ОНЛАЙН-ОБРАЗОВАНИЕ

Маршрутизация в сетях IPv4

Роутинг, рутинг, раутинг - поспорим как правильно и немного поговорим о сути

Александр Румянцев

Numbering & Renumbering

https://tools.ietf.org/html/rfc1918 https://tools.ietf.org/html/rfc6890

Серые сети

- `10.0.0.0/8`
- `172.16.12.0/12`
- `192.168.0.0/16`

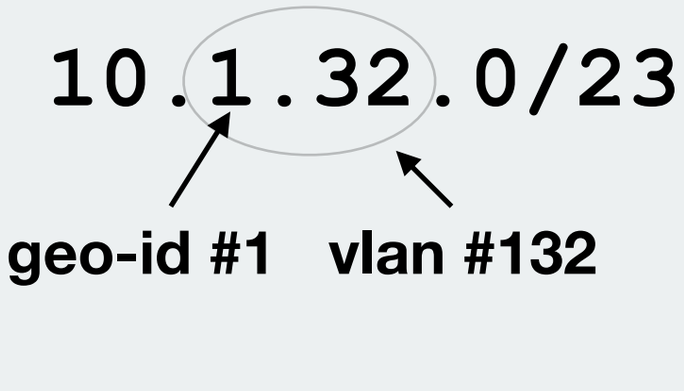
Специальные сети

- `0.0.0.0/8` - **этот хост как источник**
- `127.0.0.0/8` - **этот хост как приемник, loopback**
- `100.64.0.0/10` - **Shared Net - NAT на уровне провайдера**
- `169.254.0.0/16` - **Link Local - сеть для самоконфигурирующихся устройств**
- `240.0.0.0/4` - **Multicast**

13,8 % адресов IPv4 так или иначе зарезервированы

Numbering & Renumbering

Рекомендуемая схема нумерации для сетей



Кстати, VLANы лучше нумеровать кратно 4 или для наглядности, кратно 10 - с учетом возможного расширения сети и geo-id

Агрегаты, специфики и Null routing

Агрегат - минимально возможная сеть, включающая в себя все специфики

`10.0.0.0/13` - **Агрегат**
`10.1.0.0/16` - **Специфик**
`10.2.0.0/16` - **Специфик**
`10.6.20.0/22` - **Специфик**

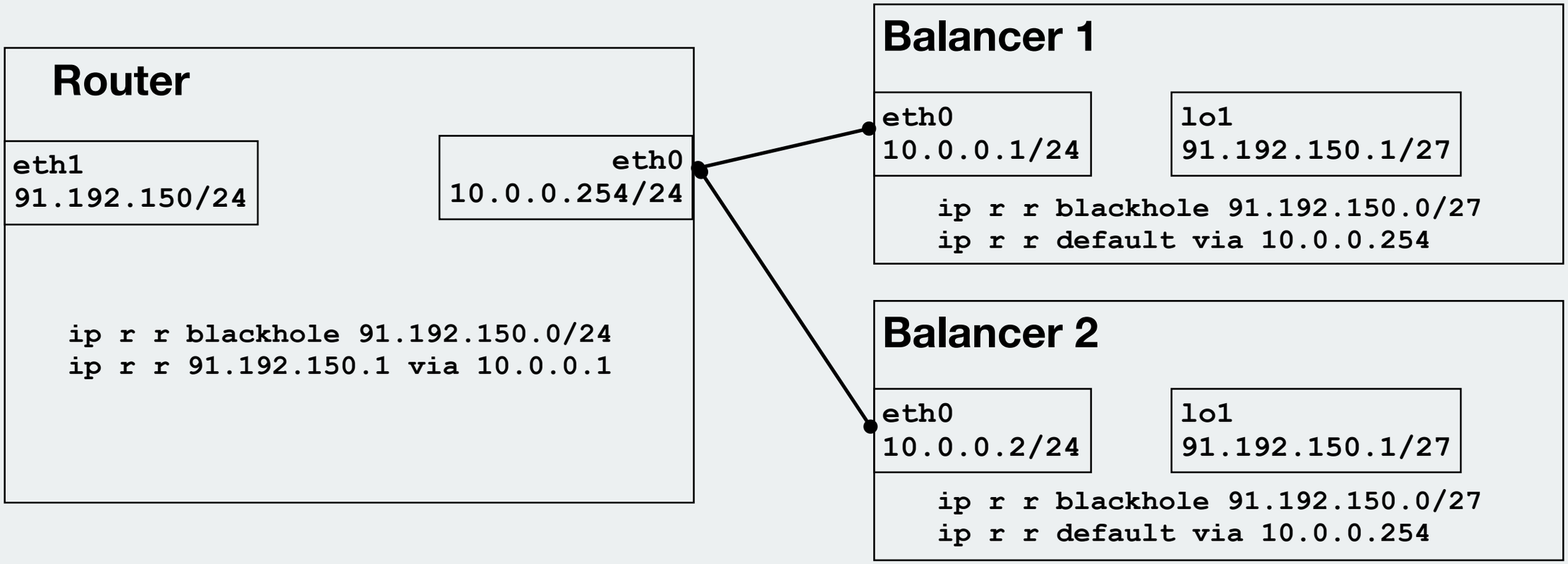
http://jodies.de/ipcalc

В роутинге специфики всегда выигрывают при поиске в FIB
Агрегаты всегда должны "заворачиваться в null", что бы избегать проблем с кольцами, броадкастами и мусорным трафиком в маршрутизации

```
ip route add blackhole 10/8
```

Кроме того, null routing - самый дешевый способ фильтрации трафика

Неочевидные возможности loopback



Динамическая маршрутизация

Автоматическое получение маршрутов и возможность их перестроения при изменении топологии

Interior Gateway Protocols

- Open Shortest Path First

- Enhanced Interior Gateway Routing Protocol

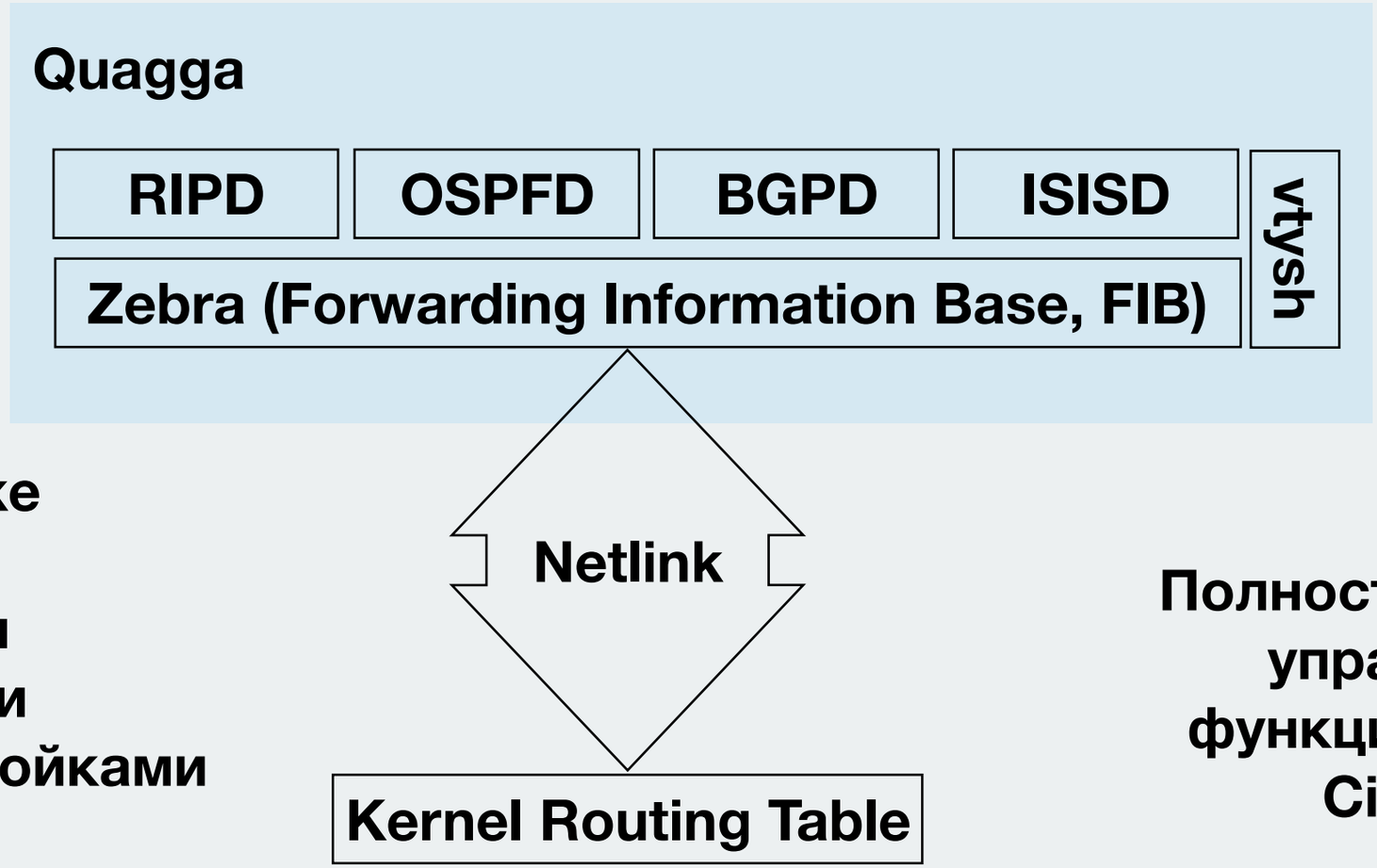
Работают на уровне L3, каждый роутер знает всю топологию своей области, отслеживается состояние соединений

Border Gateway Protocol

Работает на уровне L4 (возможно управление маршрутизаторами с отдельно-стоящего BGP Route Server), состояние пиринга определяетя по состоянию TCP-сессии. Может использоваться внутри AS/Area

Quagga (ex-Zebra)

Набор демонов, реализующий управление сетью в linux



Возможно так же управление интерфейсами и разнообразными сетевыми настройками

Полностью копирует управление и функциональность Cisco IOS

OSPF Terms

Area - интерфейсы, роутеры которых получают всю информацию о топологии сети. Для небольшой сети возможно использование единственной Area 0. Есть разновидности - backbone, standard, stub, not-so-stubby area и еще несколько

Hello - пакеты, устанавливающие и поддерживающие соеденение роутеров
LSA (Link State Announce) - единица обмена информацией между роутерами OSPF (https://habr.com/post/201794/)
BR - Border Router
ABR (ospf term) - Area Border Router
ASBR (bgp term) - Autonomous System Border Router
DR - Designated Router - ABR, являющийся основной точкой входа в Area
BDR - Backup Designated Router - ABR, являющийся резервной точкой входа
DROther - BR, не участвующий в построении графа маршрутов

OSPF Neighbourhood

Состояния соседства маршрутизаторов

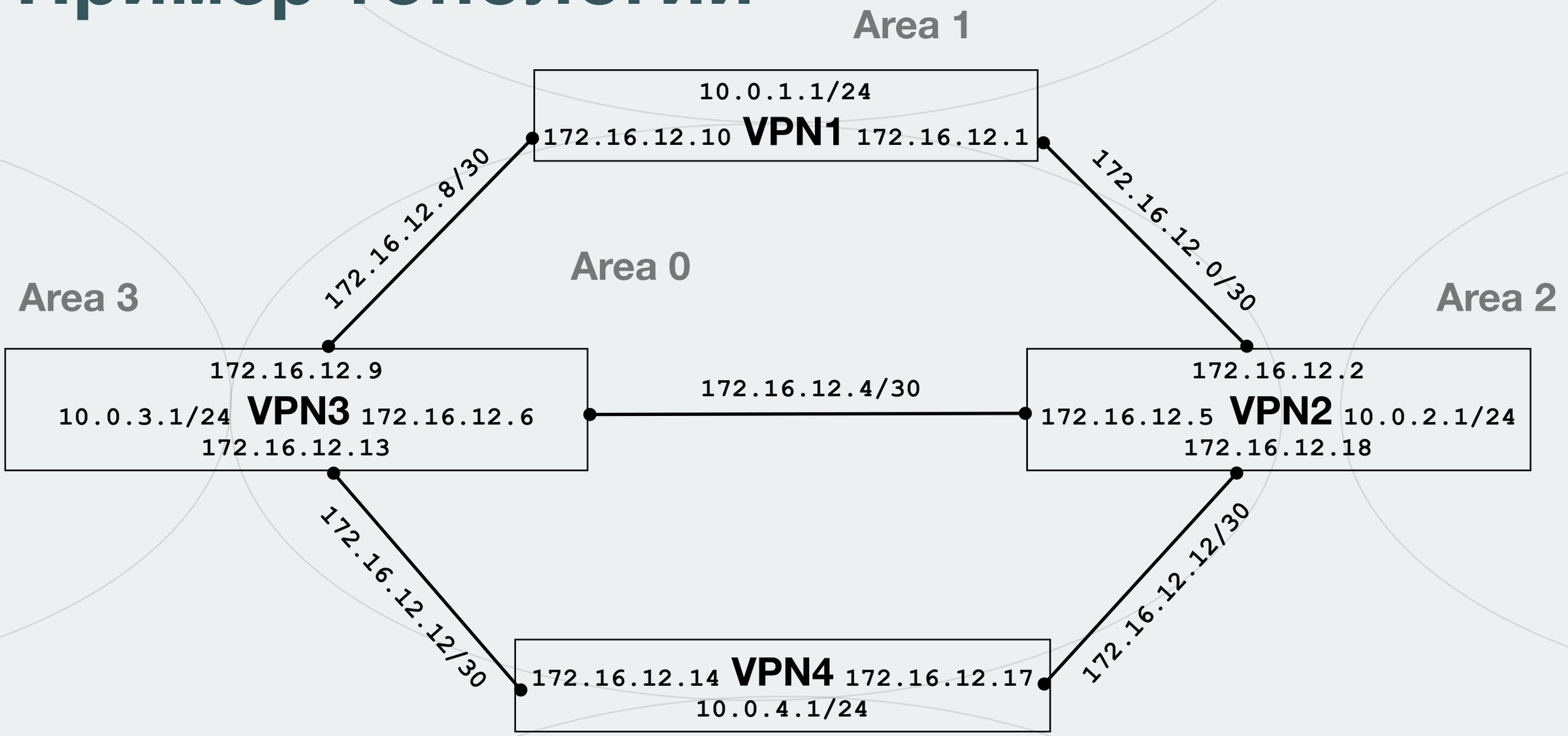
Down - соседа не видно
Attempt, Init - попытка установить соединение
2-Way - роутеры видят друг-друга
Exstart, Exchange, Loading aka Adjacent - роутеры обмениваются информацией
Full aka Full Adjacent - роутеры обменялись информацией

OSPF interface

Priority: DR/BDR election
Если в одну сеть есть два входа через два роутера, то один из них становится DR, остальные - BDR. На принятие решения влияет priority и router-id. Если по каким-то причинам роутеры не могут выбрать DR, то состояние их соседства повиснет в 2way/DRother.
Специальное значение priority 0 - отказ от выборов.

Cost - стоимость прохождения через этот интерфейс. По-умолчанию рассчитывался исходя из пропускной способности интерфейса. Теперь неактуально, потому что при bw ≥ 100 Mbps default cost = 1. Если есть два маршрута в одну сеть с одинаковой суммарной стоимостью - устанавливаются оба, в противном случае устанавливается только один с минимальной стоимостью. Параметр интерфейса роутера, а не направления в целом, поэтому для понижения приоритета всего линка должен быть установлен с обоих сторон.

Пример топологии



Linux rp_filter

https://www.kernel.org/doc/Documentation/networking/ip-sysctl.txt

По умолчанию, параметр установлен в 1, что обозначает строгую проверку маршрута источника. Это правильно до тех пор, пока мы не сталкиваемся с динамическим роутингом. При использовании IGP можно сразу выставить в 2 (нестрогая проверка) или в 0 (отсутствие проверки)

BGP

Протокол обмена маршрутной информацией уровня провайдера. Объединяет т.н. Автономные Системы (Autonomous System, AS)

AS похож на OSPF Area. Подразумевается, что все сети внутри AS так или иначе объединены с помощью IGP

В некоторых случаях BGP может использован внутри AS. Т.к. BGP в основном означает связь с провайдером, то для внутренних связей используют термин iBGP

Автономную систему хорошо иметь для независимости от ДЦ и в случае нескольких ДЦ. Минимальный размер сети, аннонсируемой через BGP не регламентируется протоколом, но в реальной жизни - /24. Поэтому минимальный размер AS тоже /24. Раньше адреса можно было получить непосредственно у RIPE, теперь ищите LIR, у которых еще остались блоки адресов.

BGP

RIPE хранит всю информацию о связях AS. Её со своей стороны надо всегда держать в актуальном состоянии, потому что есть много транзитных операторов, которые используют эту базу для конфигурации устройств в автоматическом режиме

Типичный сценарий использования: вы аннонсируете провайдеру свои сети, провайдер вам в ответ аннонсирует 0.0.0.0/0.

При пиринге всегда нужно использовать фильтрацию префиксов, что бы ни вы, ни провайдер не отослали лишнего

Вспомогательные инструменты BGP

whois - интерфейс коммандной строки ко всем базам координационных центров. Россия относится к RIPE

Looking glass - Инструмент ограниченного доступа к роутерам провайдера через web-интерфейс. Например http://lg.mtu.ru/cgi-bin/lgform_img.cgi

Спасибо за внимание!

Вопросы?