ComConsult Research

Der Netzwerk Insider

Juli 2010

Systematische Weiterbildung für Netzwerk- und IT-Professionals

Schwerpunktthema

Evolution der RZ-Netze: Konvergenz und Konzentration

Teil 2: Analyse von 10 GbE mit und ohne Priorisierung

von Dr. Franz-Joachim Kauffels

Im Rahmen des Netzwerk-Redesign Forums 2010 wurden von Dr. Suppan einige Thesen aufgestellt, die auf den ersten Blick teilweise nicht unstrittig sind und erhebliche Bedeutung für die Auswahlkriterien zukünftiger Core-Switches haben. Daher bedürfen sie einer weiteren Analyse. Eine Diskussion dieser Thesen nur auf der Grundlage mündlicher Aussagen der Beteiligten, die auf dem jeweiligen individuell geprägten Verständnis der zur Diskussion stehenden Verfahren beruht, führt nicht zu einem eindeutigen Ergebnis. Also muss teilweise auf eine mathematische Analyse zurückgegriffen werden. Es zeigt sich, dass die Thesen nicht nur verifiziert werden können, sondern eine wesentlich weiter reichende Bedeutung haben, als man zunächst annehmen könnte.

Die zur Diskussion stehenden Thesen sind:

„Die Implementierung von DCB-Funktionen ist ein wesentliches Kriterium bei der Auswahl zukünftiger Core Switches“

„Die Implementierung von DCB-Funktionen ist ein wesentliches Kriterium bei der Auswahl zukünftiger Campus-Switches“ ...

weiter auf Seite 18

Zweitthema

IP Version 6 - das Internet der nächsten Generation (Teil 3)

von Dipl.-Inform. Petra Borowka-Gatzweiler

In Teil 3 dieses Artikels lesen Sie:

- **was die Migration für Server, Endgeräte und Router bedeutet**
- **welche Alternativen für Standort-Designs möglich sind**
- **welche Kernfunktionen und Tools unterstützt werden sollten**
- **wo es mögliche Probleme mit IPv6 gibt**
- **wie sich der IPv6 Markt darstellt**
- **welches abschließende Fazit zu ziehen ist**

Adressplan

Vor Beginn der IPv6 Migration sollte ein IPv6 Adressplan erarbeitet werden, der dem Corporate WAN, den einzelnen Standorten und innerhalb der Standorte den Netzwerk-Komponenten, Servern und Hosts entsprechend abgestimmte Adressbereiche zuweist. Empfehlenswert ist der Einsatz global routbarer Adressen, die sich bei den entsprechenden Network Information Centers (RIPE für Europa) beantragen lassen. Nutzt ein Unternehmen bereits offizielle IPv4 Adressen und sind diese flächendeckend ausreichend vorhanden, so können die IPv6 Adressen mittels IPv4 Mapping erzeugt werden. Da diese Voraussetzung in den meisten Fällen nicht gegeben ist, ist eher von neu beantragten IPv6 Adressbereichen auszugehen.

weiter auf Seite 11

Sonder-Kongress

ComConsult Wireless-Forum 2010

ab Seite 4

Geleit

Wireless Netzwerke: eine echte Herausforderung kommt auf uns zu!

ab Seite 2

Neues Seminar

IPv6: Planung, Migration und Betrieb

auf Seite 8

Zum Geleit

Wireless Netzwerke: eine echte Herausforderung kommt auf uns zu!

Als wir mit der Planung von Wireless-Netzwerken begonnen hatten, war die Dichte von Teilnehmern in den einzelnen Funkzellen sehr gering. Entsprechend niedrig war die Auslastung der Wireless-Infrastruktur und damit direkt verbunden auch die Zahl der Probleme. Zwar gab es immer wieder Diskussionen über die Ablösung von Kabeln durch Wireless-Netzwerke und somit den Übergang zu einer flächendeckenden Infrastruktur mit hoher Sichte, aber dazu ist es bisher nicht gekommen. Trotzdem ist der Markt in Bewegung, die Zeit der einfachen und wenig ausgelasteten Zellen ist mehr und mehr vorbei. Wir brauchen für sichere Investitionen eine klare Prognose für die zukünftige Entwicklung.

Die zentrale Frage, die hinter dieser Betrachtung steht, ist:

Wer oder was wird die zukünftige Wireless-Infrastruktur prägen? Werden es Endgeräte sein, die vom bisherigen Kabel zum Funk konvertiert werden oder werden es neue Formen von Endgeräten und Anwendungen sein?

Die bisherige Entwicklung der Wireless-Technologie lässt sich wie folgt beschreiben:

- Erste Generation: vereinzelte Funkzellen, nicht überlappend, geringe Teilnehmer-Dichte
- Zweite Generation: höhere Teilnehmerdichten und Leistungen pro Zelle
- Dritte Generation: überlappende Zellen, mehr Teilnehmer, Handover zwischen Zellen und Mobility-Anforderungen durch Anwendungen wie Voice

Die Frage ist also, wie nun die vierte Generation aussehen wird. Kernfragen für das Design sind dabei immer:

- Sprechen wir über mobile oder stationäre Teilnehmer?
- Wie viele Teilnehmer pro Zelle?
- Wie viel Bandbreite pro Gerät oder Applikation?
- Der notwendige Grad an Redundanz

Daraus lassen sich sowohl die Zellgröße, die Zahl der notwendigen Zellen als auch die notwendige Leistung pro Zelle ableiten.

Die mit Abstand wichtigste Entwicklung findet im Bereich der mobilen Endgeräte statt. Sie werden gemäß ihrer Natur durch Funk- oder GSM-Netze versorgt bzw. in Client-Server-Architekturen eingebunden. Bisher war diese Geräteklasse auf Laptops und Funktelefone beschränkt. Doch schon seit Jahren ist dieser Markt im Umbruch. Den Anstoß hat Apple mit dem iPhone gegeben. Hier wurde ein neuer Gerätetyp geschaffen, der zum ersten Mal für mobile Applikationen über das simple telefonieren hinaus wirklich geeignet war. Apple hat den Nachweis erbracht, dass mobile Endgeräte sinnvoll in Unternehmens-Applikationen eingebunden werden können. Dabei geht es nicht nur um Anwendungen wie Email, sondern auch um den mobilen Zugriff auf Dokumente (Box.net, Dropbox u.a. in Kombination mit Anwendungen wie Documents to Go) und um Datenbank-Frontends.

Es ist mehr als wahrscheinlich, dass dieser Bereich der mobilen Endgeräte in den nächsten Jahren explodieren wird. Dafür gibt es eine Reihe von Argumenten:

- Immer mehr Applikationen kommen auf den Markt, die auch und speziell für Unternehmen nutzbar sind
- Neben Apple schafft Google mit Android eine zweite sehr tragfähige Plattform für mobile Endgeräte. Auch wenn Google hier in der Bedienung noch nachlegen muss, kann man davon ausgehen, dass bereits ab Version 2.3 oder spätestens mit Version 3 eine sehr wettbewerbsfähige Version entsteht
- Das iPad definiert nach dem iPhone die nächste Geräteklasse. Wir haben es getestet und sind davon überzeugt, dass dies eine solide Basis für eine große Menge von Unternehmensanwendungen sein wird. Zugriff und Änderungen in Dokumenten, spezielle Applikationen für den Gesundheitsbereich, für den PKW-Vertrieb, für Versicherungs-Agenturen, für den Ausbildungsbereich, ... die Liste ist fast beliebig lang. Die fehlende Offenheit ist ein Manko und der fehlende lokale Dokumentenspeicher, der von allen Applikationen genutzt werden kann. Aber das wird diesen Gerätetyp nicht davon abhalten, mindestens genauso erfolgreich zu sein wie das iPhone. Aus Unternehmenssicht wird der Erfolg noch größer sein, da die Bildschirmgröße und die Batteriekapazität eine bessere Basis für Unternehmens-Anwendungen gibt
- HTML5 wird den gesamten Markt für Webapplikationen beschleunigen. Auf Dauer wird ein erheblicher Teil der Applikationen, die heute im App-Store angeboten werden, auch direkt als Webapplikation realisiert werden können. Damit entsteht eine weitere ideale Basis zur Versorgung mobiler Endgeräte. Die technische Basis der Endgeräte wird einfacher und parallel nimmt die Nutzbarkeit von Applikationen zu.

Basierend auf dieser technischen Basis liegt die Prognose nahe, dass sich sowohl die Zahl als auch der Nutzungsgrad mobiler Endgeräte in den Unternehmen in den nächsten Jahren vervielfachen wird.

Welche Anforderungen sind zu erfüllen, um diese Geräte zu integrieren?

- Bandbreiten im Bereich von 1 bis 2 Mbit/s pro Gerät. Man kann davon ausgehen, dass alle wesentlichen Applikationen Web-Applikationen sein werden
- Die Gerätezahl pro Zelle wird sich x% der Mitarbeiterzahl in dieser Zelle annähern. X wird für jedes Unternehmen verschieden sein, aber wir reden von deutlich mehr Teilnehmern pro Zelle als bisher
- Der Grad an Flächendeckung muss höher sein als bisher
- Handover zwischen den Zellen ist wichtig, da für diese Geräteklasse sowohl Voice als auch Video in Zukunft wichtig sein werden

Zweitthema

IP Version 6 - das Internet der nächsten Generation

(Teil 3)

Fortsetzung von Seite 1

Dipl.-Inform. Petra Borowka-Gatzweiler leitet das Planungsbüro UBN und gehört zu den führenden deutschen Beratern für Kommunikationstechnik. Sie verfügt über langjährige erfolgreiche Praxiserfahrung bei der Planung und Realisierung von Netzwerk-Lösungen und ist seit vielen Jahren Referentin der ComConsult Akademie. Ihre Kenntnisse, internationale Veröffentlichungen, Arbeiten und Praxisorientierung sowie herstellerunabhängige Position sind international anerkannt.

Unternehmen, die IPv4 Multicast-Anwendungen einsetzen, müssen die Migration auf IPv6 Multicast einplanen.

Router Migration
Die Migration der Router nach IPv6 sollte in einem Unternehmen als einer der ersten Schritte vollzogen werden, da die Router im Regelfall die Schnittstelle zum Provider und damit zur öffentlichen Welt realisieren. Da die Provider entsprechend der EU-Empfehlungen bis 2011 gehalten sind, ihre Netze auf IPv6 umzustellen, ist es für Unternehmen sehr empfehlenswert, die Edge Router, die den Übergang ins Providernetz bilden, bis zu diesem Zeitpunkt ebenfalls IPv6-fähig zu machen. Router werden insbesondere mit Dual Stack und darüber hinaus als IPv4 ←→ IPv6 Gateways implementiert werden, da sie genau an der Grenze zwischen internem IPv4 Netz und externem IPv6 Netz positioniert sind. Den Routern folgen die Layer-3 Switches mit IPv6-Fähigkeit, da sie die Routing Infrastruktur des Standort-Netzes bilden. Auch für sie bietet sich der Dual Stack an, da ja die Endgeräte und Server zu diesem Zeitpunkt noch über IPv4 kommunizieren. Können nicht alle Router / Layer-3 Switches eines Standort-Netzwerks in einem Schritt IPv6-fähig gemacht werden, so dass IPv4 „Inseln“ verbleiben, so muss der Betreiber entsprechende Tunneling Verfahren über die IPv4-Inseln hinweg konfigurieren. Während der Migrationsphase fahren die Router dann sowohl IPv4 Routing als auch IPv6 Routing (z.B. OSPFv3). Die Migration der Router auf Dual Stack kann schrittweise erfolgen, so lange noch keine IPv6 Server und Endgeräte im Netz sind. Der IPv6 Stack ist dann zwar aktiviert, wird aber noch nicht genutzt.

Server Migration
Die Migration der Server wird im zweiten Schritt erfolgen, da hierdurch sowohl IPv4 als auch IPv6 Clients auf die migrierten Server zugreifen können. Insbesondere für Server ist der Parallelbetrieb beider Protokolle als Dual Stack dringend erforderlich, um auch nach der Servermigration noch alle Clients zu unterstützen, die selbst nicht IPv6-fähig sind. Zur Servermigration gehört insbesondere auch die Einrichtung eines IPv6-fähigen DNS-Dienstes, der die Servernamen zusätzlich zu IPv4 auf IPv6 Adressen mappt.

Endgeräte Migration
Die Umstellung der Clients erfolgt im Idealfall als dritter Schritt, wenn die Router- und Server-Infrastruktur schon umgestellt und getestet wurde. Kann der Netzbetreiber sicherstellen, dass die Clients nur Anwendungen/Dienste nutzen, die IPv6 unterstützen, so kann er client-seitig eine IPv6-only Installation fahren. Ist diese Voraussetzung nicht gegeben, empfehlen wir dem Betreiber, auch die Clients für eine Übergangszeit sowohl mit IPv4 als auch mit IPv6, also mit einem Dual Stack auszustatten. So wird sichergestellt, dass auch „alte“ Server und Applikationen noch nutzbar sind, die sich nicht auf IPv6 umstellen lassen. Erfolgt die Umstellung von Endgeräten für ein einzelnes IP Subnetz jeweils komplett und sind alle Server IPv6 fähig, so könnten die Router als IPv6 Router und die Endgeräte ohne Dual Stack arbeiten.

Firewall Migration
Für die Migration nach IPv6 müssen alle Filterregeln in Firewalls und Paketfilter-Komponenten (z.B. auch ACLs in Routern) neu erstellt werden, die in irgendeiner Form die IP-Adresse oder das Paketformat referenzieren - und das sind praktisch alle Filterregeln. Insbesondere gilt dies auch für ICMP bzw. ICMPv6 Filter, da sich bei ICMPv6 die Protokollnummer, Typ- und Code-Zuordnungen und auch die Funktionalität verändert haben.

NAT sehen die Standardisierer in IPv6 nicht mehr wirklich als Sicherheitsmechanismus - die eigentliche Entwicklung von NAT war ja auch schon bei IPv4 eher dem Adressraum-Problem als der Sicherheit geschuldet; hier wurde lediglich aus der Not eine Tugend gemacht. RFC 4864 beschreibt das Vorgehen, wie sich NAT auf IPv6 abbilden lässt. Zum Beispiel kann die NAT-Funktion, neue eingehende Verbindungen nicht an SOHO-Teilnetze weiterzuleiten, auf einen zustandsbasierten Paketfilter migriert werden. Ansonsten wird IPv6 als „immanent sicher“ eingeordnet, da die Nutzung von IPsec unterstellt wird.

Standort-Designs
Sowohl für Router als auch Server als auch Clientsysteme gilt natürlich, dass die parallele Aktivierung beider Stacks (IPv4 und IPv6) mehr Rechenleistung und Speicherplatz erfordert als die Aktivierung nur eines der beiden Protokolle.

In vielen Fällen wird sich die Umstellung von Routern, Servern und Endgeräten nicht so konsequent schrittweise durchführen lassen, wie es hier beschrieben wurde. Zur differenzierten Betrachtung und für Mischszenarien beschreiben wir daher nachfolgend Standort-Designs als Gesamtlösung.

Standort Design zur Migration:
Dual Stack
Ein Standort-Design mit Dual Stack wie in Abbildung 4.14 sollte für die IPv6 Adressen ein Global Unicast oder Unique Local Präfix haben ebenso wie DNS-Einträge, die Hostnamen sowohl auf IPv4 als auch auf IPv6 Adressen mappen. Anwendungen wählen je nach DNS-Antwort IPv4 oder IPv6 als Protokoll. Wie zuvor beschrieben müssen alle Router / Layer-3 Switches IPv6-fähig sein und einen Dual Stack implementiert haben, um IPv4-

und IPv6 Anwendungen parallel über den Backbone routen zu können. IPv6 Routen werden mit den IPv6 Routing Protokollen (z.B. RIPv6, OSPFv3) gelernt und propagiert. An dieser Stelle sei nochmals darauf hingewiesen, dass die Router ein doppeltes Adress-Schema, v4 und v6 Routing Protokolle sowie ausreichend Speicher und CPU-Leistung benötigen und in ihrer Hardware gegebenenfalls entsprechend aufgerüstet werden müssen.

Das Dual Stack Design ist im Vergleich mit der Hybrid oder Service Core Lösung bei weitem das übersichtlichste Design. Der Nachteil liegt darin, dass alle Router sofort auf IPv6 hochgerüstet werden müssen.

Standort-Design zur Migration: Hybridlösung

Ein Hybrides Standort-Design nutzt mehrere oder alle beschriebenen Verfahren, je nach Funktionalität der einzelnen Komponenten. Es erlaubt eine Hochrüstung und Migration auf IPv6 in fast beliebig kleinen Schritten, indem so lange Tunnel eingesetzt werden, bis eine durchgängige Dual Stack Funktionalität erreicht ist. Zudem erlaubt es, dass IPv6 Anwendungen über sehr große Netze hinweg bereitgestellt werden, die nicht durchgängig IPv6-fähig sein müssen, und bietet damit eine extreme Flexibilität für den IPv6-Einsatz. Andererseits erhöht der parallele Betrieb verschiedener Verfahren die Komplexität und den Aufwand bei der Fehlersuche. Ein typisches Einsatzszenario für ein Hybrides Design liegt vor, wenn einzelne Gebäudezugangs- / Konzentrationspunkte (Distribution Ebene) nicht IPv6-fähig sind, der Core, der RZ-Zugang und die Server jedoch schon mit Dual Stack nach IPv6 migriert sind wie im Beispiel aus Abbildung 4.15. Für ein Hybrides Standort-Design empfehlen wir den Einsatz einheitlicher Tunnel (entweder ISATAP oder manuelle Konfiguration), um die Verfahrens-Vielfalt einzuschränken.

Leider hat ISATAP den Nachteil, dass es IPv6 Multicast nicht unterstützt. Ein weiterer Nachteil ist darin zu sehen, dass ISATAP die Infrastruktur eigentlich unnatürlich „verbiegt“, indem beispielsweise der Core sich wie ein Access Layer verhält. Daher betrachten wir ein Hybrid-Design mit Dual Stack und ISATAP oder ein reines ISATAP Modell als möglichst kurzfristiges Übergangs-Szenario hin zu einem durchgängigen Dual Stack Modell für Server, Router und Layer-3 Switches.

Standort-Design: IPv6 Service Core

Dieses Standort-Design hat im Produktivnetzwerk ein Overlay Netzwerk wie das Hybride Design, jedoch sind die Serverbereiche und der Internet-Zugang tunnelfrei und komplett auf IPv6 umgestellt. Die Standort-Infrastruktur mit Core, Gebäudekonzentrationspunkten und Access Bereichen läuft wie bisher unverändert auf IPv4. Der vorhandene Core wird an einen neuen, IPv6-fähigen Service Core angebunden. Soweit schon IPv6-Endgeräte vorhanden sind, bauen diese bei Bedarf redundante ISATAP Tunnel (Primär, Sekundär) zum redundanten Service Core auf, der alle Tunnel terminiert und IPv4-Pakete nach IPv6 wandelt (Gateway-Funktion). Der Service Core kommuniziert mit dem Internet und dem RZ-Zugang (Aggregation Layer) ausschließlich über IPv6. Falls es im RZ noch IPv4 Server gibt, so muss der Service Core und RZ-Zugang anstelle reiner IPv6 Kommunikation den Dual Stack nutzen.

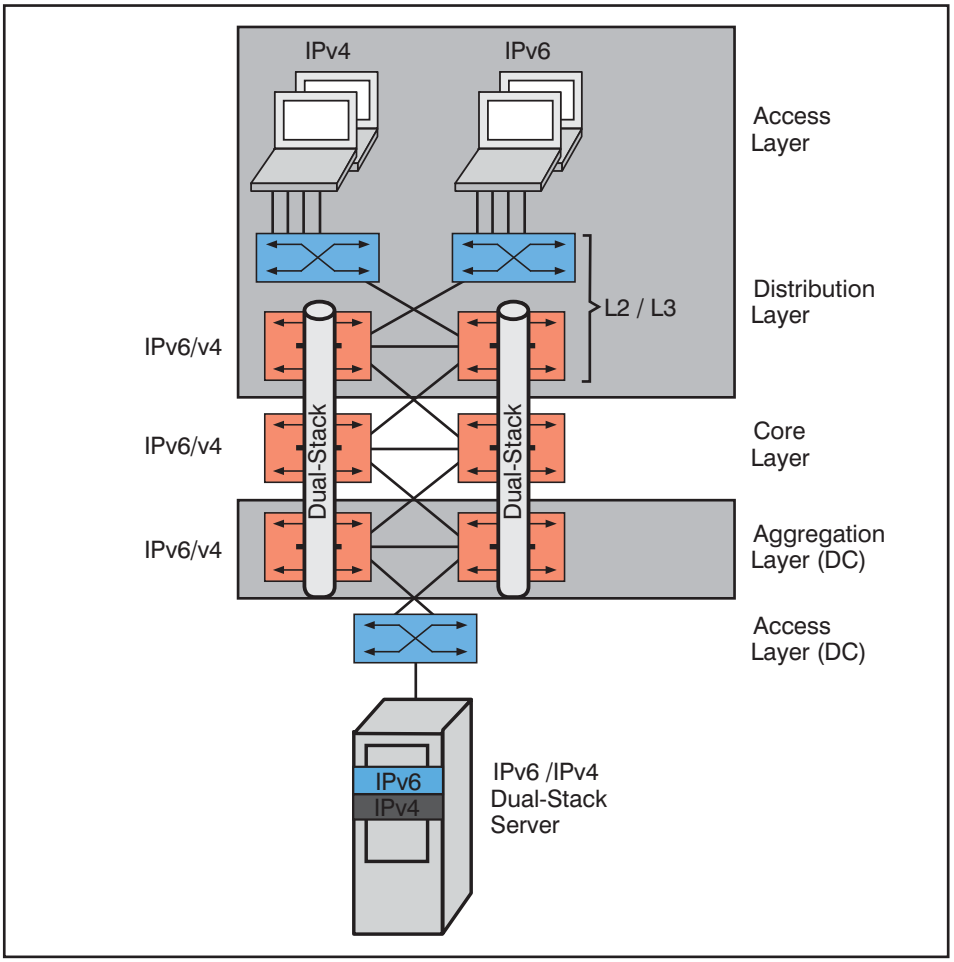

Abbildung 4.14: Standort-Design mit Dual Stack

Das Service Core Design vermeidet in den RZ-/Internet-Service-Bereichen Dual Stack Aufwände. Die vorhandene Standort-Infrastruktur wird zuerst unverändert weitergenutzt und im Nachgang schrittweise auf IPv6 umgerüstet und mit Dual Stack betrieben. Je weiter die Migration fortschreitet, umso weniger ISATAP Tunnel müssen von den Endgeräten genutzt werden. Dieses Design hat einen erhöhten Komponenten-Aufwand, da der Service Core zusätzlich zum normalen Core aufgebaut werden muss. Die Vorteile liegen darin, dass sich die vorhandene Netzwerk-Infrastruktur zuerst unverändert weiter nutzen lässt, dass der Service Core alle Tunnel terminieren kann und sich die Serverbereiche von Anfang an tunnelfrei betreiben lassen. Dieses Design erlaubt eine sehr schnelle Einführung von IPv6 ohne umfangreiche Änderungen am vorhandenen Produktivnetz im ersten Schritt.

Erforderliche Kernfunktionen

Welche Funktionen sollte eine IPv6-ready Komponente unterstützen? Zuerst einmal muss die Lizenz IPv6 „erlauben“ d.h. grundsätzlich unterstützen. Darüber hinaus beinhaltet IPv6-Readiness sicher die in Teil 1 und diesem Artikel beschriebenen Funktionen wie

- stateless Autokonfiguration für die IPv6 Adresse

Schwerpunktthema

Evolution der RZ-Netze: Konvergenz und Konzentration (Teil 2)

Fortsetzung von Seite 1

Dr. Franz-Joachim Kauffels ist einer der erfahrensten und bekanntesten Referenten der gesamten Netzwerkszene (über 20 Fachbücher und unzählige Artikel) und bekannt für lebendige und mitreißende Seminare.

„Für das Problem der Latenz in RZ-Netzen ist der Wechsel der Datenrate von 10GbE auf 100 GbE von untergeordneter Bedeutung. Bezogen auf eine produktbedingte Latenz von Switchmodellen ist die Wirkung der Erhöhung der Datenrate gering. Ausschlaggebend für die Gesamt-Latenz ist vielmehr die produktbedingte Latenz."

Wie wir sehen werden, hängen diese Thesen eng zusammen.

Schon in 2009 wurde vom Autor im Rahmen eines Forums die These aufgestellt, dass bestehende Ethernet-Switches nicht einfach durch einen Software-Upgrade auf die Fähigkeit der Verarbeitung von Speicherverkehr mittels DCB und FCoE aufgerüstet werden können. Mittlerweile hat sich herausgestellt, dass eine Implementierung dieser Funktionen mit der Notwendigkeit neuer ASICs in den Switches (oder der Einführung ganz neuer Switches) verbunden ist.

Wie auch immer lohnt es sich aufgrund des Investitionsvolumens, genau hinzuschauen, welche Funktionen man denn nun wirklich benötigt, welche „ganz nett" aber nicht entscheidend und welche schlicht überflüssig sind.

Wir haben ja noch nicht viel zu 100 GbE gesagt, das kommt im nächsten Teil dieser Reihe. Die DCB-Problematik tritt aber auch schon (und wie wir sehen werden grade) bei Netzen auf, die mit 10 GbE ausgerüstet sind. Für heute reicht es, wenn Sie sich 100 GbE einfach als zehnfach paralleles 10 GbE vorstellen.

Zur Notwendigkeit der Einführung von DCB-Funktionen

Die DCB-Funktionen wurden von Frau Borowka-Gatzweiler hervorragend dargestellt. Wer nicht auf dem Netzwerk-Redesign Forum oder der Sommerschule 2010 war, sollte sich unbedingt den neuen Report von Frau Borowka-Gatzweiler „Aktuelle Netzwerkstandards in der Analyse" zum Nachlesen beschaffen. Für die Zwecke dieser Darstellung reicht eine kurze Abstraktion.

Die DCB-Funktionen umfassen:

- Staubenachrichtigung und rückwärts gerichtete Staubehebung
- Prioritäts-basierte Flusskontrolle
- Erweiterte Verkehrssteuerung
- System zum Informationsaustausch der beteiligten Switches hinsichtlich der von ihnen unterstützten Funktionen

DCB bedeutet „Data Center Bridges". Die Funktionen wurden definiert, um der Idealvorstellung eines „lossless Ethernet" möglichst nahe zu kommen. Diese zusätzliche Eigenschaft eines Ethernets wird dann benötigt, wenn man FCoE mittels FC-BB-5/ FC-BB_E implementieren möchte.

Wir untersuchen die Funktionen der Reihe nach und geben wenn möglich Alternativen an.

DCB Congestion Notification und Congestion Control

Konventionelle Ethernet-Switches haben die unangenehme Eigenschaft, Pakete zu verwerfen, wenn die Eingangswarteschlangen in ihrer Summe so lang werden, dass der Switch annehmen muss, diese Warteschlangen nicht mehr sinnvoll abarbeiten zu können. Mögliche Gründe für das Auftreten dieser Situation sind:

- Schlecht und mit zu hoher Überbuchung konstruiertes Netz
- Zu kleine Speicher in den Switches
- Defekt im Switch
- Zu hohe Belastung des Switches aus multiplen Quellen

DCB Congestion Control arbeitet so, dass im Falle des bald zu erwartenden Auftretens einer Stausituation von den betroffenen Eingangsports (meist ist dies nur einer) eines betroffenden Switches eine Congestion Notification an den Auslöser der Überlast geschickt wird. Dieser senkt dann seine Datenrate. Die Hoffnung ist, dass sich die Stausituation dadurch auflöst.

Dabei ist zu bemerken, dass Congestion Control nur dann funktionieren kann, wenn die NIC des zu viel Verkehr erzeugenden Gerätes auch die entsprechenden Funktionen zur Senkung der Datenrate besitzt. Generell ist das vergleichbar mit der Senkung der Datenrate bei WLANs, wenn sich die Umgebungsbedingungen verschlechtern.

Praktisch ist DCB Congestion Control also auf die Endsysteme beschränkt, die das unterstützen. Das sind zurzeit nur ganz moderne Server mit entsprechend eingerichteter Virtualisierungssoftware und einer I/O-Implementierung, die die Einschränkungen differenziert an die Virtuellen Maschinen weitergibt.

An den ersten drei Punkten kann DCB Congestion Control nichts ausrichten, sondern lediglich am letzten. Das sollte man aber auch nicht übertreiben, IEEE 802.1 sagt selbst, dass Simulationen gezeigt hätten, dass ein Verlust von Paketen durch die Congestion Control nicht völlig vermeidbar ist. Das Verfahren ist überdies nicht deterministisch.

Bei den zu untersuchenden Thesen geht es ausschließlich um neue Netze, die mit

Evolution der RZ-Netze: Konvergenz und Konzentration - Teil 2

neuen Core Switches ausgerüstet werden. Aufgrund der Erfahrung und der zukünftigen Anforderungen wird man ein neues Netz direkt so designen, dass man auf Überbuchung verzichtet und es insgesamt blockierungsfrei auslegt.

Durch das neue IS-IS-basierte von der IETF standardisierte Strukturierungsverfahren TRILL stehen hier ganz andere Möglichkeiten für das Design von Netzen zur Verfügung, weil redundante Ersatzwege nicht mehr wie bei Spanning Tree einfach ausgeschaltet, sondern aktiv mitbenutzt werden. Dies kann man deshalb jetzt endlich mit den seit Jahrzehnten existierenden Erkenntnissen aus der Theorie der Mehrstufenmehrfachverbindungsnetzwerke, wie z.B. dem Clos Netz, kombinieren und kommt damit zu völlig neuartigen, optimierten und vor allem völlig blockierungsfreien Netz-Designs. Dies ist hier aber nicht das Thema. Frau Borowka-Gatzweiler wird es sicherlich bald systematisch aufgreifen, weil sie hier schon erhebliche Vorarbeit geleistet hat.

Wir können aber festhalten:

Wird ein neues Netz mit neuen Core-Switches so konstruiert, dass es frei von Überbuchungen ist und blockierungsfrei arbeitet, ist die DCB Congestion Control völlig überflüssig.

Das einzige Problem, was in einem solchen Netz noch auftauchen kann, ist ein technischer Defekt in Line Card oder Switch. Dagegen kann aber DCB Congestion Control ohnehin nichts machen, es ist in diesem Zusammenhang eine konstruktive Aufgabe, das Netz auf hinreichende Redundanz für solche Fälle auszulegen.

DCB Priority Based Flow Control

Ein weiterer Mechanismus, um Staupunkte möglichst im Vorfeld zu vermeiden, ist die Priority-based Flow Control. Das ist im Grunde genommen eine Erweiterung von 802.1p mit der Möglichkeit der Differenzierung der Prioritätensteuerung auf die einzelnen Elemente eines zusammengesetzten Verkehrsstroms.

Um die mögliche Wirkung zu untersuchen, greifen wir auf eine Warteschlangenanalyse zurück.

Abbildung 1 zeigt die Modellierungsgrundlage.

Die Sendewünsche aus allen Rand-Switches (z.B. Access Switches, ToR-Switches, Blade-Switches) bilden einen Auftragsstrom für den Core Switch mit der Ankunftsrate Lambda. Die Core-Switches bilden einen Bediener mit der Bedienrate My. Wegen der Kleinrock´schen Unabhängigkeitsannahme reicht es zunächst, einen einzigen Core Switch zu modellieren. Die Bedienzeit eines aus mehreren Switches zusammengesetzten Netzes ergibt sich wegen dieser Annahme aus der Summe der Bedienzeiten der einzelnen Switches. Das gilt sinngemäß auch für alle anderen Parameter. Wir kommen gleich nochmal darauf zurück.

Die Warteschlangentheorie bietet einige wichtige grundsätzliche Ergebnisse und Beziehungen, die bei der Bewertung extrem praktisch sind. Dazu gehört, dass wenn Lambda < 0,7 My , die Warteschlange fast immer leer ist. Wenn die Warteschlange fast immer leer ist, ist das Delay in der Warteschlange fast immer Null. Es gibt nur das Delay des Bedieners. Bei Core-Switches ist dies fast konstant. Daher ist die Varianz des Delays in diesem Fall immer gering.

Für die Modellierung machen wir einige Voraussetzungen:

- Betrachtet wird die Bedienung multipler Eingänge. Diese werden nicht weiter differenziert, sondern bilden insgesamt eine Summenlast
- Der Switch ist so konstruiert, dass er diese Summenlast problemlos bewältigt. Das ist keine besondere Anforderung, wir bewegen uns im Bereich von Core-Switches
- Ein Ethernet-Paket kann minimal 64, maximal 1526 Bytes lang werden. Bei Gleichverteilung, aber auch bei Gaussscher Verteilung wäre der Mittelwert 731 Byte, also 5848 Bit. Um den Effekt der zwangsweisen IFG oder der Verlängerung des Paketformates durch TRILL problemlos zu erfassen, runden wir das auf 6000 Bit auf. Das ist die mittlere Nachrichtenlänge $1/\mu$
- Die mittlere Überragungszeit ist $1/\mu C$ Sekunden. C = 10 Gbit/s

Wir verwenden für die einfache Analyse ein M/D/1-Modell mit einem Poisson-Prozess am Eingang. Dieser besagt, dass die Zeiten zwischen Ankünften von Paketen (Zwischenankunftszeiten) exponential verteilt und unabhängig voneinander sind. Das ist eine sehr gute Annäherung an die Realität. Der Switch arbeitet in wenigen Schritten deterministisch. Das M/D/1-Modell ist ein Spezialfall des M/G/1-Modells. Durch die Annahme des Determinismusses können die Formeln erheblich vereinfacht werden. Für die Bewertung der Funktion der Priorisierung benötigen wir allerdings leider dann doch das wesentlich komplexere M/G/1-Modell, da wir keine Aussagen darüber haben, wie ein Switch die Funktionen zu Priorisierung tatsächlich umsetzt und welche Konsequenzen das für sein Bedienverhalten hat. Also müssen wir einen „G“-Bediener annehmen, G steht für „generell, ohne vereinfachende Annahmen). Die Zahl 1 steht dafür, dass wir nur einen Bediener modellieren.

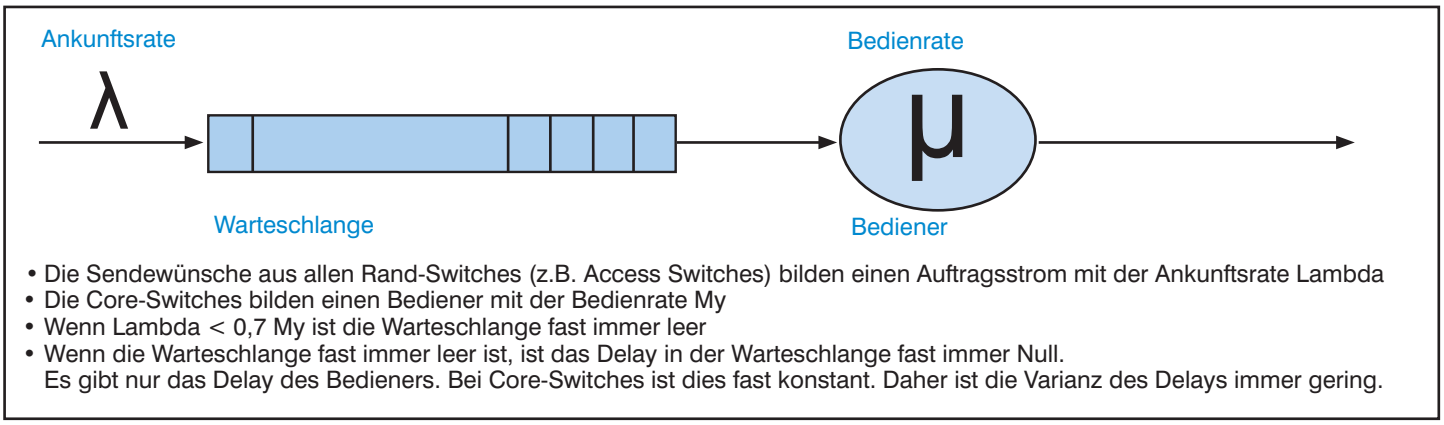

Abbildung 1: Modellierungsgrundlage