3 I/O-Techniken

Sebastian Thaele · Achim Christ · Ulf Troppens

Ziel des Kapitels

Server erzeugen, verarbeiten und löschen Daten. Auch wenn neuere Techniken wie Objektspeicher (Kap. 6) und Hyperconverged Systems (Abschnitt 9.3.6) sich auf Server mit internem Speicher (Storage-Rich Server) zurückbesinnen, verwendet ein Großteil der heutigen (2018) Installationen weiterhin externe, verteilte und gemeinsam genutzte Speichergeräte. Server legen Daten auf externen Disk- und Flashsystemen (Kap. 2) ab oder lagern sie auf Tape Libraries (Kap. 7) aus. I/O-Techniken realisieren den Datenaustausch zwischen Servern und Speichergeräten. In diesem Kapitel beschreiben wir I/O-Techniken, die heute im Einsatz sind oder in den nächsten Jahren aus unserer Sicht sehr wahrscheinlich eingesetzt werden.

Gliederung des Kapitels

Dazu betrachten wir zunächst den I/O-Pfad von der CPU zum Speichergerät und SCSI (Abschnitt 3.1). SCSI war lange Zeit die vorherrschende I/O-Technik für Server. Die Grundfunktionen des SCSI-Protokolls wurden in neuen I/O-Techniken wie Fibre Channel (Abschnitte 3.2 und 3.3) und IP Storage (Abschnitt 3.5) übernommen. Des Weiteren beschreiben wir WAN-Techniken, um Server und Speichergeräte über größere Entfernungen miteinander zu verbinden (Abschnitt 3.4). Danach wenden wir uns ausgewählten weiteren I/O-Techniken zu wie InfiniBand, RDMA und NVMe (Abschnitt 3.6). Das Kapitel schließt wie alle Kapitel mit einer Zusammenfassung des Kapitels und dem Ausblick auf die nächsten Kapitel (Abschnitt 3.7).

3.1 Grundlagen

Gliederung des Abschnitts

In diesem Abschnitt beschäftigen wir uns zunächst mit dem I/O-Pfad von der CPU zum Speichergerät (Abschnitt 3.1.1). Dann stellen wir das Small Computer System Interface (SCSI) vor, dessen Protokoll auch heute für Speichernetze immer noch von großer Bedeutung ist (Abschnitt 3.1.2).

3.1.1 Der physische I/O-Pfad von der CPU zum Speichergerät

Der I/O-Pfad

Im Server bearbeiten eine oder mehrere CPUs Daten, die im CPU-Cache oder im Hauptspeicher (Random Access Memory, RAM) vorgehalten werden. CPU-Cache und Hauptspeicher sind sehr schnell; sie können die Daten aber nicht über eine Stromabschaltung hinweg speichern. Außerdem ist Hauptspeicher teuer im Vergleich zu Flashspeicher, Festplatten und Bändern. Deshalb werden Daten vom Hauptspeicher über Systembus, Hostbus und I/O-Bus zu Speichergeräten wie Disksystemen und Tape Libraries verlagert (Abb. 3–1): Speichergeräte sind zwar langsamer als CPU-Cache und Hauptspeicher, dafür sind sie billiger und sie können Daten über eine Stromabschaltung hinweg speichern. Zudem werden erst so die gemeinsame Nutzung des Speichers durch verschiedene Server (Abschnitt 8.2) sowie moderne Hochverfügbarkeits- und Disaster-Recovery-Mechanismen möglich (Kap. 12). Der gleiche I/O-Pfad findet sich übrigens auch innerhalb eines Disksystems zwischen den Anschlussports und dem Controller des Disksystems sowie dem Controller und den internen Laufwerken (Abb. 3–2).

Systembus

Im Herz des Servers sorgt der Systembus für die schnelle Datenübertragung zwischen CPUs und Hauptspeicher. Der Systembus muss sehr hochfrequent getaktet sein, damit er die CPU hinreichend schnell mit Daten versorgen kann. Er wird in Form von Leiterbahnen auf der Hauptplatine realisiert. Aufgrund physikalischer Eigenschaften erfordern hohe Systemtakte kurze Leiterbahnen. Deshalb ist der Systembus möglichst kurz, sodass er nur CPUs und Hauptspeicher miteinander verbindet.

Host-I/O-Bus

In heutigen Servern wird versucht, möglichst viele Aufgaben auf Spezialprozessoren wie Grafikprozessoren zu verlagern, um die CPU für die Verarbeitung der Anwendung zu entlasten. Diese können wegen der oben erwähnten physikalischen Einschränkungen nur bedingt an den Systembus angeschlossen werden. Deshalb realisieren die meisten Rechnerarchitekturen einen zweiten Bus, den sogenannten Host-I/O-Bus. Sogenannte Bridge Communication Chips stellen die Verbindung zwischen Systembus und Host-I/O-Bus her. Peripheral Component Interconnect (PCI) und seine Nachfolger sind heute die verbreitetste Technik für die Realisierung von Host-I/O-Bussen. Als Alternative wird zum Beispiel InfiniBand im High-Performance-Computing (HPC) (Abschnitt 3.6.1) eingesetzt.

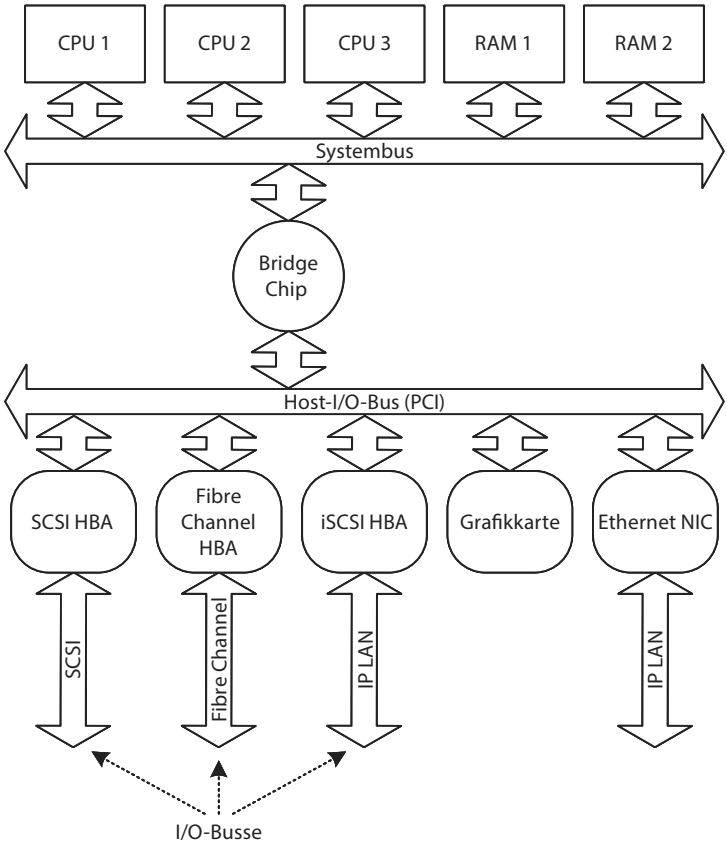


Abb. 3–1 Der physische I/O-Pfad von der CPU zum Speicher

Der physische I/O-Pfad von der CPU zum Speichergerät besteht aus Systembus, Host-I/O-Bus und I/O-Bus. Techniken wie InfiniBand, Fibre Channel und iSCSI ersetzen einzelne Busse durch ein serielles Netz. Aus historischen Gründen werden die entsprechenden Verbindungen dennoch als Host-I/O-Bus beziehungsweise als I/O-Bus bezeichnet.

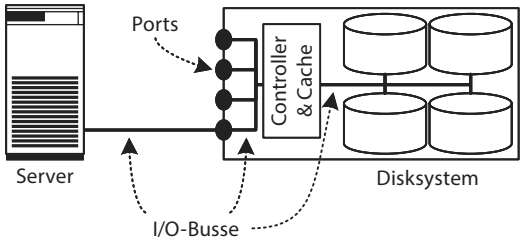


Abb. 3–2 Der physische I/O-Pfad innerhalb eines Disksystems

Innerhalb von Disksystemen werden die gleichen I/O-Techniken eingesetzt wie zwischen Disksystemen und Servern.

I/O-Bus, Gerätetreiber

Gerätetreiber (Device Driver) sind für die Steuerung von und die Kommunikation mit Peripheriegeräten aller Art zuständig. Der Gerätetreiber für Speichergeräte wird zum Teil in Software umgesetzt, die durch die CPU abgearbeitet wird. Für die Kommunikation mit Speichergeräten wird ein Teil des Gerätetreibers aber fast immer durch Firmware realisiert, die von Spezialprozessoren (Application Specific Integrated Circuits, ASICs) abgearbeitet wird. Diese ASICs werden heute zum Teil in die Hauptplatine integriert, wie zum Beispiel On-Board-SAS-Controller, oder über Zusatzkarten (PCI-Karten) mit der Hauptplatine verbunden. Diese Zusatzkarten werden meist als Hostbus-Adapter (HBA), als Netzwerkkarte (Network Interface Controller, NIC) oder einfach als Controller bezeichnet. Speichergeräte werden über den Hostbus-Adapter oder über die On-Board-Controller mit dem Server verbunden. Die Kommunikationsverbindung zwischen Controller und Peripheriegerät wird als I/O-Bus bezeichnet. Durch dieses sogenannte Offloading ist die Kommunikation mit dem Speichergerät von der Arbeit des Prozessors weitgehend entkoppelt. Auch die Fehlerbehandlung kann so parallel und gegebenenfalls ohne direkte Auswirkung auf den Hauptprozessor erfolgen.

Techniken für I/O-Busse

Die bis heute (2018) wichtigsten Techniken für I/O-Busse im Enterprise-Umfeld sind SCSI (Small Computer System Interface) und Nachfolger wie Serial Attached SCSI (SAS) sowie Fibre Channel. Das ursprüngliche SCSI-Protokoll definiert einen parallelen Bus, der bis zu 16 Server und Speichergeräte miteinander verbinden konnte. Dagegen definiert Fibre Channel verschiedene Topologien für Speichernetze, die Tausende Server und Speichergeräte miteinander verbinden können. Als Alternative zu Fibre Channel hat sich in vielen Bereichen die Möglichkeit etabliert, Speichernetze über TCP/IP und Ethernet zu realisieren (IP Storage). Bemerkenswert ist, dass die meisten aktuellen Techniken weiterhin das SCSI-Protokoll für die Kommunikation der Geräte einsetzen.

Weitere Techniken für I/O-Busse

Es sind zahlreiche andere Techniken für I/O-Busse auf dem Markt, die wir in diesem Buch nicht weiter behandeln, beispielsweise Serial Storage Architecture (SSA), IEEE 1394 (Apples Firewire, Sonys i.Link), High-Performance Parallel Interface (HIPPI), Advanced Technology Attachment (ATA)/Integrated Drive Electronics (IDE), Serial ATA (SATA), Serial Attached SCSI (SAS) und Universal Serial Bus (USB). Allen ist gemeinsam, dass sie entweder nur von sehr wenigen Herstellern verwendet werden oder für die Verbindung von Servern und Speichergeräten nicht leistungsfähig genug sind. Einige dieser Techniken können kleine Speichernetze bilden. Allerdings ist keine auch nur annähernd so flexibel und skalierbar wie die in diesem Buch beschrie-

benen Techniken Fibre Channel und IP Storage. Sie werden eher für die Vernetzung der Komponenten innerhalb von Speichersystemen oder als Peripherie-Schnittstelle im Endkundenbereich (Consumer) genutzt.

3.1.2 Small Computer System Interface (SCSI)

Small Computer System Interface (SCSI)

Das Small Computer System Interface (SCSI) war lange Zeit *die* Technik für I/O-Busse in Unix- und x86-Servern. SCSI umfasst die Übertragungstechnik (also SCSI-Kabel, SCSI-Stecker und SCSI-Hostbus-Adapterkarten) und das Kommunikationsprotokoll (Abb. 3–3). Die erste Version des SCSI-Standards wurde 1986 verabschiedet. Seitdem wird SCSI permanent vom T10-Gremium (*http://www.t10.org*) weiterentwickelt, um es an den technischen Fortschritt anzupassen. Ursprünglich konnte beispielsweise ein Server erst dann ein neues SCSI-Kommando absetzen, wenn das vorhergehende SCSI-Kommando von der Gegenstelle quittiert wurde; aber gerade die Überlappung von SCSI-Kommandos ist die Grundlage für die Leistungssteigerung durch RAID (Abschnitt 2.2.1). Heute ist es mit Asynchronous I/O üblich, an ein Speichergerät mehrere Schreib- oder Lesebefehle gleichzeitig abzusetzen.

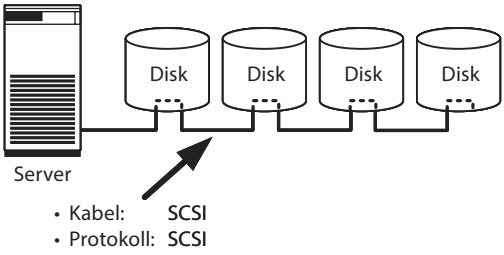


Abb. 3–3 Anschluss von Speicher über SCSI Daisy Chain
Ein SCSI-Bus verbindet mittels Daisy Chain einen Server mit mehreren Peripheriegeräten. SCSI definiert sowohl die Beschaffenheit der Verbindungskabel als auch das Übertragungsprotokoll.

SCSI-Kabel

SCSI als Übertragungstechnik wird heute nicht mehr genutzt. SCSI-Kabel und SCSI-Stecker wurden längst durch Fibre Channel, SAS und andere Übertragungstechniken ersetzt. Frühere Ausgaben dieses Buchs enthielten mehr Details über den physischen SCSI-Bus, seine Eigenschaften und seine Einschränkungen. Die obsolet gewordenen Textabschnitte haben wir in dieser Auflage aus dem Buch herausgenommen und für interessierte Leser auf unserer Webseite verfügbar gemacht (*www.speichernetze.com*).

SCSI-Protokoll

SCSI als Kommunikationsprotokoll ist aber immer noch sehr wichtig und wird dies vermutlich noch viele Jahre bleiben. Das SCSI-Protokoll definiert, wie die Geräte miteinander kommunizieren: Es legt fest, wie die Geräte die zugrunde liegende Übertragungstechnik (früher SCSI-Bus, heute beispielsweise Fibre Channel) nutzen und in welchem Format Daten übertragen werden. Der für heutige Speichernetze relevante Teil des SCSI-Protokolls behandelt das Format und die Abläufe der Kommunikation zwischen den Endgeräten, während alle Hardware-nahen Prozesse und Anforderungen in den unterliegenden Übertragungsprotokollen, wie Fibre Channel oder TCP/IP, definiert sind.

Initiator und Target

In der Kommunikation gemäß SCSI-Protokoll gibt es zwei grundlegende Rollen: den Initiator und das Target. Der Initiator ist meist ein Server und übernimmt den aktiven Teil der Kommunikation. Er beginnt diese und überwacht ihren Verlauf. Fehlererkennung und -behebung geht in aller Regel vom Initiator aus. Auch intelligente Disksysteme und andere spezielle Geräte können die Rolle eines Initiators übernehmen, wenn dies für die Bereitstellung bestimmter Dienste oder Mechanismen wie Remote Mirroring oder Speichervirtualisierung notwendig ist. Das Target ist eher passiv und in der Regel das Speichergerät. Es erwartet die Anfragen des Initiators und wird ansonsten nicht selbst aktiv. In der Kommunikation zwischen Initiator und Target ist es bis auf wenige Ausnahmen üblich, dass das Target die jeweilige Anfrage nach Abarbeitung derselbigen beendet.

SCSI ID und LUN

Teile des SCSI-Protokolls sind meist direkt im Betriebssystem oder als Gerätetreiber implementiert. Darum finden sich dort Begriffe wie SCSI ID, Target ID, Controller ID und LUN wieder (Abb. 3–4). Sie stehen für eine Betriebssystem-seitige Repräsentation der erreichbaren Speichergeräte oder beispielsweise der Fibre-Channel-Adresse von am Speichernetz angeschlossenen Geräten. Namensgebung, Bedeutung und gespeicherte Informationen variieren von Betriebssystem zu Betriebssystem. Letztendlich werden die in Abbildung 3–4. gezeigten Geräte und Teilgeräte repräsentiert.

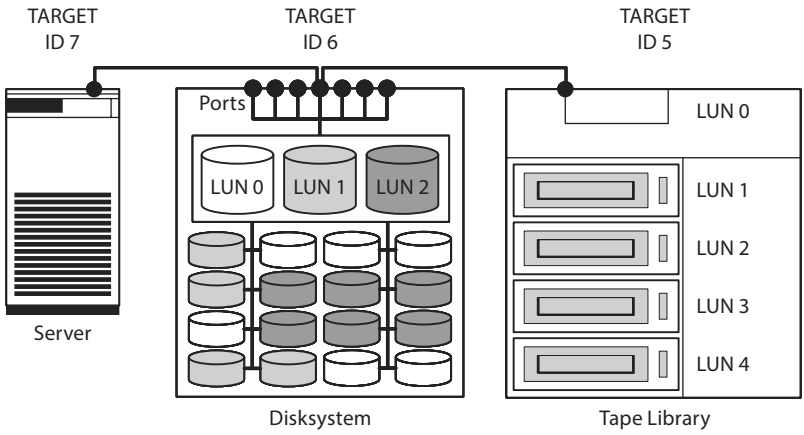


Abb. 3–4 SCSI Target IDs und SCSI LUNs

Geräte am SCSI-Bus werden durch Target IDs unterschieden. Komponenten innerhalb von Geräten (virtuelle Festplatten, Bandlaufwerke und Roboter in Tape Libraries) werden durch LUNs unterschieden. Diese Begriffe werden weiterhin verwendet.

3.2 Fibre Channel (FC)

Ziel des Abschnitts

Fibre Channel ist seit vielen Jahren die am häufigsten eingesetzte Technik für die Verwirklichung von Speichernetzen. Grundlegende Kenntnisse des Fibre-Channel-Standards helfen, die Einsatzmöglichkeiten von Fibre Channel für ein Fibre Channel SAN besser zu verstehen. In diesem Abschnitt erklären wir technische Details des Fibre-Channel-Protokolls. Dabei beschränken wir den Detaillierungsgrad auf die Teile des Fibre-Channel-Standards, die bei der Administration oder dem Entwurf eines Fibre Channel SAN hilfreich sind. Darauf aufbauend erklären wir in Abschnitt 3.3 den Einsatz von Fibre Channel für Speichernetze.

Der Fibre-Channel-Protokollturm

Der Fibre-Channel-Protokollturm untergliedert sich in fünf Schichten (Abb. 3–5). Die unteren vier Schichten, FC-0 bis FC-3, definieren die grundlegenden Kommunikationstechniken, also die physische Ebene, die Übertragung und die Adressierung. Die obere Schicht, FC-4, definiert, wie Anwendungsprotokolle (Upper Layer Protocols, ULPs) auf das zugrunde liegende Fibre-Channel-Netz abgebildet werden. Der Einsatz der verschiedenen ULPs entscheidet beispielsweise, ob ein reales Fibre-Channel-Netz als IP-Netz, als Fibre Channel SAN (also als Speichernetz) oder für beides zugleich eingesetzt wird. Quasi neben dem Fibre-Channel-Protokollturm stehen die Link Services und die

Fabric Services. Diese Dienste werden benötigt, um ein Fibre-Channel-Netz zu verwalten und zu betreiben. Im Folgenden stellen wir die verschiedenen Schichten und Dienste näher vor.

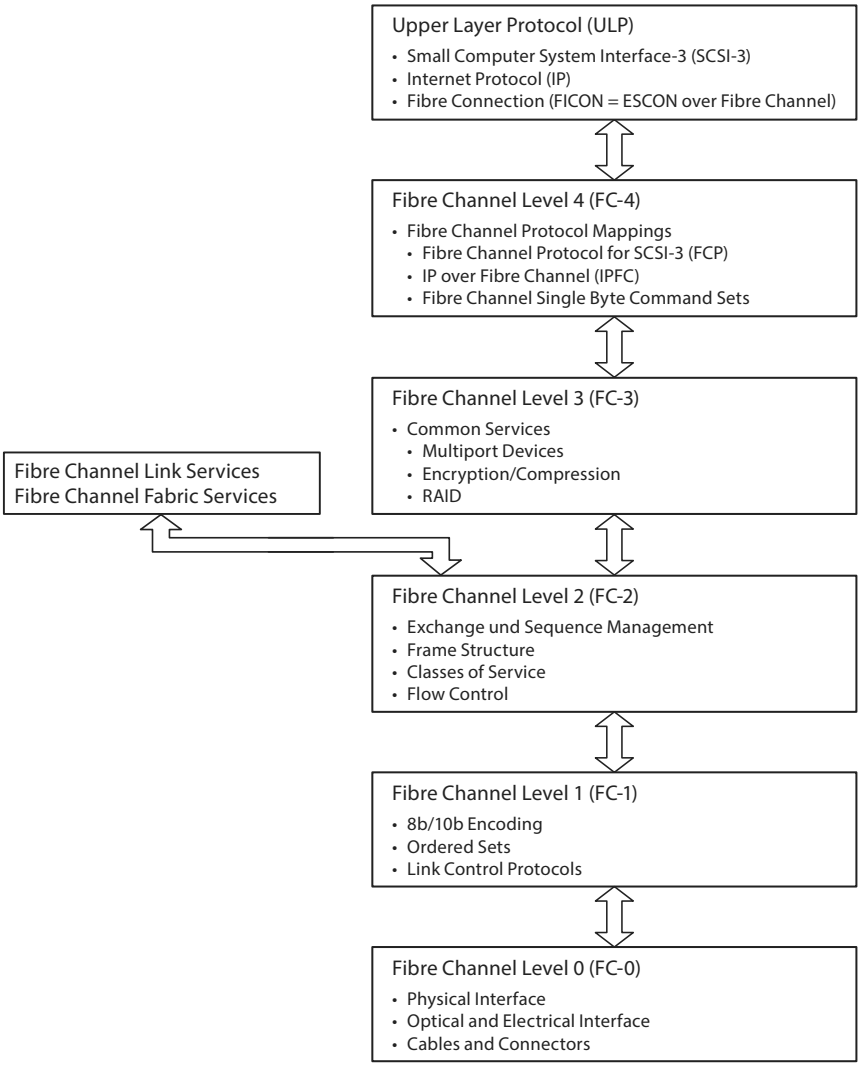


Abb. 3–5 Der Fibre-Channel-Protokollturm

Der Fibre-Channel-Protokollturm gliedert sich in zwei Teile: Die unteren vier Schichten (FC-0 bis FC-3) realisieren die grundlegende Fibre-Channel-Übertragungstechnik. Die Link Services und die Fabric Services helfen, das Fibre-Channel-Netz zu verwalten und zu konfigurieren. Darauf aufbauend definiert die obere Schicht (FC-4), wie die Anwendungsprotokolle (beispielsweise SCSI und IP) auf ein Fibre-Channel-Netz abgebildet werden.

3.2.1 Links, Ports und Topologien

Fibre-Channel-Topologien

Der Fibre-Channel-Standard definiert drei verschiedene Topologien: Fabric, Arbitrated Loop und Point-to-Point (Abb. 3–6). Point-to-Point definiert die bidirektionale Verbindung zwischen zwei Geräten. Arbitrated Loop definiert einen unidirektionalen Ring aus zwei oder mehr Geräten, in dem zu jedem Zeitpunkt aber immer nur ein Gerät mit einem anderen Gerät kommunizieren kann. Schließlich definiert Fabric ein Netz, in dem mehrere Geräte gleichzeitig mit voller Bandbreite Daten austauschen können. Eine Fabric erfordert grundsätzlich einen oder mehrere miteinander verbundene Fibre-Channel-Switches als Schaltzentrale zwischen den Endgeräten. Weiterhin erlaubt es der Standard, eine oder mehrere Arbitrated Loops an eine Fabric anzuschließen. In Speichernetzen wird heute (2018) fast ausschließlich nur noch die Fabric-Topologie eingesetzt, während die Anbindung von Arbitrated Loops an Fabrics mit der Fibre-Channel »Gen 5« (16 GBit/s Fibre-Channel) größtenteils nicht mehr unterstützt wird. Das Loop-Protokoll wird nunmehr fast ausschließlich in sehr kleinen IT-Umgebungen eingesetzt, um Speichergeräte direkt an einen Server anzuschließen. Im Folgenden legen wir darum mehr Gewicht auf die Fabric-Topologie als auf die beiden anderen Topologien.

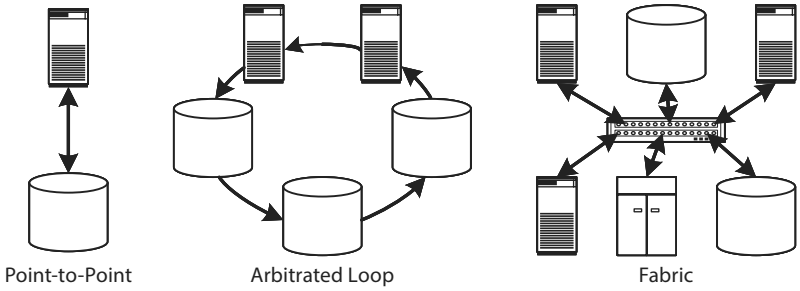


Abb. 3–6 Die drei Fibre-Channel-Topologien: Point-to-Point, Arbitrated Loop, Fabric
Die Fabric-Topologie ist die flexibelste, skalierbarste und am häufigsten eingesetzte Fibre-Channel-Topologie.

Ports

Allen Topologien ist gemeinsam, dass Geräte (Server, Speichergeräte und Switches) mit einem oder mehreren Fibre-Channel-Ports ausgestattet sein müssen. In Servern wird der Port in der Regel über sogenannte Hostbus-Adapter (HBAs, beispielsweise PCI-Karten) realisiert, die zusätzlich in den Server eingebaut werden. Ein Port besteht immer aus zwei Kanälen, einem Eingangs- und einem Ausgangskanal.

Links

Die Verbindung zwischen zwei Ports wird als Link bezeichnet. Die Links der Point-to-Point-Topologie und der Fabric-Topologie sind

bidirektional: Hier werden Eingangskanal und Ausgangskanal der beiden am Link beteiligten Ports über Kreuz miteinander verbunden, sodass jeweils ein Ausgangskanal mit einem Eingangskanal verbunden ist. Dagegen sind die Links der Arbitrated-Loop-Topologie unidirektional: Es wird jeweils der Ausgangskanal mit dem Eingangskanal des nächsten Ports verbunden, bis der Kreis geschlossen ist. Die Verkabelung einer Arbitrated Loop kann mithilfe eines Hubs vereinfacht werden. Dazu werden die Endgeräte bidirektional mit dem Hub verbunden. Die Verschaltung innerhalb des Hubs sorgt dafür, dass der unidirektionale Datenfluss innerhalb der Arbitrated Loop aufrechterhalten bleibt.

Porttypen:

N-Port

Die Topologien Fabric und Arbitrated Loop werden durch verschiedene inkompatible Protokolle realisiert. Ursprünglich wurde die Kommunikation von Fibre Channel um N-Ports und F-Ports herum entwickelt, wobei »N« für »Node« und »F« für »Fabric« steht. Ein N-Port (Node_Port) beschreibt die Fähigkeit eines Ports, als Endgerät (Server, Speichergerät), auch Knoten genannt, in der Fabric-Topologie oder als Partner in der Point-to-Point-Topologie teilzunehmen.

F-Port

F-Ports (Fabric_Port) sind im Fibre-Channel-Switch das Gegenstück zu einem N-Port. Ein F-Port weiß, wie er ein Frame (Abschnitt 3.2.4), das ein N-Port an ihn sendet, durch das Fibre-Channel-Netz an das gewünschte Endgerät weiterleiten kann. Neben dieser Weiterleitung von Ende-zu-Ende-Verkehr stellt ein F-Port einem N-Port weitere Dienste der Fabric zur Verfügung, die über vordefinierte Adressen erreicht werden können (Abschnitt 3.2.7).

L-Port

Die Arbitrated Loop verwendet andere Protokolle zum Datenaustausch als die Fabric. Ein L-Port (Loop_Port) beschreibt die Fähigkeit eines Ports, als Endgerät (Server, Speichergerät) in der Arbitrated-Loop-Topologie teilzunehmen. Neuere Geräte werden nicht mehr mit L-Ports ausgestattet, sondern mit NL-Ports oder nur noch mit N-Ports. Geräte, die ausschließlich mit L-Ports bestückt sind, kommen in modernen Rechenzentren heute (2018) nicht mehr vor.

NL-Port

Ein NL-Port (Node_Loop_Port) hat sowohl die Fähigkeiten eines N-Ports als auch die eines L-Ports. Ein NL-Port kann also sowohl in eine Fabric als auch in eine Arbitrated Loop eingebunden werden. Einige Hostbus-Adapter sind noch mit solchen Ports ausgestattet, hauptsächlich um an einen Server dedizierte Bandlaufwerke direkt anzuschließen. Einige Hersteller verzichten mittlerweile vollständig auf das Loop-Protokoll, sodass sie Hostbus-Adapter nur noch mit N-Ports anbieten.

FL-Port

Ein FL-Port (Fabric_Loop_Port) ermöglicht es, eine Fabric mit einer Loop zu verbinden. Das heißt aber noch lange nicht, dass Endgeräte in

der Arbitrated Loop mit Endgeräten in der Fabric kommunizieren können. Mehr zum Thema »Verbindung von Fabric und Arbitrated Loop« findet sich im Abschnitt 3.3.5. Hersteller von Fibre-Channel-Switches verzichten, beginnend mit der Generation »Gen 5« (16-GBit/s) des Fibre-Channel-Protokolls, mehr und mehr auf die Implementierung von FL-Ports. Aktuelle Switches erlauben die Anbindung von Arbitrated Loops nur noch über Umwege.

E-Port, Inter Switch Link (ISL)

Zwei Fibre-Channel-Switches werden über E-Ports (Expansion_Port) miteinander verbunden. E-Ports übertragen die Daten von Endgeräten, die an zwei verschiedene Fibre-Channel-Switches angeschlossen sind. Zusätzlich gleichen Fibre-Channel-Switches Informationen über das gesamte Fibre-Channel-Netz über E-Ports ab. Die Verbindung zwischen den E-Ports zweier Switches heißt dementsprechend Inter-Switch Link (ISL).

G-Port

Moderne Fibre-Channel-Switches konfigurieren ihre Ports automatisch. Solche Ports werden G-Ports (Generic_Port) genannt. Ein G-Port selbst ist nur ein temporärer Zustand während der Login-Phase (Abschnitt 3.2.6). Wenn sich ein N-Port eines Endgeräts in einen G-Port einloggt, so wird er zu einem F-Port. Wird ein anderer Switch an einen G-Port angeschlossen, so konfiguriert er sich als E-Port. Ein G-Port kann jedoch kein L-Port werden. Die Anbindung einer Arbitrated Loop ist somit nicht möglich.

Weitere Porttypen

Der Fibre-Channel-Standard definiert weitere Porttypen, die jedoch in heutigen Speichernetzen praktisch keine Verwendung finden. Des Weiteren verfügen heutige Fibre-Channel-Switches über weitere, herstellerspezifische Porttypen. Diese Porttypen stellen zusätzliche Funktionen bereit wie das Spiegeln des Datenverkehrs eines anderen Ports oder Diagnosemöglichkeiten.

3.2.2 FC-0: Kabel, Stecker und Signalcodierung

Problem: Laufzeitverzögerungen in Bussen (Skew)

FC-0 definiert das physische Übertragungsmedium (Kabel, Stecker) und spezifiziert, mit welchen physikalischen Signalen die Bits »0« und »1« übertragen werden. Im Gegensatz zum SCSI-Bus, der für jedes Bit eine eigene Datenleitung und zusätzliche Kontrollleitungen hat, überträgt Fibre Channel die Bits nacheinander über eine einzige Leitung. Allgemein kämpfen Busse mit dem Problem, dass die Signale auf den verschiedenen Datenleitungen eine unterschiedliche Laufzeit haben (Skew), sodass in Bussen die Taktrate nur begrenzt erhöht werden kann. Die unterschiedliche Signallaufzeit kann man sich wie das Handlaufband einer Rolltreppe vorstellen, das schneller oder langsamer läuft als die Rolltreppe selbst.

Lösung: serielle Übertragung

Fibre Channel überträgt deswegen die Bits seriell. Damit ist im Gegensatz zum parallelen Bus auch über weite Entfernungen eine hohe Übertragungsrate möglich. Die hohe Übertragungsrate der seriellen Übertragung gleicht die parallelen Leitungen eines Busses mehr als aus.

Übertragungsraten

Alle paar Jahre erscheinen Produkte mit einer höheren Übertragungsrate (Tab. 3–7). Heute (2018) sind von einigen Herstellern bereits Switches und Hostbus-Adapter mit 32 GBit/s erhältlich. Bei Speichersystemen setzt sich die nächsthöhere Geschwindigkeit meist erst nach mehreren Monaten bis Jahren durch. Der Großteil der aktuell laufenden Speicherumgebungen operiert mit 8 GBit/s oder schon mit 16 GBit/s. Daneben sind jedoch auch noch Altgeräte mit 4 GBit/s und in Ausnahmefällen sogar noch darunter im Einsatz. Darüber hinaus sind für die Switch-zu-Switch-Verbindung gerade über lange Strecken 10 GBit/s möglich, was einen einfacheren Anschluss an bestehendes 10-GBit/s-Equipment erlaubt. Die Fibre Channel »Gen 6« definiert zudem einen Standard für eine Quad-Verbindung aus vier gekoppelten 32-GBit/s-Faserpaaren, um so einen kombinierten Link mit 128 GBit/s zur Verfügung zu stellen. Bei der Übertragungsrate ist zu beachten, dass bei der Fabric-Topologie die Übertragung bidirektional und vollduplex ist, sodass in jede Richtung die Übertragungsrate von beispielsweise 32 GBit/s zur Verfügung steht.

Fibre-Channel-Variante	Bidirektionaler Durchsatz (MByte/s)	Übertragungsrate pro Leitung (GBit/s)	Veröffentlichungsjahr
1GFC	200	1,0625	1996
2GFC	400	2,125	2000
4GFC	800	4,25	2003
8GFC	1600	8,5	2006
16GFC	3200	14,025	2009
32GFC	6400	28,05	2013
64GFC	12800	57,8	2017
128GFC	25600	4×28,05	2014
256GFC	51200	4×57,8	2017

Tab. 3–7 Fibre Channel-Übertragungsraten
Neuere Fibre-Channel-Generationen unterstützen immer höhere Übertragungsraten und Durchsätze.

Kabel- und Steckertypen

Fibre Channel definiert verschiedene Kabeltypen (Tab. 3–8) für Kupfer (Copper) und Glasfaser (Fiber Optic). Sowohl für Kupferkabel als auch für Glasfaserkabel sind verschiedene Steckertypen definiert. Abbildung 3–9 zeigt verschiedene Steckertypen für Glasfaserkabel. Die verschiedenen Typen bringen bis auf unterschiedliche Abmessungen keine technischen Vorteile.

Medium	1 GBit/s	2 GBit/s	4 GBit/s	8 GBit/s	16 GBit/s	32 GBit/s
Kupfer intracabinet	24 m					
Kupfer intercabinet	59 m					
MM 62,5 µm OM1	300 m	150 m	70 m	21 m	15 m	
MM 50 µm OM2	500 m	300 m	150 m	50 m	35 m	20 m
MM 50 µm OM3	860 m	500 m	380 m	150 m	100 m	70 m
MM 50 µm OM4			400 m	190 m	125 m	100 m
SM 9µm OS1, OS2	10 km	10 km	10 km	10 km	10 km	10 km

Tab. 3–8 **Fibre Channel-Kabeltypen und -längen**
Fibre Channel definiert verschiedene Kabeltypen. Je nach benötigter Länge der Verbindung kann der jeweils billigste Kabeltyp ausgewählt werden. Heute (2018) unterstützen aktuelle Produkte eine Übertragungsrate von 32 GBit/s (siehe auch die Bemerkungen im Text bezüglich der unterstützten Entfernungen).

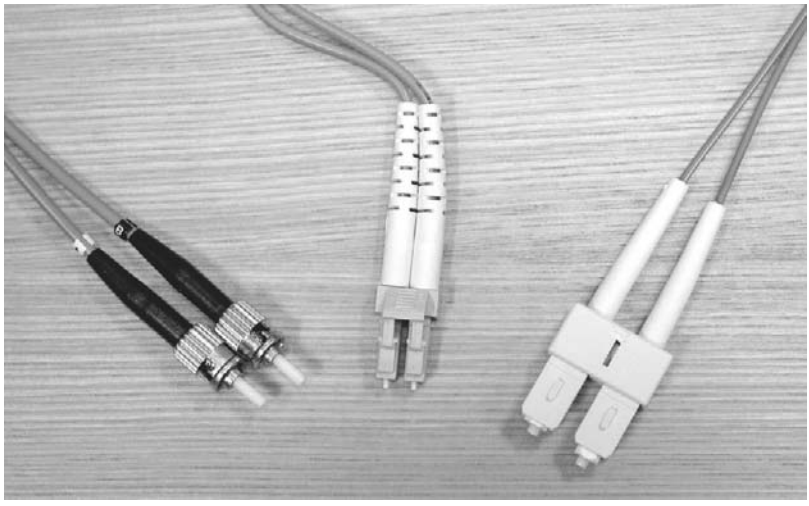


Abb. 3–9 **Drei verschiedene Steckertypen für Glasfaserkabel**

Kupferkabel

Kupferkabel werden unterschieden in »Intracabinet«-Kabel und »Intercabinet«-Kabel. Intracabinet-Kabel sind nur für die Verkabelung innerhalb eines Gehäuses gedacht. Sie sind schlechter abgeschirmt gegen elektromagnetische Störungen, wodurch sie billiger sind als Intercabinet-Kabel, mit denen Geräte über Gehäusegrenzen hinweg verkabelt werden können.

Glasfaserkabel

Glasfaserkabel sind teurer als Kupferkabel. Dafür haben sie einige Vorteile:

- größere Distanzen als mit Kupferkabeln
- Unempfindlichkeit gegenüber elektromagnetischen Störungen
- keine elektromagnetische Abstrahlung
- keine elektrische Verbindung zwischen den Geräten
- keine Gefahr des »cross talkings«

Single-Mode und Multi-Mode

Auch für Glasfaser sind verschiedene Kabel- und Steckertypen definiert: Multi-Mode-Kabel (Multi Mode Fiber, MMF) haben einen Kerndurchmesser von 62,5 µm oder 50 µm und Single-Mode-Kabel (Single Mode Fibre, SMF) haben einen Kerndurchmesser von 9 µm. Beide Kabeltypen werden mit unterschiedlichen Wellenlängen betrieben: Multi-Mode-Kabel benötigen eine Wellenlänge von 850 nm (Shortwave) und für Single-Mode Kabel werden meist Wellenlängen von 1310 nm oder 1550 nm verwendet (Longwave). Multi-Mode-Kabel und Shortwave-LEDs sind günstiger, können jedoch aufgrund physikalischer Effekte (Modendispersion) nicht so weite Strecken überbrücken wie die erheblich teureren Single-Mode-Kabel mit den ebenfalls teureren Longwave-Lasern.

Anforderungen: freie Wahl der Kabeltypten

Mit der Definition verschiedener Kabel und Lichtquellen können unterschiedliche Entfernungen mit der jeweils wirtschaftlichsten Technik überbrückt werden. Kabeltypen und die jeweils verwendeten Lichtquellen und -sensoren müssen zueinander passen. Ein Mix nicht passender Typen führt zu einer großen Anzahl an Übertragungsfehlern. Jedoch müssen für eine kostengünstige Lösung je nach Gegebenheit Server, Speichergeräte und Switches mit unterschiedlichen Kabeln verbunden werden. Deshalb müssen die Anschlussports von Switches, Hostbus-Adaptern und Speichergeräten je nach Situation mit unterschiedlichen Lichtquellen und -sensoren bestückt sein.

Lösung: Small Form-factor Pluggable (SFP)

Sogenannte Small Form-factor Pluggables (SFPs) erlauben es, Anschlussports mit den jeweils benötigten Lichtquellen- und -sensoren zu bestücken. SFPs sind standardisierte Module für Kabelanschlüsse, die flexibel getauscht werden können (Abb. 3–10). Mit SFPs können die gleichen Ports mit unterschiedlichen und somit den kostengünstigsten Kabeln betrieben werden.

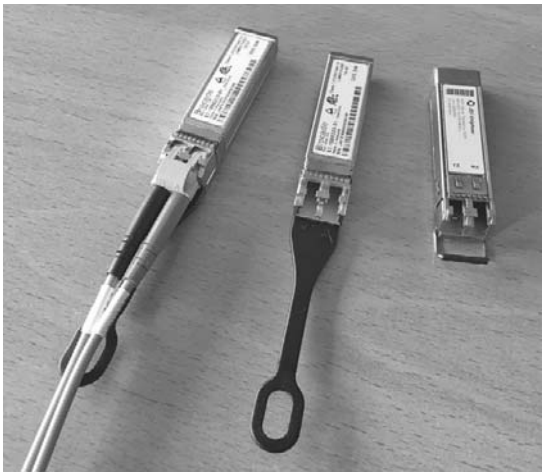


Abb. 3–10 Small Form-factor Pluggable (SFPs)
SFPs sind standardisierte Module für Kabelanschlüsse. Sie erlauben es, Anschlussports an Switches, Servern und Speichergeräten mit verschiedenen Kabeltypen zu betreiben.

Bitfehlerrate

Der Fibre-Channel-Standard fordert von allen Medien, dass ein einzelner Bitfehler höchstens einmal pro 10^{12} übertragenen Bits auftreten darf. Bei Fibre-Channel-Netzen mit 1 GBit/s oder 2 GBit/s konnten die Fehlererkennungs- und -behandlungsmechanismen der höheren Protokollschichten (wie SCSI) auftretende Bitfehler noch effektiv behandeln. Jedoch verkürzen sich mit höheren Geschwindigkeiten die Zeitabstände zwischen zwei Fehlern entsprechend. Beispielsweise würde bei einer 16-GBit/s-Verbindung unter Volllast pro Link alle 01:11 Minuten ein Bitfehler auftreten dürfen. In der Praxis hätte ein Link mit einer wie im Standard geforderten maximalen Bitfehlerrate von 10^{-12} einen erheblichen Einfluss auf die Leistung der betreffenden Endgeräte. Deshalb streben die Hersteller von Fibre-Channel-Switches und -Equipment heute (2018) Bitfehlerraten von 10^{15} und besser an, was bei 16 GBit/s immer noch einen Bitfehler alle 19,8 Stunden bedeuten kann. Daher ist bei der Installation eines Fibre-Channel-Netzes geboten, die Kabel sachgemäß zu verlegen, sodass das Signal nicht durch zu enge Biegeradien oder staubige Kontakte verschlechtert wird.

Fehlerrate versus Kabellänge

Die Entfernungsangaben in Tabelle 3–8 geben Mindestentfernungen an, mit denen die Fehlerrate beim Stand der Technik zum Zeitpunkt der Verabschiedung des entsprechenden Standards bei sachgerechter Verlegung der Kabel sicher unterschritten werden kann. Durch technische Verbesserungen und sachgerechte Verlegung der Kabel ist

es möglich, dass in konkreten Installationen auch größere Entfernungen überbrückt werden können. SFPs mit höherer Sendeleistung erlauben die Überbrückung weiterer Strecken.

Problem: Verkürzung der unterstützten Entfernungen

Werden die bisher genutzten Kabel beim Kauf neuer Fibre-Channel-Geräte weiterverwendet, kann sich durch die höheren Geschwindigkeiten die Situation ergeben, dass die im Standard definierten Strecken nicht mehr eingehalten werden können. Da es sich bei den Angaben um Mindestentfernungen handelt, kann es durchaus sein, dass beim Großteil der Verbindungen keine Probleme auftreten trotz Überschreiten der im Standard angegebenen Mindestlängen. Es ist jedoch davon abzuraten, beispielsweise eine bisher mit 2 GBit/s genutzte Verbindung zwischen zwei Gebäuden über ein OM2-Kabel von 250 m Länge mit 8 GBit/s weiter zu verwenden.

3.2.3 FC-1: Codierungen, Ordered Set und Link Control Protocol

FC-1 im Überblick

FC-1 definiert, wie Daten codiert werden, bevor sie über Fibre-Channel-Kabel übertragen werden. Weiter beschreibt FC-1 bestimmte Übertragungswörter (Ordered Sets), die zur Verwaltung einer Fibre-Channel-Verbindung benötigt werden (Link Control Protocol).

8b/10b-Codierung

Taktsynchronisation erforderlich

Bei allen Übertragungstechniken müssen Sender und Empfänger ihre Taktraten synchronisieren. In parallelen Bussen wird dazu der Bustakt über eine zusätzliche Kontrollleitung übertragen. Im Gegensatz dazu steht bei der seriellen Übertragung wie bei Fibre Channel nur die eine Leitung zur Verfügung, über die die Daten übertragen werden. Das heißt, der Empfänger muss aus dem Datenstrom den Übertragungstakt regenerieren.

Signalwechsel erforderlich

Der Empfänger kann den Takt nur an den Stellen synchronisieren, bei denen auf dem Medium ein Signalwechsel anliegt. Bei der einfachen Binärcodierung (Abb. 3–11) ist dies nur dann der Fall, wenn das Signal von »0« auf »1« oder von »1« auf »0« wechselt. Bei der Manchester-Codierung findet bei jedem übertragenen Bit ein Signalwechsel statt. Die Manchester-Codierung legt also für jedes übertragene Bit zwei physikalische Signale an. Sie benötigt deshalb eine doppelt so hohe Übertragungsrate wie die Binärcodierung. Deswegen setzt Fibre Channel wie viele andere Übertragungstechniken die Binärcodierung ein, weil sie mit der gleichen Rate an Signalwechseln mehr Bits übertragen kann als die Manchester-Codierung.

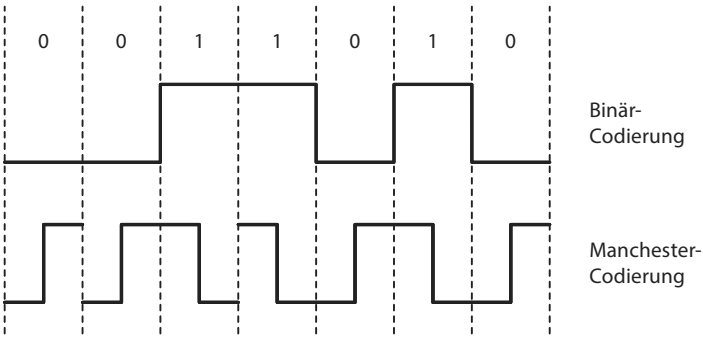


Abb. 3–11 NRZ- und Manchester-Codierung
Bei der Manchester-Codierung erfolgt mit jedem übertragenen Bit mindestens ein Signalwechsel.

Problem: Jitter

Das Problem dabei: Die Signalschritte, die beim Empfänger ankommen, sind nicht immer gleich lang (Jitter). Das heißt, beim Empfänger liegt ein Signal mal etwas länger an, mal etwas kürzer (Abb. 3–12). Im Bild der Rolltreppe bedeutet dies, dass die Rolltreppe ruckelt. Jitter kann dazu führen, dass der Empfänger die Synchronisation mit dem empfangenen Signal verliert. Schickt der Sender beispielsweise eine Folge von zehn Nullen, so kann der Empfänger nicht entscheiden, ob es sich um eine Folge von neun, zehn oder elf Nullen handelt.

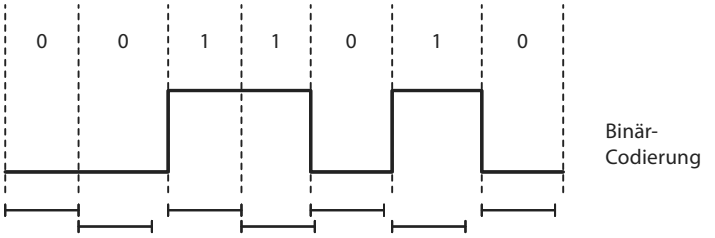


Abb. 3–12 Jitter
Aufgrund physikalischer Eigenschaften liegen beim Empfänger die Signale nicht immer gleich lange an.

Lösung: 8b/10b-Codierung

Will man trotzdem die Binärcodierung einsetzen, so muss man dafür sorgen, dass der Datenstrom oft genug einen Signalwechsel erzeugt, damit der Jitter nicht zuschlagen kann. Ein guter Kompromiss ist die sogenannte 8b/10b-Codierung. Die 8b/10b-Codierung wandelt ein zu übertragendes Achtbit-Byte in ein Zehnbit-Symbol um, das anstelle des Achtbit-Bytes über das Medium gesendet wird. Für Fibre Channel heißt dies beispielsweise, dass eine Nutzübertragungsrate von 800 MByte/s auf dem Medium nicht etwa eine Rohübertragungsrate von 6,4 GBit/s,

sondern vielmehr 8 GBit/s benötigt. Die 8b/10b-Codierung wird und wurde bei vielen anderen Übertragungstechniken, wie Enterprise System Connection Architecture (ESCON), Serial Storage Architecture (SSA), Gigabit-Ethernet, USB 3.0, SATA, SAS und InfiniBand eingesetzt.

Vorteile der 8b/10b-Codierung: Häufiger Signalwechsel

Durch die Erweiterung der Achtbit-Daten-Bytes auf Zehnbit-Übertragungssymbole werden unter anderem folgende Vorteile erreicht: Aus allen verfügbaren Zehnbit-Symbolen werden für die 8b/10b-Codierung nur solche Zehnbit-Symbole ausgewählt, die bei beliebiger Kombination eine Bitfolge erzeugen, die maximal fünf aufeinanderfolgende Nullen und maximal fünf aufeinanderfolgende Einsen enthält. Es erfolgt also spätestens nach fünf Signalschritten ein Signalwechsel, sodass die Taktsynchronisation des Empfängers gewährleistet ist.

Geringer Gleichstromanteil

Eine mit 8b/10b-Codierung erzeugte Bitfolge hat eine gleichmäßige Verteilung von Nullen und Einsen. Dies hat den Vorteil, dass in der Hardware, die 8b/10b-codierte Bitfolgen verarbeitet, nur geringe Gleichströme fließen. Dies vereinfacht und verbilligt die Realisierung von Fibre-Channel-Hardwarekomponenten. Die 8b/10b-Codierung verhindert ein Missverhältnis von Nullen und Einsen dadurch, dass für Achtbit-Bytes mit einer unterschiedlichen Anzahl Nullen und Einsen zwei verschiedene Versionen von Zehnbit-Symbolen zur Verfügung stehen: eines mit mehr Einsen und eines mit mehr Nullen. Die 8b/10b-Codierung stellt über die sogenannte Running Disparity sicher, dass zu jedem Zeitpunkt das Verhältnis von Einsen und Nullen entweder ausgeglichen oder der Überschuss höchstens 1 ist. Jedem Zehnbit-Zeichen wird mitgegeben, ob es derzeit eine negative oder positive Disparity gibt. Die Auswahl der passenden Version des nächsten Zehnbit-Symbols stellt sicher, dass die Regeln weiter eingehalten werden.

Zusätzliche Zeichen

Es stehen weitere Zehnbit-Symbole zur Verfügung, die keine Achtbit-Daten-Bytes repräsentieren. Diese sogenannten K-Words werden für die Verwaltung eines Fibre-Channel-Links, zum Beispiel zur Erkennung von Symbolgrenzen, verwendet.

Fehlererkennung

Es werden nicht alle Zehnbit-Symbole für die Codierung der Achtbit-Bytes oder als K-Words benötigt. Die übrigen Kombinationen sind nicht definiert und damit illegal. Passiert ein Bitfehler, besteht die Chance, dass dadurch ein illegales Zehnbit-Symbol entsteht. Auch eine Verletzung der Running Disparity ist ein starker Hinweis auf einen Bitfehler. In der Praxis hat sich herausgestellt, dass dadurch Bitfehler im Datenstrom sehr zuverlässig erkannt werden können. Trotzdem ist diese Methode nicht 100 % sicher, da bestimmte Kombinationen von Bitfehlern legale Zehnbit-Symbole erzeugen können. Um Fehler zumindest in Frames (Abschnitt 3.2.4) zweifelsfrei erkennen zu können, sind deshalb weitere Maßnahmen nötig wie Prüfsummen der Nutzlast.

Nachteile der 8b/10b-Codierung: Hoher Overhead

Diese Vorteile werden durch einen recht hohen Overhead erkauft: Durch die Verwendung von 10 statt 8 Bit pro Byte ergibt sich bereits in der Codierung ein Overhead von 25 % bezogen auf das Ausgangs-Byte. Ein Viertel der Daten wird also nur übertragen, um die oben genannten Aufgaben zu erfüllen. Bei elektrisch getrennten, da optisch verbundenen Fibre-Channel-Geräten spielt ein geringer Gleichstromanteil nicht mehr dieselbe kritische Rolle wie bei der Übertragung über Kupfer. Zudem kann Jitter auch bekämpft werden, wenn zwischen den regelmäßigen Signalwechseln mehr als fünf Zeichen liegen. Auch für die Fehlererkennung und die Markierung von Symbolgrenzen gibt es inzwischen effizientere Methoden. Bereits in der Sondervariante des 10-GBit/s-Fibre-Channels wurde deshalb eine Alternative verwendet: 64b/66b-Codierung.

64b/66b- und 256b/257b-Codierung

64b/66b-Codierung

Die 64b/66b-Codierung wird nicht auf jedes Byte einzeln, sondern wie der Name vermuten lässt, auf 64-Bit-Blöcke angewandt. Jedem dieser Blöcke werden zwei Bits vorangestellt. Von den vier Kombinationen, die diese erlauben (00, 01, 10, 11), sind jedoch nur 01 und 10 erlaubt. Weil hier ein Wechsel von 0 zu 1 oder von 1 zu 0 stattfindet, ist unabhängig vom Eingangssignal ein fester Signalwechsel alle 66 Bits gewährleistet. Sender und Empfänger bleiben so synchron und die Länge von Nullen und Einsen für beide Teilnehmer gleich. Die Funktion der K-Words übernehmen die beiden legalen Bitkombinationen. Sind sie 01, folgt ihnen Nutzlast. Sind sie jedoch 10, ist ihnen ein 8 Bit langes Type-Feld angehängt, dessen Inhalt darüber entscheidet, ob die restlichen 56 Bits des Blocks Kontrollinformationen oder weitere Nutzlast enthalten. Die Kontroll- und Nutzdaten laufen durch einen selbstsynchronisierenden Scrambler, der die Bits in einer bestimmten Weise verwürfelt, um zu gewährleisten, dass im Übertragungssignal zumindest statistisch weiterhin ein geringer Gleichstromanteil vorliegt. Er lässt ein höheres Ungleichgewicht an Nullen und Einsen zu als der deterministische Ansatz der 8b/10b-Codierung, ist aber für optisch gekoppelte Fibre-Channel-Geräte völlig ausreichend.

Nachteil der 64b/66b-Codierung

Einzig in der Fehlererkennung ergeben sich gegenüber der 8b/10b-Codierung Nachteile. So kann es vorkommen, dass durch den selbstsynchronisierenden Scrambler einzelne Bitfehler entweder nicht erkannt oder mehrfach gemeldet werden, da er quasi mit falschen Zwischenergebnissen weiterrechnet. Abhilfe schafft hierfür neben der CRC-Prüfsumme für Fibre-Channel-Frames (Abschnitt 3.2.4) die sogenannte Forward Error Correction (FEC), die mit Fibre Channel Gen 5 optional eingeführt und ab Gen 6 für 32 GBit/s und höher Pflicht ist.

Forward Error Correction (FEC)

Forward Error Correction (FEC) ist eine Technik, um Bitfehler in einem Datenstrom zu erkennen und zu korrigieren, ohne dass Daten neu übertragen werden müssen. Dazu fügt der Sender den zu übertragenden Daten speziell berechnete, redundante Daten bei, mit denen der Empfänger die eingehenden Daten überprüft und gegebenenfalls korrigiert. In einem Fibre Channel SAN sind normalerweise nur sehr wenige einzelne oder je nach Fehlerursache sehr kurze Bursts von Bitfehlern zu beobachten. Deshalb ist die Forward Error Correction von Fibre Channel darauf ausgelegt, in einem Fibre-Channel-Frame von 2112 Byte Länge maximal elf aufeinanderfolgende Bitfehler zu korrigieren. Sind diese Bitfehler quer über das Frame verteilt, so ist der Algorithmus nicht in der Lage, diese zu korrigieren. Dennoch kann Forward Error Correction in der Mehrheit der Fälle eine langwierige Error Recovery in höheren Protokollebenen verhindern. Die Unterstützung der Forward Error Correction ist in Fibre Channel Gen 5 (16GFC) optional und in späteren Generationen Pflicht. Sie entschärft damit das mit höheren Übertragungsraten immer dringender werdende Problem der zunehmenden Häufigkeit von Fehlern bei der im Standard konstant gebliebenen maximalen Bitfehlerrate von 10^{-12}.

256b/257b-Codierung

Die 256b/257b-Codierung ist eine Weiterentwicklung der 64b/66b-Codierung. Wie diese arbeitet 256b/257b im Gegensatz zur 8b/10b-Codierung ohne feste Übersetzungstabelle. Die 256b/257b-Codierung stellt jedem 256-Bit-Block nur noch ein Bit voran, das signalisiert, ob in dem Block nur Datenwörter oder auch Kontrollwörter enthalten sind. Die 256b/257b-Codierung garantiert den Signalwechsel über die Forward Error Correction (FEC). Zudem erlaubt sie eine Transcodierung von 64b/66b zu 256b/257b, indem sie vier aufeinanderfolgende 64b/66b-Übertragungsworte zu einem 256b/257b-Übertragungswort zusammenfasst und diesem ein Bit voranstellt. Die 256b/257b-Codierung ist eine optionale Funktion von Fibre Channel Gen 6 (32GFC und 128GFC).

Ordered Sets

Übertragungswörter, Datenwörter

Der Fibre-Channel-Standard unterscheidet zwei Arten von Übertragungswörtern: Datenwörter und Kontrollwörter (Ordered Sets). Wird die 8b/10b-Codierung verwendet, so werden vier Zehnbit-Übertragungszeichen zu einem 40-Bit-Übertragungswort zusammengefasst. Während bei 64b/66b-Codierung ein Übertragungswort 66 Bit umfasst und dabei 64 Bit Daten enthält, arbeitet Fibre Channel auch in Generation 5 und 6 (16GFC, 32GFC und 128GFC) auf Ebene von FC-2 vor der Codierung mit 4 Byte langen Datenwörtern. Dabei werden bei 64b/66b

zwei und bei 256b/257b acht dieser Datenwörter zu einem Übertragungswort zusammengefasst. Datenwörter dürfen nur zwischen einem Start-of-Frame Delimiter (SOF Delimiter) und einem End-of-Frame Delimiter (EOF Delimiter) stehen.

Ordered Set, K28.5-Zeichen und Kontrollwörter

Ordered Sets dürfen nur zwischen einem EOF Delimiter und einem SOF Delimiter stehen, wobei SOFs und EOFs selbst Ordered Sets sind. Bei 8b/10b-Codierung ist allen Ordered Sets gemein, dass sie mit einem bestimmten Übertragungszeichen, dem sogenannten K28.5-Zeichen, beginnen. Das K28.5-Zeichen enthält eine spezielle Bitfolge, die ansonsten im Datenstrom nicht vorkommt. Der Eingangskanal eines Fibre-Channel-Ports kann deshalb bei Initialisierung eines Fibre-Channel-Links oder nach Verlust der Synchronisation auf einem Link mithilfe des K28.5-Zeichens den kontinuierlich einkommenden Bitstrom in 40-Bit-Übertragungswörter unterteilen. Bei 64b/66b- und 256b/257b-Codierung übernehmen spezielle Kontrollwörter die Funktion der Ordered Sets. Die im Folgenden aufgeführten Erklärungen zu Mechanismen und Prozessen gelten analog auch für 16GFC, 32GFC und 128GFC.

Frame Delimiter, Primitive Signal, Primitive Sequences

Ordered Sets haben verschiedene Aufgaben. Sie lassen sich grob in drei Gruppen einteilen. Frame Delimiter kennzeichnen wie oben beschrieben den Beginn und das Ende von Frames (Abschnitt 3.2.4). Es gibt verschiedene Start-of-Frame Delimiter, um zu kennzeichnen, welche Serviceklasse (Abschnitt 3.2.4) genutzt wird. Auch vom End-of-Frame Delimiter gibt es verschiedene Ausführungen. Damit kann zum Beispiel markiert werden, ob ein Frame bereits als defekt erkannt wurde. Primitive Sequences sind Übertragungsworte für festgelegte Sequenzen wie die Link-Initialisierung oder das Link Failure Protocol. Dabei wird das jeweilige Ordered Set tausend- und millionenfach gesendet, bis der Kommunikationspartner mit der im Standard festgelegten Antwort – normalerweise auch in großer Zahl – reagiert. Auf das einzelne Ordered Set kommt es dabei nicht an. Anders bei den Primitive Signals. Sie unterteilen sich noch einmal in Fillwords und Non-Fillwords. Während Fillwords (Füllwörter) auch in großer Menge geschickt werden, gibt es Situationen mit festen Regeln für deren Anzahl. So müssen zwischen zwei Frames mindestens sechs Füllwörter vorkommen. Füllwörter gewährleisten die Synchronität des Links. Non-Fillwords haben Spezialaufgaben. Am prominentesten sind dabei sicherlich die R_RDYs (Receiver Ready), mit denen ein Port seinem Gegenüber einen Buffer Credit zurückgeben kann und damit die Flusskontrolle auf dem Link erst ermöglicht (Abschnitt 3.2.4). Genau wie bei seinem Verwandten, dem VC_RDY, das dieselbe Aufgabe bei Links

mit Virtual Channels erfüllt, kommt es beim R_RDY auf jedes einzelne Vorkommen an. Ein Port darf also nur genau so viele R_RDYs schicken, wie er wirklich wieder Frames empfangen kann.

Link Control Protocol

Link Control Protocol

FC-1 definiert mithilfe von Ordered Sets verschiedene Link-Level-Protokolle zur Initialisierung und Verwaltung eines Links. Die Initialisierung eines Links ist Voraussetzung für den Datenaustausch mittels Frames (Abschnitt 3.2.4).

Speed Negotiation

Hostbus-Adapter und SFPs unterstützen oft mehr als eine Geschwindigkeit. Ein 16-GBit/s-SFP ist in der Regel auch in der Lage, 8-GBit/s- und 4-GBit/s-Verbindungen aufzubauen. Ziel der Speed Negotiation ist es, die größte von beiden Seiten unterstützte Geschwindigkeit herauszufinden und den physischen Link mit dieser Geschwindigkeit aufzusetzen. Dabei schalten die sich gegenüberliegenden Ports nacheinander die von ihnen unterstützten Geschwindigkeiten durch und senden dabei einen validen Bitstrom aus Primitive Sequences, oft zum Beispiel das NOS Ordered Set. Der genaue Inhalt der Übertragungswörter ist unbedeutend. Wichtig ist, dass der vom Empfänger interpretierte Bitstrom der zuvor beschriebenen Anforderung an die Bit-Error-Rate von höchstens einem Bitfehler pro 10^{12} Bits genügt. Da ein Fibre-Channel-Link immer zwei Fasern, den Hin- und den Rückweg, umfasst, muss dieselbe Geschwindigkeit für beide Richtungen gelten. Ein Link kann nicht beispielsweise mit 4 GBit/s in die eine Richtung und 8 GBit/s in die andere Richtung betrieben werden.

Link-Initialisierung

Sobald eine gemeinsame Geschwindigkeit gefunden und auf den Link angewendet wurde, beginnt die eigentliche Link-Initialisierung (Abb. 3–13). Dabei senden sich die beiden Ports große Mengen bestimmter Primitive-Sequence-Übertragungswörter zu, bis ihre beiderseitige Reaktion zur folgenden Sequenz von Ordered Sets führt: NOS (Not Operational Sequence), OLS (Offline Sequence), LR (Link Reset), LRR (Link Reset Response) und dann IDLE, was von der Gegenseite auch mit IDLE quittiert wird.

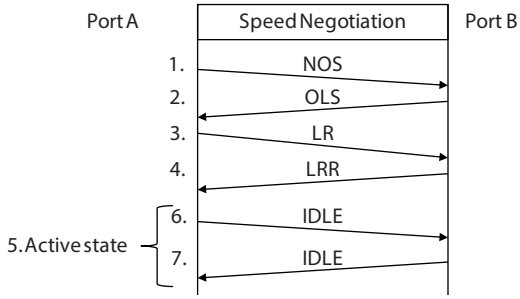


Abb. 3–13 Link-Initialisierung
Die Link-Initialisierung durchläuft mehrere Schritte. Erst danach steht der Port bereit für die weiteren Aktivitäten der nächsthöheren Protokollschicht.

3.2.4 FC-2: Datenübertragung

FC-2: das Herz des Fibre-Channel-Protokollturms

FC-2 ist die umfangreichste Schicht im Fibre-Channel-Protokollturm. Sie bestimmt, wie größere Dateneinheiten (zum Beispiel eine Datei oder das Ergebnis einer Anfrage an ein Datenbanksystem) über das Fibre-Channel-Netz übertragen werden. Sie regelt die Flusskontrolle, die dafür sorgt, dass der Sender die Daten nur so schnell sendet, wie der Empfänger sie verarbeiten kann. Und sie definiert verschiedene Dienstklassen, die auf die Bedürfnisse verschiedener Anwendungen zugeschnitten sind.

Exchange, Sequence und Frame

Exchange

FC-2 führt eine dreischichtige Hierarchie zur Übertragung von Daten ein (Abb. 3–14). Auf der obersten Ebene definiert ein sogenannter Exchange eine logische Kommunikationsverbindung zwischen zwei Endgeräten. Beispielsweise könnte jedem Prozess, der Daten liest und schreibt, ein eigener Exchange zugeordnet werden. Endgeräte (Server und Speichergeräte) können gleichzeitig mehrere Exchange-Beziehungen aufrechterhalten, auch zwischen den gleichen Ports. Verschiedene Exchanges helfen der Schicht FC-2, ankommende Daten schnell und effizient an den richtigen Empfänger in der nächsthöheren Protokollschicht (FC-3) auszuliefern. Exchanges werden über die Verkettung von den drei Informationen Absenderadresse, Zieladresse und OX_ID (Originator Exchange ID) identifiziert. Frames, die zu einem beliebigen Zeitpunkt X mit denselben Werten für diese drei Punkte in der Fabric unterwegs sind, gehören zum selben Exchange. Dabei werden die Werte für die Ziel- und die Absenderadresse natürlich wechselseitig

je nach Richtung ausgetauscht. Bei SCSI über Fibre Channel ist jeder I/O, genauer jede mit einem SCSI-Kommando startende Kommunikation, ein eigener Exchange. In FICON (Abschnitt 3.2.8), dem Kommunikationsprotokoll für Großrechner, werden Exchanges meist lang offen gehalten und können auch unidirektional benutzt werden, das heißt, die Antwort auf eine Anfrage wird in einem anderen Exchange zurückgeschickt.

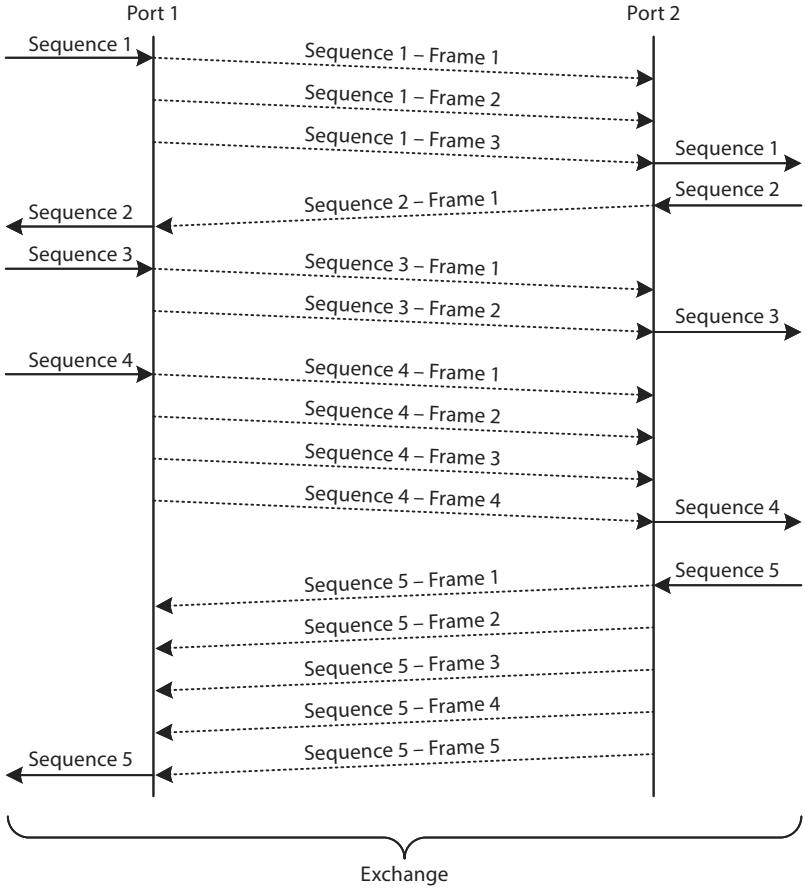


Abb. 3–14 Fibre Channel Exchange, Sequence und Frame

Innerhalb eines Exchange wird eine Sequence nach der anderen übertragen. Große Sequences werden vor der Übertragung in mehrere Frames zerlegt. Auf Empfängerseite wird eine Sequence erst an die nächsthöhere Protokollschicht (FC-3) ausgeliefert, wenn alle Frames der Sequence angekommen sind.

Sequence

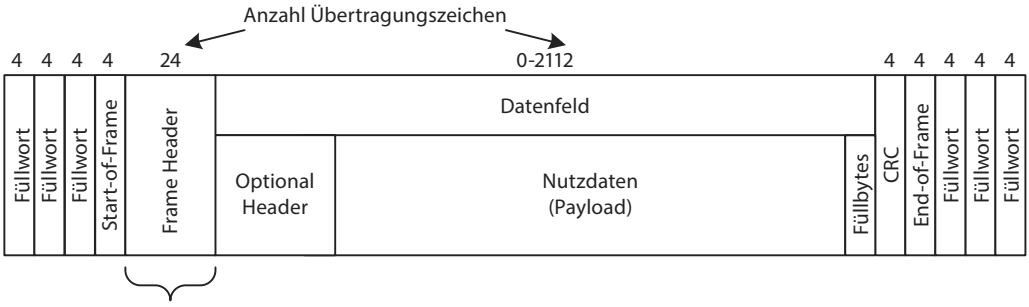
Eine Sequence ist eine größere Dateneinheit, die von einem Sender zu einem Empfänger übertragen wird. Diese sind nicht mit den Primitive Sequences aus FC-1 zu verwechseln, die lediglich eine bestimmte Art von 4-Byte-Übertragungswörtern beschreiben. Innerhalb eines Exchange kann nur eine Sequence nach der anderen übertragen werden. FC-2 gewährleistet, dass beim Empfänger Sequences in der Reihenfolge ausgeliefert werden, in der der Sender sie abgeschickt hat; daher auch der Name »Sequence«. Darüber hinaus werden Sequences nur dann an die nächsthöhere Protokollschicht ausgeliefert, wenn alle Frames der Sequence beim Empfänger angekommen sind (Abb. 3–14). Eine Sequence könnte beispielsweise die Anfrage nach einer Anzahl Datenblöcke ab einer gewissen Adresse von einer bestimmten LUN sein. Die darauffolgende Sequence käme vom Speichergerät und würde die angeforderten Daten enthalten. Sequences werden über ihre Sequence-ID innerhalb eines Exchange identifiziert.

Frame

Frame steht im Fibre-Channel-Protokoll für Übertragungspakete. Fibre Channel unterscheidet Kontroll-Frames und Daten-Frames. Kontroll-Frames enthalten keine Nutzdaten. Sie signalisieren Ereignisse wie das erfolgreiche Ausliefern eines Daten-Frames. Daten-Frames übertragen bis zu 2112 Bytes Nutzdaten. Größere Sequences müssen deshalb in mehrere Frames zerlegt werden. Theoretisch ist es zwar möglich, andere maximale Frame-Größen auszuhandeln, faktisch wird dies in der Praxis aber kaum genutzt, da alle Switches und alle angeschlossenen Endgeräte auf die gleiche Frame-Größe konfiguriert sein müssen. Jedes Frame bekommt innerhalb seiner Sequence einen Sequence-Count, um die richtige Reihenfolge der Frames sicherzustellen.

Frame-Format

Ein Fibre-Channel-Frame besteht aus einem Header, optionalen Nutzdaten (Payload) und einer CRC-Prüfsumme (Abb. 3–15). Zusätzlich wird das Frame von einem Start-of-Frame Delimiter (SOF) und einem End-of-Frame Delimiter (EOF) umklammert. Schließlich müssen über einen Link zwischen zwei Frames sechs Füllwörter übertragen werden. Fibre Channel ist im Gegensatz zu Ethernet und TCP/IP aus einem Guss: Die Schichten des Fibre-Channel-Protokollturms sind so gut aufeinander abgestimmt, dass das Verhältnis von Nutzdaten zu Protokoll-Overhead mit bis zu 98 % sehr niedrig ist. Das CRC-Prüfverfahren ist so ausgelegt, dass es alle Übertragungsfehler erkennt, wenn das zugrunde liegende Medium die vorgegebene Fehlerrate von 10^{-12} nicht überschreitet.



Unter anderem:

- Frame Destination Address (D_ID)
- Frame Source Address (S_ID)
- Sequence ID
- Nummer des Frames innerhalb der Sequence
- Exhange ID

Abb. 3–15 Das Fibre-Channel-Frame-Format

Das Fibre-Channel-Frame-Format zeichnet sich durch einen geringen Protokoll-Overhead aus.

Fehlerkorrektur auf Sequence- beziehungsweise auf Exchange-Ebene

Im Fibre-Channel-Protokoll ist eine Fehlerkorrektur auf Sequence-Ebene möglich: Wird ein Frame einer Sequence fehlerhaft übertragen, so wird die ganze Sequence neu übertragen. Mit Gigabit-Geschwindigkeit ist es effizienter, eine komplette Sequence erneut zu übertragen, als die Fibre-Channel-Hardware so zu erweitern, dass einzelne verlorene Frames erneut übertragen und in die richtige Position eingereiht werden. Die zugrunde liegende Protokollschicht muss die vorgegebene maximale Fehlerrate von 10^{-12} einhalten, damit dieses Verfahren effizient ist. Bei SCSI über Fibre Channel findet die Fehlerkorrektur jedoch praktisch auf Exchange-Ebene statt: Wird ein Frame von einem nicht korrigierbaren Bitfehler getroffen oder muss in der Fabric verworfen werden, bricht eines der Endgeräte (in der Regel der Server als Initiator) den Exchange ab und wiederholt ihn komplett. Solange einige wenige Bitfehler durch Forward Error Correction (FEC) korrigiert werden können, bleibt die CRC-Prüfsumme korrekt und das Frame kann weiterhin als intakt behandelt werden.

Flusskontrolle

Credit-Modell

Die Flusskontrolle sorgt dafür, dass der Sender Daten nur so schnell sendet, wie der Empfänger sie verarbeiten kann. Fibre Channel verwendet hierzu das sogenannte Credit-Modell. Jeder Credit repräsentiert die Fähigkeit des Empfängers, ein Fibre Channel Frame zu empfangen, weiterzuverarbeiten und gegebenenfalls zwischenzuspeichern. Gewährt der Empfänger dem Sender einen Credit von »4«, so darf der

Sender dem Empfänger nur vier Frames senden. Der Sender darf erst dann ein weiteres Frame senden, wenn der Empfänger den Empfang und die erfolgreiche Weiterverarbeitung von zumindest einem der gesendeten Frames bestätigt hat.

Flusskontrolle: End-to-End Credit, Buffer-to-Buffer Credit

FC-2 definiert zwei verschiedene Mechanismen zur Flusskontrolle: Ende-zu-Ende-Flusskontrolle und Link-Flusskontrolle (Abb. 3–16). Bei der Ende-zu-Ende-Flusskontrolle handeln zwei Endgeräte vor dem Datenaustausch den End-to-End Credit aus. Die Ende-zu-Ende-Flusskontrolle wird auf den Hostbus-Adapterkarten der Endgeräte realisiert. Im Gegensatz dazu findet Link-Flusskontrolle auf jeder physischen Verbindung statt, also auch zwischen Endgeräten und Switches, sowie zwischen den Switches selbst. Hierzu teilen sich zwei miteinander kommunizierende Ports jeweils die Anzahl ihrer Buffer mit. Ein Buffer ist ein schneller Speicher, der genau ein Frame fassen kann, unabhängig von dessen wirklicher Größe. Für den gegenüberliegenden Port ist die Anzahl der Buffer damit der Buffer-to-Buffer Credit.

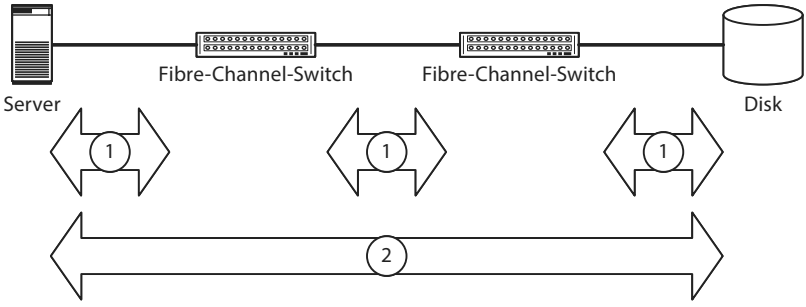


Abb. 3–16 Link-Flusskontrolle und Ende-zu-Ende-Flusskontrolle
Für die Link-Flusskontrolle handeln die Ports auf jedem Link den Buffer-to-Buffer Credit aus (1). Im Gegensatz dazu wird der End-to-End Credit für die Ende-zu-Ende-Flusskontrolle zwischen den Endgeräten ausgehandelt (2).

Dienstklassen

Dienstklassen

Der Fibre-Channel-Standard definiert sechs verschiedene Dienstklassen für den Datenaustausch zwischen Endgeräten. Drei dieser definierten Klassen (Class 1, Class 2 und Class 3) werden in auf dem Markt verfügbaren Produkten realisiert, wobei die verbindungsorientierte Class 1 eine eher selten implementierte Technik für Spezialanwendungen ist. Für Speichernetze hat sie praktisch keine Bedeutung. Die meisten Fibre-Channel-Produkte (Hostbus-Adapter, Switches, Speichergeräte) unterstützen die Dienstklassen Class 2 und Class 3, die einen paketorientierten Dienst (Datagrammdienst) realisieren. Zusätzlich dient Class F für den Datenaustausch zwischen den Switches innerhalb einer Fabric.

Class 1: verbindungsorientiert

Class 1 definiert eine verbindungsorientierte Kommunikationsverbindung zwischen zwei N-Ports: Eine Class-1-Verbindung wird vor der Übertragung von Frames geöffnet. Dabei wird ein Weg durch das Fibre-Channel-Netz festgelegt. Danach nehmen alle Frames den gleichen Weg durch das Fibre-Channel-Netz, sodass Frames in der Reihenfolge ausgeliefert werden, in der sie abgesendet wurden. Eine Class-1-Verbindung garantiert die Verfügbarkeit der vollen Bandbreite. Ein Port kann also keine anderen Frames senden, solange eine Class-1-Verbindung offen ist.

Class 2 und Class 3: paketorientiert

Class 2 und Class 3 sind dagegen paketorientierte Dienste (Datagrammdienste): Es wird keine dedizierte Verbindung aufgebaut; stattdessen werden die Frames einzeln durch das Fibre-Channel-Netz gesendet. Ein Port kann also mehrere Verbindungen gleichzeitig aufrechterhalten. Mehrere Class-2- und Class-3-Verbindungen können sich so die Bandbreite teilen. Dieses Multiplexing von mehreren Verbindungen über einen Port ermöglicht die volle Ausnutzung der Bandbreite in beide Richtungen, obwohl der Sequence-Mechanismus nur eine Halbduplex-Kommunikation pro Exchange erlaubt.

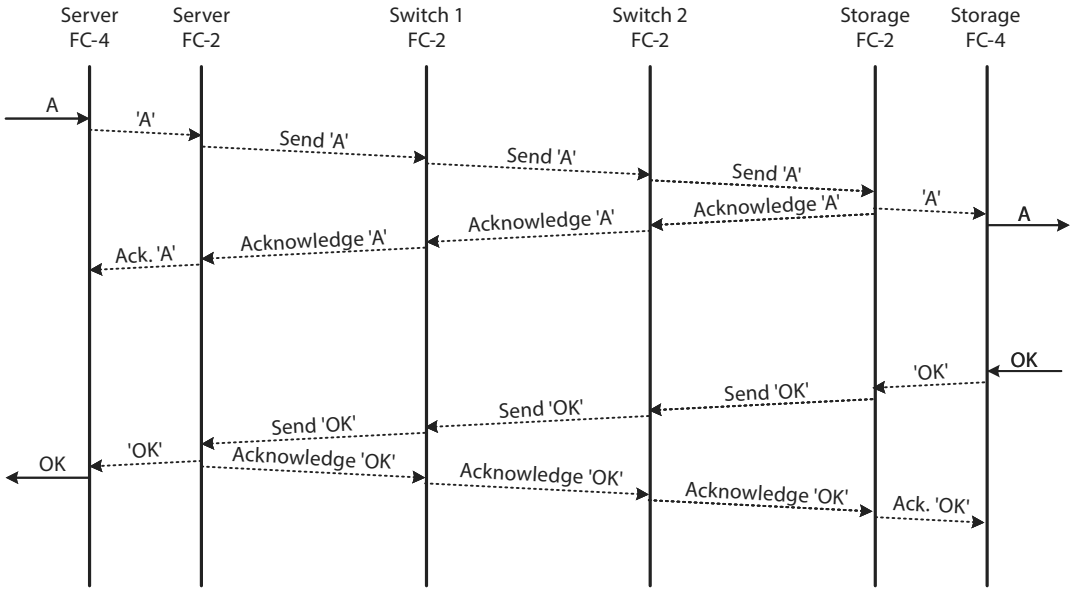


Abb. 3–17 Fibre Channel Class 2: Fehlerfreie Übertragung

Jedes übertragene Fibre-Channel-Frame wird innerhalb der Schicht FC-2 quittiert (Acknowledgement). Das Acknowledgement dient der Erkennung verlorener Frames (siehe Abb. 3–19 auf S. 112) und der Ende-zu-Ende-Flusskontrolle. Die Link-Flusskontrolle und die Umsetzung von Sequences auf Frames sind in der Abbildung nicht gezeigt.

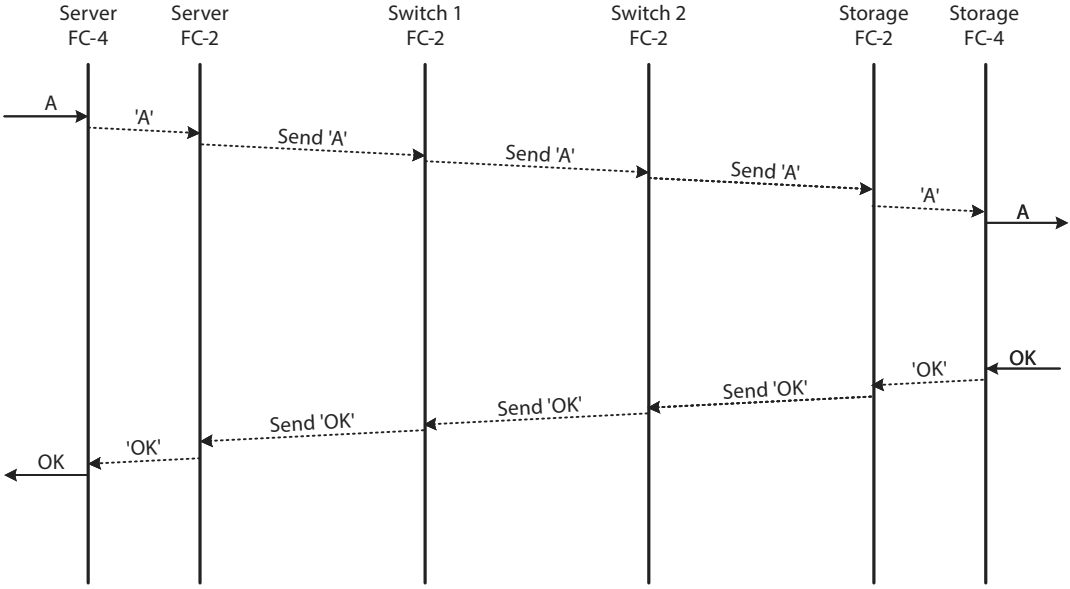


Abb. 3–18 Fibre Channel Class 3: Fehlerfreie Übertragung
Übertragene Frames werden in Schicht FC-2 nicht quittiert. Verlorene Frames müssen in den höheren Protokollschichten erkannt werden (siehe Abb. 3–20 auf S. 113). Die Link-Flusskontrolle und die Umsetzung von Sequences auf Frames sind in der Abbildung nicht gezeigt.

Class 2

Class 2 verwendet Ende-zu-Ende-Flusskontrolle und Link-Flusskontrolle. Bei Class 2 quittiert der Empfänger jedes empfangene Frame (Acknowledgement, Abb. 3–17 auf S. 110). Dieses Acknowledgement wird sowohl für die Ende-zu-Ende-Flusskontrolle als auch für die Erkennung verlorener Frames verwendet. Durch ein fehlendes Acknowledgement werden Übertragungsfehler sofort innerhalb von FC-2 bemerkt und unmittelbar den höheren Protokollschichten signalisiert. Die höheren Protokollschichten können so unmittelbar Maßnahmen zur Fehlerkorrektur einleiten (Abb. 3–19 auf S. 112). Nutzer einer Class-2-Verbindung können die Auslieferung der Frames in der richtigen Reihenfolge anfordern.

Class 3

Class 3 leistet weniger als Class 2: Frames werden nicht quittiert (Abb. 3–18 auf S. 111). Das heißt, es findet nur die Link-Flusskontrolle, aber keine Ende-zu-Ende-Flusskontrolle statt. Außerdem müssen die höheren Protokollschichten hier selbst merken, ob ein Frame verloren gegangen ist. Der Verlust eines Frames zeigt sich höheren Protokollschichten darin, dass eine erwartete Sequence nicht ausgeliefert wird, da sie von Schicht FC-2 noch nicht vollständig erhalten wurde. Ein Switch darf Class 2 und Class 3 Frames verwerfen, wenn seine Puf-

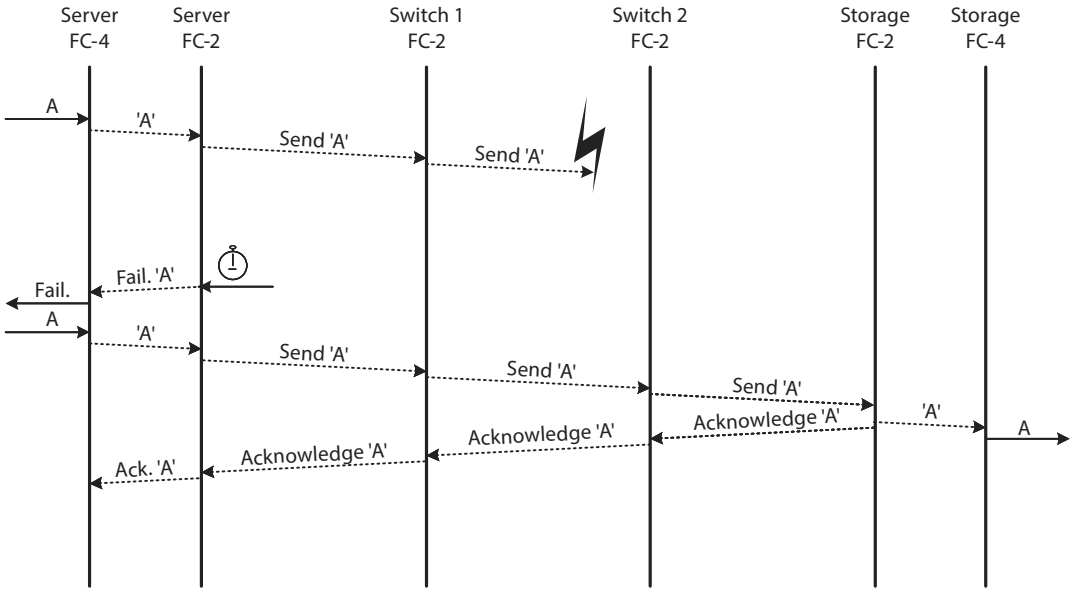


Abb. 3–19 Fibre Channel Class 2: Übertragungsfehler

Die Time-outs für Frames sind auf Schicht FC-2 relativ kurz. Fehlende Quittierungen (Acknowledgements) werden so innerhalb der Schicht FC-2 des Senders schnell erkannt und den höheren Protokollschichten signalisiert. Die höheren Protokollschichten sind für die Fehlerbehandlung zuständig. In der Abbildung wird das verlorene Frame einfach noch einmal übertragen. Die Link-Flusskontrolle und die Umsetzung von Sequences auf Frames sind in der Abbildung nicht gezeigt.

fer volllaufen. Aufgrund von höheren Time-out-Werten in den höheren Protokollschichten kann es im Gegensatz zu Class 2 wesentlich länger dauern, bis der Verlust eines Frames erkannt wird (Abb. 3–20 auf S. 113). Um häufige und aufwendige Fehlerkorrekturen zu vermeiden, ist es bei der Implementierung von Fibre Channel SANs wichtig, eine maximale Bitfehlerrate von 10^{-15} zu erreichen, auch wenn der Standard nur eine Bitfehlerrate von 10^{-12} vorsieht.

Die Dienstklassen in der Praxis

Es wurde bereits erwähnt, dass für die Praxis nur die beiden Klassen Class 2 und Class 3 von Bedeutung sind. In der Praxis wird die Dienstklasse so gut wie nie explizit konfiguriert, sodass in heutigen Fibre-Channel-SAN-Implementierungen die Endgeräte selbst aushandeln, ob sie mit Class 2 oder mit Class 3 kommunizieren. Aus theoretischer Sicht unterscheiden sich die beiden Dienstklassen darin, dass Class 3 einen Teil der Zuverlässigkeit der Kommunikation von Class 2 zugunsten eines weniger komplexen Protokolls opfert. Class 3 ist zur-

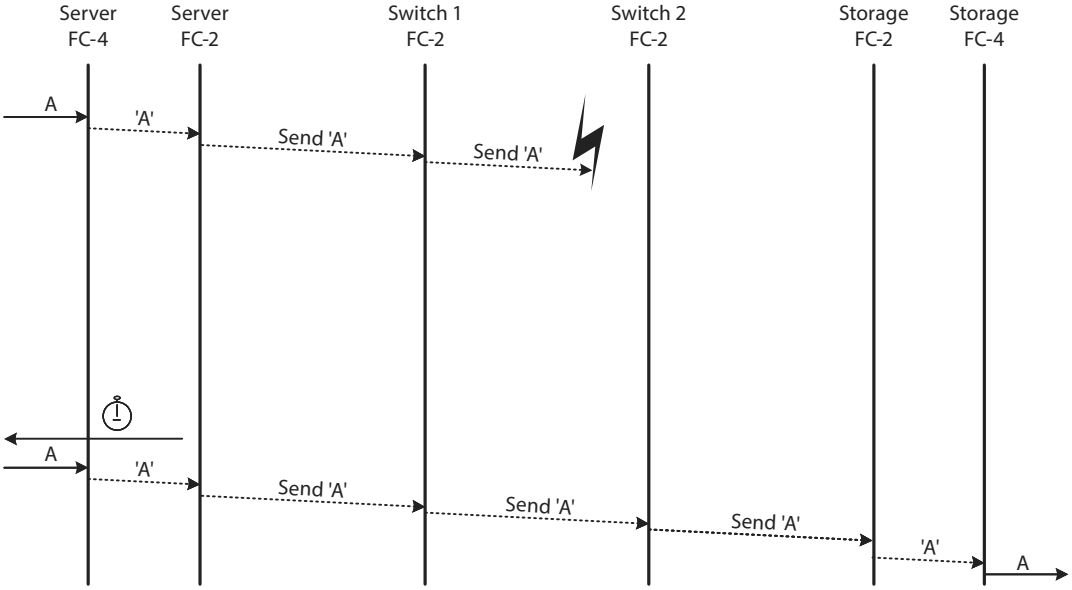


Abb. 3–20 Fibre Channel Class 3: Übertragungsfehler
Auch hier sind die höheren Protokollschichten für die Fehlerbehandlung zuständig. Die Time-outs in den höheren Protokollschichten sind im Vergleich zu den Time-outs in Schicht FC-2 relativ groß. Mit Class 3 dauert es also erheblich länger, bis auf ein verlorenes Frame reagiert wird. In der Abbildung wird das verlorene Frame einfach noch einmal übertragen. Die Link-Flusskontrolle und die Umsetzung von Sequences auf Frames sind in der Abbildung nicht gezeigt.

zeit die am häufigsten genutzte Dienstklasse. Durch den Verzicht auf das Ende-zu-Ende-Credit-Modell und den ständigen Versand von Acknowledgements sowie das Warten auf dieselbigen, ist sie effizienter als Class 2, solange kein intermittierender Fehler die Kommunikation stört. Die Vorteile von Class 2 wurden durch die Hersteller über die letzten Jahre durch zusätzliche Funktionen wie Forward Error Correction und Quality of Service aufgewogen, sodass die Verwendung von Class 2 eher abnimmt. Sie beschränkt sich auf kritische Anwendungen wie beispielsweise die Login-Phase in eine Fabric.

3.2.5 FC-3: Gemeinsame Dienste

FC-3: graue Theorie

FC-3 befindet sich seit 1988 in der Konzeptionsphase; in heute (2018) verfügbaren Produkten ist FC-3 leer. Folgende Funktionen wurden und werden für FC-3 diskutiert:

Striping
- Striping verwaltet mehrere Pfade zwischen Multiport-Endgeräten. Striping könnte die Frames eines Exchange über mehrere Ports verteilen und so den Durchsatz zwischen beiden Geräten erhöhen.

Multipathing
- Multipathing fasst mehrere Pfade zwischen zwei Multiport-Endgeräten zu einer logischen Pfadgruppe zusammen. Ausfall oder Überlastung eines Pfades können gegenüber den höheren Protokollschichten verborgen werden.

Komprimierung
- Komprimierung der zu übertragenden Daten, vorzugsweise realisiert in der Hardware auf dem Hostbus-Adapter.

Verschlüsselung
- Verschlüsselung der zu übertragenden Daten, vorzugsweise realisiert in der Hardware auf dem Hostbus-Adapter.

Mirroring
- Mirroring und andere RAID-Level sind schließlich ein letztes Beispiel, die im Fibre-Channel-Standard als mögliche Funktionen von FC-3 genannt werden.

Heute: Realisierung dieser Funktionen außerhalb des Fibre-Channel-Standards

Dass diese Funktionen nicht innerhalb des Fibre-Channel-Protokolls realisiert werden, bedeutet aber nicht, dass sie überhaupt nicht verfügbar sind. Beispielsweise werden Multipathing-Funktionen heute sowohl durch das Betriebssystem (Abschnitt 8.3.1) als auch durch Fibre-Channel-Switches (Abschnitt 3.3.4) bereitgestellt.

3.2.6 Link Services: Login und Adressierung

Link Services, Fabric Services

Link Services und die im nächsten Abschnitt beschriebenen Fabric Services (Abschnitt 3.2.7) stehen im Prinzip neben dem Fibre-Channel-Protokollturm. Sie werden benötigt, um über ein Fibre-Channel-Netz Datenverkehr zu betreiben. Aktivitäten dieser Dienste resultieren nicht aus dem Datenverkehr der Anwendungsprotokolle. Vielmehr werden diese Dienste benötigt, um die Infrastruktur eines Fibre-Channel-Netzes zu verwalten und somit den Datenverkehr auf der Ebene der Anwendungsprotokolle zu ermöglichen. Beispielsweise kennen die Switches einer Fabric zu jeder Zeit die Topologie des Gesamtnetzes.

Login

Dreistufiges Login:

Zwei Ports müssen sich miteinander bekannt machen, bevor Anwendungsprozesse über sie Daten austauschen können. Der Fibre-Channel-Standard sieht hierzu einen dreistufigen Login-Mechanismus vor (Abb. 3–21).

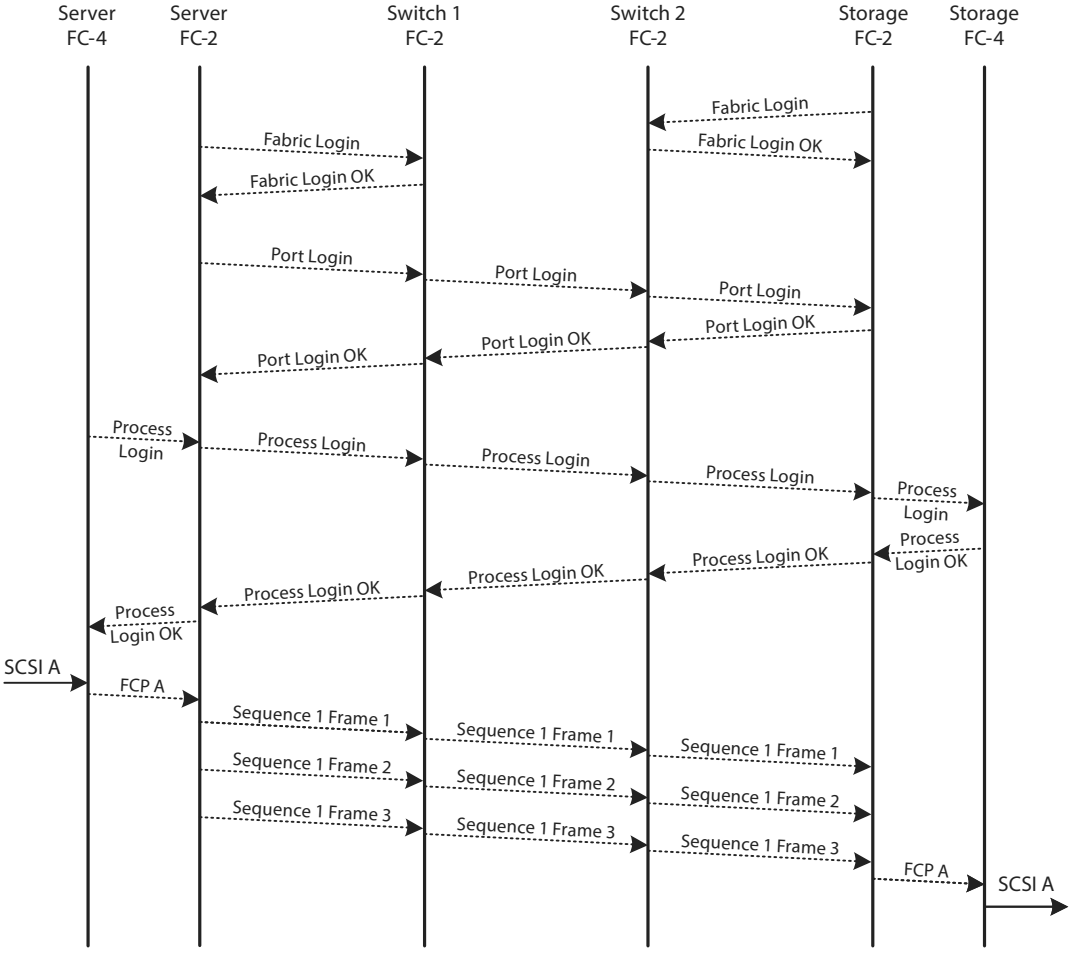


Abb. 3–21 Fabric Login, N-Port Login und Process Login
Fabric Login, N-Port Login und Process Login sind Voraussetzung für den Datenaustausch.

1. Fabric Login (FLOGI)

Das Fabric Login (FLOGI) etabliert eine Session zwischen einem N-Port und einem gegenüberliegenden F-Port. Das Fabric Login findet nach der Initialisierung des Links statt und ist zwingende Voraussetzung für den Austausch weiterer Frames. Der F-Port weist dem N-Port eine dynamische Adresse zu. Außerdem werden sich gegenseitig die konfigurierten Service-Parameter wie der Buffer-to-Buffer Credit, die maximale Frame-Größe oder die verwendeten Time-out-Werte mitgeteilt. Das Fabric Login ist für die Fabric-Topologie zwingend, wird aber auch bei der Point-to-Point-Topologie genutzt. Ein N-Port kann

aus der Antwort der gegenüberliegenden Ports ersehen, ob es sich um eine Fabric-Topologie oder um eine Point-to-Point-Topologie handelt. In der Arbitrated-Loop-Topologie war das Fabric Login optional.

2. N-Port Login (PLOGI)

Das N-Port Login (PLOGI) etabliert eine Session zwischen zwei N-Ports. Das N-Port Login findet nach dem Fabric Login statt und ist zwingende Voraussetzung für den Datenaustausch zwischen N-Ports oder zwischen einem N-Port und einem F-Port auf Ebene FC-4. N-Port Login überträgt die Service-Parameter wie den End-to-End Credit bei Class 2 sowie Informationen über die eigene Identität.

3. Process Login

Das Process Login (PRLI) etabliert eine Session zwischen zwei FC-4-Prozessen, die auf zwei verschiedenen N-Ports aufsetzen. Hierbei könnte es sich bei Unix-Systemen um Systemprozesse und bei Mainframes um Systempartitionen handeln. Process Login findet nach dem N-Port Login statt. Process Login ist aus Sicht von FC-2 optional. Allerdings erfordern einige FC-4-Protokollabbildungen einen Process Login für den Austausch FC-4-spezifischer Service-Parameter.

Adressierung

Fibre Channel Name (FCN), World Wide Name (WWN)

Fibre Channel unterscheidet zwischen Adressen und Namen. Fibre-Channel-Geräte (Server, Switches, Ports) werden durch eine 64-Bit-Kennung unterschieden. Der Fibre-Channel-Standard definiert hierzu verschiedene Namensformate. Einige Namensformate gewährleisten, dass eine solche 64-Bit-Kennung weltweit nur einmal vergeben wird. Solche Kennungen werden deshalb auch als World Wide Name (WWN) bezeichnet. Dagegen werden 64-Bit-Kennungen, die in getrennten Netzen mehrmals vergeben werden können, einfach nur als Fibre Channel Name (FCN) bezeichnet.

Sprachgebrauch »WWN« und »FCN«

In der Praxis ist dieser feine Unterschied zwischen WWN und FCN kaum bekannt: Hier werden alle 64-Bit-Kennungen grundsätzlich als WWN bezeichnet. Im Folgenden schließen wir uns dem allgemeinen Sprachgebrauch an und verwenden nur den Begriff WWN.

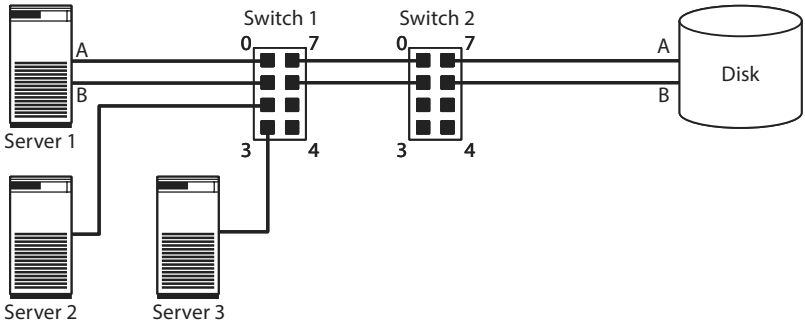
WWPN und WWNN

World Wide Names werden unterschieden in World Wide Port Names (WWPNs) und in World Wide Node Names (WWNNs). Wie die Namen schon sagen, wird jedem Port ein eigener World Wide Name als World Wide Port Name zugeordnet und zusätzlich dem gesamten Gerät ein eigener World Wide Name als World Wide Node Name. Die Unterscheidung von World Wide Node Name und World Wide Port Name ermöglicht es, im Fibre-Channel-Netz festzustellen, welche Ports zu einem gemeinsamen Multiport-Gerät gehören. Beispiele für Multiport-Geräte sind intelligente Disksysteme mit mehreren Fibre-Channel-Ports oder Server mit mehreren Fibre-Channel-Hostbus-Adapterkarten (HBAs) beziehungsweise HBAs mit mehreren Ports. WWNNs

könnten dazu genutzt werden, innerhalb des Server-Betriebssystems auf einfache Weise Funktionen wie Multipathing über redundante Pfade zu ermöglichen. Dabei ist zu beachten, dass einige Speichersysteme für jeden Port eigene WWNNs zur Verfügung stellen.

Adressierung in der Fabric: N-Port_ID

In der Fabric wird jedem 64-Bit World Wide Port Name beim Fabric Login automatisch eine 24-Bit-Portadresse (N-Port Identifier, N-Port_ID) zugeordnet (Abb. 3–22). Die 24-Bit-Portadressen werden innerhalb eines Fibre-Channel-Frames zur Kennzeichnung von Sender und Empfänger des Frames verwendet. Die Portadresse des Senders wird dabei als Source Identifier (S_ID) und die des Empfängers als Destination Identifier (D_ID) bezeichnet. Die 24-Bit-Adressen sind hierarchisch aufgebaut und spiegeln die Topologie des Fibre-Channel-Netzes in etwa wider. Dadurch kann ein Fibre-Channel-Switch einfach an der Destination ID erkennen, an welchen Port er ein ankommendes Frame weitersenden muss. Einige der 24-Bit-Adressen sind für besondere Zwecke reserviert (Abschnitt 3.2.7), sodass »nur« 15,5 Millionen Adressen für die Adressierung von Geräten übrig bleiben.



Port_ID	WWPN		WWNN		Device
010000	20000003	EAFE2C31	21000003	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

Abb. 3–22 Adressierung in der Fabric

Fibre Channel unterscheidet Endgeräte durch World Wide Node Names (WWNN). Jedem Anschlussport ist ein eigener World Wide Port Name (WWPN) zugeordnet. Für die Adressierung in der Fabric werden die WWNNs beziehungsweise WWPNs in kürzere N-Port_IDs umgesetzt, die in etwa die Netztopologie widerspiegeln.