

Ericsson (RedBack) SE100 в качестве пограничного маршрутизатора и NAT сервера

Доброго времени суток, коллеги! Хочу рассказать об опыте внедрения на нашей сети маршрутизатора фирмы [Ericsson SmartEdge 100](#). Почему именно он и как мы выбирали оборудование для апгрейда, я бы предпочел оставить за кадром, чтобы избежать ненужных размахиваний флагами. Расскажу лишь о самом роутере, впечатлениях от работы с ним и тех подводных камнях, что нас ждали.

Все, что здесь описано, актуально для SeOS 6.2.1 и соответствующей библиотеки документации.

Начну с технических характеристик, которые Вы и сами можете легко найти, но для порядка, посмотрим что же мы имеем.

Основные параметры:

- Пропускная способность 12 Гбит/с, производительность 8 Мп/с;
- Возможность маршрутизации с поддержкой IPv4 и IPv6 обеспечивается аппаратно с числом маршрутных записей до 1,5 миллиона и 1 000 элементов BGP;
- До 1M NAT трансляций.

Собственно, это все параметры, что интересовали меня в первом приближении. Ну и разумеется, интересовали гигабитные интерфейсы, а их на коробку всего лишь 6. Мало? Ну зато честно, при заявленной производительности. В итоге роутер заказали в НАГе, договорились, что будем ждать до 8 недель, но фортуна улыбнулась и желанная железка приехала быстрее — всего через две недели после оплаты счета.

Про интерфейсы и их нумерацию...

Как я уже сказал, всего на коробку 6 гигабитных интерфейсов, но это не совсем так. На «теле» есть 4 гигабитных порта, из них два могут быть использованы для форвардинга (это combo-порты медь/оптика) и еще два, исключительно для управления. При заказе мы взяли еще две MIC карты с медными гигабитными портами — MIC-SE100-2GE-TX. Есть так же вариант с SFP портами — MIC-SE100-2GE-FX.

Таким образом, на нашем роутере два гигабитных порта для управления и шесть для форвардинга. Роутер к нам пришел с уже установленными картами, поэтому процесс установки описать не смогу, но думаю он стандартный.

Как оказалось, обе карты стоят во втором слоте. Первый слот — это порты для управления.

Смотрим железо:

```
[local]nat-box#sh hardware
```

Fan(s) Status	Normal
AC Power Supply Status	Normal
Active Alarms	NONE

Slot	Type	Serial No	Rev	Ver	Mfg Date	Voltage	Temp
N/A	backplane		0	4	01-JAN-2010	N/A	N/A
1	xcrp		0	4	01-JAN-2010	OK	NORMAL
2	carrier		0	4	01-JAN-2010	N/A	NORMAL
	MIC1: ge-2-port					OK	NORMAL
	MIC2: ge-2-port					OK	NORMAL

Обе MIC карты во втором слоте. На «теле» есть маркеры, указывающие на номера портов, несколько озадачившие меня.

Как оказалось, порты имеют номера: 2/1 и 2/2 для встроенных портов, 2/3 и 2/4 для первой MIC карты, и 2/15 и 2/16 для второй MIC карты. Чем обусловлена такая нумерация, я могу только догадываться.

Базовая настройка

CLI маршрутизатора порадовал. Судя по всему, хоть об этом и скромно молчат официальные документы, ядро управления построено на NetBSD. В CLI привычно работает справка по нажатию «?» в конце командной строки и автодополнение, тоже традиционно — по «TAB».

Приглашение CLI выглядит как:

`[context-name]hostname#`, а при выполнении операций по конфигурированию:

`[context-name]hostname(config)#`

Переход в режим конфигурирования осуществляется командой `config` и далее, если конфигурирование осуществляется в пределах некоторого контекста:

`context <context-name>`

Каждое действие по конфигурированию SE100 является транзакцией и должно быть принято к исполнению командой `commit`, либо отменено командой `abort`. Текущие транзакции можно увидеть: `sh transaction`

В SE100 воплощены несколько необычных для цисковода концепций. В частности «контексты».

Контекст - это изолированный виртуальный маршрутизатор с обособленным набором пользователей, служб и сервисов. Что-то типа VRF, но с более широкими возможностями.

По-умолчанию, мультиконтекстность на роутере отключена. При необходимости включается командой: `service multiple-contexts`.

Каждый контекст имеет свой набор пользователей и сервисов. Под сервисами подразумеваются как сервисы динамической маршрутизации, обмен mpls метками, snmp, так и ssh/telnet/ftp/sftp/tftp сервер и клиенты (tftp только клиент!). Между контекстами возможна маршрутизация. Маршрутизация между контекстами включается так же, как отдельная служба, командой: `service inter-context routing`

После включения службы ssh, залогиниться в нужный контекст можно так: `ssh -l username@context-name IP`. Разумеется, предварительно создав пользователя в контексте, либо уже работая на маршрутизаторе: `context <ctx-name>`.

Интерфейсы и порты

Конфигурирование интерфейса производится в рамках контекста, в котором работает интерфейс, а конфигурирование порта — в глобальном конфигурационном режиме.

При покупке меня так же волновал вопрос работы SE100 с SFP модулями не одобренными Ericsson. Как оказалось, SE100 всеяден. Сейчас в нашем SE100 работают два модуля — от D-Link и Cisco. Вполне успешно. Порт с модулем от D-Link сходу не поднялся, помогла команда `auto-negotiate force enable`.

В рамках задачи, нам нужно настроить NAT , 3 BGP сессии и немного отфильтровать трафик с аплинков. Начнем с NAT'а.

NAT

В технических характеристиках указано, что SE100 поддерживает до 1M NAT трансляций. Как нас заверил тех. специалист продавца, роутер с этим справляется не уходя в «штопор».

Раньше мы натили наших клиентов на 16 реальных адресах и так и оставим.

Для настройки NAT необходимо:

- создать NAT пул;
- создать policy access-list по которому будем выполнять NAT;
- создать NAT политику;
- назначить NAT политику на интерфейс, смотрящий в сторону клиентов.

Так как мы успешно перешагнули гигабит, но 10GE портов на ядре сети не имеем, от SE100 в ядро сети решили кинуть EtherChannel из трех гигабитных линков. Настроили, LACP успешно поднялся, но оказалось, что NAT policy нельзя использовать на интерфейсе, к которому привязана link группа. Увы, в документации об этом ничего не было сказано.

Конфиг link-group выглядел вот так:

```
link-group users ether
description linkgroup-logical-iface
bind interface users inet
mac-address auto
lasp active
!
port ethernet 2/4
```

```

description linkgroup-users
no shutdown
link-group users
!
port ethernet 2/15
description linkgroup-users
speed 100
no shutdown
link-group users
!
port ethernet 2/16
description linkgroup-users
no shutdown
link-group users
!
interface users
ip address 192.168.3.254/30
ip mtu 1500
ip verify unicast source reachable-via rx allow-default
!
service load-balance ip layer-4

```

Зато из документации выяснилось, что роутер умеет ECMP — equal cost multi path, а саппорт Ericsson заверил, что фича работает глобально, не только для link-group.

В итоге в сторону ядра настроили три L3 линка, а также включили балансировку по Layer4.

Балансировка включается командой: *service load-balance ip...*

Вот так выглядит конфигурация интерфейсов, смотрящих в ядро сети:

```

interface users-246
description -c65k-g3/7-
ip address 192.168.3.246/30
ip mtu 1500
ip icmp suppress packet-too-big
ip arp timeout 900
!
interface users-250
description -c65k-g3/9-
ip address 192.168.3.250/30
ip mtu 1500
ip icmp suppress packet-too-big
ip arp timeout 900
!
interface users-254
description -c65k-g3/15-
ip address 192.168.3.254/30
ip mtu 1500
ip icmp suppress packet-too-big
ip arp timeout 900

```

Теперь создаем policy access-list со списком сетей, для которых будем выполнять NAT:

```

policy access-list Users
seq 10 permit ip 172.31.0.0 0.0.0.31 class NAT-0
seq 20 permit ip 172.31.0.32 0.0.0.31 class NAT-32
seq 30 permit ip 172.31.0.64 0.0.0.31 class NAT-64
seq 40 permit ip 172.31.0.96 0.0.0.31 class NAT-96
seq 50 permit ip 172.31.0.128 0.0.0.31 class NAT-128
seq 60 permit ip 172.31.0.160 0.0.0.31 class NAT-160
seq 70 permit ip 172.31.0.192 0.0.0.31 class NAT-192
seq 80 permit ip 172.31.0.224 0.0.0.31 class NAT-224

```

policy access-list отличается от обычного acl тем, что каждое правило или набор правил, могут быть классом, который потом можно использовать, в частности при настройке nat policy, forwarding policy или qos.

Есть ограничение, которое я возможно пропустил при чтении документации — в одном policy access-list не может быть более 8 классов. Для нас это оказалось важным моментом.

Создаем nat пулы:

```

ip nat pool NAT-0 napt multibind
address X.X.X.3/32 port-block 1 to 15
address Z.Z.Z.4/32 port-block 1 to 15

```

Таких пулов в нашей конфигурации — 8 штук, по количеству классов в policy acl.

napt — ключевое слово, указывающее на то, что адреса используемые в пуле, будут задействованы в NAT policy.

multibind — обязательное ключевое слово, если пул будет использоваться в политике, прикрепленной на несколько интерфейсов, а это как раз наш случай. Без этого тоже работает, но не все, что выяснилось в нашем случае, через сутки. :)

Опция *port-block* указывает на блок номеров портов которые будут использоваться для NAT. Номера портов разбиты на 16 блоков, по 4096 портов. В моем случае, используются порты с 4096 по 65535. Это сделано для того, чтобы клиентский трафик не был отфильтрован как идущий с well-known порта. Если не указывать *port-block*, будет использоваться весь диапазон номеров портов.

Ну и теперь политика:

```
nat policy NAT
connections icmp maximum 50
! Default class
ignore
admission-control tcp
admission-control udp
admission-control icmp
endpoint-independent filtering udp
! Named classes
access-group Users
class NAT-0
pool NAT-0 inet
timeout tcp 18000
timeout udp 60
timeout fin-reset 60
timeout icmp 30
timeout syn 60
admission-control icmp
endpoint-independent filtering udp
```

Привожу не всю политику, а только класс по умолчанию и один определенный мной класс, так как в дальнейшем все повторяется за исключением имени класса и пула. В моем примере, NAT выполняется для ACL Users и класса из этого ACL NAT-0 в пул NAT-0. Трафик, что не попадает под описанные классы, просто проходит без изменений(действие для Default class - ignore). Для каждого класса может быть настроен admission-control , для tcp, udp и icmp.

Внимание!!! Admission-control действует на весь класс! То есть, если в классе подсеть из 32 хостов, правило действует на них как на один объект. В нашем примере показано ограничение: 50 одновременных трансляций протокола icmp на класс. Также, для каждого класса, может быть настроено время жизни различного рода трансляций, что позволяет оптимизировать нагрузку на маршрутизатор.

Обязательно укажите в настройках политики *endpoint-independent filtering udp*, без этого могут быть проблемы с voip, и on-line играми. Кстати, об этом в документации тоже, увы, не сказано.

Ну и назначаем политику на интерфейс:

```
interface users-246
description -c65k-g3/7-
ip address 192.168.3.246/30
ip mtu 1500
ip icmp suppress packet-too-big
ip arp timeout 900
ip nat NAT
```

Как я уже выше писал, при создании пула, который будет использоваться в политике, прикрепленной на несколько интерфейсов, обязательно нужно указать *multibind*. Без этого также осуществляются NAT трансляции, но не работает *endpoint-independent filtering*.

Каждый пул имеет строго определенное кол-во трансляций. Выделение ресурсов ASIC под трансляции осуществляется на уровне порт-блоков. Как-только какой-либо пул алоцирует в ASIC последний порт-блок, SE сбрасывает об этом сообщение в syslog. Это означает, что нужно либо расширить пул доп. портами, либо уменьшить количество хостов в классе, для которого производится трансляция. Отслеживая приход этого сообщения, можно эффективно мониторить перегрузку по NAT трансляциям. Сообщение выглядит следующим образом:

```
[local]Redback#Nov 22 05:21:22: %NAT-6-INFO: Out of nat pool addr/blk for slot #num pool $PoolName ctxt
```

#ctxNum -55107104, где

#num — номер карты,

\$PoolName — имя пула,

#ctxNum — номер контекста.

Настройка BGP

Тут все достаточно стандартно. Даже скучновато было. Отличий от настройки маршрутизаторов Cisco минимум, за исключением пожалуй более логичной структуры конфигурации у Ericsson.

У нас три ап-link, от каждого мы берем FV, режим префиксы с маской длиннее /24 и анонсируем нашу AS и две наших сети.

Конфигурация выглядит примерно так:

```
router bgp XXXXX
router-id X.X.X.X
multi-paths external 2
fast-reset 10
address-family ipv4 unicast
flap-statistics
network A.A.A.A.0/22
network B.B.B.B.0/21
!
neighbor Z.Z.Z.Z external
remote-as XXXX
update-source uplink1
send community
send ext-community
address-family ipv4 unicast
route-map FullView in
prefix-list ourAS out
!
neighbor Y.Y.Y.Y external
remote-as YYYY
update-source uplink2
send community
send ext-community
address-family ipv4 unicast
route-map FullView in
prefix-list ourAS out
!
neighbor W.W.W.W external
remote-as WWWW
update-source uplink3
send community
send ext-community
address-family ipv4 unicast
route-map FullView in
prefix-list ourAS out
```

В таком простом виде, как видите, отличий нет. Возможно при более сложных конфигурациях они и появятся.

И напоследок..

Общие впечатления от работы с маршрутизатором остались скорее положительные, хотя как и у всех вендоров, не все идеально. Отмечу плюсы и минусы, которые я для себя отметил.

Плюсы:

- аппаратная платформа;
- удобный и логичный CLI, гибкость конфигурирования;
- механизм контекстов со своим набором пользователей и служб;
- всеядность в плане использования SFP модулей сторонних производителей;
- относительно невысокая цена (в сравнении с аналогичными решениями других вендоров);
- адекватная и оперативная тех. поддержка вендора и продавца.

Минусы:

- недочеты в документации;

- несколько необычный формат поставляемой документации — Active Library;
- невозможность получить некоторые параметры по SNMP, в частности количество NAT трансляций и загрузку ASIC'ов;
- невозможность выполнить NAT для протоколов туннелирования;
- отсутствие планировщика для выполнения периодических операций по расписанию(а-ля cron).

По моему мнению, маршрутизатор заслуживает внимания и найдет свое применение в малых и средних сетях. Через наш в данный момент суммарно идет примерно 4,2 Gb трафика, 700 Kpps и выполняется 250-300 тысяч NAT трансляций.

[Полный конфиг.](#)

Андрей Андрющенко, начальник отдела эксплуатации информационных систем ООО "ГорПТУС", группа компаний ["ОсколТелеком"](#).

[Ссылки по теме](#)