

# Techniques d'Optimisation BGP



**Atelier SI-F  
AfNOG 2018, Dakar**

# Techniques d'Optimisation BGP (BGP Scaling)

---

- Les spécifications et techniques de déploiement originelles de BGP étaient appropriées pour l'Internet du début des années 1990s
  - Mais Ne « scale » pas aux réseaux d'aujourd'hui
- Les problèmes rencontrés lorsque l'Internet s'est agrandi incluent:
  - Optimisation du maillage iBGP au-delà de quelques pairs?
  - Implémentation de nouvelles politiques sans causer d'oscillations (flap) et d'incohérences dans les mises à jour de routes (churn)?
  - Maintenir le réseau stable, évolutif et simple?

# Techniques d'optimisation BGP (BGP Scaling)

---

- Les meilleures techniques actuelles d'optimisation
  - Route Refresh
  - Peer-groups
  - Route reflector (et confédérations)
- Techniques d'optimisation obsolètes
  - Soft Reconfiguration
  - Route Flap Damping

# Reconfiguration dynamique



Changements non-destructifs  
de politiques

# Route Refresh

---

- Changement de politiques:
  - Réinitialisation dure avec les pairs BGP requise après chaque changement de politique car le routeur ne stocke pas les préfixes qui sont rejetés par les politiques
- Réinitialisation abrupte avec le peer BGP:
  - Casse la session BGP
  - Consomme du CPU
  - Perturbe considérablement la connectivité pour tous les réseaux
- Solution:
  - Route Refresh

# Capabilité de Route Refresh

---

- ❑ Facilite les changements de politiques - non destructifs
- ❑ Pas de configuration spécifique requise
  - Négocié automatiquement à l'établissement de la session
- ❑ Pas de mémoire additionnelle requise
- ❑ Requièrè que les routeurs pairs supportent la capabilité "route refresh" - RFC2918
- ❑ Comment dire au paire de renvoyer toutes les annonces BGP

```
clear ip bgp x.x.x.x [soft] in
```
- ❑ Comment Renvoyer toutes les annonces BGP au paire

```
clear ip bgp x.x.x.x [soft] out
```

# Reconfiguration dynamique

---

- Utiliser la capacité Route Refresh
  - Supportée par presque tous les routeurs
  - Se renseigner avec la commande “show ip bgp neighbor”
  - Non-perturbateur, “Bon pour l’Internet”
- Réinitialisation dure à utiliser seulement en dernier recours

**Considérer l’effet comme étant équivalent à un reboot du routeur**

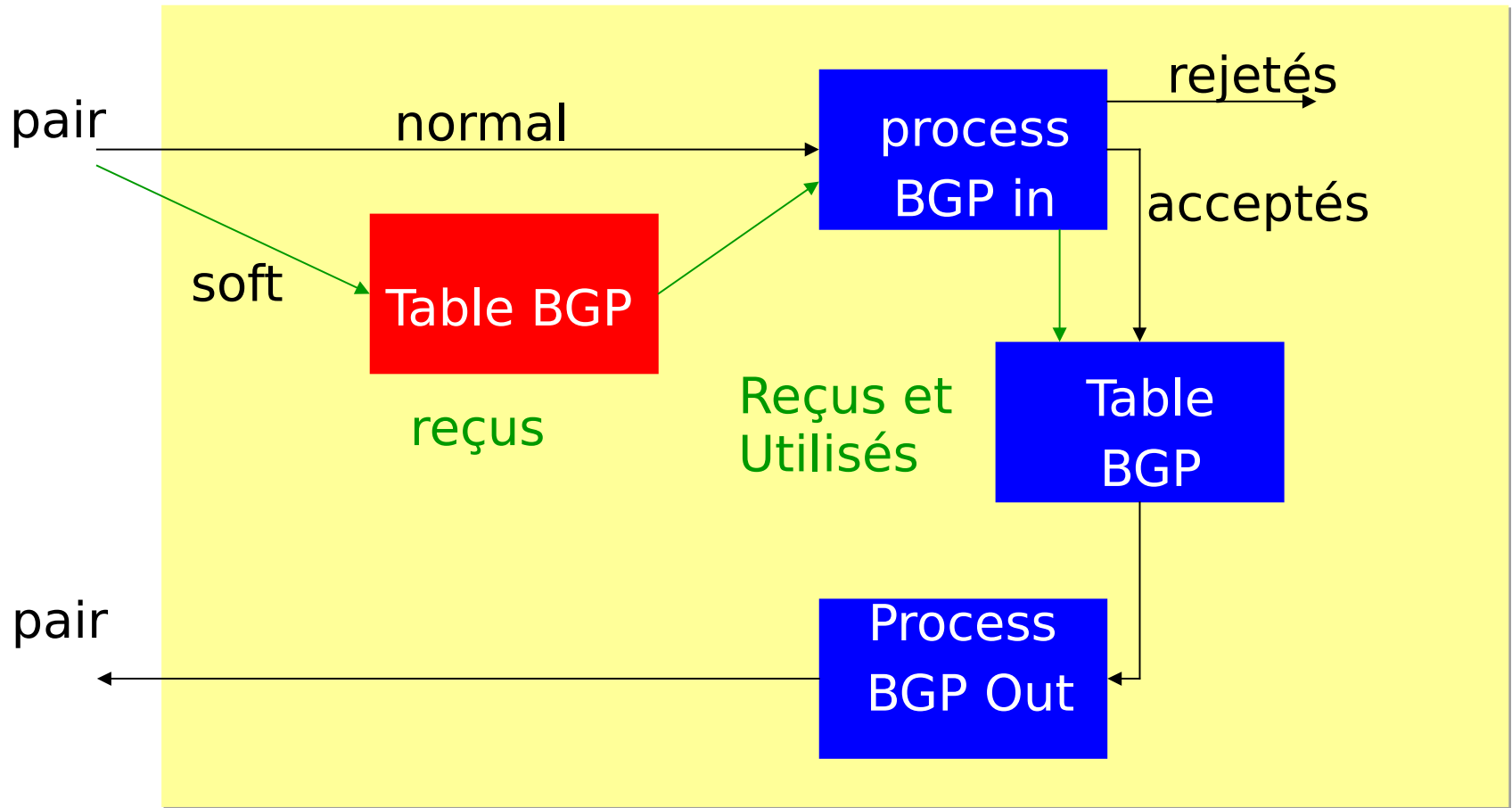
# Soft Reconfiguration de Cisco

---

- **Obsolète**— mais:
- Le routeur normalement stocke les préfixes qui ont été reçus d'un pair après application des politiques
  - Activer soft-reconfiguration signifie que le routeur stocke aussi les préfixes/attributs reçus avant d'appliquer les politiques
  - Utilise plus de mémoire pour garder les préfixes dont les attributs ont été changés ou les préfixes qui n'ont pas été acceptés
- Seulement utile lorsque l'opérateur veut savoir quels sont les préfixes qui ont été envoyés à un routeur avant que le routeur n'ait appliqué ses politiques d'entrées



# Soft Reconfiguration de Cisco



# Soft Reconfiguration

---

```
router bgp 100
  neighbor 1.1.1.1 remote-as 101
  neighbor 1.1.1.1 route-map infiltrer in
  neighbor 1.1.1.1 soft-reconfiguration inbound
! On n'a pas besoin de configurer pour
  l'Outbound!
```

- Ensuite lors de changement de politique, nous lançons la commande

```
clear ip bgp 1.1.1.1 soft [in | out]
```

- Note:

- Lorsque “soft reconfiguration” est activé, nous n’avons pas accès à la capacité route refresh
- `clear ip bgp 1.1.1.1 [in | out]` fait aussi un soft refresh

# Peer Groups



# Peer Groups

---

- Problème – comment rendre iBGP scalable
  - Le maillage complet iBGP est lente à construire
  - Les voisins iBGP recoivent le même update
  - CPU du routeur gaspillé alors que les calculs sont identiques
- Solution – peer-groups
  - Groupe les pairs avec les mêmes politiques de sortie
  - Les mises à jours (updates) sont générées une seule fois par groupe

# Peer Groups – Advantages

---

- ❑ Facilite la configuration
- ❑ Moins d'erreurs de configuration
- ❑ Configuration plus lisible
- ❑ Diminue la charge CPU du routeur
- ❑ Maillage iBGP se construit plus rapidement
- ❑ Membres peuvent avoir des politiques d'entrées différentes
- ❑ Peut être utilisé pour les voisins eBGP également!

# Configurer un Peer Group

---

```
router bgp 100
  neighbor ibgp-peer peer-group
  neighbor ibgp-peer remote-as 100
  neighbor ibgp-peer update-source loopback 0
  neighbor ibgp-peer send-community
  neighbor ibgp-peer route-map outfilter out
  neighbor 1.1.1.1 peer-group ibgp-peer
  neighbor 2.2.2.2 peer-group ibgp-peer
  neighbor 2.2.2.2 route-map infilter in
  neighbor 3.3.3.3 peer-group ibgp-peer
```

! Remarquez que 2.2.2.2 a des politiques d'entrées différentes du peer-group !

# Configurer un Peer Group

---

```
router bgp 100
  neighbor external-peer peer-group
  neighbor external-peer send-community
  neighbor external-peer route-map set-metric out
  neighbor 160.89.1.2 remote-as 200
  neighbor 160.89.1.2 peer-group external-peer
  neighbor 160.89.1.4 remote-as 300
  neighbor 160.89.1.4 peer-group external-peer
  neighbor 160.89.1.6 remote-as 400
  neighbor 160.89.1.6 peer-group external-peer
  neighbor 160.89.1.6 filter-list infilter in
```

# Peer Groups

---

- Toujours configurer des peer-groups pour iBGP
  - Même s'il n'y a que quelques pairs iBGP
  - Plus facile pour faire évoluer le réseau dans le futur
- Considérer l'utilisation des peer-groups pour eBGP
  - Particulièrement utile lorsque plusieurs clients utilisent le même AS (RFC2270)
  - Utile aussi aux points d'échanges (IXP) où les politiques sont généralement les mêmes pour tous les pairs
- Peer-groups sont obsolètes
  - Mais encore largement considérés comme bonne pratique
  - Remplacés par update-groups (codés en interne - pas configurable)
  - Amélioré avec les peer-templates (permettant des conceptions plus complexes)



# Route Reflectors



Evolution du maillage iBGP

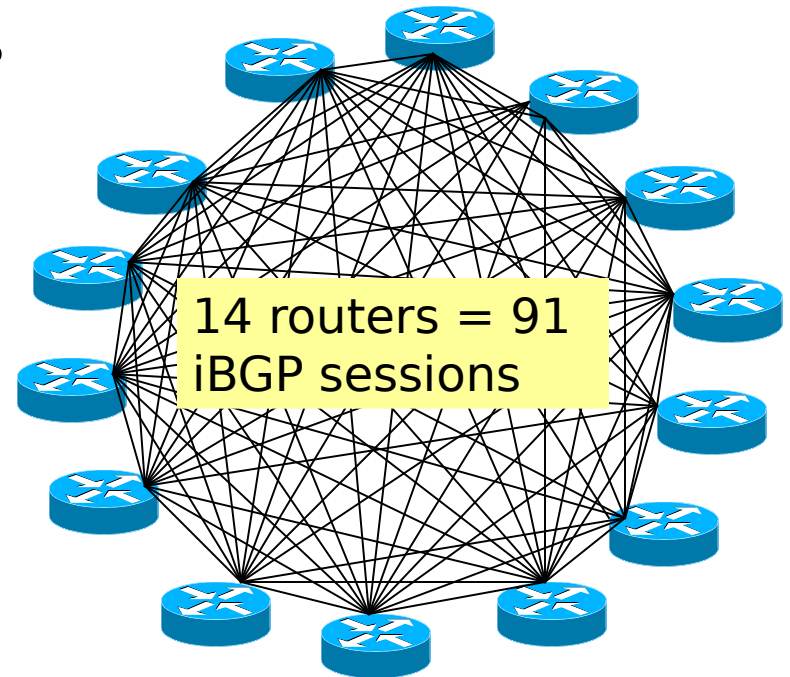
# Evolution du maillage iBGP

Eviter les  $\frac{1}{2}n(n-1)$  sessions du maillage complet iBGP

**$n=1000 \Rightarrow$  presque un demi million de sessions ibgp!**

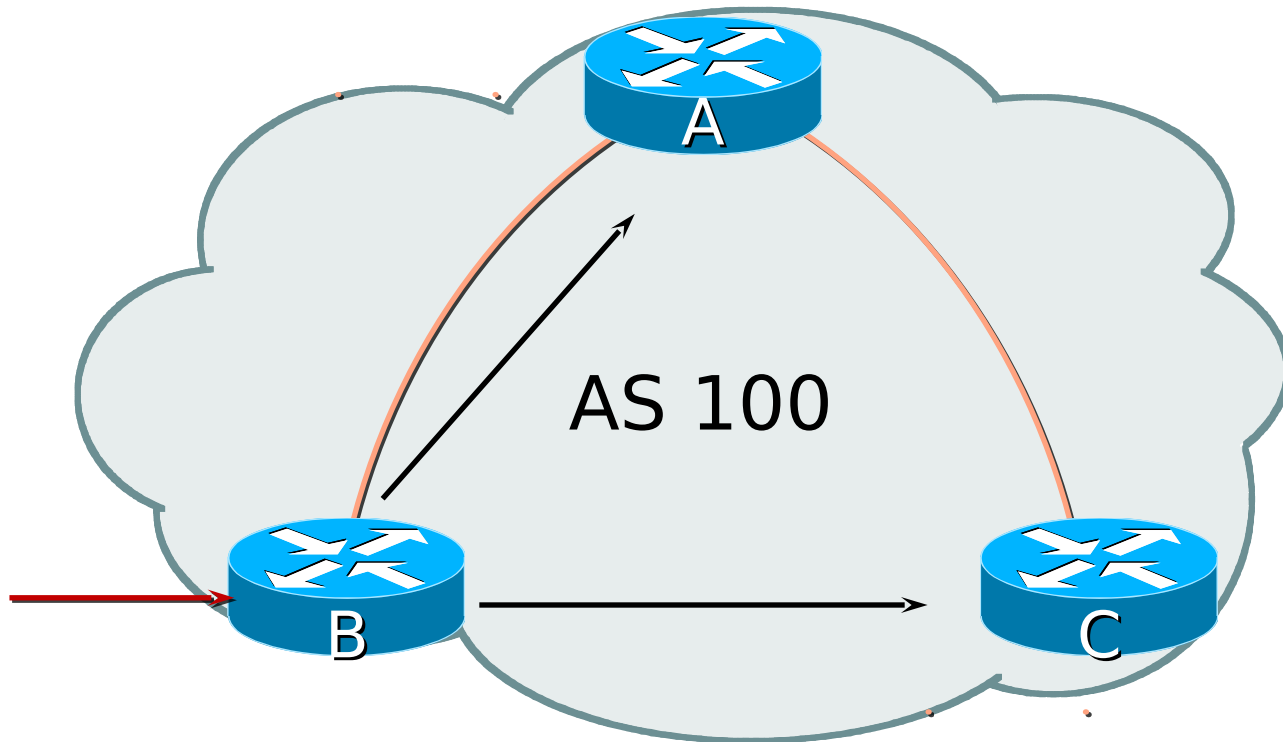
□ Deux solutions

- Réflecteurs de route – plus simples à déployer et maintenir
- Confédération – plus complexes, avantages induits



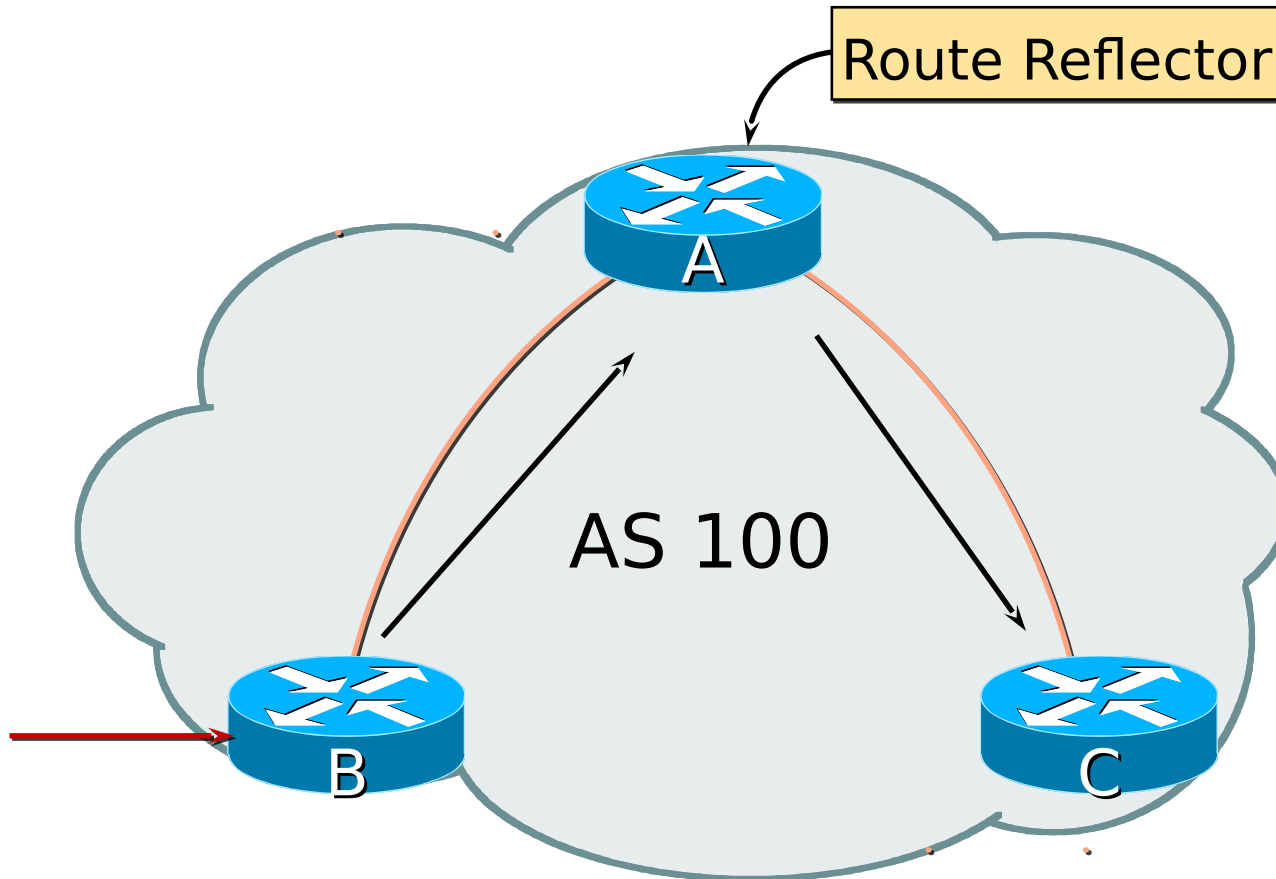
# Route Reflector: Principe

---



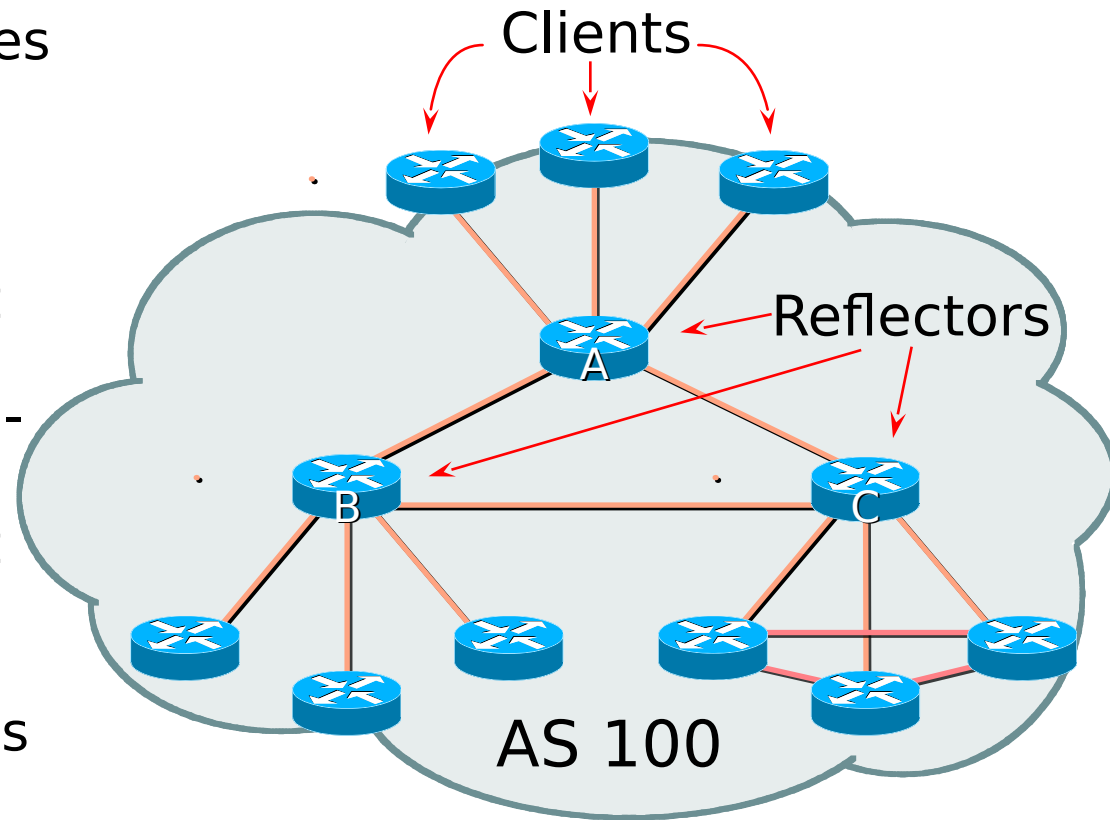
# Route Reflector: Principe

---



# Route Reflector

- ❑ Le réflecteur reçoit les chemins des clients et des non-clients
- ❑ Sélectionne le meilleur chemin
- ❑ Si le meilleur chemin est d'un client, envoyer aux autres clients et aux non-clients
- ❑ Si le meilleur chemin est d'un non-client, envoyer seulement aux clients
- ❑ Plus de maillage entre les clients
- ❑ Décrit dans RFC4456



# Topologie Route Reflector

---

- ❑ Diviser le backbone en groupes/clusters
- ❑ Au moins un route reflector et quelques clients par cluster
- ❑ Route reflectors forment un maillage complet(full-mesh)
- ❑ Clients dans un cluster peuvent former un maillage
- ❑ Un seul IGP pour transporter les next-hop et les routes locales

# Route Reflectors: Eviter les boucles

---

- Attribut Originator\_ID
  - Transporte le Router ID du routeur d'origine dans l'AS local (créé par le RR)
- Attribut Cluster\_list
  - Le cluster-id local est ajouté lorsque l'update est envoyé par le RR
  - Cluster-id est le router-id (adresse loopback)
  - **NE pas utiliser** `bgp cluster-id x.x.x.x`

# Route Reflectors: Redondance

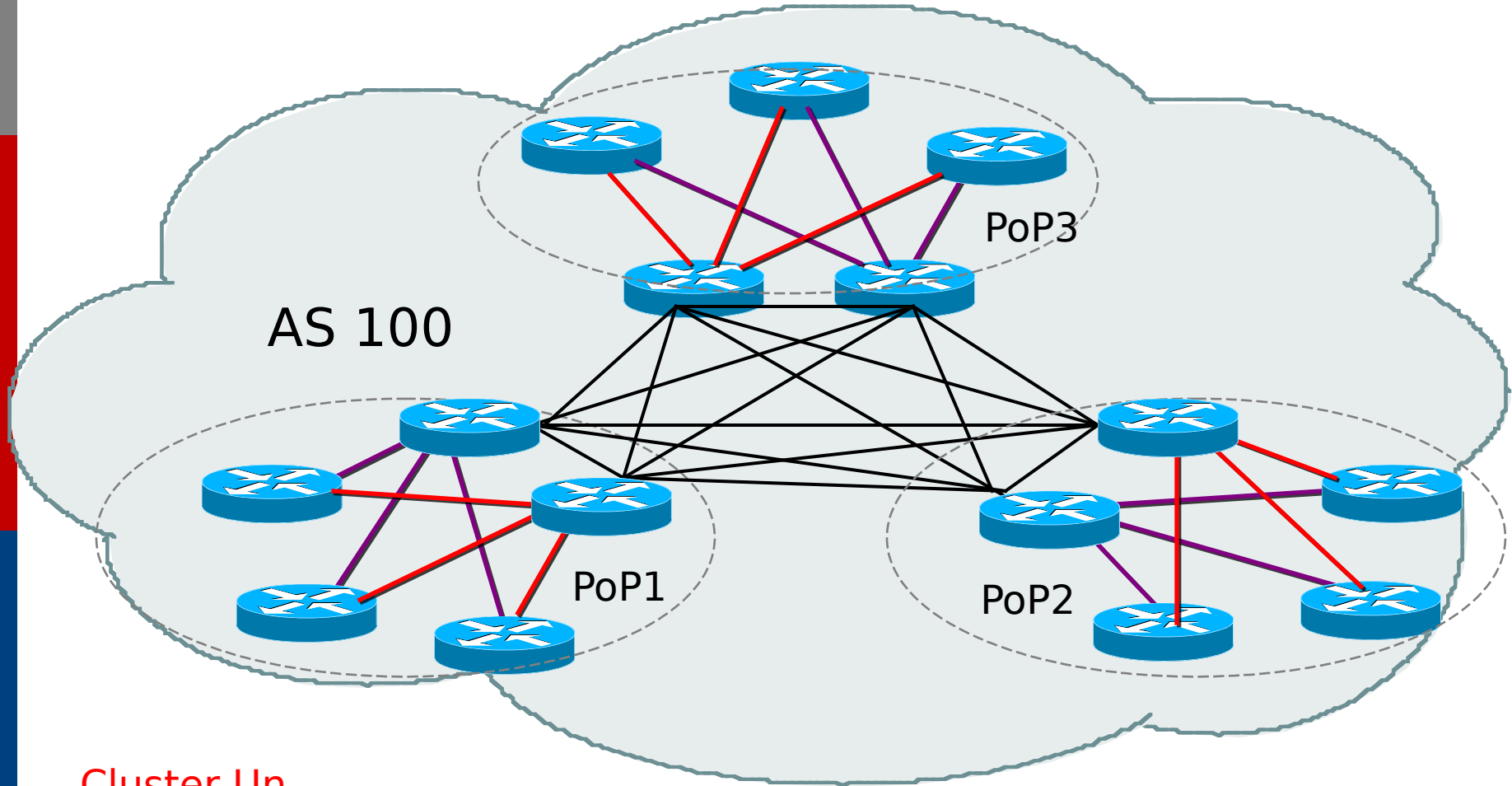
---

- Plusieurs RRs peuvent être configurés dans le même cluster – pas recommandé!
  - Tous les RRs du cluster doivent avoir le même cluster-id (sinon ils sont dans un cluster différent)
- Un routeur peut être client de RRs de clusters différents
  - Il est fréquent de voir aujourd’hui dans les réseaux d’ISP que les clusters se recouvrent – moyen pour obtenir de la redondance
  - → Chaque client a deux RRs = redondance



# Route Reflectors: Redondance

---



Cluster Un

Cluster Deux

# Route Reflector: Intérêt

---

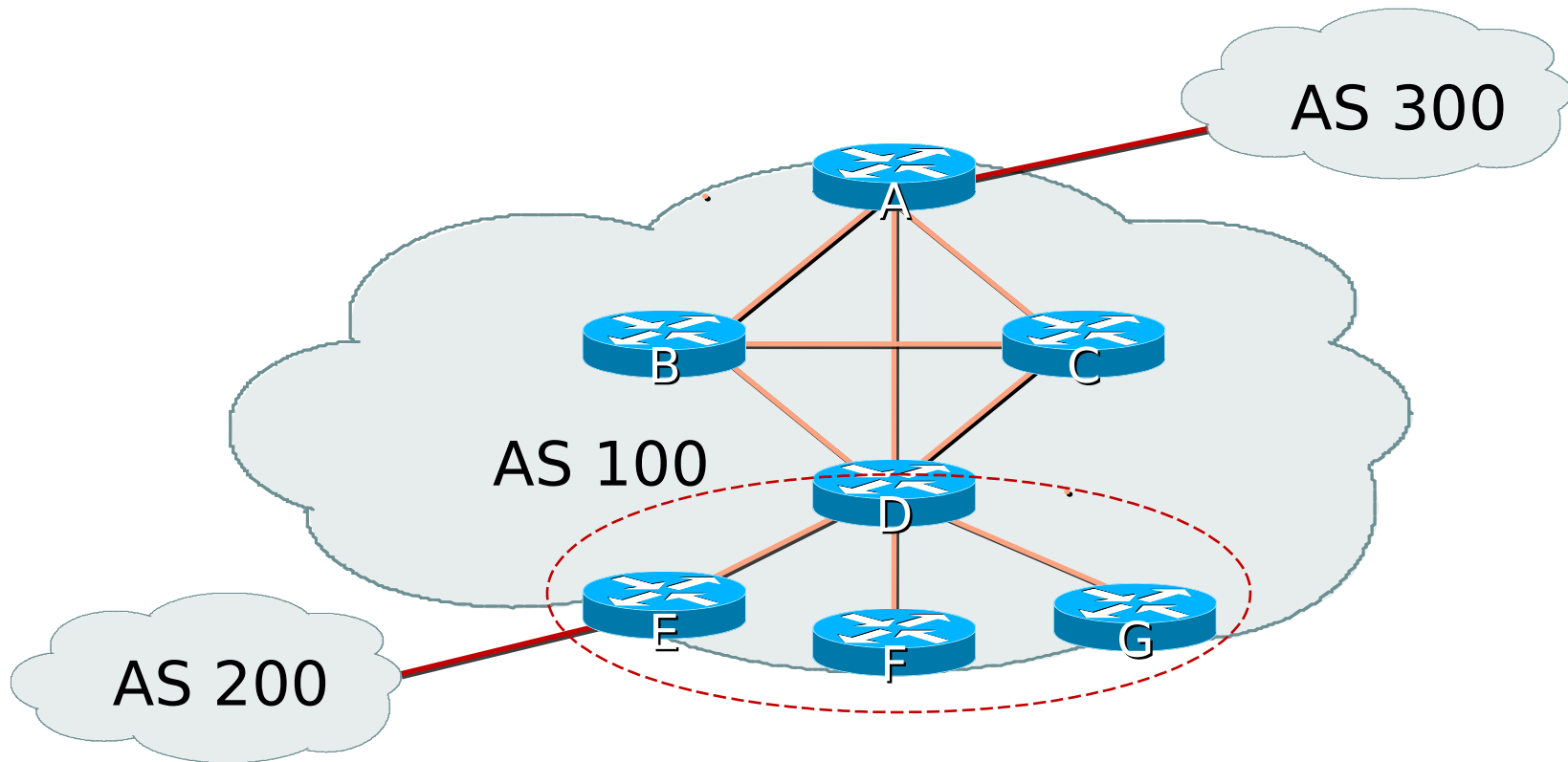
- ❑ Résout le problème du maillage iBGP
- ❑ Pas d'effet sur le transfert des paquets
- ❑ Les speakers BGP normaux co-existent
- ❑ Plusieurs reflecteurs pour un client pour avoir la redondance
- ❑ Migration facile
- ❑ Plusieurs niveaux de route reflectors

# Route Reflectors: Migration

---

- Où placer les route reflecteurs?
  - **Suivre la topologie physique!**
  - Ceci garantit que le transfert de paquets ne sera pas impacté
- Configurer un RR à la fois
  - Eliminer les sessions iBGP redondantes
  - Placer un RR par cluster

# Route Reflectors: Migration



- ❑ Migrer de petits morceaux du réseau, un morceau à la fois.

# Configurer un Route Reflector

---

- Configuration du routeur D:

```
router bgp 100
...
neighbor 1.2.3.4 remote-as 100
neighbor 1.2.3.4 route-reflector-client
neighbor 1.2.3.5 remote-as 100
neighbor 1.2.3.5 route-reflector-client
neighbor 1.2.3.6 remote-as 100
neighbor 1.2.3.6 route-reflector-client
...
```

# Techniques d'optimisation BGP

---

- Ces 3 techniques sont primordiales pour tous les ISPs
  - Route Refresh (ou Soft Reconfiguration)
  - Peer groups
  - Route Reflectors

# Confédérations BGP



# Confédérations

---

- Diviser l'AS en sous-AS
  - eBGP entre les sous-AS, mais certaines informations iBGP sont maintenues
    - Préserver NEXT\_HOP à travers les sous-AS (IGP porte cette information)
    - Préserver LOCAL\_PREF et MED
- Habituellement, un seul IGP
- Décrit dans la RFC5065



# Confédérations

---

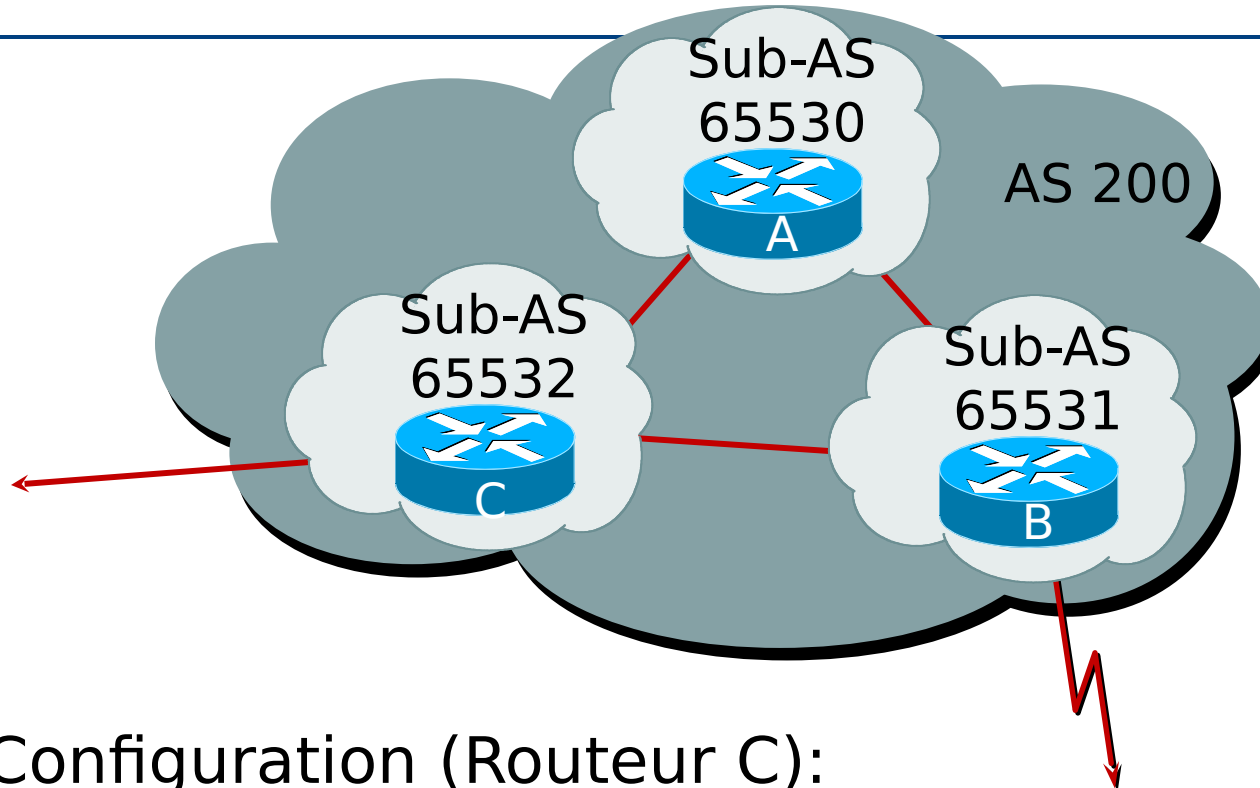
- Visible au monde extérieur comme un seul AS à travers - " identificateur de Confédération "

  - Chaque sous-AS utilise un numéro d'AS de l'espace privé (64512-65534)

- Les routeurs iBGP dans les sous-AS sont entièrement maillés
  - Le nombre total de voisins est ainsi réduit en limitant l'exigence de maillage complet qu'aux seuls pairs dans les sous-AS

# Confédérations

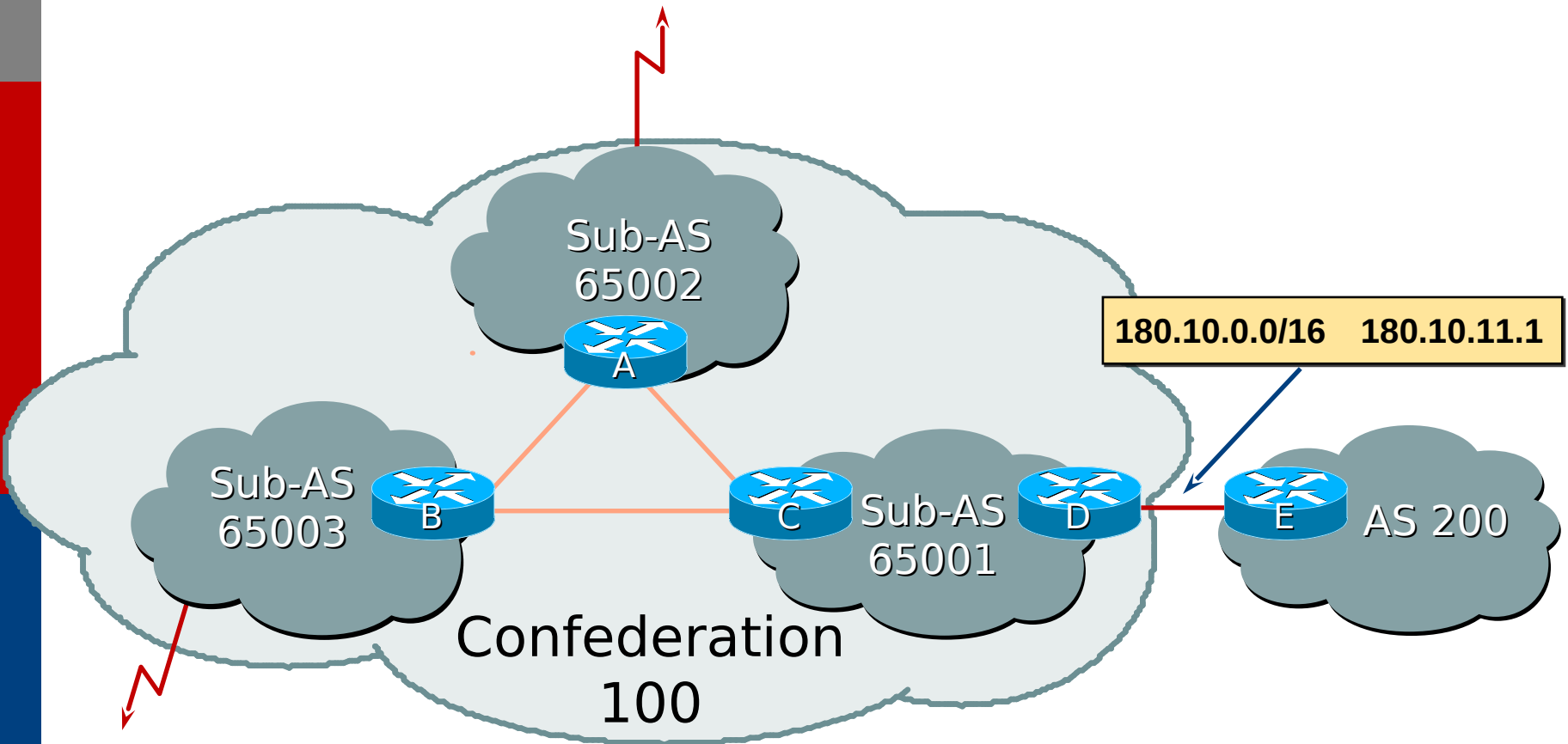
---



## □ Configuration (Routeur C):

```
router bgp 65532
  bgp confederation identifier 200
  bgp confederation peers 65530 65531
  neighbor 141.153.12.1 remote-as 65530
  neighbor 141.153.17.2 remote-as 65531
```

# Confédérations: Next Hop



# Confédération: Le Principe

---

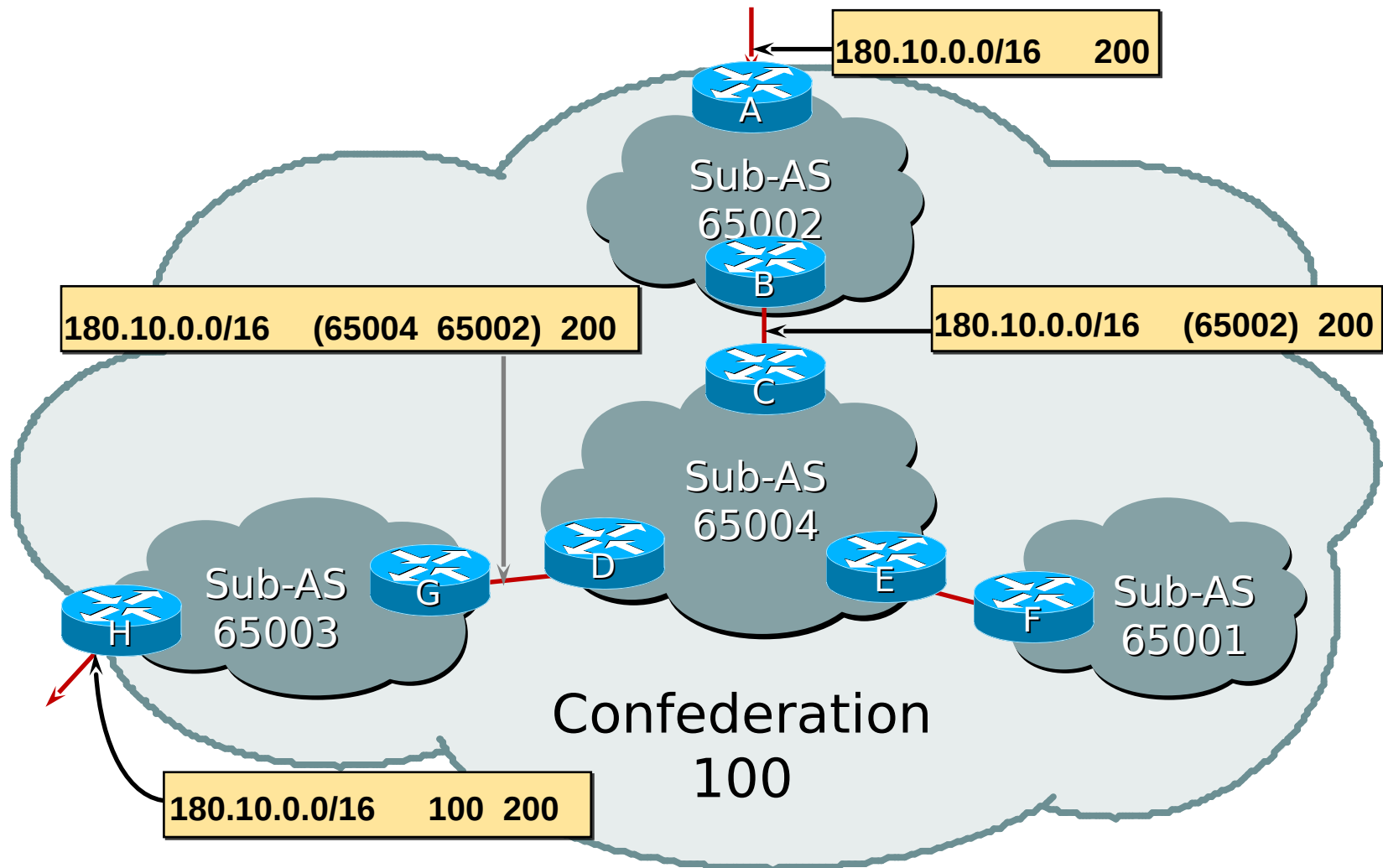
- La préférence locale et le MED influencent la sélection du chemin
- Préserver la préférence locale et MED à travers les frontières des sous-AS
- Sub-AS eBGP path administrative distance

# Confédérations: éviter les boucles (Loop Avoidance)

---

- ❑ Les sous-AS traversés sont considérés dans l'AS-path
- ❑ Transporter aussi bien la séquence AS que la longueur de l'AS-Path
- ❑ Considérer es frontières des confédérations
- ❑ Séquence AS doit être omise lors de la comparaison MED

# Confédérations: séquences d'AS



# Les décisions de propagation de route

---

- Même chose qu'avec BGP «normal»:
  - À partir de pairs dans le même sous-AS → uniquement vers des pairs externes
  - À partir de pairs externes → vers tous les voisins
- «Pairs externes» désignent
  - Des pairs en dehors de la confédération
  - Des Pairs dans un autre sous-AS
    - préservent LOCAL\_PREF, MED et NEXT\_HOP

# Confederations (suite.)

---

## □ Exemple (suite.):

BGP table version is 78, local router ID is 141.153.17.1

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    |   |   |
|----------------|--------------|--------|--------|--------|---------|---|---|
| *> 10.0.0.0    | 141.153.14.3 | 0      | 100    | 0      | (65531) | 1 | i |
| *> 141.153.0.0 | 141.153.30.2 | 0      | 100    | 0      | (65530) |   | i |
| *> 144.10.0.0  | 141.153.12.1 | 0      | 100    | 0      | (65530) |   | i |
| *> 199.10.10.0 | 141.153.29.2 | 0      | 100    | 0      | (65530) | 1 | i |



# Plus de points sur les confédérations

---

- Peut faciliter "l'absorption" d'autres ISPs dans votre réseau
  - par exemple, si un ISP achète un autre ISP
  - (peut utiliser les caractéristiques d'un AS local pour faire la même chose)
- Vous pouvez utiliser des réflecteurs de routes avec la confédération sous-AS pour réduire le maillage iBGP des sous-AS

# Confédérations : Avantages

---

- ❑ Résout les problèmes de maillage iBGP
- ❑ Le transfert de paquets n'est pas affecté
- ❑ Peut être utilisé avec des réflecteurs de route
- ❑ Des politiques pourraient être appliquées pour acheminer le trafic entre les sous-AS

# Confédérations: Mises en garde

---

- Nombre minimal de sous-AS
- Hiérarchie des sous-AS
- Inter-connectivité minimale entre les sous-AS
- Diversité de trajet
- Migration difficile
  - BGP reconfiguré en sous-AS
  - doit être appliquée à travers le réseau

# RR ou Confédérations

---

|                      | Connectivité Internet       | Hierarchie à niveaux multiples | Contrôle des Politiques | Évolutivité | Complexité de la migration |
|----------------------|-----------------------------|--------------------------------|-------------------------|-------------|----------------------------|
| Confédérations       | N'importe où dans le réseau | Oui                            | Oui                     | Moyen       | Moyen à élevé              |
| Réflecteurs de route | N'importe où dans le réseau | Oui                            | Oui                     | Très élevé  | Très faible                |

**La plupart des nouveaux réseaux fournisseurs de services désormais déploie des réflecteurs de route dès la première journée**

# Route Flap Damping



Stabilité réseau pour les années  
1990

Instabilité réseau pour le 21e  
siècle!

# Route Flap Damping

---

- Pendant de nombreuses années, Route Flap Damping était une pratique fortement recommandée
- Maintenant, elle est fortement déconseillée car elle provoque beaucoup plus d'instabilité réseau qu'elle n'en garantit.
- Mais d'abord, voyons la théorie ...

# Route Flap Damping

---

- Route flap
  - Montée et descente de trajectoire (path) ou changement dans l'attribut
    - RETRAIT d'un pair BGP suivi d'1 MISE A JOUR= 1 flap
    - Voisin eBGP oscillant down/up N'EST pas 1flap
  - Se répercute à travers tout l'Internet
  - Gaspille du CPU
- Le Damping vise à réduire la portée de propagation des " route flap "

# Route Flap Damping (suite)

---

- Exigences
  - Convergence rapide pour les changements de route normaux
  - L'histoire prédit les comportements futurs
  - Supprime les routes oscillantes
  - Annonce les routes stables
- Mise en œuvre décrit dans la RFC 2439



# Historique du Route Flap Damping

---

- Premières mises en œuvre sur l'Internet en 1995
- Paramètres par défaut des vendeurs trop rigoureux
  - Recommandations du groupe de travail sur RIPE Routing dans ripe-178, ripe-210, et ripe-229
  - <http://www.ripe.net/ripe/docs>
  - Mais de nombreux ISP se sont tout simplement positionnés sur les valeurs par défaut des vendeurs sans réfléchir

# Problèmes graves:

---

- "Route Flap Damping Exacerbates Internet Routing Convergence"
  - Zhuoqing Morley Mao, Ramesh Govindan, George Varghese & Randy H. Katz, August 2002
- "What is the sound of one route flapping?"
  - Tim Griffin, June 2002
- Divers travaux sur la convergence de routage par Craig Labovitz et Abha Ahuja il y a quelques années
- "Happy Packets"
  - Des travaux étroitement liés par Randy Bush et al

# Techniques d'Optimisation BGP



**Atelier SI-F**  
**AfNOG 2018, Dakar**