ComConsult Research

Der Netzwerk Insider

Juli 2017

Systematische Weiterbildung für Netzwerk- und IT-Professionals

Schwerpunktthema

Netzwerke in der AWS-Cloud

von Markus Schaub

Schaut man sich den gerade veröffentlichten Gartner Quadranten für IaaS-Cloud Services an, könnte einem der Gedanke kommen, dass Amazon Gartner gekauft hat. So einsam weit oben rechts findet man AWS im „Magic Quadrant“, dass man Microsoft, das es immerhin in denselben Quadranten geschafft hat, fast übersieht. Mit IaaS ist bei beiden aber schon lange keine reine Servermiete mehr gemeint.

Vielmehr umfasst das Angebot von Amazon, Microsoft, Google und Konsorten die komplette Infrastruktur eines Rechenzentrums, ergänzt um weitere Mehrwertdienste und das an Standorten rund um den Globus. Jeder Standort ist dabei redundant ausgelegt. Zu dieser virtualisierten Rechenzentrumsstruktur gehört naturbedingt ein Netzwerk. Denn all die Dienste und Server, die in der Cloud betrieben werden, müssen natürlich miteinander kommunizieren. Will man die Cloud mit dem eigenen Netz verbinden oder betreibt man Cloud-Lösungen an mehreren Standorten bzw. nutzt Dienste, die nicht weltweit an jedem Standort verfügbar sind, so kommen weitere Aufgaben bei der Netzwerkgestaltung hinzu. Es ist also höchste Zeit, sich dieses Themas auch als Netzwerker anzunehmen.

weiter auf Seite 9

Zweitthema

Ist BGP das bessere IGP?

von Markus Geller

In den letzten Jahren haben wir sehr oft über neue Layer 2 Verfahren gesprochen und geschrieben. Viele dieser neuen Möglichkeiten wie TRILL, SPB oder MC-LAG schienen ideal, um im Rechenzentrum die überkommenen Mechanismen des Spanning Tree oder der Link Aggregation abzulösen.

Und nicht nur wir bei der ComConsult haben das so gesehen, auch viele etablierte Hersteller setzen seit Jahren auf Layer 2 Fabrics für ihr Data Center Design.

- Cisco FabricPath
- Avaya Fabric Connect
- Juniper QFabric
- Brocade VCS Fabric

sind nur einige Beispiele, die ich hier nennen möchte. Der große Vorteil dieser neuen Technologien ist ein blockadefreies Netzwerk, in dem alle physikalischen Verbindungen optimal genutzt werden können.

Aber am Ende bleibt es immer noch eine Layer 2 Infrastruktur und somit eine Broadcastdomäne, die sich über viele Netzwerkelemente spannt mit all den Nachteilen und Problemen, die die Verarbeitung und Verbreitung von Broadcast und Multicast mit sich bringen.

weiter auf Seite 20

Geleit

iPad Pro: endlich tauglich für den Unternehmenseinsatz?

auf Seite 2

Standpunkt

Muss man Gebäude abschirmen?

auf Seite 19

Sonderveranstaltungen

Herausforderung Informationssicherheit

Cloud Computing, Security as a Service, Virtualisierung
IoT, Abwehr von Angriffen, rechtliche Rahmenbedingungen

auf Seite 17

Wireless und Mobility

ab Seite 6

IT-Infrastrukturen für das Gebäude der Zukunft

ab Seite 4

Geleit

iPad Pro: endlich tauglich für den Unternehmenseinsatz?

Apple positioniert das iPad seit Jahren als Laptop-Killer und als das ideale Gerät für mobile Anwendungen. Die typischen Kunden sehen das ganz unterschiedlich, das Spektrum reicht von untauglich bis hin zu perfekt. Woran kann das liegen? Und wie sieht die Zukunft aus? Apple hat gerade eine neue Version von IOS zusammen mit einer deutlich überarbeiteten Hardware angekündigt. Kommt damit der endgültige Durchbruch?

Tatsächlich hat Apple das iPad seit Jahren bis zur aktuellen Ankündigung kaum verändert. Mit den Pro-Versionen kam vor zwei Jahren der zumindest verbale Versuch, ein Gerät speziell für den Unternehmenseinsatz zu schaffen. Das wesentliche Merkmal war dabei die Kombination mit einem "pencil" und einer Tastatur. Vereinfacht dargestellt kann man sagen als Tiger gesprungen und als Maus gelandet (ich bin seit der Markteinführung leidgeprüfter Benutzer des großen Pro). Erst der Druck durch Microsoft und das kontinuierlich verbesserte Surface Pro hat Apple nun endlich dazu gebracht alle Register zu ziehen, um endlich dem Anspruch eines professionellen Gerätes zu genügen. Dazu wurde die Hardware inklusive der Grafik auf ein Leistungsniveau gezogen, das sicher weit über dem Bedarf des normalen Benutzers liegt. Ein gutes Beispiel ist LumaFusion als Video-Editor. Die Leistung, die hier im 4k-Editing erbracht wird, erfordert ansonsten einen iMAC mit einer gehobenen Ausbaustufe. Gleiches kann man für Affinity Photo auf dem iPad sagen. Dies ist die fast volle Leistung von Photoshop auf einem Touch-Endgerät mit einer Touch-Applikation (würden doch nur meine Photoshop Actions hier funktionieren, ansonsten kann man aber sagen, dass Affinity Photo auch als Desktop-Version eine mehr als ernstzunehmende Bedrohung für Adobe ist). Wesentlichen Einfluss auf die Leistungssteigerung hat die dynamische Bildwiederholrate von bis zu 120 Bilder pro Sekunde. Diese führt bei Nutzung des Pencils zu einer Verzögerung von nur noch 20ms. Schreiben oder Malen mit dem Pencil ist vom Schreiben auf Papier bis auf die glatte Oberfläche nicht mehr zu unterscheiden (ich bin absolut begeistert und mein Wacom liegt endgültig im Schrank). Das ist die neue Hardware. Dies wird ergänzt um eine neue Version von IOS, IOS 11. Auch wenn wir es noch im Test haben und die erste Public Beta gerade erst veröffentlicht wurde, kann man jetzt schon sagen, dass dieses Upgrade aus dem iPad ein neues Gerät machen wird.

Haben die langen Jahre des Wartens auf ein professionelles iPad sich also endlich gelohnt. Haben wir jetzt endlich eine Version, die professionellen Ansprüchen genügt? Und kann Apple damit den verlorenen Boden gegenüber dem Surface Pro wieder gut machen?

iPad kontra Surface Pro: ein Vergleich, der hinkt!

Microsoft hat Windows in den letzten Jahren so erweitert, dass es sowohl eine traditionelle Bedienung mit der Maus als auch eine Touch-Bedienung erlaubt. Damit lautet das Standard-Argument für den Einsatz eines Surface Pro gegenüber einem iPad, dass es die "alten" Anwendungen weiterhin unterstützt. Das ist grundsätzlich richtig und wer den Bedarf nach einer Anwendung hat, die es nicht als Touch-Anwendung gibt, der wird auf dem Surface Pro besser aufgehoben sein. Der Nachteil dabei ist, dass es kaum sinnvolle Touch-Anwendungen auf dem Surface Pro gibt. Touch verkommt unter Windows zu einem Spielzeug ohne besonderen Mehrwert.

Der Witz und der zentrale Punkt des iPad ist, dass es nur Touch-Anwendungen unterstützt. Damit muss jede Anwendung auf dem iPad neu entwickelt werden. Dementsprechend stoßen wir auch auf dem iPad auf sehr moderne Anwendungen, die es so bisher nicht gegeben hat. Dies können bahnbrechende Anwendungen wie LumaFusion oder Affinity Photo sein oder ganz "normale" Anwendungen wie Microsoft Office oder Apple Pages.

Aus Unternehmenssicht bedeutet das: wird für einen bestimmten Einsatzzweck eine neue Anwendung benötigt und hat der Bediener einen Vorteil von einer Touch-Bedienung, dann kann das zu einem signifikanten Mehrwert führen. Beispiele für diese Art von Anwendungen finden wir in Krankenhäusern mit dem iPad als elektronische Akte, im Transportwesen, in Flugzeugen oder im Kassenbereich.

Soweit so gut, aber warum tut sich das iPad im Unternehmenseinsatz dann trotzdem so schwer? Warum kann es bisher einen Laptop eben nicht ersetzen? Und warum punktet das Surface Pro bei so vielen Anwendern?

Mangelbereich 1: Multitasking

Apple hat den Bedarf für Multitasking seit Jahren schlicht ignoriert. Alle Beschwerden oder Wünsche von Anwendern wurden einfach arrogant abgewiesen. Dem Surface Pro sei Dank, dies ist jetzt vorbei. Mit IOS 11 haben wir endlich ein deutlich flexibleres Multitasking mit der Möglichkeit von Drag-and-Drop. Endlich können wir einen Anhang aus einer Email in eine andere Anwendung ziehen oder umgekehrt. Was hier harmlos klingt, ist für die Verringerung des funktionalen Abstands zum Laptop entscheidend. Wir müssen jetzt abwarten wie der Durchschnittsbenutzer diese neue Funktionalität annimmt. Sie erfordert eine Eingewöhnung. Die ist zwar minimal, aber es wird Benutzer geben, die das nicht mehr intuitiv finden. Wie auch immer, mit IOS 11 macht Apple einen Riesenschritt nach vorne! Man muss aber anmerken, dass es eine kleine Version von Drag-and-Drop innerhalb der Anwendungen der Firma Readdle schon gibt.

Reicht das aus? Im Prinzip vielleicht, wäre da nicht der Elefant im Raum.

Mangelbereich 2: das Dateisystem, der Elefant im Raum!

Wäre es nicht toll, wenn es ein Gerät gäbe, auf dem Ransomware absolut chancenlos ist? Nun, das gibt es. Ein iPad oder ein iPhone. Ein Kernmerkmal von IOS ist das Sandboxing und der damit verbundene bewusste und gewollte Verzicht auf ein Dateisystem. Jede Anwendung läuft nur in ihrer eigenen Umgebung. Soll eine Datei bearbeitet werden, dann muss sie in dieser Umgebung sein und damit ist sie für andere Anwendungen nicht zugreifbar. Wenn ich also eine Datei in einem Cloud-Speicher habe und die-

Schwerpunktthema

Netzwerke in der AWS-Cloud

Fortsetzung von Seite 1

Markus Schaub ist seit 2009 Leiter von ComConsult-Study.tv. Er verfügt über umfangreiche Berufserfahrung in den Bereichen Netzwerken und VoIP und ist seit mehr als 13 Jahren bei ComConsult beschäftigt. Seine Schwerpunkte liegen im Netzwerk-Design, IP-Infrastrukturdiensten und SIP, zu denen er viele Vorträge auf Kongressen hielt, erfolgreich Seminare durchführte und zahlreiche Veröffentlichungen schrieb.

In diesem Artikel werden zunächst die Basiselemente der Cloud-Netze bei Amazon vorgestellt. Danach wird erklärt, wie Systeme zwischen verschiedenen Subnetzen innerhalb der Cloud kommunizieren und wie man Dienste anbindet, die aus dem Internet heraus erreichbar sein sollen. Zum Abschluss wird noch auf die Koppelungsvarianten der Cloud mit dem eigenen Netz eingegangen. Auf spezifische Sicherheitsaspekte wird zwar eingegangen, sie stehen jedoch nicht im Fokus, da das den Umfang eines Artikel sprengen würden.

Gestaltungselemente

Wer Netzwerke in der Cloud verstehen will, muss sich zunächst einmal mit neuen und alten Begriffen und deren Funktion auseinandersetzen: Region, Availability Zone, VPC, Subnetz. Hat man verstanden was diese bedeuten und wie sie zusammenhängen, ist es eigentlich ganz einfach, ein Netzdesign aufzusetzen.

Region

Eine „Region“ ist bei AWS recht wörtlich, nämlich geographisch gemeint. Aktuell gibt es 14 Regionen: 4 in den USA, eine in Kanada, 3 in der EU (noch: eine ist nämlich „London“), 5 in Asien und eine weitere in Süd Amerika. Diese Regionen sind zumeist Städten oder Bundesstaaten zugeordnet, manchmal geht es aber auch durcheinander. Für die europäischen gibt es London, Frankfurt und Irland.

Diese Regionen sind weitestgehend autark: man kann zwar Verbindungen zwischen zwei Regionen konfigurieren, jedoch ist das vergleichsweise aufwendig und nicht sonderlich performant. Im Grunde ist es vergleichbar mit einem VPN-Tunnel zwischen zwei Rechenzentren an unterschiedlichen, weit entfernten Rechenzentren. Damit gilt für die meisten Cloud-Dienste, dass man sie möglichst in derselben Region betreibt, wenn zwischen ihnen viele Daten ausgetauscht werden müssen.

Was Regionen noch auszeichnet, sind die dort verfügbaren Dienste: nicht alle Dienste sind überall verfügbar. So gibt es die Amazon-eigene Datenbank „Aurora“ beispielsweise nicht in Frankfurt, wohl aber in Irland. Dasselbe gilt auch für den Mail-Dienst SES. Wenn man ein Projekt aufsetzt, sollte man also vorher klären, welche Dienste man benötigt und danach die Region auswählen. Ist einem jedoch der Standort wichtiger, so muss man ggf. andere Dienste nutzen, also mySQL statt Aurora oder diese, wenn möglich, selbst betreiben, wie beim Email-Dienst.

Availability Zones

Innerhalb einer Region gibt es „Availability Zones“. Das sind verschiedene, jedoch nah beieinanderliegende Rechenzentren. Die Availability Zonen einer Region sind mit hoch-performanten Verbindungen miteinander gekoppelt, so dass das Delay kaum ins Gewicht fällt und Bandbreite keine Rolle spielt. Alle Dienste einer Region sind in allen Availability Zones verfügbar.

Bei diesen Zonen geht es also um Verfügbarkeit: fällt in einer Zone bspw. der Strom aus, so gilt das nicht (zwangsläufig) für die andere. In jeder Region gibt es mindestens zwei Availability Zonen. In Frankfurt gibt es seit kurzem sogar drei davon.

Bei den Zonen ist zu beachten, dass sie anders heißen können als die Region, in der sie liegen. So heißt die Region „EU (Frankfurt)“, die Zonen dort eu-central-1 und nicht eu-ffm oder eu-fra, wie man hätte meinen können. Die drei Zonen sind von „a“ bis „c“ durchnummeriert, also bspw. eu-central-1a.

Bei Pro Account (Account = Vertrag, nicht User) entspricht eine Zone immer einem Standort: für alle User dieses Accounts liegt die Availability Zone eu-central-1a also im selben Rechenzentrum und eu-central-1b in einem anderen. Das ist für die Planung hochverfügbarer Anwendungen schon mal gut zu wissen. Man muss aber auch wissen, dass für einen anderen Account eu-central-1a ein anderes Rechenzentrum gemeint sein kann. Wenn bei dem Vertrag ComConsult Research bspw. eu-central-1a das Rechenzentrum 1 gemeint ist, kann eu-central-1b bei dem Vertrag ComConsult Akademie ebenfalls das Rechenzentrum 1 gemeint sein. Hat ein Unternehmen also mehr als einen Account, so kann man über die Availability Zones nicht sicherstellen, dass Anwendungen im selben bzw. in unterschiedlichen Lokalitäten laufen.

Virtual Private Cloud (VPC)

Eine „Virtual Private Cloud“ wird fast durchgängig nur abgekürzt als VPC bezeichnet. Eine VPC kann man sich als das eigene Rechenzentrumskonstrukt innerhalb einer AWS-Region vorstellen. D.h. eine VPC kann alle Availability Zonen einer Region überspannen, also ganz so als würde man an einem Standort zwei Rechenzentren betreiben. Jedoch ist eine VPC immer auf eine Region beschränkt.

Alternativ kann man VPCs aber auch als Sicherheitszonen betrachten. Denn wie später im Artikel erklärt wird, kann man zwei VPCs mit Middleboxen, wie Firewalls oder IDS/IPS, gegeneinander abschotten, innerhalb einer VPC ist das hingegen nicht möglich.

Die Verbindung zwischen zwei VPCs nennt man Peering. Um zwei VPCs peeren zu können, müssen sie in derselben Region sein und die IP-Bereiche dürfen sich nicht überlappen.

Netzwerke in der AWS-Cloud

Subnetze

Subnetze sind genau das, was sie schon immer waren. Bei der Amazon-Cloud ist ein Subnetz immer auf eine Availability Zone beschränkt. Zwischen Subnetzen wird geroutet, dabei ist es jedoch egal, ob die beiden unterschiedlichen Subnetze in derselben Availability Zone sind oder nicht.

Abbildung 1 zeigt den Zusammenhang zwischen den verschiedenen Komponenten, aus denen ein Amazon-Netz besteht.

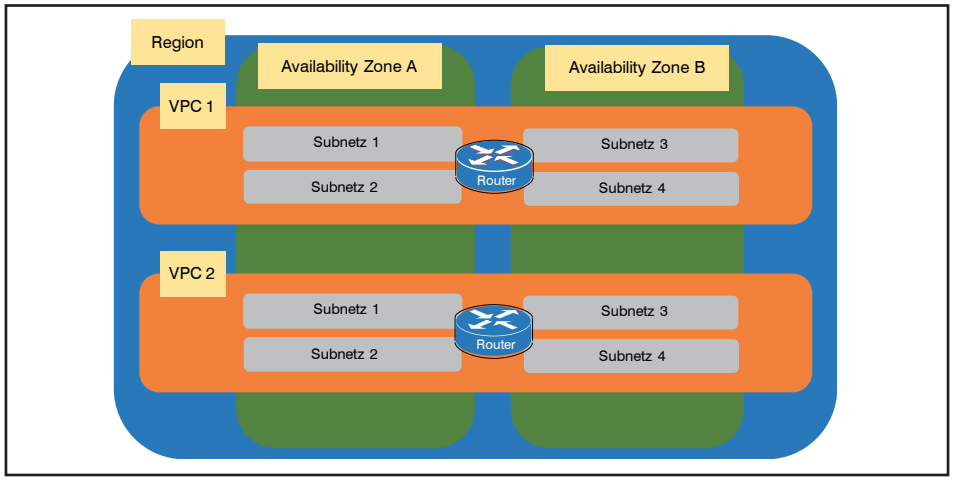

Abbildung 1: Zusammenwirken von Region, Availability Zone, VPC und Subnetze

IPv4-Adressen in der Cloud

Das Zusammenwirken von Regionen, VPCs und Subnetzen hat unmittelbare Konsequenzen für das IP-Design. Dabei müssen für IPv4 zwei verschiedene IP-Adressräume unterschieden werden:

1. Private IP-Adressen (RFC1918)
2. Elastic IP Adressen (öffentliche Adressen von Amazon)

VPCs und Subnetzen ordnet man bei der Anlage private IP-Adressen zu. Systeme, die an die Subnetze angeschlossen werden, bekommen aus diesem Bereich automatisch Adressen zugewiesen. Schließt man also bspw. einen EC2-Server an, so bekommt er eine private IP-Adresse, ebenso gilt das für einige Dienste, wie beispielsweise Datenbanken.

Mit diesen privaten IP-Adressen können die Systeme aus dem Internet nicht erreicht werden und ebenso wenig von sich aus mit dem Internet kommunizieren. Natürlich will man bestimmte Dienste und Server aber aus dem Internet erreichbar machen. Es kann auch die Anforderung geben, dass die Server selbst auf das Internet zugreifen können, bspw. für Systemupdates oder beim Betrieb von Mailservern.

Dazu gibt es drei Möglichkeiten:

1. NAT-Gateway

   Das NAT-Gateway funktioniert im Grunde wie der DSL-Router zuhause. Die Systeme können von sich aus Verbindungen ins Internet aufbauen. Verbindungen aus dem Internet hingegen benötigen eine Forwarding-Tabelle, die auf dem NAT-Gateway hinterlegt müsste. Das geht, Stand heute, jedoch nicht. Man kann jedoch selbst so genannte NAT-Instanzen betreiben, mittels derer das möglich ist. Dazu später mehr.

2. Loadbalancer

   Will man bspw. mehr als einen Web-Host betreiben, kann man einen Loadbalancer dafür nutzen.

3. Elastic IP

   Elastic IP Adressen sind öffentliche IP Adressen aus dem Vorrat von Amazon selbst. Diese kann man anfordern und einzelnen Systemen zuordnen. Die sind dann unter der Elastic IP Adresse erreichbar.

   Elastic IP-Adressen sind so lange kostenfrei, wie sie an laufende Systeme gebunden sind. Sobald sie nur vorgehalten werden, kosten sie Geld.

Es bietet sich an, das NAT-Gateways für Systeme anzubieten, die nur als Clients auf das Internet zugreifen und nur innerhalb der VPC Serverdienste anbieten. Also bspw. Datenbank-Server, die nur gelegentlich Update aus dem Internet benötigen, vom Internet aus jedoch nicht erreichbar sein sollen.

Elastic IP-Adressen wiederum nutzt man, um Serverdienste aus dem Internet erreichbar zu machen, bspw. Frontend-Web-Server.

Sämtlicher VPC-interne Datenverkehr wird über die privaten IPv4 Adressen abgewickelt, sodenn er über IPv4 transportiert wird.

Damit wären wir bei der Frage nach den

IPv6-Adressen in der Cloud

Bei Amazon gibt es die Möglichkeit, einer VPC auch einen IPv6 CIDR-Block zuweisen zu lassen (vgl. Abbildung 2). Pro VPC bekommt man dabei einen /56er Block zugewiesen. Die Größe dieses Blockes kann auch nicht geändert werden.

Damit hat man pro VPC rechnerisch also nur 256 Präfixe zur Verfügung. Für ein Rechenzentrum, was ja einem VPC entspricht, könnte das knapp werden. Für eine Sicherheitszone sollte das jedoch allemal ausreichen. Warum Amazon hier herum geizt, bleibt Amazons Geheimnis. Andererseits: wer mehr als 256 Subnetze benötigt, überweist wahrscheinlich jeden Monat so viel Geld an Amazon, dass die mit sich reden lassen, sodass diese Begrenzung nur für das Gros der Kunden gilt und denen sollten 256 Netze pro VPC genügen.

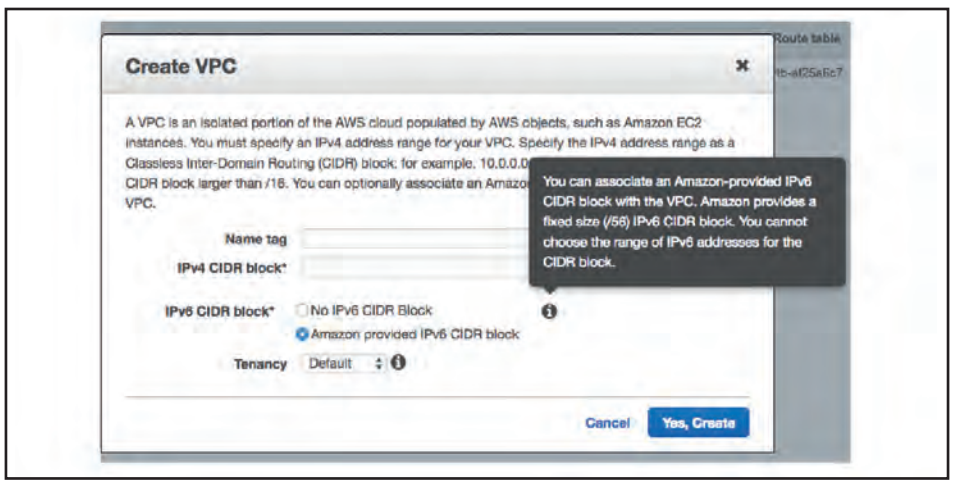

Abbildung 2: IPv6-CIDR Block für VPC bei AWS

Zweitthema

Ist BGP das bessere IGP?

Fortsetzung von Seite 1

Seit über 10 Jahren ist Markus Geller bei der ComConsult Research GmbH erster Ansprechpartner für die Themen VoIP und Lokale Netze. Der Schwerpunkt seiner Trainer Tätigkeit liegt dabei auf den Gebieten SIP, PSTN Migration, WebRTC sowie Layer 2 und 3 Techniken für MAN und LAN. Markus Geller verfügt über eine langjährige Erfahrung beim Aufbau und der Planung von Netzwerken im large Enterprise Umfeld, inkl. RZ-Netzwerken, WLAN und Multicastverfahren. In seiner über 20-jährigen IT-Laufbahn beschäftigt er sich mit der Evaluierung neuer Technologien und deren Einsatz in der Praxis. Zudem ist er als Autor diverser Fachartikel für den ComConsult Netzwerk Insider und das Wissensportal tätig.

Ein zusätzliches Problem, welches alle genannten Produkte gleichsam trifft, ist zudem der Umstand, dass die Lösungen zueinander inkompatibel sind. Da sich mit TRILL und SPB gleich zwei Standards zur Bildung von loop- und blockadefreien Layer 2 Strukturen etabliert haben, führte dies im Markt zu einer Aufspaltung in zwei Lager, so dass kaum eine Möglichkeit der Interoperabilität zwischen den Herstellern gegeben ist.

Aber auch Lösungen, die mit Multi-Chassis Link Aggregation arbeiten, haben ihre Probleme. Zum einen gibt es hierfür keine Standards im Rahmen von IEEE 802.3ad, und auch die Frage, wie viele Chassis sich zu einer virtuellen Fabric zusammenschließen lassen, ist bei jedem Hersteller anders implementiert.

Parallel zu der Entwicklung im Layer 2 Networking gab es aber auch Überlegungen, ob man die bekannten Probleme mit dem klassischen Data Center Design nicht auch mit Layer 3, also mittels Routing, lösen kann.

Vorreiter dieser Überlegungen waren die Hyperscale Rechenzentren von Microsoft und Facebook. Allen voran muss man hier Petr Lapukhov nennen, der sowohl bei Microsoft als auch bei Facebook diesen Designansatz etablierte und bei der IETF den RFC 7938 federführend betreute.

Dieser RFC 7938 beschreibt im Detail, wie ein Layer 3 Netzwerk mittels eBGP aufgebaut werden sollte. Doch dazu kommen wir etwas später.

Die primären Fragen, die wir zunächst klären sollten, lauten daher:

1. Welche Vorteile hat eigentlich ein Layer 3 Design?
2. Welche Data Center Anwendungen profitieren davon?
3. Welche Auswirkung hat die Wahl des Routing Protokolls?
4. Wie sieht die Topologie eines eBGP Layer 3 Data Center aus?
5. Lässt sich das Design auch auf den Campus übertragen?

Kommen wir damit direkt zu Frage 1:

„Wo liegen die Vorteile?“

Diese unterscheiden sich erstmal gar nicht so sehr von den Vorzügen einer Layer 2 Designs mit TRILL oder SPB.

Auch hier können wir eine Struktur aufbauen, die sowohl eine Link-, als auch eine Node-Protection ermöglicht. Dies bedeutet: Sowohl der Ausfall einer Leitung als auch eines Netzwerkknotens kann kompensiert werden, da immer mindestens ein alternativer Pfad (Link bzw. Next Hop) zur Verfügung steht. Auch die Vorzüge des ECMP (Equal Cost Multi-Path), die sowohl TRILL als auch SPB dank des zugrundeliegenden IS-IS Routing mitbringen, lassen sich in einem Layer 3 Design umsetzen.

Ebenso sind die Umschaltzeiten mehr als ausreichend. Im Vergleich zu klassischen Verfahren wie dem RSTP oder dem STP geht es sogar rasend schnell. Je nach eingesetzter Hardware können diese in Bereichen von 20 bis 50ms liegen. Schneller geht es auch bei einem MPLS Provider nicht.

Die eigentlichen Vorteile liegen vielmehr in der Beschneidung der Broadcastdomänen, da in einem solchen Designansatz das Routing bis in den Top of Rack bzw. End of Row Switch gezogen wird.

Im Ergebnis bildet jeder ToR Switch sein eigenes Subnetz, welches mittels Routing bekannt gegeben wird. Dadurch entfällt die Verarbeitung und Weiterleitung von Broadcast über den Access-Switch hinaus.

Diese Beschneidung von Layer 2 Broad- und Multicast führt uns zu einem Design, welches viel stärker skaliert als alle bisherigen Ansätze, da alle hiermit verbundenen Probleme nur singulär auf einem ToR/EoR Switch gelöst werden müssen.

In einem Designbeispiel der Firma Facebook führt dies zu einem möglichen Ausbau mit 48 ToR Switchen pro PoD (Point of Delivery oder auch Modul) und dem Zusammenschluss von 48 dieser PoDs zu einem Data Center. Gehen wir von einem 48 Port Switch im Access (ToR) aus und nutzen 4 Anschlüsse für den Uplink, so ergibt sich eine maximale Anzahl von über 100.000 Ports zur Anschaltung von Servern innerhalb eines einzigen Rechenzentrums. Eine, wie ich finde, beeindruckende Anzahl. (siehe Abbildung 1)

Da hierzu jedoch kein proprietäres Verfahren benötigt wird und auch keine Entscheidung für oder gegen ein Layer 2 Verfahrens wie TRILL oder SPB bzw. MC-LAG getroffen werden muss, ist diese Vorgehensweise auch in einer heterogenen Umgebung einsetzbar.

Denn die Entscheidung für oder gegen TRILL bzw. SPB, also einer Layer 2 Fabric, bedeutet auch, sich in die direkte Abhängigkeit eines Herstellers und seiner Lösung zu begeben.

Erstes Zwischenfazit: Ein standardisiertes Routing Verfahren erlaubt ein Netzwerk

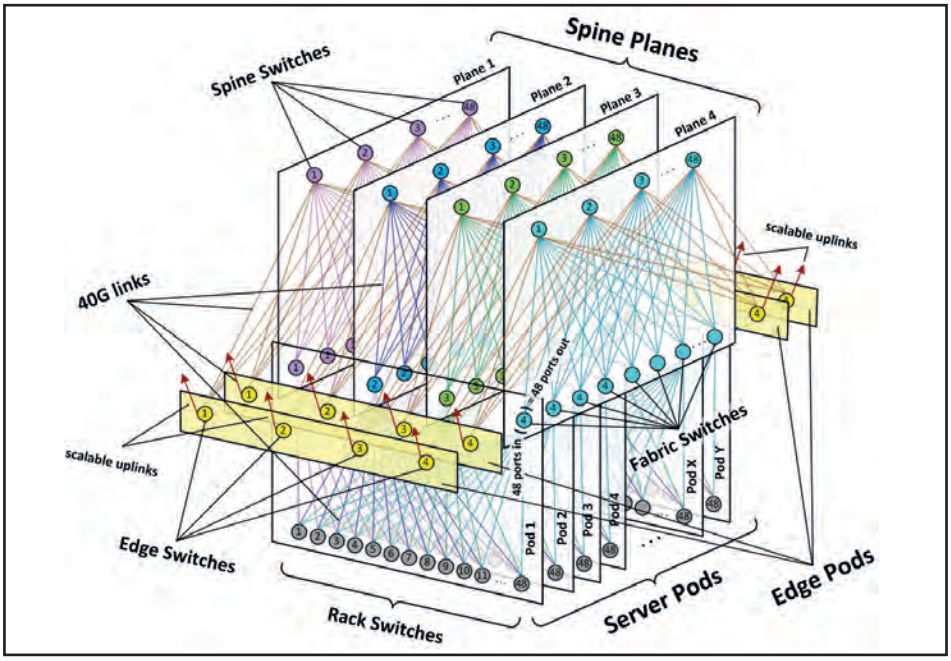

Abbildung 1: Spine-Leaf Data Center Quelle: Facebook

Design ohne Broadcastprobleme, hoher Skalierung und Herstellerunabhängigkeit.

Kommen wir somit zur 2. Frage:

„Welche Data Center Anwendungen profitieren davon?“

Zunächst einmal müssen wir natürlich feststellen, wer mit einem solchen Design ins Hintertreffen gerät, nämlich alle Anwendungen, die auf Broadcast oder Layer 2 Multicast basieren.

Ist das heute noch ein Problem? Ausschließen sollte man solche Situationen nie. Gerade bei der Migration älterer Anwendungen muss man vorsichtig agieren und sich genau darüber informieren wie eine Anwendung kommuniziert.

Betrachten wir jedoch moderne Anwendungen gerade im Web Umfeld, so sehen wir hier ausschließlich eine IP basierte Verständigung. Diese kommt naturgemäß sehr gut mit einer Layer 3 Fabric aus.

Dieser Grundgedanke der web-basierten Server Infrastruktur ist es denn auch, der den neuen Design Ansatz erst ermöglicht und vorangetrieben hat. Facebook und Microsoft setzen gezielt auf Routing und haben damit sämtliche Layer 2 Aktivitäten, soweit es geht, aus ihren Rechenzentren verbannt. Dies bedeutet: Keine Server Kommunikation mittels Broadcast oder Layer 2 Multicast, keine Layer 2 Redundanzverfahren zur Absicherung des Netzwerkes.

Ein weiterer großer Nutznießer des Routingansatzes ist die Server-Virtualisierung. Wie sicher viele von Ihnen wissen, basieren die gängigen Virtualisierungslösungen wie VMware NSX, Microsoft Hyper-V oder Cisco ACI auf einem Tunnelmechanismus. Hierzu wird auf einem bestehenden, physikalischen Netzwerk ein sogenanntes virtuelles Overlay Netzwerk gelegt.

Dieses wird zwischen den sogenannten Tunnelendpunkten, den VTEP, mittels eines Protokolls wie VXLAN, NVGRE oder Geneve gebildet. Die Tunnelendpunkte können sowohl auf einem Netzwerkknoten (z.B. Switch) als auch auf einem Serverhost liegen. Einzige Voraussetzung ist, dass sich die VTEP über IP erreichen können um, in IP gekapselt, Layer 2 Domänen zu bilden.

Und auch hier bietet uns das Layer 3 Design eine bestmögliche Lösung.

Zusammengefasst kann man also sagen, dass moderne Server Architekturen und die damit oft verbundene Virtualisierung sehr stark von einem gerouteten Netzwerkdesign profitieren.

Natürlich funktioniert Server Virtualisierung auch mit einer TRILL oder SPB Fabric, aber die Nachteile überwiegen. Da wäre zum einen der Umstand, dass neben den Overlay Tunnel zusätzlich Layer 2 Tunnel innerhalb der Layer 2 Fabric gebildet werden, was zu verschachtelten Tunnel führt und damit das Troubleshooting verkompliziert und zum anderen, dass Broadcast und Multicast sich innerhalb der Fabric wieder ungehemmt ausbreiten können.

Kommen wir damit zum nächsten Punkt:

„Welche Auswirkung hat die Wahl des Routing Protokolls?“

Diese Frage ist zugegeben ein wenig heikel und erfordert zunächst einmal einen Blick zurück in die Vergangenheit. Am Ende der Betrachtung sollte aber klargeworden sein, warum statt eines IGP Routings BGP aktuell so beliebt ist.

Die ersten Routing Protokolle, die entwickelt wurden, waren die Distance Vector Protokolle (RIP oder auch IGRP). Diese haben jedoch einige unangenehme Eigenschaften.

Da wäre zum einen der Umstand, dass sie nur sehr langsam periodische Updates zur Anzeige der Erreichbarkeit übermitteln und daher auch nur verzögert auf Netzwerkfehler reagieren können.

Ein weiteres Manko ist die nicht vorhandene Struktur. Anders als OSPF mit seinem Area-Konzept bilden alle Router, die untereinander RIP Informationen austauschen, „ein“ Netzwerk. In Kombination mit dem langsamen Verhalten bei der Erkennung von Netzwerkproblemen hat man daher die maximale Größe eines Netzes begrenzt (z.B. bei RIP auf 15 Hops). Zudem ist die Metrik meist sehr beschränkt, d.h. man betrachtet nur die Anzahl der Hops, über die ein Zielnetzwerk zu erreichen ist und beachtet dabei nicht, ob es sich um einen guten Pfad (z.B. mit hoher Bandbreite) handelt.

Außerdem haben sie die unangenehme Eigenschaft, neue Informationen schnell zu lernen, aber Fehler nur sehr langsam zu erkennen. Dies liegt daran, dass sie nicht erkennen, dass sie Teil des Problems sind. Die folgenden zwei Grafiken in Abbildung 2 sollen dies verdeutlichen.

Wie man sieht lernt der Router R1, dass sein ausgefallenes Netz über den Nachbar Router R2 scheinbar erreichbar ist. Damit nun der Hop Count nicht ins Unendliche läuft, hat man die Anzahl auf 15 Hops begrenzt. Erst beim Erreichen des max. Hop Counts verwerfen die Router das gelernte Netzwerk. Dies, in Kombination mit den langsamen Update-Intervallen, führt zu einer sehr stark verzögerten Konvergenz des gesamten Netzes.

Als Gegenentwurf zu den Distance Vector Protokollen wurden daher die Link State Protokolle entwickelt. Ihre bekanntes-