

**МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ,  
СВЯЗИ И МАССОВЫХ КОММУНИКАЦИЙ РОССИЙСКОЙ  
ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ  
УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ ИМ. ПРОФ. М.А. БОНЧ-  
БРУЕВИЧА»  
(СПбГУТ)**

---

Факультет Инфокоммуникационных сетей и систем

Кафедра Защищенных систем связи

Дисциплина Защита информации в центрах обработки данных.

**Пояснительная записка к курсовой работе**

Внедрение Openstack в виртуальную инфраструктуру VMware vSphere

---

*(тема отчета)*

11.03.02 Инфокоммуникационные технологии и системы связи

*(код и наименование направления/специальности)*

Студент:

Громов А.А. ИКТЗ - 83

Преподаватель:

Цветков А.Ю

## Оглавление

Исходные данные.....	3
Постановка задач .....	4
Алгоритм решения задачи .....	<b>Ошибка! Закладка не определена.</b>
Структура OpenStack. ....	<b>Ошибка! Закладка не определена.</b>
Обзор VMware Integrated OpenStack.....	<b>Ошибка! Закладка не определена.</b>
Основные возможности VMware Integrated OpenStack .....	<b>Ошибка! Закладка не определена.</b>
Основные сценарии использования VMware Integrated OpenStack .....	<b>Ошибка! Закладка не определена.</b>
VMware vCenter driver for OpenStack Compute.....	<b>Ошибка! Закладка не определена.</b>
Высокоуровневое представление архитектуры драйвера VMware. ....	<b>Ошибка! Закладка не определена.</b>
Применённое решение. ....	<b>Ошибка! Закладка не определена.</b>
Развёртывание vAPP шаблона OpenStack.....	<b>Ошибка! Закладка не определена.</b>
Автоматическое развёртывание OpenStack.....	<b>Ошибка! Закладка не определена.</b>
Вывод .....	<b>Ошибка! Закладка не определена.</b>

## **Исходные данные**

Исследование принципов работы протоколов SAN-сетей: FC, iSCSI и FCoE.

## **Постановка задач**

Познакомиться с типами систем хранения данных DAS, NAS и SAN.

Рассмотреть разницу между различными СХД: DAS, NAS и SAN.

Подробнее рассказать про SAN

Провести сравнение FC и iSCSI

Рассказать про NVMe.

Составить сравнительную таблицу протоколов SAN (скорость, стоимость).

## **Алгоритм решения задачи.**

### **Системы хранения данных.**

В случае отдельного ПК под системой хранения данных (СХД) можно понимать отдельный внутренний жесткий диск или систему дисков.

Традиционно можно выделить три технологии организации хранения данных: Direct Attached Storage (DAS), Network Attach Storage (NAS) и Storage Area Network (SAN). Они различаются способом организации доступа к СХД:

- DAS - система хранения, непосредственно подключаемая к серверу;
- NAS - система хранения, подсоединяемая к сети;
- SAN - сеть хранения данных. Основой SAN является выделенная специализированная сеть, которая служит исключительно для организации доступа к данным.

Основное отличие технологий связано с сетевыми интерфейсами. В случае с DAS все три компонента (приложения, файловая система и диски) размещены в одной системе.

Данные в NAS идет по сети Ethernet, используемой не только для передачи данных, но и для выхода в локальную или глобальную сеть, из-за чего с точки зрения безопасности является не безупречным вариантом. А поскольку по ethernet сети идет и другой трафик, кроме NAS, то скорость передачи данных хуже, чем у DAS, и при обращении к данным возможны заметные потери производительности. Дополнительные нагрузки, связанные с протоколом TCP/IP, также негативно влияют на быстродействие. Тем не менее, реализация iSCSI (SCSI over Ethernet) увеличивает быстродействие.

В сети SAN физические диски сосредоточены в одной сети. Они могут соединяться в огромные фермы, благодаря которым удобно масштабировать возможный хранимый объем данных.

Каждая технология хранения имеет свои преимущества и недостатки, однако редко удается найти возможность прийти к компромиссу между высокой скоростью, отказоустойчивостью и низкой ценой, поэтому при

построении систем с прицелом на будущее необходимо тщательно обдумать вопрос масштабируемости и производительности. В таблице 1 отражены особенности технологий хранения.

Таблица 1

Технологии	DAS	NAS	SAN
Сетевая технология	Нет	Да	Да
Защита от внешних помех	Нет (медь)	Нет (медь) / да (оптика)	Да (оптика)
Уровень цен	Низкий	Высокий	Очень высокий
Масштабируемость	Плохая	Хорошая	Очень хорошая
Максимальное расстояние до сервера	25м	-	10км
Основной используемый интерфейс	SCSI	Ethernet	FCP/ISCI

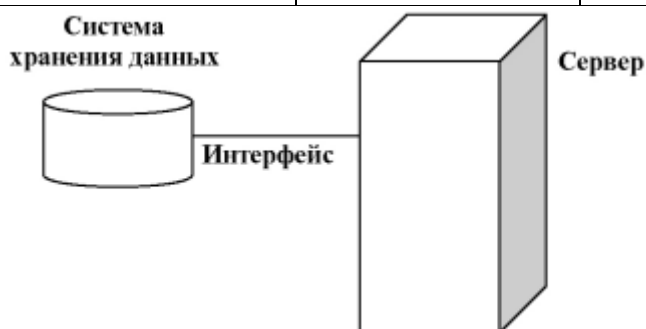


Рис Пример инфраструктуру DAS

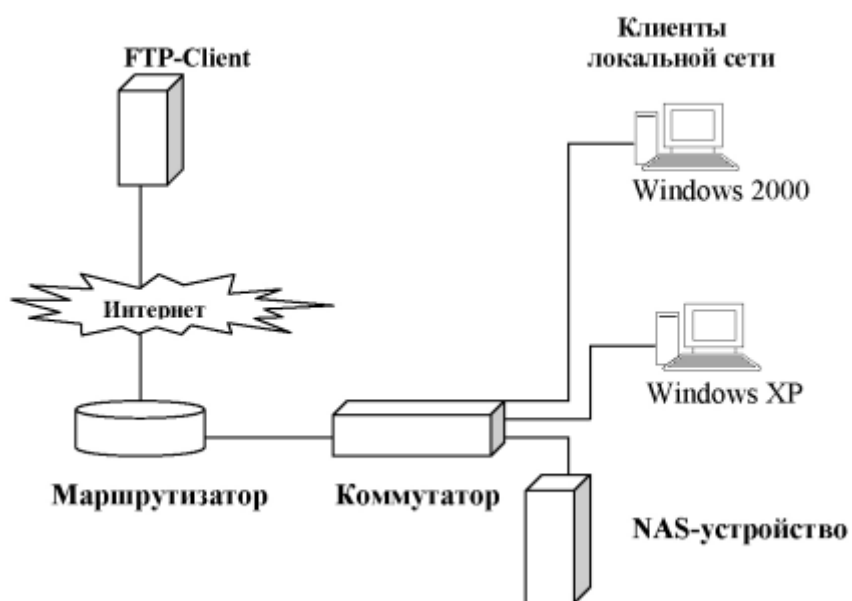


Рис Пример инфраструктуры NAS

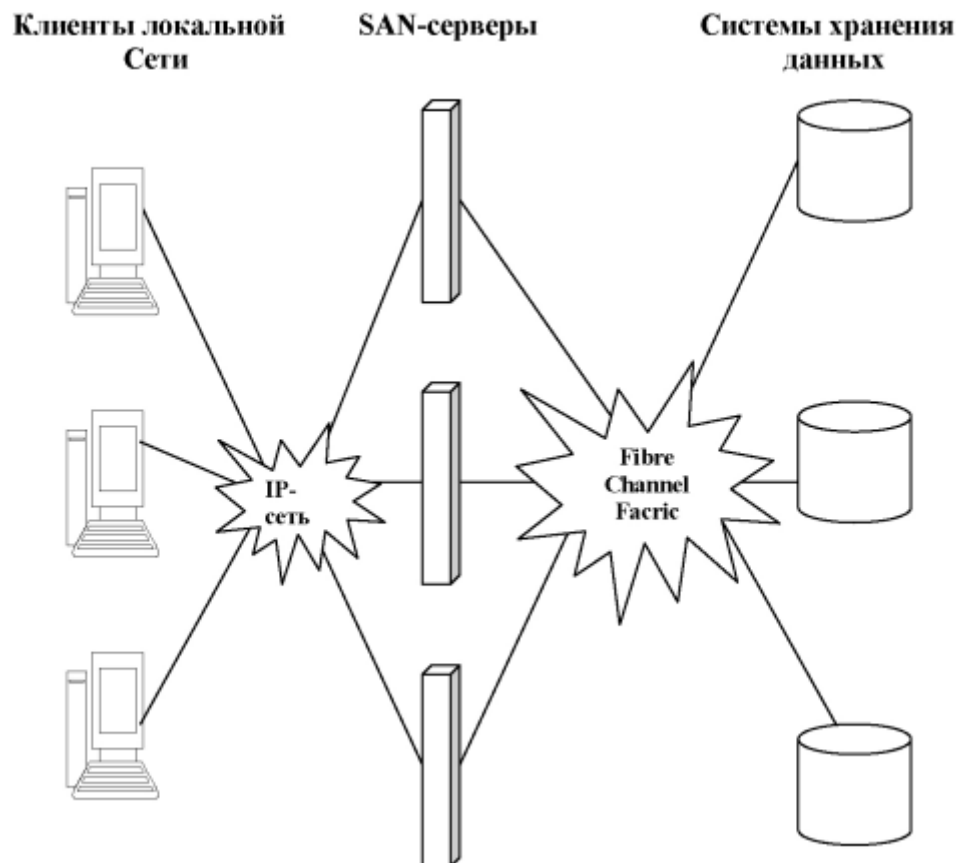


Рис Пример инфраструктуры SAN

## **SAN**

### **FC vs ISCSI**

В данный момент существует 2 основных протокола передачи информации в san-сетях:

- Fiber Channel Protocol (FCP)
- Internet Small Computer System Interface (iSCSI)

### **FC-SAN**

Fibre Channel — популярный протокол хранения, обеспечивающий низкие задержки и высокую пропускную способность за счёт своих архитектурных особенностей. Fibre Channel не требователен к ресурсам и отлично подходит для передачи большого объёма данных, так как все операции FC выполняются на стороне HBA, разгружая центральный процессор.

Новые версии протокола Fibre Channel обратно совместимы с прошлыми редакциями, что открывает хорошие перспективы для модернизации и масштабирования. Например, если внедрять FC 32Гб/с, то всё ещё можно будет использовать FC 8Гб/с и 16Гб/с, т.е. можно поэтапно менять FC-коммутаторы и FC адаптеры.

В ближайшее время FC будет обновлён до 64Гб/с и 128Гб/с (уже сейчас есть коммутаторы, поддерживающие агрегацию 4-х портов 32Гб/с в один канал 128Гб/с для соединения коммутаторов).

Простота настройки и удобство в администрировании позволили FC стать одним из наиболее распространенных протоколов хранения. Большинство администраторов SAN-сетей во всем мире знает, как он устроен и какие преимущества обеспечивает при решении различных задач. При этом FC всё ещё сложнее, чем Ethernet, хотя и обладает большим количеством средств управления и мониторинга.

Название	Скорость линии, Гбод	Пропускная способность, Мбайт/с	Год
1GFC	1	100	1997
2GFC	2	200	2001
4GFC	4	400	2004
8GFC	8	800	2005
10GFC	10	1,200	2008
16GFC	14	1,600	2011
32GFC "Gen 6"	28	3,200	2016 <sup>[10]</sup>
64GFC "Gen 7"	29	6,400	2019 <sup>[11]</sup>
128GFC "Gen 6"	28.05 ×4	12,800	2016 <sup>[10]</sup>
256GFC "Gen 7"	28.9 ×4	25,600	2019 <sup>[12]</sup>
128GFC "Gen 8"	57	12,800	Planned 2022



В Parallel FC на 128 Гбит/с используются четыре оптических волокна для передачи данных в прямом направлении и четыре — в обратном. Какие кабели и соединители потребуются для FC на 128 Гбит/с? В качестве портов могут использоваться модули QSFP28, CFP2, CFP4 или какие-то будущие четырехканальные интерфейсы, а для подключения оборудования — 12-волоконные кабели с соединителями MPO, активные волоконно-оптические кабели (АОС) длиной до 50 м или кабели прямого подключения (DAC) до 5 м.

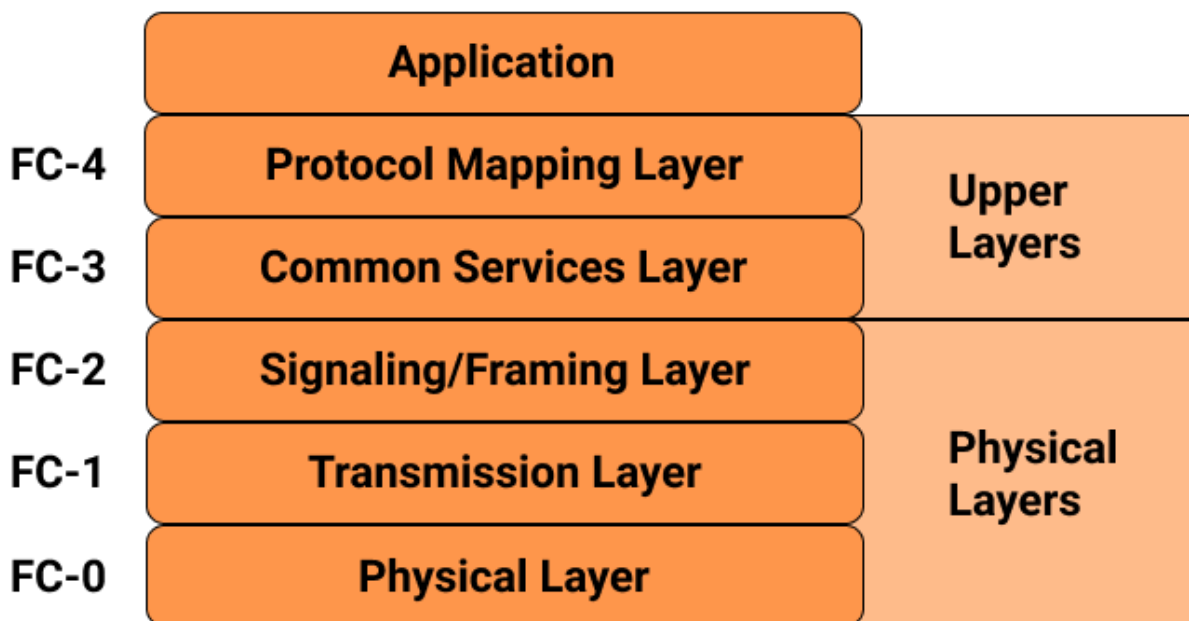
### **Что необходимо для построения Fibre Channel San.**

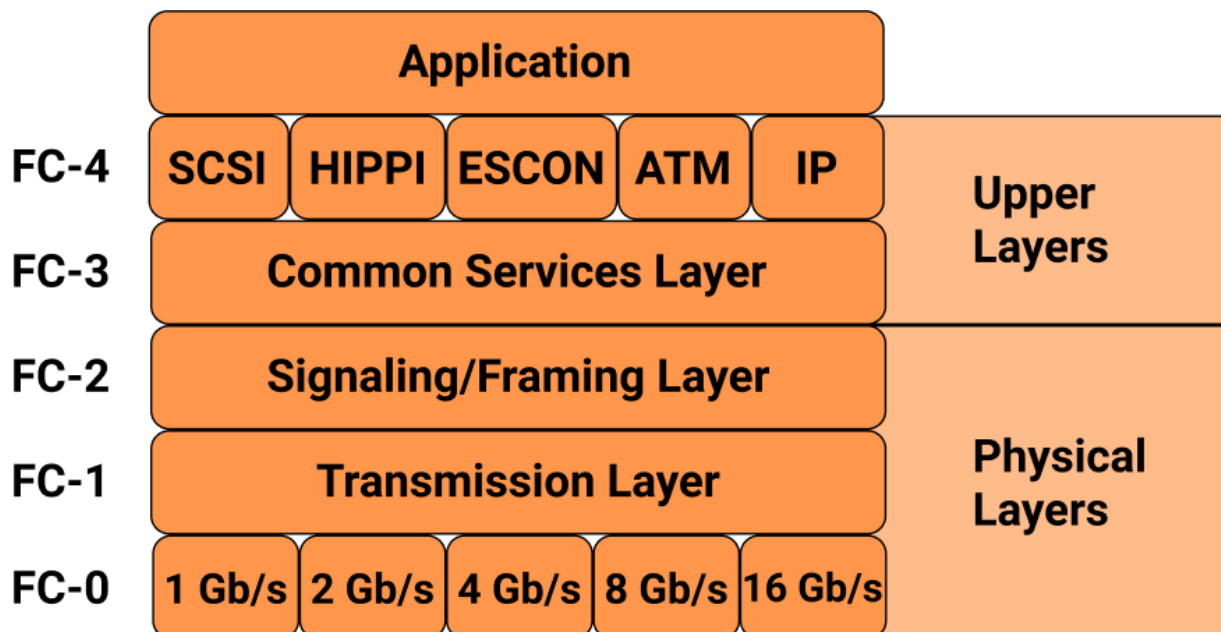
Технология Fibre Channel предшествовала стандарту iSCSI на девять лет, а протокол Fibre Channel (FCP) является единственным коммерчески используемым протоколом Fibre Channel. FCP - это стандартный протокол хранения SCSI, обернутый в кадры Fibre Channel. Существует три физических компонента сети Fibre Channel SAN. Первый компонент - это адаптер хоста Fibre Channel (HBA), известный также как инициатор. HBA представляет собой карту PCIe или чип, расположенный на материнской плате хоста. Он внедряет стандарт FC в микросхему, минимизируя ресурсы хоста. Программный драйвер также находится на хосте. Вторая часть - это коммутатор Fibre Channel. Коммутатор FC - это неблокирующий (без внутренней перегрузки пропускной способности) коммутатор второго уровня с широкими возможностями управления, соответствующий стандарту FC. Разновидностью коммутатора FC является директор FC. FC director - это неблокируемый коммутатор с большим количеством портов; он обладает высокой доступностью (HA) без единой точки отказа и функциями корпоративного класса (такими как шифрование, виртуализированные SAN, шлюзы для других сетей и т. д.). Третья часть - это цель HBA, называемая FC storage target port, или target. Это внешнее подключение хранилища к сети Fibre Channel SAN (FC fabric). Хранилище предоставляет адрес (который встроен в адаптер) или всемирное имя (WWN) адреса каждого хранилища. Каждый хост имеет свой уникальный WWN-адрес, который позволяет ему подключаться к

хранилищу для каждого сеанса чтения или записи. Каждый хост имеет свой уникальный WWN-адрес, который позволяет ему подключаться к хранилищу для каждого сеанса чтения или записи. Одним из ключевых элементов программного обеспечения является многоканальная система Fibre Channel SAN multipathing, которая находится на каждом узле, получающем доступ к внешнему хранилищу. Это позволяет каждому узлу иметь несколько путей к хранилищу для устранения отказов, а в некоторых случаях - для балансировки нагрузки и увеличения производительности. Пропускная способность портов инициатора, цели и коммутатора FC в настоящее время доступна в вариантах 8 или 16 Гбит/с, совместимых между собой.

### **Fibre Channel Protocol**

Как и сетевая модель TCP/IP, Fibre Channel состоит из пяти уровней. Каждый уровень выполняет определённый набор функций.





### Краткий обзор каждого уровня

**FC-0** — уровень физических интерфейсов и носителей. Описывает физическую среду: коннекторы, кабели, трансиверы, НВА, электрические и оптические параметры.

**FC-1** — уровень передачи и кодирования. Описывает кодирование и декодирование данных перед передачей и после получения. Уровень имеет три основные функции:

- Кодирование / декодирование
- Ordered sets
- Инициализация соединения (link initialization)

**FC-2** — уровень кадрирования и сигналов. Создает FC кадры и управляет потоком. Контроль потока нужен нам для проверки того, сколько данных мы можем отправить в определенный момент времени, чтобы принимающая сторона могла обработать все запросы. Функции, осуществляемые на этом уровне:

- Кадрирование (определение структуры кадра).
- Управление последовательностями (Sequence management)
- Управление обменом (Exchange management)

- Класс обслуживания (Class of Service)
- Управление потоком (Flow control)

**FC-3** — уровень базовых служб. Заложен, для новых функций, которые могут быть внедрены в Fibre Channel. На этом уровне обеспечивается шифрование и сжатие данных перед отправкой.

**FC-4** — уровень отображения протоколов. Обеспечивает сопоставление возможностей Fibre Channel с протоколами верхнего уровня. Наиболее популярным является SCSI. Благодаря FCP, есть возможность использовать протокол передачи данных SCSI с Fibre Channel инфраструктурой.

Рассмотрим физические уровни Fibre Channel Protocol.

#### FC-0

Тип волокна	Скорость (MB/s)	Трансмиттеры	Модификация	Расстояние
Одно-модовое волокно (SMF)	12,8	1,310 nm Длинноволновой лазер	128GFC-PSM4	0.5m - 0.5km
		1,270, 1,290, 1,310 and 1,330 nm Длинноволновой лазер	128GFC-CWDM4	0.5m - 2km
	6,4	1,310 nm Длинноволновой лазер	64GFC-LW	0.5m - 10km
	3,2	1,310 nm Длинноволновой лазер	3200-SM-LC-L	0.5m - 10km
	1,6	1,310 nm Длинноволновой лазер	1600-SM-LC-L	0.5m - 10km
		1,490 nm Длинноволновой лазер	1600-SM-LZ-I	0.5m - 2km
	800	1,310 nm Длинноволновой лазер	800-SM-LC-L	2m - 10km
			800-SM-LC-I	2m - 1.4km
	400		400-SM-LC-L	2m - 10km

		1,310 nm Длинноволновой лазер	400-SM-LC-M	2m - 4km
			400-SM-LL-I	2m - 2km
	200	1,550 nm Длинноволновой лазер	200-SM-LL-V	2m - 50km
		1,310 nm Длинноволновой лазер	200-SM-LC-L	2m - 10km
			200-SM-LL-I	2m - 2km
	100	1,550 nm Длинноволновой лазер	100-SM-LL-V	2m - 50km
		1,310 nm Длинноволновой лазер	100-SM-LL-L	2m - 10km
			100-SM-LC-L	
100-SM-LL-I			2m - 2km	
Много-модовое волокно (MMF)	12,8	850 nm Коротковолновой лазер	128GFC-SW4	0m - 100m
	6,4		64GFC-SW	0m - 100m
	3,2		3200-SN	0m - 100m
	1,6		1600-M5F-SN-I	0.5m - 125m
			1600-M5E-SN-I	0.5m - 100m
			1600-M5-SN-S	0.5m - 35m
			1600-M6-SN-S	0.5m - 15m
			800-M5F-SN-I	0.5m - 190m
	800		800-M5E-SN-I	0.5m - 150m
			800-M5-SN-S	0.5m - 50m
			800-M6-SN-S	0.5m - 21m
			400-M5F-SN-I	0.5m - 400m
			400-M5E-SN-I	0.5m - 380m
	400		400-M5-SN-I	0.5m - 150m
			400-M6-SN-I	0.5m - 70m
			200-M5E-SN-I	0.5m - 500m
			200-M5-SN-I	0.5m - 300m
	200		200-M6-SN-I	0.5m - 150m
			100-M5E-SN-I	0.5m - 860m
	100		100-M5-SN-I	0.5m - 500m
			100-M6-SN-I	0.5m - 300m
			100-M5-SL-I	2m - 500m
			100-M6-SL-I	2m - 175m

В таблице X представлены актуальные на данный момент типы волокна, скорости, модификации и расстояния.

Существует 2 типа волокна: многомодовое и одномодовое

Многомодовое волокно (Multimode Fiber, MMF) — широкое в сечении (50-62,5 микрон) относительно одномодового, предназначено для коротковолновых лазерных лучей. Слово многомодовое означает, что по волокну может идти сразу несколько лучей света, отражаясь от стенок волокна. Это позволяет быть кабелю менее чувствительным к перегибам, однако снижает силу и ухудшает качество сигнала, таким образом накладывая ограничение на дистанцию передачи данных до 500 м.

Одномодовое волокно (Singlemode Fiber, SMF) — тонкое волокно (8-10 микрон), через которое передается, не видимый человеческому глазу, сигнал длинноволновым лазером. Свет в таком волокне передается по прямой, без отражений от стенок волокна. По этой причине сигнал передается быстрее и с меньшими потерями. Однако из-за высоких цен на оборудование, передача данных по одномодовому волокну производится только для больших расстояний (до 50 км без усилителей).

Для подключения двух устройств и работы дуплексного режима, необходимо использовать два кабеля: один для передачи, второй для приема.

## **ASIC**

Переход от FC-0 к FC-1 и обратно обеспечивает ASIC это один из элементов устройств HBA, дисковых массивов и коммутаторов.

ASIC — это интегральная схема, оптимизированная для решения конкретной задачи. В отличие от интегральных схем общего назначения, специализированные интегральные схемы применяются в конкретном устройстве и выполняют строго ограниченные функции, характерные только для данного устройства; вследствие чего исполнение функций происходит значительно быстрее, что приводит к удешевлению процесса. Ярким

примером ASIC являются устройства для майнинга, единственная функция которых – считать хэши.

В устройствах Fibre Channel ASIC состоит из следующих функциональных элементов:

Encoder / Decoder — обеспечивает кодирование каждых 8 бит передаваемых данных в 10-битное представление. И декодирование обратно принимаемых данных.

SERDES (Serializer / Deserializer) — преобразует параллельный поток в последовательный поток 10-битных порций данных.

Transceiver — преобразует электрические импульсы в световые сигналы.

Трансиверы или SFP — в случае FC-коммутаторов это отдельные модули, необходимые для подключения кабеля к порту.

Подразделяются на коротковолновые (Short Wave, SW, SX) и длинноволновые (Long Wave, LW, LX). LX-трансиверы используются с многомодовым и одномодовым волокном. SW-трансиверы — только с многомодовым. И к тем и к другим кабель подключается разъёмом LC.

Также существует SFP xWDM (Wavelength Division Multiplexing), предназначенные для передачи данных из нескольких источников на большие расстояния одним световым пучком. Для подключения кабеля к ним используется разъём SC.

## **FC-1**

Первое, что происходит на этом уровне — кодирование / декодирование информации. Это довольно сложный процесс, в ходе которого каждые 8 бит информации преобразуются в 10-битное представление. Это происходит с целью улучшения контроля целостности данных, разделение данных и служебных сигналов, а также для возможности восстановления тактового сигнала из потока данных (сохранение баланса нулей и единиц).

Данные преобразования ведут к сильному снижению полезной пропускной способности, если посчитать, то получается, что 20% данных являются избыточной служебной информацией. В добавок к этому, некоторую часть потока данных может занимать служебный трафик.

Тем не менее такое кодирование используется только в оборудовании 1G, 2G, 4G и 8G. Однако в некоторых реализациях 10G, а начиная с 16G - кодирование осуществляется по принципу 64/66 бит, тем самым существенно увеличивает поток реальных данных (до 97% против 80% в случае 8/10).

## **Ordered sets**

Для объяснения данного термина необходимо знать, что такое transmission word. Transmission word – минимальная количество данных для передачи, равная четырем байтам. Если информация меньше четырех байт, то к ней добавляют специальные символы – заполняющие байты.

Ordered sets — это специальные служебные transmission words, которые делятся на три категории:

- Разделители фреймов(frame delimiters)
    - Start-of-Frame (SOF) используется для обозначения начала кадра, а также указывает, является ли этот кадр первым в последовательности и какой класс обслуживания используется.
    - End-of-Frame (EOF) используется для обозначения конца кадра, а также указывает, является ли этот кадр последним в последовательности и какой класс сервиса используется.
- Существует несколько подтипов SOF и EOF. Эти подтипы используются для передачи информации о состоянии кадра. Например:
- SOFix (SOF Initiate) предшествует первому кадру последовательности; x - число, указывающее на класс обслуживания.
  - SOFnx (SOF Normal) предшествует кадрам, отличным от первого кадра последовательности; x - число, указывающее на класс обслуживания.
  - EOFn (EOF Normal) завершает кадр, который не является последним кадром последовательности.



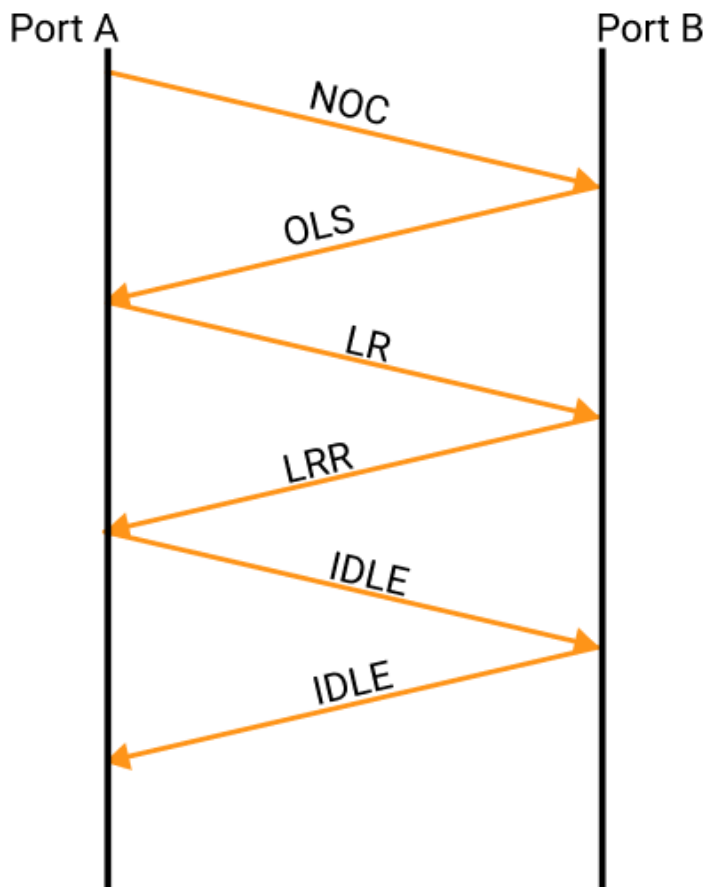
- EOFt (EOF Terminate) завершает кадр, который является последним кадром последовательности.
- EOFa (EOF Abort) завершает кадр, который был аномально прерван во время передачи.
- Примитивные сигналы (Primitive Signals) - это ordered sets, которые используются для сигнализации событий на передающем порту. Примитивные сигналы используются для управления потоком и синхронизации. В управляемых петлях примитивные сигналы также используются для управления виртуальными цепями. Выделяют 2 вида примитивных сигналов – fill words и control signals
  - (fill word) Сигнал IDLE помогает в синхронизации и повторной синхронизации. Между кадрами передается не менее шести IDLE сигналов, а для синхронизации и ретайминга между кадрами могут быть добавлены дополнительные IDLE сигналы. IDLE сигналы передаются непрерывно, когда кадры не передаются, для поддержания синхронизации.
  - (fill word) Сигнал ARB (Arbitrate) используется только в управляемых петлях. ARB - это часть механизма, который используется портами петли для принятия решения о том, какой порт контролирует петлю в определенный момент времени.
  - (control signal) R\_RDY (Receiver Ready) отправляется портом, когда порт готов к приему данных. R\_RDY используется в управлении потоком типа "buffer-to-buffer".
  - (control signal) OPN (Open Port) посылается инициатором в управляемой петле, когда инициатор хочет открыть соединение с портом.
  - (control signal) CLS (Close Port) посылается инициатором в управляемой петле, когда инициатор хочет закрыть соединение с портом.
  - (control signal) SYN (Clock Synchronization) используется для поддержания синхронизации в конфигурациях "point-to-point" и коммутируемой сети.
  - (control signal) VC\_RDY (Virtual Circuit Ready) используется для управления потоком в соединениях виртуальных цепей класса 4.
  - (control signal) DHD (Dynamic Half-Duplex) передается для изменения полнодуплексной передачи на полудуплексную между двумя L\_Ports в управляемой петле.
- Примитивные последовательности (primitive sequences) - это еще один тип ordered sets. Примитивные последовательности - это простые

протоколы уровня канала, которые используются для управления каналом.

Разница между примитивными сигналами и примитивными последовательностями заключается в том, что примитивные сигналы передаются один раз, а примитивные последовательности передаются многократно до получения ответа. Порты должны получить не менее трех последовательных повторений одной и той же примитивной последовательности, прежде чем подтвердить ее достоверность и ответить соответствующим образом. Примитивными последовательностями являются:

- NOS (Non-Operational State) и OLS (Offline State). NOS отправляется сетевым портом, который находится в отключенном состоянии. Принимающий порт отвечает OLS, чтобы начать процесс инициализации соединения.
- LR (Link Reset) и LRR (Link Reset Response). LR отправляется портом сети, который хочет инициировать операцию сброса соединения; принимающий порт отвечает LRR.
- LIP (Loop Initialization). LIP отправляется портом петли для начала процесса инициализации петли.
- LPB (Loop Port Bypass) и LPE (Loop Port Enable). Когда порт получает LPB, он переходит в режим bypass. В режиме bypass порт ретранслирует все принятые слова передачи, но не участвует в операциях цикла и не обрабатывает данные. LPE снова включает порт, находящийся в режиме bypass.

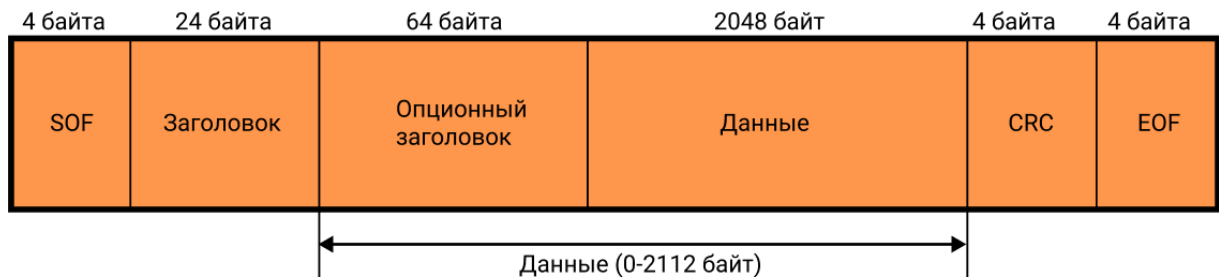
### **Инициализация соединения (Link initialization)**



## FC-2

Все данные, передаваемые в среде Fiber Channel разбиваются на кадры.

### Структура кадра



Рис

SoF – Начало кадра.

Заголовок содержит следующую информацию:

- Адрес источника.
- Адрес приёмника.
- Тип кадра: FT-0 — управляющий или FT-1 — данные.
- Номер последовательности.
- Порядковый номер кадра в последовательности.
- Служебная информация.

CRC – контрольная сумма.

EoF – Конец кадра.

Интервалы между кадрами заполняются «заполняющими словами» — fill word. Обычно, это fill word IDLE, однако начиная с FC 8G было принято решение использовать ARB(FF) вместо IDLE, в целях уменьшения электрических помех в медном оборудовании (но по-умолчанию коммутаторами используется IDLE).

### **Последовательности**

Чаще всего передающая сторона стремится отправить принимающей стороне информацию большего объема, чем 2112 байт (максимальный объём данных одного кадра). В этом случае информация делится на несколько кадров, а набор этих кадров называется последовательностью. Чтобы в логическую последовательность кадров не попадало ничего лишнего при параллельной передаче, заголовок каждого кадра имеет поля SEQ\_ID (идентификатор последовательности) и SEQ\_CNT (номер фрейма в последовательности).

### **Обмен**

По сравнению с традиционными сетевыми протоколами, уникальная способность Fibre Channel отслеживать последовательности и обмены приводит к тому, что этот протокол особенно подходит для передачи транзакционной информации, такой как данные о выполнении команд SCSI.

Обмен состоит из одной или нескольких последовательностей, а каждая последовательность внутри обмена содержит один или несколько кадров Fibre Channel. Связь между кадрами, последовательностями и обменами определяется в заголовке каждого передаваемого кадра.

Этот пример обмена между двумя узлами приводит к передаче трех последовательностей Fibre Channel. Первая последовательность содержит один кадр (кадр 1) и передается отправителем для начала обмена. Отправитель заполняет заголовок кадра информацией, которая однозначно идентифицирует обмен (OX\_ID=1), а также последовательность (SEQ\_ID=1), к которой принадлежит кадр. Дополнительно, порядок следования кадра в последовательности указывается SEQ\_CNT=1. В результате получения последовательности №1, ответчик посылает последовательность №2, которая состоит из двух кадров Fibre Channel. Опять же, заголовок каждого кадра содержит информацию, которая связывает его с исходным обменом (OX\_ID=1), но с отдельной последовательностью (SEQ\_ID=2). Наконец, каждый кадр в последовательности №2 индивидуально пронумерован (SEQ\_CNT=1, SEQ\_CNT=2). Третья и последняя последовательность этого обмена снова отправляется ответчиком. Состоящий из одного кадра, соответствующие поля заголовка указывают, что этот кадр является первым (и последним) кадром третьей (и последней) последовательности обмена №1.

### **Классы обслуживания (Classes of Services)**

Fibre Channel поддерживает функцию, называемую классом обслуживания, которая позволяет узлу запрашивать у сети определенные характеристики доставки. Эта функция вводится на данном этапе, поскольку запрос на определенный класс обслуживания неявно содержится в разделителе SoF, рассмотренном в предыдущем параграфе. Разделитель SoF представляет собой ordered set, который предшествует каждому кадру, передаваемому в сеть. Существование нескольких классов обслуживания подразумевает существование многих типов SoF в зависимости от запрашиваемого класса обслуживания. Fibre Channel поддерживает шесть классов обслуживания, которые перечислены ниже.

#### **Класс 1**

Класс 1 - это сервис, ориентированный на соединение, который характеризуется тем, что вся пропускная способность канала зарезервирована и выделена для одной сессии между двумя портами Nx\_Port. Термин «ориентированный на соединение» указывает на то, что сеть должна настроить выделенный канал для передачи кадров класса 1 между двумя портами. При необходимости этот канал может охватывать множество коммутаторов Fibre Channel и проходить через ISL, что подразумевает, что коммутаторы должны сотрудничать в создании этого канала посредством какого-либо протокола между коммутаторами. Выделенные каналы проходят по одному пути через сеть, что гарантирует доставку кадров в порядке очереди и отсутствие перегрузок. В дополнение к управлению потоком на уровне канала (система управления типа «buffer-to-buffer»), класс 1 также требует управления потоком типа «end-to-end».

## **Класс 2**

Класс 2 - это сервис без соединения, который использует управление потоком типа «buffer-to-buffer» и «end-to-end», в дополнение к подтверждению доставки (или недоставки) кадров: каждый кадр класса 2, полученный Nx\_Ports, подтверждается получение с помощью расширенной службы соединения (ELS). Термин "без соединения" означает, что сеть рассматривает каждый кадр независимо от всех других кадров и что пропускная способность не резервируется для какого-либо конкретного разговора между Nx\_Ports, что может привести к периодическим перегрузкам на загруженных ISL. Кадры с одинаковым получателем могут следовать разными путями через сеть, потенциально прибывая не по порядку. Обратите внимание, что Fibre Channel не предъявляет формальных требований к порядку доставки, хотя на практике это может вызвать проблемы и обычно избегается производителями коммутаторов.

- Класс 3: почти все порты Fibre Channel, развернутые в современных сетях SAN, используют сервис класса 3 для трафика хранения. класс 3 для трафика хранения данных. Класс 3 - это служба без подключения, которая

поддерживает только систему управления потоком типа «buffer-to-buffer». В результате класс 3 не гарантирует доставку в порядке очереди, и отправителю не возвращается информация о том, что кадр, который был потерян из-за перегрузки, повреждения или загруженного порта.

#### **Класс 4**

Класс 5 - класс услуг, использует виртуальные каналы через управляемый сетью протокол распределения полосы пропускания. Кадры гарантированно доставляются в правильном порядке, а также обеспечивается индикация доставки (или недоставки) кадров.

#### **Класс 5**

Класс 5 еще полностью не описан и не включен в стандарт. Предварительно, класс, не требующий соединения, но требующий немедленной доставки данных по мере их появления, без буферизации на устройствах.

#### **Класс 6**

Класс 6 - это однонаправленный класс обслуживания, при котором один поток кадров, исходящих от порта N\_Port, передается на коммутатор по выделенному каналу (это подразумевает один сеанс при полной пропускной способности канала). Фабрика реплицирует поток кадров в выбранную группу пунктов назначения по равному количеству выделенных каналов. Класс 6 также известен как служба многоадресной рассылки и требует наличия специального устройства в сети, называемого сервером многоадресной рассылки.

#### **Класс F**

Кадры класса F используются исключительно коммутаторами Fibre Channel, когда они общаются друг с другом по межкоммутаторным каналам с целью конфигурирования и обслуживания сети.

### **Flow Control**

Управление потоком - это механизм, который гарантирует, что передающая сторона отправит кадр только тогда, когда принимающая сторона будет готова его принять. В Fibre Channel определены два уровня управления потоком: управление потоком на уровне канала (link-level flow control) и сквозное управление потоком (end-to-end flow control). Их использование определяется классом обслуживания. Например, класс 1 использует только механизм End-to-End, класс 3 — Buffer-to-Buffer, а класс 2 — оба эти механизма.

### **Buffer-to-Buffer flow control**

Принцип технологии - отправка любого кадра должна быть обеспечена наличием кредита на отправку.

Все поступающие на вход порта кадры помещаются в специальную очередь — буфер. Количество буферов определяется физическими характеристиками порта. Один буфер (место в очереди) соответствует одному кредиту. Каждый порт имеет два счётчика кредитов:

TX BB\_Credit — счётчик кредитов передачи. После отправки каждого кадра, уменьшается на 1. Если значение счётчика стало равным нулю — передача не будет осуществляться. Как только от порта-приёмника получено R\_RDY, счётчик увеличивается на 1.

RX BB\_Credit — счётчик кредитов приёма. Как только кадр принят и помещён в буфер, значение счётчика уменьшается на 1. Когда кадр обрабатывается или пересылается дальше, счётчик увеличивается на 1, а отправителю отправляется R\_RDY. Если значение счётчика падает до 0, то в теории, приём новых кадров должен быть прекращён. На практике, из-за ошибок синхронизации кредитов может возникнуть ситуация, что источник прислал ещё один или даже несколько кадров уже после того как RX BB\_credit стал равен нулю. Данная ситуация называется buffer overflow. В большинстве реализаций порт обрабатывает такие ситуации за счёт резервных буферов. Однако, некоторое оборудование может отправить Link Reset сигнал.



Таким образом влияние расстояния между портами сильно сказывается на производительности. Чем больше расстояние и пропускная способность канала, тем больше кадров будет отправлено ещё до того, как принимающая сторона получит хотя бы первый. Но особенность архитектуры FC-коммутаторов сглаживает этот недостаток: количество буферов не закреплено жёстко за каждым портом (кроме восьми обязательных), а является общим для всех. И в случае определения длинных каналов (автоматически или вручную) количество выделяемых коммутатором буферов для этого порта увеличивается. Еще одно преимущество общей памяти заключается в том, что нет потребности переносить буферы от одного порта к другому внутри коммутатора.

### **End-to-End flow control**

Реализуется счётчиком EE\_Credit, который определяет максимум кадров, которые источник может отправить приёмнику без получения подтверждения (Acknowledge, ACK). В отличие от BB\_Credit распространяется только на фреймы с данными, а обмен/учёт происходит между конечными узлами.

## ISCSI-SAN

iSCSI строится на двух наиболее часто используемых протоколах:

SCSI — протоколе обмена блоками данных между компьютером и хранилищем

IP — сетевом транспортном протоколе, широко применяемом в корпоративных сетях Ethernet.

Сеть iSCSI SAN использует коммутацию Ethernet TCP/IP третьего уровня. Она сопоставляет протокол SCSI для хранения данных с TCP/IP и оборачивает его в пакеты Ethernet. По сути, iSCSI позволяет инициатору и объекту договариваться и обмениваться командами SCSI, используя известные и понятные сети TCP/IP. iSCSI SAN эмулирует прямое подключение устройств SCSI по локальной сети.

Подобно Fibre Channel, существует три основных аппаратных компонента iSCSI SAN. К ним относятся сетевые карты хранения iSCSI (специальные сетевые интерфейсные карты для обмена данными iSCSI) или инициаторы; стандартные сетевые коммутаторы Ethernet; и сетевые карты хранения или хранилища на внешних массивах хранения.

Сети iSCSI SAN функционально похожи на Fibre Channel, но вместо протокола Fibre Channel с требовательным детерминированным вручную протоколом второго уровня (то есть все соединения должны быть определены и отображены вручную заранее), iSCSI опирается на недетерминированный протокол TCP/IP с автоматическим обнаружением, маршрутизацией и коммутацией.

Адреса хостов и портов целевого хранилища основаны на TCP/IP. Существует три типа специальных форматов имен iSCSI для идентификации инициаторов и хранилищ. К ним относятся iSCSI qualified name (IQN), extended unique identifier (EUI) и T11 network address authority (NASA). Коммутаторы такие же, как и коммутаторы TCP/IP сети Ethernet.

Чтобы запутать все еще больше, порты инициатора и цели iSCSI также бывают трех видов. Первый, наиболее распространенный вариант реализации, использует программный инициатор, где вся обработка протокола осуществляется центральным процессором хоста. Второй вариант - разделенная стековая нагрузка TCP/IP, при которой около 80% стандартной обработки пакетов TCP/IP выполняется на сетевой карте хранилища. Только обработка исключений выполняется стеком TCP/IP хоста. Третий и последний вариант - полная выгрузка стека TCP/IP, при которой вся обработка протоколов TCP/IP выполняется сетевой картой. С ростом числа процессорных ядер потребность в механизме разгрузки TCP/IP (TOE) на базе сетевой карты резко снизилась. Стандартный программный инициатор iSCSI является, безусловно, наиболее распространенной реализацией iSCSI.

Прямого эквивалента iSCSI для коммутаторов класса Fibre Channel director не существует, даже несмотря на множество коммутаторов с большим количеством портов. Это связано с тем, что сети третьего уровня не требуют неблокируемости детерминированных сетей.

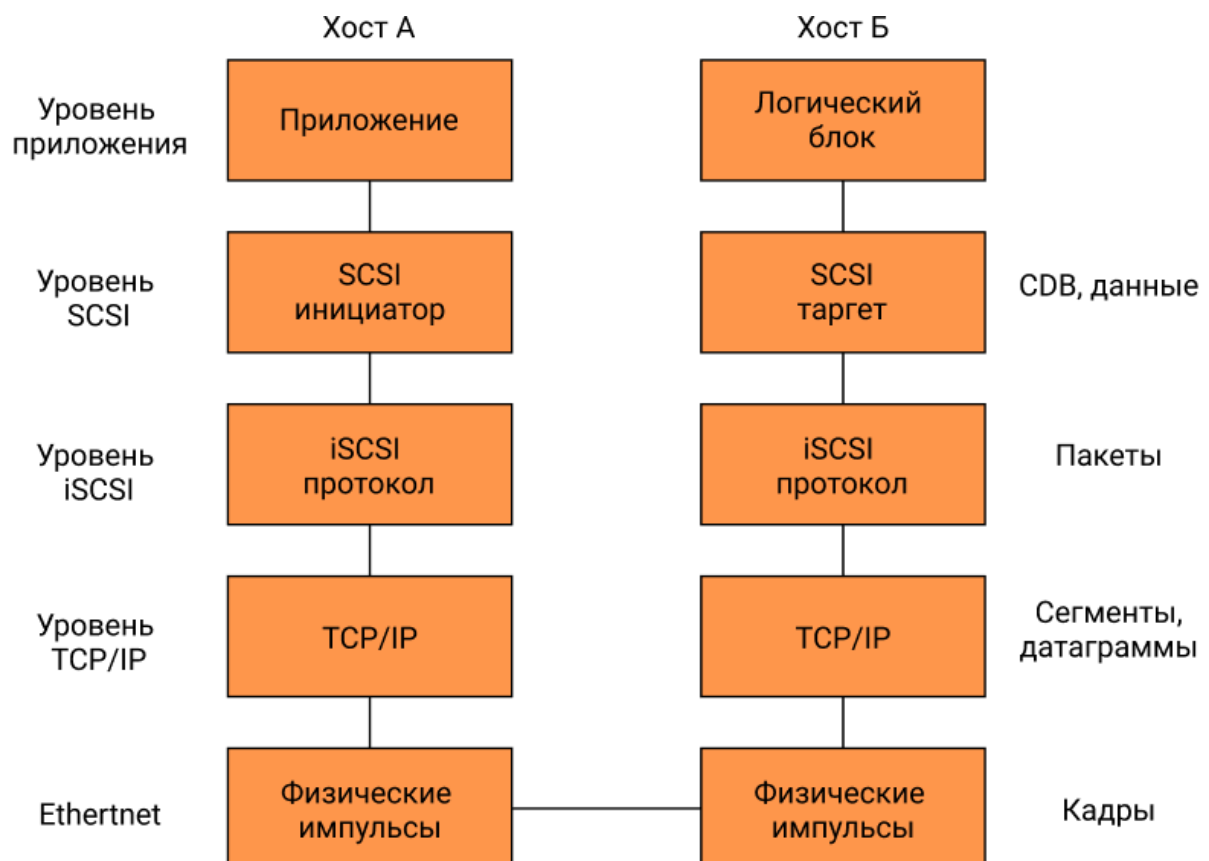
Пропускная способность портов в сети iSCSI SAN доступна в вариантах 1 Гбит/с и 10 Гбит/с, которые не являются совместимыми. Однако, поскольку iSCSI использует TCP/IP в Ethernet, несколько портов могут быть объединены для обеспечения виртуальной пропускной способности, которая намного больше, чем у отдельного порта.

Считается, что iSCSI 10Гбит обеспечивает такое же количество IOps и пропускную способность, как и FC 8Гбит, но это не совсем так. Хотя пропускная способность iSCSI и выше, но его эффективность ниже, чем у FC за счёт дополнительных накладных расходов.

Производительность iSCSI зависит от существующей инфраструктуры Ethernet (на сегодняшний день минимально рекомендованная сеть для iSCSI – 10Гбит). В ближайшем будущем стоит планировать переход на 25/40/50GbE, если будет необходимость использовать высокопроизводительные all-flash СХД.

Первоначально протокол SCSI использовался только для локального хранения данных, то есть для отдельных дисковых накопителей или систем хранения с прямым подключением (DAS). Затем, примерно в 1993 году, появился Fibre Channel, который позволил SCSI использовать протокол Fibre Channel Protocol (FCP) поверх сети Fibre Channel Storage Area Network (FC-SAN). iSCSI был представлен в качестве стандарта в 2000 году, и его популярность росла по мере того, как его поддерживало все больше операционных систем, сначала для этого требовались специальные iSCSI HBA, но позже стал использоваться программный инициатор iSCSI, который работал поверх любого типа сетевой карты Ethernet.

Специальные iSCSI HBA обеспечили iSCSI более высокую производительность, которая в то время была ближе к производительности Fibre Channel, а программный iSCSI инициатор позволил легко использовать iSCSI на многих серверах без покупки специальных HBA для каждого сервера. Вероятно, самым большим толчком к внедрению iSCSI стало появление в Microsoft Windows Server 2008 программного инициатора iSCSI.



SCSI создает блоки данных протокола (PDU), состоящие из команд SCSI, данных SCSI и ответов SCSI. PDU iSCSI вставляются в область данных сегмента TCP/IP-датаграмм. Затем IP-датаграммы перемещаются через стек TCP/IP и передаются по сети между уровнем служб SCSI хост-систем и уровнем служб SCSI целевых устройств хранения данных.

Протокол iSCSI легко работает в сетях TCP/IP, не требуя изменений в протоколах TCP/IP.

В исходящем направлении (инициатор – хранилище):

1. Уровень SCSI строит блоки дескрипторов команд(CDB) SCSI и передает их уровню iSCSI (вместе с остальными параметрами выполнения команды).
2. Уровень iSCSI создает PDU iSCSI и передает их одному или нескольким соединениям TCP.
3. TCP-соединения формируют "сессию" инициатор-хранилище (I\_T Nexus).

Во входящем направлении (Хранилище - Инициатор):

1. Уровень iSCSI получает PDU iSCSI на одном или нескольких TCP-соединениях в потоке TCP/IP.
2. Уровень iSCSI извлекает SCSI CDB из PDU iSCSI и передает их на уровень SCSI.

## Сравнение реализации FC SAN и iSCSI SAN. Цена и производительность

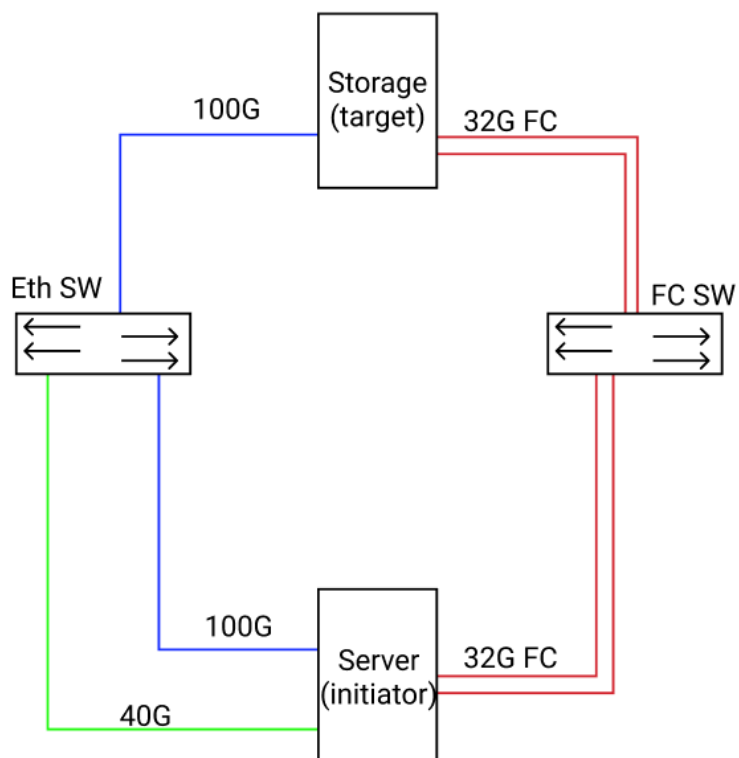
Пример оборудования с реальными ценами

Устройство	Кол-во портов	Скорость порта	Полная пропускная способность	Цена
Cisco Nexus 9200 with 48p 10/25 Gbps and 18p 100G QSFP28	48 & 18	25 & 100	6000	\$20,000.00
Cisco MDS 9148T Fibre Channel Switch (Port Side Intake)	48	32	1500	\$41,000.00
Dual Port 40G QSFP+ Network Card	1	40	40	\$500.00
Dual Port 100GbE QSFP56 Ethernet Adapter	1	100	100	\$1,000.00
32Gb Dual Port FC HBA	1	32	32	\$2,100.00

При создании сети учитываем, что для SAN сети важна отказоустойчивость, по этой причине каждое устройство продублируем.

Стоимость итогового решения минимальной SAN сети

Сеть для хранения данных	32G SAN	40G iSCSI	100G iSCSI
Коммутатор	2x32G FC SW	2x100G SW	2x100G SW
HBA / Адаптер	2x32G FC HBA	2x40G eth	2x100G eth
Общая цена	\$86,200.00	\$41,000.00	\$42,000.00



Топология

Итоги тестирования 32G FC SAN против 40G iSCSI SAN с размером блока равным 32 килобайта

Тест	Чтение(32К)	Запись(32К)
<b>iSCSI пропускная способность</b>	4832 МБ/с	4909 МБ/с
<b>FC пропускная способность</b>	2976 МБ/с	2951 МБ/с
<b>iSCSI IOPS</b>	147.0	150.0
<b>FC IOPS</b>	90.8	90.1

Итоги тестирования 64G(dual 32G) FC SAN против 100G iSCSI SAN с размером блока равным 32 килобайта

Тест	Чтение(32К)	Запись(32К)
<b>iSCSI пропускная способность</b>	10200 МБ/с	12100 МБ/с
<b>FC пропускная способность</b>	5163 МБ/с	4647 МБ/с
<b>iSCSI IOPS</b>	311.0	369.0
<b>FC IOPS</b>	158.0	142.0

Итоги тестирования 32G FC SAN против 40G iSCSI SAN с размером блока равным 4 килобайта

Тест	Чтение(4К)	Запись(4К)
<b>iSCSI пропускная способность</b>	784 МБ/с	813 МБ/с
<b>FC пропускная способность</b>	714 МБ/с	612 МБ/с
<b>iSCSI IOPS</b>	191.0	198.0
<b>FC IOPS</b>	174.0	149.0

Вопреки распространенному мнению, iSCSI вовсе не медленный. На самом деле, это одна из самых быстрых блочных сетей хранения данных на сегодняшний день. Хотя NGX Storage поддерживает в своих продуктах как FC, так и iSCSI, мы не можем принять чью-либо сторону. Поэтому наши клиенты должны прочитать и признать эту статью в блоге как технический обзор, который поможет понять цифры производительности сети хранения данных.





## **Источники**

IBM Redbook «Introduction to SAN and System Networking»

EMC «Network Storage Concepts and Protocols»

Brocade «SAN Fabric Administration Best Practices Guide»

<https://www.storagefreak.net/2015/11/fibre-channel-layers>

<https://www.sanog.org/resources/sanog8/sanog8-san-functional-overview-asimkhan.pdf>

<https://slideplayer.com/slide/13070828/>

<https://www.packetcoders.io/fibrechannel-sanprotocolsexplained/>

<https://www.snia.org/sites/default/files/news/iSCSI-Future-Cloud-Storage-Doomed-NVMe-oF.pdf>

[http://www.3kranger.com/HP3000/mpeix/en-hpux/T1452-](http://www.3kranger.com/HP3000/mpeix/en-hpux/T1452-90011/ch01s05.html?btnNext=next%A0%BB)

[90011/ch01s05.html?btnNext=next%A0%BB](http://www.3kranger.com/HP3000/mpeix/en-hpux/T1452-90011/ch01s05.html?btnNext=next%A0%BB)