NAT на PC-серверах

Тонкая настройка Vyatta

Кирилл Малеванов, ПиН Телеком

Специальное спасибо: Павел Учускин, ПиН Телеком

To NAT or not to NAT?

Предпосылки

- Структура сети с преобладанием статических приватных адресов у клиентов
- RIPE любит и рекомендует NAT 4-to-4
- Все ведущие производители телеком-оборудования предлагают Carrier Grade NAT решения

To NAT or not to NAT

Плюсы

- 80% пользователей сети все равно, какой у них IP-адрес
- Внешний IP-адрес на младших тарифах может являться фактором увеличения ARPU
- Приватный IP-адрес сокращает расходы на внешние каналы по P2P-трафику

To NAT or not to NAT

Минусы

- При использовании NAT очень важно правильно работать с ALG
- Единая точка возникновения проблем для тысяч пользователей
- Стоимость решения для всей сети

Варианты NAT

- Cisco
- Juniper
- Huawei
- Ericsson
- PC

Варианты NAT: старые Cisco

- Классический пример – 7206 или 7301

700 мбит/сек максимум, $4900

- Специальные модули: ACE, FWSM

6 гбит/сек ACE, 2 гбит/сек FWSM

$30000 ACE20-16G

- Файрволы: ASA

ASA5520, 450 мбит/сек, $3500

Варианты NAT: новые Cisco

- Продолжение классики: ASR1000

до 20 гбит/сек

- Специальные модули: ASE-SM, CGSE (CRS-1)

до 8 гбит/сек

- Файрволы: ASA

ASA5580-20, 5 гбит/сек, $25000

Средняя стоимость решения - $5К/Gbit

Варианты NAT: Juniper

- Файрволы: SRX

SRX-650, 1.5 гбит/сек

- Специальные модули: MS-DPC

до 8 гбит/сек, $120 000 GPL

Средняя стоимость решения - $5К/Gbit

Варианты NAT: Huawei

- BRAS: MA-5200

Модуль 2.5 гбит/сек half-duplex

- Софт-маршрутизаторы

AR-46, аналог Cisco 7206

- Новые маршрутизаторы

CX-series, аналог Cisco ASR1000

NE-40E, аппаратный модуль NAT/Netflow

Средняя стоимость решения - $5К/Gbit

Варианты NAT: Ericsson

- BRAS SE-series

NAT на CPU, до 8M сессий

- Специальные модули DPI

Нет внедрений

Средняя стоимость решения - ?/Gbit

Варианты NAT: PC

Плюсы

- Самый дешевый «старт»
- Куча документации и обилие обслуживающего персонала
- Низкая цена для возможностей экстенсивного и интенсивного роста
- Закон Мура на нашей стороне

Варианты NAT: PC

Минусы

- Надежность ниже, чем у аппаратных решений
- Слишком широкие возможности выбора программного обеспечения для разных задач: NAT, OSPF, BGP, firewall
- Сложность настройки для высокопроизводительных систем
- Производительность

Лавируя среди подводных камней

Надежность

- Применяем серверные решения: двойной БП, IPMI, hotswap вентиляторы.
- Низкая цена PC-решения позволяет поставить рядом резервный сервер
- Балансировка и HA перед серверами

Лавируя среди подводных камней

Широта выбора

- Версия ядра, файрвола, библиотек, компилятора, демонов маршрутизации, логирования, SSH/telnet etc для серверов
- Цельная операционная система для аппаратных решений
- Компромисс: целевая ОС для PC-маршрутизаторов

Лавируя среди подводных камней

Вопрос религии:

iOS vs Android

Лавируя среди подводных камней

Широта выбора

- Mikrotik – 512К сессий в conntrack
- Vyatta – возможность влезть «в душу»

Лавируя среди подводных камней

Сложность настройки

- Для усредненного потребителя система ставится за 20 минут, требует изменения 1-2 параметров
- Самое сложное – подбор комплектующих

Лавируя среди подводных камней

Производительность

- По сравнению с 2009 годом, когда вышла статья «NAT и Netflow на больших сетях», производительность универсальных CPU выросла в 4 раза
- Для stateful обработки 1 Mpps больше не требуются специальные сетевые карты

Vyatta: тесты

- Подбор комплектующих
- Выбор версии ОС
- Настройка ОС
- Изучение возможностей роста

Vyatta: тесты

Подбор комплектующих

- Выбор сетевой карты влияет на производительность больше, чем выбор CPU
- Чем быстрее шина, тем лучше

Vyatta: тесты

Выбор версии ОС

- Последняя версия была 6.1
- Вышла версия 6.2, но ядро оказалось хуже
- Проблема нехватки памяти
- Переход на x64 архитектуру
- Необходимость подбора версии драйвера сетевой карты

Vyatta: тесты

Настройка ОС

- netlink-buffer-size у зебры 2048576
- net.core.rmem_max = 16777216
- net.core.netdev_max_backlog = 30000
- conntrack-expect-table-size '16384'
- conntrack-hash-size '4194304'
- conntrack-table-size '10000000'
- conntrack-tcp-loose 'enable’

http://wiki.khnet.info/index.php/Conntrack_tuning

Vyatta: тесты

Нехватка памяти

- При использовании NAT выстраивается хэш-таблица. Чем длиннее ключ – тем быстрее выбор нужной сессии для каждого пакета.
- Увеличение длины ключа приводит к увеличенному потреблению памяти
- Адресация свыше 4Гбайт памяти требует x64

Vyatta: тесты

Переход на x64

- Официального релиза x64 нет
- Инструкции по сборке – на форумах

Выбор версии драйвера

- Распределение трафика по очередям
- Ручная настройка smp_affinity

Vyatta: тесты

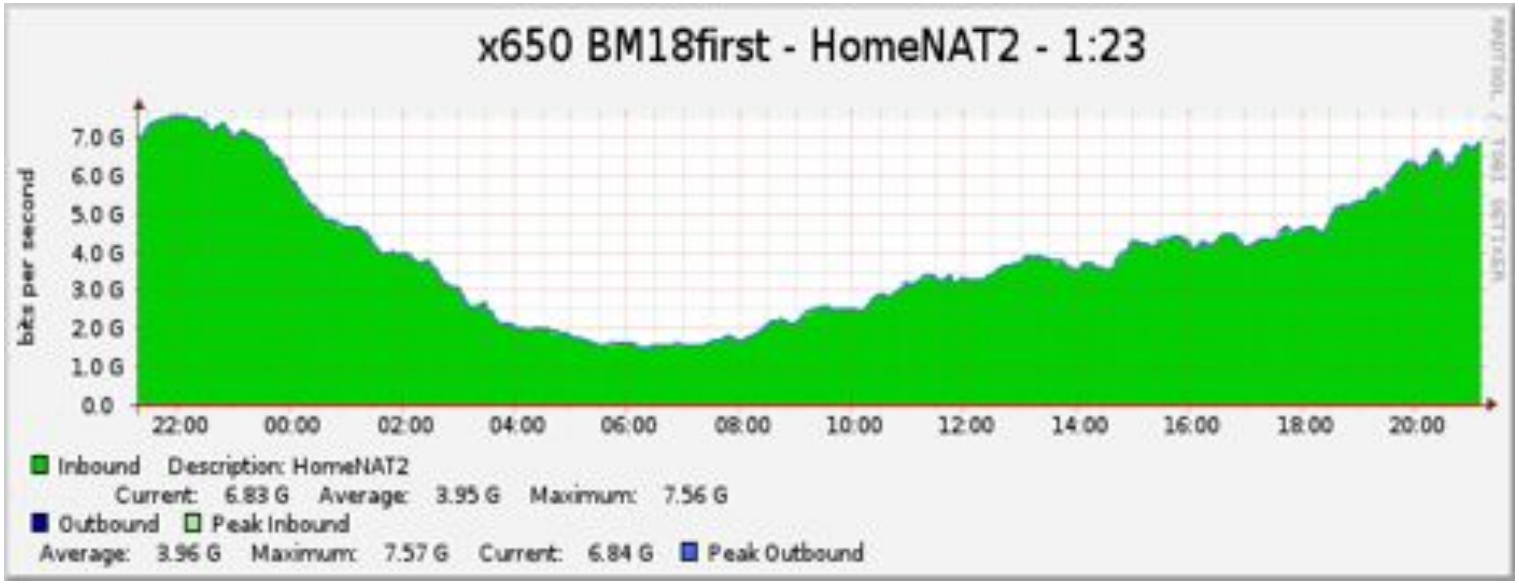
Результаты тестов:

- Xeon X3430
- NIC Intel Quad Port Pro 1000 VT
- 22% CPU на 1 Gbit/sec FD

Vyatta: production

Результаты в жизни:

- Xeon X5660
- NIC Intel X520-SR2
- 20% CPU на 1 Gbit/sec FD



Vyatta: production

Стоимость решения:

$1K/Gbit

Vyatta: post-production

Плюсы

- Очевидны

Минусы

- место в стойке
- отдельные сущности для задачи NAT
- Электропитание (хотя с чем сравнивать)