



# Multihome BGP sur REVE



## Le Réseau métropolitain d'Évry Val d'Essonne (REVE)

### *Objectifs initiaux:*

- Favoriser le développement des communications « haut débit » d'établissements implantés à Évry
- Mutualiser la sortie Internet et la supervision des infrastructures.
- Supporter des membres non éligibles Renater - > Second opérateur initialement prévu

### *Projet en trois lots:*

- Un volet «infrastructure fibre optique » qui a été livré début 2001, maintenance SPIE
- Un volet «fourniture des équipements actifs et exploitation » qui a fait l'objet d'un marché (CS)
- Un volet pour la fourniture d'un accès haut débit longue distance vers « Internet » - > Nœud Renater (Renater)

### *Opérationnel depuis le mois de juin 2001*



# Multihome BGP sur REVE

## Partenaires



- *Un groupe d'une dizaine de partenaires publics et privés*
  - l'Université d'Évry Val d'Essonne (UEVE)
    - le GIP GENOPOLE
  - le Consortium National de Recherche Génomique
  - l'Institut National des Télécommunications (GET- INT)
  - le Centre des Matériaux de l'École Nationale des Mines de Paris (ENSMP)
    - le laboratoire GENETHON
  - l'École Nationale Supérieure pour l'Industrie et l'Entreprise (ENSIIE).
- sont associés à ce projet des acteurs nationaux du monde de la recherche (comme le CNRS, l'INRA via l'unité URGV ou l'INSERM)
  - Initialement: Genset, société privée.
- *Des utilisateurs passifs (Fibre noire) non membres*
  - CNES
  - SNECMA



# Multihome BGP sur REVE Topologie



- Le réseau est constitué d'une boucle optique d'une vingtaine de kilomètres sur la ville d'Evry
- Cette boucle est constituée de 72 Fibres monomodes.
- Sur chacun des 11 sites actuels, 12 paires sont sorties sur connecteurs.
- La fibre est éclairée par du matériel cisco et implémente du GigaEthernet.
- Les connexions clientes sont redondées - > 2x1vlan par sens de boucle vers un HSRP.
- Les Points de connexion supportent du 10/ 100/ 1000 suivant les capacités des clients.
- Les services IPv4 unicast et multicast ainsi que IPv6 sont supportés.
- Les clients annoncent leurs réseaux par des sessions eBGP avec le coeur de REVE.

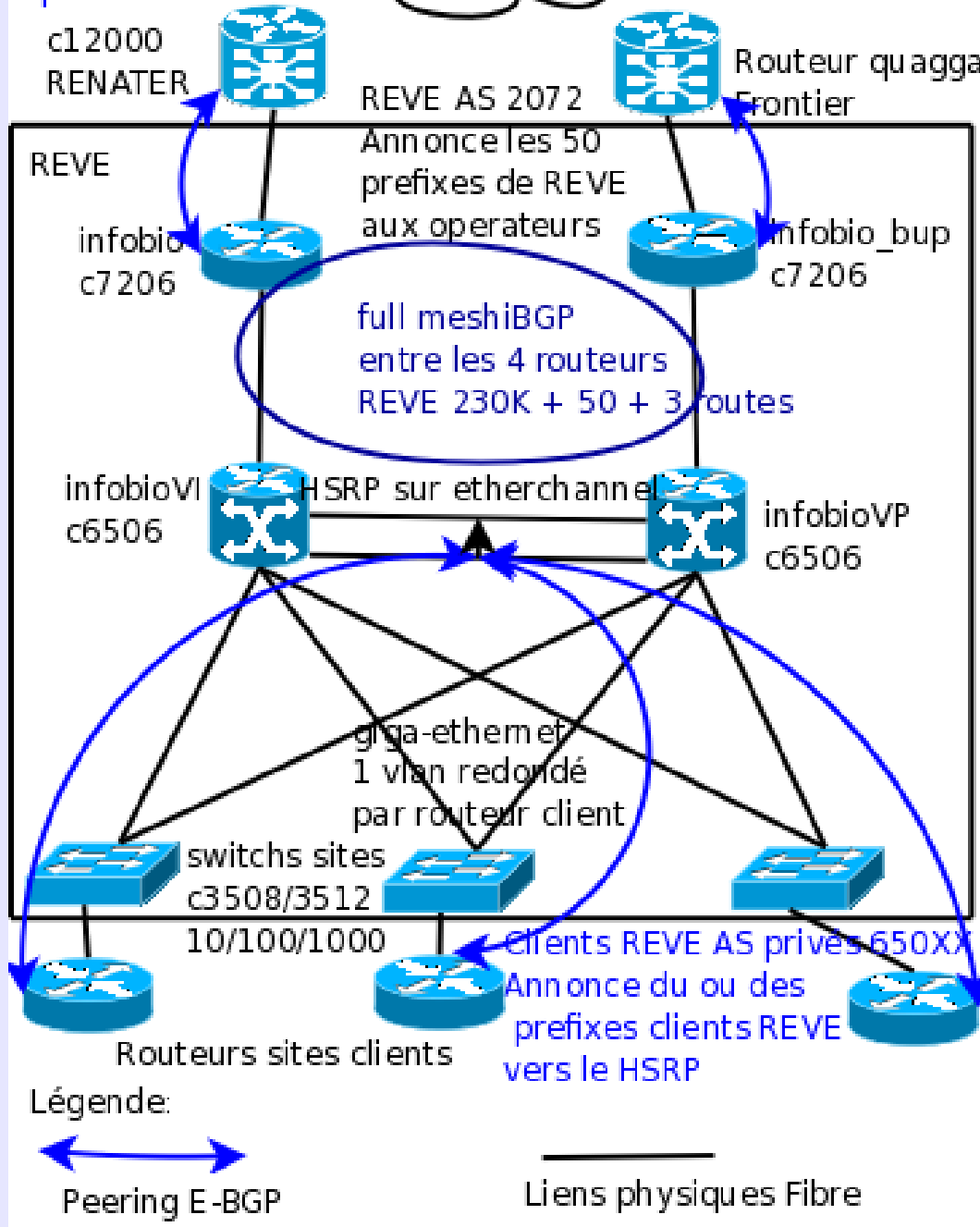


RENATER AS 2200  
Annonce 230K  
prefixes vers REVE



Internet

AS 13049  
Annonce 3 prefixes  
vers REVE



Schéma



# Multihome BGP sur REVE



## Choix de l'opérateur alternatif

Consultation non formalisée auprès d'opérateurs Internet (Colt, Cogent, Completel, NeufCegetel)

Choix d'un opérateur local - > Frontier Online

Critères:

- double connectivité (2 liens distincts Evry <- > Paris) sur Téléhouse- 2
- échange du trafic (peering) avec RENAIER au SFINX
- proximité géographique, seulement 200m de Fibre, dont 100m de GC
- rapide mise à disposition
- tarifs concurrentiels



# Multihome BGP sur REVE

## Besoins et contraintes



### Besoins d'un second opérateur:

- continuité de service: assurer un backup en cas de pannes programmées ou non de Renater
- permettre aux utilisateurs non éligibles de disposer d'un FAI.
- trafic non en adéquation avec la charte Renater (Internet commercial, perso... utopie ?)
- le déménagement du NR fin juillet 2006 a précipité ce besoin.

### Contrainte pour les membres existants:

- non renumérotation des réseaux des membres actuels
- annonce de préfixes RENATER via Frontier
- priorité RENATER en sortie et entrée pour les sites éligibles - > éviter un routage asymétrique
- routage politique dans REVE



# Multihome BGP sur REVE



## Solution retenue

La principale contrainte était de favoriser le trafic aller et retour via RENAIER pour les sites éligibles

AS Prepend : répéter artificiellement dans l'AS- PATH des routes annoncées à Frontier l'AS de REVE (2072)

```
route-map V4- REVE_FrontierOnline- OUT permit 10  
set as-path prepend 2072 2072
```

### LG Gtoyen:

*BGP routing table entry for 157.159.0.0/ 16*

*Paths: (2 available, **best #2**, table Default- IP- Routing- Table)*

[13049 2072 2072 2072 2072](#)

*Origin IGP, localpref 1000, valid, external*

*Last update: Thu Jul 20 19:15:54 2006*

[2200 2072](#)

*Origin IGP, localpref 1000, valid, external, **best***

*Last update: Thu Jul 20 17:23:15 2006*



# Multihome BGP sur REVE

## Résultat de l'AS- Prepend



L'AS Prepend fonctionne généralement bien (Constat: ~75% trafic IN)

Restent ~25% des AS (ISP) qui ne "repectent" pas l'AS- prepend - > local- pref divergentes - > trafic IN via Frontier .

lg.gitoyen.net - Looking Glass, quelques mois plus tard:

*BGP routing table entry for 157.159.0.0/ 16*

*Paths: (2 available, **best #2**, table Default- IP- Routing- Table)*

[2200 2072](#)

*Origin IGP, **localpref 1000**, valid, internal*

*Community: 20766:21000 Last update: Wed Oct 31 17:31:27 2007*

[13049 2072 2072 2072](#)

*Origin IGP, **localpref 1300**, valid, internal, **best***

*Community: 20766:23000 Last update: Wed Oct 31 16:13:33 2007*

*giga- bgpd*





# Multihome BGP sur REVE

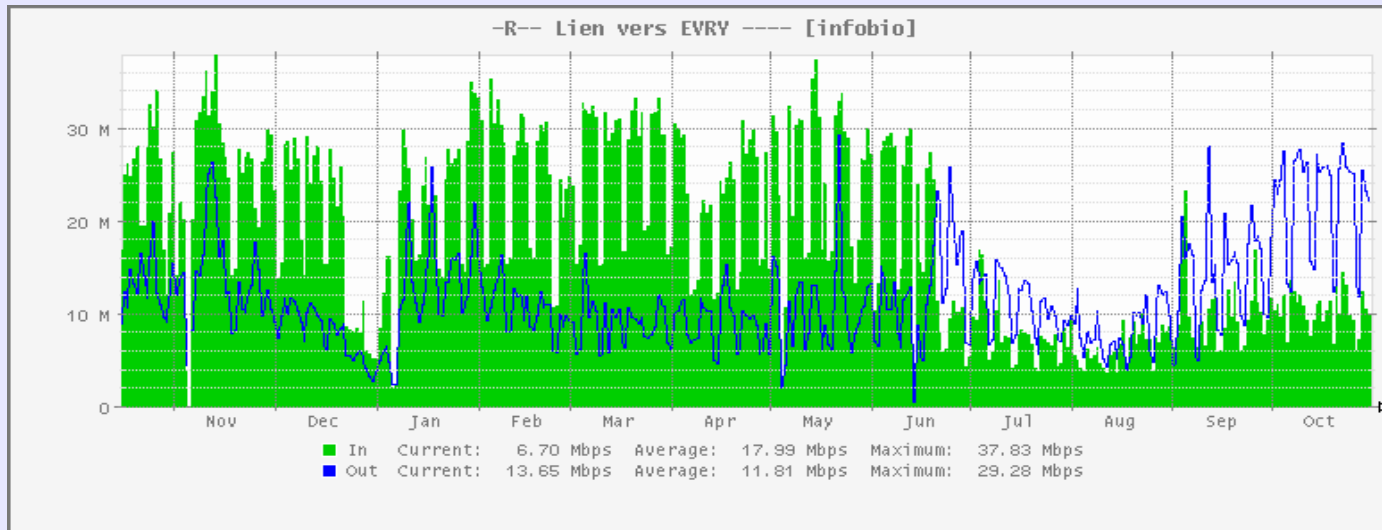


## Résultat de l'AS- Prepend

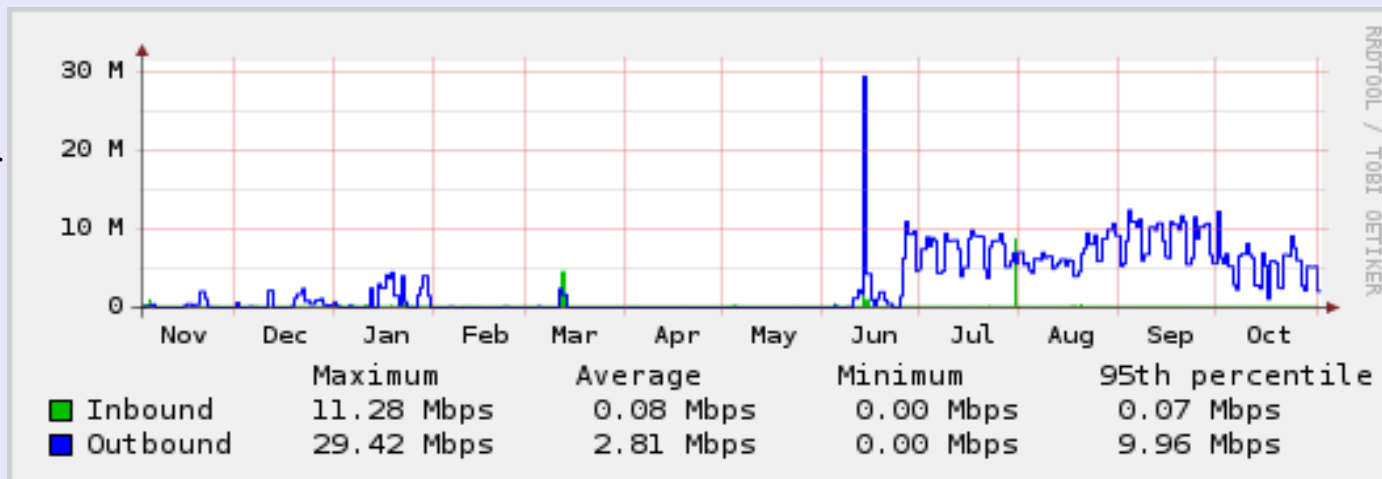
Trafic symétrique lorsque qu'un seul opérateur annonce REVE.

Autrement, trafic asymétrique - > Out: 100% Renater, IN: 75% Renater 25%Frontier

Renater



Frontier





# Multihome BGP sur REVE



## Contournement

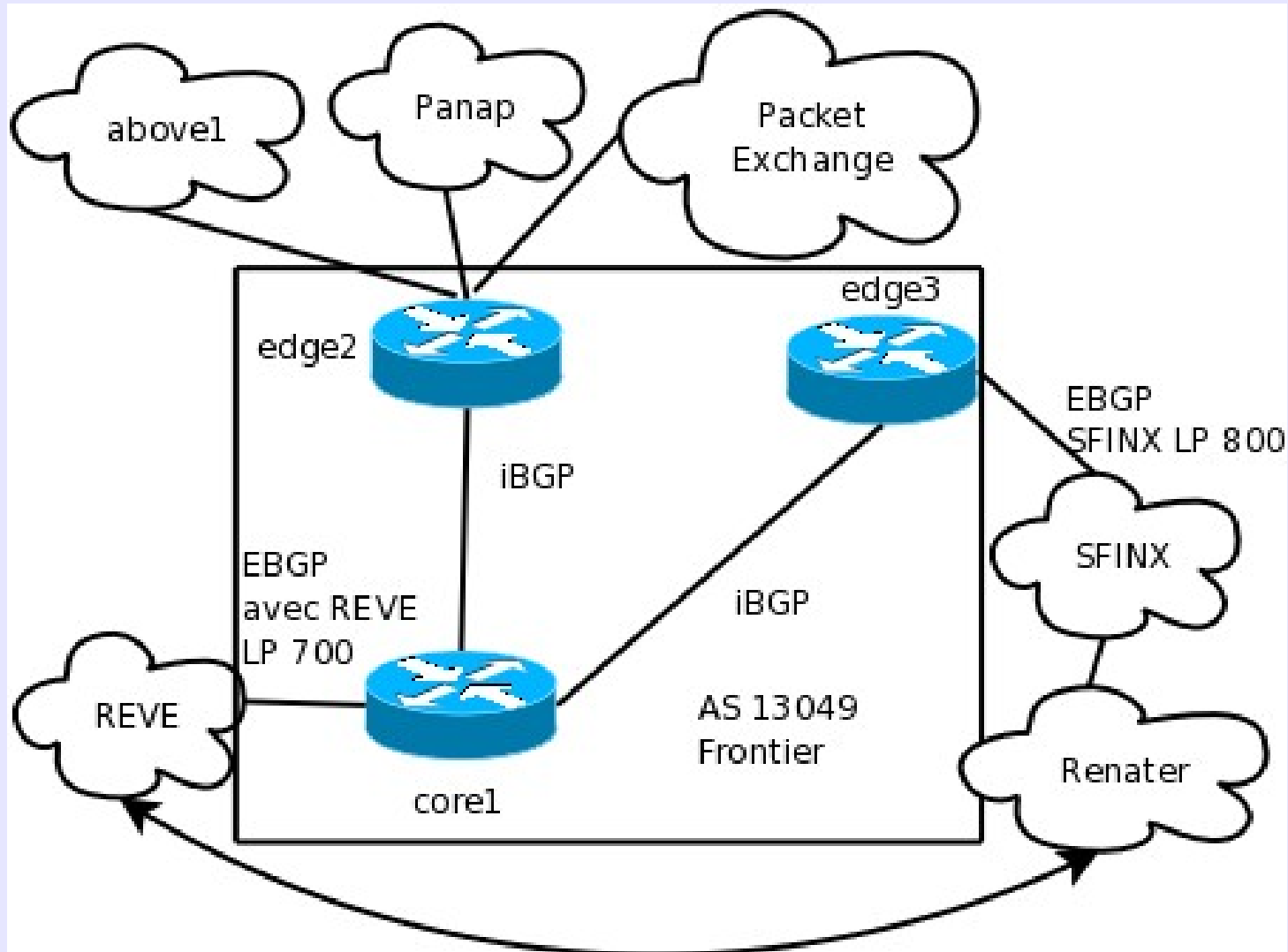
Afin de respecter le principe de trafic symétrique via Renater, Frontier annonce les préfixes de REVE uniquement lorsqu'ils ne lui sont plus visibles depuis Renater (peering au SFINX) .

### Principe technique:

- En ajustant les LocalPref internes au coeur de réseau de Frontier,
- Sur les règles du protocole BGP en terme de propagation d'annonces iBGP vs eBGP

# Multihome BGP sur REVE

Condition d'annonce des préfixes REVE par Frontier





# Multihome BGP sur REVE

## Événements imprévus



- Traitements particuliers sur les "gros" préfixes - > dampening sur 157.159.0.0/ 16
- Sous dimensionnement d'un routeur de bordure REVE (infobio\_bup ne supporte pas le full routing)
- Boucles de routage - > effet de bord lié au sous dimensionnement des 6506 et de la redondance au coeur de REVE).

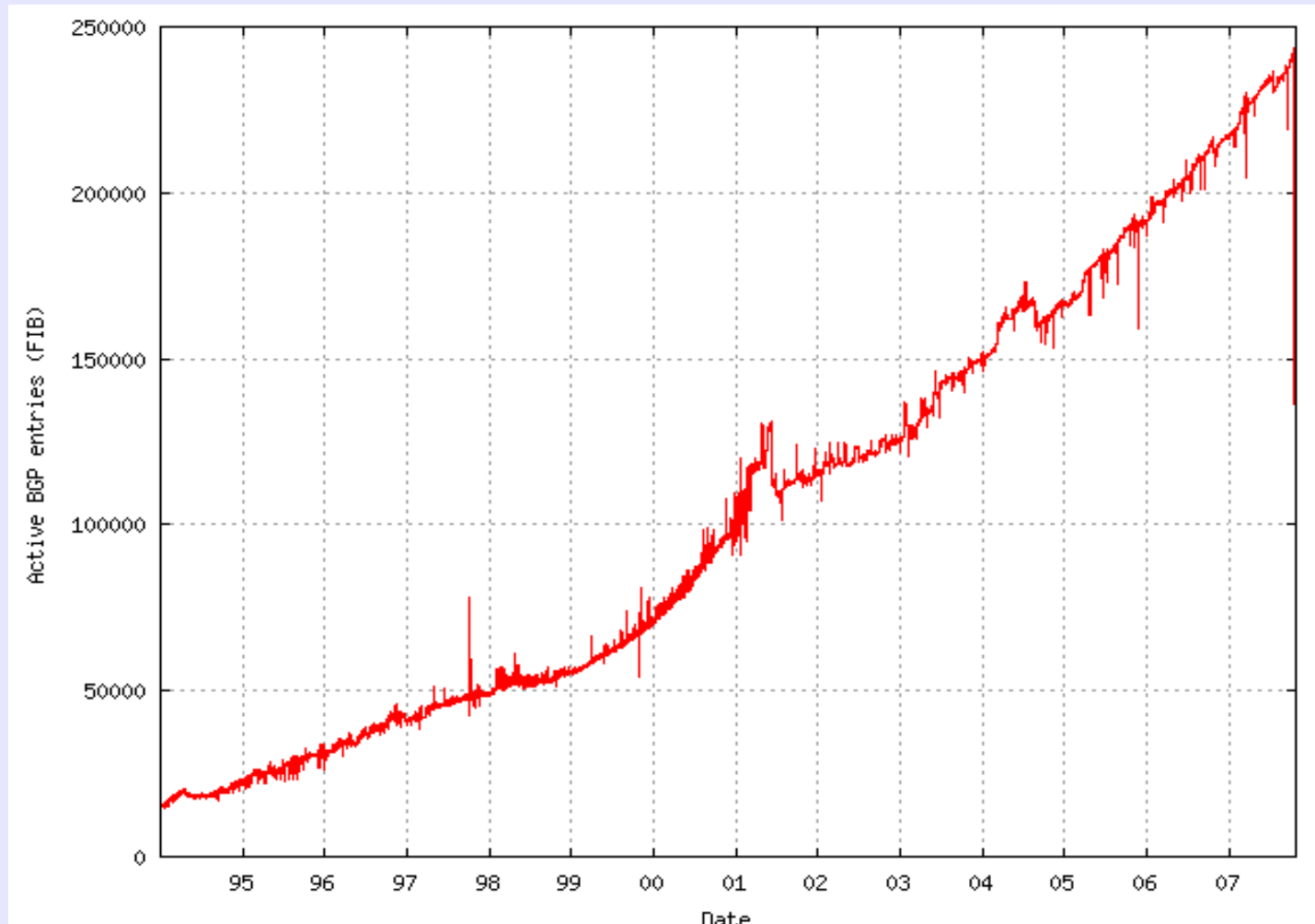
Croissance conséquente du nombre de route dans la Default Free Zone (DFZ) cf RFC4984, les 6506 ne supportent pas.



# Multihome BGP sur REVE



Evolution du nombre de préfixes BGP sur 10 ans, 241.000 le 25- 10- 2007



JRES-2007 Multihome BGP sur REVE



# Multihome BGP sur REVE

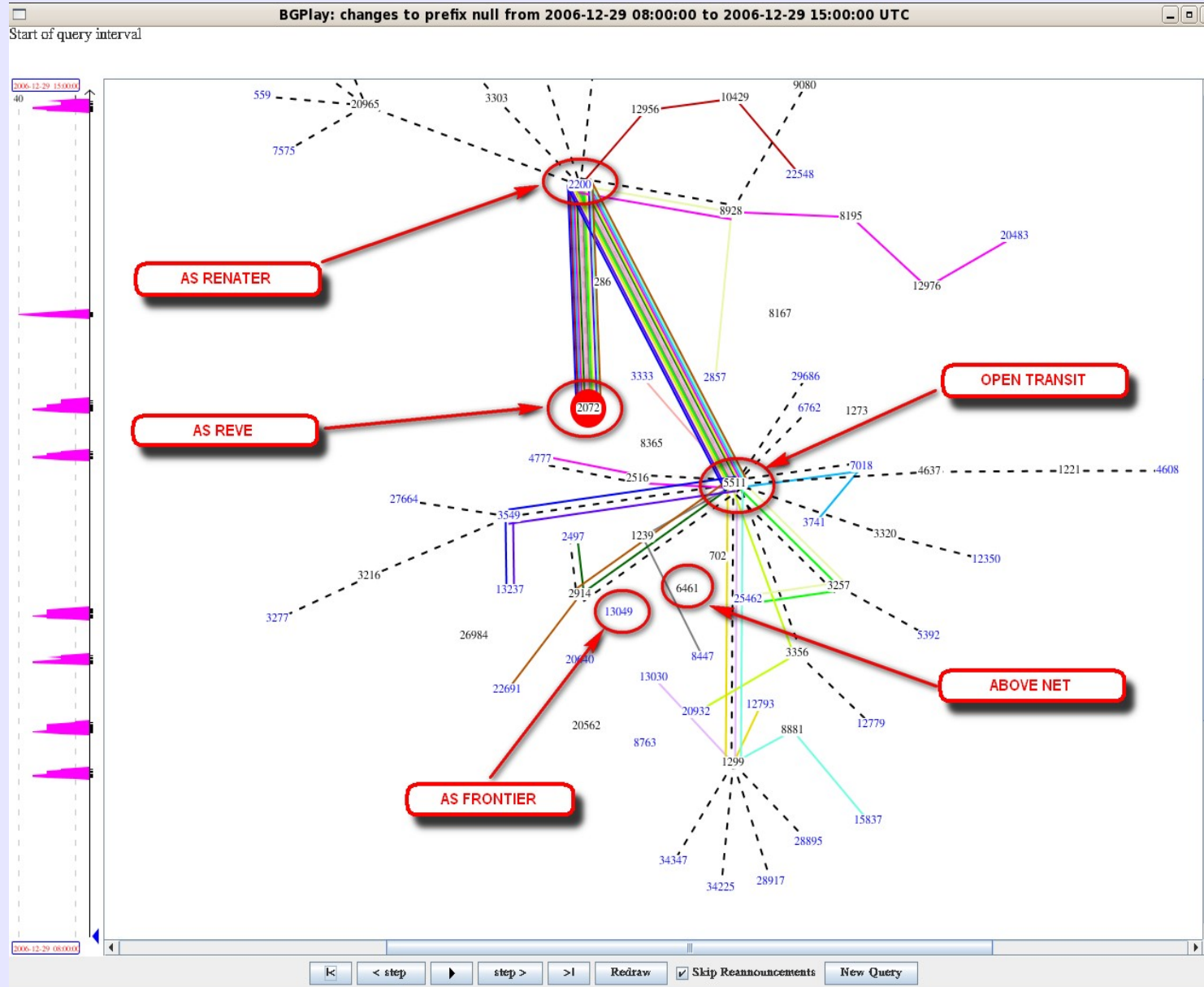


## Événements imprévus

- Perturbations sur le peering au SFINX (cisco <- > quagga !?)  
Conséquences: le contournement BGP mis en place pour éviter le trafic asymétrique n'est plus opérationnel.
- Les pertes de peering intempestives au SFINX ainsi que le débordement des tables de forward REVE (6506) provoquent de régulières bascules vues depuis l'extérieur entre Renater et Frontier pour revenir sur REVE

# Multihome BGP sur REVE

- Bascules Renater/ Frontier



# Multihome BGP sur REVE

## Animation des bascules Renater/Frontier







# Multihome BGP sur REVE Evolution de l'architecture



- Double Full Routing Renater –Frontier sur les routeurs de bordure de REVE - > éviter qu'un seul des ISP aspire le trafic via LA route par défaut (0.0.0.0/ 0).
- Relier les 2 routeurs de bordure REVE
- Décharger les routeurs internes de REVE (6506) du full routing, reception de 2 routes par défaut (0.0.0.0/ 0), choix d'une via localpref, l'IGP assurera le cheminement vers le Next Hop.





# Multihome BGP sur REVE

## Bilan



- Opérationnel en temps voulu - > déménagement du NR Juillet 2006
- Continuité de service assurée en cas de coupure d'un des opérateurs (incident, arrêts programmés - > convergence en quelques minutes)
- Dysfonctionnements imprévus lorsque les 2 ISP sont opérationnels liés à:
  - ✓ défauts d'architecture (full routing sur 6506, route par défaut d'1 ISP)
  - ✓ dimensionnement des équipements
  - ✓ pratiques originales (Multihome BGP peu commun dans nos sphères)
  - ✓ nombreuse parties en jeux (opérateurs, intégrateurs, fournisseur de matériel)
  - ✓ technologie complexe, BGP, coeur de l'Internet



# Multihome BGP sur REVE

## Bilan après un an et 1/2



- Un déménagement du NR - > 72H de bascule sur Frontier (12 H d'intervention)
- 6 bascules parfaitement opérationnelles liées à des travaux de maintenance et/ ou pannes REVE- Renater, allant de quelques minutes à plusieurs heures .
- Un cumul d'une quinzaine de jours de dysfonctionnements - > Services Internet plus ou moins opérationnels en fonction des sites distants, principalement pour l'INT .

Les améliorations envisagées en termes d'architecture, de dimensionnements et de pratiques des technologies mise en oeuvres, vont sérieusement fiabiliser la solution.



# Multihome BGP sur REVE



## Conclusion

- Expérience riche en enseignements,
- Une continuité de service améliorée, malgré les imprévus ,
- Une architecture à réviser pour de meilleures performances .