



Architecture pour la prise en compte des raccordements multiples

Description : Ce document présente l'architecture, principalement basée sur le protocole de routage BGP, pour les raccordements multiples des sites à RAP.

Version actuelle : 1.12

Date : 22/09/09

Auteur : LD

Version	Dates	Remarques
1.0	29/05/07	Création du document
...
1.12	22/09/09	Ajout « Ultime Recours »

Table des matières

Table des matières	2
Introduction	2
1 Raccordement fiabilisé de RAP sur RENATER :	3
2 Raccordement fiabilisé d'un site sur RAP :	4
2.1 Raccordement d'un site à 2 PEs :	6
2.2 Raccordement de 2 sites secourus mutuellement.....	8
2.3 Envoi de la communauté à RENATER.....	9
2.4 Accès réseau d'ultime recours :	10
2.5 Accès nominal par un autre FAI :	12
2.6 Exemples de configurations pour le routeur de site.....	13

Introduction

Le raccordement fiabilisé à RENATER permet une répartition de charge et une redondance entre les deux routeurs d'accès de RAP : pe-jussieu et pe-odeon.

La répartition du trafic entre RAP et RENATER s'obtient par le partitionnement de RAP en deux zones : la zone Odéon (O) et la zone Jussieu (J). Dans l'état normal de fonctionnement, le trafic des sites raccordés aux PoP Odéon, Malesherbes et Auteuil transite par pe-odeon, et, le trafic des sites raccordés aux PoP Cnam et Jussieu transite par pe-jussieu.

Les mécanismes permettant la redondance et la répartition de charge ont donc été généralisés pour que les sites puissent bénéficier, à leur tour, d'un raccordement fiabilisé sur RAP.

Ce document est valable pour les services IPv4 et IPv6 unicast et vient compléter, au niveau des mécanismes et de l'architecture de routage, le document « Mise en place de raccordement fiabilisé pour les sites sur RAP ». Pour les services IPv4 et IPv6 multicast, un raccordement en mode préconisé et une étude au cas par cas sont nécessaires. Cependant, la répartition de charge du trafic IPv6 entre RAP et RENATER est moins souple, puisque seuls les /48 sont acceptés par RENATER. Pour les VPN de niveau 2, il faut se référer au document http://www.rap.prd.fr/pdf/Service_L2VPN_VPLS.pdf.

1 Raccordement fiabilisé de RAP sur RENATER :

Deux *peering* BGP sont établis entre RAP et RENATER depuis pe-jussieu et pe-odeon :

- RENATER annonce à RAP toutes les routes de l'Internet (BGP Full Routing)
- RAP annonce à RENATER ses préfixes et ceux des sites raccordés

Chacun des préfixes des sites raccordés à RAP est annoncé deux fois à RENATER. RAP contrôle ses annonces BGP et utilise les valeurs de communauté¹ suivantes :

2200:590² : Chemin de secours

2200:610³ : Chemin nominal

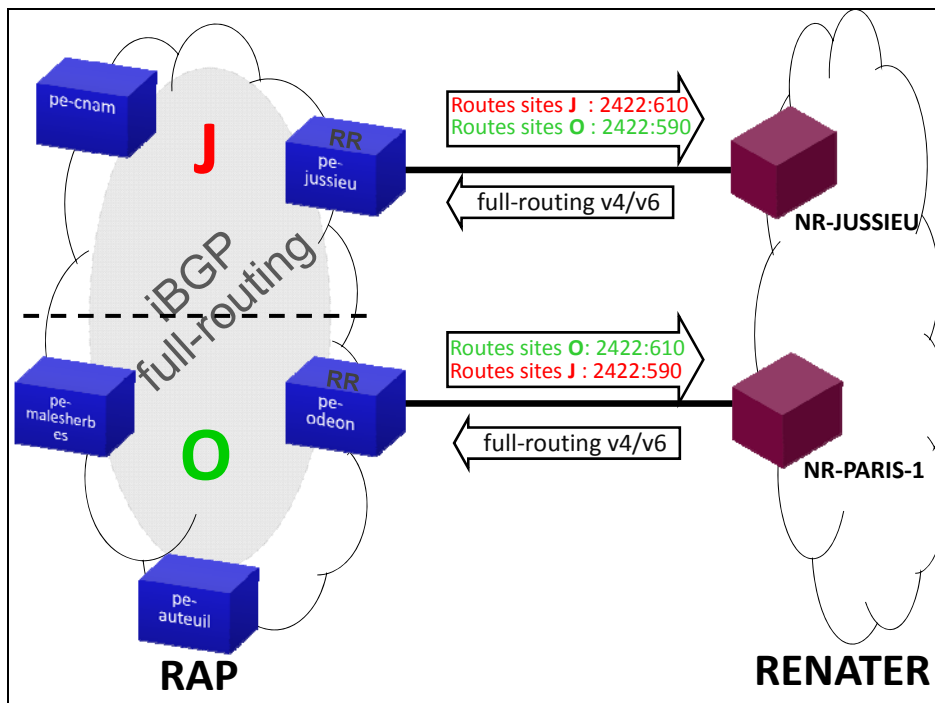


Figure 1 – Raccordement fiabilisé de RAP à RENATER

Pour un préfixe annoncé à RENATER depuis pe-jussieu avec la communauté 2200:610, par exemple, le trafic provenant de RENATER et à destination de ce préfixe transitera, dans l'état normal de fonctionnement par pe-jussieu, et, par pe-odeon en cas d'incident sur le chemin nominal.

¹ Communautés : attribut BGP qui permet de générer des groupes de destinations (préfixes IP) qui partagent des propriétés communes. Une communauté BGP est transitive, elle se propage d'AS en AS.

² Communauté spécifiée par RENATER ayant pour effet de rendre la route moins prioritaire.

³ Idem : <http://www.renater.fr/spip.php?article614>.

En ajoutant l'attribut BGP de communauté à nos annonces, on influence le chemin de retour du trafic de nos sites. Il existe donc une hiérarchie à 2 niveaux de priorités pour chaque préfixe IP, dont le schéma ci-dessus montre la répartition par zone.

Aussi, tous les PE de RAP sont en *full-mesh* iBGP, grâce aux pe-jussieu et pe-odeon qui remplissent la fonction de *route reflector* (RR) et qui redistribuent du *full-routing*. Ainsi, pour le trafic provenant des sites et à destination de RENATER, c'est par le paramétrage de l'iBGP et de l'IGP de RAP que le trafic emprunte le bon chemin de sortie et que l'on évite une asymétrie de trafic.

2 Raccordement fiabilisé d'un site sur RAP :

C'est le même principe qui est repris dans ce cas. On généralise au backbone les mécanismes de fiabilisation mis en œuvre en visant les 2 objectifs suivants :

- Le site contrôle le trafic sur ses accès à RAP
- Le routage avec RENATER s'adapte dynamiquement selon les réglages au niveau du site.

Le site choisi pour chacun de ses préfixes IP le PoP nominal et le PoP de secours en utilisant, ici aussi, l'attribut BGP *community*.

A noter, le format d'annonce d'une communauté doit suivre la recommandation du RFC 1997 :

<numéro d'AS de RAP sur 2 octets (2422) > : <valeur de la communauté sur 2 octets >

Une table des correspondances est configurée sur chacun des PE entre la communauté BGP envoyée par le site et une *local-preference* qui sera transitive dans l'iBGP de RAP. Le site agit ici sur le choix du chemin retour de son trafic, il doit par ailleurs agir sur son réseau interne pour le choix du chemin de son trafic sortant en vue de garder un trafic symétrique.

A chaque PoP ou PE est donc attribué une valeur de communauté de référence permettant de le rendre prioritaire :

Valeur de la communauté	Priorité du POP				
	Jussieu	Odéon	Auteuil	Malesherbes	CNAM
100	1	2	2	2	2
200	2	1	2	2	2
300	2	2	1	2	2
400	2	2	2	1	2
500	2	2	2	2	1

Tableau 1: Communautés de référence par PoP

Un site raccordé à 2 PoPs choisi pour chacun de ses préfixes IP par quel PoP il transite en priorité. Il annonce à RAP ses routes avec la valeur de communauté de référence du PoP choisi comme prioritaire et avec cette même valeur de communauté sur l'autre PoP choisi comme secours. Cette méthode est généralisée à tous les PE dont voici le schéma global résultant :

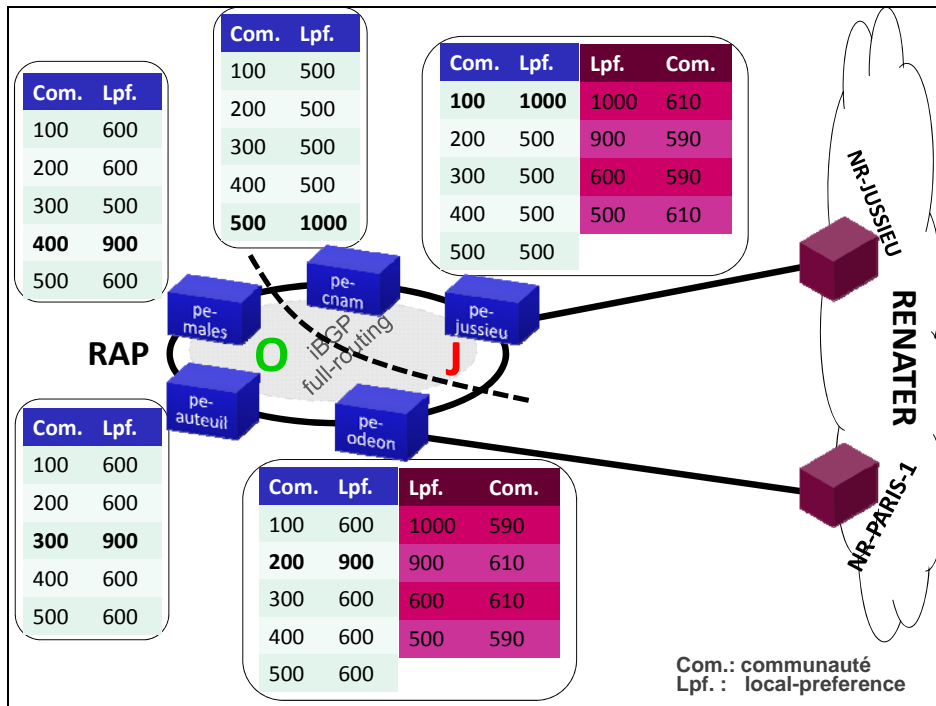


Figure 2 : Correspondances Communauté - Local-preference sur RAP

Réciproquement, une table des correspondances sur pe-odeon et pe-jussieu entre la *local-preference* obtenue et la communauté BGP envoyée à RENATER permet de gérer dynamiquement l'annonce des routes vers RENATER selon les réglages du site.

Les mécanismes qui permettent de « transformer » une communauté en *local-preference* et l'inverse sont des *policy* (en terme Juniper, *route-map* en terme cisco) :

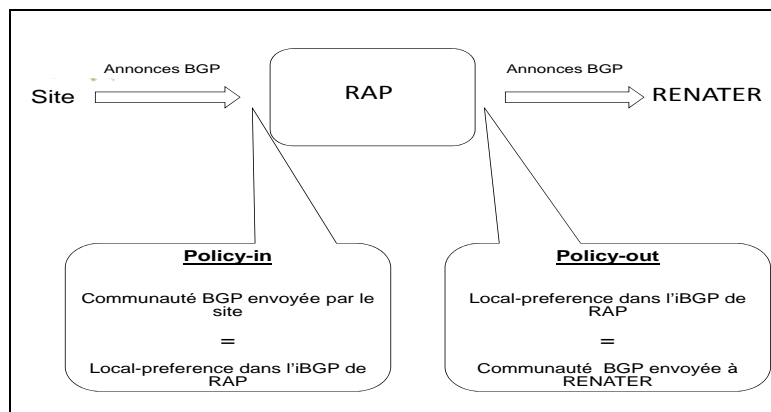


Figure 3 : Principe des correspondances *community - local-preference*

2.1 Raccordement d'un site à 2 PEs :

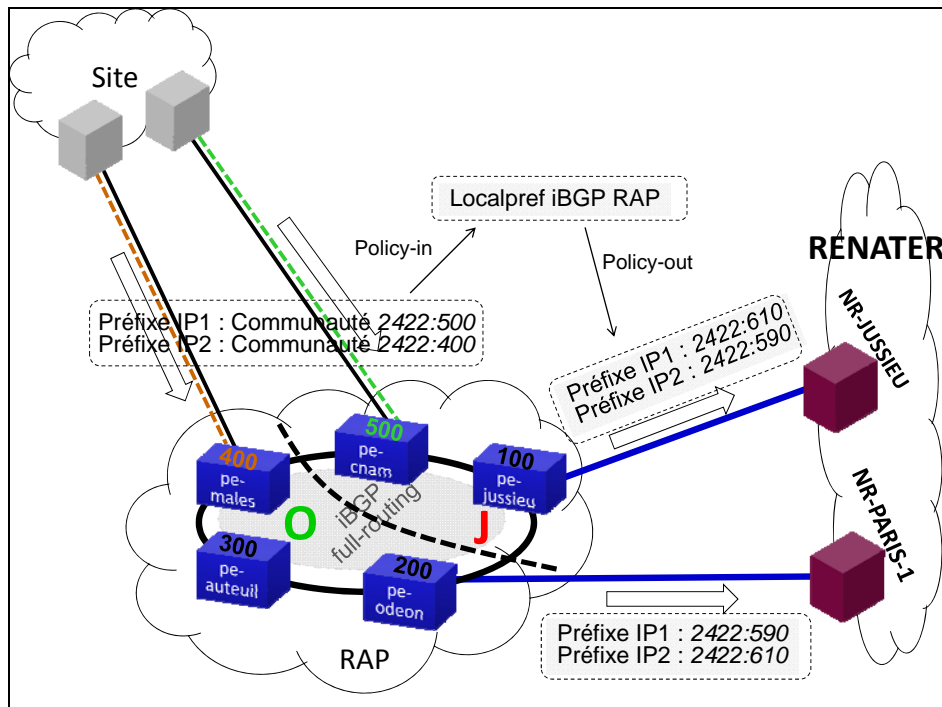


Figure 4 : Site raccordé à 2 PE

Le site envoie ses routes avec les valeurs de communautés de références.

Chaque PE met en œuvre une policy en entrée (policy-in ou route-map) sur les annonces BGP envoyées par le site, permettant de définir une *local-preference*, dont voici les caractéristiques :

- Une annonce est rejetée si elle comporte une valeur de communauté non définie sur RAP
- Seules les routes connues dans SIRAP⁴ pour le site sont acceptées.
- Seules les routes comportant le bon numéro d'AS sont acceptées.
- Allocation d'une *local-preference* pour chaque préfixe en fonction de la communauté envoyée.

Et, chaque PE met en œuvre une policy en sortie (policy-out ou route-map) afin de n'annoncer qu'une route par défaut au site.

Exemples de configurations mises en œuvre sur un PE :

- PE-JUSSIEU :
[*policy-options*]

⁴ SIRAP : Système d'Information du Réseau Académique Parisien.

```

policy-statement BGP-FILTER-SITE_A-IN {
  term ok {
    from {
      as-path AS-SITE_A;
      route-filter X.X.X.X/Y orlonger;
      ...
    }
    then next policy;
  }
  term default {
    then reject;
  }
}

```

```

policy-statement community-100-200 {
  term comm100 {
    from community RAP-community-100;
    then {
      local-preference 1000;
      accept;
    }
  }
  term comm200 {
    from community RAP-community-200;
    then {
      local-preference 500;
      accept;
    }
  }
  term default {
    then reject;
  }
}

```

```

policy-statement BGP-FILTER-SITE_A-OUT {
  term default-route {
    from {
      prefix-list default-route;
    }
    then accept;
  }
  term default {
    then reject;
  }
}

```

```

community RAP-community-100 members 2422:100;
community RAP-community-200 members 2422:200;
as-path AS-SITE_A "^AS_SITE_A.*";

```

```

[routing-options]
autonomous-system 2422;

[protocols]
bgp {
  group SITE_A {
    type external;
    local-address Y.Y.Y.Y;
    passive;
    import [ BGP-FILTER-SITE_A-IN community-100-200 ];
    family inet {
      any;
    }
  }
  export BGP-FILTER-SITE_A-OUT;
  peer-as AS_SITE_A;
  neighbor X.X.X.X;
}
}

```

2.2 Raccordement de 2 sites secourus mutuellement

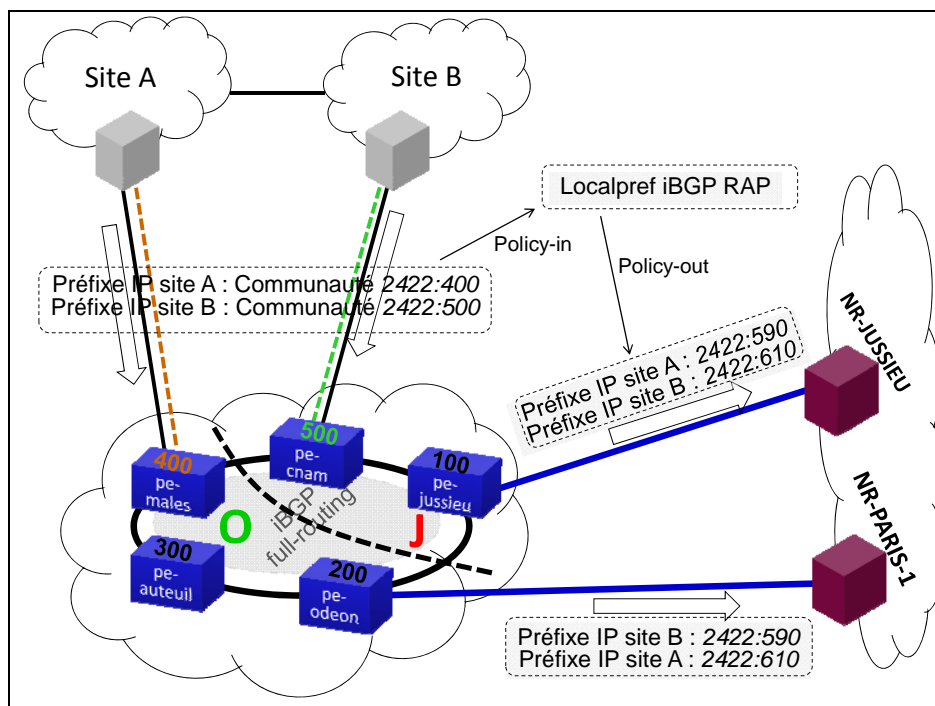


Figure 5 : Sites secourus mutuellement

Si deux sites raccordés à RAP sur 2 PoPs différents ont une liaison qui les raccorde entre eux, chacun d'eux peut devenir le backup de l'autre. Dans ce cas, il faut que la policy en entrée sur les annonces BGP envoyées par le site autorise les routes de chacun des deux sites, ainsi que l'AS des 2 sites (s'ils sont différents).

Le site peut aussi demander à ce que du trafic inter-site, passe non pas par leur liaison directe mais par RAP. Dans ce cas, il faut aussi autoriser, en plus de la route par défaut, l'annonce en BGP des routes du site A vers le site B et l'inverse. On modifie pour cela la policy en sortie des annonces BGP vers le site. En interne, le site doit donc mettre en œuvre les bons mécanismes de priorisation des routes (local-preference, coût IGP, etc...) afin de bien organiser sa répartition de trafic inter-site.

2.3 Envoi de la communauté à RENATER

La *local-preference* transitant dans l'iBGP de RAP dont la valeur est déterminée par la communauté envoyée par le site est analysée par pe-odeon et pe-jussieu afin d'en déduire la communauté à envoyer vers RENATER.

- Les communautés admises sur RENATER sont les suivantes :

2200:610 (Communauté libre d'utilisation) : Cette communauté permet de déclarer un chemin primaire dans le cas d'un réseau avec plusieurs attachements sur RENATER.

2200:590 (Communauté libre d'utilisation) : Cette communauté permet de déclarer un chemin secondaire dans le cas d'un réseau avec plusieurs attachements sur RENATER.

2200:290 (Sous réserve de validation, faire la demande au NOC-RENATER) : Un préfixe reçu avec cette communauté ne sera pris en compte que s'il n'existe pas de route alternative. Cette communauté permet donc à un établissement multi-homé de prioriser un accès via un autre opérateur que RENATER. Cette communauté n'est disponible que pour les sites ayant des adresses PI.

2200:3000 (Communauté libre d'utilisation) : Un préfixe reçu avec cette communauté ne sera pas annoncé sur GEANT, les transits et les points d'échange. Cette communauté n'est disponible que pour les sites ayant des adresses PI. Pour les préfixes dérivés des blocs RENATER (Adresses PA), il n'est pas possible d'utiliser cette fonctionnalité puisque les préfixes agrégés sont annoncés.

On les déclare dans le contexte [policy-options] :

```
community RENATER-Community-Backup members 2200:590;  
community RENATER-Community-Main members 2200:610;  
community RENATER-Community-Passive members 2200:290;
```

Une table des correspondances local-preference/communauté est configurée à l'aide d'une policy :

- o PE-JUSSIEU :
[policy-options]

```

policy-statement bgp-renater-out-new {
    term localpref1000-to-community {
        from local-preference 1000;
        then {
            community set RENATER-Community-Main;
            next term;
        }
    }
    term localpref900-to-community {
        from local-preference 900;
        then {
            community set RENATER-Community-Backup;
            next term;
        }
    }
    term localpref600-to-community {
        from local-preference 600;
        then {
            community set RENATER-Community-Backup;
            next term;
        }
    }
    term localpref500-to-community {
        from local-preference 500;
        then {
            community set RENATER-Community-Main;
            next term;
        }
    }
}

```

2.4 Accès réseau d'ultime recours :

Un site peut recourir à l'établissement d'un raccordement dit « d'ultime recours » en vue d'atteindre un degré très élevé de disponibilité de son accès réseau à RAP. De manière générale, le site met en place cet accès pour ses services critiques, cependant le choix des préfixes IP et des débits à supporter sur ce lien est à son entière appréciation.

Ce raccordement peut être un accès direct sur RAP avec un accès bas-débit de 2 à 20 Mbit/s (cela concerne les sites haut-débit) ou bien un accès par VPN sous la forme d'un tunnel IPSEC (cela concerne les sites haut-débit comme bas-débit) terminé dans un équipement de cœur de RAP. Dans ce deuxième cas, le site est seul responsable de la mise en place et du fonctionnement du raccordement à Internet utilisé pour son accès d'ultime recours, et devra par ailleurs disposer de l'équipement terminateur du tunnel sur son site.

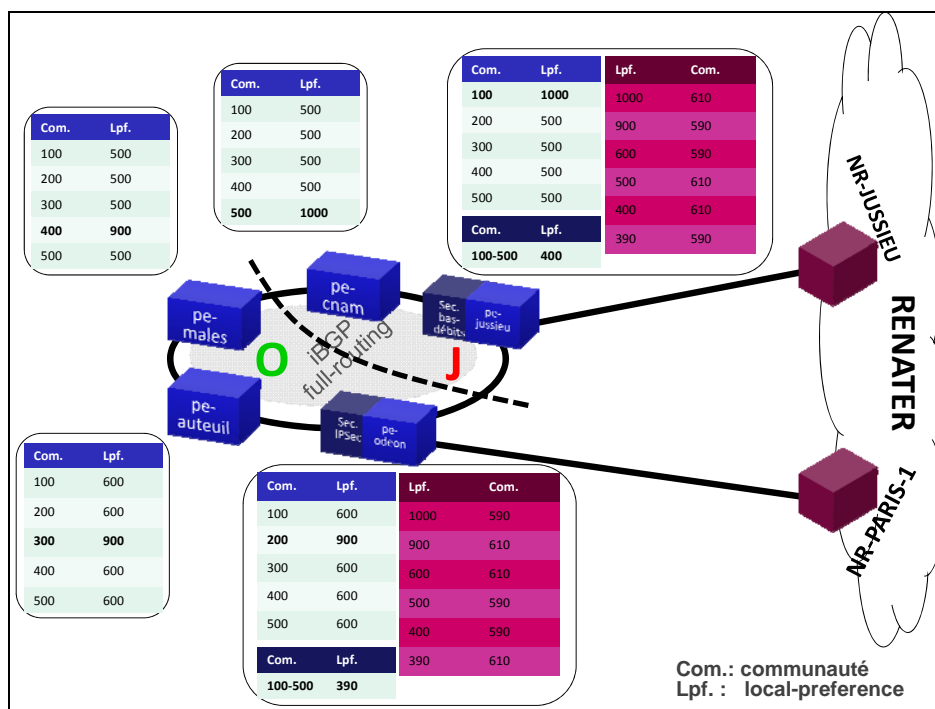


Figure 6 : Ajout de local-preference sur RAP pour les accès d'ultime recours

Chaque site RAP ne peut disposer que d'un accès d'ultime recours au maximum et de part la configuration BGP fixée sur RAP, sa priorité est strictement moins élevée que celles de l'accès nominal (site bas-débit ou site en raccordement haut-débit simple) et de l'accès secondaire (site haut-débit en raccordement fiabilisé). En temps normal, aucun trafic de production ne passe par cet accès d'ultime recours. Pour bénéficier pleinement du dispositif, un site haut-débit aura au préalable migré vers un raccordement fiabilisé avant d'avoir recours à cette solution.

Pour rendre homogène les configurations BGP, toutes les valeurs de communauté [100-200-300-400-500] sont acceptées, cependant, de part la nature strictement moins prioritaire des routes reçues, RAP fixe en statique la *local-preference* correspondante sur RAP.

Exemple pour un site haut-débit en raccordement fiabilisé avec un ultime recours par un accès bas-débit :

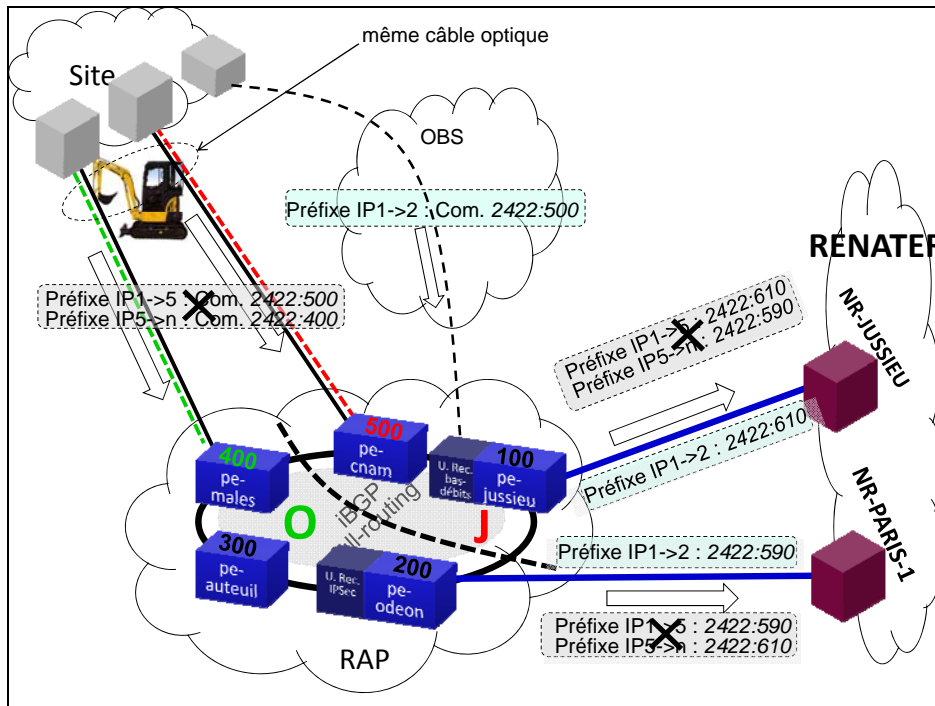


Figure 7 : Un accès réseau d'ultime recours par le service bas-débit de RAP

2.5 Accès nominal par un autre FAI :

Un site raccordé à un FAI peut choisir ce dernier comme chemin nominal. Pour cela, la communauté 2200:290 sera transmise par RAP à RENATER. Dans le cas où le FAI tombe et que RAP prend le relais, et puisque les routes du site sont transmises par les 2 accès à RENATER, il faut rendre l'un des 2 accès moins prioritaire. RENATER n'a pas prévu d'autres communautés pour ce faire, la configuration se basera donc sur du BGP prepending : on « gonfle » l'as-path.

Prenons l'exemple d'un site raccordé sur la partition Jussieu (faire l'inverse pour un site de la partition Odeon) :

- On transmet depuis le pe-jussieu la communauté 2200:290, pour cela on positionne les configurations suivantes dans le term qui accepte les routes pour l'établissement dans le *policy-statement bgp-renater-out-new* :

```
[policy-options]
policy-statement bgp-renater-out-new {
....
  term <nom de l'établissement> {
    from {
      route-filter A.B.C.D/E orlonger;
    }
    then {
```

```

        community set RENATER-Community-Passive;
        next term;
    }
}
...
}

```

- On transmet depuis le pe-odeon la communauté 2200:290 et on gonfle l'as-path pour ses routes avec la méthode de BGP prepending, pour cela on positionne les configurations suivantes :

```

[policy-options]
policy-statement bgp-renater-out-new {
....
    term <nom de l'établissement> {
        from {
            route-filter A.B.C.D/E orlonger;
        }
        then {
            community set RENATER-Community-Passive;
            then as-path-prepend "2422 2422";
            next term;
        }
    }
}
...
}

```

2.6 Exemples de configurations pour le routeur de site

Se référer au document « Mise en place de raccordement fiabilisé pour les sites sur RAP » en ligne sur le portail de RAP.