Сетевые Хранилища Данных –
Storage Area Network

Проф. Смелянский Р.Л. Доп. главы компьютерных сетей

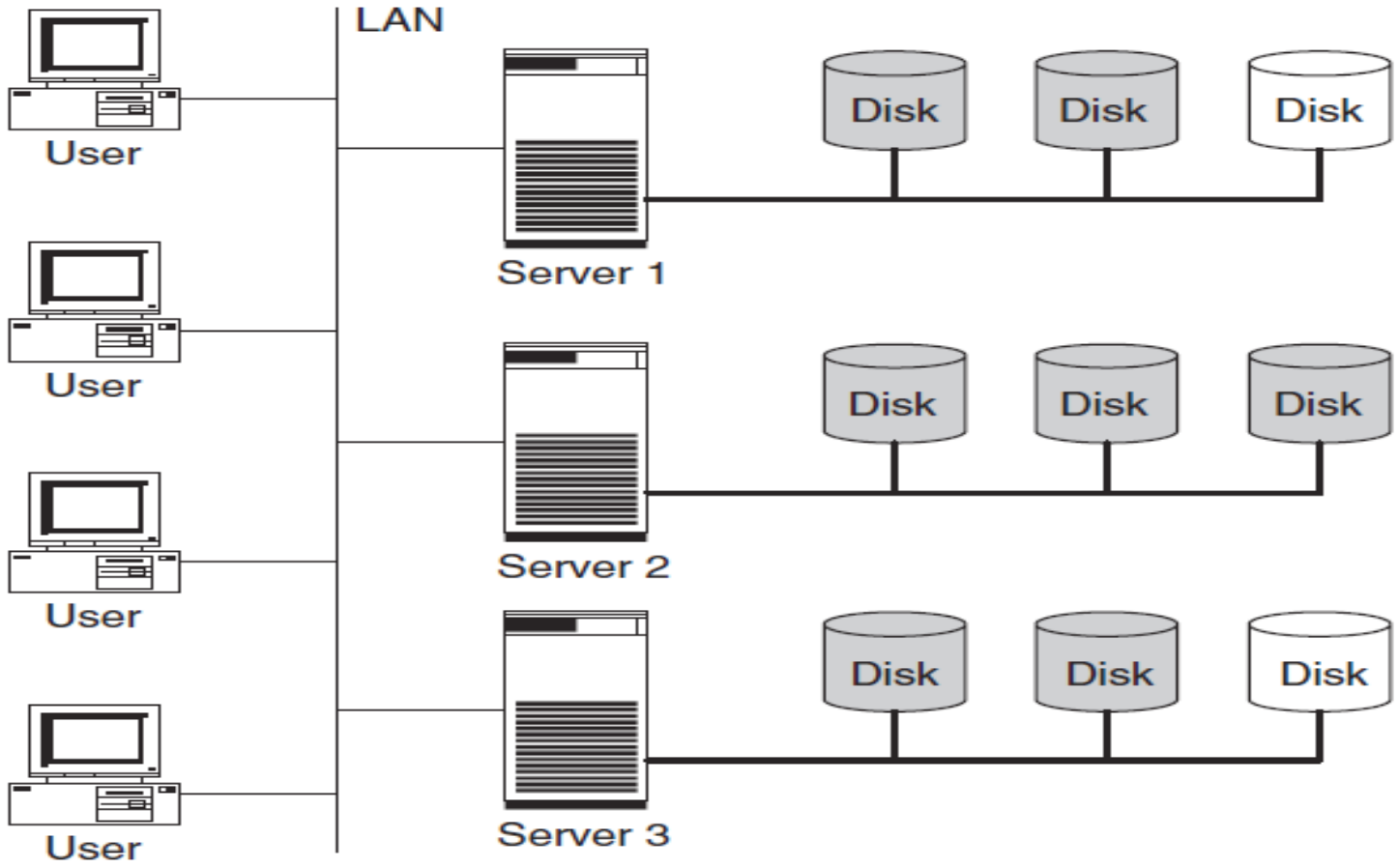
Сетевые Хранилища Данных

Storage Networks Explained: Basics and Application of Fibre Channel SAN, NAS, iSCSI, InfiniBand and FCoE, Second Edition.

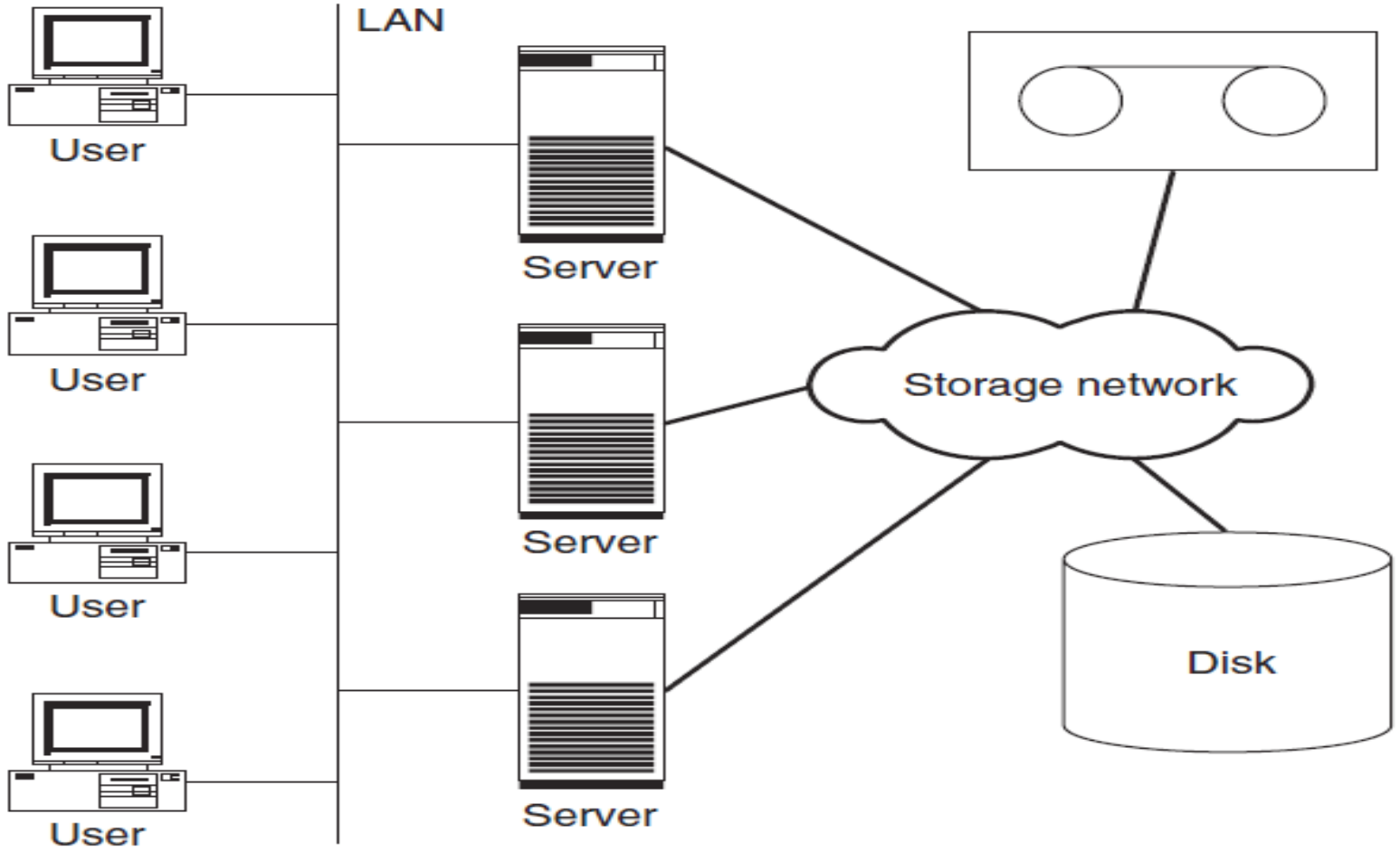
U. Troppens, W. Müller-Friedt, R. Wolafka, R. Erkens and N. Haustein © 2009 John Wiley & Sons Ltd. ISBN: 978-0-470-74143-6

- Архитектура информационной инфраструктуры
- Дисковые подсистемы и их организация
- JBOD
- RAID
- От CPU до ДПС

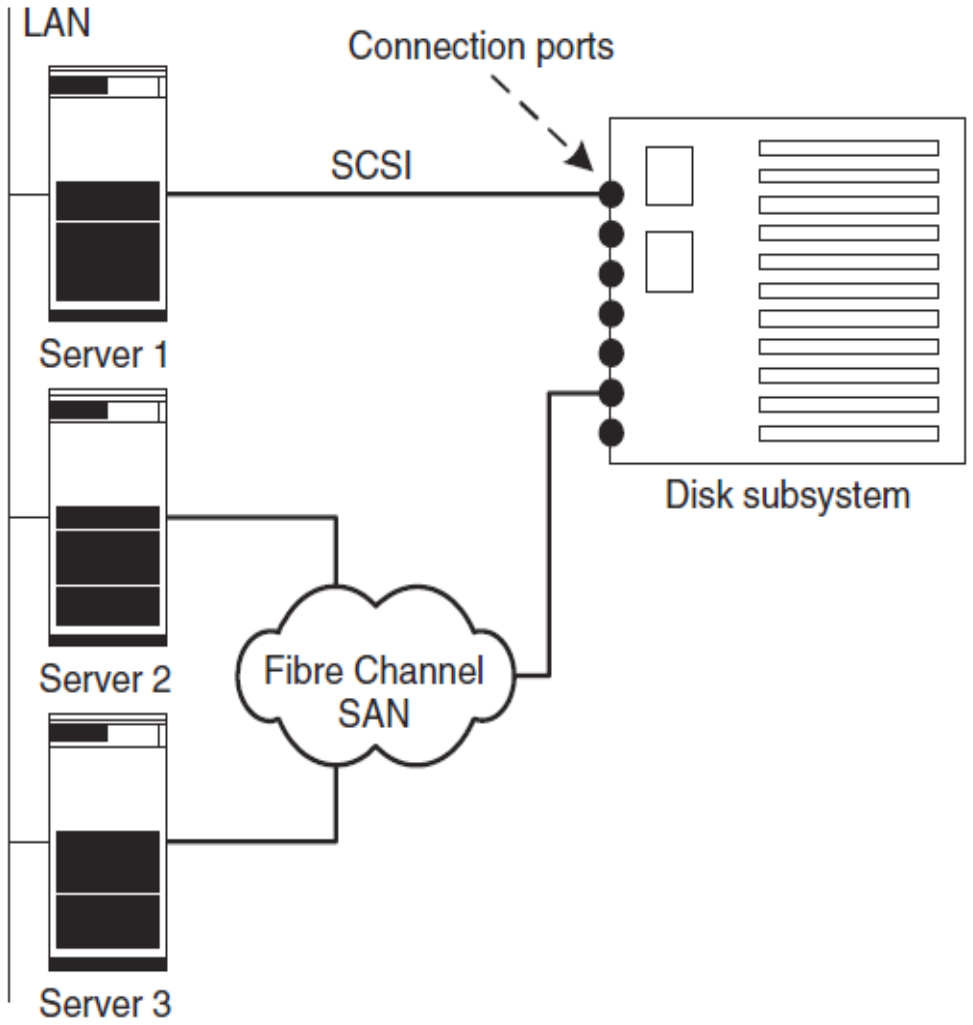
Server Centric Architecture

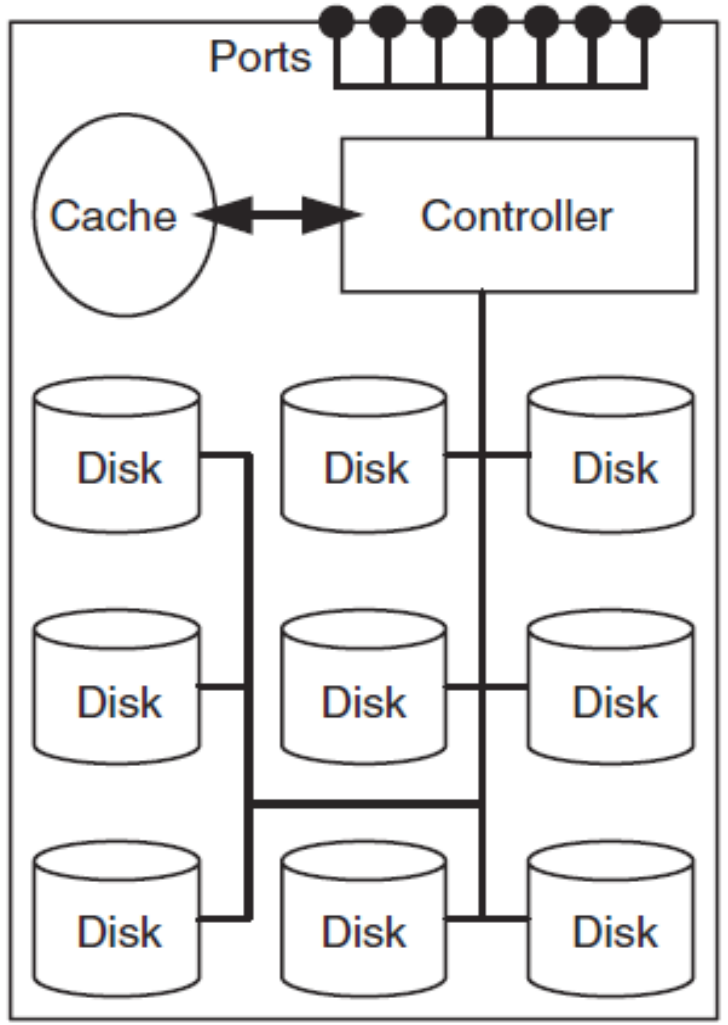


Storage Centric Architecture

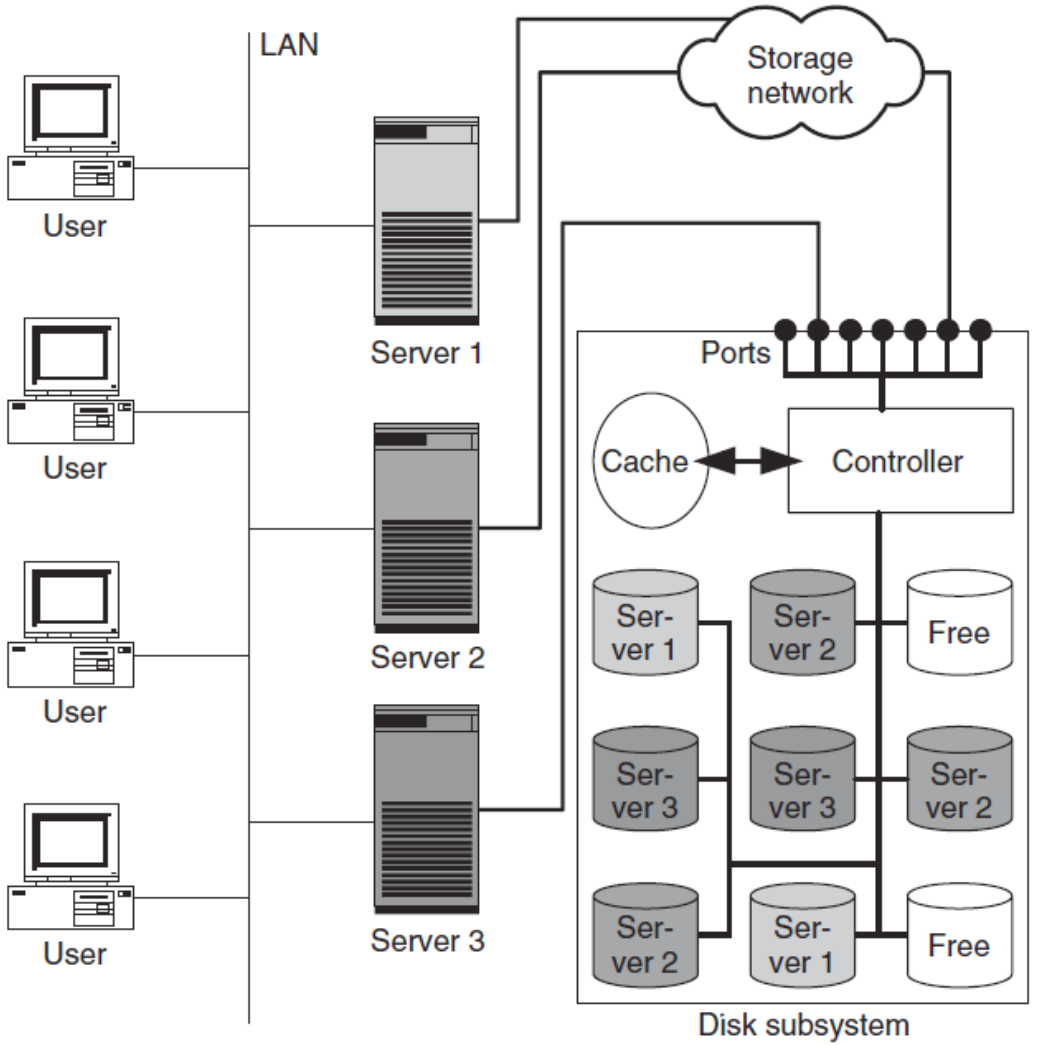


Архитектура Дисковой подсистемы (ДПС)

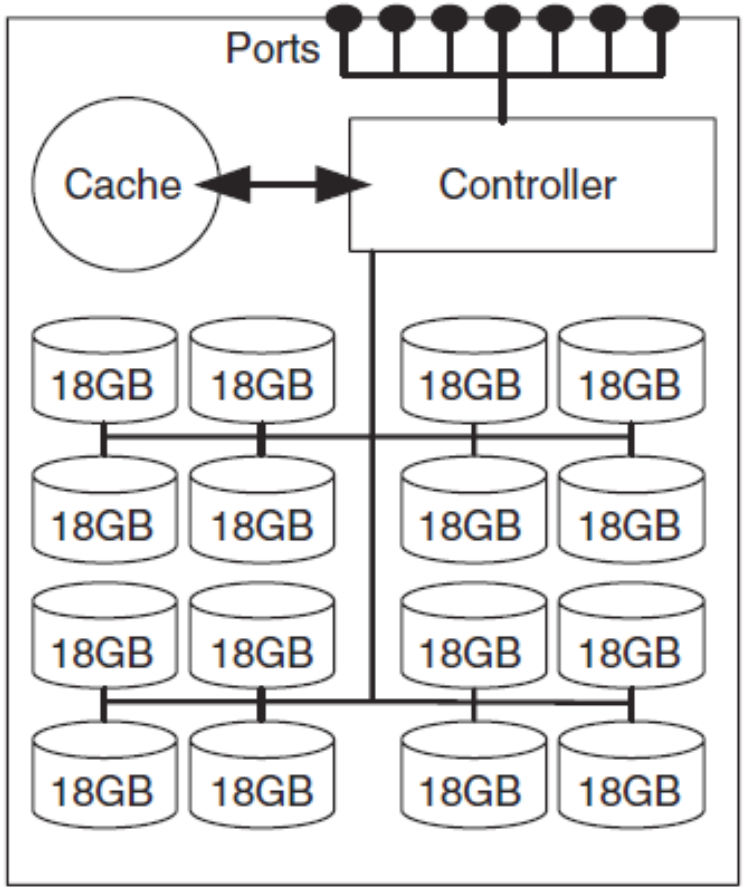




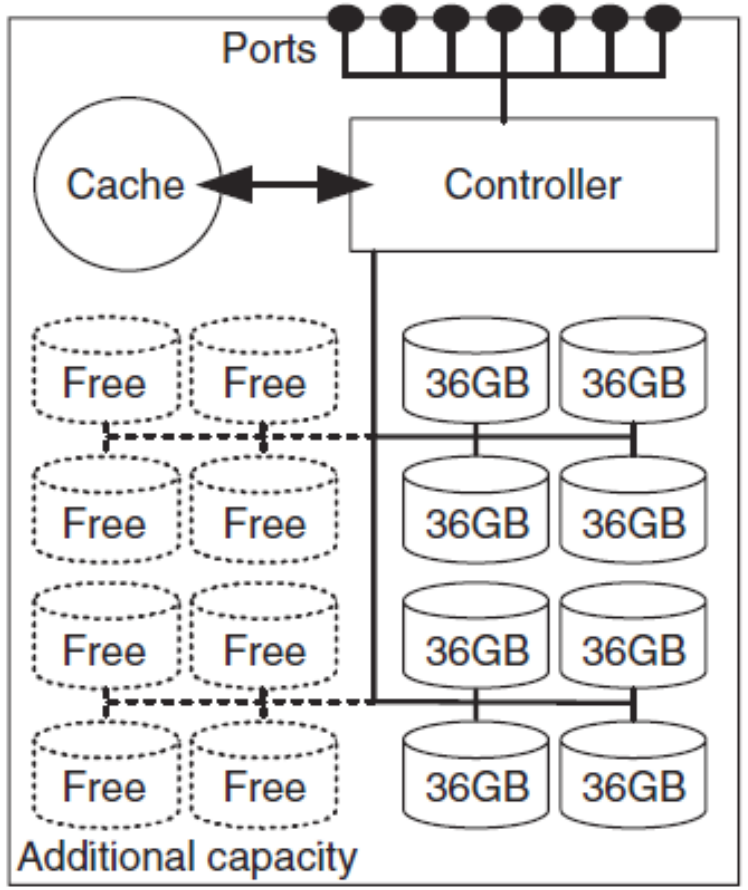
Пример использования ДПС



Внутренняя организация и емкость дисков

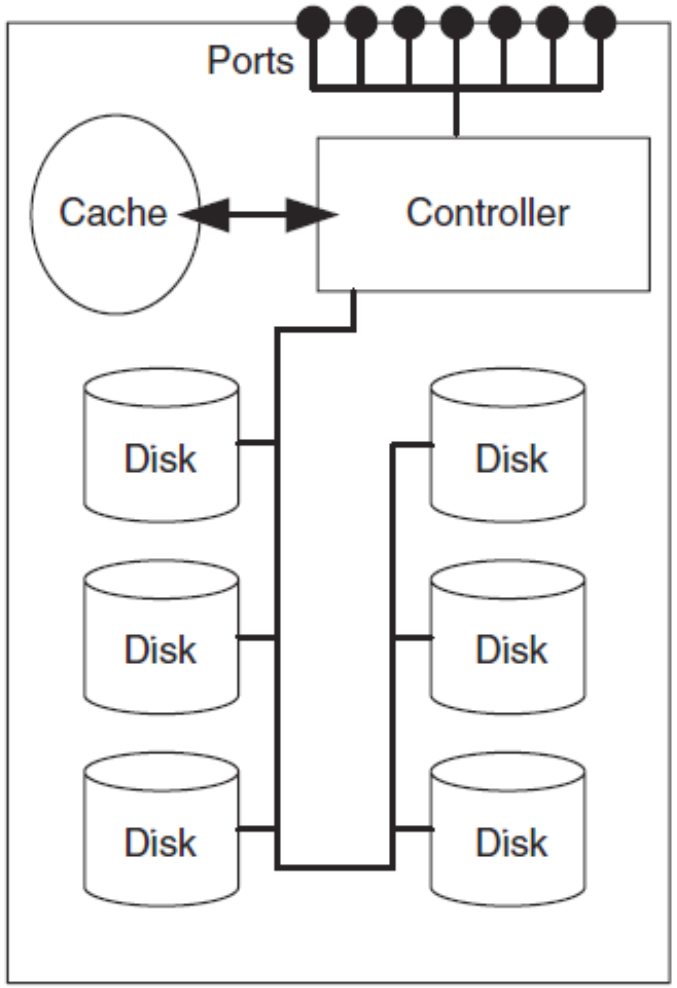


Маленькие диски

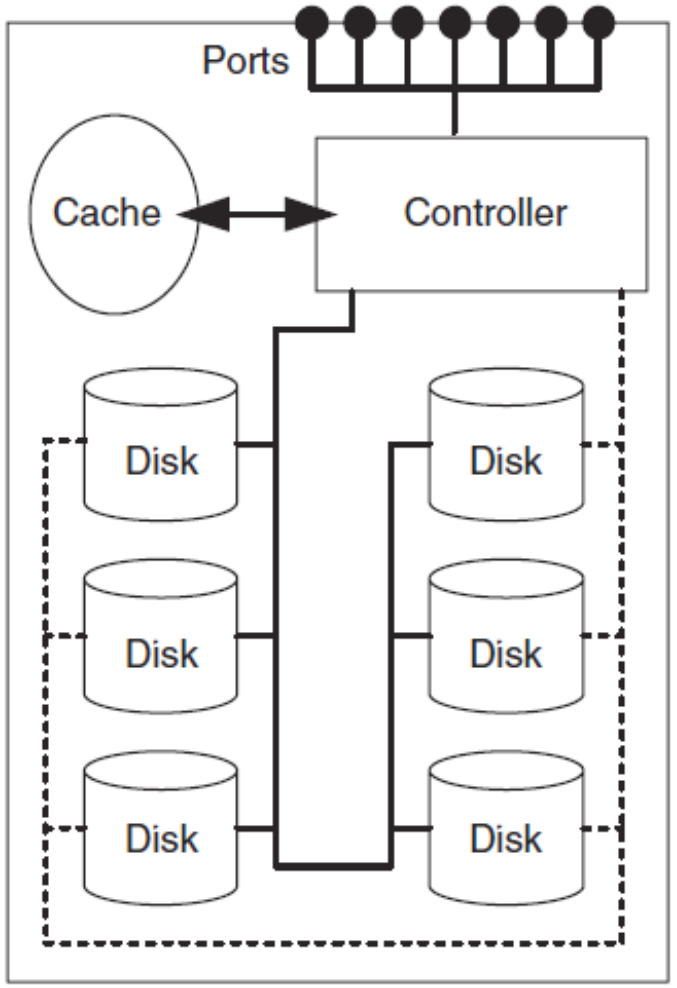


Большие диски

Внутренняя организация ДПС

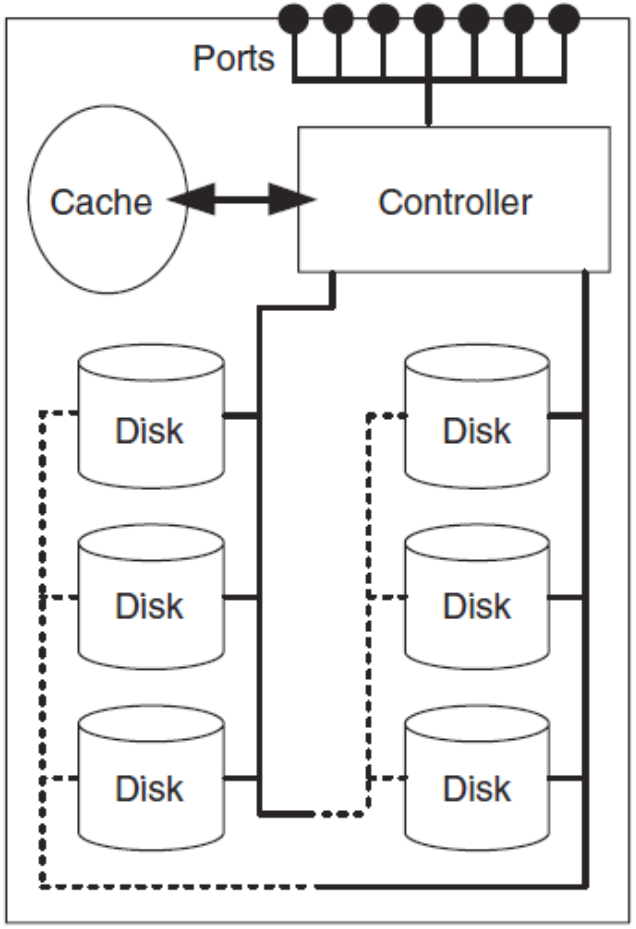


Одинарная

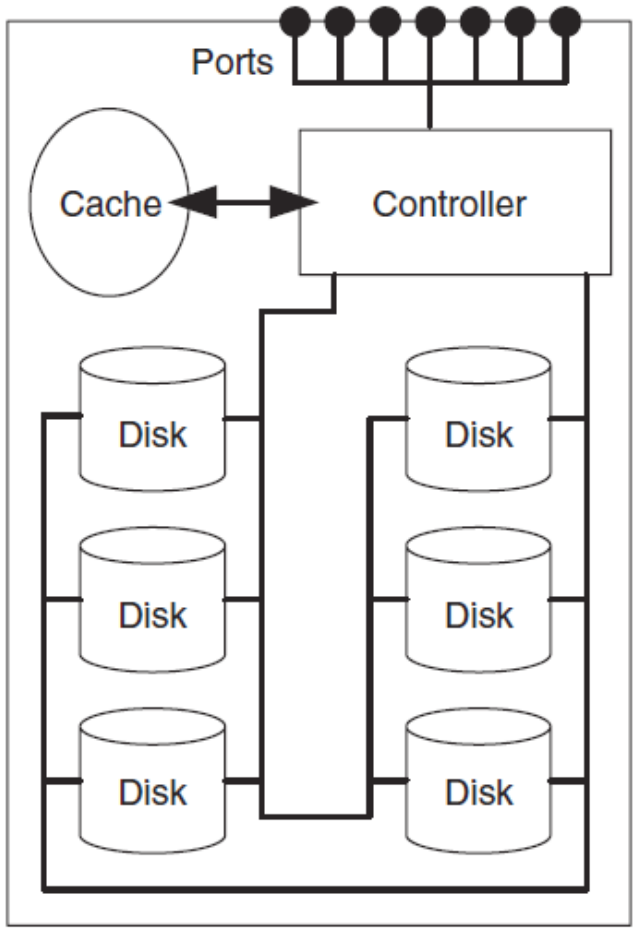


Дублированная

Активное дублирование

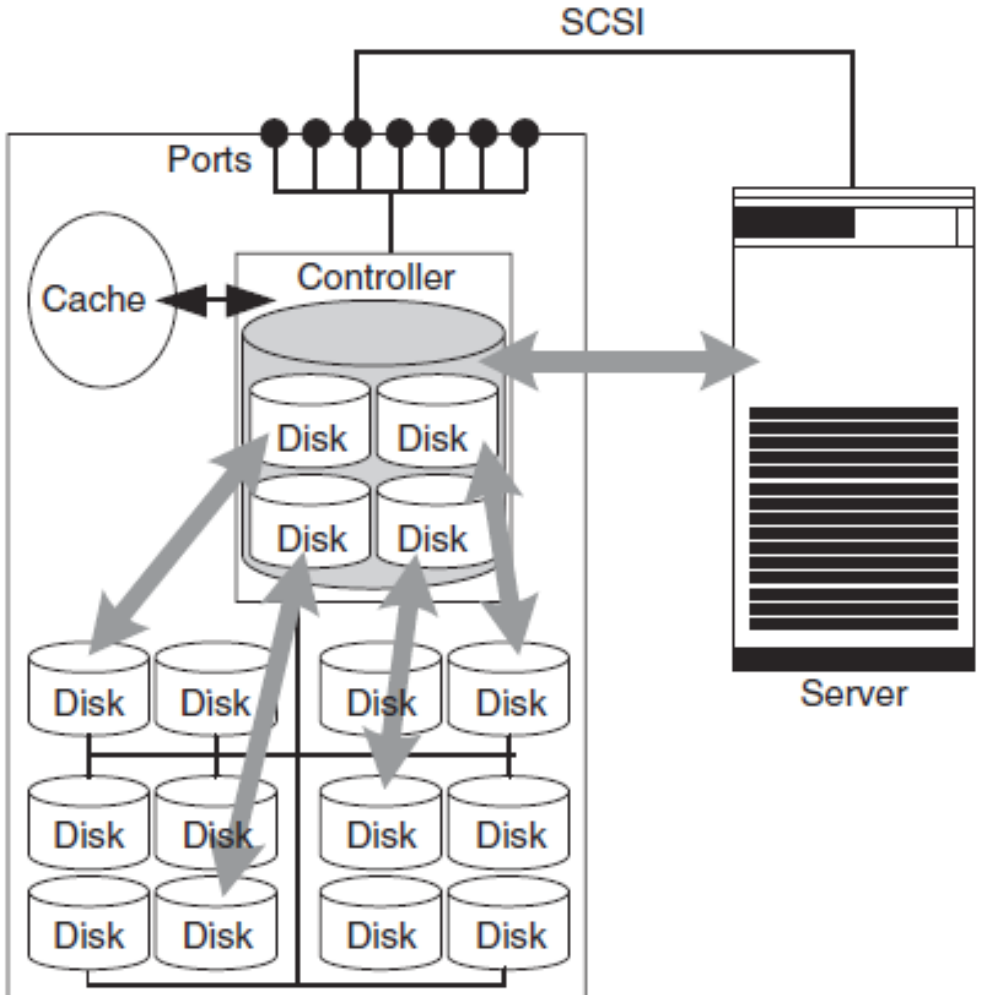


С фиксированным распределением дисков

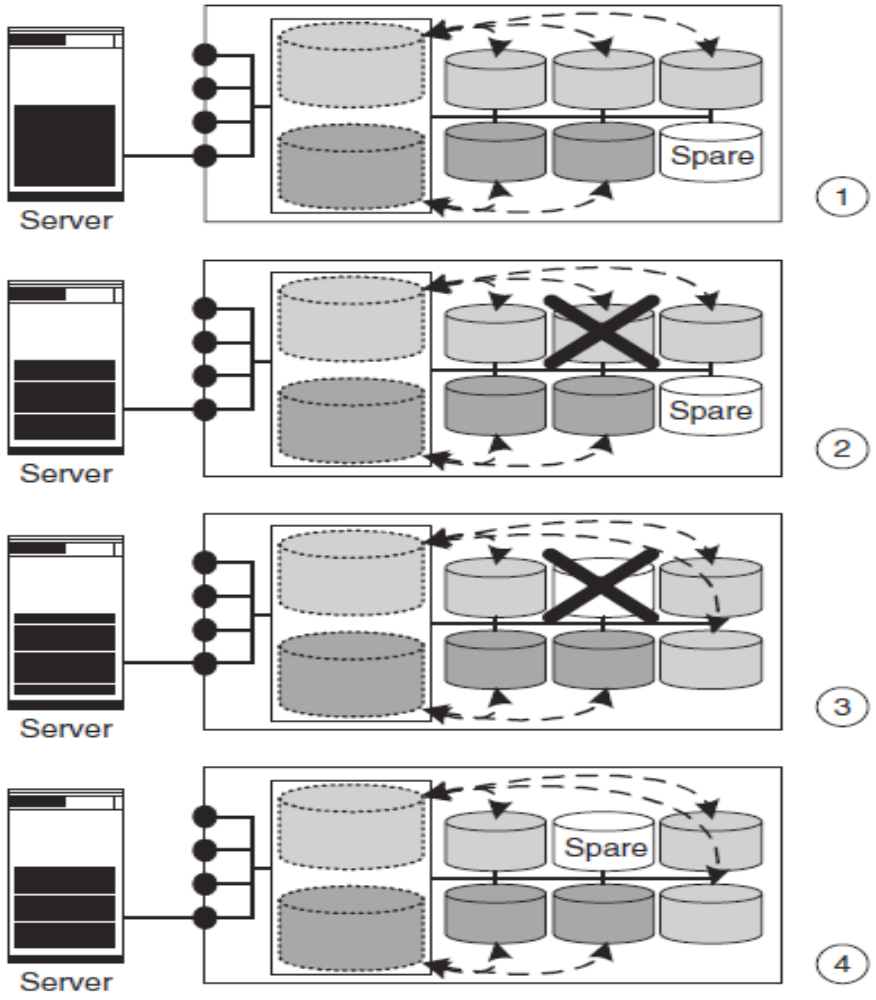


С динамическим распределением дисков

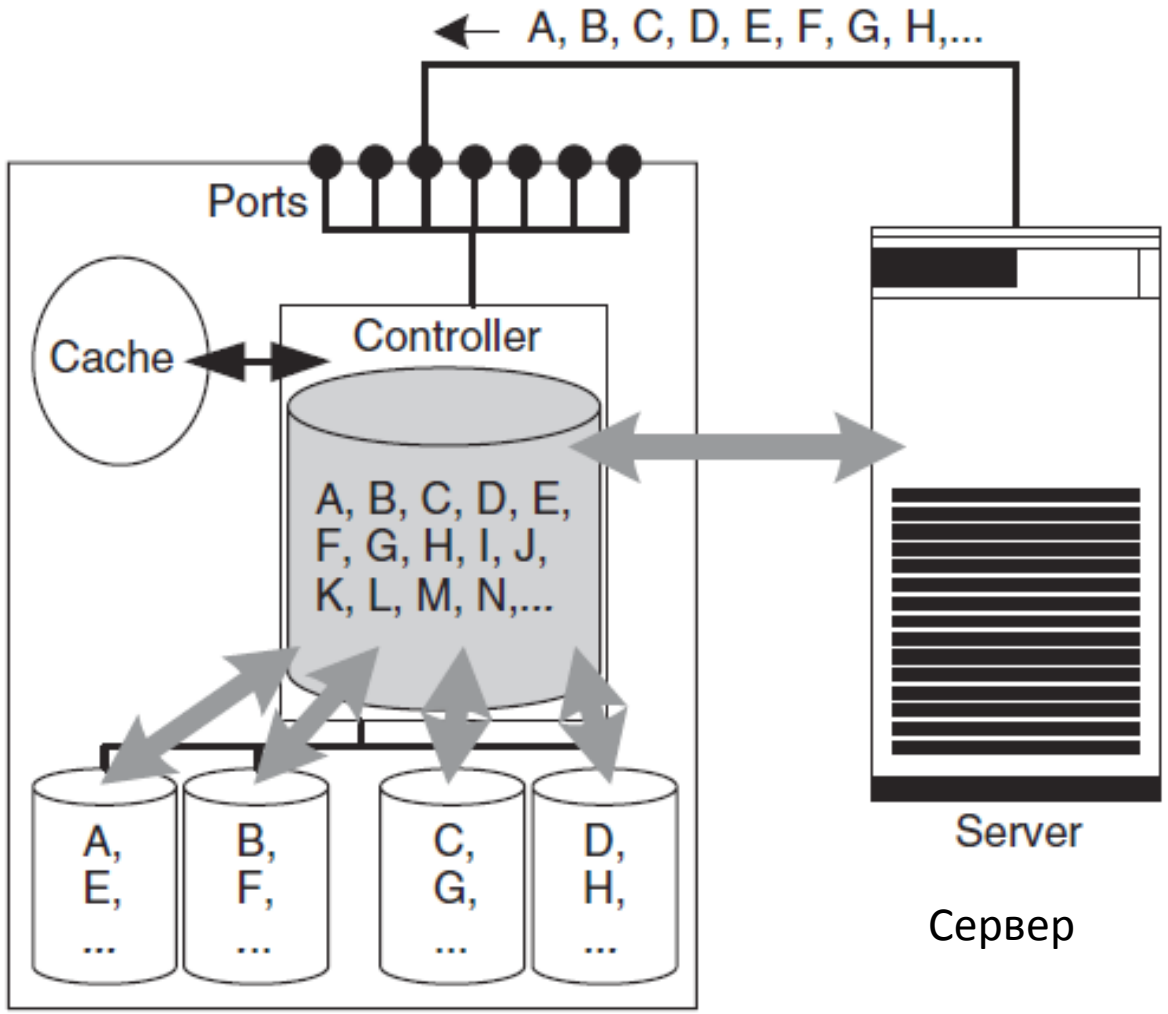
RAID - Redundant Array of Independent Disks



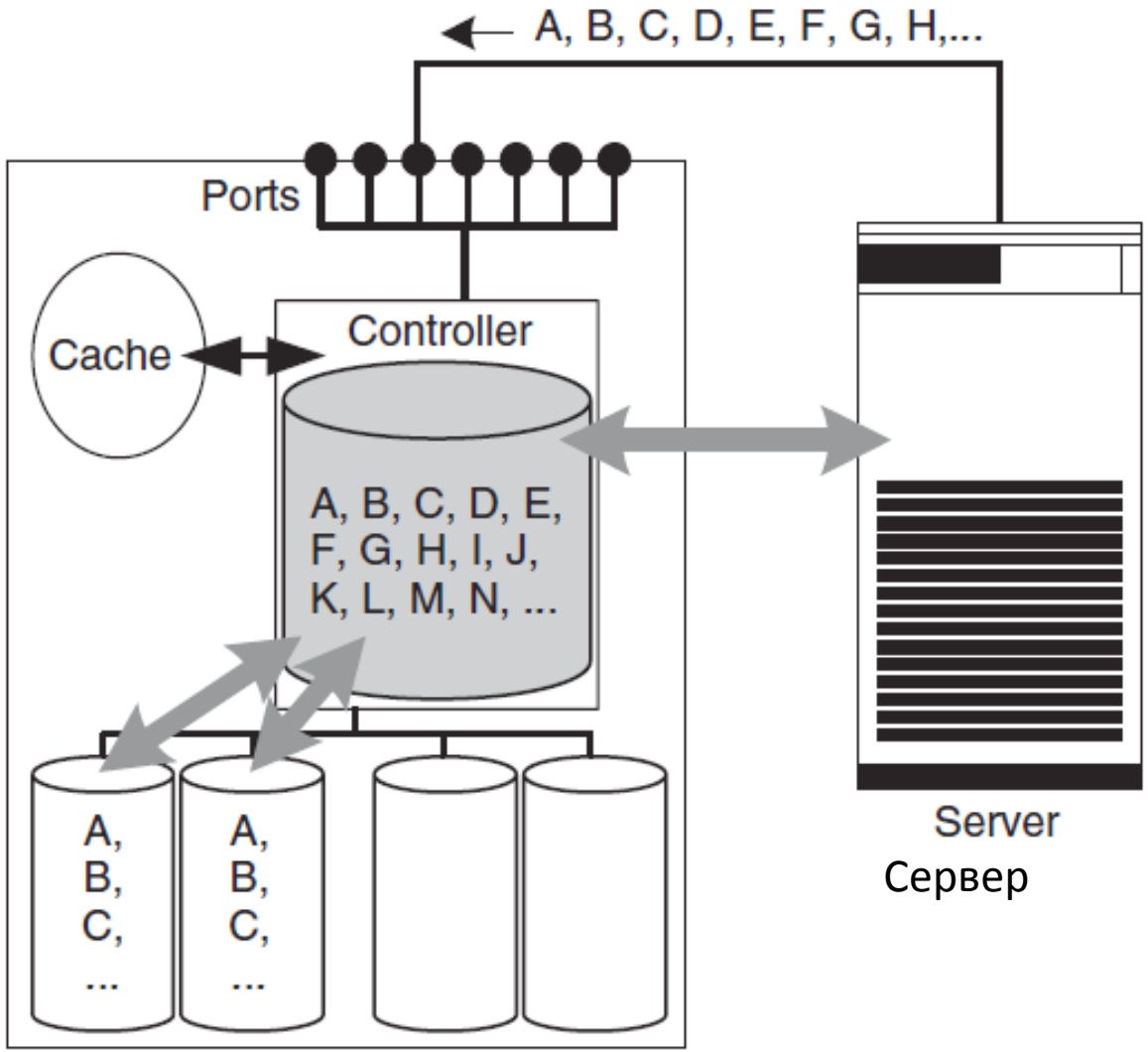
Горячий дисковый резерв



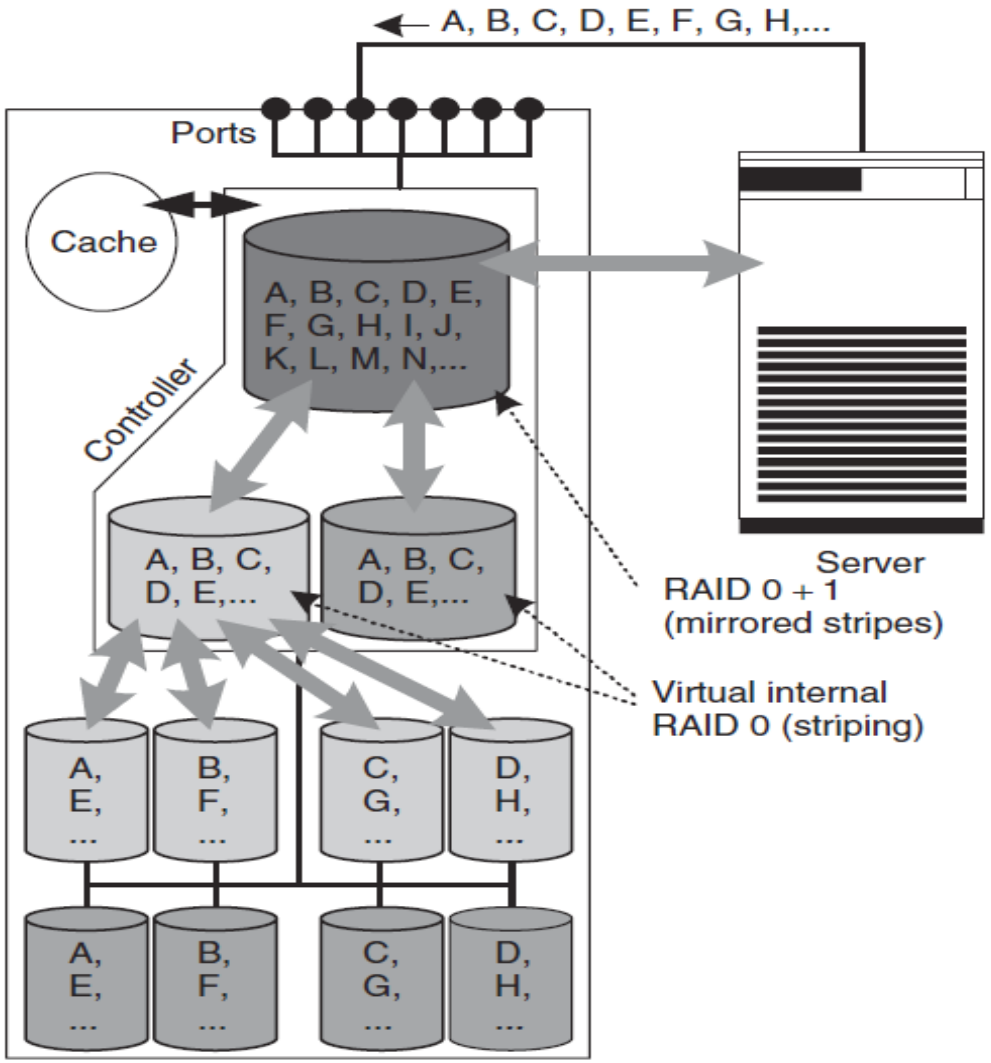
Уровни RAID = 0



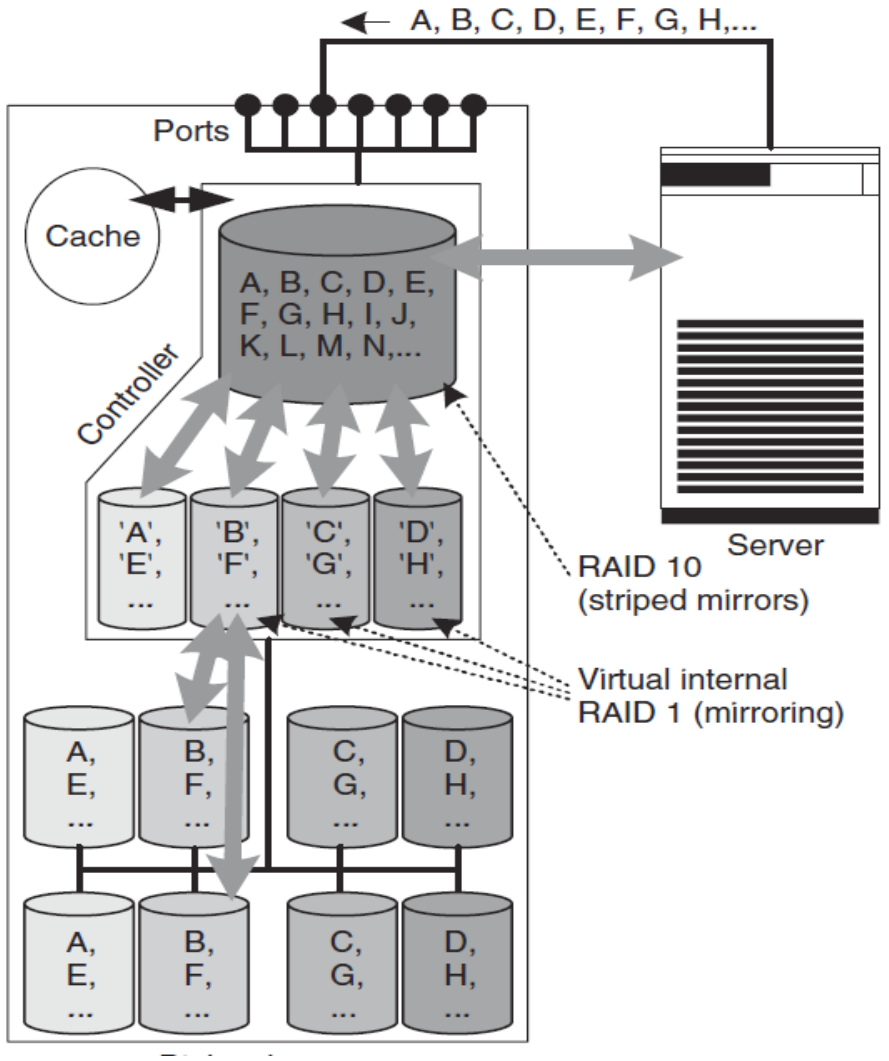
RAID = 1


A, B, C, D, E, F, G, H,...
Ports
Cache
Controller
A, B, C, D, E,
F, G, H, I, J,
K, L, M, N, ...
A,
B,
C,
...
A,
B,
C,
...
Server
Сервер

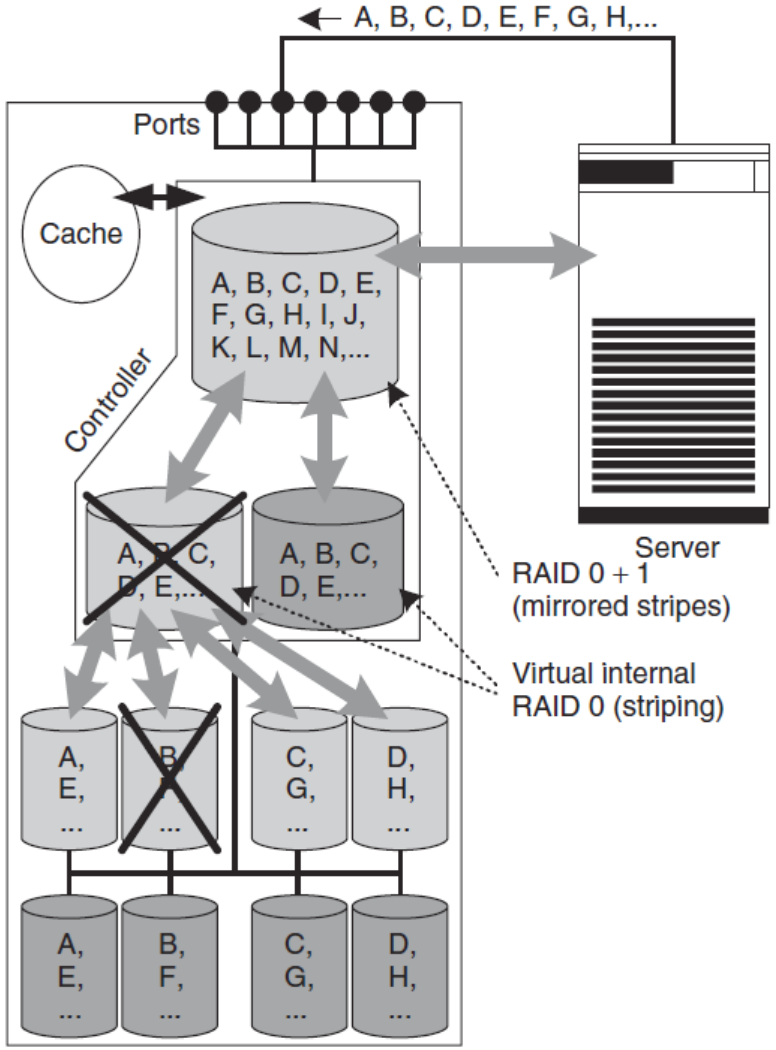
RAID = 0+1

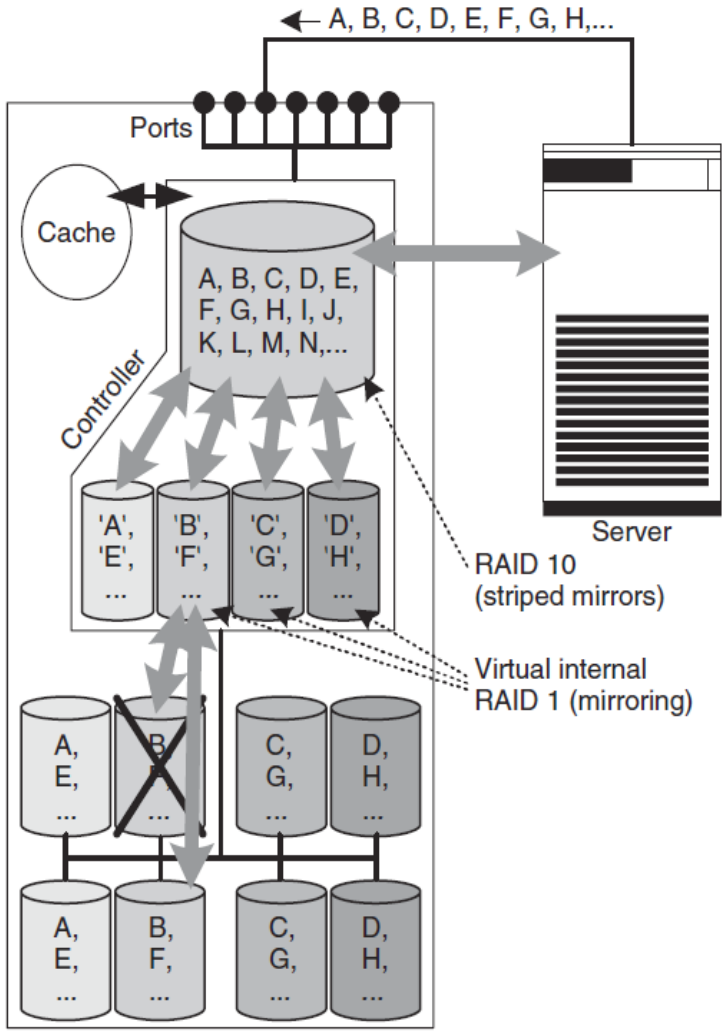

A, B, C, D, E, F, G, H,...
Ports
Cache
A, B, C, D, E, F, G, H, I, J, K, L, M, N,...
Controller
A, B, C, D, E,...
A, B, C, D, E,...
Server
RAID 0 + 1
(mirrored stripes)
Virtual internal
RAID 0 (striping)
A, E, ...
B, F, ...
C, G, ...
D, H, ...
A, E, ...
B, F, ...
C, G, ...
D, H, ...

RAID = 10

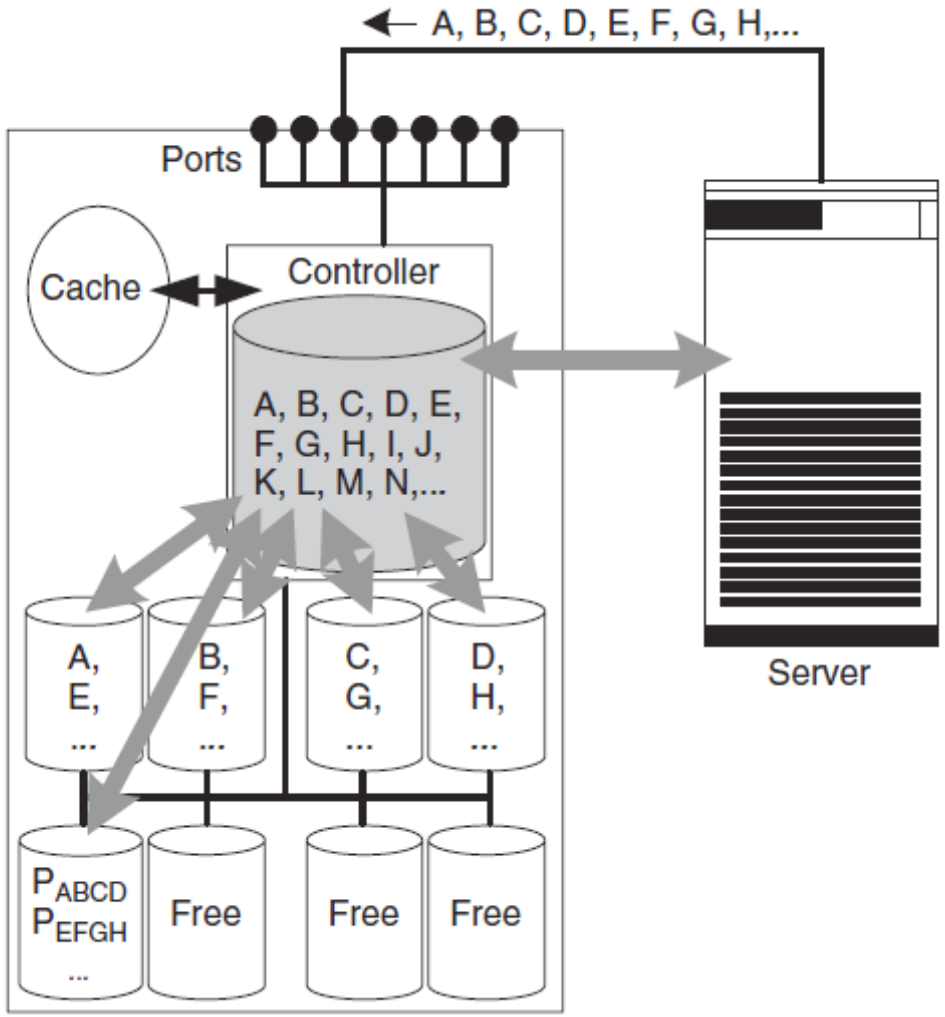

A, B, C, D, E, F, G, H,...
Ports
Cache
A, B, C, D, E,
F, G, H, I, J,
K, L, M, N,...
Controller
'A',
'E',
...
'B',
'F',
...
'C',
'G',
...
'D',
'H',
...
Server
RAID 10
(striped mirrors)
Virtual internal
RAID 1 (mirroring)
A,
E,
...
B,
F,
...
C,
G,
...
D,
H,
...
A,
E,
...
B,
F,
...
C,
G,
...
D,
H,
...

Сравнение RAID 0+1 и 10





RAID = 4 vs RAID = 5



A изменили на ~A

Δ = A xor ~A

~P = Δ xor P

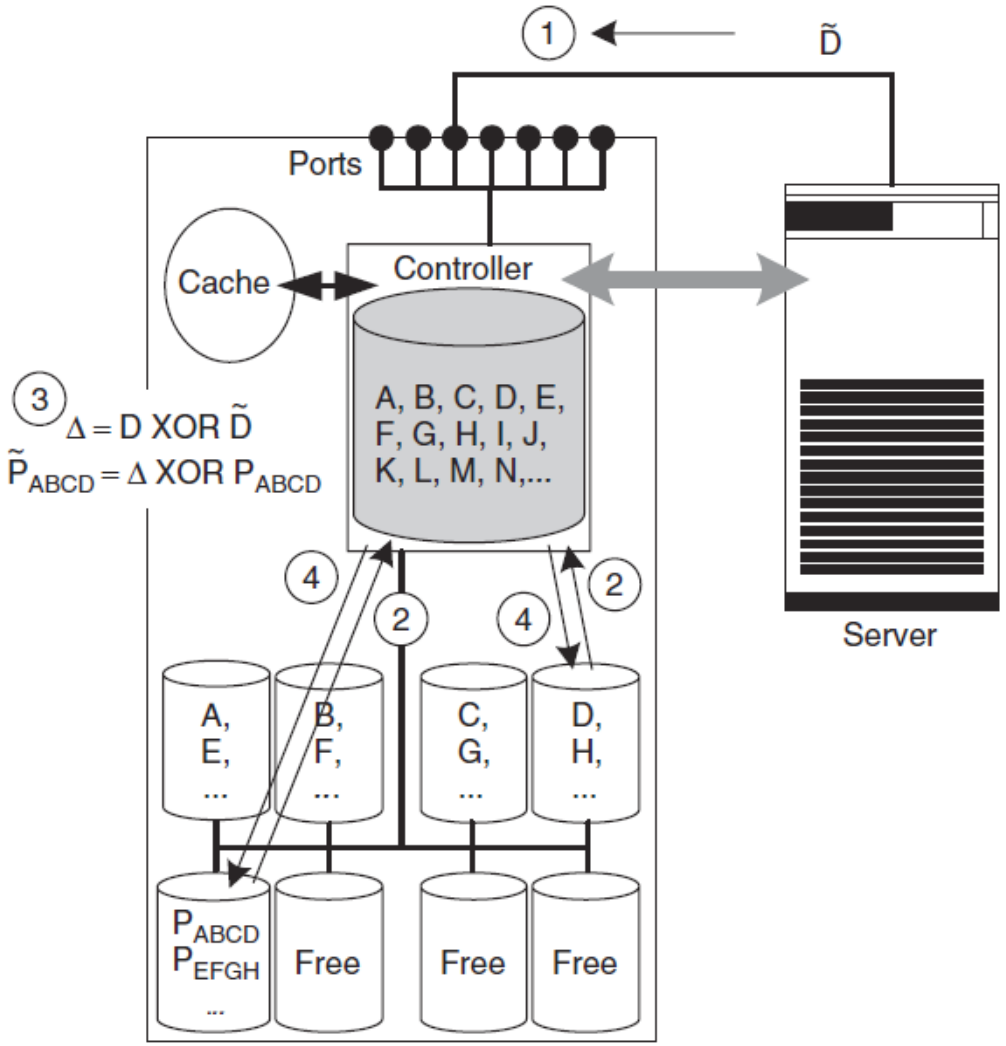
Если изменился только блок A,

То легко пересчитать P_{ABCD}, не зная BCD.

Однако надо считать старый блок A,

чтобы рассчитать Δ

Накладные расходы на запись в RAID = 4 и 5



RAID = 6

- Современные диски 1TB с BER 10^-15 => 100 TB одним сектором без ошибок не считать
- 10 дисковых массивов по 16X1TB будут терять один массив 1 раз в год+
- Режим эксплуатации теперь 7X24
- RAID 6 использует дополнительный диск четности
  - Увеличение затрат
  - Увеличение времени операции записи и коррекции

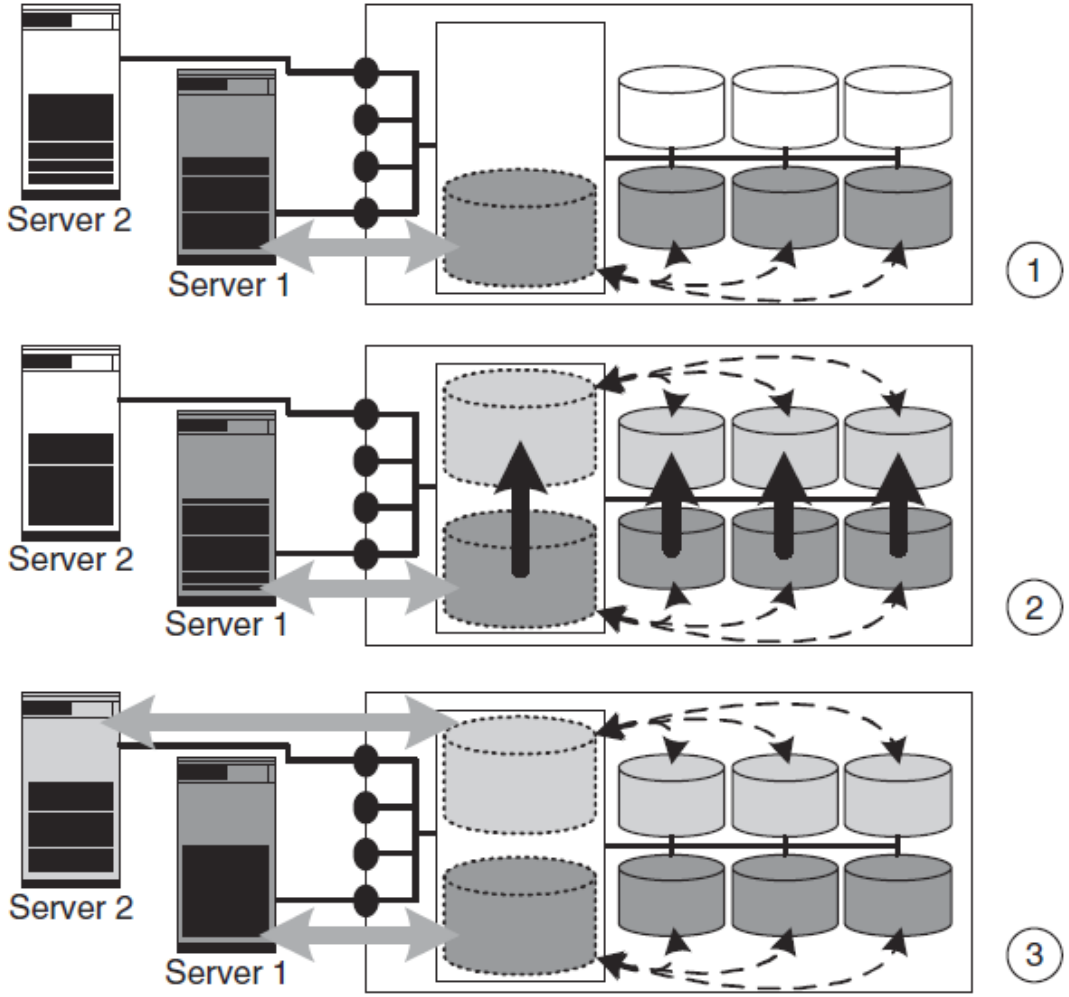
Сравнение схем RAID массивов

RAID level	Fault-tolerance	Read performance	Write performance	Space requirement
RAID 0	None	Good	Very good	Minimal
RAID 1	High	Poor	Poor	High
RAID 10	Very high	Very good	Good	High
RAID 4	High	Good	Very very poor	Low
RAID 5	High	Good	Very poor	Low
RAID 6	Very high	Good	Very very poor	Low

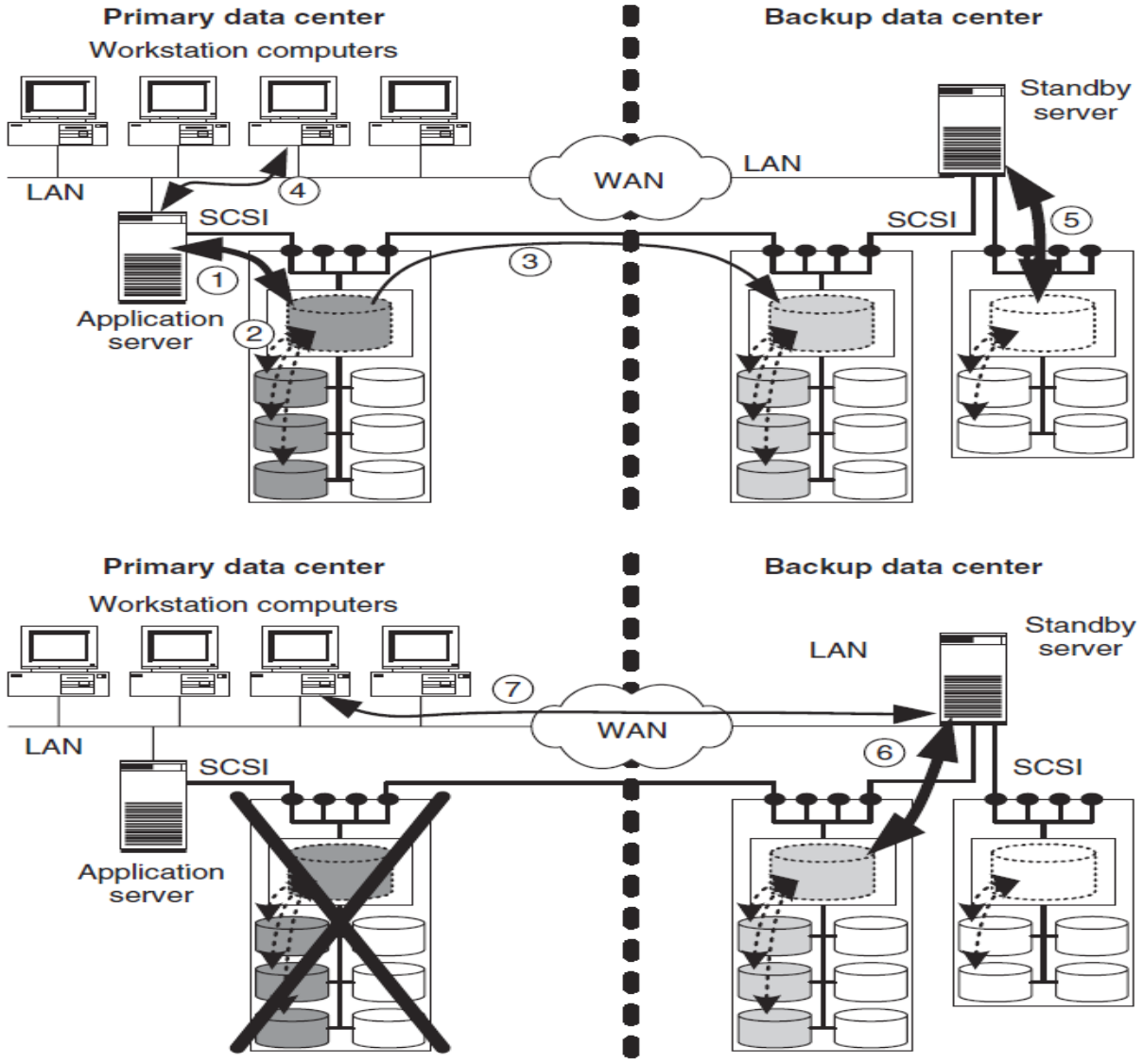
Caching – ускорение дисковых хранилищ

- Кэш на уровне HD
- Кэш на уровне контроллера ДС при записи
  - ГБ кэш
  - Здесь главное сохранить данные в кэш даже при отключении питания (UPS)
  - Важно для блочных приложений
- Кэш для ускорения чтения контроллером ДС

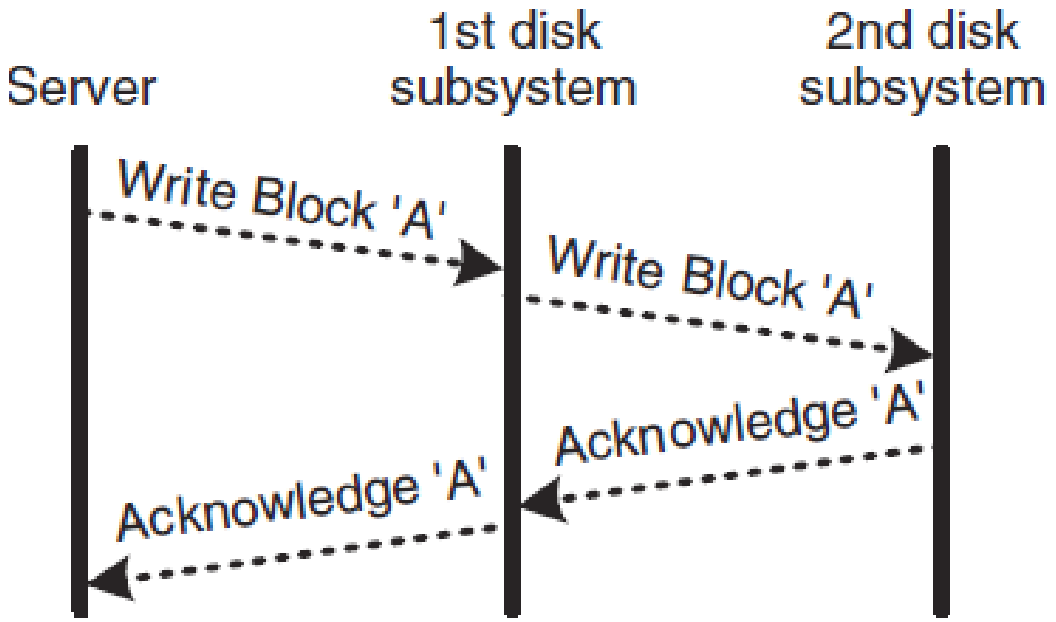
Интеллектуальная дисковая подсистема

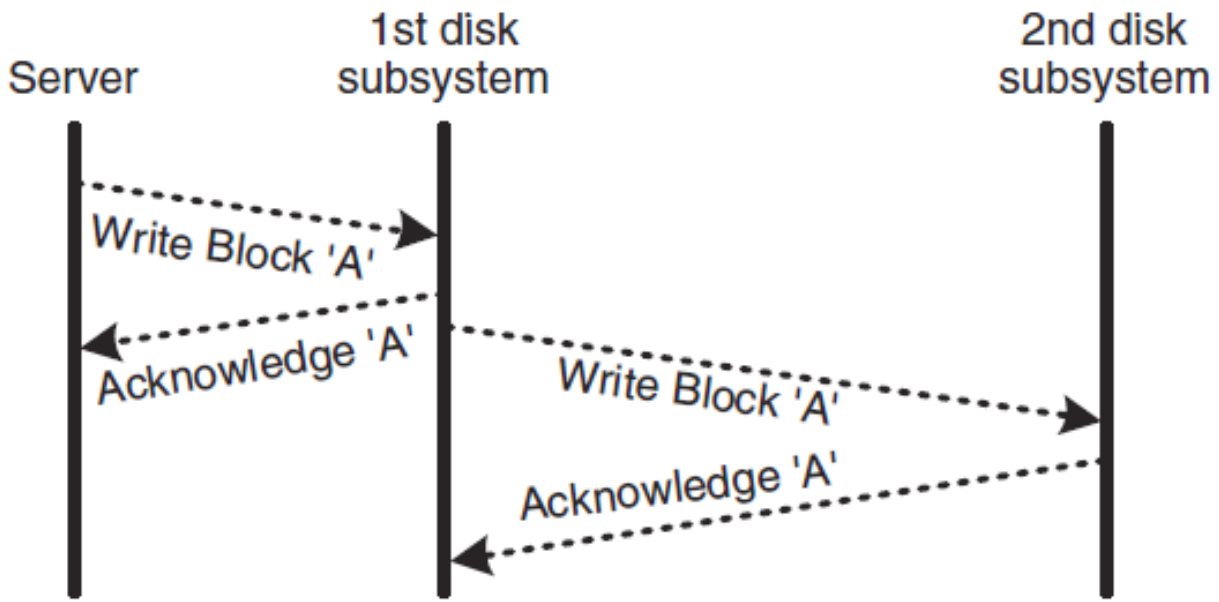


Удаленное зеркалирование

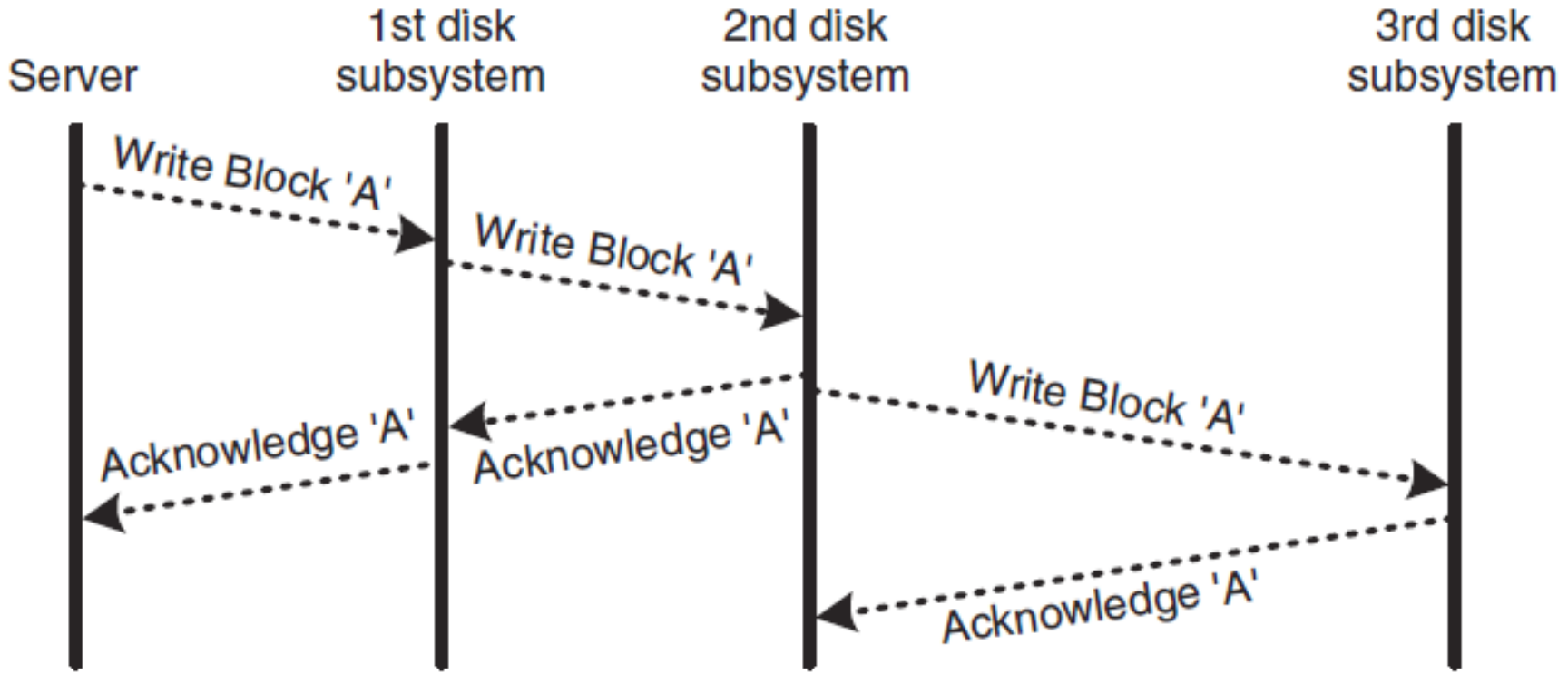


Синхронное и асинхронное зеркалирование

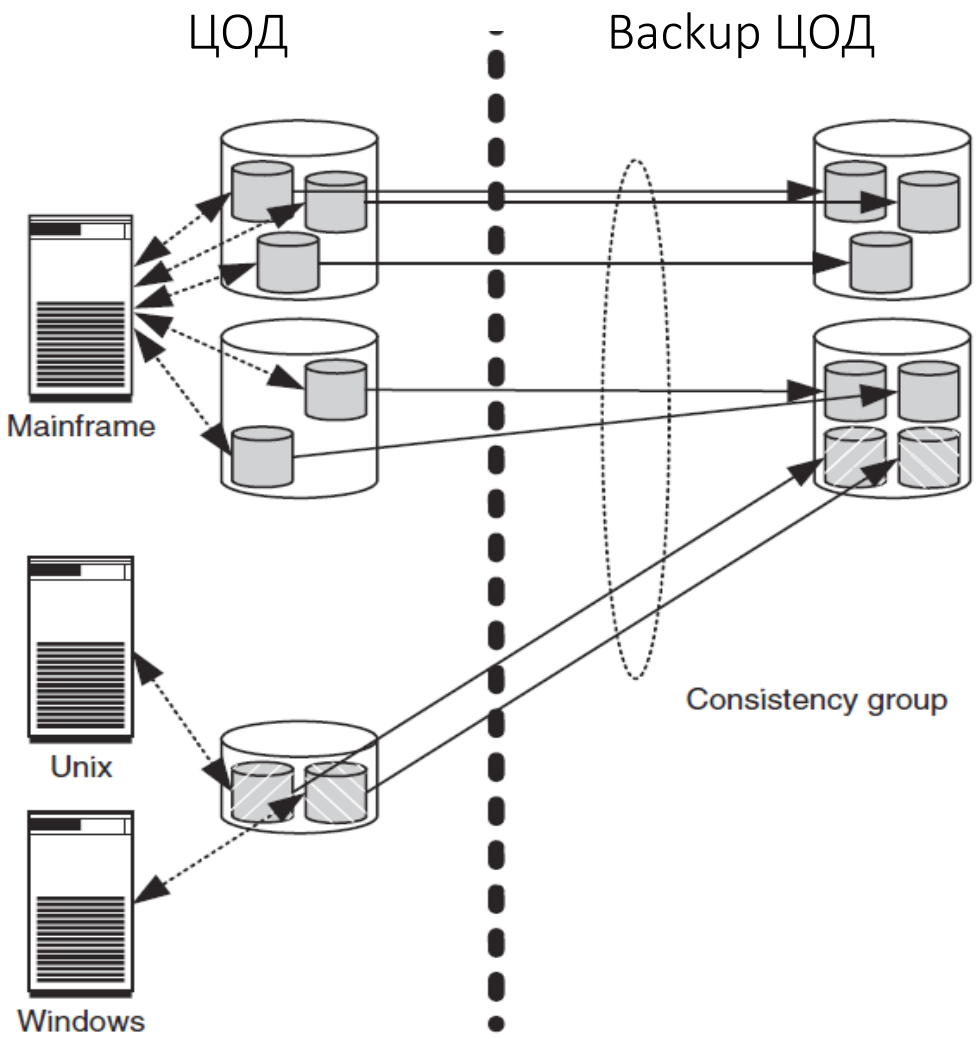




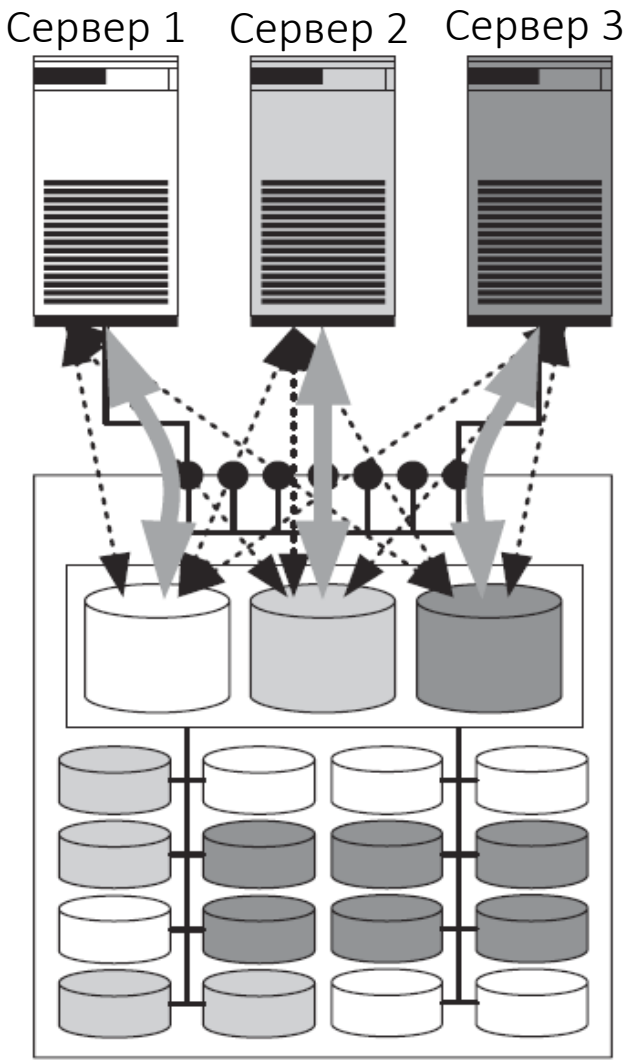
Комбинированная схема удалённого зеркалирования



Групповая консистентность



LUN маскирование

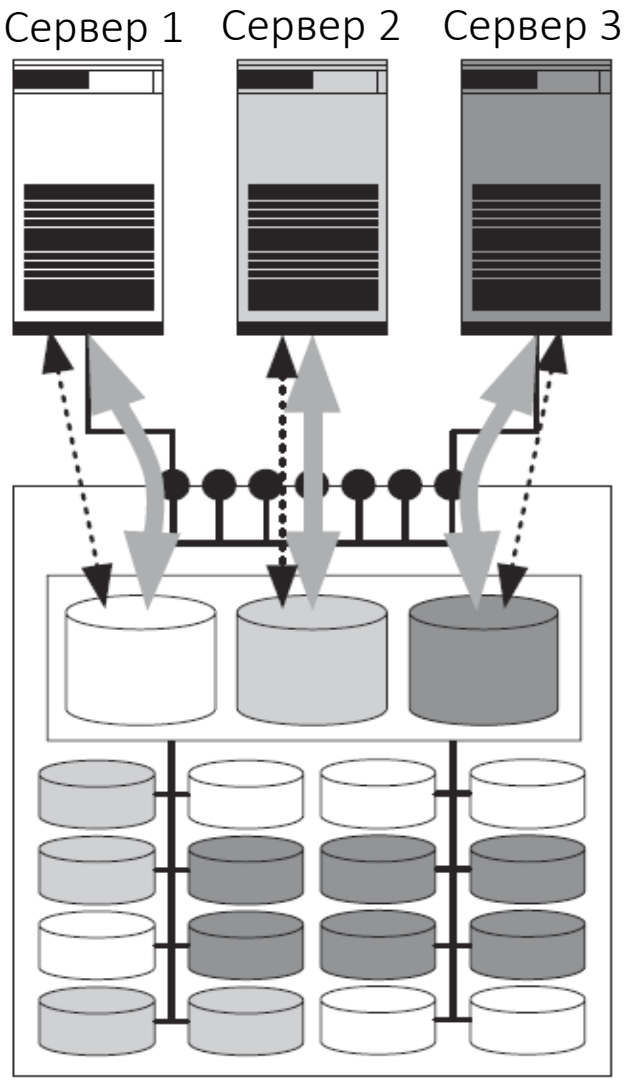


Дисковая подсистема

Local
Unit
Number

Сервер использует LUN

Сервер видит LUN



Дисковая подсистема

Устойчивость работоспособности ДПС

• Данные распределяют по нескольким дискам с помощью механизмов RAID и снабжают избыточными данными (блоки четности).

• На каждом физическом диске данные закодированы кодом Хемминга. Кроме этого диск оснащен подсистемой самодиагностики, которая контролирует частоту ошибок, вибрацию шпинделя и т.д. Это позволяет проактивно прогнозировать отказы диска.

• Каждый диск подсоединен к контроллеру хотя бы через две внутренние шины.

• Контроллер дисковой подсистемы может быть продублирован. Выход одного экземпляра, автоматически будет активизировать следующий экземпляр.

• Дублируются UPS, системы охлаждения так же. ДС подключают к разным электрическим сетям

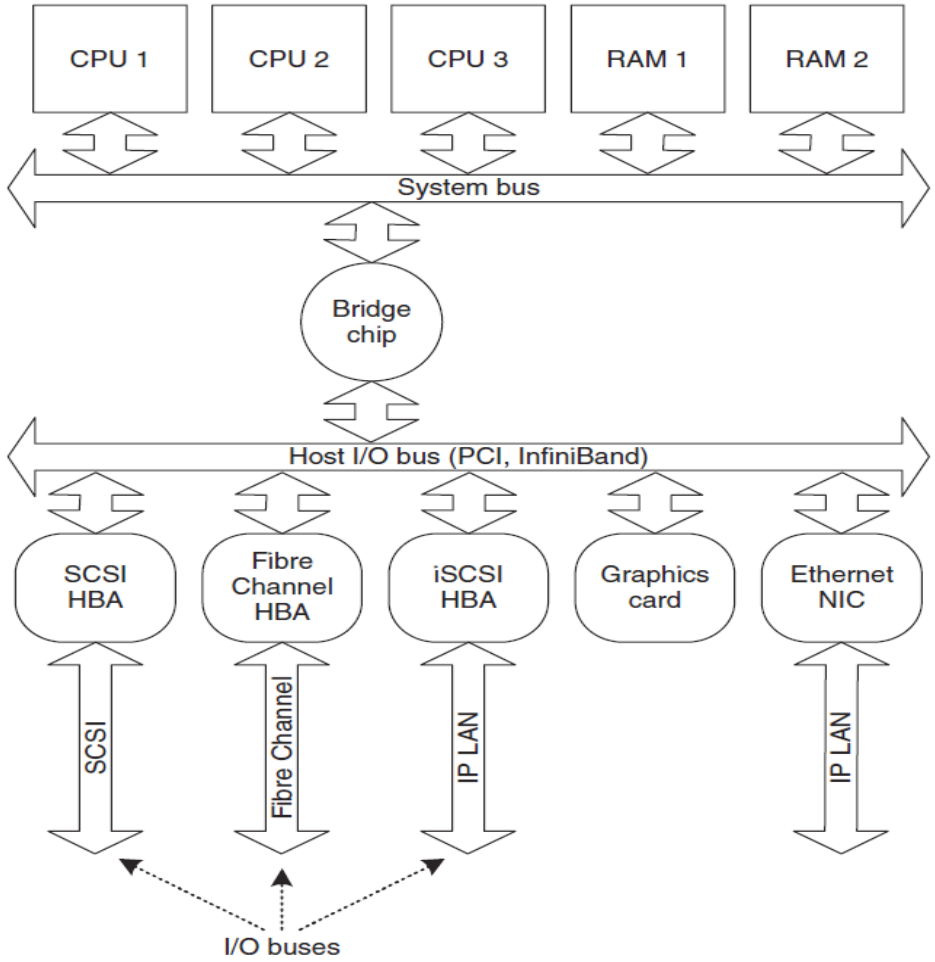
• Сервер соединяют с ДС через несколько линий.

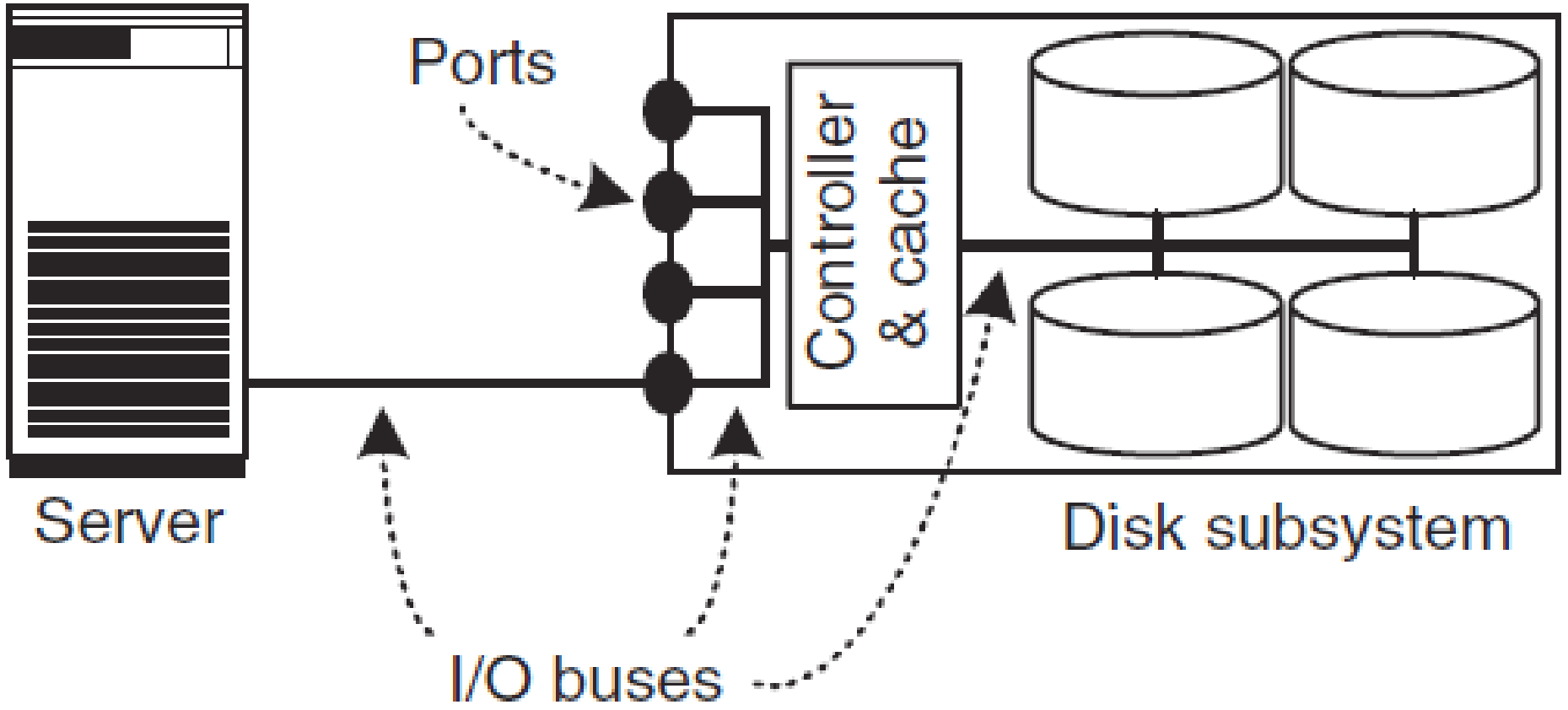
• Мгновенное копирование используют от логических ошибок. Например, создание мгновенной копии данных через каждый час. Тогда в случае сбоя и уничтожения какой-то таблицы, она может быть восстановлена.

• Удаленное зеркалирование используют от физического уничтожения или повреждения оборудования. В сочетании с мгновенным копирование эти сервисы гарантируют сохранение и консистентность данных даже для нескольких виртуальных дисков или дисковых подсистем.

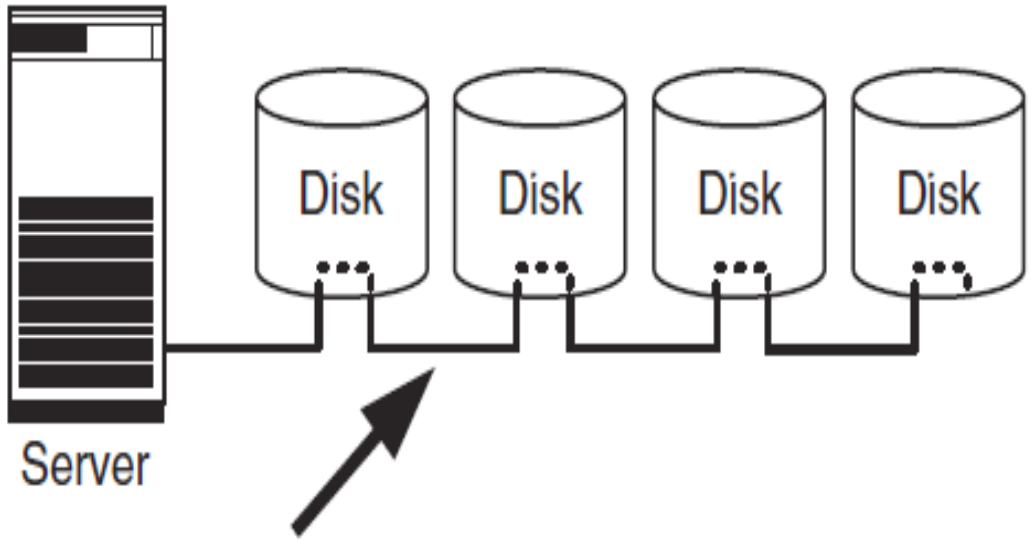
• LUN маскирование защищает от несанкционированного доступа, упрощает работу системного администратора, защищает от случайных сбоев в работе приложений серверов и их оборудования.

Тракт от CPU до СХД



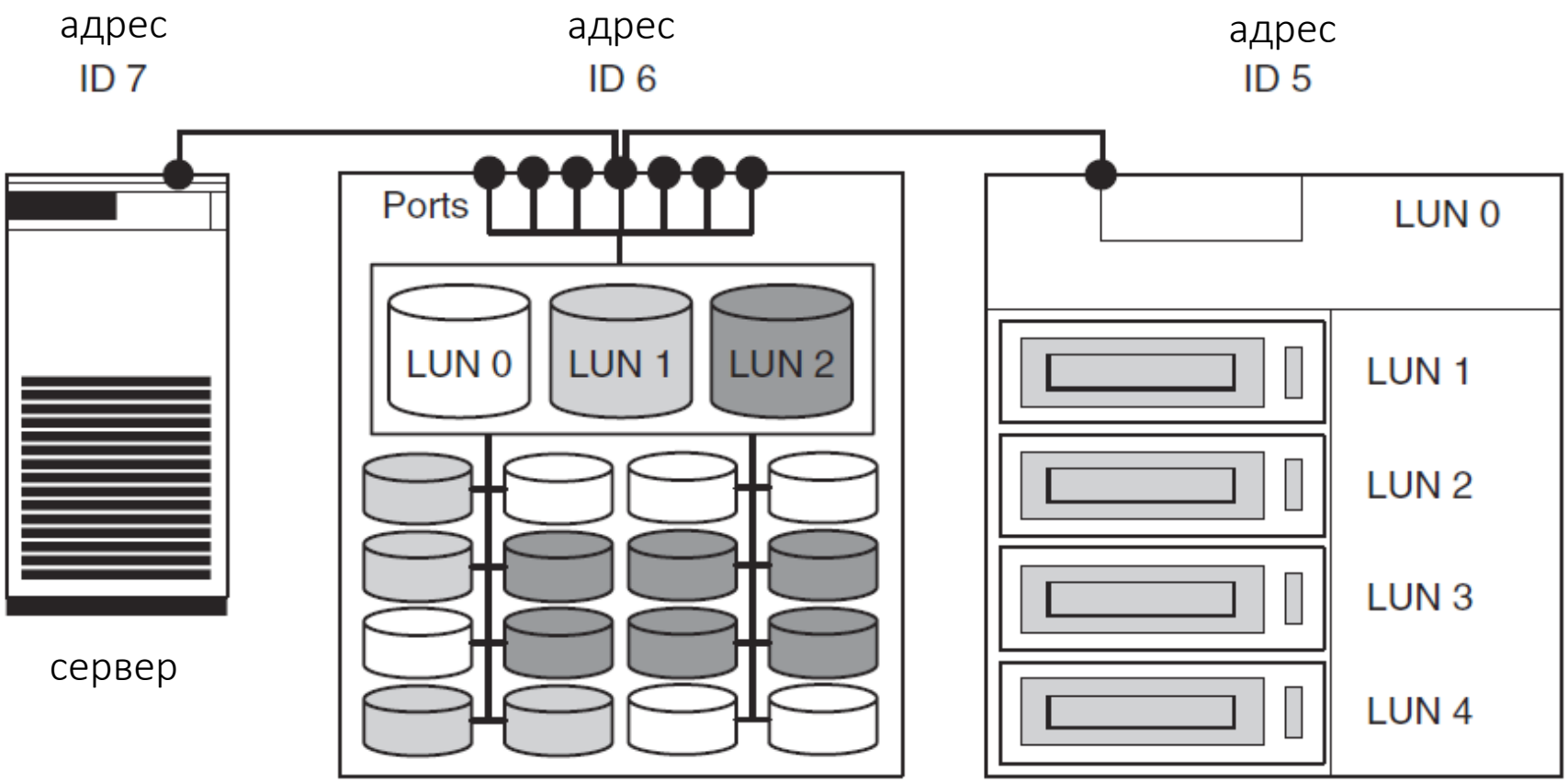

Ports
Controller
& cache
Server
Disk subsystem
I/O buses

Small Computer System Interface (SCSI)

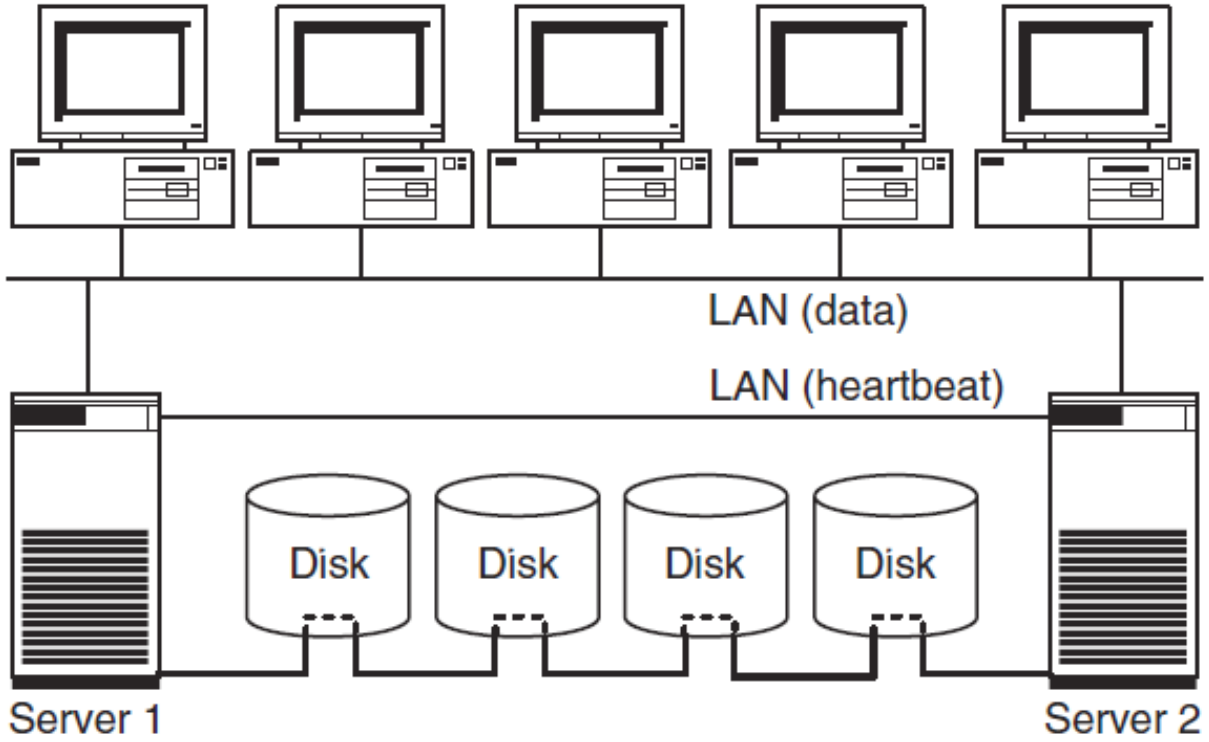


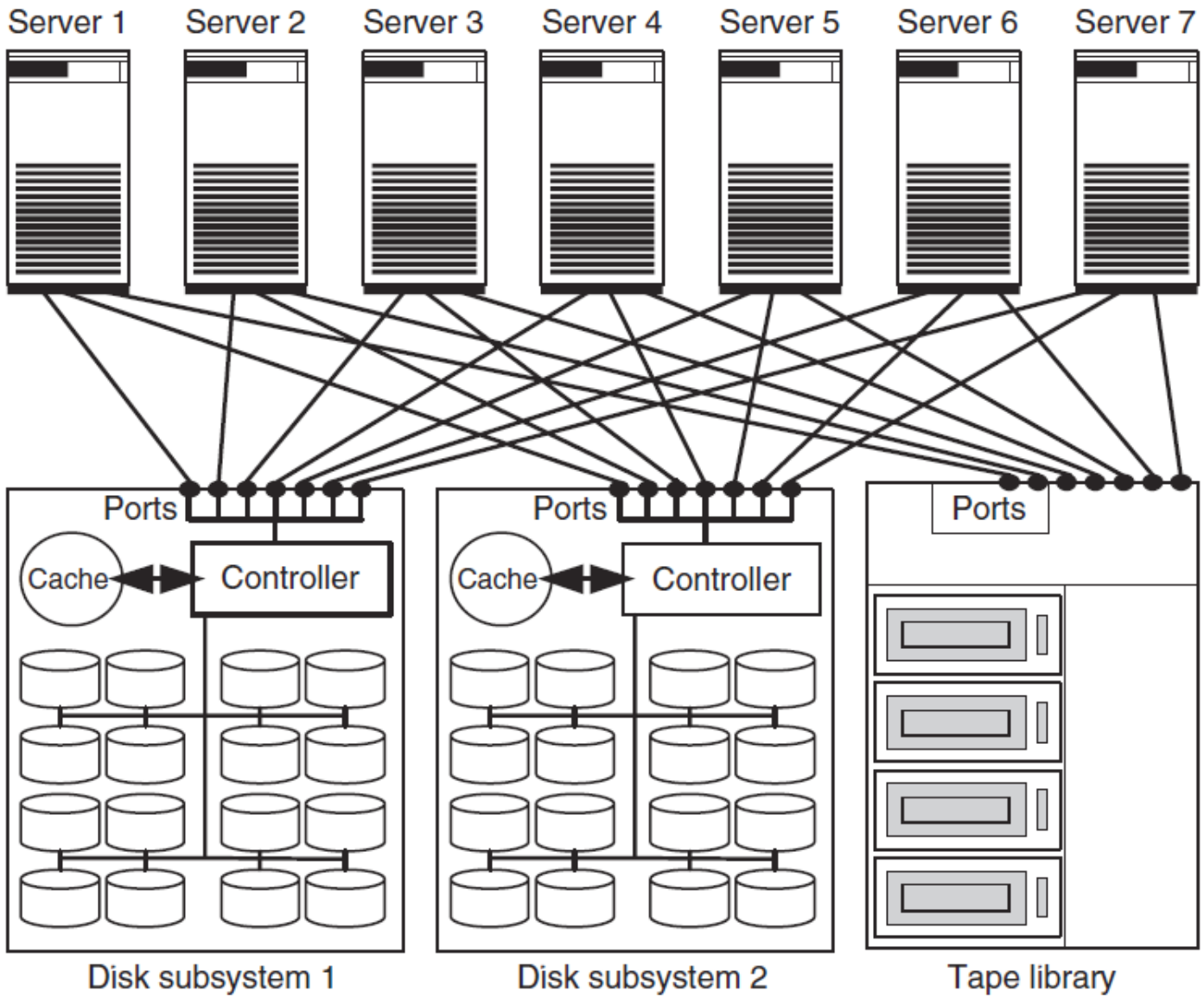
SCSI version	MByte/s MByte/s	Bus width	Max. no. of devices
SCSI-2	5	8	8
Wide Ultra SCSI	40	16	16
Wide Ultra SCSI	40	16	8
Wide Ultra SCSI	40	16	4
Ultra2 SCSI	40	8	8
Wide Ultra2 SCSI	80	16	16
Ultra3 SCSI	160	16	16
Ultra320 SCSI	320	16	16

Адресация устройств на шине SCSI



SCSI и сеть хранения




Server 1
Server 2
Server 3
Server 4
Server 5
Server 6
Server 7
Ports
Cache
Controller
Ports
Cache
Controller
Ports
Disk subsystem 1
Disk subsystem 2
Tape library

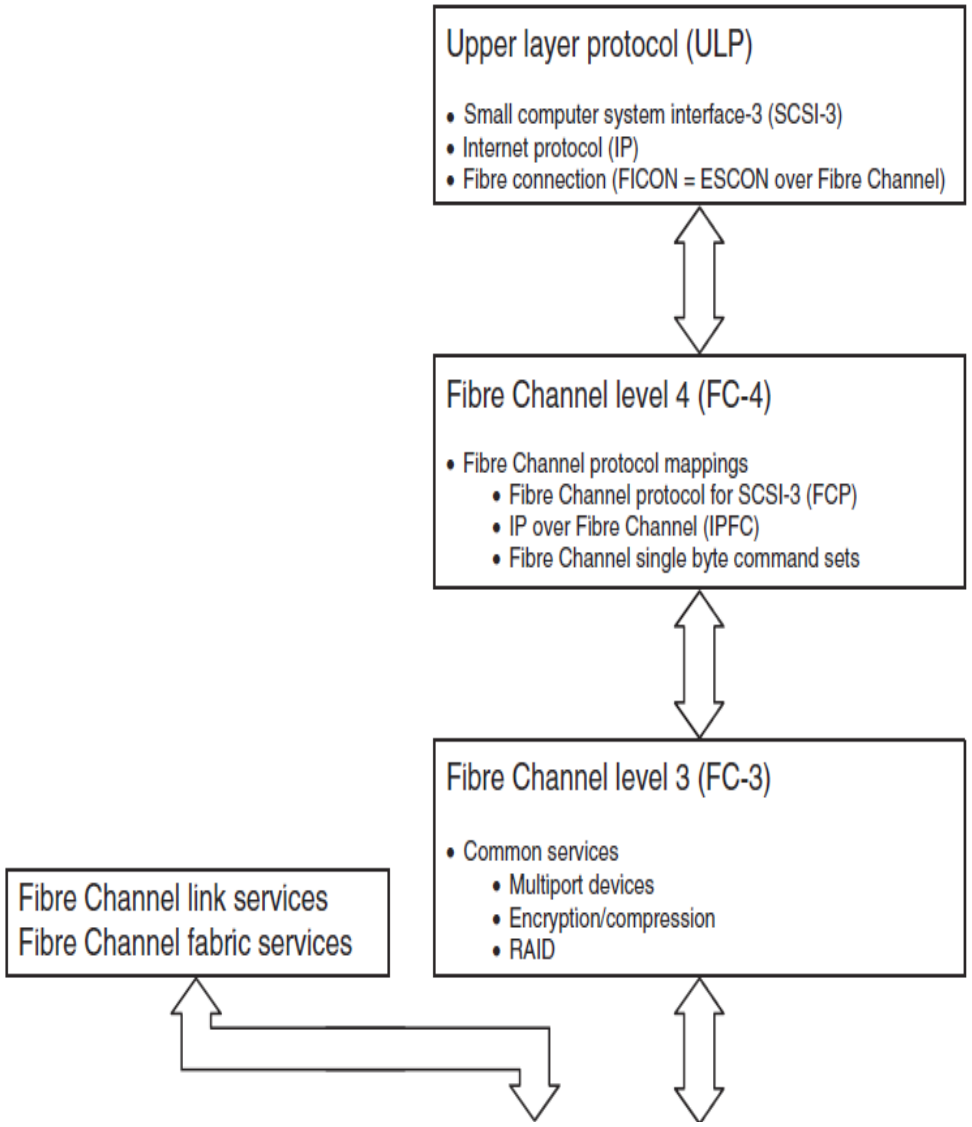
Fibre Channel

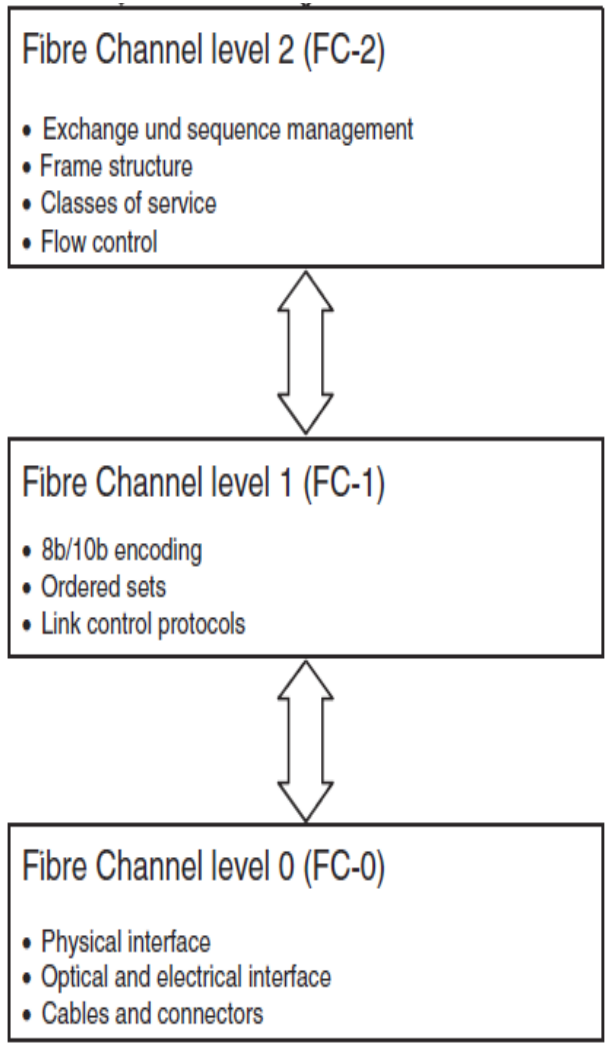
- I/O канал
- Сетевое взаимодействие
- Fiber Channel сопрягаем с IPI (Intelligent Peripheral Interface), SCSI, HIPPI (High Performance Parallel Interface), ATM, IP и 802.2.
- Fibre Channel - n x100МБ/с при длинах канала 10 км и более, где *n* – число каналов. Предельная скорость передачи - 4,25 Гбод.
- В качестве физической среды может использоваться одномодовое или мультимодовое оптическое волокно. Допускается применение медного коаксиального кабеля и витых пар (при скоростях до 200 МБ/с).

Fiber Channel

- FC-0 определяет физические характеристики интерфейса и среды, включая кабели, разъемы, драйверы (ECL, LED, лазеры), передатчики и приемники. Вместе с FC-1 этот уровень образует физический слой.
- FC-1 определяет метод кодирования/декодирования (8B/10B) и протокол передачи, где объединяется пересылка данных и синхронизирующей информации.
- FC-2 определяет правила сигнального протокола, классы услуг, топологию, методику сегментации, задает формат кадра и описывает передачу информационных кадров.
- FC-3 определяет работу нескольких портов на одном узле и обеспечивает общие виды сервиса.
- FC-4 обеспечивает реализацию набора прикладных команд и протоколов вышележащего уровня (например, для SCSI, IPI, IEEE 802, SBCCS, HIPPI, IP, ATM и т.д.)

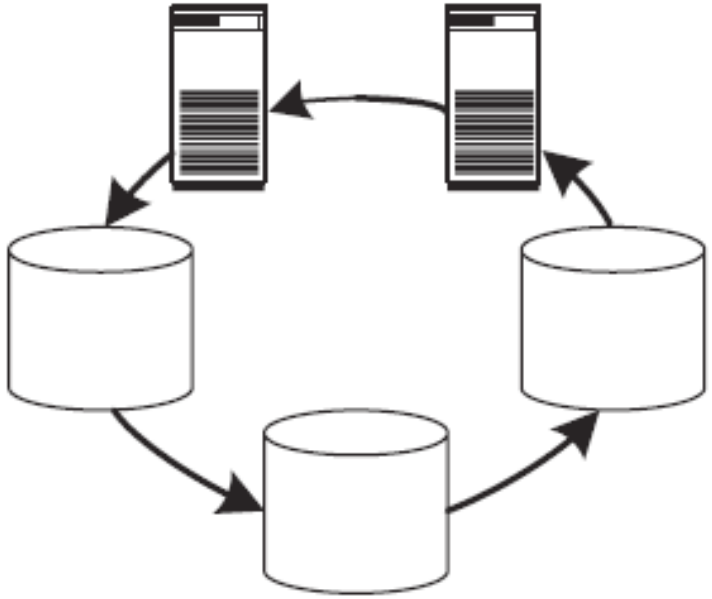
FIBRE CHANNEL



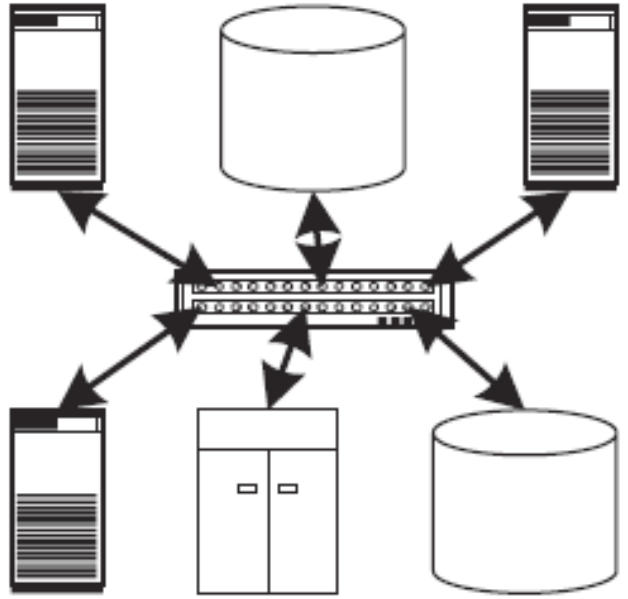


Топологии FC

Точка-Точка



Кольцо с арбитражем



Коммутатор

Fibre Channel: типы портов

- N-Port – порт определяет характеристики для соединения через коммутатор или P2P
- F-Port – порт для подключения к коммутатору
- L-Port – порт для КсА
- NL-Port – работает как N Port или как L-Port. Можно подключать порт через коммутатор или в КсА.
- FL-Port – для подключения коммутатора в КсА
- E-Port – для соединения двух FC коммутаторов
- G-Port – может настраиваться как E или FL в зависимости от подключения.
- B-Port – для соединения двух FC коммутаторов через ATM, SDH, Ethernet или IP. Например две FC SAN могут быть соединены через WAN.

FC-0: разъемы, кабели и кодировка

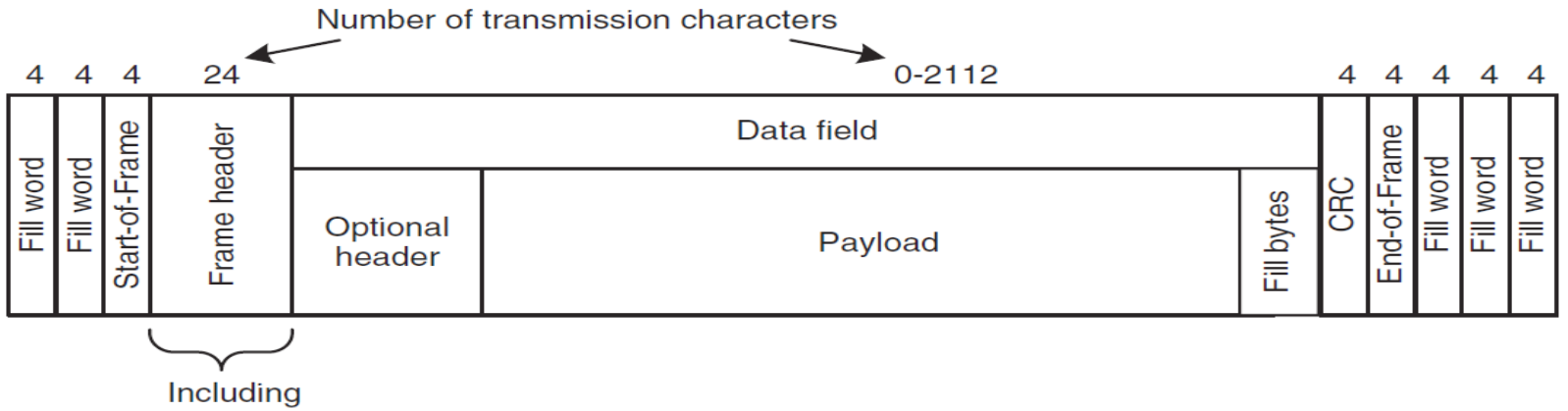
- Скорость от 100 МБ/с до 10ГБ/с ожидается до 20ГБ/с в одном направлении
- Есть два вида оборудования Base 2 и Base 10
- Передача последовательная
- BER = 10^12, т.е. для линии 100Мб/с ошибка не чаще чем раз в 16,6 мин. Протоколы верхнего уровня для обнаружения ошибок оптимизированы под эту величину BER
- Преимущества оптического кабеля

FC-1: кодировка, упорядоченные наборы, управление линией

- 8b/10b кодирование
- Transmission words
  - Data word: SOF, 4 bytes, EOF
  - Ordered set: EOF, K28.5, SOF
- Управление линией

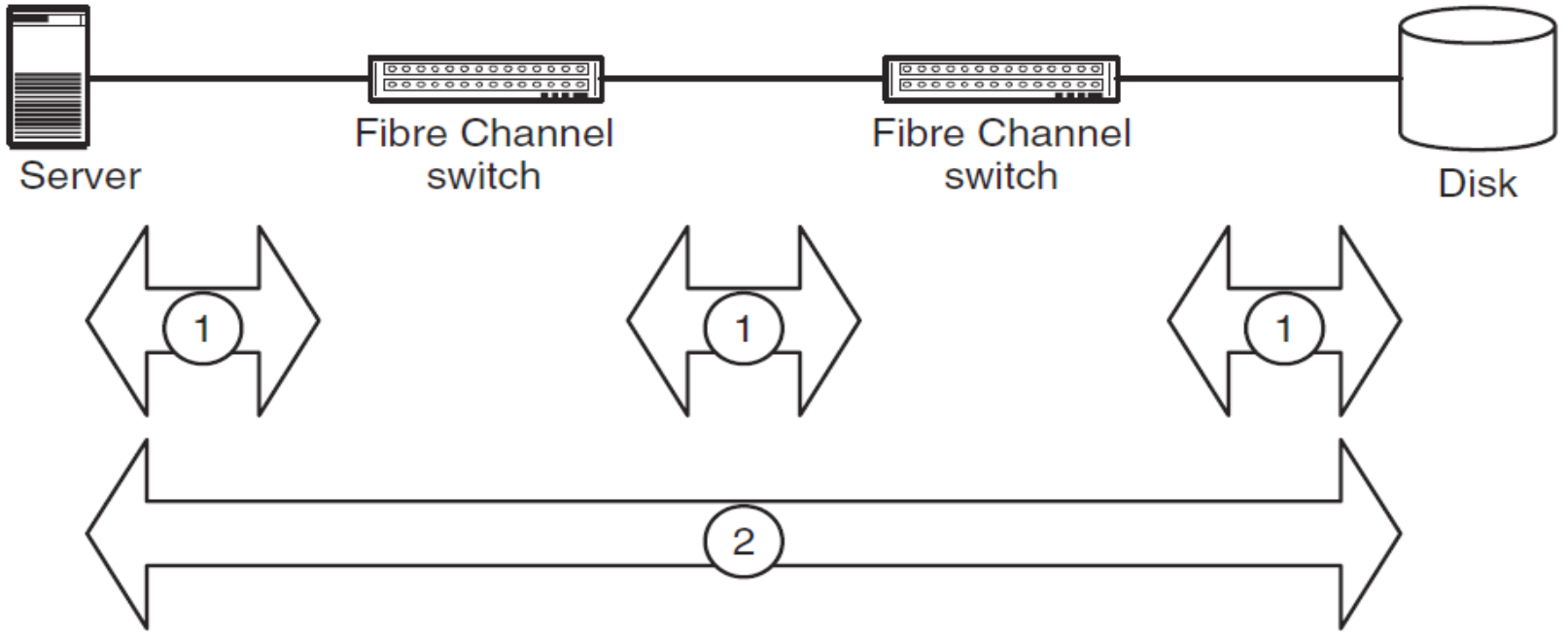
FC-2: передача данных

- Определение размера передаваемых данных
  - Exchange – сессия между логическими сущностями (процессы)
  - Sequence – последовательность кадров
  - Frame – управления и данных (2 112 Б)
- Управление потоком
- Классы сервиса



Управление потоком

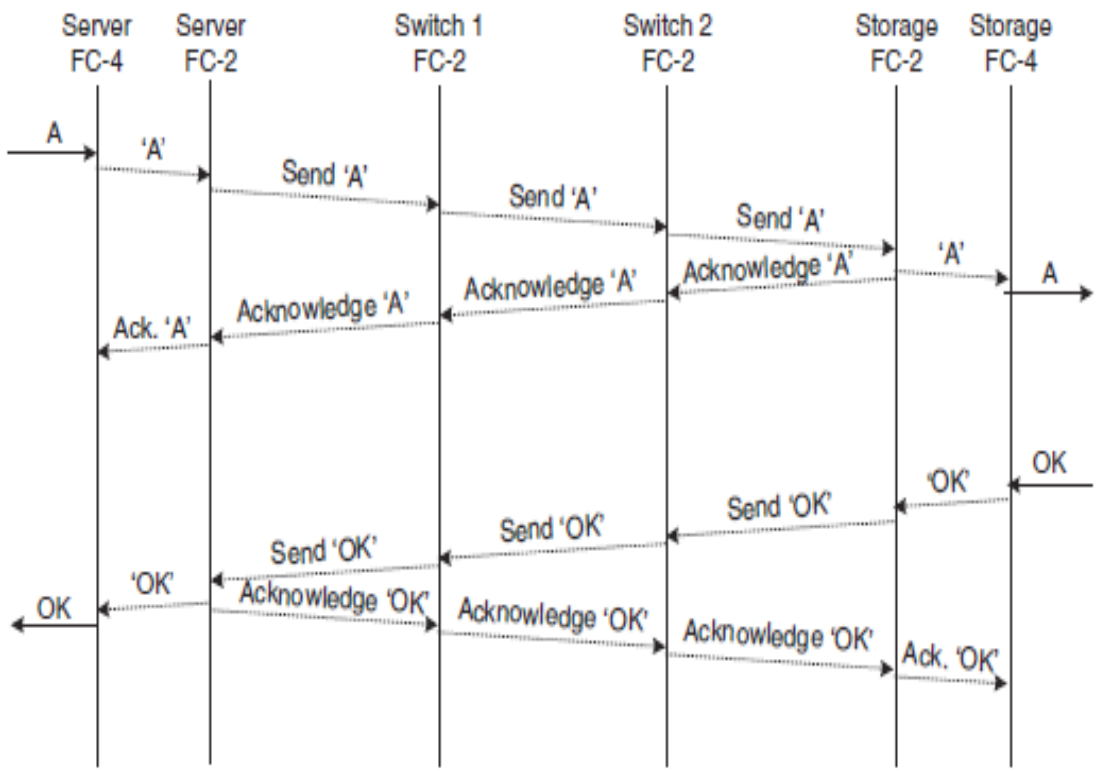
- Кредитная схема
- E2E flow control (2)
- Link flow control (1)

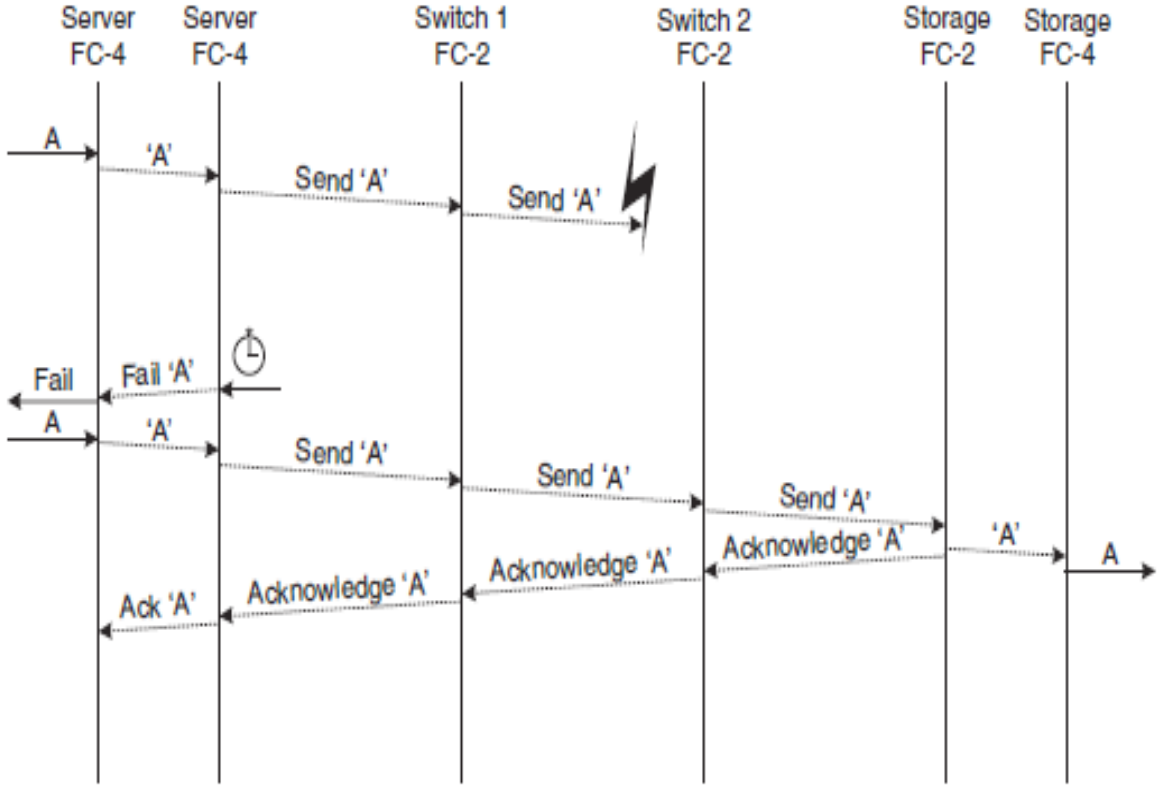


Классы обслуживания

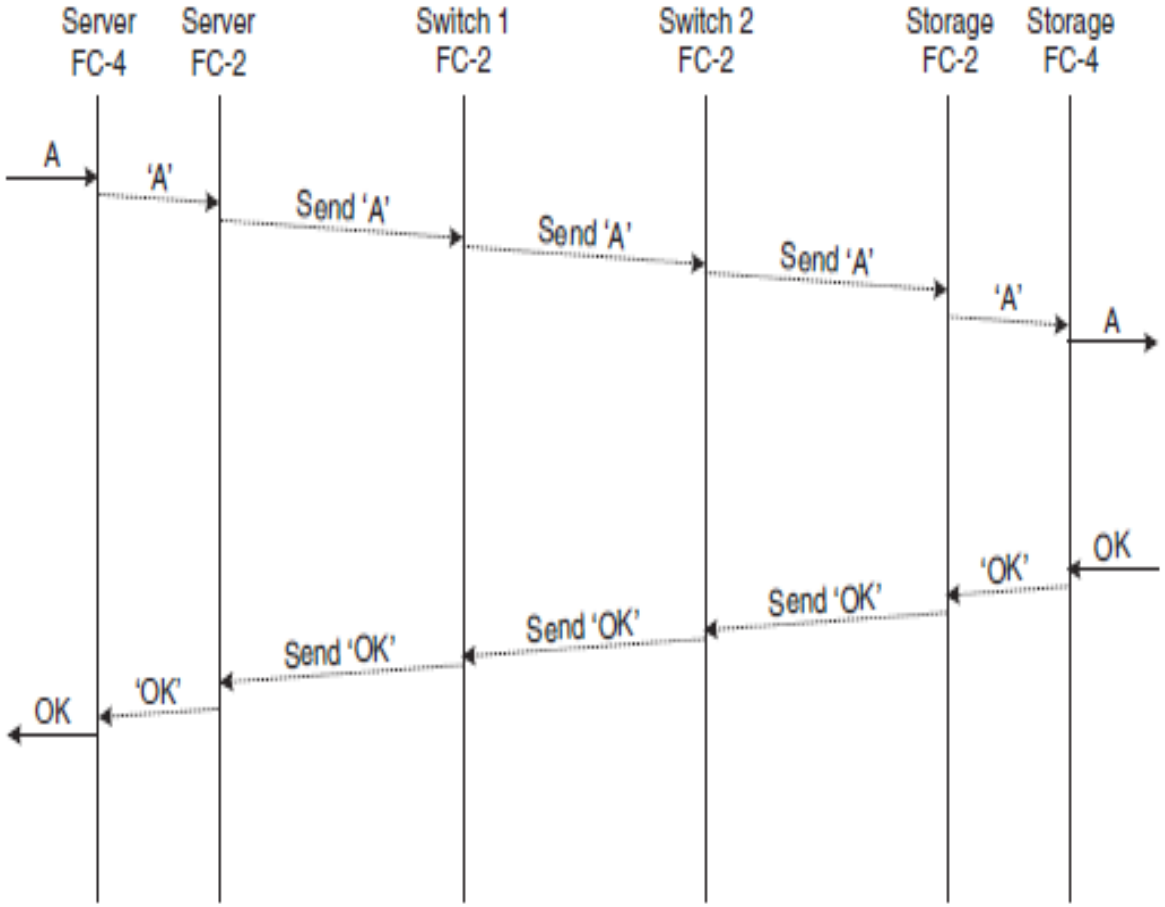
Класс 1	Соединение точка-точка (end-to-end) между портами типа n_port через коммутацию каналов. Класс удобен для аудио и видео приложений, например, видеоконференций. После установления соединения используется вся доступная полоса пропускания канала. При этом гарантируется, что кадры будут получены в том же порядке, в каком они были посланы. Есть управление потоком.
Класс 2	Без установления соединения с коммутацией пакетов, гарантирующий доставку данных. Порт может взаимодействовать одновременно с любым числом портов типа n_port в режиме дуплекс. Не гарантируется порядок доставки кадров, кроме соединения P2P или KcA. Есть управление потоком. Этот класс характерен для локальных сетей, где время доставки данных не является критическим.
Класс 3	Обмен дейтограммами без установления соединения и без гарантии доставки. Есть управление потоком. Применяется для каналов SCSI.
Класс 4	Обеспечивает выделение определенной доли пропускной способности канала с заданным качеством обслуживания (QoS). Только для топологии структура матрица с n_port. Гарантируется порядок доставки кадров.
Класс 5	Регламентирующие документы находятся в процессе подготовки.
Класс 6	Предусматривает групповое-обслуживание с коммутацией.

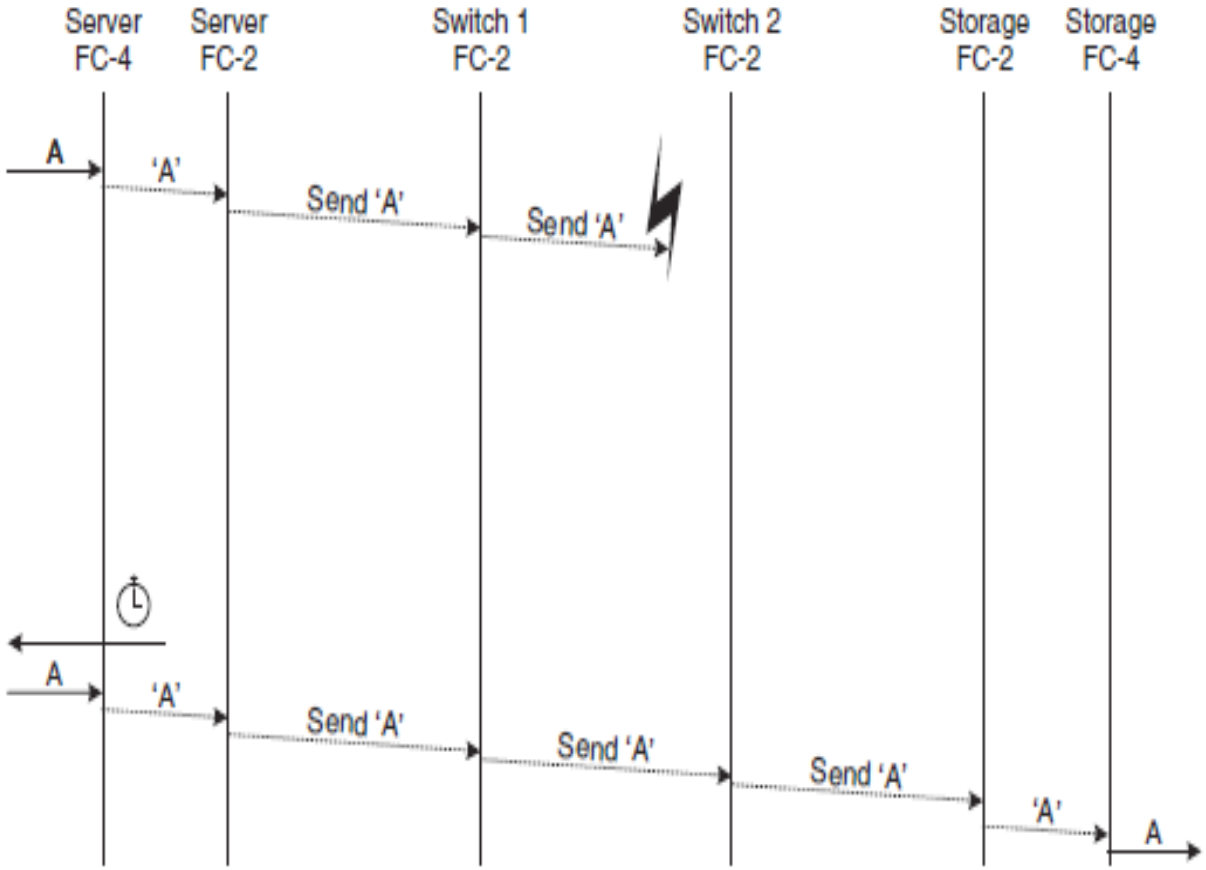
FC-2 класс 2


Server FC-4
Server FC-2
Switch 1 FC-2
Switch 2 FC-2
Storage FC-2
Storage FC-4
A
'A'
Send 'A'
Send 'A'
Send 'A'
'A'
A
Acknowledge 'A'
Acknowledge 'A'
Acknowledge 'A'
Ack. 'A'
OK
'OK'
Send 'OK'
Send 'OK'
Send 'OK'
'OK'
OK
Acknowledge 'OK'
Acknowledge 'OK'
Acknowledge 'OK'
Ack. 'OK'


Server FC-4
Server FC-4
Switch 1 FC-2
Switch 2 FC-2
Storage FC-2
Storage FC-4
A
'A'
Send 'A'
Send 'A'
Fail
Fail 'A'
A
'A'
Send 'A'
Send 'A'
Send 'A'
'A'
A
Acknowledge 'A'
Acknowledge 'A'
Ack 'A'
Acknowledge 'A'

FC-2: класс 3


Server FC-4
Server FC-2
Switch 1 FC-2
Switch 2 FC-2
Storage FC-2
Storage FC-4
A
'A'
Send 'A'
Send 'A'
Send 'A'
'A'
A
OK
'OK'
Send 'OK'
Send 'OK'
Send 'OK'
'OK'
OK

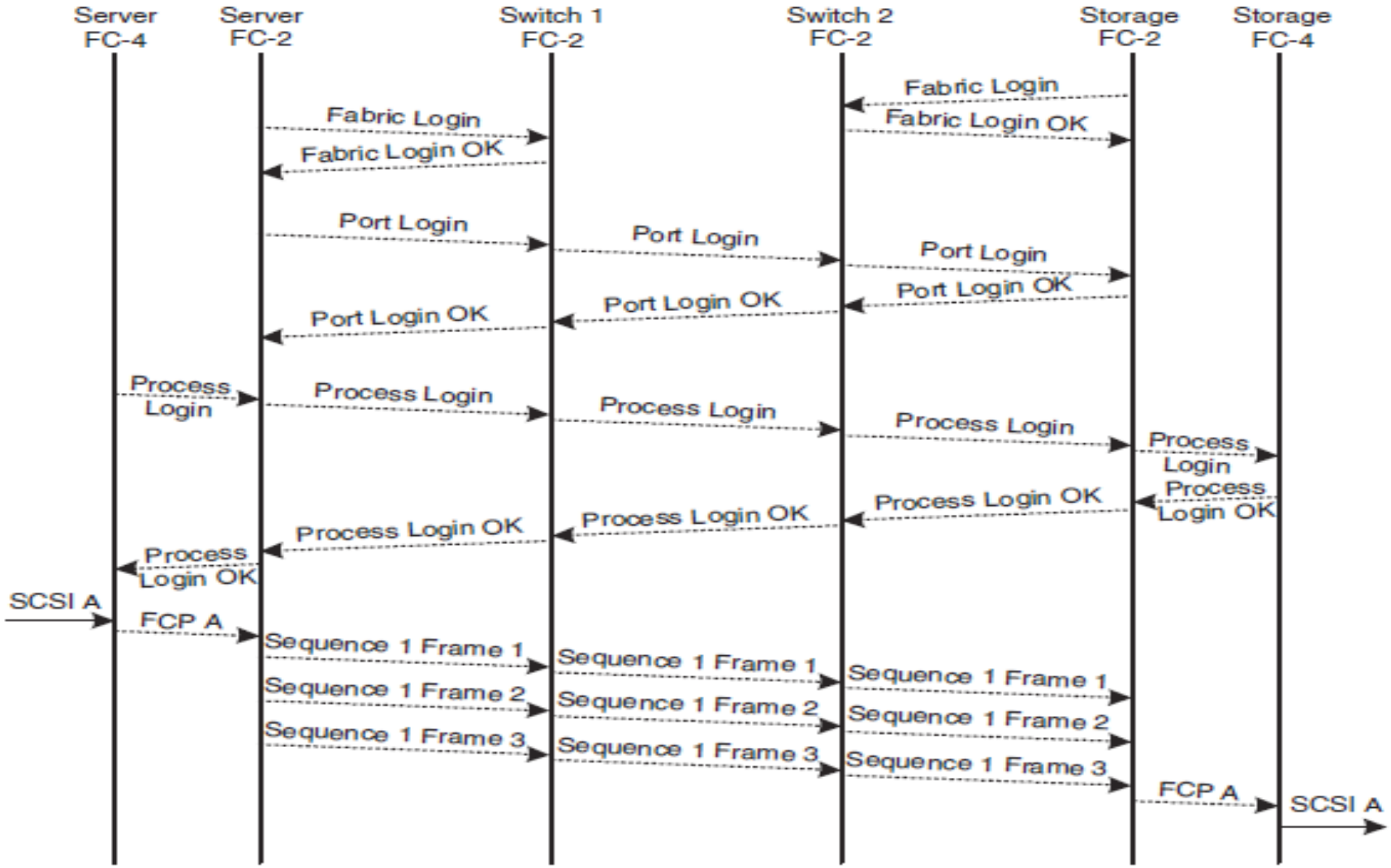

Server FC-4
Server FC-2
Switch 1 FC-2
Switch 2 FC-2
Storage FC-2
Storage FC-4
A
'A'
Send 'A'
Send 'A'
A
'A'
Send 'A'
Send 'A'
Send 'A'
'A'
A

FC-3: сервисы

- Распределение кадров по маршрутам между много портовыми устройствами для увеличения пропускной способности
- Формирование логических групп маршрутов для управления переполнения на маршруте или сбоя, чтобы не загружать верхние уровни стека.
- Компрессия передаваемых данных (на HBA)
- Шифрование данных (на HBA)
- Зеркалирование

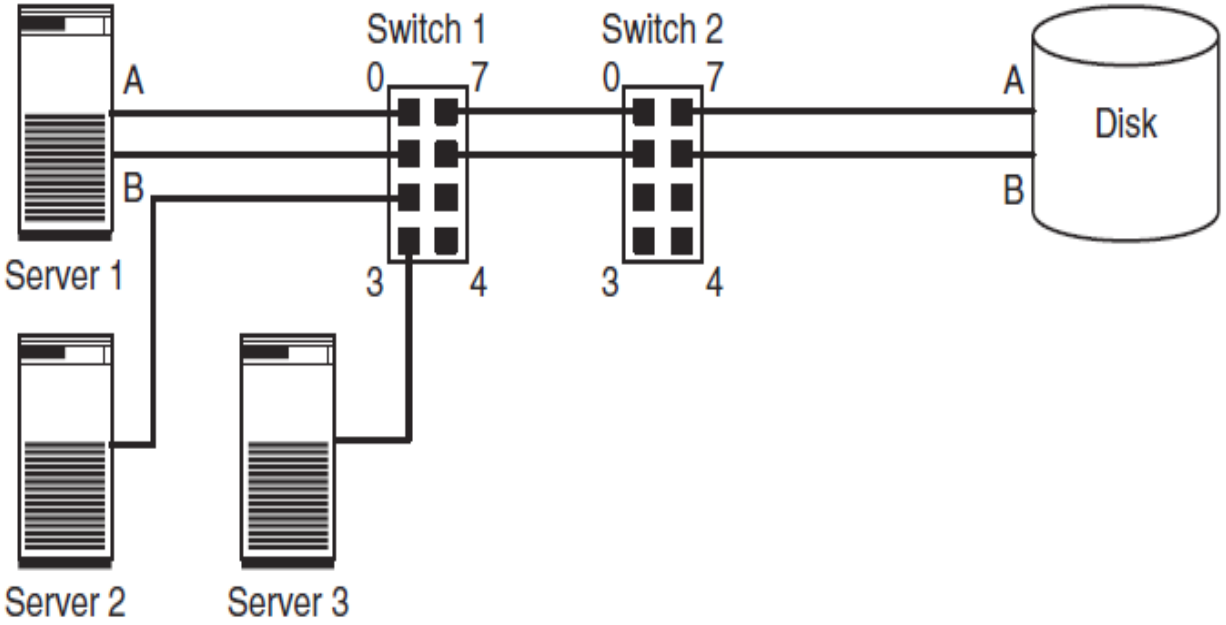
Пока это в планах

Службы линии: идентификация и адресация



Адресация

- Имена и адреса в FC
- У всех устройств FC сети есть 64 бит. имена
- WWN vs FCN
- WWN: WWPN. WWNN
- FLOG – 24 bit port address
- S_ID vs D_ID
- КсА 8 bit AL_PA (Arbitrated Loop Physical Address)



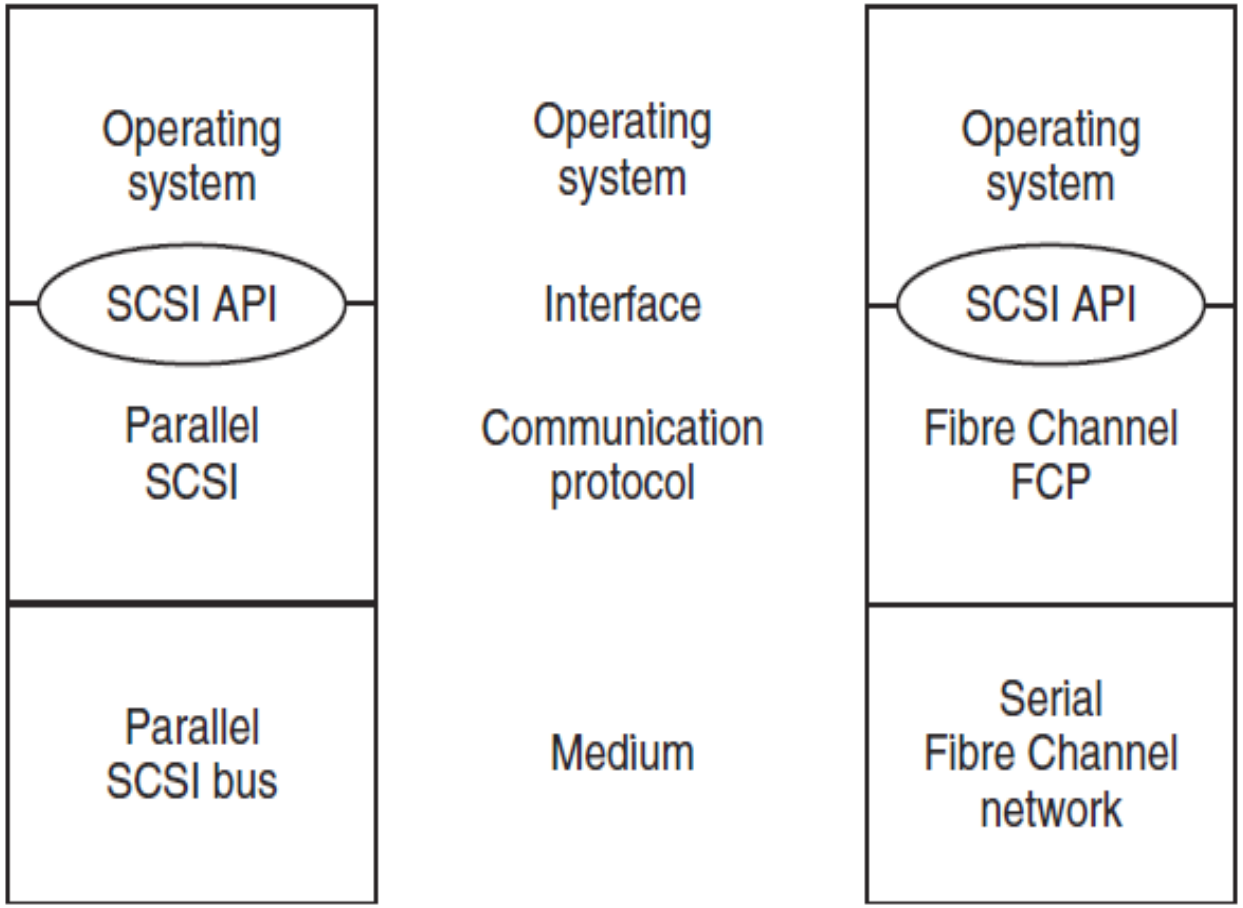
Port_ID	WWPN	WWNN	Device
010000	20000003 EAFE2C31	2100000C EAFE2C31	Server 1, Port A
010100	20000003 C10E8CC2	2100000C EAFE2C31	Server 1, Port B
010200	10000007 FE667122	10000007 FE667122	Server 2
010300	20000003 3CCD4431	2100000A EA331231	Server 3
020600	20000003 EAFE4C31	50000003 214CC4EF	Disk, Port B
020700	20000003 EAFE8C31	50000003 214CC4EF	Disk, Port A

Сервисы коммутационной среды

- КСС нужны для управления инфраструктурой и потоками в FC
- Все сервисы имеют реализуют определённые сервера, которые имеют строго определенные адреса.
- FLOG сервер отвечает за обработку всех входящих fabric login request
- За всеми изменениями в FC сети следит fabric controller
- Name server – отвечает за БД все имен N_Port’ов

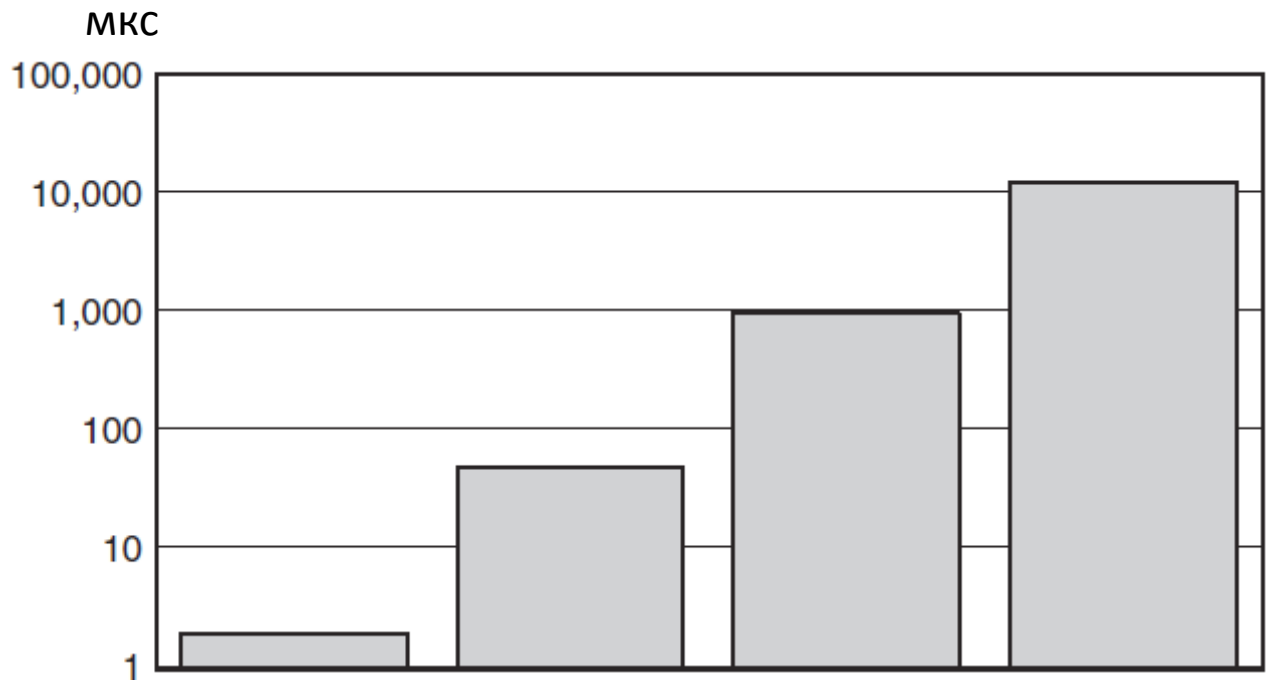
Address	Description
0×FF FF FF	Broadcast addresses
0×FF FF FE	Fabric Login Server
0×FF FF FD	Fabric Controller
0×FF FF FC	Name Server
0×FF FF FB	Time Server
0×FF FF FA	Management Server
0×FF FF F9	Quality of Service Facilitator
0×FF FF F8	Alias Server
0×FF FF F7	Security Key Distribution Server
0×FF FF F6	Clock Synchronisation Server
0×FF FF F5	Multicast Server
0×FF FF F4	Reserved
0×FF FF F3	Reserved
0×FF FF F2	Reserved
0×FF FF F1	Reserved
0×FF FF F0	Reserved

FC-4: ULP


Operating system
SCSI API
Parallel SCSI
Parallel SCSI bus
Operating system
Interface
Communication protocol
Medium
Operating system
SCSI API
Fibre Channel FCP
Serial Fibre Channel network

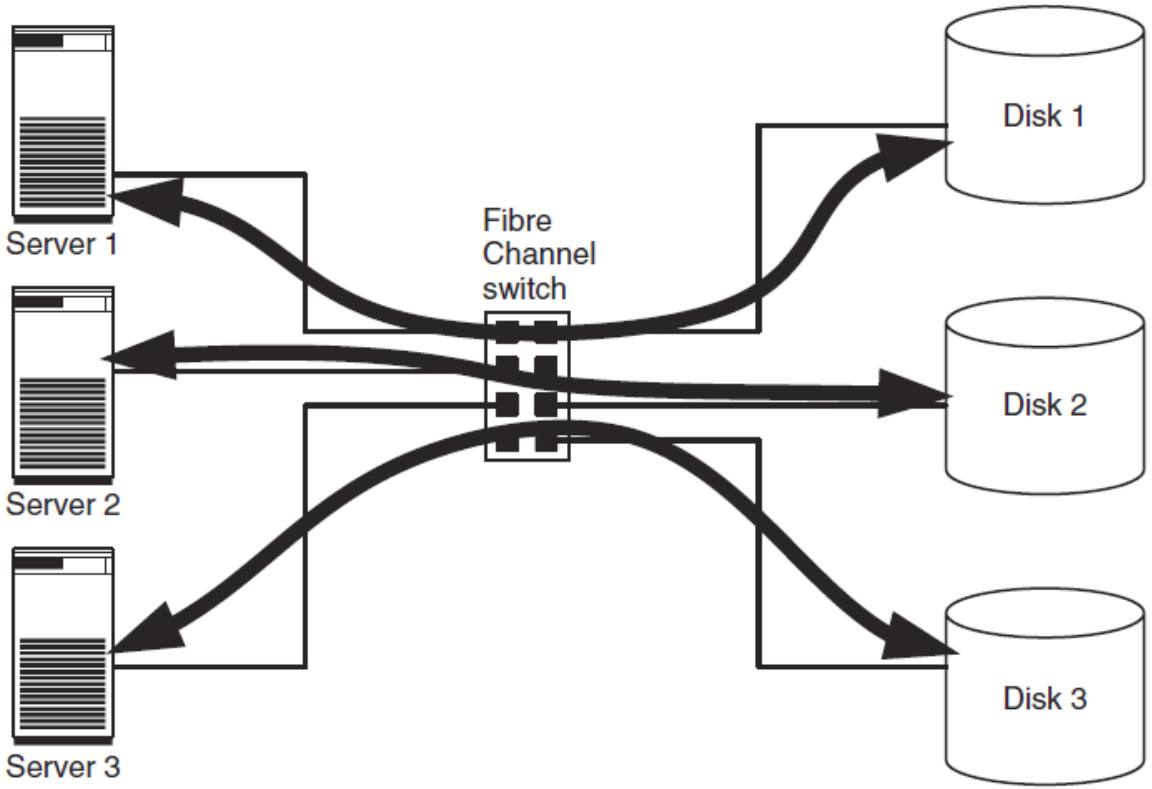
Fibre Channel SAN

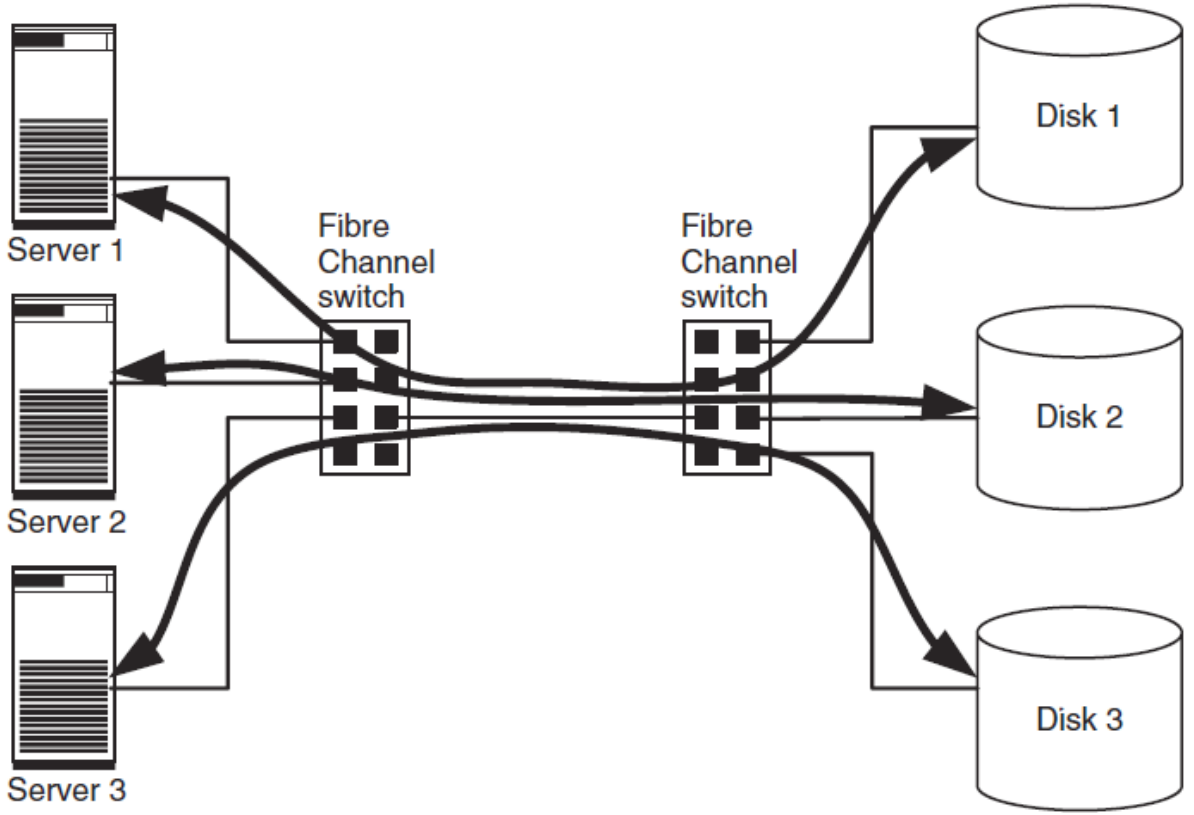
- P2P – FC до 10 км, SCSI не более 25 метров
- SCSI – медь, FC - оптика



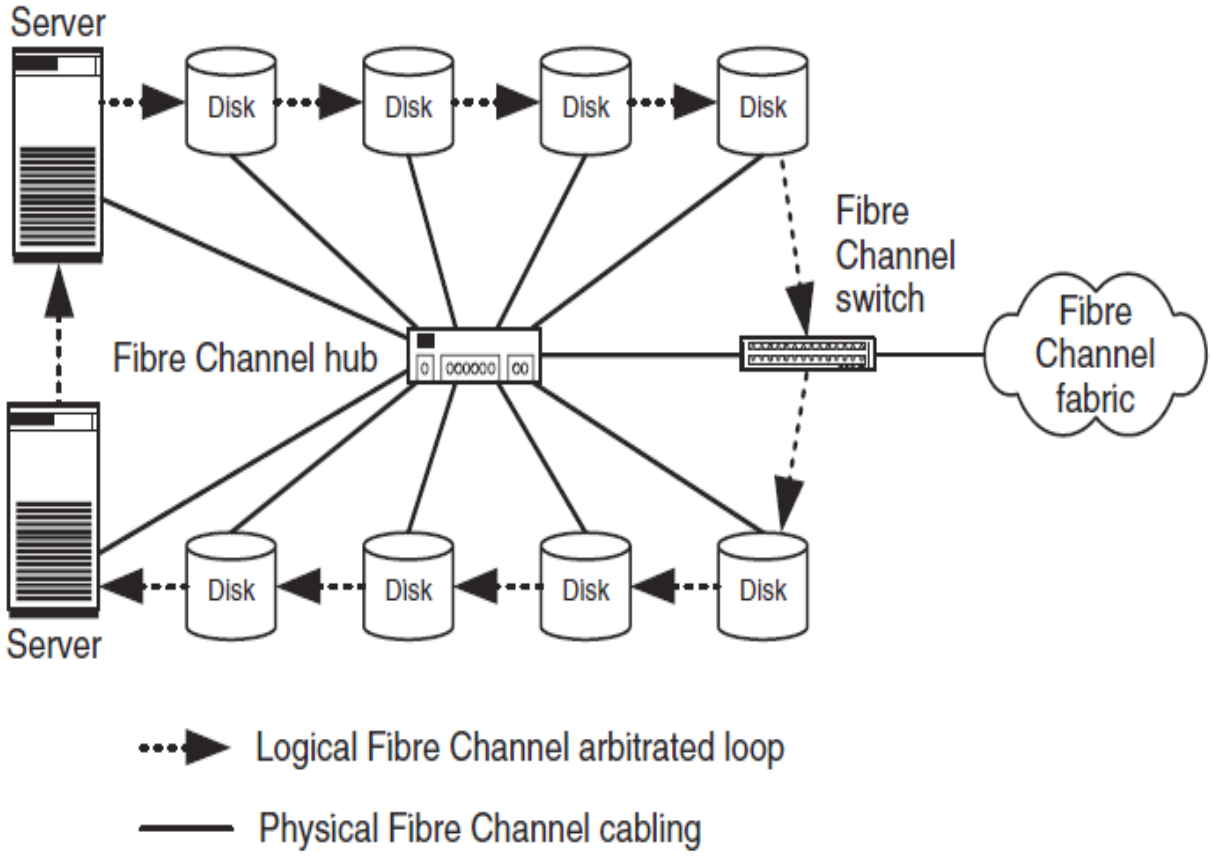
Коммутатор 10 км линия ДПС –промах ДПС попал

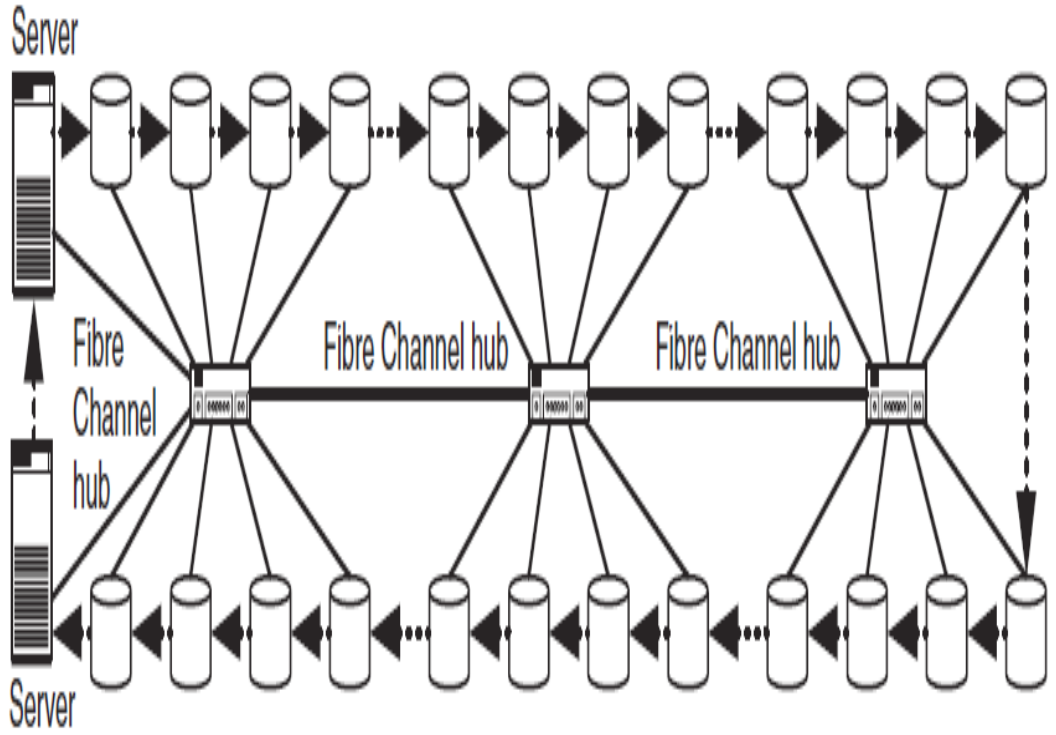
Топология коммутационной среды (fabric)



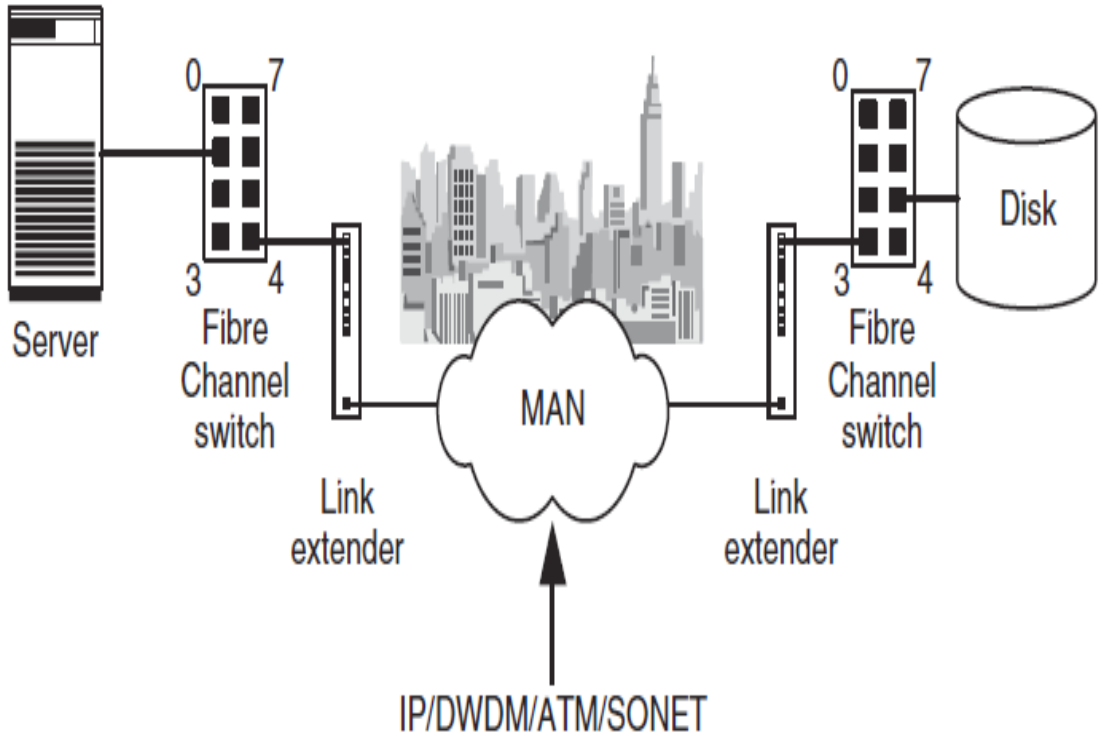


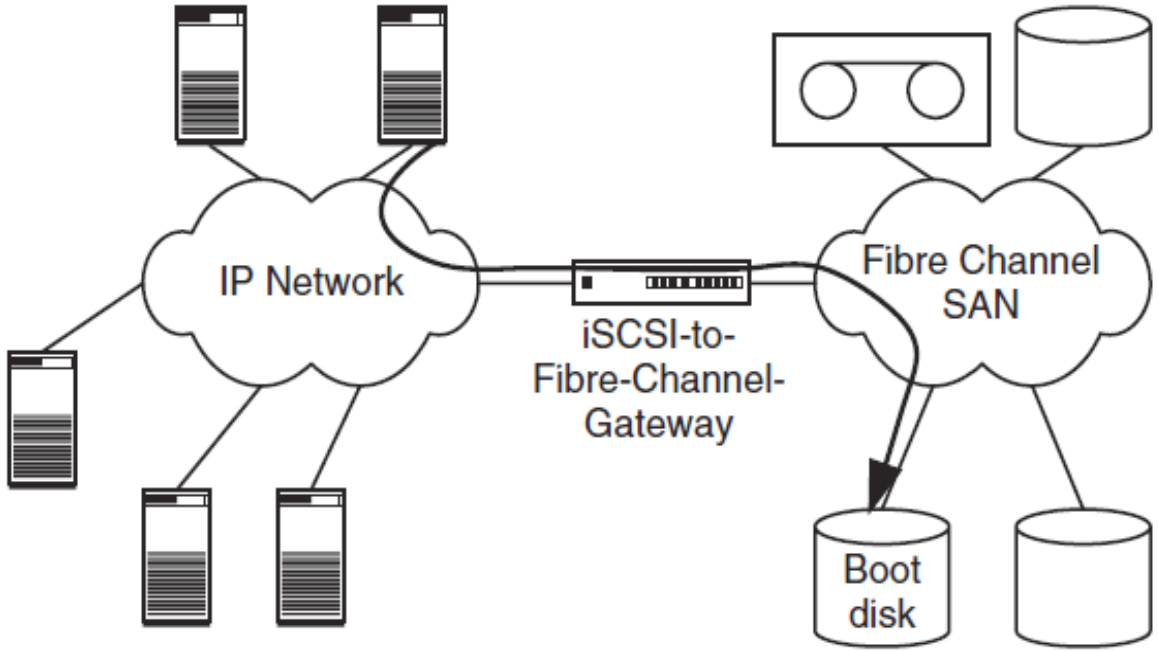
КсА топология



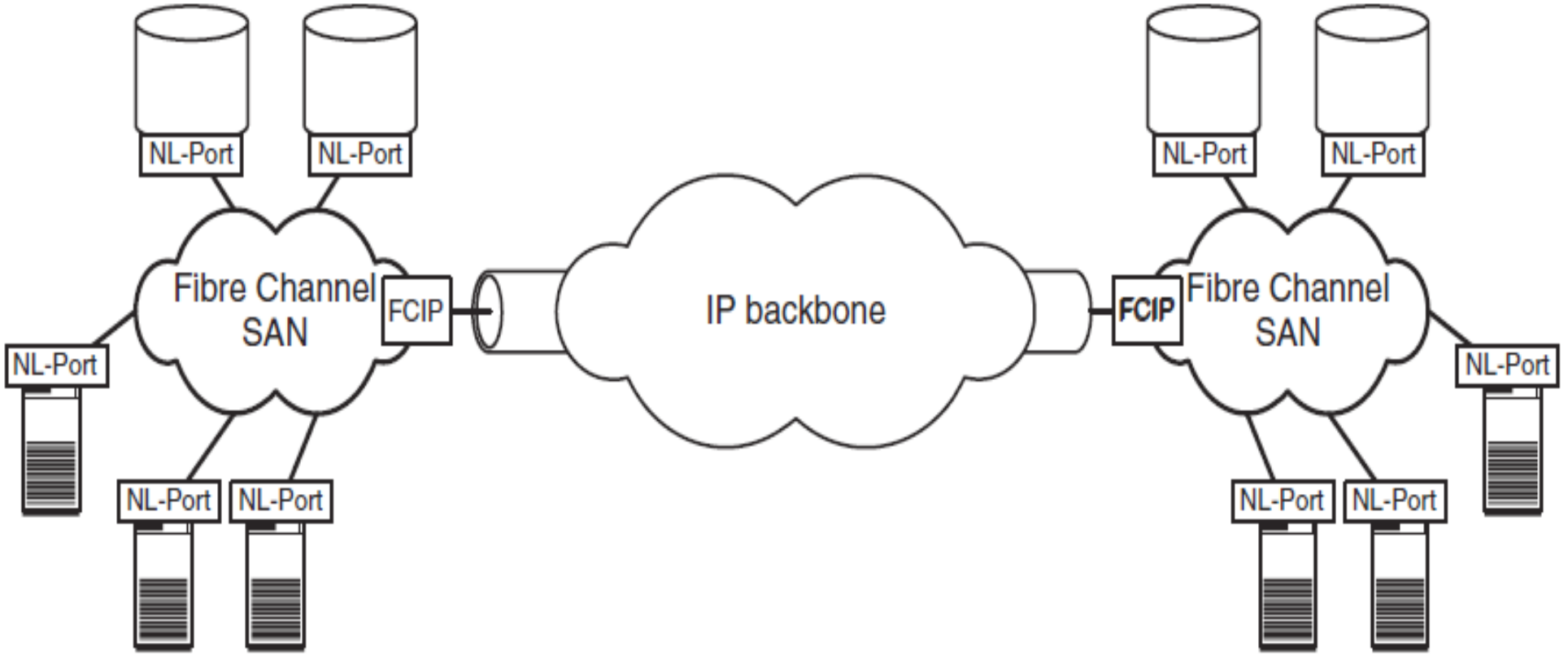


Metro SAN и IP_FC SAN





Соединение FC_SAN через TCP/IP магистраль

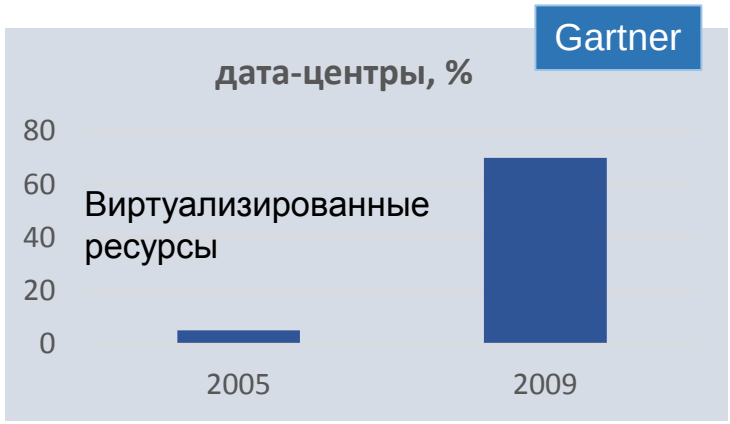


В качестве заключения: новые горизонты

Тенденции развития рынка информационных технологий

Мобильность

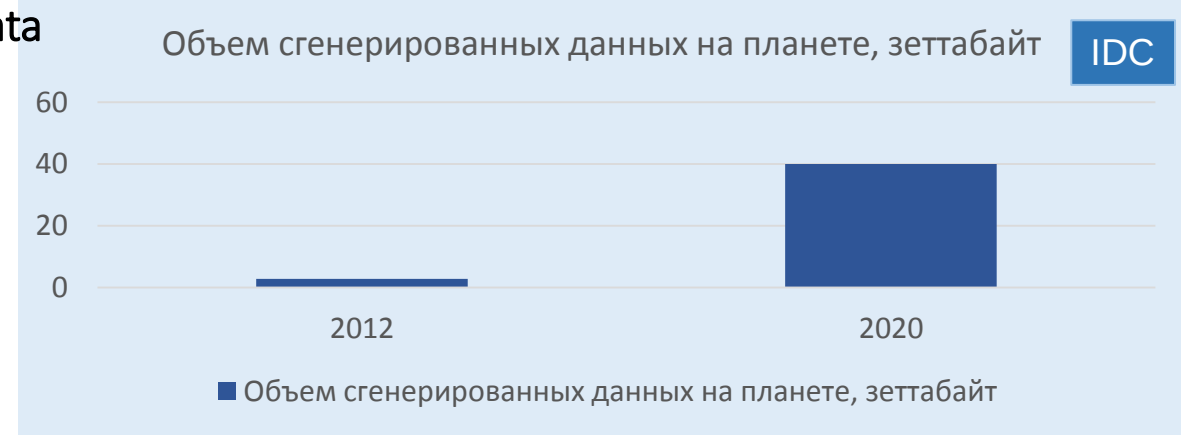
Виртуализация



Консолидация инфраструктуры

Тенденции развития рынка информационных технологий

Big Data



Всего с начала 2010 г. объем хранимых данных вырос в 50 раз

Центры обработки данных

- В 2014 году объем мирового рынка колокации в ЦОДах составил $22,8 млрд. Общая площадь размещения оборудования достигла 10,13 кв. км. (451 Research)
- Общее число дата-центров всех типов в 2017 г. вырастет до 8,6 млн (IDC)

Телеком

- Каждый из пользователей глобальной сети генерирует больше трафика, чем вся Всемирная паутина 30 лет назад
- В 2014 году интернет-трафик вырос, по сравнению с 1984 годом, в 2,7 миллиардов раз (Cisco)

Изменение бизнес-модели

РЕНТАБЕЛЬНОСТЬ БИЗНЕСА

Промышленное производство
Энергодобыча
Тяжелая и легкая промышленность

РЕНТАБЕЛЬНОСТЬ БИЗНЕСА

Оператор связи
Интернет провайдер

Доступ к «транспорту» должен быть бесплатным, платным должен стать контент и услуги.

Информация о пользователе:

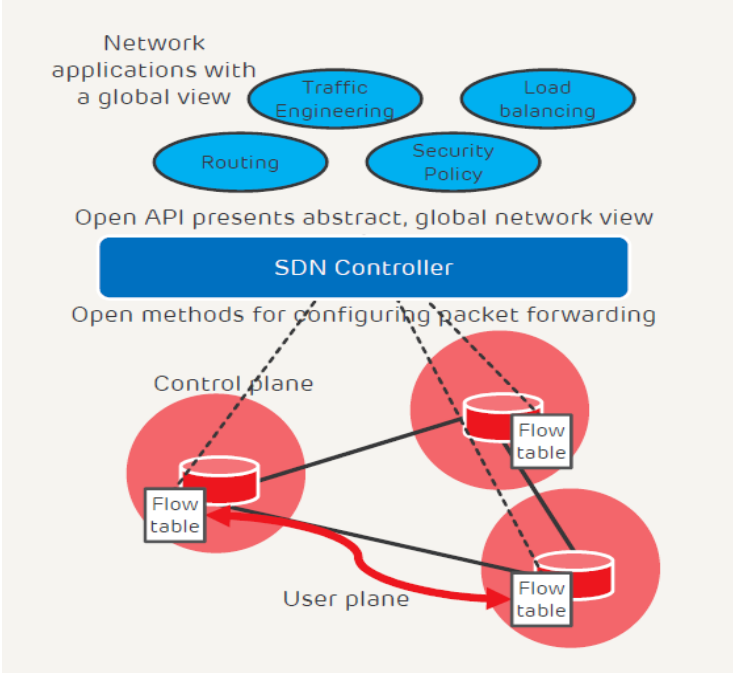
- Использование: посещаемые сайты, звонки и сообщения (включая тип сообщений и их частоту);
- География: где находится мобильное устройство в конкретный момент (уровень точности может разниться от района к району);
- Демография: доход домохозяйства, число и возраст проживающих детей;
- Уровень дохода: тарифный план, история платежей, паттерн совершения покупок;
- Мультиплатформенность: использование данных на разных устройствах и типах подключения к сети (3G, WiFi и т.п.).

- 2011 - AT&T – запуск подразделения AdWorkds: поддержка целевой рекламы в web, мобильной среде и ТВ.
- 2013 – AdWorks открывает доступ к анализу данных 70 млн.пользователей.
- 2012 – Verizon - запуск инициативы Precision Market Insights – доступ к мобильным данным пользователей для маркетинговых и рекламных компаний.

SDN и NFV : схожести и различия

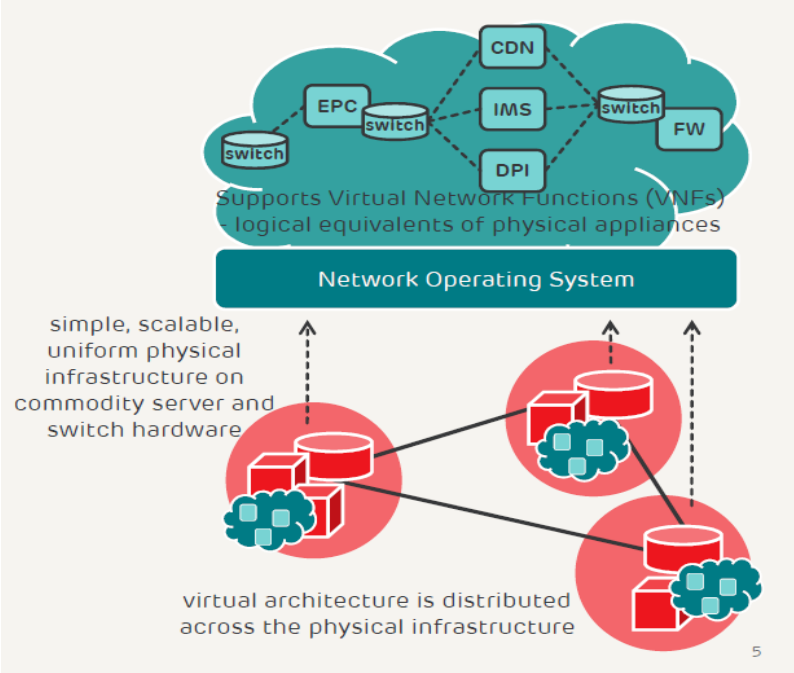
с 2007 года

ИТ → SDN



с 2012 года

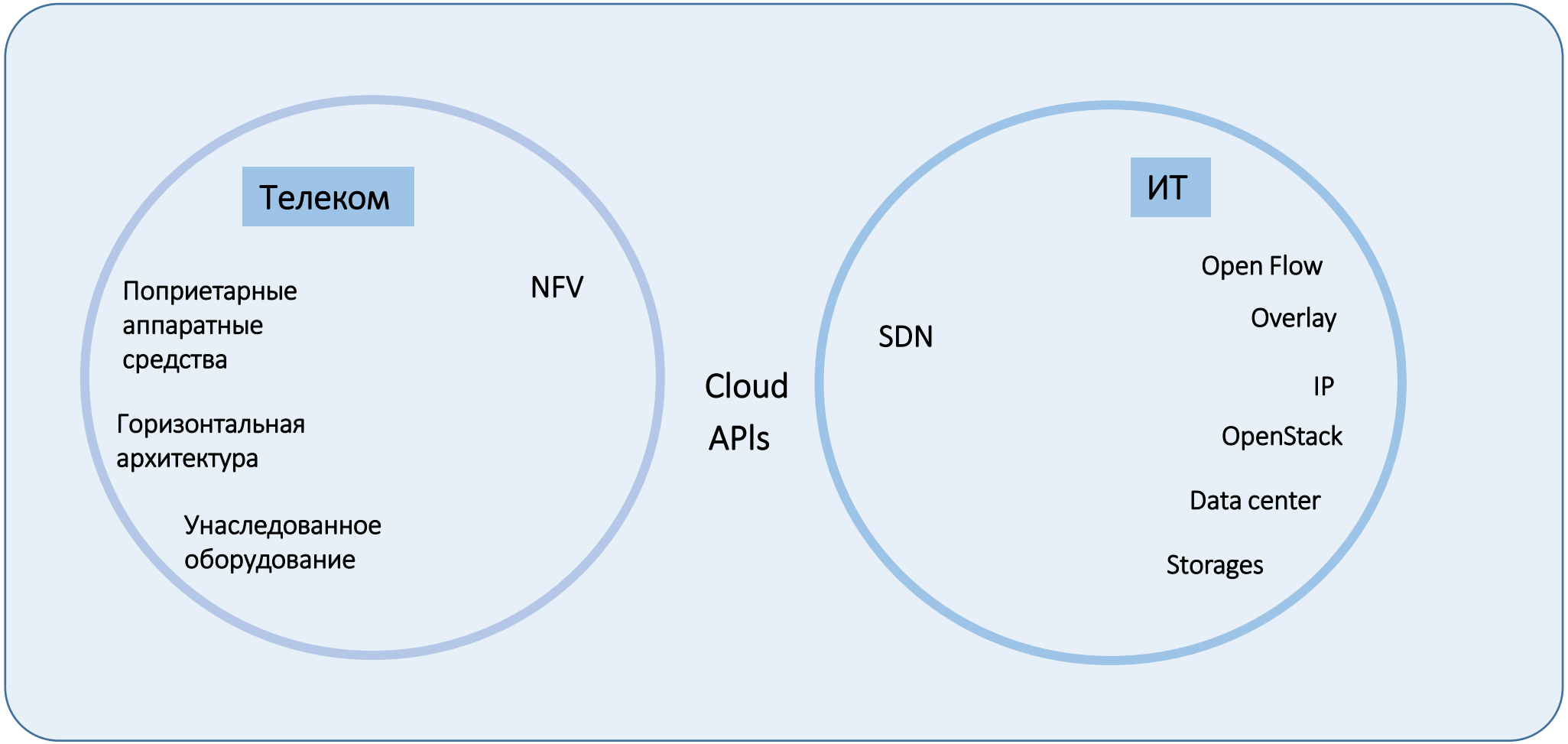
Телеком → NFV



IBM, HUAWEI, NEC, hp, CISCO, BROCADE, extreme networks, Ethernet Everywhere, big switch networks, JUNIPER NETWORKS

NTT, at&t, Deutsche Telekom, AllOt communications, CONTEXTREAM, Sonus NETWORKS, Metaswitch Networks, nuagenetworks

Симбиоз SDN и NFV



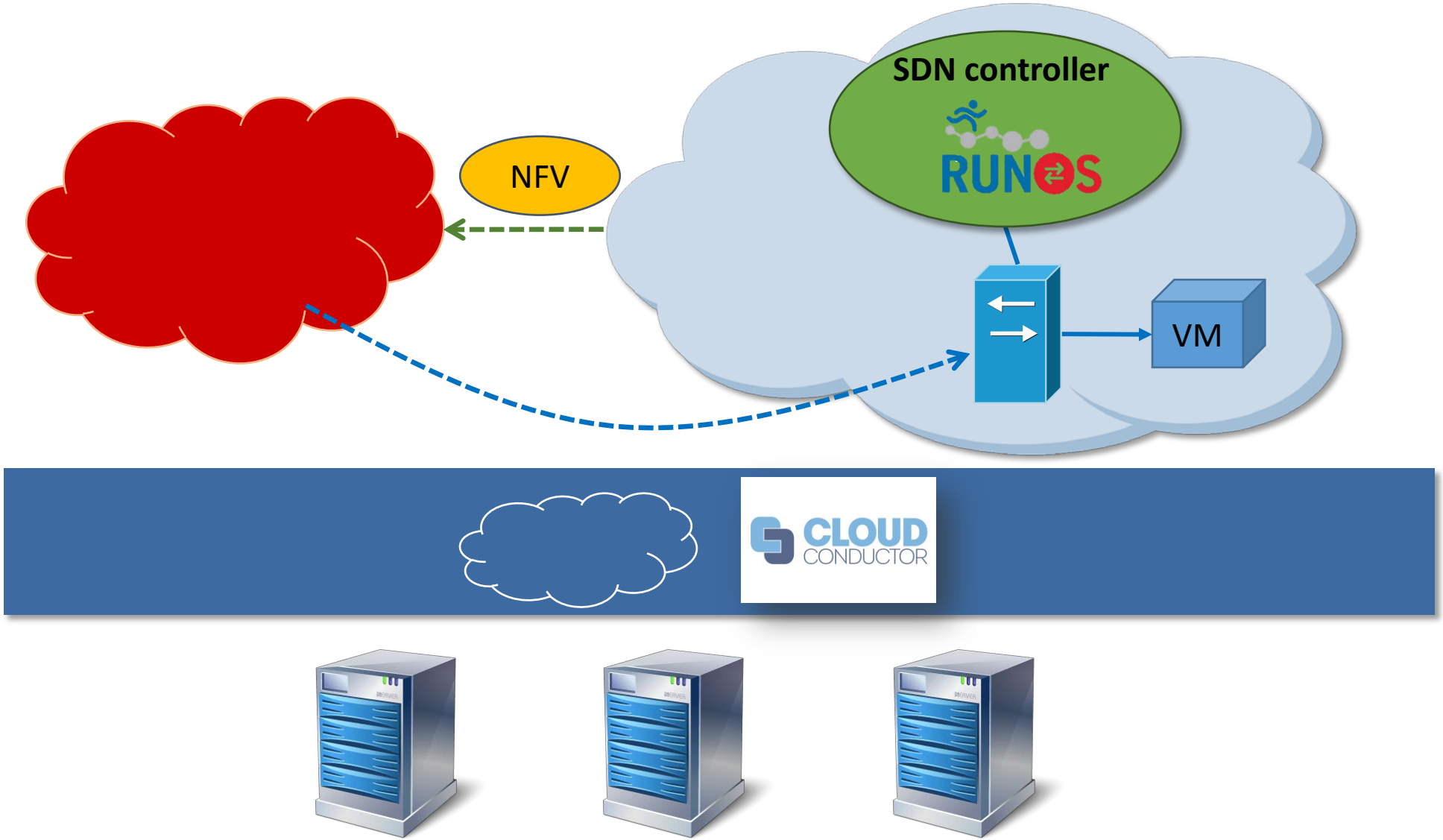
1 NFV_SDN

2 SDN_NFV

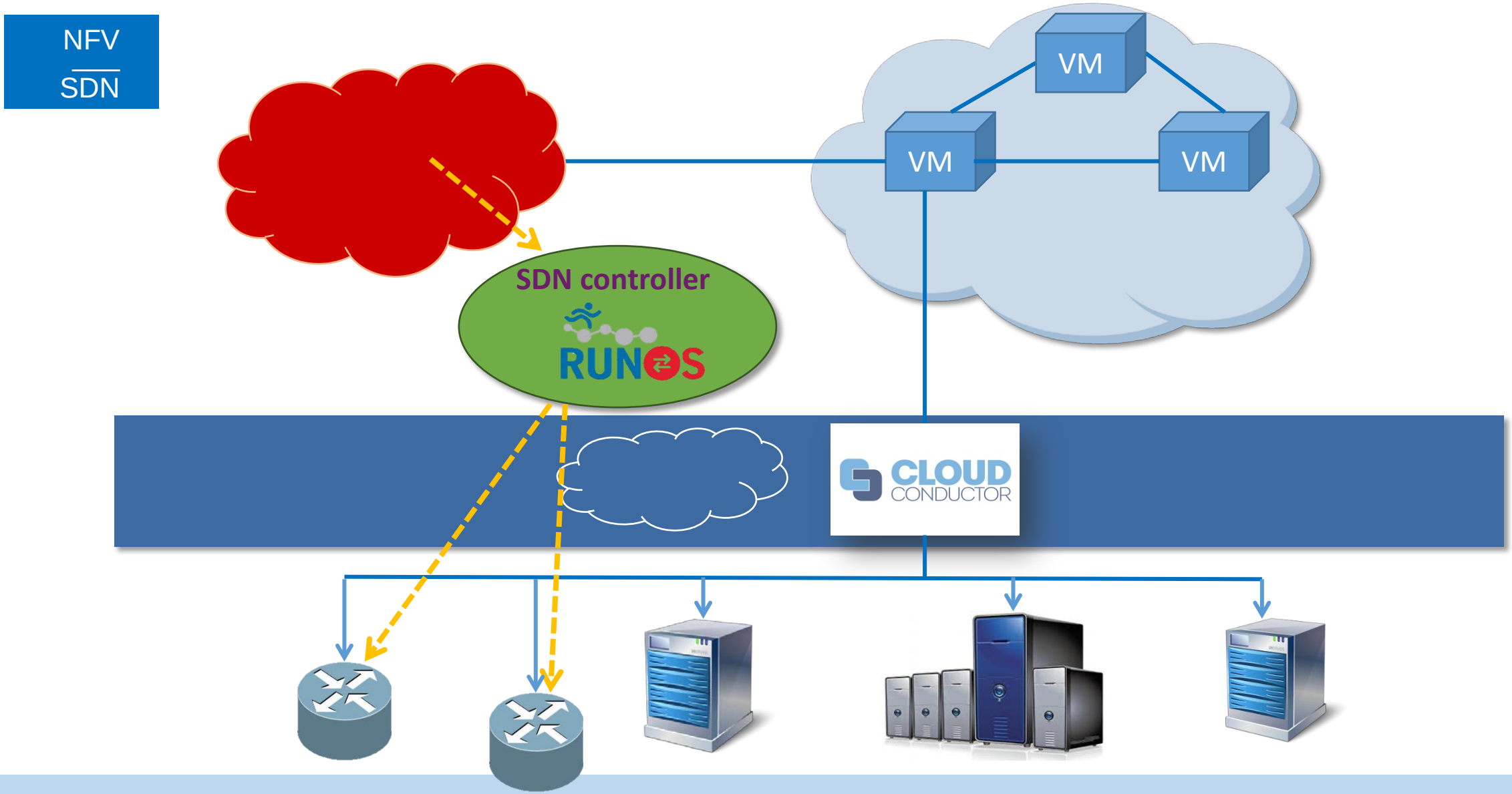
3 SDN NFV

Пример 1 сценария

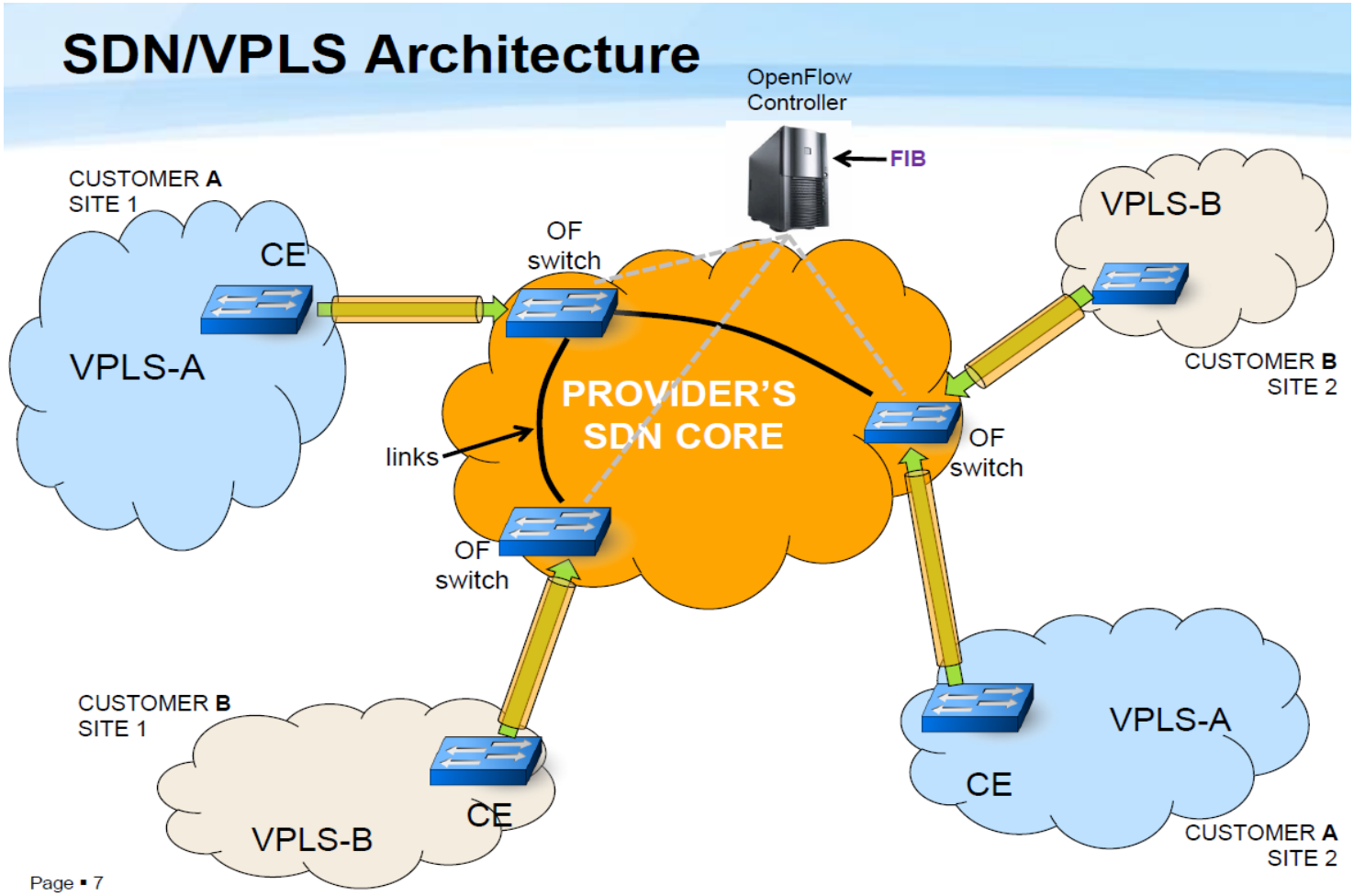
SDN
NFV



Пример 2 сценария

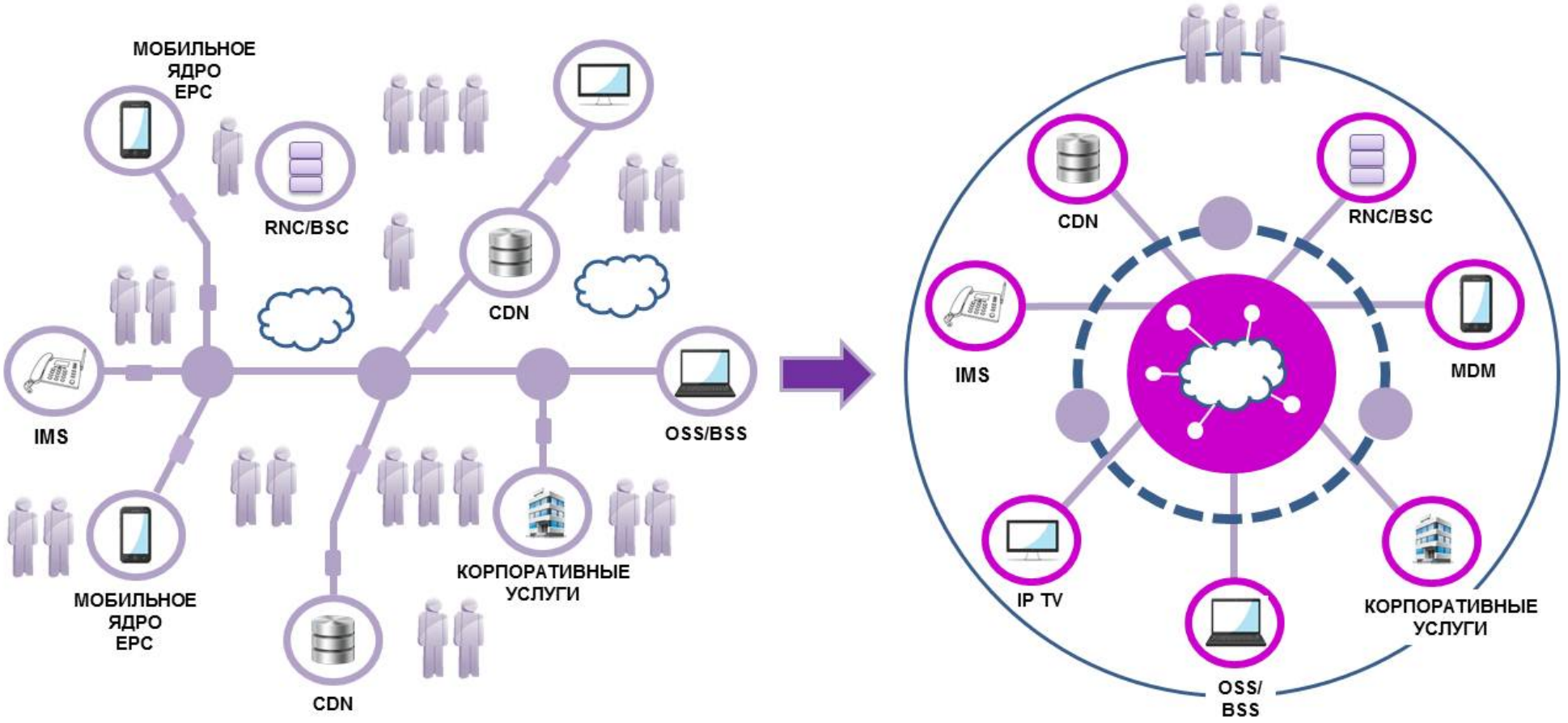


SDN & VPN



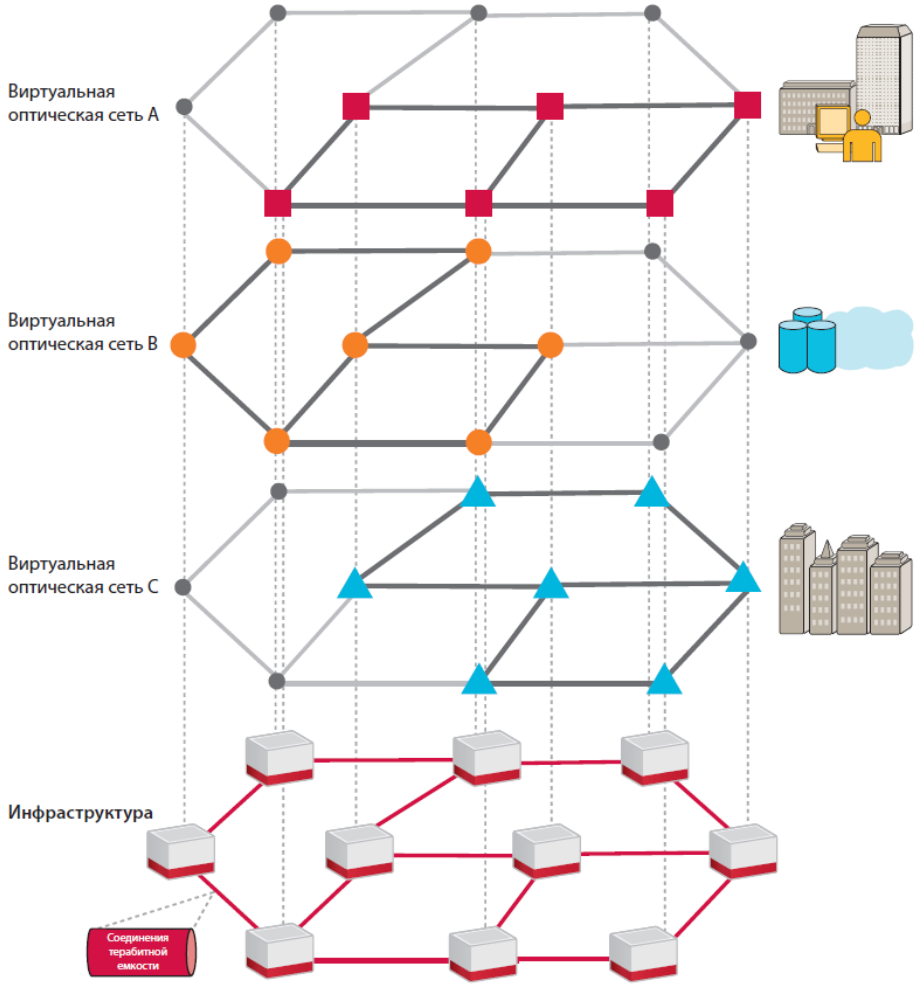
Источник: S.Konstantares, G.Thesolonikets Software Defined VPNs. University of Amsterdam, 2014

Пример 3 сценария **SDN и NFV**

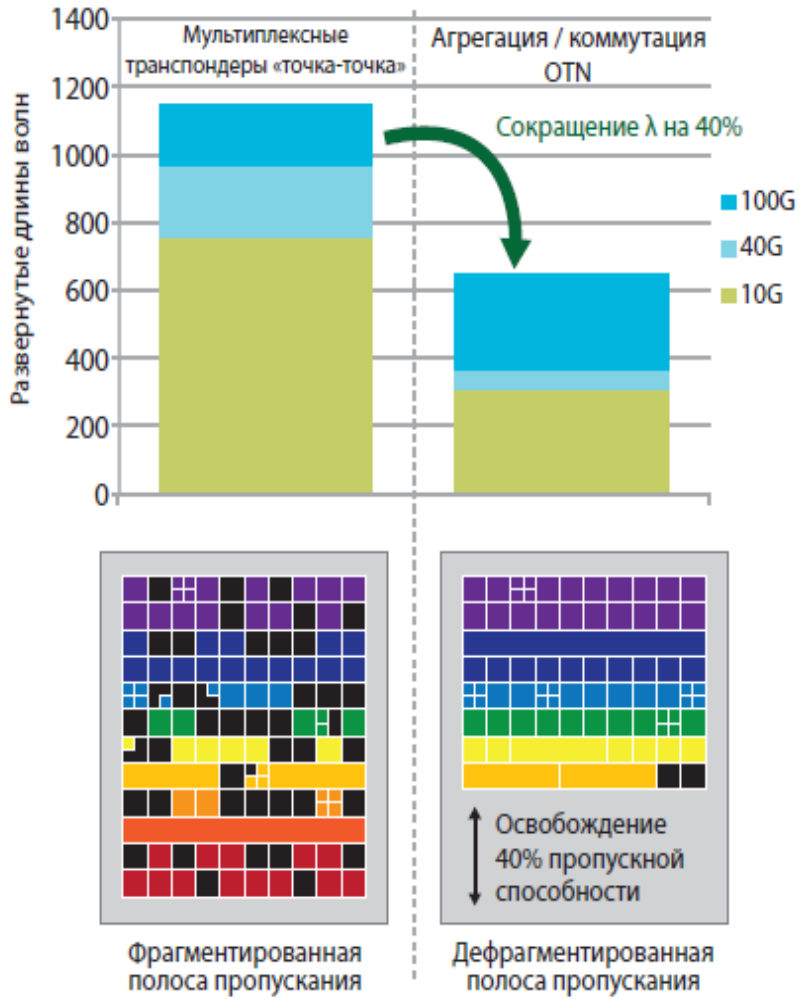


Optical Transport Networks

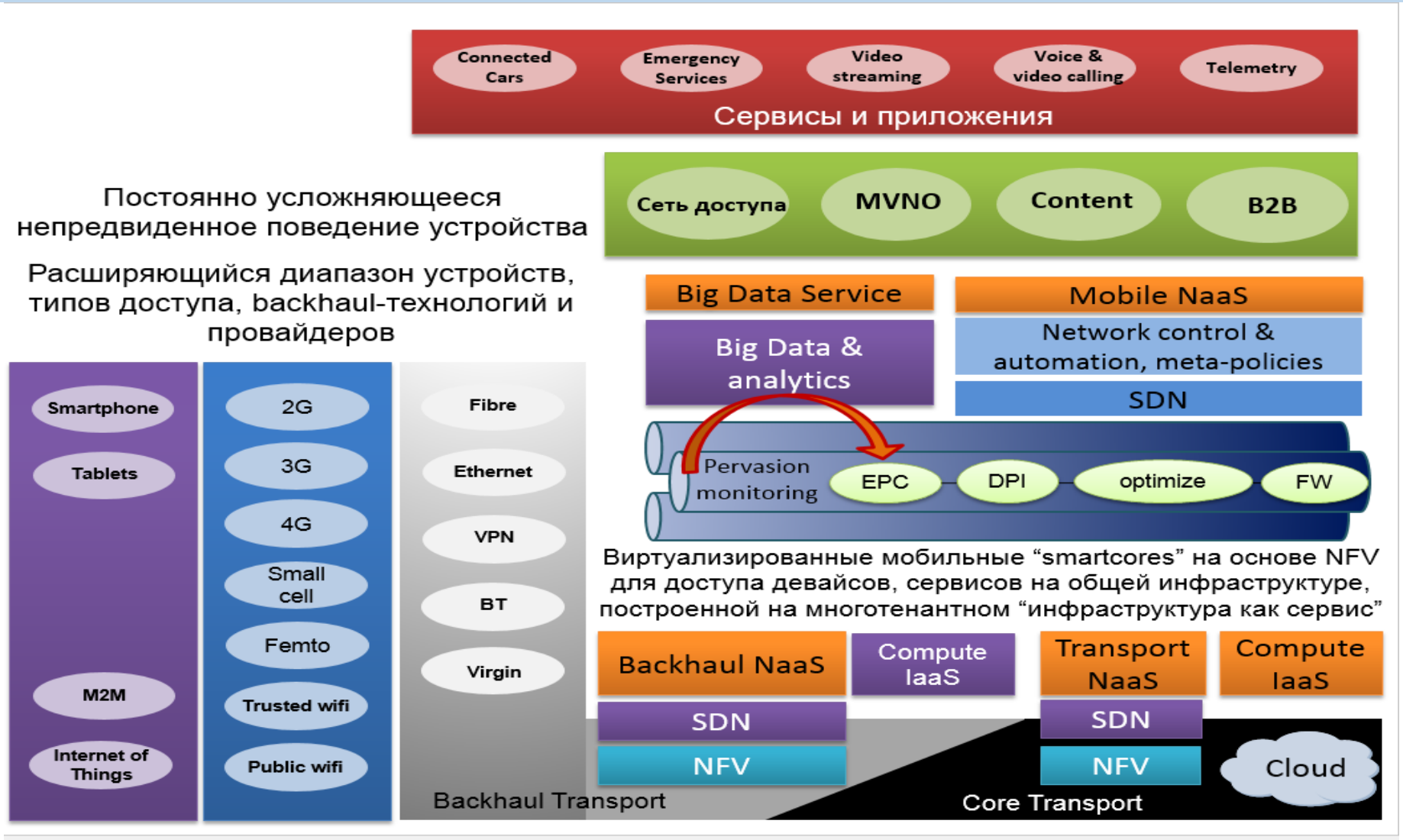
Виртуальные сети на база OTN



Дефрагментирование загрузки линий



Use-Case: **Mobile SDN**



Источник: Revolutionisingmobile networks with SDN and NFV / Cambridge Wireless Virtual Networks SIG, Philip Bridge, Senior Network Architect at EE, 2014

Use-case: Transport – SDN

Источник: Telefonaktiebolaget LM Ericsson 2015 | MPLS SDN World Congress & NFV/SDN Summit 2015, Paris

Use-case: **SDN-безопасность**

Источник: Gartner (2015)

SDN - безопасность

```
-------------------------------------------------
 LINK INFO FROM DB : Count = 1
-------------------------------------------------
 LINK_TABLE_NAME      = controller_link
 LINK_ID              = 00:00:00:00:00:00:00:01-2-00:00:00:00:00:00:00:02-2
 LINK_SRC_SWITCH      = 00:00:00:00:00:00:00:01
 LINK_SRC_PORT        = 2
 LINK_SRC_PORT_STATE  = 0
 LINK_DST_SWITCH      = 00:00:00:00:00:00:00:02
 LINK_DST_PORT        = 2
 LINK_DST_PORT_STATE  = 0
 LINK_VALID_TIME      = 1390964347029
 LINK_TYPE            = internal
-------------------------------------------------

-------------------------------------------------
 LINK INFO FROM DB : Count = 2
-------------------------------------------------
 LINK_TABLE_NAME      = controller_link
 LINK_ID              = 00:00:00:00:00:00:00:02-2-00:00:00:00:00:00:00:01-2
 LINK_SRC_SWITCH      = 00:00:00:00:00:00:00:02
 LINK_SRC_PORT        = 2
 LINK_SRC_PORT_STATE  = 0
 LINK_DST_SWITCH      = 00:00:00:00:00:00:00:01
 LINK_DST_PORT        = 2
 LINK_DST_PORT_STATE  = 0
 LINK_VALID_TIME      = 1390964347026
 LINK_TYPE            = internal
-------------------------------------------------
```

Link 1

Link 2

```
ow - [ATTACK]-------------------------------------------------
ow - [ATTACK] LINK INFO FROM DB : Count = 1
ow - [ATTACK]-------------------------------------------------
ow - [ATTACK] LINK_TABLE_NAME      = controller_link
ow - [ATTACK] LINK_ID              = 00:00:00:00:00:00:00:02-2-00:00:00:00:00:00:00:01-2
ow - [ATTACK] LINK_SRC_SWITCH      = 00:00:00:00:00:00:00:02
ow - [ATTACK] LINK_SRC_PORT        = 2
ow - [ATTACK] LINK_SRC_PORT_STATE  = 0
ow - [ATTACK] LINK_DST_SWITCH      = 00:00:00:00:00:00:00:01
ow - [ATTACK] LINK_DST_PORT        = 2
ow - [ATTACK] LINK_DST_PORT_STATE  = 0
ow - [ATTACK] LINK_VALID_TIME      = 1390964347026
ow - [ATTACK] LINK_TYPE            = internal
ow - [ATTACK]-------------------------------------------------
ow -
 [ATTACK] Access InternalDB : delete Link Information
```

Link 2 Only
Link 1 has been deleted

```
2014-05-12 09:26:33.219 PDT [Statistics Collector] DEBUG o.o.c.p.o.i.InventoryServiceShim - Connection service
accepted the inventory notification for OF|00:00:00:00:00:00:00:02 CHANGED
2014-05-12 09:26:34.217 PDT [Statistics Collector] DEBUG o.o.c.c.internal.ConnectionManager - updateNode: OF|00
:00:00:00:00:00:00:03 type CHANGED props [Description[None]]
2014-05-12 09:26:34.217 PDT [Statistics Collector] DEBUG o.o.c.s.internal.SwitchManager - updateNode: OF|00:00:
00:00:00:00:00:03 type CHANGED props [Description[None]] for container default
2014-05-12 09:26:34.217 PDT [Statistics Collector] DEBUG o.o.c.p.o.i.InventoryServiceShim - Connection service
accepted the inventory notification for OF|00:00:00:00:00:00:00:03 CHANGED
2014-05-12 09:26:51.791 PDT [SwitchEvent Thread] DEBUG o.o.c.h.internal.HostTracker - Received for Host: IP 10.
0.0.1, MAC 000000000001, HostNodeConnector [nodeConnector=OF|1@OF|00:00:00:00:00:00:00:01, vlan=0, staticHost=f
alse, arpSendCountDown=0]
2014-05-12 09:26:51.794 PDT [Thread-37] DEBUG o.o.c.h.internal.HostTracker - New Host Learned: MAC: 00000000000
1  IP: 10.0.0.1
2014-05-12 09:26:51.794 PDT [Thread-37] DEBUG o.o.c.h.internal.HostTracker - Notifying Applications for Host 10
.0.0.1 Being Added
2014-05-12 09:26:51.795 PDT [Thread-37] DEBUG o.o.c.h.internal.HostTracker - Notifying Topology Manager for Hos
t 10.0.0.1 Being Added
2014-05-12 09:26:51.796 PDT [SwitchEvent Thread] INFO  o.o.controller.attack.crash.Crash - [ATTACK.CRASH] Packe
t Received
2014-05-12 09:26:51.796 PDT [SwitchEvent Thread] INFO  o.o.controller.attack.crash.Crash - [ATTACK.CRASH] Syste
m.exit() called
2014-05-12 09:26:51.798 PDT [Listener:59957] DEBUG com.arjuna.ats.arjuna - Recovery listener existing com.arjun
a.ats.arjuna.recovery.ActionStatusService
2014-05-12 09:26:51.798 PDT [Thread-11] DEBUG org.jgroups.stack.GossipRouter - ConnectionHandler[peer: /127.0.0
.1, logical_addrs: localhost-12306] is being closed
2014-05-12 09:26:51.805 PDT [Thread-11] DEBUG org.jgroups.stack.GossipRouter - router stopped
$
```

Call a System Exit Function

OpenDayLight has been crashed

```
in] DEBUG n.f.core.internal.Controller - OFListeners for PACKET_IN: net.floodlightcontroller.attack
in] INFO  n.f.core.internal.Controller - Listening for switch connections on 0.0.0.0/0.0.0.0:6633
w I/O server worker #1-1] INFO  n.f.core.internal.Controller - New switch connection from /127.0.0.1
w I/O server worker #1-2] INFO  n.f.core.internal.Controller - New switch connection from /127.0.0.1
w I/O server worker #1-1] DEBUG n.f.core.internal.Controller - This controller's role is null, not s
w I/O server worker #1-2] DEBUG n.f.core.internal.Controller - This controller's role is null, not s
w I/O server worker #1-2] INFO  n.floodlightcontroller.attack.Crash - [ATTACK] Crash Application
~/floodlight-0.90#
```

App calls the System.exit function

```
19:52:44.229 [New I/O server worker #1-1] INFO  n.f.attack.MemoryLeak - [ATTACK] MemoryLeak Application
19:52:44.361 [New I/O server worker #1-1] ERROR n.f.core.internal.Controller - Error while processing me
java.lang.OutOfMemoryError: Java heap space
        at net.floodlightcontroller.attack.MemoryLeak.receive(MemoryLeak.java:59) ~[floodlight.jar:na]
        at net.floodlightcontroller.core.internal.Controller.handleMessage(Controller.java:1285) ~[flood
        at net.floodlightcontroller.core.internal.Controller$OFChannelHandler.processOFMessage(Controlle
        at net.floodlightcontroller.core.internal.Controller$OFChannelHandler.messageReceived(Controller
        at org.jboss.netty.handler.timeout.IdleStateAwareChannelUpstreamHandler.handleUpstream(IdleState
        at org.jboss.netty.handler.timeout.ReadTimeoutHandler.messageReceived(ReadTimeoutHandler.java:18
```

FloodLight - Out of Memory Error

```
def handle_PacketIn(event):
   packet = event.parsed
   inport = event.port
   .....
def launch():
  core.openflow.addListenerByName("PacketIn", handle_PacketIn)
  print '[ATTACK] Crash Application'
  sys.exit(0)
```

```
openflow@openflowtutorial:~/pox$ ./pox.py monitoring crash
POX 0.0.0 / Copyright 2011 James McCauley
[ATTACK] Crash Application
openflow@openflowtutorial:~/pox$ _
```

Crash App Kills monitoring App and POX

```
osgi> 18:22:24.972 [SpringOsgiExtenderThread-4] TRACE n.b.learningswitch.LearningSwitch - Starting
18:22:28.999 [pool-2-thread-1] TRACE n.b.learningswitch.LearningSwitch - [ATTACK] Crash Applicatio
mininet@mininet-vm:
```

App calls the System.exit function

```
18:14:16.526 [pool-2-thread-1] TRACE n.b.learningswitch.LearningSwitch - [ATTACK] Memory Leak Application
18:14:16.549 [pool-2-thread-1] TRACE n.b.learningswitch.LearningSwitch - [ATTACK] allocated mem_size: 104857600 tot
18:14:16.550 [pool-2-thread-1] TRACE n.b.learningswitch.LearningSwitch - [ATTACK] Memory Leak Application
18:14:28.536 [pool-2-thread-1] TRACE n.b.learningswitch.LearningSwitch - [ATTACK] allocated mem_size: 104857600 tot
18:14:28.537 [pool-2-thread-1] TRACE n.b.learningswitch.LearningSwitch - [ATTACK] Memory Leak Application
Exception in thread "pool-2-thread-1" java.lang.OutOfMemoryError: Java heap space
        at net.beaconcontroller.learningswitch.LearningSwitch.receive(LearningSwitch.java:78)
        at net.beaconcontroller.core.internal.Controller.handleMessages(Controller.java:387)
        at net.beaconcontroller.core.internal.Controller.handleSwitchEvent(Controller.java:199)
        at net.beaconcontroller.core.internal.Controller.handleEvent(Controller.java:138)
        at net.beaconcontroller.core.io.internal.IOLoop.doLoop(IOLoop.java:122)
        at net.beaconcontroller.core.internal.Controller$2.run(Controller.java:541)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1145)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:615)
        at java.lang.Thread.run(Thread.java:724)
```

Java Out of Memory Error

Источник: KAIST Atto Research Korea, 2014

SDN - безопасность

Устройства OF сети – Data Plane

- Уязвимости программного обеспечения
  a. неустойчивость кода к внешним воздействиям
  b. код с уязвимостями
- Атаки с использованием вредоносного кода
- DDoS атаки
- Атака сетевых устройств изнутри сети
- Вредоносные устройства в OF-сети

Каналы связи

- В каналах Open Flow используются SSL/TLS, но данные протоколы не являются обязательными
- Аутентификация между контроллером и OF устройствами
- DDoS атаки – поддержание насыщенности канала

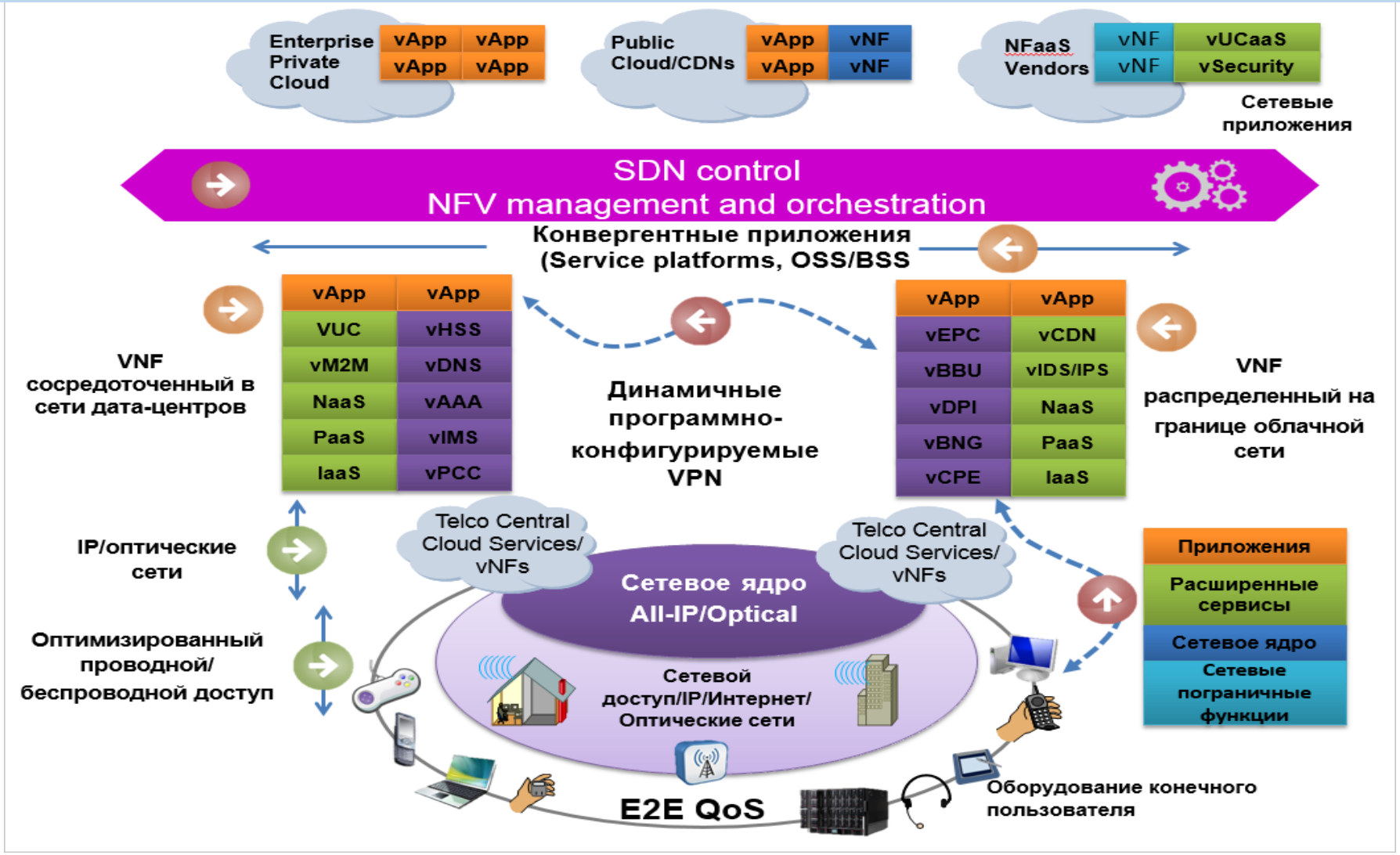
Контроллер

- Обеспечение безопасности контроллера
- Компрометация контроллера позволяет атакующим управлять всей сетью
- DDoS атаки на контроллер
- Поддельный контроллер может изменять топологию сети
- Строгий механизм аутентификации для доступа к SDN-контроллеру
- Целостность контроллера
- Внедрение нежелательной информации в контроллер

Control Plane

- Требуется обеспечение безопасности control plane, управление авторизацией доступа для сетевых приложений
- Требуется аутентификация доступа приложений на control plane
- Сеть должна обслуживать требования бизнес приложений, и логика данный приложений определяет способы обеспечения безопасности

Конвергентная сеть



Источник: Bell Labs Alcatel Lucent «Reshaping the future with NFV and SDN», 2015

Информационная сеть vs Компьютерная сеть

Information Centric Networking

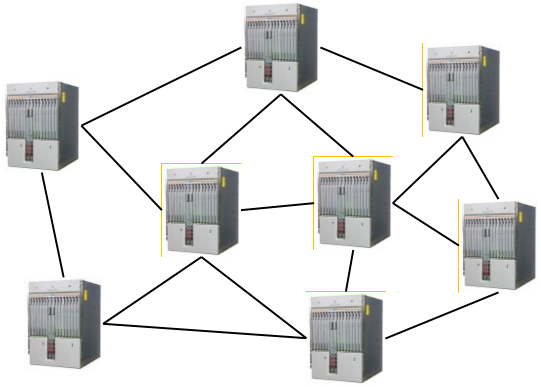
Интернет сегодня

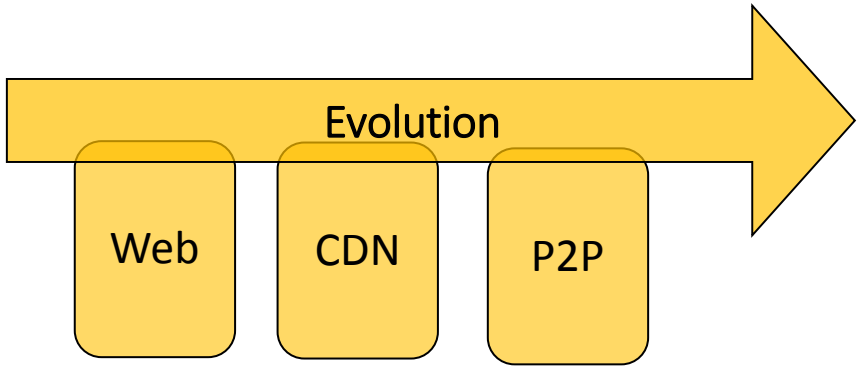
Акцент на узлах

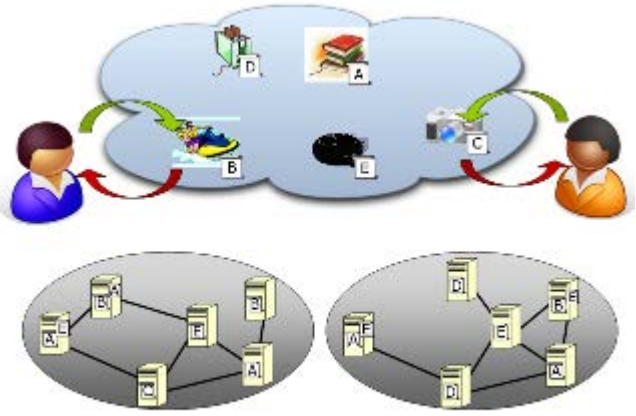
В современном интернете доминирующая функция – доступ к информации!

Information Centric Network

Фокус на объектах информации







Важные требования:

- доступ к названным ресурсам, а не хостам
- масштабируемое распределение через репликацию и кэширование
- хороший контроль разрешающей способности маршрутизации и доступа

С повсеместным кэшированием, НО для всех приложений и для всех пользователей и провайдеров контента!

Потенциальные новые ниши для SDN и NFV

Динамическое обеспечение услуг	Новые расширенные услуги
▪ Предоставление пропускной способности по требованию	▪ Контекстная оптимизация качества сервисов в режиме реального времени
▪ Расширение возможности управления пользователем виртуальными сетями (тенант)	▪ Сервисы безопасности: firewalls, IPS. IDS и безопасность конечных пользователей
▪ Эластичная пропускная способность, учитывающая «взрывной» трафик	▪ Iaas: вычисление, хранение и «рабочий стол» как сервис
▪ Оптимизация качества обслуживания сервисов в режиме реального времени и в зависимости от контекста	▪ Сетевые функции как сервис (vIMS и vEPC)
▪ Быстрое развертывание и конфигурирование информационных ресурсов предприятия	▪ Связность подключенного предприятия: SD-VPN и виртуальные CPE сервисы
▪ Возможность быстрой костомизации сервиса	
▪ Федерация динамических виртуальных сетей от разных операторов	

«Если в 80-е годы главным было качество, а в 90-е – реинжиниринг, то в 2000-е главное - **скорость»**.

Bill Gates, Microsoft

«То, насколько быстро вы можете **адаптировать** свои цели, лучше всего характеризует вашу компанию. Поэтому надо прививать людям **вкус к переменам.** Надо говорить о переменах постоянно».

Jack Welch, General Electric

Источник: Bell Labs Alcatel Lucent «Reshaping the future with NFV and SDN», 2015

ВОРОСЫ?

ЦЕНТР
ПРИКЛАДНЫХ
ИССЛЕДОВАНИЙ
КОМПЬЮТЕРНЫХ
СЕТЕЙ

http://arccn.ru/

smel@arccn.ru


+7 (495) 240-50-63


@ArccnNews