

Cisco Expo 2011



Нехватка IPv4 адресов – практический подход к решению от Cisco Systems

Денис Коденцев
dkodents@cisco.com

Системный инженер Cisco

innovate *together*

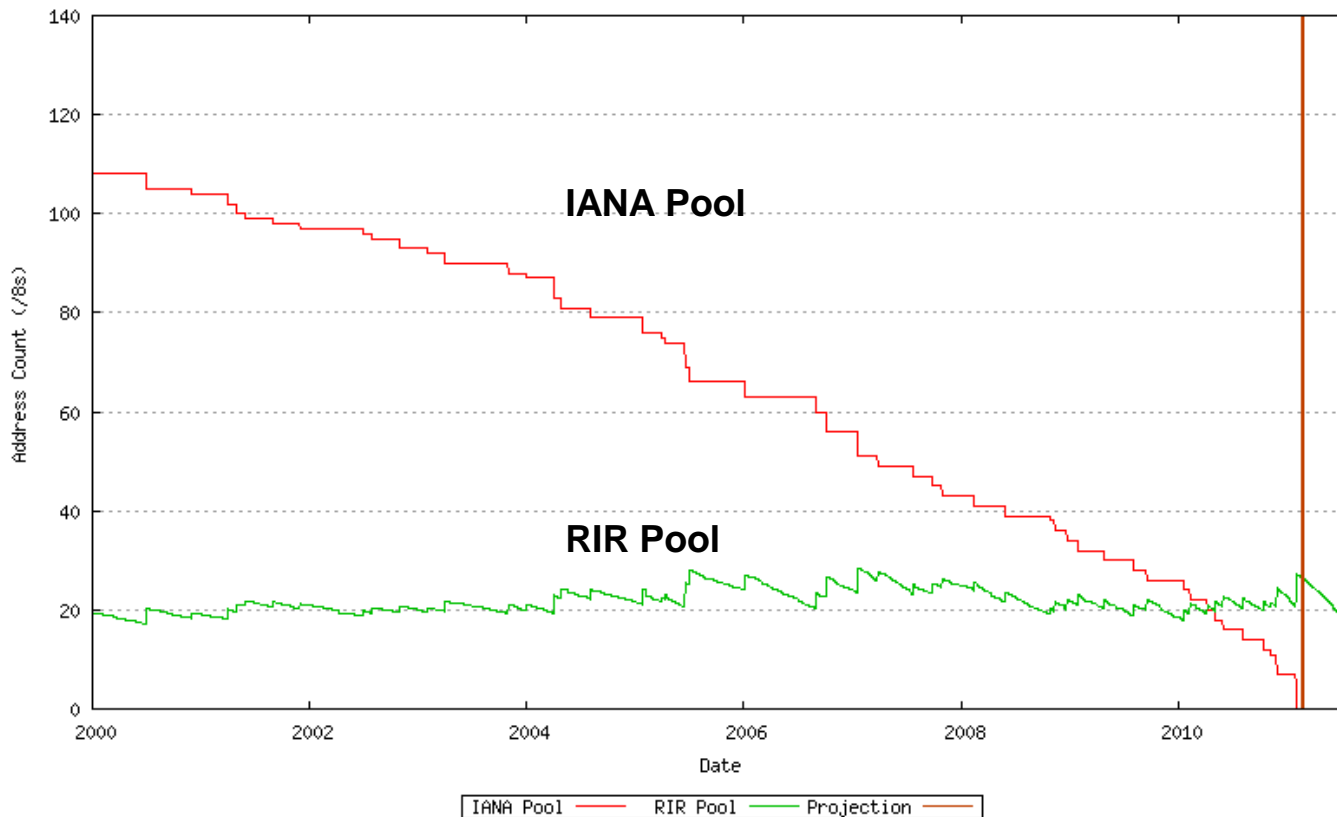
Что предлагается обсудить

- Проблема – наступила эпоха дефицита IPv4 адресов!
- Что такое операторский NAT?
- Реализация в решениях Cisco
- Схемы резервирования и результаты испытаний



Исчерпание IPv4 адресов

IANA IPv4 пулы адресов закончились в Феврале 2011
RIR пулы адресов могут закончиться уже в Августе....



IPv4 & IPv6 Statistics

RIR v4 /24s Left

AfrNIC 267,258

APNIC 317,811

ARIN 549,736

LACNIC 262,793

RIPE 327,431

v6 ASNs

8% (3,289/37,090)

v6 Ready TLDs

83% (256/306)

v6 Glues

4,628

v6 Domains

1,467,051

0

days remaining
IANA exhausted

HURRICANE ELECTRIC
INTERNET SERVICES

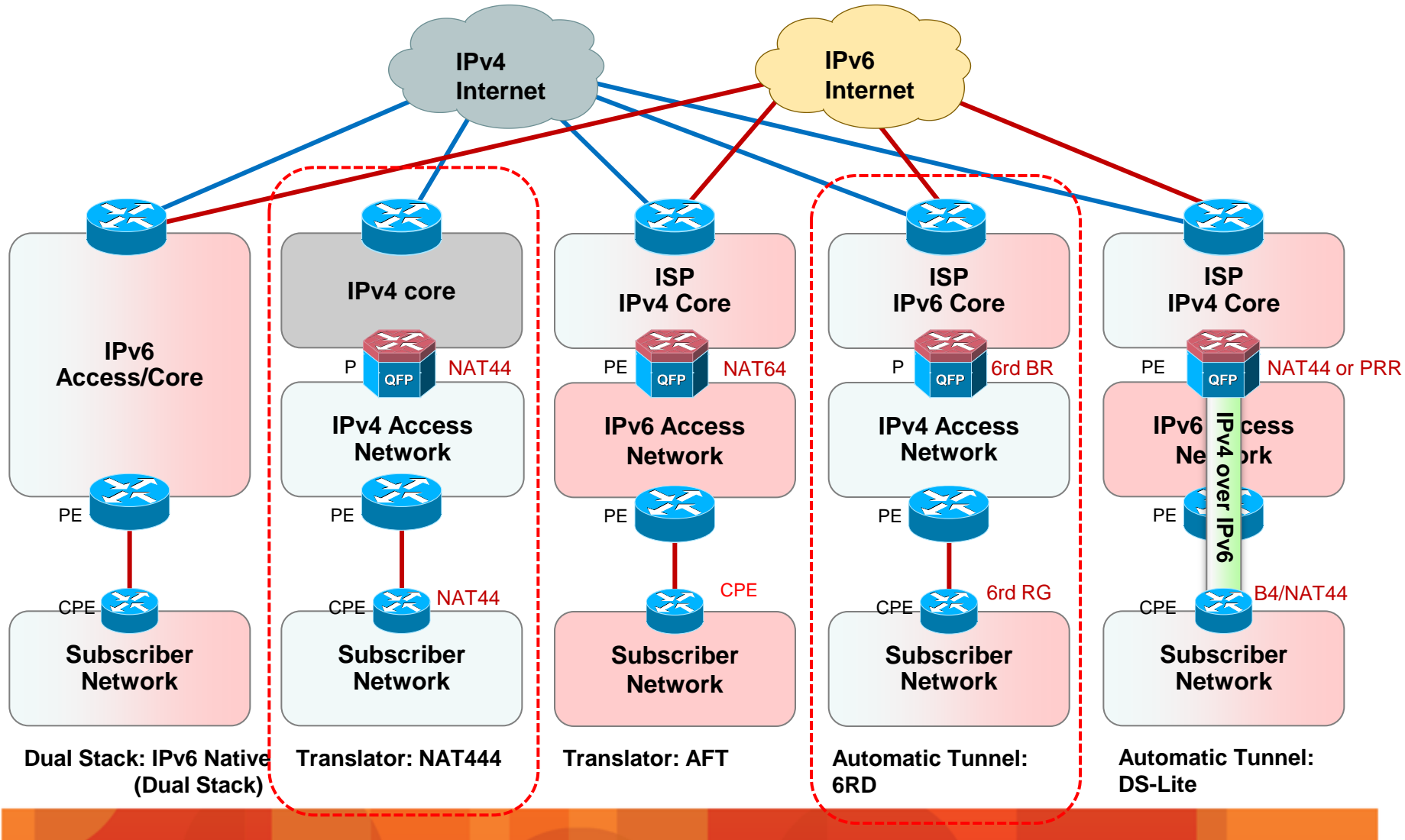
Главные задачи

Решение Cisco Carrier-Grade NAT

- **Сегодня** – Carrier Grade NAT44
- **Завтра и ближайшие годы** – IPv6 тунелирование, IPv4/IPv6 трансляция, Dual-Stack
- **В будущем** – Облачные услуги, ЦОД, Сенсоры, Smart Grid и прочие услуги на IPv6



Возможные сценарии



Что такое операторский NAT?



Операторский NAT – в чем отличия?

- По сути, это обычный NAT44
 - Только большой и операторского класса
- Требуются «операторские» функции, например:
 - Ограничение кол-ва трансляций на абонента
 - Слишком малое число – проблемы с приложениями
 - Google maps – классический пример
 - Протоколирование трансляций
- Как нумеровать пространство между абонентом и LSN?
 - RFC1918 может конфликтовать с домашней адресацией абонента



CGN – Требования к NAT44

- NAT Функционал (RFC4787, RFC5382, RFC5508)
 - Endpoint Independent Mapping/Filtering
 - Paired IP address pooling behavior
 - Port Parity preservation for UDP
 - Hairpinning behavior
 - Static Port Forwarding
 - Active FTP ALG
- Управление
 - Port Limit per subscriber
 - Mapping Refresh
 - NAT logging (важно для СОРМ)
- Отказоустойчивость (Intra-box Active/Standby, Inter-box Active/Active)



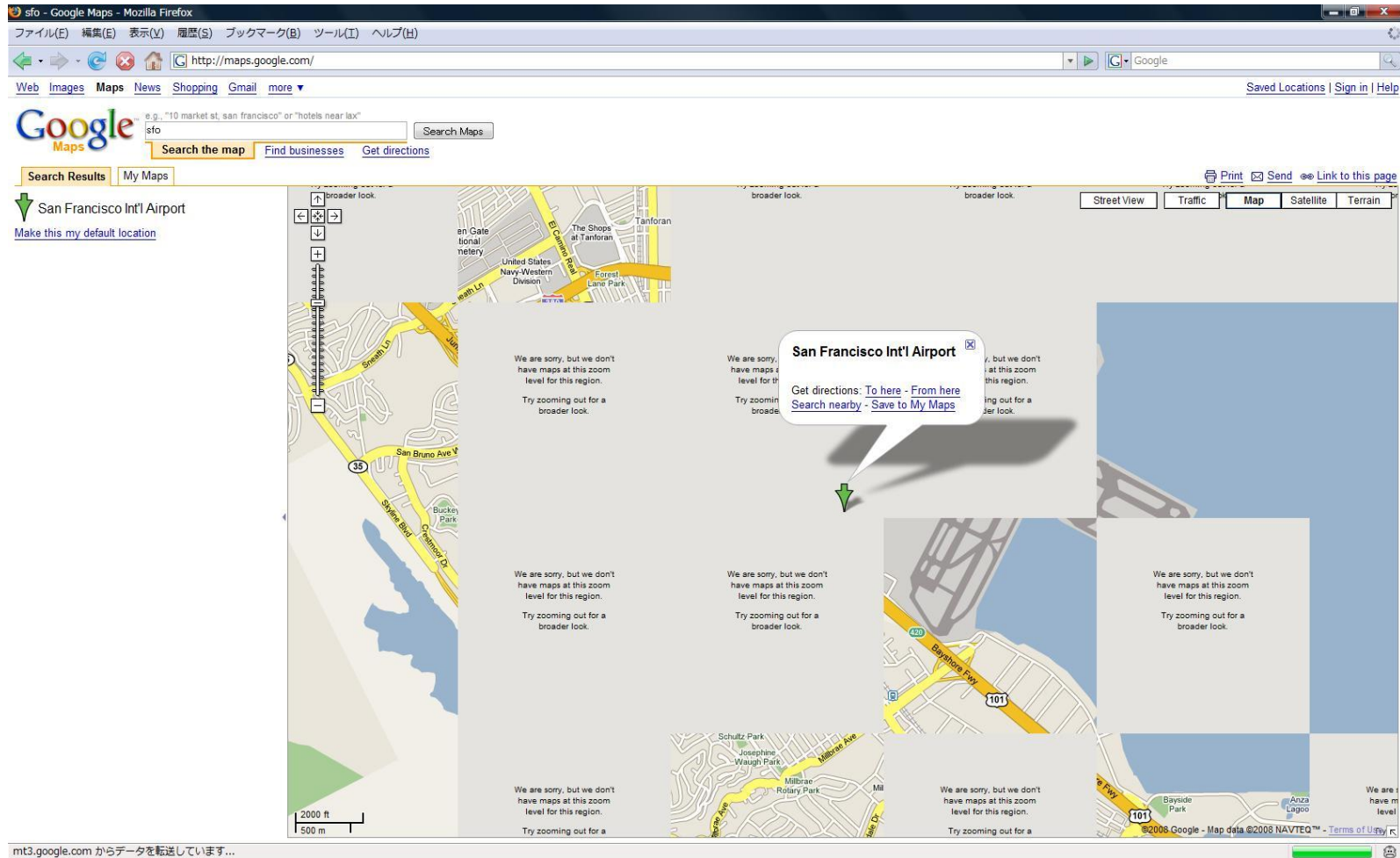
NAT: проблемы и их решения



Дополнительные требования к NAT44 оборудованию

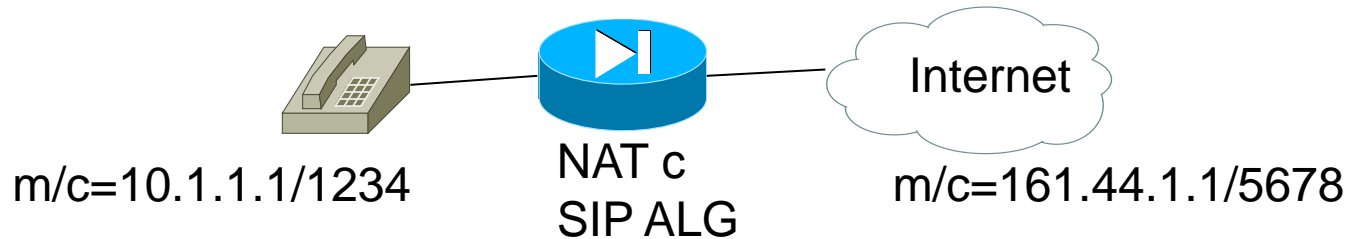
- Известная проблема Large Scale NAT – нехватка хлате на абонента
- Контекст абонента. Например:
 - Rapidshare – ограниченное кол-во скачиваний с одного IP в сутки
- COPM
 - Какой абонент запостил сообщение на www.example.com в 20 час. 34 мин.?
 - LSN должен протоколировать трансляции
 - WWW сервер должен протоколировать номера портов

Google maps – слишком мало трансляций на абонента



Application Layer Gateway (ALG)

- Осведомленность NAT механизма о транслируемом трафике приложений
- Функции ALG:
 1. Модификация IP адресов и портов внутри передаваемых данных приложений
 2. Создание NAT соответствий
- Каждое приложение требует отдельного ALG алгоритма
 - FTP, SIP, RTSP, RealAudio,...



Проблемы с ALG механизмом

- ALG необходим для каждого приложения
- Для каждого приложения, его версии необходим специфичный ALG механизм
 - Частные расширения, отклонения от стандартов реализации;
 - Новые версии протоколов
- ALG требует для своей работы:
 - Нешифрованную сигнализацию (!!);
 - Видеть и сигнализацию, и трафик данных приложений
 - Что просто для сетей с одним подключением, но значительно усложняется в случае нескольких соединений с Internet, балансировки нагрузки



Альтернативы

- FTP PASV, дата соединения всегда создается в сторону сервера
- ICE, STUN, TURN
 - Протоколы, добавляющие средства обнаружения/обхода NAT клиентам;
 - Используются «запрос/ ответ» протоколами (SIP, XMPP и т.д.);
 - Стандартизуются в рамках работ MMUSIC и BEHAVE IETF рабочих групп;
- RTSPv1, эффективно заменяется Flash поверх HTTP
- RTSPv2, ICE подобное решение
- Skype, криптозащищенная передача с собственными средствами обхода NAT



Session Traversal Utilities for NAT (STUN), ICE, TURN

- Протокол «запрос/ ответ» используется:
 - Самим STUN (узнать внешний IP адрес и порт);
 - ICE (для проверки связности);
 - TURN (для конфигурации TURN сервера);
- Ответ содержит IP адрес и порт пришедшего запроса
 - Работает поверх UDP (обычно) или TCP, порт 3478
- Аналогичен <http://whatismyip.com>

STUN, Interactive Connectivity Establishment (ICE), TURN

- Процедура оптимизации потоков передачи медиа данных
- Определяет SDP синтаксис для индикации 'адреса кандидата'
- Использует STUN сообщения для проверки СВЯЗНОСТИ
 - Посылаются RTP соседу, используя **те же самые порты**, что и передаваемый RTP трафик
- Выбирается первый работающий путь
- Уроженно, шаги:
 - 1.Собрать все свои IP адреса (внешние)
 - 2.Послать их соседу
 - 3.Проверить связность

Реализации ICE

- Google chat (XMPP)
- Microsoft MSN (SIP inside of XML)
- Yahoo (SIP)
- Counterpath softphone (SIP)



STUN, ICE, Traversal Using Relays around NAT (TURN)

- Media Relay Protocol и Media Relay Server
- Используются в случае:
 - Обе стороны находятся за ‘Address and Port-Dependent Filtering NAT’ согласно RFC 4787(встречается редко, менее 25% NAT инсталляций), или
 - Одна из сторон не реализует ICE и находится за ‘Address and Port-Dependent Filtering NAT’

NAT64 - AFT



Address Family Translation (AFT)

Терминология

- **Stateful AFT64**
 - Для каждого потока/ flow создается отдельное состояние;
 - Поддерживаются только IPv6 инициируемые соединения;
 - Количество состояний - $O(\# \text{ трансляций})$;
 - N:1 соответствие (как и в случае NAPT для NAT44) (1:1 соответствие также безусловно возможно)
- **Stateless AFT64**
 - Поток/ flow не создает никаких состояний;
 - Алгоритмическая операция, проводимая над заголовком пакетов;
 - 1:1 соответствие (один IPv4 адрес используется для каждого IPv6 хоста) ;
 - Поддерживаются как IPv6, так и IPv4 инициируемые соединения

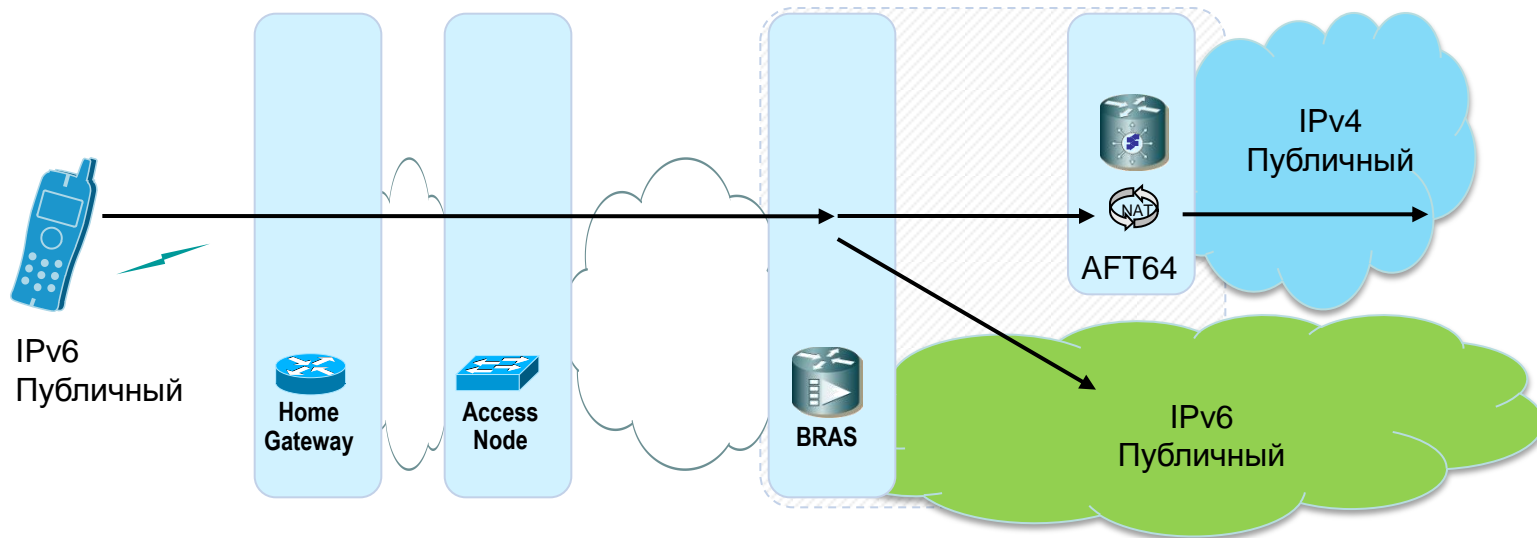
[draft-ietf-behave-v6v4-xlate-stateful](#)

[draft-ietf-behave-v6v4-xlate](#)

IPv6-only абоненты:

взаимодействие с IPv4 Internet

- AFT64 технология применима в случае, когда IPv6 хосты общаются с IPv4 хостами
- AFT64 для трафика из IPv6 области в IPv4



- AFT64:= “stateful v6 to v4 translation” или “stateless translation”
 - Рекомендации draft-ietf-behave-v6v4-framework, draft-ietf-behave-v6v4-xlate, draft-ietf-behave-v6v4-xlate-stateful

CGN в продуктах Cisco Systems



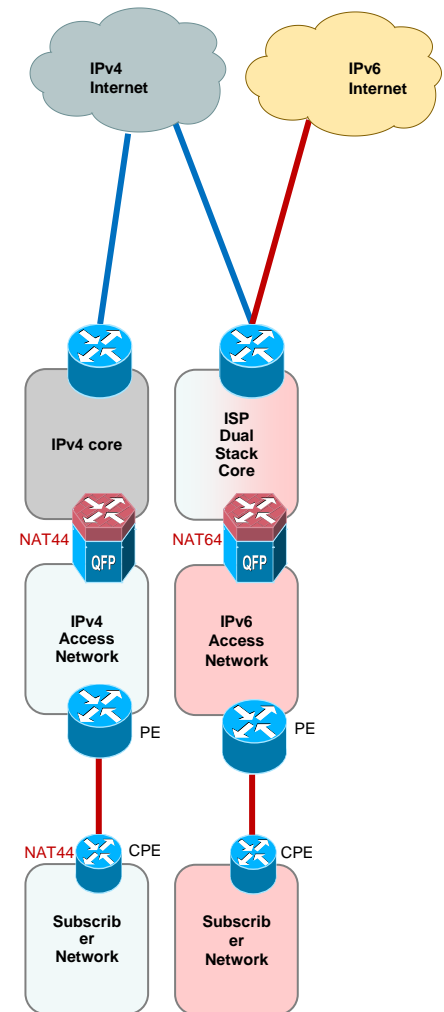
CGN функционал ASR1000

- Поддерживается сейчас (IOS XE 3.2S)
 - Dual-stack (PPP sessions)
 - NAT 44
 - 6rd
 - L2TP (IPv4 и IPv6
oPPPoL2TP)
 - 6PE
 - 6VPE
 - NAT64 – Stateless
- Поддержка в следующих версиях
 - ISGv6 (3.4S)
 - Stateful NAT64 (3.4S)
 - DS-Lite



ASR1000 Carrier Grade NAT

- **Сценарии:**
 - NAT44/NAT444,
 - NAT64 stateless
 - NAT64 stateful (3.4S)
- **Ограничение кол-ва трансляций для одного абонента**
- **Протоколирование трансляций посредством Netflow**
- **Inter-VRF NAT через VASI**
- **Резервирование:**
 - Intra-box Active/Standby,
 - Inter-box Active/Active
- **Производительность (для ESP40):**
 - 40Gbps
 - 2М трансляций
 - 270К трансляций в секунду



Method2: NAT44 or NAT444
Method3: NAT64

Carrier-Grade Services Engine (CGSE)

Решение для **крупных** внедрений Cisco CGN



Cisco CGSE

- 20+ миллионов активных трансляций
 - 100ни тысяч абонентов
- 1+ миллион соединений в секунду
 - 20Гбит/с на модуль **CGSE**

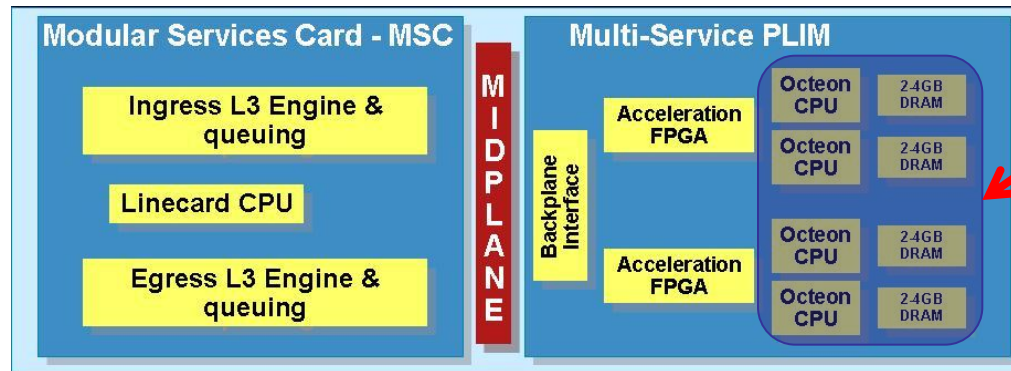


Cisco CRS

CGSE PLIM и CGN

- **Аппаратное обеспечение**

- CGv6 функционал выполняется на CGSE PLIM
- Многоядерная архитектура Quad Oocteon, 64 ядра
- Стандартный интерфейс к MSC, производительность 20 Гбит/с



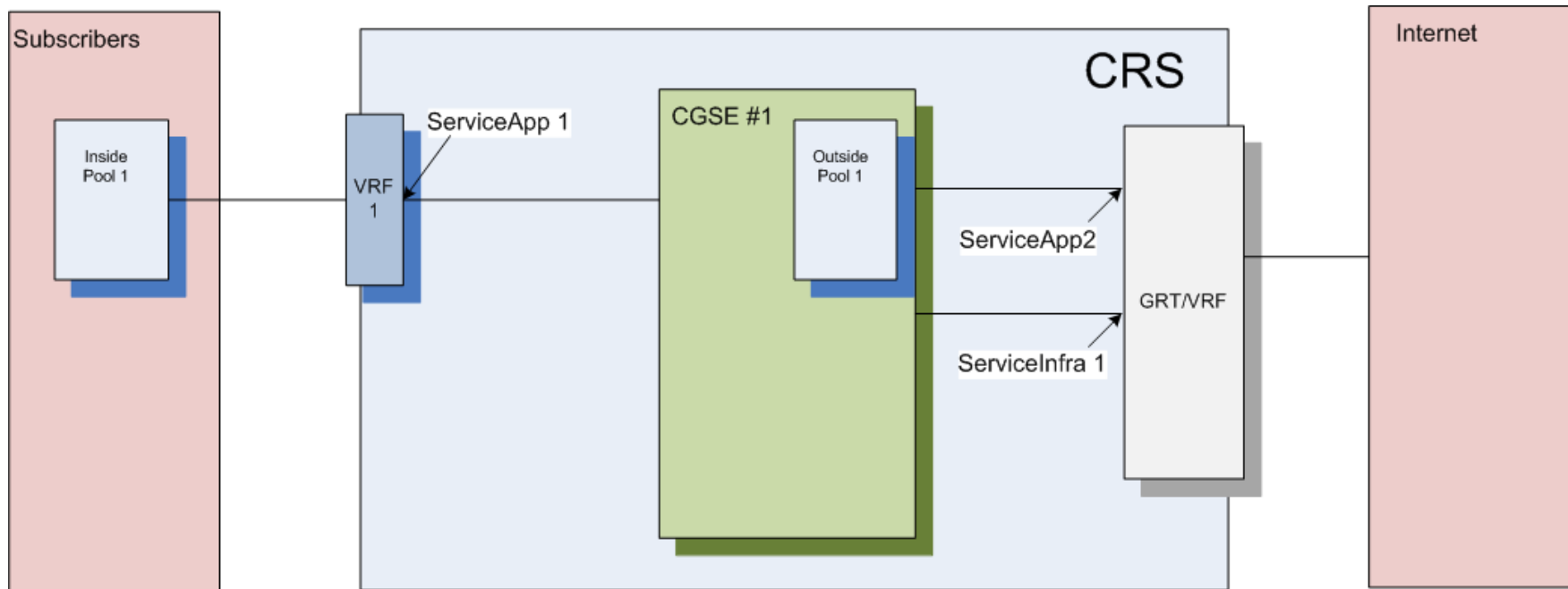
CGN и будущие приложения

- **Программное обеспечение**

- IOS-XR на MSC, Linux на Oocteon CPU
- Integrated configuration & management via IOS XR

CGSE – основы архитектуры

- Взаимодействие с «внешним миром» через логические интерфейсы ServiceApp и ServiceInfra
- Inside ServiceApp обязательно в VRF
- Outside ServiceApp в GRT/VRF
- В CLI указывается только Outside pool
- ServiceInfra для экспорта Netflow



CGSE Производительность и масштабируемость

Функциональность NAT44, 6RD в IOS XR 3.9.1



- **20+ миллионов** активных трансляций
- **1+ миллион** соединений в секунду
- **20 Гбит+** пропускная способность одного CGSE
- Управление
 - Ограничение числа портов на частный IPv4 адрес
 - TCP/UDP/ICMP таймеры сессий
 - Netflow логирование
- Резервирование
 - 1:1 Active/Standby (Warm)
- Дополнение
 - CGN Bypass
 - Балансировка нагрузки

CGSE и COPM



- Для IPv6 абонентов с 6rd прямое соответствие IPv4 адреса и 6rd IPv6 снимает все вопросы
- Сегодняшняя система COPM не требует модернизации
- Для IPv4 абонентов с NAT44 требуется снимать логи всех трансляций, что решается с помощью оптимизированного Netflow v9
- Сегодняшняя система COPM не требует модернизации

Netflow v9 логирование NAT

Template 256 (создание соединения)

Field ID	Attribute	Value
234	Incoming VRF ID	32 bit ID
235	Outgoing VRF ID	32 bit ID
8	Source IP Address	IPv4 Address
225	Translated Source IP Address	IPv4 Address
7	Source Port	16 bit port
227	Translated Source Port	16 bit port
4	Protocol	8bit value

Template 257 (удаление)

Field ID	Attribute	Value
234	Incoming VRF ID	32 bit ID
8	Source IP Address	IPv4 Address
7	Source Port	16 bit port
4	Protocol	8bit value

Netflow Объем данных

```
[root@lulycgse cgn-test2]# ls -l
total 164
-rw-r--r-- 1 root root 62071 Mar 16 21:48 172.16.1.1.2148 (1000 add-event)
-rw-r--r-- 1 root root 36486 Mar 16 21:51 172.16.1.1.2151 (1000 delete-event)
-rw-r--r-- 1 root root 6417 Mar 16 22:00 172.16.1.1.2200.gz (1000 add-event)
-rw-r--r-- 1 root root 3354 Mar 16 22:03 172.16.1.1.2203.gz (1000 delete-event)
-rw-r--r-- 1 root root 31044 Mar 16 22:23 172.16.1.1.2223.gz (5000 add-event)
-rw-r--r-- 1 root root 16139 Mar 16 22:27 172.16.1.1.2227.gz (5000 delete-event)
```

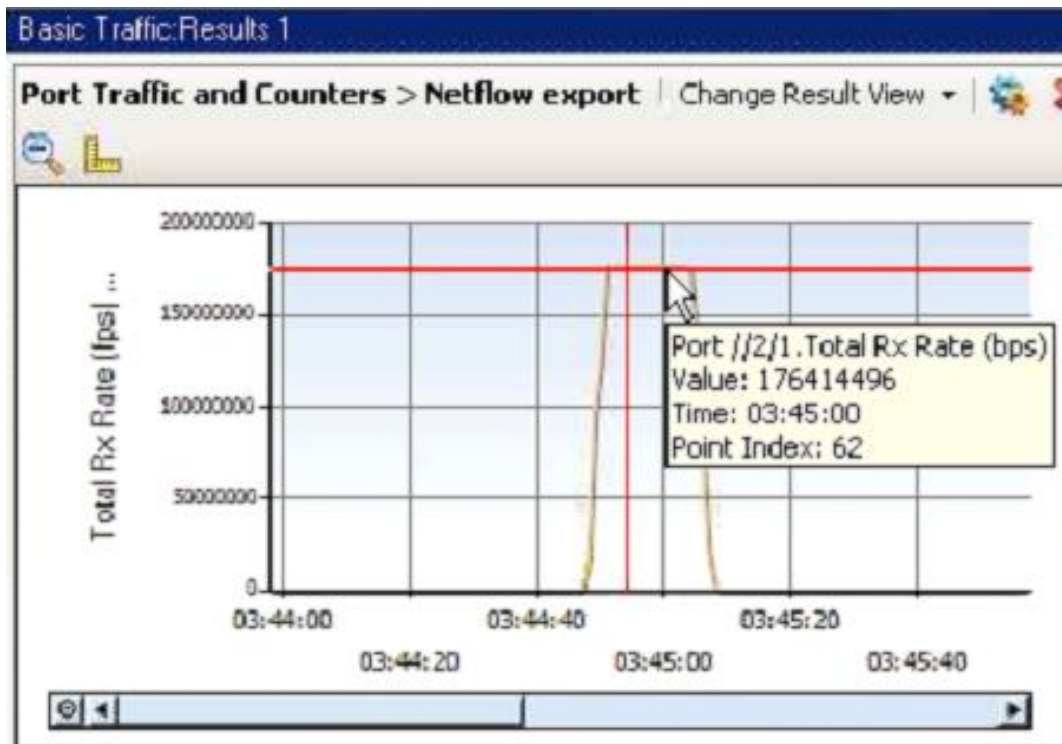
Cisco NFC 6.0.0

В среднем 6.5 байт при создании и 3.5 при удалении сессии (компрессия)

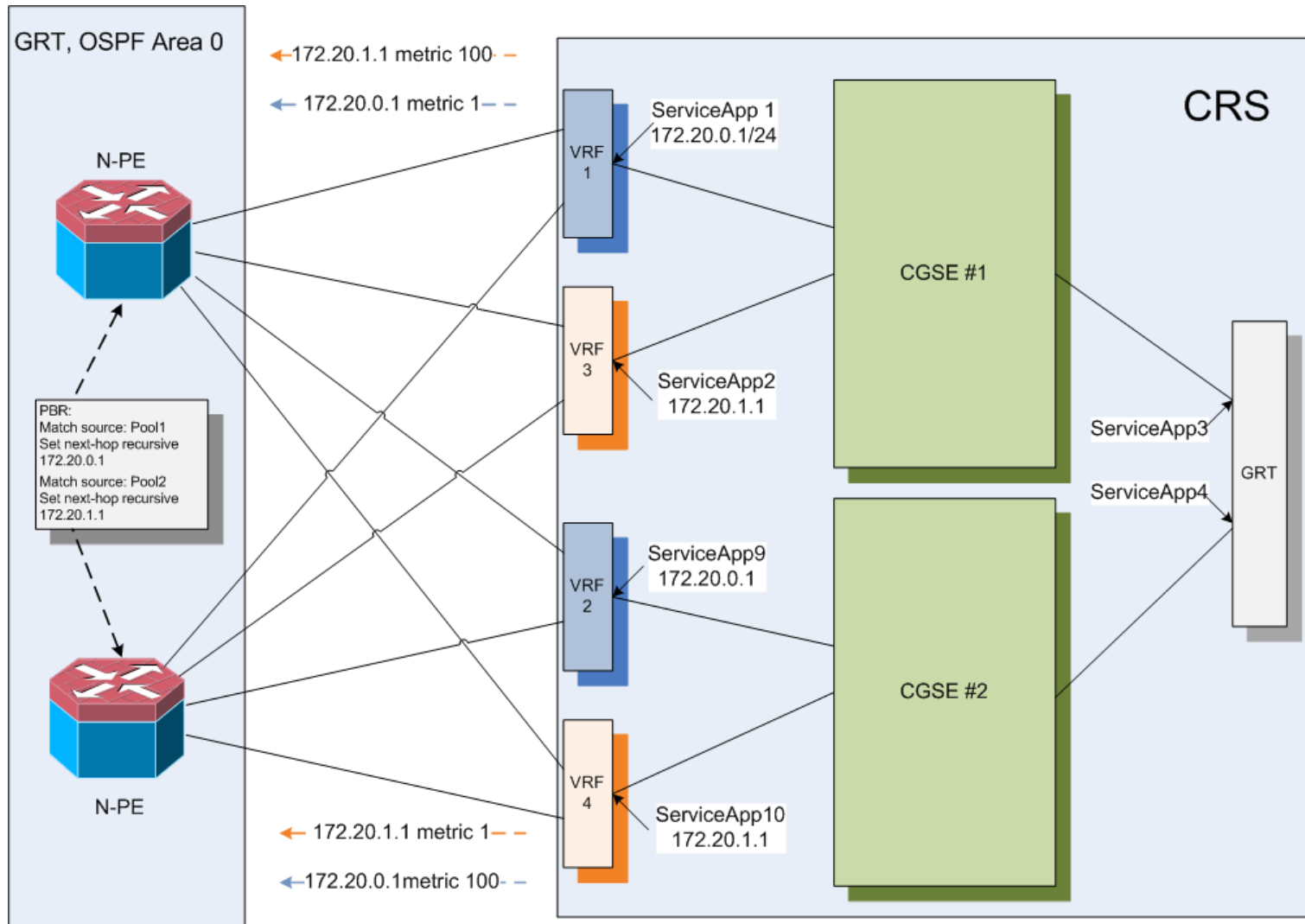
Количество трансляций	Размер на диске
500k	5 Мбайт
432М (5k в сек. * 24 часа)	4 Гбайт
4320М (50k в сек. * 24 часа)	43 Гбайт
43200М (500k в сек. * 24 часа)	432 Гбайт

Netflow Объем данных

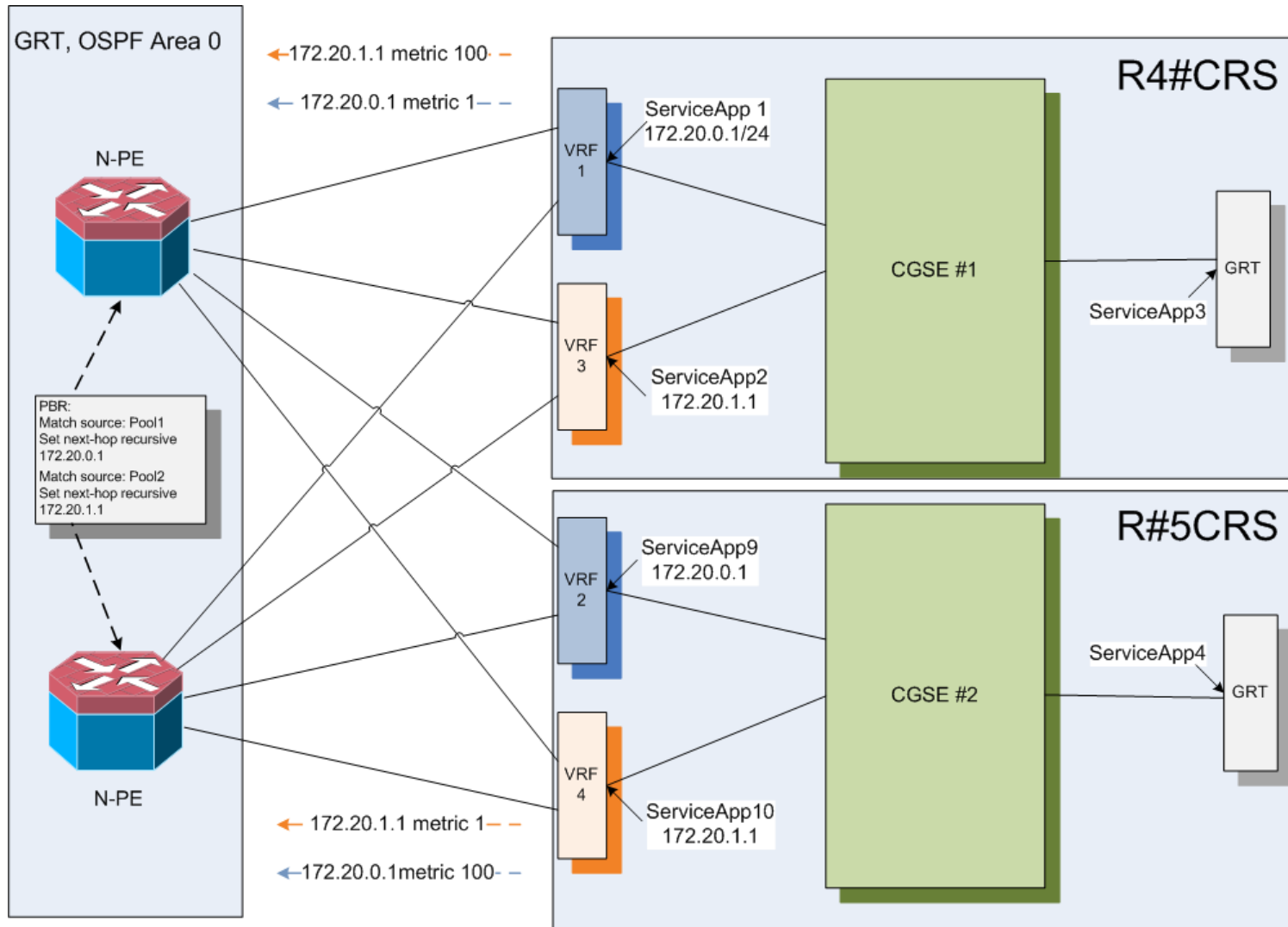
- создание 1М трансляций в секунду
- 176Mbps Netflow экспорта



Резервирование active-active - Intra-chassis



Резервирование active-active - Inter-chassis



Вариант внедрения Cisco CGN

- PBR конструкция на PE маршрутизаторе:

```
route-map TO_CGSE permit 10
  match ip address 101
  set ip next-hop recursive 172.20.0.1
!
route-map TO_CGSE permit 20
  match ip address 102
  set ip next-hop recursive 172.20.1.1
!
```

- Изменения/сбои на CGN узле не влияют на IGP в сети оператора



Заключение



IPv6 в ВАШЕЙ IPv4 сети?

- Легко проверить!
- С помощью IPv4 NetFlow
 - Protocol 41: IPv6 поверх IPv4 или 6to4 туннели
 - IPv4 address: 192.88.99.1 (6to4 anycast server)
 - UDP 3544, - Teredo
- 'ISATAP' в логах DNS запросов

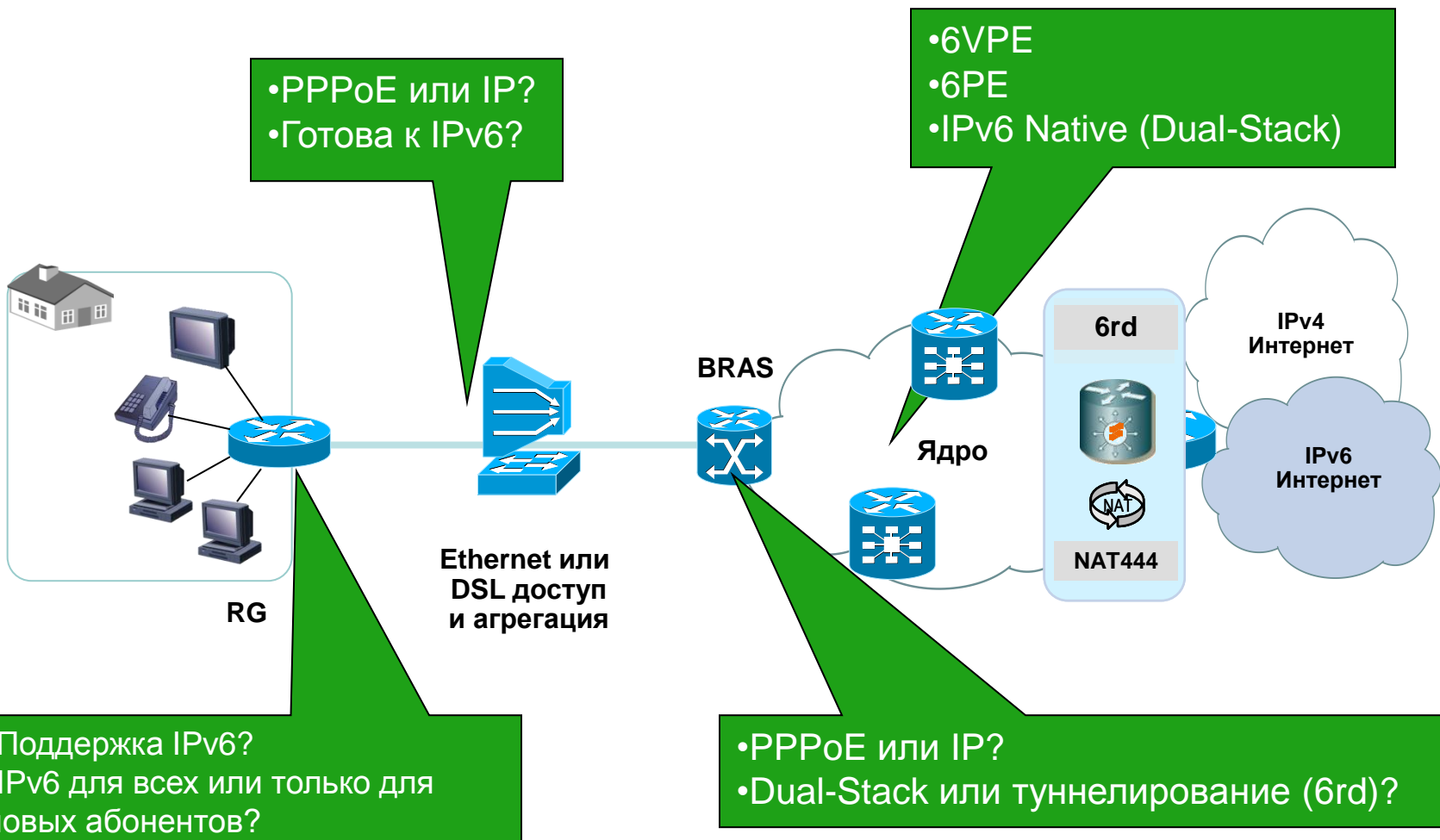


Размышления: так где же делать NAT?

Критерий	NAT на шлюзе (например, ASR 5000)	NAT на маршрутизаторе (например, CSR-1 с CGSE)
Масштабируемость	> 120М соединений > 1М/сек скорость установления соединений	~240М соединений(CRS- 16) > 1М/сек
Управление NAT трансляциями	Для каждого абонента; Для всей системы	Для всей системы
Сбор NAT статистики	Для каждого абонента; Общий	Общий
Высокая доступность (HA)	1:1 Intrabox HA 1:1 Interbox HA	1:1 Intra-box hot standby (Будущее: 1:1 Interbox HA)
Конвергенция (FMC)	NAT специфичен шлюзу, сети доступа	NAT закрывает несколько сегментов сети
Управление публичными IPv4 адресами	Распределенное	Централизованное
Решение нехватки частных IPv4 адресов	Разделение сети на сегменты: - локальный пул частных IPv4 адресов для каждого шлюза;	Разделение сети: - локальный пул для каждого VPN; - будущее: GI-DS-lite;

Технологии

Что выбрать?



Cisco Expo 2011



Спасибо!

Просим Вас заполнить анкеты.
Ваше мнение очень важно для нас.

innovate *together*