Leitfaden iSCSI – Alternative und Ergänzung zum Fibre Channel Protocol

Stand 5. September 2005

BITKOM
Connecting visions - creating solutions

Impressum

Herausgeber:
BITKOM
Bundesverband Informationswirtschaft,
Telekommunikation und neue Medien e.V.
Albrechtstraße 10
10117 Berlin-Mitte

Tel.: 030/27 576 – 0
Fax: 030/27 576 – 400
bitkom@bitkom.org
www.bitkom.org

Redaktion:	Manfred Buchmann Ulrich Hamm Dirk Hampel Dr. Ralph Hintemann, BITKOM e.V. Heiko Schrader
Redaktionsassistenz:	Jeannette Beyer
Stand:	5. September 2005, 1. Auflage

Dieses Informationspapier gibt einen Überblick über das Thema „iSCSI“.Die Inhalte dieses Leitfadens sind sorgfältig recherchiert. Sie spiegeln die Auffassung im BITKOM zum Zeitpunkt der Veröffentlichung wider. Die vorliegende Publikation erhebt jedoch keinen Anspruch auf Vollständigkeit. Sie soll eine erste Hilfestellung bei der Beschäftigung mit dieser Thematik darstellen. Wir übernehmen trotz größtmöglicher Sorgfalt keine Haftung für den Inhalt.

Ansprechpartner:
Dr. Ralph Hintemann, BITKOM e.V.
Tel: +49 (0)30 / 27576 – 250
E-Mail: r.hintemann@bitkom.org

Inhaltsverzeichnis

Vorwort

Sehr geehrte Leserinnen und Leser,

nur wenige Begriffe wurden in den letzten Jahren in der Speicherindustrie so häufig verwendet wie *Virtualisierung* und *iSCSI.* Leider blieb auch bei iSCSI aus Anwendersicht manchmal der Eindruck, mit einem „Hype“ oder negativer mit „Slideware“, konfrontiert zu werden. Dies spiegelte sich letztlich in den Verkaufszahlen wieder: laut IDC betrug im Jahre 2004 der Gesamtumsatz mit iSCSI Disk Arrays gerade $45 Millionen, was einer Verdreifachung des Umsatzes gegenüber 2003 entsprach.[1]

In der Tat hat sich iSCSI, also die Blockübertragung von Speicherdaten mit Hilfe des SCSI-Massenspeicherprotokolls über IP, schwer getan. Zum einen verzögerte sich die Ratifizierung eines herstellerübergreifenden Industriestandards auf Grund der Komplexität länger, als erwartet (Verabschiedung 2003). Zum anderen waren die Positionierungsversuche einiger Anbieter, Analysten und Fachmagazine nicht gerade glücklich: Botschaften wie „iSCSI löst den Fibrechannel ab“, „iSCSI ist viel preiswerter als FC SANs“ etc. beherrschten leider zu oft als dominierender Tenor die einschlägige Presse.

Dabei hat sich durch die breite Unterstützung der (Speicher-) Industrie also z.B. von Microsoft, Hewlett Packard, EMC, Fujitsu-Siemens, CISCO, Mc Data, Network Appliance, Brocade oder IBM die Situation inzwischen grundlegend geändert! Heute liegt eine Palette von Lösungen vor und die Leistung bzw. Interoperabilität der verschiedenen Technologieoptionen (NICs, TOEs, Treiber für Betriebssysteme, Initiators, Targets etc.) hat sich entscheidend verbessert.

Also, wo stehen wir jetzt und wie sind die Fakten? Hier setzt der vorliegende BITKOM iSCSI – Leitfaden an.
Kurz, prägnant doch umfassend wurde über die technologischen Grundlagen, Leistungsdaten und Einsatzszenarien von ausgewiesenen Fachleuten ein iSCSI Kompendium entwickelt. Nicht vergessen wurden auch verschiedene Anwendungsbeispiele die deutlich machen, dass iSCSI für eine Vielzahl unterschiedlicher Anwendungen und Aufgaben im Unternehmen eingesetzt werden kann.

Zusammengefasst kann man sagen: iSCSI ist eine Alternative zu Fibre Channel, die auf Ethernet Technik aufbaut. Entscheidend aus Anwendersicht ist aber wie bei jeder Technik das perfekte Zusammenspiel von Infrastruktur mit Speicherapplikationen (z.B. für Disaster Recovery, Datenspiegelung, oder Backup-to-Disk). Erst damit kann der eigentliche Nutzen der IT für das Unternehmen auch sichtbar zum Tragen kommen.

Möge dieser Leitfaden dazu beitragen!

Viel Spaß beim Studium dieses Kompendiums wünscht Ihnen,

Ihr Norbert Deuschle.

Deuschle Storage Business Consulting
Chairman Storage Consortium GY
Chief Editor StorageWelt

1 *IDC, Worldwide Disk Storage Systems Forecast and Analysis, 2003–2007.*

1 Einleitung

Seit ihrer Einführung im Jahre 1996 erfreuen sich Storage Area Networks (SANs), also Speichernetzwerke einer zunehmenden Beliebtheit. Bis vor einiger Zeit war die extra für diesen Zweck entwickelte Fibre Channel Technologie hierfür die einzige Option für offene Systeme (Mainframes nutzen schon länger die ESCON-Technologie).

Allerdings waren die Investitionen, die der Einsatz dieser Technologie erfordert, anfänglich relativ hoch, auch wenn Fibre Channel wie viele andere Technologien einem starken Preisverfall unterliegt. Diese Investitionen und die Tatsache, dass eine zusätzliche Netzwerktechnologie auch zusätzliches Know-How erfordert, haben die Verbreitung von SANs heute auf vorwiegend mittlere und große Rechenzentren begrenzt. Auch die Vorsicht, mit der viele Anwender einer neuen und insbesondere derartig komplexen Technologie gegenübertreten, hat die weitere Verbreitung von SANs beeinträchtigt. Aber auch bei großen und mittleren Rechenzentren sind häufig vorrangig nur die kritischen Systeme in SANs integriert, wogegen oft eine Vielzahl von Servern noch nicht in das SAN integriert ist.

Viele Klein- und mittelständische Unternehmen (KMUs) haben bislang noch gar keine SAN-Technologie im Einsatz, sondern arbeiten i.d.R. mit so genanntem Direct Attached Storage (DAS), d.h. entweder in die Server integrierter Plattenspeicher oder direkt am Server angeschlossene Plattensysteme.

Mit der Verabschiedung der Spezifikationen für die neue iSCSI-Technologie durch die Storage Networking Industry Association (SNIA) im Jahr 2003 wurden zahlreiche Hoffnungen und vielleicht auch Illusionen geweckt, was mit iSCSI erreicht werden kann. Diese gehen bis zu der Behauptung, dass iSCSI mittelfristig Fibre Channel komplett ablösen wird.

Der vorliegende Leitfaden soll zum besseren Verständnis die Technologien iSCSI und Fibre Channel eingehender beleuchten, sie technisch und kostenmäßig miteinander vergleichen und somit als Grundlage für eine individuelle Entscheidung dienen, welche Technologie für welchen Einsatzzweck die geeignete ist.

Der Leitfaden wurde von Mitarbeitern von Unternehmen erstellt, die alle sowohl im Fibre Channel als auch im iSCSI-Umfeld aktiv sind und dort teilweise als Mitbewerber zueinander auftreten. Hiermit soll eine möglichst objektive Betrachtung beider Technologien erreicht werden.

2 Definition und Abgrenzung

Internet SCSI (iSCSI) ist ein Protokoll zur Verbindung von Servern mit Speichersystemen, ähnlich dem Fibre Channel Protocol (FCP). iSCSI bedeutet praktisch „SCSI over Internet Protocol“ so wie FCP für „SCSI over Fibre Channel“ steht.

iSCSI ist von anderen IP-basierten wie „Fibre Channel over IP (FCIP)“ oder „Internet Fibre Channel Protocol (iFCP)“ abzugrenzen, die dazu dienen, FCP basierte Netzwerke über IP miteinander zu verbinden.

Zwar erlaubt Fibre Channel auch, IP-Pakete zu transportieren („IP over Fibre Channel“), was aber nicht für die Kopplung von Servern mit Speichersystemen genutzt wird. Daher ist auch diese nicht Bestandteil dieser Betrachtung.

Dieses Dokument beschäftigt sich ausschließlich mit iSCSI („SCSI over IP“) als Alternative bzw. Ergänzung zu FCP („SCSI over Fibre Channel“).

3 Historie

3.1 Das Small Computer System Interface (SCSI) Protocol

Als das SCSI Protokoll in der Definition SCSI-1 im Jahr 1986 veröffentlicht wurde, unterstützte es bei einer Übertragungsrate von 5MB/s maximal 8 Geräte auf einem Bus. Im folgenden verdoppelte sich die Leistung im Fünf-Jahres-Takt: SCSI-2 Ausprägungen wie Fast SCSI oder Fast Wide SCSI steigerten die Leistung auf 10MB/s beziehungsweise 20MB/s. Mit Ultra-SCSI oder SCSI-3 wurden bereits vor einigen Jahren 40MB/s erreicht. Heute gibt es SCSI 160 und SCSI 320 mit bis zu 160 bzw. 320MB/s. Hierbei ist allerdings zu berücksichtigen, dass die maximal realisierbaren Entfernungen bei den schnellen Verbindungen deutlich sinken. Die maximale Anzahl der Geräte auf einem Bus erhöhte sich auf 16. Rückwärtskompatibilität war gegeben, so dass neuere und ältere Geräte denselben Bus nutzen können.

Das SCSI Protokoll ist eine seit vielen Jahren bewährte Technologie, die die Kommunikation über den SCSI Bus festlegt. Da sich der parallele SCSI Bus nicht für den Aufbau von Storage-Netzwerken eignet, wurden alternative Übertragungstechnologien (Medien) wie Fibre Channel und Ethernet entwickelt.

3.2 Das Fibre Channel Protocol (FCP)

Das Fibre Channel Protocol (FCP) wurde entwickelt, um die physikalischen Beschränkungen der SCSI-Übertragungstechnik wie Längen, Anzahl anschließbarer Endgeräte, Durchsatz usw. zu überwinden.

Damit stellt es eine Technologie dar, die auf einen speziellen Anwendungsfall, nämlich Speichernetzwerke, optimiert wurde.

Die grundlegende Standardisierung wurde 1995 durch das ANSI T11 Gremium abgeschlossen. Seitdem wurden zahlreiche Erweiterungen zu diesen Standards verabschiedet.

Der Vorteil des Fibre Channel Protocol ist die Kombination aus hoher Geschwindigkeit (u.a. durch geringen Protokoll Overhead) und der Fähigkeit, verschiedene Upper-Layer-Protokolle wie IP, SCSI und ESCON (Enterprise Systems Connectivity) transportieren zu können.

Es gibt folgende Fibre-Channel-Topologien:

- Fibre Channel Point-to-Point
  - Nur zwei Endgeräte sind verbunden: Initiator (typischerweise der Hostbus Adapter im Server) und Target (typischerweise der Fibre Channel Controller im Speicher-Subsystem).
- Fibre Channel Arbitrated Loop (FCAL)
  - Bis zu 127 Endgeräte können in einer Loop verbunden sein. Aber immer nur zwei kommunizieren zu einem Zeitpunkt, d.h. die Bandbreite wird geteilt.
  - Als Verbindungskomponenten kommen Fibre Channel Hubs (Private Loop) oder Fibre Channel Switches mit Loop Ports (Private oder Public Loop) zum Einsatz.
- Switched Fabric
  - Durch den Einsatz von Fibre Channel Switchen wächst die Flexibilität. Beginnend mit 8 und 16 Ports pro Gerät, sind heute Geräte mit bis zu 256 adressierbaren Ports verfügbar.

Um die Restriktionen von SCSI zu umgehen, wurden im ersten Schritt die Server direkt über Fibre Channel an die Storage Subsysteme angeschlossen (Point-to-Point). Dies hatte jedoch zur Folge, dass man aufgrund der fixen Verkablung auf den wesentlichen Vorteil von Fibre Channel, nämlich die Flexibilität und das Ressourcen-Sharing, verzichten musste.

Im nächsten Schritt konnten über Fibre Channel Hubs mehrere Server an einen Storage Port angeschlossen werden (FCAL). Größter limitierender Faktor dieser Topologie ist allerdings die Bandbreitenbegrenzung, da zu einer Zeit immer nur 2 Teilnehmer in einer Loop miteinander kommunizieren können.

Erst durch den Einsatz von Fibre Channel Switches und Direktoren (Switche mit einer sehr hohen Verfügbarkeit) wird die volle Flexibilität erreicht und bietet sich die Möglichkeit, komplexere Storage Area Network (SAN) Designs zu implementieren.

Durch die Vergrößerung des Speichers in den Geräten (Buffer Credits) konnte man schnell Entfernungen von 100 km und mehr, ohne Einbußen im Durchsatz, überwinden und somit Disaster-Recovery- (DR) und High-Availability- (HA) -Konzepte realisieren.

Der redundante Anschluss der Endgeräte an eine oder zwei parallele Netzwerke (Fabrics) erhöht nochmals die Ausfallsicherheit.

Die hohe Verfügbarkeit, die Flexibilität der Infrastruktur, der hohe Datendurchsatz, die leichte Implementierung komplexer DR/HA-Konzepte und die daraus resultierenden Kosteneinsparungen durch Konsolidierung sind Grund für den heutigen breiten Einsatz des Fibre Channel Protocol. Den neuen Anforderungen nach noch mehr Durchsatz und Protokoll Flexibiltät werden folgende Trends und Entwicklungen gerecht:

- Zur Überwindung von großen Distanzen von 1000 km und mehr macht man sich die IP-Netzwerke zu nutze und transferiert FC Frames mittels FCIP oder iFCP zwischen Fabrics an verschiedenen Lokationen.
    - Es gibt auch Fibre Channel Switches, die diese Distanzen ohne IP Gateways überwinden können. Dafür werden wieder größere Speicher in den Geräten gefordert sein (Buffer Credits).

Die Interoperabilität der Switche verschiedener Hersteller wird in letzter Zeit durch das Thema SAN - Routing bestimmt. Mit verschieden Konzepten können Daten zwischen Fabrics desselben oder verschiedener Hersteller transferiert werden.

Zukünftige Standards wie Virtual Fabrics und SAN-Routing werden die Basis für immer größere und dann wirklich unternehmensweite Speichernetzwerke sein.

3.3 Das Internet Protocol (IP)

Das TCP/IP (Transmission Control Protocol / Internet Protocol) ist heute der Defakto-Standard bei der Daten-Kommunikation. TCP sorgt für die gesicherte Übertragung zwischen den Endpunkten, während das IP Protokoll für die Übertragung bzw. Wegewahl auf den Teilstrecken zuständig ist. Die wichtigsten Hardware Elemente in einem IP-Netzwerk sind Router und Switches, die für die Wegewahl und Übertragung der IP-Pakete sorgen. TCP/IP ist für eine Any-to-Any-Kommunikation ausgelegt und unterstützt sehr viele Übertragungsmedien wie z.B. Ethernet, ATM (Asynchronous Transfer Mode), Serielle Verbindungen usw. Das TCP/IP-Protokoll wird durch die IETF (Internet Engineering Task Force) standardisiert. Die Standardisierungsdokumente sind s.g. RFCs (Requests for Comments).

3.4 Technologischer Vergleich

3.4.1 Protokoll-Vergleich

Das TCP/IP-Protokoll wird über einen entsprechenden Protocol Stack und Geräte-Treiber auf den Client und Server Systemen implementiert. Der TCP-Teil sorgt für die gesicherte Übertragung zwischen den Endpunkten, ist also z.B. für die Fehlerkorrektur beim „In-Order Delivery“ zuständig. In-Order Delivery ist aus Gründen des Multi TCP Connection Support zusätzlich auch im iSCSI Protokoll implementiert. Diese Funktionen sind immer in den Endgeräten verankert und nicht in den Netzwerk-Komponenten. Die Flow Control erfolgt über einen Window-Mechanismus, über den festgelegt wird, wie viele Pakete / Bytes in der Übertragung sein dürfen, bevor eine Bestätigung erfolgen muss. Im IP-Netz selbst stehen viele Funktionen für Lastausgleich, unterbrechungsfreies Re-Routing, Quality of Service usw. zur Verfügung.

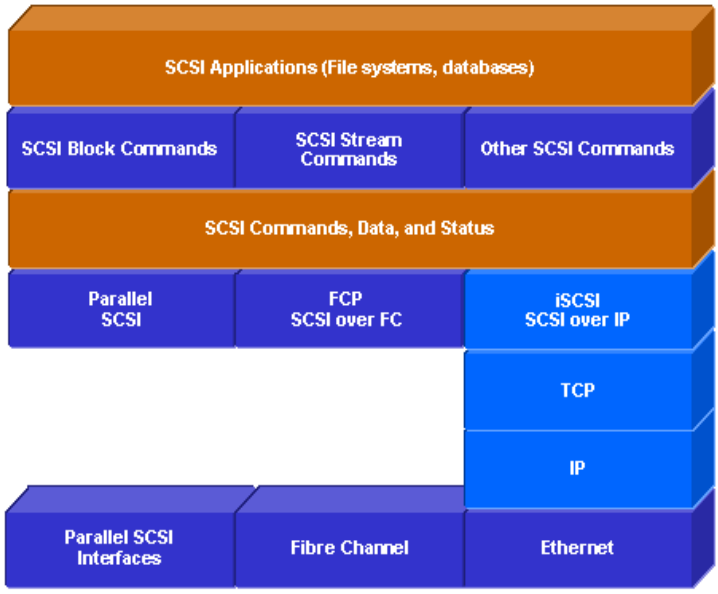

Abbildung 1: Protokoll Vergleich

Da der TCP/IP Protocol Stack in der Regel in Software implementiert ist, werden die Instruktionen auch von der jeweiligen lokalen CPU abgearbeitet. D.h. größerer TCP/IP Traffic hat auch Auswirkung auf die CPU-Last. Dieser Faktor tritt aber bei den heute verfügbaren CPUs immer mehr in den Hintergrund.

3.4.2 Leistung

3.4.2.1 Effektiver Durchsatz

Bereits bei der Implementierung von Client-Server-Architekturen und der Integration des IBM SNA (System Network Architecture) Protocol wurde TCP/IP als Transport-Protokoll für das SNA Protokoll verwendet. Dabei wurde das SNA-Protokoll in TCP/IP eingekapselt und über eine TCP Verbindung übertragen. Bei iSCSI wird grundsätzlich der gleiche Ansatz verwendet. In diesem Fall werden SCSI Kommandos und Daten in TCP/IP eingekapselt und über eine TCP Verbindung zwischen den Endpunkten übertragen. Natürlich entsteht bei einem Einkapselungsverfahren ein gewisser Overhead durch das Transport Protokoll. Dieser ist jedoch abhängig von der zu übertragenden Paketgröße. D.h. je größer das Paket, desto kleiner der Overhead.

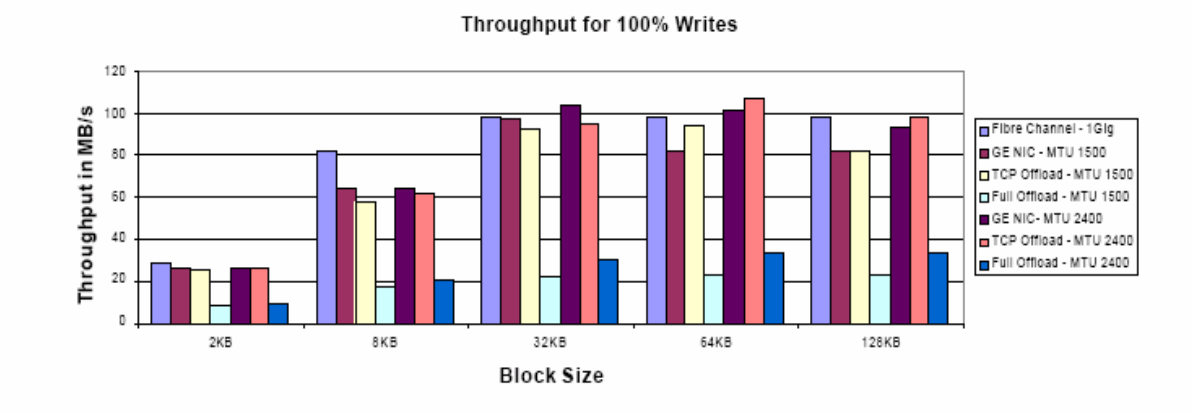

Abbildung 2: iSCSI vs Fibre Channel Durchsatz 100% schreiben

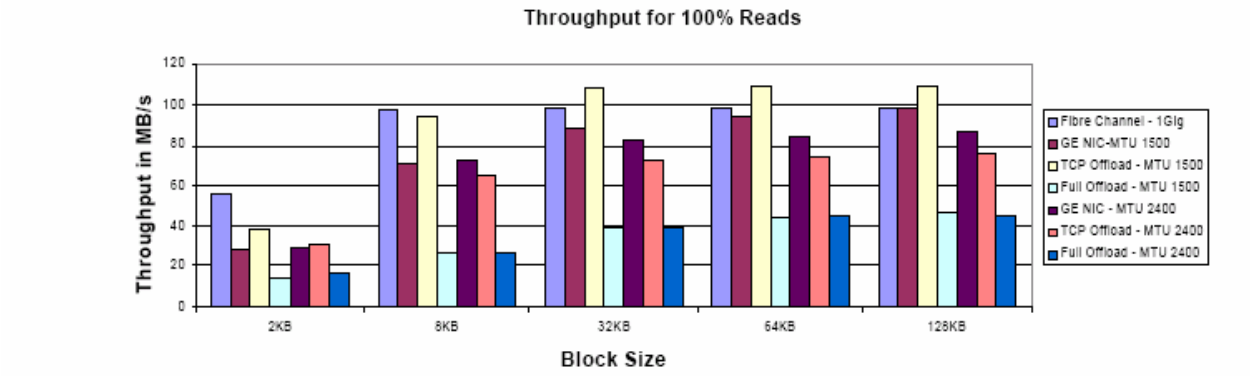

Abbildung 3: iSCSI vs Fibre Channel Durchsatz 100% lesen

Die Messergebnisse in Abbildung 2 und 3 zeigen, dass bei Schreibzugriffen der Unterschied zu native Fibre Channel nicht sehr groß ist. Bei den Lesezugriffen ist deutlich zu erkennen, dass bei kleineren Blockgrößen ein geringerer Durchsatz erreicht wird. Je größer die Blöcke sind desto besser ist der Durchsatz der erreicht wird. Dies sind für die Planung und die Auswahl der Applikationen für die Nutzung von iSCSI wichtige Punkte. Die Messergebnisse zeigen weiterhin das die Nutzung von TOEs (TCP Offload Engines) keinen positiven Einfluss auf den Durchsatz haben müssen.

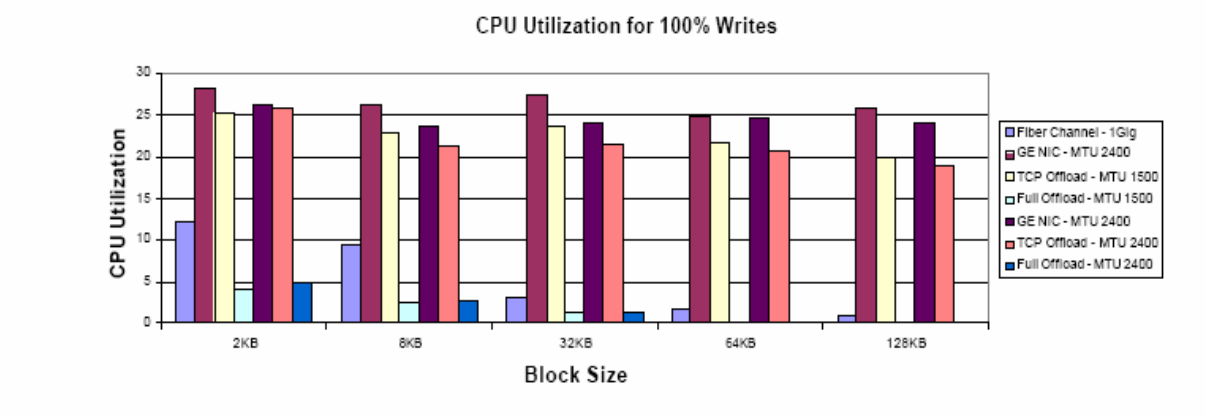

Abbildung 4: CPU Auslastung bei 100% schreiben mit unterschiedlichen Blockgrößen

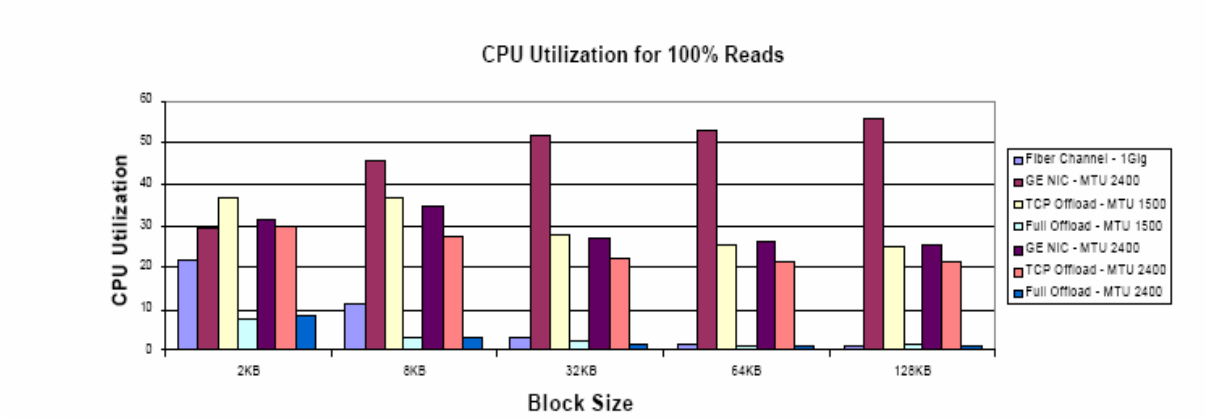

Abbildung 5: CPU Auslastung bei 100% lesen mit unterschiedlichen Blockgrößen

Die Messergebnisse in Abbildung 4 und 5 zeigen den Einfluss auf die CPU Last der Server bei iSCSI mit Software Treibern, Nutzung von TOEs (TCP Offload Engines) mit unterschiedlichen Rahmengrößen im Vergleich zu native Fibre Channel. Hier ist sicher deutlich zu sehen, dass bei Einsatz von iSCSI mit Software Treibern, eine größere CPU Last zu erwarten ist. Mit der Nutzung von TOEs kann die CPU Last reduziert werden. Der Einfluss der Blockgrößen ist auch hier bei Lesezugriffen größer als bei den Schreibzugriffen. Aufgrund des dargestellten Protokollaufwands bieten Fibre-Channel-Netzwerke bei gleicher Basis-Geschwindigkeit gegenüber iSCSI-Netzwerken etwa 10-15% mehr effektive Bandbreite.
Die Frage von Durchsatz und iSCSI ist direkt verbunden mit der physikalischen Infrastruktur der vorhandenen Lösung. Heutzutage gibt es mehrere Komponenten auf dem Markt, die es erlauben, eine iSCSI Umgebung so zu planen, wie der Nutzer es sich vorstellt: von hoher Leistung bis zu Backup-Anwendungen. Generell kann man den Durchsatz als einen Zusammenhang von iSCSI-Initator und iSCSI-Target betrachten.

Für die iSCSI-Initiatoren gibt es Software-implementierte (Treiber auf Network Interface Cards (NIC) oder TCP Offload Engine Network Interface Card (TNIC)**)** und Hardware-implementierte (iSCSI Host Bus Adapter (HBA)) Lösungen. Diese haben verschiedene Vorteile bezüglich der CPU-Belastung, I/O Operationen und der Konfiguration. Beim iSCSI-Target gibt es ebenfalls eine breite Auswahl von Komponenten. Inwiefern sie den gesamten Durchsatz beeinflussen, hängt stark von der Implementierung im iSCSI Gateway und Disk Array ab.

Es ist wichtig zu bedenken, dass in jedem Fall eine Encapsulation und De-Encapsulation stattfindet und dass diese Prozesse einen direkten Einfluss auf den Durchsatz haben werden. Bei der De-Encapsulation stellt sich die Frage ob diese am

besten am Control Processor des Speichersubsystemes oder am Gateway stattfindet. Der Vorteil eines Gateway-Switches ist, dass der Konvertierungsprozess unabhängig von dem Speichersystem geschieht. Dies erlaubt dem Speichersubsystem-Controller, seine IO- Operationen effizienter durchzuführen. Der geringere Overhead hat direkten Einfluss auf die Leistung. iSCSI Gateway Switches haben bei einer 1:1 Beziehung zwischen Host und Storage Controller und Gigabit-Ethernet-Anschluß, einen Full-Duplex-, Wire-Speed-Durchsatz von bis zu 219 MB/s) bei verscheidenden Blockgrößen bewiesen.

3.4.2.2 1, 2, 4, 8 und 10 Gbit/s

Während iSCSI auf den für IP-Netzwerke standardisierten Geschwindigkeiten 1 Gbit/s und 10 Gbit/s (und mit Einschränkungen auch auf 100 Mbit/s) eingesetzt werden kann, sind Speichernetzwerke auf Fibre-Channel-Basis historisch mit 1 Gbit/s, 2 und neuerdings auch 4 Gbit/s verfügbar. Auch 10 Gbit/s steht heute im Fibre–Channel-Umfeld mit Einschränkungen zur Verfügung.

8-Gbit/s-Fibre-Channel-Technologie ist gerade als neuer Standard für Speichernetzwerke durch die Fibre Channel Industry Association (FCIA) verabschiedet worden. Mit einer Verfügbarkeit von entsprechenden Produkten ist allerdings nicht vor 2006/2007 zu rechnen.

Die derzeit noch hohen Kosten von 10-Gbit/s- Netzwerken sind sowohl im IP als auch im Fibre-Channel-Umfeld ein wesentlicher Faktor, der viele Anwender vom Einsatz dieser Technologien abhält und daher auch auf Herstellerseite zu Zurückhaltung führt. Inter Switch Link (ISL) oder Backbone-Verbindungen haben jedoch höhere Anforderungen als z.B. Server-Storage oder Client-Server-Verbindungen. Für derartige Verbindungen kann die 10 Gbit/s Technologie interessant sein.

Auch ist zu beachten, dass 1 Gbit/s im IP-Umfeld sowie 1, 2, 4 und 8 Gbit/s im Fibre-Channel-Umfeld mit entsprechenden 10 Gbit/s Technologien nicht interoperabel sind, d.h. an einen 10 Gbit/s Switch Port kann auch nur ein anderer Switch mit einem 10 Gibt/s Port oder ein Endgerät (Server oder Speichersystem) mit 10 Gbit/s angeschlossen werden.

Gbit/s-Komponenten im Fibre-Channel-Umfeld sind abwärtskompatibel mit 2 und 1 Gbit/s, so wie 2-Gbit/s-Komponenten sind abwärtskompatibel mit 1 Gbit/s. Auch 8-Gbit/s-Komponenten werden abwärtskompatibel mit 4, 2 und 1 Gbit/s sein.

Nicht zu vernachlässigen ist ferner beim Einsatz von 10-Gbit/s- Technologien eine eventuell erforderliche Änderung der Kabel-Infrastruktur von Multimode-Verkabelung (50 oder 62,5µ) auf Monomode-Verkabelung (9µ).

3.4.2.3 Latenzzeit

Die Latenzzeit, d.h. die Verzögerung, die ein Datenpaket bei der Durchleitung durch ein Netzwerk erfährt, ist in Speichernetzwerken eine kritische Größe. Wird die Latenzzeit eines Datenpakets auf seinem Weg durch das Speichernetzwerk zu groß, so führt dieses zu einem Abbruch der Übertragung und damit u.U. bis hin zu einem Absturz der Anwendung oder des ganzen Servers.

Die Latenzzeit wird von mehreren Faktoren beeinflusst:

- der „Schaltgeschwindigkeit“ der involvierten Komponenten (Switches & Direktoren, Router usw.)
- der Anzahl der involvierten Komponenten
- der Kabellänge

- des Kabeltypes und der Übertragungsform (Kupferkabel mit elektrischer Übertragung oder Glasfaser mit optischer Übertragung)
- Übertragungsprotokoll
- Übertragungsbandbreite

Das Fibre Channel Protocol ist, da es als Protokoll von vornherein auf eine Optimierung als Speichernetzwerk entwickelt wurde, auf eine niedrige Latenzzeit ausgelegt. Zwar kann das Fibre Channel Protocol auch auf Kupferkabeln gefahren werden, dieses ist jedoch insbesondere aufgrund der dadurch bedingten Längenbeschränkung eher die Ausnahme und wird nur in Sonderfällen (z.B. innerhalb von Speichersystemen) angewendet.

Als typisches „Latenzzeit-Problem“ kann man auch die Längenbeschränkungen für so genannten „synchrone Anwendungen“ (wie z.B. synchrone Spiegelung oder der direkte Zugriff eines Servers auf seinen Speicher) über jede Art von Verbindung sehen. Diese liegt, je nach Art der involvierten Komponenten, heute bei ca. 100 km bzw. auch etwas darüber. Darüber hinaus ist nur eine asynchrone Verarbeitung (z.B. asynchrone Spiegelung) möglich. Da hier die Laufzeit des Lichts in einer Glasfaserleitung (und somit die Kabellänge) den Hauptfaktor darstellt, ist aus heutiger Sicht auch in absehbarer Zukunft nicht mit einer Änderung dieser Begrenzung zu rechnen.

In der Welt des IP-Protokolls spielt Latenzzeit normalerweise nur eine sehr geringe Bedeutung. Dieses ändert sich allerdings mit dem iSCSI-Protokoll, denn für die übergeordneten Betriebssysteme und Anwendungen ist es völlig transparent, welches Protokoll im Speichernetzwerk verwendet wird. Die Anforderungen ändern sich durch Einsatz des iSCSI Protokolls natürlich nicht. Daher muss bei Einsatz von iSCSI sehr stark auf die Latenzzeit geachtet werden.

Dieses ist auch der Grund, warum iSCSI kaum geeignet ist, um Server über öffentliche Netze mit Speichersystemen zu verbinden. Gerade in öffentlichen Netzen ist die Latenzzeit nicht vorhersagbar und kann mit der Netzauslastung und möglichen Fehlern im Netzwerk sehr stark schwanken. Der Weg eines Datenpakets durch ein öffentliches Netz kann in aller Regel nicht bestimmt werden.

So kann es passieren, dass die Datenpakete völlig unterschiedliche Wege gehen. Bevor sie jedoch zum Endgerät (Speichersystem oder Server) kommen, müssen sie aber wieder in die richtige Reihenfolge gebracht werden („In-Order-Delivery“). Werden hier die Laufzeitunterschiede zu groß, so wird auch die Verzögerung, mit der z.B. der Datenstrom bei der Anwendung ankommt, zu groß und diese stürzt ab.

iSCSI ist daher in aller Regel nur in privaten Netzen anwendbar, in denen alle involvierten Komponenten bekannt und der Datenpfad möglichst fest definierbar und damit für alle Datenpakete einheitlich ist.

Bei den heute zum Einsatz kommenden Netzwerkkomponenten (Router / Switches) sind viele Funktionen in Hardware abgebildet und verursachen daher keine größere Latenzzeit. Zusätzlich stehen weitere Priorisierungs- und QoS-Funktionen (Quality of Service) zur Verfügung, mit denen sichergestellt werden kann dass für Storage-Daten über iSCSI genügend Bandbreite zur Verfügung steht. Storage-Verkehr kann gegenüber anderem Daten-Verkehr wie z.B. Web, FTP, Telnet usw. entsprechend priorisiert werden.

3.4.2.4 Flow Control

Netzwerke, egal ob auf Fibre Channel- oder TCP-Basis, bestehen aus mehreren Komponenten, deren Zusammenarbeit über Protokolle geregelt ist. In einem Netzwerk gibt

es zum einen typische Lastprofile und zum anderen haben die Komponenten und Verbindungselemente eine definierte Leistung als Obergrenze. Erschwerend kommt hinzu, dass in einem Netzwerk zeitlich parallel viele Datenströme mit zum Teil unterschiedlichen Profilen abzuwickeln sind. Die Protokolle, die auf den einzelnen Netzwerken zum Einsatz kommen, versuchen den optimalen Arbeitspunkt für die Gesamtkonfiguration unter Berücksichtigung des typischen Lastprofils in einem komplizierten Optimierungsprozess zu erreichen. Der Optimierungsprozess ist in der Regel nicht statisch, sondern wird dynamisch je nach Lastsituation permanent neu berechnet.
Wesentliche Parameter für den Optimierungsprozess sind:

- Möglichst hoher Gesamtdurchsatz des gesamten Netzwerkes
- Möglichst geringe Verweilzeit aller Datenpakete eines Datentransfers im Netzwerk
- Faire Behandlung aller Teilnehmer am Netzwerk
- Stabiles Verhalten in Überlastsituation
- Schnelle Anpassung an sich ändernde Lastprofile
- Bandbreitenerhöhung durch parallele Nutzung mehrer Komponenten / Verbindungen
- Fehlertoleranz, bei Ausfall redundanter Komponenten
- Aufwandsarme Recovery-Prozesse zur Beseitigung von Fehlersituationen

Ein wichtiger Punkt in der Diskussion ist, wie ein Protokoll Grenzsituation in einem Netzwerk behandelt. In jedem Netzwerk kommerzieller Prägung kann es passieren, dass von der Quelle mehr Datenpakete gesendet werden, als sie das Netzwerk transportieren bzw. die Quelle abnehmen kann. Das heißt, das Protokoll muss dem Sender in irgendeiner Form mitteilen, dass er die Paketübertragungsrate reduzieren muss. Dazu gibt es unterschiedliche Algorithmen für die Flusskontrolle. Hier unterscheiden sich Fibre Channel Netzwerke und TCP-Netzwerke (insbesondere wie es im Internet verwendet wird), erheblich.

In Fibre Channel Netzwerken ist überwiegend eine "Credit-based flow control" vorzufinden. Sender und Empfänger handeln im Voraus die an einem Stück übertragbaren Pakete aus. Die Netzwerkkomponenten und die Quelle muss entsprechende Ressourcen in Form von Puffern bereitstellen, um die ankommenden Daten auch garantiert abnehmen zu können. Wesentlichen Einfluss auf die Anzahl der notwendigen Puffer hat die zu überbrückende Entfernung. Je größer die Entfernung ist, desto mehr Puffer müssen bereitgestellt werden, um alle Datenpakete, die sich noch auf der Leitung befinden, abnehmen zu können. Grundsätzlich gilt, parallele Last, auch wenn sie hochprior ist, darf die einmal reservierte Ressourcen nicht verwenden.

TCP-Netzwerke hingegen verwenden in der Regel die "Window-based flow control". Das Grundprinzip liegt hier in der Rückkopplung, wenn an irgendeiner Stelle im Netz ein Engpass auftritt. In diesem Konstrukt kann es passieren, dass eine Komponente u.U. keine neuen Pakete abnehmen kann, da die Puffer nicht rechtzeitig gelehrt werden konnten. In Folge müssen Pakte verworfen werden. Aufwändige Recovery-Maßnahmen sind die Folge. Recovery-Maßnahmen binden zum einen erhebliche Ressourcen und zum anderen erhöht sich die Verweilzeit wegen der erneuten Anforderung einzelner Pakete. Ob dieser Verfahren als nachteilig zu betrachten ist, hängt wiederum vom Lastprofil ab. Angenommen ein Datentransfer lässt sich immer innerhalb eines einzigen Paketes (z.B. kleiner ca. 1.500 Bytes) abbilden, dann wird dieses Verfahren nicht so schwer in Gewicht fallen, als wenn riesige Datentransfer z.B. mit Datenblöcken von 64.000 Bytes oder sogar größer über eine Netzwerk geschickt werden müssen,

welches „Window-based flow control“ verwendet und erst relativ spät erkannt wird, dass die Pufferanzahl auf der Netz- oder Empfängerseite nicht ausreichen.

Ein weiteres Problem ist, wie schnell sich ein TCP-Netzwerk auf sich sprunghaft verändernden Datenverkehr einstellt. Wenn zu Beginn der Übertragung eines großen Blockes das Netzwerk grundsätzlich nur wenige Pakete zulässt und der maximale Wert der an einem Stück übertragbaren Pakete erst nach mehreren Zyklen sich einstellt, dann hat dies erhebliche Nachteile auf die Verweilzeit sprunghaft wechselnder Lasten, wie sie im Datenbereich anzutreffen sind. Kontinuierliche Lastprofile, wie z.B. Streaming-Daten sie darstellen, sind somit besser für TCP-Netzwerke geeignet, da nach einer Einschwungphase ein optimaler Durchsatz gewährleistet werden kann. Eine Eigenheit der „windows-based flow control“ liegt auch darin, dass bei Überlastsituationen das Window (an einem Stück übertragbare Datenpakete) drastisch reduziert wird. Wenn wieder Ressourcen frei sind, dann nähert sich das Window, wie oben beschrieben, erst langsam wieder an das Maximum an. Ein weiterer Vorteil für ein datenorientiertes Profil ist, wenn der gesamte Datentransfer in einer einzigen Session abgewickelt werden kann. Das heißt, es wird nur ein einziger Pfad verwendet. Dies garantiert die Ankunftsreichenfolge an der Quelle. Datenpakete müssen nicht zwischengepuffert werden, um die Reihenfolge sicherzustellen. Zwischenspeicherung bedeutet entsprechende Verzögerung und Ressourcenbelegung, da am Schluss die Daten nochmals in den Anwenderbereich kopiert werden müssen.

Für iSCSI-Umgebungen wird daher der Anwendungs-spezifische Aufbau von dedizierten „switched“ TCP/IP-Netzwerken mit vergleichbaren Bandbreiten, wie sie in FC-Netzwerken gegeben sind, empfohlen.

3.4.2.5 Host Bus Adapter, Network Interface Card und TCP/IP Offload Engine

Host Bus Adapter (HBAs) sind die Schnittstelle des Servers in das Speichernetzwerk. In einem IP-Netzwerk übernimmt normalerweise die Network Interface Card (NIC) diese Funktion. Bei iSCSI-basierten Speichernetzwerken gibt es im Prinzip 3 Arten von Netzwerk-Schnittstellen: einfache Netzwerkkarten (NICs), so genannte TCP/IP Offload Engines (TOEs) sowie iSCSI Host Bus Adapter. Einfache NICs sind die billigste Variante für iSCSI, allerdings erfolgt hier die Abwicklung des IP Protocol auf der Host CPU, was bei stärkerer I/O-Last zu hohen Interrupt-Zahlen und damit zu zusätzlicher CPU-Belastung führt. Um die Host CPU zu entlasten und die Latenzzeiten zu verringern, wurden TOE Karten entwickelt, bei denen das TCP/IP-Protokoll in Form von Hardware auf der Karte „gerechnet“ wird. Mit diesen Karten kann die Host CPU von jeglichem TCP/IP Traffic entlastet werden, wobei auch andere Protokolle außer iSCSI unterstützt werden. Zur Auslagerung des normalen TCP/IP Traffic ist allerdings Betriebssystemunterstützung erforderlich, was nicht überall gegeben ist. Die alleinige Auslagerung des iSCSI Protokolls ist unkritischer. Hier läuft TCP/IP und iSCSI als Hardware-Implementierung und entlastet die CPU komplett. Nur diese iSCSI HBAs sind in Bezug auf die Aspekte der CPU-Belastung und der Latenzzeit mit Fibre Channel HBAs vergleichbar.

iSCSI HBAs sind in Ihrer Funktionalität dem Server gegenüber auch mit herkömmlichen SCSI-Adaptern vergleichbar. Bei der Auswahl eines iSCSI HBA sollten Funktionalität wie Booten (siehe 3.4.3) und Leistung beachtet werden.

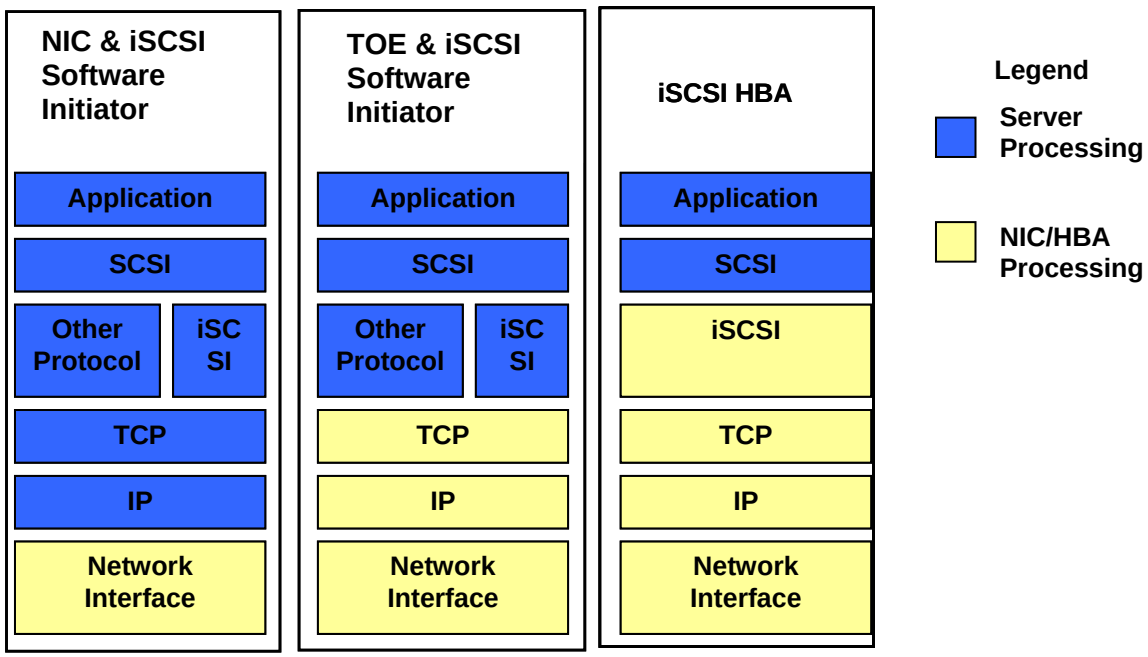

Abbildung 6: Vergleich von iSCSI-Varianten

3.4.3 Externes Booten

Die Standard-Variante für das Booten von Intel-basierten Servern ist der „INT 13“ Boot, wie von IDE-, SCSI-Adaptern und Fibre Channel HBAs benutzt. Diese Art des Bootens lässt sich momentan nur durch iSCSI HBAs und nicht durch reine TOE-Karten realisieren.

Allerdings ist diese Funktionalität zurzeit noch nicht bei allen verfügbaren iSCSI HBAs implementiert, was bei einer Beschaffung gegebenenfalls berücksichtigt werden sollte.

Für die Betriebssysteme Windows und Linux besteht auch die Möglichkeit, über iSCSI-Treiber zu booten. Voraussetzung dafür ist die PXE (Pre-Boot Execution Environment) auf dem Server. Zusätzlich werden noch ein Dynamic Host Configuration Protocol (DHCP)- und ein Trivial File Transfer Protocol (TFTP)-Server benötigt. Mit der iSCSI-Boot-Unterstützung können somit auch Server, die über iSCSI ohne iSCSI HBA mit dem Speichersystem verbunden sind, diskless betrieben werden.

3.5 Verfügbarkeit

Anders als bei „normalen“ Netzwerken ist bei Speichernetzwerken die Verfügbarkeit ein ganz wesentlicher Faktor, da schon kürzeste Unterbrechungen zu kompletten Systemausfällen und (bei größeren Systemen) stunden- oder gar tagelangen Recovery-Maßnahmen führen können.

Um die höchstmögliche Verfügbarkeit eines Speichernetzwerks sicherzustellen, müssen zahlreiche Punkte beachtet werden, die nachfolgend eingehender ausgeführt sind.

3.5.1 Trennung von Daten- und Speichernetzwerken

Es könnte der Gedanke nahe liegen, bestehende Netzwerk-Infrastrukturen, die scheinbar noch ungenutzte Kapazitäten bieten, für das Speichernetzwerk mit zu benutzen.

Hiervon muss dringend abgeraten werden! Die Mischung von „normalen“ Benutzerdaten, wie sie in einem Local Area Network (LAN) vorkommen, mit I/O-Daten, wie sie für Speichernetzwerke typisch sind, kann dramatische Auswirkungen auf das I/O-Verhalten der entsprechenden Systeme bis hin zum Systemabsturz haben.

Hintergrund hierfür ist die hohe Sensibilität von Speichernetzwerken, wie sie schon mehrfach beschrieben wurde. Die Benutzerdaten, für die die Latenzzeit des Netzwerkes nur geringere Rolle spielt, können z.B. eben genau diese für die I/O-Daten so kritische Latenzzeit drastisch beeinträchtigen.

Speichernetzwerke sollte daher immer in einer eigenen Infrastruktur unabhängig von sonstigem Netzwerk-Verkehr betrieben werden.

Beim Fibre Channel Protocol bedeutet dieses immer eine eigene, physische Netzwerkinfrastruktur.

Bei iSCSI kann, sofern die bestehende IP-Infrastuktur ausreichende Kapazitäten bietet, auch in diese integriert werden. Dabei muss aber darauf geachtet werden, dass die erforderlichen Bandbreiten zu jeder Zeit garantiert sind (z.B. durch Virtuelle LANs (VLANs)) und dass auch eine hohe Belastung mit dem parallelen LAN-Verkehr bei den involvierten Komponenten (Switches, Direktoren, Router usw.) keine Auswirkungen auf die Latenzzeit hat.

3.5.2 Redundante Netzwerke / Multipathing

Netzwerke zwischen Rechnersystemen (Server und Workstations) sind häufig nur einfach ausgelegt. Dieses liegt darin begründet, dass das Netzwerk hier nur eine vermittelnde Wirkung hat, die Rechnersysteme untereinander aber autark sind. D.h. das einzelne Rechnersystem kann auch ohne die Verbindung zu den anderen Systemen „überleben“:

Obwohl in solchen Netzwerken auch Redundanzen durch alternative Pfade vorhanden sind, kommt es dennoch oft zu einem oder mehreren „Single Points of Failure“, deren Ausfall eine Trennung der Verbindung zur Folge hat. Auch kann in einem solchen Netzwerk ein Administratorfehler zu einer Trennung von Verbindungen führen.

Speichernetzwerke hingegen sind substantieller Bestandteil eines „Systems“, das typischerweise aus Rechner (CPU, Memory usw.), Massenspeicher und eben genau der Verbindung dazwischen besteht. Eine auch nur kurzzeitige Unterbrechung kann (IT-technisch) zur Katastrophe führen.

Aus diesem Grund werden Speichernetzwerke fast immer nicht nur in sich redundant, sondern tatsächlich doppelt ausgelegt („Dual Fabric“). Somit kann dann z.B. auch der Fehler eines Administrators abgefangen werden, da sich ein solcher Fehler zwar auf ein komplettes Netzwerk, nicht aber auf das zweite, redundante Netzwerk ausdehnen kann. Somit besteht ggf. immer noch ein alternativer Pfad zwischen Rechner und Massenspeicher.

Nur bei Systemen mit sehr niedriger Priorität (z.B. Entwicklungs-, Testsysteme, absolut unkritische Anwendungen usw.) kann eventuell eine einpfadige Anbindung in Frage kommen.

Die Frage, ob eine ein- oder zweipfadige Anbindung erforderlich ist, ist unabhängig vom verwendeten Speichernetzwerkprotokoll (FCP oder iSCSI).

Für eine zweipfadige Anbindung ist allerdings auf den entsprechenden Rechnern auch der Einsatz eines sogenannten „Multipathing-Treibers“ erforderlich. Dieses liegt darin begründet, dass Betriebssysteme und Anwendungen normalerweise nur einen Weg zu den Daten kennen. Leistungsstarke Multipathing-Treiber können Betriebssystem und Anwendung diesen einen Pfad darstellen lassen, tatsächlich aber mehrere Pfade nutzen, die I/O-Last darüber verteilen, im Fehlerfall „umleiten“ (Failover) und nach Behebung eventueller Probleme den verlorenen Pfad wieder herstellen (Recovery).

Einige Betriebssysteme bieten heute eine solche Multipathing-Funktionalität bereits ohne zusätzliche Treiber. Im Einzelfall ist aber immer zu prüfen, ob die angebotene Funktionalität ausreichend ist und ob der Storage-Hersteller dies unterstützt. Alternativ sollte dann der Einsatz eines Produktes eines anderen Anbieters für solche Multipathing-Funktionalität geprüft werden. Zu den Anbietern gehören sowohl die Hersteller von Speichersystemen als auch Hardware-unabhängige Softwarehäuser.

In Bezug auf iSCSI gibt es mehrere Varianten für eine redundante Auslegung. Die Möglichkeiten hängen u.a. auch von der Art der Implementierung (iSCSI-Treiber, TOE-Karte oder iSCSI HBA) ab.

Das iSCSI-Protokoll bietet die Möglichkeit, eine iSCSI-Session über mehrere IP-Verbindungen parallel zu betreiben. Dabei können mehrere logische TCP-Verbindungen über eine oder oder mehrere Netzkarten (NICs oder TOE-Karten) aufgebaut werden. Zum gegenwärtigen Zeitpunkt ist diese Funktionalität aber noch nicht bei allen Anbietern von iSCSI-Treibern implementiert. Auch wird eine identische Funktionalität auf der „Target“-Seite (sprich: dem Speicher-Subsystem bzw. der iSCSI-FC-Bridge) erforderlich.

Redundante Netzwerke kann man bei Einsatz von iSCSI-Treibern auch durch Einsatz des sogenannten „Trunking“ auf dem Ethernet Layer erreichen. Dies ist Stand heute mit TOE oder iSCSI-Karten nicht möglich, da diese Karten zur Zeit nur einen Ethernet Port verwenden, was sich aber voraussichtlich in Zukunft ändern wird. Allerdings sind diese TOE-Karten bzw. iSCSI HBAs wieder ein „Single Point of Faliure (SPoF), da bei ihrem Ausfall alle Verbindungen zum Speicher verloren gehen.

3.5.3 Redundanz bei den Hardwarekomponenten

Um die Verfügbarkeit eines Speichernetzwerks weiter zu erhöhen, sollte darauf geachtet werden, bei allen kritischen Komponenten weitestgehende Hardware-Redundanz vorzusehen. Hierbei ist „redundant“ nicht automatisch mit „doppelt“ gleichzusetzen. Vielmehr bedeutet redundant, dass die entsprechende Komponenten mindestens einmal mehr vorhanden ist, als sie für einen störungsfreien Betrieb erforderlich ist („n+1“). Braucht ein Switch z.B. 2 Lüfter im Dauerbetrieb, so sollten mindestens 3 vorhanden sein. Somit kann ein Lüfter ausfallen, ohne den Betrieb zu beeinträchtigen.

Neben redundanten Netzteilen und Lüftern können auch die so genannten Control Processors, d.h. die Steuerungseinheiten der Switches, Direktoren und Router in einem Speichernetzwerk redundant ausgelegt sein.

3.5.4 Firmware Upgrades im laufenden Betrieb

Wie praktisch alle Systeme in einem Rechenzentrum benötigen auch die Komponenten eines Speichernetzwerks von Zeit zu Zeit ein Update der internen Software (Firmware). Dies ist z.B. der Fall, um Fehler zu beheben, um neue Funktionalitäten zur Verfügung zu stellen oder um die Interoperabilität mit neuen oder geänderten angeschlossenen Endgeräten zu gewährleisten.

Unterstützen die Komponenten einen Firmware Upgrade im laufenden Betrieb („hot code activation“), so bedeutet dies, dass neue Firmware im laufenden Betrieb und ohne jegliche I/O-Unterbrechung eingespielt und aktiviert werden kann.

Ist dieses nicht gegeben, so muss für einen Firmware-Update ein komplettes Speichernetzwerk (eine Fabric) außer Betrieb genommen werden. Ist kein zweites, redundantes Netzwerk vorhanden oder sind an diesem Netzwerk Systeme nur einpfadig angeschlossen, so müssen diese im Vorfeld heruntergefahren und nach Abschluss des Firmware Upgrade neu gestartet werden.

Die Möglichkeit von Firmware Upgrades im laufenden Betrieb kann somit geplante Ausfallzeiten („planned downtimes“) verkürzen oder ganz vermeiden und damit die Gesamtverfügbarkeit von Systemen und Anwendungen erhöhen.

3.6 Internet Storage Name Service (iSNS)

iSNS ergänzt iSCSI um ein Dictionary, indem u.a. Informationen über iSCSI-Instanzen gespeichert werden. Die Verwaltung und Bereitstellung der Informationen erfolgt über entsprechende standardisierte Services. Damit stehen analoge Funktionen wie in einem Fibre Channel Netzwerk zur Verfügung. iSNS erlaubt z.B., dass sich auf der einen Seite iSCSI-Targets darin registrieren, so dass sie dann von den iSCSI-Initiatoren gefunden werden können. Mit Hilfe des iSNS können iSCSI-Initatoren und iSCSI-Targets auch überprüfen, ob der jeweilige Partner noch aktiv ist. Falls der Partner z.B. nicht mehr aktiv ist, kann die entsprechende iSCSI-Sitzung beendet werden. Damit iSCSI-Targets die Überprüfung durchführen können, ist es erforderlich, dass sich auch die iSCSI-Initiatoren registrieren. iSNS kann auch ereignisorientierte Meldungen verschicken. Konfigurationsänderungen im iSNS-Dictionary lösen z.B. derartige Meldungen aus. Clients, die diese Meldungen abonniert haben, können damit neu hinzugefügte Speicherbereiche sofort erkennen. iSNS bildet die Basis für iSCSI-Gateways, indem sie eine Abbildung der iSCSI-Namen auf das Fibre Channel Netzwerk-Adressierung vornehmen. Das iSNS bildet eine zentrale Informationsbasis für Überwachungs- und Managementwerkzeuge. Unter der Voraussetzung, dass sich alle Teilnehmer ordnungsgemäß registrieren, braucht ein Überwachungswerkzeug kein aufwändiges Discovery durchzuführen. Alle Basisinformationen können über das iSNS abgegriffen werden.

Neben iSNS gibt es noch andere Mechanismen für das iSCSI-Management. Beispiele sind das Service Location Protocol (SLP), das Domain Name System (DNS) etc.

iSNS ist heute noch nicht weit verbreitet.

3.7 Management

Speichernetzwerke, egal ob Fibre-Channel- oder iSCSI-basiert, müssen natürlich auch administriert werden. Speziell mit iSCSI wird hier auch sehr oft die Frage aufkommen, wer ein „iSCSI Network“ administriert: die Netzwerk- oder die Storage-Administratoren im Unternehmen. Aufgrund der spezifischen Anforderungen von Speichernetzwerken steht dabei weniger die Integration in die bekannten Netzwerk-Management-Systemen, sondern vielmehr die Integration in das System- und insbesondere das Speichermanagement im Vordergrund. Das Management von iSCSI Devices sollte eine wichtige Bedeutung bekommen, da mit Hilfe von iSCSI-Speicher, der bislang direkt an den Systemen angeschlossen war (Direct Attached Storage = DAS) in Netzwerk-Storage umwandelt werden kann, was speziell Low und Midrange Server betrifft. Einfaches Storage Management ist in diesem Umfeld sehr wichtig, da die Administratoren dieser Server sehr oft mit vielen Aufgaben betraut sind und dadurch Generalisten und keine Storage Spezialisten sind.

Nachdem in der Vergangenheit hersteller-spezifische, proprietäre Management-Systeme vorherrschten, hat sich mittlerweile die Storage Mangement Interface Specification (SMI-S) als Standard mit breiter Unterstützung durchgesetzt.

Hinsichtlich des Managements müssen die zwei verschiedenen Möglichkeiten von iSCSI-Implementationen unterschieden werden. Diese sind:

- iSCSI Native auf dem Speichersystem: Es besteht eine durchgehende iSCSI Verbindung vom Server bis zum Speichersystem
- iSCSI Gateway Funktion: Hier ist der Server mit einem iSCSI Gateway im SAN verbunden, das die Umsetzung von iSCSI zu Fibre Channel vornimmt.

Bei der iSCSI Native Anbindung ist es wichtig, dass die Management-Funktionen des Speichersystems eine Integration in vorhandene Netzwerk-, Server- und Speichermanagement-Funktionen anbietet.

Bei der iSCSI-Gateway-Anbindung müssen die entsprechenden Funktionen auf der iSCSI-Gateway-Seite vorhanden sein, um eine Integration in die vorhandene Management–Umgebung zu ermöglichen. Da aber in aller Regel eine Umsetzung von iSCSI auf Fibre Channel im Gateway erfolgt, also IP-Adressen Fibre-Channel-Adressen zugeordnet werden, ist dies für vorhandene Management Systeme transparent. Es ist allerdings wichtig, dass der iSCSI-Verbindungsteil im Management entsprechend erkennbar und konfigurierbar ist.

Die Möglichkeit, über Ethernet Storage-Konsolidierung zu betreiben, kann zu einer sehr hohen Anzahl von Servern mit zentralisiertem Storage führen. Diese Zahl kann, je nach Umfeld, u.U. höher sein als die Zahl der Systeme beim gleichen Anwender, die über ein Fibre Channel SAN angebunden sind. Im Gegenzug sind aber diese System typischerweise weniger geschäftskritisch als die in der Fibre Channel SAN-Umgebung. In diesem Umfeld wird es dann immer wichtiger, dass die Management Tools sich als Betriebssystem-Tools darstellen und die Administratoren sich nicht um das spezielle Storage Management kümmern. Die Aufgabe des Administrators hier ist, sich um die Anwendung und Betriebsystem zu kümmern und nicht um Storage Management.

3.8 Zertifizierung / Freigabe

Eine der komplexesten Herausforderungen von Speichernetzwerken ist die Sicherstellung der Interoperabilität auch von heterogenen Umgebungen. Hierzu betreiben alle namhaften Hersteller aufwendige Tests mit abschließenden Zertifizierungen.

Das Erfordernis solcher Tests und Zertifizierungen gilt gleichermaßen für Fibre Channel als auch iSCSI basierte Speichernetzwerke.

Mit fortschreitender Standardisierung, sowohl im Bereich von Fibre Channel als auch iSCSI, ist eine Vereinfachung dieser Interoperabilitätsthematik zukünftig zu erwarten. Allerdings lässt die weitreichende Standardisierung im IP-Umfeld nicht gleichzeitig den Rückschluss zu, dass iSCSI hier einen Vorsprung hat. Ein Schwerpunkt der Tests und Zertifizierung betreffen heute die SCSI-Protokollebene und somit sowohl das Fibre-Channel-als auch das iSCSI-Protokoll.

Im Unterschied zu Fibre Channel gibt es allerdings durch die Software iSCSI Treiber eine direkte Betriebssystemunterstützung.

Folgende Betriebsysteme unterstützen iSCSI direkt durch eigene Treiber:

- HP UX Version
- IBM AIX Version

- Microsoft
- Novell
- Suse Linux

Hier sollten immer die Version geprüft werden.

Die Betriebsystemhersteller wie zum Beispiel Microsoft zertifizieren hier die iSCSI Implementierungen.

Die neueren Microsoft Betriebsysteme beinhalten einen iSCSI Treiber. Es ist damit zu rechnen, dass dieser Treiber zukünftig auch Multipathing unterstützt. Kosten und Komplexität sind bei einem Einsatz von iSCSI Software Treibern sicher geringer als mit TOEs. Der Nachteil der etwas größeren CPU Last ist bei den heutzutage zur Verfügung stehenden CPU Kapazitäten sicher kein großes Problem mehr.

3.9 Sicherheit

Wie bei allen Netzwerken ist auch bei Speichernetzwerken das Thema Sicherheit nicht zu vernachlässigen.

Fibre Channel ist dabei von seiner eigentlichen Konzeption nur bedingt gefährdet. Eine Manipulation der eigentlichen transportierten Daten ist nach heutigem Stand der Technik praktisch nicht möglich, ohne bemerkt zu werden bzw. ohne zu erheblichen Auswirkungen wie Systemabstürzen zu führen. Ein „Abhören“ von Daten ist zumindest bei Glasfaser-Verkabelung technisch nicht möglich, ohne entsprechende Komponenten („Splitter“) in den Datenpfad einzuhängen. Diese führt dann allerdings zu einer vorübergehenden I/O-Unterbrechung, die normalerweise nicht unbemerkt bleibt.

Daher sollte die Intaktheit eines Speichernetzwerks konstant überwacht, eventuelle Störungen automatisch protokolliert und die Ursachen von Störungen grundsätzlich erforscht werden.

Auch dehnen sich Speichernetzwerke in den meisten Fällen nur in Rechenzentren sowie u.U. über ein Firmengelände aus, sind also für Firmenfremde nicht zugänglich.

Kritischer ist der Zugriff auf die involvierten Komponenten im Datenpfad wie Switches und Direktoren. Da diese Komponenten zu Administrationszwecken ebenfalls an ein Local Area Network (LAN) angeschlossen werden sollten oder eventuell sogar müssen, muss für dieses Management-Netzwerk eine erhöhte Sicherheitsanforderung gestellt werden. Es muss sichergestellt sein, dass nur berechtigte Administratoren Zugriff auf diese Komponenten haben.

Zu den möglichen Maßnahmen zählen u.a.:

- Physische oder logische Trennung des Management-Netzwerks von anderen LANs
- Benutzername/Passwort beim Login auf den Komponenten
- Authentifizierung über z.B. RADIUS (Remote Authentication Dial-In User Service)
- Benutzerabhängige Rechte
- Access Control Lists (ACLs)
- Verschlüsselter Zugriff (Secure Shell (SSH), Secure Socket Layer (SSL) usw.)

Zahlreiche Funktionalitäten sind bereits standardmäßig seitens der Komponenten-Anbieter berücksichtigt. Höhere Anforderungen können über optional zusätzlich erhältliche Lösungen realisiert werden.

Einen Sonderfall stellen noch Verbindungen über öffentliches bzw. Fremdgelände dar. Um hier einen Eingriff von Außen zu verhindern, bieten sich die nachfolgenden Möglichkeiten an:

- Verschlüsselung der transportierten Daten
- Digitale Zertifikate zwischen den Komponenten

Bei der Verschlüsselung der transportierten Daten ist aber wieder darauf zu achten, dass die Latenzzeit nicht und nur in geringem Maße beeinträchtigt wird. Aus diesem Grunde kommen hier nur leistungsstarke, hardware-basierte Lösungen in Frage.

Alle oben genannten Punkte sind auch bei einer auf iSCSI basierenden Speichervernetzung anwendbar. Es ist sicher klar, dass es wesentlich mehr Know How und Werkzeuge gibt um TCP/IP Traffic abzuhören bzw. aufzuzeichnen. Daher hat die IETF ein Rahmenwerk zur Sicherung von Block-Speicher Protokollen über IP Netzwerke entwickelt – draft-ietf-ips-securitry.

Der iSCSI Traffic kann grundsätzlich mit den vorhandenen IP-Sicherheitsfunktionen geschützt werden wie Verschlüsselungs-Access-Listen usw. Ein Beispiel wäre die Nutzung von IPSec (IP Security Protocol) Tunnels für iSCSI Traffic. Allerdings ist dabei zu berücksichtigen, dass eine typische IPSec-Verschlüsselung implementiert z.B. auf einem Router signifikante Latenzzeit-Auswirkungen hat, so dass deren Einsetzbarkeit im Einzelfall zu prüfen ist!

Es sind mittlerweile Produkte verfügbar die IPSec in Hardeware implementiert haben, so dass ohne zusätzliche Latenzzeit ein Verschlüsseln des iSCSI Verkehrs möglich ist.

IP Security (IPSec) ist ein Rahmenwerk von offenen Standards von der IETF Entwickelt. IPSec bietet privacy, integrity und authenticity für Informationen die über ein IP Netzwerk transportiert werden. IPSec arbeitet auf dem Netzwerk Layer (Layer 3) schützt und authentiziert IP Pakete die zwischen IPSec Teilnehmer (IPSec Peers) ausgetauscht werden. IPSec bietet die folgenden Sicherheitsfunktionen:

- **Data Confidentiality**. Der IPSec Sender kann die Daten vor der Übertragung im Netzwerk verschlüsseln.
- **Data Integrity**. Ein IPSec Empfänger kann IP Pakete des Senders vor dem Empfang authentifizieren und damit sicherstellen das die Paket während der Übertargung nicht geändert worden sind.
- **Data Origin Authentication**. Der IPSec Empfänger kann die Quelle der Pakete authentifizieren.
- **Anti-Replay**. Der IPSec Empfänger kann Replayed Pakte erkennen und zurückweisen.

Mit IPSec können Daten über ein Netzwerk sicher transportiert werden, ohne dass diese abgehört und / oder geändert werden können. IPSec eignet sich daher auch vorzüglich für die Sicherung von Speicherdaten über eine IP Verbindung.

Ein wichtiger Punkt ist die Authentifizierung. Darüber kann sichergestellt werden dass nur berechtigte Server eine iSCSI Verbindung aufbauen können. Dafür kann iQN (iSCSI qualified name) – Registrierung über CHAP (Challenge Handshake Authentication Protocol) genutzt werden. Diese Funktionen – Verwaltung von iSCSI Accounts – können auf einem RADIUS Server zentralisiert werden.

Die Zuordnung der iSCSI Targets kann dynamisch oder statisch erfolgen. Eine statische Zuordnung bietet eine etwas größere Sicherheit als eine dynamische Zuordnung durch bessere Kontrolle.

3.10 Kosten

Vier Anforderungen verändern das Gesicht der Fibre Channel Welt maßgeblich.

- Kosten
- Durchsatz/Geschwindigkeit
- Distanz und
- Interoperabilität.

Die Kosten der Anbindung (Fibre Channel HBAs, FC-Switch Ports usw.) kleinerer Server an zentralen Storage sind im Vergleich zu den eigentlichen Anschaffungskosten des Server immer noch relativ hoch, auch wenn der Preisverfall bei HBAs und Switches in den vergangenen Jahren stark zur Attraktivität dieser Technologie auch für mittlere und kleinere Server beigetragen hat. Daher rückt iSCSI als weniger performante Variante immer mehr in den Vordergrund.

Hier ist allerdings zu beachten, dass die Netzwerkbandbreite und die Art der Netzwerkkarten (NIC, TOE-Karte oder iSCSI-HBA) von entscheidender Wichtigkeit bei der Kalkulation der Bandbreite und der Kosten sind.

Die Kosten von iSCSI sind so flexibel wie das Protokoll selbst und lassen sich an die verschiedenen Anforderungen der Umgebung anpassen. Benötigt ein Server hohe Leistung, muss man eine TCP Offload-Engine (TOE), NIC (TNIC) oder einen iSCSI HBA einsetzen. Diese liegen in ähnlichen preislichen Regionen wie Fibre Channel HBA. Ist die Leistung eines Rechners nicht kritisch, kann man iSCSI mit einem sehr niedrigen Kostenfaktor implementieren. Die meisten Server sind bereits mit zwei NICs ausgestattet und verfügen über ausreichend CPU-Leistung, um die benötigten Anforderungen zu erfüllen. iSCSI ermöglicht eine problemlose Integration in eine bestehende LAN-Infrastruktur, wenn diese ausreichend Bandbreite bietet. Die netzwerkseitigen Kosten von IP Ports sind im Augenblick noch wesentlich geringer als die Kosten von FC Ports im SAN.

Am Speichersubsystem muss man die Kosten unter den Gesichtspunkten Skalierbarkeit, Lösungsflexibilität und I/O-Operationen betrachten:

- Macht es Sinn, am Storage-Subsystem einen FC Controller gegen einen iSCSI-Zugang zu tauschen und dafür weniger FC Ports zu haben?
- Wird der gesamte Durchsatz des Speichers durch den Overhead der De Encapsulation verringert?
- Ist es sinnvoller, sich die Fibre-Channel-Flexibilität am Storage-Subsystem zu erhalten und ein iSCSI Gateway statt eines iSCSI Controller im Subsystem einzusetzen?

Insgesamt ist davon auszugehen, dass die Kosten mit den Anforderungen an Durchsatz, Antwortzeiten und Verfügbarkeit steigen.

4 Unterstützung

Die Unterstützung für iSCSI muss durch die kooperative Arbeit zwischen den drei verschiedenen IT-Gruppen erbracht werden. Server-, Speicher- und die Netzwerk-Administratoren müssen sich hierbei ergänzen. Wichtig ist zu bemerken, dass die gesamte Lösung mehrere Ebenen beinhaltet.

Wenn es zum Beispiel bei einer Anwendung zu Schreibfehlern kommt oder es Probleme beim Zugriff auf eine Logische Units (LUN) gibt, kann der Netzwerk-Administrator dieses Problem nicht alleine lösen.

Des Weiteren muss auch sicher gestellt sein, dass die angestrebte iSCSI-Lösung auch von der Anwendung selbst (bzw. von dem Hersteller) unterstützt wird.

Gerade bei kleineren und mittleren Unternehmen können Probleme nicht immer detailliert vorqualifiziert werden, da die verantwortlichen Administratoren eine sehr große Breite an Themen abzudecken haben. Hier ist die klare Empfehlung, möglichst homogene Lösungen aus einer Hand einzusetzen um das so genannte „Finger Pointing“ zu vermeiden und kurze Kommunikations- und Eskalationswege zu garantieren.

5 Schlussfolgerungen

iSCSI ist ein leistungsfähiges Protokoll für Speichernetzwerke auf Basis des de-facto Standards TCP/IP. Der Einsatz kann je nach Anwendungsfall eine interessante Alternative zu Fibre Channel sein.

Besonders geeignet ist iSCSI für den Einsatz im Bereich von Low-Cost-Servern mit geringen I/O-Anforderungen. Für unternehmenskritische Anwendungen mit hohen I/O- und Verfügbarkeitsanforderungen bietet Fibre Channel jedoch die besseren Möglichkeiten. Der Bereich zwischen diesen beiden Klassen ist fließend und muss in jedem Fall einzeln betrachtet und bewertet werden.

Die Frage, ob ein SAN auf Basis von iSCSI tatsächlich, wie oft behauptet, billiger als auf Basis von Fibre Channel ist, hängt stark von den Anforderungen und Gegebenheiten des Einzelfalls ab und kann so nicht pauschal beantwortet werden.

Die Kombination von Fibre Channel und iSCSI birgt zusätzliche Komplexitäten, so dass sie nur bei einer entsprechenden Größenordnung (d.h. hohen Anzahl von Servern sowohl für Fibre Channel als auch iSCSI) eine interessante Option darstellt.

6 Anwendungsbeispiele reine iSCSI-Infrastruktur bzw. iSCSI Gateway

6.1 Anwendungsszenario: Universitätsklinik

Eine Universitätsklinik hatte bereits eine Storage-Gigabit-Ethernet-Struktur mit NAS Storage für seine Picture Archiving and Communication Systeme (PACS). Im Anschluss wurden zusätzlich SQL-Server- und Exchange-Systeme mit iSCSI auf einer zentralen Storage-Plattform konsolidiert. Die Klinik nutzte das vorhandene physikalische Ethernet-Netzwerk in Verbindung mit iSCSI, allerdings wurde eine logische Trennung des iSCSI-Traffics vom übrigen NAS- und sonstigen IP-Traffic mit Hilfe der VLAN-Technologie. Zudem greift ein ERP System auf die IP-SAN-Lösung zu; zahlreiche kleine Anwendungen – wie etwa eine Datenbank mit Mikroskopaufnahmen oder das Klinik-Intranet – wurden von ihren Speichersubsystemen auf zentralen Speicher mit iSCSI migriert.

Möglich wurde dieses durch die relativ geringen I/O-Anforderungen der jetzt iSCSI-basierten Anwendungen.

Die wichtigsten Vorteile liegen hier in der Nutzung der bestehenden physischen Ethernet-Netzwerkinfrastruktur und der Zentralisierung der Storage-Ressourcen, so dass ein einheitliches Netzwerk und einfacheres Storage-Management sowie eine geringere Total Cost of Ownership (TCO) erreicht werden.

6.2 Anwendungsszenario: Fertigungsindustrie

Ein Kunde aus der Fertigungsindustrie setzt für seine ERP Anwendung zentralisierte Storage-Ressourcen in Form eines FC-SAN ein und nutzt gleichzeitig NAS für das File Serving.

iSCSI ist hier die ideale Ergänzung, um die Verfügbarkeit der Fertigungssteuerungsservern in den Fertigungshallen zu erhöhen, ohne dazu mit hohem Aufwand zusätzliche Fibre Channel Verkabelung in die Produktionshallen zu verlegen.

In den Produktionshallen wird die bestehende Gigabit Ethernet Infrastuktur benutzt, um über iSCSI die Fertigungsteuerungsserver zu booten und damit „diskless“ zu betreiben. Somit sind weder lokale Daten noch Konfigurationen auf den Servern in den Produktionshallen, was die Datensicherheit gegenüber einer Lösung mit lokalen Platten an den Produktionssteuerungsservern deutlich erhöht.

iSCSI bietet hier in mehrerlei Hinsicht Kostenvorteile. Der Kunde kann mit iSCSI auf der Ethernet-Infrastruktur in den Fertigungshallen aufsetzen und vermeidet so Investitionen in eine Fibre-Channel-Struktur. Darüber hinaus erübrigen sich zusätzliche Kosten beim Storage Management: Die Verwaltung des verwendeten, zentralen Storage Arrays macht zwischen iSCSI und Fibre-Channel keinen Unterschied, so dass das IT-Personal nicht zusätzlich belastet wird. Der Faktor Performance war in diesem Fall nachrangig, da die Server in den Fertigungshallen sehr geringe I/O-Anforderungen haben.

7 Links

BITKOM: www.bitkom.org

SNIA IP Storage Forum: http://www.snia.org/ipstorage/home

Storage Networking Solutions Europe (SNS Europe): http://www.snseurope.com/

IETF (Internet Engineering Task Force): http://www.ietf.org/html.charters/ips-charter.html

8 Abkürzungsverzeichnis

ACL	Access Control Lists
ANSI	American National Standard Institute
ATM	Asynchronous Transfer Mode
CHAP	Challenge Handshake Authentication Protocol
CPU	Central Processing Unit
DAS	Direct Attached Storage
DHCP	Dynamic Host Configuration Protocol
DNS	Domain Name System

DR	Desaster Recovery
ERP	Enterprise Resource Planing
ESCON	Enterprise Systems Connectivity
FC	Fibre Channel
FCAL	Fibre Channel Arbitrated Loop
FCIA	Fibre Channel Industry Association
FCIP	Fibre Channel over IP
FCP	Fibre Channel Protocol
FTP	File Transfer Protocol
HA	High Availability
HBA	Host Bus Adapter
IETF	Internet Engineering Task Force
iFCP	Internet Fibre Channel Protocol
I/O	Input/Output
IP	Internet Protocol
IPSec	IP Security Protocol
iQN	iSCSI qualified name
iSCSI	Internet SCSI
ISL	Inter Switch Link
iSNS	Internet Storage Name Service
LAN	Local Area Network
LUN	Logical Unit Number (Hinweis: Es hat sich eingebürgert die Logischen Einheiten selbst als LUN zu bezeichnen
NAS	Network Attached Storage
NIC	Network Interface Card
PACS	Picture Archiving and Communication System
PXE	Pre-Boot Execution Environment
RADIUS	Remote Authentication Dial-In User Service
RFC	Requests for Comments
SAN	Storage Area Network
SLP	Service Location Protocol
SMI-S	Storage Management Interface Specification
SNA	System Network Architecture
SNIA	Storage Networking Industry Association
SPoF	Single Point of Failure
SQL	Sturctured Query Language
SSH	Secure Shell

SSL	Secure Socket Layer
TCO	Total Cost of Ownership
TCP	Transmission Control Protocol
TFTP	Trivial File Transfer Protocol
TNIC	TOE Network Interface Card
TOE	TCP/IP Offload Engine
VLAN	Virtual LAN

Bundesverband Informationswirtschaft,
Telekommunikation und neue Medien e.V.
Albrechtstraße 10
10117 Berlin-Mitte

Tel.: 030/27 576 - 0
Fax: 030/27 576 - 400

bitkom@bitkom.org
www.bitkom.org