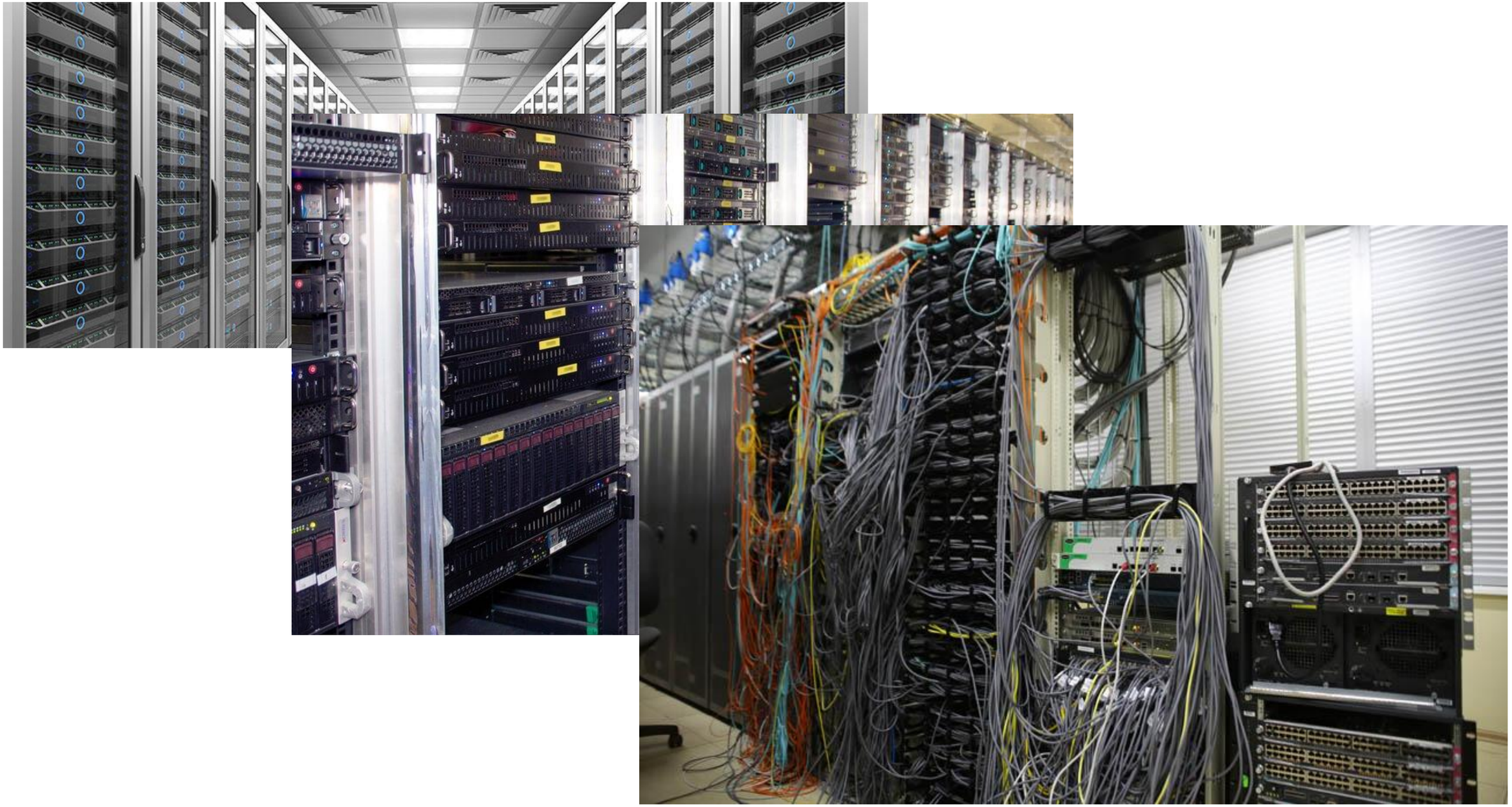
Вычислительные сети и Телекоммуникации

Сети датацентров и виртуализация

2019/2020, 8 семестр

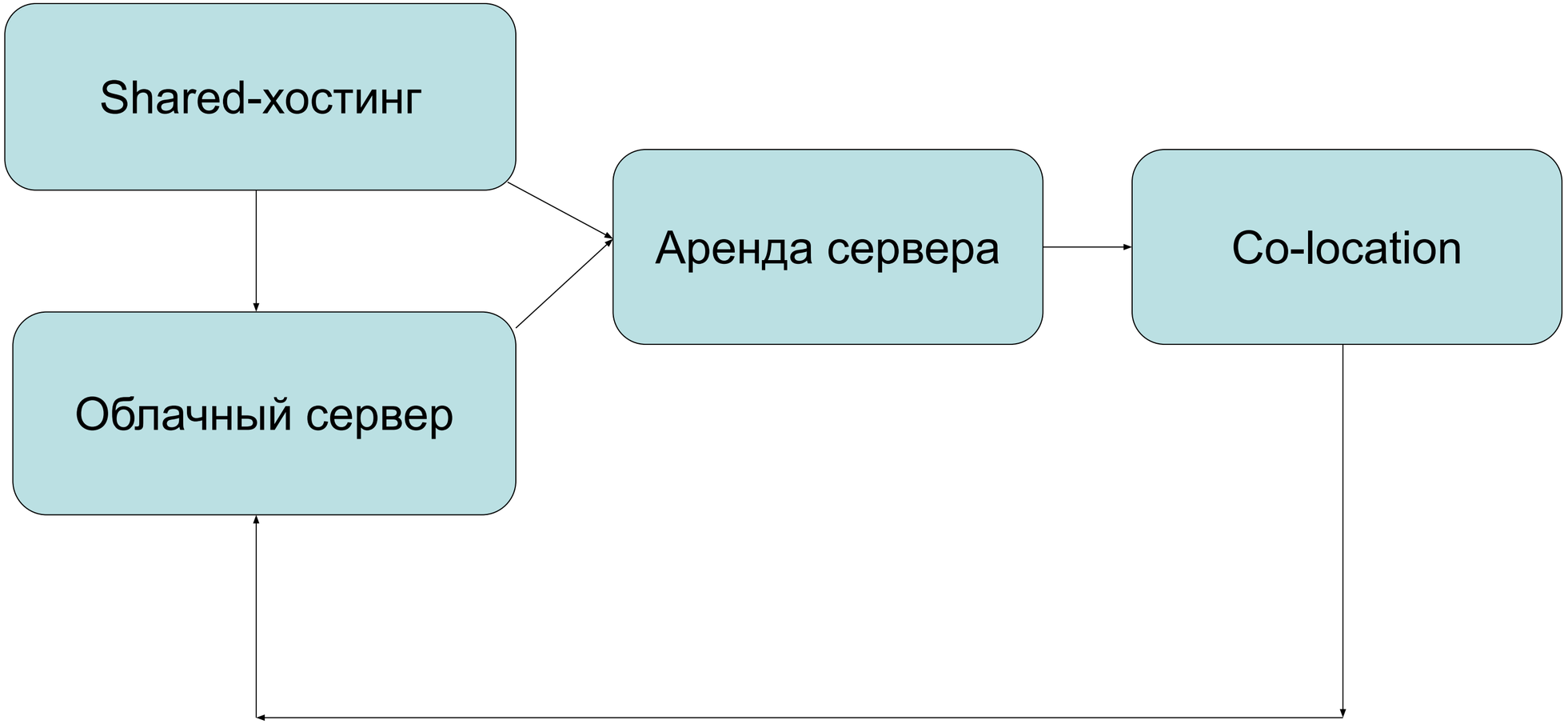
Внутренний вид среднестатистического датацентра



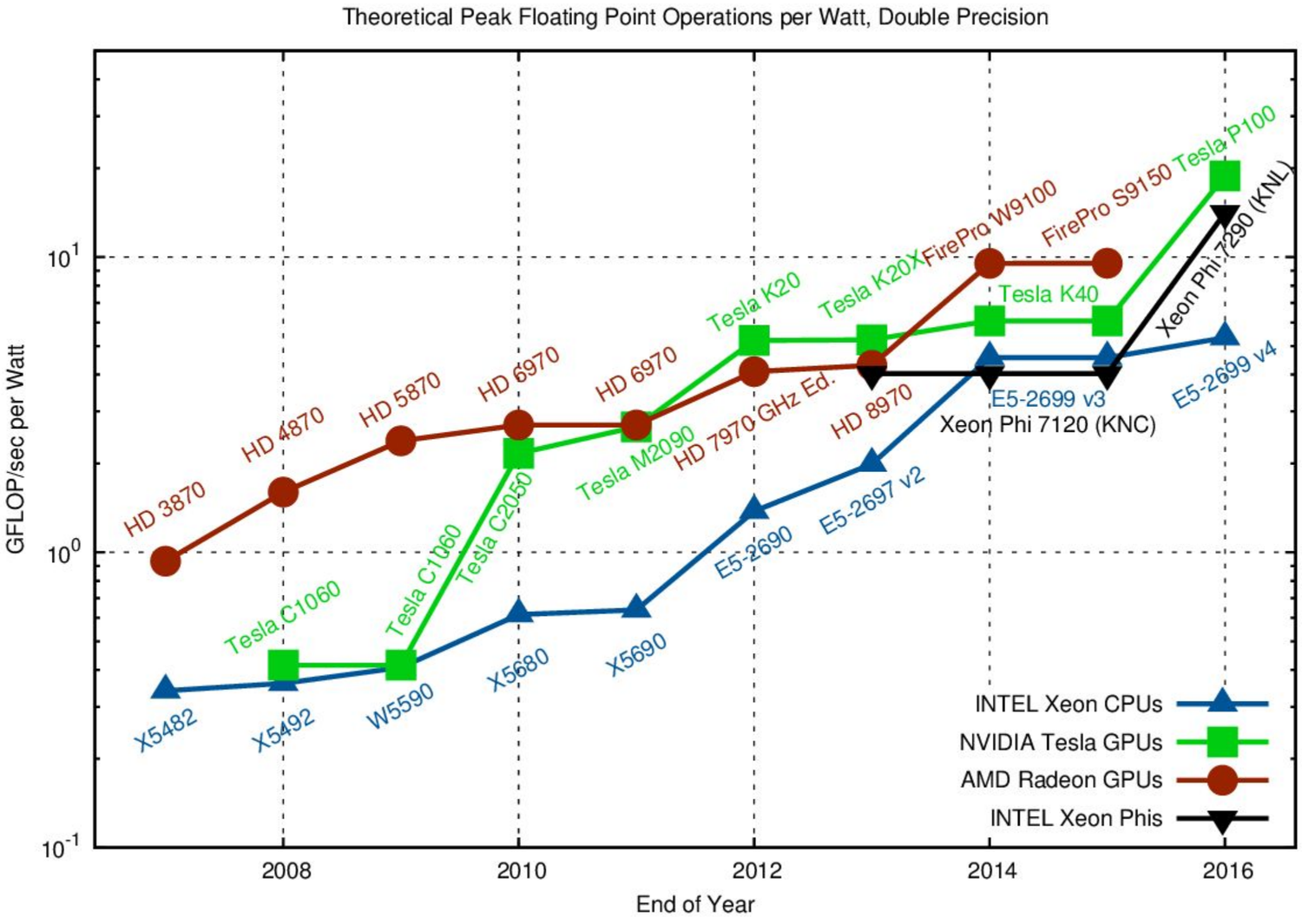
Основные технические вызовы в классическом ДЦ

1. Электропитание (300 Вт на юнит, 10-12 КВт на стойку,резерв)
2. Охлаждение (1 КВт ~ 3400 BTU/час, резерв)
3. Пожаротушение (фреон, углекислота, порошок уходи)
4. Управление трафиком
5. Контроль доступа
6. Борьба с владельцами здания

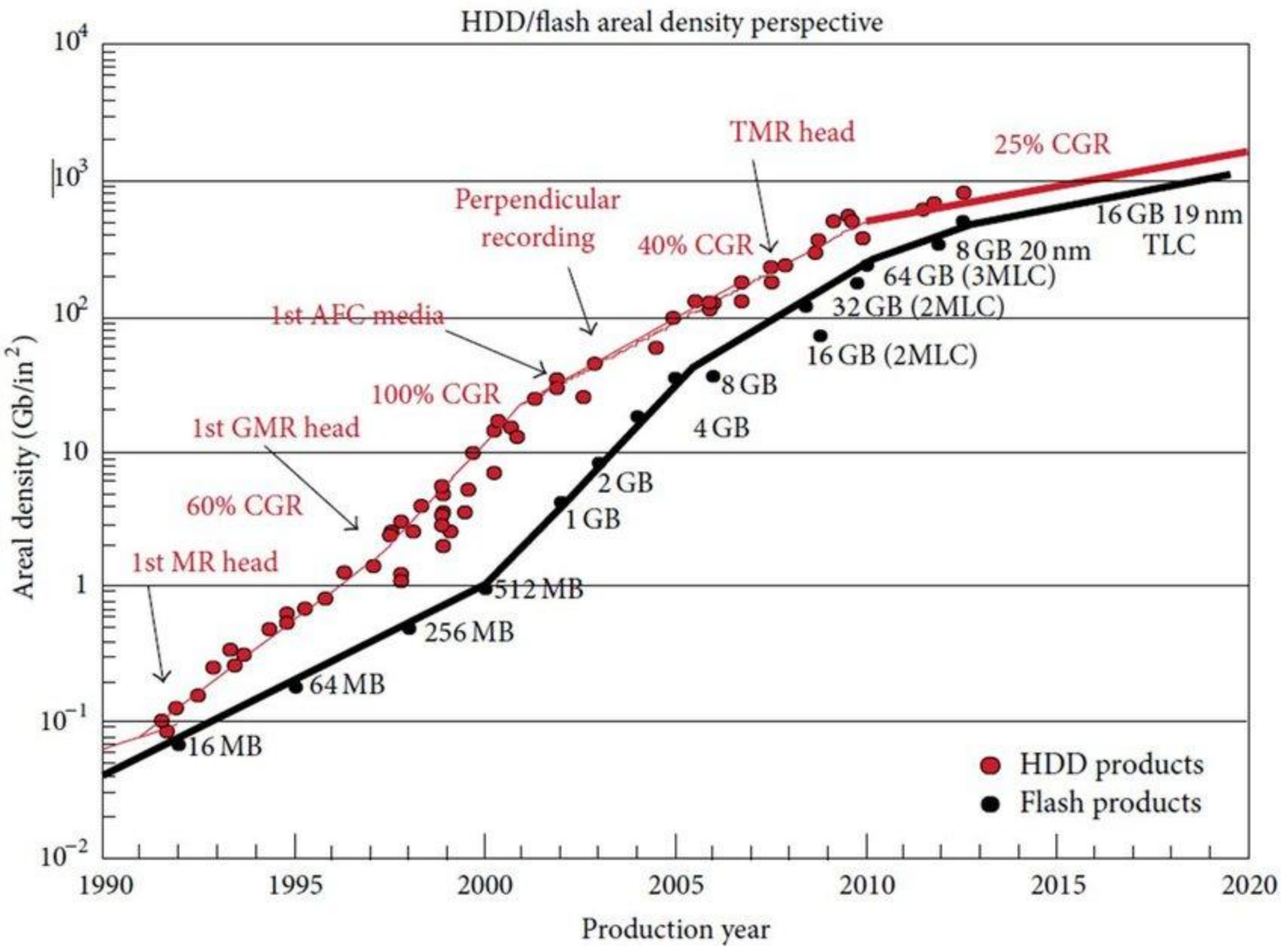
Жизненный цикл среднестатистического клиента



Рост вычислительной мощности на ватт



Рост плотности записи HDD и SSD

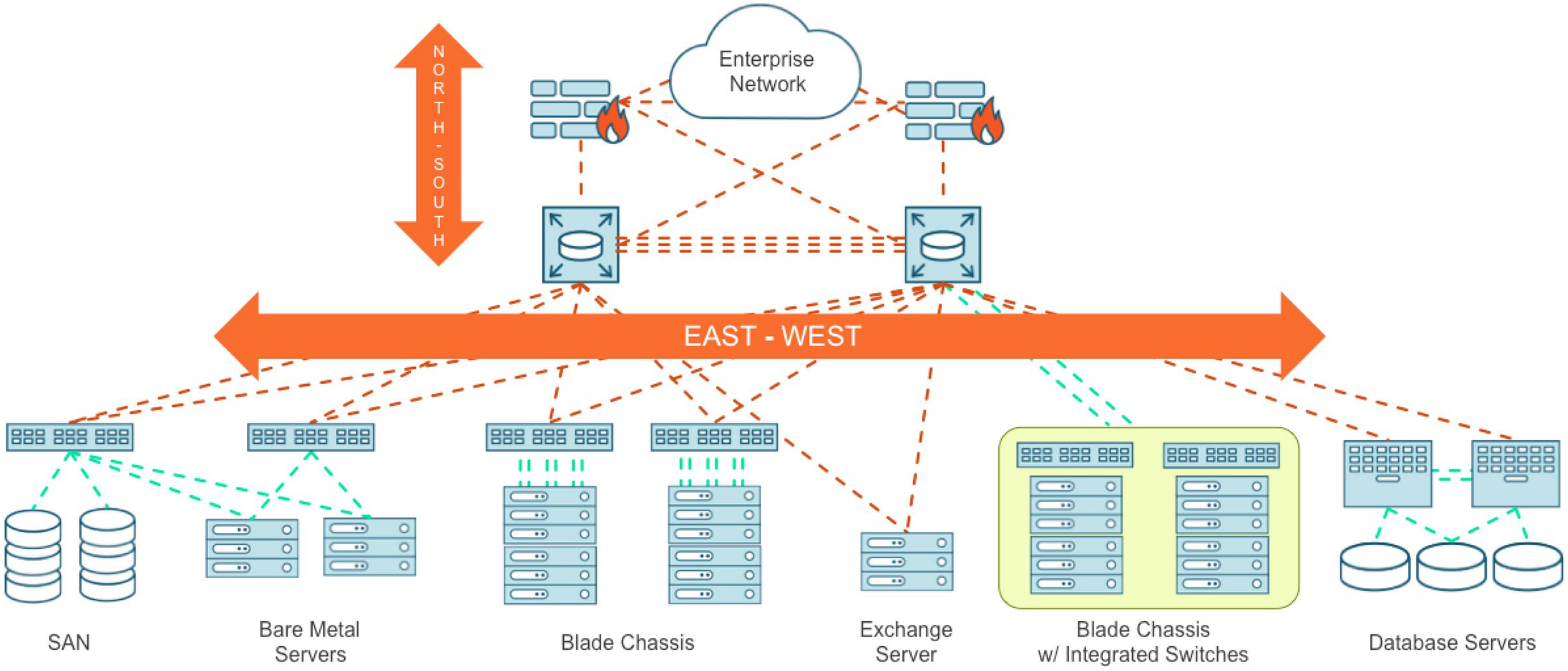


Увеличение плотности компоновки устройств

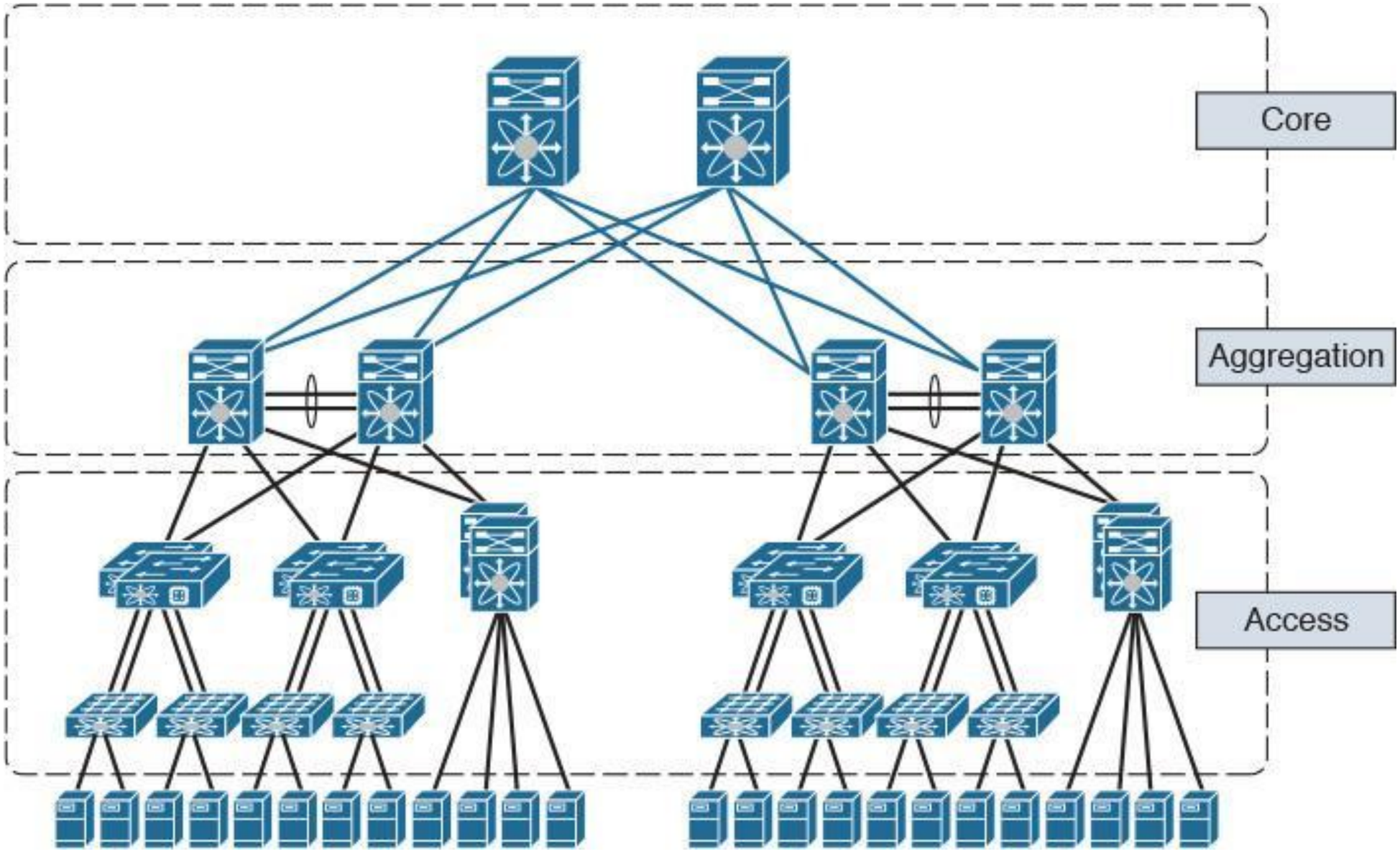
HP DL380G4
2004 год

HP DL380 Gen9
2014 год

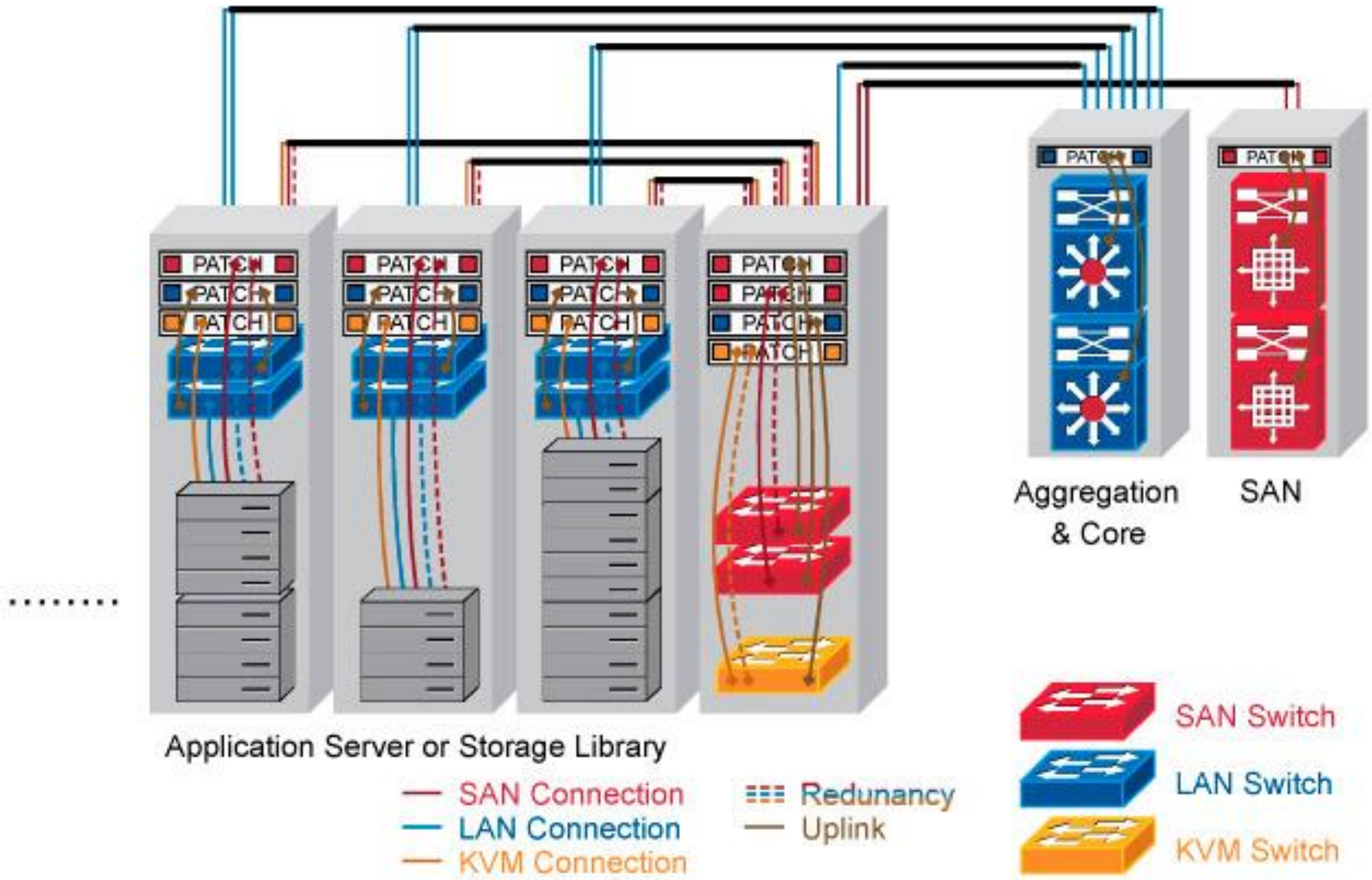
Изменение направления трафика



Стандартная трехуровневая архитектура



Классический ДЦ с архитектурой Top of Rack (ToR)



Классический ДЦ с архитектурой Top of Rack (ToR)

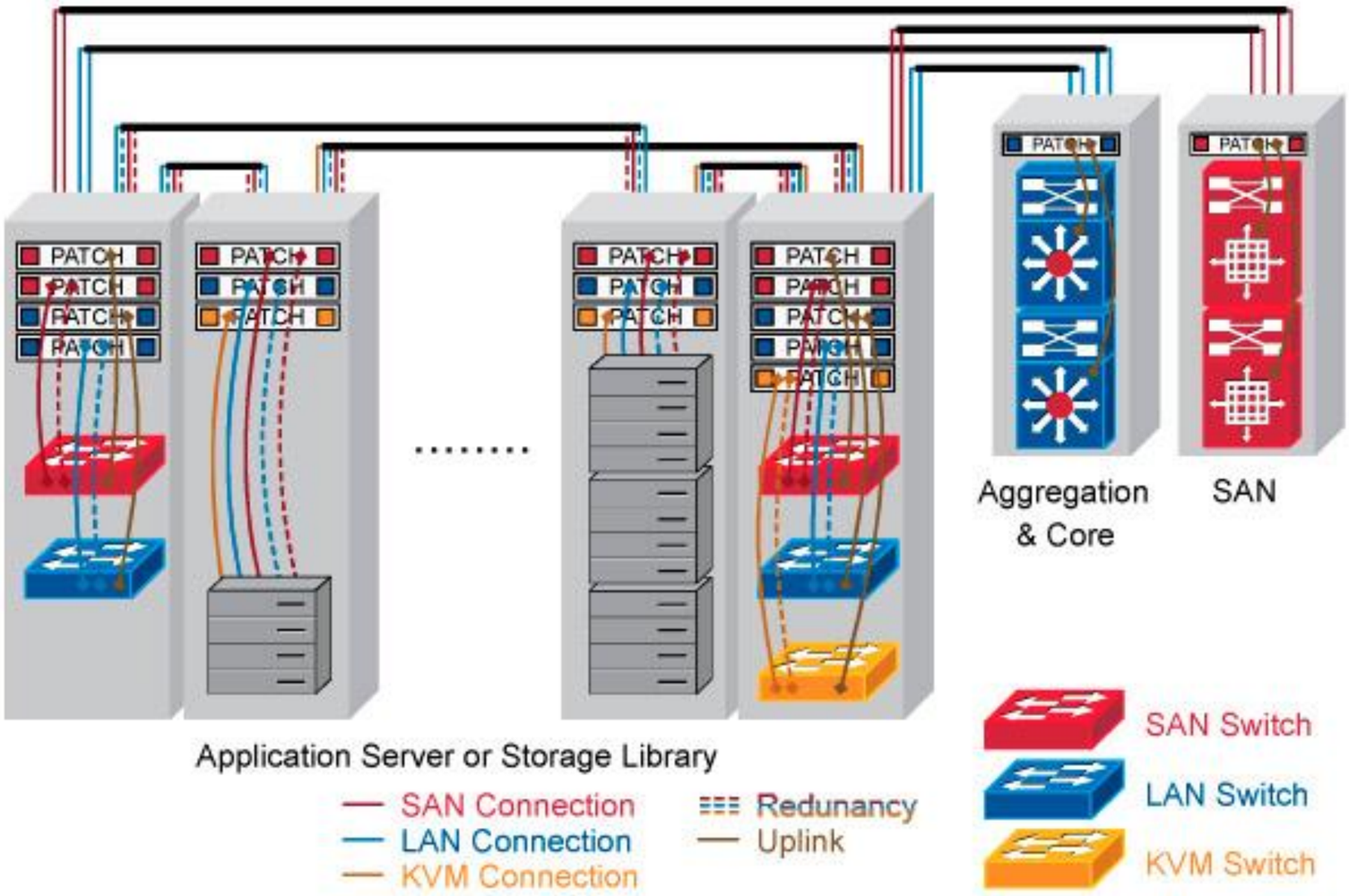
Достоинства

- Медная проводка сосредоточена в пределах стойки
- Меньшая стоимость и большая организованность кабелей
- Возможность увеличения скоростей в будущем
- Легкость добавления стоек

Недостатки

- Большее количество коммутаторов
- Худшая масштабируемость с точки зрения STP
- Чистая L2-сеть до уровня aggregation
- Сложность управления конфигурацией коммутаторов

Классический ДЦ с архитектурой End of Row (EoR)

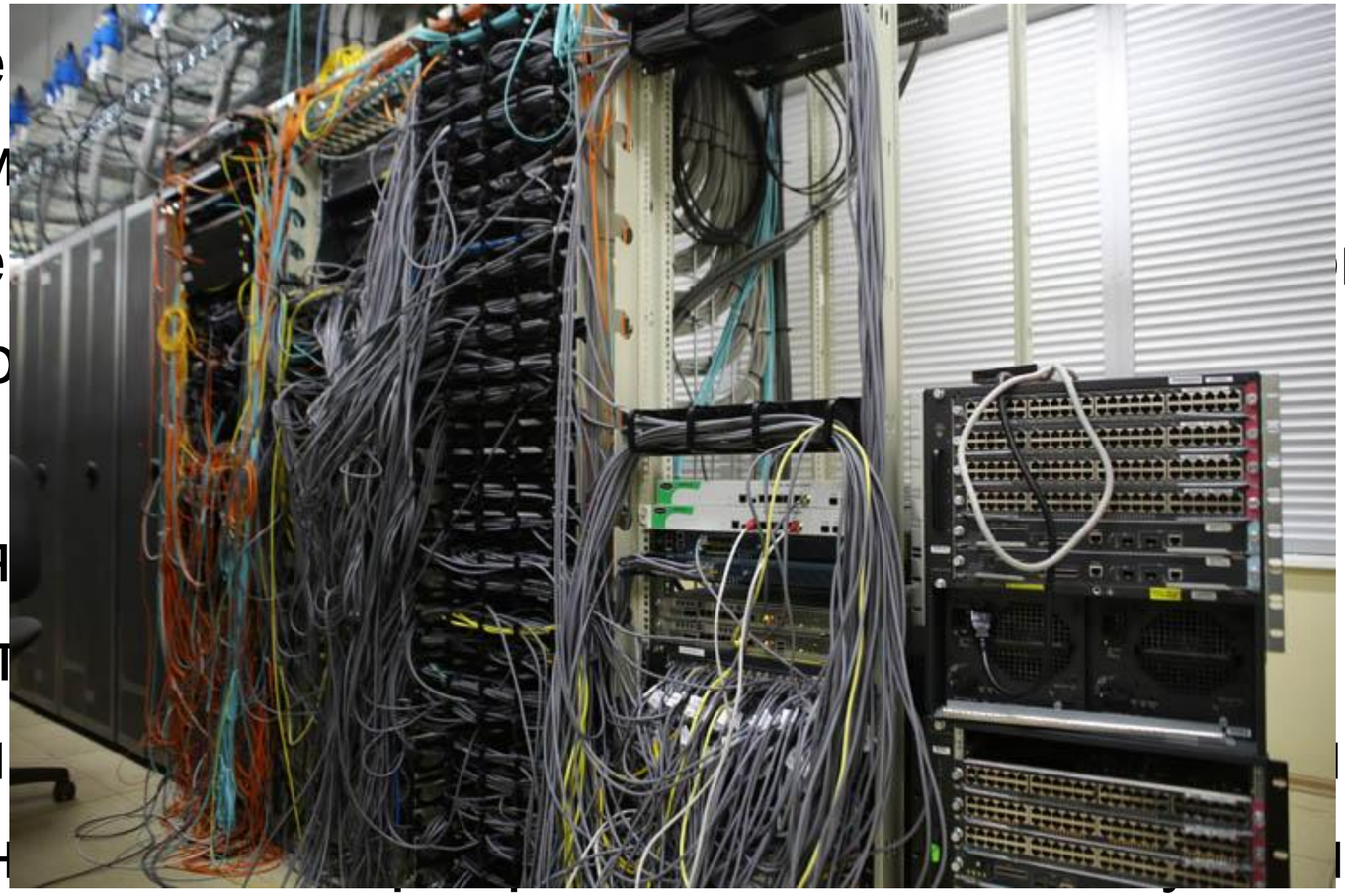


Классический ДЦ с архитектурой End of Row (EoR)

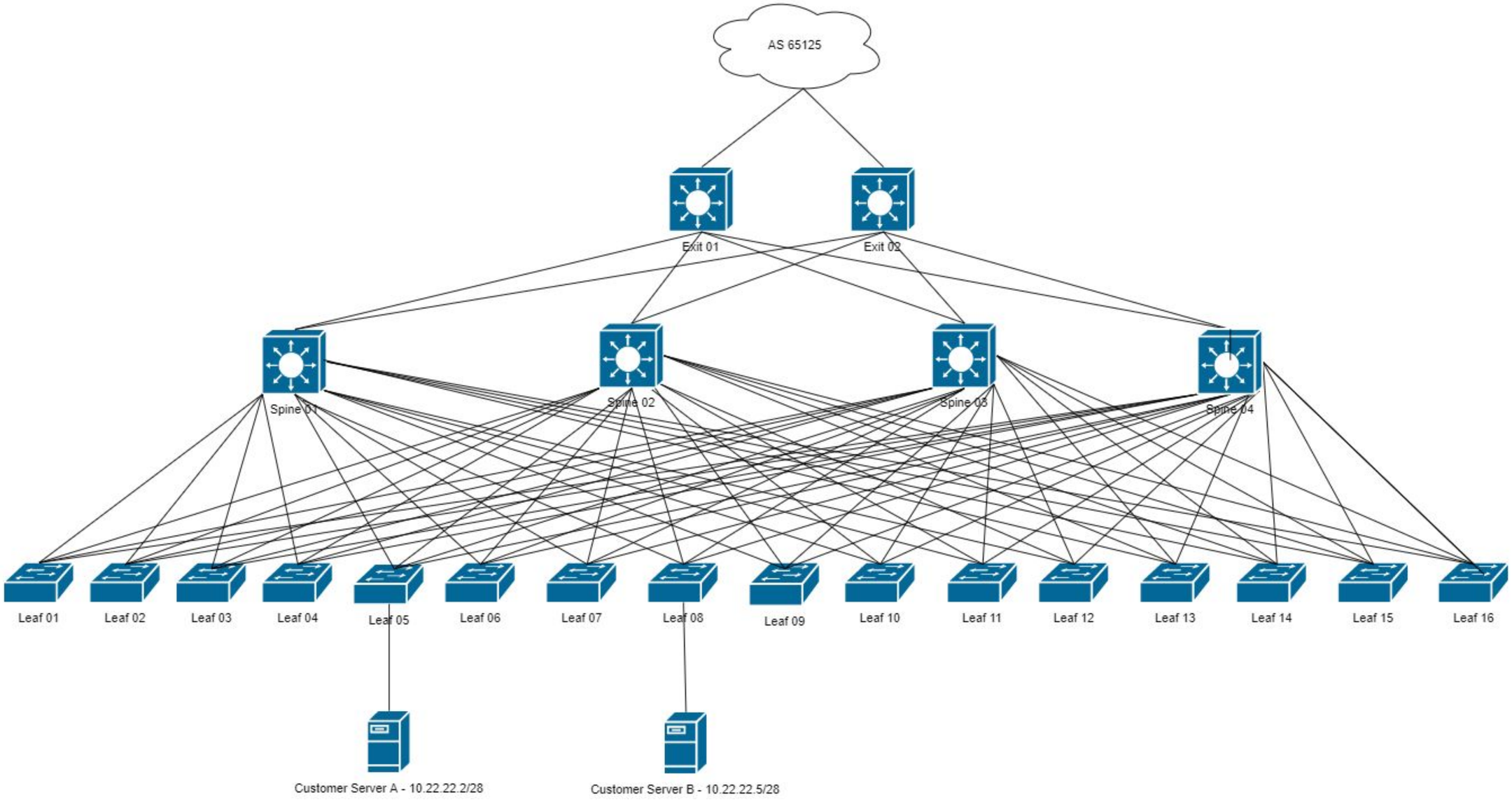
Достоинства

- Меньшее
- Лучшая м
- Меньшее n
- Общая то

- Безумная
- Сложност
- Увеличен атично
- Добавлен ерацию



Архитектура Leaf-Spine (сети Клоза)



Архитектура Leaf-Spine (сети Клоза)

Достоинства

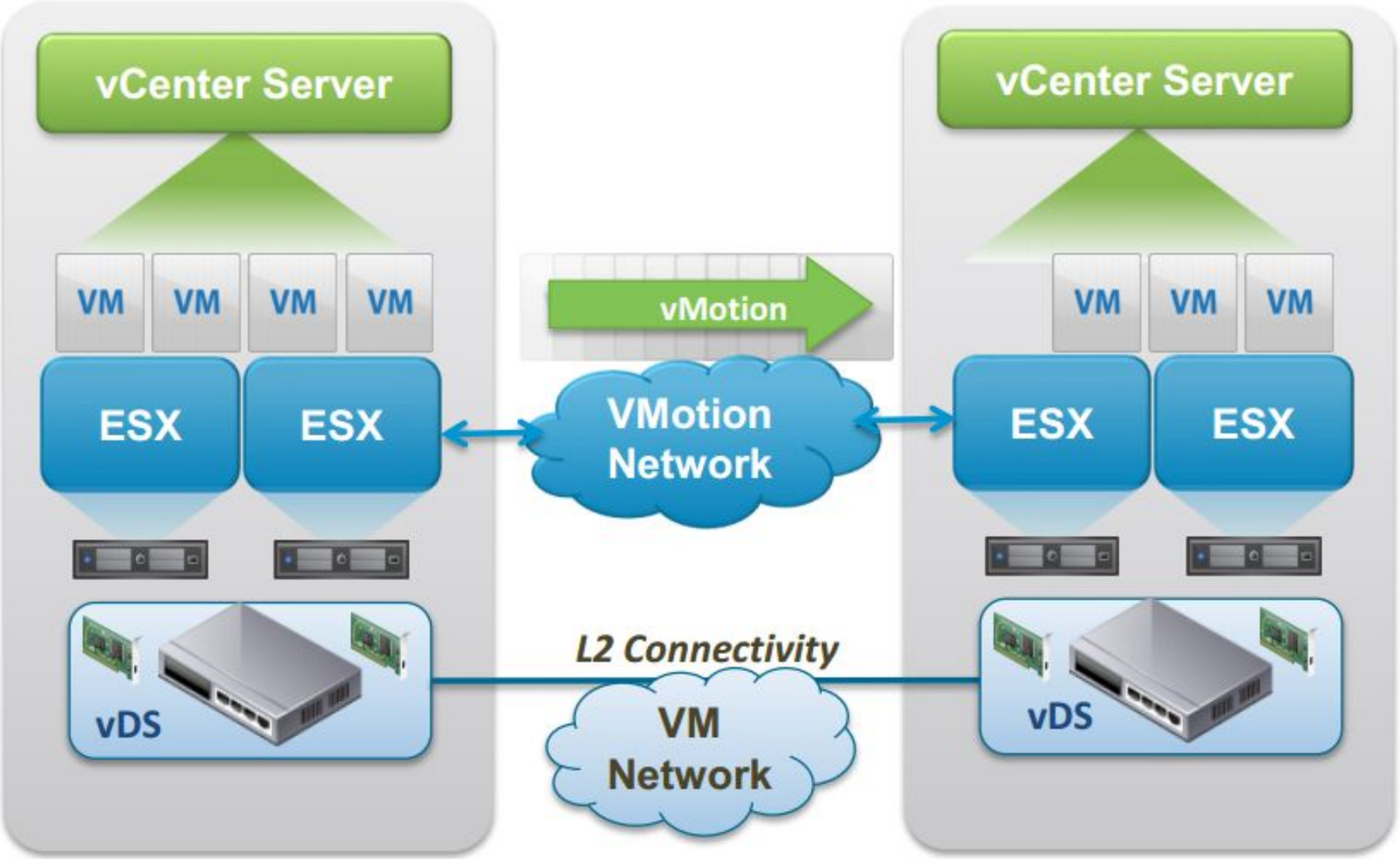
- Высокая отказоустойчивость
- Ориентация на горизонтальные потоки трафика
- Высокая производительность за счет отказа от STP
- Мгновенная реконфигурация при сбоях

Недостатки

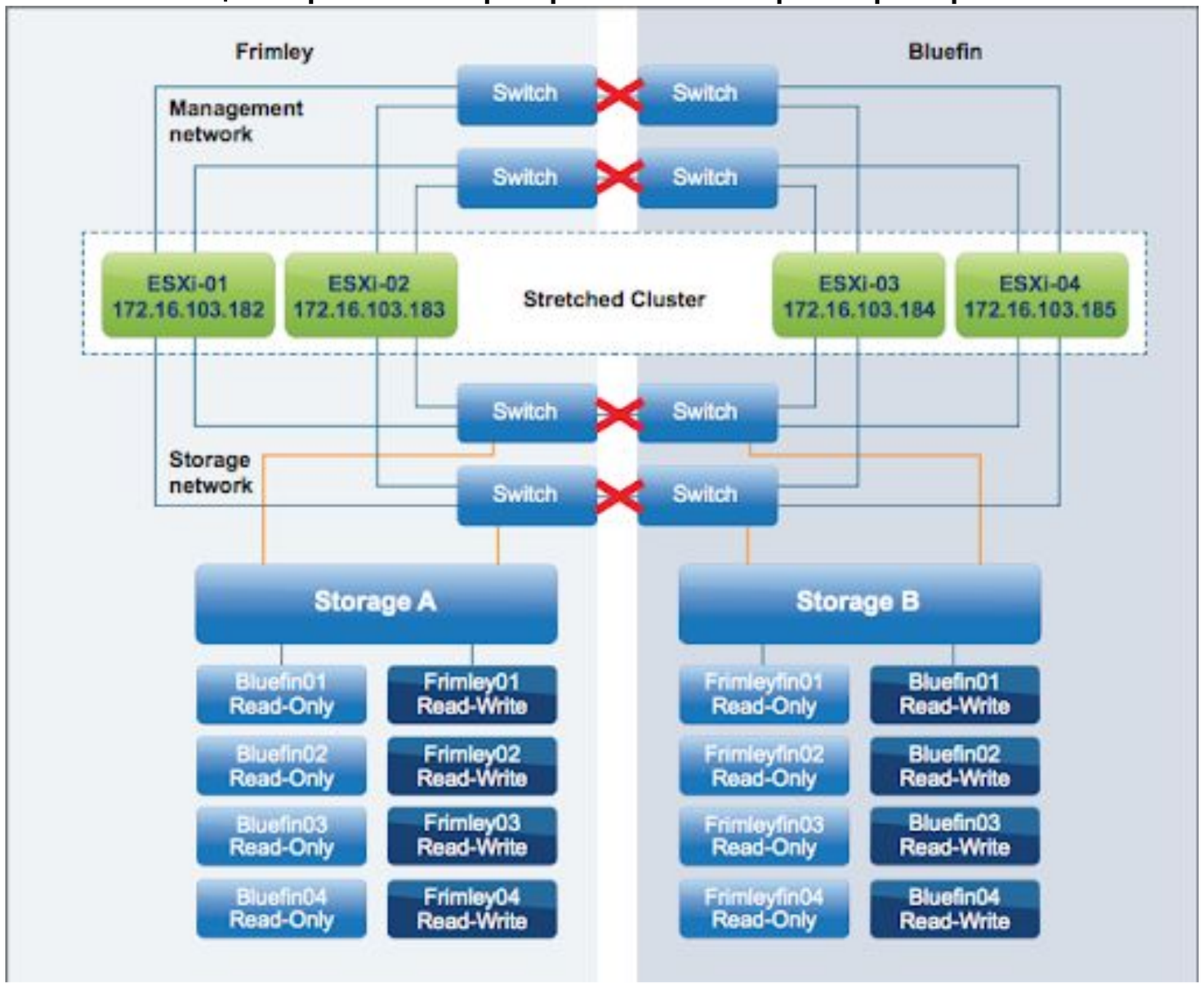
- Отсутствует L2-домен как класс
- Увеличение стоимости коммутаторов и проводки
- Сложный технологический стек
- Неочевидный процесс поиска решения проблем

А теперь со всем этим мы попытаемся взлететь..

Классический сценарий балансировки/резервирования нагрузки



Классический сценарий географического резервирования кластера



Общие особенности сетей виртуальных машин

Чего хочется

- Как правило требуется L2-связность на уровне гипервизоров
- Желательна L2-связность для СХД
- Клиенты хотят строить собственные сети поверх сетей ДЦ
- Клиенты хотят самостоятельно конфигурировать эти сети
- Клиентские сети должны тоже быть L2-прозрачны

Что есть

- L2 только между гипервизором и leaf-коммутатором
- Кругом L3-маршрутизация с ECMP
- Много быстрых процессоров и толстых каналов связи

В железе

- Кадр L2 покидает гипервизор как есть
- Leaf-коммутатор инкапсулирует L2 в L3, т.е. кадр Ethernet помещается внутрь пакета IP (yeah, baby!)
- L3-пакет вычурным образом маршрутизируется внутри ДЦ и, возможно, через WAN
- На Leaf-коммутаторе назначения исходный L2-кадр вынимается из пакета L3
- Кадр L2 передается гипервизору как есть
- Реализации – TRILL, SPB, FabricPath

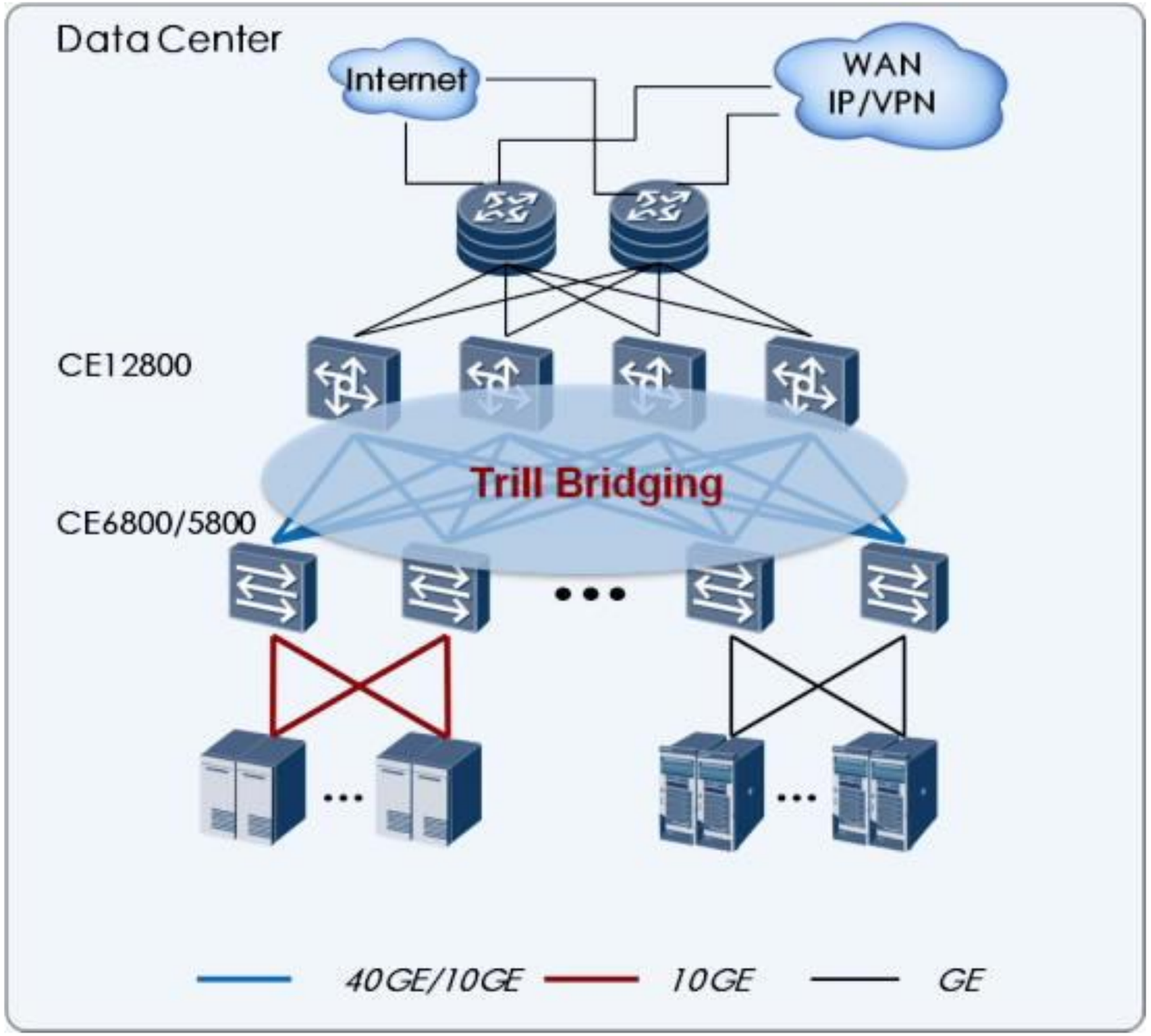
В софте

- Кадр L2 инкапсулируется в пакет L3 непосредственно на гипервизоре внутри виртуального коммутатора
- L3-пакет вычурным образом маршрутизируется через Leaf-Spine внутри ДЦ и, возможно, через WAN
- Кадр L2 вынимается из пакета L3 непосредственно на гипервизоре назначения
- Реализации – VXLAN, NVGRE

Общие проблемы

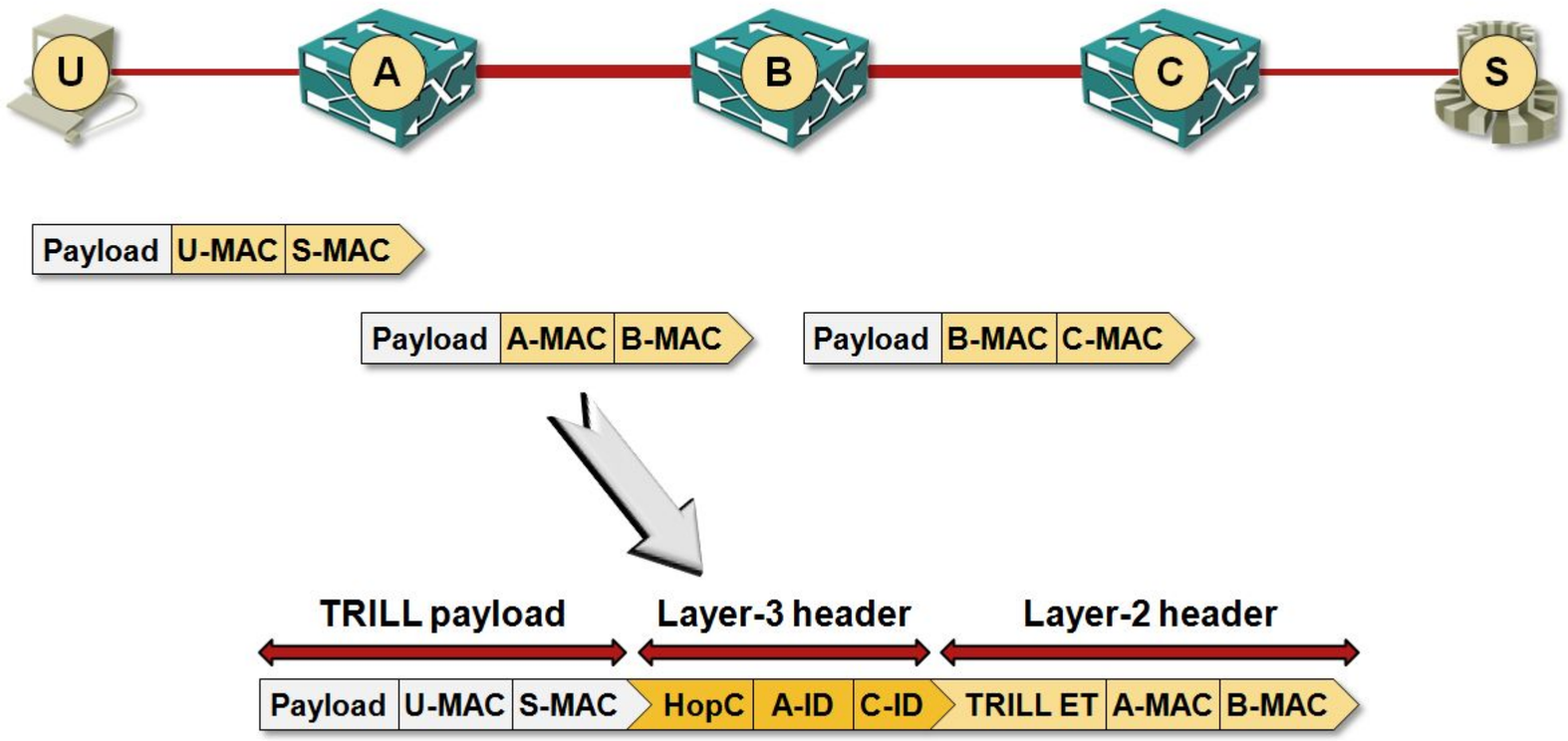
- Необходимость построения топологии конечных и промежуточных узлов. TRILL использует IS-IS, VXLAN использует BGP
- Таблица MAC-адресов становится распределенной
- Broadcast и multicast трафик становится опасным
- Значительное сокращение полезного MTU
- Возможно появление медленных и слабо выявляемых петель

TRILL (Transparent Interconnection of Lots of Links)

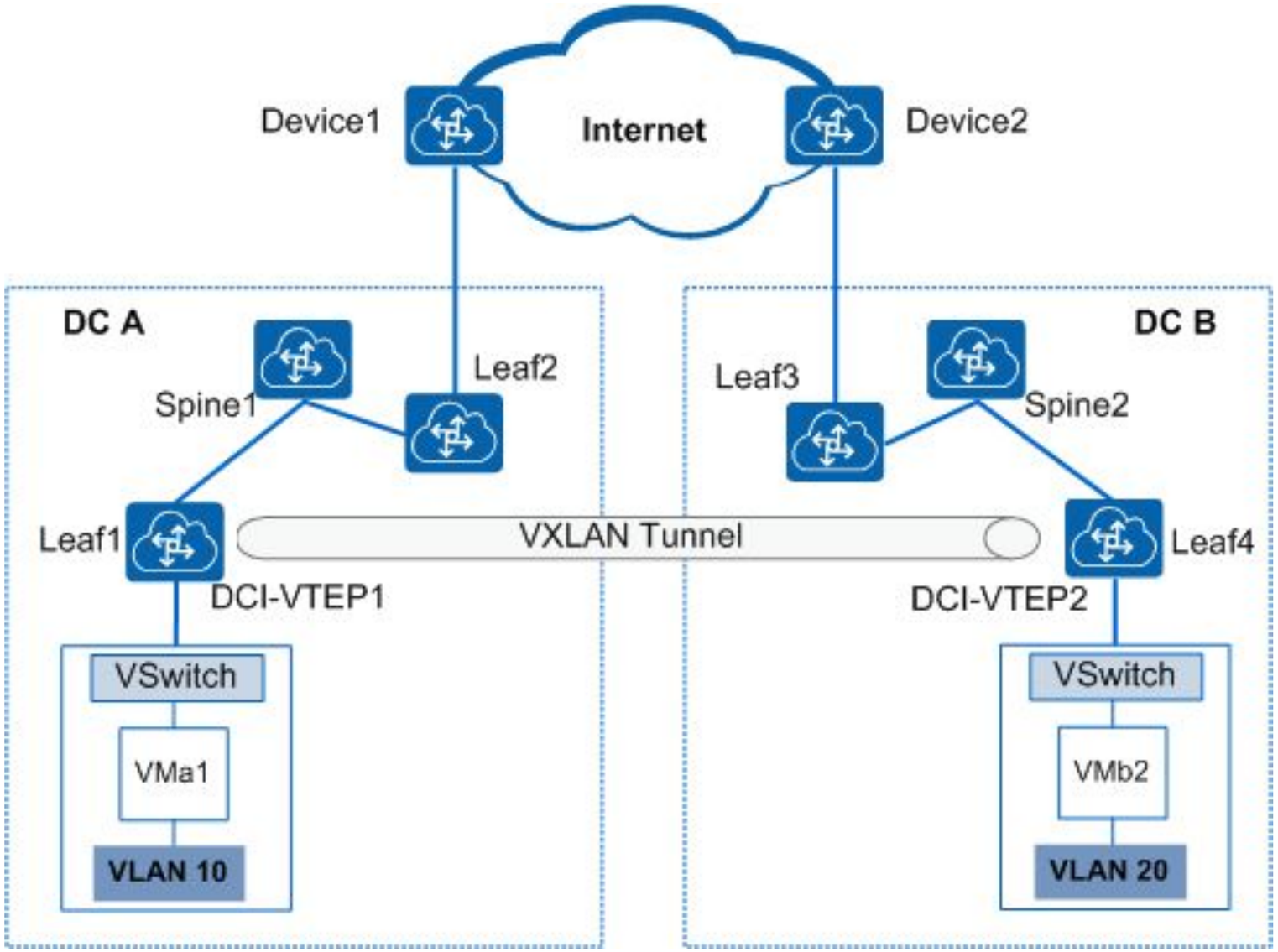


TRILL (Transparent Interconnection of Lots of Links)

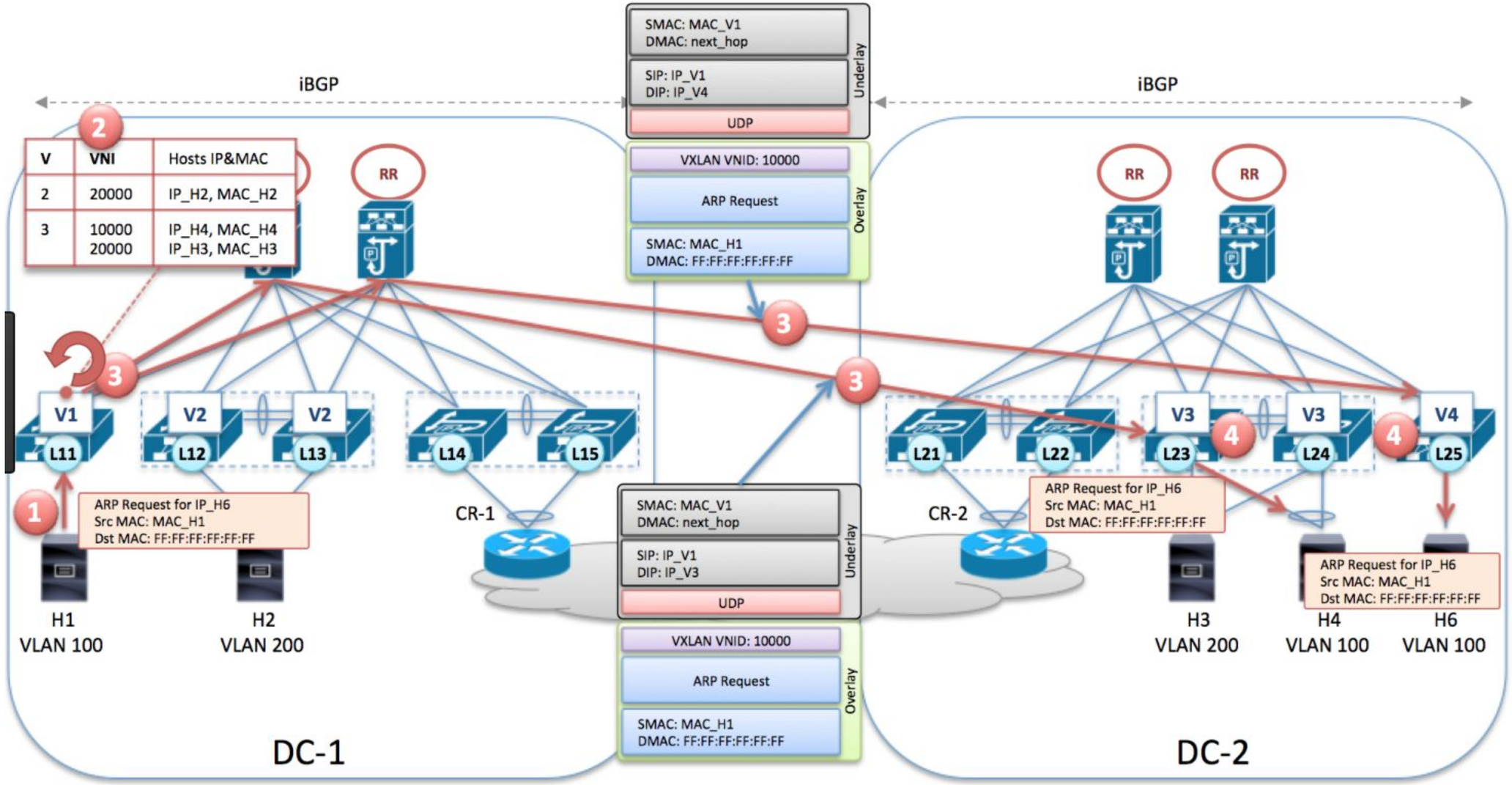
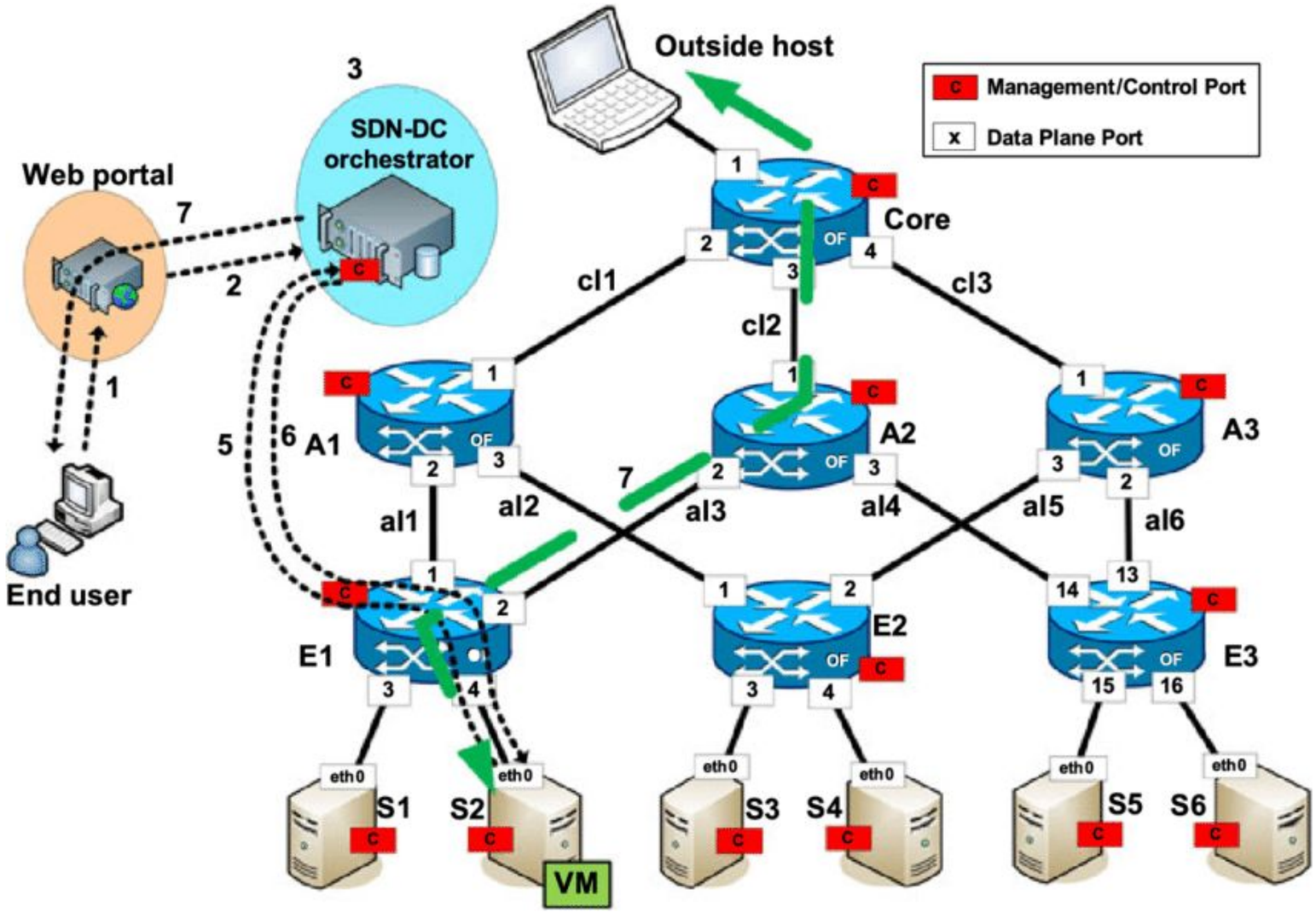
TRILL: Forwarding Paradigm



VXLAN


Device1
Internet
Device2
DC A
DC B
Spine1
Leaf2
Leaf3
Spine2
Leaf1
VXLAN Tunnel
Leaf4
DCI-VTEP1
DCI-VTEP2
VSwitch
VMa1
VLAN 10
VSwitch
VMb2
VLAN 20

VXLAN


iBGP
iBGP
SMAC: MAC_V1
DMAC: next_hop
SIP: IP_V1
DIP: IP_V4
UDP
Underlay
VXLAN VNID: 10000
ARP Request
SMAC: MAC_H1
DMAC: FF:FF:FF:FF:FF:FF
Overlay
V | VNI | Hosts IP&MAC
2 | 20000 | IP_H2, MAC_H2
3 | 10000 | IP_H4, MAC_H4
20000 | IP_H3, MAC_H3
RR
RR
RR
V1
V2
V2
V3
V3
V4
L11
L12
L13
L14
L15
L21
L22
L23
L24
L25
ARP Request for IP_H6
Src MAC: MAC_H1
Dst MAC: FF:FF:FF:FF:FF:FF
CR-1
CR-2
SMAC: MAC_V1
DMAC: next_hop
SIP: IP_V1
DIP: IP_V3
UDP
Underlay
VXLAN VNID: 10000
ARP Request
SMAC: MAC_H1
DMAC: FF:FF:FF:FF:FF:FF
Overlay
H1
VLAN 100
H2
VLAN 200
H3
VLAN 200
H4
VLAN 100
H6
VLAN 100
DC-1
DC-2

Будущее здесь – Программно определяемые сети (SDN)