



Networking  
For everyone

# CCIE Enterprise Infrastructure

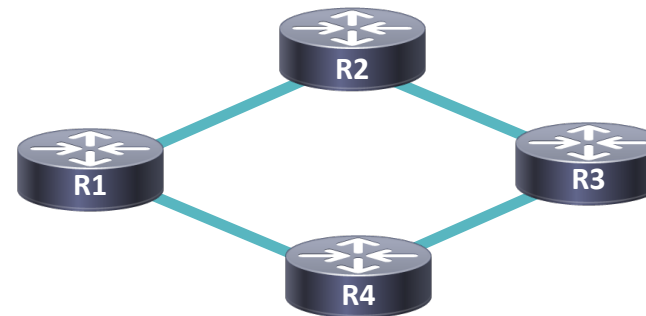
## OSPF

---

# OSPF

- Стандартный открытый протокол
- Находит соседей обменом hello-пакетами
- Обменивается с установленными соседями LSA (Link State Announcement), в которых передает:
  - С какими маршрутизаторами установлено соседство
  - Как устроена топология
  - Какие IP-сети анонсируются

R1 LSDB	LSA	Установленные соседи
	R2	R1, R3
	R3	R2, R4
	R4	R1, R3



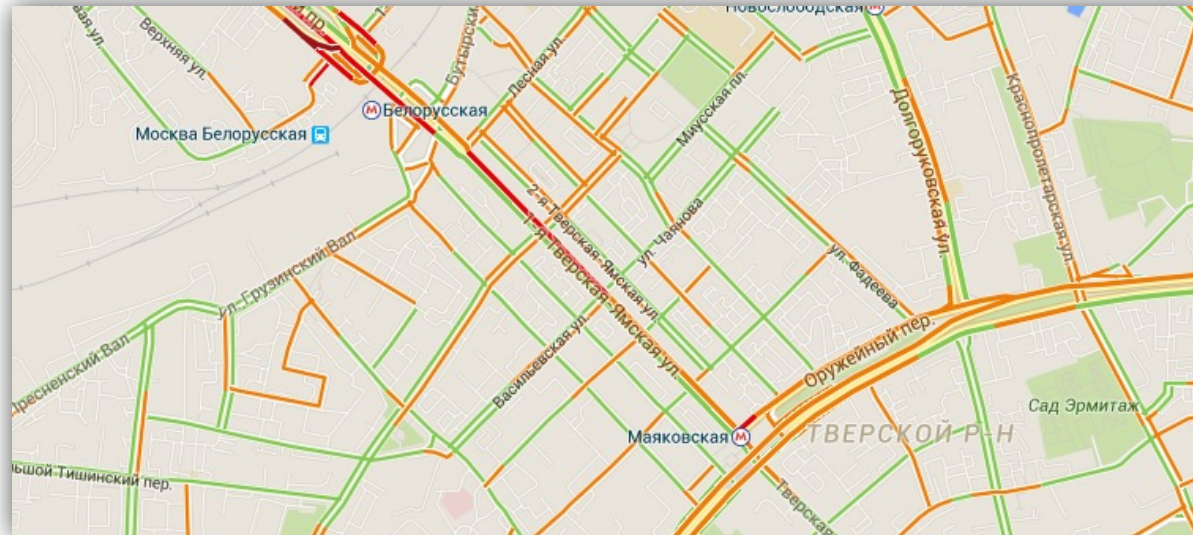
# Версии протокола OSPF

- OSPFv2:
  - Работает только поверх IPv4
  - Обменивается только маршрутами IPv4
  - RFC 2328
    - Устаревшие RFC для OSPFv2: 1131, 1247, 1583, 2178
    - 50+ дополнительных RFC для различных костыльных механизмов
  - Multicast 224.0.0.5 (Hello) и 224.0.0.6 (DR), вложение IP 89
- OSPFv3:
  - Работает только поверх IPv6
  - Обменивается маршрутами как IPv4, так и IPv6
  - RFC 5340
  - Multicast ff02::5 (Hello) и ff02::6 (DR), вложение IP 89



# Отслеживание состояний каналов

- Каждый маршрутизатор строит топологию сети
  - Обнаруживает непосредственно подключенных соседей
  - Отсылает им и получает от них информацию о связях маршрутизаторов
  - Строит карту связей и просчитывает кратчайшие маршруты до каждого узла
  - Пользуется полученными маршрутами для нахождения шлюза для каждой анонсированной в топологии сети



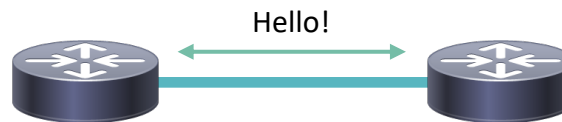
# Особенности протоколов Link State

- Преимущества:
  - Эффективно используют полосу пропускания
  - Быстро реагируют на изменения в сети
  - Гарантируют защиту от петли
- Недостатки
  - Вычислительно существенно более сложны по сравнению с ДВ-протоколами
  - Плохо масштабируются в сетях с произвольной топологией
  - Рассчитаны на работу с топологией, а не с маршрутной информацией IP
    - Обросли немыслимым количеством костылей



# Hello

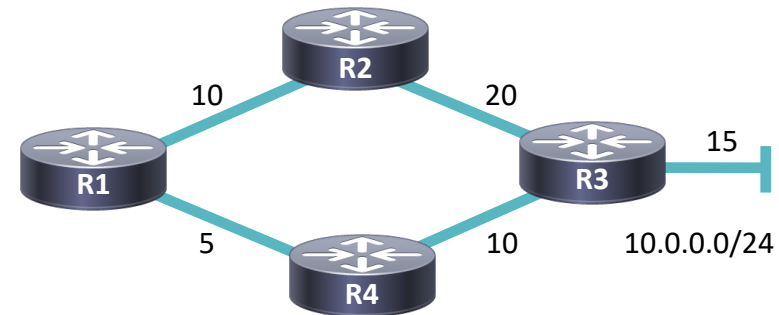
- Hello-пакеты отсылаются по таймеру на всех интерфейсах OSPF
  - Используется local multicast (224.0.0.5)
  - В каждом пакете указывается информация о конфигурации маршрутизатора
  - С непротиворечиво настроенными соседями устанавливается соседство
    - Должны совпасть таймеры Hello Interval и Dead Interval
    - Должны совпасть номер региона и флаг stub
    - Пакеты должны пройти процедуру аутентификации



# Алгоритм SPF

- OSPF использует алгоритм Дейкстры для нахождения кратчайшего пути в топологии между двумя маршрутизаторами
  - Метрика – стоимость пути до удаленной сети
  - Стоимость пути = сумма стоимостей интерфейсов, составляющих путь
  - Меньше стоимость пути – более вероятно прохождение трафика по нему

R1 LSDB	LSA	Соседи	Сети
	R2	R1 (10), R3 (20)	
	R3	R2 (20), R4 (10)	10.0.0.0/24 (15)
	R4	R1 (5), R3 (10)	
R1 RIB	Сеть	Маршрут	Стоимость
	10.0.0.0/24	R3-R2-R1	45
		R3-R4-R1	30



# Router ID

- Каждый маршрутизатор выпускает LSA, которые подписываются его уникальным идентификатором
  - ID обязан быть уникальным в пределах всей топологии
  - И не меняться в течение всего срока работы OSPF
- Идентификатор – 32-битное число
  - Часто записывается в форме IP-адреса
- Возможные способы назначения
  - Вручную
  - Взять самый маленький IP-адрес на устройстве
  - Взять какой-нибудь еще IP-адрес на устройстве
    - Например, маршрутизаторы Cisco предпочитают брать самый большой IP-адрес с виртуальных loopback-интерфейсов





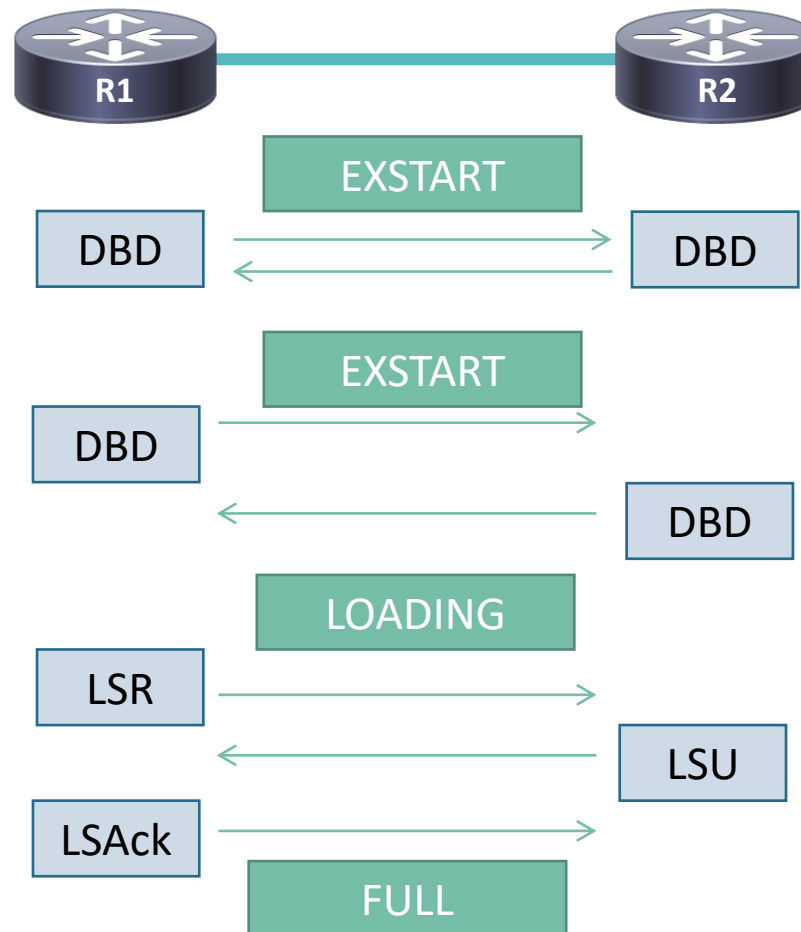
# Состояние соседства

- Hello protocol
  - INIT: принимаются Hello без указания «нас» как соседа
  - 2WAY: принимаются Hello с указанием «нас» как соседа
- Database Exchange
  - EXSTART: согласование стартовых Sequence Number, выбор Master
  - EXCHANGE: обмен пакетами Database Description
  - LOADING: обмен пакетами Link State Request и Link State Update
  - FULL: синхронизированные LSDB
- Также существуют состояния ATTEMPT и DOWN
- Детально процесс разбирается далее в курсе



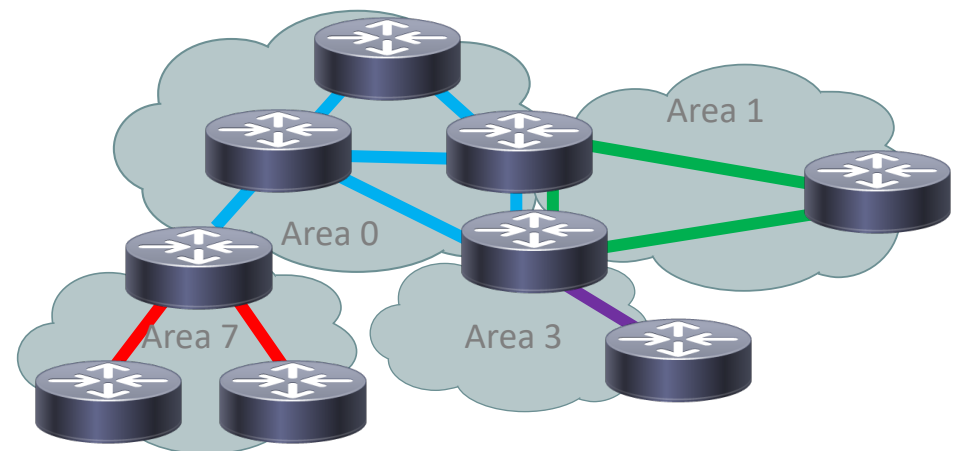
# Синхронизация таблиц топологии

- Состояние 2WAY и FULL стабильные, остальные - переходные



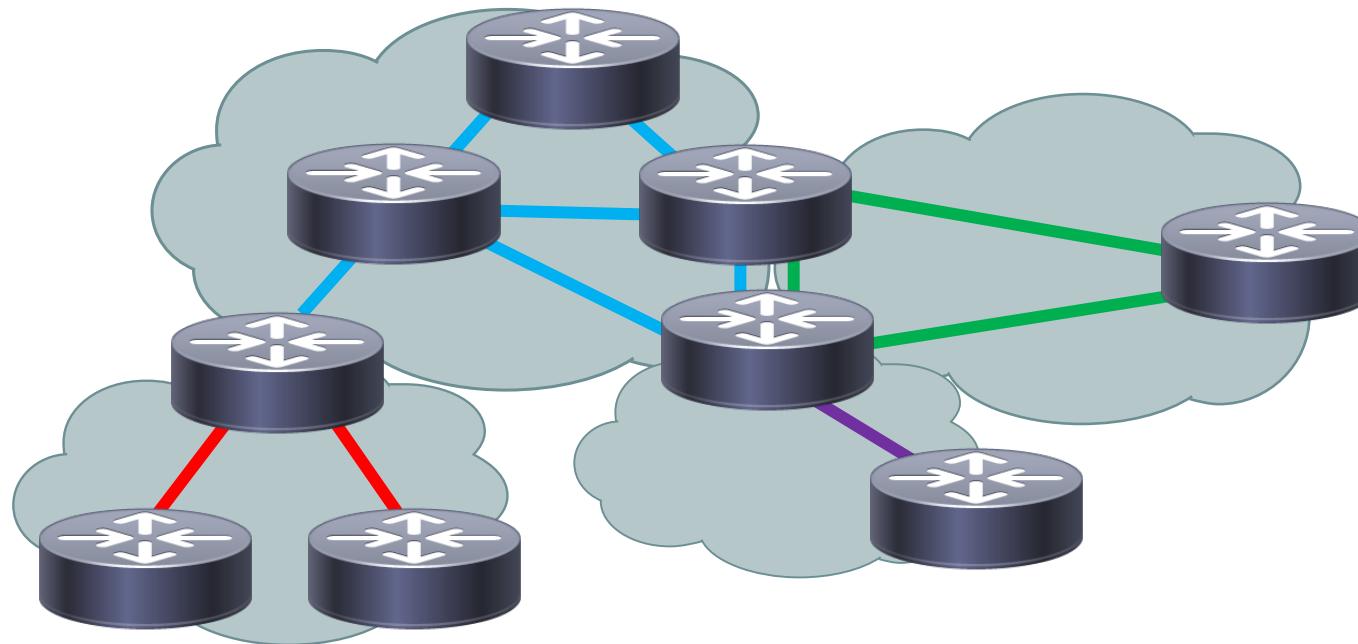
# Регионы в OSPF

- Маршрутизаторы описывают в выпускаемых LSA топологию сети
  - При изменениях в топологии все маршрутизаторы пересчитывают ее
  - Если топология большая и сложная, пересчет занимает много ресурсов
- OSPF позволяет разбить автономную систему на регионы (area)
  - Внутри региона топология считается «по-честному»
  - За пределы региона адресная информация передается в ДВ-стиле
- Классификация маршрутизаторов
  - Internal Router
  - Area Border Router
  - Backbone Router
  - AS Boundary Router



# Иерархическая топология

- OSPF гарантирует защиту от петель
  - Внутри региона – нативно: кратчайшие пути петель содержать не могут
  - Между регионами используется топология "звезда"
- Центральный регион имеет особую роль (0.0.0.0, backbone, transit)





Networking  
For everyone

# OSPF: типы LSA

---

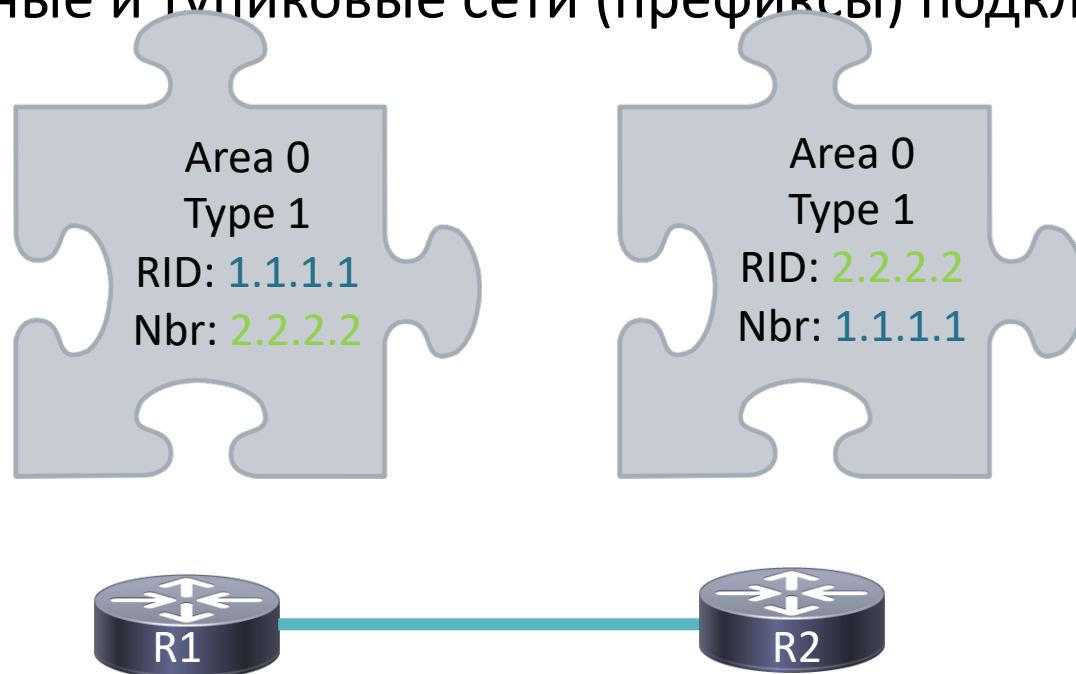
# LSA в OSPFv2

- Тип 1, Router LSA: описывает топологию и подключенные префиксы
- Тип 2, Network LSA: оптимизирует топологию в multiaccess сетях
- Тип 3, Summary LSA: синхронизирует маршруты между регионами
- Типы 4,5,7: используются для инъекции внешних маршрутов
- Типы 9,10,11: Opaque LSA для передачи иной информации
- Типы 6,8: экзотика, не поддерживаются Cisco



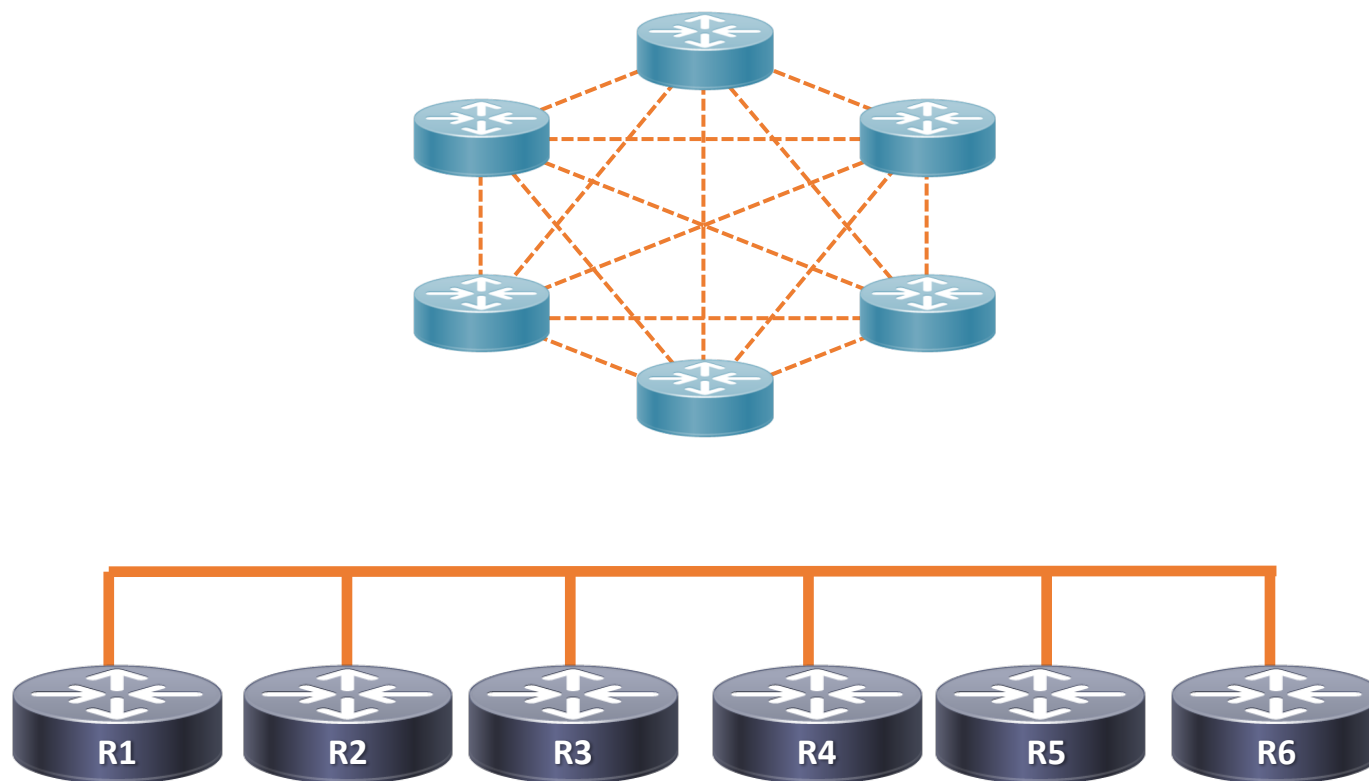
# Router LSA (тип 1)

- Выпускаются каждым маршрутизатором, по одной в каждый регион
- Содержат топологическую и адресную информацию:
  - С какими маршрутизаторами (RID) установлено P2P-соседство
  - Какие транзитные и тупиковые сети (префиксы) подключены



# Проблема: multiaccess-каналы

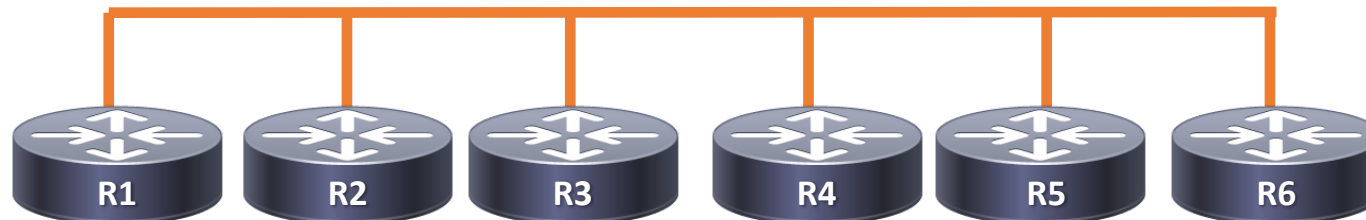
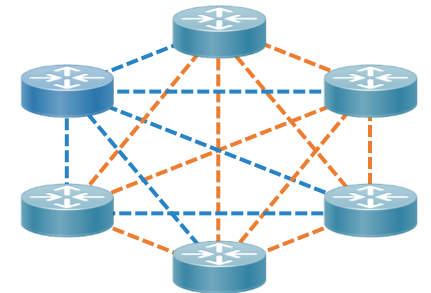
- Если в канале возможно наличие более чем двух маршрутизаторов, полная синхронизация LSDB «каждый с каждым» неэффективна





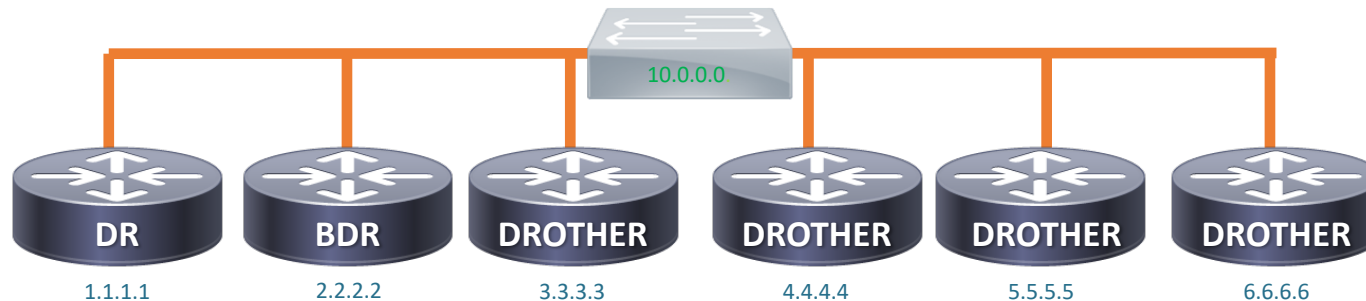
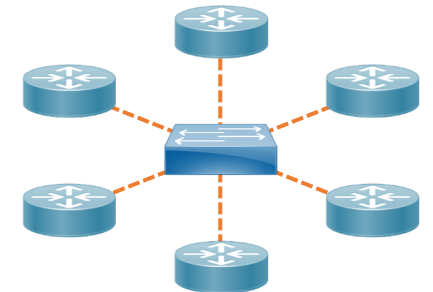
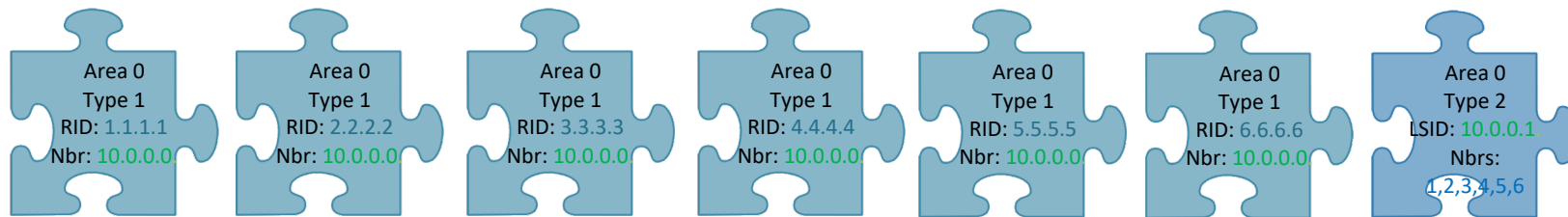
# Designated Router

- Оптимизация: в канале выбирается Designated Router (DR)
  - Максимальный Priority+RID, схема без преемства
  - DR синхронизирует LSDB со всеми маршрутизаторами в канале
  - DROTHER между собой остаются в 2WAY
- Если есть связь с DR, то есть и связь с остальными
  - У всех маршрутизаторов синхронизирована LSDB
  - Для отказоустойчивости выбирается Backup DR (BDR)



# Network LSA (тип 2)

- Designated Router выпускает Network LSA (тип 2) для общего канала:
  - С какими маршрутизаторами DR полностью синхронизировал LSDB
  - Какие подключенные префиксы используются в канале
- В Router LSA (тип 1) вместо информации о соседях указывается информация о подключенной Network LSA



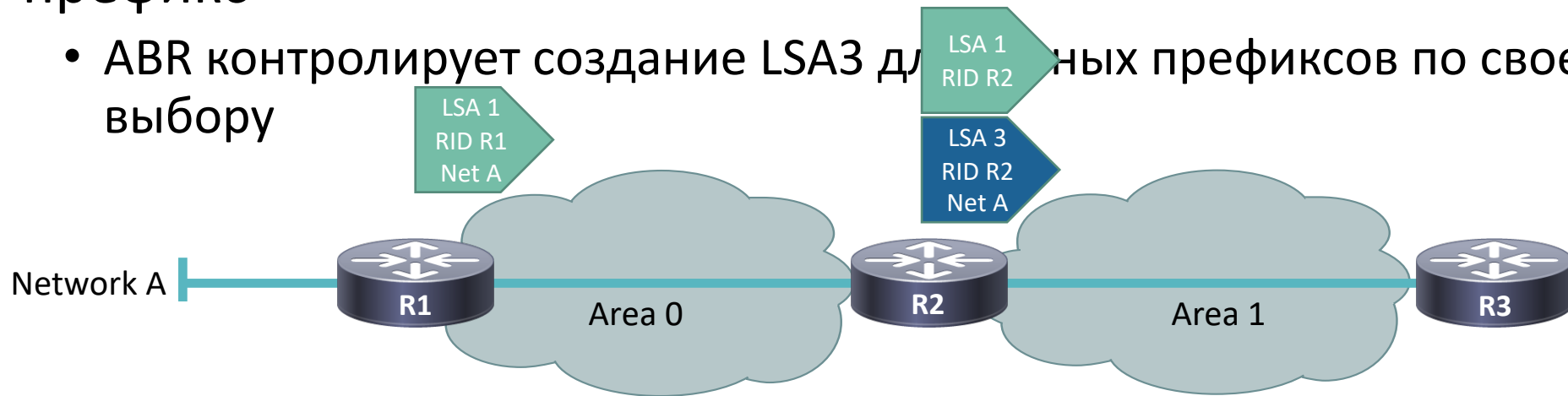
# Проблема: несколько регионов

- Область распространения LSA 1 и 2 типов – регион
  - Все маршрутизаторы в регионе синхронизируют LSDB между собой
  - Маршрутизаторы в других регионах не получают информацию о топологии
  - Вообще говоря, им она не особенно интересна
- Решение: передавать адресную информацию между регионами
  - Маршрутизатор на границе между регионами (ABR, Area Border Router) формирует в другие регионы анонсы об имеющихся префиксах



# Summary LSA (тип 3)

- ABR транслирует в регион префиксы, известные в других регионах
  - Из LSA типов 1 и 2 ненулевых регионов – в LSA типа 3 региона 0
  - Из LSA типов 1, 2 и 3 региона 0 – в LSA типа 3 ненулевых регионов
    - Область распространения LSA3 - регион
  - В LSA3 указывается суммарная стоимость пути от ABR до анонсируемой сети
- Каждый ABR генерирует по одной Summary LSA на каждый префикс
  - ABR контролирует создание LSA3 для своих префиксов по своему выбору



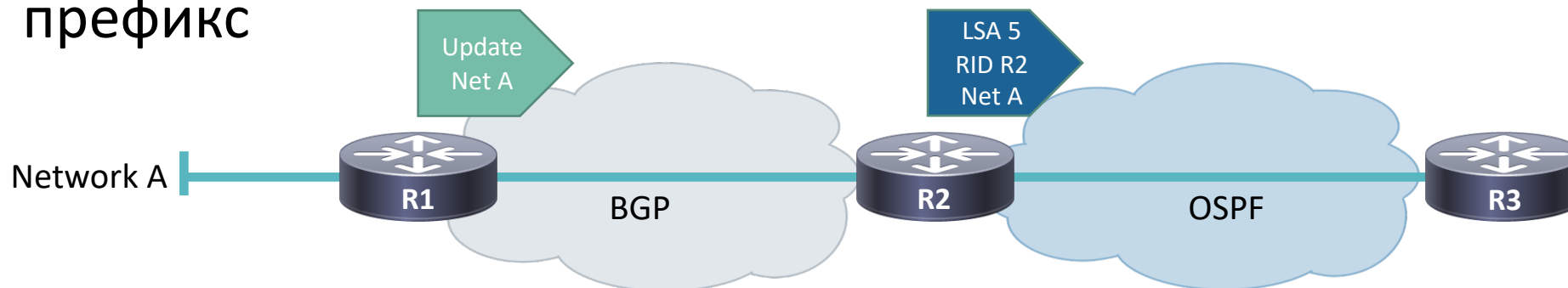
# Проблема: внешние для OSPF маршруты

- Маршрутизатор OSPF в LSA1/2 может анонсировать соседям только подключенные (directly connected) префиксы с интерфейсов OSPF
- Соседям может быть интересно знать про другие префиксы из RIB
  - Статические маршруты
  - Динамические маршруты, полученные в другом протоколе маршрутизации
  - Подключенные префиксы на интерфейсах, не включенных в OSPF
- Решение: распространять эту информацию по аналогии с LSA3
  - Маршрутизатор на границе автономной системы (ASBR, Autonomous System Boundary Router) формирует анонс об имеющихся внешних префиксах



# External LSA (тип 5)

- ASBR транслирует в AS префиксы, известные из других источников
  - Область распространения LSA5 – вся автономная система
  - В LSA5 указывается ASBR RID и некоторая стоимость пути от ASBR до сети
    - Может также указываться Forwarding Address – IP-адрес, через который пойдет трафик
  - LSA5 бывают двух типов:
    - Тип 1: указывается стоимость, посчитанная по схожей с OSPF методике
    - Тип 2: указывается стоимость, посчитанная по принципиально отличной методике
- Каждый ASBR генерирует по одной External LSA на каждый префикс



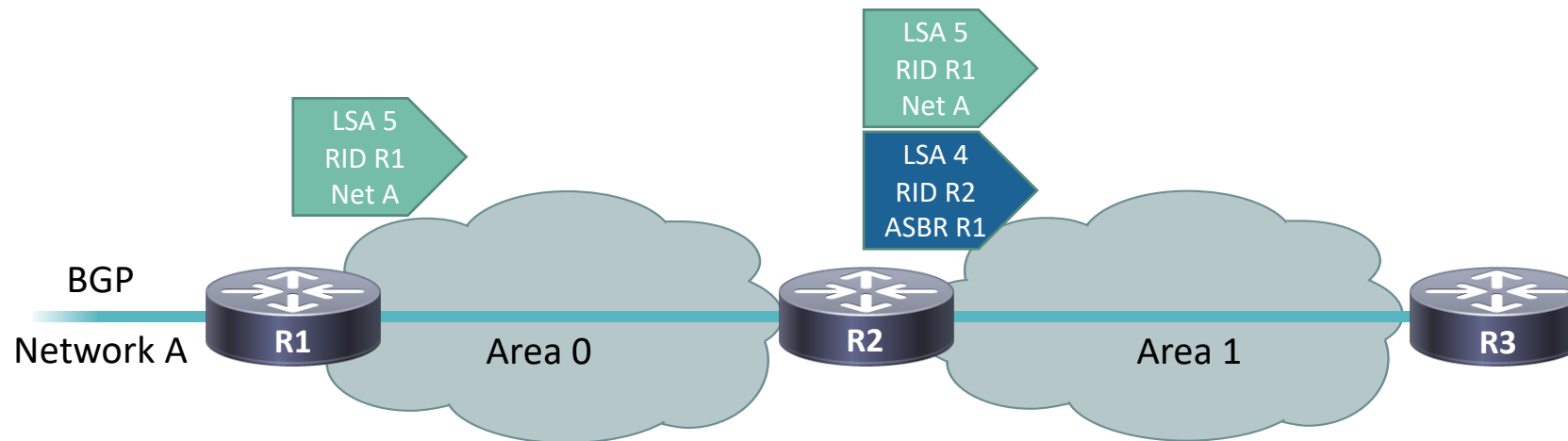
# Проблема: как добраться до ASBR?

- В LSA5 указывается внешняя стоимость маршрута и RID ASBR
  - Router ID – топологическая информация, видна в LSA1
  - За пределами региона маршрут до ASBR по LSA1/2 построить невозможно
- Решение: распространять информацию об ASBR по аналогии с LSA3
  - Маршрутизатор на границе между регионами (ABR, Area Border Router) формирует в другие регионы анонсы об известных ему ASBR



# ASBR Summary LSA (тип 4)

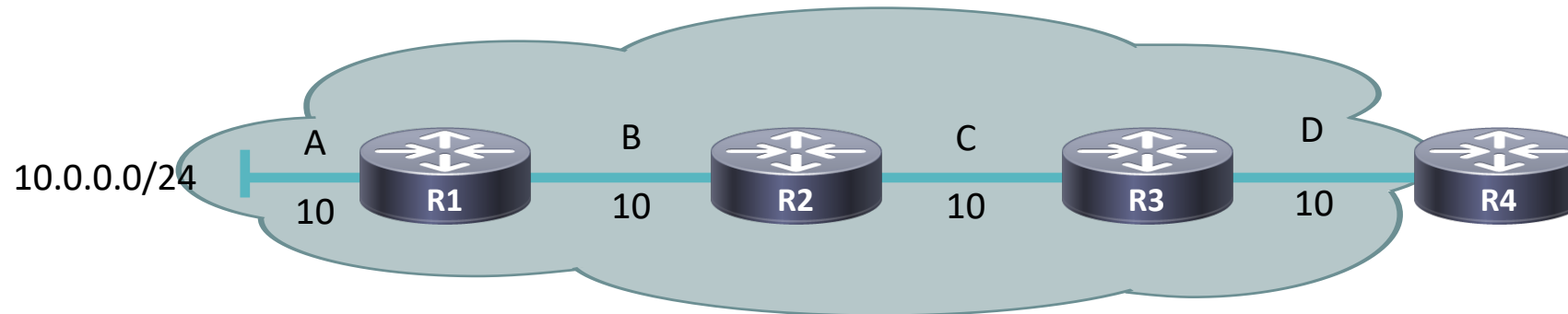
- ABR транслирует в регион информацию об известных ASBR
  - ASBR Router ID
  - Суммарную стоимость пути от ABR до ASBR
  - Область распространения LSA4 - регион
- Каждый ABR генерирует по одной LSA4 на каждый известный ASBR





# Метрика intra-area маршрута OSPF

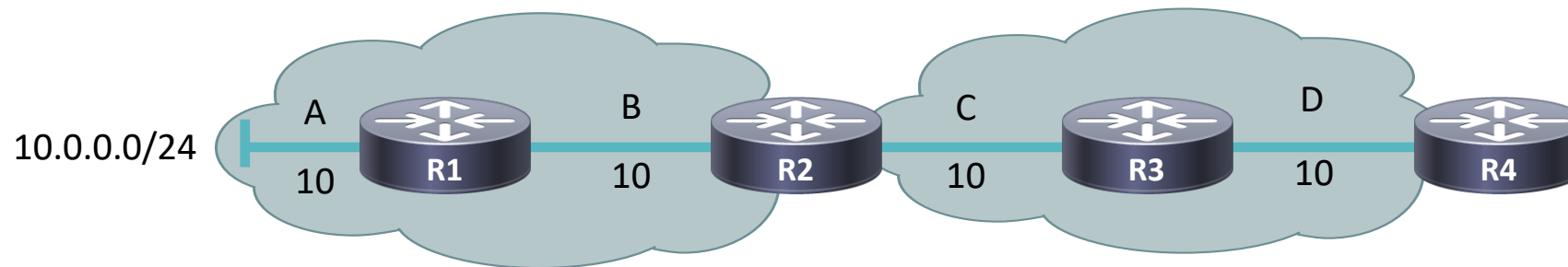
- Складывается из двух компонентов:
  - Стоимость анонсированной сети на LSA-источника (0 для LSA2)
  - Суммарная стоимость пути до LSA-источника (вычисляется из LSA1)



Маршрутизатор	Метрика сети 10.0.0.0/24
R1	A (подключенная сеть)
R2	A (анонсированная стоимость на R1) + B (суммарная стоимость пути до R1)
R3	A (анонсированная стоимость на R1) + B+C (суммарная стоимость пути до R1)
R4	A (анонсированная стоимость на R1) + B+C+D (суммарная стоимость пути до R1)

# Метрика inter-area маршрута OSPF

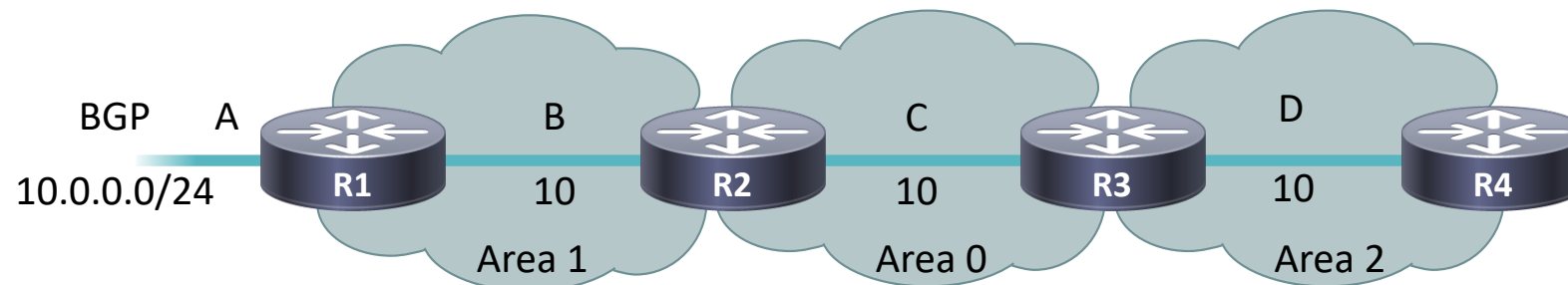
- Складывается из двух компонентов:
  - Стоимость анонсированной сети на ABR (указывается в LSA3)
  - Суммарная стоимость пути до ABR (вычисляется из LSA1)



Маршрутизатор	Метрика сети 10.0.0.0/24
R1	A (подключенная сеть)
R2	A (анонсированная стоимость на R1) + B (суммарная стоимость пути до R1)
R3	A+B (стоимость LSA3 от R2) + C (суммарная стоимость пути до R2)
R4	A+B (стоимость LSA3 от R2) + C+D (суммарная стоимость пути до R2)

# Стоимость пути до ASBR

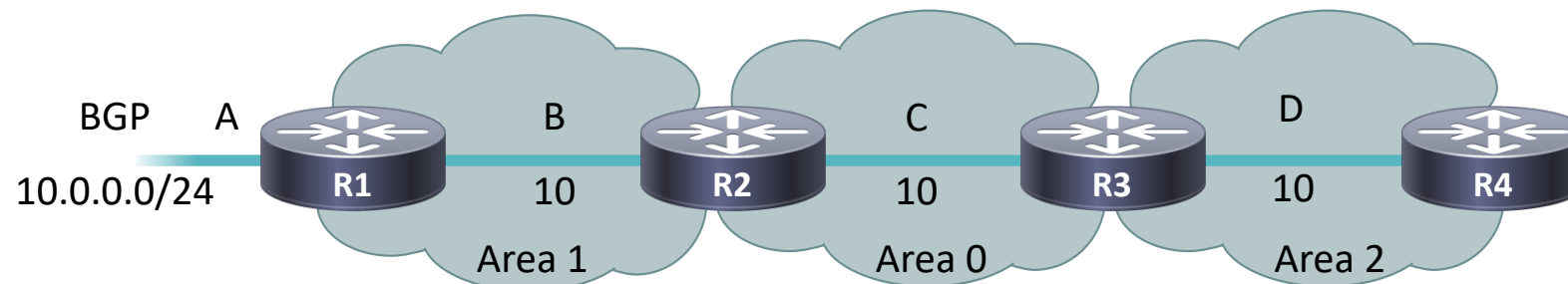
- Складывается из двух компонентов:
  - Суммарная стоимость пути от ABR до ASBR (указывается в LSA4)
  - Суммарная стоимость пути до ABR (вычисляется из LSA1)



Маршрутизатор	Стоимость пути до ASBR (R1)
R1	-
R2	B (суммарная стоимость пути до R1)
R3	B (стоимость LSA4 от R2) + C (суммарная стоимость пути до R2)
R4	B+C (стоимость LSA4 от R3) + D (суммарная стоимость пути до R3)

# Метрика external маршрута OSPF

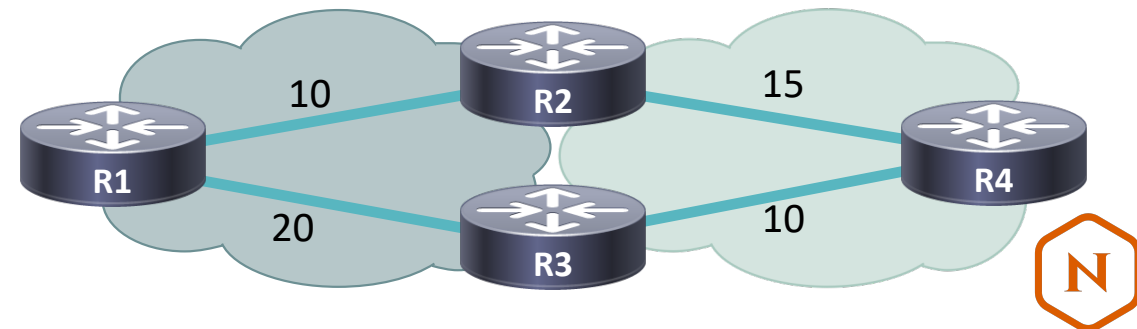
- Состоит из двух независимых компонентов:
  - Стоимость пути от ASBR до внешней сети (указывается в LSA5)
  - Суммарная стоимость пути до Forwarding Address или (при FA=0) ASBR RID (вычисляется из LSA4 и LSA1)



Маршрутизатор	Внешняя стоимость	Стоимость пути до ASBR (R1)
R1	-	-
R2	A (стоимость в LSA5)	B (суммарная стоимость пути до R1)
R3	A (стоимость в LSA5)	B (стоимость LSA4 от R2) + C (суммарная стоимость пути до R2)
R4	A (стоимость в LSA5)	B+C (стоимость LSA4 от R3) + D (суммарная стоимость пути до R3)

# Метрика E1 и E2

- LSA5 бывают двух типов:
  - Тип 1: внешняя стоимость считается по схожей с OSPF методике
  - Тип 2: внешняя стоимость считается по отличной методике
- Компоненты метрики в LSA5 типа 1 можно складывать друг с другом
  - Если два ASBR предлагают LSA5 типа 1, OSPF предпочтет маршрут с минимальной суммой внешней и внутренней стоимости
- Компоненты метрики в LSA5 типа 2 нельзя складывать друг с другом
  - Если два ASBR предлагают LSA5 типа 2, OSPF предпочтет маршрут с минимальной внешней стоимостью



# Предпочтения маршрутов по RFC 2328

- LSA1: intra-area маршруты
- LSA3: inter-area маршруты
- LSA5 тип 1: внешние маршруты со сравнимой с OSPF стоимостью
- LSA5 тип 2: внешние маршруты с несравнимой с OSPF стоимостью



# LS Type

- Point-to-point neighborhood (соседняя LSA1)



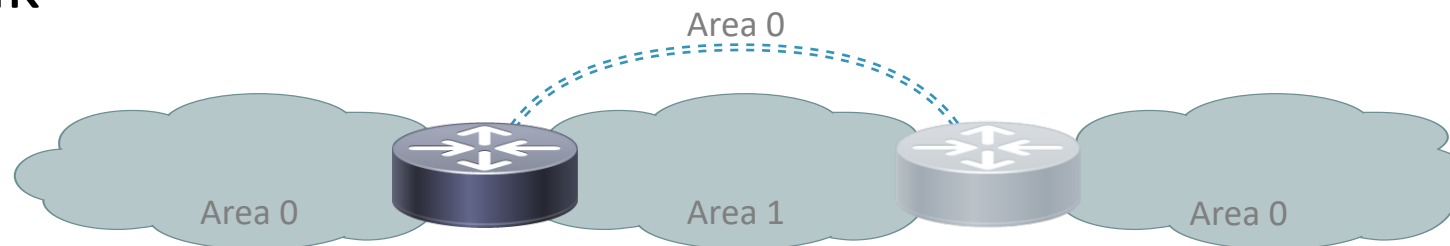
- Transit network (соседняя LSA2)



- Stub network (адресная информация)



- Virtual Link





Networking  
For everyone

# OSPF: Convergence and Scalability

---



# В этом разделе

- Loop Free Alternate
- Nonstop Forwarding
- Nonstop Routing
- BFD
- Prefix Suppression
- Stub Router





Networking  
For everyone

Loop Free Alternate

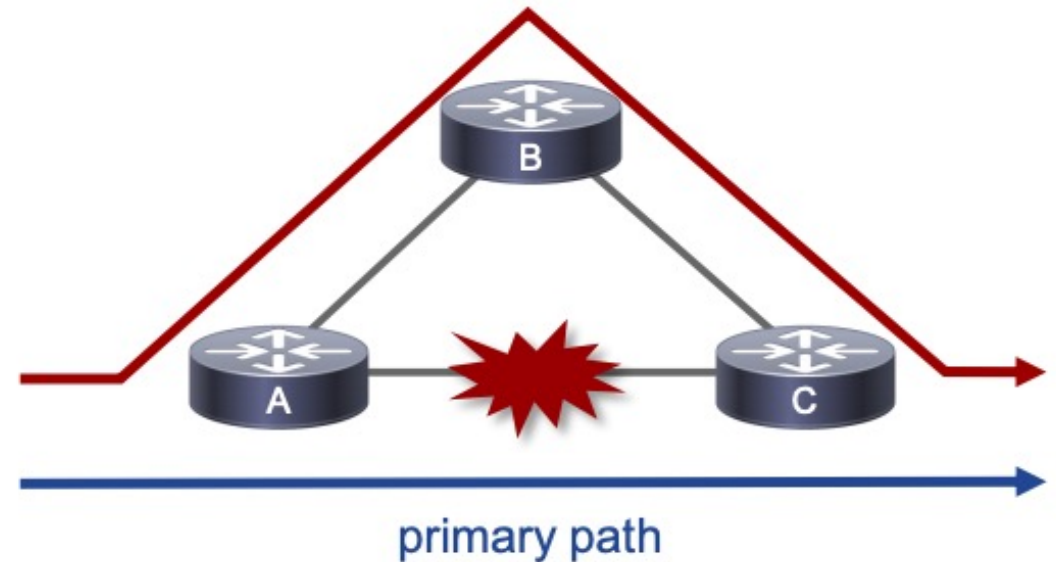
# Loop Free Alternative

- OSPF LFA FRR позволяет быстро (в течение  $\sim 50$  мс) переключаться на резервный путь
- В обычной ситуации, OSPF должен пересчитать весь граф в случае выхода интерфейса из строя
- С LFA FRR, OSPF делает **предварительный расчёт**
  - Резервный next-hop устанавливается внутрь FIB



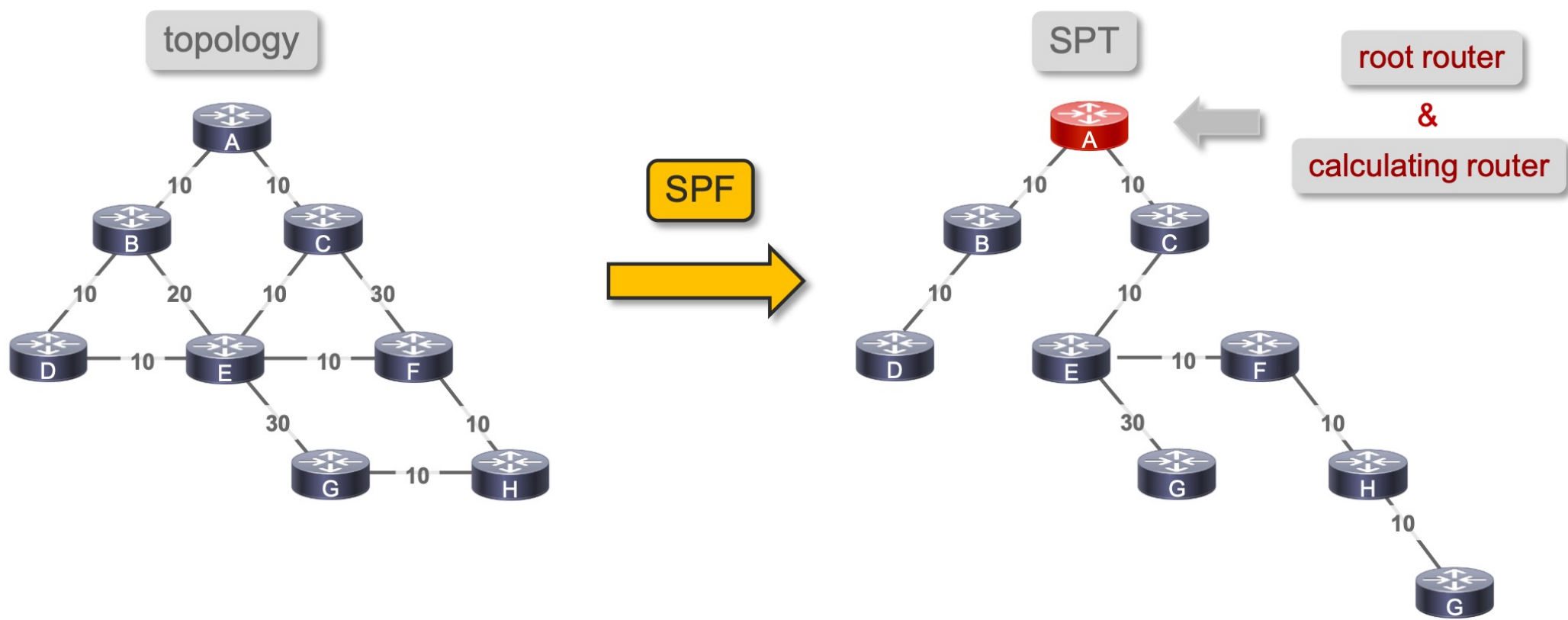
# Основная идея

- Маршрутизатор А делает всю калькуляцию
- Другие маршрутизаторы не вовлечены в процесс
- Repair Path (LFA):
  - Трафик от В не должен вернуться к А
  - Трафик должен миновать упавший интерфейс



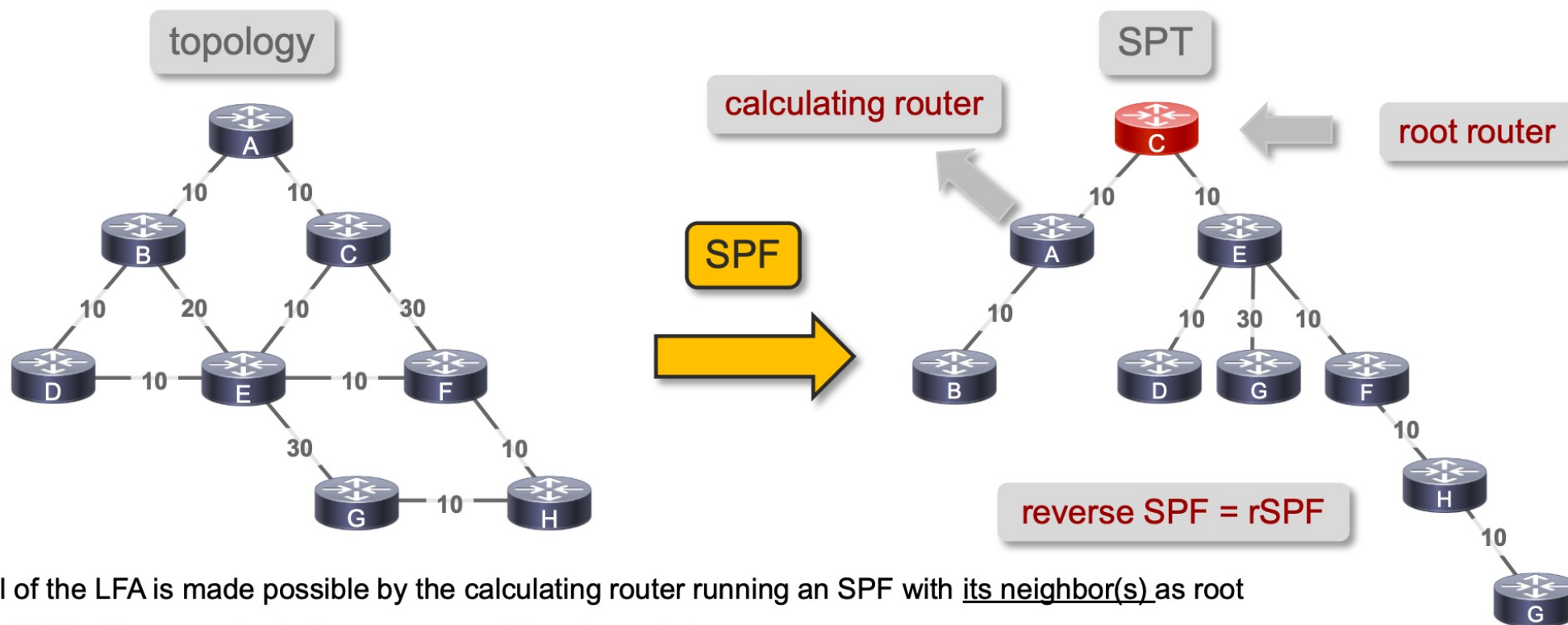
# Классический SPF

- Необходимо запустить SPF и в качестве корневого устройства поставить **себя**



# Хак SPF для LFA (rSPF)

- Необходимо запустить SPF, но в качестве корневого устройства поставить **соседа**

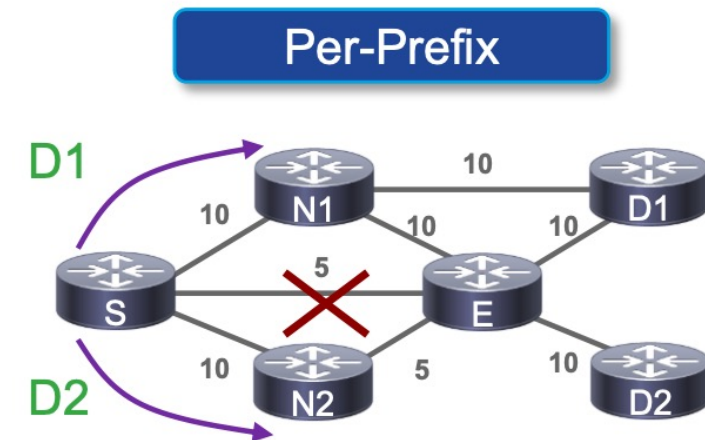
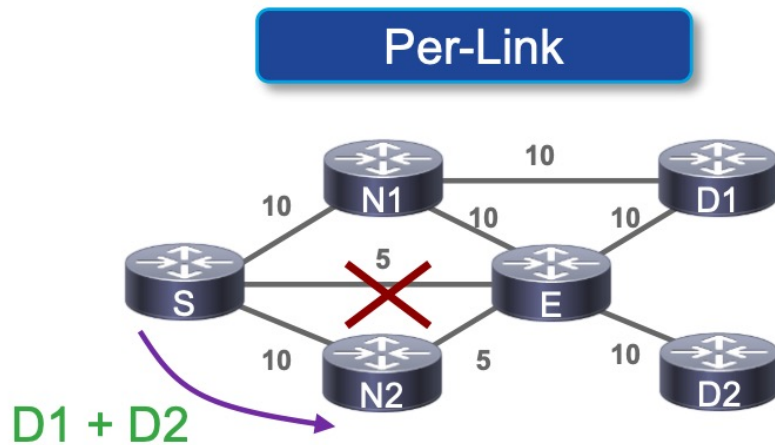


All of the LFA is made possible by the calculating router running an SPF with its neighbor(s) as root

An SPF with any router in the area as root is not needed

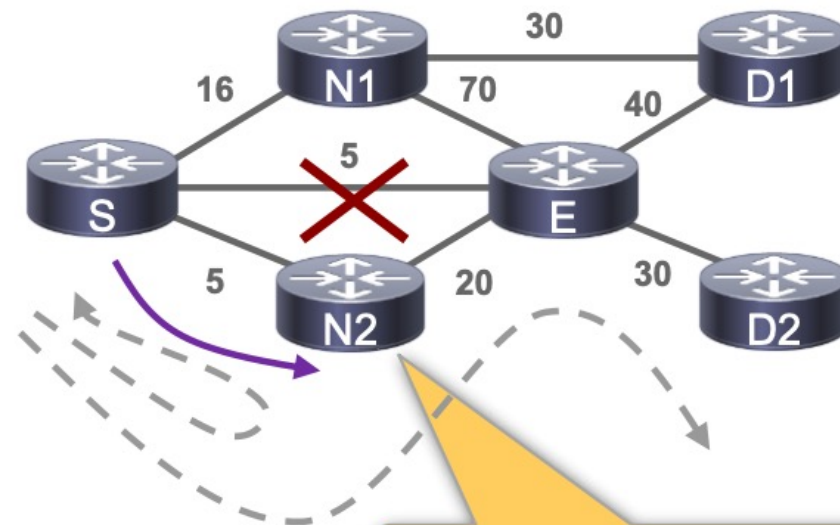
# LFA методы

- IGP может запускать LFA в одном из двух режимов
  - Per prefix
    - Резервный путь для каждого префикса считается независимо
  - Per link
    - Резервный путь для всех префиксов через один next-hop
    - не поддерживается в IOS-XE



# А что если ...

- Трафик к D2 передаётся на E от N2
- Трафик к D1 возвращается обратно к S



Prefix from D1, NH = S, cost = 51  
Prefix from D2, NH = E, cost = 50



# Per-link vs Per-prefix

## Per-link

- Простой расчёт, один rSPF для соседа
- Всё или ничего

## Per-prefix

- Расчёт для каждого префикса через каждого соседа

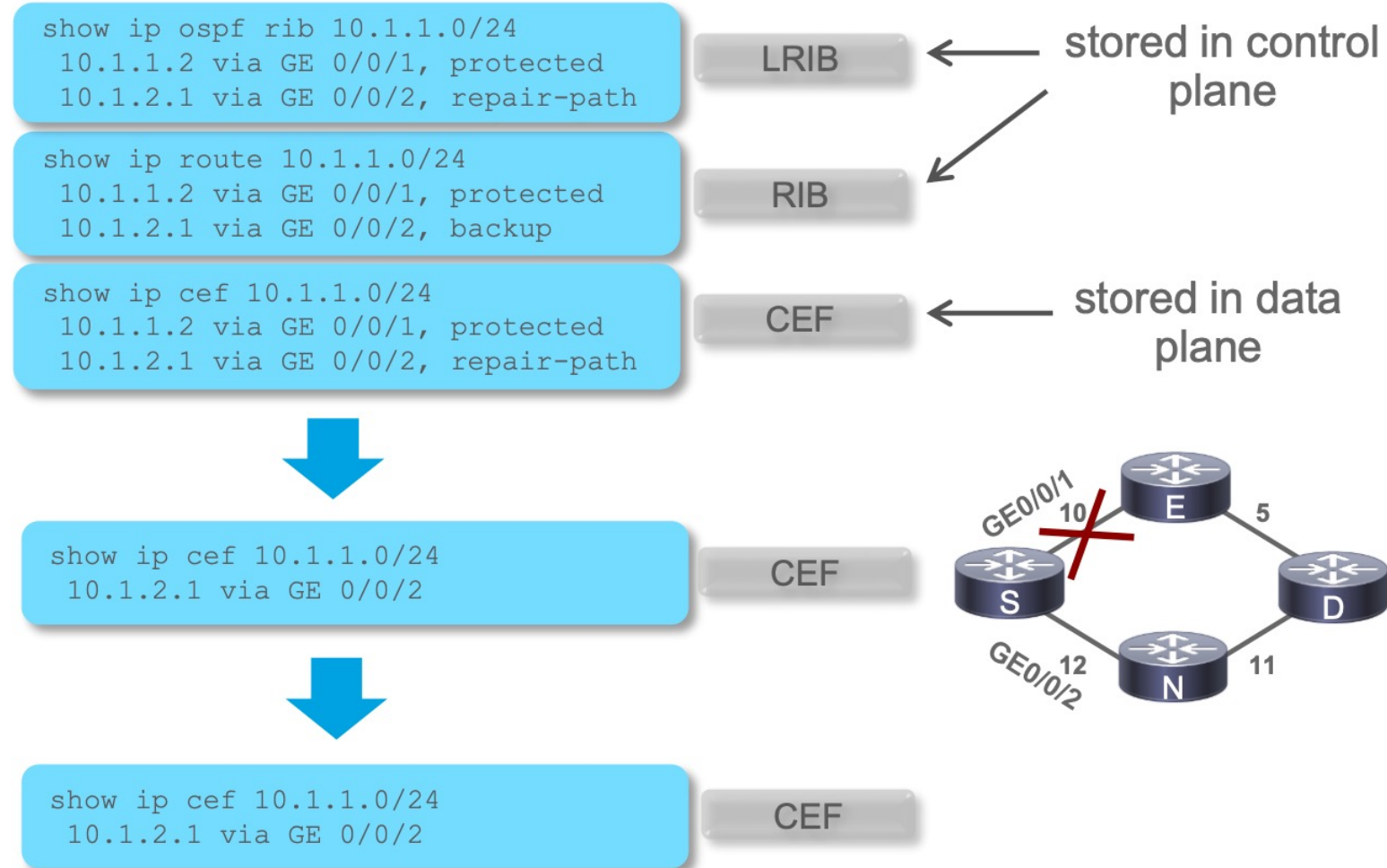


# Основные блоки

- До аварии
  - Альтернативный NH устанавливается в RIB и IGP local RIB (LRIB)
  - Альтернативный NH устанавливается в FIB (CEF)
- Во время аварии
  - Детектирования потери интерфейса/соседа
  - Триггер для IP-FRR LFA: переключить префиксы в FIB
- После аварии
  - Обычная конвергенция (SPF)



# Основные блоки



# А если альтернатив несколько?

- Когда OSPF должен выбрать резервный путь, он может посмотреть не только на наименьшую метрику, но и учитывать дополнительные параметры
  - SRLG (Shared Risk Link Groups)
  - Interface Protection
  - Broadcast Interface Protection
  - Node Protection
  - Downstream Path
  - Line-Card Disjoint Interfaces
  - Metric
  - Equal-Cost Multipath



# Shared Risk Link Group (SRLG)

- Ручная настройка
- Если два интерфейса подключены к одному коммутатору, идут через одну физическую трассу, то логично назначить им одинаковые SRLG



# Primary Path



# Interface Disjoint

- Предпочесть альтернативный next-hop, который располагается за другим интерфейсом
- Ethernet суб-интерфейсы считаются ***разными*** интерфейсами



# Lowest-Metric

- Предпочесть путь с наименьшей метрикой
- У команды нет ключевого значения “required”
  - Т.к. метрика присутствует всегда 😊





# Linecard-disjoint

- Предпочесть путь, который использует интерфейс, находящийся на другой линейной карте
  - В виртуальной лаборатории не удастся эмулировать



# Node protecting

- Предпочесть путь, который не проходит через тот же маршрутизатор, который используется в качестве основного next-hop



# Broadcast interface disjoint

- Понизить приоритет альтернативным маршрутам, которые используют тот же широковещательный домен, что и путь через основной next-hop



# Downstream

- Выключен по-умолчанию
- По сути поведение очень похоже на EIGRP Feasible Condition
- Мне не очень понятно, зачем такая опция вообще нужна в LS протоколе



# Secondary-Path

- Выключен по-умолчанию
- Предпочесть пусть, который не является частью ESMR



# Основные шаги конфигурации

- Включить FRR для зоны или глобально
- Включить FRR prefix-priority

```
(config-router)#fast-reroute per-prefix enable prefix-priority { high | low }
```

- Настроить приоритет префиксов (route-map | RPL)
  - /32 = “high” на IOS
  - только match tag | route-type | ip address

```
(config-router)#prefix-priority high route-map { ROUTE-MAP }
```

- Добавить/изменить tie-breakers

```
(config-router)#fast-reroute per-prefix tie-break { TIE } [required] index { INDEX }
```



# Пример конфигурации

```
interface Ethernet1/0
  srlg gid 100
!
interface Ethernet6/0
  ip ospf fast-reroute per-prefix candidate disable
!
router ospf 1
  prefix-priority high route-map lfa-ospf
  fast-reroute per-prefix enable prefix-priority high
  fast-reroute per-prefix tie-break srlg index 10
  fast-reroute per-prefix tie-break node-protecting index 20
  fast-reroute keep-all-paths
!
ip prefix-list lfa-high seq 5 permit 10.0.0.0/8 ge 30
!
route-map lfa-ospf permit 10
  match ip address prefix-list lfa-high
```



# Пример конфигурации

```
R9#show ip cef 5.5.5.5
5.5.5.5/32
  nexthop 10.5.9.5 GigabitEthernet1.59
    repair: attached-nexthop 10.9.5.5 GigabitEthernet1.95
  nexthop 10.9.5.5 GigabitEthernet1.95
    repair: attached-nexthop 10.5.9.5 GigabitEthernet1.59
```

```
R9#show ip route repair-paths 5.5.5.5
Routing entry for 5.5.5.5/32
  Known via "ospf 1", distance 110, metric 2, type intra area
  Last update from 10.5.9.5 on GigabitEthernet1.59, 00:01:40 ago
  Routing Descriptor Blocks:
    10.9.5.5, from 5.5.5.5, 00:01:40 ago, via GigabitEthernet1.95
      Route metric is 2, traffic share count is 1
      Repair Path: 10.5.9.5, via GigabitEthernet1.59
  * 10.5.9.5, from 5.5.5.5, 00:01:40 ago, via GigabitEthernet1.59
    Route metric is 2, traffic share count is 1
    Repair Path: 10.9.5.5, via GigabitEthernet1.95
```







Networking  
For everyone

Non-Stop Forwarding  
(Graceful Restart)

# NSF? Wtf?

- Мы уже познакомились с понятием FRR – быстрая конвергенция в случае изменения сети
- Часто в ядре стоят устройства, у которых есть два управляющих модуля (Supervisor)
- К чему приведёт переключение SUP?
- Основной идеей NSF (Nonstop Forwarding) является минимизация времени, в течение которого сеть не способна доставлять трафик в место назначения после события переключения SUP
- Основной целью NSF является продолжать пересылку IP-пакетов, после того, как на устройстве произошло переключение SUP



# Логика работы NSF

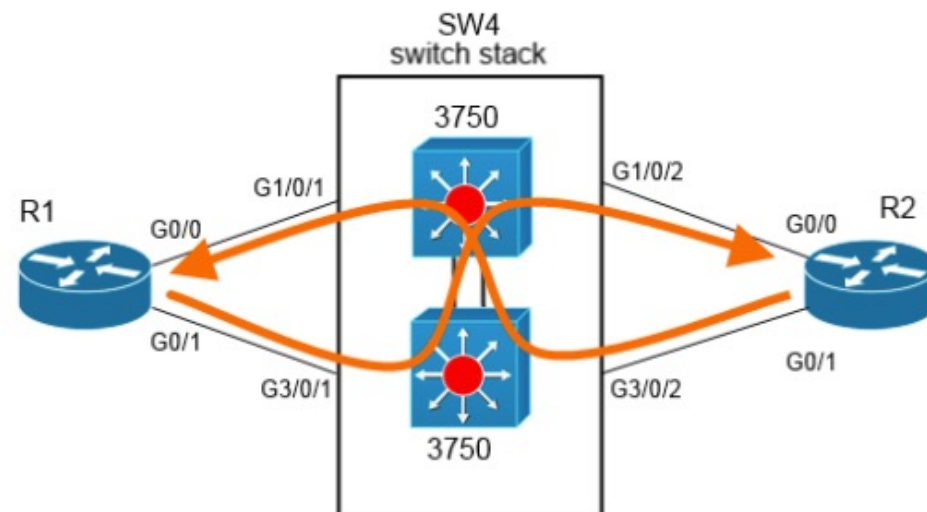
- В случае события переключения SUP (или RP switchover), новый RP перехватывает управление устройством и запускает новый процесс маршрутизации
- Новый процесс OSPF пытается восстановить соседство (adjacency) со всеми предыдущими соседями
- Если NSF настроен для прокола, CEF (FIB) фиксируется (для того, чтобы устройство продолжало пересылку трафика, не смотря на то, что RIB (таблица маршрутизации) пуста
- Во время восстановления соседства, ни одно из устройств не извещает остальную сеть о событии
- Только когда прокол маршрутизации завершает передачу всей необходимой информации, данные попадают в RIB и (оттуда) могут обновить CEF



# NSF может таить в себе проблемы

- Падает SW4 (1 мембер)
- длительное переключение вызвано фиксацией CEF на время работы процесса NSF

```
%LINK-3-UPDOWN: Interface GigabitEthernet0/0, changed state to down  
TACKMGR-4-SWITCH_REMOVED: Switch 1 has been REMOVED from the stack  
12:26:03.996: OSPF: IETF NSF complete check for area 0 process 1  
OSPF: will poll [count 10] interface status for GigabitEthernet3/0/1  
OSPF: Graceful Restart timer expired for process 1, terminating IETF NSF
```



# Основные шаги

- Устройство информирует соседей, что процесс OSPF перегружается
  - Graceful restart mode
- Отправляется grace LSA (LSA 9-го типа)
- Соседи отправляют LS ACK и переходят в Helper Mode
- В течение grace периода, соседи работают так, как если бы restarting router не сообщал об изменениях
- В течение перезагрузки OSPF процесса, RIB/FIB не изменяется
- После перезагрузки, переустанавливается OSPF соседство



# Основные шаги

- Когда GR завершён, restarting router удаляет Grace LSA
- Пересоздаются все LSA, порождённые маршрутизатором
- Запускается SPF чтобы освежить таблицу маршрутизации



# Примечания

- Для работы GR необходима аппаратная поддержка функционала
- Устройства, которые поддерживают Helper Mode = NSF-aware
- Устройства, которые поддерживают GR = NSF-capable





Networking  
For everyone

Non-Stop Routing



# Проблемы с GR

- Все маршрутизаторы должны поддерживать механизм GR для конкретного протокола
- Switchover может быть наиболее неприятным на PE
  - Маленькие CE могут не поддерживать GR
- В некоторых ситуациях GR может замедлить конвергенцию сети



# Nonstop Routing, о чём это?

- NSR использует внутренние процессы маршрутизатора для поддержания копии Control Plane на резервном управляющем модуле в актуальном состоянии
- Switchover абсолютно прозрачен для всех соседей



# Ограничения

- OSPF NSR может потребовать большого количества памяти
- Переключение между управляющими модулями занимает ~ 2 сек. В течение этого времени OSPF не может отправлять Hello сообщения
  - Аккуратнее с маленькими таймерами

```
Router# show ip ospf 1 nsr
Standby RP
Operating in duplex mode
Redundancy state: STANDBY HOT
Peer redundancy state: ACTIVE
ISSU negotiation complete
ISSU versions compatible
Routing Process "ospf 1" with ID 10.1.1.100
NSR configured
Checkpoint message sequence number: 3290
Standby synchronization state: synchronized
```





Networking  
For everyone

# Bidirectional Forwarding Detection

# BFD

- BFD – очень легковесный и быстрый протокол, предназначенный для определения нарушения сетевой связности
- BFD работает независимо
- Может работать в двух режимах
  - асинхронный
  - по-требованию (demand)
    - не уверен, что кто-либо из вендоров его реализовал



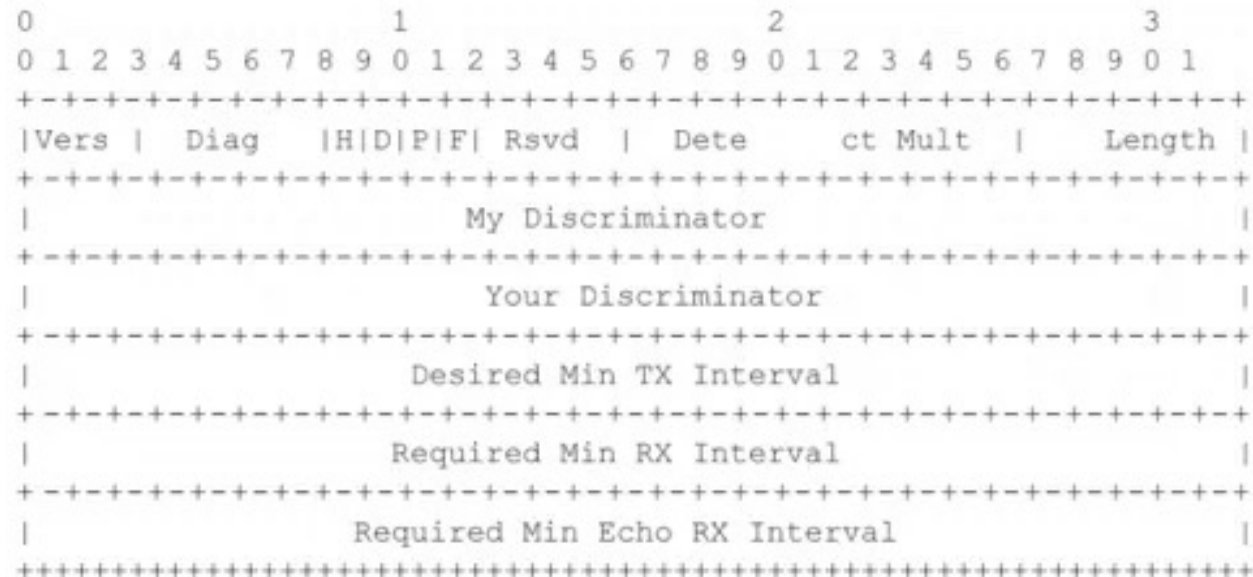
# Асинхронный режим

- Наиболее классический режим Hello/Holddown
- BFD отправляет Hello. И если не видит сообщений от соседа — регистрируется факт недоступности
- Данный триггер передаётся всем протоколам, которые подписаны на BFD
- Возможно включение функционала Echo
  - Сосед не обрабатывает прилетающие BFD Hello, а просто отправляет их обратно



# Формат пакета

- Diag – код, описывающий причину перехода состояния сессии из UP во что-либо другое
- Дискриминатор – мультиплексирование сессий
- Min Interval:
  - desired TX = предпочитаемый интервал
  - required RX = минимально поддерживаемый
- БИТЫ:
  - H = “I Hear You”
  - P = “Poll”
  - F = “Final”



# Настройка

```
(config-if)#bfd interval 300 min_rx 600 multiplier 3
```

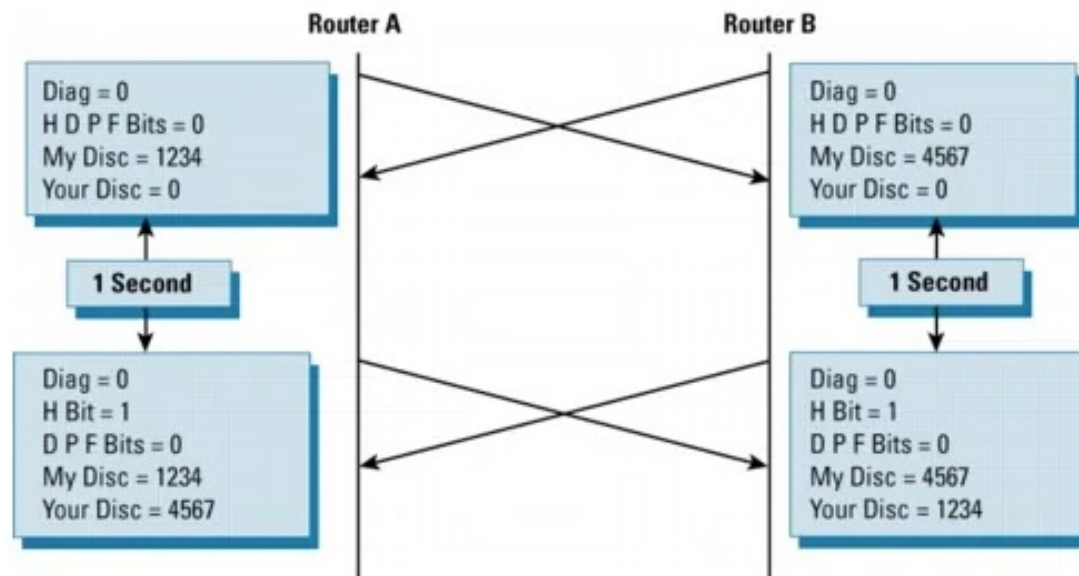
- interval = как часто устройство отправляет BFD пакеты
- min\_rx = как часто ожидаем приём BFD пакетов





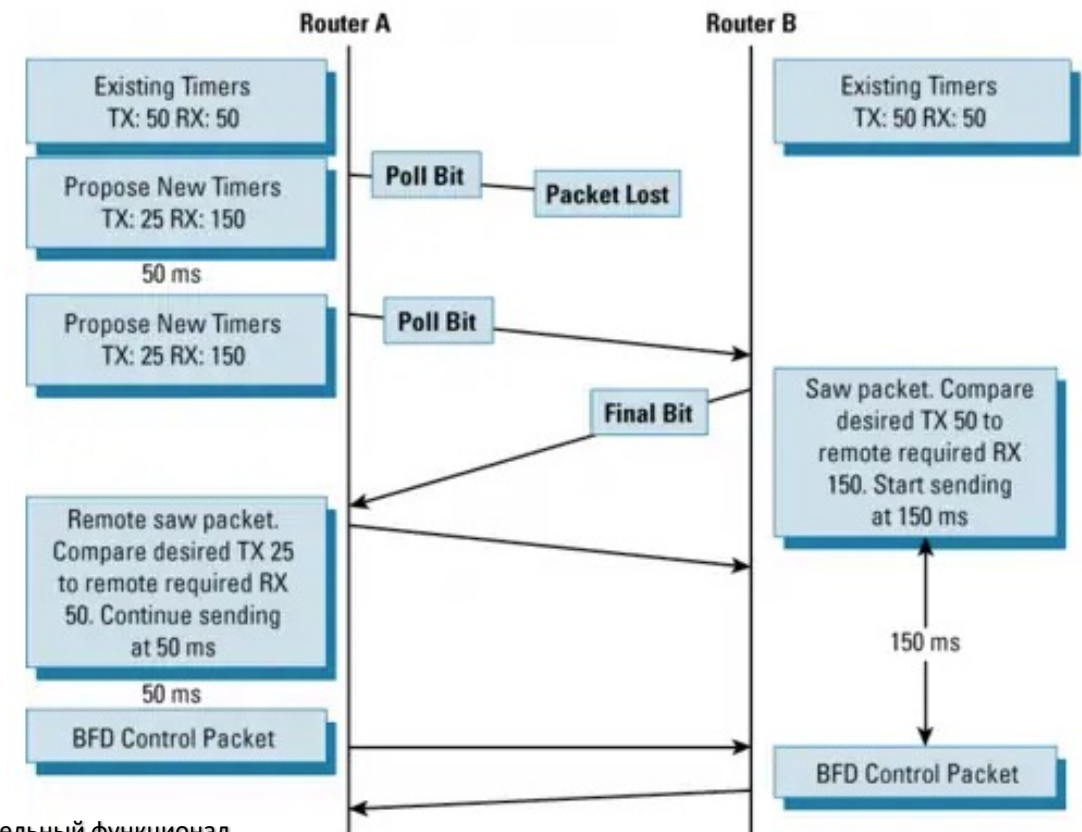
# Установка сессии

- BFD не изучает IP адреса соседей
  - их сообщает привязанный протокол (напр. OSPF)
- Все пакеты передаются с помощью UDP
- Зарезервированный Destination Port
- Если видим H бит и Your Disc поле, то сессия установлена



# Изменение таймеров

- При изменении таймеров, устройство выставляет Р бит
- Если удалённый маршрутизатор увидел Р бит, в ответном сообщении выставляется F бит
  - это не подтверждение принятия изменений
- Не требует переустановления



# Потеря BFD соседа

- Если не приходит контрольный пакет в течение detect-timer  $[(Required\ Minimum\ RX\ Interval) * (Detect\ Multiplier)]$ , то сосед помечается как потерянный
- Сам BFD на это никак не реагирует
- Факт потери соседа передаётся привязанному протоколу, который реагирует на данное событие



# BFD на модульных платформах

- На RSP располагается BFD сервер, а на линейной карте BFD агент
- BFD сервер получает информацию об IP адресах соседей
- BFD агент создаёт все сессии
- Все BFD пакеты отправляются на CPU линейной карты
  - если включен HW Offload, то BFD обрабатывается на Network Processor (NP)
    - позволяет увеличить количество поддерживаемых BFD сессий



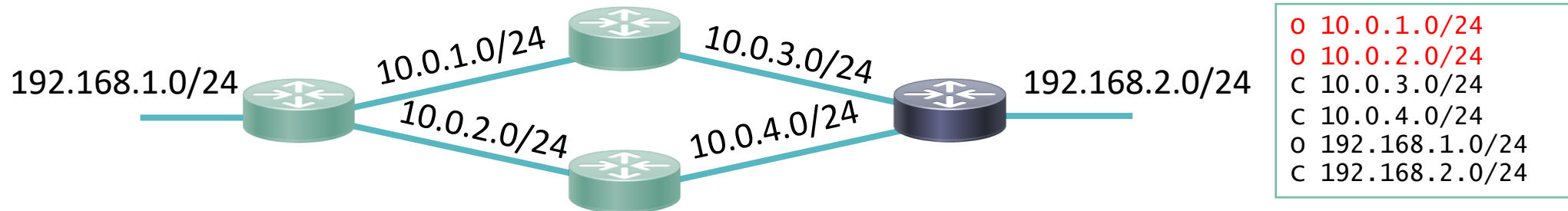


Networking  
For everyone

Соккрытие транзитных  
сетей

# Соккрытие транзитных сетей

- Штатно в OSPFv2 префиксы транзитных сетей попадают в RIB



- RFC 6860 позволяет OSPF сократить количество записей в RIB
  - На интерфейсах P2P и P2MP не анонсируются connected-сети
    - В LSA1 создается point-to-point link до RID, но не создается stub-запись для адреса
  - На broadcast и NBMA интерфейсах адреса анонсируются в LSA2
    - В LSA2 отправляется DR IP с маской /32
    - Новые роутеры не устанавливают в RIB маршруты из LSA2 с маской /32



Networking  
For everyone

Тупиковый  
маршрутизатор

# Тупиковый маршрутизатор

- Иногда (напр. во время проведения миграционных работ) необходимо отвести трафик от маршрутизатора
- Сделать это можно, выставив максимальную метрику для OSPF интерфейсов
- В определённых ситуациях надо дождаться конвергенции BGP

```
R1(config-router)#max-metric router-lsa on-startup ?
```

```
<5-86400>  Time, in seconds, router-LSAs are originated with max-metric
```

```
wait-for-bgp  Let BGP decide when to originate router-LSA with normal metric
```







Networking  
For everyone