

Extreme Networks BlackDiamond 10808

Der Netzwerkdiamant

Highend-Switching in der 10-Gigabit-Ethernet-Klasse verspricht Extreme Networks Unternehmen, die ihren »BlackDiamond 10808« im Core einsetzen. In einem Real-World-Labs-Test von Network Computing musste der »schwarze Diamant« zeigen, was in ihm steckt und ob er lupenrein ist.

Highend-Switches der 10-Gigabit-Ethernet-Klasse sorgen seit einigen Monaten verstärkt für Furore – zumindest auf dem Papier. Wir wollten wissen, wie gut 10-Gigabit-Ethernet-Switches heute schon in aktuellen Szenarien arbeiten und ob es mit ihnen möglich ist, auch die gewünschte »Quality of Service« zu realisieren, die in modernen konvergenten Netzen heute gefordert wird. Aus diesem Grund haben wir einen Vergleichstest von 10-Gigabit-Ethernet-Switches ausgeschrieben. Um unseren Test wie gewohnt schon vor Beginn sauber zu strukturieren, haben wir zur Definition unserer Test-Spezifikation, die wir an alle einschlägigen Hersteller gesandt haben, wieder auf unser Modellunternehmen HighFair zurückgegriffen.

Unser Modellunternehmen HighFair möchte neben den klassischen Datenapplikationen und Voice-over-IP weitere Real-Time-Applikationen in ihr Unternehmensnetz integrieren. Ein geeigneter Vergleichstest sollte evaluieren, welche Switches für diese Aufgaben auch unter entsprechender Last geeignet sind. Dabei galt es, verschiedene CoS-Queuing-Mechanismen, wie Strict-Priority-Queuing, Weighted-Fair-Queuing oder Weighted-Round-Robin, auf ihre Eignung

für das geplante Szenario zu untersuchen.

Folgende Dienste sollten im LAN der HighFair integriert werden:

- ▶ Videokonferenzen (Video-over-IP, bidirektional, unicast),
- ▶ Voice-over-IP (Call-Center),
- ▶ SAP-Anwendungsdaten sowie
- ▶ übrige Datenanwendungen und Updates.

Um die möglichst absolute Störungsfreiheit der Kommunikations- und Arbeitsprozesse in unserem Modellunternehmen zu garantieren, sind eine vierstufige Daten-Priorisierung sowie eine intelligente Queuing-Policy erforderlich. Gefordert haben wir neben der Datenpriorisierung ein intelligentes Bandbreitenmanagement, das es ermöglicht, von vorneherein eine Überlastung des Backbones zu vermeiden. Aus diesem Szenario ergaben sich im Detail die folgenden Anforderungen an die Teststellungen.

Teststellung 10-Gigabit-Ethernet-Switch:

- ▶ Layer-3-Ethernet-Switch,
- ▶ Mindestens fünf 10-Gigabit-Ethernet-Ports,
- ▶ Datenpriorisierung nach IEEE 802.1p/Q auf Layer-2,
- ▶ Diffserv-Datenpriorisierung nach RFC 2474 oder
- ▶ Type-of-Service-Datenpriorisierung nach RFC 791 und/oder 1349 auf Layer-3 sowie
- ▶ mindestens 2 CoS-Queuing-Mechanismen wie Strict-Priority-Queuing, Weighted-Fair-Queuing oder Weighted-Round-Robin, die softwareseitig konfiguriert werden können.

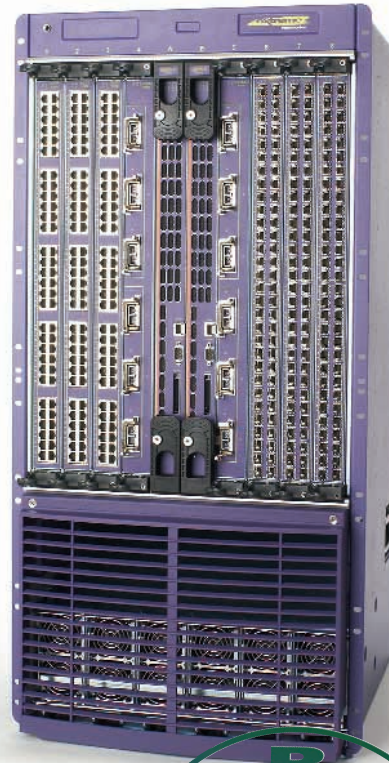
Als Testverfahren haben wir Messungen nach RFC 2544 (Many-to-One) festgelegt, die die Parameter Performance, Packet-Loss, Latency und Jitter ermitteln. Analysiert wird dann das unterschiedliche Verhalten der Systeme in den verschiedenen CoS-Queuing-Modi.

Die unserem Test zu Grunde liegende Testsuite setzt sich aus Performance-Tests sowie einigen Funktionstests zusammen. Die Performance-Tests untersuchen insbesondere die Echtzeitfä-

higkeit des Systems unter Last beziehungsweise Überlast und die Umsetzung verschiedener Class-of-Service-Strategien. Da heute LAN-Switches durch die Integration von Voice- und Video-over-IP auch im direkten Vergleich mit klassischen großen TK-Anlagen konkurrieren müssen, haben wir darüber hinaus einige Funktionstests durchgeführt, die insbesondere über die Verfügbarkeit solcher Systeme aussagefähig sind. So erlauben Systeme wie der Blackdiamond-10808 beispielsweise, Software-Updates im laufenden Betrieb durchzuführen, eine Funktionalität, die klar auf die TK-Konkurrenz zielt, die sogar den Austausch kompletter Betriebssysteme ohne Unterbrechungszeiten erlauben.

Unsere Testspezifikation haben wir an die einschlägigen Hersteller geschickt und diese eingeladen, an unserem Real-World-Labs-Test in den Labs des

Messgeräteherstellers Spirent in Kalifornien teilzunehmen. Nach anfänglich großem Interesse seitens der Hersteller häuften sich dann die Absagen. Wegen angeblich mangelnder Ressourcen oder anderer Unpässlichkeiten sahen sich dann unter anderem Alcatel, Allied Telesyn, Enterasys Networks, Force 10 Networks, Foundry Networks oder Hewlett-Packard mehr oder weniger kurzfristig nicht in der Lage, einen passenden Switch zum Test zur Verfügung zu stellen. Von allen eingeladenen Herstellern trafen dann noch Cisco und Extreme Networks tatsächlich mit einer Teststellung in den Labs in Kalifornien ein. Unglücklicherweise erwies sich der Cisco-Switch als schadhaft. Jedenfalls gelang es den Cisco-Ingenieuren nicht, das System so weit in Betrieb zu nehmen, dass die entsprechenden Messungen durchgeführt werden konnten. Auch sah sich Cisco außer Stande, innerhalb weniger Tage



Steckbrief

BlackDiamond 10808

Hersteller: Extreme Networks

Charakteristik:

10-Gigabit-Ethernet-Switch

Kurzbeschreibung: Der Blackdiamond-10808 ist ein Ethernet-Core-Switch der neuen 10-Gigabit-Ethernet-Generation. Er beheimatet maximal 48 10-Gigabit-Ethernet- sowie 480 Gigabit-Ethernet-Ports. Als gesamte Switching-Kapazität gibt Extreme 1,6 TBit/s an. Die Class-of-Service-Mechanismen nach IEEE 802.1p/Q auf Layer-2 sowie Diffserv-Datenpriorisierung nach RFC 2474 sorgen für die Sicherung der notwendigen Übertragungsqualitäten im konvergenten Unternehmensnetz. Eine Vielzahl an Security- und Management-Features vervollständigt die umfassende Funktionalität des Blackdiamond-10808. Als Switch-Betriebssystem wird eine Linux-Variante verwendet.

Web: www.extremenetworks.com

Preis: ca. 360 000 Euro
je nach Ausstattung

mit einem intakten System zu einem zweiten Messtermin zu kommen. Von daher hatte Extreme Networks den »Verglechtest« schon früh gewonnen – sie war die einzige, die es schaffte, eine funktionierende Teststellung zur Verfügung zu stellen. Wie sich der Extreme-Switch im Test verhalten hat, steht weiter unten. Zunächst ist ein wenig Theorie erforderlich, um die Testszenarien besser zu erläutern.

Qualitätsprobleme im LAN

Lastspitzen sorgen in vielen auf Ethernet basierenden Unternehmensnetzen immer wieder dafür, dass es eng wird, und dann kommt es je nach Aus- und Überlastung zu teils erheblichen Datenverlusten. Aber auch andere übertragungstechnische Parameter wie Latency oder Jitter können – wenn sie gewisse Toleranzwerte überschreiten, weil beispielsweise das Netzwerk und seine Komponenten überfordert sind, zu Störungen einzelner Services oder auch der gesamten Kommunikation führen. Im Zeitalter konvergenter Netze, die Echtzeitapplikationen wie Voice- und Video-over-IP, aber auch Produktionssteuerungsdaten in das klassische Datenetz integrieren, führen Datenverluste und andere Pannen in der Praxis zu empfindlichen Kommunikationsstörungen und Produktionsausfällen. So haben in realen Projekten schon mangelhaft priorisierende Switches ganze Call-Center außer Betrieb gesetzt.

Bei klassischen Dateitransfers arbeitet das System mit möglichst großen Datenrahmen. Bei Echtzeit-Applikationen teilt sich das Feld. Video-Übertragungen nutzen ähnlich den Dateitransfers relativ große Datenrahmen. Voice-over-IP bewegt sich dagegen im Mittelfeld. Messungen mit Ethernet-LAN-Phones der ersten Generation in unseren Real-World Labs haben beispielsweise ergeben, dass diese Voice-over-IP-Lösung die Sprache mit konstant großen Rahmen von 534 Byte überträgt. Aktuelle Lösungen überlassen es dem IT-Verantwortlichen, selbst festzulegen, mit welchen Frame-Größen die Systeme arbeiten sollen. Dabei sollte er berücksichtigen, dass der Paketierungs-Delay mit kleiner werdenden Datenrahmen kleiner wird. Dagegen wächst der Overhead, der zu Lasten der Nutzdatenperformance geht, je kleiner die verwendeten Pakete sind. Generell kann man bei der IP-Sprachübertragung davon ausgehen, dass mittelgroße Frames verwendet werden. Und auch die meisten Web-Anwendungen nutzen mittelgroße Datenrahmen. Kurze Frames von 64 Byte sind dagegen beispielsweise bei den TCP-Bestätigungspaketen oder interaktiven Anwendungen wie Terminalansitzungen zu messen.

Für eine realitätsnahe und aussagefähige Auswertung der Messergebnisse ist es darüber hinaus entscheidend zu wissen, welche Framegrößen in welchen Verteilungen in realen Netzwerken vorkommen. Die Analyse der Verteilung der Framegrößen, die für das MCI-Backbone dokumentiert sind, sowie die Ergebnisse der Analyse typischer Business-DSL-Links haben ergeben, dass rund 50 Prozent aller Datenrahmen in realen Netzwerken 64 Byte groß sind. Die übrigen rund 50 Prozent der zu transportierenden Daten-

rahmen streuen über alle Rahmengrößen von 128 bis 1518 Byte.

Für die Übertragung von Real-Time-Applikationen ist zunächst das Datenverlustverhalten von entscheidender Bedeutung. Für Voice-over-IP gilt beispielsweise: Ab 5 Prozent Verlust ist je nach Codec mit deutlicher Verschlechterung der Übertragungsqualität zu rechnen, 10 Prozent führen zu einer massiven Beeinträchtigung, ab 20 Prozent Datenverlust ist beispielsweise die Telefonie definitiv nicht mehr möglich. So verringert sich der R-Wert für die Sprachqualität gemäß E-Modell nach ITU G.107 schon bei 10 Prozent Datenverlust um je nach Codec 25 bis weit über 40 Punkte, also Werte, die massive Probleme im Telefoniebereich sehr wahrscheinlich machen. Auf Grund ihrer Bedeutung für die Übertragungsqualität haben wir daher das Datenrahmenverlustverhalten als primäres K.o.-Kriterium für unseren Test definiert. Die Parameter Latency und Jitter sind dann für die genauere Diagnose und weitere Analyse im Einzelfall wichtig. Sind jedoch die Datenverlustraten von Hause aus schon zu hoch, können gute Werte für Latency und Jitter die Sprachqualität auch nicht mehr

Info

Real-Time-Switches in Network Computing

- ▶ Gigabit-Ethernet-Switches, Switch as Switch can, in NWC Special Infrastruktur 6/04, S.4 ff.
- ▶ Real-Time-Core- und -Edge-Switches, Klassengesellschaft im Netz, Teil 1 Core-Switches, in NWC 4/04, S.14 ff.,
- ▶ Real-Time-Core- und -Edge-Switches, Klassengesellschaft im Netz, Teil 2 Edge-Switches, in NWC 5/04, S.16 ff.,
- ▶ Klassengesellschaft im Netz – die Theorie, in NWC 4/04, S.48 ff.,
- ▶ Pilottest Class-of-Service-Queuing: Vorfahrt im Netz, in Special Konvergenz, S.II in NWC 20/03.

retten. Dafür, dass es zu solchen massiven Datenverlusten im Ethernet-LAN erst gar nicht kommt, sollen entsprechend gut funktionierende Priorisierungsmechanismen sorgen. Bei entsprechender Überlast im Netz sind Datenverluste ganz normal, jedoch sollen sie durch die Priorisierungsmechanismen in der Regel auf nicht echtzeitfähige Applikationen verlagert werden. Arbeitet diese Priorisierung nicht ausreichend, kommt es auch im Bereich der höher priorisierten Daten zu unerwünschten Verlusten.

Qualitätssicherungsmechanismen

Um solchen Problemen vorzubeugen, rüsten die Ethernet-Hersteller ihre Switches mit einer zusätzlichen Funktion aus, die es ermöglichen soll, bestimmten Applikationen die Vorfahrt im Netzwerk einzuräumen, wenn es einmal eng wird. Diese Priorisierungs-Mechanismen werden allgemein als Class-of-Service oder – verfälschend in Anlehnung an ATM – als Quality-of-Service

bezeichnet. Standardisiert ist eine achtstufige Priorisierung auf Layer-2 nach IEEE 802.1p/Q und auf Layer-3 nach RFC 1349/2474/2475. Die jeweils zugeordnete Priorisierung lesen die Systeme aus den Headern der Datenpakete aus.

Wie Switches die Datenpakete dann gemäß ihrer Priorität behandeln, hängt von den jeweils implementierten Queuing-Mechanismen ab. So besteht beispielsweise die Möglichkeit, bestimmten Daten absolute Vorfahrt einzuräumen oder auch für niedrigere Prioritäten Mindestdurchsatzraten zu garantieren. Eine gute Queuing- oder Scheduling-Strategie sollte folgende Voraussetzungen erfüllen:

- ▶ Sie muss die faire Verteilung der Bandbreite auf die verschiedenen Serviceklassen unterstützen. Dabei sollte auch die Bandbreite für besondere Dienste berücksichtigt werden, sodass es zu bestimmten Gewichtungen bei der Fairness kommen kann.
- ▶ Sie bietet Schutz zwischen den verschiedenen Serviceklassen am Ausgangsport, so dass eine Serviceklasse mit geringer Priorität nicht die anderen Serviceklassen anderer Queues beeinflussen kann.
- ▶ Wenn ein Dienst nicht die gesamte Bandbreite verwendet, die für ihn reserviert ist, dann sollte diese Überkapazität auch anderen Diensten zur Verfügung stehen, bis der eigentliche Dienst diese Kapazitäten wieder benötigt. Alternativ soll die Bandbreite für diesen Dienst absolut begrenzt werden.
- ▶ Ein schneller Algorithmus, der hardwaremäßig implementiert werden kann, muss für diese Strategie existieren. Nur dann kann diese Strategie auch auf Switches eingesetzt werden, die mit hoher Geschwindigkeit arbeiten. Algorithmen, die nur softwareseitig implementiert werden können, sind in der Regel ungeeignet, da man die Priorisierung bei hoher Last braucht, und gerade dann reicht die Performance der Softwarelösung in der Regel nicht aus.

Eine intelligente Ausnutzung der in einem konvergenten Netzwerk vorhandenen Bandbreiten und die Garantie für angemessene Service-Qualitäten der einzelnen Anwendungen setzen eine durchgängige Switching-Policy voraus, die nicht nur blind bestimmten Daten die Vorfahrt anderen gegenüber einräumt, sondern ein sinnvolles Bandbreitenmanagement für das gesamte Netzwerk realisiert. So ist es möglich, bei auftretenden Überlasten die Störungen im Betrieb gering und lokal begrenzt zu halten. Aus diesem Grund haben wir für unseren Vergleichstest Switches gefordert, die CoS-Queuing-Mechanismen sowie Bandbreitenmanagement unterstützen und damit erlauben, nicht nur bestimmten Applikationen Vorfahrt anderen gegenüber einzuräumen, sondern auch durch die Einräumung von Mindestbandbreiten die Service-Qualität niedriger Prioritäten und somit das Funktionieren der entsprechenden Applikationen zu garantieren und dafür zu sorgen, dass Dienste nicht mehr senden als vorgesehen.

Die Hersteller von Switches verwenden oft eigene Namen für die CoS-Queuing-Strategien oder ändern die eigentliche Strategie nach ihren Vorstellungen ab. Oft werden auch verschiedene

Strategien miteinander kombiniert, um die Ergebnisse zu verbessern. Die ursprünglichen Queuing-Strategien sind:

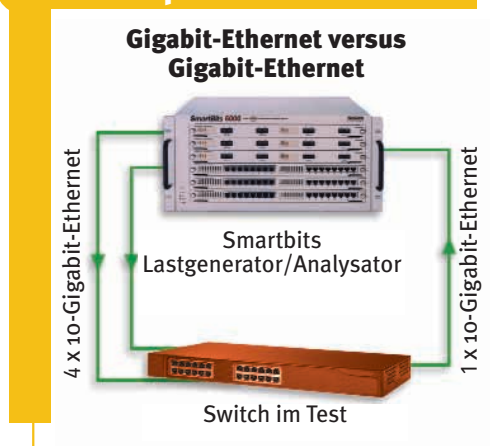
- ▶ First-In First-Out (FIFO),
- ▶ Strict- und Rate-Controlled-Priority-Queuing (PQ),
- ▶ Fair-Queuing (FQ),
- ▶ Weighted-Fair-Queuing (WFQ),
- ▶ Weighted-Round-Robin-Queuing (WRR), auch als Class-Based-Queuing (CBQ) bezeichnet und
- ▶ Deficit-Weighted-Round-Robin-Queuing (DWRR).

Wie diese Verfahren im Einzelnen arbeiten, haben wir in einem separaten Grundlagenartikel in Network Computing 4/04 ab Seite 48 dargestellt.

Die Testergebnisse

Messtechnisch sind die einzelnen CoS-Queuing-Verfahren zum Teil schlecht auseinander zu halten, da sie unter entsprechenden Lasten zu einem ähnlichen Verhalten der Systeme führen. Dieser Fakt ist aber auch nicht weiter problematisch, da für einen möglichst störungsfreien Netzwerkbetrieb das konkrete Switching-Verhalten der Systeme und nicht die dahinter stehenden Mechanismen und Theorien entscheidend sind. Konkret haben wir zwei Policies isoliert und messtech-

Test-Setup



nisch untersucht. Zunächst sollten die Switches eine Strict-Priority-Policy umsetzen. Hier kam es vor allem darauf an, dass die Daten der höchsten Priorität unter allen Umständen weitergeleitet werden sollten. Dieses Verhalten ist im Backbone-Bereich sinnvoll. Als zweites TestszENARIO haben wir dann mit Bandbreitenmanagement gearbeitet. Hier sollten die Systeme den Datenströmen aller vier Prioritäten Maximalbandbreiten garantieren, um einem Zusammenbruch der Anwendungen niedrigerer Prioritäten bei Überlastungen im Core-Bereich entgegen zu wirken. Deshalb wird dieses Verhalten im Edge-Bereich bevorzugt.

Aus den Ergebnissen von Performance-Messungen wie den von uns durchgeführten ist gut zu erkennen, ob, und wenn ja, in welchem Bereich, das jeweilige System Schwierigkeiten hat. Arbeitet der so belastete Switch korrekt, muss er in allen Fällen gemäß den »Class-of-Service-Regeln« die niedrig priorisierten Daten zugunsten

der höher priorisierten verwerfen. Ein Datenverlust in der höchsten Priorität dürfte theoretisch bei keinem unserer Strict-Priority-Tests theoretisch vorkommen, nur so würde der jeweilige Switch die fehlerfreie Übertragung der am höchsten priorisierten Echtzeitanwendung, beispielsweise einer Video-Konferenz, garantieren.

Für die Switches sind pro Zeiteinheit desto mehr Header-Informationen auszuwerten, je kleiner die einzelnen Datenrahmen sind. Ein Switch wird also zuerst Probleme mit 64-Byte-Datenströmen bekommen, wenn er bei der internen Verarbeitungsgeschwindigkeit an seine Grenzen stößt. Bei großen Datenrahmen können je nach Design dagegen schneller Probleme mit dem Speichermanagement beziehungsweise mit der Größe des überhaupt verfügbaren Pufferspeichers entstehen.

Strict-Priority-Switching

In unserem 10-Gigabit-Ethernet-Switch-Test haben wir ausschließlich die 10-Gigabit-Ethernet-Ports des Systems eingesetzt. Um die notwendigen Lasten zu erzeugen, haben wir mit unseren Smartbits auf vier 10-Gigabit-Ethernet-Eingangs-Ports gesendet und einen Gigabit-Ethernet-Ausgangs-Port adressiert. In diesem Test-Setup beträgt die maximale Überlast am Switch-Ausgang 400 Prozent.

Da wir den Switch systematisch überlastet haben, kam es bei einer maximalen Last von 100 Prozent auf den Eingangsport zu einer vierfachen Überlastung der Ausgangsport. Dadurch ist es natürlich normal, dass der Switch im Test nur wenige Frames übertragen konnte und somit viele Frames verloren hat. Anhand der Verteilung der einzelnen Prioritäten oder der einzelnen resultierenden Bandbreiten konnten wir dann erkennen, ob, und wenn ja, wo der jeweilige Testkandidat Probleme hatte.

In einer ersten Messreihe haben wir zunächst Strict-Priority-Switching gefordert und untersucht. In unseren Tests haben wir jeweils mit unseren Smartbits-Lastgeneratoren Datenströme auf die Eingangs-Ports gesendet und diese Datenströme an die Ausgangsport adressiert. Hierbei haben wir Datenströme in den vier Layer-2-Prioritäten – VLAN 7, 5, 3 und 1 nach IEEE 802.1p/Q – sowie in vier Layer-3-Prioritäten – Diffserv 6, 4, 2 und 0 – erzeugt. Die Eingangslast wird hierbei schrittweise erhöht, sodass die Last an den Eingangsport 25, 33,33, 50 und 100 Prozent betrug, was bei einer vierfachen Bündelung einer Last am Ausgangsport von 100, 133, 200 und 400 Prozent entspricht. Die Datenströme bestanden aus konstant großen Frames von jeweils 64, 128, 256, 512, 1024, 1280 und 1518 Byte. Als Burst-Size haben wir 1 Frame verwendet. Für die Ergebnisse haben wir die für CoS wichtigen Parameter Frame-Loss, Latency und Jitter ausgewertet. Im Mittelpunkt unserer Analysen steht dabei wegen seiner Bedeutung für die Übertragungsqualität das Datenverlustverhalten.

Verhält sich ein Switch anforderungsgerecht, dann verliert er bei 100 Prozent Last am Ausgangsport noch keine Daten. Bei 133 Prozent Last sollte er dann Totalverlust der niedrigsten Priorität erzeugen, die anderen Streams sollten

ohne Verluste ankommen. Bei 200 Prozent Last sollte der Switch dann die Daten der beiden niedrigen Prioritäten komplett verlieren und die beiden hohen Prioritäten ungehindert passieren lassen. Bei Vollast ist dann bei einer Last am Ausgangsport von 400 Prozent ein Totalverlust aller Prioritäten mit Ausnahme der höchsten erforderlich, damit die höchste Priorität noch verlustfrei verarbeitet werden kann. Die Daten der höchsten Priorität sollten in allen Fällen unbeschadet die Systeme passieren.

Sowohl bei den Strict-Priority-Messungen mit Layer-2- wie mit Layer-3-Priorisierung arbeitete der Blackdiamond-10808 nahezu mustergültig. Abweichungen vom Sollwert konnten wir dem großen Extreme-Switch – wenn überhaupt – dann erst in der zweiten Stelle hinter dem Komma nachweisen. Da, wo es gemäß Policy gewünscht war, lieferte der Blackdiamond Wire-speed – und zwar bei allen Frame-Größen von 64 bis 1518 Byte.

Switching mit Bandbreitenmanagement

Auch in unseren Tests mit Bandbreitenmanagement haben wir jeweils mit unseren Smartbits-Lastgeneratoren Datenströme auf die Eingangsport gesendet und diese Datenströme an die Ausgangsport adressiert. Hierbei haben wir Datenströme in den vier Layer-2-Prioritäten – VLAN 7, 5, 3 und 1 nach IEEE 802.1p/Q – erzeugt. Die Eingangslast wird hierbei schrittweise erhöht, sodass die Last an den Eingangsport 25, 33,33, 50 und 100 Prozent betrug, was bei einer vierfachen Überlast einer Last am Ausgangsport von 100, 133, 200 und 400 Prozent entspricht.

Die Datenströme bestanden aus konstant großen Frames von jeweils 64, 128, 256, 512, 1024, 1280 und 1518 Byte. Die Burst-Size betrug hierbei 1 Frame. Als Maximalbandbreiten haben wir für die höchste Priorität VLAN 7 10 Prozent, für VLAN 5 20 Prozent, für VLAN 3 30 Prozent und für VLAN 1 40 Prozent gefordert. Für die Ergebnisse haben wir dann die für CoS wichtigen Parameter Frame-Loss, Latency und Jitter ausgewertet. Im Mittelpunkt auch dieser Analysen steht wegen seiner Bedeutung für die Übertragungsqualität das Datenverlustverhalten.

Bei der Priorisierung mit Bandbreitenmanagement sind zwei grundsätzlich unterschiedliche Verhaltensweisen der Switches zu unterscheiden: Rate-Limited-Switching und Weighted-Switching. Beim Rate-Limited-Switching garantiert der Switch den entsprechend konfigurierbaren Bandbreitenanteilen der einzelnen Prioritäten nicht nur eine Mindestbandbreite, er »deckelt« quasi auch die Durchsätze, indem er übrige Bandbreiten, die ein Dienst, dem sie zur Verfügung stehen, derzeit nicht benötigt, auch nicht für andere Dienste verfügbar macht. Eine solche Funktion ist insbesondere für den Edge-Bereich je nach Policy unverzichtbar, weil es so möglich ist, der Entstehung von Überlasten bereits in der Netzwerkperipherie vorzubeugen. Switches im Core-Bereich sollten dagegen die Mindestbandbreiten für die ihnen zugeordneten Dienste reservieren. Wenn diese Dienste die ihnen zustehenden Bandbreiten aber nicht be-

nötigen, dann sollten andere Dienste freie Bandbreiten über die ihnen selbst zustehenden hinaus ruhig nutzen können. Ansonsten wird die insgesamt im Core-Bereich zur Verfügung stehende Bandbreite unnötig verringert. Ein solches Verhalten kann aber auch im Rahmen der Policy erwünscht sein. Um den verschiedenen Mechanismen gerecht zu werden, haben wir hier ausschließlich das Verhalten der Systeme bei Volllast gewertet, da dann unabhängig vom verwandten Mechanismus die gleichen Maximalbandbreiten für die jeweilige Priorität eingehalten werden müssten.

Um das gewünschte Verhalten des Blackdiamond-10808 zu erhalten, mussten die Extreme-Ingenieure aufwändige Konfigurationsarbeiten vornehmen, die nur mit Hilfe nicht dokumentierter, erweiterter Kommandos möglich waren. So präpariert arbeitete der Netzwerkdiamant recht ordentlich.

Dabei kam er am besten mit großen Datenrahmen zurecht, so pendelten die Abweichungen vom Sollwert bei Volllast mit 1518-Byte-Paketen um die 0,2 Prozent. Mit abnehmender Frame-Größe arbeitete der Extreme-Switch dann tendenziell unpräziser, so verlor er beispielsweise in der zweitniedrigsten Priorität bei der Messung mit 64-Byte-Paketen 3,4 Prozent Daten zuviel. In

Multicast-Betrieb

Hat es sich bei den obigen Messungen generell um Punkt-zu-Punkt-Verbindungen gehandelt, bei denen der Switch Datenströme zwischen den verschiedenen Ports vermitteln muss, so sendet der Switch im Multicast-Betrieb den Datenstrom, der an einem Port ankommt, an eine festgelegte Zahl von Ports weiter. Im Fall unseres Testaufbaus bedeutete das, dass der Smartbits-Lastgenerator auf einen Port des zu testenden Switches mit Volllast gesendet hat. Der Switch im Test musste dann den Datenstrom vervielfältigen und an 18 Ausgangsports senden, die wiederum mit dem Smartbits-Analysator verbunden waren. Den 20. Gigabit-Ethernet-Port haben wir zur Kontrollmessung ebenfalls mit dem Smartbits-System verbunden, um zu sicherzustellen, dass das System nicht in den Broadcast-Betrieb wechselt und die Datenströme einfach an alle Ports sendet. Diese Messung haben wir dann wieder nacheinander mit 64, 128, 256, 512, 1024, 1280 und 1518 Byte großen Frames durchgeführt. Für den Multicast-Betrieb haben wir das auf IP basierende Internet-Group-Management-Protocol, kurz IGMP, genutzt, um im Trainingsverkehr den Switch lernen zu lassen, an welche Ports er den Multicast-Traffic senden soll. Stößt der Switch im Test in diesem Szenario an seine Grenzen, dann

wir bei den Tests jeweils nur zwei der vier Ports benutzt. Die Aussagen dieses Tests beziehen sich also nur auf Szenarien, in denen lediglich die Hälfte der Ports pro Blade aktiv verwendet werden. Die anderen Ports kann man dann als Hot-Standby-Redundanz-Ports einsetzen.

Recovery-Time

In einem zweiten Testaufbau haben wir dann eine Reihe weiterer Features des Blackdiamond-10808 untersucht und überprüft. Hierzu haben wir zunächst ein Testnetzwerk aufgebaut, das aus zwei baugleichen Blackdiamond-Systemen bestand, die wir via 10-Gigabit-Ethernet-Trunk-Ports miteinander verbunden haben. Dann haben wir beide Switches an einen Lastgenerator/Analysator vom Typ »Ixia 1600T« angeschlossen und bidirektionale Ethernet-Datenströme mit einer konstanten Datenrate und einer Frame-Größe von 64 Byte erzeugt. Die jeweilige Network-Recovery-Time konnten wir dann basierend auf den gemessenen Datenverlusten ermitteln. Hierzu haben wir den Primary-Link unterbrochen, indem wir den entsprechenden Stecker aus dem Port herausgezogen haben. Im Fall eines solchen Link-Fehlers schaltet das System mit Hilfe des Layer-2-Protection-Protokolls EAPS auf einen Secondary-Link um. Der Mittelwert aus fünf Messungen für die Recovery-Time betrug 40 ms. Im Bereich der klassischen TK-Anlagen ist eine Umschaltzeit von unter 50 ms ein guter Wert, hier kann der Blackdiamond-10808 recht gut mithalten.

Dann haben wir in einem identischen Testaufbau einen Fehler des Management-and-Switch-Fabric-Modules, kurz MSM, simuliert, indem wir das aktive Management-Board einfach herausgezogen und dann wieder die Recovery-Time ermittelt haben. Als Mittelwert von fünf Messungen konnten wir 270 ms feststellen. Gemessen am Stand der LAN-Technik ist auch diese Umschaltzeit state-of-the-art. Allerdings sind hier die LAN-Komponenten vom Standard im Telefoniebereich noch ein Stück entfernt. Klassische TK-Anlagen liegen auch in der Umschaltzeit zwischen Modulen bei unter 50 ms.

ACL-Performance

Als Nächstes haben wir die Performance der 10-GBit/s-Access-Control-Lists, kurz ACL, untersucht. Der »4GNSS«-ASIC des Blackdiamond-10808 soll den Datenverkehr des Core-Switches mitlesen und filtern und so das erforderliche Traffic-Monitoring durchführen und für die notwendige Sicherheit sorgen. Damit besitzt der Switch Paketfiltereigenschaften, eine Teilfunktionalität einer Firewall. Hierzu wertet das System Informationen auf Layer-2 bis -4 aus. Dabei können IT-Administratoren auch eigene Regeln erstellen. Insgesamt unterstützt der Blackdiamond-10808 maximal 64 000 ACL-Regeln und -Counter.

Zur Messung der ACL-Performance haben wir wieder zwei baugleiche Blackdiamond-Systeme mit 10-Gigabit-Ethernet-Links verbunden und in einer Zangenmessung mit dem Ixia-Chassis verbunden. Mit dem Ixia-Lastgenerator haben wir dann bidirektionalen Datenverkehr erzeugt, der fünf Verkehrsklassen angehörte. Beide Black-

Info

Messergebnisse mit Bandbreitenmanagement Layer-2

| | Ausgang | Prio7 | Prio5 | Prio3 | Prio1 | Summe | Durchschnittliche Abweichung (absolut) |
|-----------|---------|-------|-------|-------|-------|--------|--|
| Sollwert | 400% | 90,00 | 80,00 | 70,00 | 60,00 | | |
| 64 Byte | 400% | 89,50 | 79,02 | 73,40 | 58,05 | 299,97 | |
| | | 0,50 | 0,98 | -3,40 | 1,95 | | |
| | | 0,50 | 0,98 | 3,40 | 1,95 | | 1,71 |
| 128 Byte | 400% | 89,70 | 79,41 | 72,00 | 58,85 | 299,95 | |
| | | 0,30 | 0,59 | -2,00 | 1,15 | | |
| | | 0,30 | 0,59 | 2,00 | 1,15 | | 1,01 |
| 256 Byte | 400% | 89,82 | 79,66 | 71,07 | 59,37 | 299,91 | |
| | | 0,18 | 0,34 | -1,07 | 0,63 | | |
| | | 0,18 | 0,34 | 1,07 | 0,63 | | 0,56 |
| 512 Byte | 400% | 89,86 | 79,78 | 70,54 | 59,65 | 299,83 | |
| | | 0,14 | 0,22 | -0,54 | 0,35 | | |
| | | 0,14 | 0,22 | 0,54 | 0,35 | | 0,31 |
| 1024 Byte | 400% | 89,85 | 79,81 | 70,22 | 59,78 | 299,66 | |
| | | 0,15 | 0,19 | -0,22 | 0,22 | | |
| | | 0,15 | 0,19 | 0,22 | 0,22 | | 0,19 |
| 1280 Byte | 400% | 89,84 | 79,80 | 70,14 | 59,81 | 299,58 | |
| | | 0,16 | 0,20 | -0,14 | 0,19 | | |
| | | 0,16 | 0,20 | 0,14 | 0,19 | | 0,17 |
| 1518 Byte | 400% | 89,81 | 79,76 | 70,04 | 59,77 | 299,37 | |
| | | 0,19 | 0,24 | -0,04 | 0,23 | | |
| | | 0,19 | 0,24 | 0,04 | 0,23 | | 0,18 |

der höchsten Priorität betrug die Differenz dagegen auch hier lediglich 0,5 Prozent. Insgesamt konnte das Priorisierungsverhalten mit Bandbreitenmanagement auf Layer-2 durchaus überzeugen.

Überhaupt nicht überzeugend endete dagegen der gleiche Test mit Layer-3-Priorisierung. Die geplanten Messungen mit Layer-3-Priorisierung waren schlicht nicht möglich, da die Ingenieure von Extreme auch mit Hilfe erweiterter Kommandos nicht in der Lage waren, das System zum gewünschten Verhalten zu bringen.

kommt es zu Datenverlusten, und die maximale Durchsatzrate sinkt.

Im Multicast-Szenario verhielt sich der Blackdiamond mustergültig und patzte bei keiner Frame-Größe. Zu messen waren in allen Fällen konstant 100 Prozent Leistung an allen Ausgangsports. Und auch bei der Messung mit 9000 Bytes großen »Jumbo-Frames« konnten wir dem schwarzen Diamanten keine Schwächen nachweisen. Zu beachten ist jedoch, dass die Bandbreite von den 4-Port-10-Gigabit-Blades zu der Backplane nur 20 GBit/s beträgt. Deshalb haben

diamonds mussten dann die Datenströme filtern. Wir haben eine Mischung aus erlaubtem und verbotenen Datenverkehr erzeugt, um normale wie potentiell gefährliche Daten zu simulieren. Die Blackdiamonds zeigten hier eine souveräne Leistung, sie leiteten den erwünschten Traffic zu 100 Prozent weiter, filterten den unerwünschten zu 100 Prozent aus und arbeiteten tatsächlich mit 10-GBit/s Datendurchsatz brutto, boten also Wirespeed.

Dynamische Software-Updates

Das modulare Betriebssystem auf Linux-Basis der Blackdiamond-Switches »ExtremeWare XOS« erlaubt das Laden neuer Software-Module ohne dass der Switch außer Betrieb genommen werden muss. So können IT-Administratoren jederzeit neue Funktionalität in ihr Netzwerk bringen, ohne Ausfallzeiten zu riskieren. Wir erproben dieses Feature, indem wir einen Blackdiamond-Switch ohne SSH2-Modul in Betrieb nahmen und dann im laufenden Betrieb das SSH2-Modul luden und starteten. Der Blackdiamond arbeitete währenddessen störungsfrei weiter, ein Neustart oder Ausfallzeiten war nicht erforderlich. Der Aufbau einer Verbindung via SSH2 war problemlos möglich, und auch das Deinstallieren des SSH2-Moduls im laufenden Betrieb verlief völlig unauffällig und störte den Netzwerkbetrieb nicht.

Selbstheilungs-Software

Die Blackdiamond-Switches sind mit einer »selbstheilenden« Software ausgestattet, die auf Prozess-Level arbeitet. Unter der Selbstheilungsfunktion versteht Extreme Networks die Fähigkeit der Extremeware-XOS-Software, im Fall von Prozess-Abstürzen die betroffenen Prozesse automatisch neu zu starten, ohne andere Prozesse zu beeinträchtigen oder den Switch neu zu booten. Diese Funktionalität erhöht die Verfügbarkeit der Systeme deutlich und schützt kritische Applikationen. Außerdem ist der Switch hierdurch gegenüber Denial-of-Service-Angriffe weniger empfindlich.

Wir begannen den Test dieser Selbstheilungsfunktion mit einem Blackdiamond-Switch, dessen Telnet-Funktion aktiviert war. Mit Userzugriffen via Telnet stellten wir sicher, dass Telnet aktiv war und fehlerfrei arbeitete. Dann haben wir den Telnet-Service abgeschossen und so einen Prozess-Crash simuliert. Die Software Extremeware-XOS erkannte das Problem und startete den Telnet-Service neu, ohne dass ein manueller Eingriff erforderlich gewesen wäre.

Virtuelles Routing

Die Blackdiamond-Switches ermöglichen die Einrichtung bis zu acht virtueller Router auf einem Switch. Hierzu pflegt jeder virtuelle Router seine eigenen Routing-Tabellen. Das Aufteilen der Routing-Informationen hat Vorteile in Bezug auf Netzwerkstabilität, Skalierbarkeit und Sicherheit. Wir begannen den Test mit einem Blackdiamond-Switch in Default-Konfiguration und erzeugten dann zwei virtuelle Router. Der erste virtuelle Router VR1 sollte OSPF als dynamisches Routing-Protokoll nutzen, der zweite Router

VR2 sollte mit RIPv2 arbeiten. Dann haben wir VLANs auf den virtuellen Routern eingerichtet und die virtuellen Router mit anderen OSPF- und RIP-Routern verbunden. Die beiden virtuellen Router haben dann ihre jeweiligen Routing-Tabellen aktualisiert. Dann haben wir die Routing-Tabellen der beiden virtuellen Router miteinander verglichen. Beide virtuellen Router hatten ihre eigene Routing-Tabelle korrekt gepflegt und sich keine Routing-Information geteilt. Damit hat man auf Ebene 3 ein dem VLAN der

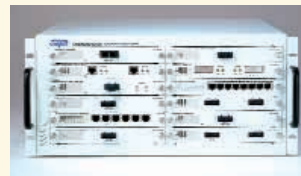
trieb mit Bandbreitenmanagement auf Layer-3 war nicht möglich, hier nutzen auch die erweiterten Kommandos nicht viel.

Im Multicast-Betrieb – beispielsweise relevant für die Distribution von Video- oder Voice-Streams – vermochte der Extreme-Switch zu überzeugen. Generell ist aber bei allen 10-Gigabit-Ethernet-Performance-Tests zu beachten, dass die 4-Port-10-Gigabit-Ethernet-Module jeweils nur mit 20 GBit/s an die Backplane angebunden sind. Folglich können 50 Prozent der

Info

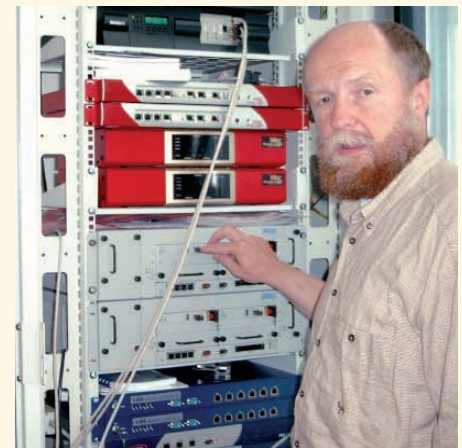
So testete Network Computing

Als Lastgenerator und Analytiker haben wir in unseren Real-World Labs einen »Smartbits 6000B Traffic Generator/Analysator« von Spirent Communications eingesetzt. Das in dieser Konfiguration rund 800 000 Euro teure System ist mit der Software »SmartFlow« ausgestattet und mit acht 10-Gigabit-Ethernet-Fiber-Ports bestückt. Alle Ports können softwareseitig als Lastgeneratorausgang und/oder als AnalySATOREINGANG eingesetzt werden. Die Class-of-Service-Eigenschaften der Switches im Testfeld haben wir in verschiedenen Testreihen gemäß RFC 2544



(vgl.: www.ietf.org/rfc/rfc2544.txt) gemessen. In diesen Tests haben wir die Priorisierung auf Layer-2 nach IEEE

802.1p/Q sowie auf Layer-3 nach RFC 2474 (Diffserv) untersucht. In unseren Test-szenarien »10-Gigabit-Ethernet-Switches« haben wir verschieden priorisierte Datenströme von mehreren Eingangsport auf einen Ausgangsport geschickt und das System so gezielt überlastet. Die die Priorisierung festlegenden Bits haben wir im Header der Datenrahmen mit drei Bits nach IEEE 802.1p und auf Layer-2 und nach Diffserv auf Layer-3 festgelegt. Durch eine gezielte Überlastung der Switches in diesen Tests ist es möglich, das genaue Datenverlustverhalten sowie weitere Testparameter wie Latency oder Jitter zu ermitteln, das Leistungspotential der unter-



suchten Switches zu analysieren und deren Eignung für bestimmte Einsatzszenarien zu prüfen. Ein Durchsatztest im Multicast-Modus unter Einsatz des Internet-Group-Management-Protocol (IGMP) in der Trainingsphase hat unseren Test ergänzt.

Abschließend haben wir einige neuere herstellereigene Features des Systems verifiziert, die insbesondere die Verfügbarkeit des Systems erhöhen sollen.

Ebene 2 vergleichbares Verhalten und erhält eine deutlich höhere Flexibilität im Netzwerkdesign.

Fazit

Sieger im aktuellen 10-Gigabit-Ethernet-Test ist Extreme Networks, der Hersteller punktete schon allein durch die Tatsache, dass er letztendlich der einzige war, der ein funktionierendes 10-Gigabit-Ethernet-System in unserer Testumgebung präsentieren konnte. Und im Class-of-Service-Performance-Teil unseres Tests vermochte sich das System auch recht gut zu schlagen. Bei unseren Strict-Priority-Messungen arbeitete der Blackdiamond-10808 ohne Fehl und Tadel. Auch das Bandbreitenmanagement beherrschte der »schwarze Diamant« gut, ganz »lupenrein« ist er allerdings noch nicht. So konnten die Extreme-Ingenieure das System nur auf Ebene 2 und unter Zuhilfenahme erweiterter Kommandos und aufwändiger Konfigurationsarbeiten zum Arbeiten bringen. Der Be-

ports bei entsprechenden Lasten nur als aktive Ports genutzt werden. Die anderen Ports kann man dann als Hot-Standby-Redundanz-Ports einsetzen.

Die Prüfung ausgewählter Funktionen, die insbesondere die Verfügbarkeit des LAN-Switches untersucht hat, hat ergeben, dass der Blackdiamond-10808 in dieser Disziplin durchaus mit aktuellen klassischen TK-Anlagen konkurrieren kann, auch wenn er noch nicht in allen Disziplinen die dort üblichen Standards erreicht hat.

Insgesamt hat der vorliegende Test gezeigt, dass 10-Gigabit-Durchsätze inzwischen machbar und LAN-Switches für den Einsatz in leistungsfähigen konvergenten Netzen geeignet sind. Allerdings ist den IT-Verantwortlichen nach wie vor anzuraten, auf unabhängige Tests zu setzen, denn es ist immer noch nicht überall Gigabit und Quality-of-Service drin, wo es drauf steht.

Prof. Dr. Bernhard G. Stütz, [dg]