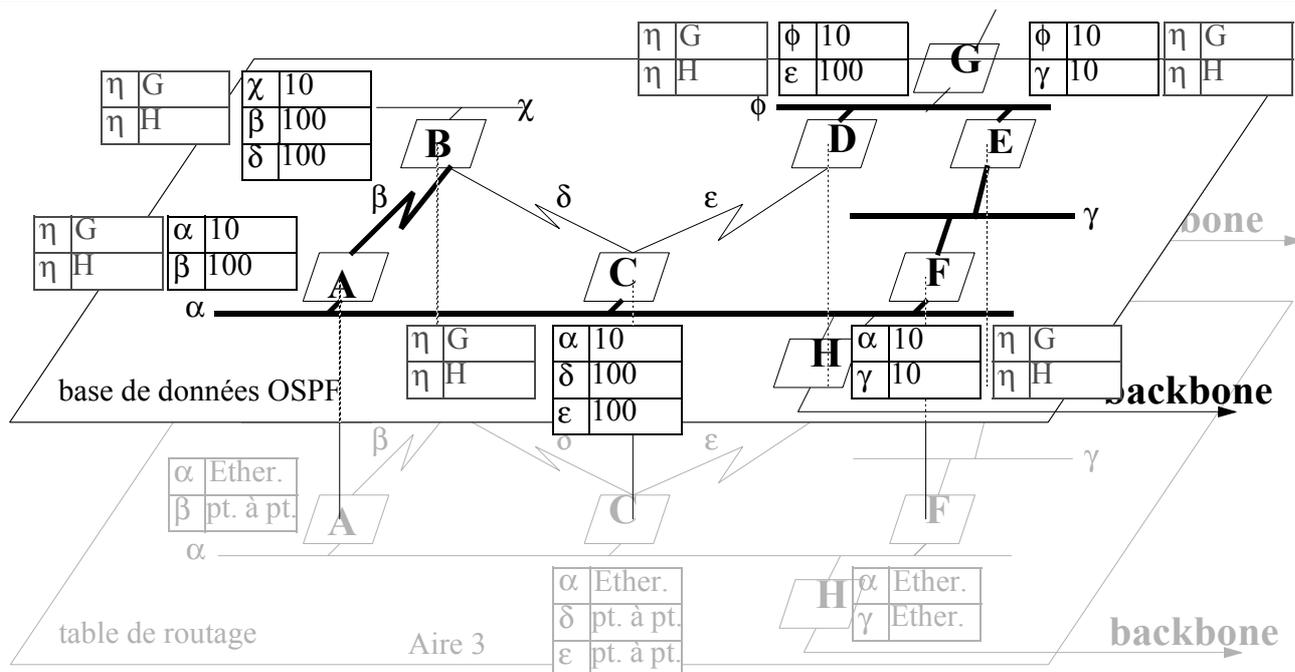
EXEMPLE

η : backbone



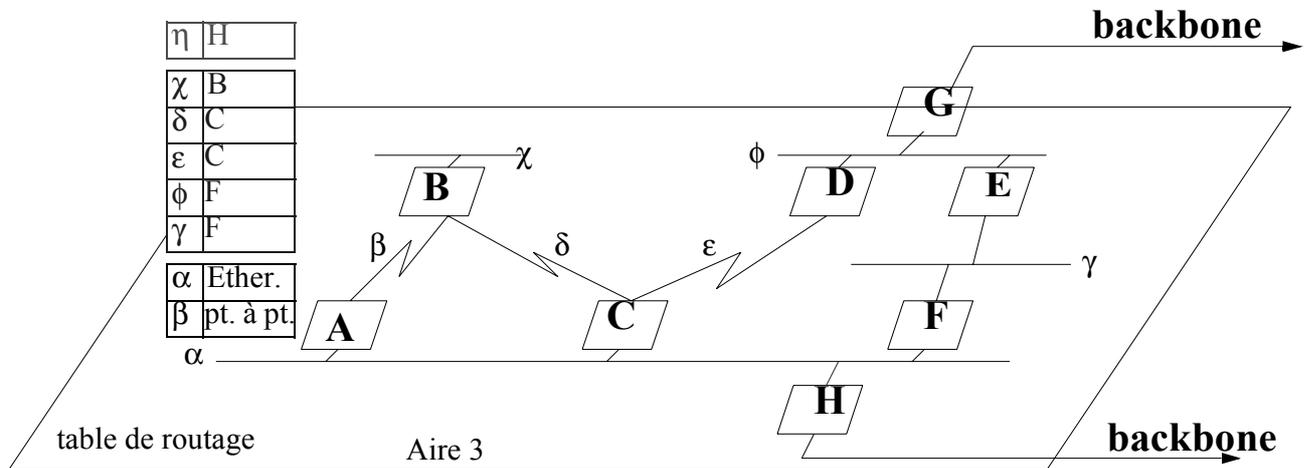
- Les routeurs G et H appartiennent au backbone, ils connaissent le réseau η .

DIFFUSION DES ROUTES EXTERIEURES À L'AIRE



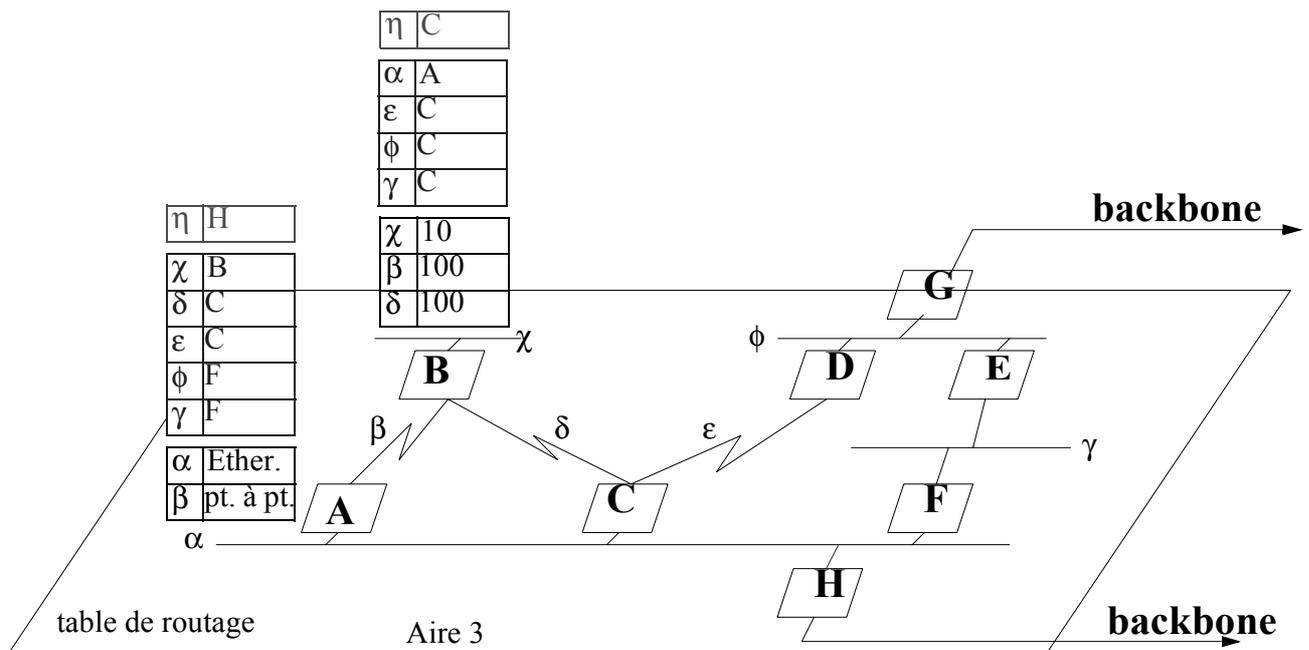
- Les routeurs calculent le meilleur chemin pour joindre les routeurs annonçant

MISE À JOUR DE LA TABLE DE ROUTAGE DE A



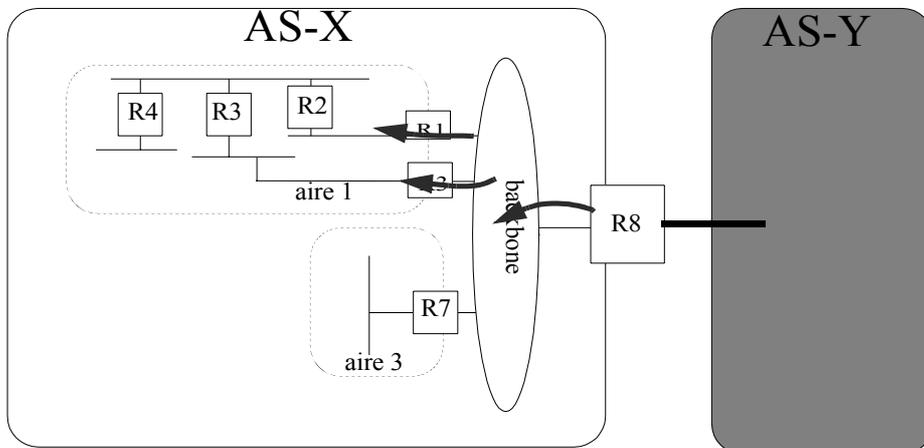
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Routing_Interne.fm

MISE À JOUR DE LA TABLE DE ROUTAGE DE B



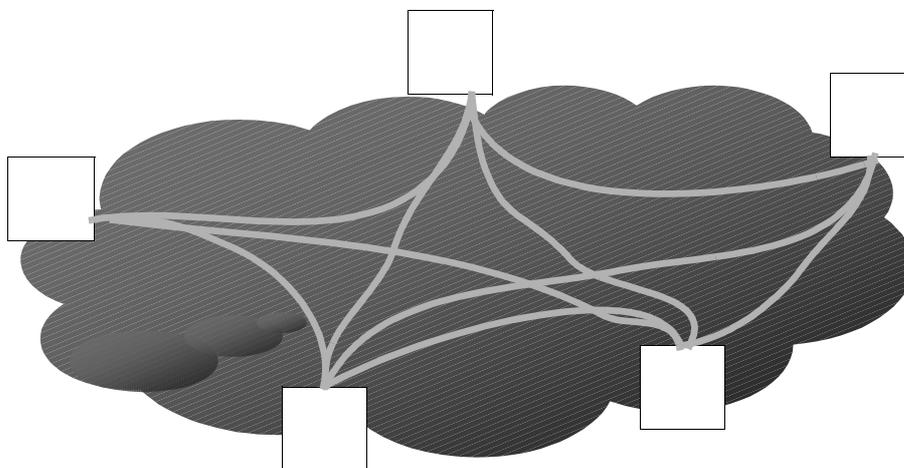
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Routing_Interne.fm

- Marquées différemment pour ne pas les réexporter
- Importées par les routeurs en frontière d'aire



- Deux types de routes :
 - type 1 (préfééré) : coût = interne + externe
 - type 2 : coût = externe

- Des routeurs du lien : pour les réseaux NBMA (Frame Relay, X.25,...)
 - permet de se synchroniser pour connaître tous les routeurs qui forment le lien,
 - détermine le routeur désigné,
 - doit connaître au moins un des éléments par configuration statique,



LE PROTOCOLE OSPF

- Meilleur que RIP
- Au dessus d'IP (proto = 89)
- Utilise les adresses de Multicast
 - 224.0.0.5 tous les routeurs du lien
 - 224.0.0.6 le routeur désigné et celui en secours
- Nombre illimité de routeurs
- Faible trafic (seules les modifications sont transmises)
- Authentification

Mais

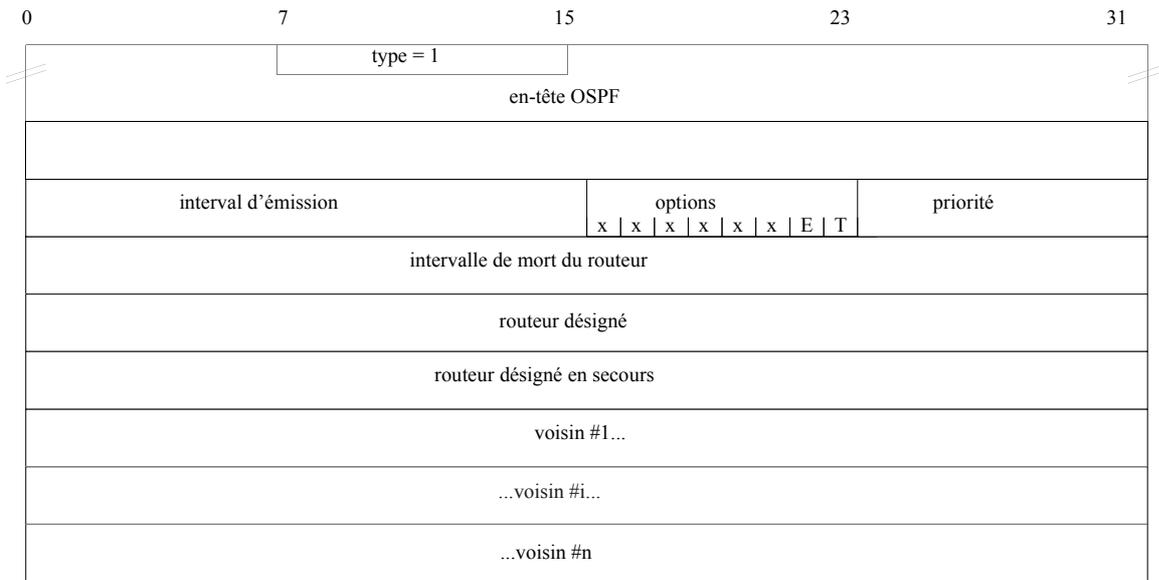
- Demande des routeurs puissants
- Difficile à configurer (nombreux paramètres)

FORMAT DES EN-TÊTE OSPF

0	7	15	23	31
version	type	longueur du paquet		
identité du routeur				
Indicateur de zone				
checksum		type d'authentificateur		
authentification				

- Paquet Hello : émis périodiquement pour identification des voisins
- Paquet description de la base de donnée : contient l'information sur la topologie du réseau
- Paquet demande, mise à jour, acquittement de l'état de la liaison : permettent de mettre à jour les tables de routeurs

PAQUET D'HELLO



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\ROUTAGE_Interne.fm

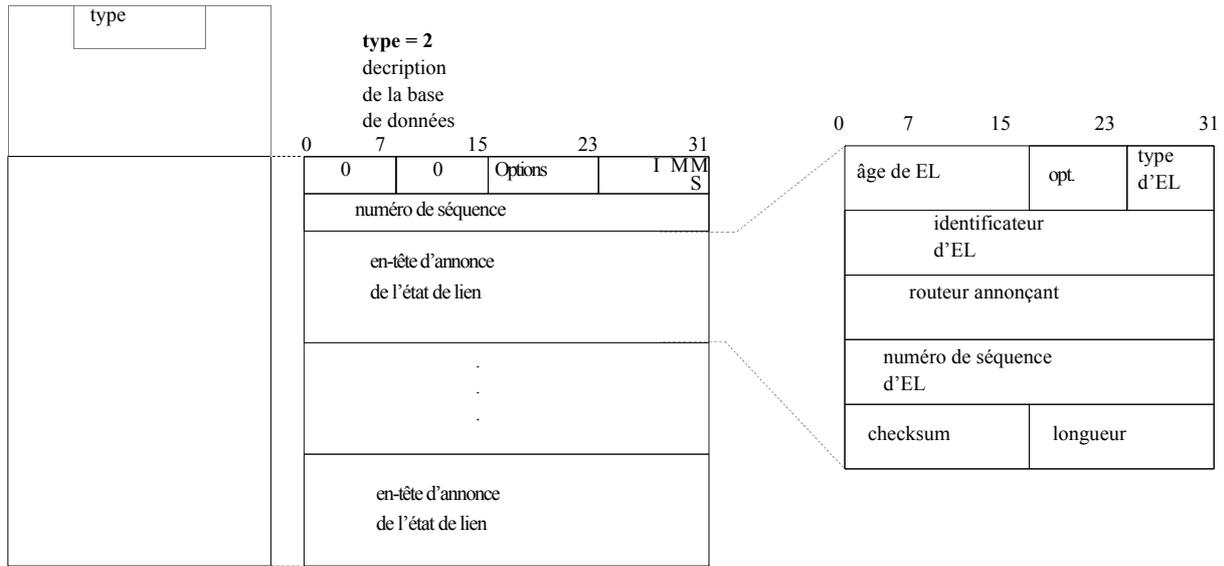
ALGORITHME DU LINK STATE

L'algorithme du *Link State* est le suivant :

- chaque routeur identifie ses voisins immédiats.
- un routeur principal (*designated router*) et de secours (*backup designated router*) sont désignés sur chaque réseau grâce à un mécanisme d'élection.
- le routeur acquière la base de donnée du routeur désigné.
- chaque routeur construit un message contenant la liste de ses voisins immédiats ainsi que le coût associé à la liaison. Ce message sera appelé LSP pour *Link State Packet*.
- ce paquet est transmis à tous les autres routeurs du réseau avec un mécanisme de diffusion qui limite la propagation des messages et évite les boucles.
- chaque routeur met à jour sa base de donnée ce qui lui donne une vision globale du réseau et il peut en déduire ses tables de routage.

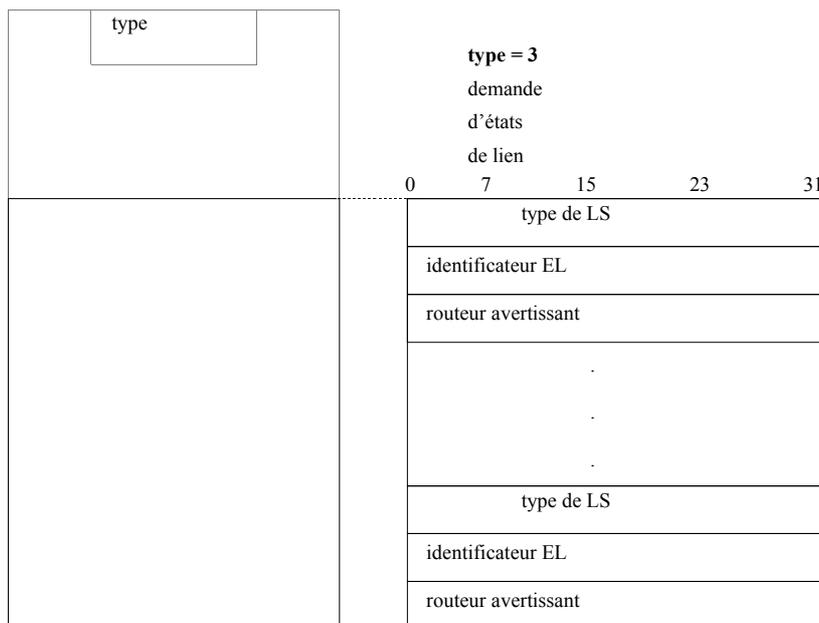
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\ROUTAGE_Interne.fm

PAQUET DE DESCRIPTION DE LA BASE DE DONNÉES



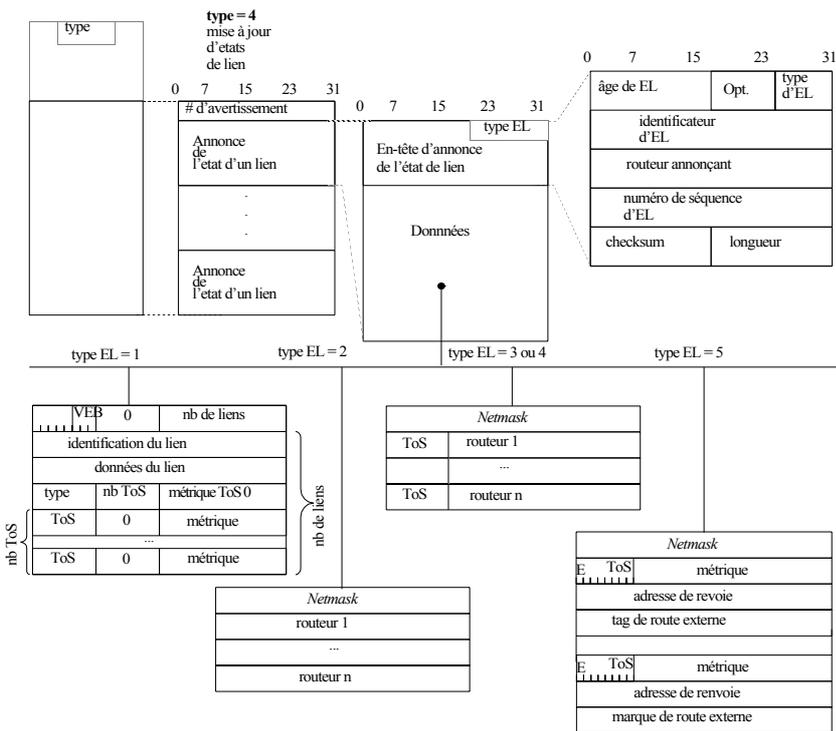
17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\ROUTAGE_Interne.fm

PAQUET DE DEMANDE D'ÉTAT DE LIEN



17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\ROUTAGE_Interne.fm

MISE À JOUR DE L'ÉTAT DE LIEN

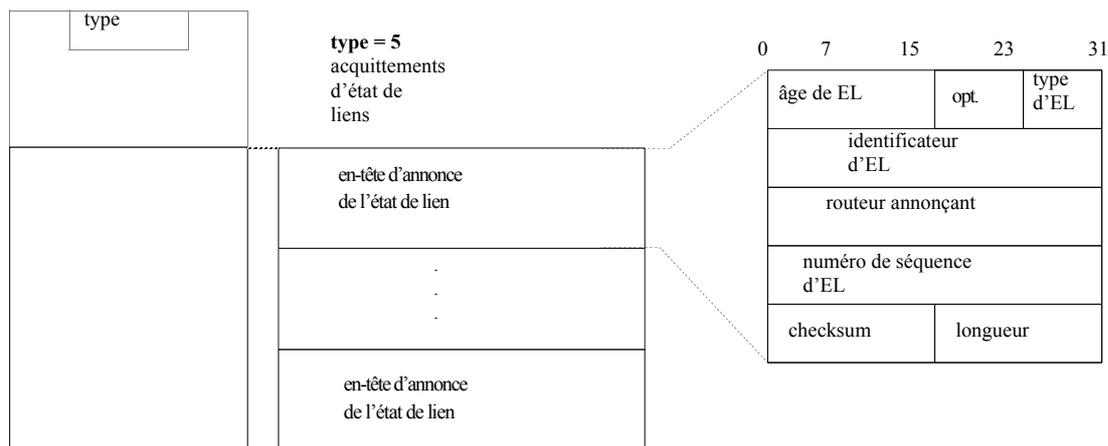


+ ces paquets inondent le réseau

+ la transmission doit être fiable

17 janvier 2003 - C:\Documents and Settings\ouain\mes documents\Support_ENST\ROUTAGE_Interne.fm

ACQUITTEMENT.



17 janvier 2003 - C:\Documents and Settings\ouain\mes documents\Support_ENST\ROUTAGE_Interne.fm

EXEMPLE

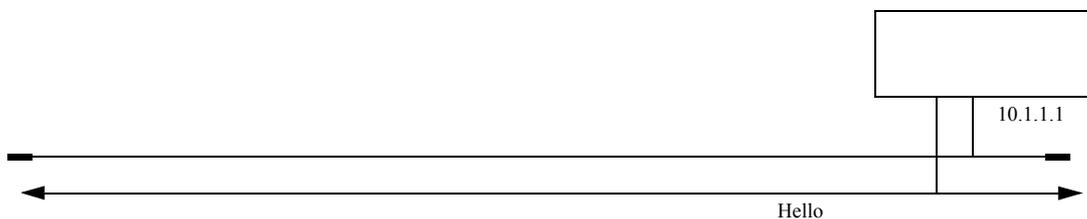


- On ne sait rien du réseau,
- Découverte de la topologie en écoutant le trafic OSPF

Exemple

272

EXEMPLE : HELLO (ÉQUIPEMENT ISOLÉ)



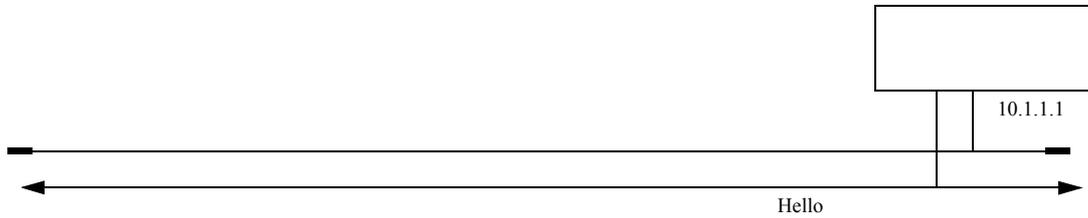
```
10.1.1.1 > 224.0.0.5:      OSPFv2-hello 44:
                           area 0.0.0.1 E mask 255.255.255.0 int 10 pri 5 dead 40 dr 10.1.1.1
                           nbrs
```

- Un routeur présent (10.1.1.1) de priorité 5
- Le préfixe est 10.1.1.0/24
- On se trouve dans l'aire 1
- Ce routeur ne connaît pas d'autres voisins sur ce lien

Exemple

273

EXEMPLE HELLO (ÉQUIPEMENT ISOLÉ)



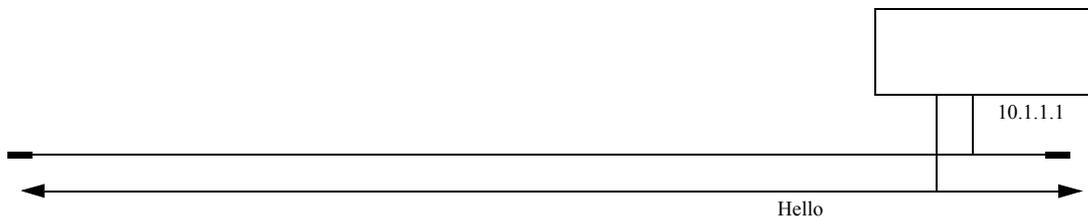
```
10.1.1.1 > 224.0.0.5:      OSPFv2-hello 44:
                           area 0.0.0.1 E mask 255.255.255.0 int 10 pri 5 dead 40 dr 10.1.1.1
                           nbrs
```

- La période d'annonce des Hello est de 10 secondes
- Si on ne reçoit pas de message Hello pendant 40 seconde, le routeur est considéré comme mort.
- Priorité 5 utilisée pendant l'élection du routeur désigné

Exemple

274

EXEMPLE : HELLO (ÉQUIPEMENT ISOLÉ)



```
10.1.1.1 > 224.0.0.5:      OSPFv2-hello 44:
                           area 0.0.0.1 E mask 255.255.255.0 int 10 pri 5 dead 40 dr 10.1.1.1
                           nbrs
```

```
10.1.1.1 > 224.0.0.5:      OSPFv2-hello 44:
                           area 0.0.0.1 E mask 255.255.255.0 int 10 pri 5 dead 40 dr 10.1.1.1
                           nbrs
```

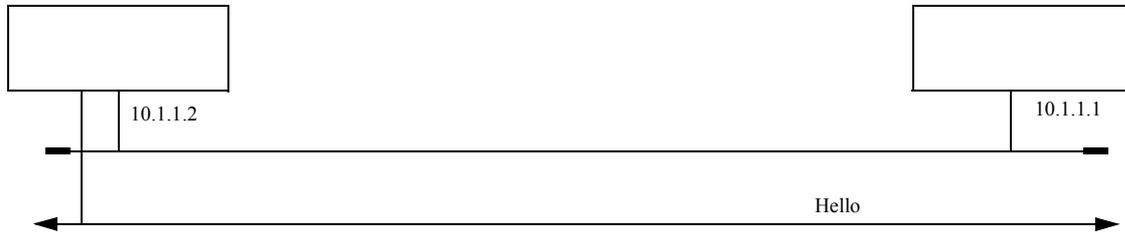
```
10.1.1.1 > 224.0.0.5:      OSPFv2-hello 44:
                           area 0.0.0.1 E mask 255.255.255.0 int 10 pri 5 dead 40 dr 10.1.1.1
                           nbrs
```

....

Exemple

275

EXEMPLE : ECHANGE DE HELLO



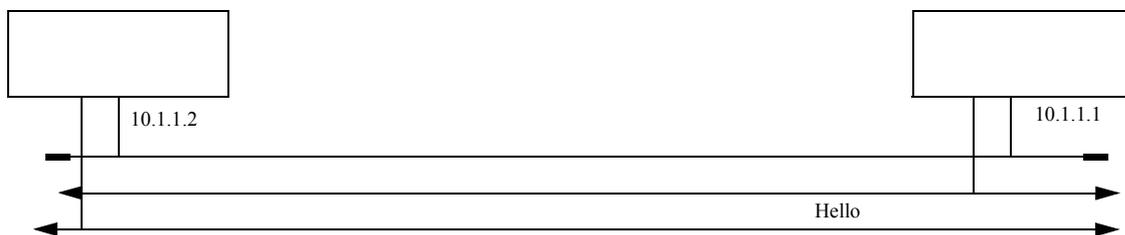
```
10.1.1.2 > 224.0.0.5:      OSPFv2-hello 44:  
                           area 0.0.0.1 E mask 255.255.255.0 int 10 pri 27 dead 40  
                           nbrs
```

- Un deuxième routeur émet un message Hello

Exemple

276

EXEMPLE : ECHANGE DE HELLO



```
10.1.1.1 > 224.0.0.5:      OSPFv2-hello 48:  
                           area 0.0.0.1 E mask 255.255.255.0 int 10 pri 5 dead 40  
                           dr 10.1.1.1 nbrs 10.1.1.2
```

- Le routeur 10.1.1.1 devient routeur désigné
- Le routeur 10.1.1.2 apparaît dans sa liste de voisins

Exemple

277

EXEMPLE HELLO

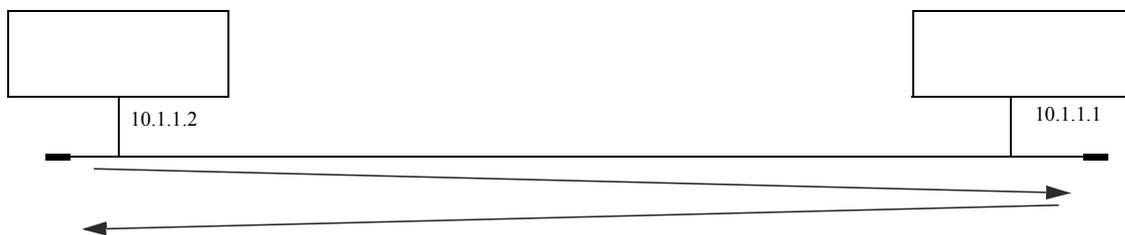


- La connectivité est bi-directionnelle
- Le routeur 10.1.1.2 peut se synchroniser avec le routeur désigné.

Exemple

278

EXEMPLE : DESCRIPTION DE LA BASSE DE DONNÉES



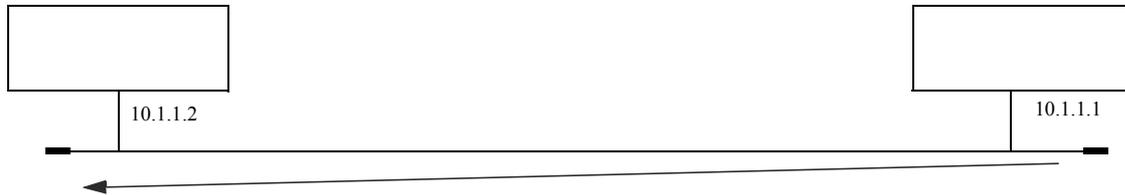
```
10.1.1.2 > 10.1.1.1: OSPFv2-dd 32: area 0.0.0.1 E I/M/MS S B  
10.1.1.1 > 10.1.1.2: OSPFv2-dd 32: area 0.0.0.1 E I/M/MS S 1973
```

- 10.1.1.2 a comme numéro de séquence B
- 10.1.1.1 a comme numéro de séquence 1973
- I : premier paquet, M : d'autres paquets vont suivre, MS : maître.
- 10.1.1.2 va gagner (n° de séquence plus petit)

Exemple

279

EXEMPLE : DESCRIPTION DE LA BASE DE DONNÉES



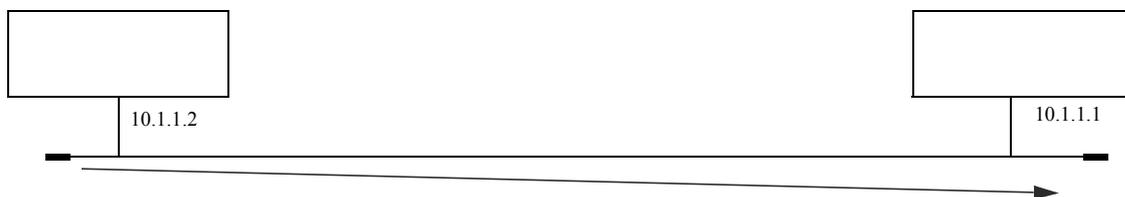
```
10.1.1.1 > 10.1.1.2:      OSPFv2-dd 112: area 0.0.0.1 E M S B
                          { E S 80000002 age 3:09 rtr 10.1.1.1 }           TYPE 1
                          { E S 80000001 age 2:49 sum 10.1.2.0 abr 10.1.1.1 } TYPE 3
                          { E S 80000003 age 2:44 sum 10.1.100.0 abr 10.1.1.1 } TYPE 3
                          { E S 80000001 age 2:59 abr 10.1.1.1 rtr 10.1.1.1 } TYPE 4
```

- 10.1.1.1 décrit sa base de donnée :
 - TYPE 1 : liaison que connaît le routeur, TYPE 3 : réseaux présents à l'exterieur de l'aire, TYPE 4 : routeurs faisant l'agrégation pour cette aire.

Exemple

280

EXEMPLE : DESCRIPTION DE LA BASE DE DONNÉES



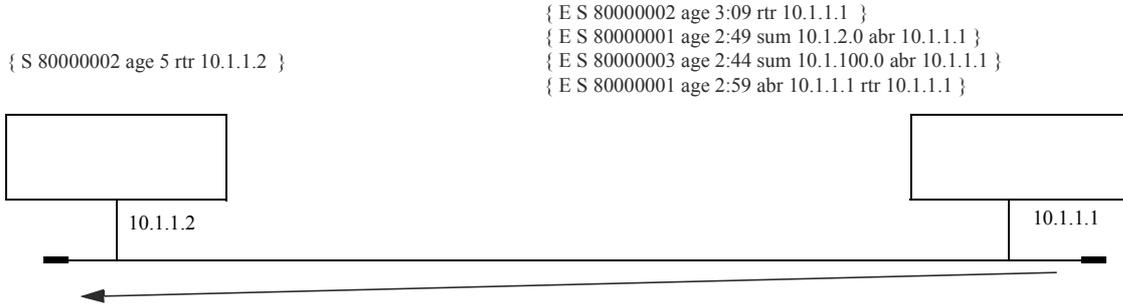
```
10.1.1.2 > 10.1.1.1:      OSPFv2-dd 52: area 0.0.0.1 E M S S C
                          { S 80000002 age 5 rtr 10.1.1.2 }           TYPE 1
```

- 10.1.1.2 acquitte le message précédent en incrémentant le champ séquence et décrit ses connaissances :
 - les réseaux auquel il est attaché.
- plus d'autres informations (le bit M n'est plus présent)

Exemple

281

EXEMPLE : DESCRIPTION DE LA BASE DE DONNÉES



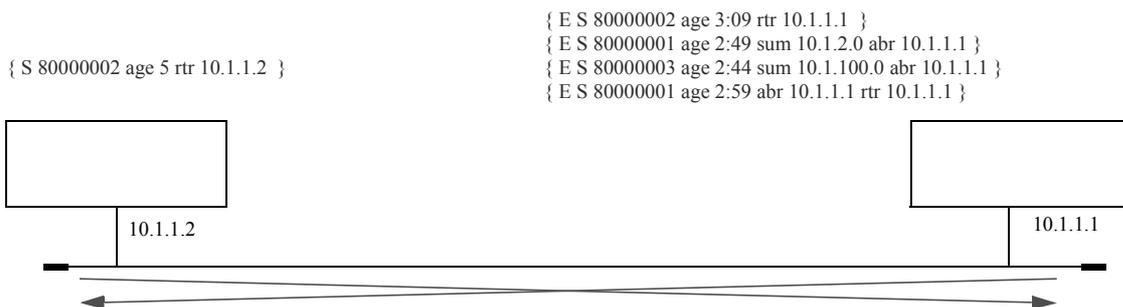
15:14:50.472224 10.1.1.1 > 10.1.1.2: OSPFv2-dd 32: area 0.0.0.1 E S C

- 10.1.1.1 n'a plus rien à décrire,
- fin de la phase de description de la base de données.
- chaque routeur sait ce que connaît l'autre.

Exemple

282

EXEMPLE : DEMANDE D'INFORMATION



10.1.1.2 > 10.1.1.1: OSPFv2-ls_req 72: area 0.0.0.1
 { rtr 10.1.1.1 } { sum 10.1.2.0 abr 10.1.1.1 } { sum 10.1.100.0 abr 10.1.1.1 }
 { abr 10.1.1.1 rtr 10.1.1.1 }

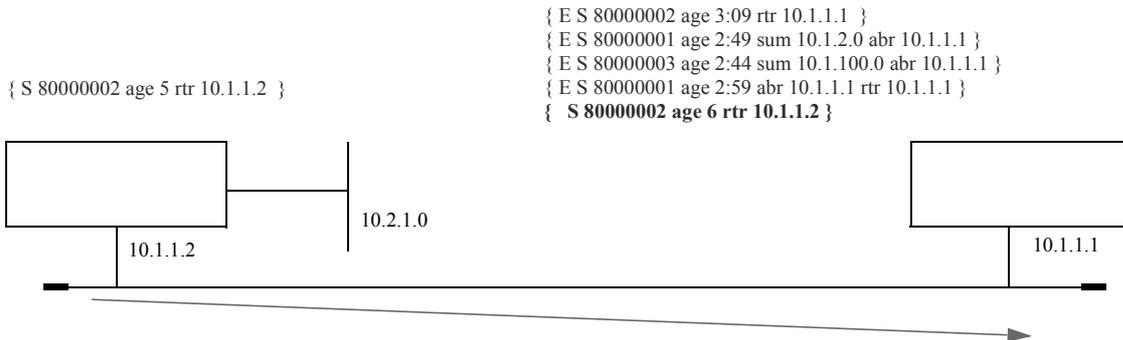
10.1.1.1 > 10.1.1.2: OSPFv2-ls_req 36: area 0.0.0.1
 { rtr 10.1.1.2 }

- Les routeurs veulent les informations qui :
 - leur manquent ou
 - qui sont plus récentes (pas dans cet exemple)

Exemple

283

EXEMPLE : MISE À JOUR DE LA BASE DE DONNÉES



```

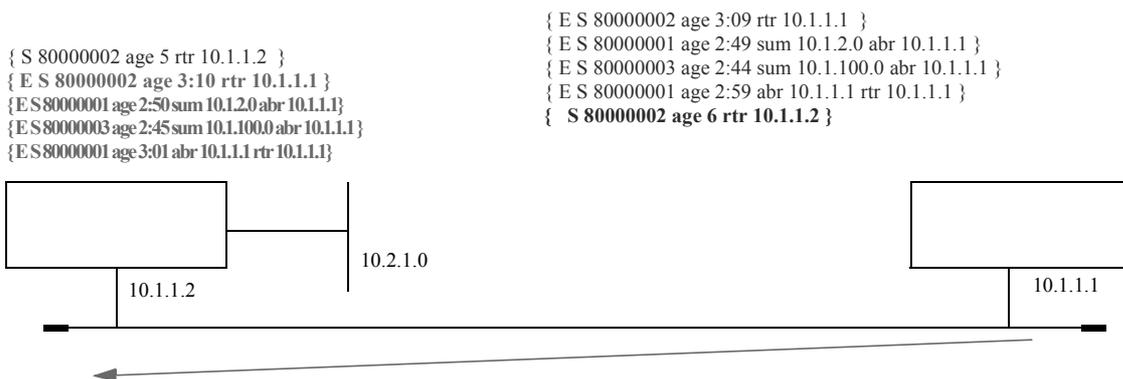
10.1.1.2 > 10.1.1.1: OSPFv2-ls_upd 76: area 0.0.0.1
{ S 80000002 age 6 rtr 10.1.1.2
  { net 10.1.1.0 mask 255.255.255.0 tos 0 metric 1 }
  { net 10.2.1.0 mask 255.255.255.0 tos 0 metric 1 } }
    
```

- Enregistrement complet du type 1 décrivant les attachements locaux du routeur
- L'âge de l'enregistrement a évolué

Exemple

284

EXEMPLE : MISE À JOUR DE LA BASE DE DONNÉES



```

10.1.1.1 > 10.1.1.2: OSPFv2-ls_upd 148: area 0.0.0.1
{ E S 80000002 age 3:10 rtr 10.1.1.1 B
  { net 10.1.1.0 mask 255.255.255.0 tos 0 metric 10 } }
{ E S 80000001 age 2:50 sum 10.1.2.0 abr 10.1.1.1 mask 255.255.255.0 tos 0 metric 20 }
{ E S 80000003 age 2:45 sum 10.1.100.0 abr 10.1.1.1 mask 255.255.255.0 tos 0 metric 10 }
{ E S 80000001 age 3:01 abr 10.1.1.1 rtr 10.1.1.1 tos 0 metric 16777215 }
    
```

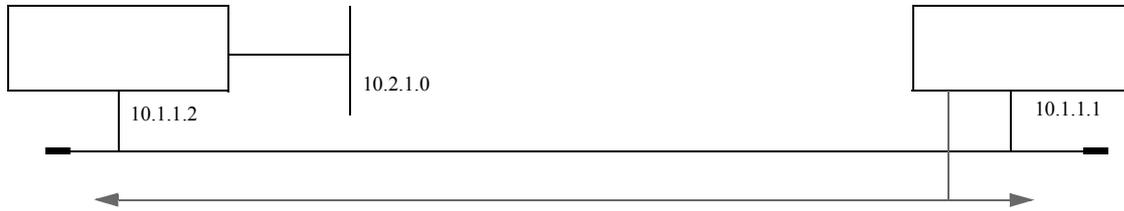
- Informations du routeur 10.1.1.1 :
 - Un seul attachement dans l'aire 1 et deux réseaux (10.1.2.0/24 et 10.1.100.0/24) à l'extérieur de l'aire.

Exemple

285

EXEMPLE : MISE À JOUR DE LA BASE DE DONNÉES

```
{ S 80000002 age 5 rtr 10.1.1.2 }
{ E S 80000002 age 3:10 rtr 10.1.1.1 }
{ ES 80000001 age 2:50 sum 10.1.2.0 abr 10.1.1.1 }
{ ES 80000003 age 2:45 sum 10.1.100.0 abr 10.1.1.1 }
{ ES 80000001 age 3:01 abr 10.1.1.1 rtr 10.1.1.1 }
{ E S 80000002 age 3:09 rtr 10.1.1.1 }
{ E S 80000001 age 2:49 sum 10.1.2.0 abr 10.1.1.1 }
{ E S 80000003 age 2:44 sum 10.1.100.0 abr 10.1.1.1 }
{ E S 80000001 age 1:00:00 abr 10.1.1.1 rtr 10.1.1.1 }
{ S 80000002 age 6 rtr 10.1.1.2 }
```



```
10.1.1.1 > 224.0.0.5: OSPFv2-ls_upd 56: area 0.0.0.1
{ E S 80000002 age 1:00:00 abr 10.1.1.1 rtr 10.1.1.1 tos 0 metric 16777215 }
```

- 10.1.1.1 rend l'enregistrement obsolète

EXEMPLE : MISE À JOUR DE LA BASE DE DONNÉES

```
{ S 80000003 age 1 rtr 10.1.1.2 }
{ E S 80000002 age 3:10 rtr 10.1.1.1 }
{ ES 80000001 age 2:50 sum 10.1.2.0 abr 10.1.1.1 }
{ ES 80000003 age 2:45 sum 10.1.100.0 abr 10.1.1.1 }
{ ES 80000001 age 3:01 abr 10.1.1.1 rtr 10.1.1.1 }
{ E S 80000002 age 3:09 rtr 10.1.1.1 }
{ E S 80000001 age 2:49 sum 10.1.2.0 abr 10.1.1.1 }
{ E S 80000003 age 2:44 sum 10.1.100.0 abr 10.1.1.1 }
{ E S 80000001 age 1:00:00 abr 10.1.1.1 rtr 10.1.1.1 }
{ S 80000002 age 6 rtr 10.1.1.2 }
```



```
10.1.1.2 > 224.0.0.5: OSPFv2-ls_upd 76: area 0.0.0.1
{ S 80000003 age 1 rtr 10.1.1.2
  { dr 10.1.1.1 if 10.1.1.2 tos 0 metric 1 }
  { net 10.2.1.0 mask 255.255.255.0 tos 0 metric 1 } }
```

- Le routeur émet une nouvelle version de son enregistrement

EXEMPLE : MISE À JOUR DE LA BASE DE DONNÉES

```
{ S 80000003 age 1 rtr 10.1.1.2 }
{ E S 80000002 age 3:10 rtr 10.1.1.1 }
{ E S 80000001 age 2:50 sum 10.1.2.0 abr 10.1.1.1 }
{ E S 80000003 age 2:45 sum 10.1.100.0 abr 10.1.1.1 }
{ E S 80000001 age 3:01 abr 10.1.1.1 rtr 10.1.1.1 }
{ E S 80000003 age 1 rtr 10.1.1.1 }
{ E S 80000001 age 2:49 sum 10.1.2.0 abr 10.1.1.1 }
{ E S 80000003 age 2:44 sum 10.1.100.0 abr 10.1.1.1 }
{ E S 80000001 age 1:00:00 abr 10.1.1.1 rtr 10.1.1.1 }
{ S 80000002 age 6 rtr 10.1.1.2 }
{ E S 80000001 age 1 net dr 10.1.1.1 }
```



```
10.1.1.1 > 224.0.0.5:      OSPFv2-ls_upd 64: area 0.0.0.1
                          { E S 80000003 age 1 rtr 10.1.1.1 B
                            { dr 10.1.1.1 if 10.1.1.1 tos 0 metric 10 } }
```

```
10.1.1.1 > 224.0.0.5:      OSPFv2-ls_upd 60: area 0.0.0.1                                TYPE 2
                          { E S 80000001 age 1 net dr 10.1.1.1 if 10.1.1.1 mask 255.255.255.0 rtrs 10.1.1.1 10.1.1.2 }
```

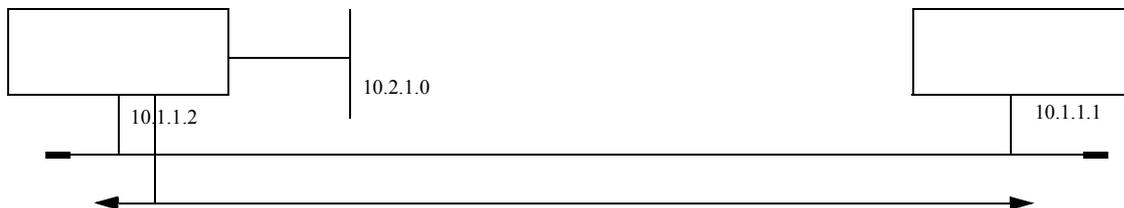
- Remet à jour l'enregistrement sur les liaisons du routeur.
- Emet un nouvel enregistrement de type 2.

Exemple

288

EXEMPLE

```
{ S 80000003 age 1 rtr 10.1.1.2 }
{ E S 80000002 age 3:10 rtr 10.1.1.1 }
{ E S 80000001 age 2:50 sum 10.1.2.0 abr 10.1.1.1 }
{ E S 80000003 age 2:45 sum 10.1.100.0 abr 10.1.1.1 }
{ E S 80000001 age 3:01 abr 10.1.1.1 rtr 10.1.1.1 }
{ E S 80000003 age 1 rtr 10.1.1.1 }
{ E S 80000001 age 2:49 sum 10.1.2.0 abr 10.1.1.1 }
{ E S 80000003 age 2:44 sum 10.1.100.0 abr 10.1.1.1 }
{ E S 80000001 age 1:00:00 abr 10.1.1.1 rtr 10.1.1.1 }
{ S 80000002 age 6 rtr 10.1.1.2 }
{ E S 80000001 age 1 net dr 10.1.1.1 }
```



```
10.1.1.2 > 224.0.0.5:      OSPFv2-hello 48:
                          area 0.0.0.1 E mask 255.255.255.0 int 10 pri 27 dead 40
                          dr 10.1.1.1 bdr 10.1.1.2 nbrs 10.1.1.1
```

Exemple

289

EXEMPLE DE CONFIGURATION OSPF

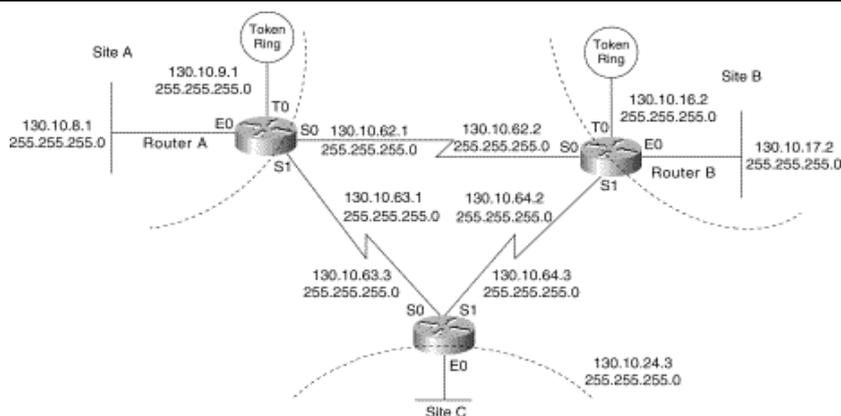
```

Router#sh conf
Using 508 out of 32762 bytes
!
version 11.3
no service password-encryption
!
hostname Router
!
enable secret 5 $1$IB2a$HxkclsFbROv7zvnIa8s6A.
enable password root
!
!
!
interface Ethernet0
 ip address 192.108.119.137 255.255.255.0
 ip ospf cost 4
!
interface Serial0
 no ip address
 no ip mroute-cache
 shutdown
!
interface Serial1
 no ip address
 shutdown
!
router ospf 1
 network 192.108.119.0 0.0.0.255 area 0.0.0.0
!
ip classless
!
line con 0
line aux 0
line vty 0 4
 password root
 login
!
end
  
```

Exemple de configuration OSPF

290

EXEMPLE PASSAGE DE RIP À OSPF



Router A :

```

interface serial 0
 ip address 130.10.62.1 255.255.255.0
interface serial 1
 ip address 130.10.63.1 255.255.255.0
interface ethernet 0
 ip address 130.10.8.1 255.255.255.0
interface tokenring 0
 ip address 130.10.9.1 255.255.255.0
  
```

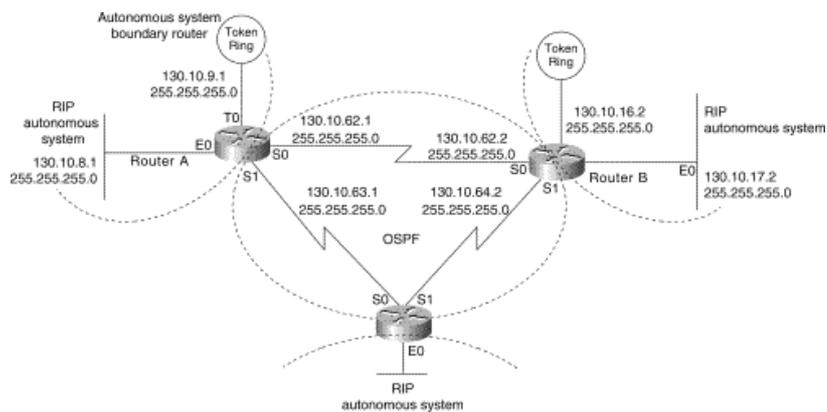
```

router rip
 network 130.10.0.0
  
```

Exemple de configuration OSPF

291

EXEMPLE PASSAGE DE RIP À OSPF



Router A :

! même définition des interfaces

```

router rip
default-metric 10
network 130.10.0.0
passive-interface serial 0
passive-interface serial 1
redistribute ospf 109 match internal external 1
external 2
!
!

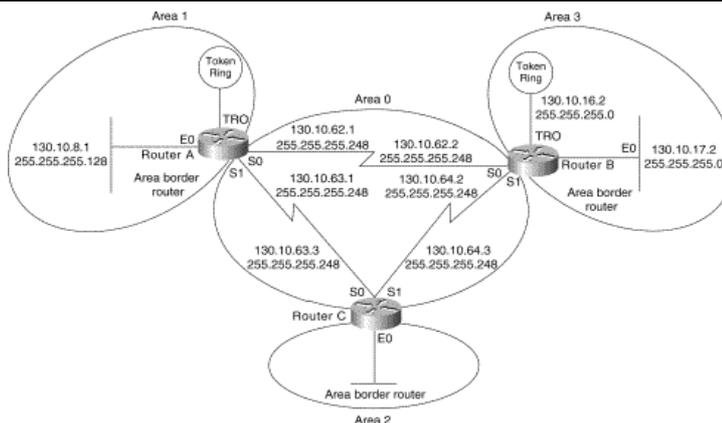
```

```

router ospf 109
network 130.10.62.0 0.0.0.255 area 0
network 130.10.63.0 0.0.0.255 area 0
redistribute rip subnets
distribute-list 11 out rip
!
access-list 11 permit 130.10.8.0 0.0.7.255
access-list 11 deny 0.0.0.0 255.255.255.255

```

EXEMPLE PASSAGE DE RIP À OSPF



Router A :

```

interface serial 0
ip address 130.10.62.1 255.255.255.248
interface serial 1
ip address 130.10.63.1 255.255.255.248
interface ethernet 0
ip address 130.10.8.1 255.255.255.0
ip irdp

```

```

interface tokenring 0
ip address 130.10.9.1 255.255.255.0
ip irdp
router ospf 109
network 130.10.62.0 0.0.0.255 area 0
network 130.10.63.0 0.0.0.255 area 0
network 130.10.8.0 0.0.7.255 area 1
area 1 range 130.10.8.0 255.255.248.0

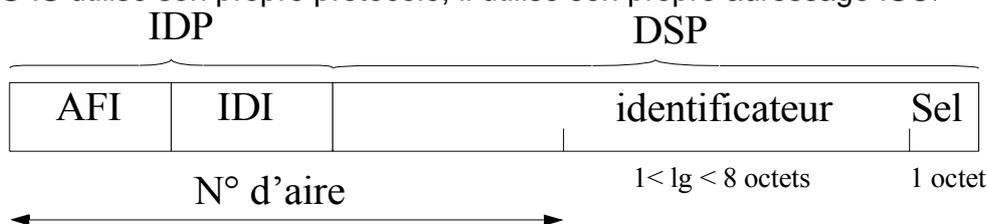
```

IS-IS

- Intermediate System to Intermediate System
- Protocole défini par l'ISO et par le groupe de travail IS-IS de l'IETF
- Utilisé à l'origine par les opérateurs car :
 - permet la gestion des réseaux ISO et IP avec le même protocole
 - opérationnel avant OSPF
- Opérateurs actuels utilisent plutôt OSPF
- Groupe de travail ISIS à l'IETF :
 - permet d'ajouter plus facilement un nouveau protocole (par exemple : IPv6)
- Principe très similaire à OSPF (deux niveaux de réseau : niveau 1 : équivalent aux aires d'OSPF et niveau 2 : équivalent au backbone d'OSPF)

ADRESSAGE

Comme IS-IS utilise son propre protocole, il utilise son propre adressage ISO:



- IDP (*Initial Domain Part*) valeurs sont attribués par l'UIT-T ou l'OSI :
 - AFI (*Authority and Format Identifier*), sur un octet, identifie le plan d'adressage ,l'organisme chargé d'attribuer les adresses ainsi que son format.
 - IDI (*Initial Domain Identifier*) l'autorité responsable de l'attribution des adresses

+ Pour IS-IS, la valeur peut être 49 (adressage local)

- DSP (*Domain Specific Part*) la partie locale de l'adresse :
 - numéro d'aire
 - l'identificateur d'interface (généralement 6 octets) construit avec l'adresse IP ou MAC
- SEL : vaut 0 pour IS-IS

NIVEAUX

Il existe deux types de routeurs :

- les routeurs de niveau 1 ont une connaissance de l'ensemble de préfixes de leur aire, le trafic vers une autre aire est systématiquement envoyé au routeur de niveau 2 le plus proche (au sens du routage),
- les routeurs de niveau 2 ont une connaissance complète des préfixes présents dans les différentes aires.

Comme pour OSPF, les routeurs de niveau 2 doivent former un backbone contigu.

Trois types de routeurs:

- les routeurs uniquement de niveau 1 qui ne dialoguent qu'avec des routeurs appartenant à la même aire,
- les routeurs uniquement de niveau 2 qui ne dialoguent qu'avec des routeurs de ce niveau,
- les routeurs à la fois de niveau 1 et 2 qui exécutent deux instances du protocole de routage pour chaque niveau.

PROTOCOLES

Plusieurs protocoles cohabitent pour permettre aux bases de données de se synchroniser :

- le protocole Hello (appelé également IIH : *IS-IS Hello*) permet de découvrir les voisins sur un lien donné et d'élire le routeur désigné qui maintiendra la base de données d'état de lien de référence.
Le protocole Hello permet également de vérifier la connectivité bi-directionnelle du lien.
- les routeurs s'échangent les LSP :
 - Le routeur désigné émet périodiquement des message contenant un "sommaire" de l'information stockée dans la base (paquets CSN : *Complete Sequence Number*).
 - Les autres routeurs peuvent demander une retransmission d'une information ou acquitter les informations reçues (paquets PSN : *Partial Sequence Number*).

Le format des messages est codé TLV, il est facile d'ajouter des types.

EXEMPLE

```
router isis
 redistribute connected
 redistribute rip level-1
 redistribute bgp 65525 metric 20 metric-type external
 default-information originate
 net 49.0001.1921.0811.9190.00

interface Vlan1
 description switched default VLAN
 ip address 192.108.119.190 255.255.255.192
 ip router isis
```

Le routeur ayant l'adresse IPv4 192.108.119.190 et étant dans l'aire 1

Il redistribue les informations de routage apprises dans :

- RIP au niveau 1
- les routes directement connectées et apprises par BGP au niveau 2
- une route par défaut au niveau 2.

EXEMPLE

```
alcmene#sh isis database detail
IS-IS Level-1 Link State Database
LSPID          LSP Seq Num  LSP Checksum  LSP Holdtime  ATT/P/OL
1921.0811.9150.00-00* 0x00000066   0x7EEB        780           0/0/0
Area Address: 49.0001
NLPID:         0xCC
IP Address:    192.108.119.150
Metric: 10 IP 192.108.119.128 255.255.255.192
Metric: 10 IS 1921.0811.9150.01
Metric: 0 ES 1921.0811.9150
1921.0811.9150.01-00* 0x00000065   0xDE33        1003          0/0/0
Metric: 0 IS 1921.0811.9150.00
Metric: 0 IS 1921.0811.9190.00
1921.0811.9190.00-00 0x00001482   0x8848        1192          0/0/0
Area Address: 49.0001
NLPID:         0xCC
Code: 137 Length: 4
IP Address:    192.108.119.190
Metric: 10 IP 192.108.119.128 255.255.255.192
Metric: 10 IS 1921.0811.9150.01
IS-IS Level-2 Link State Database
LSPID          LSP Seq Num  LSP Checksum  LSP Holdtime  ATT/P/OL
1921.0811.9150.00-00* 0x00001482   0x9B62        1194          0/0/0
Area Address: 49.0001
NLPID:         0xCC
IP Address:    192.108.119.150
Metric: 10 IS 1921.0811.9150.01
Metric: 10 IP 192.108.119.128 255.255.255.192
1921.0811.9150.01-00* 0x00000063   0x722A        490           0/0/0
Metric: 0 IS 1921.0811.9150.00
Metric: 0 IS 1921.0811.9190.00
1921.0811.9190.00-00 0x00000068   0x8B8A        907           0/0/0
Area Address: 49.0001
NLPID:         0xCC
Code: 137 Length: 4
IP Address:    192.108.119.190
```

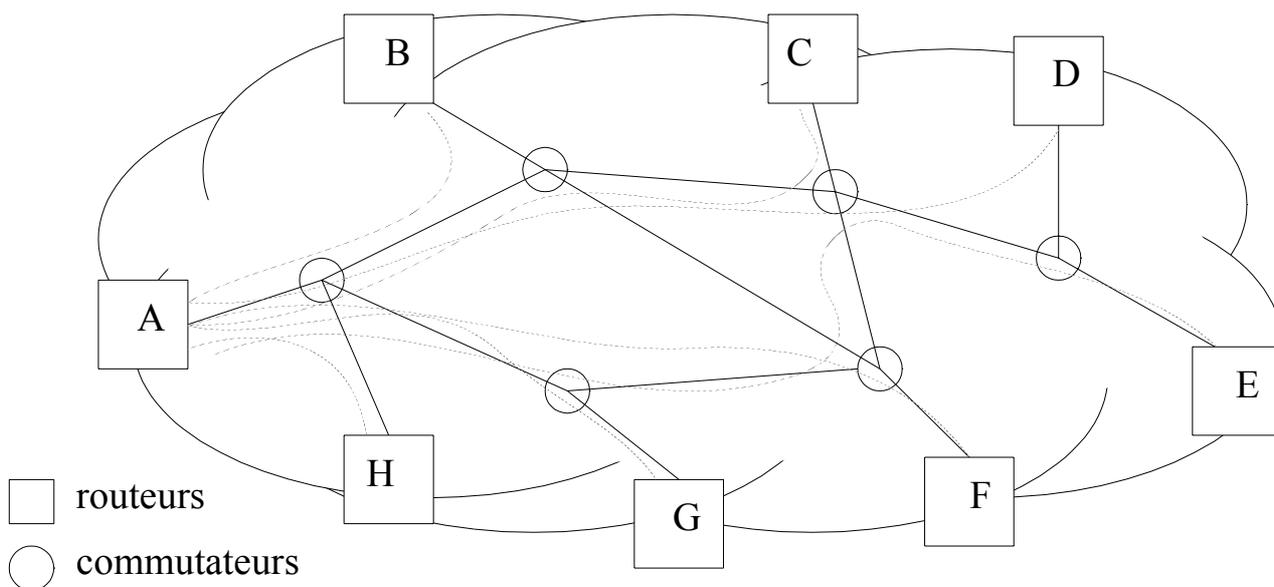
MPLS

Limites des IGP :

- les tables de routages (ou plutôt la FIB : *Forwarding Information Base*) de plus en plus grandes.
 - pour chaque paquet plus de 100 000 entrées doivent être consultées dans cette base.
 - Du point de vue de l'opérateur, les paquets doivent uniquement être dirigés vers un des routeurs de sortie de son réseau,
- le routage au niveau IP ignore la topologie physique du réseau.
 - Mettre des routeurs en périphérie du réseau et des commutateurs ATM/FR dans le cœur du réseau.
 - Les routeurs en périphérie choisissent en fonction de l'adresse de destination un circuit virtuel sur lequel les paquets seront transmis.
 - Le traitement lié à la table de routage n'est effectué qu'une fois pour chaque paquet dans le réseau de l'opérateur.
 - Par contre le problème de scalabilité lié au nombre élevé de messages de signalisation demeure quand l'architecture logique diffère de l'architecture physique.

17 janvier 2003 - C:\Documents and Settings\loutain\Mes documents\Support_ENST\Routage_Interne.fm

MANQUE DE SCALABILITÉ DE VP/DLCI.



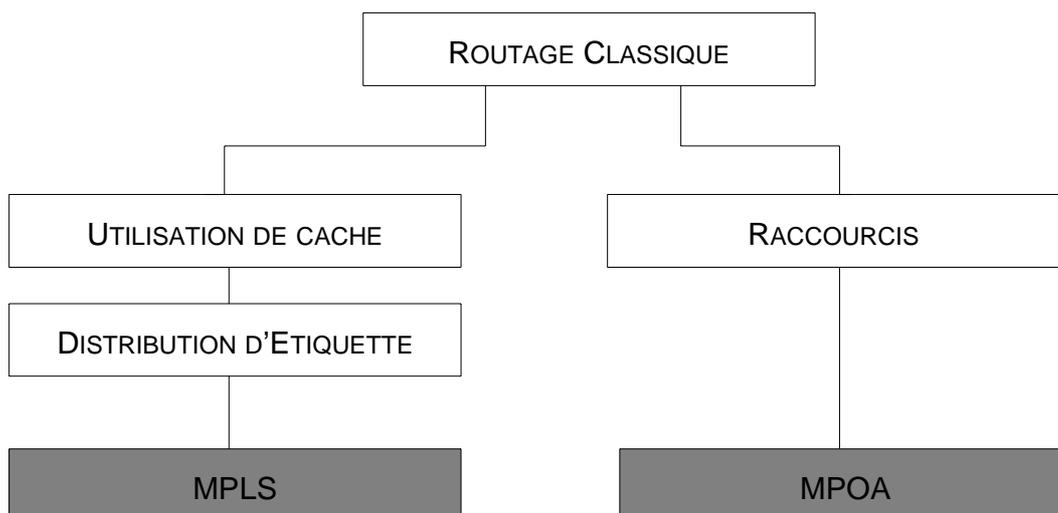
17 janvier 2003 - C:\Documents and Settings\loutain\Mes documents\Support_ENST\Routage_Interne.fm

LIMITE DES IGP (SUITE)

- Les algorithmes de plus courts chemins ont tendance à faire converger le trafic sur les mêmes artères et ignorer des routes qui sont peu chargées.
 - Même si des possibilités de partage de charge (*load balancing*) sont envisageables avec les algorithmes de routage traditionnels, le choix de la route est largement automatisé et laisse peu de possibilité pour différencier finement la route en fonction de la nature des flux.
- ATM ou Frame Relay permettent de construire des circuits virtuels entre deux équipements du réseau, or si le protocole de niveau deux ne permet pas le multiplexage de plusieurs circuits (comme par exemple PPP sur une liaison SHD), il est difficile de construire plusieurs réseaux IP distincts sur une même architecture physique.

17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\Routage_Interne.fm

AMÉLIORER L'EFFICACITÉ DU ROUTAGE



17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\Routage_Interne.fm

PRINCIPE DE MPLS (*MULTI PROTOCOL LABEL SWITCHING*)

- MPLS permet la commutation de l'information en fonction d'une étiquette relativement courte.
 - Le processus de retransmission dans le cœur de réseau est simplifié permettant une augmentation en débit.
 - Les routeurs en entrée situés en périphérie du réseau calculent cette étiquette en fonction de l'adresse de destination,
 - les routeur en sortie retirent l'étiquette.
- MPLS différencie fortement le plan donné (correspondant au transfert des paquets IP) et celui de signalisation (correspondant au déroulement des protocoles de routage).
 - Les équipements capables de traiter les en-têtes MPLS sont appelés des LSR (*Label Switch Router*).
 - Les équipements en entrée de réseau (LER : *Label Edge Router*) sont des routeur MPLS particuliers chargés d'insérer une étiquette au début des données en fonction de données contenues dans l'en-tête de niveau 3 du paquet.
 - Les LER vont analyser les paquets pour déterminer la FEC (*Fowarding Equivalence Class*) à laquelle correspondra une étiquette MPLS.

AVANTAGES DE MPLS (SUITE)

- Chaque LSR, même ceux situés en cœur du réseau participent au processus de routage, même s'ils n'effectuent pas de relaying au niveau 3 des paquets.
 - Cela permet une meilleure correspondance entre l'infrastructure logique (vue d'IP) et physique (vue de la commutation).
 - Le chemin pris par les paquets définis par les contextes de commutation dans les LSR s'appelle un LSP (*Label Switched Path*)
 - le calcul des chemins suivit par les paquets peut être automatisé grâce aux informations récoltées par les protocoles de routage,
 - il est également possible de forcer un chemin pour un flux particulier en créant un circuit virtuel pour celui-ci.
- +
- MPLS permet aux opérateurs une meilleure gestion de leurs flux facilitant l'ingénierie de trafic.

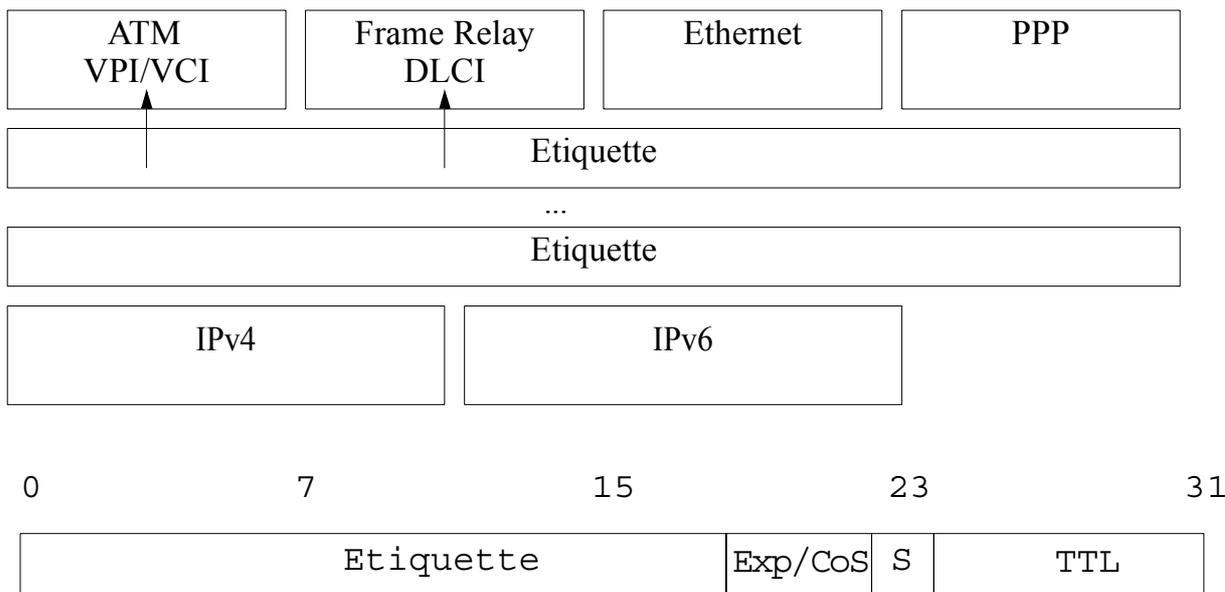
AVANTAGES DE MPLS

- MPLS permet une meilleure agrégation du trafic. Avec une table de routage classique, seul le prochain routeur est identifié. Avec MPLS, il est possible de faire des choix de chemin notamment en fonction :
 - du routeur de sortie du réseau,
 - de routeur d'entrée du réseau,
 - de flux particuliers.
 - l'étiquette permet de retrouver les fonctionnalités de création de circuits virtuels présente dans ATM ou Frame Relay, mais absente d'autres protocoles de niveau 2 comme PPP.
- + MPLS a longtemps été présenté comme la seule technologie permettant le haut débit,
- La commutation de niveau 3 avec des composants électroniques spécialisés pour atteindre des débits de plusieurs Gigabits par seconde.

AVANTAGES DE MPLS

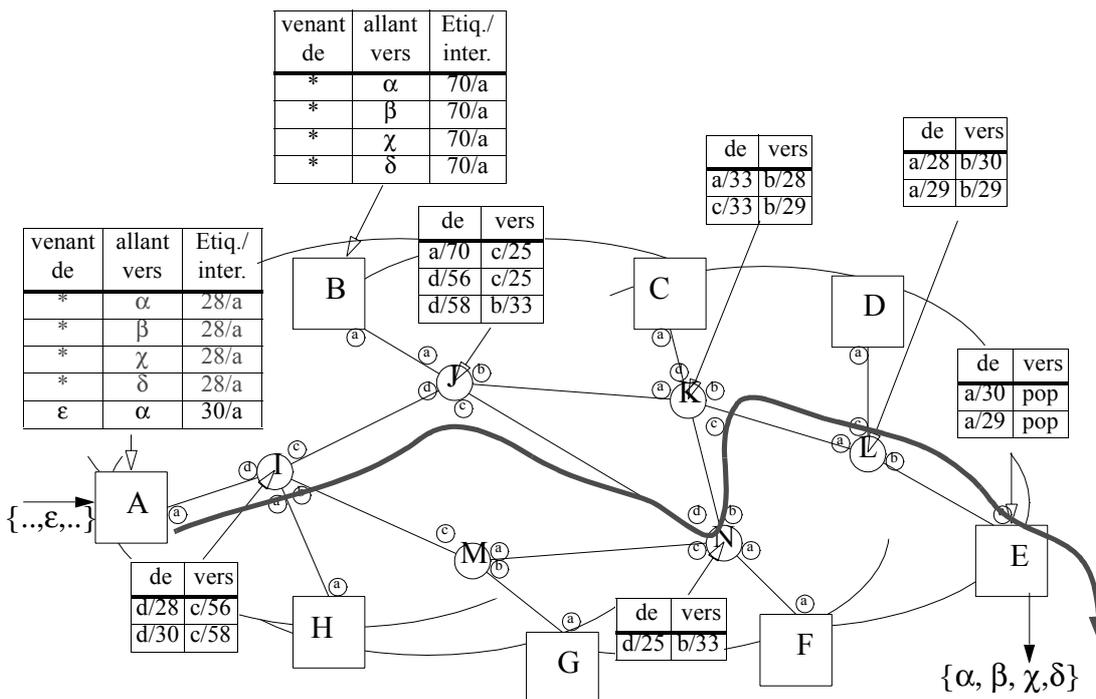
- Les arguments les plus en faveur de MPLS sont :
 - la meilleure gestion du trafic à l'intérieur d'un réseau d'opérateur,
 - la possibilité de créer des réseaux privés virtuels,
 - de relier des Intranets utilisant un adressage privéregroupés sous le terme d'ingénierie de trafic (*Traffic Engineering*)
- Les techniques employées par MPLS s'adaptent bien aux réseaux IP sur fibres optiques utilisant le multiplexage de longueurs d'ondes WDM (*Wavelength Division Multiplexing*).
 - Des routeurs rapides peuvent être construits pour brasser les longueurs d'ondes aux lieu d'étiquettes dans les trames, conduisant au concept de MPλS, λ représentant les longueurs d'ondes.

PRINCIPE DE FONCTIONNEMENT

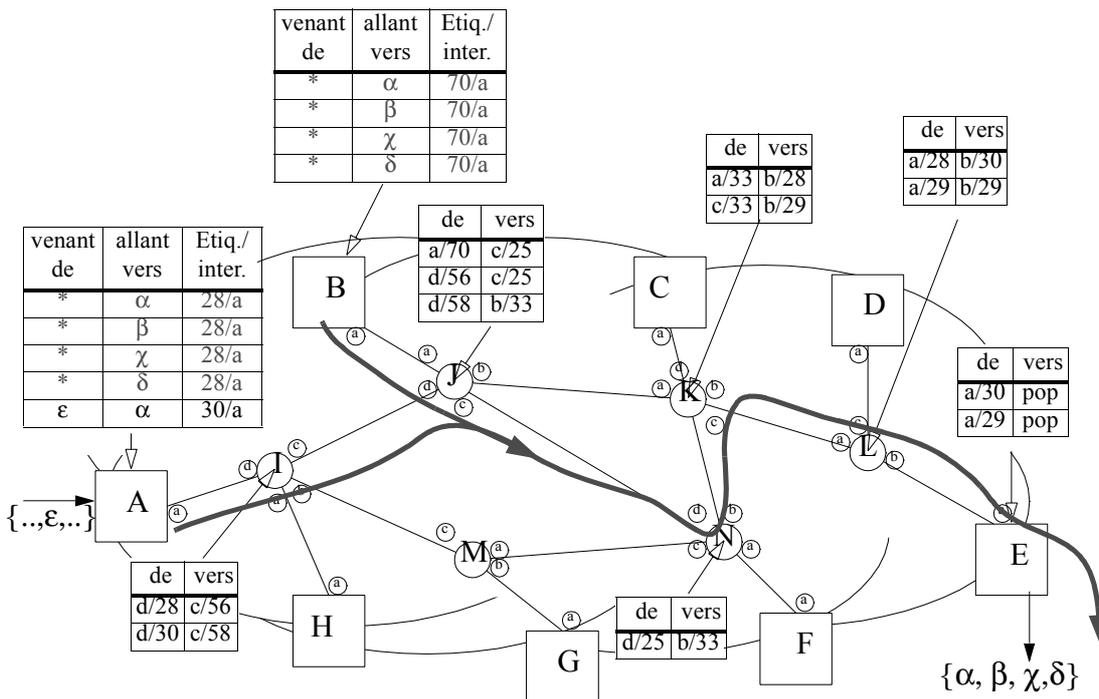


- recopie du TTL de la couche supérieure
 - même sémantique que pour IPv4 ou IPv6

EXEMPLE



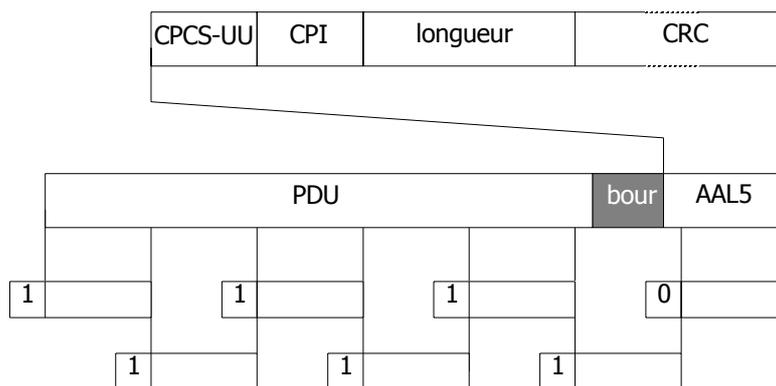
EXEMPLE : FUSION DE CIRCUIT (VC MERGING)



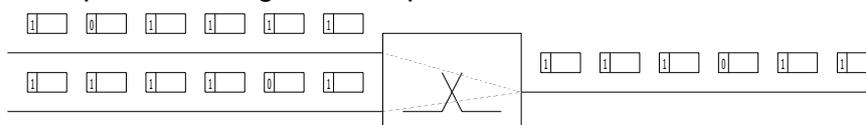
17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\ROUTAGE_Interne.fm

PROBLÈME AVEC ATM

Encapsulation des données avec ATM et l'AAL5.

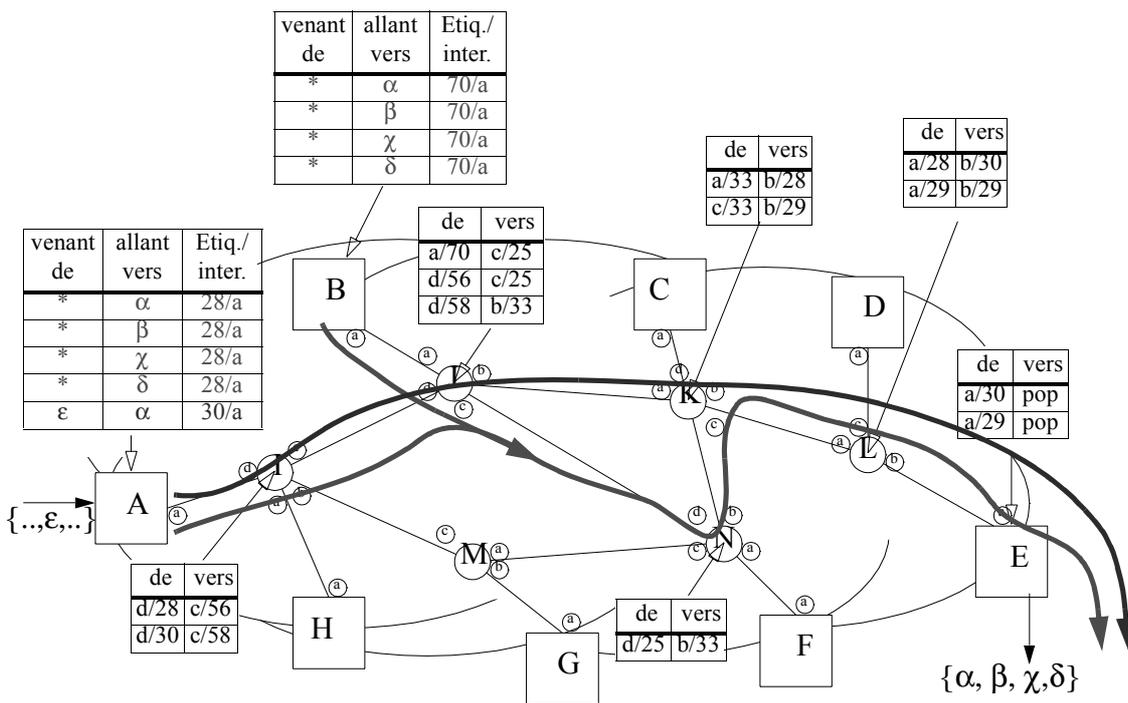


Un seul bit ne permet pas de distinguer entre plusieurs PDU initiaux



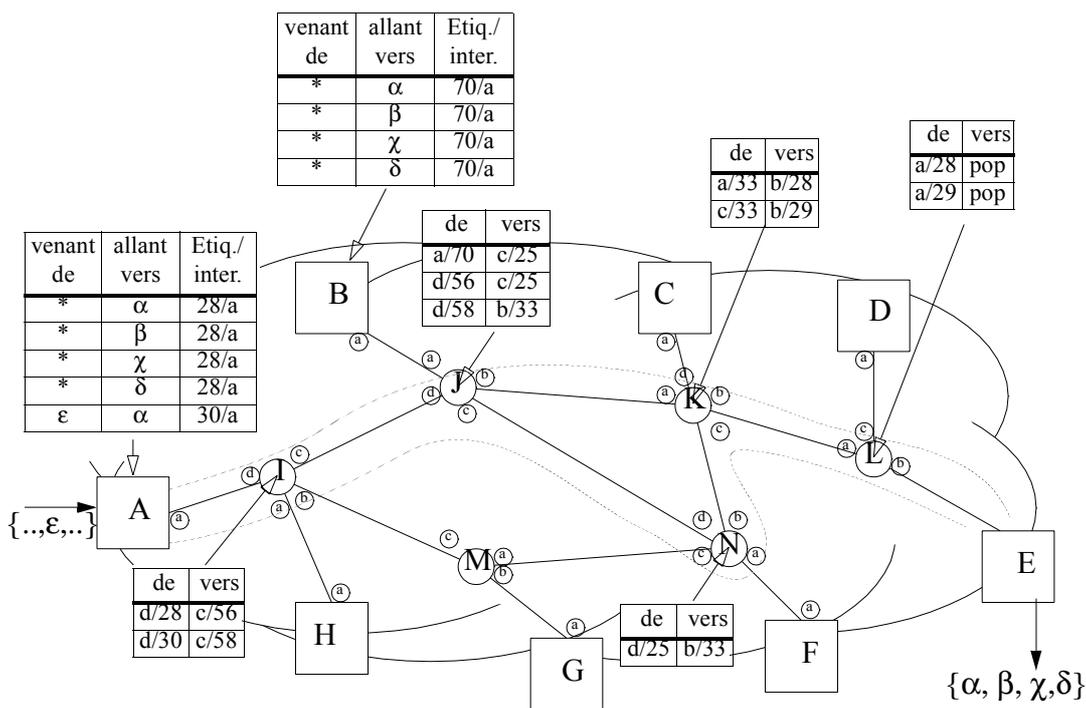
17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\ROUTAGE_Interne.fm

EXEMPLE : INGÉNIÉRIE DE TRAFIC



17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\ROUTAGE_Interne.fm

OPTIMISATION DU PENULTIEME ROUTEUR MPLS



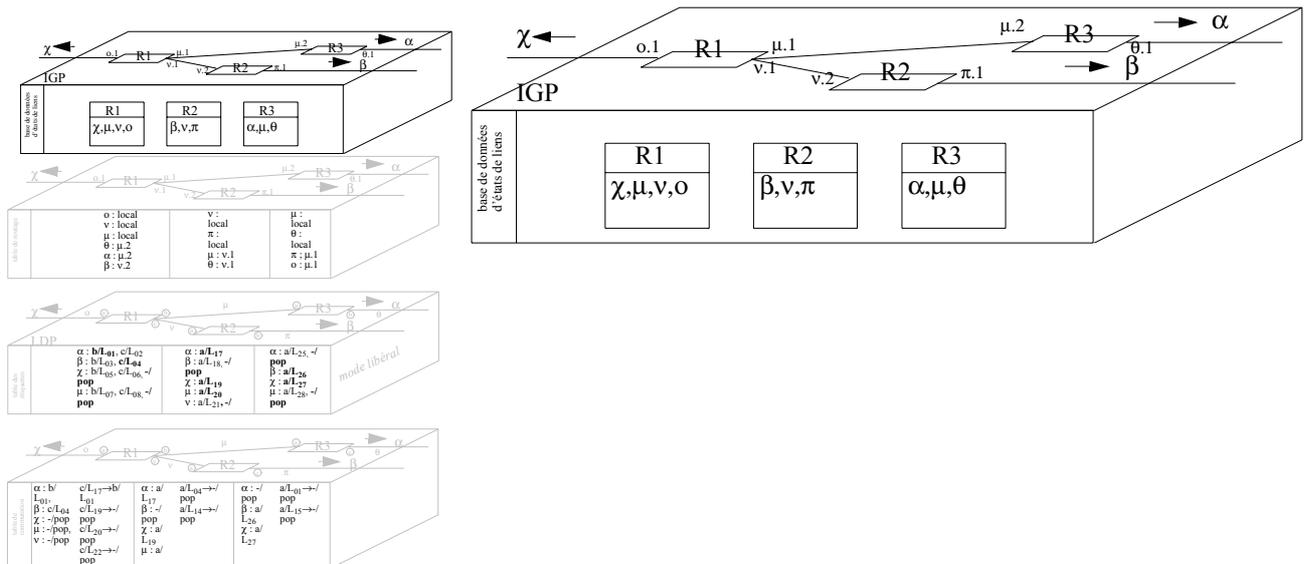
17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\ROUTAGE_Interne.fm

DISTRIBUTION DES ETIQUETTES

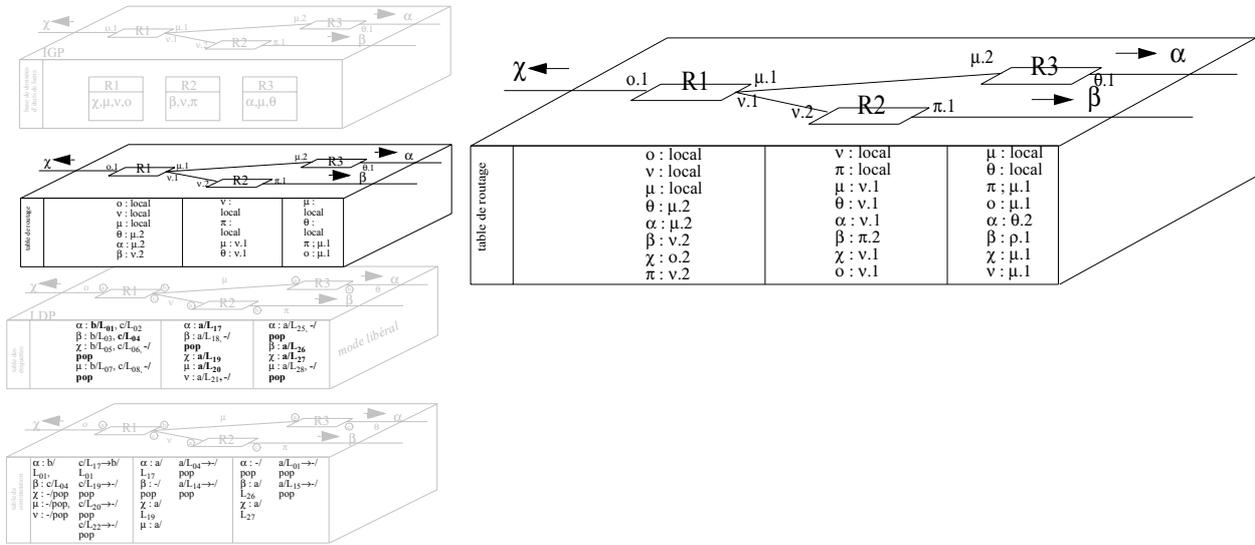
- La distribution des étiquettes se fait toujours du routeur en aval vers le routeur en amont :
 - non-sollicité : le routeur en aval apprend une nouvelle route par le protocole de routage, il attribue une étiquette et en informe les routeurs en amont. Si un routeur sélectionne le routeur en aval comme prochain saut, l'étiquette sera utilisée.
 - à la demande : le routeur en amont, quand son processus de routage sélectionne un prochain saut pour émettre des paquets vers une destination donnée, il doit posséder une étiquette MPLS. Le routeur envoie une requête au routeur en aval pour que ce dernier lui attribue une étiquette.
- le routeur en amont en :
 - mode libéral, mémorise toutes les étiquettes transmises par les routeurs en aval.
 - mode conservateur, ne garde en mémoire que les étiquettes pour lesquelles, il a choisi l'équipement comme prochain saut.

Dans le mode libéral, l'ouverture d'un LSP est plus rapide puisque tous les équipements intermédiaires disposent des étiquettes nécessaires pour construire le chemin, par contre, il faut toutes les étiquettes

TECHNIQUES DE DISTRIBUTION DES ETIQUETTES

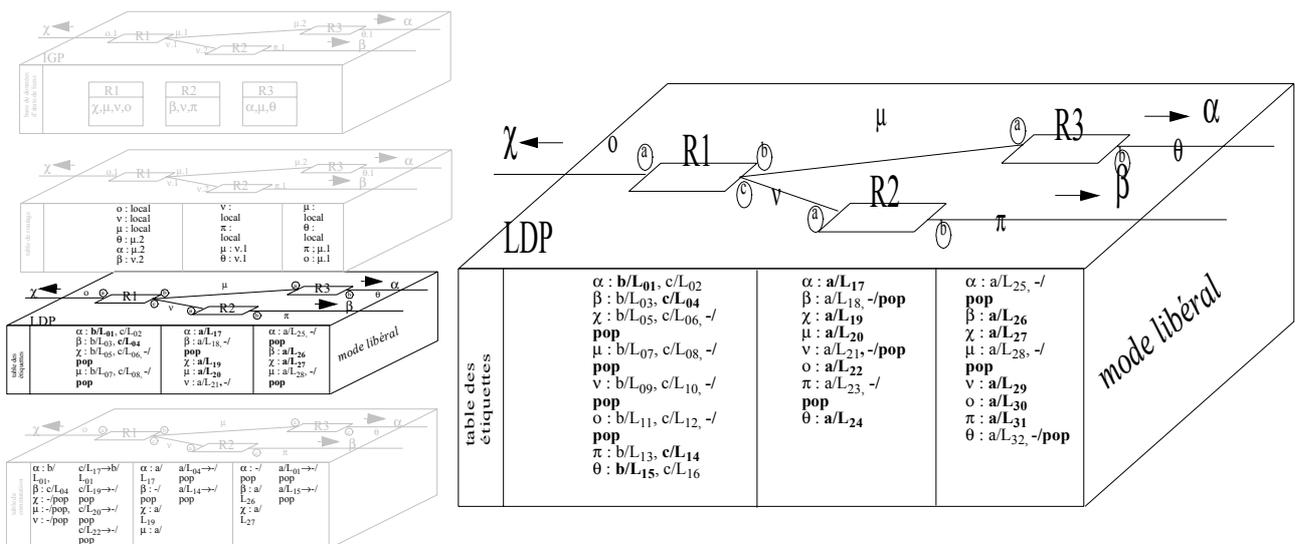


TECHNIQUES DE DISTRIBUTION DES ÉTIQUETTES



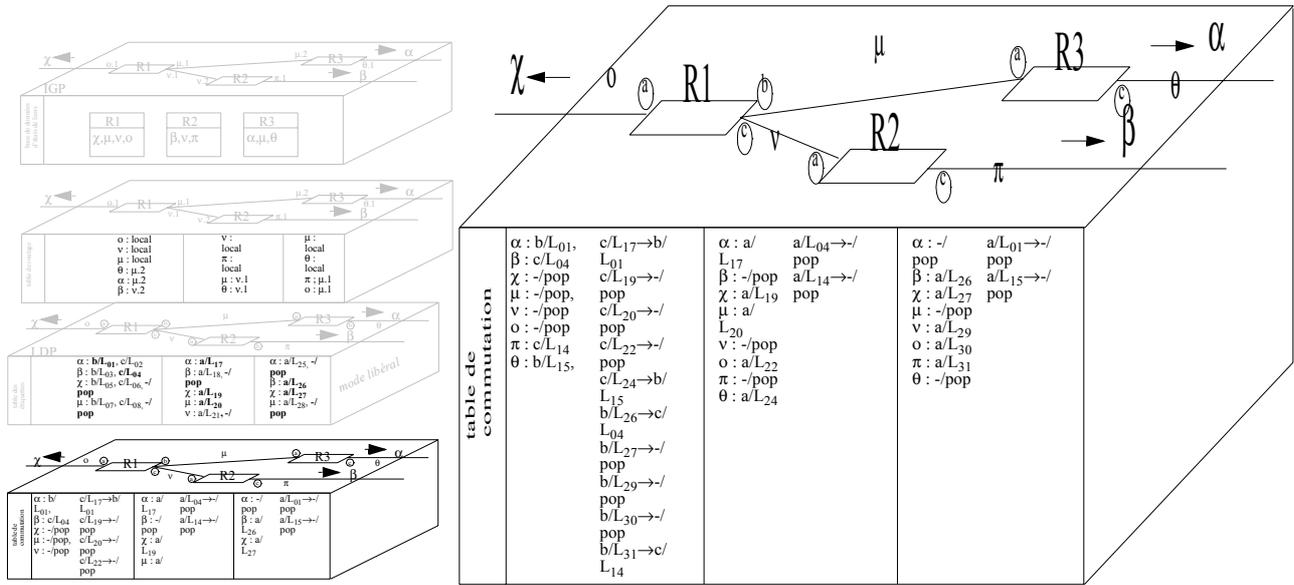
17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\ROUTAGE_Interne.fm

TECHNIQUES DE DISTRIBUTION DES ÉTIQUETTES



17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\ROUTAGE_Interne.fm

TECHNIQUES DE DISTRIBUTION DES ÉTIQUETTES



17 janvier 2003 - C:\Documents and Settings\loutain\mes documents\Support_ENST\ROUTAGE_Interne.fm

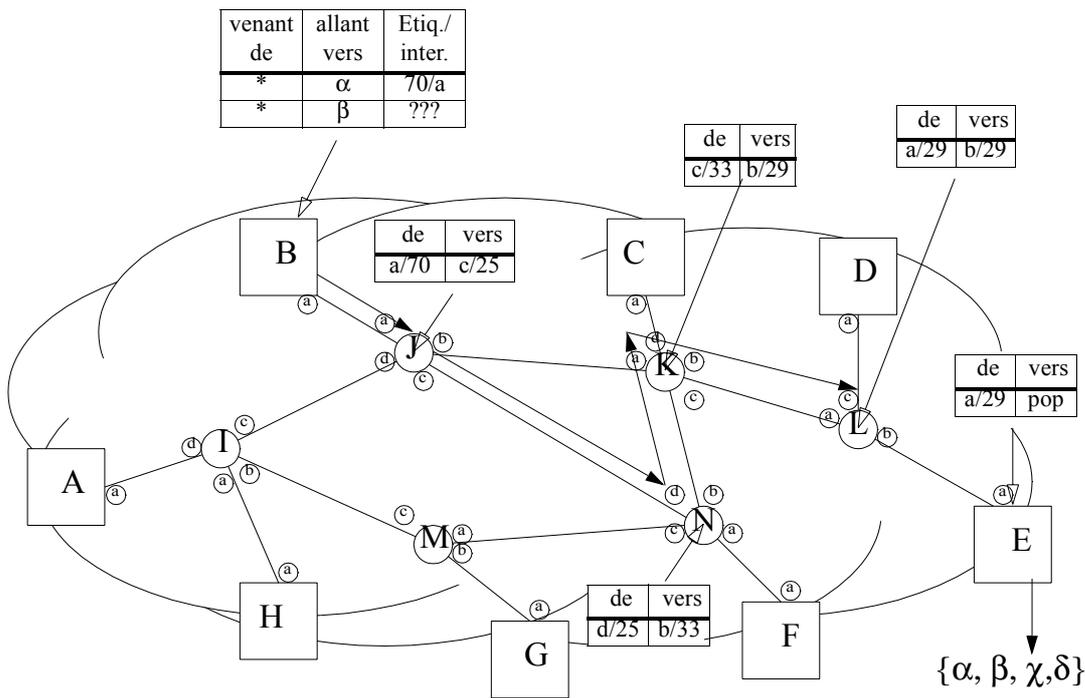
DEUX MODES DE DISTRIBUTION

- La distribution des étiquettes peut être dans le :
 - *mode ordonné*, c'est-à-dire qu'un routeur ne peut attribuer une étiquette à une FEC donnée que s'il a déjà reçu du routeur en aval une étiquette.
 - Dans le *mode indépendant*, quand un routeur reconnaît une nouvelle classe d'équivalence, il envoie une étiquette à ses voisins.

Le mode ordonné favorise l'agrégation des chemins mais l'établissement du circuit peut être plus long.

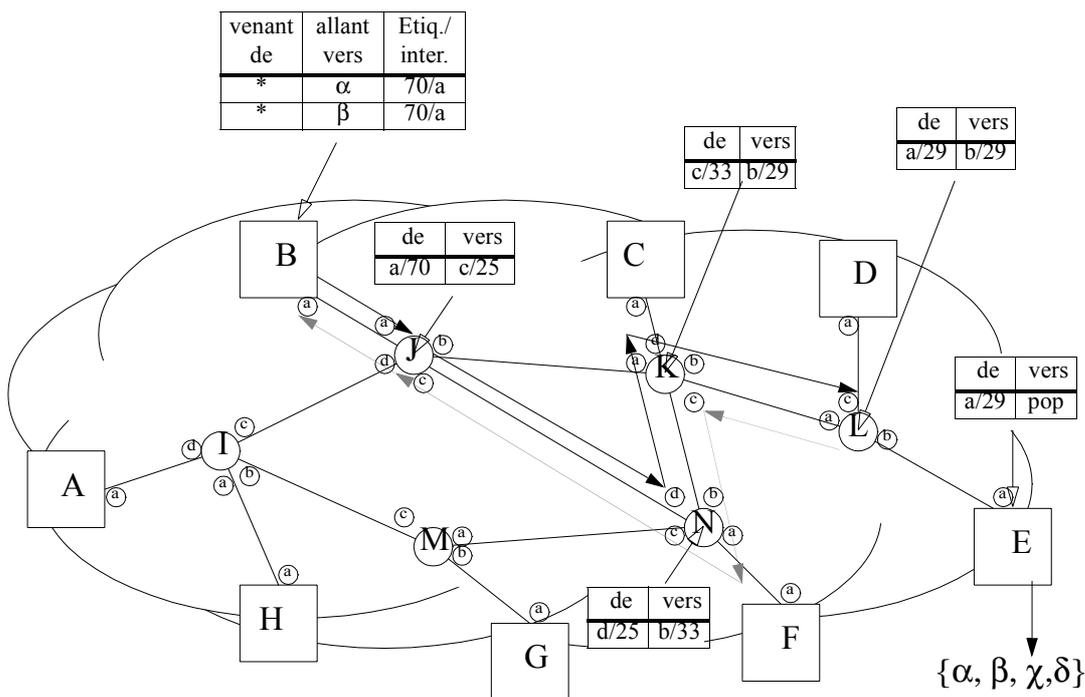
17 janvier 2003 - C:\Documents and Settings\loutain\mes documents\Support_ENST\ROUTAGE_Interne.fm

EXEMPLE DE MODE ORDONNÉ



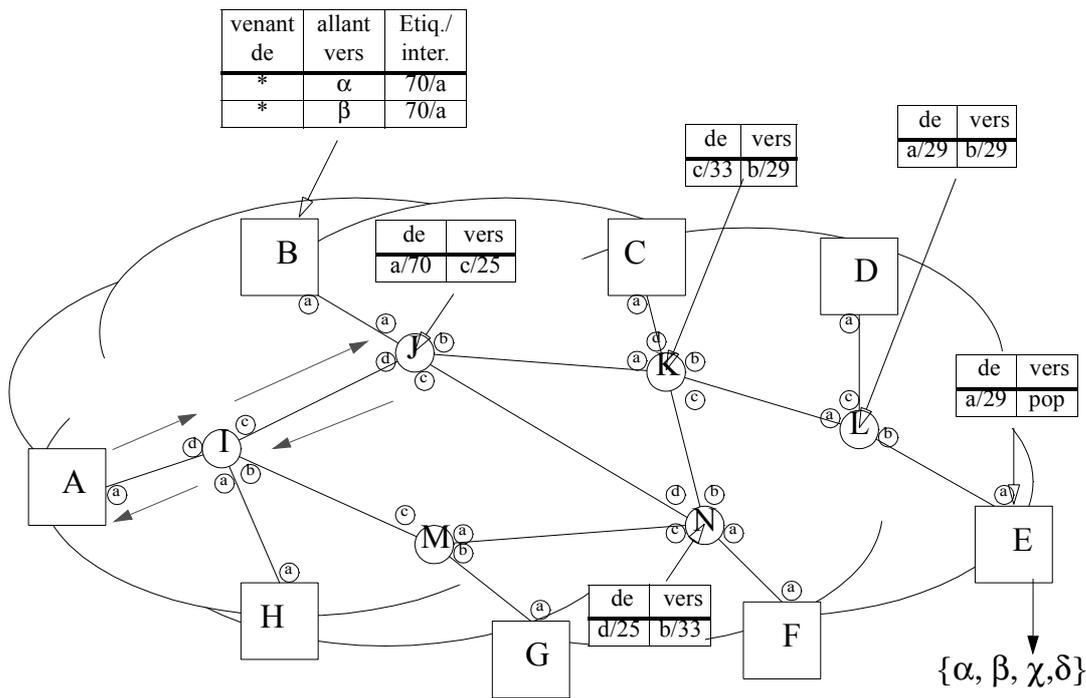
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\ROUTAGE_Interne.fm

EXEMPLE DE MODE ORDONNÉ



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\ROUTAGE_Interne.fm

EXEMPLE DE MODE ORDONNÉ



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\ROUTAGE_Interne.fm

PROTOCOLES DE DISTRIBUTION DES ÉTIQUETTES

Il n'existe pas de méthode unique standardisée pour distribuer les étiquettes :

- LDP (*Label Distribution Protocol*) définit par le groupe de travail MPLS
- Extension au protocole de routage externe BGP

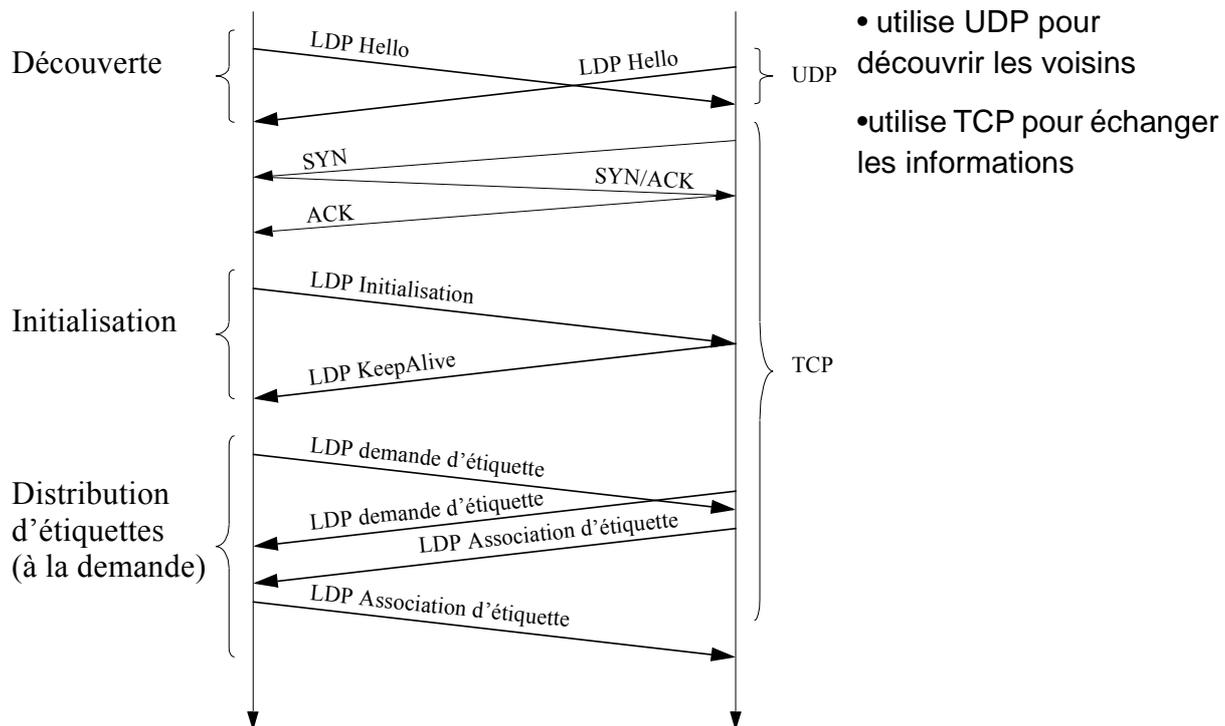
Création manuelle des tables.

Pour l'ingénierie de trafic et la réservation de ressources :

- CR-LDP
- RSVP

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\ROUTAGE_Interne.fm

LDP : ECHANGES PROTOCOLAIRES



17 janvier 2003 - C:\Documents and Settings\jclain\Mes documents\Support_ENST\ROUTAGE_Interne.fm

SYSTÈMES AUTONOMES

- Impossible de gérer le réseau Internet de manière centralisée.
- Un système autonome dispose d'une autonomie de gestion
 - numérotation des machines
 - plan de routage interne (manuel ou IGP)
 - ...
- Un système autonome est représenté par un entier sur 16 bits
- Alloué par les RIR
- Il faut des points de centralisation de l'information.
- Certains réseaux possèdent des routes par défaut dans leur table, certains non.

⇒ Comment trouver le premier routeur qui n'a pas de route par défaut ?

TRACEROUTE 1.2.3.4

tracertoute to 1.2.3.4 (1.2.3.4), 30 hops max, 40 byte packets

```
1 mgs-rsm (192.44.77.2) 4 ms 1 ms 1 ms
2 ft-renater (193.52.72.1) 3 ms 2 ms 2 ms
3 rennes3.or-br.ft.net (193.48.78.41) 4 ms 15 ms 8 ms
4 rennes1.or-br.ft.net (193.48.78.25) 13 ms 4 ms 4 ms
5 192.93.43.210 (192.93.43.210) 14 ms 12 ms 12 ms
6 stamand2.renater.ft.net (192.93.43.138) 13 ms 12 ms 12 ms
7 stamand1.renater.ft.net (192.93.43.34) 24 ms 12 ms 16 ms
8 stamand3.renater.ft.net (192.93.43.17) 23 ms 20 ms 17 ms
9 * rbs1.renater.ft.net (192.93.43.121) 25 ms 36 ms
10 * rbs1.renater.ft.net (192.93.43.121) 26 ms !H^C
```

LA COMMANDE WHOIS

>whois AS1717

aut-num: AS1717
 descr: RENATER
 descr: Reseau National de telecommunications pour la Technologie
 descr: l'Enseignement et la Recherche
 descr: FR
 as-in: from AS1755 100 accept ANY
 as-in: from AS5511 100 accept ANY
 as-in: from AS2470 100 accept AS2470
 as-in: from AS789 50 accept AS789
 as-in: from AS786 50 accept AS786
 as-in: from AS1899 50 accept AS1899
 as-in: from AS2917 50 accept AS2917
 as-in: from AS3215 50 accept AS3215
 as-out: to AS5511 announce AS1717
 as-out: to AS2470 announce AS1717
 as-out: to AS1755 announce AS1717 AS2470
 as-out: to AS786 announce AS1717
 as-out: to AS789 announce AS1717
 as-out: to AS1899 announce AS1717
 as-out: to AS2917 announce AS1717
 as-out: to AS3215 announce AS1717

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Routeage_Externe.fm

default: AS1755 10
 admin-c: Michel Lartail
 tech-c: Isabelle Morel
 tech-c: Marie-Helene Guilmin
 mnt-by: AS1717-MNT
 changed: rensvp@renater.fr 960423
 source: RIPE

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Routeage_Externe.fm

NOM DES AS

AS	Nom
AS1755	EBONE
AS5511	TRANSPAC
AS2470	Ile de la Réunion Département d'Outre-Mer
AS789	IN2P3
AS786	The JANET IP Service
AS1899	Fnet, EUnet-France
AS2917	OLEANE - PIPEX International

AS	Nom
AS3215	RAIN (Réseau d'Accès à l'Internet TRANSPAC)

TRACEROUTE AVEC SYSTEME AUTONOME

```

>./traceroute -A -h whois.ripe.net -n rki.kbs.co.kr
traceroute to rki.kbs.co.kr (210.115.193.23): 1-30 hops, 38 byte packets
 1 193.52.74.2 [AS1717 - FR-OR-ENSTRE] 2.95 ms 2.31 ms 2.29 ms
 2 193.51.128.81 [AS1717 - RENATER] 3.7 ms 3.23 ms 3.13 ms
 3 193.48.78.41 [AS1717-proxy aggregate done by ENONE of RENATER (AS1717) prefixes] 4.23 ms 4.21 ms 4.12 ms
 4 193.48.78.25 [AS1717-proxy aggregate done by ENONE of RENATER (AS1717) prefixes] 4.45 ms 4.63 ms 4.48 ms
 5 193.55.253.170 [AS1717 - RENATER] 4.57 ms 4.64 ms 4.79 ms
 6 195.220.180.153 [AS1717 - FR-RENATER-RNI] 11.1 ms 10.3 ms 10.5 ms
 7 195.220.180.34 [AS1717 - FR-RENATER-RNI] 12.7 ms 15.5 ms 11.2 ms
 8 195.220.180.29 [AS1717 - FR-RENATER-RNI] 13.3 ms 12.3 ms 11.6 ms
 9 193.55.152.70 [AS5511 - OPENTRANSIT.NET] 87.4 ms 87.6 ms 89.7 ms
10 144.232.8.209 88.7 ms 88.0 ms 87.6 ms
11 144.232.7.142 88.2 ms 88.9 ms 88.2 ms
12 206.157.77.33 90.2 ms 89.8 ms 98.0 ms
13 166.48.14.1 159 ms (ttl=240!) 160 ms (ttl=240!) 160 ms (ttl=240!)
14 166.48.15.246 163 ms (ttl=239!) 160 ms (ttl=239!) 159 ms (ttl=239!)
15 203.233.35.249 332 ms (ttl=238!) 328 ms (ttl=238!) 329 ms (ttl=238!)
16 210.120.128.4 327 ms (ttl=237!) 325 ms (ttl=237!) 326 ms (ttl=237!)
17 203.233.37.146 374 ms (ttl=236!) 347 ms (ttl=236!) 340 ms (ttl=236!)
18 210.115.193.23 345 ms (ttl=108!) 333 ms (ttl=108!) 353 ms (ttl=108!)
    
```

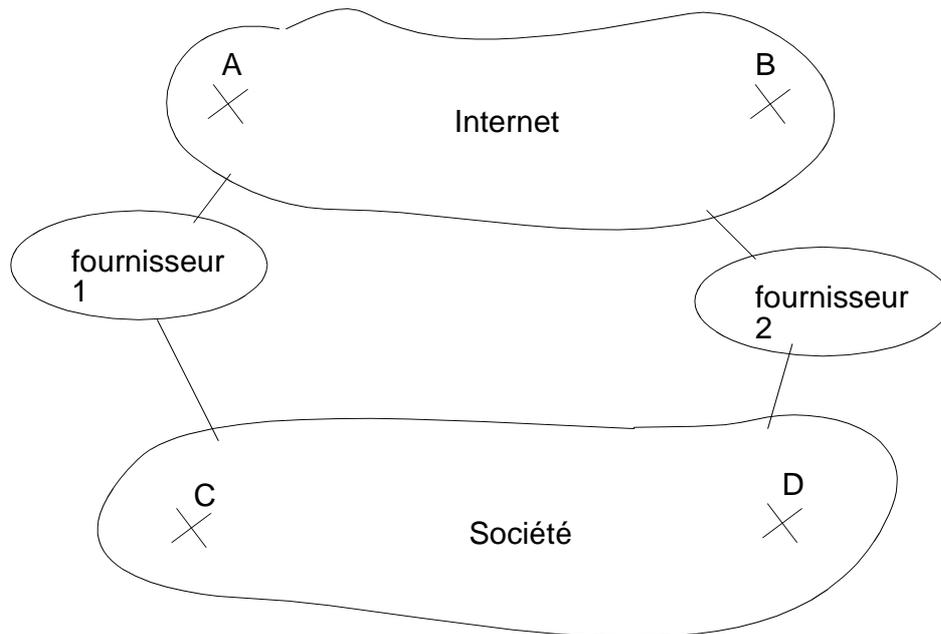
✦ Limité à l'Europe



ROUTER ADVISERY

```
>tracert -A -h whois.ra.net www.yahoo.com.cn
tracert to cn.yahoo.com (61.135.128.37): 1-30 hops, 38 byte packets
 1  homere2.rennes.enst-bretagne.fr [AS1717 - Ecole Nationale Supérieure des Telecommunications de Bretagne]
 2  vthd (192.108.119.190) [AS1717 - Ecole Nationale Supérieure des Telecommunications de Bretagne]
 3  193.50.69.73 (193.50.69.73) [AS1717 - RENATER] 3.06 ms 13.0 ms 6.60 ms
 4  PAO-Rennes2.rrb.ft.net (193.48.78.89) [AS1717 - RENATER] 4.31 ms 4.92 ms 4.80 ms
 5  peering-GIP.rrb.ft.net (195.101.145.6) [AS3215 - RAIN] 5.00 ms 5.40 ms 4.81 ms
 6  194.214.109.193 (194.214.109.193) [AS2200 - FR-RENATER2] 5.14 ms 7.04 ms 5.48 ms
 7  nio-nl.cssi.renater.fr (194.214.109.165) [AS2200 - FR-RENATER2] 12.0 ms 13.9 ms 12.3 ms
 8  nio-i.cssi.renater.fr (193.51.206.10) [AS1717 - RENATER] 15.1 ms 13.1 ms 11.6 ms
 9  opentransit.cssi.renater.fr (193.51.206.42) [AS1717 - RENATER] 20.2 ms 14.0 ms 12.8 ms
10  P2-0.PASBB2.Pastourelle.opentransit.net (193.251.128.65) [AS5511 - France Telecom] 12.7 ms 12.9 ms
11  P8-0.NYKBB4.New-york.opentransit.net (193.251.240.142) [AS5511 - France Telecom] 92.8 ms 91.5 ms
12  P7-0.NYKBB1.New-york.opentransit.net (193.251.154.249) [AS5511 - France Telecom] 96.1 ms 88.3 ms
13  P2-0.NYKBB2.New-york.opentransit.net (193.251.154.94) [AS5511 - France Telecom] 89.8 ms 86.9 ms
14  if-4-0-0.bb5.NewYork.Teleglobe.net (64.86.90.129) [AS6453 - Globeinternet - customers] 89.1 ms 87.7 ms
15  if-3-0.core2.NewYork.Teleglobe.net (207.45.221.66) [AS6453 - Teleglobe Inc.] 86.0 ms 86.9 ms 86.4 ms
16  if-9-0.core3.NewYork.Teleglobe.net (64.86.83.158) [AS6453 - Globeinternet -customers] 168 ms (ttl=238!)
17  if-8-0.core2.LosAngeles.Teleglobe.net (64.86.83.174) [AS6453 Globeinternet customers] 178 ms (ttl=235!)
18  64.86.255.10 (64.86.255.10) [AS14080 - AT&T Colombia] 188 ms (ttl=234!) 179 ms (ttl=234!)
19  202.97.48.10 (202.97.48.10) [AS4134 - ChinaNET backbone] 345 ms 343 ms 343 ms
20  p-13-0-r1-c-bjbj-1.cn.net (202.97.33.9) [AS4134 - ChinaNET backbone] 341 ms (ttl=235!)
21  p-2-0-r1-a-bjbj-2.cn.net (202.97.38.50) [AS4134 - ChinaNET backbone] 336 ms (ttl=234!)
22  202.96.12.50 (202.96.12.50) [AS4134 - DATA COMMUNICATIONS BUREAU] 350 ms (ttl=230!) 353 ms (ttl=230!)
23  202.106.193.182 (202.106.193.182) [AS4808 - CHINANet Beijing Site] 344 ms (ttl=234!) 359 ms (ttl=234!)
24  202.108.41.178 (202.108.41.178) [AS4808 - CHINANet Beijing Site] 349 ms (ttl=227!) 361 ms (ttl=227!)
25  202.108.61.6 (202.108.61.6) [AS4808 - CHINANet Beijing Site] 338 ms (ttl=230!) 343 ms (ttl=230!)
26  w4.cn.yahoo.com (61.135.128.37) [AS4808 - CHINANet Beijing Site] 343 ms (ttl=229!) 339 ms (ttl=229!)
```

CAS DU MULTIHOMING



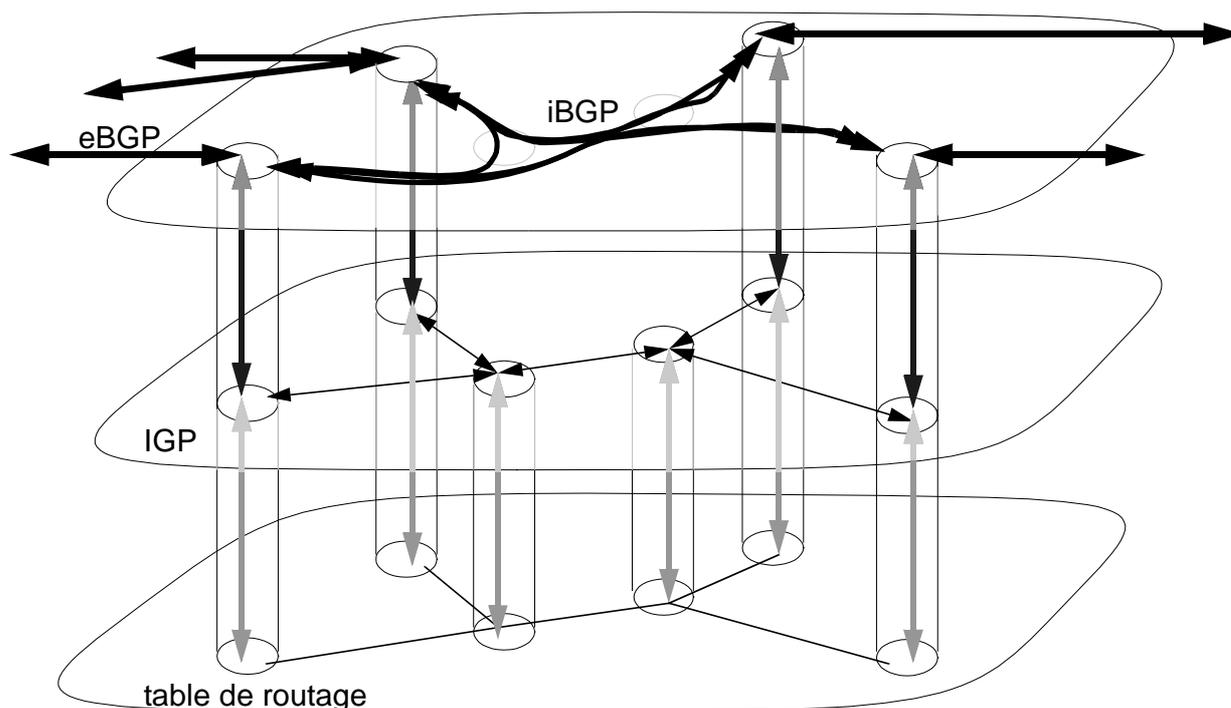
ANNONCES

- La machine C veut joindre la machine B
 - Si utilisation de routes par défaut, le paquet sort au plus vite du réseau pour aller via le fournisseur 1 vers l'Internet
 - Si les annonces du fournisseur 2 sont diffusées dans le réseau de la société, le paquet passera par le fournisseur 2 (si la métrique est meilleure).
- La machine A veut joindre la machine C
 - Si la société annonce que les routes "proches" du fournisseur 1, le paquet transitera via l'Internet.
 - Si la société annonce toutes ses routes, le paquet passera par le fournisseur 1.
- Si l'utilisateur de A utilise l'option de Source Routing pour joindre B
 - il peut utiliser les ressources privées de la société
 - filtrer les options de Source Routing en frontière de zone

BGP

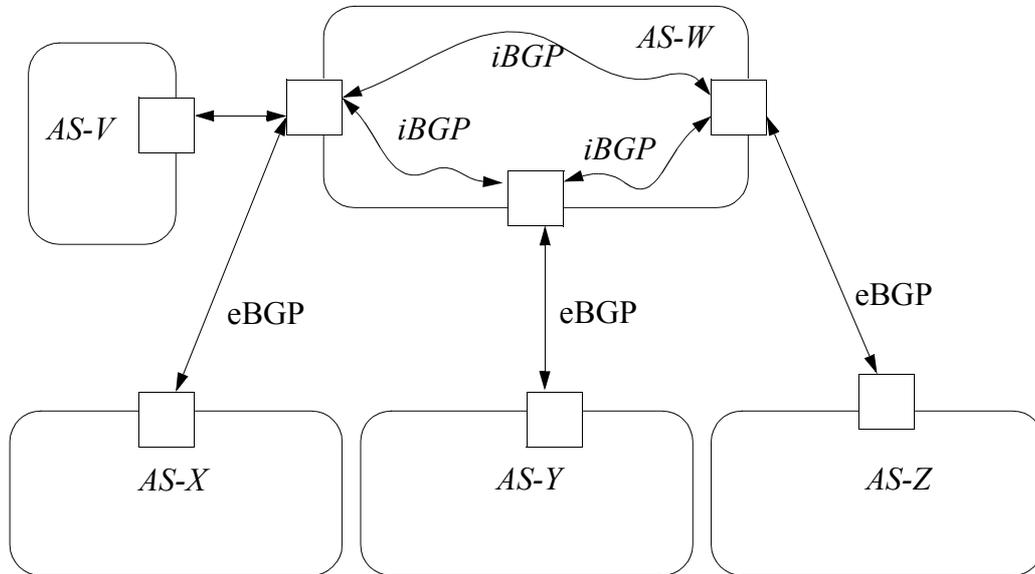
- Propagation des annonces de routes
- Evite des boucles par maintien du chemin des AS traversés (semblable au Source Routing de l'anneau à jeton).
- BGP 4 permet l'adressage CIDR
- La future version sera le protocole IDRP de l'ISO
- Utilisation de TCP (EGP utilise UDP) port 179 pour faciliter :
 - ouverture
 - mise à jour (échange d'information de routage)
 - notification (détection d'une erreur de protocole)
 - sonde (vérifier la vie de la connexion)

PRINCIPES



IBGP

- Protocole de routage Externe (mais utilisé en interne)
- Synchronisation des bases de données
- Pose des problèmes de synchronisation avec les protocoles de routage internes.

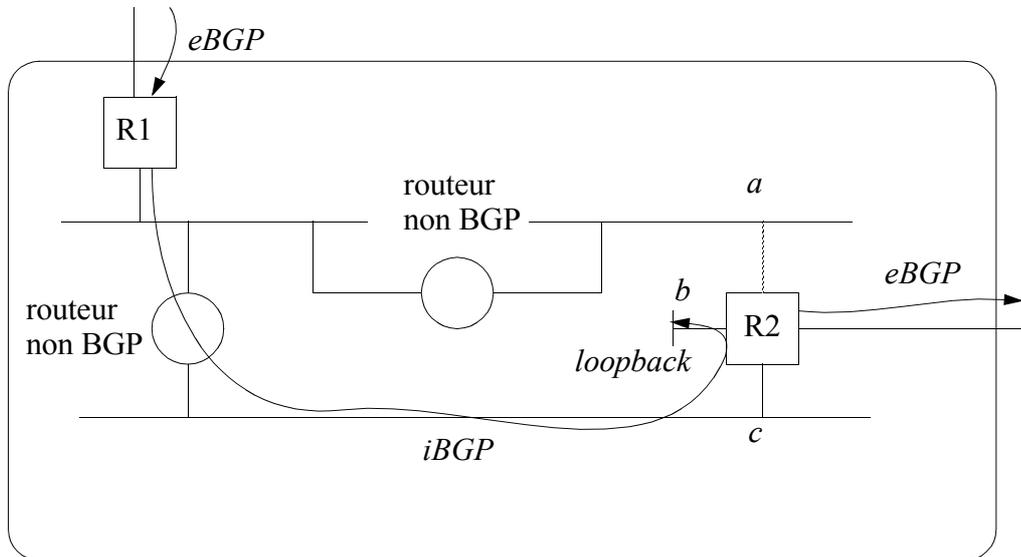


ATTRIBUTS.

drapeau	code	Attribut
bien connu obligatoire	1	ORIGIN
bien connu obligatoire	2	AS_PATH
bien connu obligatoire	3	NEXT_HOP.
optionnel, non transitif	4	MULTI_EXT_DISCR (BGP-4) ou INTER-AS (BGP-3).
bien connu facultatif	5	LOCAL_PREF
bien connu facultatif	6	ATOMIC_AGGREGATE.
optionnel transitif	7	AGGREGATOR.
optionnel transitif	8	COMMUNITY.t.
optionnel non-transitif	9	ORIGINATOR_ID.
optionnel non-transitif	10	CLUSTER_LIST
	11	DPA
optionnel non-transitif	12	ADVERTISER
optionnel non-transitif	13	RCID_PATH / CLUSTER_ID

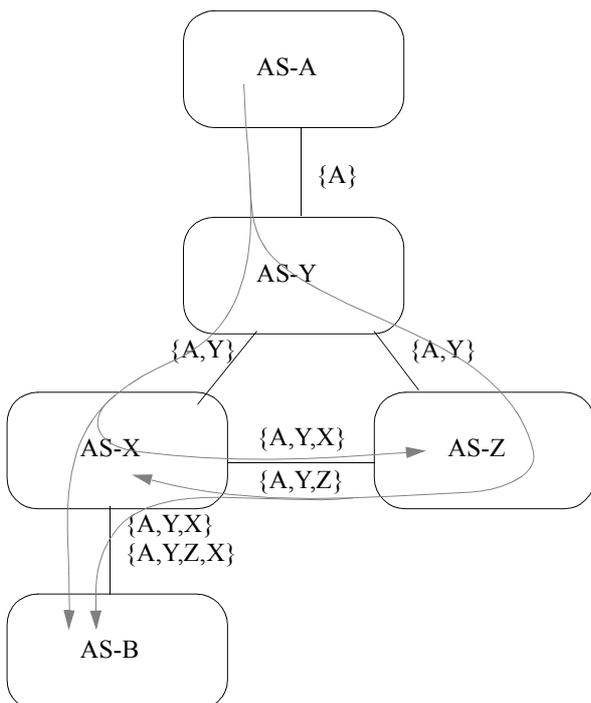
- bien connu : mis en œuvre dans tous les routeurs
- transitif : peut sortir de l'AS

UTILISATION D'UNE ADRESSE LOOPBACK



17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\ROUTAGE_Externe.fm

AS_PATH



Evite les boucles dans les annonces

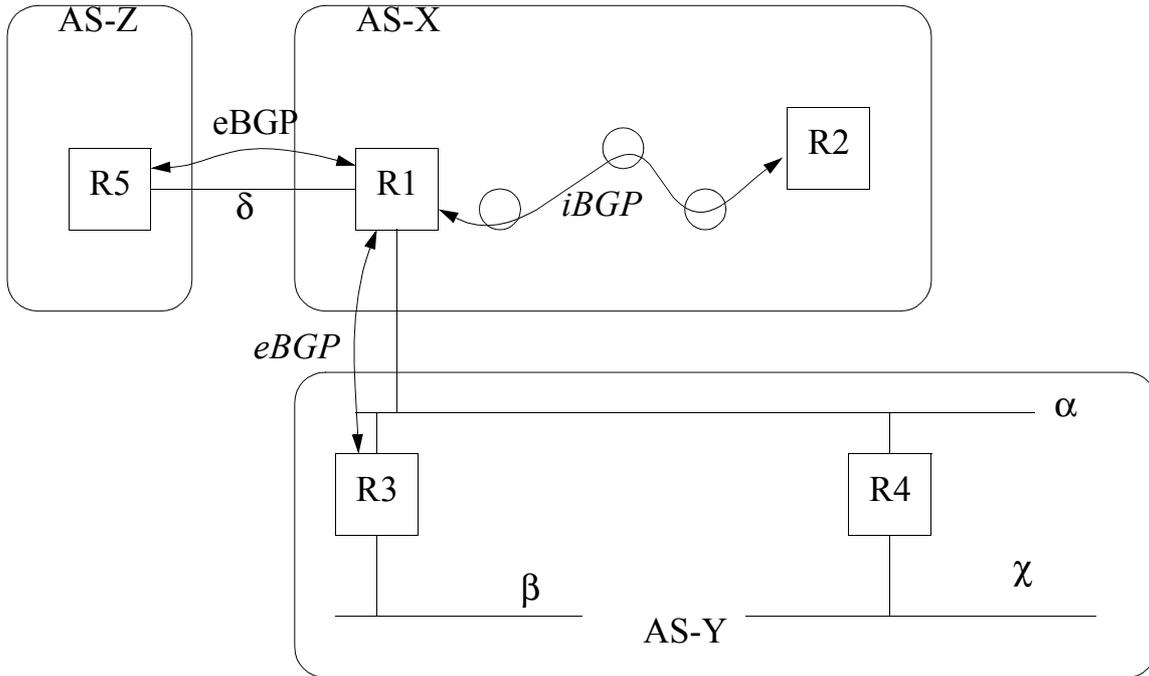
Permet de choisir le chemin le plus court

+ Certains opérateurs ajoutent plusieurs fois leur numéro d'AS pour défavoriser cette route

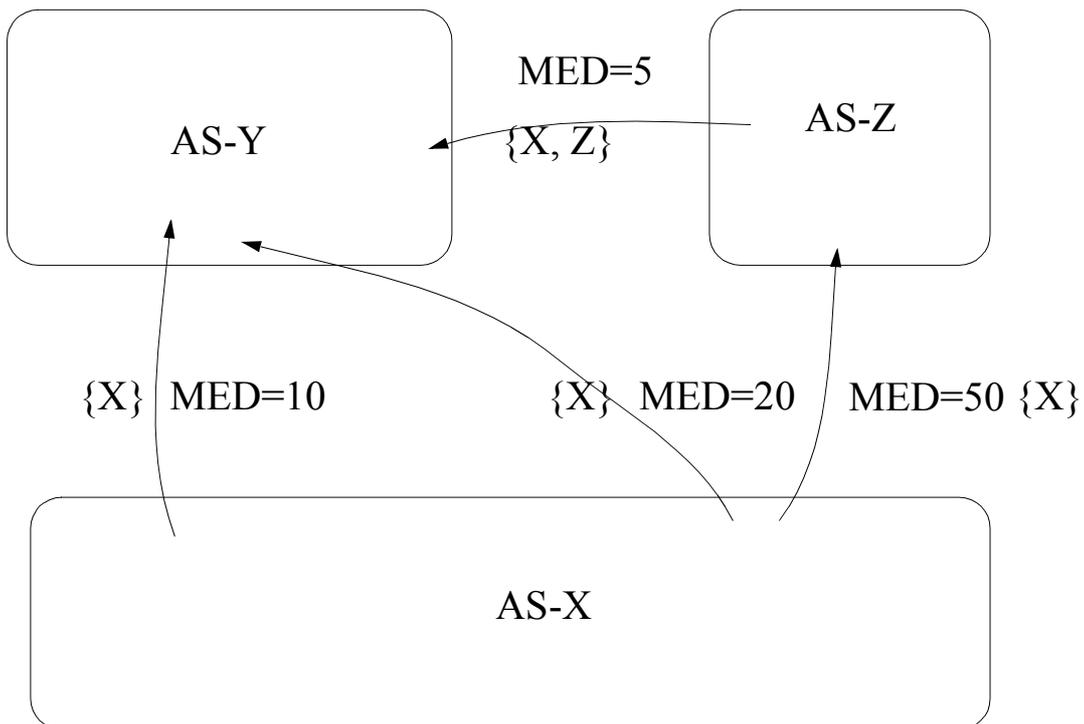
+ Problème en iBGP car les boucles ne sont pas détectées avec cet attribut

17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\ROUTAGE_Externe.fm

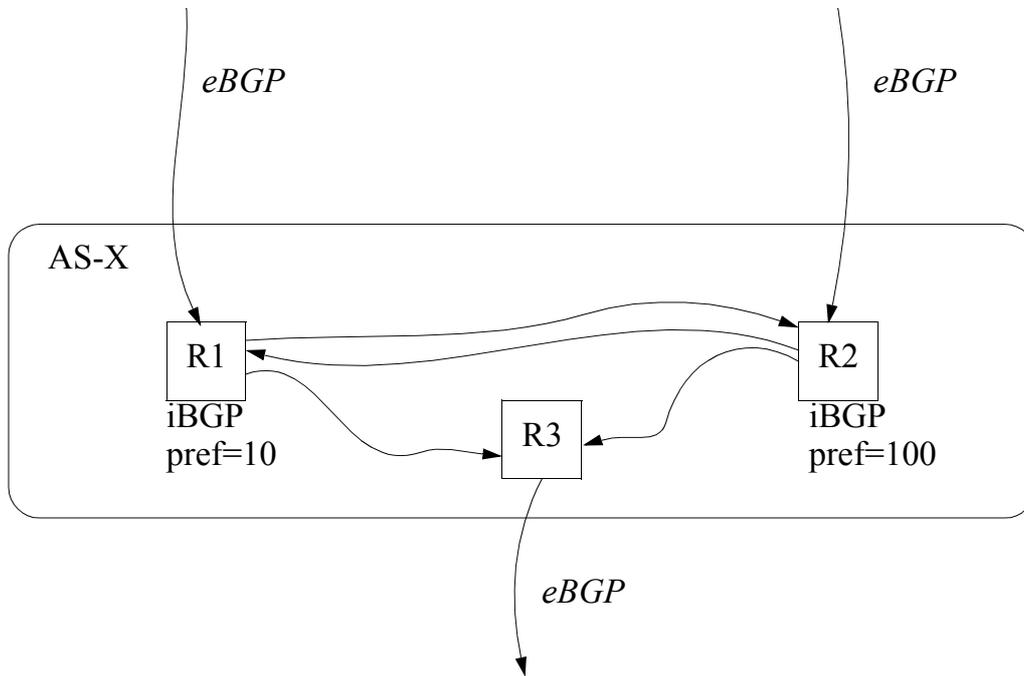
NEXT_HOP



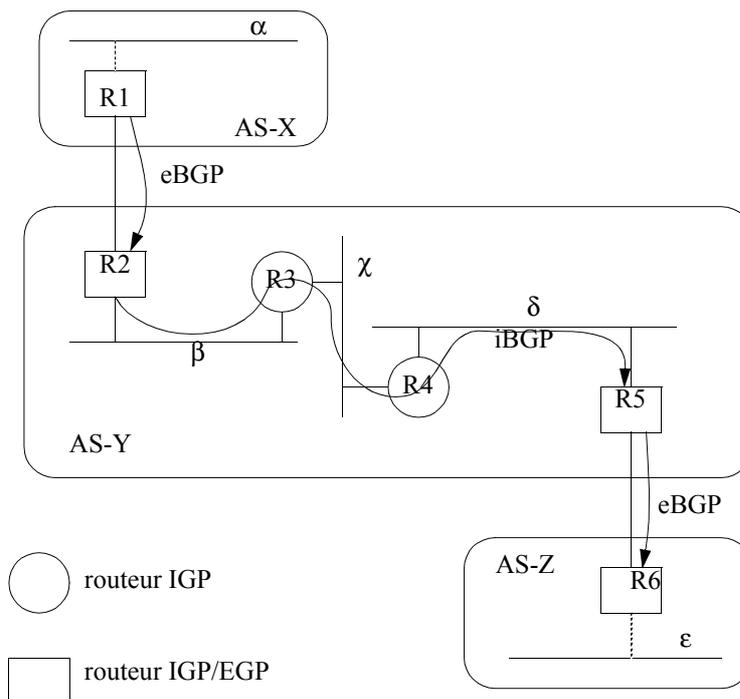
MULI_EXIT_DISC



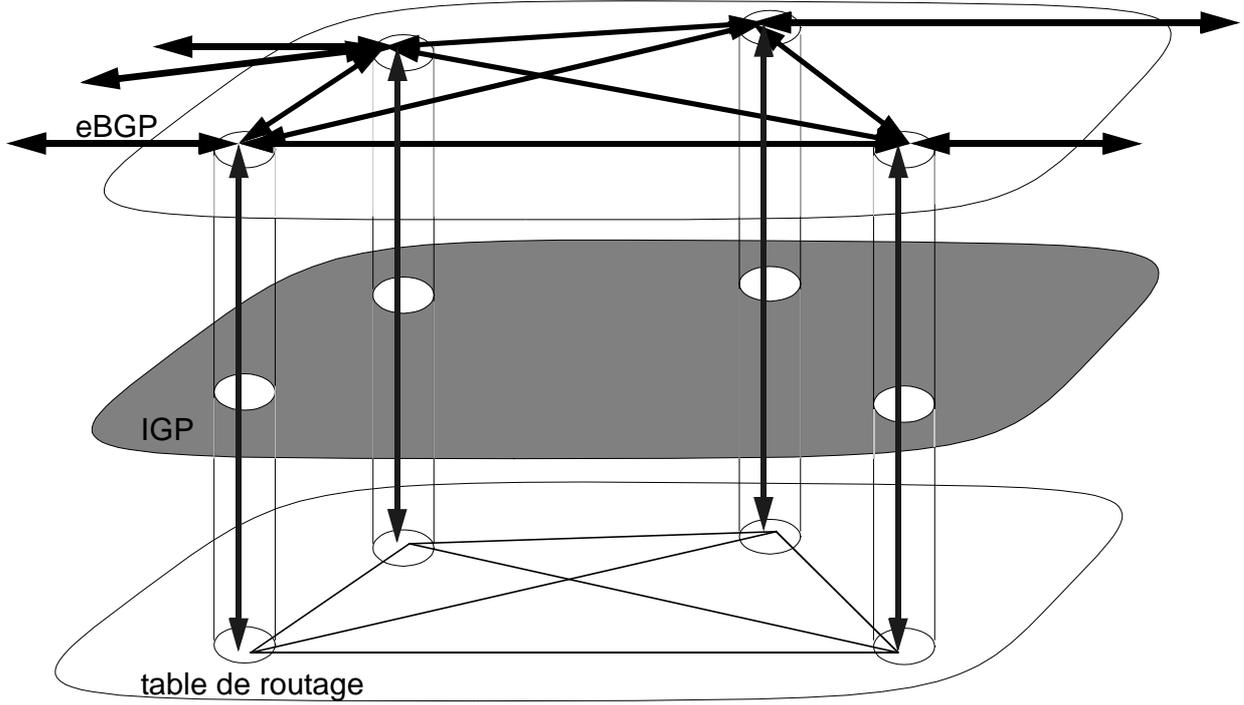
LOCAL_PREF



SYNCHRONISATION

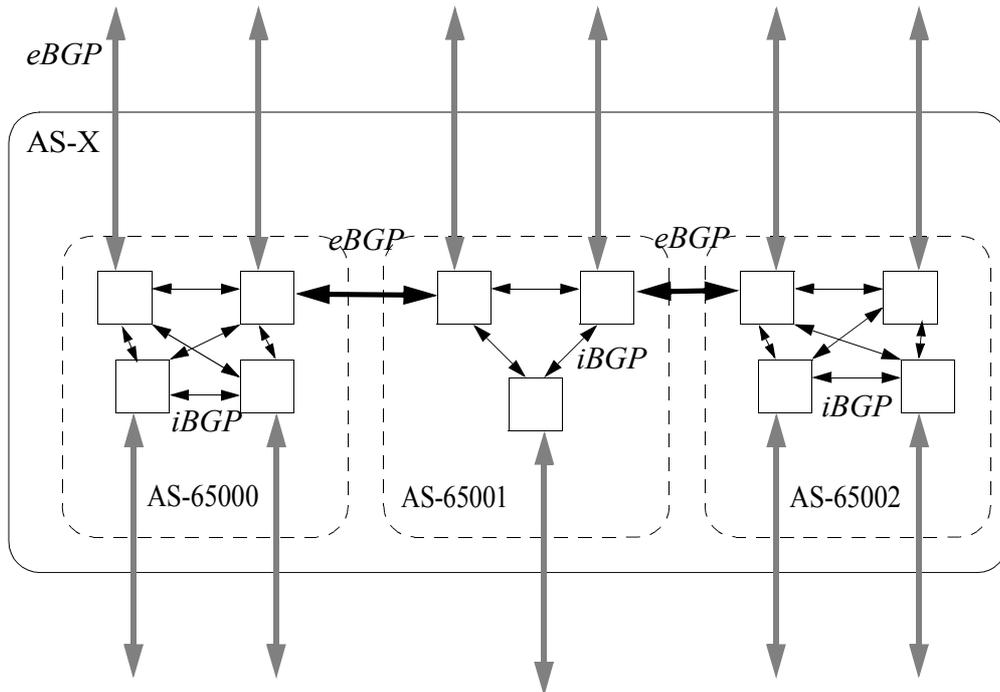


SYNCHRONISATION



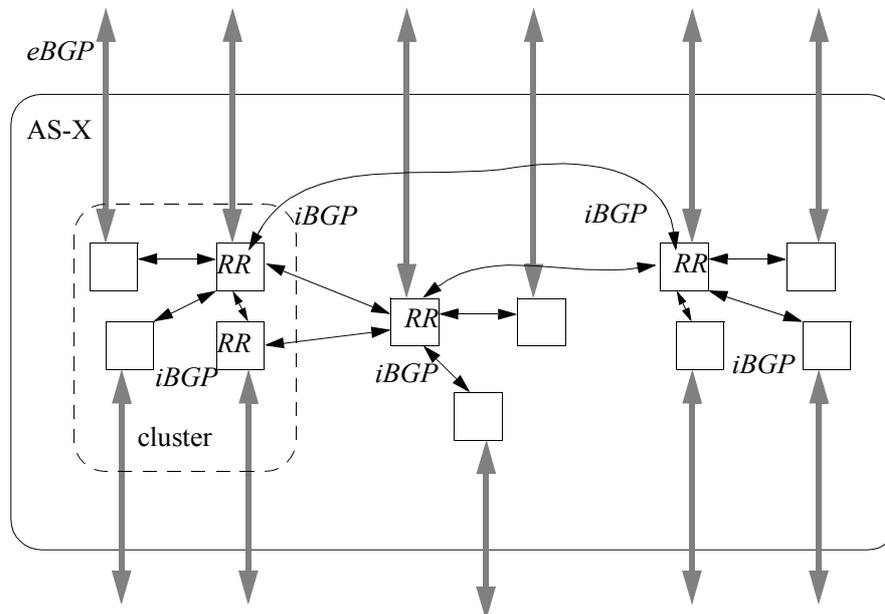
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Routing_Externe.fm

ATTRIBUTS POUR IBGP



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Routing_Externe.fm

ATTRIBUTS POUR IBGP



17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\Routing_Extene.fm

RÈGLES DE SELECTION

La stratégie est mise en œuvre :

- le routeur peut être configuré pour mettre en œuvre la politique de routage :
 - filtres AS : expressions régulières sur les chemins d'AS,
 - filtres sur les préfixes;
- poids local à une route en fonction soit de sa provenance (quel routeur a fait l'annonce) soit de sa composition (systèmes autonomes traversés);
- si les poids sont égaux, choisir l'annonce dont l'attribut LOCAL_PREF est le plus grand ;
- si les préférences locales identiques, choisir l'annonce avec le chemin d'AS le plus court ;
- si les chemins d'AS sont de même longueur, l'attribut ORIGIN est pris en considération : les annonces provenant d'un protocole de routage interne sont préférées à celles provenant d'un protocole de routage externe, elles-mêmes préférées à celles qui sont d'origine incomplète ;

17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\Routing_Extene.fm

- si les origines sont les mêmes, l'attribut MED est pris en considérations (si les annonces proviennent du même système autonome);
- sinon la métrique du protocole de routage interne vers le routeur indiqué dans l'attribut NEXT_HOP est prise en compte pour départager les routes identiques ;
- si la métrique est identique, les annonces provenant d'un protocole de routage interne sont préférées à celles venant de eBGP, elles mêmes préférées à celles venant de iBGP ;
- si les routes proviennent du même protocole de routage, arbitrairement, le routeur ayant le plus petit identificateur est pris en compte.

17 janvier 2003 - C:\Documents and Settings\loutain\Mes documents\Support_ENST\Routage_Externe.fm

EXEMPLE

```

router rip
  version 2
  redistribute connected
  passive-interface Vlan22
  passive-interface Vlan108
  passive-interface Vlan219
  passive-interface Vlan606
  passive-interface Vlan666
  network 192.108.119.0
  no auto-summary
!
router bgp 65525
  no synchronization
  bgp log-neighbor-changes
  network 10.35.0.0 mask 255.255.0.0
  network 192.108.119.0
  network 193.252.113.48 mask 255.255.255.252
  neighbor 192.108.119.136 remote-as 65525
  neighbor 192.108.119.136 description Aricie
  neighbor 193.50.69.74 remote-as 65525
  neighbor 193.50.69.74 description c6500-R11
  neighbor 193.252.113.49 remote-as 20603
  neighbor 193.252.113.49 description VTHD
  neighbor 193.252.113.49 prefix-list BGP-to-VTHD out
!
!
ip classless
ip route 0.0.0.0 0.0.0.0 193.50.69.73
ip route 10.35.30.0 255.255.255.0 192.108.119.160
ip route 192.168.0.0 255.255.0.0 193.50.69.74
no ip http server
!
!
ip prefix-list BGP-to-VTHD seq 5 permit 193.52.74.0/24
ip prefix-list BGP-to-VTHD seq 10 permit 192.108.119.0/24
ip prefix-list BGP-to-VTHD seq 15 permit 192.44.77.0/24
ip prefix-list BGP-to-VTHD seq 20 permit 192.108.117.192/26
ip prefix-list BGP-to-VTHD seq 25 permit 192.44.75.0/24
ip prefix-list BGP-to-VTHD seq 26 permit 192.44.76.0/24
ip prefix-list BGP-to-VTHD seq 30 permit 192.108.115.0/24
ip prefix-list BGP-to-VTHD seq 35 permit 192.108.116.0/24
ip prefix-list BGP-to-VTHD seq 36 permit 192.108.117.0/24
ip prefix-list BGP-to-VTHD seq 40 permit 192.108.118.0/24
ip prefix-list BGP-to-VTHD seq 45 permit 193.50.97.128/25

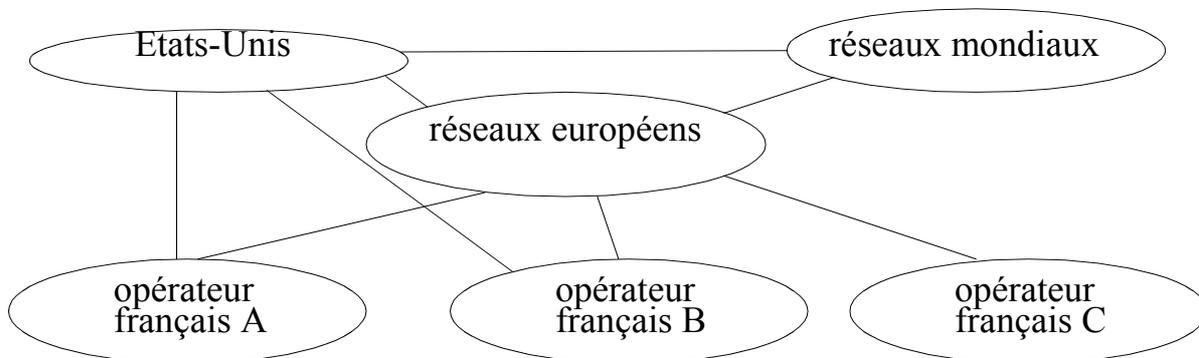
```



Pourquoi no synchronization puisqu'il y a un protocole de routage RIP qui tourne ?

17 janvier 2003 - C:\Documents and Settings\loutain\Mes documents\Support_ENST\Routage_Externe.fm

POINTS D'ÉCHANGE

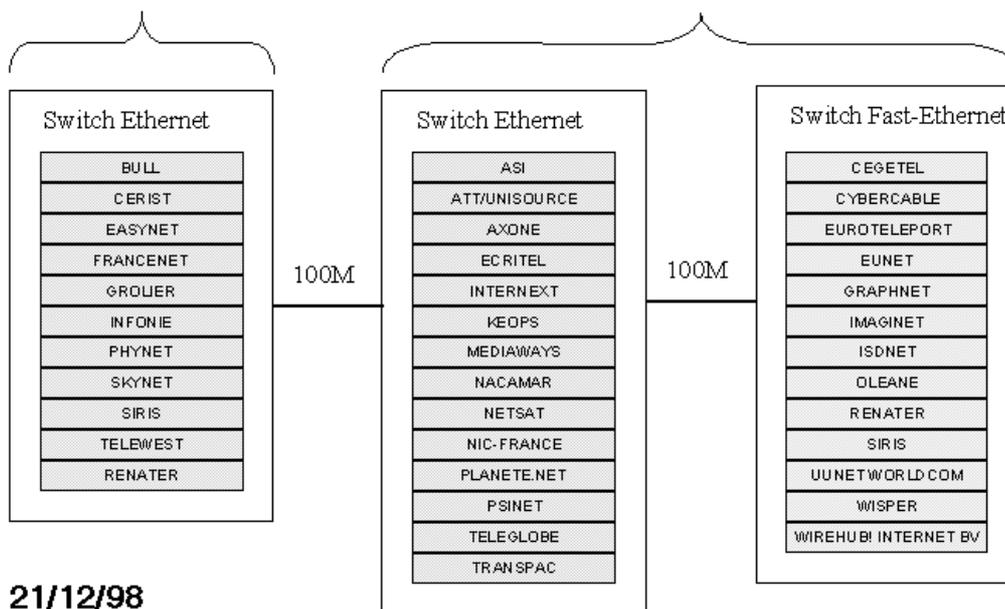


+ Eviter d'utiliser des liaisons internationales pour le trafic national

POINT D'INTERCONNEXION : FRANCE SFINX

SFINX géré / Managed SFINX

SFINX hébergé / Hosted SFINX



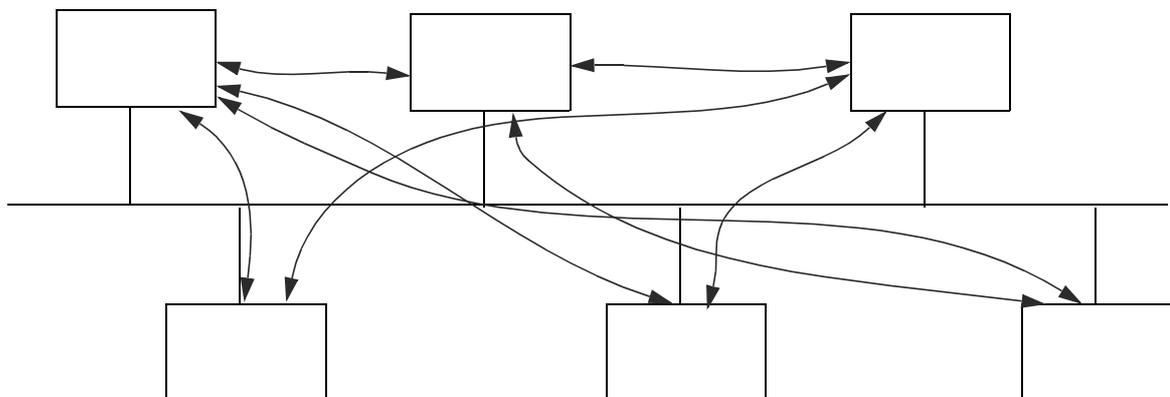
ECHANGE D'ANNONCES SUR LE SFINX

11/01/99			ASI	ATT	AXO	BUL	CER	CYB	CEG	EAS	ECR	EUN	EUR
			AS 6804	AS 3300	AS 2686	AS 6	AS 3208	AS 6678	AS 8228	AS 6727	AS 8304	AS 1899	AS 6771
ASI	ASI	AS 6804								X			
AT&T UNISOURCE	ATT	AS 3300											
AXONE IBM	AXO	AS 2686				X		X		X	X	X	X
BULL IBT	BUL	AS 6			X			X	X	X	X	X	X
CERIST	CER	AS 3208											
CYBERCABLE	CYB	AS 6678			X	X				X			X
CEGETEL	CEG	AS 8228				X				X	X		X
EASYNET	EAS	AS 6727	X		X	X		X	X		X		X
ECRITEL	ECR	AS 8304			X	X			X	X			X
EUNET France	EUN	AS 1899			X	X							
EUROTELEPORT	EUR	AS 6771			X	X		X	X	X	X		
FRANCENET	FRA	AS 5436	X		X					X	X		X
GRAPHNET	GRA	AS 7148								X	X		
GROLIER INTERACTIVE	GRO	AS 5410			X	X		X	X	X	X	X	X
IMAGINET	IMA	AS 6675						X	X	X	X	X	
INFONIE	INF	AS 8332	X		X			X			X		X
INTERNEXT	INE	AS 6761				X		X		X	X		X
ISDNET	ISD	AS 5534			X	X	X	X	X	X	X		X
KEOPS	KEO	AS 8740								X	X		

Points d'échange

353

NÉGOCIATION ENTRE OPÉRATEUR

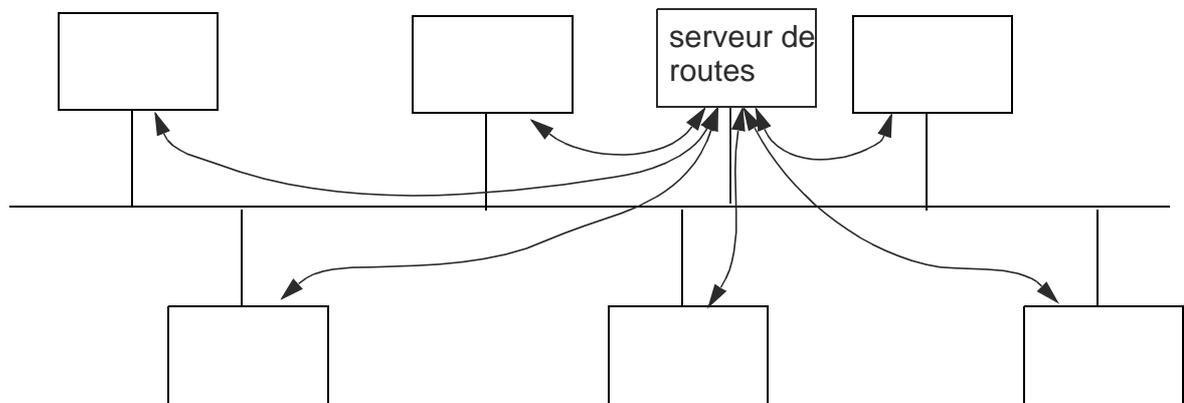


- le peering BGP est de la responsabilité de chaque opérateur

Points d'échange

354

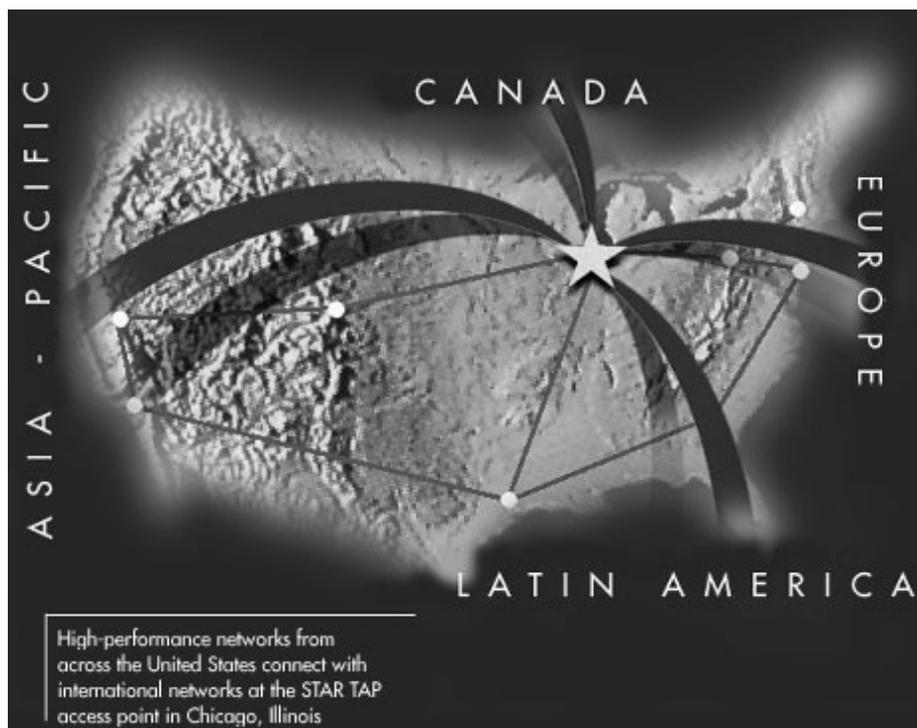
SERVEUR DE ROUTES



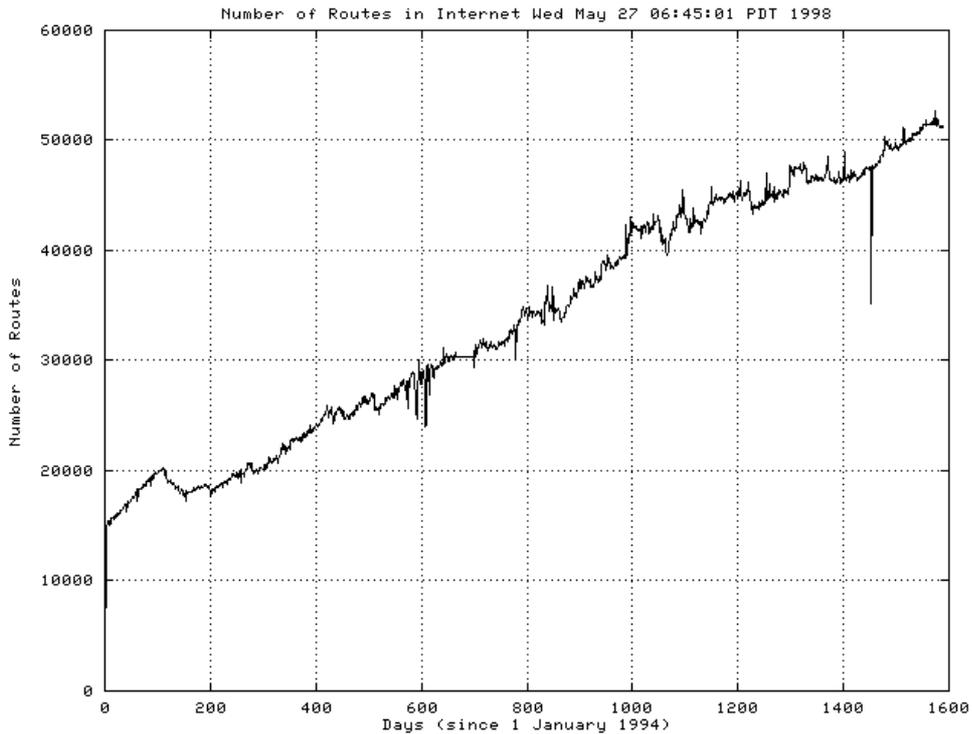
- Le serveur de route connaît la politique de routage des opérateurs
 - dérivée des bases de données comme RIPE

⇒ Est ce que les paquets de données passent par le serveur de route ?

LE CENTRE DU MONDE UNIVERSITAIRE



AGRÉGATION



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\ROUTE_Extene.fm

AGRÉGATION TRÈS CONTRÔLÉE

1) Gains by aggregating at the origin AS level

--- 12Feb99 ---

ASnum	NetsNow	NetsCIDR	NetGain	% Gain	Description
AS271	385	139	246	63.9%	BCnet Backbone
AS2493	275	157	118	42.9%	iSTAR Internet, Inc.
AS174	656	540	116	17.7%	Performance Systems International
AS11305	116	8	108	93.1%	UNKNOWN

...

List of possibly interesting aggregates

Aggregate | Origin | AS Description | Netname

166.49.146.0/23	AS9335	Concert Internet Plus	*
166.49.148.0/24	AS9335	Concert Internet Plus	
168.108.2.0/24	AS7741	OPCO-AS	*
168.108.3.0/24	AS7741	OPCO-AS	*
168.108.4.0/24	AS7741	OPCO-AS	*
168.108.5.0/24	AS7741	OPCO-AS	*
168.108.6.0/24	AS7741	OPCO-AS	*
168.108.7.0/24	AS7741	OPCO-AS	*

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\ROUTE_Extene.fm

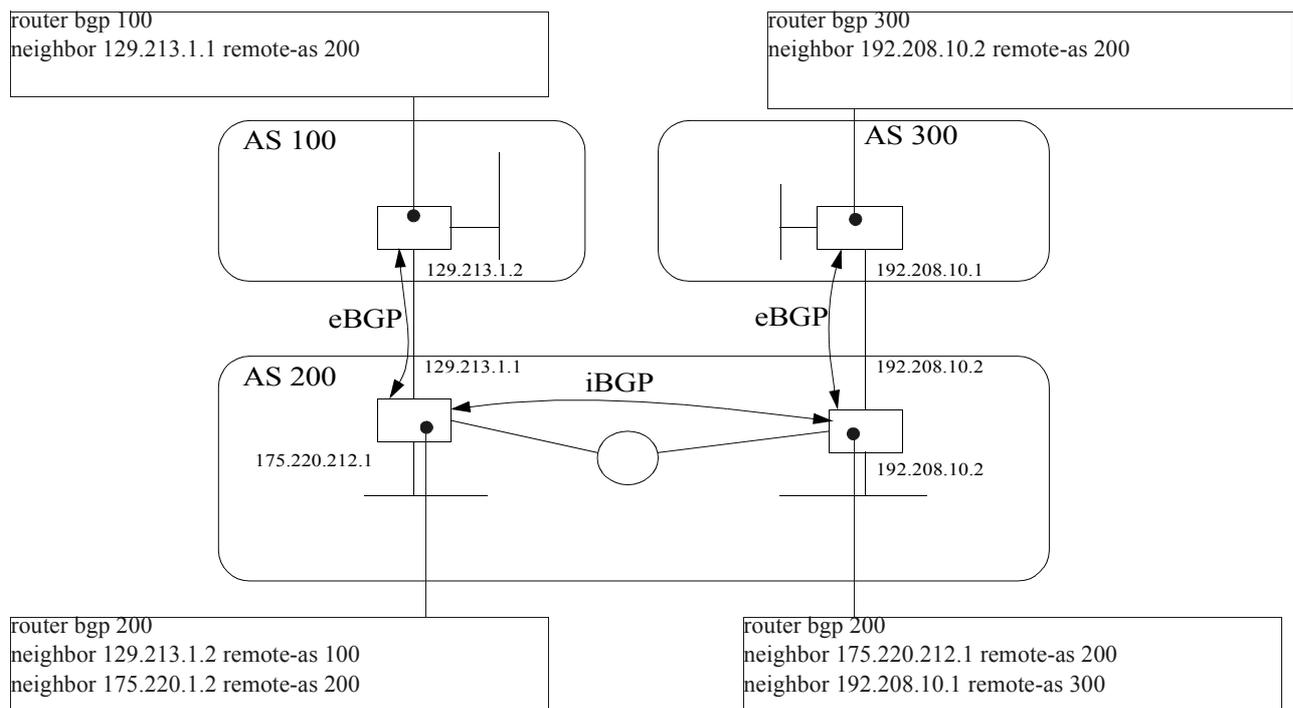
CONFIGURATION DE BGP SUR UN CISCO

+ Contrairement aux IGP, les EGP doivent être configurés :

- définir l'adresse du routeur (et le numéro de système autonome) vers lequel les NLRI seront envoyés
- accepter les NLRI d'un autre routeur (adresse IP et système autonome)
- définir des filtres pour n'émettre que les NLRI ayant certaines caractéristiques
 - utilisation d'expressions régulières
- définir des filtres pour n'accepter que les NLRI ayant certaines caractéristiques

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Routing_Externe.fm

EXEMPLE SIMPLE



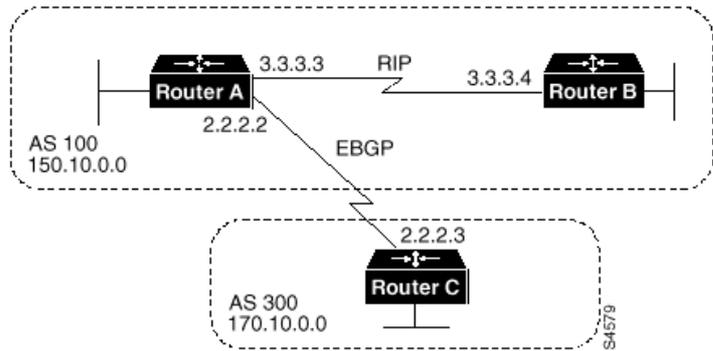
17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Routing_Externe.fm

TRAITEMENT DES ATTRIBUTS

!Router A

```

router rip
network 3.0.0.0
network 2.0.0.0
network 150.10.0.0
passive-interface serial 0
redistribute bgp 100 route-map SETMETRIC
!
router bgp 100
neighbor 2.2.2.3 remote-as 300
network 150.10.0.0
!
route-map SETMETRIC permit 10
match ip-address 1
set metric 2
!
route-map SETMETRIC permit 20
set metric 5
!
access-list 1 permit 170.10.0.0 0.0.255.255
    
```

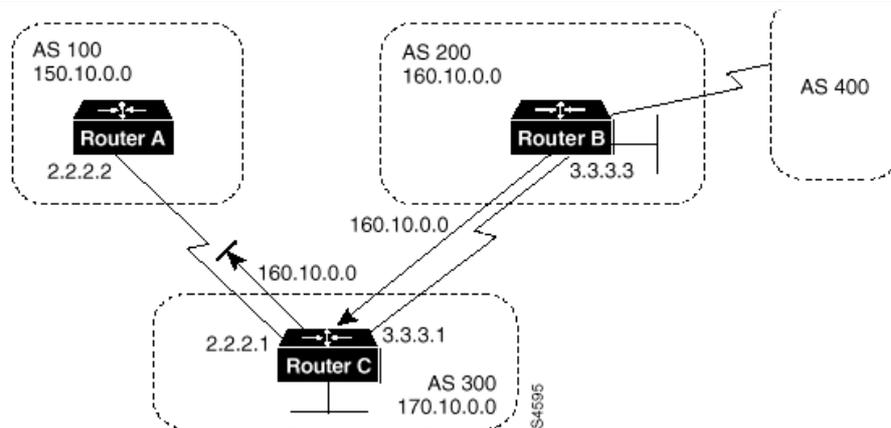


!Router C

```

router bgp 300
network 170.10.0.0
neighbor 2.2.2.2 remote-as 100
neighbor 2.2.2.2 route-map SETCOMMUNITY out
!
route-map SETCOMMUNITY permit 10
match ip address 1
set community 300
!
access-list 1 permit 0.0.0.0 255.255.255.255
    
```

FILTRAGE DES NRLLI

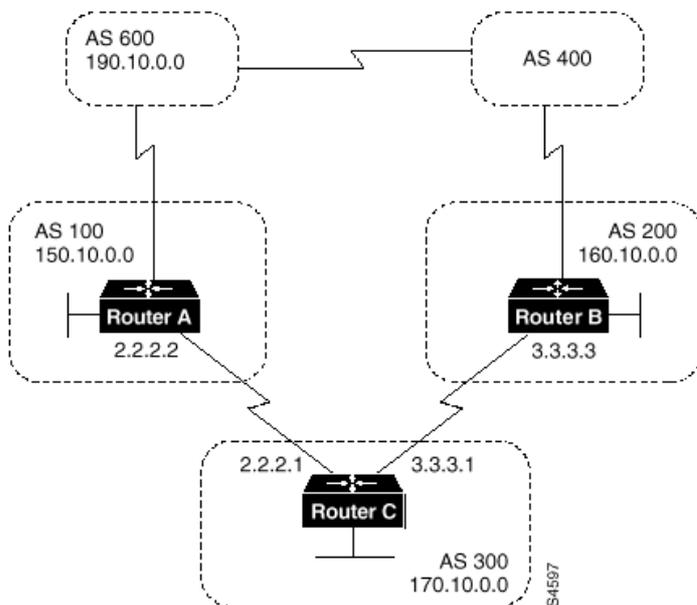


!Router C

```

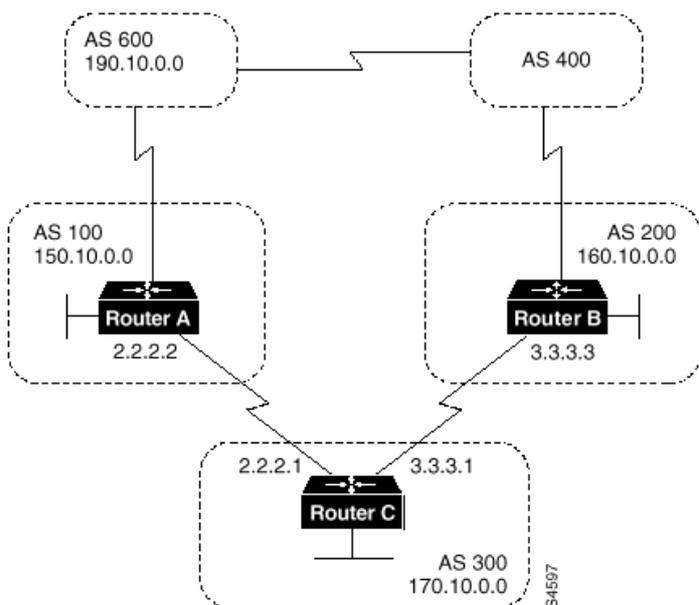
neighbor 3.3.3.3 remote-as 200
neighbor 2.2.2.2 remote-as 100
neighbor 2.2.2.2 filter-list 1 out
!
ip as-path access-list 1 deny ^200$
ip as-path access-list 1 permit .*
    
```

FILTRAGE DES NRLI



```
!Router C
router bgp 300
network 170.10.0.0
neighbor 3.3.3.3 remote-as 200
neighbor 3.3.3.3 route-map STAMP in
!
route-map STAMP permit 10
match as-path 1
set weight 20
!
ip as-path access-list 1 permit ^200$
```

FILTRAGE DES NRLI



```
!Router C
router bgp 300
network 170.10.0.0
neighbor 3.3.3.3 remote-as 200
neighbor 3.3.3.3 route-map STAMP in
!
route-map STAMP permit 10
match as-path 1
set weight 20
!
route-map STAMP permit 20
match as-path 2
!
route-map STAMP permit 30
set weight 10
!
ip as-path access-list 1 permit ^200$
ip as-path access-list 2 deny _400_
```

⇒ Que fait cette configuration ?

LES EXPRESSIONS RÉGULIÈRES

symbole		actions
point	.	correspond à n'importe quel caractère y compris un espace.
astérisque	*	correspond à 0 ou plus séquence du modif.
plus	+	correspond à 1 ou plus séquence du motif.
point d'interrogation	?	correspond à 0 ou 1 occurrence du motif.
circonflexe	^	correspond au début de la chaîne de caractères.
dollar	\$	correspond à la fin de la chaîne de caractères.
souligné	_	correspond aux caractères , { } () ^ \$ espace
crochet	[]	définit un ensemble d'un seul caractère
moins	-	sépare un interval de caractères
parenthèses	()	mémorise la séquence pour un traitement #1, #2

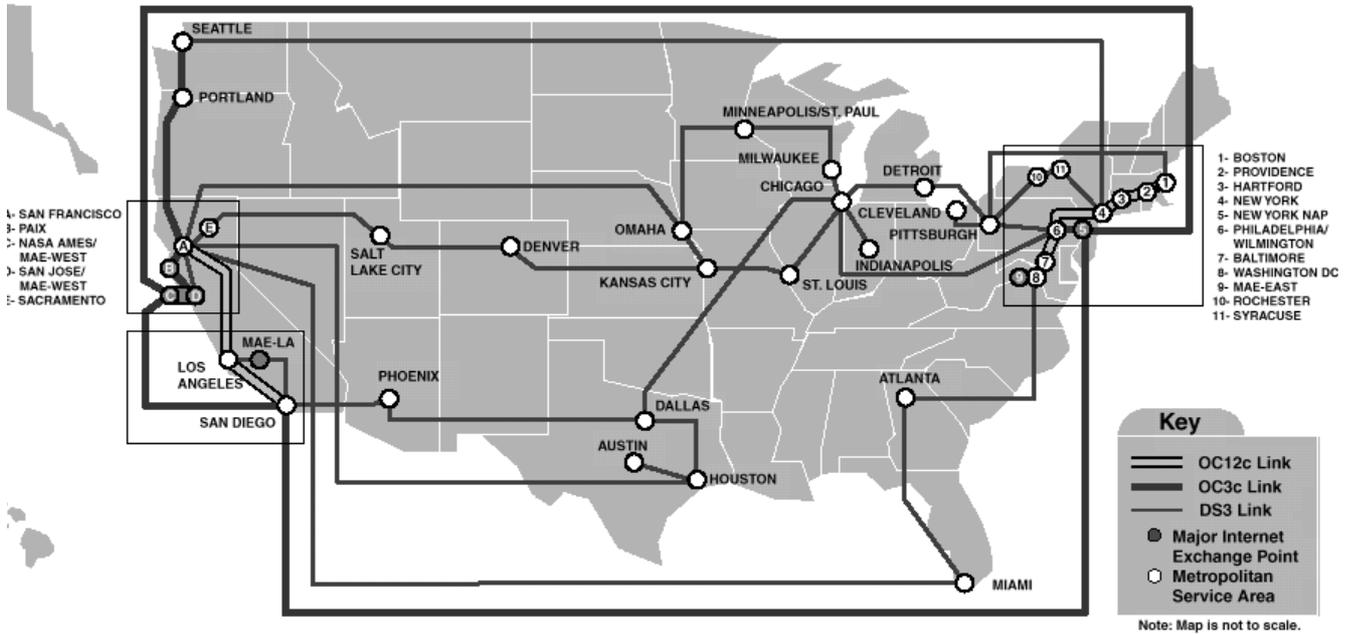
- `_100_` (via AS100)
- `^100$` (peering direct de l'AS100)
- `^.* 100` (venant de l'AS100)

LES OUTILS DE CONTRÔLE ET DE MESURE

+ Les Looking Glass :

- Possibilité de visualiser la configuration des routeurs (ou des serveurs de routes)
- Accessible :
 - telnet
 - page web.
- Par exemple :
 - <http://www.merit.edu/ipma/tools/lookingglass.html>
 - telnet://route-server.cerf.net
 - telnet://route-server.ip.att.net

EXEMPLE SUR ROUTE-SERVER.CERF.NET



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Routeage_Exeme.fm

EXEMPLE SUR ROUTE-SERVER.CERF.NET

ROUTE-SERVER.CERF.NET

CERFnet BGP Route Monitor

Reflects CERFnet's Peering Arrangements at the New York Sprint NAP, PAIX, MAE-West, Private Peers, and MAE-LA. This router should have the global BGP4 routing table.

The ip forwarding table used by this router does not always reflect the routing policy of CERFnet AS 1740. This router is in AS 1838 and maintains ebgp multihop peers with key CERFnet border routers.

192.157.69.5 == New York Sprint NAP 198.32.136.5 == MAE-West (NASA Side)
 198.32.176.25 == PAIX 198.32.136.55 == MAE-West (MFS Side)
 198.32.146.20 == MAE-LA

The corresponding topology is at:
<http://www.cerf.net/cerfnet/about/Bbone-map.html>

This route-server is off on a private ethernet off our San Diego nodes.
 Questions? email pushp@cerf.net

ROUTE-SERVER.CERF.NET

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Routeage_Exeme.fm

EXEMPLE SUR ROUTE-SERVER.CERF.NET

```
route-server.cerf.net>sh ip bgp summary
```

```
BGP table version is 8855242, main routing table version 8855242
53576 network entries (267670/267876 paths) using 14534372 bytes of memory
8467 BGP path attribute entries using 1161132 bytes of memory
0 BGP route-map cache entries using 0 bytes of memory
0 BGP filter-list cache entries using 0 bytes of memory
Dampening enabled. 397 history paths, 164 dampened paths
```

```
Neighbor    V  AS MsgRcvd MsgSent  TblVer  InQ  OutQ  Up/Down  State
192.157.69.5 4 1740 2569023 104319 8855231 0 0 4d00h
198.32.136.5 4 1740 2658338 104318 8855238 0 0 4d00h
198.32.136.55 4 1740 2658774 104264 8855238 0 0 4d00h
198.32.146.20 4 1740 2551705 104330 8855242 0 0 4d00h
198.32.176.25 4 1740 2512480 104318 8855231 0 0 4d00h
route-server.cerf.net>
```

EXEMPLE SUR ROUTE-SERVER.CERF.NET

```
route-server.cerf.net>sh ip bgp neighbors
```

```
BGP neighbor is 192.157.69.5, remote AS 1740, external link
Index 1, Offset 0, Mask 0x2
BGP version 4, remote router ID 134.24.127.201
BGP state = Established, table version = 8855381, up for 4d00h
Last read 00:00:02, hold time is 180, keepalive interval is 60 seconds
Minimum time between advertisement runs is 30 seconds
Received 2569071 messages, 0 notifications, 0 in queue
Sent 104322 messages, 0 notifications, 0 in queue
Outgoing update network filter list is 10
Connections established 4; dropped 3
External BGP neighbor may be up to 255 hops away.
Connection state is ESTAB, I/O status: 1, unread input bytes: 0
Local host: 192.215.254.5, Local port: 179
Foreign host: 192.157.69.5, Foreign port: 14070
```

```
Enqueued packets for retransmit: 0, input: 0, saved: 0
```

```
Event Timers (current time is 0x175660BA8):
```

Timer	Starts	Wakeups	Next
Retrans	5806	11	0x0
TimeWait	0	0	0x0
AckHold	36490	25431	0x0
SendWnd	0	0	0x0
KeepAlive	0	0	0x0
GiveUp	0	0	0x0

iss: 2275911518 snduna: 2276021615 sndnxt: 2276021615 sndwnd: 15928
irs: 2275848384 revnxt: 2286517440 revwnd: 16341 delrevwnd: 43

SRTT: 382 ms, RTTO: 1061 ms, RTV: 148 ms, KRTT: 0 ms
minRTT: 60 ms, maxRTT: 860 ms, ACK hold: 300 ms
Flags: passive open, nagle, gen tcbs

Datagrams (max data segment is 536 bytes):
Rcvd: 52815 (out of order: 43), with data: 47388, total data bytes: 10669055
Sent: 48132 (retransmit: 11), with data: 5794, total data bytes: 110096

⇒ Quel est le débit moyen de la connexion BGP ?

17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\ROUTE_Extene.fm

EXEMPLE SUR ROUTE-SERVER.CERF.NET

```
route-server.cerf.net>sh ip bgp
BGP table version is 8854815, local router ID is 39.7.46.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete
```

Network	Next Hop	Metric	LocPrf	Weight	Path
* 3.0.0.0	198.32.136.55	0	1740	701	80 i
*	198.32.136.5	0	1740	701	80 i
*	192.157.69.5	0	1740	701	80 i
*>	198.32.146.20	0	1740	701	80 i
*	198.32.176.25	0	1740	701	80 i
* 4.0.0.0	198.32.136.55	0	1740	1	i
*	198.32.136.5	0	1740	1	i
*	192.157.69.5	0	1740	1	i
*>	198.32.146.20	0	1740	1	i
* 6.0.0.0	198.32.136.55	0	1740	568	721 1455 i
*	198.32.136.5	0	1740	568	721 1455 i
*	192.157.69.5	0	1740	568	721 1455 i
*>	198.32.146.20	0	1740	568	721 1455 i
*	198.32.176.25	0	1740	568	721 1455 i
* 9.2.0.0/16	198.32.136.55	0	1740	1673	1675 i
*	198.32.136.5	0	1740	1673	1675 i
*	192.157.69.5	0	1740	1673	1675 i
*>	198.32.146.20	0	1740	1673	1675 i

17 janvier 2003 - C:\Documents and Settings\clouain\Mes documents\Support_ENST\ROUTE_Extene.fm

EXEMPLE SUR ROUTE-SERVER.CERF.NET

```

route-server.cerf.net>sh ip bgp regexp _1717_
BGP table version is 8774930, local router ID is 39.7.46.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

  Network      Next Hop      Metric LocPrf Weight Path
* 128.93.0.0   198.32.136.55      0 1740 1239 1800 5511 1717 1309 ?
*              198.32.136.5      0 1740 1239 1800 5511 1717 1309 ?
*              192.157.69.5      0 1740 1239 1800 5511 1717 1309 ?
*>             198.32.146.20     0 1740 1239 1800 5511 1717 1309 ?
*              198.32.176.25     0 1740 1239 1800 5511 1717 1309 ?
* 129.20.0.0   198.32.136.55      0 1740 1239 5511 1717 1938 ?
*              198.32.136.5      0 1740 1239 5511 1717 1938 ?
*              192.157.69.5      0 1740 1239 5511 1717 1938 ?
*>             198.32.146.20     0 1740 1239 5511 1717 1938 ?
*              198.32.176.25     0 1740 1239 5511 1717 1938 ?
* 129.88.0.0   198.32.136.55      0 1740 1239 5511 1717 2477 ?
*              198.32.136.5      0 1740 1239 5511 1717 2477 ?
*              192.157.69.5      0 1740 1239 5511 1717 2477 ?
*>             198.32.146.20     0 1740 1239 5511 1717 2477 ?
*              198.32.176.25     0 1740 1239 5511 1717 2477 ?

```

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\ROUTE_Support_ENST\ROUTE_Externe.fm

EXEMPLE SUR ROUTE-SERVER.CERF.NET

```

route-server.cerf.net>sh ip bgp regexp ^.* 1717$
BGP table version is 8853425, local router ID is 39.7.46.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

```

Network	Next Hop	Metric	LocPrf	Weight	Path
* 132.166.0.0/15	198.32.136.55	0	1740	1239	5511 1717 i
*	198.32.136.5	0	1740	1239	5511 1717 i
*	192.157.69.5	0	1740	1239	5511 1717 i
*>	198.32.146.20	0	1740	1239	5511 1717 i
*	198.32.176.25	0	1740	1239	5511 1717 i
* 132.168.0.0/15	198.32.136.55	0	1740	1239	5511 1717 i
*	198.32.136.5	0	1740	1239	5511 1717 i
*	192.157.69.5	0	1740	1239	5511 1717 i
*>	198.32.146.20	0	1740	1239	5511 1717 i
*	198.32.176.25	0	1740	1239	5511 1717 i
* 161.104.0.0/15	198.32.136.55	0	1740	1239	5511 1717 i
*	198.32.136.5	0	1740	1239	5511 1717 i
*	192.157.69.5	0	1740	1239	5511 1717 i
*>	198.32.146.20	0	1740	1239	5511 1717 i
*	198.32.176.25	0	1740	1239	5511 1717 i
* 164.81.0.0	198.32.136.55	0	1740	1239	5511 1717 ?
*	198.32.136.5	0	1740	1239	5511 1717 ?
*	192.157.69.5	0	1740	1239	5511 1717 ?

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\ROUTE_Support_ENST\ROUTE_Externe.fm

EXEMPLE SUR ROUTE-SERVER.CERF.NET

- Afficher les annonces de routes passant par OPEN TRANSIT (As 5511) et originaire de Renater (AS1717)

```
route-server.cerf.net>sh ip bgp regexp ^.* 5511 .*1717$
BGP table version is 8853957, local router ID is 39.7.46.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete
```

Network	Next Hop	Metric	LocPrf	Weight	Path
* 132.166.0.0/15	198.32.136.55		0	1740 1239 5511 1717	i
*	198.32.136.5		0	1740 1239 5511 1717	i
*	192.157.69.5		0	1740 1239 5511 1717	i
*>	198.32.146.20		0	1740 1239 5511 1717	i
*	198.32.176.25		0	1740 1239 5511 1717	i
* 132.168.0.0/15	198.32.136.55		0	1740 1239 5511 1717	i
*	198.32.136.5		0	1740 1239 5511 1717	i
*	192.157.69.5		0	1740 1239 5511 1717	i
*>	198.32.146.20		0	1740 1239 5511 1717	i
*	198.32.176.25		0	1740 1239 5511 1717	i
* 161.104.0.0/15	198.32.136.55		0	1740 1239 5511 1717	i
*	198.32.136.5		0	1740 1239 5511 1717	i
*	192.157.69.5		0	1740 1239 5511 1717	i
*>	198.32.146.20		0	1740 1239 5511 1717	i
*	198.32.176.25		0	1740 1239 5511 1717	i

EXEMPLE SUR ROUTE-SERVER.CERF.NET

- Pour un NLRI :

```
route-server.cerf.net>sh ip bgp 192.108.119.0
BGP routing table entry for 192.108.119.0/24, version 8842821
Paths: (5 available, best #4)
 1740 1239 5511 1717 1938
 198.32.136.5 from 198.32.136.5 (134.24.127.29)
  Origin incomplete, valid, external
 1740 1239 5511 1717 1938
 198.32.136.55 from 198.32.136.55 (134.24.127.27)
  Origin incomplete, valid, external
 1740 1239 5511 1717 1938
 192.157.69.5 from 192.157.69.5 (134.24.127.201)
  Origin incomplete, valid, external
 1740 1239 5511 1717 1938
 198.32.146.20 from 198.32.146.20 (134.24.127.23)
  Origin incomplete, valid, external, best
 1740 1239 5511 1717 1938
 198.32.176.25 from 198.32.176.25 (134.24.127.35)
  Origin incomplete, valid, external
route-server.cerf.net>
```

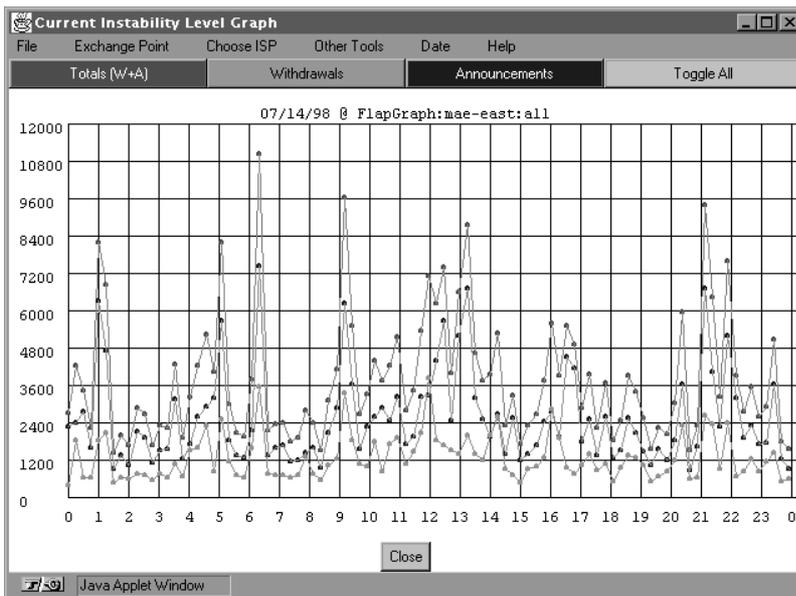
EXEMPLE SUR ROUTE-SERVER.CERF.NET

- Pour un NLRI :

```
route-server.cerf.net>sh ip bgp 24.48.0.0
BGP routing table entry for 24.48.0.0/18, version 8543708
Paths: (5 available, best #4)
 1740 1673 {1324,1329,1327,1325}, (aggregated by 1673 140.223.220.62)
   198.32.136.55 from 198.32.136.55 (134.24.127.27)
   Origin incomplete, valid, external, atomic-aggregate
 1740 1673 {1324,1329,1327,1325}, (aggregated by 1673 140.223.12.62)
   198.32.136.5 from 198.32.136.5 (134.24.127.29)
   Origin incomplete, valid, external, atomic-aggregate
 1740 1673 {1324,1329,1327,1325}, (aggregated by 1673 140.223.217.62)
   192.157.69.5 from 192.157.69.5 (134.24.127.201)
   Origin incomplete, valid, external, atomic-aggregate
 1740 1673 {1324,1329,1327,1325}, (aggregated by 1673 140.223.220.62)
   198.32.146.20 from 198.32.146.20 (134.24.127.23)
   Origin incomplete, valid, external, atomic-aggregate, best
 1740 1673 {1324,1329,1327,1325}, (aggregated by 1673 140.223.220.62)
   198.32.176.25 from 198.32.176.25 (134.24.127.35)
   Origin incomplete, valid, external, atomic-aggregate
```

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Routing_Externe.fm

STATIBILITÉ DES ROUTES



Pour 42 000 routes, entre 3 et 6 millions d'annonces et de retraits par jour.

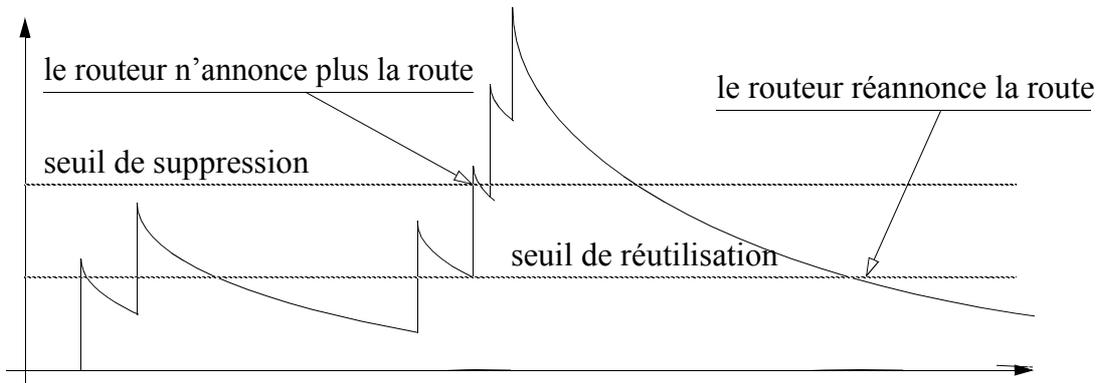
Comportement oscillatoire mal compris

Réductions des oscillations par rejet des routes instables

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Routing_Externe.fm

DAMPERING

- Décroissance exponentielle



17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\ROUTAGE_Externe.fm

EXEMPLE SUR ROUTE-SERVER.CERF.NET

```
route-server.cerf.net>sh ip bgp dampened-paths
BGP table version is 8859804, local router ID is 39.7.46.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete
```

Network	From	Reuse	Path
*d 203.55.55.0	198.32.136.55	00:00:1	1740 2828 3840 4433 4740 i
*d 203.55.55.0	192.157.69.5	00:00:1	1740 2828 3840 4433 4740 i
*d 203.55.55.0	198.32.146.20	00:00:2	1740 2828 3840 4433 4740 i
*d 203.55.54.0	198.32.176.25	00:00:4	1740 2828 3840 4433 4740 i
*d 203.26.142.0	198.32.176.25	00:01:0	1740 2828 3840 4433 4740 i
*d 203.93.64.0/19	198.32.136.55	00:01:3	1740 1 5727 4799 4799 4799 4799 4799 i
*d 203.55.54.0	198.32.136.5	00:01:4	1740 2828 3840 4433 4740 i
*d 203.26.142.0	192.157.69.5	00:02:1	1740 2828 3840 4433 4740 i
*d 203.26.142.0	198.32.136.5	00:02:3	1740 2828 3840 4433 4740 i
*d 202.54.73.0	192.157.69.5	00:02:4	1740 701 6453 4755 i
*d 202.54.73.0	198.32.146.20	00:02:5	1740 701 6453 4755 i
*d 194.110.137.0	192.157.69.5	00:02:5	1740 701 6453 5515 i
*d 194.110.137.0	198.32.146.20	00:03:0	1740 701 6453 5515 i
*d 202.54.73.0	198.32.136.55	00:02:5	1740 701 6453 4755 i
*d 194.110.137.0	198.32.136.55	00:03:0	1740 701 6453 5515 i
*d 194.110.137.0	198.32.176.25	00:03:0	1740 701 6453 5515 i
*d 194.110.137.0	198.32.136.5	00:03:1	1740 701 6453 5515 i
*d 194.42.64.0/19	192.157.69.5	00:03:3	1740 174 8218 i

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\ROUTAGE_Externe.fm

EXEMPLE SUR ROUTE-SERVER.CERF.NET

route-server.cerf.net>**sh ip bgp flap-statistics**

BGP table version is 8860218, local router ID is 39.7.46.1

Status codes: s suppressed, d damped, h history, * valid, > best, i - internal

Origin codes: i - IGP, e - EGP, ? - incomplete

Network	From	Flaps	Duration	Reuse	Path
*d 24.232.0.0/19	198.32.136.55	5	00:06:18	00:21:40	1740 7018 3561 6813 4926 4926 4926 4926 7908 10318
*d	198.32.136.5	5	00:06:07	00:21:40	1740 7018 3561 6813 4926 4926 4926 4926 7908 10318
*d	192.157.69.5	5	00:06:06	00:21:40	1740 7018 3561 6813 4926 4926 4926 4926 7908 10318
*d	198.32.146.20	6	00:06:01	00:29:00	1740 7018 3561 6813 4926 4926 4926 4926 7908 10318
*d	198.32.176.25	5	00:06:23	00:21:40	1740 7018 3561 6813 4926 4926 4926 4926 7908 10318
h 53.122.48.0/20	198.32.176.25	1	00:13:13		1740 1800 1755 1273 6805 6878
h	198.32.136.5	1	00:13:16		1740 1800 1755 1273 6805 6878
h	192.157.69.5	2	00:13:19		1740 1239 1800 1755 1273 6805 6878
h	198.32.146.20	2	00:13:33		1740 1239 1800 1755 1273 6805 6878
h	198.32.136.55	2	00:13:20		1740 1239 1800 1755 1273 6805 6878
* 62.172.0.0/16	198.32.136.5	2	00:21:46		1740 2548 2856
*>	198.32.146.20	2	00:22:04		1740 2548 2856
* 128.8.0.0	198.32.136.55	2	00:21:54		1740 2548 27
*	198.32.136.5	2	00:21:46		1740 2548 27
*	192.157.69.5	2	00:21:51		1740 2548 27
*>	198.32.146.20	2	00:22:04		1740 2548 27
*	198.32.176.25	2	00:21:49		1740 2548 27
* 128.164.0.0	198.32.136.55	3	00:20:57		1740 3951 6350 11039

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Routing_Externe.fm

BGP MULTIPROTOCOLE

- La solution ISO IDRIP (Inter Domain Routing Protocol) ne sera jamais déployée dans l'Internet
- Le passage de BGP3 à BGP4 a *traumatisé* les opérateurs
- Evolution en douceur de BGP
- BGP Multiprotocol (RFC 2858)
- Deux extensions :
 - Négociation des capacités des deux équipements,
 - Nouvel attribut MP_REACH_NLRI
- Utilisé par IPv6 (BGP4+), le Multicast, MPLS,...

17 janvier 2003 - C:\Documents and Settings\louain\Mes documents\Support_ENST\Routing_Externe.fm

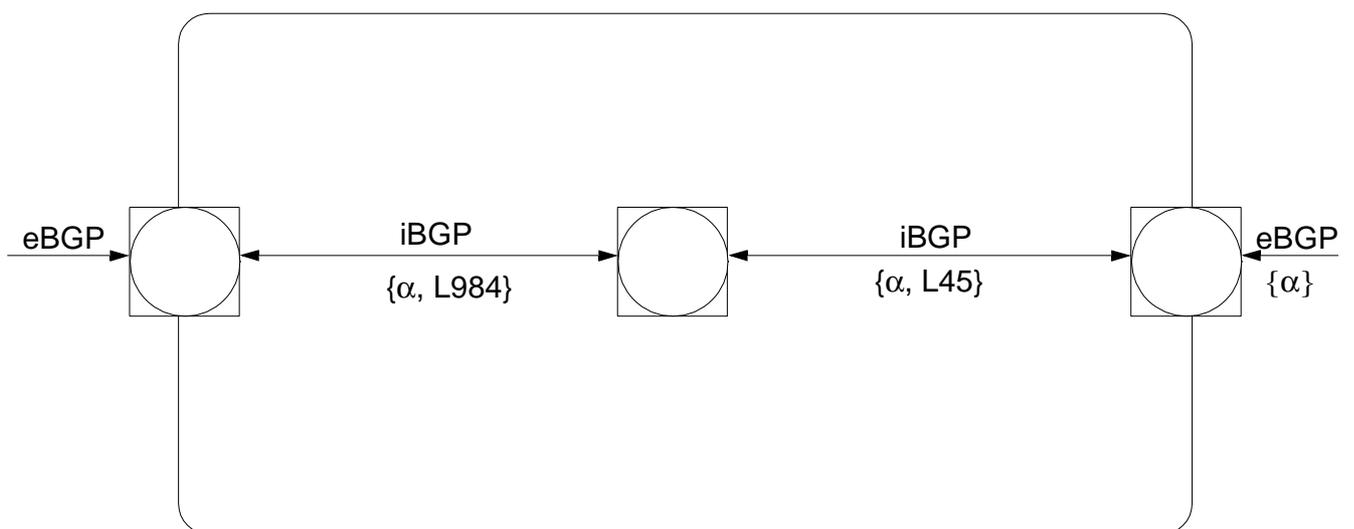
BGP & MPLS

Evite l'utilisation d'un IGP :

- pour y injecter les 100 000 routes de l'Internet,
- l'IGP peut servir pour les routes internes de l'opérateur.
- Si deux routeurs BGP adjacents sont aussi des LSR pas besoin de protocole de distribution d'étiquettes.
- L'attribut multiprotocole permet de distribuer la pile d'étiquettes et le préfixe associé.

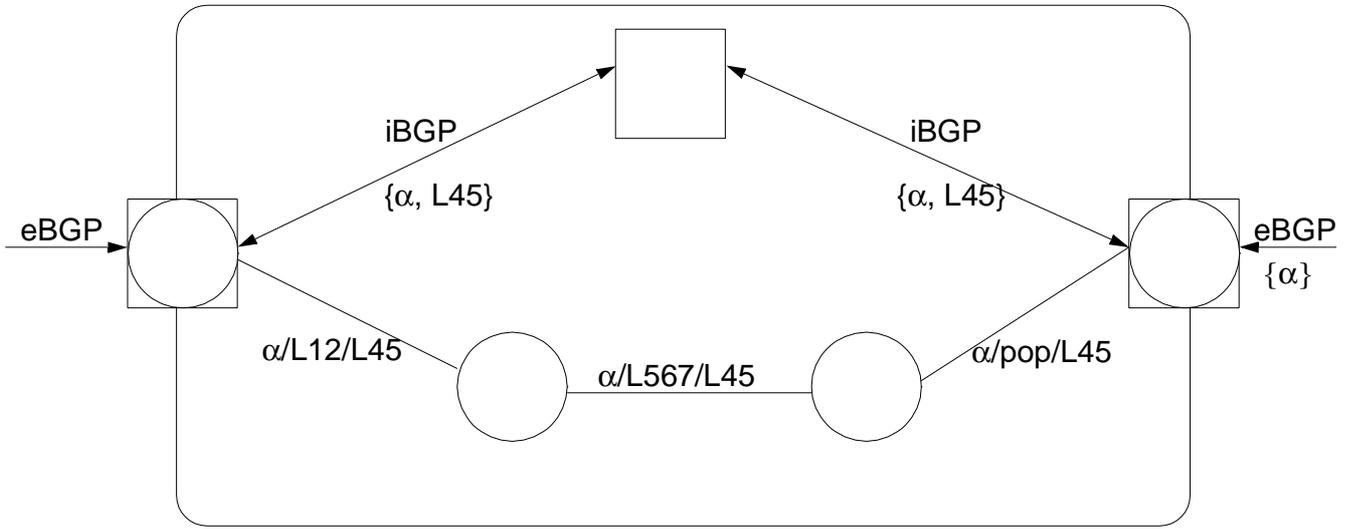
17 janvier 2003 - C:\Documents and Settings\loutain\Mes documents\Support_ENST\Routing_Externe.fm

CAS SANS IGP



17 janvier 2003 - C:\Documents and Settings\loutain\Mes documents\Support_ENST\Routing_Externe.fm

CAS AVEC IGP



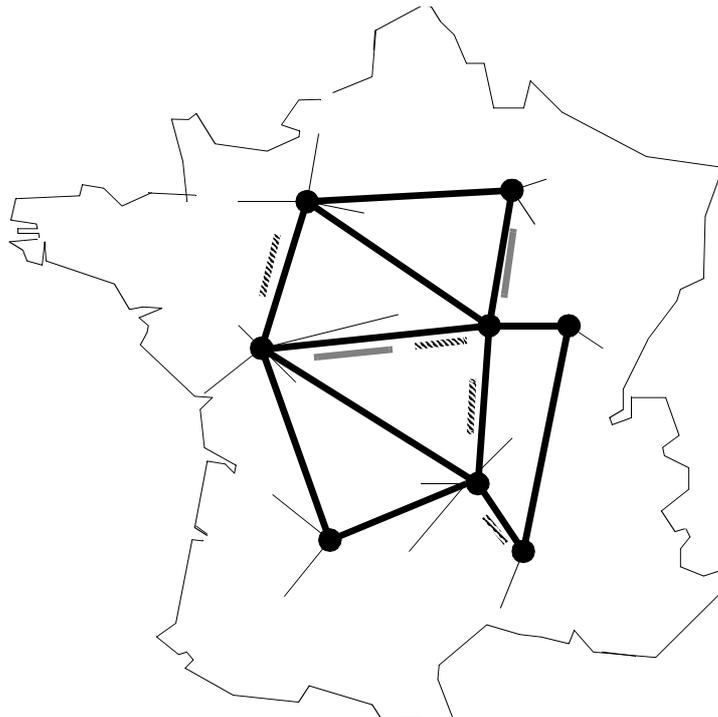
17 janvier 2003 - C:\Documents and Settings\clair\Mes documents\Support_ENST\ROUTAGE_Extreme.fm

Le multimédia et les réseaux

Laurent Toutain

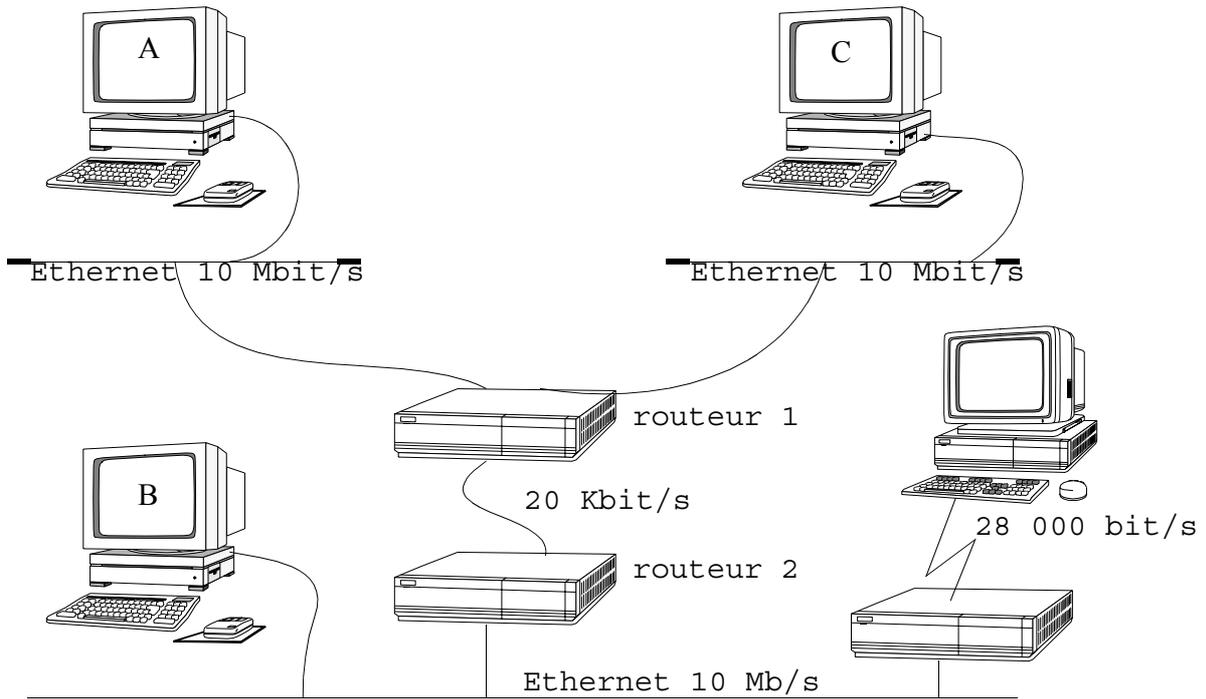
386

Rappel sur les architectures de réseau



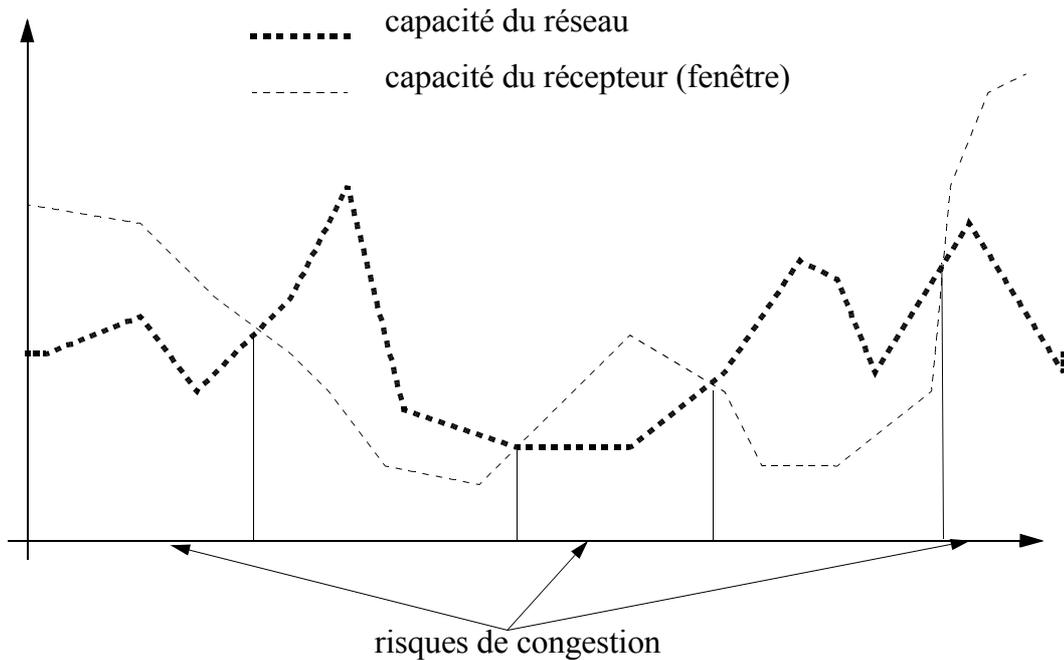
387

Internet : réseau "best-effort"



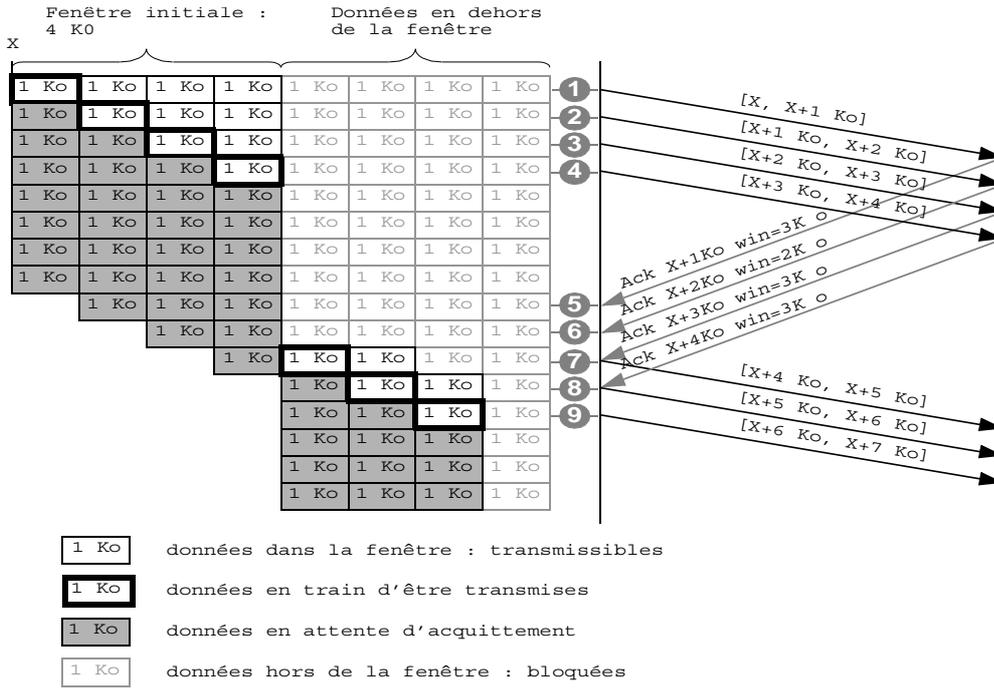
17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fm

Deux fenêtres de Congestion



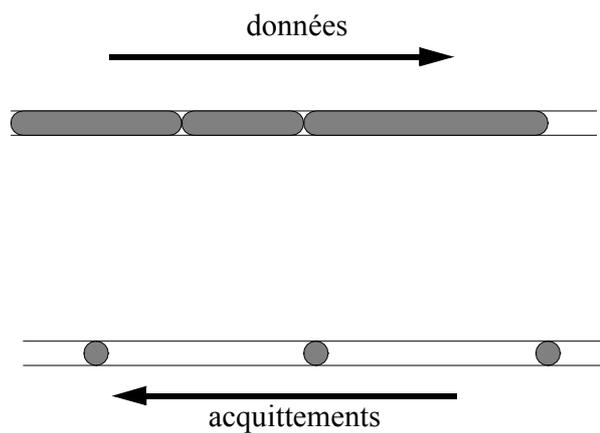
17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fm

Fenêtre d'anticipation



17 janvier 2003 - C:\Documents and Settings\vaunain\Mes documents\Support_ENST_QoSDE4.fm

Régime Stationnaire



17 janvier 2003 - C:\Documents and Settings\vaunain\Mes documents\Support_ENST_QoSDE4.fm

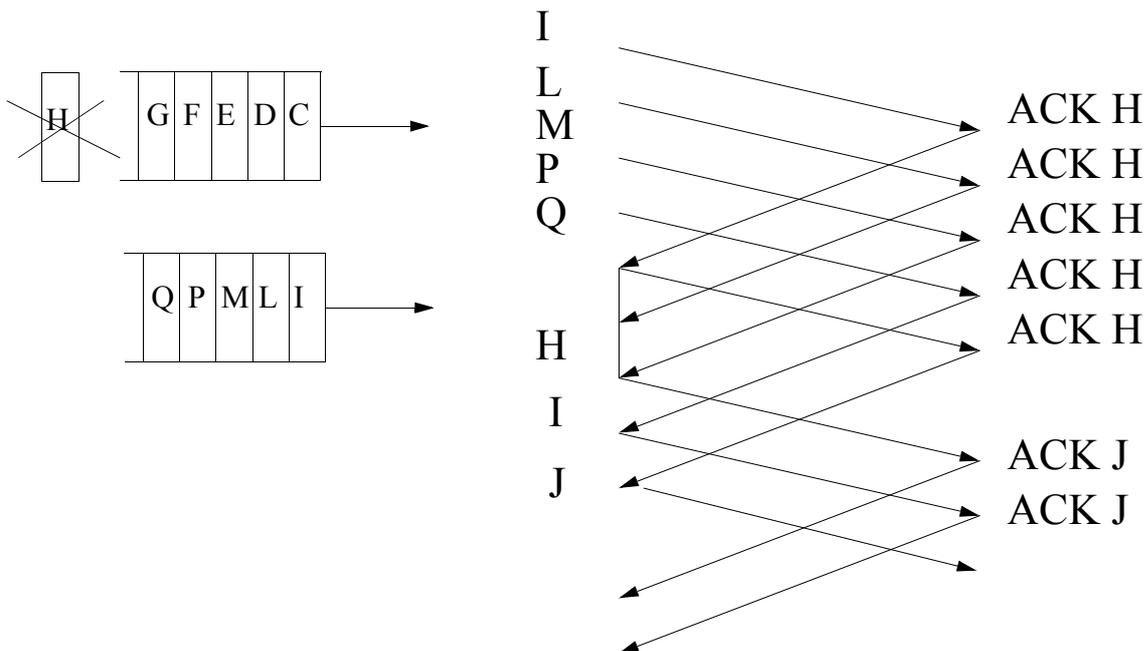
Fenêtre de congestion (algorithme du slow start)

- Soit la variable `seuil` initialement à 65 535 et la `taille` de la fenêtre initialement à 1.
- A chaque réception d'un acquittement, ajouter 1 à la `taille` de la fenêtre. La `taille` de la fenêtre croît exponentiellement.
- Si la `taille` de la fenêtre atteint le `seuil`, incrémenter linéairement sa `taille`.
- Si un paquet est perdu (réception de trois acquittements identiques) alors :
 $taille \text{ de la fen\^etre} = taille \text{ de la fen\^etre} / 2$
- Si pertes de plusieurs paquets (déclenchement du timer) alors :
 $seuil = taille \text{ de la fen\^etre} / 2$ (le `seuil` ne peut pas être inférieur à 2 segments)
 et
 $taille \text{ de la fen\^etre} = 1$

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fm

392

Exemple

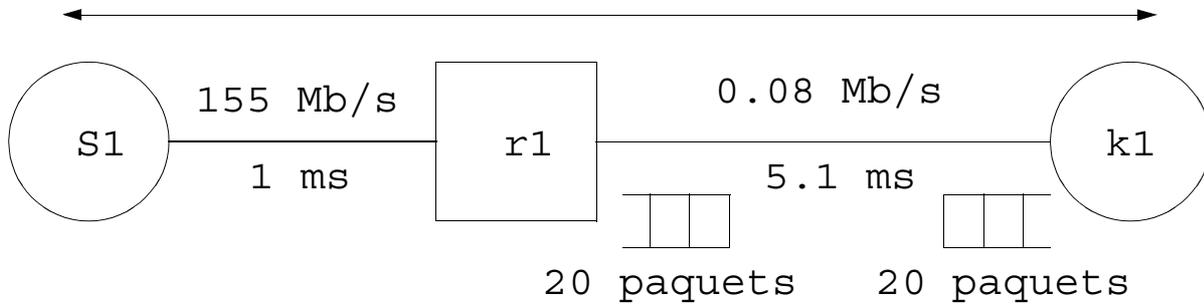


17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fm

393

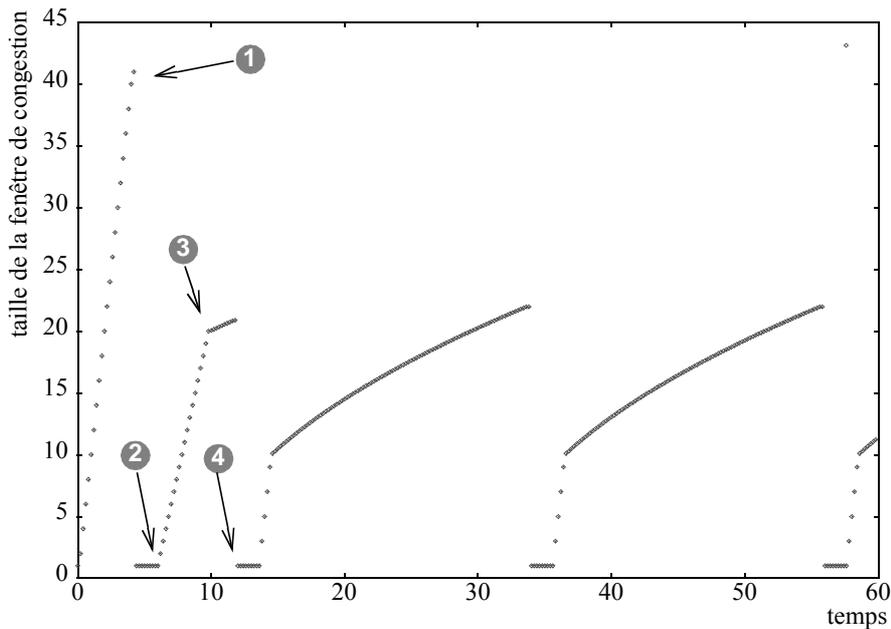
Exemple par simulation

tcp Tahoe



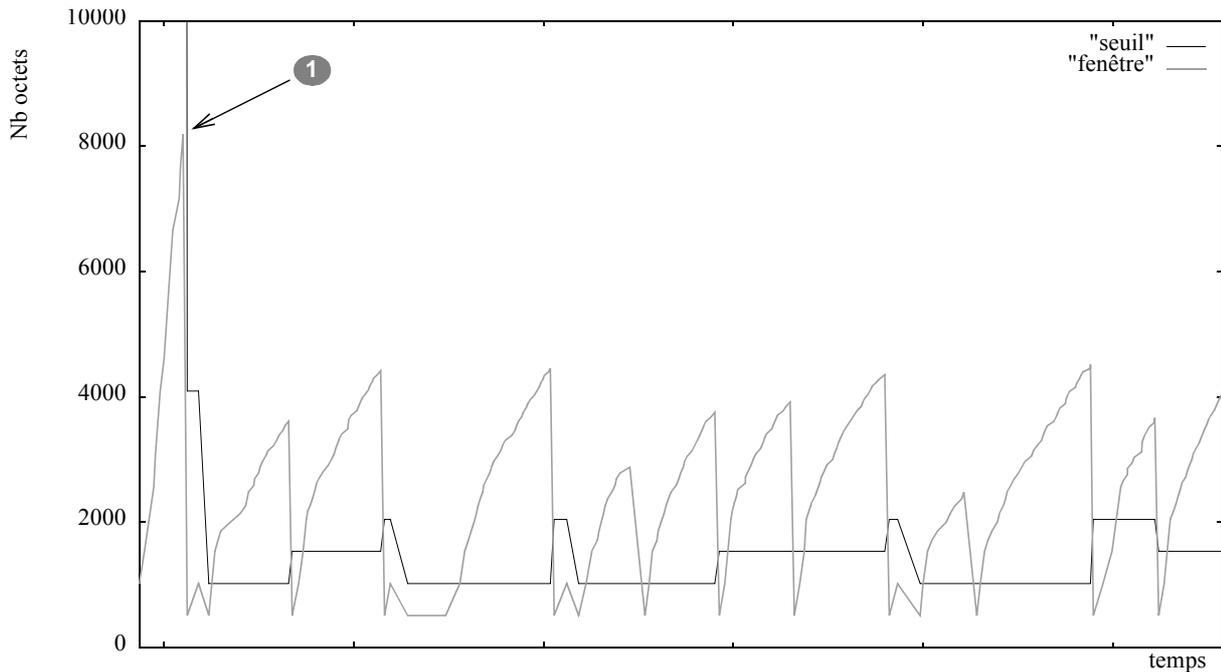
17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoS\DEA_fm

Exemple par simulation



17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoS\DEA_fm

Exemple par mesure



17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA_fm

396

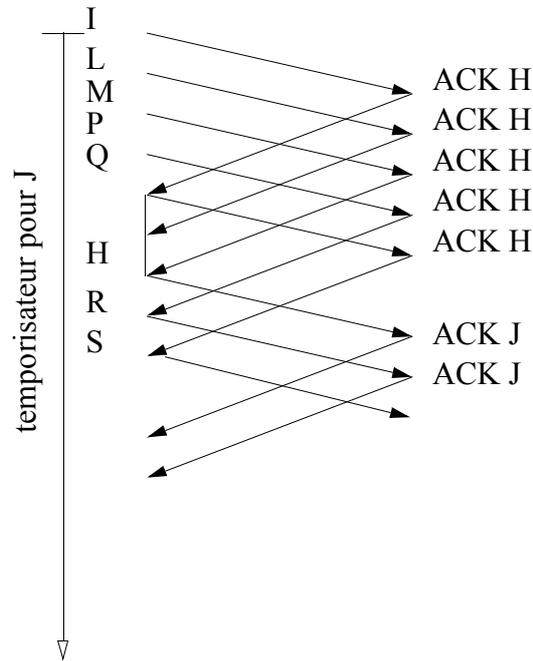
Plusieurs versions de TCP

- Nom liée à la version de BSD dans laquelle elle est parue
 - Tahoe : initiale de Van Jacobson
 - Reno : Une perte unique (i.e. un acquittement dupliqué), n'est pas une congestion grave.
 - Pas de phase de Slow Start,
 - Division de la fenêtre par 2 et phase de Congestion Avoidance,
 - Retransmission uniquement du paquet perdu.
 - new Reno :
 - Correction des bugs de Reno.

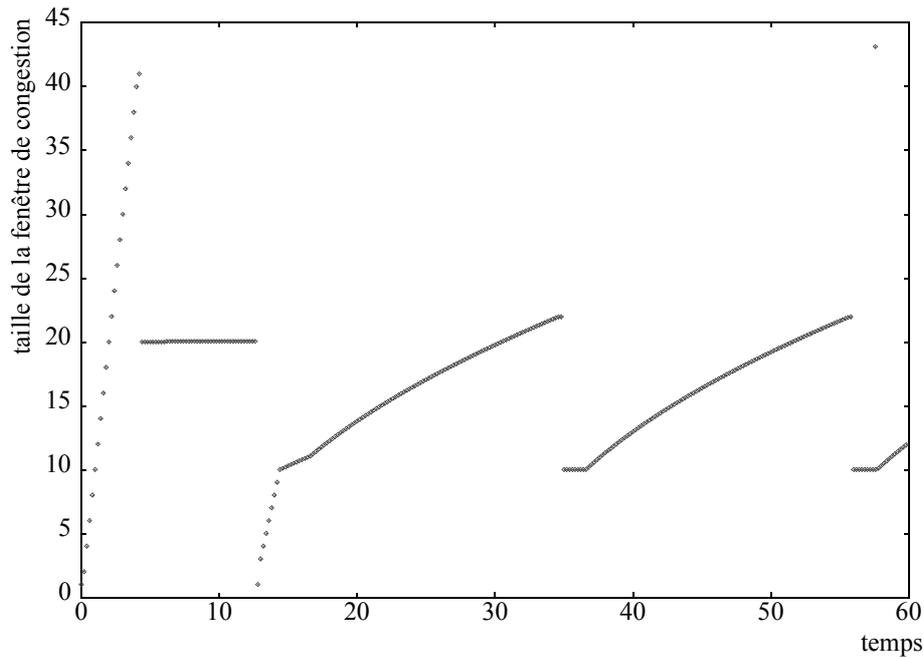
17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA_fm

397

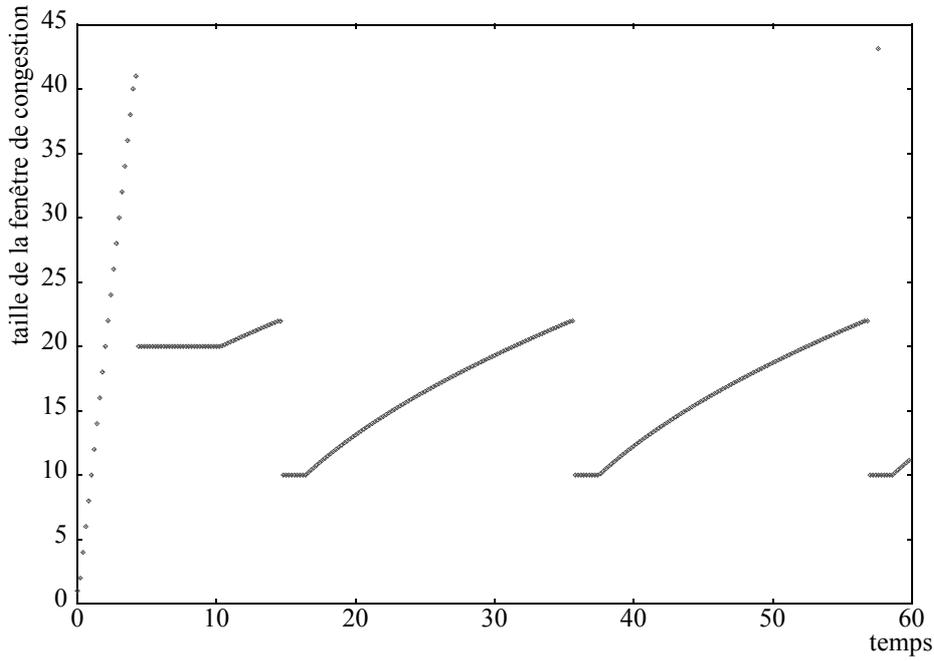
Reno



Reno

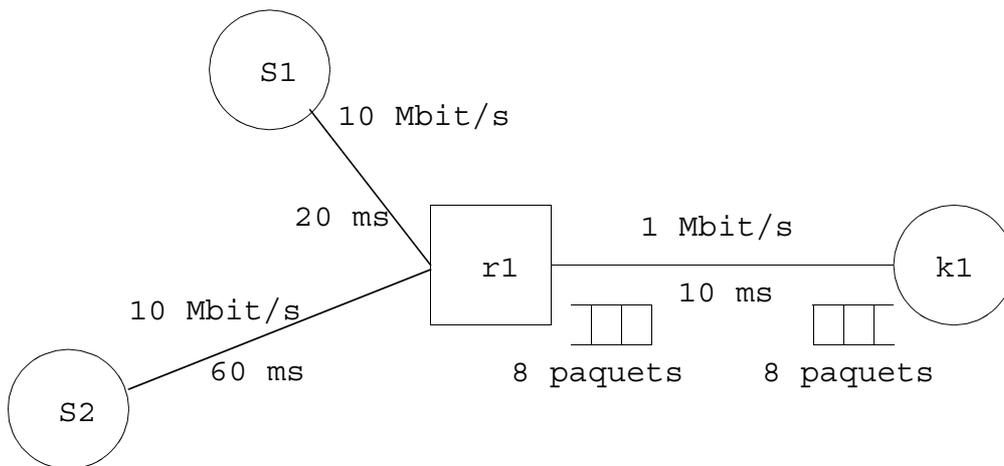


New Reno



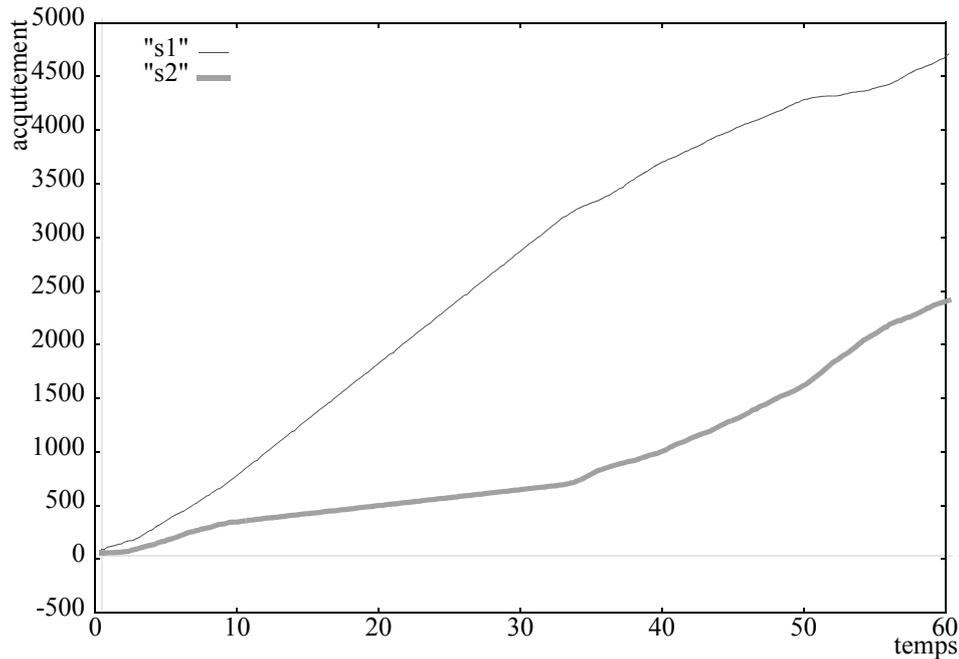
17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fn

TCP Equitable ?



17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fn

TCP Equitable ?



17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

402

Problèmes de TCP

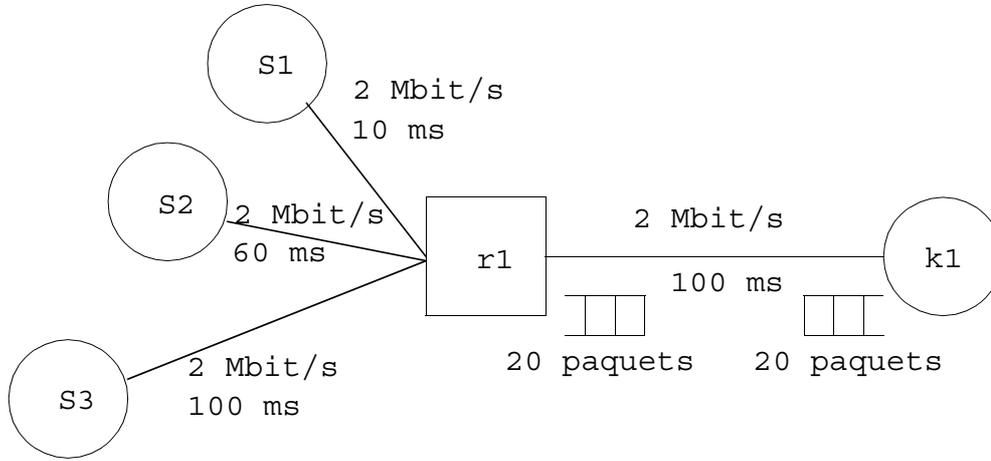
+ Synchronisation :

- Goulot d'étranglement :
 - saturation de la file d'attente du routeur
 - toutes les connexions perdent des paquets
 - toutes les connexions réduisent en même temps leur débit
- Solution :
 - RED
 - probabilité de perte proportionnelle à l'occupation de la file d'attente
 - les plus gros flux sont pénalisés

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

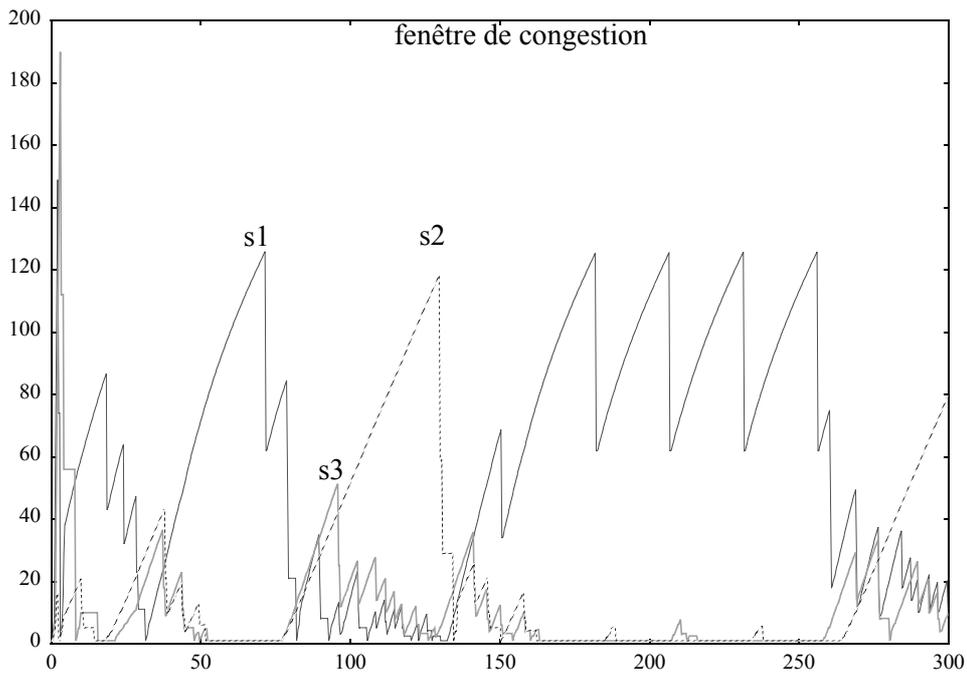
403

Exemple



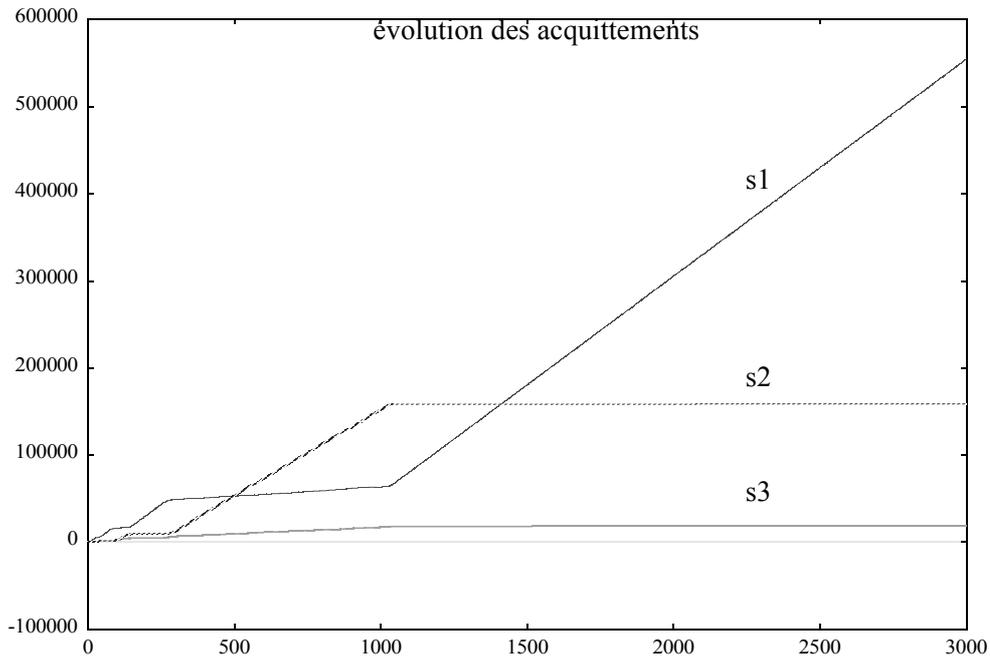
17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fm

Exemple FIFO



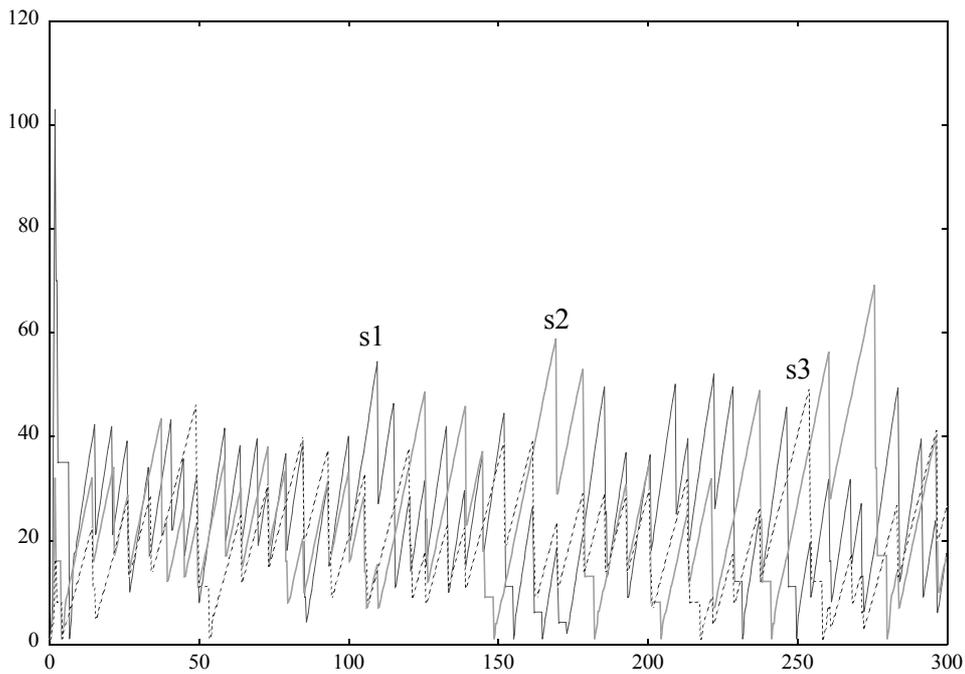
17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fm

Exemple FIFO



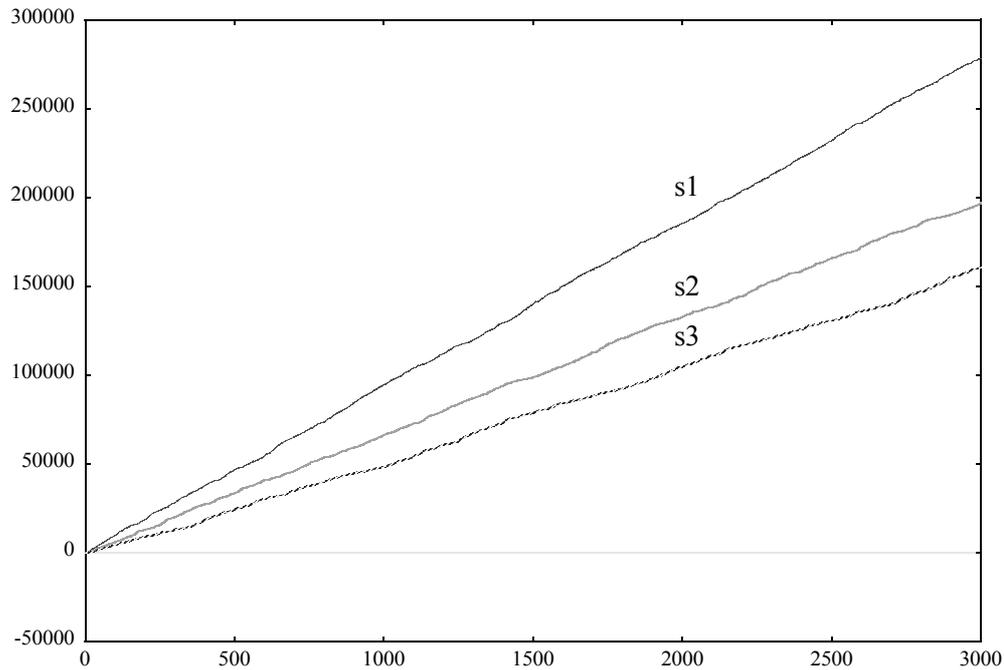
17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fm

Exemple RED



17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fm

Exemple RED



17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

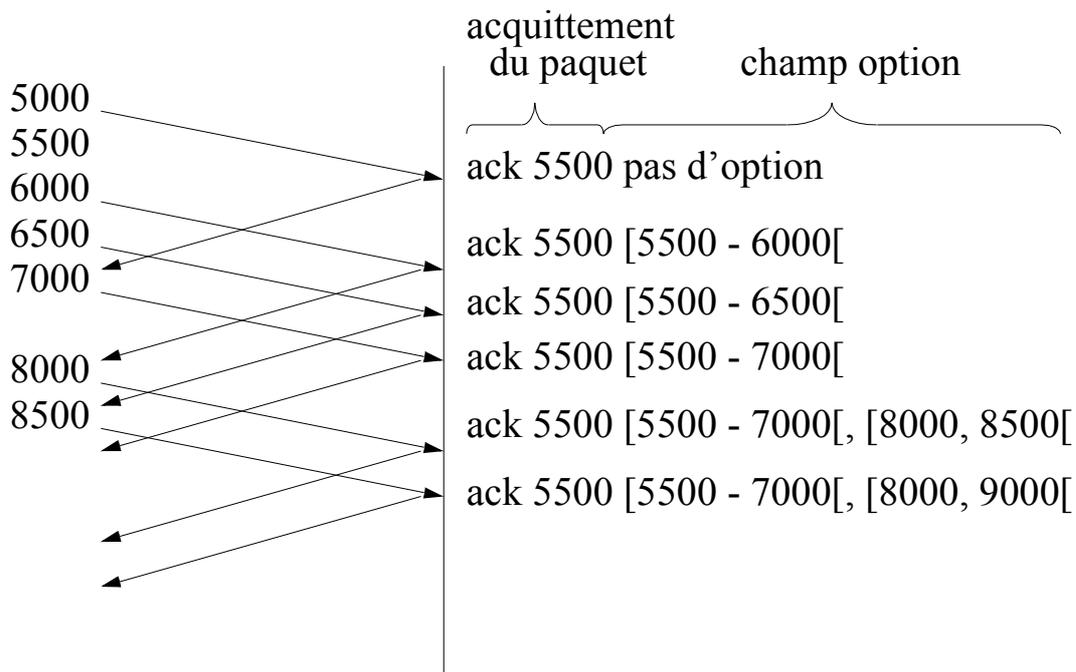
Problèmes de TCP

+ Liens satellites ou radio :

- Beaucoup d'erreurs de transmission
- Algorithme de gestion des congestions prend les pertes de paquets pour des congestions
 - Réduction du débit d'émission
- Solution :
 - Acquittements sélectifs
 - Dans le champ option des messages TCP
 - Evite de tomber dans le déclenchement du temporisateur de retransmission

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

Exemple



17 janvier 2003 - C:\Documents and Settings\vaivain\Mes documents\Support_ENST_QoSDEA.fm

Problèmes de TCP

+ Liens asymétriques

- ADSL, réseaux câblés, satellites,...
- La perte de paquet a lieu sur les acquittements pas sur les données
- Réduction du débit d'émission
- Solution ?
 - retarder les acquittements
 - enlever des acquittements dans la file d'attente
 - faire une autre connexion TCP (~proxy)

+ HTTP 1.1

17 janvier 2003 - C:\Documents and Settings\vaivain\Mes documents\Support_ENST_QoSDEA.fm

Autres problèmes de TCP

Problème d'équité

- ouvrir plusieurs connexions (ex Netscape)
- mauvaises mises en œuvre de TCP
- petits transferts
- multimédia

Solutions

- Modifier les applications (ex. p-HTTP)
- Mettre un contrôle de flux au niveau applicatif (ex : RTCP)
- Problèmes de cohérence
- Mettre un contrôle dans le réseau
 - dans chaque routeur.
 - à l'entrée.
 - pour chaque flux ?

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

Nouveaux besoins

Transports d'autres types d'informations :

- vidéo
- téléphonie
- jeux interactifs distribués

Contraintes de :

- débit
 - mécanismes de contrôle de flux de TCP inappropriés
- pertes
- délai
 - interactivité (téléphonie)

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

La qualité de service dans Internet

- Internet : interconnexion de réseau
- Absence d'état global
- La réservation nécessite une connaissance de l'état global
- Principe fondamental : Best Effort et Equité (en théorie)



+ Absence de garantie de bande passante

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fn

Mécanismes de contrôle de flux par Van Jacobson

TCP	UDP
<ul style="list-style-type: none"> • Estimation du délai : $\text{erreur} = \text{mesure} - \text{moyenne}$ $\text{moyenne} = \text{moyenne} - \alpha \times \text{erreur}$ $\text{dev} = \text{dev} + \beta \times (\text{erreur} - \text{dev})$ $\text{RTO} = \text{moyenne} + 4 \times \text{dev}$ • Slow start : <p><i>si pertes de plusieurs paquets (déclenchement du timer) alors :</i></p> $\text{seuil} = \text{taille de la fenêtre} / 2 \text{ (le seuil ne peut pas être inférieur à 2 segments)}$ <p>et</p> $\text{taille de la fenêtre} = 1$ 	<pre> From list_mgr@ISI.EDU Tue Jan 10 22:35:19 1995 Received: by venera.isi.edu (5.65c/5.61+local-20) id <AA25275>; Wed, 11 Jan 1995 06:31:42 -0800 Received: by rx7.ee.lbl.gov for mbone-na@isi.edu (5.65/1.44r) id AA01872; Wed, 11 Jan 95 06:35:21 -0800 Message-Id: <9501111435.AA01872@rx7.ee.lbl.gov> To: mbone-na Subject: jpuckett@168.18.130.249 sendig 600kb/s video Date: Wed, 11 Jan 95 06:35:19 PST From: Van Jacobson <van@ee.lbl.gov> Someone named jpuckett in Georgia (domain peachnet.edu) is sending ridiculous amounts of video (400-600kb/s) to the "Live from Antarctica" session (and most other mbone video sessions). The reverse dns mapping for this address doesn't work so I haven't been able to send email to this individual. Could someone get in touch with them & explain to them that this behavior is antisocial? Thanks. - Van </pre>

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fn

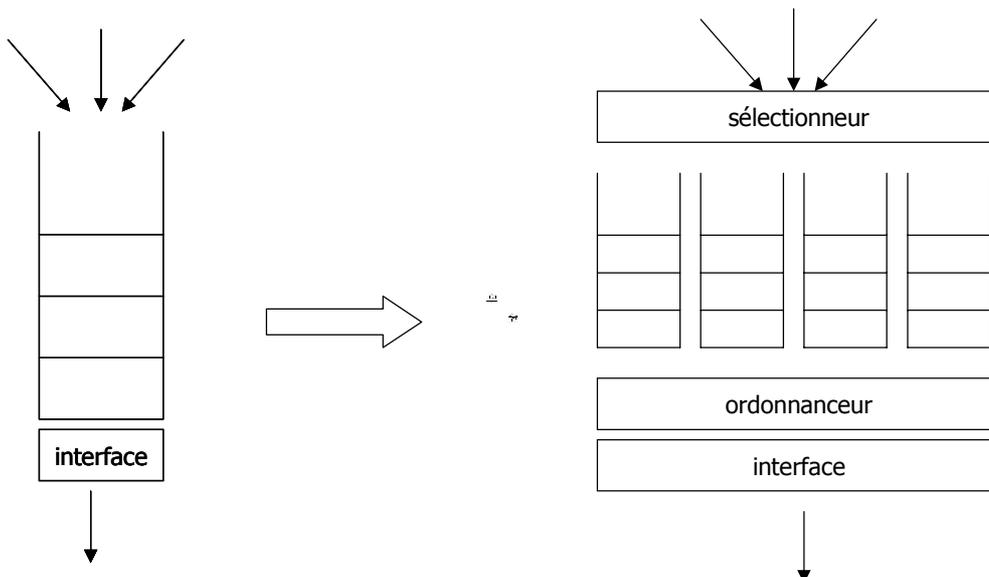
Peut-on modifier TCP ?

- Protocole mal compris :
 - pas de modèle analytique
 - simulations limitées
- Comment maintenir le compromis entre optimisation de la bande passante pour :
 - un flux donné
 - le débit total du réseau
- Protocole très agressif :
 - d'autres mécanismes de contrôle de flux n'ont jamais pu s'imposer

+ Il y a eu un IP Next Generation, mais toutes les tentatives pour faire un TCPng ont échoué

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDE1.fm

Différencier les traitements : Oublier la FIFO



17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDE1.fm

Problèmes

- Qu'est ce qu'un flux ?
 - protocole, adresse source, adresse destination, port source, port destination
 - champs communs dans l'en-tête :
 - permet l'agrégation
- Quel mécanisme d'ordonnancement ?
- Quelles contraintes temps réels ?
- Quand mettre en œuvre ces mécanismes ?

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

ordonnanceur, pas une question de vitesse

2 messages :

A : Longueur 270 octets - Echéance $t=320$ s

B : Longueur 15 octets - Echéance $t= 21$ s

**Routeur
Tortue:**

Vitesse : 1 o/s

Contexte : 1 s

Ordonnancement : EDF

**Routeur
Lièvre:**

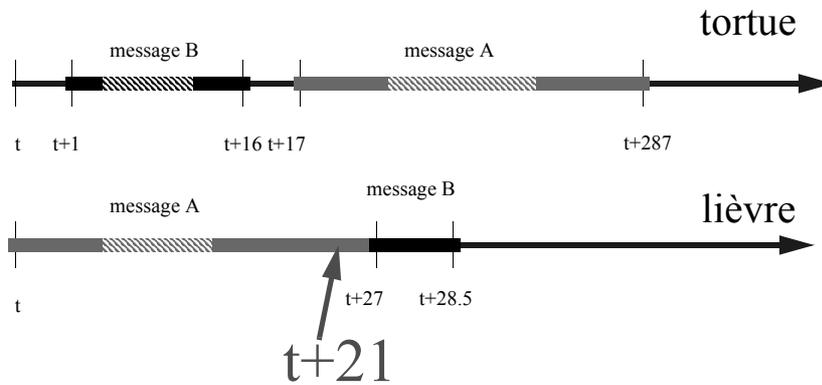
Vitesse : 10 o/s

Contexte : 0 s

Ordonnancement : PAPS

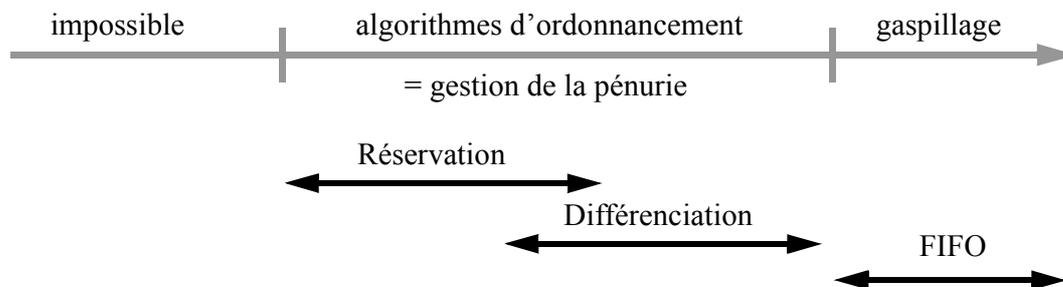
17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

résultat de l'ordonnancement



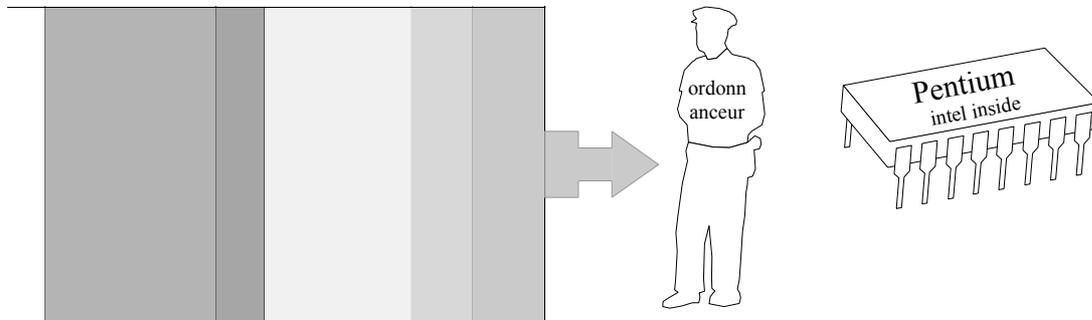
17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDE4.fm

Trois types de systèmes



17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDE4.fm

Mêmes problème dans les machines....



- + Sauf qu'une tâche peut être interrompue, pas la transmission d'un paquet (**modèle fluide**)

17 janvier 2003 - C:\Documents and Settings\vainain\Mes documents\Support_ENST_QoSDE1.fm

Modification des files d'attente

- Priorité stricte :
 - plusieurs files d'attentes classées par ordre de priorité
 - servir la file la plus prioritaire tant qu'elle contient des données
 - mise en œuvre simple et rapide
 - risque de famine
- Round Robin :
 - Servir alternativement chacune des files d'attente
 - Equitable ?
 - Doit-on être équitable ?

17 janvier 2003 - C:\Documents and Settings\vainain\Mes documents\Support_ENST_QoSDE1.fm

Equité pondérée

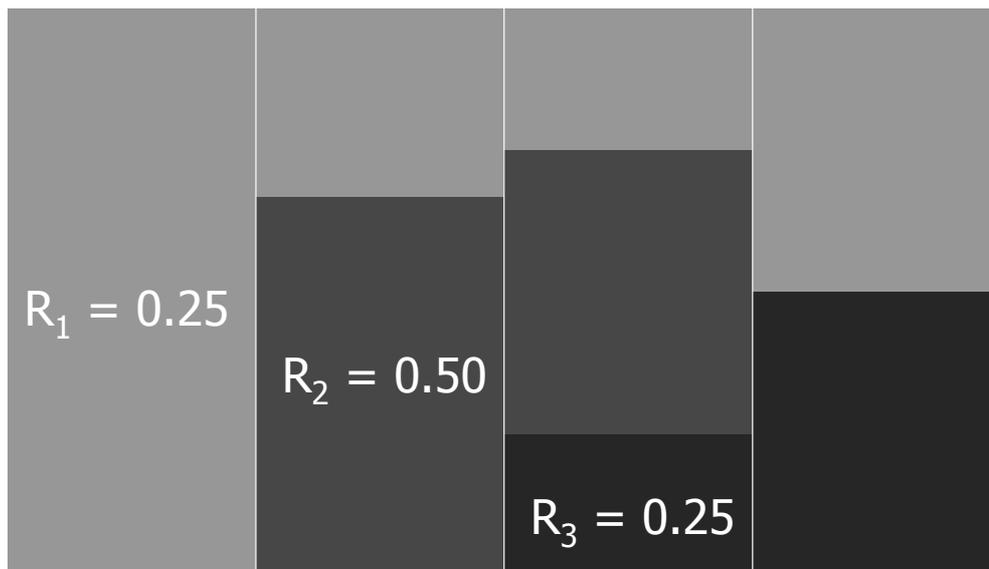
+ Le Round Robin : équité en paquets, pas en octets.

+ Définition d'une équité :

$$\forall t_1 t_2 \quad \frac{w_k(t_1, t_2)}{w_j(t_1, t_2)} = \frac{r_k}{r_j}$$

- $W_i(t_1, t_2)$: nombre de bits émis par le flux i entre les instants t_1 et t_2
- Définit une proportion de bande passante entre les flux,
- Ce n'est pas un mécanisme d'ordonnancement,
- Impossible à réaliser (modèle fluide)

Exemple



Remarques

- Pour un flux f donné
- Pour un lien de débit d
 - équivalent dans le pire des cas à un lien vide de débit :

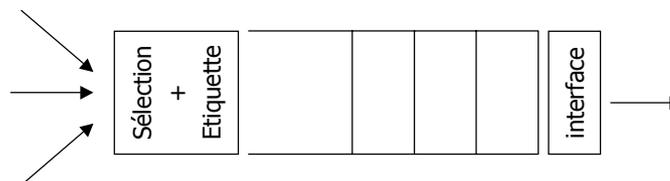
$$d_f = d \cdot \frac{r_f}{\sum_{\forall i} r_i}$$

+ le débit peut être supérieur si certains des flux ne sont pas présents

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fm

Weighted Fair Queueing

Une seule file d'attente, classée par estampilles.



- Exemple de calcul de l'estampille : $F_{i,k} = 1/r_k L_{i,k} + F_{i-1,k}$
 - Où $F_{i,k}$ représente l'estampille du k^{ieme} paquet du flot i
 - et $L_{i,k}$ la longueur du k^{ieme} paquet du flot i

+ Cet ordonnancement naïf est non équitable.

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fm

Preuve

Flux $k = 0; r_0 = 2$

$L_{i,0}$	10	25	10	10	25	15	15
	A	B	C	D	E	F	G
$F_{i,0}$	5	17.5	22.5	27.5	40	47.5	55

Flux $k = 1; r_1 = 1$

$L_{i,1}$	10	25	10	10	25	15	15
	L	M	N	O	P	Q	R
$F_{i,1}$	10	35	45	55	80	95	110

Ordonnement

10	10	25	10	10	25	25	10	15	15	10	25
A	L	B	C	D	M	E	N	F	G	O	P
5	10	17.5	22.5	27.5	35	40	45	47.5	55	55	80

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fm

Preuve (suite)

Flux $k = 0; r_0 = 2$

L^i_0	10	25	10	10	25	15	15
	A	B	C	D	E	F	G
F^i_0	105	117.5	122.5	127.5	140	147.5	155

Flux $k = 1; r_1 = 1$

L^i_1	10	25	10	10	25	15	15
	L	M	N	O	P	Q	R
F^i_1	10	35	45	55	80	95	110

Ordonnement

10	10	25	10	10	25	15	15	25	10	10	25
A	L	M	N	O	P	Q	R	B	C	D	E
105	10	35	45	55	80	95	110	117.5	122.5	127.5	140

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fm

Equité

- L'estampille d'un flux inactif n'évolue pas :
 - équitable sur une très longue période
 - loin du modèle fluide

Formule exacte du WFQ :

$$F_k^i = \frac{1}{r_k} L_k^i + \max(v(a_k^i), F_k^{i-1})$$

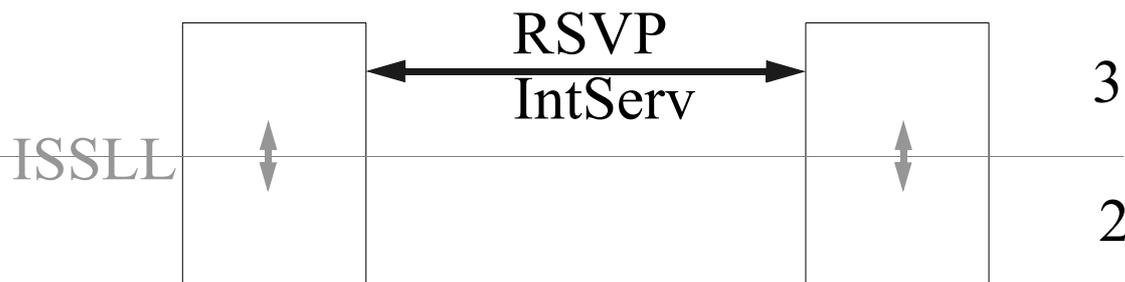
- max permet de faire avancer le temps, même si le flux est inactif
- $v(a_k^i)$
 - temps virtuel (accélération du débit d'émission du à l'absence d'autres flux)
 - difficile à calculer car demande de connaître l'ensemble des flux présents dans la file d'attente

+ Heuristiques pour approcher le temps virtuel

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

La réservation de ressources dans l'Internet

- Modifier l'architecture de l'Internet pour permettre une amélioration de certains flux par rapport au "Best Effort"
- Architecture modulaire :
 - plusieurs groupes de travail de l'IETF
 - possibilité d'ajouter des éléments
 - prend en compte le point-à-point et le multicast



17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

Réservation de ressources

+ Basée sur une théorie mathématique : Le Network Calculus :

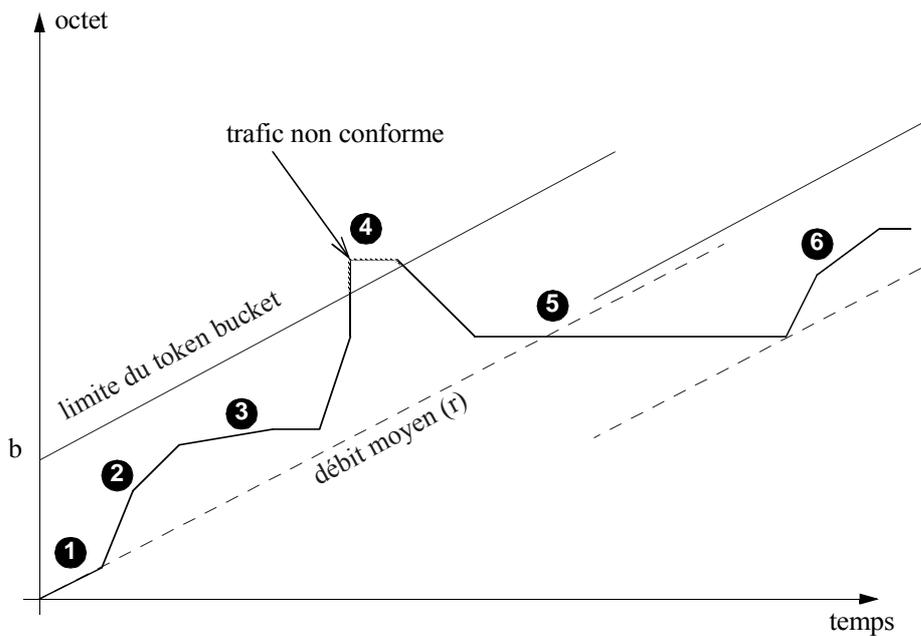
Donne une borne Max sur la transmission de paquets

+ Modèle de trafic : Token Bucket

- Permet de caractériser un trafic
- Paramètres :
 - capacité : b jetons
 - débit : r jetons
- Pour émettre un paquet de k octets, le bucket doit contenir au moins k jetons.

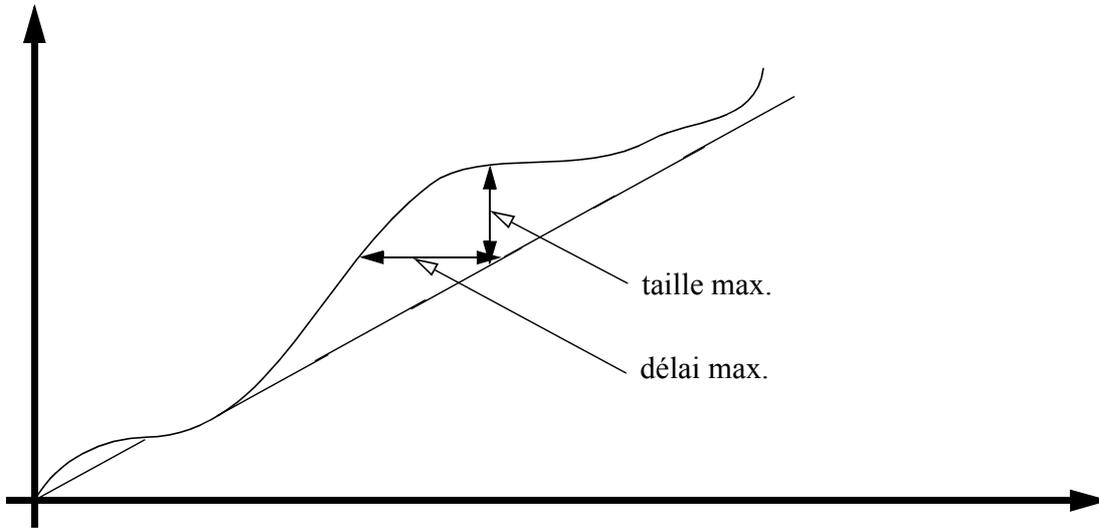
17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fm

Exemple



17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fm

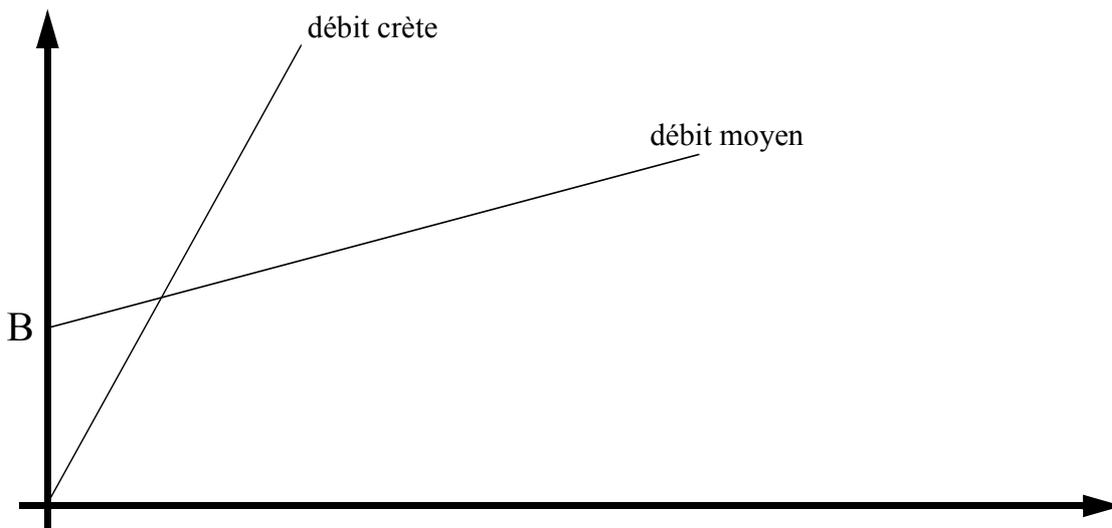
Network Calculus¹



17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

1. voir : *An Introduction to Network Calculus* - Jean-Yves Le Boudec - <http://ica1www.epfl.ch/netCalSlides.pdf>

Network Calculus



17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

+ token bucket : c débit crête, B taille du bucket, r débit moyen

Network calculus

- permet de calculer la taille du buffer pour éviter les pertes
- le temps de traversée d'un élément de réseau
- permet de combiner plusieurs éléments de réseaux

- Inconvénient : bornes très pessimistes

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

RSVP

- ReSeRvation Protocol
- Participe à la réservation
- Définition de mécanismes génériques
- Dynamique, par flux
- Point-à-point et multicast

+ Peut être remplacé par SNMP

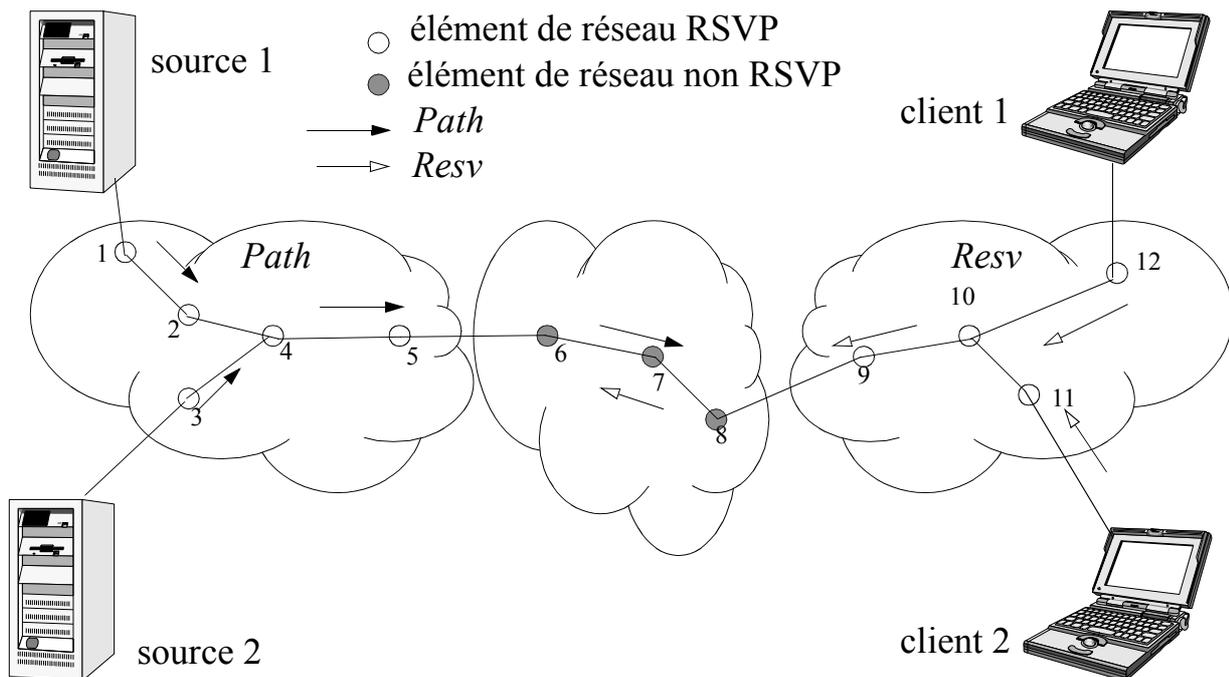
17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

RSVP

- Principes
 - Réserve simplexe
 - Demande de réservation initiée par le récepteur
 - Filtrage des paquets, dissocié de la réservation
 - Offre plusieurs styles de réservation (multicast)
- Maintien de contexte « soft-state » dans le réseau
- Ce que RSVP ne fait pas :
 - Transmission de paquets
 - Routage (QOSR)
 - Contrôle d'admission dans les routeurs
 - Ordonnancement des paquets

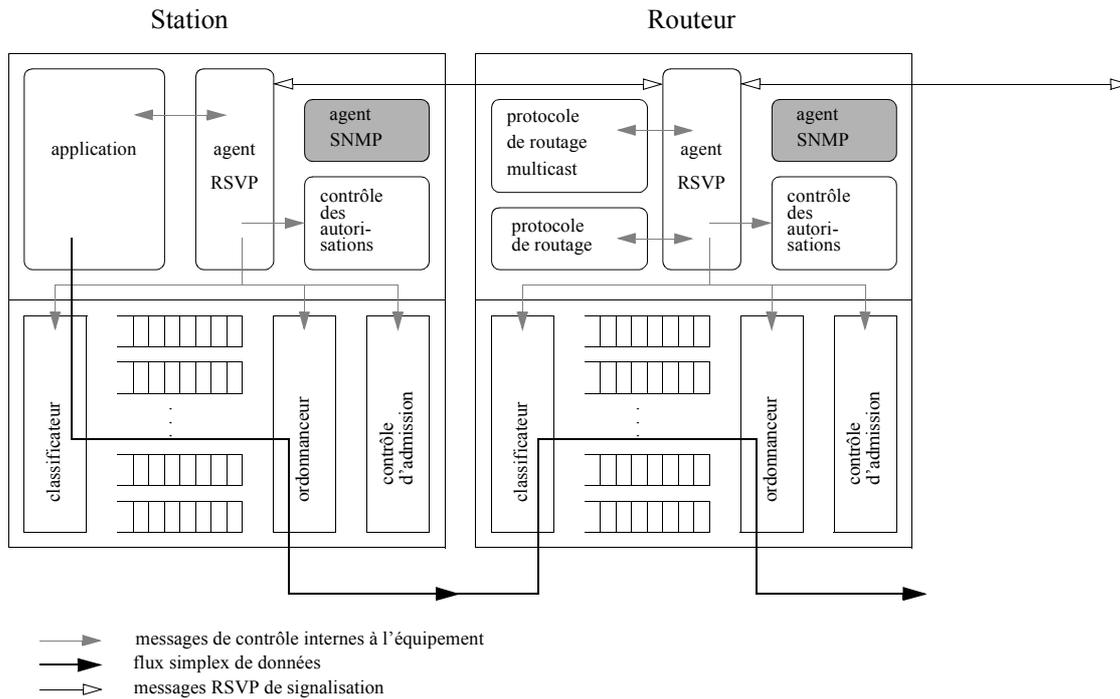
17 janvier 2003 - C:\Documents and Settings\vaunain\Mes documents\Support_ENST_QoSDEA.fm

RSVP



17 janvier 2003 - C:\Documents and Settings\vaunain\Mes documents\Support_ENST_QoSDEA.fm

RSVP



17 janvier 2003 - C:\Documents and Settings\vaunain\Mes documents\Support_ENST_QoSDEA\fm

RSVP : Message Path

- La source émet des T-specs précisant ses caractéristiques de débit
 - débit moyen et débit de pointe
 - mesuré à l'aide d'un Token Bucket
- Le réseau ajoute ses paramètres propres (AD-spec)
 - bande passante minimale disponible
 - délai de latence minimal MTU
- Chaque routeur (ou élément de réseau va introduire un retard) :
 - Paramètre C (en octets) : dépend de la vitesse réservée pour ce flux
 - Paramètre D (en ms) : indépendant de la vitesse réservée pour le flux par exemple : erreur due à l'approximation du modèle fluide

- + Aucune réservation n'est faite pour ce flux
- + Un contexte est mis dans le routeur pour ce flux (pour déterminer le chemin inverse)

17 janvier 2003 - C:\Documents and Settings\vaunain\Mes documents\Support_ENST_QoSDEA\fm

RSVP : Message RESV

- Les clients en fonction des paramètres reçus peuvent demander une réservation de ressources
- Les requêtes suivent le chemin inverse.
- Les routeurs font un contrôle d'admission
- Plusieurs classes de services définies par l'IntServ :
 - Délai et débit maximal garantis
 - Charge contrôlée ~ Internet peu chargé
 - Best Effort

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

Exemple de service garanti

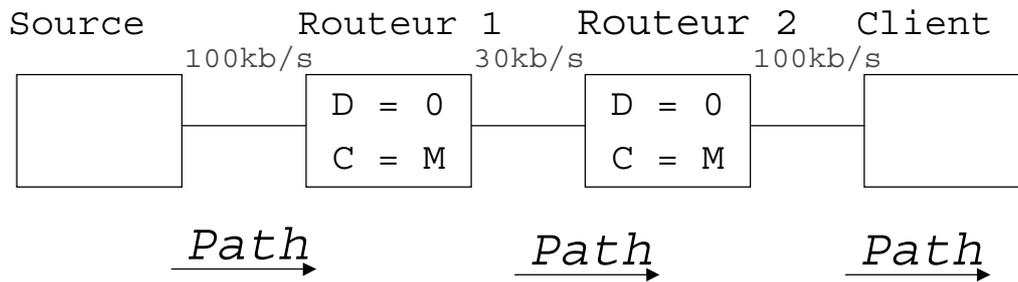
Une application autorise un délai d'une seconde :

- mesures psycho-visuelles,
- taille des buffers de réception
- ...

Elle demande un service garanti.

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

Message Path reçu par le client



Sender T_SPEC :	Sender T_SPEC :	Sender T_SPEC :
r = 22 kb/s	r = 22 kb/s	r = 22 kb/s
b = 10 kb/s	b = 10 kb/s	b = 10 kb/s
p = ∞	p = ∞	p = ∞
AD_SPEC :	AD_SPEC :	AD_SPEC :
∅	D = 0	D = 0
	C = M	C = 2.M

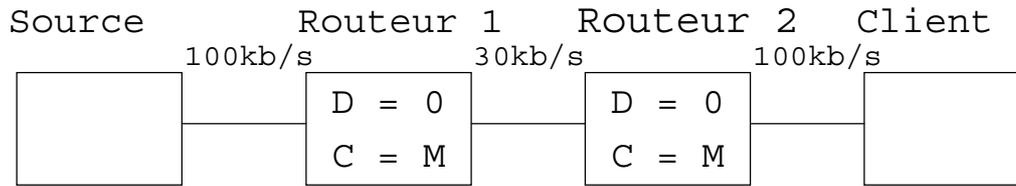
Calcul par le client de la réservation

- Quel est le débit nécessaire pour que le temps de traversée d'un paquet de taille M=1000 soit d'une seconde ?
 - $(b + Ct_{tot}) / r + Dt_{tot} = 1 \text{ s}$
 - $r = 12 \text{ kb/s}$

- + Si on ne réserve que 12 kb/s, il y a création d'un goulot d'étranglement
- + Le débit réservé minimum doit être de 22 kb/s.

- Le temps de traversée d'un paquet est donc de 0,55 s
- L'application peut donc supporter d'un réseau une tolérance (Slack) de 0,45 s

Message RESV



Resv

$$R = 22 \text{ kb/s}$$

$$S = 0,45 \text{ s}$$

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fn

Calcul de C et D

- Soit un réseau composé de n nœuds, et ayant un débit de d kb/s pour tous les liens.
- Le temps de traversée du réseau est de $n \cdot M/d$
- Si WFQ en modèle fluide, un paquet a l'impression d'avoir un débit $r < d$.
- Le délai est donc de $n \cdot M/r$.

+ Le modèle n'est pas fluide donc avant l'émission d'un paquet le délai d'attente est le temps d'émission d'un paquet de taille M , soit M/d .

+ $C = M$ octets

+ $D = M/d$ ms

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fn

Service contrôlé

- Comportement identique à un internet peu chargé
- Part de l'observation que les applications multimédia marchent quand l'Internet est peu chargé
- Pas de garantie quant au temps de traversée maximum des paquets
- Définition plus vague donc plus simple à mettre en œuvre :
 - Deux files d'attente
 - Un WFQ avec deux classes
 - Nécessite surtout un contrôle d'admission dans les routeurs.

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

RSVP

448

RSVP (suite)

Principales critiques à RSVP :

- comment facturer la réservation ?
- mécanisme coûteux en calcul dans les routeurs
- bornes du service garanti trop pessimistes
- nécessité d'un état par flot dans les routeurs :
 - problème d'extensibilité
 - impensable sur un lien trans-atlantique

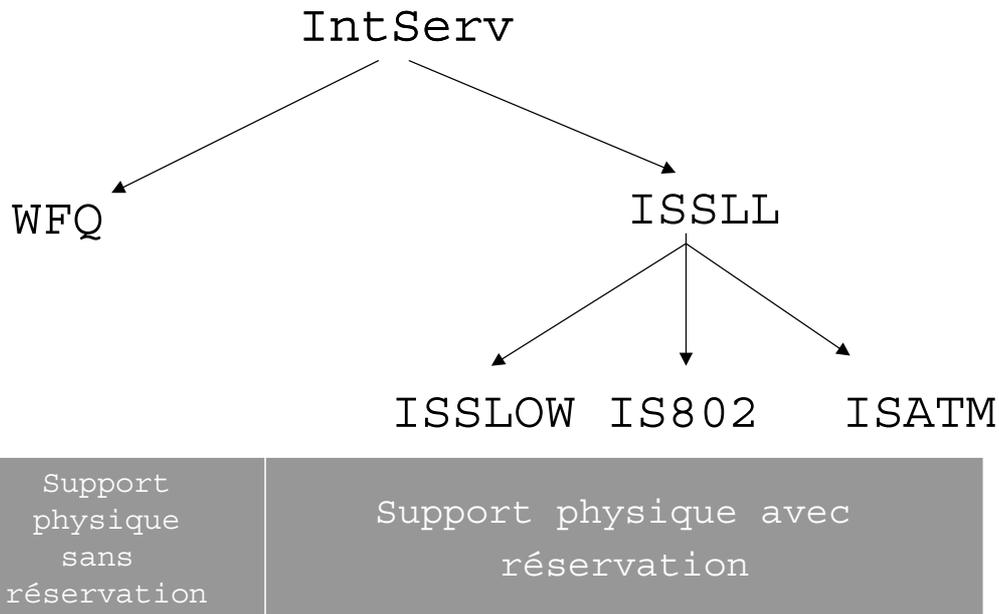
- + RSVP limité aux systèmes autonome ?
- + Travaux sur l'agrégation
- + Peut servir de protocole de signalisation

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

RSVP

449

Sur le niveau 2



17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDE1.fm

ISSLOW

- Liaison par modem, le goulot d'étranglement.
- Modem 28,800 Kbit/s, Téléphonie 13 Kbit/s
- Paquets longs => éloignement du modèle fluide
- Paquets courts => perte de bande passante avec les en-têtes
- Compression d'en-tête pour TCP+IP
- ISSLOW propose :
 - étendre la compression à RTP + UDP + IP
 - interrompre la transmission d'un paquet, par un paquet plus prioritaire.

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDE1.fm

IS802

- Utilise le standard IEEE 802.1 p
- Introduction de priorités dans les réseaux locaux.
- Gestionnaire de la bande passante
- Utilisation de priorités :
 - 0 moins que le Best Effort
 - 1 Best Effort
 - 2 Control load
 - 4 Garanti 100 ms
 - 5 Garanti 10 ms

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

—Sur le niveau 2

452

La différenciation de services

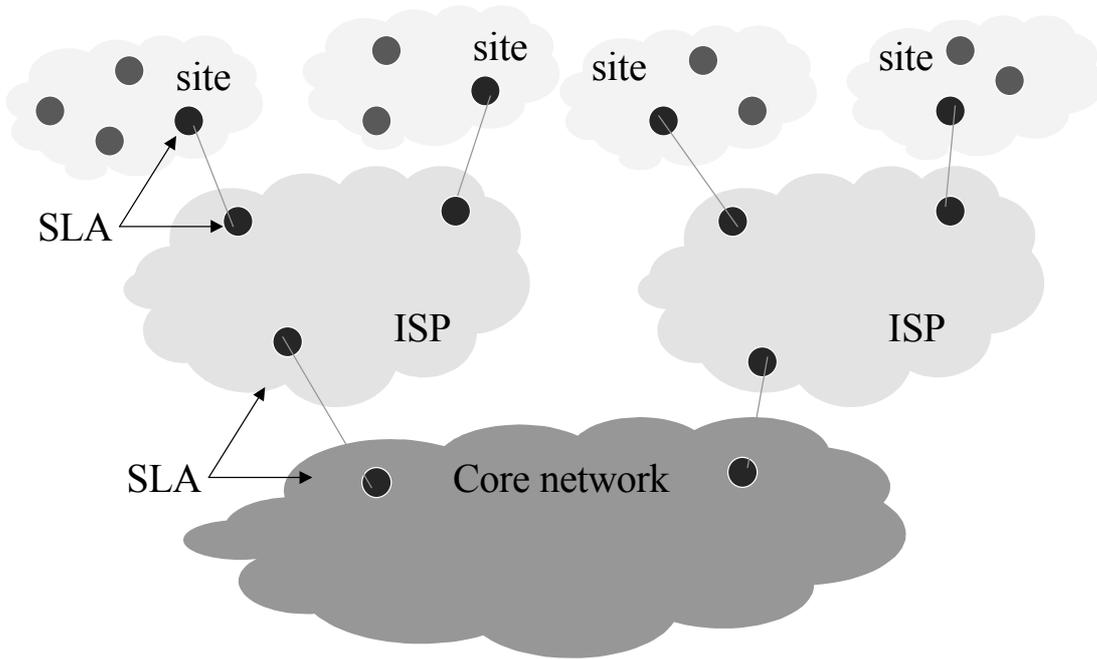
- + Les applications multimédias vont être de plus en plus utilisées.
- + Les applications multimédias cassent le principe d'équité de l'Internet.
- + Les utilisateurs peuvent changer leurs mécanismes de contrôle d'accès pour gagner plus de bande passante.
- + Les industriels sont prêts à payer pour une meilleure "qualité de service"
- + Différencier les services en fonction des abonnements (première classe, classe économique,...)

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

—La différenciation de services

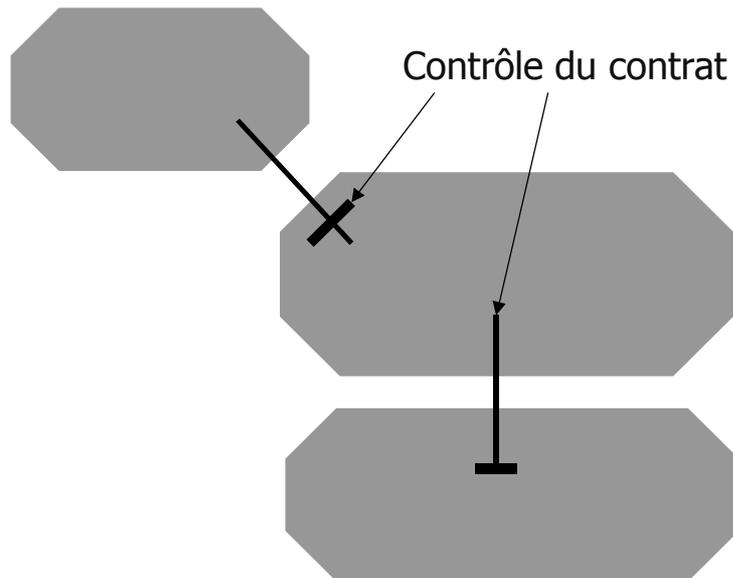
453

Internet à 3 niveaux



17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fn

Service Level Specification



17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA\fn

+ Ce schéma permet l'agrégation des flux

Comportement des routeurs

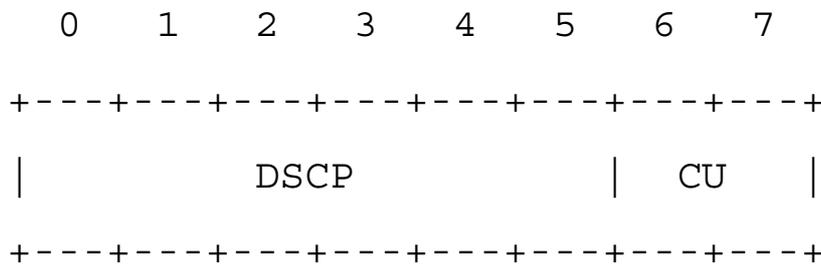
Deux types de routeurs :

- en frontière de réseau : mesure le trafic en fonction d'un contrat (par exemple un token bucket). En cas de non respect d'un contrat :
 - retarder les paquets pour rendre conforme,
 - marquer les paquets hors-contrat
 - rejeter les paquets hors-contrat
- dans le cœur du réseau :
 - différenciation entre les différentes classes de services,
 - taux de rejet plus ou moins important
 - délai plus ou moins important

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

Octet DiffServ (DSCP)

- réutilisation de l'octet ToS des paquets IPv4
- dans le champ identificateur de flux d'IPv6



DSCP: differentiated services codepoint

CU: currently unused

17 janvier 2003 - C:\Documents and Settings\voitain\Mes documents\Support_ENST_QoSDEA.fm

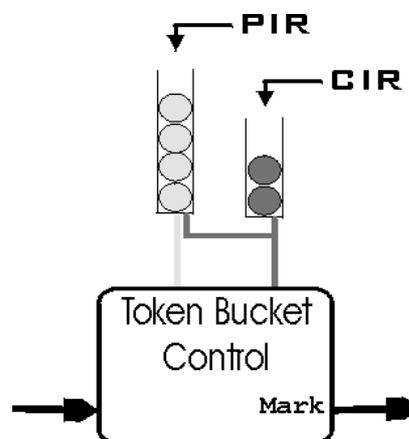
Classes de services EF

- Expedited Forwarding :
 - faible proportion du trafic,
 - délai de traversée très court,
 - traitement prioritaire par rapport aux autres types de trafic,
 - les flux non conformes à un token bucket sont retardés.
 - problèmes d'interconnexion entre AS : Bandwidth Broker ?

+ Service équivalent à une liaison spécialisée

Classes de services AF

- Assured Forwarding,
- 4 classes, 3 priorités,
- Three Color Marker :



Conclusion

- La qualité de service passe par le dimensionnement des réseaux
- TCP est trop mal compris pour être modifié facilement
- Beaucoup de travaux de recherche sur la qualité de service :
 - bande passante, ressource rare => gestion de la pénurie
- La bande passante croît plus vite que la puissance de calcul :
 - bientôt des réseaux Internet à 40 Giga Octets
 - possibilité de gaspillage de bande passante ?
 - Introduction de mécanismes de gestion des files d'attente plus raffinés,
 - Introduction de mécanismes de facturation basés sur la nature des flux.

+ La qualité de service passe par une différenciation des flux

Conclusion

- Internet : réseau simple
 - Les opérateurs ne traitent plus des flux mais des agrégats de flux (MPLS,...)
 - La réservation ne dépassant pas un système autonome : lier DiffServ à RSVP
 - La réservation peut être utilisée dans certains cas critiques (où la bande passante ne peut pas être augmentée, par exemple les réseaux sans fils).
 - Mieux répartir la charge dans le réseau d'opérateur : Ingénierie de Trafic.
 - DiffServ permet de séparer les flux pour créer des Internet à plusieurs vitesses
 - DiffServ permettrait de protéger certains flux (TCP, Multimédia,...)

Conclusion

- Internet = réduction des états (faire simple)
- Les mécanismes de compression de données multimédia prennent en compte que le réseau n'est pas un médium entièrement fiable.