ComConsult Research

Der Netzwerk Insider

Juli 2009

Systematische Weiterbildung für Netzwerk- und IT-Professionals

Schwerpunktthema

Neue Anforderungen ändern das Netzdesign:

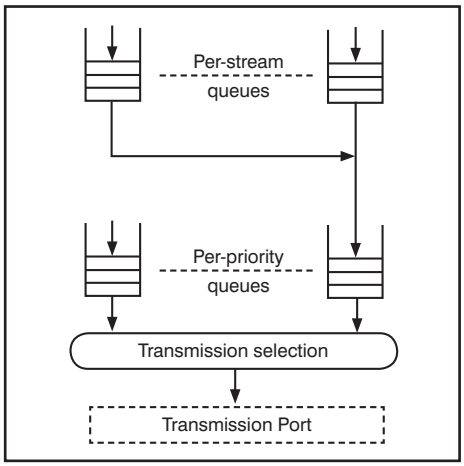
Neue Funktionen und Standards bei IEEE und IETF (Teil 1)

von Dipl.-Inform. Petra Borowka

1. Neue Anwendungen stellen neue Anforderungen an das Netzwerk

Sowohl im Nutzer-/Endgeräte-Bereich als auch im RZ/Backend werden aktuell neue Anwendungen in Ethernet-/IP-Netzen etabliert. Die Basis-Standards für Redundanz und Datenraten bis 10 Gbit sind fertiggestellt, die Skalierung der Ethernet-Datenrate auf 40 Gbit und 100 Gbit ist absehbar.

Somit können sich die Standard-Gremien weiteren ehrgeizigen Zielen zuwenden, die helfen sollen, die neuen Echtzeit-Anwendungen zu optimieren und die neuen Funktionen, die die LAN-Switches hierfür



in den letzten Jahren von den Herstellern erhalten haben, zu standardisieren. Dahinter steht wiederum der Ansatz, neue Funktionen herstellerübergreifend kompatibel und somit geeignet für Multivendor-Umgebungen implementieren zu können.

Neue Anforderungen im Front End

Im Echtzeit-Umfeld etabliert sich nicht nur VoIP als Massenmarkt, sondern auch Video über IP/Ethernet, insbesondere Videokonferenzen kommen zunehmend zum Einsatz. Sowohl Microsoft als auch Cisco stellen deutliche Pushfaktoren für dieses Thema dar.

weiter auf Seite 7

Zweitthema

Der Kampf ums RZ:

die nächste Runde - Teil 2: Matrix, FCoE und EoFC

von Dr. Franz-Joachim Kauffels

Nicht nur Ethernet hat das Potenzial, das führende Netzwerk im Rechenzentrum zu sein. Auch der Fibre Channel hat sich in den letzten Jahren deutlich weiter entwickelt. Dies steht im Zentrum dieses Artikels. Wir werden sehen, dass sich dadurch eine weitere echte Alternative für die Konstruktion eines RZ-Kern-Netzes ergibt und diese an einem Produktbeispiel belegen. Damit das aber richtig wirkt, betrachten wir zuerst mögliche Kostenfallen und mit der HP Blade Server Matrix ein Beispiel dafür, was moderne, für die Virtualisierung ausgelegte Systeme eigentlich von einem RZ-Netz erwarten.

Ethernet hat sich in den letzten Jahren scheinbar zum Universal-Netzwerk für alle denkbaren Anwendungen entwickelt. Doch spätestens bei der Integration von Speichernetzwerken im Rechenzentrum stößt es an Grenzen, die mit der bisherigen Technologie nicht zu durchbrechen sind. Dies hat dazu geführt, dass verschiedene Ansätze zur Erweiterung des Ethernet-Standards entwickelt worden sind, die die bekannten Mängel abstellen sollen. Und da stellt sich eben die Frage, ob das immer ausgedehntere Hinzufügen von Zusatzfunktionen und -Standards, wie es DCB vorsieht, wirklich der richtige Weg ist.

weiter auf Seite 18

Neues Seminar

Virtualisierungstechnologien in der Analyse

ab Seite 4

Geleit

System-Management: vergessene Disziplin?

ab Seite 2

Neuer Report

Konsolidierung im Rechenzentrum

ab Seite 15

Zum Geleit

System-Management: vergessene Disziplin?

Der Betrieb eines Rechenzentrums ist eine komplexe und kostspielige Angelegenheit. Ein Maximum an unterschiedlichen Technologien kommt an einem Ort zusammen und beeinflusst sich gegenseitig. So ist es nicht überraschend, dass Optimierungen im Betrieb das zentrale Argument der Verkäufer im Rahmen des Redesigns von Rechenzentren sind. Speziell mit der Welle der neuen Serverprodukte sind die Begriffe Integration und Konsolidierung untrennbar verbunden. Der Auslöser dieser Entwicklung war Ciscos Ankündigung des Cisco Unified Computings. HP und IBM haben nicht lange gebraucht, um nachzuziehen.

Nun steht der Kunde vor der Situation, dass alle Hersteller im Rahmen der Konsolidierung und Integration speziell das Argument strapazieren, dass diese neue Form von Integration speziell auch im Bereich Management eine neue Dimension von Vereinfachung im Betrieb schafft. Da werden schnell auch Größenordnungen von 30% und mehr an Einsparungen in den Raum gestellt.

Die Ausgangslage und Aufgabenstellung ist klar. Im direkten Umfeld der Server kommen mindestens folgende Technologien zusammen:

- Anwendungen
- Middleware (Datenbanken, Webserver, TCP/IP ...)
- Server-Betriebssysteme
- Server-Hardware
- Daten-Netzwerke
- Speicher-Netzwerke
- Speicher (NAS; SAN; iSCSI; ...)

Die Abhängigkeiten zwischen diesen Technologien sind erheblich. So können Performance-Probleme in einer Anwendung ihre Ursachen in der Anwendung haben, in der Middleware, in der Server-Hardware, im Server-Betriebssystem, im Netzwerk, im Speicher: sprich eine Abgrenzung zur Fehlerlokalisierung ist schwierig. Naturgemäß wird jeder Kunde in dieser Situation mit Begeisterung auf Verbesserungen in den Management-Applikationen reagieren. So gibt jede Verbesserung doch zur immer wiederkehrenden Hoffnung Anlass, dass der Betrieb optimiert werden könnte.

Nun sind Netzwerk- und System-Management relativ alte Disziplinen. Sie hatten ihre Hype-Phase in der zweiten Hälfte der 90er Jahre, doch im Endeffekt sind sie so alt wie die Technologien, denen sie zugeordnet werden können.

Umso mehr überrascht es, dass in regelmäßigen Abständen wesentliche Kern-Erfahrungen im Umgang mit Management-Lösungen irgendwie verloren gehen. Aus diesem Grund im Folgenden eine nicht vollständige Erinnerung an einige typische Erfahrungswerte.

1. Die Ablauforganisation legt den Stellenwert der verschiedenen Management-Disziplinen fest

   Betriebsabläufe können sich zwischen verschiedenen Unternehmen erheblich unterscheiden. Speziell die Größe der Unternehmen, die Zahl der Standorte und die Betriebszeiten haben großen Einfluss auf die Ablauforganisation. Je komplexer ein Unternehmen ist, umso formaler müssen Abläufe beschrieben werden (siehe ITIL). Kerndisziplinen des Managements sind ohne Anspruch auf Vollständigkeit:

   - Konfigurations-Management
   - Operating / Überwachung
   - User-Helpdesk
   - Fehlbearbeitung
   - Änderungs-Management
   - Performance-Management
   - SLA-Management

   Diese Disziplinen oder Management-Bereiche stellen sehr unterschiedliche Anforderungen an technische Hilfsmittel (Tools). Ein kleines Unternehmen, bei dem der Gesamtbetrieb von 10 Personen geleistet wird, hat einen anderen Bedarf als eine verteilte Organisation mit 200 Personen.

2. Der Element-Manager ist das unverzichtbare Hilfsmittel für die Konfiguration und das Trouble-Shooting

   Kein noch so tolles und grafisch ansprechendes System-Management-System kann den der jeweiligen Technologie zugeordneten Element-Manager ersetzen. Seine Qualität entscheidet über die Einfachheit der Konfiguration, die Vermeidung typischer Konfigurationsfehler und die Geschwindigkeit der Fehler-Ermittlung.

3. Probleme werden durch Menschen gelöst

   Die vollständige und perfekte Automatisierung wird es nie geben. Auch die Abbildung von Topologien und Zusammenhängen in entsprechenden Datenbanken ist kein Garant dafür, dass Fehlerursachen automatisch ermittelt werden können. Tatsächlich kommen Top-Spezialisten der Fehlersuche in der Praxis mit sehr einfachen Hilfsmitteln zurecht. Viele der ansprechenden und teuren Top-Tools dienen eher der emotionalen Beruhigung der Führungskräfte. Auch sind sie gut geeignet, Besuchern ein Gefühl von Professionalität zu suggerieren.

4. Hardwareausfälle sind kein Problem

   Fast alle Management-Lösungen haben ihren Ursprung in der Erkennung und Ermittlung von Hardware-Ausfällen. Topologie-Datenbanken helfen in diesem Zusammenhang betroffene Anwendungen und Benutzer zu ermitteln. Aber: Hardware-Ausfälle sind selten und in vielen Fällen ohnehin schnell einzukreisen.

5. „Weiche“ Fehler sind das Kernproblem des Managements

   Performance-Störungen, schwankende Betriebszustände und ähnliche Erscheinungen sind seit dem Beginn von Management das Kernproblem. Dabei ist

zu unterscheiden zwischen der Erkennung des Problems und der Ursachenermittlung. Bei der Erkennung wird jeder Betreiber das natürliche Ziel haben, ein Problem vor dem betroffenen Anwender zu kennen. Das ist dummerweise gar nicht so trivial. Immerhin sitzt der Anwender quasi direkt an der Quelle. Aber auch wenn das Problem erkannt ist, ist gerade bei den weichen Problemen die Ermittlung des Auslösers zum Teil wirklich problematisch.

6. Mehr Information ist nicht immer die Lösung

   Schon immer waren Anbieter von System-Management-Lösung schnell mit dem Argument dabei, dass man nur genügend Detailinformationen aus den betriebenen Technologien absaugen muss und schon wäre man in der Lage, auch alle Probleme zu erkennen. Teilweise stimmt das, teilweise aber auch nicht. Das typische Gegenbeispiel ist die Messung der CPU-Last oder der Speicherauslastung. Kann man wirklich aus einer 90% belasteten CPU einen Rückschluss auf die Antwortzeit einer Datenbank-Applikation ziehen?

7. Einfache Lösungen bringen den größten Zugewinn

   Das Design und die Umsetzung scheinbar guter Management-Lösungen kann extrem teuer werden. Doch statt immer mehr Information und Regeln in ein System zu füttern, kann eine einfachere Lösung, die die Betriebssituation des Anwenders ermittelt, die Erkennung von Problemen deutlich vereinfachen. Es hat sich dabei nicht bewährt, direkt auf den Endgeräten der Anwender zu messen, der Aufwand wird relativ hoch und die Abhängigkeit von der Konfiguration und Nutzung ist immer gegeben. Einfach und schnell umzusetzen sind Robot-Systeme, die typische Transaktionen künstlich ausführen und dabei die Antwortzeiten etc. ermitteln. Dies hilft nicht bei der Ermittlung der Fehlerursache, aber die Kenntnis einer Störung ist die Basis der Bearbeitung.

8. Drill-Down ist ein fragwürdiger Mehrwert

   Fast alle großen Management-Lösungen von BMC, HP und IBM arbeiten mit der Verdichtung von Information. Auf der obersten Informationsebene werden nur elementare Betriebsdaten, am liebsten in bunten Icons angezeigt. Bei Störungen erlauben die Systeme, gezielt in die Tiefe zu gehen und mehr Detail-Informationen abzurufen. Die Erstellung dieser Systeme ist sehr aufwendig und führt nicht immer zu einem Erfolg. Auch hier gilt: der größte Erfolg kommt durch den Menschen und sein Knowhow. Kein technisches System kann einen erfahrenen Trouble-Shooter ersetzen.

9. Normalisierung ist unverzichtbar

   Gerade in der Ermittlung von Performance-Werten oder in der Bearbeitung weicher Fehler hat jedes Unternehmen eine andere Ausgangslage. In jedem Fall sind Informationen über den als normal angesehenen Betriebszustand die Basis für jede Fehlerbearbeitung. Ein gutes Berichtswesen ist deshalb die Basis der Störungsbearbeitung.

10. 24/7 ist eine Herausforderung

    Schichtbetrieb und 7 Tage-Betriebszeiten sind die zentrale Herausforderung des System-Managements. So sehr der erfahrende Trouble-Shooter unersetzbar ist, so wenig wird er 24/7 anwesend sein. Die Auswahl der Betriebsinformation, die im 24/7-Überwachungs-Management angezeigt wird und die Werkzeuge zur Abgrenzung der Fehler zwischen Technologien sind entscheidend.

Die Menge dieser Regeln und Erfahrungswerte könnte ohne Probleme fast beliebig erweitert werden. Tatsache ist, dass eine Integration der verschiedenen Technologien eines Rechenzentrums auf Toolebene nicht wirklich glaubhaft ist. Gerade im Rechenzentrum sind gute Element-Manager in der Konfiguration von Servern und SANs unverzichtbar.

Was wirklich hilft, ist die Reduzierung der Menge zu managender Komponenten und die Reduzierung der Komplexität auf dieser Ebene. Hier hat Cisco mit seiner Unified Computing Initiative der Branche einen guten Dienst erwiesen. Die Vereinfachung der technischen Architektur und die Reduzierung der Komponentenzahl bringt mehr als scheinbare Integrationen auf der Tool-Ebene.

Ihr
Dr. Jürgen Suppan

Neues Seminar

Virtualisierungstechnologien in der Analyse

Die ComConsult Akademie veranstaltet vom 14.09. - 15.09.09 ihr neues Seminar „Virtualisierungstechnologien in der Analyse“ in Köln.

Virtualisierung ist die wichtigste IT-Technologie des Jahres 2009. Sie beinhaltet das Potenzial, in allen Kostenbereichen des Rechenzentrums, aber auch der Clients zu erheblichen Einsparungen zu kommen. Betroffen sind die Server-Hardware, der Platz-, Strom und Klimatisierungsbedarf und vor allem die Konsolidierung von Speicher. Ein weiterer wichtiger Gesichtspunkt bei der Umsetzung gerade größerer Lösungen ist die Vereinfachung des Betriebs von Rechenzentren. Virtualisierung kann als die führende Zukunfts-Technologie für optimierte Rechenzentren angesehen werden.

Virtualisierung erfolgt in 3 Grundstufen:
Stufe 1: Basis-Virtualisierung einzelner Server
Stufe 2: Ausbau der Servermenge und Einführung geeigneter Management-Software
Stufe 3: Aufbau einer virtuellen Infrastruktur für Server und Speicher mit Lastverteilung, Verlagerung von virtuellen Maschinen und Hochverfügbarkeit

- Wie sieht die grundsätzliche Architektur einer Virtualisierungslösung aus (Hyper-Visor, virtueller Switch, HBA, Speicher, Treiber)?

- Wie unterscheiden sich die 3 marktgrößten Virtualisierungslösungen VMware, Citrix, Microsoft Hyper-V?
  - Architekturunterschiede
  - Hochverfügbarkeit
  - Lizenzmodelle
  - Management-Applikationen
  - Bewertung der verfügbaren Produkte
- Welche Konzepte zur Lastverteilung existieren in virtualisierten Umgebungen?
  - CPU, Hauptspeicher, I/O
- Wie funktioniert das Verschieben von virtuellen Maschinen und Speicher zwischen physikalischen Hosts (sowohl manuell als auch dynamisch)?
- Welche Hardware-Plattformen sind am Markt verfügbar (klassische Server vs. Blades)?
  - Welche Mindestanforderungen an die Hardware gibt es?
- Welche Hardware ist in der Zukunft zu erwarten, insbesondere von Seiten der CPU-Hersteller? Welchen Einfluss wird dies auf die Virtualisierung haben?
- Welche Software ist in Zukunft zu erwarten (z.B. Cisco Nexus 1000V)
- Welche Möglichkeiten gibt es zur SAN-Anbindung (Fibre Channel vs. iSCSI vs. FCoE)?
  - Technische Konzepte
  - Leistungsgrenzen
  - Bewertung
- Wie erfolgt die Verbindung mit dem Datennetz?
- Welche Ansätze gibt es zur Virtualisierung des Datennetzes (z.B. VRF)?
- Welche Unterschiede gibt es bei der Virtualisierung von Sicherheitskomponenten, wie z.B. Firewalls?
- Worin bestehen Vorteile und Gefahren von Sicherheitskomponenten, die als virtuelle Maschine auf den Server Host Systemen implementiert sind?
- Wie sicher sind virtuell getrennte Netze im Vergleich zu physikalisch getrennten Netzen?
- Welche Sicherheitsmerkmale besitzt ein virtueller Switch, der auf einem Server Host System konfiguriert ist? Worin bestehen seine Grenzen hinsichtlich Sicherheit?

Fax-Antwort an ComConsult 02408/955-399

Anmeldung
Virtualisierungstechnologien in der Analyse

Ich buche das Seminar
Virtualisierungstechnologien in der Analyse

14.09. - 15.09.09 in Köln
zum Preis von € 1.390,- zzgl. MwSt.

☐ Bitte reservieren Sie für mich ein Hotelzimmer

vom_____bis_____09

Buchen Sie über unsere Web-Seite
www.comconsult-akademie.de

Vorname	Nachname
Firma	Telefon/Fax
Straße	PLZ,Ort
eMail	Unterschrift

Neuer Report

Neuer Report im Bundle mit aktuellem Kongress

Konsolidierung im Rechenzentrum

Welche Netzwerk-Technologie wird gewinnen?

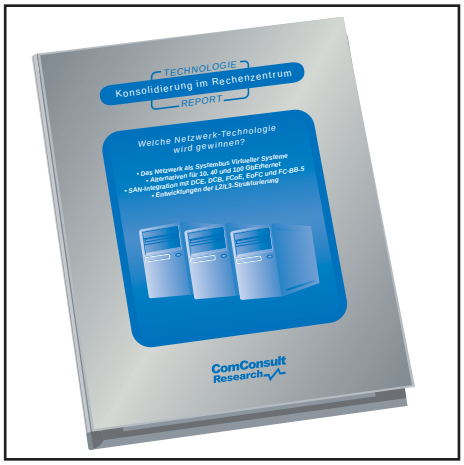
Das Netzwerk als Systembus Virtueller Systeme - Alternativen für 10, 40 und 100 GbEthernet - SAN-Integration mit DCE, DCB, FCoE, EoFC und FC-BB-5 - Entwicklungen der L2/L3-Strukturierung

Nach den Sommerferien erscheint der brandaktuelle Technologie-Report „Konsolidierung im Rechenzentrum: welche Netzwerk-Technologie wird gewinnen“.

Mit dem aktuellen Redesign von Rechenzentren werden die dort bisher eingesetzten Kommunikations-Technologien in Frage gestellt:

- die Anzahl der virtuellen Maschinen pro Hardware-Server nimmt immer weiter zu (Faktor 10),
- parallel nimmt die Dichte von Blade-Servern zu, multipliziert mit der wachsenden Anzahl virtueller Maschinen pro Blade entsteht eine enorme Leistungsdichte, die geeignet auf eine Kommunikations-Infrastruktur abgebildet werden muss
- 10 Gigabit-Ethernet ist jetzt die Standard-Datenrate, die zum Einsatz kommt, damit entfällt die bisher übliche physikalische Zuordnung von virtuellen Maschinen zu Netzwerk-Adaptern, die gemeinsame Nutzung der 10-Gigabit-Adapter muss organisiert werden
- große Mengen von virtuellen Maschinen generieren große Layer-2-Strukturen, für die auch eine passende Strukturform gefunden werden muss
- die Fähigkeit des Wanderns virtueller Maschinen erfordert das „Mitwandern“ der Netzwerk- und Speicher-Zugänge
- der bisherige Parallelbetrieb unterschiedlicher Technologien für Inter-Prozessor-Kommunikation, Datenverkehr und Speicherzugriff wird mit der hohen Dichte problematisch und auch teuer, das Ziel der Konsolidierung aller Übertragungstechnologien in einer Technologie rückt immer mehr in den Vordergrund

Eine Zukunfts-sichere Netzwerk-Technologie in Rechenzentren muss also mindestens folgende Kriterien erfüllen:



- Generell die gezielte Unterstützung virtueller Server-Umgebungen
- Dynamische Zuordnung von virtuellen Maschinen zu Daten- und Speichernetzwerken
- Abbildung verschiedener Formen von Netzwerk-Technologien auf eine zentrale Konsolidierungs-Technologie
- Architektonische Integrations-Fähigkeit in die hohe Dichte von Blades und virtuellen Maschinen bis hin zu einem geeigneten Layer-2-Redundanz-Verfahren

Jeder Betreiber eines Rechenzentrums ist gefordert, hier die richtige Entscheidung zu treffen. Dies wird aber erheblich erschwert durch

- Verschiedene Formen und Ausprägungen von Virtualisierung
- Verschiedene Strategien und Entwicklungsrichtungen der führenden Netzwerk- und Speicher-Hersteller
- Technische Probleme in der Konsolidierung auf eine zentrale Technologie

Typische Themen sind:

- Einsatz des brandneuen Fibre Channel over Ethernet-Standards
- Die zukünftige Bedeutung des Fibre Channels
- Data Center Ethernet, Data Center Bridging
- Neue Strukturkonzepte und Redundanzverfahren
- Virtuelles Switching über physikalische Server hinweg

Dieser hochaktuelle Report von Dr. Franz-Joachim Kauffels greift diese Problematik auf und bietet dem Leser folgende wichtige Hilfestellungen für seine Projekt-Entscheidungen:

- Er analysiert, wo und warum akuter Handlungsbedarf besteht
- Er geht auf die verschiedenen Ausprägungen von Virtualisierung ein
- Alle wesentlichen in der Diskussion befindlichen Kommunikations-Technologien werden beschrieben, analysiert und diskutiert
- Vor allem die auch mit neuen Standards bestehenden Schwachstellen werden heraus gearbeitet und aufgezeigt
- Er analysiert die Struktur-Problematik und zeigt Wege zu optimalen Netzwerk-Strukturen auf

Dieser Report liefert eine unverzichtbare und elementare Hilfe in der Analyse der verschiedenen Technologien und ebnet den weg zu einer Zukunfts-sicheren Entscheidung für die richtige Kommunikations-Technologie im Rechenzentrum. Er sollte in keinem Unternehmen fehlen.

Der Autor Dr. Franz-Joachim Kauffels ist einer der erfahrensten und bekanntesten Referenten der gesamten Netzwerkszene und bekannt für lebendige und mitreißende Seminare und unzählige Veröffentlichungen.

Paketangebot

Sparen Sie 260,-- € bei einer Paketbuchung bis zum 30.07.2009

Report

„Konsolidierung im Rechenzentrum„

im Paket mit dem Forum

„Rechenzentrum Infrastruktur-Redesign Forum 2009“

**Wir bieten Ihnen diesen Report bei der Buchung dieses Kongresses zu einem Sonderpreis an. Statt regulär € 398,-- zahlen Sie nur € 338,--
Der bestellte Report wird Ihnen bei der Veranstaltung vor Ort von der Betreuerin ausgehändigt. Wenn Sie noch bis zum 30.07.2009 buchen, profitieren Sie obendrein noch von der Frühbucherphase und sparen insgesamt € 260,-- gegenüber der Summe der regulären Einzelpreise. (alle Preise zzgl. MwSt.)**

Fax-Antwort an ComConsult 02408/955-399

Anmeldung

Rechenzentrum Infrastruktur-Redesign Forum 2009

Ich buche den Kongress
Rechenzentrum Infrastruktur-Redesign Forum 2009

☐ 16.11. - 18.11.09 in Königswinter
Frühbucher-Preis € 1.690,--* zzgl. MwSt.
*gültig bis 31.07.09

☐ inkl. kostenpflichtigem Report
☐ ohne Report

☐ Bitte reservieren Sie mir ein Zimmer

vom _____ bis_____ 09
im Maritim Hotel

Buchen Sie über unsere Web-Seite
www.comconsult-akademie.de

_____ Vorname
_____ Nachname
_____ Firma
_____ Telefon/Fax
_____ Straße
_____ PLZ,Ort
_____ eMail
_____ Unterschrift

Zweitthema

Neue Anforderungen ändern das Netzdesign:

Neue Funktionen und Standards bei IEEE und IETF - Teil 1

Fortsetzung von Seite 1

Dipl. Inform. Petra Borowka 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.

Video-Anwendungen bedeuten erhöhten Bandbreitenbedarf im Echtzeit-Umfeld, besonders Telepräsenz als High End Videokonferenz mit hohem Wirklichkeitsfaktor.

Fassen wir an dieser Stelle die typischen Echtzeit-Anforderungen für Verbindungen bei guter bis sehr guter Qualität zusammen: Unkomprimierte Sprache benötigt eine moderate Bandbreite von ca. 85 kbit (unverschlüsselt) bis 105 kbit (verschlüsselt); die Antworzeit sollte 100 ms (LAN) bis 150 ms (WAN) als One-Way Delay nicht überschreiten, der Jitter sollte im LAN maximal 1 ms bis 5 ms betragen, im WAN maximal 20 ms bis 30 ms. Für Streaming-Anwendungen sind die Delay- und Jitter-Anforderungen etwas entspannter, da die Jitterpuffer mit 200 bis 500 ms erheblich größer sein können als bei interaktiven Anwendungen. Die Verlustrate (inkl. out-of-time Frames) sollte im LAN auf nahezu 0%, im WAN maximal 3% bis 5% eingegrenzt werden, abhängig von eingesetzter Kompression und Verlustcharakteristik (Kompression ist verlustsensitiv, Burst-Verluste können die Qualität auch schon bei insgesamt niedrigem Verlustraten-Prozentsatz erkennbar verschlechtern).

Video-Anwendungen haben bis auf den Bandbreitenbedarf identische Anforderungen wie Voice-Anwendungen. Die Bandbreite einer Vollbild-Videosession kann von 2 Mbit (Desktop-Video) über 8 Mbit bis 45 Mbit bei einer Multipunkt-Verbindung und 300 Mbit z.B. in Fernsehstudios für Preview von Sendungen betragen. Während Voice üblicherweise als kritischer Dienst mit sehr hoher Verfügbarkeits-anforderung eingestuft wird, gilt dies für Video meist nur bei Telepräsenzanwendungen. Da die Endgeräte im Regelfall 1 Gbit-fähig sind und zunehmend 10 Gbit unterstützen, müssen die Netze hier nachziehen: entweder mit Verfahren zur Regulierung von Hoch- und Überlasten oder mit erhöhter Summenbandbreite für Etagen- und Backbone-Anbindungen.

Die weiteren Echtzeitfunktionen für UC: Instant Messaging und Erreichbarkeits-Dienste, sind in ihren Anforderungen deutlich weniger kritisch: der Durchsatz von IM (ohne Attachments) dürfte sich unterhalb von 256 kbit bewegen, die Delay-Anforderung liegt im Sekundenbereich, ein Jitter von mehreren Hundert ms ist verkraftbar.

Die genannten Anforderungen gelten auch für die aktuell als Shared LAN Technologie genutzten Wireless LANs! Für sie könnten die neuen Funktionen ebenfalls genutzt werden, da alle diese Funktionen unter IEEE 802.1 (MAC-Layer-übergreifende Standards) spezifiziert werden.

Hieraus resultieren erweiterte Anforderungen an die Switching Infrastruktur:

- Implementierung von Durchsatzgarantie für kritische Anwendungen
- Garantierte kurze Antwortzeiten
- Sehr schnelle Fehlerumschaltungen
- Effiziente Ausnutzung der im LAN vorhandenen Bandbreiten

Neue Anforderungen im Back End

Zusätzlich zu zentralen TK- und Video-Komponenten, die für die Signalisierung Echtzeit-ähnliche Anforderungen stellen, werden in steigendem Maße neue Dienste für Server-Server Kommunikation und Server-Speicher-Kommunikation im zentralen RZ-/Backend-Bereich eingerichtet, die hohe Anforderungen sowohl an die Bandbreite als auch die Antwortzeit stellen und zudem eine Verlustrate von 0% zwingend erfordern:

- Cluster-Lösungen
- HPC
- HPCC
- IPC
- Virtualisierung
- SAN

Diese neuen Dienste schaffen den Bedarf nach größeren Layer-2-Bereichen, da sie eine Layer-2 Verbindung zwischen den kommunizierenden Systemen erfordern. Gleichzeitig benötigen sie sehr hohe Durchsatzraten, was den Bedarf und Wunsch nach einer effizienteren Ausnutzung der LAN-Switching-Infrastrukturen erzeugt, als es die aktuellen Spanning Tree Varianten ermöglichen.

Dies führt zu weiteren Anforderungen an die Switching Infrastruktur:

- Echtzeit-Anforderungen im RZ (SAN, Voice, Video)
- Konsolidierung von Produktiv- und Backend-Netzen
- Steigender Einsatz von Layer-2 Verfahren
- Zunehmende Größe von Layer-2 Bereichen im RZ / Data Center
- Für Katastrophen-Eignung: Ausdehnung von Layer-2 Redundanz über Standortgrenzen hinweg

Neue Funktionen und Standards

Wie tragen die Standardisierungs-Gremien diesen neuen Anforderungen Rechung? Sowohl ANSI/INCITS als auch IEEE als auch IETF arbeiten an neuen Verfahren zur Erweiterung der Layer-2-Switching Funktionalität. Die IEEE hat unter 802.1 zwei neue Task Groups (TG) ins Leben gerufen: Audio/Video Bridging Systeme (AVB) und Data Center Bridging (DCB). Beide TG‘s sind zwar nicht absolut abhängig davon, aber favorisieren eine effizientere Layer-2 Forwarding- und Redundanz-Struktur als RST/MST, die unter dem

Neue Anforderungen ändern das Netzdesign: Neue Funktionen und Standards bei IEEE und IETF - Teil 1

Begriff „Shortest Path Bridging“ erarbeitet wird. Leider konnten sich IEEE und IETF hier nicht einigen, so dass derzeit IEEE unter 802.1Qaq und IETF unter der TRILL WG/“RBridge“ unterschiedliche Lösungen spezifizieren.

2. Die Task Group AVB: Funktionen für Voice und Video

Die AVB TG will die Implementierung unterschiedlicher Anwendungsklassen, respektive Dienstklassen für QoS und non-Qos detaillierter spezifizieren als das der aktuelle Standard 802.1Q mit der dort zwar nicht explizit vorgeschriebenen, aber doch implizit unterstellten strikten Priorisierung leistet. Die Klassen sollen VLANs unterstützen, d.h. je VLAN steuerbar sein. Im Einzelnen entwickelt die AVB TG vier Unterstandards, von denen zwei in das Dokument 802.1Q (Virtual Bridged Local Area Networks) integriert werden. Allerdings ist nicht vor Mitte bis Ende 2010 mit formalen WG Ballots zu rechnen. Die Einzelstandards von Audio/Video Bridging befassen sich mit den Bereichen

- Profile, Defaults, A/V Erkennung → IEEE 802.1BA
- Zeitgeber und Synchronisation → IEEE 802.1AS
- Flowbasierte Reservierung: Stream Reservation Protokoll → IEEE 802.1Qat
- Forwarding und Queueing für zeitsensitive Anwendungen → IEEE 802.1Qav

Eine Übersicht der neuen Standard-Drafts zeigt Abbildung 1.

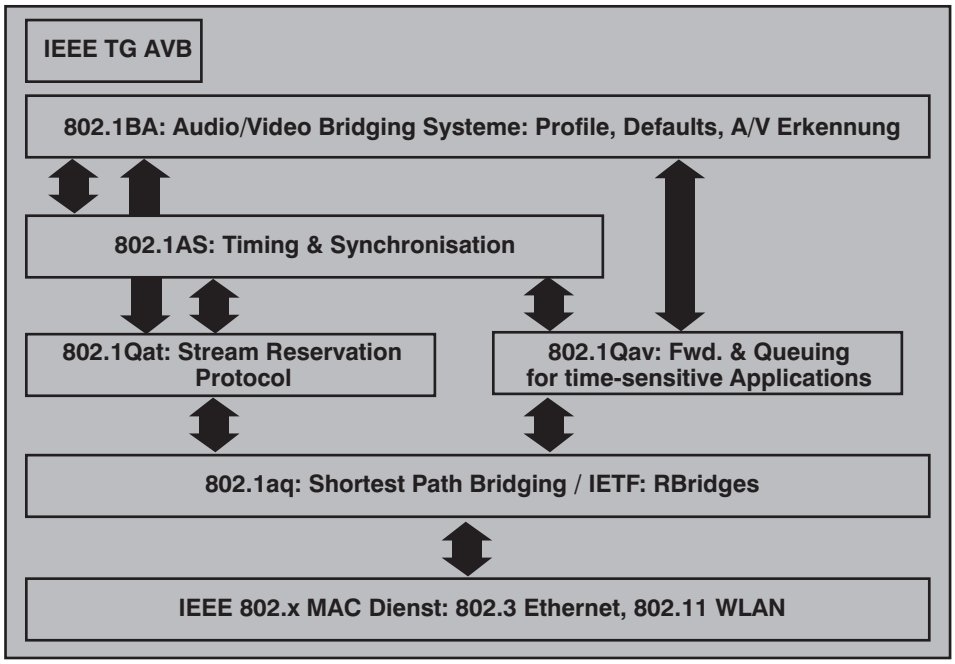


Abbildung 1: Standard-Drafts der IEEE Task Group „Audio/Video Bridging Systems“

IEEE 802.1BA: Profile für AV-Bridging Komponenten

Der Standard definiert Profile für Features, Optionen, Konfigurationen und Default-Werte, die Switches und Endgeräte netzwerktechnisch Audio-/Video-kommunikationsfähig machen. Hierzu gehört auch die gegenseitige Erkennung als „A/V-aware“ oder „nicht A/V-aware“ zwischen benachbarten Komponenten (Endgerät - Switch, Switch - Switch), die bei A/V-aware Nachbarn zu einem Parameterabgleich führen soll. Aus Performancegründen zielt der Standard auf die Nutzung geeigneter Defaults ab, um eine länger dauernde Auswahl (Negotiation) von Parametern zu vermeiden. Da 802.1BA sich auf die anderen Unterstandards abstützt, ist zuerst deren Fertigstellung erforderlich.

IEEE 802.1AS: Zeit-Synchronisierung

Da die Zeitsynchronisierung von RTP auf Layer-2 transparent ist, soll eine separate Synchronisierung das zeitsensitive Audio-/Video-Forwarding auf Layer-2 optimieren. Dies wird seit März 2006 unter IEEE 802.1AS standardisiert, nach der ersten Zielsetzung soll die Fertigstellung Mitte 2010 erfolgen. Die Synchronisierung auf Layer-2 soll nicht nur den normalen Betriebsfall abdecken sondern auch die Anbindung neuer Komponenten, Entfernung vorhandener Komponenten und Behandlung von Störfallsituationen. Zur Synchronisierung wird ein externes Zeitsignal genutzt, Zeitstandards hierfür sind UTC oder TAI. Systeme, die eine solche Synchronisierung unterstützen, werden „time-aware“ genannt.

Die gesamte Zeitsynchronisierung ist verteilt auf mehrere Protokoll-Module (Entities). Ein time-aware System beinhaltet ein Funktionsmodul, das den jeweils besten Zeitmaster findet und eine Synchronisierungs-Funktion. Hierzu gehören sowohl portspezifische Parameter als auch solche, die das gesamte System betreffen.

Ein externer (Standard-)Zeitgeber (TAI, UTC), genannt „ClockSource“, informiert über eine Anwendungs-Schnittstelle den ClockMaster. Dieser gibt die Zeitinformation an das SiteSync Modul weiter. Das arbeitet als „Grandmaster“ und Zeitgeber für die ClockSlaves (synchronisierte Komponenten), die diese Informationen für zeitsensitive Applikationen wie Voice verwenden. Dazu nutzen Sie eine weitere externe Anwendungs-Schnittstelle.

Das PortSync Modul empfängt die Zeitinformation der time-aware Komponente am anderen Ende der Verbindung und vergleicht sie mit der Best Master Information. Das Ergebnis wird wiederum an die SiteSync gegeben. Diese handhabt die Best Master Clock Auswahl für das Gesamtsystem, d.h. sie nutzt die Best Master Informationen, die sie von allen Ports erhalten hat, um zu bestimmen, welcher Port die tatsächlich beste Information hatte und sorgt für ein entsprechendes Update aller Ports.

Abbildung 2 zeigt das Protokoll-Modell mit dem Informationsfluss zwischen synchronisierten Komponenten mittels Portsynchronisierung (PortSyncSync), medienabhängigem Senden (MDSyncSend) und medienabhängigem Empfangen (MDSyncReceive).

IEEE 802.1Qat: Flow-basierte Reservierungen

Viele Hersteller haben in ihren Layer-2 Switches Funktionen für Priorisierung in Kombination mit Bandbreitenreservierung/Limitierung für priorisierte Frames implementiert. Die Arbeitsgruppe IEEE 802.1Qat will diese Funktionen standardisieren und nimmt so einen erneuten Anlauf, das zu tun, womit ATM und auch RSVP seinerzeit als zu komplex gescheitert sind: Die Implementierung von Quality of Service mittels definitiver Ressourcen-Garantie, sprich: Bandbreitenmanagement mittels Reservierung. Reservierungen sollen nun auf Layer-2 für Echtzeitanwendungen, respektive interaktives Audio und Video oder Audio-/Video-Streams auf der Basis von Registrierungen dynamisch geschaltet werden. Dynamisch bedeutet: ein aktiver Stream registriert sich, um Res-

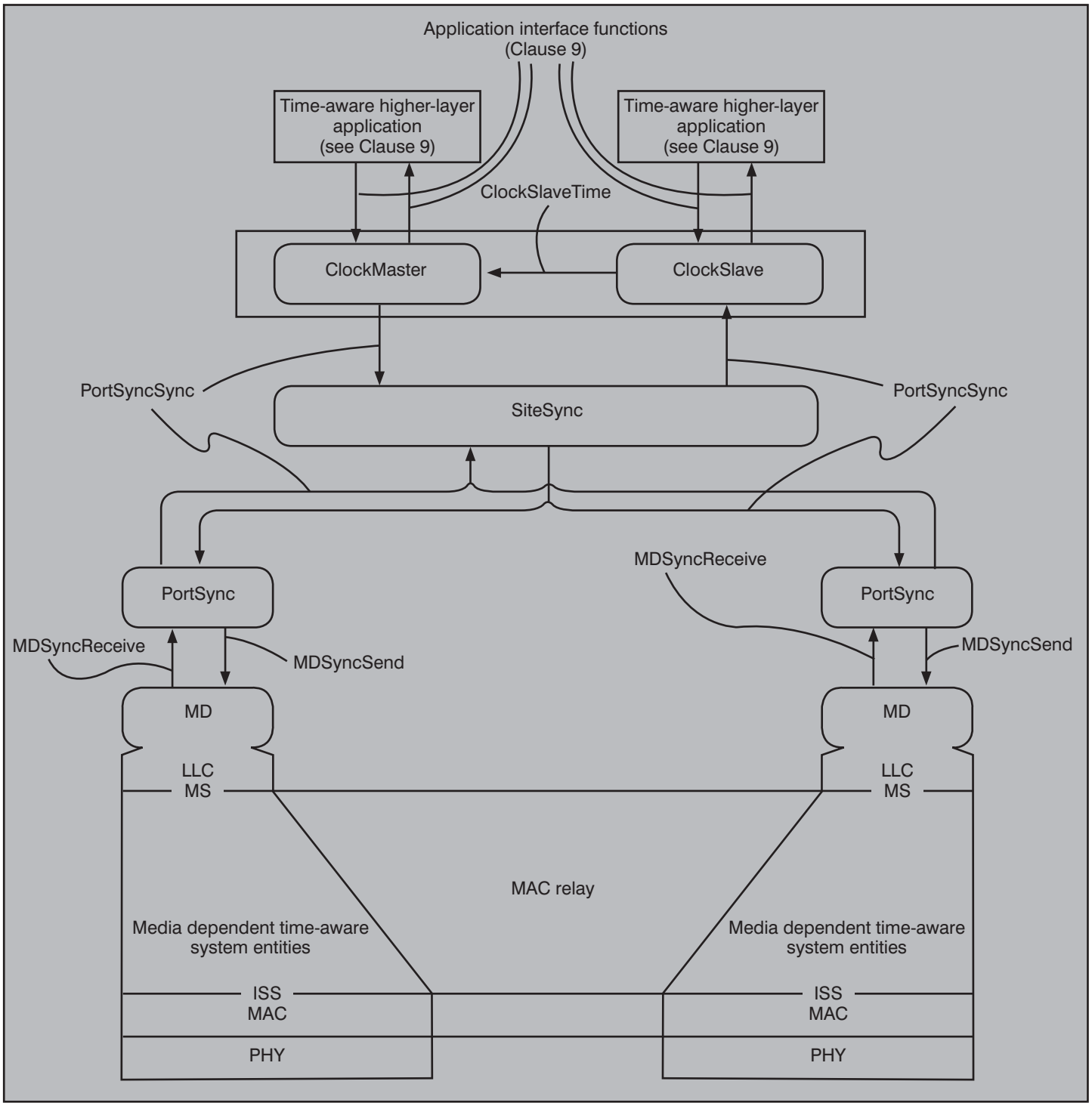


Abbildung 2: Protokoll-Modell für den medienunabhängigen (Sub)Layer „Synchronisierung“

sourcen reservieren zu können. Wenn er endet, deregistriert er sich, damit werden die Ressourcen wieder frei. Das entsprechende Protokoll, das die Reservierung signalisiert, heißt SRP (Stream Reservation Protocol) bzw. in seiner Erweiterung auf verschiedene VLAN‘s MSRP (Multiple SRP). Hierbei wird unterstellt, dass eine typische A/V-Session (Durchschnittsdauer eines Telefonats wird allgemein mit 1,5 bis 3 Min kalkuliert, eine Videokonferenz dauert länger) lang genug ist, damit sich der Aufwand für die Reservierung lohnt. Immerhin wird hier nicht versucht, für jeden Stream eine Reservierung zu fahren, sondern dies zu Stream-Klassen zu aggregieren.

Das (M)SRP nutzt dabei den Nachfolger des guten alten GVRP (GARP VLAN Registration Protocol). Dieser heißt Multiple Registration Protocol (MRP, IEEE 802.1Qak / 2007) und definiert einen Protokollrahmen um diverse „Registrierungs-Anwendungen“ zu handhaben: VLAN Registrierung (MVRP), Multicast Registrierung (MMRP) und eben auch Stream Registrierungen für

MSRP. Für jedes dieser Protokolle wurde ein spezieller eindeutiger MAC Multicast definiert, wie in Abbildung 3 dargestellt. Ein Modell der grundsätzlichen MSRP-Arbeitsweise zeigt Abbildung 4: Die Attribute dieser MRP Anwendung werden innerhalb einer Komponente deklariert (MAD) und mit der Relay Funktion komponentenübergreifend propagiert (MAP). Die Attribute steuern eine Filter-Datenbank, über deren Filter die Weiterleitung, das Verwerfen oder die bevorzugte Bearbeitung als Filteraktionen programmiert werden können.

Das MSRP kann für unterschiedliche VLANs und deren jeweils unterschiedliche Prioritätswerte jeweils eigene Stream-(Reservierungs-)Klassen (SR-Klassen) registrieren, wobei jede Stream-Klasse mittels so genannter TSpecs d.h. Traffic Stream Beschreibungen die für sie erforderlichen Ressourcen (für alle beteiligten Layer-2 Switches) bekanntgibt oder diese für die Klasse fest konfiguriert werden. Mittels SRP können solche Stream-Klassen erkannt und anhand der zugehörigen TSpec die entsprechenden Ressourcen bestimmt werden, die die beteiligten Endgeräte und Switches für diese Klasse dynamisch handhaben müssen. Die zu reservierenden Ressourcen werden entlang des gesamten Layer-2 Weges zwischen Sender und Empfänger mit SRP signalisiert.

Eine Komponente, die (M)SRP unterstützt, besteht immer aus einem Talker und einem Listener; über diese beiden Prozesse wird die Reservierung von Ressourcen VLAN- und Prioritäten-spezifisch mit TSpecs gehandhabt.

Anwendung	Wert
MMRP	01-80-C2-00-00-20
MVRP	01-80-C2-00-00-21
MSRP	01-80-C2-00-00-22

Abbildung 3: MRP Anwendungen und zugeordnete Multicast-Adressen

IEEE 802.1Qav: Forwarding und Queueing für zeitsensitive Anwendungen

Diese Arbeitsgruppe wurde im Februar 2007 ins Leben gerufen und spezifiziert ergänzend zu IEEE 802.1Qat Priorisierung und Warteschlangen-Steuerung. Zeit-sensitive und verlust-sensitive Echtzeit-Audio/Video Datenströme sollen Leistungsgarantien für Antwortzeit und Verlustrate erhalten können. Ein Layer-2 Switch implementiert hierfür einige neue Funktionen: Er misst die Eingangslast je Prioritätsklasse (ingress priority metering), setzt bei Bedarf den Prioritätswert neu (Priority Regenerationln, Remarking) und hat eine Warteschlangen-Steuerung, die die Antwortzeitrestriktionen dieser Anwendungen berücksichtigt. Die notwendige Zeitsynchronisierung wird mit dem zuvor beschriebenen Standard IEEE 802.1AS implementiert. Diese Leistungsgarantien werden für solche Streams gegeben, die das SRP/MSRP nutzen; 802.1Qav setzt also den Einsatz von 802.1Qat voraus.

Bei IEEE 802.1Qav werden Prioritätswerte bereitgestellt und im VLAN-Tag codiert,

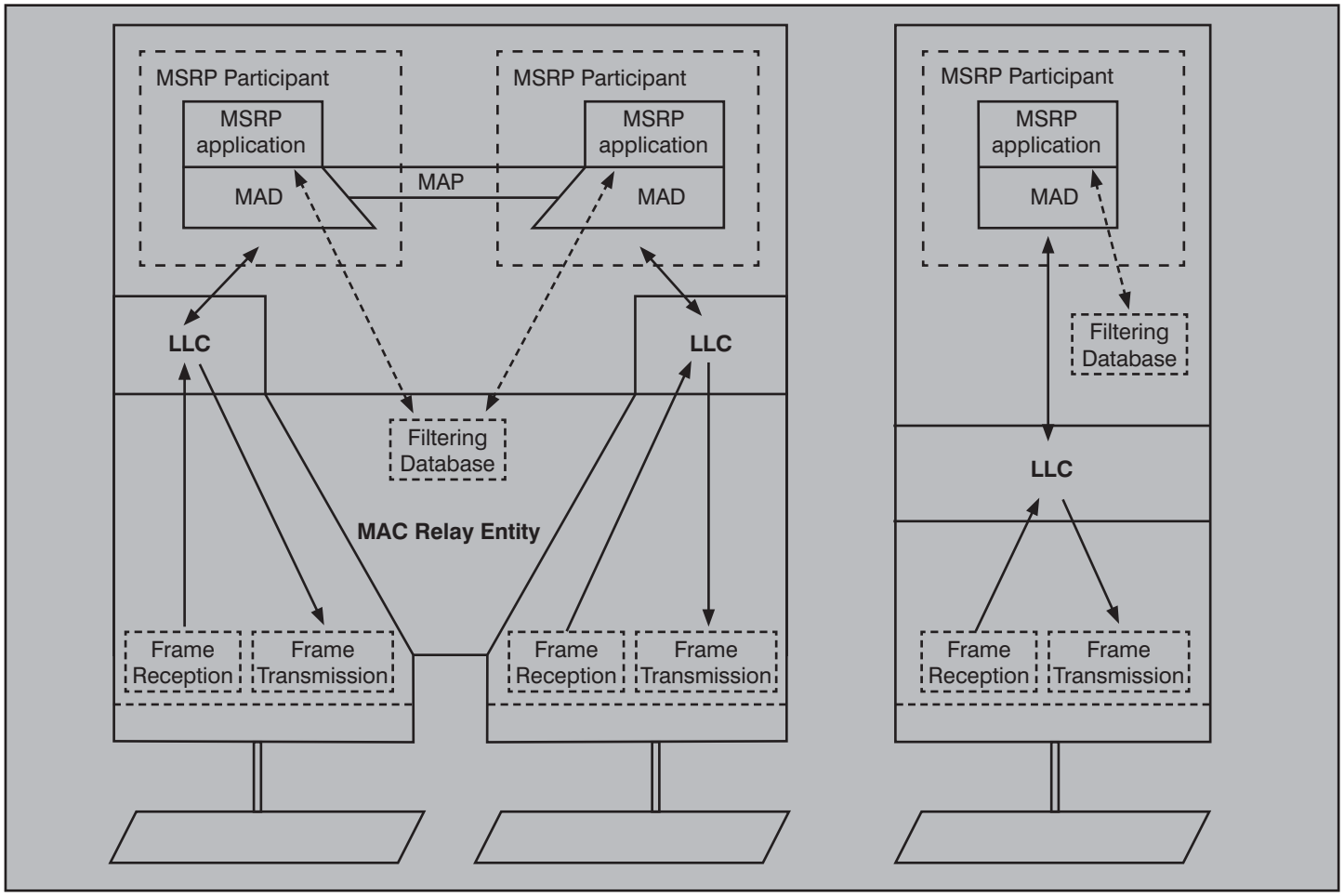


Abbildung 4: Arbeitsweise von MSRP

anhand derer die Warteschlangen-Steuerung den ankommenden Verkehr insgesamt auf überwachte und nicht-überwachte Queues aufteilt. So wird ein paralleles Weiterleiten von kritischem A/V-Verkehr und unkritischem (Daten-)Verkehr über Layer-2 Switches mit verkabelten und drahtlosen LANs ermöglicht. Der Standard wird insbesondere auch A/V-Streaming Anwendungen berücksichtigen, die ja ein gleichmäßigeres Lastaufkommen als interaktive Echtzeit-Anwendungen haben. Zu den speziellen Funktionen von 802.1Qav gehört die Harmonisierung von Jitter und Verlustrate – zwei Faktoren, die immer wieder in Echtzeitumgebungen zu Problemen führen, wenn sie nicht vernünftig limitiert werden können, das gilt insbesondere für starke Jitter-Schwankungen. Auch hier wird die Harmonisierung ehrgeizig sowohl für verkabelte als auch drahtlose als auch gemischte LAN-Netzwerke angestrebt. Ist diese Funktion einmal implementiert, können auch die weniger kritischen non-Echtzeit-Anwendungen davon profitieren.

IEEE 802.1Qav geht ebenso wie IEEE 802.1Qat nicht von einzelnen Streams aus, sondern ordnet die Streams Verkehrsklassen zu, die in der Warteschlangen-Steuerung entsprechend berücksichtigt werden, z.B.

- A/V Verkehr
- (Data Center Verkehr?)
- non-AV Verkehr

Für die sensitiven Anwendungen reserviert das Protokoll in den beteiligten Layer-2 Switches Bandbreiten, die jedoch für andere Anwendungen freigegeben werden, soweit die sensitiven Anwendungen sie nicht nutzen. Dies erfordert jedoch ein preemptive Verhalten: Wird eine sensitive Anwendung aktiv, so bekommt unter Umständen eine nicht-sensitive Anwendung spontan Bandbreite weggenommen. Das ist aber ja auch nicht weiter schlimm, sofern die nicht-sensitive Anwendung dabei nicht verhungert. Letzteres verhindert ein Credit-based Shaper Algorithmus, der für jede Verkehrsklasse N die konfigurierten Bandbreitenparameter misst und überwacht:

Damit nicht-priorisierte Verkehrsklassen nicht verhungern, darf die Summe aller deltaBandwidth-Werte 75% der Gesamtbandbreite eines Switchports nicht überschreiten.

Die 802.1Qat SR-Klassen für sensitive Anwendungen können nun auf 802.1Qav Verkehrsklassen gemappt werden, die den Credit-based Shaper Algorithmus als Verkehrssteuerung auswählen (Transmission Selection Verfahren). Es können auch mehrere Stream Reservation Klassen auf dieselbe Priority-Queue gemappt werden. Hierbei gilt:

- Verkehrsklassen, die den Credit-based Shaper unterstützen, haben immer höhere Priorität als solche, die dies nicht tun
- Es muss mindestens eine Verkehrsklasse definiert werden, die den Credit-based Shaper nutzt
- Es muss mindestens eine Verkehrsklasse definiert werden, die keinen Credit-based Shaper nutzt (und dann „den Rest“ der vorhandenen Bandbreite konsumieren kann)

Freie Bandbreite wird wie zuvor beschrieben von Klassen mit niedrigerer Priorität mitgenutzt. Eine grafische Übersicht des Mappings von Streams auf Priority-Queues zeigt Abbildung 5.

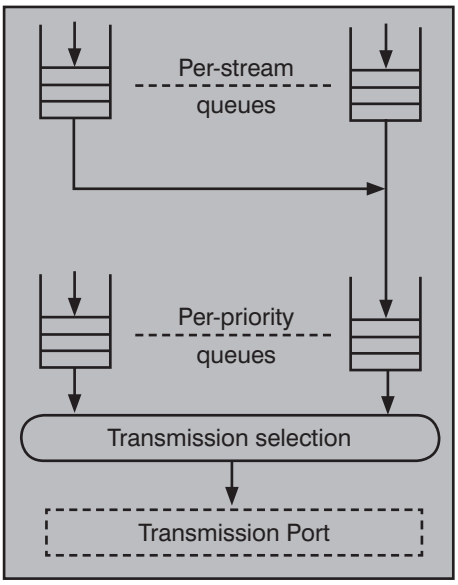


Abbildung 5: Warteschlangen-Steuerung in IEEE 802.1Qav 802.1Qav Computing

3. Basisverfahren für die neuen Standards: Shortest Path Bridging

Die Vorteile von Layer-2 Netzen und Spanning Tree als Redundanzverfahren liegen auf der Hand: Innerhalb eines Layer-2 Netzes können Umzüge mit einer gleichbleibenden IP Konfiguration automatisch mittels Neulernen der MAC Adresse am Layer-2 Switch durchgeführt werden. Der Spanning Tree konfiguriert sich automatisch ohne bzw. mit nur sehr geringen Administrationsvorgaben. Auf der anderen Seite hat Spanning Tree auch einige wesentlichen Nachteile: Die wenig effiziente Frame-Weiterleitung des STP und RST, die auch durch MSTP nur bedingt verbessert werden konnte, ist sowohl den Anwendern als auch den Standardisierern inzwischen ein Dorn im Auge. Anwender setzen teilweise getunte vermaschte Link-Aggregierungs-Verfahren ein, um Spanning Tree zu vermeiden.

Fassen wir die STP-Problematik hier nochmals zusammen: Innerhalb eines VLANs wird eine vollständige und loopfreie Verbindungsstruktur dadurch erreicht, dass zwischen allen Systemen genau ein Weg erhalten bleibt und alle redundanten Wege für den aktiven Datentransport abgeschaltet werden. Im Regelfall führt dies bei vernünftigen Netzdesigns zu einer aktiven Sternstruktur, in deren Zentrum die Spanning Tree Root liegt. Die Verbindung zweier Edge-Komponenten erfolgt dann über mehrere Hops und nutzt den Weg über die Root, selbst wenn es eine kürzere Querverbindung gäbe – denn diese wurde vom Spanning Tree für den Datentransport deaktiviert! Ein Beispiel zeigt Abbildung 6. Der Weg von A nach C verläuft immer über B-D, der Weg von B nach C verläuft über D, obwohl es in beiden Fällen eine direkte Verbindung A-C und B-C gibt, die für den Datentransport genutzt werden könnte. Somit ist der Summendurchsatz auf den Durchsatz der Kaskade A-B-D-C eingeschränkt, die zusätzliche Kapazität von A-C und B-C bleibt im Normalfall ungenutzt, sie wird nur im Fehlerfall aktiviert.

deltaBandwidth (N) := Obere Grenze

- d.h. max. Prozentsatz der Gesamtbandbreite, die von der Queue genutzt werden kann, die mit Verkehrsklasse N assoziiert ist
- zuzüglich freier Delta-Bandbreite höher-priorisierter Verkehrsklassen

reservedBandwidth (N) := Aktuelle Bandbreite in bps (!)

- d.h. die Bandbreite, die aktuell für die Queue reserviert ist, die mit Verkehrsklasse N assoziiert ist

Zielsetzung und Übersicht von Shortest Path Bridging (SPB)

IEEE und IETF arbeiten an einer verbesserten Layer-2 Forwarding Variante, die nicht nur einen singulären Weg je VLAN

Neue Anforderungen ändern das Netzdesign: Neue Funktionen und Standards bei IEEE und IETF - Teil 1

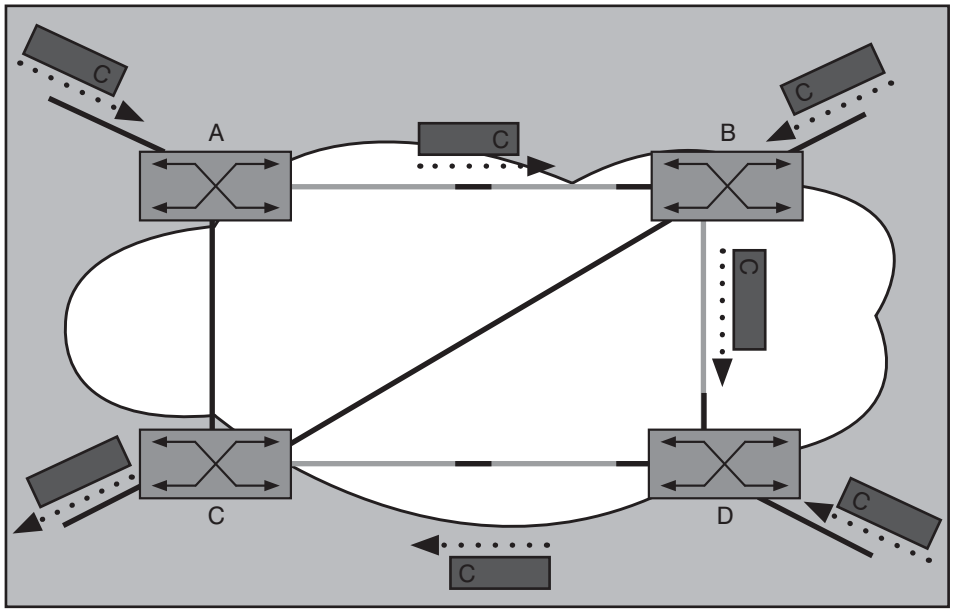


Abbildung 6: Aktive Wege in einer Spanning Tree Konfiguration

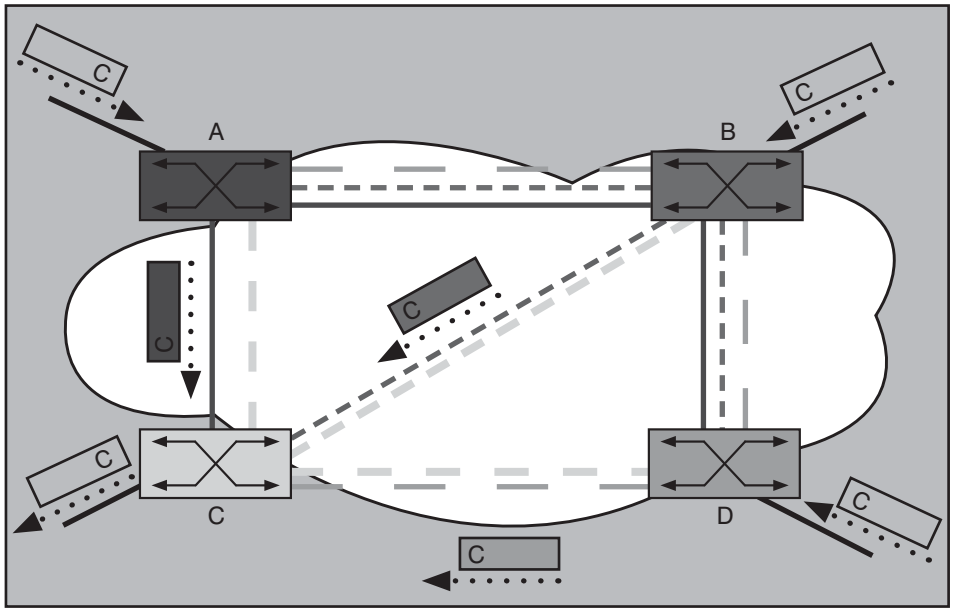


Abbildung 7: Optimierte Weiterleitung beim Shortest Path Bridging

nutzt wie im Spanning Tree, sondern alle vorhandenen Wege nutzt, soweit sie von einem bestimmten Sender zu einem bestimmten Ziel die kürzesten Wege sind. Die neuen Verfahren arbeiten ähnlich wie die Link State Verfahren OSPF und IS-IS mit einer Link- bzw. Port-gebundenen Kosten-Metrik, um den jeweils kürzesten Weg zu einem bestimmten Ziel zu berechnen. Wobei das Ziel eine MAC-Adresse und die Wurzel eines SPF-Baumes jeder Layer-2 Switch ist. So können selbst innerhalb eines VLANs oder auch wenn nur ein einzelnes VLAN betrieben wird, in einer vermaschten Switching-Infrastruktur unterschiedliche Wege zu unterschiedlichen Zielen genutzt werden, die gleichzeitig loopfrei sind und implizite Lastverteilung erreichen, wie Abbildung 7 zeigt: Frames mit MAC Destination C, die bei A ankommen, werden über den direkten Weg A-C geleitet, Frames mit MAC Destination B, die bei A ankommen, werden über den kürzesten Weg A-B und nicht A-C-B weitergeleitet. Frames mit MAC Destination D, die bei B ankommen, werden über den direkten Weg B-D geleitet und nicht B-C-D. Dies gilt auch jeweils in umgekehrter Richtung.

Insbesondere bei Ringkonfigurationen ist der vom Spanning Tree berechnete aktive Weg oft ungünstig (siehe Abbildung 8): Der Ring wird an einer Stelle für den Datentransport deaktiviert. Unterstellen wir, dass RB_A die Root Bridge ist, dann sind die Wege A-B-C-D und A-G-F-E aktiv, E-D ist an Bridge D geblockt. Falls die Verbindung A-B ausfällt, aktiviert RB_D den geblockten Port, aber das kann zu einer kompletten Neuberechnung des Spanning Tree führen. Der noch ungünstigere Fall ist der, dass RB_R die Root ist und bei Ausfall der Verbindung R-A eine neue Root berechnet werden muss. BPDU Updates an die alte und neue Root können in diesem Fall kurzfristig zu einem Denial of Service (Count to Infinity) führen, das bedeutet bei RST einige Sekunden Ausfallzeit.

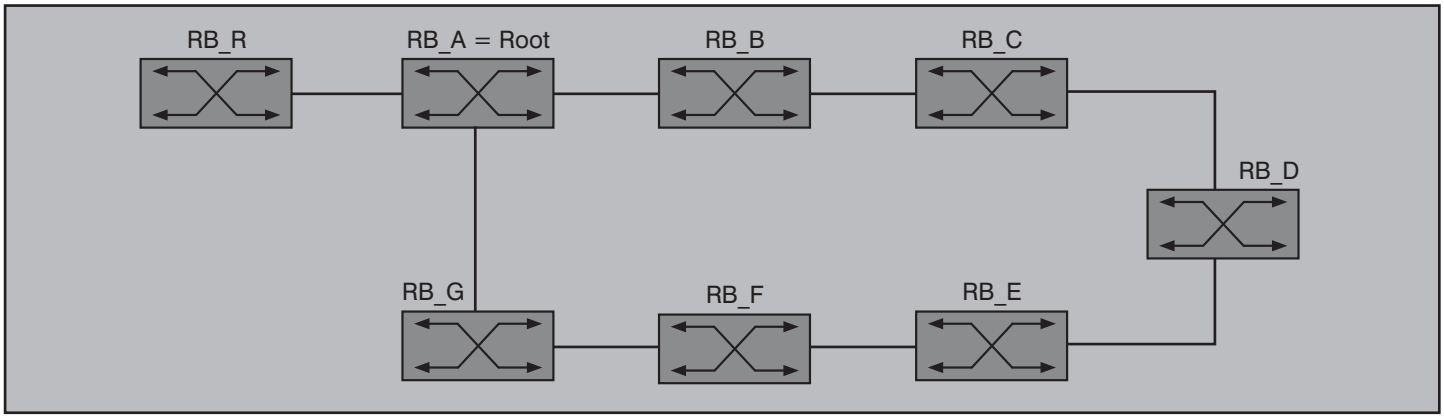


Abbildung 8: Spanning Tree in einer Ring Konfiguration

Neue Anforderungen ändern das Netzdesign: Neue Funktionen und Standards bei IEEE und IETF - Teil 1

Welcher Layer-2 Switch („Bridge“) welche MAC-Adresse angebunden hat, lernen die Switches anhand der MAC-Source-Adressen, die für ein bestimmtes VLAN auf einem Port empfangen werden, und/oder die Switches teilen sich in speziellen Informations-Paketen die angebundenen Adressen mit.

Da auf Layer-2 keine Hierarchie von IP Netzen und Subnetzen vorliegt, muss das Shortest Path Bridging (SPB) auf der Basis von Portkosten die jeweils günstigsten Wege zu bestimmten MAC-Zieladressen berechnen. Eine grundsätzliche Kenntnis von Link State Verfahren, die sich ja seit 1990 im Routing Umfeld etabliert haben, setzen wir an dieser Stelle voraus (für eine detaillierte Beschreibung verweisen wir auf den Report „Design-Varianten Lokaler Netzwerke im Vergleich“ auf der Homepage www.comconssult-research.de.)

Shortest Path Bridging bei IEEE und IETF

Für SPB gibt es leider aktuell konkurrierende Ansätze bei IEEE und IETF. Die IEEE Arbeitsgruppe 802.1Qaq definiert ein Verfahren, dass insbesondere MST, Provider Backbone Bridging und Shared VLAN Learning noch weiter unterstützt und ohne Enkapsulierung arbeitet. Die IETF definiert in der TRILL WG ein SPB Verfahren, das im Verbund mit RST, MST betrieben werden kann, aber im Grunde mit den Spanning Tree Verfahren nicht mehr viel zu tun hat. Die am IETF-SPB beteiligten Switches sind ein neuer Komponententyp und werden „RBridges“ (Routing Bridges) genannt. Da die ursprüngliche MAC-Information erhalten bleiben muss und nicht wie beim Routing Hop by Hop ausgetauscht werden kann, nutzt IETF eine Enkapsulierung zur MAC-Adressierung des Next Hop Switches auf dem Weg zum Ziel. Da die IETF-Lösung in der Fertigstellung weiter fortgeschritten ist als das IEEE-Gegenstück, befassen wir uns zuerst mit den „Rbridges“.

3.1 TRILL WG: Shortest Path Bridging mit RBridges

Die TRILL WG hat im Prinzip den Ansatz, die Arbeitsweise von gerouteten IP Subnetzen und den hierfür eingesetzten dynamsichen Routing Verfahren auf Layer-2 Netze zu übertragen. Daher leitet sich auch der neue Name „RBridge“ (Routing Bridge) für die entsprechenden Switchkomponenten ab. So wie ein Router lernt, welche IP Netze über welche Next Hop Router auf dem günstigsten Weg erreichbar sind, so lernt eine RBridge, welche MAC Adressen über welche Next Hop RBridge(s) auf dem kürzesten Weg erreichbar sind. Der günstigste Weg berechnet sich nach dem Link State Verfahren mit einer Kostenmetrik, wobei jeder Port mit Kosten in ausgehender Richtung belegt wird und die Summe der Kosten der Einzelverbindungen die Kosten für den Gesamtweg ergibt.

Unterstützung vorhandener IEEE Standards

Die TRILL Lösung berücksichtigt folgende Bridging und Spanning Tree Protokolle, mit denen sie in Kombination lauffähig sein soll:

- IEEE 802.1D STP
- IEEE 802.1D RST
- IEEE 802.1Q VLAN Prioritäten und VLAN Unterstützung
- IEEE 802.1Qv Layer-3/4 Protokollbasierte VLANs (IP und TCP/UDP Portnummer)
- IEEE 802.1Qs MST

Es ist keine Berücksichtigung vorgesehen für

- IEEE 802.1Qad Provider Bridges
- IEEE 802.1Qah Provider Backbone Bridges

Selbstverständlich beachtet die TRILL Lösung, dass der normale MAC-Dienst für Unicast, Multicast und Broadcast Frames erhalten bleibt. Gleiches gilt für die Beibehaltung der Frame-Reihenfolge im Normalbetrieb (Ausnahme sind Fehlerumschaltungen). Loopfreiheit ist ebenfalls im Normalbetrieb gegeben, für Schaltsituationen wird eine so genannte Loop-Minimierung (Loop Mitigation) durch Hop Count Limits erreicht - so löst das die IEEE in ihrem Ansatz auch.

Stabile Optimierung für Multicast-Transport

Obwohl es eigentlich eine Layer-2-Verletzung ist, gehört es zu den etablierten Funktionen von Layer-2 Switches, IP Multicast Kontrollpakete mitzulesen (IGMP Snooping). Jeder Switch optimiert dadurch das Fluten von Multicasts an jedem seiner Ports. Ergibt sich eine STP Neuberechnung, so muss diese Optimierung an allen Ports erneut durchgeführt werden. Während der STP Neuberechnung kann kurzfristig ein Denial of Service für Multicasts entstehen, nach der Neuberechnung wird kurzfristig komplett geflutet, bis die Optimierung mit IGMP Snooping wieder hergestellt ist. Würde hier für die Berechnung so genannter Multicast Trees ein Link State Verfahren genutzt, könnte eine globale Sicht des kompletten Layer-2 Netzes mit allen Multicast Group Memberships und allen vorhandenen Multicast Routern in der Link State Datenbank gehalten werden, die nach einem LAN Topologiewechsel eine sofortige Anpassung der Multicast-Weiterleitung ermöglichen würde.

Zielsetzung des TRILL Protokolls

Für RBridges bzw. das TRILL Protokoll ergibt sich folgende Zielsetzung: RBridges leisten paarweise optimierte Frame-

Weiterleitung, kombiniert mit Autokonfiguration (Zero Configuration), sichere Weiter-leitung auch in Schaltsituationen sowie Unterstützung von Lastverteilung für Unicast und Multicast Verkehr. Um diese Zielsetzung zu erreichen, nutzen RBridges IS-IS Routing und eine Enkapsulierung, deren Header einen Hop Count enthält, dessen Funktion in der TRILL-Übersicht detaillierter beschrieben wird.

Layer-3/IP ist zwar für Layer-2 Switches eigentlich transparent, die Unterstützung von sowohl IPv4 als aus IPv6 Routern und Endgeräten wird jedoch explizit als Zielsetzung aufgeführt.

Übersicht über das TRILL Protokoll

Als generischer Ansatz beschreibt das TRILL Protokoll das Verhalten von RBridges in einem Netz, in dem es auch noch Teilnetze mit normalen Spanning Tree Switches gibt. TRILL bezeichnet diese Teilnetze salopp als „Ethernet Wolke“, obwohl es auch WLAN Netze sein könnten (siehe Abbildung 3.4). Zugegeben, da ein WLAN Access Point meistens weder Spanning Tree noch gar Rapid Spanning Tree unterstützt, scheint es nicht sehr wahrscheinlich, dass wir das TRILL Protokoll bald in WLAN Access Points sehen.

Das TRILL Protokoll unterteilt Layer-2 Frames hinsichtlich der Weiterleitung in drei Kategorien: Erstens Kontroll-Frames mit einer Multicast-Adresse 01-80-C2-00-00-00 bis 80-C2-00-00-0F oder 80-C2-00-00-21. High Level Kontroll-Frames sind BPDUs (80-C2-00-00-00) und VRP Frames (MVRP, 80-C2-00-00-21). Low Level Kontroll-Frames sind alle anderen Kontroll-Frames. Zweitens TRILL Frames; diese haben entweder den TRILL Ethertype Wert im Ethernet Typfeld eingetragen oder den TRILL-alloquierten MAC Multicast (dieser muss noch von der IEEE Registration Authority festgelegt werden). Drittens „native“ Frames, das sind alle anderen Frames außer Kontroll- und TRILL-Frames.

RBridges fahren untereinander das IS-IS Link State Protokoll, um paarweise (zwischen jeweils zwei RBridges) den günstigsten Weg für Unicast sowie Baumstrukturen (Distribution Trees) für die Weiterleitung von Multicasts, Broadcasts und unbekannten Zieladressen zu berechnen. Die erste RBridge (RB_1) in einer Ethernet Wolke, die ein Unicast Frame erhält, enkapsuliert dieses in einen TRILL Header, der die Ziel-RBridge (RB_2, letzte RBridge vor der Ethernet-Wolke) enthält, die das Frame wieder dakapsulieren muss. RB_1 heißt Ingress RBridge, RB_2 Egress RBridge. Um im TRILL Header Platz zu sparen, verwendet der TRILL Header die aus IS-IS bekannten 2-Byte Nicknames anstelle der 6-Byte IS-IS System ID (6-Byte MAC Adresse).

Um das Loopverhalten bei Schaltsituationen zu minimieren, enthält der TRILL-Header einen Hop Count. Der Hop Count ist ein 6-Bit integer Wert. Jede RBridge muss vor der Weiterleitung eines Frames dieses Feld prüfen und das Frame vernichten, wenn der Wert 0 ist. Ist der Wert 1 oder höher, muss er im weitergeleiteten Frame um 1 erniedrigt sein. Außerdem tragen RBridges bei der Weiterleitung von Unicast Frames auf ein Shared Medium (z.B. LWAN!) die Next Hop RBridge als Destination in den Header ein, um unnötiges Weiterleiten bei einem temporären Loop zu vermeiden. Für Multi-Destination Ziele (MC, BC, unbekannte MAC) führen sie unter anderem eine Reverst Path Forwarding Prüfung durch, um potenzielle Loops zu kontrollieren.

Für Multi-Destination Frames (MC, BC, unbekannte MAC) wird Lastverteilung (Multipathing) untersützt, sofern es mehrere Distribution Tree Roots gibt, für Unicast Frames wird Multipathing mit ECMP (mehrere kostengleiche Wege) unterstützt.

Mit dem IS-IS Protokoll wird für jedes Layer-2 LAN (im wesentlichen ein Link) eine Designierte RBridge (DRB) ausgewählt. Wie der IS-IS Router, so kann auch die DRB dem Link einen Pseudoknoten-Namen geben, versendet CSNPs (Complete Sequence Number PDU) auf dem Link und legt fest, welches VLAN das Designierte VLAN für die Kommunikation von RBridges auf diesem Link ist.

Entweder enkapsuliert/dekapsuliert die DRB sämtlichen Datenverkehr zu oder von diesem Link oder delegiert dies im Fall von Lastverteilung für ein oder mehrere VLANs an eine andere RBridge. Diejenige RBridge, die für den Verkehr eines VLANs zuständig ist, heißt „appointed VLAN-x Forwarder).

Für Management sollen RBridges SNMPv3 unterstützen, eine separate MIB wird die TRILL WG als separates Dokument erarbeiten.

Wie werden die Endgeräte-Adressen gelernt?

Eine RBridge (z.B. RB_1), die VLAN-x Forwarder ist, muss sowohl auf ihren direkten Verbindungen ins VLAN-x als auch auf allen anderen Backbone-Verbindungen lernen, wo (d.h. an welchem Port) welches VLAN-x-Endgerät zu finden ist. Dies geschieht wie bei normalen Layer-2 Switches auch: die RBridge schaut nach, welche MAC Source-Adressen in den MAC Frames eingetragen sind, die zu VLAN-x gehören. Im WLAN werden die Adressen über 802.11 Assoziierung und Authentifizierung gelernt. Zum Verständnis, wie eine RBridge MAC-Adressen von Endgeräten lernt, die sie nicht direkt angebunden hat, betrachen wir die grobe Struktur des TRILL Headers, wie in Abbildung 3.5 gezeigt. Die VLAN-x Endgeräte, die über entfernte RBridges (z.B. RB_2) angebunden sind, lernt RB_1, indem sie im TRILL Header den Nickname von RB_2 sieht und diesen mit dem VLAN und der MAC-Source im inneren Ethernet Header (originäres Ethernet Frame) verbindet: Steht dort z.B. das VLAN 0815 und die MAC Source 04-35-67-11-22-33, so weiß RB_1 jetzt, dass das Endgerät 04-35-67-11-22-33 im VLAN 0815 über die RBridge RB_2 erreichbar ist.

Zusätzlich hat eine RBridge noch eine weitere Möglichkeit, MAC Adressen zu lernen und bekanntzugeben: sie kann das ESADI Protokoll (End Station Address Distribution Information) nutzen. Adressen, die über explizite ESADI Frames gelernt wurden, könnten auch als authentischer eingestuft werden als Adressen, die über native Daten-Frames oder TRILL Frames gelernt wurden. Zudem unterstützt ESADI die verschlüsselte Authentifizierung von Nachrichten (gem. RFC5304). Wenn eine RBridge ESADI nutzt, bedeutet das jedoch nicht, dass sie die Informationen der TRILL Frames ignorieren darf – diese muss sie wei-

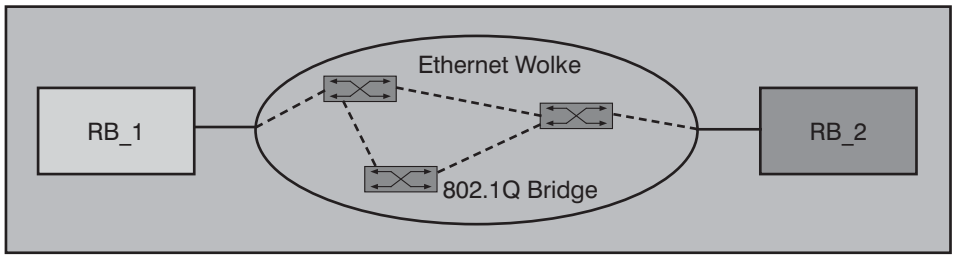


Abbildung 9: Mischkonfiguration mit RBridges und STP Bridges

terhin lernen, so lange das Lernen in der RBridge nicht durch administrative Konfiguration deaktiviert wird. Die Unterstützung von ESADI ist im TRILL Protokoll keine MUSS-Forderung sondern optional.

Die Enkapsulierungs-Architektur von RBridges

Zur Handhabung der Routing Informationen und zur Loop-Vermeidung packen RBridges die ursprünglichen Ethernet Frames in einen zusätzlichen TRILL Header ein. Damit das ursprüngliche Frame ordentlich über Ethernet weitergeleitet werden kann, führt dies zu folgendem Frame Format: Den Anfang bildet ein äußerer Ethernet Header, gefolgt vom TRILL Header, anschließend der innere Ethernet Header (originäres Ethernet Frame) inklusive VLAN Tag, danach die Nutzdaten (Payload) und am Ende der Trailer mit der neuen FCS. Sobald eine RBridge eine Enkapsulierung durchführt, muss sie natürlich eine neue FCS berechnen. Die FCS des inneren Ethernet Frames bleibt dabei nicht erhalten, da der Trailer nur einmal und nicht doppelt angelegt wird. Da die Relay-Funktion oberhalb von Layer-2 angesiedelt ist, können auch andere Layer-2 Verfahren als Ethernet genutzt werden. In diesem Fall wird der äußere Ethernet Header durch die jeweilige Link Schicht ersetzt, z.B. PPP, wie in Abbildung 3.6 gezeigt ist. Der aktuelle Draft hat jedoch TRILL mit Ethernet Enkapsulierung im Fokus.

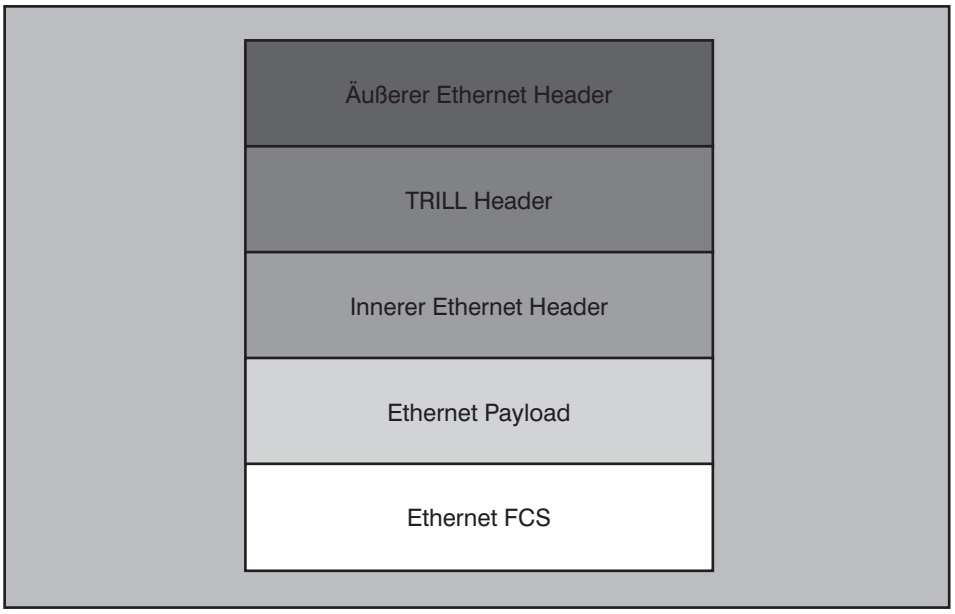


Abbildung 10: TRILL Header für Ethernet Verbindungen

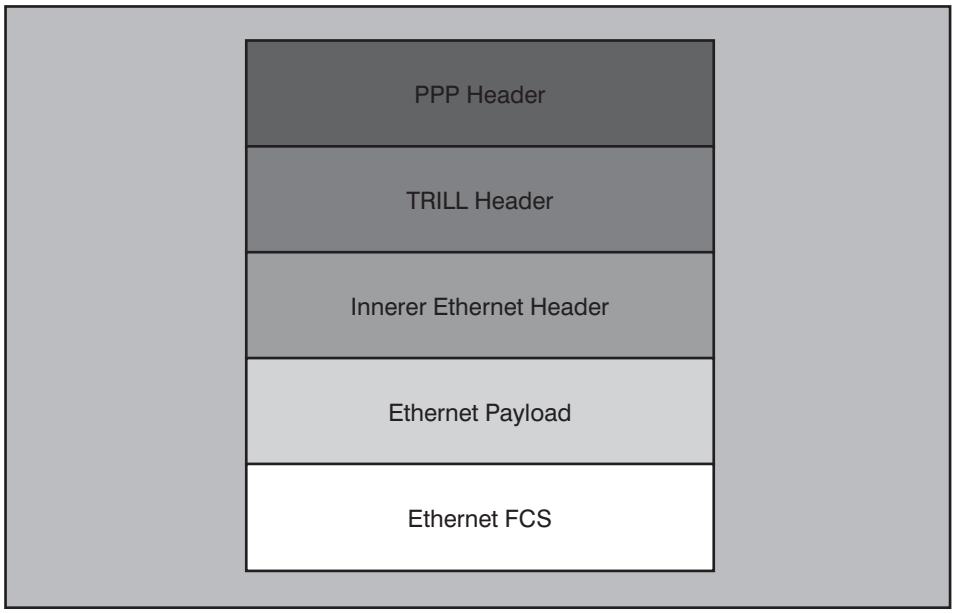


Abbildung 11: TRILL Header für PPP Verbindungen

Die Nutzung des TRILL Headers bietet einige Vorteile: Erstens wird das Loop Problem durch ein Hop Count Feld gelöst. Zweitens müssen Transit RBridges, die nur Backbone-Verbindungen haben und keine VLAN-x Forwarder sind, keine MAC Adressen von Endgeräten mehr lernen. Drittens verkleinert die Weiterleitung der Frames mit Zieladresse „Egress RBridge“ im äußeren Ethernet Header die Größe der Adresstabelle von Transit RBridges von der Anzahl Endgeräte auf die Anzahl RBridges im Netz. Viertens wird der Verkehr zwischen RBridges mittels eigenem VLAN Tag über ein VLAN gesendet, dass von den VLANs für den Nutzdatenverkehr völlig unabhängig ist.

Die Weiterleitung von Frames über RBridges

Für Frames, deren innere MAC Zieladresse eine bekannte/gelernte Unicast-Adresse ist, kennt die Ingress RBridge die Egress RBridge. Diese Frames werden wie beim Routing Hop by Hop zur Egress RBridge weitergeleitet, wobei jeweils im äußeren Ethernet Header die MAC Source der aktuell sendenden und die MAC Destination der Next Hop RBridge stehen. Für Multi-Destination Frames werden Distribution Trees gebildet. Diese werden jedoch nach unbekannten MAC-Adressen/Broadcasts und speziellen Multicast-Gruppen unterschieden. Verschiedene Multicast-Gruppen sind IGMP Membership Reports (RFC 3376), IGMP und MLD Queries (RFC 2710), MRD (RFC 4286) Announcement Nachrichten und andere von IP Multicasts abgeleiteten MAC-Multicasts (01-00-5E-XX-XX-XX). Für jeden Distribution Tree erfolgt ein eigenes Pruning (Bereinigen des Flutens), um unnötiges Fluten zu minimieren.

Für unbekannte Unicast Frames muss die Ingress RBridge den Hop Count mindestens auf die Anzahl der (erwarteten) RBridge Hops bis zur Egress RBridge setzen. Die Ingress RBridge sollte jedoch einen höheren Wert nehmen, um alternative Wege mit höherem Hop Count ebenfalls zu ermöglichen, die sich an einem weiter auf dem Pfad liegenden Punkt ergeben könnten. Für Multi-Destination Frames muss der Hop Count mindestens auf die Anzahl Hops zu der am weitesten entfernten RBridge gesetzt werden,

denn man weiß ja nie … ob sich nicht dort eine Broadcast-Empfänger oder ein Multicast Listener auftun wird. Um diese Hopzahl zu bestimmen, berechnet die RBridge die maximale Hopzahl in jedem Distribution Tree.

RBridges, VLANs und andere 802.1 Layer-2 Protokolle

Per Default, d.h. für ungetagte Frames bestimmt der Switchport, auf dem ein Frame empfangen wird, das VLAN, zu dem es gehört. Dieses ist im Regelfall das jeweilige Port-VLAN, dem der Switchport zugeordnet ist (in Ausnahmefällen ein dynamisches VLAN, das anhand der IP-Adresse oder der TCP/UDP Portnummer zugeordnet wird). Wurde keine explizite Zuordnung konfiguriert, gehört das Frame automatisch in das Default VLAN mit der Nummer 1. Für Backbone Verbindungen legt IEEE 802.1Q fest, dass mittels Tagging Frames aus mehreren VLANs über denselben Link weitergeleitet werden können. TRILL nutzt dasselbe VLAN Tag Format wie IEEE 802.1Q. Ein VLAN Tag kann im äußeren und/oder inneren Ethernet Header vorhanden sein. Im äußeren Header bezeichnet der VLAN Tag das VLAN, das zwischen zwei RBridges gilt, im inneren Header das VLAN des ursprünglichen Ethernet Frames. RBridges berücksichtigen den VLAN Tag des inneren Ethernet Headers, indem sie dieses Frame nur auf solche Verbindungen weiterleiten, die diesem VLAN zugeordnet sind bzw. auf denen die Weiterleitung getagter Frames allgemein erlaubt ist.

Einige Switches unterstützen so genanntes VLAN Mapping, bei dem ein eingehendes VLAN ausgehend auf ein anderes VLAN (andere Nummer) gemappt wird. Dies kommt insbesondere bei Provider Bridges zum Einsatz. TRILL berücksichtigt VLAN Mapping jedoch nicht konkret sondern erwähnt es lediglich als optionale Funktion.

Zusätzlich zu IEEE 802.1Q unterstützen RBridges 802.1X Port Security und 802.3ad Link Aggregierung als Layer-2 Dienste. RBridges nutzen jedoch keinen Spanning Tree und blocken keine Ports nach Art des Spanning Tree. Damit ergibt sich ein Port-Modell für RBridges und die Bearbeitung von Protokollen, wie es Abbildung 12 zeigt: Die obere Schnittstelle der Port/Link Control Logik korrespondiert mit der ISS (Internal Sublayer Service) Schnittstelle von IEEE 802.1Q. BPDUs und (M)VRP Frames werden über diese Schnittstelle bearbeitet. Die obere Schnittstelle der Port VLAN Bearbeitung korrespondiert mit der EISS (Extended Internal Sublayer Service) Schnittstelle von IEEE 802.1Q und handhabt die Bearbeitung von VLAN- und Priorisierungs-Funktionen. Die Relay-Funktion für native Datenframes und TRILL Frames liegt oberhalb von EISS, diese Frames werden in RBridges oberhalb der EISS Schnittstelle bearbeitet.

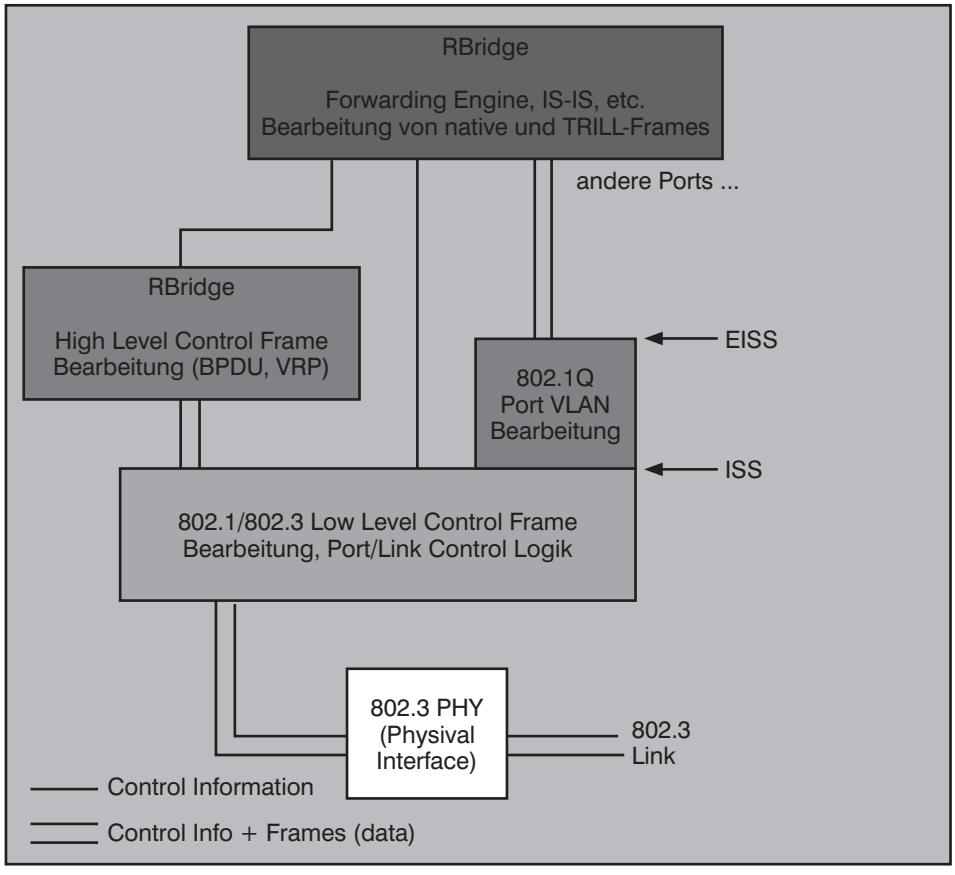


Abbildung 12: RBridge Port-Modell

Die Sache mit den Nicknames

Der Nickname ist 2 Byte lang. Dies ermöglicht 2^{16} RBridges in einem Layer-2 Verbund, was wohl bei weitem ausreichen sein dürfte. Um die aktiven Nicknames verwenden zu können, fahren die RBridges ein Nickname Aquisition Protokoll, mit dem sie die standortweit eindeutigen Nicknames aller RBridges lernen. Für dieses Protokoll gelten folgende Rahmenbedingungen:

- Das Nickname Aquisition Protokoll läuft quasi als Piggyback zusammen mit dem Link State Protokoll IS-IS, da der Nickname zusammen mit einem Prioritätswert zur Nutzung / Gültigkeit in den IS-IS TLVs steht. Die Nickname Aquisition erzeugt also keinen erhöhten Overhead. Jede RBridge wählt ihren eigenen Nickname.

- Der Nickname ist administrativ konfigurierbar. In diesem Fall hat er einen höheren Prioritätswert als alle nicht-konfigurierten Nicknames.

- Die Werte 0x0000 und 0xFFC0 bis 0xFFFF sind reserviert und dürfen weder automatisch ausgewählt noch gar konfiguriert werden. Der Wert 0x0000 steht dafür, dass der Nickname einer RBridge unbekannt ist, somit wäre es fatal, diesen zu konfigurieren.

- Der Prioritäts-Wert ist ein 8-Bit Wert, wobei das Most Significant Bit (0x80) anzeigt, dass der Name konfiguriert wurde. Die unteren 7 Bits haben den Default 0x40. Hat eine RBridge den Nicknamen für eine längere Zeitspanne gehalten, so darf sie den Prioritätswert erhöhen (sofern er nich schon auf 0x7F steht, denn dann ist der nächsthöhere Wert 0x80, der nur für konfigurierte Nicknames verwendbar ist).

- Hat eine RBridge einmal einen Nicknamen erfolgreich erobert, so sollte sie nach einem Reboot versuchen, diesen weiter zu nutzen.

- Jede RBridge ist dafür verantwortlich, die Eindeutigkeit ihres Nickname si-

cherzustellen. Nehmen wir an, RB_1 wählt Nickname X und entdeckt dann bei Erhalt eines LSP von RB_2, dass diese denselben Namen gewählt hat. Dann darf die RBridge mit dem höheren Prioritätswert den Namen behalten, bei gleicher Priorität gewinnt die RBridge mit der höheren IS-IS ID (48 Bit). Die RBridge, die das Rennen verloren hat, muss sich einen neuen Nickname suchen. Das kann auch einer RBridge mit einem konfigurierten Nickname passieren, wenn der Name der anderen RBridge ebenfalls konfiguriert ist.

- Um Namenskollisionen zu minimieren, berechnet eine RBridge üblicherweise für einen neuen Nickname einen Hash-Wert aus einigen ihrer eigenen Parameter wie Port-MAC-Adresse, Zeit, Datum und weitere Parameter, die in RFC 4086 beschrieben sind. Es ist übrigens nicht notwendig, dass alle RBridges ihre Nicknames mit demselben Algorithmus berechnen. Dies gibt unterschiedlichen Herstellern hier mehr Freiheitsgrade.

- Wenn zwei RBridge-Verbundnetze zusammengeschaltet werden, entsteht eine gewisse Wahrscheinlichkeit für Nickname Kollisionen, die auch dazu führen können, dass eine RBriddge sich mehrmals einen neuen Nickname suchen muss.

- Um die Wahrscheinlichkeit einer Namenskollision zu minimieren, sollte eine RBridge, die keinen konfigurierten Namen hat und sich einen neuen Zufalls-Namen suchen muss, warten, bis sie die Link State Datenbank eines ihrer Nachbarn erhalten hat (in der die bereits genutzten Namen stehen!), bevor sie ihren neuen Namen bekanntgibt.

Eine RBridge, die kein Ingress, Egress oder Root eines Distribution Tree ist, benötigt keinen Nickname.

Schrittweiser Einisatz von RBridges

Da RBridges grundsätzlich kompatibel zu Spanning Tree Bridges sein sollen, können Netze schrittweise von 802.1D Switches auf RBridges umgestellt werden. 802.1D Switches sind aus Sicht der RBridges transparent, sie bilden zusammen mit den physikalischen Links zu den RBridges so genannte Bridged LANs. Solche Bridged LANs sind aus Sicht der RBridges Multi-Access Links. Allerdings räumt der RBridge Draft an dieser Stelle ein, dass das Standort-LAN am besten funktionieren wird, wenn die Spanning Tree Bridges komplett auf RBridges migriert wurden.

Abkürzungen Teil 1

ANSI	American National Standards Institute
ATM	Asynchroner Transfer Modus
AVB	Audio / Video Bridging
BC	Broadcast
BPDU	Bridge PDU
CSNP	Complete Sequence Number PDU
DA	Destination Address
DCB	Data Center Bridging
DRB	Designated RBridge
DS	Differentiated Services
DS	Distribution System
DSCP	Differentiated Services Code Point
ECMP	Equal Cost Multipath
EISS	Extended Internal Sublayer Service
ESADI	End Station Address Distribution Information
FCS	Frame Check Sequence
GARP	Generic Attribute Registration Protocol
GVRP	GARP VLAN Registration Protocol
HPC	High Performance Computing
HPCC	High Performance Cluster Computing
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IGMP	Internet Group Management Protocol
IM	Instant Messaging
IP	Internet Protocol
IPC	Inter Process Communication
IS-IS	Intermediate System to Intermediate System
ISS	Internal Sublayer Service
LAN	Local Area Network
LB	Load Balancing
LLC	Logical Link Control
LSP	Link State PDU
MAC	Media Access Control
MAD	MRP Attribute Declaration
MAP	MRP Attribute Propagation
MC	Multicast
MDI	Media Dependent Interface
MII	Media Independent Interface
MLD	Multicast Listener Discovery
MMRP	Multiple MAC Adress Registration
MRD	Multicast Router Discovery
MRP	Multiple Registration Protocol
MSRP	Multiple Stream Reservation Protocol
MVRP	Multiple VLAN Registration Protocol
OSPF	Open Shortest Path First
PDU	Protocol Data Unit
PPP	Point-to-Point Protocol
QoS	Quality of Service
Rbridge	Routing Bridge
RFC	Request for Comments
RST	Rapid Spanning Tree
RSTP	Rapid Spanning Tree Protocol
RTCP	Real Time Control Protocol
RTP	Real Time Protocol
SA	Source Address
SNMP	Simple Network Management Protocol
SPB	Shortest Path Bridging
SPF	Shortest Path First
SR	Stream Reservation
SRP	Stream Reservation Protocol
TAI	Temps Atomique International (international atomic time)
TCP	Transmission Control Protocol
TG	Task Group
TK	Telekommunikation
TLV	Type, Length, Value
TRILL	TRansparent Interconnection of Lots of Links
TTL	Time to Live
UC	Unified Communications
UTC	Coordinated Universal Time
VLAN	Virtual Local Area Network
VoIP	Voice over IP
VRP	VLAN Registration Protocol
WAN	Wide Area Network
WG	Working Group
WLAN	Wireless LAN

Links

http://www.ieee802.org/1/
http://www.ieee802.org/3/
http://www.ieee802.org/11/
http://www.ietf.org

Literatur

Timing and Synchronization for Time-Sensitive Applications in Brdiged Local Area Networks, IEEE Standard Drafts; 2008

MSRP Attribute Declaration, Propagation and Bandwidth Allocation, IEEE 802.1 Intermin – Seoul, Craig Gunther; 2008

Transparent Interconnection of Lots of Links: Problem and Applicability Statement, Network Working Group, J. Touch & R. Perlman; 2009

Rbridges: Base Protocol Specification <draft-ietf-trill-rbridge-protocol-12.txt> TRILL Working Group, R. Perlman; 2009

Virtual Bridged Local Area Networks - Amendment 9:, Shortest Path Bridging, Interworking Task Group of IEEE 802.1; 2008

Schwerpunktthema

Der Kampf ums RZ: die nächste Runde -

Teil 2: Matrix, FCoE und EoFC

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.

In diesem Artikel schauen wir uns zunächst die Kostenfalle an. Dann kommen wir zu Aspekten der Vernetzung des ersten verfügbaren für die Virtualisierung konzeptionierten Gesamtkonzeptes: der Blade System Matrix von HP.

1. Die Kostenfalle

Allgemein gibt es folgende Herausforderungen an ein Rechenzentrum in einem Unternehmen:

- Stromverbrauch, Kühlung und Dichte
- Integriertes und Technologie-übergreifendes Management
- Cost of Ownership, ausgedrückt in Kapital- und Betriebskosten

Es zeigt sich, dass die jährlichen Kosten bei einer 10-jährigen Nutzungsdauer auf das 2,3-fache des Systempreises steigen können.

Das ist an und für sich schon sehr dramatisch. Es gibt aber tatsächlich eine Entwicklung, die von der Kostenseite her noch dramatischer ist: die Server Management und -Verwaltungskosten.

Nach einer Untersuchung von IDC, diesmal bezogen auf die weltweit installierte Server-Basis, zeigt ein erschreckendes Bild: in jedem Jahr muss bezogen auf die Kosten für die Neuanschaffungen bei Servern relativ immer mehr Geld für deren Management ausgegeben werden, auf die nächsten Jahre hochgerechnet bis zum 3,5-fachen! (Abbildung 2)

Im Klartext: selbst massiv weiter steigende Stromkosten vernichten nicht soviel Geld wie mangelhafte Systeme für das System-Management!

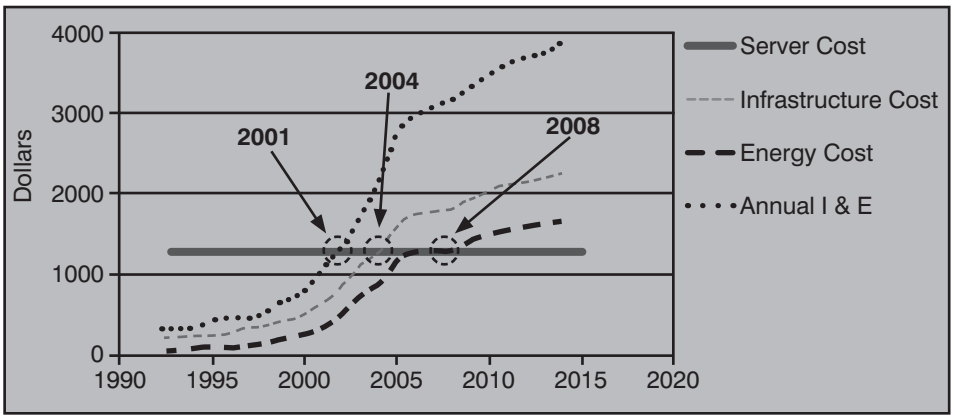


Abbildung 1: Jährliche amortisierte Kosten eines vollkonfigurierten 1U-Servers in einem missionskritischen RZ

Das Ganze ist noch unter Berücksichtigung der herkömmlichen Struktur hochgerechnet. Bei dieser sind hauptsächlich zu verwalten:

- Server
- Betriebssysteme
- Anwendungen
- Speichersysteme
- Netzkomponenten

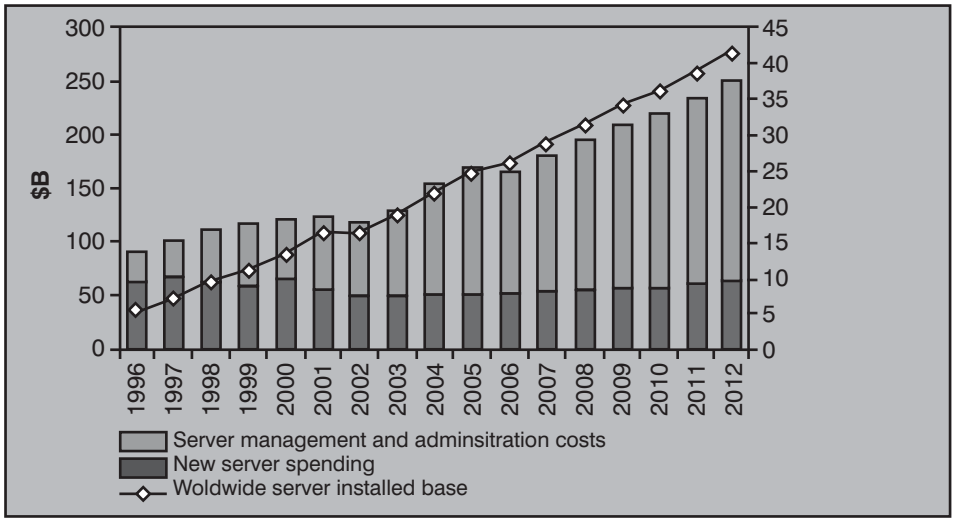


Abbildung 2: Kosten von Server Hardware in Relation zu Gesamt-Management-Kosten Quelle: IDC

Der Kampf ums RZ: die nächste Runde - Teil 2: Matrix, FCoE und EoFC

Im Rahmen von Virtualisierung kommt noch einiges hinzu:

- Virtualisierungssoftware
- Virtuelle Switches
- Virtuelle Speicher

Die Virtualisierung ist eines der Kerninstrumente zur Senkung der Betriebskosten. Wichtige Aktivitäten wie Backup/Restore werden im Aufwand massiv gesenkt und die Gesamtzahl der Komponenten sinkt. Allerdings senkt Virtualisierung die Betriebskosten nur dann wirkungsvoll, wenn entsprechende Betriebs-Instrumente bereitstehen.

In virtualisierten Systemumgebungen muss ein Management-System existieren, welches alle wichtigen Bereiche einheitlich abdeckt. Ist dies nicht gegeben, könnten die durch Virtualisierung und I/O-Konsolidierung möglicherweise einzusparenden Kosten durch stark wachsende Betriebskosten aufgezehrt werden.

Was bedeutet das für unser eigentliches Kernthema, die Netze? Viel wichtiger als die Diskussion einzelner Protokollmerkmale ist es, sich anzuschauen, wie weit ein Hersteller von Netzkomponenten diese in ein Gesamtsystem zur Systemverwaltung einbinden kann.

2. HP Blade System Matrix

Cisco Systems hat als erster Hersteller mit UCS den Gedanken einer Integrierten Lösung vorgestellt. Mit der Blade System Matrix kann jedoch HP als erster Hersteller überhaupt ein Konzept ausliefern, in dem alle wichtigen Dimensionen, nämlich Server, Speicher, Netzwerk, Virtualisierung und Verwaltung einheitlich und zusammenhängend betrachtet werden. Die Analyse der Eigenschaften wird an dieser Stelle auf die Integration der Netzwerk-Technologie reduziert, weitere Eigenschaften des Produktes sind außerhalb der Zielsetzung dieses Artikels.

Cisco und HP stehen mit dieser Vorgehensweise nicht allein. Auch IBM folgt dieser Linie auf der Basis einer Kooperation mit Brocade und kann so ebenfalls eine Technologie-integrierte Gesamt-Lösung anbieten.

HP gibt an, u.a. folgende Ziele in einem zukünftigen RZ besonders unterstützen zu wollen:

- Discovery Funktionen und Zustandsinformationen. Administratoren müssen Informationen über ihre infrastrukturellen Elemente und deren Zustand und die Beziehungen zwischen ihnen sammeln können, bevor sie die Struktur intelligent steuern können. Viele dieser Informationen sind in grundsätzlichen Management-Möglichkeiten vorhanden, doch zusätzliche Informationen, besonders bezüglich der Interaktionen, sind einzigartig für die Kontrolle dieser dynamischen Umgebungen.

- Verwaltung virtueller und physischer Betriebsmittel. Historisch gesehen haben sich physikalische und virtuelle Betriebsmittel unterschiedlich entwickelt und stellen verschiedene Sichten auf die Elemente eines RZs dar. Die Konvergenz des Verhaltens und der Verwaltung virtueller und physischer Betriebsmittel ist kritisch für die Implementierung einer dynamischen RZ-Infrastruktur. In einer idealen Welt würden virtuelle und physische Server, Speicherpools und Netzwerk-Betriebsmittel als äquivalente Elemente mit vergleichbarem Verhalten angesehen. Administratoren sollten dafür spezielle Tools besitzen, die es ihnen ermöglichen, entsprechende Elemente gleichartig zu betrachten und in spezifische Lösungen zu überführen.

- Isolation und Kapselung. Dies sind fundamentale Prinzipien der Software-Entwicklung. Eine ideale Systemarchitektur sollte in der Lage sein, ausgewählte Regionen der Infrastruktur zu kapseln, so dass Änderungen auch isoliert und gekapselt bleiben und somit keine Auswirkungen auf den Rest der Infrastruktur haben. Z.B. werden heute Server, Speicher und Netze oftmals in verschiedenen administrativen Rollen behandelt. Alle diese Elemente interagieren aber in einem RZ, was die Komplexität des Managements erhöht. Betrachtet man die Infrastruktur als eine Sammlung isolierter Regionen mit spezifizierten Interaktionen, kann man die Komplexität der Verwaltung und damit deren Kosten erheblich senken.

- Automation. Nach Ansicht von HP ist Automation ein heute oftmals überstrapazierter Begriff. Der Hersteller sieht die Rolle der Automation vor allem darin, ein Kontinuum von Techniken zu schaffen, die die Reaktionszeit auf bestimmte geplante oder außerplanmäßige Ereignisse reduziert.

- Unverwüstlichkeit und Verfügbarkeit. Hier stehen vor allem Fail-over Clustering und Site Disaster Recovery im Vordergrund. Der Übergang zu physikalisch verteilteren und skalierenden Architekturen im Zusammenhang mit der Virtualisierung schafft hier neue und elegante Möglichkeiten, die von der Umleitung einer Anwendung von einer fehlerhaften Umgebung auf eine andere benachbarte virtuelle Maschine bis hin zur vollständigen Replikation über ein via WAN abgesetztes Ersatz-RZ reicht.

Es gibt zu diesen Themen eine so genannte Adaptive Infrastruktur Vision, siehe www.hp.com/go/ai. (siehe Abbildung 3)

Der letzte Absatz ist ein Zitat der Absichten des Herstellers. Die Fähigkeit zur Discovery und die Anordnung von Management-Funktionen in einer Hierarchie ist nicht neu. Auch das, was als Isolation und Kapselung bezeichnet wird, ist das seit 20 Jahren hinlänglich bekannte Konzept des Element-Managers. Das Grundproblem der immer schon bestehenden Management-Hierarchie ist nicht der Element-Manager, der in jedem Fall am Bedarf der Spezialisten der jeweiligen Technologie ausgerichtet ist. Er ist unverzichtbar, weil er ins tiefste Detail gehen muss. Das Grundproblem ist, was man auf der obersten Hierarchiestufe sieht und wie aussagekräftig das ist. Hier stellt sich vor allem die Frage nach dem Umgang mit Informationen des Element-Managers wenn der Spezialist nicht im Hause ist. In der Vergangenheit haben uns alle Hersteller hier

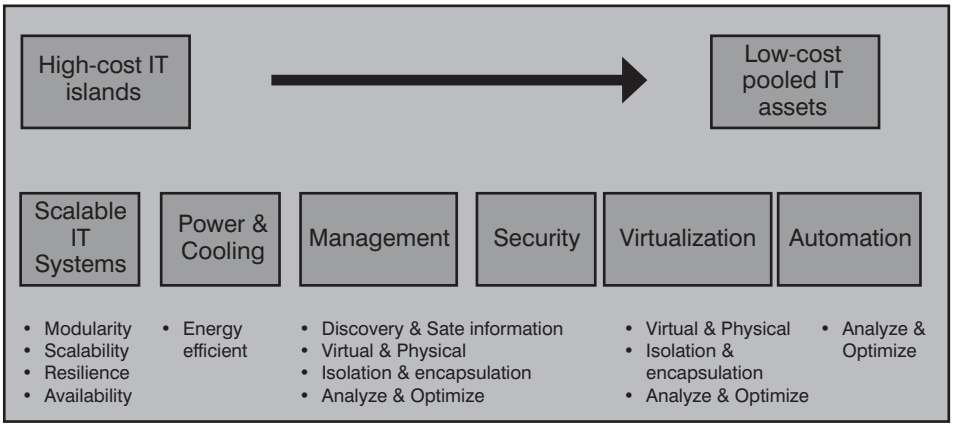


Abbildung 3: Kernelemente der adaptiven Infrastrukturvision von HP Quelle: HP

eher enttäuscht. Ob sich das geändert hat, lässt sich nur in Projekten klären und nicht hier auf dem Papier.

Die Blade System Matrix bezieht sich (zunächst?) auf Systeme der so genannten c-Class. Ein BladeSystem der c-Class besteht aus folgenden Elementen:

- Enclosure, das kann man vornehm formulieren, aber das ist der Kasten, der Blade Systeme aufnimmt. Er hat eine passive Mittelebene für die Datensignale, eine passive Backplane für den Strom, eine verteilte Stromversorgungs- und Kühlsystemarchitektur und Module für das lokale Management

- Server Blades in verschiedenen Ausführungen mit AMD oder Intel x86-Prozessoren und Storage Blades. Sie unterstützen die HP Non-Stop-Architektur

- Interconnects: das ist eine große Auswahl von Modulen für die Interaktion der c-Class Blades mit externen und internen Speicher- und Netzwerksystemen. Das umfasst auch Ethernet- und Fibre Channel Switches, die entweder standardmäßig oder mit HP Virtual Connect verwaltet werden können sowie NICs für die Blades mit standardisierten oder spezialisierten Funktionen

- Virtual Connect: das ist eine HP-Technologie für die Virtualisierung, Isolation und Kapselung vielfacher Aspekte von Servern, Speichern und Netzen

Es gibt nun unterschiedliche Ausführungen der Enclosures und Blades und sozusagen unendlich viele mögliche Kombinationen. Für den Netzwerker ist aber interessant, dass z.B. das größte Modell Nr. 7000 auf der Mittelebene für die Datensignale pro Verbindungsspur (Lane) 10 Gbps zulässt. Bei voller Bestückung bedeutet das eine aggregierte Bandbreite von bis zu 5,12 Tbps. Konstruktiv bedingt sind eine Reihe von zusätzlichen Maßnahmen zur Steigerung der Zuverlässigkeit möglich. Natürlich gibt der Hersteller seitenweise Dinge an, die zur Energieeffizienz beitragen, das ist aber hier uninteressant.

Kommen wir lieber zu Funktionen, die die Virtualisierung direkt unterstützen. Hier ist vor allem die Virtual Connect Technologie interessant, die Server-zu-Netzwerk I/O-Verbindungen partitioniert und abstrahiert (diese Technologie nutzt HP auf der Basis eines OEM-Vertrags mit Brocade). Der Hersteller ist der festen Überzeugung, dass es nicht ausreicht, nur Server zu virtualisieren, sondern die Virtualisierung muss auch das gesamte I/O-Verhalten umfassen.

Die primären Charakteristika der I/O-Virtualisierung umfassen:

- Isolation von Änderungen bei Verbindungen zwischen Servern und Netzen
- Kompatibilität mit der extrernen Netzwerkumgebung des RZs
- Reduktion von Kabeln ohne die Erhöhung der Management-Komplexität

Die dazu wichtigsten Werkzeuge sind Virtual Connect und Flex-10. Virtual Connect virtualisiert die Verbindungen zwischen dem HP Blade System und den LANs und SANs im RZ. Das ermöglicht Administratoren, Ethernet und Fiber Channel Verbindungen zu poolen und aufzuteilen. Änderungen bei den Servern kann das Netz nicht mehr sehen. Im Kern ist Virtual Connect eine Abstraktionstechnologie für Geräte der physischen Ebene, die in ihrer Wirkungsweise der Abstraktionstechnologie der Virtuellen Maschinen auf der logischen Ebene entspricht. Dadurch kann sie eine ähnliche Flexibilität hinsichtlich der Workload und der Mobilität auf dieser Ebene erreichen. Genau wie eine Hypervisor-Software physikalische Server in Virtuelle Maschinen abstrahiert, abstrahiert die HP Virtual Connect Technologie Gruppen von physikalischen Servern in einer Virtual Connect Domäne in eine anonyme physikalische Maschine. Sobald der Server-Pool mit einem LAN oder SAN verbunden wird, benutzt der Server Administrator eine Virtual Connect Benutzerschnittstelle, um ein I/O-Verbindungsprofil für jeden Server zu erstellen. Anstelle von default MAC-Adressen für alle NICs default WorldWide-Names für alle Host Bus Adapter erzeugt der Virtual Connect Manager spezifische Serverprofile, ordnet diesen dann eindeutige MAC-Adressen und WWNs zu und verwaltet sie lokal. Virtual Connect verwaltet diese MACs und WWNs dann sicher durch den Onboard Administrator und die entsprechenden Schnittstellenkarten auf den Server Blades. Das erlaubt, dass man diese Profile verändern und auch migrieren kann, ohne die Sichtweise des Administrators auf die Verbindungen zwischen Servern und dem externen Netz zu verändern, ein c-Class Blade System erscheint als eine Sammlung von Servern mit statischen MAC und WWN-Zuordnungen. (siehe Abbildung 4)

Intuitiv betrachtet leistet Virtual Connect dasselbe wie VM-Tagging von Cisco. Die Frage ist nun, wo die Unterschiede liegen. Virtual Connect ist eine Technologie, die aus dem Umfeld des Fibre Channel entwickelt wurde. Weiter unten sehen wir, wie es im Rahmen eines Brocade-DCX-Umfeld funktioniert. Beide Systeme haben die Aufgabe, dafür zu sorgen, dass eine wandernde Virtuelle Maschine am Ziel ihrer Wanderung wieder die gleichen Verbindungen hat wie vor ihrer Wanderung. Dafür müssen eine Reihe von Parametern mitgenommen werden. Es gibt ein absolut vergleichbares Problem aus dem Umfeld von Providernetzen. Versorgungsbereiche, wie z.B. 1000 Haushalte mit DSL oder Mandanten im Zusammenhang eines mandantenfähigen Netzes müssen bezogen auf das Kernnetz eines Provi-

ders auch manchmal „umziehen“, z.B. immer dann, wenn der Provider seine innere Netzstruktur optimiert oder wenn die Mandanten tatsächlich umziehen. Hierfür benötigt man eine systematische Lösung für die Trennung von Kernnetz und Versorgungsbereichen. In der Praxis hat sich dafür weitest gehend MPLS durchgesetzt. MPLS definiert über einem Netzkern sozusagen einen Rand, an dem alle Einheiten stehen, die Kommunikation benötigen. Das Kernnetz muss die Verbindungen bereitstellen. Dafür hat Cisco vor langer Zeit das Tagging erfunden, mit dem den Netzknoten des Kernnetzes eine optimierte Arbeitsweise vorgegeben wird. Im VM-Umfeld sieht das nun genau so aus: Es gibt eine physikalisch/logische Struktur bestehend aus einem Netz, den Server, den Speichern, den lokalen Basis-Betriebssystemen, den Hypervisoren und den Virtuellen Maschinen. Das Netz, wie wir es kennen, wird dabei zum Systembus für diese Virtualisierte Infrastruktur. Auch dieses hat einen Rand, an dem stehen jetzt die Virtuellen Maschinen und wollen verbunden werden. Cisco hat einfach eine bewährte Technologie genommen, um die Wanderung der VMs zu unterstützen. Ich kenne in einem Ethernet-Umfeld auch keine Alternative mit vergleichbarem Leistungsumfang.

Jetzt bekommen wir aber ein Problem: wir benötigen für eine VM-Wanderung oft Komponenten aus zwei Welten, nämlich immer dann, wenn die VM sich auf Ethernet- und FC-SAN-Daten stützt. Die Aufgabe ist also nicht, sich zwischen VM-Tagging und der FC-Variante zu entscheiden, sondern diese zu integrieren. Dafür gibt es einen neuen Standard, FC-BB-5 (Fibre Channel BackBone) der von Cisco, Brocade und QLogic entwickelt wurde und z.Zt. ANSI zur Verabschiedung vorliegt. In diesem Standard gibt es einen Sub-Standard FC-BB_E, der die konvergierte Kommunikation für FC- und Ethernet-Daten beschreibt und eine Weiterentwicklung der Gedanken zu FCoE ist. Es wird darüber einen weiteren Artikel geben, wo sich die Einzelheiten dann klären werden.

Auf längere Sicht ist dann interessant, wie und mit welcher Leistung diese Funktionen in den Hypervisor integriert werden können. Das müssen wir aufmerksam beobachten.

Das neueste technische Element der HP Virtual Connect Strategie ist Flex-10. Flex-10 erlaubt dem Kunden eine Partitionierung einer 10 GbEthernet-Verbindung und die Regulierung von Größe bzw. Übertragungsgeschwindigkeit einer jeden Partition. Administratoren können einen einzelnen 10 GbE-Port so konfigurieren, dass er bis zu vier physikalische NIC-Geräte repräsentiert, die dann den Namen Flexi-NIC bekommen und zusammen eine Bandbreite von 10 Gbps belegen können. Jedes Dual-Port Flex-10 Gerät unterstützt bis zu acht Flexi-NICs, vier an jedem Port. Jedes Flex-10 Interconnect Modul kann bis zu 64 Flexi-NICs unterstützen.

Die Flexi-NICs erscheinen dem Betriebssystem wie diskrete NICs, jedes mit seinem eigenen Treiber. Obwohl sich die Flexi-NICs den gleichen physikalischen Port teilen, wird der Verkehr für jede Flexi-NIC mit einer eigenen MAC-Adresse isoliert. Außerdem gibt es VLAN-Tags zwischen einem Flexi-NIC und einem VC Flex-10 Interconnect-Modul.

Speziell in einer Umgebung einer virtuellen Maschine mit dem Bedarf nach multi-

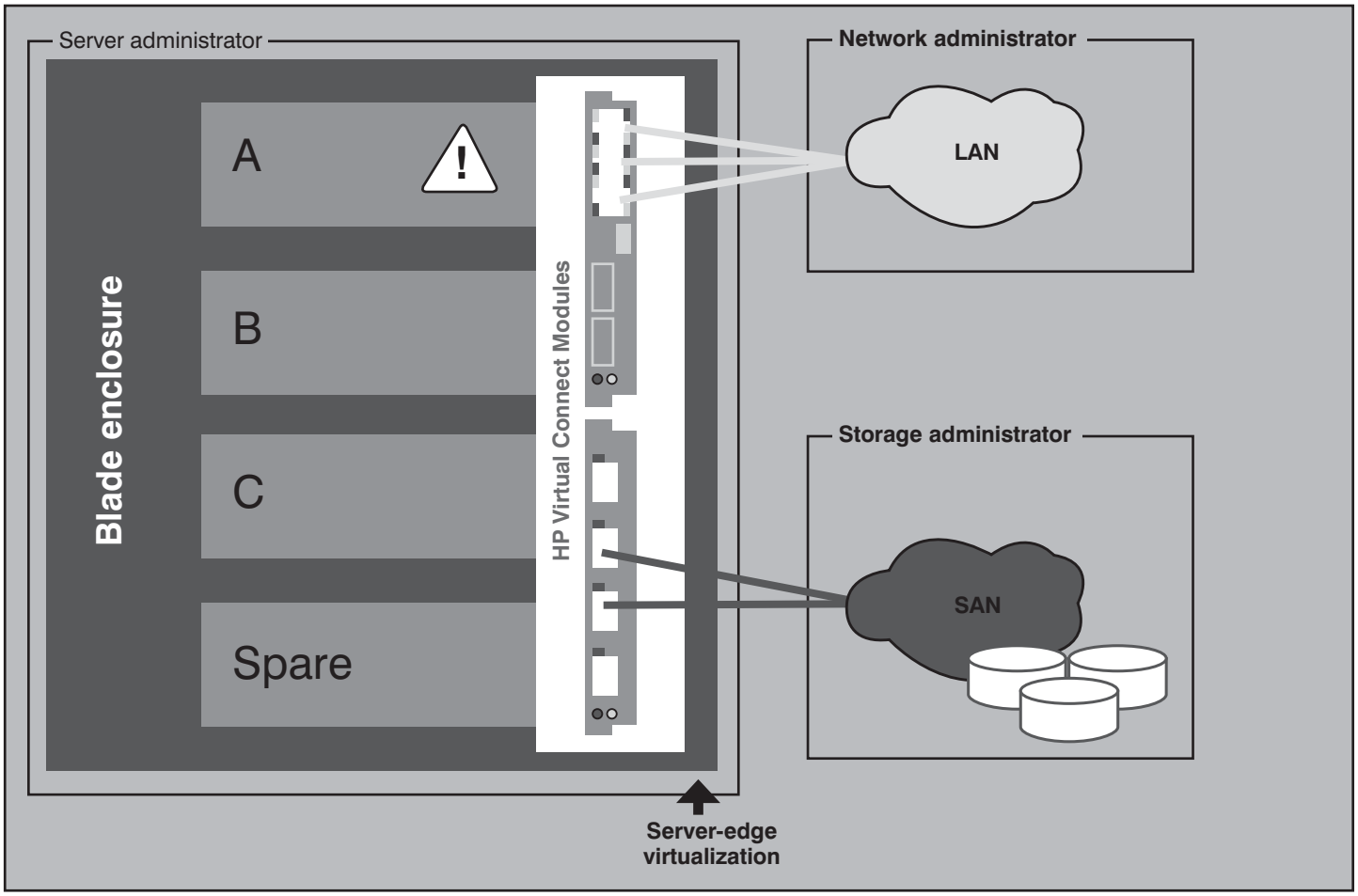


Abbildung 4: HP Virtual Connect

Quelle: HP

Der Kampf ums RZ: die nächste Runde - Teil 2: Matrix, FCoE und EoFC

plen NICs kann man damit viel Geld sparen, weil keine zusätzlichen Server NIC Mezzanine Karten und dazugehörige Interconnect Module mehr benötigt werden. Wegen der nativen 10 Gb-Leistung können Administratoren ultraschnelle Serverwechsel und Recoveries innerhalb einer Blade System Enclosure und zwischen Blades vornehmen und den Verkehr der virtuellen Server präzise während Backup, Wanderung Virtueller Maschinen usw. kontrollieren. (siehe Abbildung 5)

Jedes Flexi-NIC kann auf ein anderes Virtual Connect Network (vNet) abgebildet werden. Der Verkehr für jedes Flexi-NIC wird durch VLAN-Tags isoliert. Pakete, die durch VC und die Flexi-NICs getaggt und isoliert wurden, wandern über einen einzigen Weg, der vermöge IEEE 802.3ap 10 GBase-KR-Backplane gebildet wird, von einer Flex-10-Einheit (LOM oder Mezzanine-Karte) zum Flex-10 Enet Modul. (siehe Abbildung 6)

Flex-10 unterstützt Tunneling und Mapping von OS VLAN-Tags. Im Tunneling Modus leitet VC getaggte OS-Daten direkt durch das Flex-10 Enet-Modul, ohne sie sich anzusehen. (siehe Abbildung 7)

Ein Flexi-NIC kann aber auch so konfiguriert werden, dass die OS VLAN-Tags auf unterschiedliche VC-Netze abgebildet werden. (siehe Abbildung 8)

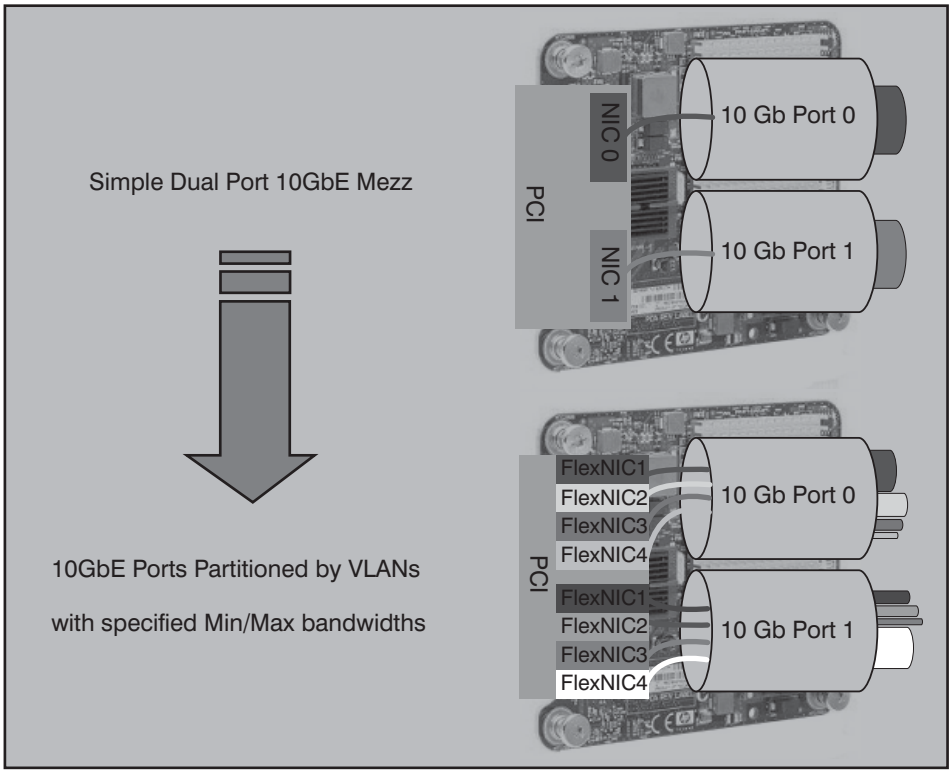


Abbildung 5: Bandbreitezuordnung durch Flex-10 Quelle: HP

In den Bildern sieht man deutlich, dass Flex-10-Server natürlich auch friedfertig mit älteren Servern zusammenarbeiten, die nur einen 1 GbE-Anschluss haben. Der Übergang wird durch VC konfiguriert und verwaltet.

Das HP Virtual Connect Flex-10 10 Gb Ethernet Modul ist mit 16 internen 10 GbE-Ports und acht externen 10 GbE SFP+

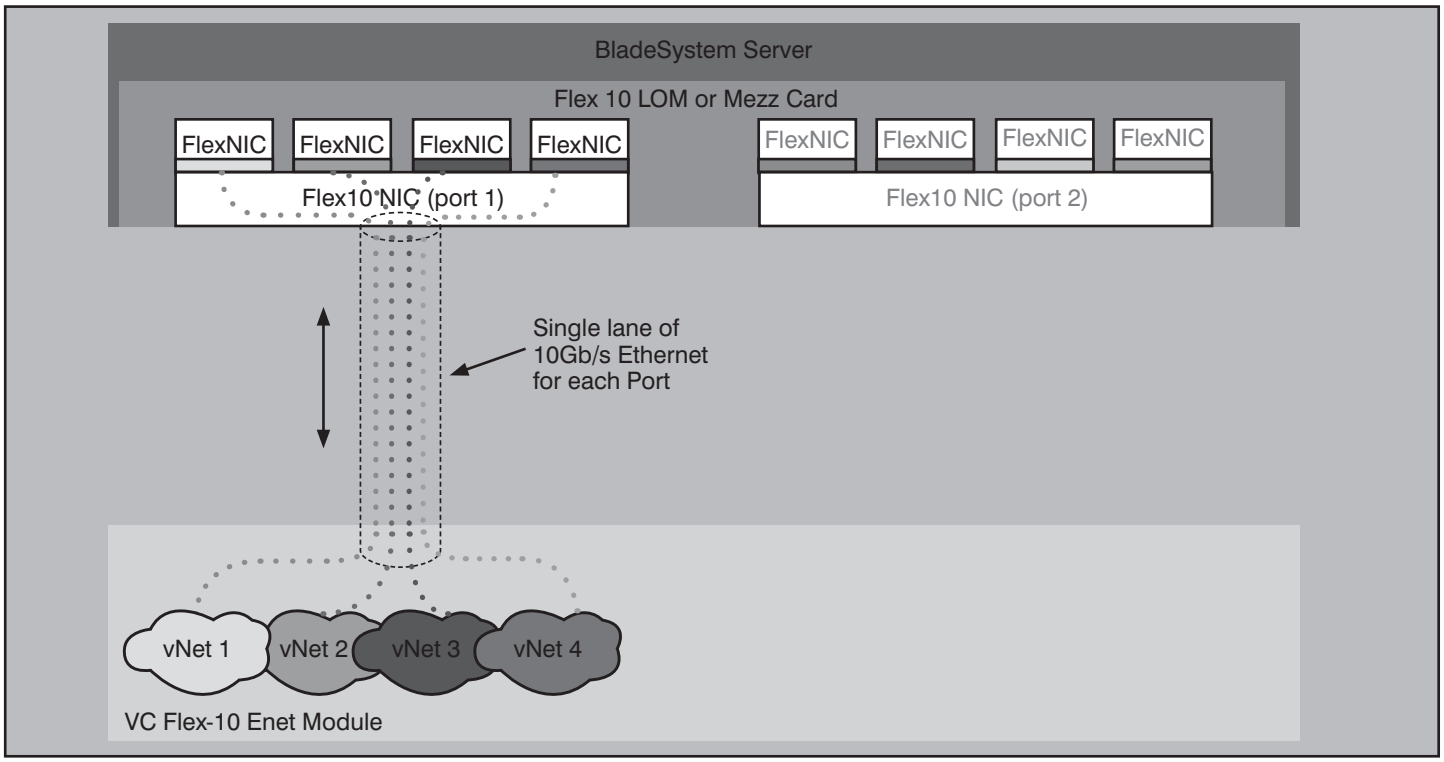


Abbildung 6: Flexi-NICs auf einer einzelnen physikalischen Verbindung Quelle: HP

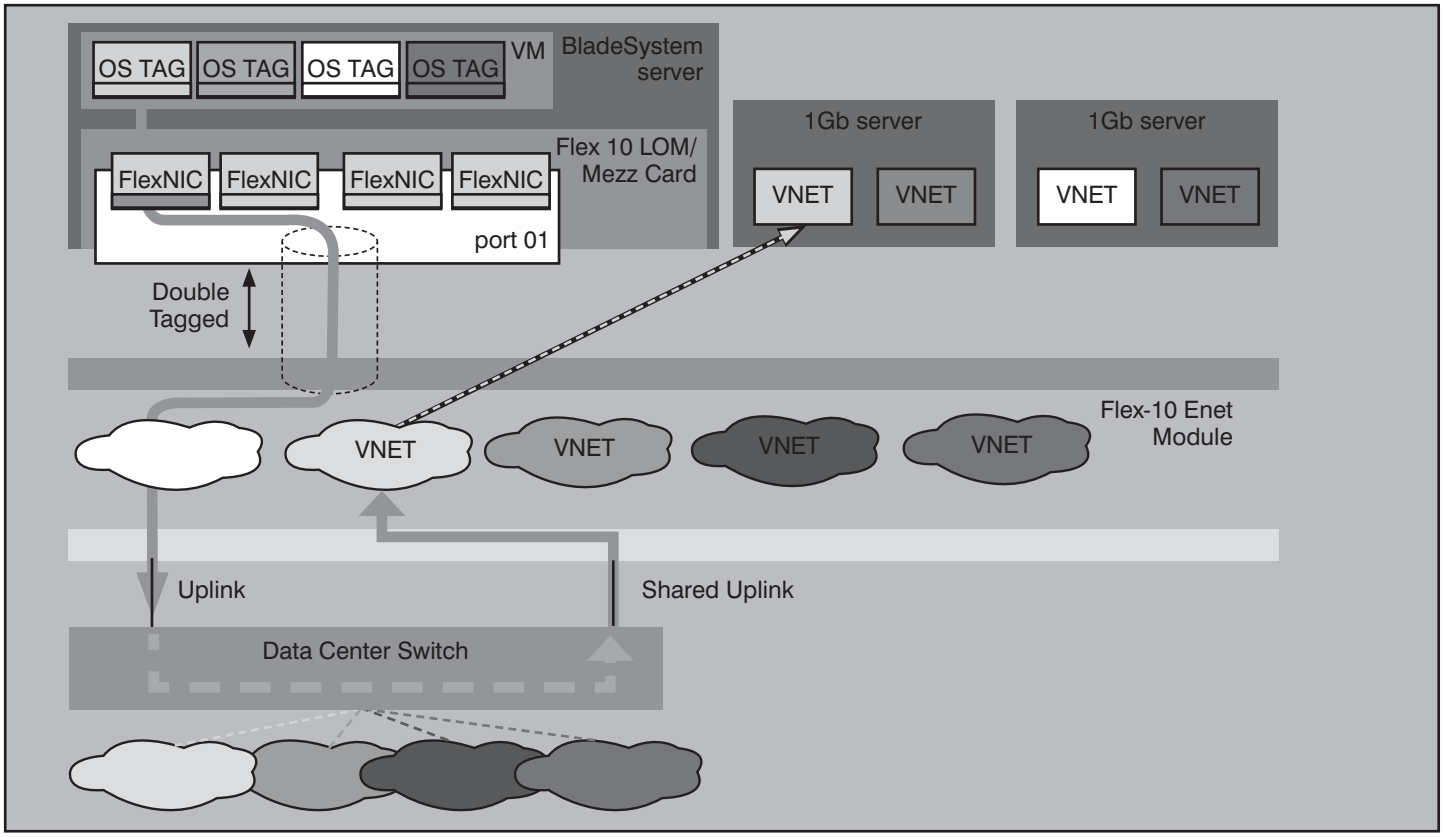


Abbildung 7: VLAN-Tunneling Mode

Quelle: HP

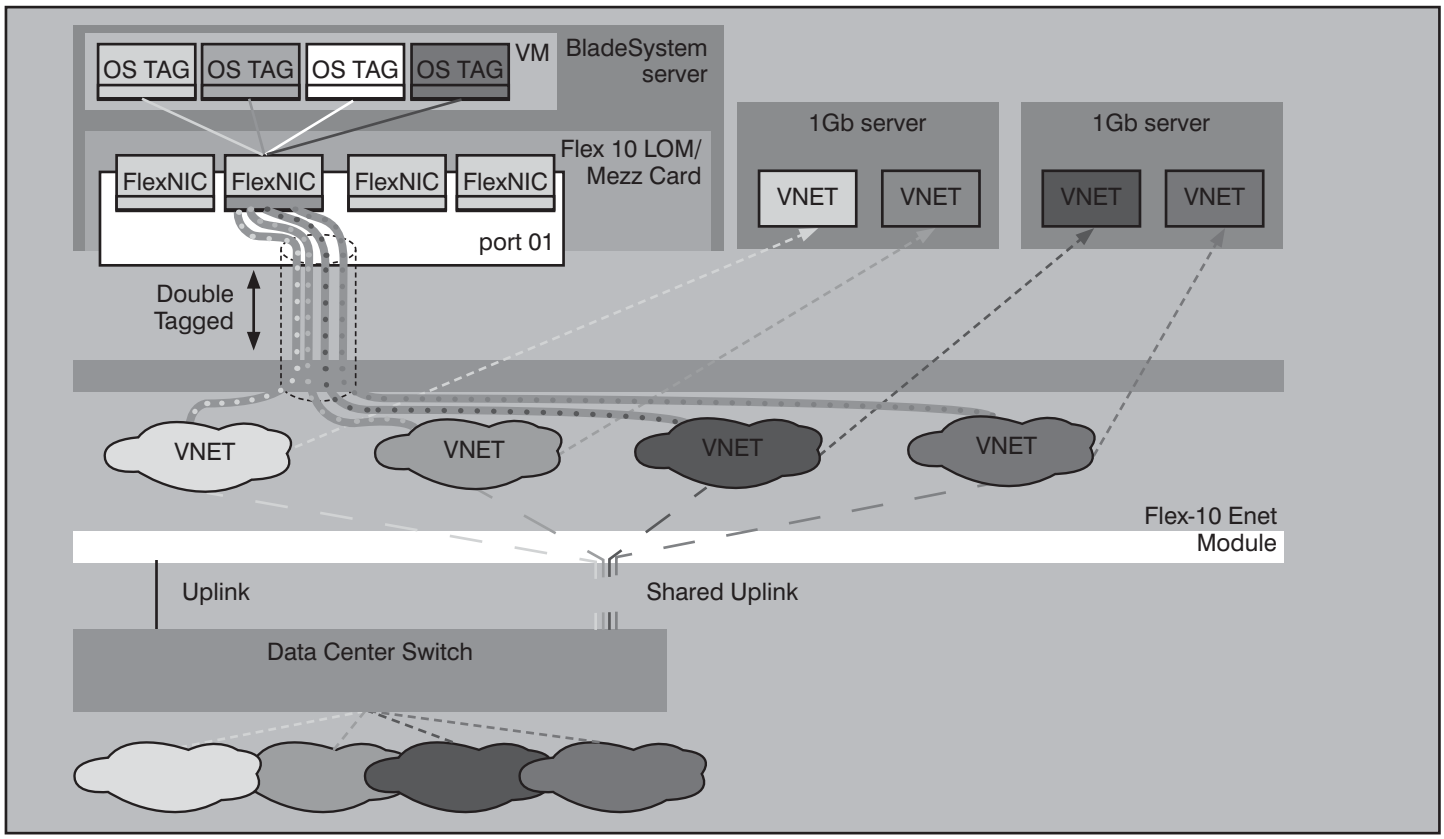


Abbildung 8: Doppelt getaggte Pakete im VLAN-Mapping Mode

Quelle: HP

Der Kampf ums RZ: die nächste Runde - Teil 2: Matrix, FCoE und EoFC

Ports und entsprechenden weiteren Ports für Crosslink Stacking die Verbindung zur Außenwelt, zwischen Enclosure und RZ-Netz.

Normalerweise wird die 10 GbE-Verbindung über einen CX-Port hergestellt, das ist mit Abstand die preiswerteste Variante. Außerdem gibt es noch Unterstützung für die 10 GbE-Varianten SR, LR und LRM.

Zusammenfassend kann man sagen, dass Flex-10 innerhalb der Enclosure genau das macht, was DCB eigentlich können sollte, nämlich eine sichere und deterministische Aufteilung der Bandbreite für die Koexistenz von Verkehrsströmen mit unterschiedlichen Anforderungen!

Die Virtual Connect Lösung umfasst den optionalen Virtual Connect Enterprise Manager VCEM. Diese Software unterstützt eine zentrale Konsole zur Verwaltung multipler VC-Domänen und effektiver LAN- und SAN-Verbindungen über die Domäne. VCEM erlaubt gruppen-basiertes Konfigurationsmanagement. Administratoren können damit Server-zu-Netzwerk-Verbindungen für bis zu 150 Enclosures mit insgesamt maximal 2400 Servern zuordnen, bewegen und Failover durchführen. VCEM hat einen zentralen Pool von VC LAN und SAN Adressen der es den Kunden erlaubt, multiple Enclosures physisch oder logisch miteinander zu verbinden und Server-zu-Netzwerkverbindungen in großen Mengen vorzudefinieren.

Virtual Connect erlaubt die physikalische Abstraktion von Servern in einem Ressourcen-Pool. So können Administratoren das Konzept eines logischen Servers dazu benutzen, entweder eine virtuelle oder eine physische Maschine zu beschreiben. Unter den Begriffen HP Insight Dynamics-VSE und HP Insight Orchestration finden Sie mehr zu diesem Themenbereich. Schauen Sie mal rein, das ist wirklich interessant.

Was ist nun die Blade System Matrix?

Die Blade System Matrix ist eine integrierte adaptive Infrastruktur-Lösung, die den kompletten Management-Software Stack und eine BladeSystem Software enthält.

Sie besteht aus folgenden Elementen:

- BladeSystem c-Class Infrastruktur. Das ist eine voll konfigurierte c7000 Enclosure mit Netzteilen, Lüftern, einem redundanten Onboard Administrator sowie dualen Flex-10 Ethernet und dualen 8 Gb Fibre Channel Virtual Connect Modulen
- Software und Lizenzen für bis zu 16 Server für Insight Control Management, HP Systems Insight Manager, Insight Dynamics-VSE, Insight Orchestration und VCEM

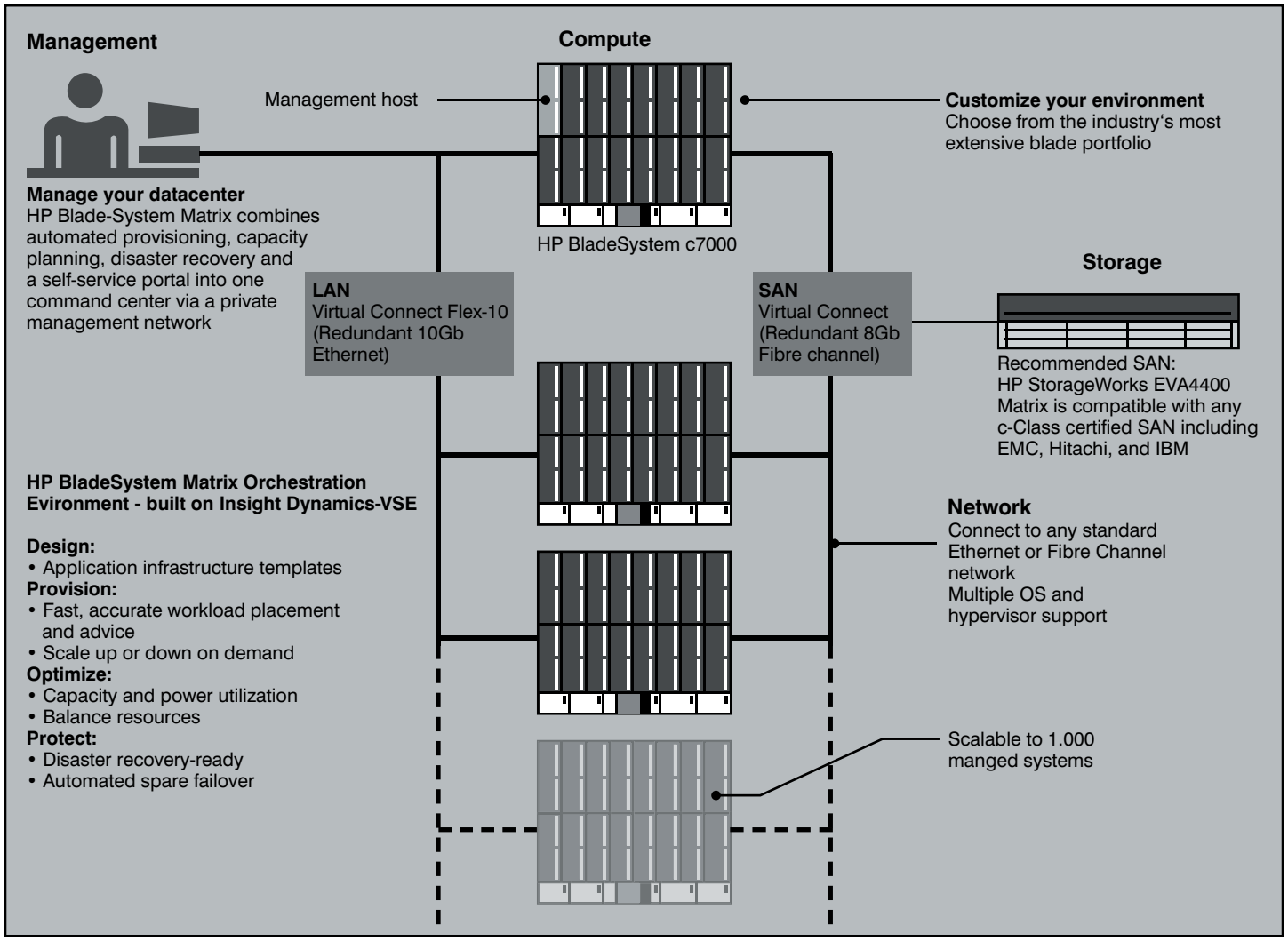


Abbildung 9: HP Blade System Matrix

Quelle: HP

- Speicher. Optional ist EVA 4400 SAN-Speicher, genauso gut wird aber auch SAN-Speicher von Dritten unterstützt

- Fabrik- und Onsite-Services für Installation, Konfiguration und Setup

Eine einzige Blade System Matrix mit einem Enclosure bildet sozusagen einen RZ-Mikrokosmos aus Servern, Speichern, Netzen und übergreifendem Management.

Diese Vorstellungen können wir durchaus aus der Enge der Enclosure befreien, denn einer virtualisierten Lösung ist es völlig gleichgültig, ob die Server und Speicher nun Blades sind oder Geräte mit eigenen Gehäusen und ob das Netz mit NICs und Switches nun durch Mezzanine-Karten oder ebenfalls durch eigenständige Karten und Geräte realisiert wird.

Bezüglich des Netzes sehen wir Folgendes:

- Die aggregierte Gesamtleistung liegt im Bereich mehrerer Terabit/s.

- Es gibt ZWEI gleichberechtigte Netzwerksysteme: Flex-10 und Fibre Channel

- Flex-10 kann das, was Ethernet mit DCB können sollte (und immer noch nicht deterministisch schafft): flexible Zuordnung fester Bandbreiten im Rahmen einer konsolidierten 10 GbE-Verbindung

- Flex-10 und FC werden einheitlich und gleichberechtigt verwaltet

- Alle Netzwerkressourcen werden ausnahmslos virtualisiert

- Die Verwaltung der Netzwerkressourcen ist in ein allgemeines Management-System für Server, Speicher und Netze eingebettet

- Es gibt KEIN NORMALES ETHERNET mehr

Erstaunlich ist aber vor allem: *der erfahrene Server- und Netzwerkhersteller HP wirft das Fibre Channel SAN keineswegs über Bord, obwohl doch Flex-10 in der Lage sein sollte, FCoE-Verbindungen zu unterstützen!*

Natürlich bleiben noch offene Fragen, z.B. wie es sich HP hinsichtlich des technischen Netzes genau vorstellt, wenn man 100 oder mehr Enclosures zu einer großen Matrix zusammenschalten möchte (die Management-Software kann das ja nach Aussagen des Herstellers schon verwalten) oder ob das Flex-10 auch für normale 10 GbE-ProCurve-Switches zur Verfügung stehen wird.

3. Roundup Fibre Channel

Fibre Channel (FC) ist eine serielle Übertragungstechnologie für den High-Speed-Datentransfer. Fibre Channel ist ein offener Standard, definiert durch ANSI und OSI und unterstützt alle wichtigen höheren Protokolle wie IP, ATM, HIPPI, SCSI und sogar IEEE 802.1/3 Ethernet.

Der FC verfügt über keinen eigenen Befehlssatz, sondern stellt lediglich den Datentransfer zwischen den einzelnen FC-Geräten her. Er ist allerdings nicht auf die Übertragung von optischen Signalen durch Glasfasern beschränkt, sondern kann auch mit kostengünstigen Kupferkabeln wie Twisted Pair- oder Koax-Kabeln realisiert werden.

Gegenüber Ethernet hat der Fibre Channel eine Reihe von Alleinstellungsmerkmalen, wie z.B. die Möglichkeit isochroner Übertragung. Darüber hinaus ist es durch die Unterstützung verschiedener Topologien sehr flexibel und auf kleinen und großen Systemen anwendbar beziehungsweise skalierbar. Die Installation ist einfach, eine ID-Vergabe erfolgt automatisch. Zusätzlich lässt sich FC einfach integrieren und arbeitet extrem zuverlässig, da es alle wichtigen Protokolle unterstützt und eingebaute Korrekturalgorithmen besitzt.

Einer der großen Vorteile von FC ist die enorme Vielseitigkeit der möglichen Konfigurationen. Sie reicht von einfachen Strukturen, bei denen lediglich zwei Geräte miteinander verbunden werden, bis hin zu komplexen Netzwerken mit über 16 Millionen Teilnehmern.

FC unterstützt mehrere unterschiedliche Topologien, die in der Praxis auch gemischt werden können. Die Point-to-Point-Verbindung ist die einfachste Verbindung von zwei FC-Geräten zum Beispiel zwei Servern oder ein Server- und ein Festplatten-Subsystem. Die FC Geräte wie etwa den FC-Controller bezeichnet man als Nodes. Diese wiederum haben einen oder auch mehrere so genannte N_Ports – die eigentliche FC-Schnittstelle. Jeder N_Port hat jeweils einen Sender (Transmitter) und einen Empfänger (Receiver).

Sender und Empfänger zwischen den einzelnen Geräten sind über Kreuz miteinander durch so genannte Links verbunden. Diese Links können bei FC entweder Kupferkabel oder Glasfasern sein. Bei der Point-to-Point-Verbindung steht die gesamte Bandbreite der FC-Verbindung von aktuell 8 oder 10 Gbps exklusiv für die Kommunikation der zwei Nodes zur Verfügung.

In der Arbitrated-Loop-Topologie können bis 127 Ports in einer Ringstruktur zusammengeschaltet werden. Die Ports in einer Arbitrated Loop werden als NL_Ports bezeichnet. In dieser Konfiguration sind jeweils zwei Ports gleichzeitig aktiv. Die anderen Ports arbeiten als Repeater und reichen lediglich die Signale weiter. Das heißt natürlich, dass sich die Bandbreite von 8 oder 10 Gbps auf alle Teilnehmer verteilt. Ähnlich wie beim Token Ring „sieht“ jeder Teilnehmer der Arbitrated

Der Kampf ums RZ: die nächste Runde - Teil 2: Matrix, FCoE und EoFC

Loop alle Messages und sendet diejenigen, die nicht für ihn bestimmt sind, einfach weiter.

Die Switched Fabric ist eine vermaschte Topologie. FC-Geräte sind über sogenannte F_Ports oder FL_Ports an dieses Netzwerk angeschlossen – je nachdem, ob es sich um einfache oder Loop-fähige Ports handelt. Die Verbindung zwischen den einzelnen Ports wird durch das Netzwerk geschaltet. Das FC-Gerät wird über einen F_Port oder FL_Port an die Switched Fabric angeschlossen. Möchte nun zum Beispiel Node A mit Node B kommunizieren, wird die entsprechende Adresse der Switched Fabric übergeben. Die Switched Fabric schaltet eine entsprechende Verbindung vom Initiator zum Target, wobei jedoch beide nicht wissen, welchen Weg die Signale nehmen. Die Verbindung wird vollständig durch die Switched Fabric hergestellt und ist vollkommen transparent für die Teilnehmer.

Zur Adressierung stehen 24 Bits zur Verfügung (bei der Arbitrated Loop werden nur die untersten 8 Bit genutzt). Das reicht für 16 Millionen Teilnehmerstationen.

Es gibt im FC fünf Serviceklassen:

1	verbindungsorientiert
2	verbindungslos, mit Bestätigung
3	verbindungslos, ohne Bestätigung
4	parallele Übertragung
5	parallel und isochron

Die Serviceklasse 1 stellt eine dedizierte Verbindung zwischen Sender und Empfänger her. Alle gesendeten Pakete werden vom Empfänger quittiert. Während des Bestehens dieser Verbindung können keine anderen Teilnehmer die verbundenen Partner ansprechen. Serviceklasse 2 ist eine „verbindungslose“ Methode mit Bestätigung des Datentransfers. Dies bedeutet, der Weg, den die Datenpakete nehmen, ist unbestimmt. Die verfügbare Bandbreite kann hierbei unter mehreren Teilnehmern aufgeteilt werden. Serviceklasse 3 ist ähnlich wie Klasse 2, jedoch ohne Bestätigung des Datentransfers. Diese Verbindungsklasse wird in der Regel bei Massenspeichersystemen genutzt. Durch die Möglichkeit der Aufteilung der Bandbreite können mehrere FC-Geräte parallel miteinander kommunizieren. Zum Beispiel kann der RAID-Controller Daten sehr schnell aufeinanderfolgend an mehrere Festplatten senden, ohne auf die Bestätigung der einzelnen Datenpakete warten zu müssen. Da die Empfangsbestätigung durch das höhere SCSI-Protokoll ausgeführt wird, ist auf der unteren FC-Protokollebene keine Empfangsbestätigung notwendig.

In Serviceklasse 4 werden Datenpakete zwischen zwei Teilnehmern in einem Netzwerk unter Ausnutzung mehrerer Verbindungsmöglichkeiten bei garantierter Bandbreite ausgetauscht. Serviceklasse 5 ist ähnlich wie Klasse 4, jedoch bei zusätzlicher isochroner Datenübertragung.

An diesen Serviceklassen sieht man sofort, dass die Nachbildung von FC über Ethernet mit FCoE und DCB schwierig ist, denn man geht ja einfach davon aus, die FC-Pakete in Ethernet-Pakete einzupacken, die Logik an den Enden aber wiederum FC-Funktionseinheiten zu überlassen. Eine FC-Funktionseinheit, die aber jetzt z.B. eine isochrone Klasse-5-Übertragung bereitstellen möchte, weiß ja nicht, dass statt des bisher üblichen Weges nunmehr eine Ethernet-Switching Fabric durchlaufen werden soll, die auch beim besten Bemühen keine Isochronität hinbringt.

FC unterstützt unterschiedliche Paketlängen von prinzipiell 0 Bytes bis 2048 Bytes pro Paket und ist deshalb einerseits optimal geeignet für kleine IOs, wie sie typischerweise bei Datenbanken auftreten; andererseits ist der FC aber auch in der Lage, größere Datenmengen, wie sie zum Beispiel in Video-Applikationen vorkommen, effektiv und ohne viel zusätzlichen Overhead zu übertragen.

Die Definition des FC-Protokolls ist in fünf verschiedene Protokollschichten untergliedert. Die Definitionen der physikalischen Verbindungen (Kabel, Stecker, Transmitter und Receiver) werden in der so genannten **FC-0-Schicht** zusammengefasst. Die **FC-1-Schicht** beschreibt die Bitübertragung und steuert die 8/10-Bit-Kodierung-/Dekodierung. Diese ermöglicht eine extrem niedrige Bit-Fehlerrate, erhöht aber die zu transportierende Datenmenge um 25 Prozent. Die **FC-2-Schicht** ist für die Steuerung des Datenflusses verantwortlich. Hier werden die einzelnen Pakete mit Adresse, Daten und CRC-Information zusammengestellt. Diese Schicht übernimmt auch das ACK-Handling. In der **FC-3-Schicht** werden gemeinsame Funktionalitäten von Gruppen von Netzwerkteilnehmern definiert. Die Unterstützung der höheren Protokolle (IP, IEEE 802, HIPPI oder SCSI) wird schließlich in der **FC-4-Schicht** geregelt. Insgesamt können in dieser Schicht bis zu 255 verschiedene höhere Protokolle definiert werden, was noch genügend Raum für zukünftige Protokolle lässt.

Die Verbindungsstrecken zwischen den einzelnen FC-Geräten bei Massenspeichersystemen sind normalerweise nicht sehr lang. Deshalb kommen hier bevorzugt Kupferkabel zum Einsatz. Für FC sind insgesamt drei verschiedene Kupferkabel (Video Coax, Miniature Coax und Shielded Twisted Pair) mit FC-DB9-Stecker (ähnlich denen der seriellen PC Schnittstelle) definiert. Mit diesen Kabeln können Strecken von bis zu 25 Metern zwischen den einzelnen Geräten realisiert werden.

Bei größeren Entfernungen nimmt man Glasfasern. Die 50-μm-MMF erreicht eine Kabellänge von 500 Metern, die 62,5-μm-MMF bis zu 175 Metern. Stecker: SC-Duplex. Mit 9-μm-SMF können bis zu 10 Kilometern überbrückt werden.

Um aufgeteilte Ressourcen zu vermeiden, werden heute für FC-SANs echte Switches eingesetzt. Ihre Backbone-Architektur kann zwischen den angeschlossenen Systemen gleichzeitig mehrere, voneinander unabhängige Verbindungen mit voller Bandbreite schalten. Mit FC-Switches lassen sich vermaschte oder kaskadierte SANs mit vielen Endgeräten konstruieren. Zwischen den Endgeräten können die Daten zumindest theoretisch frei fließen. Damit agieren FC-Switches wie Ethernet-Switches in LAN-Infrastrukturen.

Eine FC-Switching-Fabric darf heute aus maximal 239 Switches bestehen. Jeder Switch unterstützt wiederum maximal 256 Loops und 256 Ports sowie 128 Nodes pro Loop. An die Ports von FC-Switches lassen sich auch alte Endgeräte, die auf Arbitrated-Loop-Technologie optimiert sind, anschließen. Jedes Gerät im FC-SAN ist durch einen World-Wide-Node-Name (WWN) und eine 24 Bit lange Fibre-Channel-Adresse eindeutig gekennzeichnet. Die FC-Adresse setzt sich aus einem 8 Bit langen Abschnitt für die Domain und einem genauso langen Abschnitt für das Areal (den Loop) sowie 8 Bit für den Port zusammen. Die Adressen werden durch einen Fibre-Channel-spezifischen Domain Name Service (DNS) entdeckt und festgelegt.

Eine wichtige Funktion von Switches ist die Zugriffssteuerung auf die einzelnen Speichersysteme mittels Zoning. Zoning kann Hardware- und Software-basiert erfolgen. Beim Softzoning erhalten Geräte lediglich Informationen über die Systeme, mit denen sie reden dürfen. Beim Hardzoning überprüft eine Hardware alle durchlaufenden Pakete und leitet sie nur an erlaubte Adressen weiter. Eine weitere Zoning-Methode bezieht sich auf die WWN-Namen jedes Geräts oder Ports und bietet demzufolge mehr Eindeutigkeit. Neue Ansät-

ze erlauben das Zoning bis auf die Logical Units (LUN) der angeschlossenen Speichereinheiten hinab.

Enterprise-Switches haben zwischen 16 und 32 Ports. Sie sind ebenfalls meist fest konfiguriert, erlauben aber aufgrund ihrer hohen Portzahl den Aufbau kaskadierter oder vermaschter Infrastrukturen.

Direktoren sind die Entsprechung zu Highend-Core-Switches in LANs. Sie verfügen derzeit über zwischen 64 und 256 Ports und eine modulare Architektur mit Karten, die in ein Chassis eingebaut werden. Diese Geräte bieten Redundanz aller wesentlichen Komponenten, weiter gehende Management-Features und eine breitbandige Backplane für viele gleichzeitige Verbindungen mit voller Bandbreite. Die Backplane ist bei Highend-Systemen häufig doppelt ausgelegt, um Ausfälle auch an diesen kritischen Stellen zu verhindern.

Fabric Services sind Dienste, die Geräte im FC-SAN beanspruchen können. Die wichtigsten Dienste sind Name Server Support (ein Namensverzeichnis für die gesamte Fabric, das über alle Switches verteilt und somit überall verfügbar ist), Management Services, Statuskontrolle und Time Server.

4. Die Zukunft von FC und EoFC

FC-Hersteller sind immer recht bescheiden, denn traditionell wird bei FC nicht die Nominaldatenrate angegeben, sondern die tatsächlich nutzbare. Wir haben ja gesehen, dass FC immer die stabile 8B/10B-Codierung verwendet. Das zieht nach sich, dass bei einem 8 GbFC 10 Gbps auf der Leitung sind.

Schaut man es sich genau an, verwenden 10 GbEthernet und 8 GbFC eine praktisch identische Übertragungstechnik. So ist es auch zu erklären, dass führende FC-Hersteller für den 8 GbFC die gleichen technischen Schnittstellen angeben, wie dies führende Hersteller auch für 10 GbE machen. Sie haben nur etwas andere Namen. So entsprechen die Bezeichnungen SWL, LWL und ELWL (Short Wave Length Laser, Long Wave Length Laser und Extended Long Wavelength Laser) den aus dem 10 GbE-Universum bekannten SR, LR und ER (Short, Long und Extended Reach). Zugrunde liegen die Wellenlängenbereiche um 850nm, 1310nm und 1550nm, so dass man bei FC eben einen Bezug zu diesen Wellenlängen herstellt und sich bei Ethernet lieber auf die mit Lasern dieser Wellenlängen erreichbaren Distanzen, ein geeignetes Kabel vorausgesetzt, bezieht. Es sind aber die gleichen Transceiver z.B. in SFP- oder XFP-Gehäusen. Der einzige wirkliche Unterschied ergibt sich eigentlich nur bei Kupferkabeln. Während sowohl FC als auch 10 GbE eine Koaxvariante haben, wird STP-Verkabelung nur von 10 GBASE-T unterstützt.

Damit halten wir etwas Wichtiges fest:

Welche Strategie ein RZ-Betreiber in der Zukunft auch wählen möchte, mit OM-3-MMF-Kabeln und Standard-SMF ist man immer auf der sicheren Seite. Die CX-Verkabelung in Racks kann man leicht durch aktive optische Kabel ersetzen.

Eine STP-Variante gibt es für FC nicht.

Die Behandlung von FC-Daten außerhalb der Transceiver kann ganz normal mit ASICs oder Netzwerk-Prozessoren geschehen. In einem FC-Director moderner Bauart ist eigentlich auch nichts anderes drin als in einem vergleichbaren High End Ethernet Core-Switch.

Die Nutzung identischer Komponenten im 10 GbE und im 8GbFC führt zur zweiten wichtigen Aussage:

Es gibt keinen nachvollziehbaren Grund, warum sich ein FC-Core Director und ein Ethernet Core-Switch erheblich im Preis unterscheiden sollten.

Heute gibt es ja noch große Preisdifferenzen zwischen FC- und Ethernet-Komponenten, z.B. bei den Adaptern. Das muss und wird sich ändern.

Nun wäre das ja weiter noch nicht wirklich aufregend, wenn der FC-Director nach wie vor nur ein SAN und der Ethernet Core-Switch eben das Ethernet im RZ bilden würde. Mit FCoE versucht man ja aber grade, die SANs auf Ethernet-Strukturen abzubilden, um I/O-Konsolidierung zu erzielen.

Die interessante Frage ist: kann man das nicht auch „andersherum“ machen, also einen Kern aus FC-Directors bilden und den Ethernet-Verkehr darauf abbilden? Dann wäre das Ziel, ein einziges Kernnetz zu schaffen, ebenfalls erreicht.

Dazu würde man EoFC benötigen, aber das gibt es längst. Unter der US-Patentnummer USPTO 20080159277 gibt es ein Patent auf EoFC von den Herren S. Vabbilisetty und J.M.Terry. Gleichermaßen hat auch ANSI entsprechende Verfahren entwickelt.

In dieser Richtung ist die Abbildung viel einfacher. Bei FCoE benötigt man (bis zum heutigen Tage nicht standardisierte) Jumbo-Frames, um es zu schaffen, ein FC-Paket in nur ein Ethernet-Frame einzupacken. Würde man bei der normalen Ethernet Frame Größe bleiben, käme es zu der weiteren Komplikation, dass man die FC-Frames auch noch teilen und wieder zusammenbauen müsste.

Umgekehrt passen Ethernet-Pakete ohne Probleme in den FC. Sind die Ethernet-Pakete einmal in einem FC-Frame, merken sie nichts mehr, denn die FC-Verkehrsdurchschaltung geschieht automatisch. Aus der Perspektive des Ethernet-Frames

ist die ganze FC Fabric ein Tunnel. Das Einzige, was notwendig ist, ist die Abbildung der MAC-Ziel-Adressen der Ethernet-Pakete in entsprechende FC World Wide Names.

Man könnte jetzt wie bei Carrier Ethernet hingehen, und wesentliche Ethernet-Strukturen nachbilden. Viel einfacher ist es aber, um den FC Fabric Kern, in dem es nur FC gibt, sozusagen einen Rand zu schaffen, der die außerhalb des Kerns liegende Ethernet-Infrastruktur abbilden und umsetzen kann. Die Hauptfunktion der Randswitches ist neben dem Auspacken der Ethernet-Päckchen im Grunde die einer Ethernet Remote Bridge nach IEEE 802.1, also durchaus zu schaffen.

Möchte man noch mehr Qualität erreichen, kann man eine bei den Ethernet-Paketen vorhandene Priorisierung natürlich auf die unterschiedlichen FC-Verkehrsklassen abbilden. Manche FC-Hersteller haben aber für die FC-SANs auch ein eigenes QoS-Konzept, auf das man dann ebenfalls zurückgreifen kann.

Damit kommen wir zur dritten wichtigen Aussage:

Die Gestaltung eines RZ-Core Netzes mit Fibre Channel als Basistechnologie ist möglich. Die Abbildung von Ethernet via EoFC ist deterministisch und hat keine zusätzlichen Problemfelder wie das Gegenstück FCoE.

Es geht hier ausschließlich um die Konsolidierung des RZ-Cores, die ja nicht nur den Zweck haben sollte, Kabel und Adapter zu sparen, sondern eine notwenige Voraussetzung für eine erfolgreiche Unterstützung der Virtualisierung ist. Sie erinnern sich: in diesem Zusammenhang wird das RZ-Netz zum Systembus! Ethernet alleine kann das nicht leisten, darum gibt es ja die Bemühungen von FCoE bis DCB. In einem RZ gibt es aber meist schon ein FC-SAN und man hat entsprechende Investitionen auch in Funktionen gemacht, die das Ethernet im Speicherumfeld einfach nicht bietet. Manche Hersteller sagen uns jetzt, man solle das FC-SAN über Bord werfen und weitgehend durch ein Ethernet mit FCoE ersetzen. Sie garantieren uns aber immer noch nicht, dass das in größeren und/oder hoch belasteten Umgebungen ordentlich funktioniert, mal ganz abgesehen von den Problemen mit der Netzwerksteuerung. (siehe Abbildung 10)

Dann ist es doch nur legitim, darüber nachzudenken, ob man nicht einen anderen Weg, eben mit FC als Core-Netz

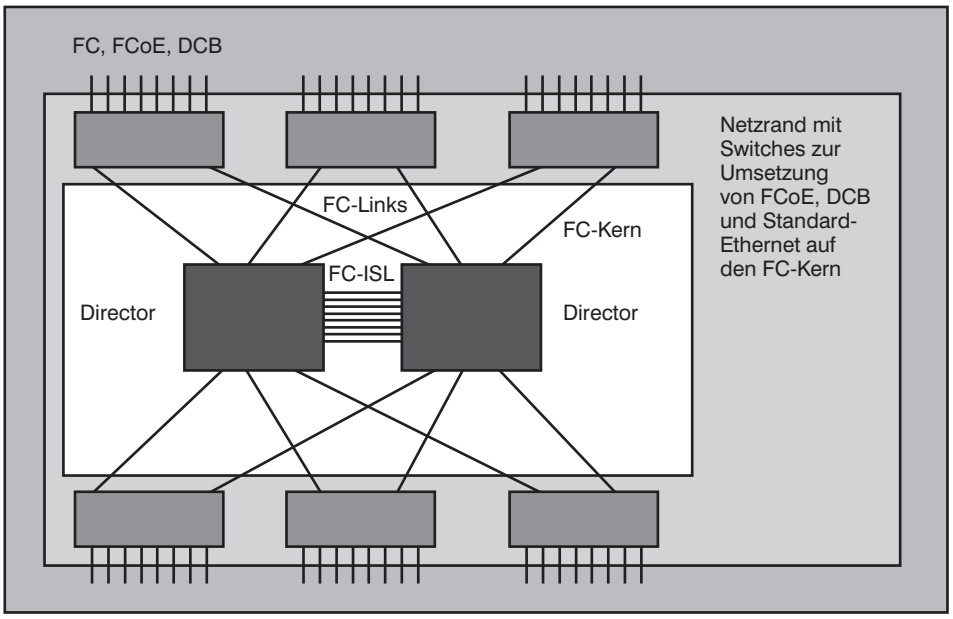


Abbildung 10: RZ-Kern-Netz auf FC-Basis

und Ethernet-Umwandlung am Rande via EoFC, gehen kann. In jedem RZ wird es eine Menge von Geräten geben, die eine Ethernet-Schnittstelle besitzen und deren Aufrüstung auf FC schlicht unsinnig wäre. Mit einer EoFC-Umwandlung könnte man sie leicht in ein FC-Kernnetz integrieren.

5. Brocade DCX-Backbone

Brocade ist der Hersteller, der mit seinen Produkten die Brücke zwischen Ethernet und Fibre Channel schlägt. Durch die enge Kooperation mit HP und IBM ist Brocade zu einem der Schlüsselhersteller für die Zukunft der Rechenzentren-Infrastrukturen aufgestiegen.

Es gibt für den DCX-Backbone zwei erheblich erweiterte SAN-Directors mit unterschiedlichen Leistungsklassen: einen mit 3 Terabit/s aggregater Chassis-Gesamtleistung für 384 8 GbFC-Ports und einen mit „nur“ 1,5 Terabit/s. für 192 Ports. Außerdem gibt es beide Varianten als Doppelchassis-Ausführung mit je zwei indentischen Maschinen. Wir werden gleich sehen, wozu das gut ist. Insgesamt kann man bis zu 239 Switches in einer Fabric zusammenführen. Mitte 2009 sind folgende Eckdaten zertifiziert:

- 6000 aktive Knoten, 56 Switches, 19 Hops mit Brocade FOS-Fabrics
- 31 Switches und 3 Hops mit Brocade M-EOS Fabrics

Für größere Netze muss man beim Hersteller eine Zertifizierung anfordern.

Natürlich besteht der Wunsch nach Inter Switch Links, z.B. für redundanten Gesamtaufbau. Hier unterstützt der große DCX bis zu 512 Gbit/s, die durch 4 ICLs mit je 16 8 GbFC-Ports aufgebaut werden können, der kleine wieder genau die Hälfte. Das gibt bei voller Nutzung ein ordentliches Kabelbündel, aber man kann überlegen, das z.B. bei Fernkopplung über DWDM zu realisieren. Sehr elegant ist in diesem Zusammenhang die Verwendung des Doppelchassis. Hier besteht zwischen beiden Maschinen direkt eine 512 oder 256 Gbi/s. ISL.

Auf allen Switches von Brocade ist Brocades FOS (Fibre Channel Operating System) implementiert. Darin sind Funktionen wie das Switch-Management über die Anwendung WebTOOLs und Zoning enthalten. Zusätzlich können Anwender Lizenzen für Erweiterungsanwendungen erwerben. So können sie zum Beispiel via Advanced Performance Management die SCSI-Lese-Schreibleistung jedes einzelnen Ports kontrollieren. Trunking ermöglicht, besonders leistungsstarke Verbindungen zu Speichersystemen aufzubauen. Remote Management macht die Systemsteuerung unabhängig von Vor-Ort-Einsätzen, so lange es um Softwareprobleme geht. Besonders für den Backup-Bereich und Remote-SANs eignen sich Entfernungserweiterungen (Extended Distance) bis auf rund zehn Kilometer auf Port-Level. Auch Brocades Fabric Manager läuft als optionale Lösung auf Switches, die Fabric Management ermöglichen.

Der Kampf ums RZ: die nächste Runde - Teil 2: Matrix, FCoE und EoFC

Damit erwartet die Kunden gegenüber einer ausschließlich externen Switching-Lösung keinerlei Nachteil. Vielmehr sparen sie Hardware und Verkabelungsaufwand. Zudem vereinfacht sich das Management und es ergibt sich wegen der großen Verbreitung von Brocade-Switches für die bisher erhältlichen Lösungen ein Kompatibilitätsvorteil, der vielen Unternehmen sehr gelegen kommen dürfte. Das Konzept des Embedded Switching in Blade-Chassis erscheint wegen all dieser Eigenschaften prinzipiell so überzeugend, dass man sich fragen muss, wann weitere OEMs auf den fahrenden Zug aufspringen und wann Embedded Switching zum allgemein akzeptierten Standard für Blade-Systeme werden wird.

M-EOS ist eine andere Betriebssystemvariante von Brocade, die im Zusammenhang mit dem Director 48000 steht. Die Ausrichtung dieses Geräts ist etwas anders als die des DCX, für die Integration heterogener Systeme wird hier primär FCIP verwendet. Wir wollen hier nicht das gesamte Produktspektrum von Brocade ausbreiten. Die Investitionen in intelligente SAN-Blades, die im Zusammenhang mit dem 48000 gemacht wurden, lassen sich bei der Migration von 48000 auf DCX retten.

Für die Entwicklung der DCX Data Center Fabric Platforms gibt der Hersteller vor allem folgende Ziele an:

- Konsolidierung der Fabric-Technologie und Erweiterung auf virtuelle Speicher und Server sowie abgesetzte Rechenzentren
- Integration virtueller Verbindungen zwischen virtuellen Servern und virtuellen Speichern über unterschiedliche Protokollwelten mittels der Virtual Channel Technologie von Brocade
- Einführung hoher Intelligenz in die Fabric mit Anwendungs-bewussten Service-Leveln durch die Brocade Adaptive Networking Services
- Integration einer Anwendungsplattform in die Fabric für die Realisierung unterbrechungsfreier Skalierbarkeit und Anwendungen, die Speicherdaten bewegen (Replikation, Migration, Virtueller Speicher…)
- Integration von Fabric Diensten mit der Bereitstellung virtueller Server. Diese erweitern die Anwendungs-bewussten Service-Level und Methoden zur Sicherung virtueller Speicherdaten

Die Adaptive Networking Services umfassen QoS, Traffic Management, Fabric Dynamic Profiling und Resource Recovery. QoS erlaubt dem Administrator, den für bestimmte Anwendungen erforderlichen Service-Level zu quantifizieren und definiert in diesem Zusammenhang unterschiedliche Prioritätsklassen. Macht man das, benötigt man auch Traffic Management. Hier geht es ebenfalls um Congestion Control und Congestion Notification, aber aus einer anderen Perspektive als bei DCB. Während dort Priorisierung und diese Verfahren eingesetzt werden müssen, um den Ethernet-Verkehr einigermaßen zuverlässig zu gestalten, hat ja FC einen genau umgekehrten inneren Mechanismus: Daten werden erst dann übertragen, wenn der Empfänger bestätigt, dass er sie aufnehmen kann. Eigentlich reicht deshalb ein Rate Limiter in Verbindung mit hinreichend großen Pufferspeichern völlig aus, um Stausituationen in FC-Fabrics zu vermeiden, aber im Zuge des scharfen Wettbewerbs muss man das natürlich dramatischer formulieren. Bei einfacheren FC-Netzen kann es schon mal zu Congestion kommen, wenn aufgrund falscher Konfiguration Server, die zu viel Bandbreite verbrauchen, zusammen auf einen Director zugreifen wollen. (siehe Abbildung 11)

Brocade sieht offensichtlich auch das Problem, dass diese Situation ungewollt auftreten kann, wenn virtuelle Maschinen über physikalische Maschinen wandern. Dann muss es im Kern-Netz eine Automatik geben, die das nachverfolgt und ggf.

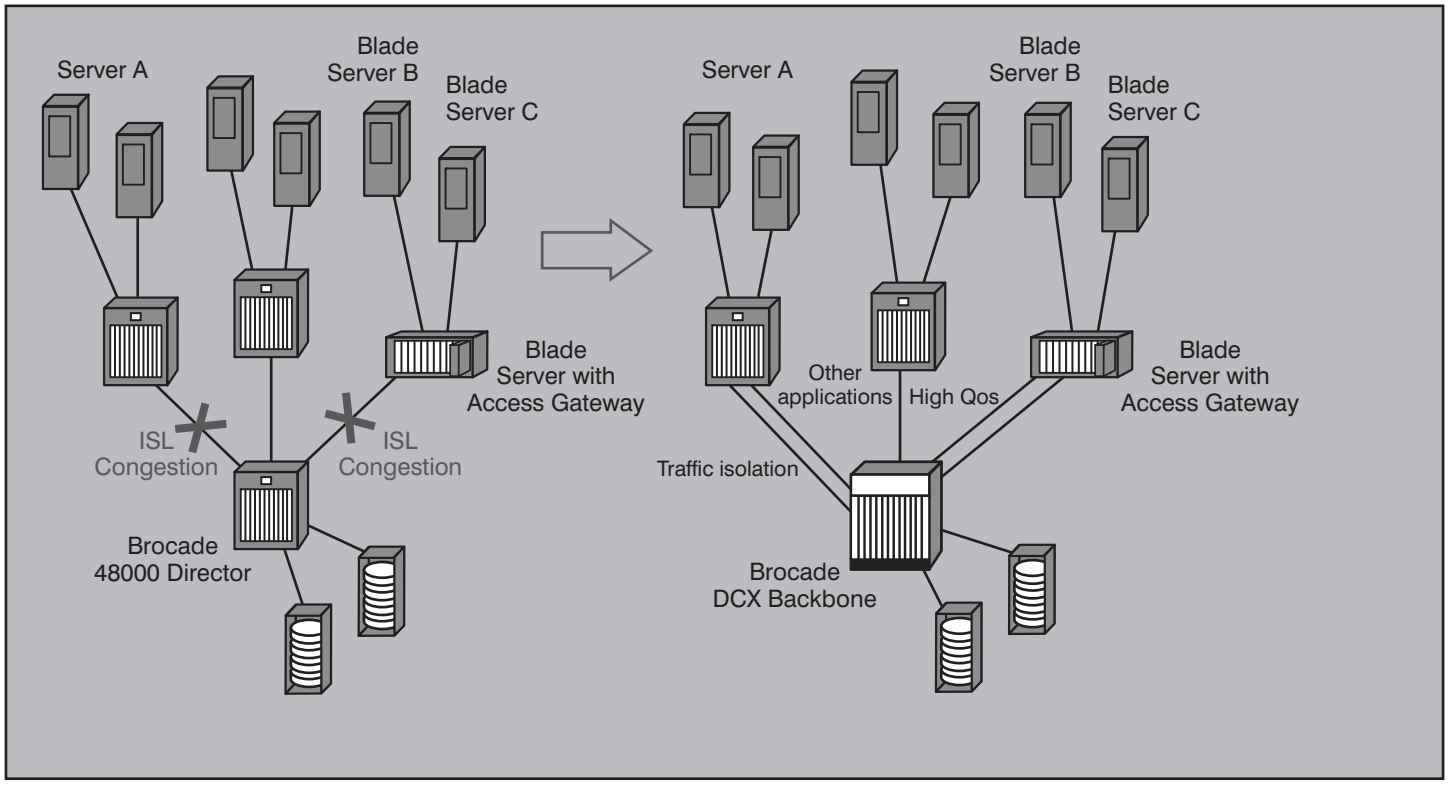


Abbildung 11: Vermeidung von ISL-Congestion mit DCX

Quelle: Brocade

Der Kampf ums RZ: die nächste Runde - Teil 2: Matrix, FCoE und EoFC

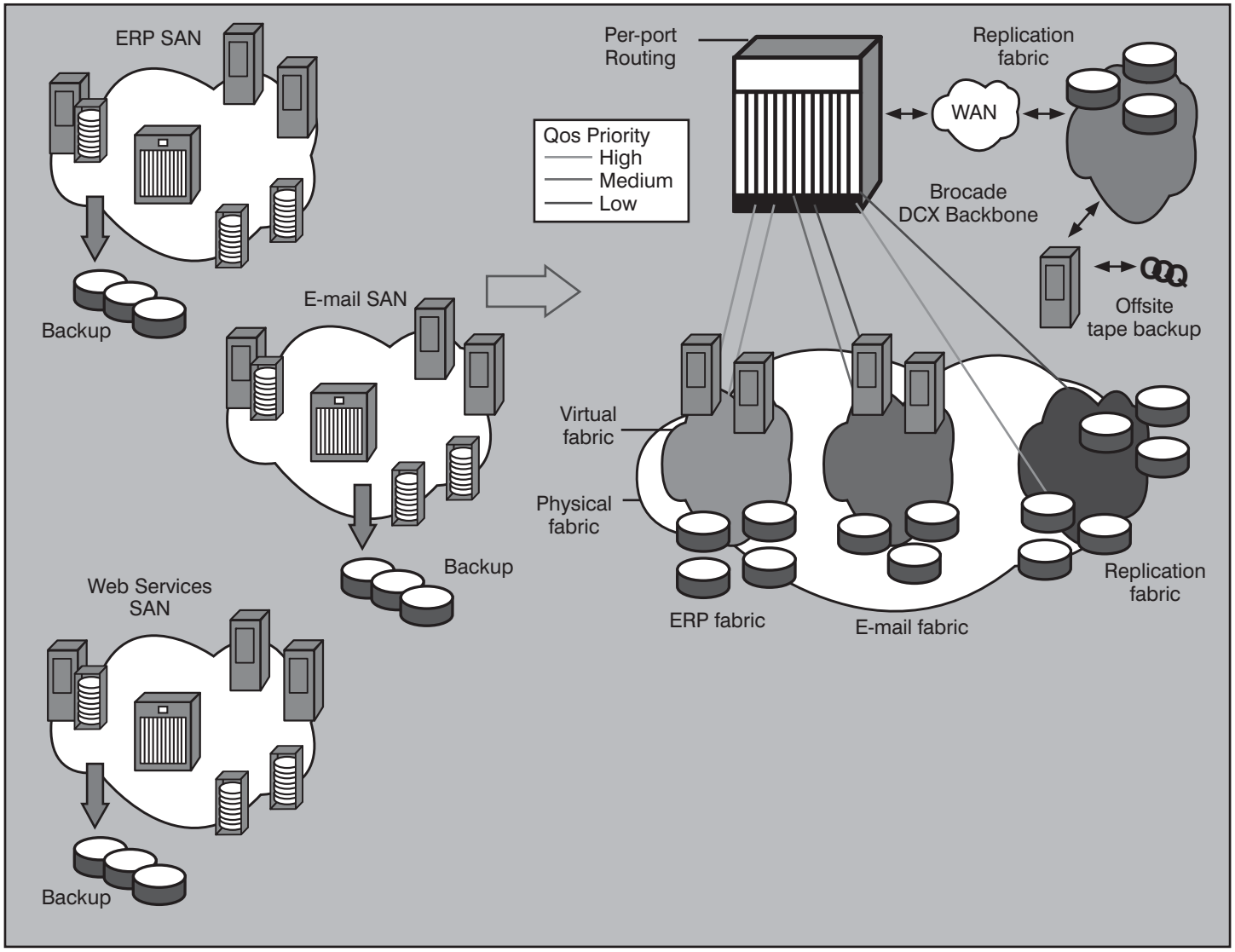


Abbildung 12: DCX-Backbone mit Virtual Fabrics

Quelle: Brocade

die Rate, mit der die auf einem physikalischen Server sitzenden virtuellen Maschinen Daten austauschen wollen, soweit begrenzt, dass der Kern geschützt wird. Da IEEE DCB auch dieses Problem noch nicht abschließend behandelt hat, muss ein Hersteller zu einer eigenen Lösung greifen. Die DCX-Directoren sind jedenfalls dafür gerüstet.

Es gibt bedingt durch die historische Entwicklung eine Reihe von Eigenheiten bei Speichersystemen und SANs, die wir hier nicht weiter aufgreifen wollen. Diese haben aber in der Vergangenheit dazu geführt, dass für verschiedene Anwendungen, z.B. ERP, Email oder Web-Services unterschiedliche SANs aufgebaut wurden. Um diese nunmehr in ein System zu integrieren, unterstützt DCX das so genannte portbasierte SAN-Routing. Es hat weniger mit Routing in dem Sinne zu tun, wie wir es gewohnt sind, sondern dient vielmehr der Abbildung der unterschiedlichen älteren Formate ineinander. Wer das Problem hat, wird dies zu schätzen wissen, für uns ist es aber nur insofern interessant, dass es prinzipiell die Möglichkeit bietet, auch andere artfremde Kommunikation, wie Ethernet-Päckchen oder FICON-Datenströme (FICON ist der leistungsfähigere Nachfolger des IBM ESCON), einfach zu integrieren. Darum werden Sie bei den Ausführungen des Herstellers auch nirgendwo das Kürzel EoFC finden. Das portbasierte SAN-Routing kann mit Priorisierung verbunden werden.(siehe Abbildung 12)

Beeindruckend ist, dass man in einen DCX-Director einfach eine oder mehrere Verschlüsselungskarten einstecken kann. Sie müssen nicht alle Datenströme verschlüsseln, sondern nur die, auf die es ankommt. Sie sind für eine inline-Verschlüsselung für 8 GbFC oder 10 GbE mit minimaler Latenz ausgelegt. Eckdaten sind:

- AES 256 oder DataFort Verschlüsselung
- 16 Ports für 8 GbFC
- Bis zu vier Blades pro DCX-Gerät

Brocade gibt in Veröffentlichungen mittlerweile an, auch FCoE unterstützen zu können. Das ist auch durchaus glaubhaft, weil sie genau das mit Fibre Channel Switches implementieren. Die DCX Direktoren sind ganz klar reine FC-Switches. Kommt dann ein FCoE-Paket an, gibt es mehrere Möglichkeiten: man kann dieses Paket einfach nochmal in ein FC-Paket einpacken, also FCoEoFC, doppelt gemoppelt hält besser. Man könnte aber auch das FCoE-Paket auspacken und den Inhalt als FC-Paket weiterleiten. Wie auch immer: der

Der Kampf ums RZ: die nächste Runde - Teil 2: Matrix, FCoE und EoFC

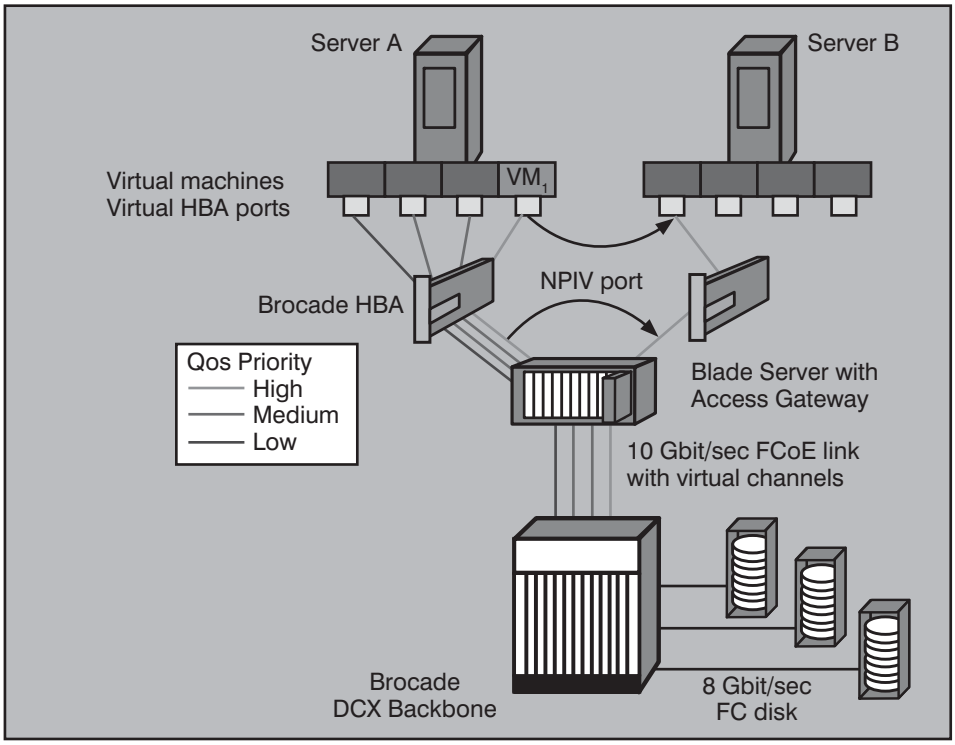


Abbildung 13: Behandlung wandernder Virtueller Maschinen in DCX Quelle: Brocade

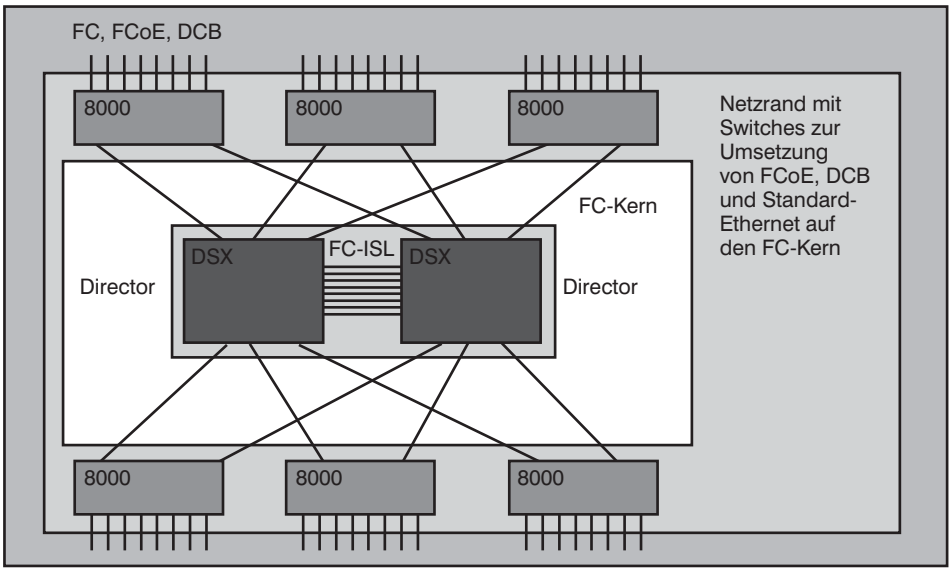


Abbildung 14: RZ-Kern-Netz auf FC-Basis mit Brocade

Inhalt wird deterministisch, verlustfrei und auf Wunsch auch noch isochron behandelt. Also, die beste Implementierung von FCoE ist die mit FC-Switches.

Es gibt in diesem Zusammenhang noch ein weiteres Problem: die wandernden Virtuellen Maschinen. Auf einem Server A werden virtuelle Maschinen definiert. Diese bekommen natürlich auch virtuelle HBAs, die wiederum auf einen realen HBA abgebildet werden. In diesem Zusammenhang bekommen die virtuellen HBAs auch eine QoS-Zuordnung, die für sie z.B. eine bestimmte Bandbreite bedeutet. Nun gibt es neuere Konzepte von VMware, die vorsehen, dass eine virtuelle Maschine auf einen anderen physikalischen Server, sagen wir Server B, wandern kann. Das ist ein Problem für das Netzwerk. In Zusammenhang mit FC regelt Brocade das so, dass für jede Anwendung, die vermöge eines virtuellen HBAs einen Virtuellen FC N-Port hat, eine N-Port ID Virtualisation NPIV erzeugt wird, mit der man diesen Port und seine assoziierten Eigenschaften (QoS, Bandbreite) eindeutig beschreiben kann. Wandert nun eine Anwendung vom Server A auf den Server B, wandert NPIV mit und wird auf der Zielmaschine der Wanderung vom dort vorhandenen HBA wieder so etabliert, dass die assoziierten Eigenschaften erfüllt werden. So ist es z.B. gar kein Problem, wenn eine Anwendung nach der Wanderung wieder auf die gewohnten Speicher-Ressourcen zugreifen möchte.

Eine Übertragung dieses Mechanismus auf FCoE-Ports ist durchaus möglich, wird aber vom Hersteller z.Zt. nicht angesprochen. (siehe Abbildung 13)

Im letzten Absatz habe ich davon gesprochen, dass man einen RZ-Netz-Kern mit FC-Directoren bauen kann und dazu einen Rand gestalten muss, der die Übergänge zu den angeschlossenen Systemen schafft.

Zu diesem Zweck drängt sich der Brocade 8000 ToR-Switch auf. Das ist ein kompaktes Multiprotokoll-Gerät für Fibre Channel, FCoE, CEE (DCB) und gewöhnliches Ethernet. Er hat 8 FC-Ports mit jeweils echten 8 GbFC und 24 10 GbE-CEE-Ports. Er kann vollständig in die DCX-Fabric integriert werden und unterstützt damit auch ISLs. Er unterstützt Zoning mit Access Control Lists und Active Directory mit LDAP. Neben der Standard IEEE 802.3ad Link Aggregation auf der Ethernet-Seite unterstützt er auch die FC-Aggregation von bis zu 8 Leitungen mit bis zu 64 Gbit/s.

Kommen FCoE-Pakete an, werden diese ausgepackt und zielgerichtet weitergeleitet, also die intelligentere Variante. Ansonsten werden alle von SNIA festgelegten Standards erfüllt. Der 8000 implementiert auch schon das im letzten Artikel angesprochene DCBX für die CEE/DCB-Seite mit Priority-based Flow Control IEEE 802.1Qbb und Enhanced Transmission selection ETS. Dazu kommt natürlich das gesamte gewöhnliche Ethernet-Protokolluniversum mit STP, MSTP, RSTP, VLANs, QoS usf. Sie brauchen also auf nichts zu verzichten, was Sie bisher gewohnt sind.

Neuerdings geben sie sogar die Emmisionswerte an: der Brocade 8000 emittiert 1,05 kg CO2 pro Gigabit/sec. pro Jahr. (siehe Abbildung 14)

Der 8000 ist nur ein Beispiel für ein mögliches Gerät am Rand des FC-Kerns. Durch die Übernahme von Foundry gibt es auch größere Varianten. Die kennen Sie bereits als NetIron, nur dass diese jetzt auch hinsichtlich der FC-Unterstützung so auf-

gerüstet wurden, dass sie an einen DCX-Kern angegliedert werden können.

6. Konsequenzen

Die Mitte 2008 aufgekommene Werbebotschaft „werft den teuren FC weg, nehmt jetzt FCoE“ vernachlässigt wichtige Entwicklungen im Bereich FC. Durch die im ersten Teil dargestellten Erweiterungen ist DCB zwar auf einem richtigen Weg. Eine wirkliche Garantie für eine zufriedenstellende dauerhaft deterministische Funktion von FCoE in einer größeren Umgebung gibt es nicht, auch wenn es in übersichtlichen Szenarien funktioniert.

Für die Entwicklung eines konvergierten RZ-Kerns stehen jetzt also zwei Alternativen zur Verfügung:

- Ethernet-Kern mit FCoE/DCB
- Fibre Channel Kern mit EoFC

Die Verfechter des Ethernet-Kerns setzen vor allem auf zwei Argumente: Ethernet ist billiger als FC und Ethernet verfügt über ein nachgelagertes Protokolluniversum, welches man seit vielen Jahren gewohnt ist, und welches es sonst in dieser Art nicht gibt. Die Integration des FC-SANs wird als notwendiges Übel betrachtet und soll mit FCoE abgefackelt werden.

Bei der Entscheidung sollte man bedenken, dass Fibre Channel die deutlich bessere Übertragungstechnik ist und war. Heute kann man folgende Eckdaten ausmachen:

- Echte 8 oder 10 Gbps auf den Links
- Trunking mit bis zu 64 Gbps
- ISL mit bis zu 512 Gbps
- Extrem schnelle Umschaltzeiten weit unter 50 msec
- Möglichkeit isochroner Übertragung
- Integration aller wesentliche modernen optischen Übertragungsalternativen
- Überwindung von bis zu 10 km ohne weitere Geräte

Genau wie 40 GbE die nächste Evolutionsstufe von Ethernet ist, wird 32 GbFC die nächste FC-Stufe sein.

Mancher wird jetzt einwenden, dass in dieser Liste nicht berücksichtigt ist, dass es auch die ER-Schnittstelle bei 10/40 GbE gibt, die doch 40 km überwinden soll. Ich hatte schon in einem früheren Artikel darauf hingewiesen, dass ein FC-Signal plesiochron (oder in der Stufe 5 sogar isochron) ist. Packt man es mit FCoE noch in Ethernet-Pakete ein, wird die Toleranz der Plesiochronität schon so massiv beansprucht, dass man damit nur geringe Entfernungen im Bereich unter 1 km überwinden kann. Lässt man es weiter laufen, fällt das Signal buchstäblich auseinander. Eine theoretische Reichweite von 40 km für Ethernet ist daher nicht relevant für den Speicherverkehr.

Für einen Ethernet-Kern spricht die größere Auswahl hinsichtlich der Hersteller.

Allerdings gibt es auch hier wieder eine neue Entwicklung. Die Hersteller Cisco Systems, Brocade und QLogic haben gesehen, dass die Bemühungen von IEEE 802.1 DCB Mitte 2009 immer noch nicht zu einem wirklich verlustfrei arbeitenden Ethernet geführt haben, was man ja für die Realisierung von FCoE dringend benötigt. Andererseits drängt der Markt nach einer Lösung. Durch eine Standardisierung im Rahmen von INCITS, die letztlich zu einem ANSI-Standard führen wird, haben sie die Problematik entkoppelt. Der neue Standard FC-BB-5 (Fibre Channel Back Bone) umfasst generelle Möglichkeiten zum Transport von FC-Daten über andere Netze. Neben FCIP, einer generalisierten Form für WAN-Übertragung und Pseudowire PW wurde auch FCoE betrachtet. Dabei wurde FCoE so definiert, dass es eine Reihe dauernd laufender Kontrollprozesse mit hoher Reaktionsgeschwindigkeit gibt, die Fehler im Ethernet insofern abfangen, dass die kommunizierenden FC-Komponenten davon nicht negativ beeinflusst werden. Im Extremfall bleibt die FC-Kommunikation einfach kurzzeitig stehen, was ihr an und für sich nichts ausmacht. Gleichzeitig werden die Szenarien, in denen FCoE überhaupt eingesetzt werden kann, eingeschränkt. Mit diesen Erweiterungen muss auch der größte FCoE-Skeptiker (wie der Autor) konstatieren, dass man FCoE in passenden Umgebungen sinnvoll einsetzen und extrem risikoarm laufen lassen kann.

Es gibt eine Reihe von Fragen, die immer wieder auftauchen, wie „brauche ich ein Produkt, wie den Nexus 1000 V, der den Hypervisor ersetzt ?“, „was muss ein HBA in Zukunft leisten, um virtuelle NICs zu unterstützen?“ oder „wo sind die konzeptionellen Unterschiede zwischen dem HP und dem Cisco-Ansatz?“ und schließlich „wozu brauche ich das alles?“

Die letzte ist am einfachsten zu beantworten: ich brauche das immer genau dann, wenn neben dem Ethernet eine gewachsene FC-SAN-Infrastruktur existiert, die eine Reihe von speicherorientierten Zusatzfunktionen unterstützt, die man gerne benutzt und in die man Investitionen getätigt hat, die es zu schützen gilt.

Die konzeptionellen Unterschiede zwischen Cisco und HP, Brocade oder anderen Herstellern werden hinsichtlich der Technik durch die Implementierung von FC-BB-5 wesentlich minimiert und integriert. In diesem gibt es ein so genannte ENode-Element, welches die Aufgaben eines HBAs klärt. Wir behandeln FC-BB-5 in einem weiteren Artikel. Ein virtueller Switch wie der Nexus 1000V wird eigentlich nur dann benötigt, wenn der Hypervisor diese Aufgaben nicht selbst erledigen kann. Meiner Ansicht nach muss das in Zukunft integriert werden, hier sind sicherlich auch die neuen Ansätze von Citrix sehr interessant, es mag zur Integration allerdings auch Alternativen geben. Das werden wir dann sehen.

Das Beispiel der HP Blade System Matrix hat gezeigt, welche Erwartungen Rechner der aktuellen Generation an Infrastruktur-Netze stellen. Man kann hier keinesfalls mit einer Abschaffung des FC-SANs aus dieser Richtung rechnen.

Wie auch immer die weitere Entwicklung aussehen wird, es bleibt spannend.

ComConsult Veranstaltungskalender

Aktuelle Veranstaltungen

Lokale Netze für Einsteiger, 31.08. - 04.09.09 in Frankfurt

Dieses Seminar vermittelt kompakt und intensiv innerhalb von 5 Tagen die Grundprinzipien des Aufbaus und der Arbeitsweise Lokaler Netzwerke. Dabei werden sowohl die notwendigen theoretischen Hintergrundkenntnisse vermittelt als auch der praktische Aufbau und der Betrieb eines LANs erläutert. Ausgehend von einer Darstellung von Themen der Verkabelung und der grundlegenden Übertragungsprotokolle werden die wichtigen Zusammenhänge zwischen der Arbeitsweise von Switch-Systemen, den darauf aufsetzenden Verfahren und der Anbindung von PCs und Servern systematisch erklärt.

Preis: € 2.290,-- zzgl. MwSt.

IP-Wissen für TK-Mitarbeiter, 07.09. - 08.09.09 in Aachen

Dieses Seminar vermittelt kompakt und effizient das IP-Wissen, das TK-Mitarbeiter ohne Vorkenntnisse zur Planung und zum Betrieb von IP-basierten Telefonie-Lösungen benötigen. Alle Seminarinhalte werden von einem Referenten mit hoher Praxiserfahrung betreut. Ziel ist dabei bewusst, statt einer umfassenden Theorieschulung gezielt die Aspekte vorzustellen und unter Praxis-relevanten Gesichtspunkten zu beleuchten, die erfahrungsgemäß aus Sicht einer IP-basierten Telefonielösung wichtig sind. Preis: € 1.390,-- zzgl. MwSt.

IP-Telefonie und Unified Communications erfolgreich planen und umsetzen, 14.09. - 16.09.09 in Köln

Dieses Seminar behandelt die Projektschritte, Einsatz- und Migrations-Szenarien, einsetzbare Basis-Technologien, Komponenten und erweiterte TK-Anwendungen, Bewertungskriterien für eine TK-Lösung und gibt eine Übersicht über den bestehenden TK-Markt etablierter Hersteller wie Alcatel-Lucent, Avaya, Cisco, Nortel und Siemens aber auch des Newcomers Microsoft.

Preis: € 1.690,-- zzgl. MwSt.

Virtualisierungstechnologien in der Analyse, 14.09. - 15.09.09 in Köln

Dieses Seminar analysiert die verfügbaren Virtualisierungstechnologien der führenden Anbieter. Sie lernen, welche Gestaltungselemente virtuelle Umgebungen haben, angefangen von einfachen und überschaubaren Lösungen bis hin zu komplexen und umfassenden Rechenzentrums-Gesamt-Architekturen. Dabei wird auch der Bedarf an Infrastruktur-Leistung insbesondere auf der Netzwerkseite untersucht.

Preis: € 1.390,-- zzgl. MwSt.

Sicherheit 1: Grundlagen und Kernbausteine zur erfolgreichen IT-Sicherheit, 14.09. - 18.09.09 in Köln

Bedrohungen der IT-Sicherheit bestehen für praktisch alle Elemente einer vernetzten IT-Infrastruktur. Die Quelle der Bedrohung kann sowohl von außen auf das Netz wirken als auch von innen stammen. Sicherheit entsteht erst, wenn alle signifikanten Gefahrenbereiche systematisch verriegelt werden. Dieses Seminar identifiziert die wesentlichen Gefahrenbereiche und zeigt effiziente und wirtschaftliche Maßnahmen zur Umsetzung einer erfolgreichen Lösung auf. Dabei wird jeder einzelne Baustein detailliert erklärt und anhand typischer Einsatzszenarien wird der Weg zu einer erfolgreichen Sicherheits-Lösung aufgezeigt.

Preis: € 2.290,-- zzgl. MwSt.

Sicherheit im LAN mit IEEE 802.1X, 21.09. - 22.09.09 in Köln

Dieses 2-tägige Seminar vermittelt den optimalen Umgang mit IEEE 802.1X, erläutert die Einsatzvarianten, beschreibt die gegebenen Fallstricke und liefert die ideale Basis zur Vorbereitung eines Einsatzes.

Preis: € 1.390,-- zzgl. MwSt.

Projekt-Erfahrungsbericht: Cisco CallManager Rollout und Migration CUCM Version 6, 21.09. - 22.09.09 in Köln

Dieses 2-tägige Seminar beschreibt Planung, Installation und den Betrieb einer großen verteilten IP-Telefonie-Lösung auf der Basis des Cisco CallManagers. Es macht deutlich, in welchem Umfang die Standard-Installation angepasst und erweitert werden musste, um den Anforderungen der Teilnehmer zu entsprechen. Auch die Umstellung traditioneller Betriebsabläufe im Änderungs-Management und deren Auswirkung auf die Konfiguration des CallManagers wird beschrieben. In diesem Zusammenhang werden insbesondere auf die Akzeptanz der Benutzer und die damit notwendigen Änderungen in der Bedienung der Telefone eingegangen.

Preis: € 1.390,-- zzgl. MwSt.

Unified Communications mit Siemens - HiPath 8000 & OpenScape im Überblick, 21.09. - 22.09.09 in Köln

Mit der Zusammenführung der rein SIP-basierten TK-Lösung HiPath 8000 und der Applikation-Suite OpenScape präsentiert Siemens ein umfangreiches Kommunikationsprodukt, das verspricht, im Sinne von Unified Communications alle modernen Kommunikationstechnologien unter einer gemeinsamen Struktur für den Endanwender steuerbar und nutzbar zu machen. So wurden neben der in der Tradition der bekannten HiPath-Telefonanlagen stehenden Sprachlösung weitere Dienste und Leistungsmerkmale wie Präsenzanzeige, Erreichbarkeitsanzeige, regelbasierte automatische Steuerung der Erreichbarkeit, Instant Messaging, Fax und E-Mail sowie Webkollaboration und Videokonferenzsysteme integriert.

Preis: € 1.390,-- zzgl. MwSt.

TCP/IP und SNMP, 21.09. - 25.09.09 in Köln

LAN-, WLAN- und WAN-Netzwerke sind heutzutage IP-Netze, und ein Verzicht auf Nutzung des IP-basierten Internet undenkbar. Auch für früher nur mit herstellerspezifischen Protokollen in Verbindung gebrachte Anwendungsgebiete wie Telefonie oder Produktionsumgebungen gibt es mittlerweile geeignete IP-basierte Lösungen. Hersteller und Dienstanbieter versuchen den Eindruck zu vermitteln, die Nutzung sei kinderleicht, fast schon plug and play - man trägt ein paar Adressen ein (wenn überhaupt), und es kann losgehen. Falsch!

Preis: € 2.290,-- zzgl. MwSt.

Zertifizierungen

ComConsult Certified Network Engineer

Lokale Netze
31.08. - 04.09.09 in Frankfurt
23.11. - 27.11.09 in Hamburg

TCP/IP und SNMP
21.09. - 25.09.09 in Bonn

Internetworking
05.10. - 09.10.09 in Frankfurt

Paketpreis für alle drei Seminare € 6.183,-- zzgl. MwSt. (Einzelpreise: je € 2.290,--)

ComConsult Certified Trouble Shooter

Trouble Shooting 1
06.10. - 09.10.09 in Aachen

Trouble Shooting 2
03.11. - 06.11.09 in Aachen

Paketpreis für beide Seminare, eine digitale Stromzange, die Prüfung und den Report „Fehlersuche in konvergenten Netzen“ € 4.120.-- zzgl. MwSt.
(Seminar-Einzelpreis € 2.190,--, mit Prüfung € 2.370,--)

ComConsult Certified Security Expert

Sicherheit 1: Grundlagen und Kernbausteine zur erfolgreichen IT-Sicherheit
14.09. - 18.09.09 in Köln

Sicherheit 2: Erarbeitung und Umsetzung von Sicherheitskonzepten
26.10. - 30.10.09 in Aachen

Sicherheit 3: Praxis-Intensiv-Seminar zur erfolgreichen Konfiguration von Firewalls, VPNs, Windows Clients und WLANs
23.11. - 27.11.09 in Aachen

Paketpreis für alle drei Seminare und die beiden Reports „VPN-Technologien: Alternativen und Bausteine einer erfolgreichen Lösung“ und „Sicherheit in Wireless LANs“. € 6.183,-- zzgl. MwSt. (Einzelpreise: je € 2.290,--)

ComConsult Certified Voice Engineer

Basis-Seminar: Session Initiation Protocol-Basis-Technologie der IP-Telefonie
28.09. - 30.09.09 in Bad Neuenahr
23.11. - 25.11.09 in Hamburg

Basis-Seminar: Sicherheitsmechanismen für Voice over IP
05.10. - 06.10.09 in Frankfurt

Alternative 1: IP-Telefonie und Unified Communications erfolgreich planen und umsetzen
14.09. - 16.09.09 in Köln
02.11. - 04.11.09 in Frankfurt

Alternative 2: IP-Telefonie: Vorbereitung, Migration, Management
26.10. - 28.10.09 in Berlin

Optionales Einsteiger-Seminar: IP-Wissen für TK-Mitarbeiter
07.09. - 08.09.09 in Aachen
09.11. - 10.11.09 in Königswinter

Basis-Paket Alternative 1: Beinhaltet die zwei Basis-Seminare und Seminar „Alternative 1“
Grundpreis: € 4.250,-- zzgl. MwSt. statt € 4.770,-- zzgl. MwSt.

Basis-Paket Alternative 2: Beinhaltet die zwei Basis-Seminare und Seminar „Alternative 2“
Grundpreis: € 4.250,-- zzgl. MwSt. statt € 4.770,-- zzgl. MwSt.

Optionales Einsteigerseminar: Aufpreis € 990,-- zzgl. MwSt. statt € 1.390,-- zzgl. MwSt.

Impressum

Verlag:
ComConsult Technology Information Ltd.
ComConsult Research
64 Johns Rd
Christchurch 8051
GST Number 84-302-181
Registration number 1260709
German Hotline of ComConsult-Research:
02408-955300

E-Mail: insider@comconsult-akademie.de
http://www.comconsult-research.de

Herausgeber und verantwortlich
im Sinne des Presserechts:
Dr. Jürgen Suppan
Chefredakteur: Dr. Jürgen Suppan
Erscheinungweise: Monatlich,
12 Ausgaben im Jahr

Bezug: Kostenlos als PDF-Datei
über den eMail-VIP-Service
der ComConsult Akademie