TD : Routage inter-domaine avec BGP

Principes de fonctionnement du protocole

- Le protocole de routage BGP (*Border Gateway Protocol*) est un protocole permettant de communiquer des informations de routages entre différents systèmes autonomes (AS pour *Autonomous System*). Il est dit externe aux systèmes autonomes. Dans un protocole de routage externe ce sont les routeurs aux frontières des AS, appelés routeurs de bordure, qui s'échangent les informations de routage. Le protocole de routage externe va permettre l'échange des adresses contenues dans les AS entre ces routeurs de bordure. Il va aussi propager des routes apprises depuis un autre AS. Le passage des informations de routage se fera de routeur de bordure en routeur de bordure et elles seront éventuellement propagées dans les routeurs internes aux AS par une redistribution dans les protocoles de routages internes.
- Le but d'un tel protocole est de pouvoir propager (comme les protocoles de routages internes) des routes connues vers d'autres AS en pouvant appliquer des restrictions décidées par l'administrateur de chaque AS. Il va falloir faire la distinction entre les routes apprises de façon internes et celles apprises depuis l'extérieur des AS dans ce type de protocole.
- Le protocole BGP repose sur TCP (port 179). Les échanges se font toujours entre 2 routeurs. La version actuelle est la v4 décrite dans la nouvelle RFC 4271 qui est récupérable ici http://www.ietf.org/rfc/rfc4271.
- On peut distinguer 2 types de dialogue BGP :
    - Entre deux routeurs de bordure de deux AS différents, dénommé eBGP (external BGP).
    - Entre les routeurs d'un même AS dénommé iBGP (internal BGP).
- Pour qu'un dialogue BGP s'établisse entre deux routeurs, on les déclarera « voisins » (au sens BGP). Deux voisins d'AS différents sont forcément sur le même réseau local. Deux voisins du même AS peuvent être sur des réseaux différents. C'est le protocole de routage interne qui maintient leur connectivité. On peut filtrer à volonté les routes à diffuser à l'extérieur. Les routeurs BGP vont prendre leur décision de routage au vue des attributs des adresses qu'ils auront pu recevoir de divers AS et des restrictions/préférences locales. Ces attributs vont spécifier pour une adresse destination donnée @ :
  • le prochain routeur à qui envoyer *(next hop)* pour atteindre @
  • l'origine de l'apprentissage de cet @ (interne, externe ou statique)
  • des préférences locales de poids affectés au entrées/sorties d'un AS
  • des métriques associées aux adresses, ...
- Pour qu'une route vers un réseau donné soit propagée, il faut qu'elle soit connue de BGP (c'est à dire présente dans la table BGP), mais aussi que le réseau en question apparaisse dans la table de routage IP. Cette vérification évite que le routeur de bordure reçoive du trafic dont il ne saurait pas quoi faire.

Mise en place

1. Ce TD utilise **`GNS3`** et **`dynamips`** qui sont installés sur les machines. Configurez **`GNS3`** dans les préférences pour qu'il trouve **`dynamips`** dans **`/usr/bin`**. Testez son lancement.
2. Récupérez une image pour le **Cisco 7200** et configurez **`GNS3`** pour qu'il la trouve localement sur votre machine.
3. Dans chaque routeur configurez le **`slot 0`** avec un module **`C7200-IO-FE2`** et le **`slot 1`** avec un module **`PA-8T`**.
4. Les commandes **Cisco** données ci-dessous sont **à compléter correctement**, pour BGP voir ici http://www.cisco.com/en/US/docs/ios/12_0/np1/configuration/guide/1cbgp.html.

Réseau avec chemins uniques

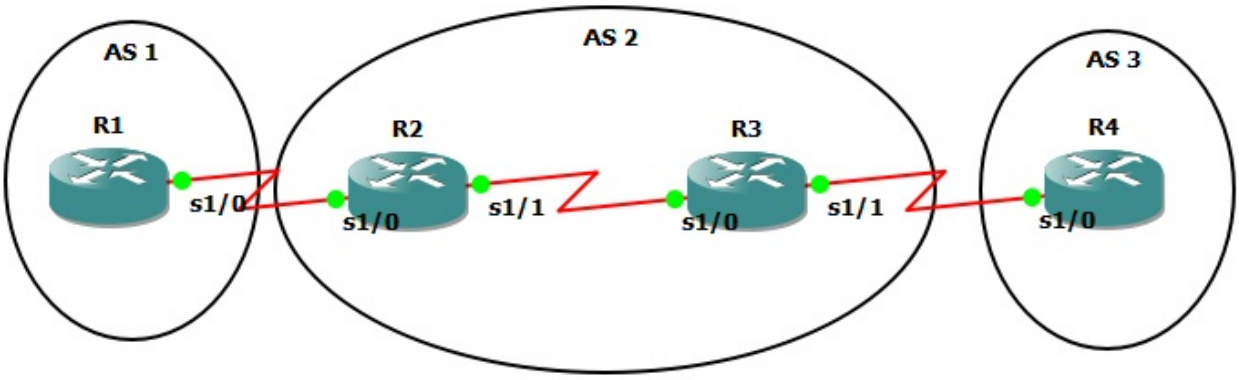

Figure 1. Topologie réseau simple.

5. Créez la topologie ci-dessus dans **GNS3** en connectant les routeurs par des liens série. Démarrez et connectez-vous à ces routeurs.
6. Configurez bien la valeur *Idle PC* des routeurs sinon vos processeurs seront sollicités en permanence à 100%.
7. Créez des interfaces de *loopback* sur R1 et R4 pour simuler des LANs avec **`(config) interface loopback`**.
8. Configurez les adresses IP de toutes les interfaces avec **`(config-if) ip address`** en utilisant le plan d'adressage suivant :
    a. 11.0.0.0/8 sur **`lo`** de R1 (préfixe réseau 11 noté P11)
    b. 12.0.0.0/8 entre R1 et R2
    c. 23.0.0.0/8 entre R2 et R3
    d. 34.0.0.0/8 entre R3 et R4
    e. 44.0.0.0/8 sur **`lo`** de R4
9. Sauvegarder la configuration de chaque routeur avec **`copy running-config startup-config`**. Ensuite dans **GNS3** sauvegardez votre projet complet. Pour cela, allez dans ***`file->save project as…`*** puis tapez un nom de fichier et cochez la case ***`save IOS startup files`*** afin de conserver toute votre topologie avec les configurations des routeurs.
10. Vérifiez les routes directes dans les réseaux internes de chaque AS avec **`show ip`**.
11. Démarrez **`wireshark`** sur chaque lien série. Choisissez l'encapsulation HDLC. Vous pourrez observer les paquets échangés. Quels types de paquets sont échangés ?
12. Configurez BGP sur chaque routeur avec **`(config) router bgp.`**
13. Déclarez R1 et R2 comme voisins BGP avec **`(config-router) neighbor.`** Faites de même avec R3 et R4.
14. Exportez les préfixes des LANs de R1 et R4 avec **`(config-router) network`**.
15. Regardez les tables de routage internes et externes avec **`show ip`** et à l'aide des traces dans **`wireshark`** expliquez les étapes de leur remplissage.
16. Testez la connectivité du réseau pour toutes les adresses avec **`ping`** et si besoin corrigez les configurations.

Réseau avec chemins multiples

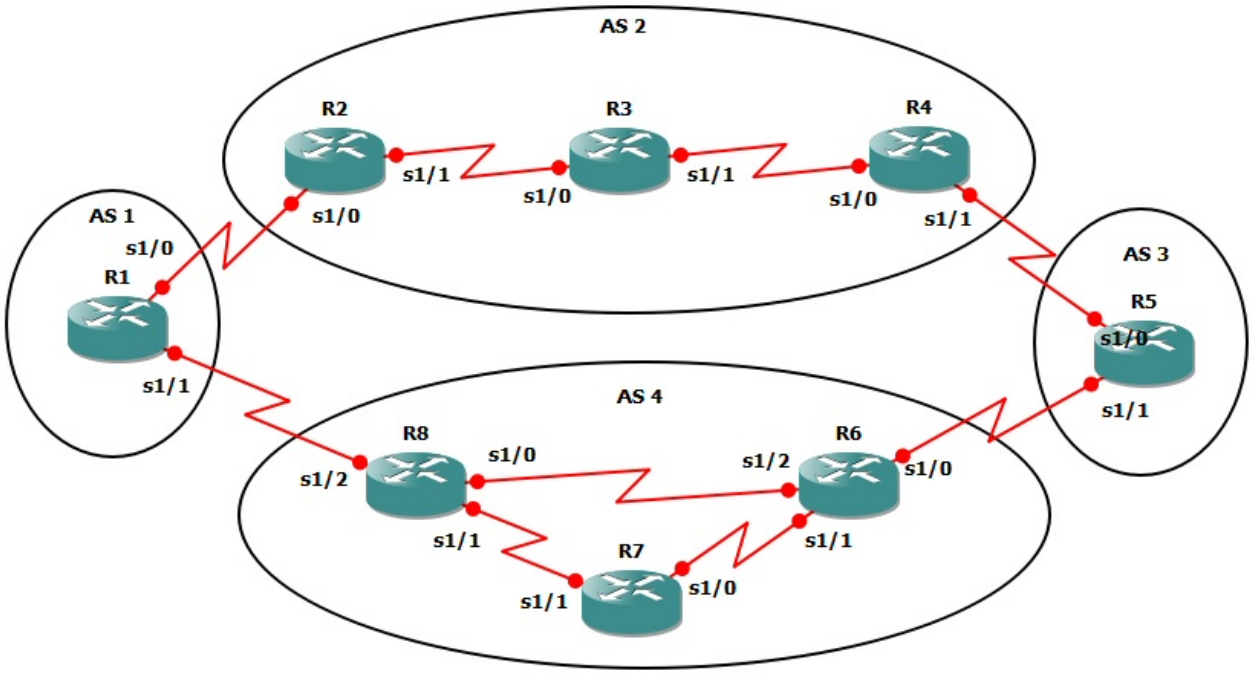

Figure 2. Topologie avec chemins multiples.

17. Créez un **nouveau** réseau pour qu'il soit identique à la figure 2. Créez des LANs avec des interfaces de *loopback* sur R1, R3, R5 et R7. Utilisez le plan d'adressage suivant :
    a. 11.0.0.0/8 sur **`lo`** de R1 (préfixe réseau 11 noté P11)
    b. 12.0.0.0/8 entre R1 et R2
    c. 18.0.0.0/8 entre R1 et R8
    d. 23.0.0.0/8 entre R2 et R3
    e. 33.0.0.0/8 sur **`lo`** de R3
    f. 34.0.0.0/8 entre R3 et R4
    g. 45.0.0.0/8 entre R4 et R5
    h. 55.0.0.0/8 sur **`lo`** de R5
    i. 56.0.0.0/8 entre R5 et R6
    j. 67.0.0.0/8 entre R6 et R7
    k. 68.0.0.0/8 entre R6 et R8
    l. 77.0.0.0/8 sur **`lo`** de R7
    m. 78.0.0.0/8 entre R7 et R8
18. Sauvegarder la configuration de chaque routeur avec **`copy running-config startup-config`**. Ensuite dans **`GNS3`** sauvegardez votre projet complet. Pour cela, allez dans ***`file->save project as…`*** puis tapez un nom de fichier et cochez la case ***`save IOS startup files`*** afin de conserver toute votre topologie avec les configurations des routeurs.
19. Démarrez **`wireshark`** sur les liens R1-R2, R1-R8, R8-R6 et R8-R7. Vous pourrez observer les paquets échangés.
20. Configurez un routage intra domaine dynamique en utilisant RIP dans l'AS 2 avec **`(config) router rip`** et OSPF dans l'AS 4 avec **`(config) router ospf`**, puis définissez chaque interface supportant OSPF avec **`(router-ospf) network`** ***`IP_@_interf`*** **`0.0.0.0 area 0.`**
21. Configurez BGP sur tous les routeurs de bordure avec **`(config) router bgp`**. Configurez R8 et R6 comme voisins intra-AS.

22. Redistribuez les routes externes à l'intérieur de l'AS 4 avec **(router-ospf) redistribute bgp**. Observez le résultat sur R7 avec **show ip**.
23. Testez la connectivité pour toutes les adresses. Si elle est incomplète, expliquez pourquoi et corrigez le problème.
24. Regardez les tables de routage internes et externes avec **show ip** et définissez toutes les routes existant entre les LANs. Quel est le chemin traversé pour atteindre P55 à partir de P11 ? Existe-t-il des routes asymétriques ?
25. Désactivez le lien R5-R6. Observez les échanges de messages avec **wireshark** et l'évolution des tables de routage. Pingez P55 à partir de R1. Combien de temps faut-il pour récupérer la connectivité ? Quel est le nouveau chemin pour atteindre P55 à partir de P11 ?
26. Désactivez le lien R6-R8. Observez les échanges de messages avec **wireshark** et l'évolution des tables de routage. Pingez P55 à partir de R1. Combien de temps faut-il pour récupérer la connectivité ? Quel est le nouveau chemin pour atteindre P55 à partir de P11 ?
27. Désactivez le préfixe P55. Observez les échanges de messages avec **wireshark** et l'évolution des tables de routage.

Filtrage par route

- On peut affecter à chaque voisin une liste de permission d'apprentissage ou divulgation de préfixes. On peut affecter la liste d'accès de numéro ***num*** au voisin d'adresse ***IP@*** et cela soit en entrée (apprentissage depuis ce voisin), soit en sortie (divulgation d'information vers ce voisin) avec **neighbor *IP@* distribute-list *num* in/out**.
- Les listes d'accès sont ensuite définies par **access-list *num* permit/deny *IP@ non_significants_bits_mask.*** Par exemple **access-list 1 deny 160.10.0.0 0.0.255.255.** On devra lui ajouter une autorisation de toutes les autres adresses par la commande **access-list 1 permit 0.0.0.0 255.255.255.255.**
- Pour accélérer la prise en compte de ces suppressions de route, on peut nettoyer les tables avec **clear ip bgp ***. On peut visualiser les listes d'accès avec **show access-list *num*.**

Filtrage par AS ou *AS path*

- On peut filtrer la propagation par BGP de l'ensemble des routes apprises depuis un AS ou une suite d'AS avec **ip as-path access-list num deny/permit regu-expr** où l'expression régulière permet de spécifier un AS ou un chemin d'AS. Elle se compose de numéros d'AS et de caractères spéciaux aux significations suivantes :
  - **^** : début de chemin
  - **$** : fin de chemin
  - **.** : n'importe quel caractère
  - ***** : un nombre quelconque de fois
- On peut vérifier que l'expression est bonne avec **show ip bgp regexp regu-expr**.
- Exemples :
  - **^200$** (toutes les adresses venant directement de l'AS 200)
  - **.*** (spécifie tout AS)
  - **^200 300$** (spécifie le chemin AS300 (source) puis AS200)
  - **^200.*** (spécifie toute route passant en dernier lieu par l'AS 200 mais dont le chemin antérieur peut être quelconque)

Gestion de routes multiples

On peut les pondérer de diverses manières afin de déterminer laquelle sera mise dans la table de routage par BGP. Cette pondération peut être locale à un serveur, locale à un AS ou diffusée d'un AS à l'autre. Pour les diffusions de ces pondérations entre routeur BGP, des attributs particuliers

sont associées aux adresses transmises dans les paquets BGP. Le choix d'une route se fait suivant l'ordre des critères suivant :

1. Poids
2. Préférence locale (à un AS)
3. Longueur du chemin d'AS
4. Origine : protocole interne préféré à un protocole externe
5. Métrique de BGP
6. Métrique du protocole interne vers le *next hop*

1. Informations locales à un routeur
On peut associer un poids à un voisin avec **neighbor *IP@* weight *wgt***
Dans le cas de routes multiples le passage par le voisin de poids le plus élevé sera utilisé. Attention ce poids est une information qui n'est pas transmise de routeur en routeur. Elle sert à sélectionner les routes au niveau d'un routeur donné. à l'opposé, les méthodes de sélection suivantes (préférence et métrique) sont transportées par BGP.

2. Préférences locales à un AS
Cette information est stockée dans les messages dans l'attribut LOCAL PREF des paquets BGP, son code est 5. Par défaut cet attribut vaut 100. Un routeur diffusant l'attribut LOCAL PREF le plus grand pour une destination sera choisi pour l'atteindre. Il permet ainsi de privilégier une entrée/sortie d'un AS par rapport à une autre.
La commande **bgp default local-preference *value*** peut être utilisée pour changer cet attribut pour toutes les adresses diffusées depuis un routeur. D'autres commandes permettent de spécifier cet attribut seulement pour certaines routes en utilisant des listes d'accès.

3. Informations entre AS
Un AS va pouvoir influer sur les choix de ses voisins par la pondération des routes qu'il leur diffuse. C'est l'attribut METRIC qui permet de diffuser cette pondération entre AS. Il est appelé MED (Multi Exit Discriminator) dans BGP version 4. Cet attribut n'est pas transitif, il n'est pas propagé : l'indication n'est donc valable que pour les routeurs qui sont immédiatement connectés à un AS. Il permet à un AS d'associer une métrique à une destination qu'il diffuse à un autre AS. Un routeur qui reçoit différentes possibilités pour accéder une destination prendra celle de métrique la plus faible. Un AS peut donc décider de la pondération qu'il associe à sa traversée par exemple. Les commandes suivantes permettent de spécifier cette pondération (pour toutes les adresses émises) :

```
neighbor IP@ route-map my-route-map out
route-map my-route-map permit 10
set metric m1
```

Réseau avec chemins multiples et routage politique

28. Sauvegarder la configuration de chaque routeur avec **copy running-config startup-config**. Ensuite dans **GNS3** sauvegardez votre projet complet. Pour cela, allez dans ***file->save project as…*** puis tapez un nom de fichier et cochez la case ***save IOS startup files*** afin de conserver toute votre topologie avec les configurations des routeurs.
29. On considère le réseau de la figure 2 avec tous les liens opérationnels. Implémentez un filtrage par route sur les préfixes P33 et P77 afin qu'ils ne soient pas connus des AS 1 et 3. Vérifiez que ce filtrage a bien eu lieu par des *pings* et la visualisation des tables de routage et tables BGP.
30. Supprimez les listes d'accès précédentes par **no neighbor *IP@* distribute-list**. Activer la liste de filtrage par AS avec **neighbor *IP@* filter-list *num* in/out.** Définissez un filtrage par AS dans l'AS 3 afin que R5 ne prenne pas en compte les adresses venant de l'AS 2 mais qu'il prenne en compte celle venant de l'AS 4. Est-ce que P11 est visible dans R5 ?

31. Supprimez le filtrage précédent. L'AS 3 souhaite que le trafic avec l'AS 1 passe (dans les deux sens) de préférence par l'AS 2 mais en gardant la connectivité par l'AS 4 en cas de coupure avec l'AS 2. Configurez le routage politique de l'AS 3 en conséquence et vérifier son bon fonctionnement dans tous les cas.
32. Modifiez maintenant la politique de l'AS 3 afin que le trafic sortant à destination du préfixe P11 passe via l'AS 2 et que le trafic entrant en provenance du préfixe P11 passe via l'AS 4. Comment R3 et R7 ont-ils appris ces contraintes BGP ? Vérifiez le bon fonctionnement de la politique de routage dans tous les cas et si besoin corrigez la configuration.
33. Configurez 2 préfixes supplémentaires P111 et P112 dans l'AS 1 et modifiez la politique de l'AS 3 afin que le préfixe P111 soit accessible de préférence via l'AS 2, le préfixe P112 de préférence via l'AS 4, et le préfixe P11 uniquement par l'AS 4. Vérifiez le bon fonctionnement de la politique de routage dans tous les cas et si besoin corrigez la configuration.

Des observateurs de tables BGP

34. Allez sur le site http://www.route-views.org/. De quoi s'agit-il ?
35. Téléchargez une archive récente de la table d'un des routeurs observateurs et consultez la. Prenez de préférence une table ASCII provenant d'un *dump* de **`sh ip bgp`**. Si c'est un *dump* de *Zebra*, il faudra convertir le fichier binaire MRT v1/v2 en texte avec un parseur adéquat. Combien de préfixes y sont stockés ? Quelle est la taille maximale des *AS paths* ? Quels sont les plus gros AS de l'Internet ?
36. Un autre site intéressant http://showipbgp.com/. De quoi s'agit-il ?

Travail à rendre

A la fin des séances de ce TD, vous rendrez un rapport de TD par binôme, au format PDF, que vous enverrez par e-mail à votre chargé de TD. Ce rapport contiendra les réponses aux questions posées dans ce sujet (en souligné) en y incluant tous les justificatifs nécessaires :

- Extraits pertinents des **`running-configuration`** des routeurs ou listings des commandes IOS utilisées pour résoudre les questions
- Extraits des tables de routage intra- (**`sh ip ro`**) et inter- (**`sh ip bgp`**) domaines pertinentes
- Extraits pertinents des captures de trames prises par **`wireshark`**
- Sorties des commandes **`ping`** et **`traceroute`**

Références

Ce TD est dérivé des sujets de P. Sicard, M. Heusse et J.J. Pansiot.