IPv6

Nicolas Prunier

Inspiré de la formation IPv6 du CRIHAN

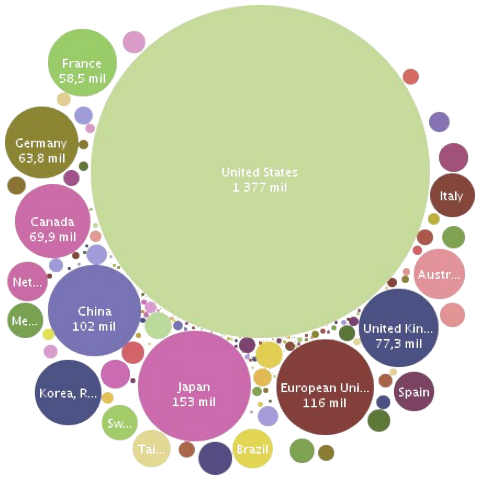
Plan

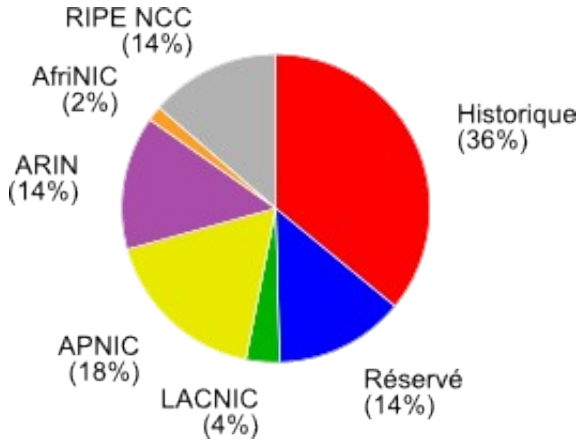
- Pourquoi IPv6
- Le protocole IPv6
- Les adresses IPv6
- Les Protocoles associés

Pourquoi IPv6

IPv4 gestion des adresses

- Rappel sur les adresses IPv4
  - Elles sont gérées par l'Internet Assigned Numbers Authority (IANA) qui les fournit aux Registres régionaux (Regional Internet registry - RIR)
  - 5 classes d'adresses historiques
  - Répartition inégale sur la planète





La fin des adresses IPv4

- Protocole de 1983 (une centaine d'hôtes sur Internet)
- Début de l'utilisation commerciale en 1992
- Pénurie d'adresses IPv4 prévue pour 1994
  - La montée en puissance d'Internet n'a pas été anticipée, les adresses IP ont été distribuées trop vite et le nombre d'adresses est trop limité
- Dès 1993, les adresses de classes B sont quasiment épuisées
  - En urgence, la RFC 1519 introduit le CIDR (Classless Inter-Domain Routing)

La fin des adresses IPv4

- En 1995, la RFC 1918 définit les classes IP non routables sur Internet :
  - 10.0.0.0/8, 172.16.0.0/20, 192.168.0.0/16
- Grâce à ces 2 RFC et au NAT, la pénurie d'adresses n'est plus estimée qu'en ... 2010
- 3 février 2011, comme décidé courant 2010, l'IANA fournit un des 5 derniers /8 à chacun des RIR
  - Courant 2012 les RIR n'auront plus d'adresses à fournir

Les limitations d'IPv4

- Le NAT rend difficile le déploiement de certains services (FTP, VOIP, téléphonie, p2p, ..)
- Explosion des tables de routage (de 100 000 entrées en 2001 à plus de 350 000 aujourd'hui)
- Problème de conception :
  - Multicast difficile à mettre en œuvre
  - Configuration complexe des postes nomades
  - Sécurité non prise en charge nativement

IPv6 un peu d'histoire

- 1995 : spécifications initiales (RFC 1883)
- 1998 : spécifications « finales » (RFC 2460)
- 1998 : déploiement de 6bone
  - Réseau IPv6 entre la France, le Japon et le Danemark
- 2004 : Fin des expérimentations, distribution de préfixes définitifs
- 2007 : Les RIR alertent le grand public sur la nécessité de migration

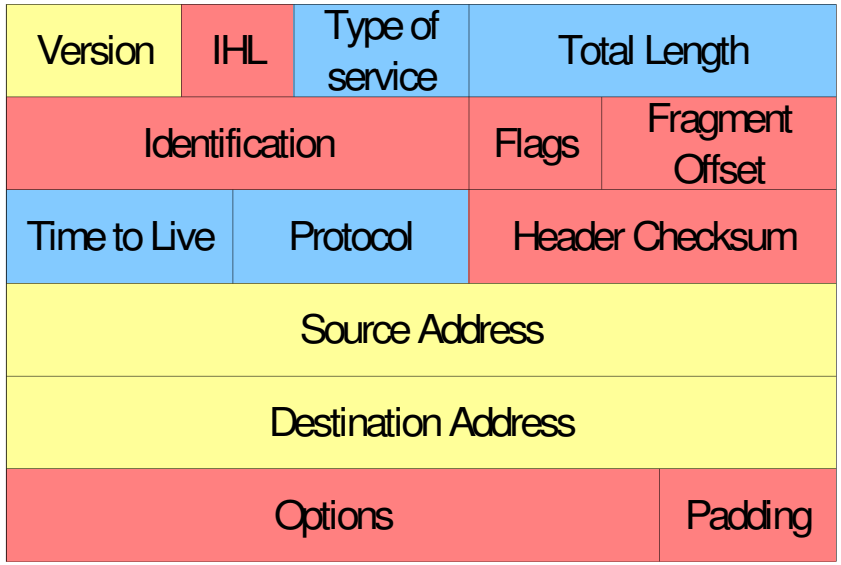
Le protocole IPv6

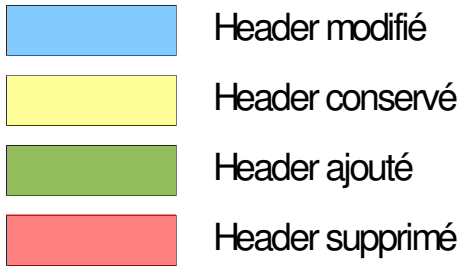
Les bases

- Protocole de couche 3 (évidemment), pas de modification pour TCP/UDP/...
- Adresses sur 128 bits contre 32 en IPv4
- $3,4.10^{38}$ adresses théoriques (au moins 1564 adresses réellement disponibles par métre carré sur la planète, océans compris)
- Simplification du routage (entête de taille fixe, adressage hiérarchisé et PMTU)
- Qos et IPsec natif, meilleure gestion du multicast
- Autoconfiguration des clients (mobilité)

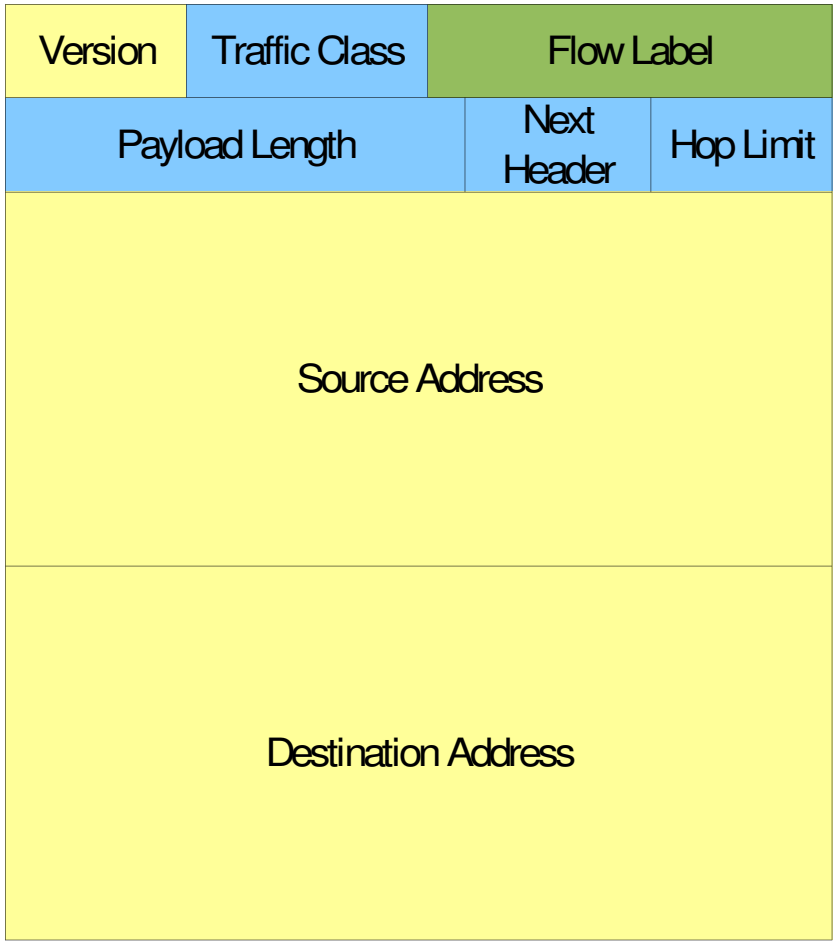
Entête IPv6

IPv4





IPv6



Entête IPv6

- Plusieurs champs sortent de l'entête, d'autres sont modifiés :
  - Next Header peut contenir des options IPv6 ou le type de payload :
    - Hop by hop / routing (seules options analysées par les routeurs)
    - Authentication/fragment/confidentiality/...
    - IPv4/IPv6/TCP/UDP/ICMPv6/...
- Flow Label apparaît :
  - Le triplet IPsrc/IPdst/FL permet d'identifier les paquets appartenant à un même flow
  - Optimisation du hop-by-hop/routing

Les adresses IPv6

Notation

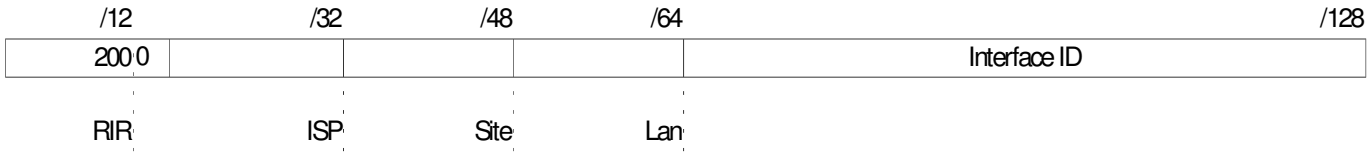
- Notation :
  - 8 champs de 16bits, notés en hexa, séparés par ':'
  - Insensible à la casse
  - Les zéros à gauche d'un champ peuvent être omis
  - Une suite de champs à 0 peut être représentée par '::' (mais une seule)
- Exemple :
  - 2001:0660:7401:02C1:0000:0000:0000:ABCD

    => 2001:660:7401:2C1::abcd

Terminologie

- Générale :
  - Node (Nœud) : N'importe quel élément du réseau
  - Router (Routeur) : Nœud qui forward les paquets qui ne lui sont pas destinés
  - Host (Hôte) : Noeud qui détruit les paquets qui ne lui sont pas destinés
  - Local-link (lien-local) : Canal de communication de niveau 2
  - Neighbor (voisin) : Nœuds sur un même local-link
- Type d'adresses :
  - Unicast (one-to-one) : Globale (internet) ou locale (~= RFC 1918 en IPv4)
  - Multicast (one-to-many) : identique à IPv4
  - Anycast (one-to-nearest) : unicast sur plusieurs nœuds, seul le nœud le plus proche répond

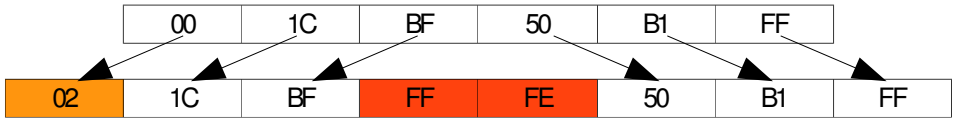
Allocation



- Plusieurs modèles d'allocation existent, tous sur une même base :
  - L'IANA utilise le 2000::/3 pour les IPv6 unicast (seulement 1/8 des adresses disponibles)
  - L'interface d'un node (Interface ID) est sur 64 bits (le préfixe est donc aussi sur 64 bits)
- Une des propositions :
  - Un /12 par RIR
  - Un /32 par LIR (FAI)
  - Un /48 pour l'utilisateur final (soit 2^{16} sous-réseaux) ou un /64 (pour les particuliers)

L'identifiant d'interface

- L'identifiant d'interface est toujours sur 64bits, il est :
  - Configuré manuellement (non recommandé)
  - Attribué par un serveur DHCPv6 (autoconfiguration avec état)
  - Généré de manière cryptographique (CGA)
  - Issu de l'autoconfiguration sans état (SLAC) à partir de l'adresse MAC (7ième bit a 1, identifiant EUI-64) :

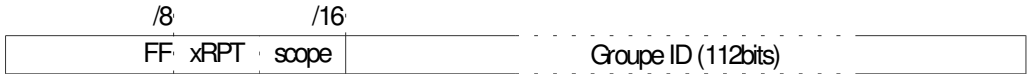


  - Autoconfiguré avec tirage pseudo aléatoire
- Dans tous les cas, cet identifiant doit être unique pour un même préfixe IPv6

Les adresses particulières

- Adresse non spécifiée (unspecified address) :
  - ::/128 (0000:0000:0000:0000:0000:0000:0000/128)
  - Equivalent de 0.0.0.0/32 en IPv4
  - Utilisé pour les mécanismes d'autoconfiguration et pour la détection d'adresses dupliquées (DAD)
- Adresse de loopback :
  - ::1/128 (0:0:0:0:0:0:0:1/128)
  - Equivalent de 127.0.0.1/32 en IPv4
- Adresse local-link :
  - Concaténation du préfixe FE80::/10 et de l'identifiant d'interface
    - FE80::21C:BFFF:FE50:BIFF
  - Permet de communiquer uniquement sur le lien de niveau 2

Les adresses multicasts



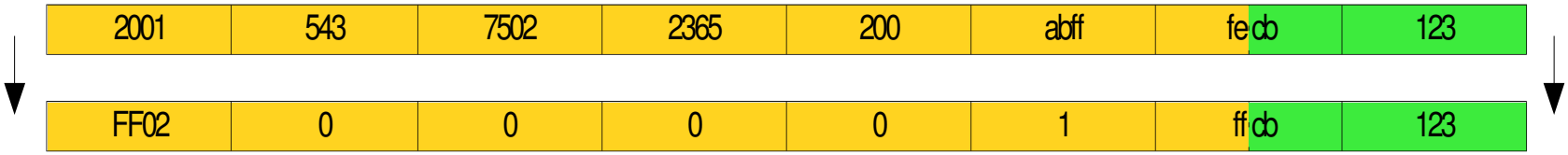
- Les adresses multicast IPv6 sont en préfixe FF00::/8
- Le 3ième octet contient les flags xRPT :
  - x toujours à 0
  - T (0 si adresse permanente attribuée par l'IANA, 1 sinon)
- Le 4ième octet définit le périmètre de diffusion (scope) :
  - Node-local (1)
  - Link-local (2)
  - Site-local (5)
  - Organization-local (8)
  - Global (E)

Les adresses multicasts

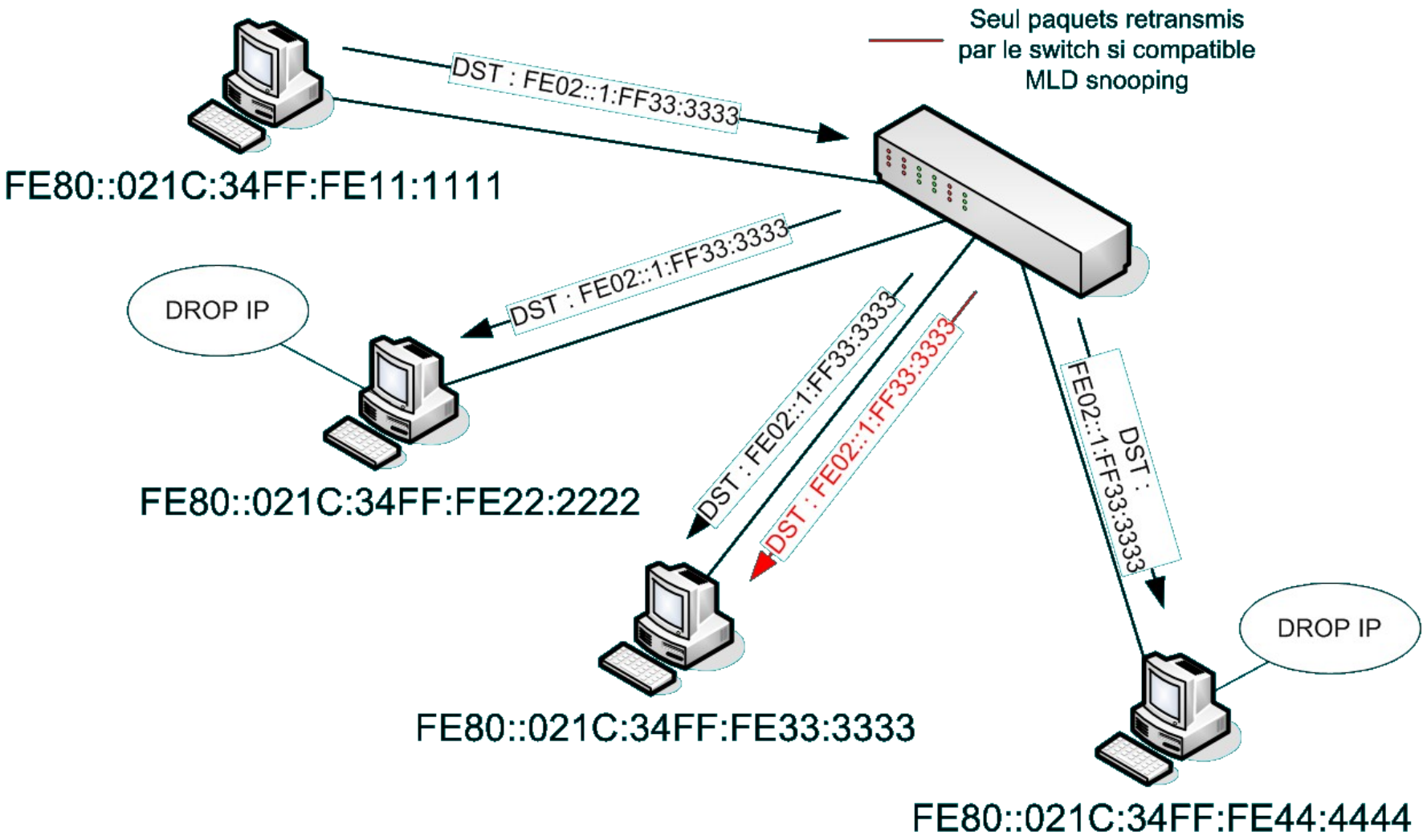
- Les 112 bits restant sont notamment utilisés pour définir le groupe unicast (et d'autres informations en fonction des flags), parmi les groupes prédéfinis :
  - Groupe 1 : Tous les nœuds (scope 1 ou 2)
  - Groupe 2 : Tous les routeurs (scope 1, 2 ou 5)
  - Groupe 9 : les noeuds RIP (scope 1 ou 2)
  - Groupe 101 : les nœuds NTP (scope 1, 2, 5, E)
- Autrement dit :
  - FF01::101 : tous les serveurs NTP du nœud
  - FF02::101 : tous les serveurs NTP du lien local
  - FF05::101 : tous les serveurs NTP du site
  - FF0E::101 : tous les serveurs NTP d'internet

L'adresse multicast solicited-node

- Elle est notamment utilisée pour l'équivalent IPv6 des requêtes arp
  - Plus restrictive que le broadcast IPv4, elle limite le traitement par les hôtes non concernés (drop dès la couche 3)
  - Les switchs implémentant le MLD snooping ne les retransmettent que sur les ports nécessaires
- Elle est construite en concaténant le préfixe FF02::1:FF00:0/104 et les 24 derniers bits de l'adresse unicast que l'on veut joindre
- Exemple :
  - Adresse unicast : 2001:543:7502:2365:200:abff:fecb:123/128
  - Adresse multicast solicited-node : FF02::1:FFcb:123/128



L'adresse multicast solicited-node



Les adresses anycast

- Ne peuvent être distinguées d'une adresse unicast
- Entre multicast et unicast, seul le nœud le plus proche reçoit le paquet et y répond
- Une seule utilisation aujourd'hui :
  - les routeurs doivent posséder l'adresse anycast «subnet prefix»::/128
  - Dans le réseau 2001:AB:12:CD:34:EF:56:76::/64, le routeur le plus proche est joignable à l'adresse 2001:AB:12:CD:34:EF:56:76::/128
- Leur utilisation reste assez flou

Plusieurs adresses par interface

- Exemple d'une interface sur un hôte:
  - Une adresse de type local-link (obligatoire):
    - **FE80::**200:abff:fecb:123/128
  - Une ou plusieurs adresse(s) unicast :
    - Globale : **2001:**543:7502:2365:200:abff:fecb:123/128
    - De site : **FD00:**543:7502:2365:200:abff:fecb:123/128
- La sélection de l'adresse à utiliser est faites par l'OS en fonction d'une politique définit (RFC 3484)
  - Prefer Same Address, Prefer Appropriate Scope, Prefer Public Address, ...

Plusieurs adresses par interface

- Exemple d'une interface sur un routeur :
  - Une adresse de type local-link (obligatoire):
    - **FE80::**0200:abff:fe:cb:9876/128
  - Une ou plusieurs adresse(s) unicast :
    - Global :

      2001:0543:7502:2365:0200:abff:fecb:9876/128
    - Et/ou de site :

      FD00:0543:7502:2365:0200:abff:fecb:9876/128
  - Les adresses anycast correspondantes (pour les interfaces sur lesquels il est routeur):
    - FE80::/128
    - 2001:0543:7502:2365::/128
    - FD00:0543:7502:2365:0200::/128

Les adresses oligatoires

- Un hôte doit à minima se reconnaître :
  - Sur son adresse de loopback (::1/128)
  - Ses adresses unicast (2000::/3 ;FD00::/8)
  - Ses adresses local-link (FE80::/10)
  - Ses addresses multicast solicited-node
  - Les addresses multicast « All-Nodes »(FF01::1/128, FF02::1/128)
  - Les adresses multicast des groupes spécifiques qu'il a rejoint
- Un routeur doit en plus se reconnaître dans:
  - Ses adresses anycast (« prefix unicast »::/128)
  - Les adresses multicast « All-routers »(FF01::2/128, FF02::2/128, FF05::2/128)
  - Les adresses liées au protocole de routage (exemple, si RIP : FF01::9/128, FF02::9/128)

Les protocoles associés

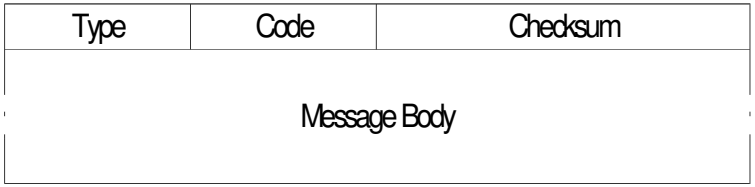
DNS

- Actuellement, configuration proche de celle d'IPv4 (RFC 1886)
  - Enregistrement type AAAA pour la résolution de noms : équivalent de l'enregistrement A en ipv4, l'adresse complète est définie :

    pc1.truc.fr 3600 IN AAAA2001:abc:123::2ca:11FF:FE22:3344
  - Enregistrement PTR classique pour le reverse, définition spécifique pour ipv6 (.ip6.arpa), chaque suite de 8 bits est représentée :

    4.4.3.3.2.2.E.F.F.F.1.1.a.c.2.0.0.0.0.0.3.2.1.0.c.b.a.0.1.0.0.2.ip6.arpa IN PTR pc1.truc.fr
- La RFC 2874 définit les entrées A6 qui doivent simplifier la syntaxe des fichiers (la résolution des préfix est déléguée). Celle-ci est conservée pour usage futur, mais la RFC 1886 (AAAA) est à privilégier pour le moment.

ICMPv6



- Encore plus important en v6 qu'en v4
- Intègre les fonctionnalités d'ICMPv4, ARP et IGMP (multicast IPv4) via MLD (Multicast Listener Discovery)
- Protocole de contrôle d'IPv6, partie intégrante de l'architecture, il ne faut surtout pas filtrer ICMP en IPv6
- Modification des messages par rapport à ICMPv4, suppression de messages obsolètes et ajout de nouveaux messages
- ICMPv6 est encapsulé dans un datagramme IPv6 (next header = 58)

Rôle d'ICMPv6

- Rapporter les erreurs (type < 128):
  - 1 Destination Unreachable
  - 2 Packet Too Big (PMTUD, Path MTU Discovery)
  - 3 Time Exceeded
  - 4 Parameter Problem (erreur dans l'entête IPv6)
- Effectuer les diagnostiques :
  - 128 Echo Request
  - 129 Echo Reply

Rôle d'ICMPv6

- Découverte des voisins (ND, Neighbor Discovery)
  - 133 Router Solicitation
  - 134 Router Advertisement
  - 135 Neighbor Solicitation
  - 136 Neighbor Advertisement
- Gestion du Multicast (MLD, Multicast Listener Discovery)
  - 130 Group Membership Query (recensement des récepteurs)
  - 131 Group Membership Report (rapport d'abonnement)
  - 132 Group Membership Reduction (résiliation d'abonnement)

PMTUD (Path MTU Discovery)

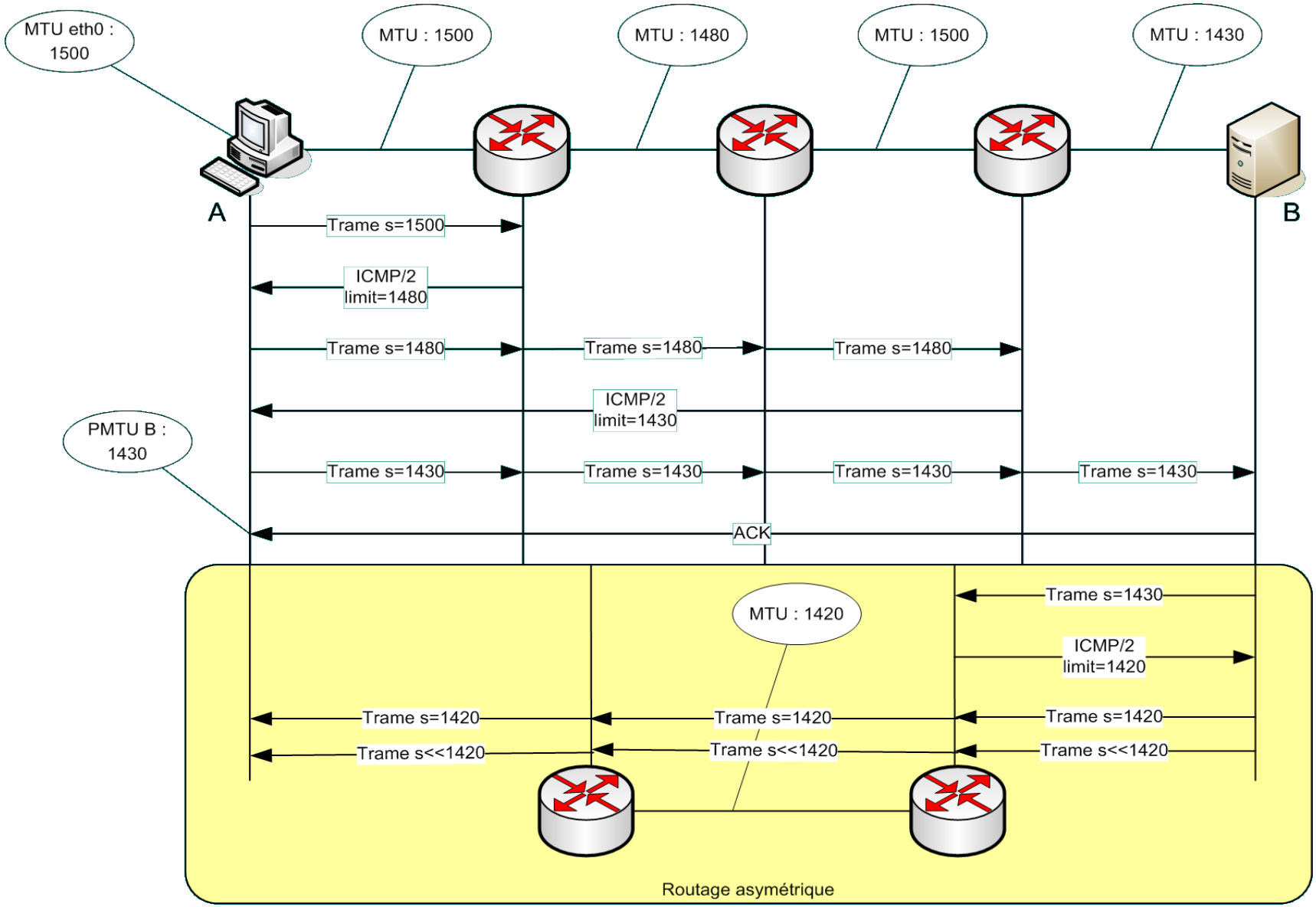
- Rappel : Dans les simplifications de routage d'IPv6, la fragmentation a été retirée
- La fragmentation des paquets est à la charge de l'émetteur

  => Celui-ci doit connaître, et donc décourir le MTU du lien utilisé
- Mis en place avec IPv4, mais fonctionne mal (filtrage ICMP)
- Pose des problèmes, notamment par son intéraction avec le MSS de TCP dans le cas de routage asymétrique
- Risque de DoS par annonces malveillantes de faible MTU

PMTUD (Path MTU Discovery)

- Principe :
  - La source envoie un premier datagramme de la taille maximale qu'elle peut utiliser
  - Dès qu'un routeur ne peut faire passer ce message, il envoie à la source un message ICMP/2 (Packet too big)
  - La source réemet le datagramme avec cette nouvelle taille
  - Dès que le paquet a traversé tout le réseau, la source reçoit la réponse de la destination
  - Elle dispose alors du MTU pour ce chemin
  - La source peut régulièrement tenter un datagramme de taille supérieure pour détecter une modification du PMTU

PMTUD (Path MTU Discovery)



ND (Neighbor Discovery)

- Permet :
  - L'association IP/Mac (remplace ARP en IPv4)
  - La découverte des routeurs (ICMPv4 router discovery / router redirect)
  - La détection des adresses dupliquées (DAD, Duplicated Address Detection)
  - La détection des voisins inaccessibles (NUD, Neighbor Unreachability Detection)
- Fortement utilisé dans l'autoconfiguration

ND (Neighbor Discovery)

- Les types de messages :
  - ICMP 133 : Router Solicitation (RS) :
    - Lancé au démarrage de la machine pour obtenir des informations des routeurs du lien
    - Utilise l'adresse source indéterminée si aucune adresse n'est définit (::/128)
    - Emis vers l'adresse multicast des routeurs du lien (FF02::2)
  - ICMP 134 : Router Advertisement (RA) :
    - Emis périodiquement par le routeur (ou en cas de solicitation)
    - Utilise l'adresse source local-link du routeur (FE80::/10)
    - Emis vers l'adresse multicast de tous les nœuds du lien (FF02::1)

ND (Neighbor Discovery)

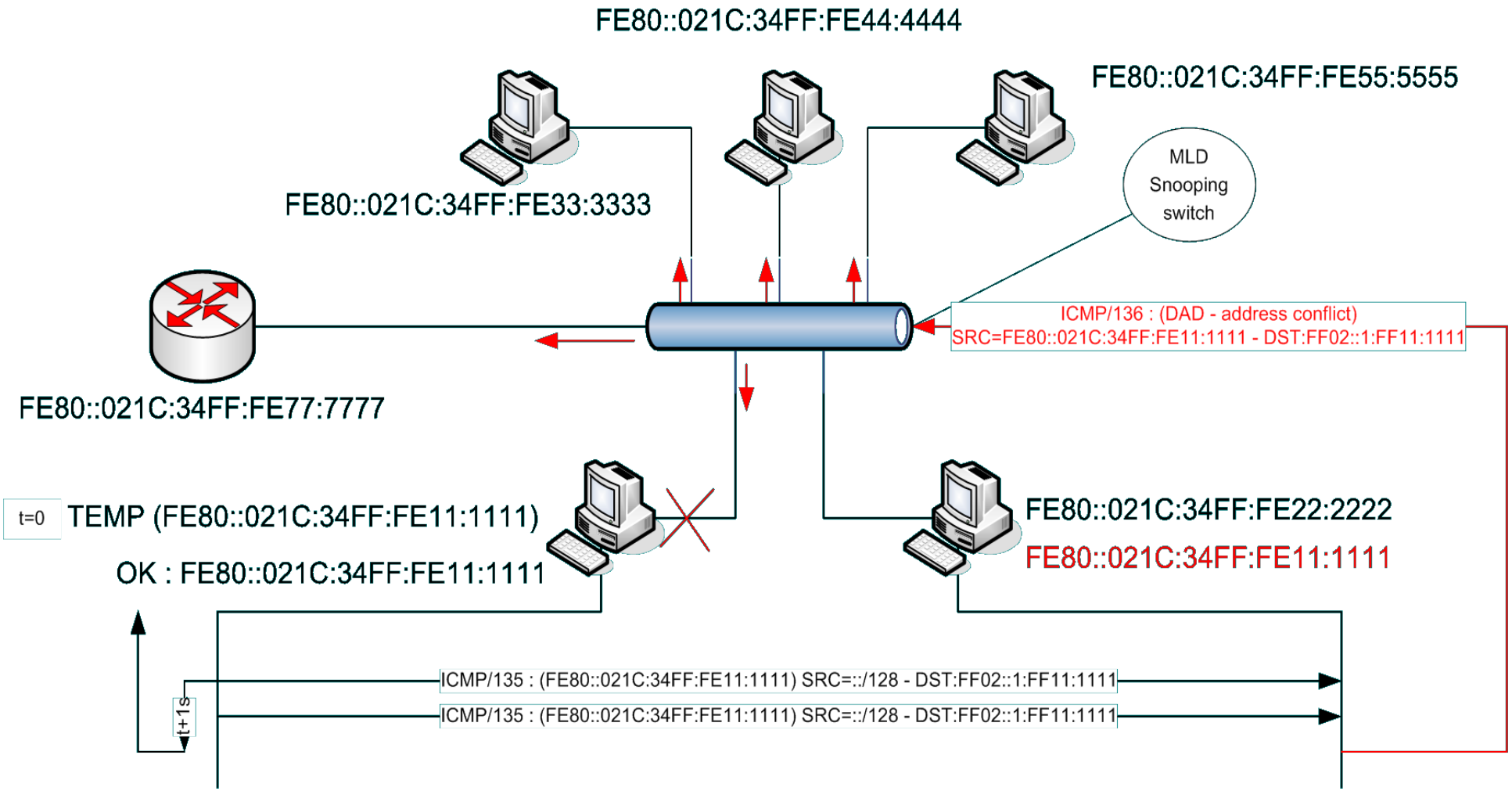
- Les types de messages (suite) :
  - ICMP 135 : Neighbor Solicitation (NS)
    - Pour obtenir des informations d'un voisin sur le lien local
    - Utilise l'adresse lien-local, l'adresse unicast global ou l'adresse indéterminée suivant les cas
    - Emis soit vers l'adresse multicast solicité (solicited multicast address) de l'adresse recherchée (ARP en IPv4), soit directement vers l'adresse du nœud (NUD)
  - ICMP 136 : Neighbor Advertissement (NA)
    - Réponse à une solicitation (NS)
    - Spontané (changement d'état de l'interface sur le lien)

DAD (Duplicated Address Detection)

- Permet de vérifier que l'adresse (auto)configurée est unique
- Tant que cette vérification n'a pas eu lieu, l'adresse est provisoire (les messages ne sont pas acceptés)
- Principe :
  - Utilisation de l'adresse indéterminée (::/128)
  - Envoie de paquets NS vers l'adresse multicast solicité correspondant à l'adresse provisoire (au moins 2 fois)
  - Si pas de réponse après 1s, l'adresse devient valide
  - Si un voisin possède l'adresse, il prévient tous les voisins du lien (NA) de la colision d'adresses
  - L'adresse ne peut être affectée, l'intervention humaine est nécessaire

DAD (Duplicated Address Detection)

RFC 4429 : Even if the node moves into networks of 50000 nodes once per minute for 100 years, the probability of it causing or suffering a collision at any point are a little over 1 in a million.



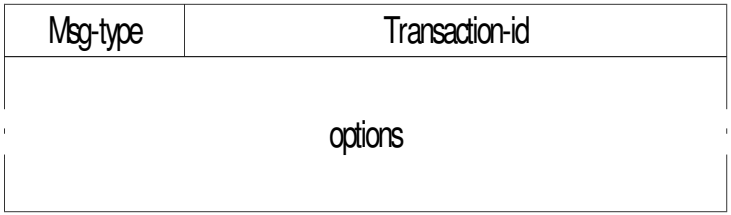
Autoconfiguration

- Généralement précédé de la création d'une adresse « local-link » (associé à une DAD)
- Permet la configuration automatique d'une interface
- Deux méthodes :
  - Autoconfiguration sans état (stateless autoconfiguration)
  - Autoconfiguration avec état (statefull autoconfiguration)
- Rôle important du routeur, il annonce aux machines la méthode de configuration à utiliser
- En l'absence de routeur, le nœud tente une configuration avec état, si celle-ci n'aboutit pas, il n'aura que son adresse local-link (sauf problème de DAD)

Autoconfiguration sans état

- Utilisée lorsque l'affectation d'adresse IPv6 ne nécessite pas d'attention particulière
- Principe :
  - L'hôte démarre et envoie un paquet type RS (::/128 => FE02::2/128)
  - Le routeur renvoie un paquet type RA (FE80::/10 => FE02::1/128) pouvant contenir :
    - Le préfix
    - Le subnet
    - La passerelle par défaut
    - Le mode d'autoconfiguration sur le lien (statefull / stateless)
    - D'autres options (serveur DNS, MTU, ...)
  - L'hôte va alors :
    - générer son adresse IP temporaire (prefix + identifiant d'interface)
    - Effectuer une vérification d'adresse dupliqué (DAD) avant de valider son adresse
- Enfin, si nécessaire, l'hôte contacte un serveur DHCPv6 stateless pour les informations complémentaires (DNS, …) si le routeur ne lui a pas fourni

DHCPv6



- Basé sur UDP
- Utilisation d'un identifiant DHCP unique (DUID, DHCP Unique Identifier) coté client et coté serveur.
  - 3 méthodes de génération définit en fonction du type de matériel
  - Dans tous les cas, cet identifiant doit être aussi stable que possible (dans l'idéal ne jamais changer)
  - L'adresse MAC disparaît des messages DHCP, c'est ce DUID qui permet d'associer un client à des informations particulières
- Permet l'affectation :
  - D'une adresse IP prédéfinit (en fonction du DUID)
  - D'autres informations :
    - Passerelle
    - Serveur DNS
    - Beaucoups d'autres options (serveur ntp, ...)
- Cas particulier de DHCPv6 stateless (pour l'autoconfiguration sans état) qui fournit uniquement les informations complémentaires

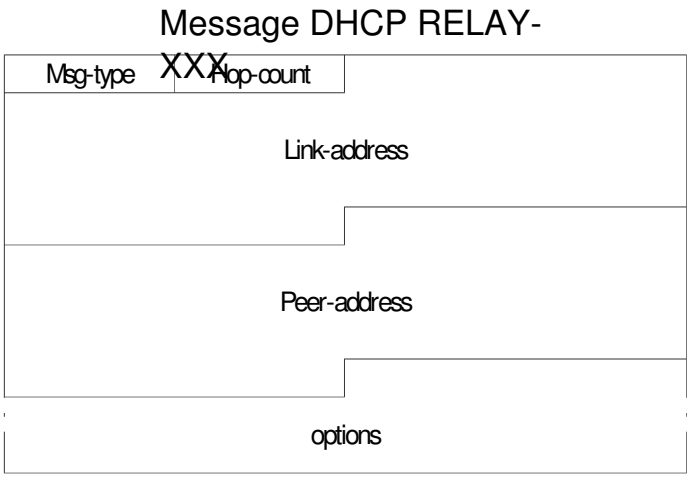
Autoconfiguration avec état

- Principe :
  - Envoie d'un message DHCP SOLICIT (1) depuis l'adresse local-link (FE80::interface_ID) vers l'adresse multicast FF02::1:2 (tous les serveurs DHCP du lien) pour identifier les DHCP disponible en mesure de fournir une adresse
  - Réponse du (ou des) serveur(s) DHCP type ADVERTISE (2) depuis l'adresse local-link du serveur vers l'adresse local_link du client. Une option priority peut être défini
  - Si des priorités ont été fournis au client, il choisit le serveur avec la plus faible priorité et effectue un DHCP REQUEST (3) (local-link => local-link) pour obtenir ses informations de configuration
  - Le serveur interrogé lui fournit ces informations via un message DHCP REPLY (7)
- Les étapes 2 et 3 peuvent être supprimées si le client précise l'option « Rapid Commit » dans son message de solicitation
- Le client doit effectuer une DAD sur l'adresse avant de la considérer valide
- Dans le cas d'une autoconfiguration sans état, le client utilise un message DHCP INFORMATION-REQUEST (11) pour obtenir uniquement les informations complémentaires

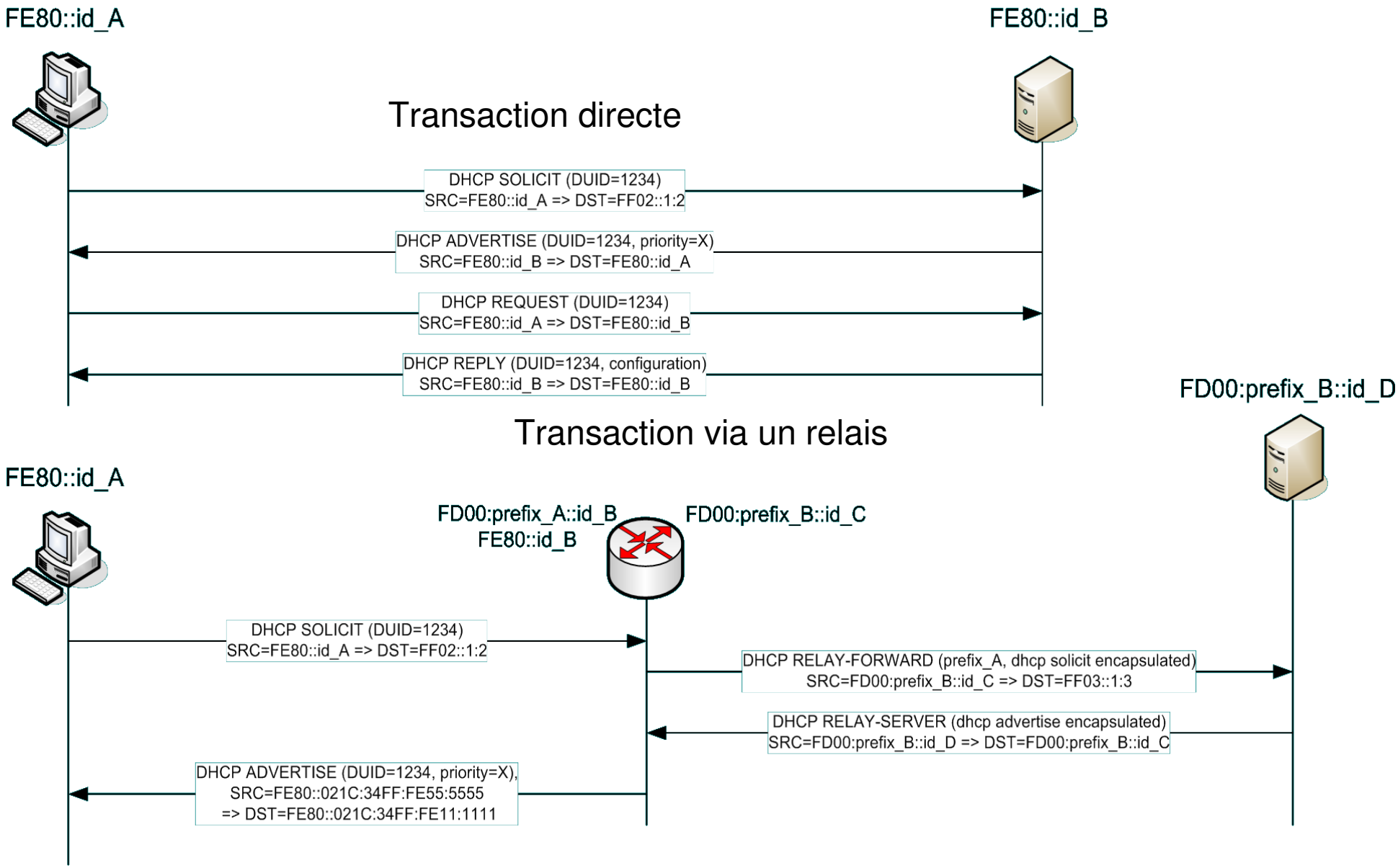
Autoconfiguration avec état

- Si le serveur DHCP n'est pas sur le lien local, un relais peut se charger de transmettre les requetes :
  - Il joue le rôle du serveur coté client et encapsule les requêtes dans des paquets RELAY-FORWARD (12). Il ajoute le préfix de l'interface ou il a reçu la demande (et d'autres options en fonction des cas) et envoit ces paquets soit à l'adresse unicast d'un serveur DHCP, soit à l'ensemble des serveurs DHCP du site (FF05::1:3)
  - Le serveur lui répond par des paquets RELAY-REPLY (13) encapsulant le message DHCP à transmettre au client
- Les autres messages :

CONFIRM (4)	Demande à tous les serveurs DHCP la confirmation de validité des informations
RENEW (5)	Demande au serveur DHCP ayant fournit l'adresse de prolonger celle-ci
REBIND (6)	Demande à tous les serveurs DHCP la prolongation de l'adresse
RELEASE (8)	Déclare la libération de l'adresse au serveur l'ayant fournit
DECLINE (9)	Signale un problème de DAD sur l'adresse proposée et le refus de celle-ci



Autoconfiguration avec état



Durée de vie d'une adresse

- Une adresse IPv6 a une durée de vie limitée (sauf l'adresse local-link), 4 états sont possibles en fonction des timer fournis lors de la configuration (par le routeur ou le serveur DHCPv6) :
  - Tentative : Avant validation par la DAD
  - Preferred : L'adresse est totalement valide :
    - t < preferred lifetyme
  - Deprecated : L'adresse ne peut plus être utilisé que pour finir les communications en cours :
    - preferred lifetime < t < valid lifetime
  - Invalid : L'adresse ne peut plus être utilisée
    - t > valid lifetime