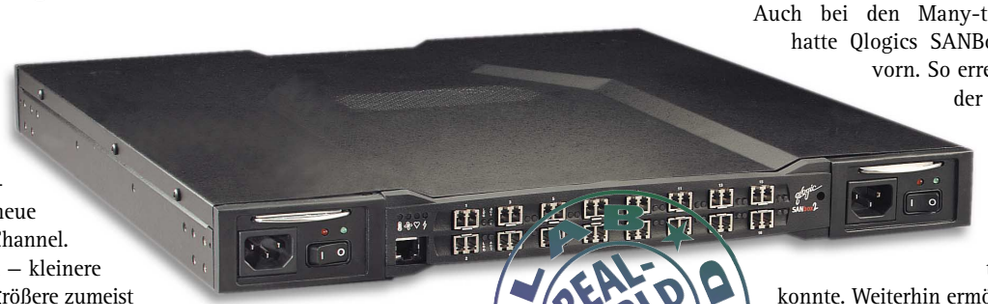
Qlogic SANBox 2-16

Formel 1 der Storage-Netze

Echte Highspeed-Perfomance bieten SAN-Switches mit 2 GBit/s Nennleistung pro Port. Wie schnell die Fibre-Channel-Switches der neuen Generation unter Last wirklich sind, musste Qlogics »SANBox 2-16« in einem Test unserer Real-World Labs beweisen.



Speichernetzwerke, die so genannten Storage-Area-Networks oder kurz SANs, revolutionieren die IT-Landschaft in den Unternehmen. Und mit ihnen hält eine für viele IT-Verantwortliche neue Netzwerktechnologie Einzug: Fibre-Channel. Aktive Fibre-Channel-Komponenten – kleinere Komponenten werden als Switches, größere zumeist als Directors gehandelt – sollen für die hochperformante Verbindung der unterschiedlichen Storage-Systeme und der Server sorgen. Wir wollten wissen, wie durchsatz- und leistungsstark diese Systeme wirklich sind und wie sie sich verhalten, wenn es im Speichernetz mal eng wird. Für unsere Tests stellte uns Hitachi Data Systems ihr Storage-Lab in Feldkirchen bei München zur Verfügung. Im Mittelpunkt unseres ersten FC-Switch-Tests stand die »SANBox 2-16« von Qlogic.

Best-Case-Durchsatz – Fully-Meshed und Many-to-one

Das Test-Setup Fully-Meshed ermöglicht es, den maximalen Datendurchsatz eines Switches zu messen. Dazu haben wir die Spirent-Smartbits so konfiguriert, dass der Traffic den gesamten Switch belastet. Jeder der hierbei eingesetzten 16 Ports sendet zu allen anderen Ports seine Frames. Das heißt, es besteht pro Port eine 1:n-Beziehung zu allen weiteren verwendeten Ports. Dabei haben wir automatisch die Bandbreite jedes Ports inkrementell von 25 Prozent bis auf 100 Prozent Last angehoben. Dies geschah in einer Schrittweite von 25 Prozent. Da alle Ports gleichzeitig senden und empfangen mussten, ergibt sich ein bidirektionaler Datenverkehr. Dieser würde im Idealfall für jeden einzelnen Port 100 Prozent des theoretisch erreichbaren Maximalwertes an Nutzdaten betragen. Messtechnisch ermittelt wird bei diesem Test-Setup der Gesamtdurchsatz aller eingesetzten Ports. Diesen haben wir dann auf die Pro-Port-Leistung umgerechnet und zur Ermittlung eines hypothetischen Nutzdatendurchsatzes unterstellt, dass 36 Byte je Frame für den Header erforderlich sind. Der tatsächliche Nutzdatendurchsatz kann je nach Communication-Model und diversen anderen Faktoren noch deutlich geringer ausfallen.

Mit diesem Test haben wir die Performance der zu testenden Geräte untersucht. Zum einen wird hier der Prozessor des Switches getestet, indem man kleine Framegrößen verwendet, zum anderen wird der Speicher durch die großen Frames geprüft. In dieser Messung haben wir Frames von der Größe von 64 Byte, 1024 und 2148 Byte generiert. Dabei werden mit kleinen Frames die CPUs beziehungsweise die Switching-ASICs in den Switches am meisten belastet. Da meist mehrere Ports auf eine CPU geschaltet werden, kann es in diesem Bereich zu Engpässen kommen. Wenn im SAN viele Datenbank-Anfragen stattfinden, ist mit kleinen Frames zu rechnen. Hat ein Switch Probleme mit kleinen Frames, können diese zu nicht vorhergesehenen Engpässen im SAN führen.

Bei allen drei Datenrahmengrößen bewies sich Qlogics SANBox-2-16 als sehr performant. Bei der Messung mit 64-Byte-Paketen kam der Switch auf einen maximalen Bruttodurchsatz von 148,31 MByte/s und einen Nutzdatendurchsatz von 64,89 MByte/s und verfehlte somit das theoretische Maximum von brutto 154,55 MByte/s und netto 67,61 MByte/s nur knapp. Bei den größeren Datenrahmen verringerte sich der Abstand der gemessenen Durchsätze zu den theoretisch möglichen weiter. Beispielsweise schaffte das Qlogic-System bei der Messung mit 2148-Byte-Paketen einen Bruttodurchsatz von 210,09 MByte/s gegenüber einem theoretischen Maximalwert von 210,15.

Das Many-to-One-Szenario ähnelt dem vorhergehenden Fully-Meshed-Test-Setup, allerdings senden hierbei 15 Eingangs-Ports unidirektional auf einen einzigen Ausgangs-Port. Bei diesem Test-Setup haben wir automatisch die Bandbreite jedes Ports inkrementell von 5 Prozent bis auf 20 Prozent Last mit einer Schrittweite von 5 Prozent angehoben, was für den Ausgangsport eine Last von 75 bis 300 Prozent bedeutet. Auch dieser Test ermittelt die Performance der zu testenden Geräte, variiert aber deren Belastung. Auch in diesem Test-Setup haben wir Frames in Größen von 64, 1024 und 2148 Byte generiert. Messtechnisch ermittelt dieser Test den maximalen Bruttodatendurchsatz am Ausgangsport des Switches, an den alle 15 Eingangsports senden. Auch bei den Many-to-One-Messungen hatte Qlogics SANBox-2-16 die Nase vorn. So erreichte er schon bei der Messung mit 64-Byte-Rahmen die theoretische Wirespeed, die er auch bei den größeren Datenrahmen halten konnte. Weiterhin ermöglichst das Many-to-One-Test-Setup eine Analyse der Durchsatz-Fairness. Der Switch von Qlogic behandelte alle Ports gleich und verteilte den Datendurchsatz somit gleichmäßig auf alle Ports.

Latency und Datendurchsatz – die Worst-Case-Falle

Die ermittelten Durchsatzraten oben basieren auf der Annahme, dass die Datenquelle sendet, ohne dass sie auf weitere Signalisierungen wartet, die sie veranlasst, den nächstfolgenden Datensatz zu senden. Steuern Applikationen auf diese Weise den Datenfluss, dann spielt die Latency bei heutigen FC-Switches keine allzu große Rolle – vorausgesetzt es stehen optimale Applikationen und optimale Datenstrukturen zur Verfügung. Die ermittelten Latency-Werte liegen alle deutlich unter den für Echtzeitapplikationen kritischen Grenzwerten, so dass die Verteilung beispielsweise von Videostreams über das SAN keine Probleme machen sollte. Wartet die Datenquelle jedoch auf ein Feedback-Signal, mit dem die Applikation erst den nächsten Datensatz anfordert, dann ist der limitierende Faktor für den erzielbaren Datendurchsatz die Laufzeit der Datenpakete. Solche Applikationen rufen Daten blockweise ab, diese Datenblöcke können auch einzelne Frames sein. Ein solches Verhalten zeigen beispielsweise Warenwirtschaftssysteme und diverse statistische Anwendungen, weil diese Applikation nicht »vorhersehen« können, auf welche Daten als nächstes zugegriffen werden soll, und deshalb kein sequentieller Zugriff möglich ist. Andere Applikationen sind einfach schlecht programmiert und verfügen über keine entsprechende Prozessoptimierung. Diese besteht entweder in einem Read-Ahead-Mechanismus oder in einem intelligenten Caching. Aber auch die intelligenteste Applikation kann nicht immer »vorhersehen« welche Daten als nächstes erforderlich sind, daher kann es auch bei entsprechend optimierten Anwendungen immer wieder zu Situationen kommen, in denen die Appli-

Steckbrief

Qlogic SANBox 2-16

Hersteller: Qlogic

Charakteristik: Fibre-Channel-Switch

Kurzbeschreibung: Performanter 16-Port-Fibre-Channel-Switch mit Unterstützung für 1- und 2-GBit/s-Technologie sowie Auto-Sensing-Funktionalität.

Web: www.qlogic.com

Preis: 17 995 Dollar

Info

So testete Network Computing

Damit wir für unseren Fibre-Channel-Test das erforderliche Equipment nutzen konnten stellte uns Hitachi Data Systems ihr Storage-Lab in Feldkirchen bei München zur Verfügung. Ein Teil der getesteten FC-Switches gehörten zur Ausstattung der HDS-Labs. Für die Konfiguration der FC-Systeme sorgten HDS-Techniker und -Ingenieure sowie Experten der Hersteller, die wir zum Test eingeladen hatten. Als Lastgenerator und Analysator haben wir einen »Smartbits 6000B Traffic Generator/Analyser« von Spirent eingesetzt.

Er war mit 24 Fibre-Channel-Ports, die jeweils wahlweise 1-GBit/s- oder 2-GBit/s-Datenströme erzeugen und analysieren können, bestückt. Das rund 300 000 Euro teure System war mit der Software »SmartFabric 1.20« ausgestattet.
Unser Test gliederte sich in drei Bereiche oder Test-Setups. Als erstes haben wir einen einfachen Performancetest im Fully-Meshed-Setup durchgeführt. Hier findet ein bidirektionaler All-to-all-Datenverkehr statt. Dabei senden und empfangen alle 16 Ports mit ansteigender Last Datenströme an alle anderen Ports. Im Many-to-one-Test-Setup senden dann 15 Eingangsports auf einen Ausgangsport des jeweiligen FC-Switches. Diese beiden Test-Setups ermitteln die maximal erreichbaren Datendurchsatzraten sowie das Latency-Verhalten unter Stress. Als drittes Test-Setup haben wir dann die Übertragung von Daten über Inter-Switch-Links (ISL) getestet.

kation kurzfristig Datensätze anfordern muss und somit die Latency den Datenstrom zum »stottern« bringen kann. Je nach Applikation und Art der Daten ist die Latency ein größeres oder kleineres Problem, ignorieren kann man sie in keinem Fall, wie auch die derzeitige Erfahrung in der Praxis zeigt.

Um dem Latency-Effekt nachzugehen haben wir daher als zweiten Parameter die Latency im Many-to-One-Szenario näher betrachtet. Dabei haben wir die Latenzzeiten gemessen und unterstellt, dass sie der gesamten Round-Trip-Time des Systems entspricht. Dieser Ansatz idealisiert das Szenario dahingehend, dass sie von einer technisch nicht möglichen Signallaufzeit für die Bestätigung von 0 µs entspricht. Somit ist der zeitliche Mindestabstand zwischen zwei Paketen in unserem Modell die Latency-Dauer, die tatsächliche Round-Trip-Time liegt in allen Fällen noch darüber, ist aber in unserem Test-Setup nicht exakt zu bestimmen. Die Latency wirkt sich um so kritischer aus, je kürzer die verwendeten Frames sind. Die Grenzwerte, ab deren Überschreitung die Latency den Datendurchsatz negativ beeinflusst, liegen recht niedrig. So ist von einer Reduzierung des Durchsatzes bei 64-Byte-Paketen ab 0,3 µs, bei 1024-Byte-Paketen ab 5 µs und bei den 2148-Byte-Paketen ab rund 10 µs auszugehen. Oberhalb von diesen Schwellwerten werden entsprechende Anwendungen in Folge ihrer Round-Trip-Time schnell zu Killerapplikationen, die das SAN nachhaltig ausbremsen.

Um diesem Effekt weiter nachzugehen haben wir die Messwerte der Many-to-One-Messungen auf ihre Latency-Werte und die durch die Latency zu erwartenden maximalen Durchsatzraten bei den Messungen mit den verschiedenen Frame-Größen ermittelt. Als Basis haben wir die Mittelwerte der maximalen und die Mittelwerte der minimalen Latency-Werte bei 100 Prozent Last zu Grunde gelegt und dann noch den Mittelwert der beiden Mittelwerte gebildet. Einen recht hohen Nutzdatendurchsatz erreichte in dieser Disziplin und mit 2148-Byte-Paketen Qlogics SANBox-2-16 mit 48,82 MByte/s – wohlgemerkt bei der Auswertung der Mittelwerte der Latency-Minimalwerte. Die Auswertung der jeweiligen Latency-Spitzenwerte entspricht schon bei den Messungen mit 2148-Byte-Paketen einem völligen Datentransferkollaps. Qlogic erreicht hier noch 1,15 MByte/s. Die Werte für die Messung mit 1024-Byte-Paketen führen zu ähnlichen Resultaten. Die Ergebnisse bei den Messungen mit 64-Byte-Paketen liegen noch deutlich darunter. Zum unter Umständen massiven Problem wird die Latency-Falle hier in vielen Szenarien. Hinzu kommt, dass sich die Latency addiert, wenn die Datenströme mehrere Switches durchlaufen müssen, die beispielsweise über ISL miteinander verbunden sind. Verdoppelt oder potenziert sich aber die Latency, dann verschlechtert sich der Datendurchsatz des Gesamtsystems um den entsprechenden Faktor.

ISL-Tests – fair play

Da es in der Praxis in einem SAN nicht nur zum Einsatz eines Switches kommt, sondern sich die Fabric über mehrere Koppelelemente erstreckt, ist es von Interesse, das Handling der Ports innerhalb des Switches zu testen. Das Inter-Switch-Link-Test-Setup trifft eine Aussage über die Fairness der einzelnen Ports bei gezielter Überlastung. Wir haben jeweils zwei baugleiche Switches in zwei Messreihen auf sieben beziehungsweise acht Fibre-Channel-Datenports beschaltet und mit dem Spirent-Lastgenerator/Analysator verbunden. Vier ISL-Leitungen stellten die Verbindung zwischen den beiden Switches her. Die Kommunikation fand unidirektional statt. Dabei haben wir die Bandbreite jedes Ports inkrementell von 25 bis auf 100 Prozent angehoben. Wir haben wieder eine Schrittweite von 25 Prozent sowie je 64-, 1024- und 2148-Byte-Rahmen verwendet. Die theoretische Überlastung der ISL sollte bei einem Verhältnis von acht beziehungsweise sieben zu vier liegen, da je nur vier ISL-Ports zu Verfügung standen. Bei einer Überlastung der ISL-Strecke treten auch hier keine Frameverluste auf, da wieder die Flusskontrolle den Input beschränkt, jedoch ist es interessant zu wissen, wie die Last auf den einzelnen Ports verteilt wird. Hier stellt sich heraus, ob die Lastverteilung fair erfolgt oder ob einige Ports konstruktionsbedingt bevorzugt werden. Ferner kann die Performance dieser Zusammenschaltung analysiert werden.

Bei den Messungen mit sieben Input- und sieben Output-Ports fällt bei allen Framegrößen auf,

dass Stream D mit rund dem doppelten Durchsatz arbeitet wie die anderen Streams. Dieser Befund zeigt, dass der Qlogic-Switch keine echte Lastverteilung beherrscht, sondern nur die zu befördernden Streams auf die verschiedenen Links verteilt und so das gesamte System nicht immer optimal auslasten kann. – Oder anders formuliert, sechs der sieben Streams teilen sich je einen Link, dem siebten Stream steht exklusiv ein eigener Link zur Verfügung, seine Daten haben »Vorfahrt« im Switch. Dieses Design ist aus Sicht des Herstellers verständlich, da es zumeist nicht nur wichtig ist, Datenverluste zu vermeiden, sondern auch sicherzustellen, dass der Switch mit In-Order-Delivery arbeitet, die Frames also auch in der richtigen Reihenfolge ankommen sollen. Bei den Messungen mit acht Input- und acht Output-Ports behandelte das Qlogic-System dann alle Datenströme gleich. Je zwei Datenströme mussten sich einen Link teilen, dies erfolgte fair, die Durchsatzraten erreichten bei allen Messungen für jeweils alle Streams innerhalb einer Messung praktisch identische Werte.

Head-of-Line-Blocking

Fibre-Channel-Systeme arbeiten mit Input-Pufferspeichern, die die eingehenden Datenströme zwischenspeichern und dafür sorgen, dass der Input-Port nicht mehr Daten weiterleiten, als die adressierten Output-Ports verarbeiten können. Drohen diese Input-Pufferspeicher »überzulaufen«, reduziert der Switch über den Flow-Control-Mechanismus den Output der Datenquelle, also im Fall unseres Test-Setups des Smartbits-Lastgenerators. Durch diesen Mechanismus vermeiden die Fibre-Channel-Geräte, dass sie unter Überlast Datenrahmen verlieren. Diese Funktionsweise führt unter bestimmten Umständen zum so genannten Head-of-Line-Blocking. Die Fibre-Channel-Ports 1 und 2 am Switch arbeiten als Eingangsports, die Ports 3 und 4 als Ausgangsports. Port 1 empfängt zwei Datenströme, die zu gleichen Teilen an die Ports 3 und 4 adressiert sind. Port 2 empfängt einen Datenstrom, der an Port 3 adressiert ist.

Bei 100 Prozent Eingangslast auf den Ports 1 und 2 würde dieses Szenario eine Überlast von 150 Prozent an Port 3 und eine Last von 50 Prozent an Port 4 bedeuten. Die drohende Überlast und damit verbundene Datenverluste verhindert das System, indem es Input-regulierend eingreift. Arbeitet der Switch non-blocking, dann schöpft er die zur Verfügung stehenden Bandbreiten soweit möglich aus. Das bedeutet, dass er auf Output-Port 3 die maximale Durchsatzrate erreicht, Port 4 arbeitet mit maximal 50 Prozent der maximalen Durchsatzrate, ein höherer Durchsatz ist hier technisch nicht möglich, da sich Port 4 die Datenraten des Input-Ports 1 mit dem Output-Port 3 teilen muss. Kommt es zum Head-of-Line-Blocking, reduziert die Flow-Control den ohnehin nur mit maximal 50 Prozent Leistung belasteten Port 4 weiter im Durchsatz.

Das Head-of-Line-Blocking-Verhalten ist für das Design von SANs durchaus wichtig. Schreibt in einem solchen Szenario beispielsweise ein Server gleichzeitig auf einem schnellen Festplattensystem und auf einem deutlich langsameren Backup-System, die entsprechend angeschlossen sind, dann kann die Performance des Plattensystems unter Umständen nicht mehr voll ausgenutzt werden, weil es quasi durch das deutlich langsamere Backup-System indirekt ausgebremst wird.

Für Qlogics SANBox-2-16 war Head-of-Line-Blocking kein Thema mehr, dieses Verhalten konnten wir dem Switch in keiner Messung signifikant nachweisen. Seine maximalen Durchsatzraten entsprachen den Werten, die auf Grund der vorhergehenden Many-to-One-Messungen zu erwarten gewesen sind.

Fazit

Effektiv zu erzielende Durchsatzraten von Fibre-Channel-Switches hängen von einer ganzen Reihe von Faktoren ab. Faktoren wie die Auslegung von Pufferspeicher oder die Leistungsfähigkeit von Prozessoren limitieren die erzielbaren Ergebnisse. Hinzu kommen technologiebedingte Faktoren, die die Nutzdatenbandbreite zwangsläufig reduzieren, wie die 8-Bit-10-Bit-Codierung oder der Overhead, der auf Header und Signalisierungsverkehr zurückzuführen ist.

Geringe Probleme hatte der Qlogic-Switch allenfalls in den Durchsatztests bei den Messungen mit 64-Byte-Paketen. Ansonsten kam er den theoretischen Maximaldurchsätzen sehr nah. Auch in Bezug auf die Fairness bei der Kommunikation über Inter-Switch-Links ist gegenüber unseren vorhergehenden Tests diverser FC-Switches ein Fortschritt feststellbar, das Qlogic-System ließ sich keine Unfairness in der Behandlung der Datenströme nachweisen. Diese Aussage gilt aber nur für den Fall, dass eine gerade Zahl von Links auf eine gerade Zahl von Inter-Switch-Links sendet. Ist dies nicht der Fall, werden einzelne Links bevorzugt behandelt. Grund hierfür ist, dass der Switch keine echte Lastverteilung beherrscht, sondern nur die zu befördernden Streams auf die verschiedenen Links verteilt und so das gesamte System nicht immer optimal auslasten kann. Auch das Head-of-Line-Blocking führte in unserem Test zu keinen Problemen, sollte aber als möglicher Effekt nicht aus den Augen verloren werden.

Ein Problem der besonderen Art kann die Latency bereiten. Je nach Applikation kann es im Worst-Case zu massiven Performance-Einbußen kommen. Applikationsabhängig können Geräte wie der Qlogic-Switch im Test voll in die Latency-Falle laufen. Um diese zu vermeiden gilt es, alle Applikationen auf ihr Verhalten genau zu untersuchen und unbedingt einen gründlichen messtechnisch gestützten Testbetrieb der endgültigen Installation voranzustellen.

Insgesamt gilt auch im SAN-Umfeld, dass IT-Verantwortliche deutlich mehr Bandbreite einkaufen sollten, als sie von der reinen Papierform der Systeme her benötigen würden, sonst ist die Gefahr groß, dass es im Betrieb zu Durchsatzproblemen kommen kann. Denn von den maximalen Durchsatzraten gehen nicht nur technologiebedingte Raten wie durch Codierung oder Signalisierung ab, auch designbedingte Effekte wie die interne Lastverteilung auf die Inter-Switch-Links können für Überraschungen sorgen. Und dafür, dass die Latency-Falle nicht zuschlägt, müssen die IT-Verantwortlichen und die SAN-Designer unbedingt in der Planungsphase ihres neuen Speichernetzwerks sorgen, sonst droht im Worst-Case das SAN zum Nadelöhr zu werden.

Prof. Dr. Bernhard G. Stütz, [dg]