

Meilleures pratiques courantes de BGP



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

Configuration BGP



Par où commençons-nous?

Meilleures pratiques avec IOS

- Les FAI devraient commencer avec les commandes BGP suivantes comme modèle de base:

```
router bgp 64511
  bgp deterministic-med
  distance bgp 200 200 200
  no synchronization
  no auto-summary
```

Remplacer par l'ASN publique

Rendre ebgp et ibgp la même distance

- Pour supporter plus que les voisins IPv4 unicast
 - `no bgp default ipv4-unicast` est également très important et nécessaire

Bonnes pratiques avec IOS Cisco

- BGP dans Cisco IOS est **permissif** par défaut
- Configuration de peering BGP sans utiliser de filtres signifie:
 - Tous les meilleurs chemins sur le routeur local sont transmises au voisin
 - Toutes les routes annoncées par le voisin sont reçues par le routeur local
 - Peut avoir des conséquences désastreuses
- **La bonne pratique consiste à s'assurer que chaque voisin eBGP a des filtres entrant et sortant appliqués:**

```
router bgp 64511
  neighbor 1.2.3.4 remote-as 64510
  neighbor 1.2.3.4 prefix-list as64510-in in
  neighbor 1.2.3.4 prefix-list as64510-out out
```

À quoi sert BGP??



Ce qu'un IGP n'est pas?

BGP par rapport OSPF / ISIS

- Protocoles de routage interne (IGP)
 - exemples : ISIS et OSPF
 - utilisé pour le transport des adresses de **l'infrastructure**
 - **N'est PAS** utilisé pour transporter des préfixes d'Internet ou de préfixes de client
 - L'objectif de la conception est de **minimiser** le nombre de préfixes en IGP pour faciliter l'évolutivité et la convergence rapide

BGP par rapport OSPF / ISIS

- BGP utilisé en interne (iBGP) et externe (eBGP)
- iBGP utilisé pour transporter
 - certains / tous les préfixes d'Internet à travers le backbone
 - préfixes des clients
- eBGP utilisé pour
 - échanger les préfixes avec d'autres AS
 - mettre en œuvre la politique de routage

BGP par rapport OSPF / ISIS

- NE PAS:
 - distribuer des préfixes BGP dans un IGP
 - distribuer des routes IGP dans BGP
 - utiliser un IGP pour transporter les préfixes du client
- **SINON VOTRE RÉSEAU NE S'ADAPTERA PAS**

Agrégation



Agrégation

- L'agrégation consiste à annoncer le bloc d'adresse reçu du RIR aux autres ASs connectés à votre réseau
- Les sous-préfixes de cet agrégat peuvent être:
 - Utilisé en interne dans le réseau du FAI
 - Annoncé à d'autres AS pour aider avec le multi-homing
- Malheureusement, trop de gens pensent toujours à la classe C, ce qui entraîne une prolifération de /24s dans la table de routage Internet
 - Remarque : Même chose se passe pour /48s avec IPv6

Configuration de l'agrégation – IOS Cisco

- FAI a 101.10.0.0/19 bloc d'adresses
- Pour mettre en BGP comme un agrégat:

```
router bgp 64511
  network 101.10.0.0 mask 255.255.224.0
  ip route 101.10.0.0 255.255.224.0 null0
```
- La route statique est une route de “pull up”
- des préfixes plus spécifiques au sein de ce bloc d'adresse assurent la connectivité aux clients du FAI
 - “Longest match first”

Agrégation

- Le bloc d'adresses doit être annoncé à l'Internet comme un agrégat
- Les sous-préfixes de bloc d'adresses **NE** doivent **PAS** être annoncés à Internet sauf pour de l'ingénierie du trafic
 - Voir les présentations BGP Multi-homing (si le temps le permet)
- l'Agrégat doit être généré en interne
 - Non pas sur les frontières du réseau!

L'annonce de l'agrégat – IOS Cisco

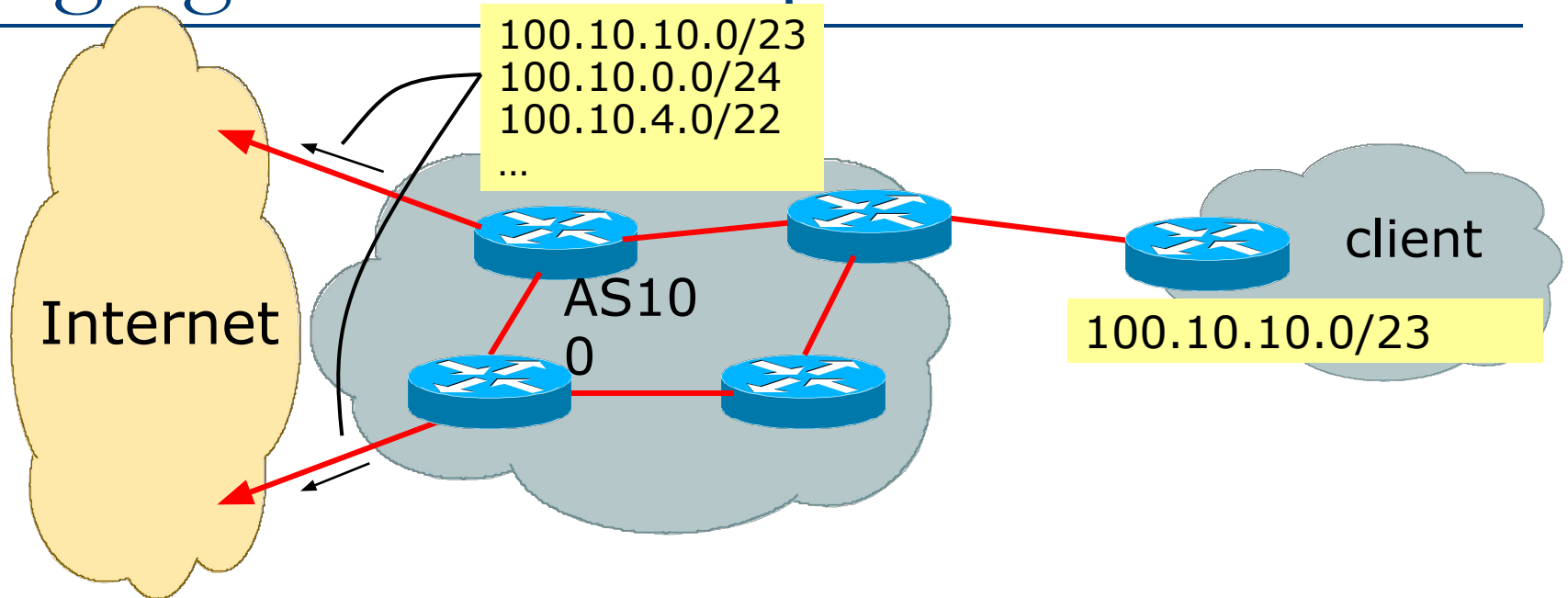
- Exemple de configuration

```
router bgp 64511
  network 101.10.0.0 mask 255.255.224.0
  neighbor 102.102.10.1 remote-as 101
  neighbor 102.102.10.1 prefix-list out-filter out
!
ip route 101.10.0.0 255.255.224.0 null0
!
ip prefix-list out-filter permit 101.10.0.0/19
ip prefix-list out-filter deny 0.0.0.0/0 le 32
```

Annnonce d'un agrégat

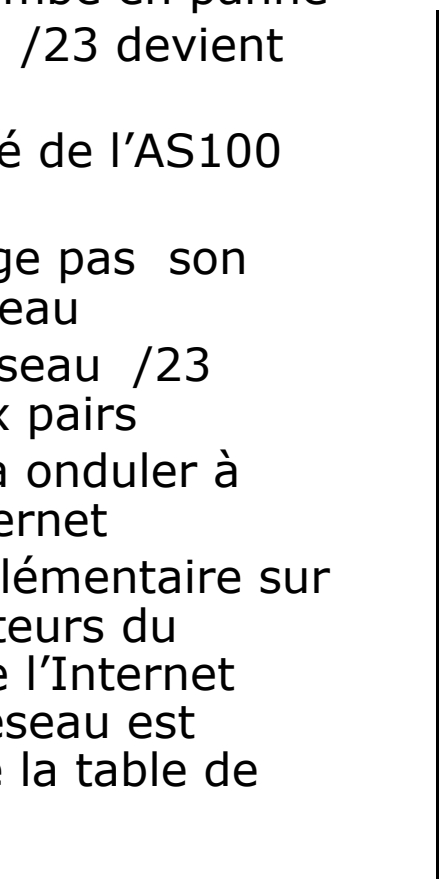
- Les FAI qui n'agrègent pas et ne vont pas le faire sont mals vus par la communauté
- Les RIR publient leur taille d'allocation minimale
 - Pour IPv4
 - Allant de / 20 à / 24 selon le RIR
 - Différentes tailles pour différents blocs d'adresses
 - APNIC a changé en Octobre 2010 sa taille d'allocation minimale à /24
 - Pour IPv6
 - Des assignations de /48 et des allocations de /32
- Jusque récemment il ne devrait pas y avoir de préfixes plus longs que /22 dans Internet
- Certainement l'épuisement des adresses IPv4 commence à avoir un impact

Agrégation - Exemple

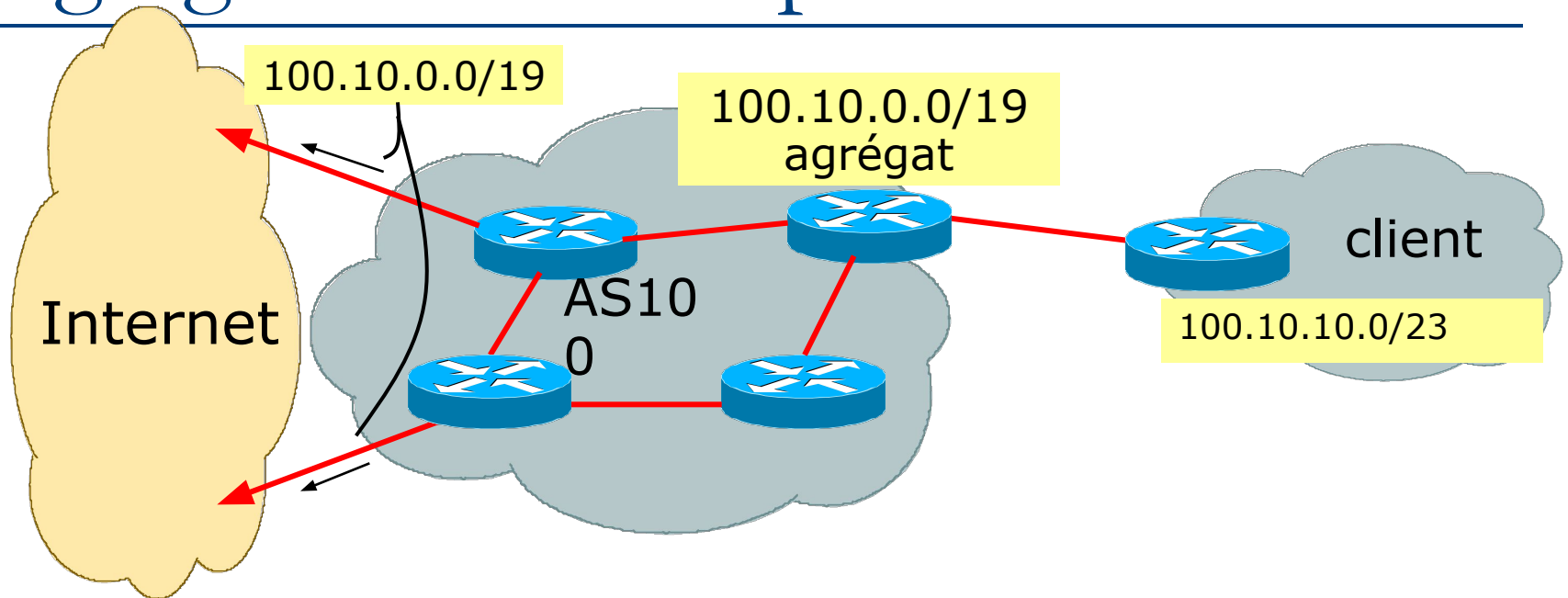


- Le client dispose d'un réseau /23 affecté du bloc d'adresses AS100 /19
- AS100 annonce des réseaux individuels des clients à l'Internet

Agrégation - Mauvais exemple


- Lien du client tombe en panne
 - Leur réseau /23 devient inaccessible
 - /23 est retiré de l'AS100 iBGP
 - Si le FAI n'agrège pas son bloc /19 du réseau
 - retrait de réseau /23 annoncé aux pairs
 - commence à onduler à travers l'Internet
 - charge supplémentaire sur tous les routeurs du backbone de l'Internet comme le réseau est supprimé de la table de routage
 - Le lien du client remonte
 - Le réseau /23 est désormais visible à leur FAI
 - Le réseau /23 est republié aux pairs
 - commence à onduler à travers l'Internet
 - Charge sur les routeurs du backbone de l'Internet tant que réseau est réinséré dans la table de routage
 - Certains FAI suppriment les flappings
 - Internet peut prendre 10-20 minutes ou plus pour être visible
 - Où est la qualité de service???
- 

Agrégation - Exemple



- Le client dispose d'un réseau /23 affecté du bloc d'adresses AS100 /19
- AS100 annonce l'agrégat /19 à Internet

Agrégation - Bon exemple

- Le lien du client tombe en panne
 - leur réseau /23 devient inaccessible
 - /23 est retiré de l'AS100 iBGP
 - l'agregat /19 est toujours annoncé
 - no BGP hold down problems
 - no BGP propagation delays
 - no damping by other ISPs
- 
- Le lien du client remonte
 - Le /23 est de nouveau visible
 - Le /23 est réinjecté dans le iBGP de AS100
 - Tout l'Internet devient immédiatement visible
 - Client dispose de la perception de la qualité de service

Agrégation - Résumé

- Le bon exemple c'est ce que tout le monde devrait le faire!
 - Ajoute à la stabilité de l'Internet
 - Réduit la taille de la table de routage
 - Réduit le taux de désabonnement de routage
 - Améliore la qualité de service Internet pour **tout le monde**
- Le mauvais exemple c'est ce que beaucoup font encore malheureusement!
 - Pourquoi? Manque de connaissances?
 - Paresse?

Séparation de iBGP et eBGP

- De nombreux FAI ne comprennent pas l'importance de séparer iBGP et eBGP
 - iBGP est l'endroit où tous les préfixes des clients sont propagés
 - l'eBGP est employé pour annoncer l'agrégat à l'Internet et pour l'ingénierie du trafic
- **NE PAS** faire d'ingénierie de trafic avec le client originant de préfixes iBGP
 - Conduit à une instabilité similaire à celle mentionnée dans le mauvais exemple précédent
 - Même s'il est agrégé (avec plusieurs opérateurs) un préfixe de routage instable générera à son tour de l'instabilité pour le client concerné.
- **Générer les préfixes utilisés pour l'ingénierie du trafic sur le routeur de bordure**

L'Internet Aujourd'hui (Mai 2018)

- Statistiques actuelles de la table de routage
 - Les entrées de table de routage BGP: 690376
 - Préfixes après agrégation maximale: 267537
 - Préfixes uniques dans Internet: 336141
 - Préfixes plus petits que l'allocation des registres: 171043
 - AS en utilisation: 60948

<http://www.cidr-report.org/as2.0/>

Les efforts pour améliorer l'agrégation

- Le rapport CIDR
 - Initié et géré depuis de nombreuses années par Tony Bates
 - Maintenant, combiné à l'analyse de routage de Geoff Huston
 - www.cidr-report.org
 - (couvre à la fois les tables BGP de IPv4 et IPv6)
 - Résultats envoyés par courriel sur une base hebdomadaire pour les listes de la plupart des opérations dans le monde entier
 - Dresser la liste des 30 meilleurs fournisseurs de services qui pourrait faire mieux à l'agrégation
- Recommandation de l'agrégation RIPE routage WG
 - RIPE-399 — www.ripe.net/ripe/docs/ripe-399.html

CIDR REPORT

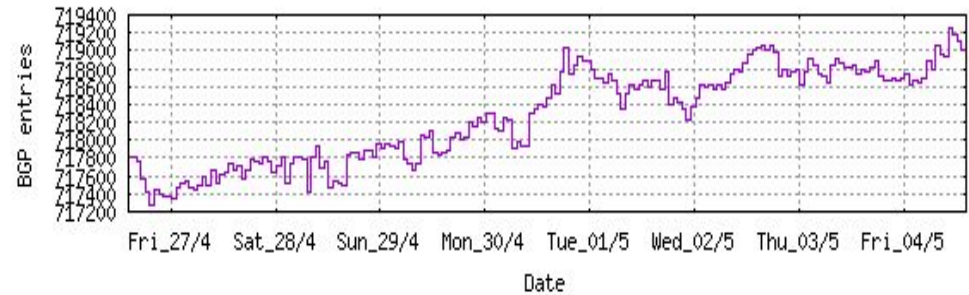
- Calcule aussi la taille de la table de routage en supposant que les FAI effectuent une agrégation optimale
- **Ce site Web permet des recherches et des calculs d'agrégation effectués sur la base des AS**
 - Outil flexible et puissant pour aider les FAI
 - Destiné à montrer comment une plus grande efficacité en termes de taille de la table BGP peut être obtenu sans perte de routage et d'information politique
 - Montre quelles sont les formes d'origine d'agrégation AS qui devraient être utilisées et les avantages potentiels de ces actions sur la taille totale de la table BGP Internet
 - Défie très efficacement l'excuse ingénierie de trafic

Status Summary

Table History

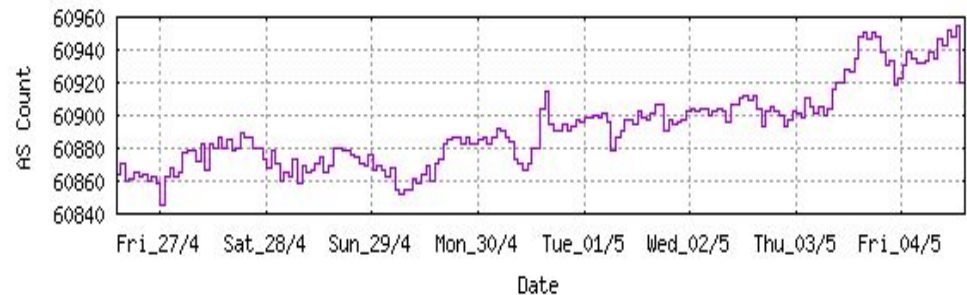
Date	Prefixes	CIDR Aggregated
27-04-18	717379	388556
28-04-18	717632	389064
29-04-18	717965	388571
30-04-18	718193	389000
01-05-18	718886	388973
02-05-18	718373	389153
03-05-18	718796	389434
04-05-18	718702	389443

Plot: [BGP Table Size](#)



AS Summary

60948	Number of ASes in routing system
22751	Number of ASes announcing only one prefix
5390	Largest number of prefixes announced by an AS AS4538 : ERX-CERNET-BKB China Education and Research Network Center, CN
119177216	Largest address span announced by an AS (1/22)



Aggregation Suggestions

This report does not take into account conditions local to each origin AS in terms of policy or traffic engineering requirements, so this is an approximate guideline as to aggregation possibilities.

Rank	AS	AS Name	Current	Withdw	Aggte	Annce	Redctn	%
2	AS6389	BELLSOUTH-NET-BLK - BellSouth.net Inc.	3411	3256	41	196	3215	94.25%

Prefix	AS Path	Aggregation Suggestion
12.81.90.0/23	4777 2516 3356 7018 6389	
12.83.3.0/24	4777 2516 3356 7018 6389	
12.83.5.0/24	4777 2516 3356 7018 6389	
12.83.7.0/24	4777 2516 3356 7018 6389	
65.0.0.0/12	4777 2516 3356 7018 6389	
65.0.0.0/18	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.0.40.0/22	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.0.50.0/23	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.0.64.0/18	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.0.128.0/18	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.0.192.0/19	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.0.224.0/19	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.1.0.0/19	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.1.32.0/19	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.1.64.0/19	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.1.128.0/18	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.1.224.0/20	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.1.240.0/20	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.2.0.0/16	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.2.0.0/17	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.2.128.0/17	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.3.224.0/19	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.4.32.0/19	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.4.64.0/18	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.1.0/24	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.12.0/22	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.16.0/22	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.20.0/23	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.21.0/24	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.22.0/23	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.24.0/22	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.28.0/22	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.32.0/20	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.34.0/24	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.46.0/24	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389
65.5.57.0/24	4777 2516 3356 7018 6389	- Withdrawn - matching aggregate 65.0.0.0/12 4777 2516 3356 7018 6389

1232:30 remaining

AS Report

Aggregation Suggestions

This report does not take into account conditions local to each origin AS in terms of policy or traffic engineering requirements, so this is an approximate guideline as to aggregation possibilities.

Rank	AS	AS Name	Current	Withdw	Aggte	Annce	Redctn	%
6	AS18566	COVAD - Covad Communications Co.	2092	1687	301	706	1386	66.25%

Prefix	AS Path	Aggregation Suggestion
64.81.7.0/24	4777 2516 3356 18566	
64.81.16.0/22	4777 2516 3356 18566	
64.81.20.0/22	4777 2516 4565 18566	
64.81.22.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.20.0/22 4777 2516 4565 18566
64.81.24.0/21	4777 2516 3356 18566	+ Announce - aggregate of 64.81.24.0/22 (4777 2516 3356 18566) and 64.81.28.0/22 (4777 2516 3356 18566)
64.81.24.0/22	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.28.0/22 (4777 2516 3356 18566)
64.81.28.0/22	4777 2516 3356 18566	- Withdrawn - aggregated with 64.81.24.0/22 (4777 2516 3356 18566)
64.81.32.0/20	4777 2516 4565 18566	
64.81.32.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.33.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.34.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.35.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.36.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.37.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.38.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.39.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.40.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.44.0/24	4777 2516 4565 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 4565 18566
64.81.48.0/20	4777 2516 3356 18566	
64.81.48.0/24	4777 2516 3356 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.49.0/24	4777 2516 3356 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.50.0/24	4777 2516 3356 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.51.0/24	4777 2516 3356 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.52.0/24	4777 2516 3356 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.53.0/24	4777 2516 3356 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.54.0/24	4777 2516 3356 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.55.0/24	4777 2516 3356 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.56.0/24	4777 2516 3356 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.57.0/24	4777 2516 3356 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.58.0/24	4777 2516 3356 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.59.0/24	4777 2516 3356 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.60.0/24	4777 2516 3356 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.61.0/24	4777 2516 3356 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3356 18566
64.81.64.0/20	4777 2516 3356 18566	
64.81.64.0/24	4777 2516 3356 18566	- Withdrawn - matching aggregate 64.81.64.0/20 4777 2516 3356 18566
64.81.65.0/24	4777 2516 3356 18566	- Withdrawn - matching aggregate 64.81.64.0/20 4777 2516 3356 18566

Importance de l'agrégation

- Taille de la table de routage
 - La mémoire des routeur n'est pas tellement un problème tel qu'il était dans les années 90
 - Les routeurs actuels peuvent être conçus pour transporter plus de 1 million de préfixes
- Convergence du système de routage
 - C'est un problème
 - Un plus grand tableau prend plus de temps pour le processeur a etre traité
 - Mises à jour BGP prend plus de temps à etre traitées
 - Rapport d'instabilité de routage BGP suit l'activité de mise à jour du système
 - bgpupdates.potaroo.net/instability/bgpupd.html

The BGP Instability Report

The BGP Instability Report is updated daily. This report was generated on 01 June 2012 06:15 (UTC+1000)

50 Most active ASes for the past 7 days

RANK	ASN	UPDs	%	Prefixes	UPDs/Prefix	AS NAME
1	383	192192	8.46%	126	1525.33	AFCONC-BLOCK1-AS - 754th Electronic Systems Group
2	8452	91007	4.00%	1375	66.19	TE-AS TE-AS
3	8402	60332	2.65%	1941	31.08	CORBINA-AS OJSC "Vimpelcom"
4	9829	54914	2.42%	1298	42.31	BSNL-NIB National Internet Backbone
5	19647	47156	2.07%	20	2357.80	HPOD20001 - Hewlett-Packard Operation Division
6	38142	33384	1.47%	16	2086.50	UNAIR-AS-ID Universitas Airlangga
7	12479	27393	1.21%	697	39.30	UNI2-AS France Telecom Espana SA
8	35994	26629	1.17%	73	364.78	AKAMAI-AS - Akamai Technologies, Inc.
9	7908	25800	1.14%	79	326.58	Comsat Argentina S.A.
10	5800	23102	1.02%	268	86.20	DNIC-ASBLK-05800-06055 - DoD Network Information Center
11	41661	21899	0.96%	356	61.51	ERTH-CHEL-AS CJSC "ER-Telecom Holding"
12	1257	20680	0.91%	190	108.84	TELE2
13	13118	20040	0.88%	48	417.50	ASN-YARTELECOM OJSC Rostelecom
14	28306	19103	0.84%	34	561.85	TC Net Informática e Telecomunicações LTDA
15	10455	18155	0.80%	65	279.31	LUCENT-CIO - Lucent Technologies Inc.
16	17813	17513	0.77%	133	131.68	MTNL-AP Mahanagar Telephone Nigam Ltd.
17	3549	16826	0.74%	1007	16.71	GBLX Global Crossing Ltd.
18	36856	16593	0.73%	7	2370.43	MOZILLA-CORP - Mozilla Corporation
19	25543	15053	0.66%	55	273.69	FasoNet-AS
20	24560	13716	0.60%	1028	13.34	AIRTELROADBAND-AS-AP Bharti Airtel Ltd., Telemedia Services
21	2118	12271	0.54%	1312	9.35	RELCOM-AS OOO "NPO Relcom"
22	19318	12159	0.53%	47	258.70	NJIIX-AS-1 - NEW JERSEY INTERNATIONAL INTERNET EXCHANGE LLC

1235:57 remaining

The BGP Instability Report

50 Most active Prefixes for the past 7 days

RANK	PREFIX	UPDs	%	Origin AS -- AS NAME
1	109.161.64.0/19	19671	0.81%	13118 -- ASN-YARTELECOM OJSC Rostelecom
2	63.245.221.0/24	16573	0.68%	36856 -- MOZILLA-CORP - Mozilla Corporation
3	168.87.176.0/24	14222	0.58%	19647 -- HPOD20001 - Hewlett-Packard Operation Division
4	168.87.128.0/21	14171	0.58%	19647 -- HPOD20001 - Hewlett-Packard Operation Division
5	130.36.34.0/24	10737	0.44%	32528 -- ABBOTT Abbot Labs
6	69.31.106.0/23	9520	0.39%	35994 -- AKAMAI-AS - Akamai Technologies, Inc.
7	155.72.152.0/21	9383	0.38%	19647 -- HPOD20001 - Hewlett-Packard Operation Division
8	155.72.158.0/24	9376	0.38%	19647 -- HPOD20001 - Hewlett-Packard Operation Division
9	41.43.147.0/24	9035	0.37%	8452 -- TE-AS TE-AS
10	23.2.6.0/23	8522	0.35%	35994 -- AKAMAI-AS - Akamai Technologies, Inc.
11	23.65.27.0/24	8520	0.35%	35994 -- AKAMAI-AS - Akamai Technologies, Inc.
12	62.36.252.0/22	8022	0.33%	12479 -- UNI2-AS France Telecom Espana SA
13	91.202.212.0/22	7750	0.32%	44798 -- PERVOMAYSK-AS PP "SKS-Pervomaysk"
14	62.36.249.0/24	6535	0.27%	12479 -- UNI2-AS France Telecom Espana SA
15	62.36.241.0/24	6218	0.25%	12479 -- UNI2-AS France Telecom Espana SA
16	62.36.210.0/24	6060	0.25%	12479 -- UNI2-AS France Telecom Espana SA
17	59.177.48.0/20	5860	0.24%	17813 -- MTNL-AP Mahanagar Telephone Nigam Ltd.
18	194.63.9.0/24	4646	0.19%	1273 -- CW Cable and Wireless Worldwide plc
19	182.64.0.0/16	4607	0.19%	24560 -- AIRTELBROADBAND-AS-AP Bharti Airtel Ltd., Telemedia Services
20	192.11.177.0/24	3616	0.15%	10455 -- LUCENT-CIO - Lucent Technologies Inc.
21	192.11.147.0/24	3616	0.15%	10455 -- LUCENT-CIO - Lucent Technologies Inc.
22	135.245.152.0/23	3610	0.15%	10455 -- LUCENT-CIO - Lucent Technologies Inc.
23	135.244.152.0/21	3610	0.15%	10455 -- LUCENT-CIO - Lucent Technologies Inc.
24	135.245.152.0/21	3610	0.15%	10455 -- LUCENT-CIO - Lucent Technologies Inc.
25	202.56.215.0/24	3170	0.13%	24560 -- AIRTELBROADBAND-AS-AP Bharti Airtel Ltd., Telemedia Services
26	137.14.144.0/21	3127	0.13%	246 -- ASIFICS-GW-AS - 754th Electronic Systems Group 383 -- AFCONC-BLOCK1-AS - 754th Electronic Systems Group

Réception des préfixes



Réception des préfixes

- Il existe trois scénarii pour la réception de préfixes d'autres ASN
 - Client parlant BGP
 - Pair parlant BGP
 - Opérateur en amont / Transit parlant BGP
- Chacun a une différente configuration de filtrage et doit être considéré séparément

Réception des préfixes: des clients

- Les FAI ne devraient accepter les préfixes qui ont été assignés ou affectés à leur client en aval
- Si le FAI a attribué l'espace d'adressage à son client, alors le client est en droit de l'annoncer de nouveau à son FAI
- Si le FAI n'a pas attribué l'espace d'adressage à son client, alors:
 - Vérifier dans les cinq bases de données RIR pour voir si cet espace d'adressage a vraiment été affecté au client
 - L'outil: `whois -h jwhois.apnic.net x.x.x.0/24`
 - (jwhois interroge toutes les bases de données RIR)

Réception des préfixes: des clients

- Exemple d'utilisation du whois pour vérifier si le client est en droit d'annoncer l'espace d'adressage:

```
$ whois -h whois.apnic.net 202.12.29.0
inetnum:          202.12.28.0 - 202.12.29.255
netname:          APNIC-AP
descr:           Asia Pacific Network Information Centre
descr:           Regional Internet Registry for the Asia-Pacific
descr:           6 Cordelia Street
descr:           South Brisbane, QLD 4101
descr:           Australia
country:         AU
admin-c:         AIC1-AP
tech-c:          NO4-AP
mnt-by:          APNIC-HM
mnt-irt:         IRT-APNIC-AP
changed:         hm-changed@apnic.net
status:          ASSIGNED PORTABLE
changed:         hm-changed@apnic.net 20110309
source:          APNIC
```

Portable - signifie que son
assignation au client est correcte,
le client peut vous l'annoncer

Réception des préfixes: des clients

- Exemple d'utilisation du whois pour vérifier si le client est en droit d'annoncer l'espace d'adressage:

```
$ whois -h whois.ripe.net 193.128.0.0
inetnum:          193.128.0.0 - 193.133.255.255
netname:          UK-PIPEX-193-128-133
descr:           Verizon UK Limited
country:         GB
org:             ORG-UA24-RIPE
admin-c:         WERT1-RIPE
tech-c:          UPHM1-RIPE
status:          ALLOCATED UNSPECIFIED
remarks:         Please send abuse notification to abuse@uk.uu.net
mnt-by:          RIPE-NCC-HM-MNT
mnt-lower:       AS1849-MNT
mnt-routes:     AS1849-MNT
mnt-routes:     WCOM-EMEA-RICE-MNT
mnt-irt:         IRT-MCI-GB
source:         RIPE # Filtered
```

ALLOUÉ - signifie qu'il s'agit d'espace d'adressage du fournisseur et ne peut être annoncé que par le FAI tenant l'allocation (dans ce cas Verizon Royaume-Uni)

Réception des préfixes: Cisco IOS

- Par Exemple:
 - L'opérateur en aval a le bloc 100.50.0.0/20
 - doit l'annoncer que sur les opérateurs amonts
 - les opérateurs amonts devraient seulement accepter ceci de eux
- Configuration sur l'opérateur amont

```
router bgp 100m
  neighbor 102.102.10.1 remote-as 101
  neighbor 102.102.10.1 prefix-list customer in
!
ip prefix-list customer permit 100.50.0.0/20
```

Réception des préfixes: des pairs

- Un pair est un FAI avec lesquels vous êtes d'accord pour échanger des préfixes que vous avez créés dans la table de routage d'Internet
 - Préfixes, que vous acceptez d'un pair, ne sont que ceux qu'ils ont indiqué qu'ils annonceront
 - Les préfixes que vous annoncez à votre pair sont seulement ceux vous avez indiqués que vous annoncerez

Réception des préfixes: des pairs

- Accepter ce que chacun va annoncer à l'autre:
 - Échange de courrier électronique de documentation dans le cadre de l'accord de peering, puis mises à jour en cours
OU
 - L'utilisation du registre routage de l'Internet et des outils de configuration tel que le IRRToolSet

www.isc.org/sw/IRRToolSet/

Réception des préfixes de pair:

Cisco IOS

- Par Exemple:
 - Le Pair a les blocs d'adresses 220.50.0.0/16, 61.237.64.0/18 et 81.250.128.0/17
- Configuration sur le routeur local

```
router bgp 100
  neighbor 102.102.10.1 remote-as 101
  neighbor 102.102.10.1 prefix-list my-peer in
!
ip prefix-list my-peer permit 220.50.0.0/16
ip prefix-list my-peer permit 61.237.64.0/18
ip prefix-list my-peer permit 81.250.128.0/17
ip prefix-list my-peer deny 0.0.0.0/0 le 32
```

Réception des préfixes:

De fournisseur en amont / Transit

- Le Fournisseur en amont/Transit est un FAI à qui vous payez pour vous donner un accès **ENTIER** à l' Internet
- Recevoir la totalité des préfixes n'est pas souhaitable

(sauf si vraiment nécessaire)

- Ingénierie de trafic – voir présentation de BGP
Multi-homing (si le temps le permet)
- Demander à votre fournisseur en amont / transit soit:
 - d'annoncer une route par défaut
 - OU
 - annoncer un préfixe, que vous pouvez utiliser par défaut

Réception des préfixes:

De fournisseur en amont / Transit

- Configuration du routeur en aval

```
router bgp 100
```

```
network 101.10.0.0 mask 255.255.224.0
```

```
neighbor 101.5.7.1 remote-as 101
```

```
neighbor 101.5.7.1 prefix-list infilter in
```

```
neighbor 101.5.7.1 prefix-list outfilter out
```

```
!
```

```
ip prefix-list infilter permit 0.0.0.0/0
```

```
!
```

```
ip prefix-list outfilter permit 101.10.0.0/19
```


Réception des préfixes:

De fournisseur en amont / Transit

- Configuration du routeur en amont

```
router bgp 101
  neighbor 101.5.7.2 remote-as 100
  neighbor 101.5.7.2 default-originate
  neighbor 101.5.7.2 prefix-list cust-in in
  neighbor 101.5.7.2 prefix-list cust-out out
!
ip prefix-list cust-in permit 101.10.0.0/19
!
ip prefix-list cust-out permit 0.0.0.0/0
```

Réception des préfixes:

De fournisseur en amont / Transit

- nécessaire pour recevoir les préfixes de n'importe quel fournisseur, des soins sont nécessaires.
 - Ne pas accepter la route par défaut (sauf si vous en avez besoin)
 - Ne pas accepter vos propres préfixes
- Pour IPv4:
 - Ne pas accepter les adresses privées (RFC1918) et certains préfixes d'usage spécial:
<http://www.rfc-editor.org/rfc/rfc5735.txt>
 - Ne pas accepter les préfixes plus long que /24
- Pour IPv6:
 - Ne pas accepter certains préfixes d'usage spécial:
<http://www.rfc-editor.org/rfc/rfc5156.txt>
 - Ne pas accepter les préfixes plus long que /48(?)

Réception des préfixes:

De fournisseur en amont / Transit

- Vérifier la liste de Team Cymru de "Bogon"
www.team-cymru.org/Services/Bogons/http.html
- Pour IPv4 consulter aussi:
www.rfc-editor.org/rfc/rfc6441.txt
- Pour IPv6 consulter aussi:
www.space.net/~gert/RIPE/ipv6-filters.html
- Serveur de Route de Bogon:
www.team-cymru.org/Services/Bogons/routeserver.html
 - Délivre un flux BGP (IPv4 et / ou IPv6) des blocs d'adresses qui ne doivent pas apparaître dans la table BGP

Réception des préfixes IPv4

```
router bgp 100
 network 101.10.0.0 mask 255.255.224.0
 neighbor 101.5.7.1 remote-as 101
 neighbor 101.5.7.1 prefix-list in-filter in
!
ip prefix-list in-filter deny 0.0.0.0/0          ! Défaut
ip prefix-list in-filter deny 0.0.0.0/8 le 32    ! Network Zero
ip prefix-list in-filter deny 10.0.0.0/8 le 32   ! RFC1918
ip prefix-list in-filter deny 100.64.0.0/10 le 32 ! RFC6598 shared address
ip prefix-list in-filter deny 101.10.0.0/19 le 32 ! Local prefix
ip prefix-list in-filter deny 127.0.0.0/8 le 32 ! Loopback
ip prefix-list in-filter deny 169.254.0.0/16 le 32 ! Auto-config
ip prefix-list in-filter deny 172.16.0.0/12 le 32 ! RFC1918
ip prefix-list in-filter deny 192.0.2.0/24 le 32  ! TEST1
ip prefix-list in-filter deny 192.168.0.0/16 le 32 ! RFC1918
ip prefix-list in-filter deny 198.18.0.0/15 le 32 ! Benchmarking
ip prefix-list in-filter deny 198.51.100.0/24 le 32 ! TEST2
ip prefix-list in-filter deny 203.0.113.0/24 le 32 ! TEST3
ip prefix-list in-filter deny 224.0.0.0/3 le 32   ! Multicast
ip prefix-list in-filter deny 0.0.0.0/0 ge 25     ! Prefixes >/24
ip prefix-list in-filter permit 0.0.0.0/0 le 32
```

Réception des préfixes IPv6

```
router bgp 100
  network 2020:3030::/32
  neighbor 2020:3030::1 remote-as 101
  neighbor 2020:3030::1 prefix-list v6in-filter in
!
ipv6 prefix-list v6in-filter permit 2001::/32          ! Teredo
ipv6 prefix-list v6in-filter deny 2001::/32 le 128     ! Teredo subnets
ipv6 prefix-list v6in-filter deny 2001:db8::/32 le 128 ! Documentation
ipv6 prefix-list v6in-filter permit 2002::/16          ! 6to4
ipv6 prefix-list v6in-filter deny 2002::/16 le 128     ! 6to4 subnets
ipv6 prefix-list v6in-filter deny 2020:3030::/32 le 128 ! Local Prefix
ipv6 prefix-list v6in-filter deny 3ffe::/16 le 128     ! Old 6bone
ipv6 prefix-list v6in-filter permit 2000::/3 le 48     ! Global Unicast
ipv6 prefix-list v6in-filter deny ::/0 le 128
```

Réception des préfixes

- Attention aux préfixes provenant des clients, des pairs et des fournisseurs de transit
- Responsabilité de tous les FAI d'être de bons citoyens d'Internet

Préfixes dans iBGP



Injection de préfixes dans iBGP

- Utiliser iBGP pour transporter les préfixes des clients
 - Ne pas utiliser IGP
- Pointer une route statique vers l'interface client
- Utiliser l'instruction network du BGP
- Aussi longtemps que la route statique existe (interface active), le préfixe sera dans BGP

Configuration du Routeur: l'instruction du réseau

- Exemple:

```
interface loopback 0
  ip address 215.17.3.1 255.255.255.255
!
interface Serial 5/0
  ip unnumbered loopback 0
  ip verify unicast reverse-path
!
ip route 215.34.10.0 255.255.252.0 Serial 5/0
!
router bgp 100m
  network 215.34.10.0 mask 255.255.252.0
```

Injection de préfixes dans iBGP

- L'instabilité de l'interface se traduira par le retrait, l'annonce à nouveau du préfixe
 - Utiliser "IP route . . . permanent"
- De nombreux fournisseurs de services Internet redistribuent des routes statiques dans BGP, plutôt que d'utiliser l'instruction network
 - Faites seulement ceci si vous comprenez pourquoi

Configuration du Routeur: redistribuer statique

- Exemple:

```
ip route 215.34.10.0 255.255.252.0 Serial 5/0
```

```
!
```

```
router bgp 100m
```

```
  redistribute static route-map static-to-bgp
```

```
<snip>
```

```
!
```

```
route-map static-to-bgp permit 10
```

```
  match ip address prefix-list ISP-block
```

```
  set origin igp
```

```
<snip>
```

```
!
```

```
ip prefix-list ISP-block permit 215.34.10.0/22 le 30
```

Injection de préfixes dans iBGP

- La Route Map peut être utilisée pour beaucoup de choses:
 - régler les communautés et d'autres attributs
 - régler le code d'origine du IGP, etc
- Être prudent avec les sous-préfixes et les route-map
 - l'absence de l'un ou des deux signifie que tous les préfixes statique vont dans iBGP

Résumé

- Meilleures pratiques couvertes:
 - Quand utiliser BGP
 - Quand utiliser ISIS/OSPF
 - Agrégation
 - Réception des préfixes
 - Préfixes dans BGP

Meilleures pratiques courantes de BGP



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