Достижения в области технологий пиринга

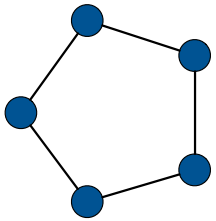
ENOG9 Казань, июнь 2015

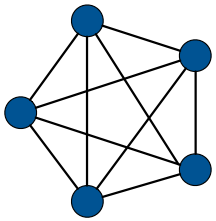
О чем пойдет речь

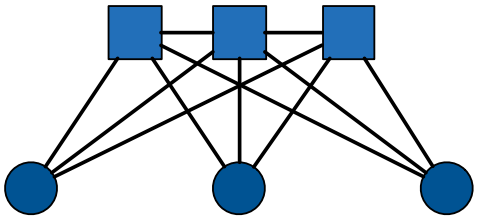
- Платформа IX будущего. Критерии выбора.
- Сложности и тонкости перехода на «Двойное ядро MSK-IX»
- Порты 100Gbit/sec. Так ли это хорошо и удобно?
- Что делать в случае перегрузки 100G?
- Протокол BFD или новое качество пиринга для IX
- SDN (Openflow) – имеет ли смысл и не пора ли его внедрять?
- Атаки через IX – выше скорости, сложнее бороться. Как облегчить жизнь оператору связи и спастись от напасти DDoS?

MSK IX

Масштабирование сетей коммутации



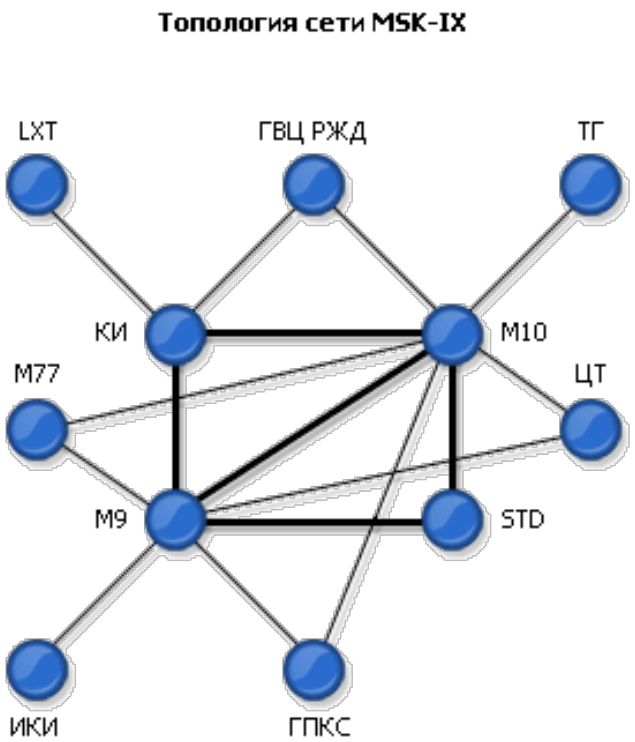


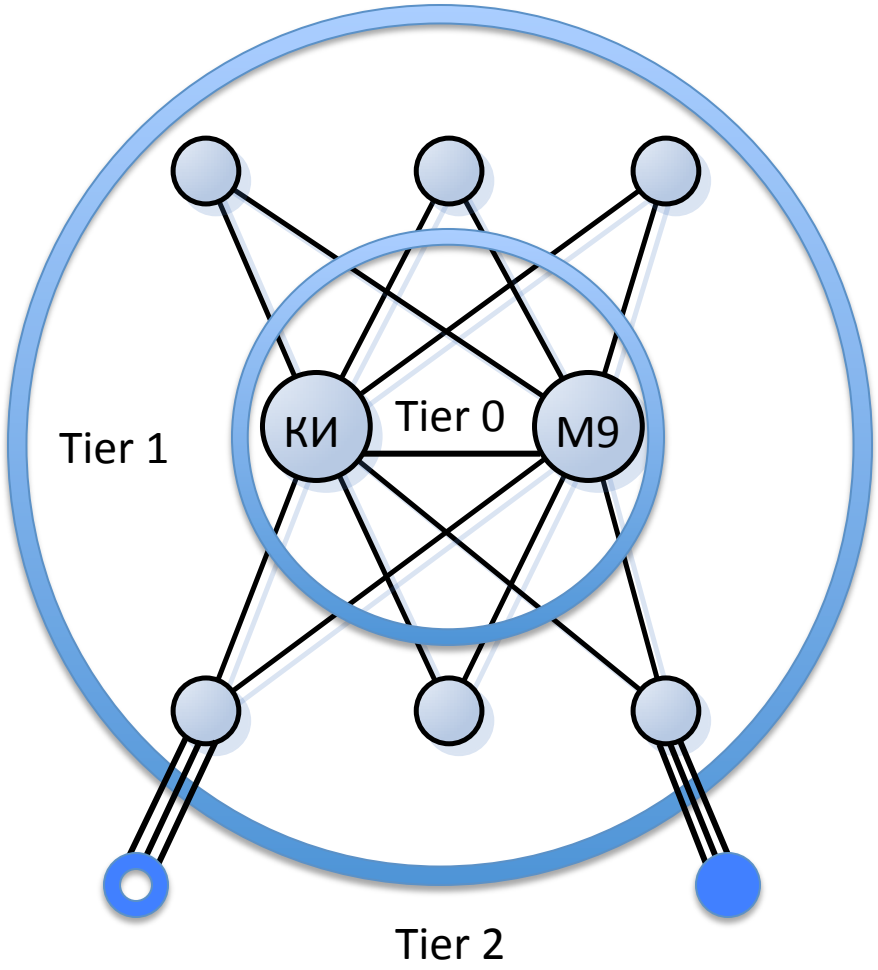


Топология	Ring	Full Mesh	C-Core
Число каналов	N	N(N-1)/2	NC+C-1
Емкость магистрали	N	N(N-1)/2	NC
диаметр	N/2	1	2
Двойной отказ	Разрыв	Не влияет	Не влияет
Масштабируемость	Проблема отказоустойчивости	Дорого	Компромисс

MSK IX

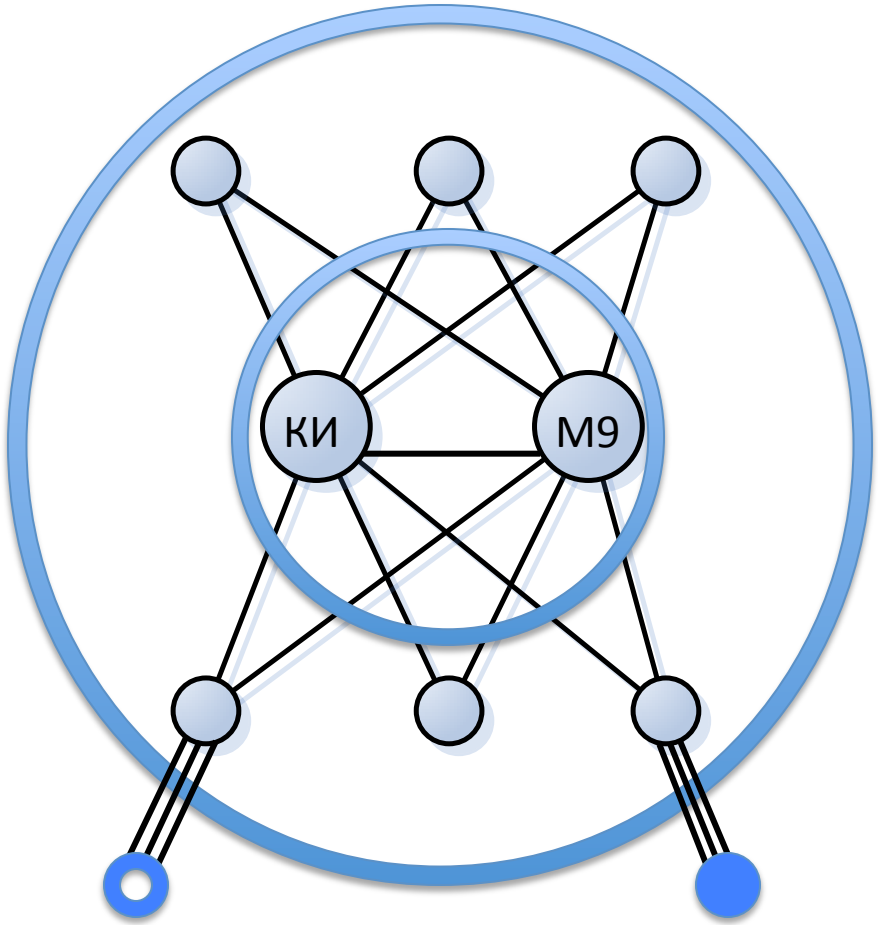
Переезд на «ДВОЙНОЕ ЯДРО»





Этапы большого пути

- Монтаж двух новых ядер
- Организация стыковки по протоколу MLAG
- Организация DWDM между ядрами, изменение всей оптической инфраструктуры
- Стык между ядрами и существующей инфраструктурой
- Поочередное переключение одного из коммутаторов на ядро с сохранением резерва на старую схему
- Полный переход на ядро
- Разбор старых линий связи

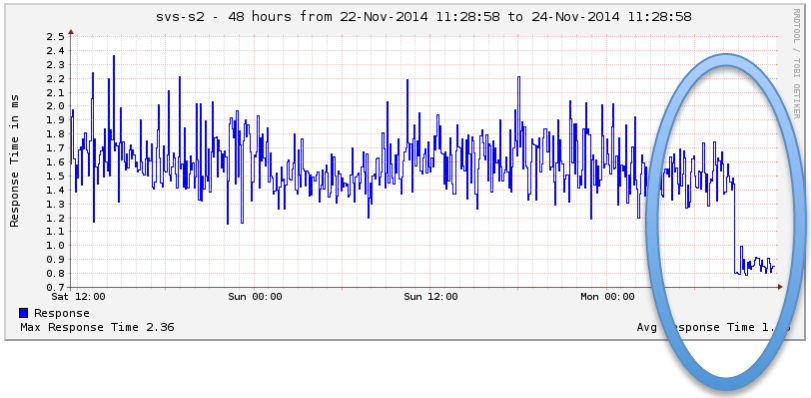


MSK IX

Специфика перехода

Технологический опыт:

- Сложности перехода с Rapid PVST+ на двойное ядро без возможных петель
- Необходим глубокий анализ внутрисетевого трафика
- Нужен детальный анализ физической инфраструктуры

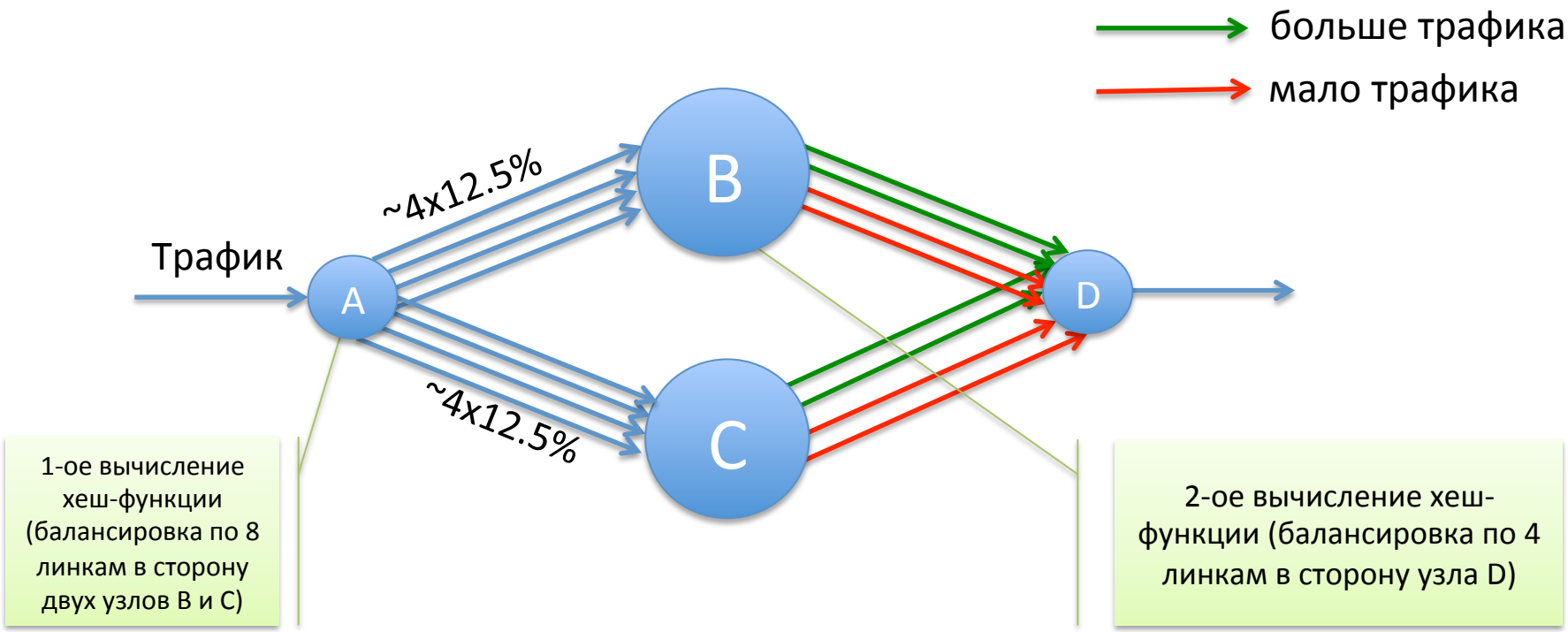


Преимущества в результате:

- Полное резервирование инфраструктуры
- Уменьшение времени отклика
- Уход от протоколов STP
- Максимальная эффективность линий связи
- Упрощение модели прогнозов роста трафика
- Возможность включения новых площадок и точек выноса

Поляризация хэш-функции и ядро

Балансировка по составляющим Etherchannel может привести к проблеме поляризации на сети, когда одинаковый алгоритм на разных участках сети.

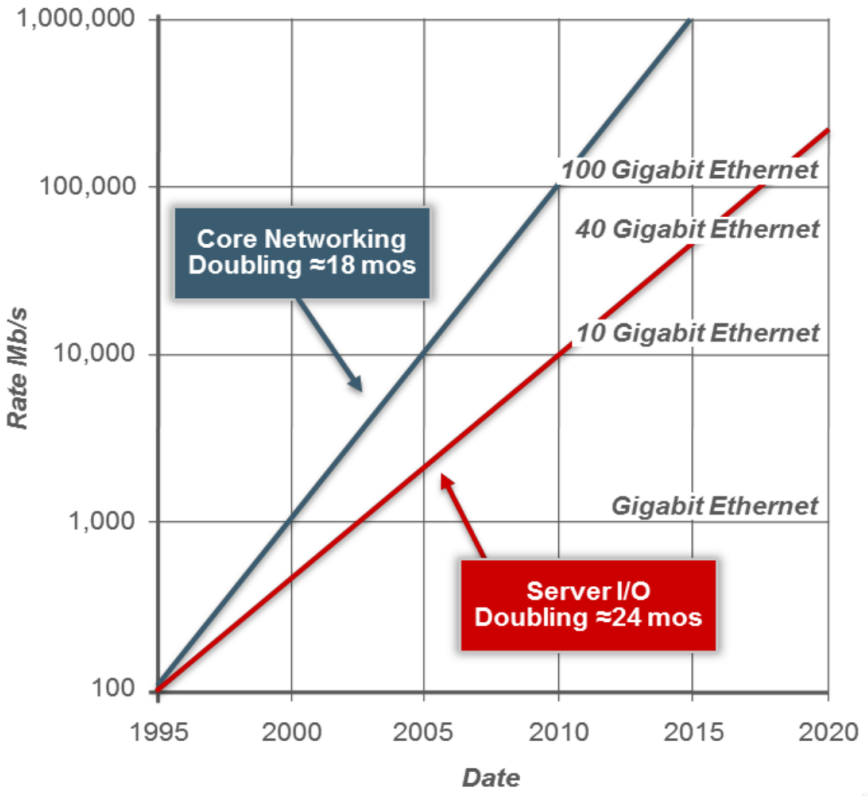


Поляризация хэш-функции и ядро

Допустимые решения проблемы (сильно зависит от аппаратных возможностей)

- Подмешивание случайного значения (seed) к алгоритму балансировки
- Применение разных алгоритмов балансировки на разных устройствах
- Использование разного кол-ва линий связи для балансировки трафика на разных участках

Рост трафика и новые скорости



Скорость	Wireless	Сервера	Сети
1Gbit/s	✔	✔	✔
2.5Gbit/s	✔		
5Gbit/s	✔		
10Gbit/s	✔	✔	✔
25Gbit/s		✔	
40Gbit/s		✔	✔
100Gbit/s			✔
400Gbit/s			✔

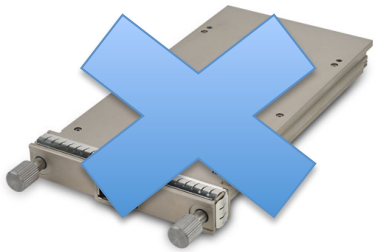
Источники:http://www.ieee802.org/3/hssg/public/nov07/HSSG_Tutorial_1107.zip
Презентация Chief Ethernet Evangelist John Ambrosia (26 Euro-IX)

MSK IX

100G @ MSK-IX

Запущены порты 100G для участников MSK-IX

- Второе поколение трансиверов (CFP2)
- Доступен Etherchannel
- 32 порта 100G на коробке
- Меньшее энергопотребление (8W вместо 80W)
- Меньший форм-фактор (в 2 раза меньше)
- Сниженное тепловыделение
- Пока только доступны 100GBASE-LR4, 100GBASE-SR10



CFP

CFP2

Плюсы и минусы 100G

Плюсы:

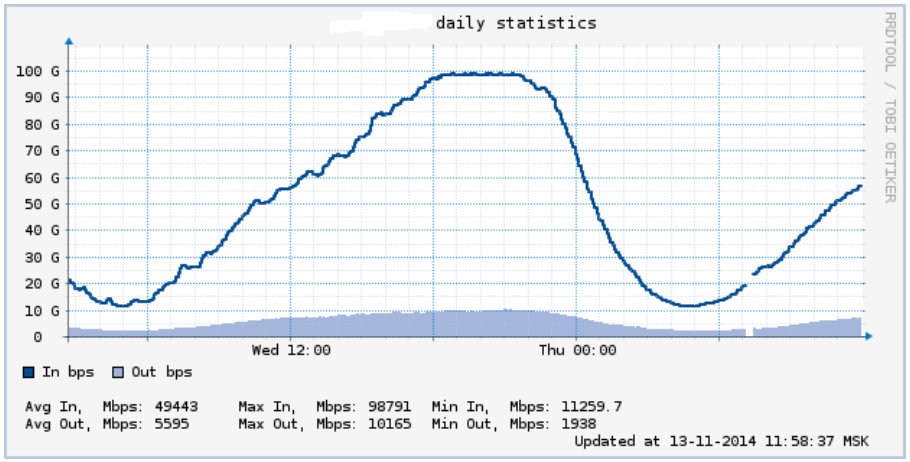
- Выше скорость
- Новая технология
- Высокая энергоэффективность
- Существенное (в 5-10 раз) снижение затрат на оптику

Минусы:

- Ограниченная дальность (100GBASE-LR4=10км)
  - Сложно развиваться в масштабах города – нужно решение на 40 км
  - Усилители, активный DWDM – дорого
  - Подбор параметров оптических вставок – надежность (?), сложность замены
- Высокая стоимость оптических вставок
- Высокая стоимость оборудования и отсутствие фиксированных 1-2U свитчей со 100G

MSK IX

Специфика перегрузки 100G



– Подключить второй 100G (дорого, высокая концентрация трафика в одной коробке в случае Etherchannel)

– 10G Etherchannel в виде отдельного стыка с IX,
что нельзя объединить со 100G портами в Etherchannel (сложно поддерживать и балансировать трафик)

SDN или не SDN?

SDN(Openflow) помогает вынести «голову» на выделенное оборудование

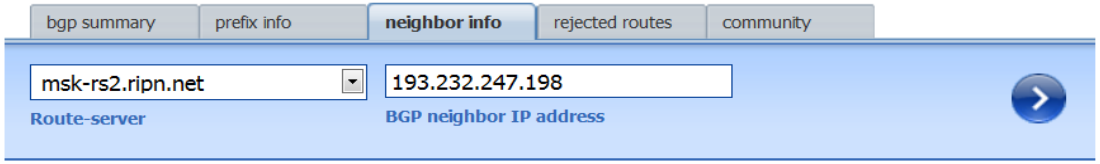
Особенности для IXP:

- Возможно снижение надежность платформы
- Отделяет пуговицы от пальто
- Проблемы совместимости разных версий
- Возможности снижения «паразитного» трафика
- Возможно выделение «нужного» трафика для анализа
- Унификация логики в одном месте
- Критичность доступности контроллера для всей сети, необходимость в резервировании
- Кому предъявлять претензии в случае ошибок?

MSK IX

Протокол BFD

- Запущен на всех точках обмена трафиком MSK-IX
- Ускоряет реакцию на возможные проблемы между клиентами MSK-IX, между клиентами и Route Server.
- Существенно повышает качество обмена, оперативно решая проблему попадания трафика в «черную дыру»
- В 6 раз ускоряет реакцию на потерю пиринг партнера (5 секунд вместо 30 для BGP в реализации MSK-IX)
- Работает по UDP
- Новый уровень пиринга



msk-rs2.ripn.net

```
BGP information for neighbor 'ivi_2' (IP: 193.232.247.198 AS: 57629)
BGP protocol state: Established (since 2014-09-12 15:10:14)
Keepalive/Holdtime: 10/30
Number of outgoing  announced  routes  after  all filters: 19253
Number of incoming accepted/best routes after all filters: 2 / 2

BFD protocol state: Configured
Operational  state: Up (since 2014-09-12 15:10:14)
Parameters: interval '1.000', timeout '5.000'

Accepted BGP routes detailed info:
Status codes: * valid, > best
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network              Next Hop         Metric LocPrf Weight Path
*> 91.233.216.0/22      193.232.247.198       0    100        57629 i
*> 91.233.218.0/24      193.232.247.198            100        57629 i
```

Route Server & BFD

Специфика реализации BFD:

- Выше требования к надежности IXa, более чувствительно к падениям (недостаток)
- Таймеры – можно и меньше, но опасения высокой нагрузки на CPU (в том числе со стороны пиринг-партнеров) и еще выше требования к падениям
- BFD & STP Topology Change несовместимы (любое изменение топологии будет «дергать» BFD)
- На Cisco не поддерживается для Secondary IP на интерфейсе (SR631566485). Планируется доп. Команда ’ bfd neighbor multihop-ipv4’ для задания SRC IP в IOS начиная с 152-1.SY0a

MSK IX

Развитие BFD

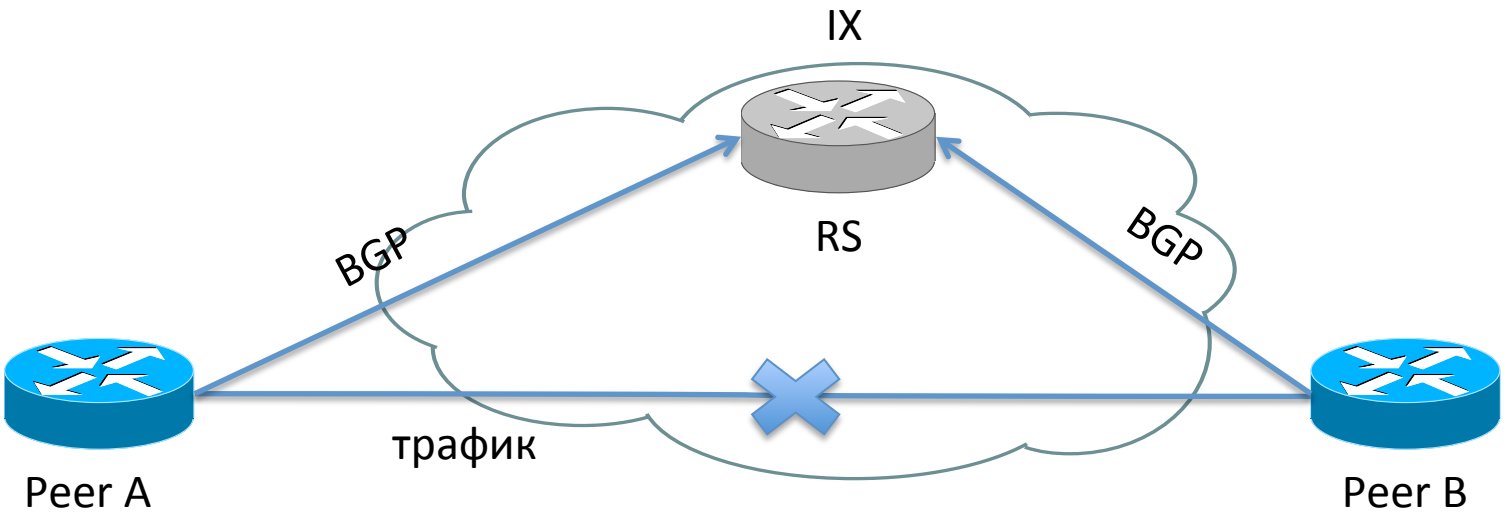
Идея:

реализовать сигнализацию Route Server о проблемах доступности между двумя операторами

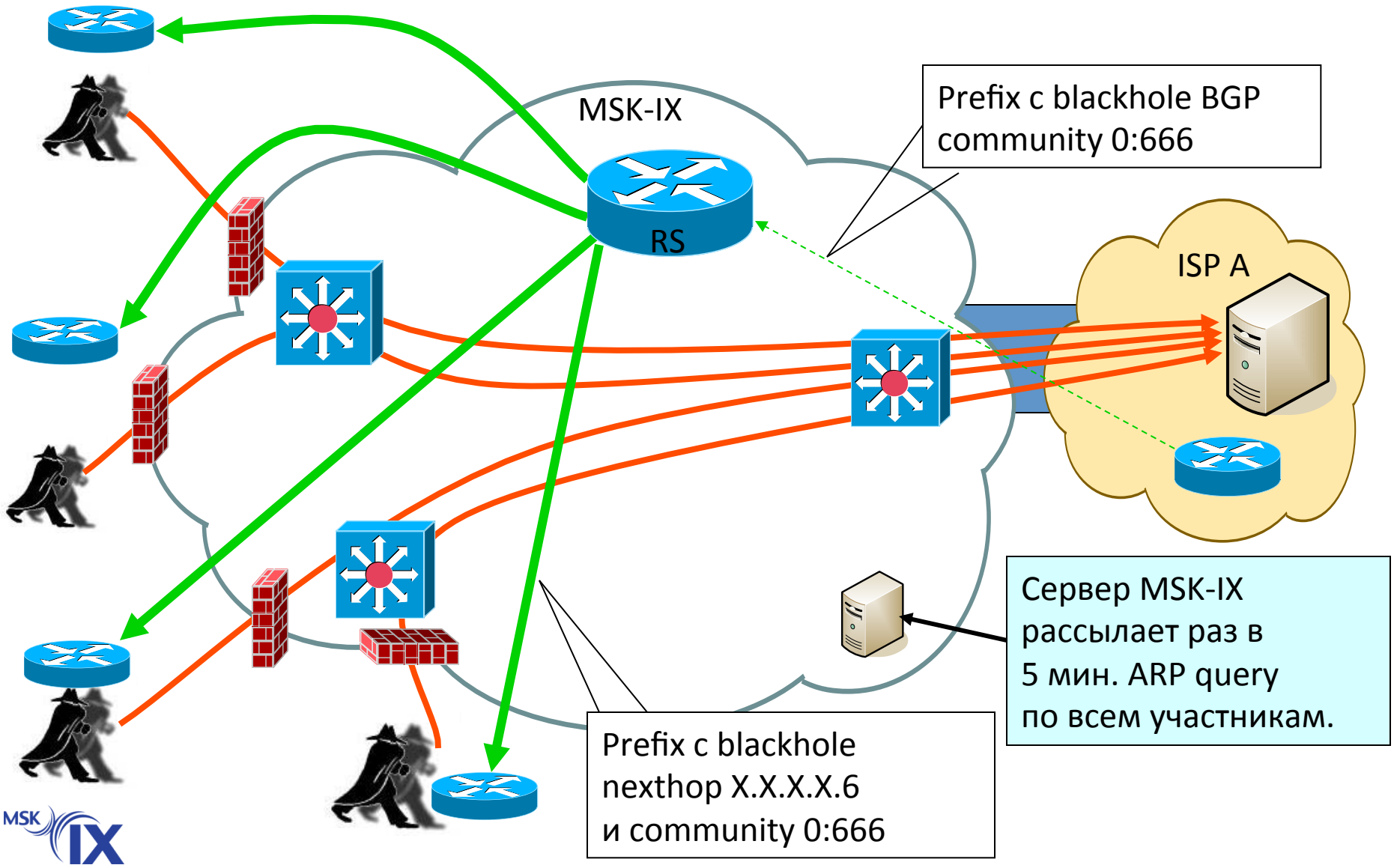
Предлагаемое решение:

- BFD между операторами (RFC 5880)
- Передача Next-Hop Cost Information по BGP (Internet Draft)

http://datatracker.ietf.org/doc/draft-ymbk-idr-rs-bfd/



Blackhole filtering



Blackhole filtering

- Не требует «сложных» конфигураций
- IX контролирует принадлежность blackhole к блоку сетей пиринг партнера
- Позволяет оперативно «прикрыть» атакуемый хост
  - мы фиксировали атаки свыше 40Gbit/sec, которые блокировались с помощью blackhole на MSK-IX)
- **Нужно, чтобы оператор принимал маршрут /32 от MSK-IX**
  - по нашим оценкам около 85% трафика подпадает под blackhole

- Трафик блокируется на входе в магистраль MSK-IX на аппаратном уровне не нагружая устройства

- Возможно использование на private peering (важно проставлять BGP next-hop X.X.X.6 для атакуемых маршрутов)

MSK IX

AntiDDoS

MSK-IX провел тестирования аппаратных решений для нужд очистки трафика/защиты. Некоторые итоги:

- функционал защиты с некоторыми доработками реализуем
- Сложности разделения ресурса защиты в условиях IX
- Возможна интеграция с Route Server
- Высокая стоимость и низкая готовность рынка «платить» за защиту
- Открыты к сотрудничеству и поиску решений

BGP Flowspec

Возможность передачи фильтров защиты от DDoS соседу по BGP

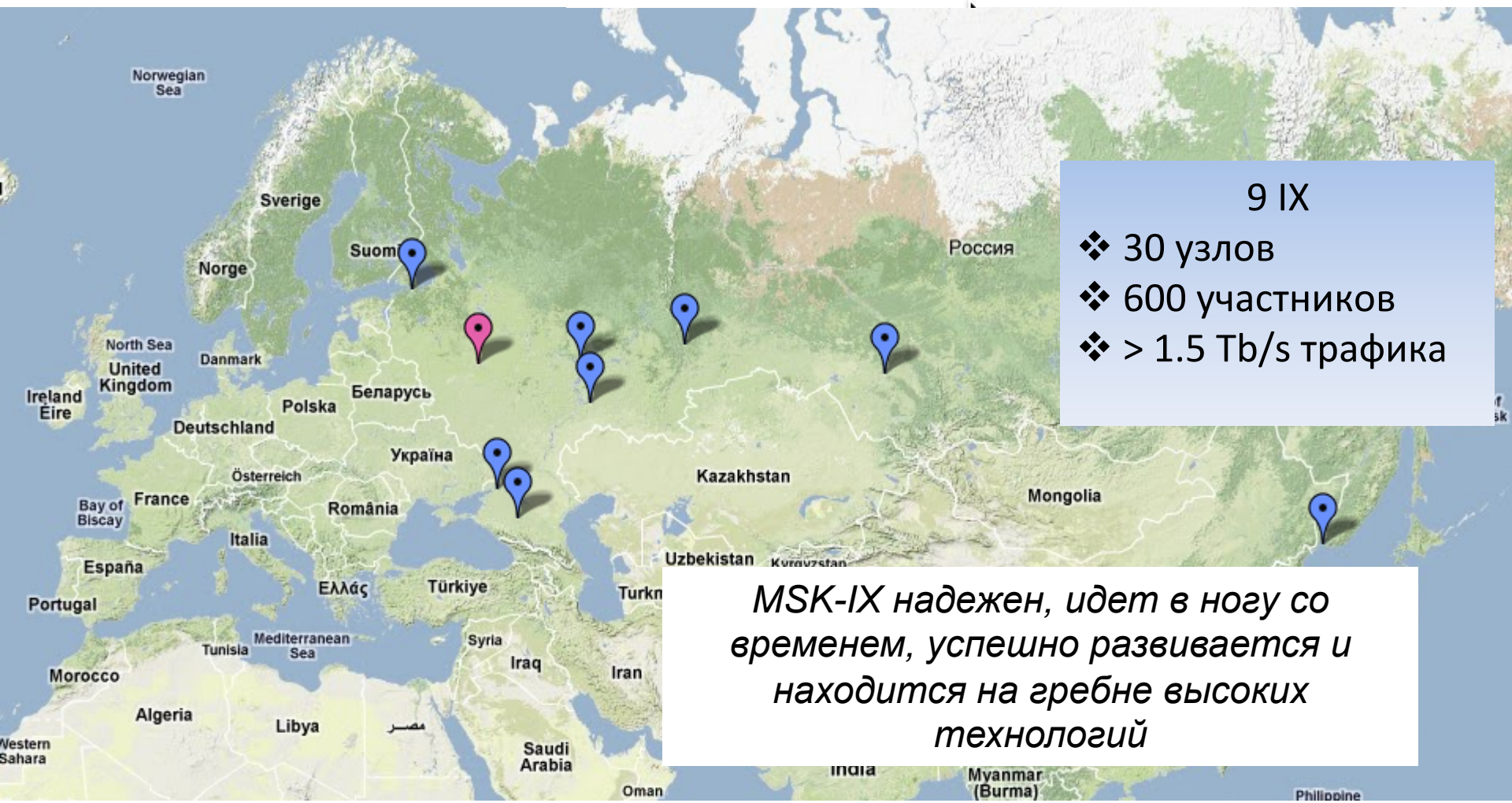
- На основе RFC 5575 с расширением до IPv6
- Изначально поддерживается Juniper, работает на связке Arbor-Juniper
- Поддерживается в Cisco на части оборудования
- Передается посредством BGP NLRI
- Полный набор фильтров L4 и возможности лимитирования или выкидывания трафика, смены VRF, маркировки DSCP, смены next-hop

MSK IX

BGP Flowspec

- Не терпит человеческих ошибок и сложных фильтров
- Имеются вопросы надежности промежуточных устройств (при больших фильтрах они могут игнорировать правило и пропустить трафик)
- Применяется механизм BGP persistence (XR 5.2.2) когда в случае атаки фильтры не сразу пропадают если внутри сети потеряна связанность (полезно в случае изменения профилей атак)
- Пока не поддерживается BIRD Route Server, но имеются планы по внедрению. Есть интерес со стороны сообщества в последнее время.
- Сложности взаимодействия (доверия пиринг-партнеру в случае применения Flowspec фильтров)
- Возможно в будущем «аппаратная» фильтрация на базе Flowspec

MSK IX 20 ЛЕТ



9 IX

- 30 узлов
- 600 участников
- > 1.5 Tb/s трафика

MSK-IX надежен, идет в ногу со временем, успешно развивается и находится на гребне высоких технологий

MSK IX

Александр Ильин

Технический директор

MSK-IX

123154, г. Москва, ул. Маршала Тухачевского, д. 37/21
+7 (495) 737-9295