

# DATA CENTER ETHERNET: DIE CISCO INNOVATION FÜR DATA CENTER-NETZWERKE

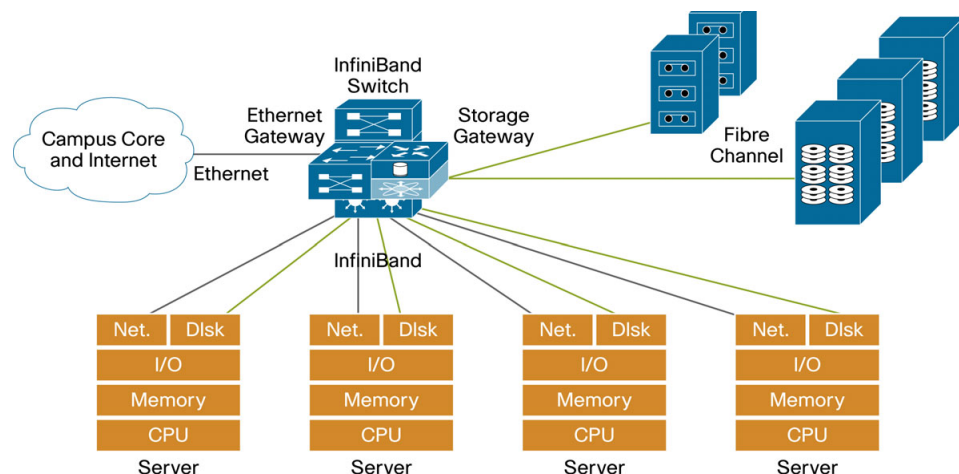
## Was Sie hier erfahren

Cisco® hat Ethernet-Erweiterungen auf der Basis von Standards entwickelt, und zwar speziell für Data Center-Netzwerke, um neue Funktionen unterstützen zu können. Dieses Dokument informiert Sie über die kontinuierlichen Innovationen von Cisco im Bereich Ethernet Data Center-Netzwerke, um hier die Anforderungen neuer Anwendungen zu erfüllen. Es geht hier um die Erweiterungen der Cisco Data Center Ethernet-Architektur, mit denen die Belastbarkeit von Ethernet verbessert und neue Konzepte wie Unified I/O und Unified Fabric unterstützt werden. Außerdem erfahren Sie hier, wo andere Ansätze sich als unzulänglich erwiesen haben und warum Cisco Data Center Ethernet ganz wie die Architektur aussieht, die den Anforderungen der nächsten Generation von Ethernet-Netzwerken im Data Center entsprechen wird.

## Einleitung

Ethernet ist in der Regel das Netzwerk der Wahl für die Vernetzung von Ressourcen im Data Center. Es ist überall zu finden und wird von Netzwerkingenieuren und Entwicklern auf der ganzen Welt verstanden. Ethernet hat sich über lange Zeit bewährt und sich gegen viele Konkurrenten durchgesetzt, die es als System der Wahl für Netzwerkumgebungen in Rechenzentren ablösen wollten. Die neuesten sich abzeichnenden Anforderungen der Anwendungen setzen zusätzliche Fähigkeiten in den Netzwerkinfrastrukturen voraus, was oft zur Implementierung von mehreren separaten und anwendungsspezifischen Netzwerken führt. Häufig implementieren Data Center in großen Unternehmen ein Ethernet-Netzwerk für IP-Verkehr, ein oder zwei SANs (Storage Area Networks) für Block Mode Fibre Channel-Verkehr und ein InfiniBand-Netzwerk für Hochleistungs-Cluster-Computing (Abb. 1). Die Kapital- und Betriebskosten für Implementierung und Management von drei verschiedenen Netzwerktypen sind hoch. Hieraus ergibt sich die Gelegenheit zur Konsolidierung auf ein Unified Fabric.

**Abbildung 1:** Drei verschiedene Data Center-Netzwerke



Werden die drei Arten von Netzwerken technisch evaluiert, so zeigt Ethernet das größte Potenzial, um den meisten Anforderungen aller drei Netzwerktypen zu entsprechen. Es sind jedoch einige zusätzliche Fähigkeiten notwendig. Netzwerkkonsolidierung mit Ethernet ist einer der wichtigsten geschäftlichen Vorteile dieses Ansatzes. Die Cisco Data Center Ethernet-Architektur bietet Ethernet-Erweiterungen auf der Basis von Standards. Diese eröffnen die Möglichkeit, Netzwerk-Infrastrukturen auf ein Unified Fabric zu konsolidieren.

Bei Cisco Data Center Ethernet handelt es sich um eine Reihe von Erweiterungen der Ethernet-Architektur, mit denen die Rolle von Ethernet-Networking und -Management speziell im Data Center verbessert und erweitert werden soll. Die Cisco Data Center Ethernet-Architektur hat drei wichtige Aspekte: Erweiterungen des Ethernet mit Unterstützung von I/O-Konsolidierung auf ein Unified Fabric, mit Trennung und Beibehaltung von separaten Verkehrsklassen über das Fabric hinweg; Unterstützung eines „No-Drop Service“, so dass Verkehr, der garantiert übermittelt werden muss, jetzt über Lossless Fabrics transportiert werden kann; außerdem mehr bisektionale Bandbreite durch Aktivierung von Multipathing auf Layer 2.

Einer der geschäftlichen Vorteile der Netzwerkkonsolidierung ist Kosteneinsparung. Eine homogene Ethernet-Infrastruktur wäre unter betrieblichen Aspekten einfacher, da sie die existierenden Fähigkeiten von Ethernet-Netzwerkingenieuren nutzen könnte. Es wären weniger Management-Tools notwendig, und die Einführungszeit für neue Netzwerke könnte verkürzt werden. Außerdem würde ein konsolidiertes Cisco Data Center Ethernet-Netzwerk alle existierenden Funktionen des Layer 2-Netzwerks bieten, das es ersetzt. Im Falle von Ethernet umfasst dies Multicast- und Broadcast-Verkehr, VLANs, Link Aggregation etc.; für Fibre Channel umfasst dies die Bereitstellung aller Fibre Channel Services, wie Zoning und Name Server, sowie die Unterstützung von virtuellen SANs (VSANs) Inter-VSAN Routing (IVR) etc.

Diese Erweiterungen der klassischen Ethernet-Fähigkeiten führen zu einer Netzwerk-Infrastruktur im Data Center, die zu Folgendem in der Lage ist:

- Unterstützung von Multiprotokoll-Transport über ein Unified Fabric
- bietet Skalierbarkeit für größere Layer 2 Domains
- Steigerung der verfügbaren Bandbreite im Netzwerk durch Aktivierung existierender physikalischer Routen, die gerade nicht genutzt werden

### **Die Entwicklung des Data Center-Netzwerks**

Ethernet und auch das Data Center-Netzwerk entwickeln sich immer weiter. Gesteuert wird das heute durch die Art und Weise, wie Anwendungen das Netzwerk als Ressource verwenden. Die Anforderungen des Netzwerks haben sich verändert – es wird nicht länger einfach nur für traditionelle Client-to-Server-Transaktionen verwendet. So nimmt z.B. die Implementierung von Server-Clustern zu, was zu gesteigertem Server-to-Server-Verkehr führt. Grid Computing hat ebenfalls dazu beigetragen, den Verkehr von Server zu Server zu intensivieren. Höhere, periodische Storage Backup-Anforderungen haben zu mehr Datenverkehr zwischen Server-Farmen und Storage Devices auf SANs geführt. Außerdem sind Serverless Backups zwischen Storage Devices heute ganz alltäglich, was den Disk-to-Disk- und den Disk-to-Tape-Verkehr erhöht. Netzwerkverkehr im Data Center wird jetzt von Client zu Server, von Server zu Server, von Server zu Storage und von Storage zu Storage übertragen.

Diese Zunahme des allgemeinen Netzwerkverkehrs und die Veränderung der Verkehrsmuster hat dazu geführt, dass man sich mehr und mehr auf das Netzwerk verlässt, um den notwendigen Durchsatz für die Unterstützung von Server Cluster-Anwendungen zu liefern. Anwendungsleistung wird jetzt gemeinsam mit Netzwerkleistung gemessen, sodass Bandbreite und Latenz wichtig sind. Auch bei den verschiedenen Verkehrsarten gibt es Unterschiede. Client-to-Server- und Server-to-Server-Transaktionen umfassen kurze, stoßartige Übertragungen. Die meisten Server-to-Storage- und die reinen Storage-Anwendungen erfordern lange, gleichmäßige Datenflüsse. Eine neue Data Center Ethernet-Architektur muss flexibel sein und über die notwendige Netzwerkintelligenz verfügen, um Änderungen der Netzwerkdynamik zu unterstützen, zu identifizieren und entsprechend darauf zu reagieren.

Außerdem bestehen Unterschiede in der Fähigkeit von Anwendungen, mit Packet Drops umzugehen. Packet Drops haben einmalige Auswirkungen auf verschiedene Protokolle, wobei Anwendungen auf ganz unterschiedliche Art und Weise reagieren: einige Anwendungen können Drops tolerieren und reagieren mit Neusendung. Ethernet unterstützt diese Fälle, während andere Anwendungen Paketverlust einfach nicht tolerieren können und garantierte Ende-zu-Ende-Übertragungen ohne Drops benötigen. Fibre Channel-Verkehr über Ethernet ist so ein Beispiel für eine Anwendungsanforderung nach „No Drop“-Service. Damit Ethernet-Netzwerke Anwendungen mit „No Packet Drop“-Anforderung unterstützen können, muss eine Methode für die Bereitstellung einer Lossless Service-Klasse über Ethernet etabliert werden. Cisco Data Center Ethernet-Erweiterungen für Verkehrsmanagement bieten eben diese Fähigkeit.

Data Center-Netzwerke müssen auch mit großen, flachen Designs umgehen können. Data Center-Netzwerke werden immer mehr erweitert, während die Kunden mehr und mehr Server und Switches hinzufügen. Große existierende Layer 2 Network-Domains werden somit noch größer – und das ist die Regel, nicht die Ausnahme.

### **Weitere Optionen zur Netzwerkkonsolidierung**

Ethernet ist nicht die einzige Option zur Data Center-Netzwerkkonsolidierung, bietet aber im Vergleich zu anderen Möglichkeiten die größten Erfolgchancen. Fibre Channel erfordert zuverlässigen Transport mit „No-Drop“-Service während eines Netzwerkstaus, um Neuübertragungen zu vermeiden. Eine echte Herausforderung für Ethernet war die Übertragung von Fibre Channel-Verkehr über Ethernet, und zwar „native“ und ohne Drops. PFC (Priority-based Flow Control) ermöglicht Lossless Ethernet, während FCoE (Fibre Channel over Ethernet) Native Fibre Channel-Verkapselung erlaubt.

### **iSCSI**

Ethernet-basiertes Small Computer System Interface over IP (iSCSI) wurde als Ersatz für Fibre Channel in Betracht gezogen, da es die Konsolidierung von Block Storage Transfers über Ethernet erlaubt. Obwohl iSCSI in vielen Speicheranwendungen nach wie vor sehr beliebt ist, insbesondere bei kleinen und mittleren Unternehmen (SMBs), kann es doch den weit verbreiteten Einsatz von Fibre Channel für geschäftskritische Storage Media Transfers im Unternehmen nicht übertreffen oder ersetzen. Ein Grund hierfür ist das Zögern, geschäftskritischen Storage Media Transfer einer Ethernet-Infrastruktur anzuvertrauen, die bisher keinen „No-Packet-Drop“-Service garantieren konnte. Ein weiterer Grund dafür, dass Fibre Channel nicht durch iSCSI ersetzt wurde, ist der, dass iSCSI die „native“ Fibre Channel Services oder Tools nicht unterstützt, auf die sich SAN-Administratoren verlassen.

Manche glauben, dass allein das Hinzufügen von 10 Gigabit Ethernet es iSCSI erlauben wird, FCoE zu verdrängen, und zwar rein in punkto Leistung. Aus dieser Perspektive wird 10 Gigabit Ethernet auch mehr Geschwindigkeit bieten, wenn FCoE darüber transportiert wird. Zuverlässigkeit und Integrität der Daten sind wichtiger für Fibre Channel als die Geschwindigkeit der Verbindung. Deshalb wäre Lossless Ethernet durchaus attraktiv. Eine weitere Herausforderung in punkto iSCSI war der TCP/IP Overhead, der die CPU-Nutzung auf dem Server steigert. iSCSI verlässt sich auf TCP/IP, um zuverlässigen, ordnungsgemäßen Storage-Verkehr zu bieten. TCP/IP Offload Engines wurden in Network Adapter Hardware eingesetzt, aber dieser Ansatz kann die Kosten der Schnittstelle steigern und spezielle ASICs (Application-specific Integrated Circuits) erfordern.

### **InfiniBand**

InfiniBand-Technologie wurde ebenfalls als potenzieller Kandidat für die Netzwerkkonsolidierung im Data Center positioniert. InfiniBand erfordert niedrige Latenz. Obwohl InfiniBand Gateways zu Fibre Channel- und Ethernet-Netzwerken bietet, erfordert es doch den Aufbau eines weiteren parallelen Netzwerks. Da Ethernet-Netzwerke so weit verbreitet sind, ist es unwahrscheinlich, dass IT-Abteilungen ihre Ethernet-basierten Infrastrukturen auf InfiniBand verschieben werden. Dies wäre einfach nicht kosteneffektiv. Schließlich wäre dies ein zusätzlicher Ausbau mit entsprechendem Administrationsaufwand. Von betrieblicher Seite wäre dies nur mit erheblichem InfiniBand-Training der Ethernet IP Networking-Mitarbeiter zu erreichen. Eine weitere Hürde, die InfiniBand zu überwinden hat, ist der Mangel an Vernetzung zwischen verschiedenen InfiniBand-Subnetzen. Während Datenzentren skalieren und gemeinsam genutzte Ressourcen benötigen, müssten InfiniBand-Subnetze über InfiniBand Router miteinander verbunden werden, die heute noch nicht existieren. 10 Gigabit Ethernet ist ein Standard mit höherer Bandbreite als 10-Gbps InfiniBand. Da InfiniBand 10 Bits verwendet, um 8 Datenbits zu verschlüsseln, bedeutet dies einen Verlust der Leitungsrate von 20 %. Die brauchbare Bandbreite des Data Link Layers wird somit auf 8 Gbps beschränkt. 10 Gigabit Ethernet kann 10 Gbps guten Durchsatz senden und die tatsächliche Leitungsrate eines 10 Gigabit Ethernet Links erreichen.

Wenn man bedenkt, dass 80 % aller Server Cluster heute über Ethernet-Infrastrukturen verwirklicht werden, so dürfte die wahrscheinlichste Option die sein, dass Ethernet so ausgebaut wird, dass es einen großen Prozentsatz von Server Clustern und Grid Computing-Anwendungen versorgen kann. Die Erstellung von RDMA (Remote Direct Memory Access)-Treibern für 10 Gigabit Ethernet ist eine Entwicklung, die geradezu unausweichlich erscheint. Es wird Ethernet mit niedriger Latenz und hohem Durchsatz für die direkte Vernetzung von Speicherressourcen benötigt. Die Verfügbarkeit der Lossless Transport-Klasse in Cisco Data Center Ethernet wird ebenfalls für Server Cluster-Anwendungen von Vorteil sein.

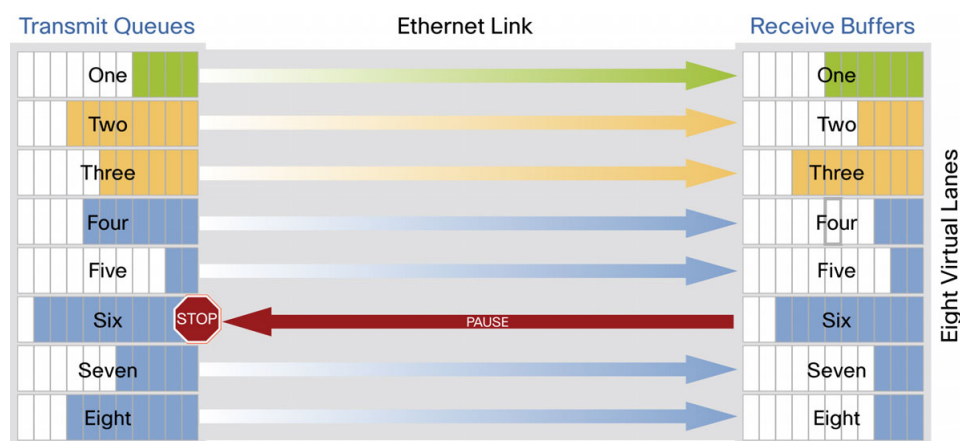
## Cisco Data Center Ethernet

Die Cisco Data Center Ethernet-Architektur wurde wohlgedacht, um klassische Ethernet-Stärken zu nutzen, mehrere entscheidende Erweiterungen hinzuzufügen und so die Infrastruktur der nächsten Generation für Data Center-Netzwerke zu bieten, und um das Unified Fabric bereitzustellen, das in der Cisco Data Center 3.0-Architektur versprochen wurde. Der Rest dieses Dokuments beschreibt die Cisco Data Center Ethernet-Architektur, eine Layer 2-Architektur für das Data Center, und schildert, wie jede der wichtigsten Komponenten der Architektur zu einem robusten Ethernet-Netzwerk beiträgt – einem Netzwerk, das dazu in der Lage ist, den wachsenden Anforderungen der Anwendungen von heute zu entsprechen und auf den zukünftigen Bedarf der Data Center-Netzwerke von morgen zu reagieren.

### Priority-based Flow Control

Die gemeinsame Nutzung von Links ist für die I/O-Konsolidierung entscheidend. Damit dieses Link Sharing erfolgreich ist, dürfen sich Häufungen oder „Bursts“ eines Verkehrstypen nicht auf andere Verkehrstypen auswirken, lange Verkehrsschlangen eines bestimmten Verkehrstypen dürfen nicht die Ressourcen anderer Verkehrstypen verbrauchen, und die Optimierung eines Verkehrstypen darf nicht zu hoher Latenz für kürzere Nachrichten anderer Verkehrstypen führen. Der Ethernet Pause-Mechanismus kann verwendet werden, um die Auswirkungen eines Verkehrstypen auf einen anderen zu kontrollieren. PFC (Priority-based Flow Control) oder die Datenflusskontrolle auf der Basis von Prioritäten ist eine Erweiterung des Pause-Mechanismus (Abbildung 2).

Abbildung 2: Priority-based Flow Control



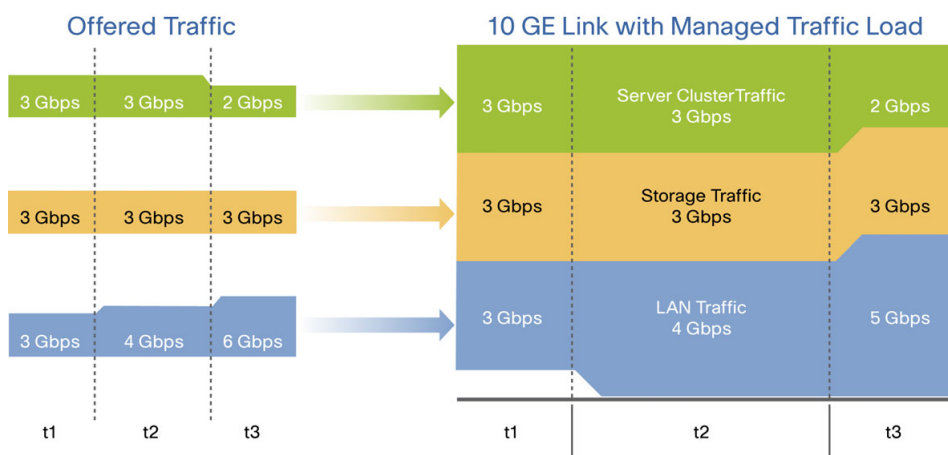
Die aktuelle Ethernet Pause-Option stoppt allen Verkehr auf einem Link; im Grunde handelt es sich hierbei um eine Link-Pause für die gesamte Verbindung. PFC erstellt acht separate virtuelle Links auf der physikalischen Verbindung und erlaubt, dass jeder dieser Links unabhängig voneinander angehalten und neu gestartet werden kann. Dieser Ansatz ermöglicht es dem Netzwerk, eine „No-Drop“-Serviceklasse für einen individuellen virtuellen Link zu erstellen, die neben anderen Verkehrstypen auf der gleichen Schnittstelle existieren kann. PFC erlaubt differenzierte QoS (Quality of Service)-Strategien für die acht einmaligen virtuellen Links. Außerdem spielt PFC eine wichtige Rolle, wenn es als Arbitrer für Intraswitch Fabrics verwendet wird und Ingress-Ports mit Egress Port-Ressourcen verbindet (siehe „Lossless Fabric“ weiter unten in diesem Dokument; IEEE 802.1Qbb und <http://www.ieee802.org/1/files/public/docs2007/new-cm-barrass-pause-proposal.pdf>).

### Enhanced Transmission Selection

PFC kann auf einer physikalischen Verbindung acht separate virtuelle Link-Typen erstellen. Es kann durchaus von Vorteil sein, innerhalb jedes virtuellen Links verschiedene Verkehrsklassen zu definieren. Verkehr innerhalb der gleichen PFC IEEE 802.1p-Klasse kann zusammengruppiert und trotzdem innerhalb jeder Gruppe unterschiedlich gehandhabt werden. ETS (Enhanced Transmission Selection) bietet priorisierte Verarbeitung auf der Basis von Bandbreitenzuweisung, niedriger Latenz oder „Best Effort“, was zu Verkehrsklassenzuweisung pro Gruppe führt. Als Erweiterung des Konzepts des virtuellen Links bietet der NIC (Network Interface Controller) virtuelle Schnittstellen-Warteschlangen: eine für jede Verkehrsklasse. Jede virtuelle Schnittstellen-Warteschlange ist für das Management der zugewiesenen Bandbreite für ihre Verkehrsgruppe zuständig, hat aber eine gewisse Flexibilität innerhalb der Gruppe, um den Verkehr dynamisch zu verwalten. So könnte z.B. der virtuelle Link 3 für die IP-Verkehrsklasse eine Bezeichnung „Hohe Priorität“ und einen „Best Effort“ innerhalb der gleichen Klasse haben, wobei die Virtual Link 3-Klasse einen Prozentsatz des Links insgesamt mit anderen Verkehrsklassen teilt. ETS erlaubt Differenzierung des Verkehrs der gleichen Prioritätsklasse, was zur Bildung von Prioritätsgruppen führt (Abb. 3).

Die heutige IEEE 802.1p-Implementierung spezifiziert eine strikte Zeitplanung der Warteschlangen auf der Basis von Priorität. Mit ETS kann ein flexibler „drop-free“ Scheduler für die Warteschlangen den Verkehr priorisieren, und zwar nach den IEEE 802.1p-Verkehrsklassen und der Hierarchie für die Verkehrsbehandlung, die in jeder Prioritätsgruppe festgelegt ist. Die Fähigkeit, verschiedene Verkehrsarten innerhalb der gleichen Prioritätsklasse unterschiedlich zu behandeln, wird durch die Implementierung von ETS aktiviert (siehe IEEE 802.1Qaz und <http://www.ieee802.org/1/pages/802.1az.html>).

**Abbildung 3:** Enhanced Transmission Selection

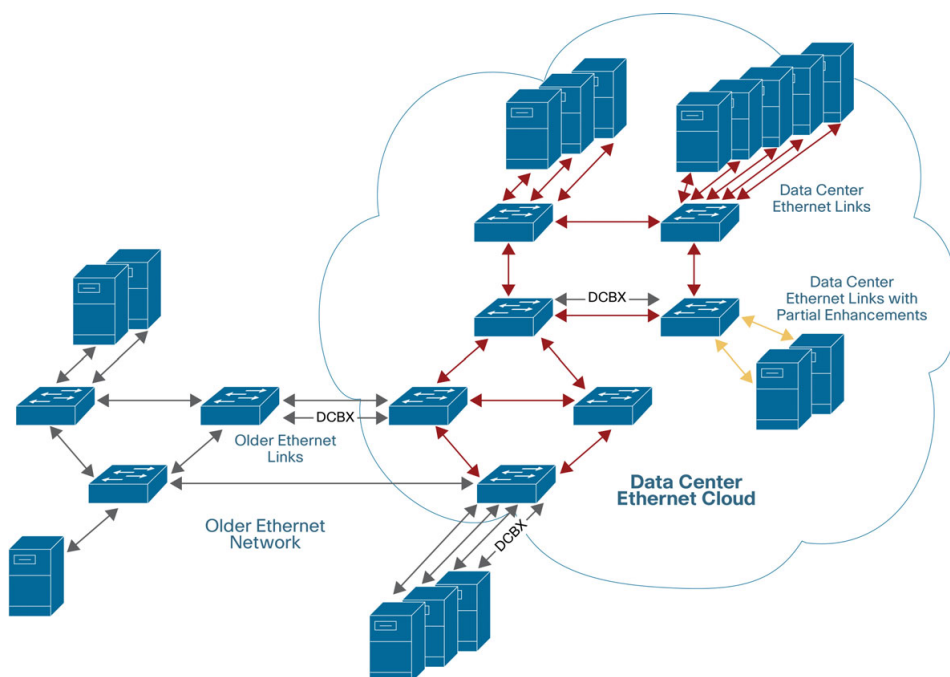


### Data Center Bridging Exchange Protocol

Das Data Center Bridging Exchange (DCBX) Protocol ist ein Discovery- und Capability Exchange-Protokoll, das von Cisco, Nuova und Intel entwickelt wurde. Es wird von Cisco DCE™-Produkten dazu verwendet, Peers in Cisco Data Center Ethernet-Netzwerken zu entdecken und Konfigurationsinformationen zwischen Cisco DCE™ Switches auszutauschen (Abb. 4). Die folgenden Parameter der Cisco Data Center Ethernet-Funktionen können mit DCBX ausgetauscht werden (siehe <http://www.ieee802.org/1/files/public/docs2008/az-wadekar-dcbcxp-overview-rev0.2.pdf>):

- Prioritätsgruppen in ETS
- PFC
- Benachrichtigung bei Netzwerkstau
- Anwendungen
- Logical link-down
- Virtualisierung der Netzwerkschnittstelle

Abbildung 4: Data Center Bridging Exchange Protocol

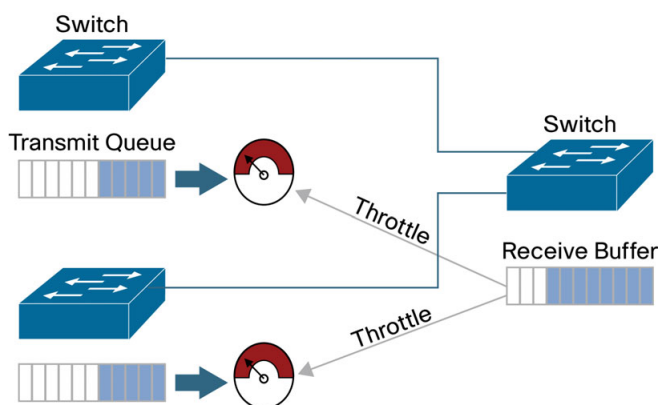


**Benachrichtigung bei Netzwerkstau**

Benachrichtigung bei Netzwerkstau oder „Congestion Notification“ ist Verkehrsmanagement, das Stauungen an den Rand des Netzwerks verschiebt, und zwar indem es Rate Limiter anweist, den Verkehr, der den Stau verursacht, zu bearbeiten. Die IEEE 802.1Qau-Arbeitsgruppe akzeptierte den Vorschlag Ciscos für Congestion Notification. Dieser definiert eine Architektur für das aktive Management von Verkehrsflüssen, um Stauungen zu vermeiden.

Stauungen werden am Congestion Point gemessen. Liegt ein Stau vor, so werden Rate Limiting oder Back Pressure am Reaction Point angewandt, um den Verkehr zu beeinflussen und die Auswirkungen des Staus auf den Rest des Netzwerks zu reduzieren. In dieser Architektur kann ein Aggregation Level Switch Control Frames an zwei Access Level Switches senden und diese dazu auffordern, ihren Verkehr einzuschränken (Abb. 5). Dieser Ansatz schützt die Integrität des Network Cores und wirkt sich nur auf die Teile des Netzwerks aus, die den Stau verursachen, und zwar in der Nähe der Quelle (siehe IEEE 802.1Qau und <http://www.ieee802.org/1/pages/802.1au.html>).

Abbildung 5: Congestion Notification



### Layer 2 Multipathing

ECMP (Equal-Cost Multipath) Routing wird heute auf Layer 3 durchgeführt. Die Organisationen zur Definition von Standards schlagen mehrere Alternativen vor, um ECMP auch auf Layer 2 zu erreichen. TRILL (Transparent Interconnection of Lots of Links) ist eine Lösung, die in der IETF Standards Group vorgeschlagen wird, während Shortest-Path Bridging (IEEE 802.1Qat) momentan von der IEEE untersucht wird. Beide suchen eine Lösung, um höhere Vernetzung zwischen Switches auf Layer 2 zu ermöglichen. Layer 2 Multipathing (L2MP) steigert die bisektionale Bandbreite, indem es mehrere parallele Pfade zwischen Knoten ermöglicht. Dies führt zu höherer Bandbreite im Interconnect-Netzwerk mit niedrigeren Latenzen. Auf der Basis der Verkehrsmuster von großen Server-Farmen wird L2MP die Leistungsfähigkeit dieser Netzwerke steigern. Das Load Balancing des Verkehrs zwischen alternativen Equal Cost Paths wird die Anwendungsleistung und die Widerstandsfähigkeit des Netzwerks verbessern. Cisco Data Center Ethernet mit L2MP wird den Gebrauch aller verfügbaren Verbindungen zwischen Knoten ermöglichen, Single Path Constraints vermeiden und es den Data Center Operators erlauben, dynamische Topologie-Änderungen vorzunehmen, ohne sich Gedanken über Konvergenzeffekte machen zu müssen, wenn ein Pfad entfernt oder hinzugefügt wird.

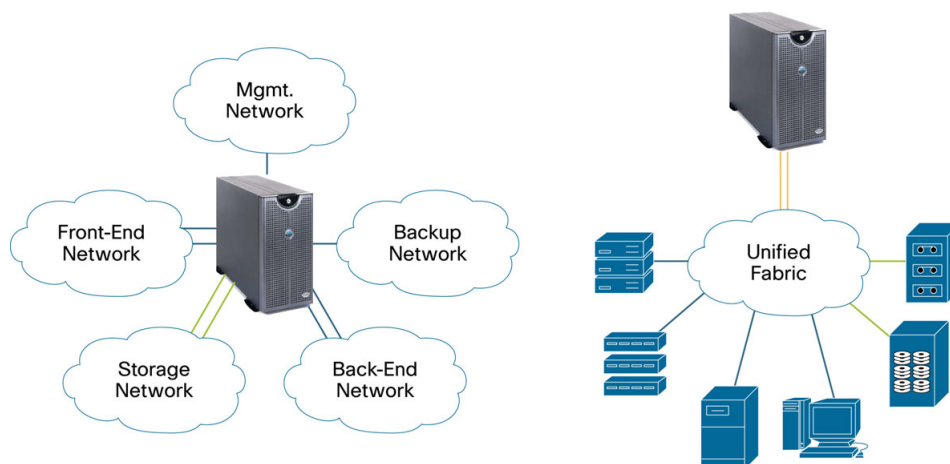
### I/O-Konsolidierung

Die kontinuierliche Erweiterung von 10 Gigabit Ethernet unterstützt die Vermischung von Verkehrstypen zwischen Servern und Switched Networks. Cisco Data Center Ethernet-Erweiterungen (PFC, ETS, DCBX und Congestion Notification) versetzen eine 10 Gigabit Ethernet-Verbindung in die Lage, mehrere Verkehrstypen gleichzeitig zu unterstützen und dabei die jeweiligen Verkehrsbehandlungen beizubehalten. Mit diesen Erweiterungen wird die gleiche 10 Gigabit Ethernet-Verbindung auch in der Lage sein, Fibre Channel Storage-Verkehr zu unterstützen, und zwar mit einer No-Drop-Funktion für FCoE-Verkehr. Konsolidierte I/O am Server mit Unterstützung von FCoE erlaubt es allen Hosts, auf Storage-Ressourcen zuzugreifen, und zwar über das gemeinsame Unified Fabric und unter Einsatz einer gemeinsamen, vereinheitlichten I/O-Schnittstelle (Abb. 6).

Änderungen der Server-Architekturen wirken sich ebenfalls auf den Trend hin zu vereinheitlichtem I/O aus. Die Einführung von Peripheral Component Interconnect Express (PCI-Express) über PCI und PCI-X hat Server in die Lage versetzt, den I/O-Engpass am PCI-Bus zu überwinden. Diese grundlegende Änderung erlaubt es den Servern, die volle 10 Gigabit Ethernet-Schnittstelle zu nutzen. Gleichzeitig verwenden Server Chips mit hoher Dichte, Quad Cores und Multiprozessor-Plattformen, was zu mehr Bedarf für hohe Bandbreite von und zu den Servern führt. Mit mehreren Prozessoren, Cores und virtuellen Maschinen auf einem einzigen Server wird 10 Gigabit Ethernet vermehrt eingesetzt werden. Eine Methode für das Management von mehreren Verkehrstypen gleichzeitig wird von entscheidender Bedeutung sein, soll Verkehr auf größeren, konsolidierten I/O-Verbindungen gemeinsam genutzt werden.

Ein konsolidierter I/O Link über ein vereinheitlichtes I/O-Kabel kann Multiprotokoll-Verkehr auf einem einzigen Kabel einem Unified Fabric präsentieren. Ein Unified Fabric ist eine einzige Ethernet-Mehrzweckverbindung, die IP- und Fibre Channel-Verkehr gleichzeitig über die gleiche Schnittstelle und das gleiche Switch Fabric übertragen und dabei differenzierte Serviceklassen beibehalten kann. Zu den Anwendungsfällen gehören Multiprotokoll-Transport, FCoE und RDMA über Ethernet mit niedriger Latenz.

Abbildung 6: Unified Fabric



Hat ein Produkt die erforderlichen Cisco Data Center Ethernet-Standardspezifikationen implementiert, so sollte es mit anderen Produkten kompatibel sein, in denen die gleichen Spezifikationen implementiert wurden. Die Mindestanforderungen für Cisco DCE™-Endpunkte (Hosts, Targets, Server etc.) sind PFC, ETS und DCBX. Ein Cisco DCE™-fähiger Switch wird auch eine Lossless Intraswitch Fabric-Architektur enthalten, die einen No-drop-Service und L2MP bieten kann.

### **Lossless Fabric**

Ein Cisco DCE™ Switch muss ein Lossless Fabric für eine Lossless Transmission Service-Klasse bieten, bei dem kein Frame verloren geht. Um FCoE zu unterstützen, ist ein Lossless Fabric unbedingt erforderlich. Es sorgt dafür, dass Storage-Verkehr einen No-Drop-Service hat. Um ein Lossless Ethernet Fabric mit Multiprotokoll-Support zu erstellen, sind zwei Elemente erforderlich: ein Pause-Mechanismus auf der Basis von Prioritäten (PFC) und ein intelligenter Switch Fabric Arbitration-Mechanismus, der Ingress Port-Verkehr mit Egress Port-Ressourcen verbindet, um Pausenanforderungen zu erfüllen.

Der heutige Standard-Pausemechanismus in Ethernet hält alle Verkehrstypen auf dem Link an. PFC ist für die I/O-Konsolidierung entscheidend, da es bis zu acht separate logische Links über den gleichen physikalischen Link erstellt. So können beliebige der acht Verkehrstypen unabhängig voneinander angehalten werden. Andere Verkehrstypen können ungestört weiterfließen. PFC verwendet einen Pause-Mechanismus auf der Basis von Prioritäten, um den IEEE 802.1p-Verkehrstyp auszuwählen, der auf einem physikalischen Link angehalten werden soll. Die Fähigkeit, Pausen für unterschiedliche Verkehrstypen auszulösen, bedeutet, dass der Verkehr über eine einzelne Schnittstelle konsolidiert werden kann. Dies führt zu einer einzigen Ethernet-Verbindung für vereinheitlichte I/O-Verbindungen. Einzelne, vereinheitlichte I/O-Verbindungen führen mehrere Verkehrstypen zusammen und stellen diese einem Unified Fabric zur Verfügung.

PFC bietet auf jedem logischen Link eine No-drop-Option, und zwar mit der Fähigkeit, unabhängige logische Verkehrstypen anzuhalten. PFC (ebenso wie der standardmäßige Pause-Mechanismus) macht einen Link verlustlos, aber das reicht nicht aus, um aus einem Netzwerk ein Lossless Fabric zu machen. Zusätzlich zu No-drop-Service auf dem Link ist eine Möglichkeit erforderlich, um das Pauseverhalten des Ingress-Ports mit den Egress Port-Ressourcen zu verbinden, und zwar über das Intraswitch Fabric und mit PFC. Um das Netzwerk verlustlos zu machen, muss jeder Switch die Ressourcen der Ingress-Links mit den Ressourcen der Egress-Links assoziieren. Die logische Verbindung der Verfügbarkeit von Egress Port-Ressourcen mit Ingress Port-Verkehr ermöglicht eine Arbitrierung, damit keine Pakete verloren gehen – und das ist die Definition einer Lossless Switch Fabric-Architektur. Das Verhalten des Lossless Ethernet Intraswitch Fabric bietet den erforderlichen No-drop-Service, der das Buffer Credit Management System emuliert, das heute in Fibre Channel Switches zu sehen ist.

### **Fibre Channel over Ethernet**

Um Fibre Channel Storage-Verkehr oder jede andere Anwendung, die verlustlosen Service benötigt, über ein Ethernet-Netzwerk zu transportieren und ein Unified Fabric zu erreichen, ist eine Lossless Service-Klasse erforderlich. Fibre Channel Storage-Verkehr benötigt No-drop-Fähigkeit. Ein No-drop-Verkehrstyp kann mit Cisco Data Center Ethernet und einem Lossless Ethernet Switch Fabric erstellt werden.

Das FCoE-Protokoll ordnet Fibre Channel Frames über Ethernet zu, und zwar unabhängig vom Ethernet Forwarding Scheme. Dies ermöglicht einen evolutionären Ansatz für die I/O-Konsolidierung durch die Beibehaltung aller Fibre Channel-Konstrukte.

INCITS T11 schreibt den Standard für FCoE. Diese Gruppe wird vorgeben, dass ein Lossless Ethernet-Netzwerk erforderlich ist, um FCoE zu unterstützen. Der standardmäßige Pause-Mechanismus (ebenso wie PFC) macht einen Link verlustlos, aber das reicht nicht aus, um ein Netzwerk verlustlos zu machen. Um das Netzwerk verlustlos zu machen, muss jeder Switch die Zwischenspeicher der eingehenden Links mit den Zwischenspeichern der Egress-Links in Übereinstimmung bringen und sie mit der Pause-Implementierung verbinden. Dies ist eine Fähigkeit einer Plattformarchitektur, die nichts mit Protokollen zu tun hat. Die Switches der Cisco Nexus 5000 Serie bieten diese Fähigkeit schon heute. Die Switches der Cisco® Nexus 7000 Serie werden diese Lossless Network-Fähigkeit in Zukunft mit Cisco DCE™ Modulen bieten.

Nicht alle Cisco Produkte werden Cisco Data Center Ethernet implementieren, da viele Cisco Produkte nicht nur für das Data Center entwickelt wurden. Cisco Data Center Ethernet wurde in erster Linie mit dem Ziel entwickelt, Data Center-Netzwerke zu erweitern, obwohl es durchaus auch in anderen Umgebungen seine Vorteile hat. Die meisten dieser Erweiterungen werden in Cisco Produkten geboten, die im Data Center zu finden sind. Da Cisco Data Center Ethernet aus zahlreichen Komponenten besteht, wird zumindest ein Teil der Erweiterungen wahrscheinlich zu bestimmten Plattformen hinzugefügt. So könnte ein Cisco Switch z.B. I/O-Konsolidierung unterstützen, dafür aber keine Congestion Notification implementieren. In diesem Falle würde dieser Switch als kompatibel mit anderen Cisco DCE™-Produkten angesehen, und zwar ohne die optionale Congestion Notification. Außerdem werden nicht alle IT-Abteilungen einen Lossless Service über ein konvergentes Ethernet-Netzwerk laufen lassen oder ein Unified Fabric implementieren wollen. In diesen Fällen wird nach wie vor das klassische Ethernet implementiert werden.

## Schlussfolgerung

Cisco Data Center Ethernet liefert die Architektur für ein Unified Fabric und erfüllt die Versprechungen der Cisco Data Center 3.0-Vision schon heute. Ethernet ist das System der Wahl für ein einzelnes, konvergentes Fabric, das mehrere Verkehrstypen unterstützen kann. Ethernet ist in Datenzentren auf der ganzen Welt vertreten und verfügt somit über eine breite Basis von technischer und betrieblicher Expertise weltweit. FCoE ist die erste konkrete Anwendung eines Unified Fabric. Mit Cisco Data Center Ethernet-Erweiterungen kann eine Lossless Ethernet-Fähigkeit geschaffen werden, um die „No-drop“-Anforderung von Fibre Channel zu erfüllen. Cisco Data Center Ethernet unterstützt sowohl iSCSI als auch FCoE und bietet Kunden die Option, eine oder beide dieser Möglichkeiten zu benutzen; beide werden von einem robusteren Ethernet Fabric profitieren. Die Schaffung eines verlustlosen Service über ein Ethernet Fabric wird FCoE, RDMA und Anwendungen wie Echtzeitvideo ermöglichen, um Übertragungsgarantien von virtuellen No-drop Links gemischt mit anderen konkurrierenden Anwendungen zu erhalten. Von der L2MP-Innovation wird jedes Data Center Ethernet-Netzwerk profitieren, und zwar unabhängig davon, ob SAN-Verkehr darauf konvergiert ist. Höhere Bandbreite und niedrigere Latenz mit L2MP wird zu Leistungssteigerungen für alle Anwendungen führen. Cisco Data Center Ethernet-Erweiterungen, Lossless-Funktionalität und L2MP werden in einer Familie von Cisco-Switches der nächsten Generation bereitgestellt. Hierzu gehören die Cisco Nexus 7000 und Cisco Nexus 5000 Serien, integrale Bestandteile der Data Center Unified Fabric-Architektur.

Die Cisco Data Center Ethernet-Architektur fügt Innovationen speziell für Ethernet im Data Center zu einer leistungsstarken existierenden Basis von Ethernet-Implementierungen hinzu. Diese neuen Data Center-Erweiterungen bieten IT-Abteilungen mehrere Vorteile: eine neue, flexible Methode für die Konsolidierung von I/O über Ethernet auf dem gleichen Netzwerk-Fabric, im Gegensatz zur Unterstützung von separaten Netzwerken; eine Methode für die Bereitstellung einer Lossless Traffic-Klasse auf Ethernet; intensivere Ausnutzung der Bandbreite zwischen Knoten mit Hilfe von Equal-cost Multipathing auf Layer 2 für höheren Support von bisektionalem Verkehr. Die Art und Weise, wie IT-Abteilungen unterschiedliche Verkehrstypen über die gleiche Schnittstelle, das gleiche Kabel und den gleichen Switch konsolidieren, wird sich im Laufe der Zeit weiterentwickeln. Cisco Data Center Ethernet bietet Ihnen die Flexibilität, selbst auszuwählen, was über eine konsolidierte Schnittstelle, einen Link und einen Switch Fabric läuft, und wann dies geschehen soll.

## Hier finden Sie weitere Informationen

Besuchen Sie **Cisco Data Center Ethernet**: <http://www.cisco.com/go/dce>

**Erläuterung von DCBX**: <http://www.ieee802.org/1/files/public/docs2008/az-wadekar-dcbcxp-overview-rev02.pdf>



Cisco Systems GmbH  
Kurfürstendamm 21-22  
10719 Berlin

Cisco Systems GmbH  
Neuer Wall 77  
20354 Hamburg

Cisco Systems GmbH  
Hansaallee 249  
40549 Düsseldorf

Cisco Systems GmbH  
Ludwig-Erhard-Straße 3  
65760 Eschborn

Cisco Systems GmbH  
Wilhelmsplatz 11 (Herold Center)  
70182 Stuttgart

Cisco Systems GmbH  
Am Söldnermoos 17  
85399 Hallbergmoos

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at [www.cisco.com/go/offices](http://www.cisco.com/go/offices).

©2006 Cisco Systems, Inc. All rights reserved. CCVP, the Cisco logo, and the Cisco Square Bridge logo are trademarks of Cisco Systems, Inc.; Changing the Way We Work, Live, Play, and Learn is a service mark of Cisco Systems, Inc.; and Access Registrar, Aironet, BPX, Catalyst, CCDA, CCDP, CCIE, CCIIP, CCNA, CCNP, CCSP, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Cisco Unity, Enterprise/Solver, EtherChannel, EtherFast, EtherSwitch, Fast Step, Follow Me Browsing, FormShare, GigaDrive, GigaStack, HomeLink, Internet Quotient, IOS, IP/TV, iQ Expertise, the iQ logo, iQ Net Readiness Scorecard, iQuick Study, LightStream, Linksys, MeetingPlace, MGX, Networking Academy, Network Registrar, Packet, PIX, ProConnect, RateMUX, ScriptShare, SlideCast, SMARTnet, StackWise, The Fastest Way to Increase Your Internet Quotient, and TransPath are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries.

All other trademarks mentioned in this document or Website are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (0609R)